

Leveraging Existing Data Sources to Obtain Performance Measures for a Multi-modal Transportation System

Final Report

Prepared for:



Prepared by:



June 2025



ACKNOWLEDGMENTS

This research project is sponsored by the Pima Association of Governments (PAG), funded by federal transportation source Carbon Reduction Program (CRP) and led by the University of Arizona team. The University of Arizona team would like to acknowledge PAG staff and the City of Tucson and Pima County for their invaluable support and collaboration and for providing traffic data. This team would also like to express gratitude to Dr. Xiaofeng Li for providing technical advice.

DISCLAIMER

The content of this report reflects the views of the authors, who are responsible for the facts and the accuracy of the information presented herein and do not necessarily reflect the official view or policies of the sponsoring organization. These contents do not constitute a standard, specification, or regulation. The research work of this study was completed before June 2024. Any new products/modules and the information released after the completion date of this study are not included in this study. This document is disseminated under the sponsorship of the Pima Association of Governments in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

AUTHORS

The University of Arizona and Pima Association of Governments:

Main Authors:

Yeji Jeon	University of Arizona	Graduate Research Assistant
Danial Chekani	University of Arizona	Graduate Research Assistant
Jun Zhao, Ph.D.	University of Arizona	Postdoctoral Researcher

Supervision, Review, Revision:

Ali Shamshiripour, Ph.D.	University of Arizona	Assistant Professor (Co-Principal Investigator)
Yao-Jan Wu, Ph.D., P.E.	University of Arizona	Professor (Principal Investigator)
Xiaobo Ma, Ph.D.	Pima Association of Governments	Senior Transportation Data Scientist
James Tokishi	Pima Association of Governments	Transportation Data Science Manager
Hyunsoo Noh, Ph.D.	Pima Association of Governments	Data Science Administrator

EXECUTIVE SUMMARY

The Pima Association of Government (PAG) is committed to delivering a sustainable, equitable, and efficient transportation system that improves mobility for all users while also advancing regional and national goals for greenhouse gas (GHG) emission reductions. This project, funded through the federal Carbon Reduction Program (CRP), develops and applies a data-driven framework for assessing transit and active transportation performance across the Tucson metropolitan area. It builds on prior PAG efforts that primarily examined motorized vehicle traffic, extending the scope to include transit, micromobility, and walking modes that are critical to achieving both accessibility and climate objectives.

By leveraging state-of-the-art data science techniques, this project establishes new capabilities to archive, prepare, and process various data sources available in the PAG region to generate a plethora of practice-ready, multi-modal performance measures. These measures are tailored to achieve two objectives synergistically: (i) to create a medium for continuous and low-cost monitoring of multimodal performance, and (ii) to identify unprecedented insights on improvement priorities. Following a comprehensive exploration of all data sources accessible at the time, the following were selected and used as the main sources in this project: the General Transit Feed Specification (GTFS) feeds for the transit system; the General Bike Feed Specification (GBFS) feeds for shared micromobility including Tugo bike-share and e-scooters; regional traffic sensor data collected from the MaxView and Miovision sensors, alongside PAG's turning movement counts for the walking trips. Together, these sources provide comprehensive coverage of the multimodality landscape in the region.

The analytical framework developed in this project combines the high- to low-level performance measures to evaluate the regional multimodal services through several lenses.

For the transit users, it evaluates various serviceability aspects, including speed and travel time, reliability and regularity, and headways and on-time performance.

For the pedestrians, this analytical framework evaluates pedestrian volumes, intersection delays, and pedestrian level of service (PLOS).

Lastly, for micromobility service users, the framework evaluates utilization, availability, and geographic service coverage.

To ensure reliability and replicability of the outputs, the data collection pipeline and all performance measures were carefully designed and validated against ground truth data and performance targets are derived from a comprehensive review of literature. This project demonstrates a scalable multimodal monitoring framework that positions PAG to achieve its objectives improving mobility and reducing the carbon footprint of the transportation system as highlighting the performances of transit and active transportation.

Key findings of the study are listed below:

- Transit:
 - SunTran demonstrated reliable transit service, with travel time reliability consistently at or above 0.9 based on 0-to-1 scale and operating speeds exceeding 15 mph for most of the day. Average headways were approximately 26 minutes throughout the day, including during the AM and PM peak periods.
 - SunLink showed slightly lower travel time reliability, ranging from 0.8 to 0.9. This is likely due to its rail service operation in mixed traffic conditions. Headways were generally well-maintained at around 15 minutes, with deviations between 2 and 6 minutes.
 - SunExpress provided strong performance, particularly in outlying areas where fewer stops and higher speed limits allowed for faster service. Travel time reliability was above 0.9 in the morning and around 0.85 in the afternoon. This pattern may reflect the nature of express service, especially in the afternoon outbound direction when most passengers are returning home.
- Pedestrian:
 - Based on the developed performance measures, regional pedestrian services experienced average delays of 30 to 40 seconds at signalized intersections, corresponding level of service (LOS) C, and 8 to 10 seconds at HAWK crossings (LOS A) during daytime hours, including AM and PM peak periods. HAWK signals facilitated a quicker pedestrian crossing by providing dedicated pedestrian-focused service within the MaxView (City of Tucson) service area.
 - Within Miovision service area (Pima County), observed average delays of 40 to 45 seconds at signalized intersections (LOS D) and approximately 12 seconds at HAWK crossings (LOS B) during weekday daytime hours.
- Micromobility (E-Scooters and Bike-Share):
 - Shared e-scooter services covered approximately 2.29% to 2.88% of the city area. Usage typically ranged from 1% to 4 % on weekdays and 2% to 5% on weekends during daytime hours including AM and PM peak periods. Notably nighttime activity was higher, ranging from 2% to 7%, which may be attributed to the relocation of e-scooters to designated areas particularly after midnight.
 - Shared bike service (Tugo Bike) covered about 1.94% of the City of Tucson boundary. Its usage ranged from 0.75% to 2.5% on weekdays and 1% to 3% on weekends during daytime hours, including AM and PM peak periods.

The following conclusions are drawn from the project:

- The resulting framework provides PAG with a replicable, low-cost toolbox to continuously monitor performance across the modes, transit, pedestrian, and micromobility.

- Multimodal performance measures can be reliably extracted from existing GTFS, GBFS, and pedestrian datasets paired with rigorous data quality assessment and improvement techniques.
- Investments in further improving the transit service, walkability, and micromobility coverage offer the combined potential to improve multimodality significantly.

Based on the findings and conclusions above, we recommend the following:

- Focus on monitoring the regional multimodal transportation services with the developed performance measures.
- Establish comprehensive frameworks that link multimodal performance measures to air quality outcomes, capturing the relationship between transportation and air quality.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	I
DISCLAIMER	II
AUTHORS	III
EXECUTIVE SUMMARY	IV
TABLE OF CONTENTS	VII
LIST OF TABLES	XI
LIST OF FIGURES	XIV
CHAPTER 1. INTRODUCTION	1
1.1. BACKGROUND	1
1.2. ORGANIZATION OF THE REPORT	1
CHAPTER 2. LITERATURE REVIEW	4
2.1. LITERATURE REVIEW	4
2.1.1 MODE-SPECIFIC PERFORMANCE INDICATORS.....	4
2.1.2 EXAMPLES OF MULTIMODAL PERFORMANCE INDICATOR APPLICATIONS.....	17
CHAPTER 3. TRANSIT DATA COLLECTION AND ANALYSIS	23
3.1. DELIVERABLES	23
3.2. GENERAL TRANSIT FEED SPECIFICATION (GTFS)	24
3.2.1 GTFS GENERAL OVERVIEW.....	24
3.2.2 GTFS TRIP UPDATE.....	26
3.2.3 GTFS STATIC DATA.....	28
3.2.4 GTFS DATA ANALYSIS.....	33
CHAPTER 4. PEDESTRIAN DATA COLLECTION AND ANALYSIS	43
4.1. DELIVERABLES	43
4.1.2 MAXVIEW DATA.....	45
4.1.3 MIOVISION DATA.....	68
4.1.4 PAG TURNING MOVEMENT COUNT (TMC) PEDESTRIAN DATA.....	74
CHAPTER 5. MICROMOBILITY (BIKE SHARING AND E-SCOOTER) DATA COLLECTION AND ANALYSIS	77
5.1. DELIVERABLES	77
5.2. E-SCOOTER DATA	79
5.2.1 E-SCOOTER GBFS (GENERAL BIKESHARE FEED SPECIFICATION).....	79
5.2.2 DATA ANALYSIS.....	84

5.3. BIKE SHARING DATA	88
5.3.1 TUGO BIKE SHARE GBFS	88
5.3.2 DATA ANALYSIS	93
5.3.3 STREETLIGHT DATA	96
CHAPTER 6. TRANSIT PERFORMANCE MEASURE EXTRACTION.....	102
6.1. DELIVERABLES.....	102
6.2. SPEED	106
6.2.1 SUN TRAN SYSTEM LEVEL SPEED PERFORMANCE MEASURE	106
6.2.2 SUN TRAN ROUTE LEVEL SPEED PERFORMANCE MEASURE	108
6.2.3 SUN EXPRESS SYSTEM LEVEL SPEED PERFORMANCE MEASURE	109
6.2.4 SUN EXPRESS ROUTE LEVEL SPEED PERFORMANCE MEASURE	110
6.2.5 SUN LINK SYSTEM LEVEL SPEED PERFORMANCE MEASURE.....	112
6.2.6 SUN LINK STOP LEVEL SPEED PERFORMANCE MEASURE	113
6.3. IN-VEHICLE TRAVEL TIME RELIABILITY	114
6.3.1 SUN TRAN SYSTEM-LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY PERFORMANCE MEASURE	115
6.3.2 SUN TRAN ROUTE LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY PERFORMANCE MEASURE	116
6.3.3 SUN EXPRESS SYSTEM LEVEL TRAVEL TIME RELIABILITY PERFORMANCE MEASURE	117
6.3.4 SUN EXPRESS ROUTE LEVEL TRAVEL TIME RELIABILITY PERFORMANCE MEASURE.....	118
6.3.5 SUN LINK SYSTEM LEVEL TRAVEL TIME RELIABILITY PERFORMANCE MEASURE.....	119
6.3.6 SUN LINK STOP LEVEL TRAVEL TIME RELIABILITY PERFORMANCE MEASURE	120
6.4. HEADWAYS	122
6.4.2 SUN TRAN SYSTEM LEVEL HEADWAYS PERFORMANCE MEASURE	122
6.4.3 SUN TRAN ROUTE LEVEL HEADWAYS PERFORMANCE MEASURE	124
6.4.4 SUN LINK SYSTEM LEVEL TRAVEL TIME RELIABILITY PERFORMANCE MEASURE.....	125
6.4.5 SUN LINK STOP LEVEL HEADWAY PERFORMANCE MEASURE	125
6.5. HEADWAY IRREGULARITY	128
6.5.1 SUN TRAN SYSTEM LEVEL HEADWAY STANDARD DEVIATION PERFORMANCE MEASURE.....	128
6.5.2 SUN TRAN ROUTE LEVEL HEADWAYS IRREGULARITY PERFORMANCE MEASURE	130
6.5.3 SUN LINK STOP LEVEL IRREGULARITY PERFORMANCE MEASURE.....	131
6.6. HEADWAY RELIABILITY	132
6.6.1 SUN TRAN SYSTEM LEVEL HEADWAY RELIABILITY PERFORMANCE MEASURE	133
6.6.2 SUN TRAN ROUTE LEVEL HEADWAY RELIABILITY PERFORMANCE MEASURE	134
6.6.3 SUN LINK STOP LEVEL HEADWAY RELIABILITY PERFORMANCE MEASURE	135
6.6.4 ON-TIME PERFORMANCE	137
CHAPTER 7. PEDESTRIAN PERFORMANCE MEASURE EXTRACTION.....	139
7.1. DELIVERABLES.....	139
7.2. PEDESTRIAN VOLUME.....	142
7.2.1 INTRODUCTION	142
7.2.2 METHODOLOGY.....	142

7.2.3	RESULTS.....	150
7.3.	PEDESTRIAN DELAY	152
7.3.1	INTRODUCTION	152
7.3.2	METHODOLOGY.....	152
7.3.3	RESULTS.....	154
7.4.	PEDESTRIAN LEVEL OF SERVICE	158
7.4.1	INTRODUCTION	158
7.4.2	METHODOLOGY.....	158
7.4.3	RESULTS.....	159
CHAPTER 8.	MICROMOBILITY PERFORMANCE MEASURE EXTRACTION.....	161
8.1.	DELIVERABLES.....	161
8.2.	E-SCOOTER PERFORMANCE MEASURES	162
8.2.1	E-SCOOTER SYSTEM LEVEL VEHICLE UTILIZATION RATE	162
8.2.2	E-SCOOTER GEOGRAPHIC TIME AVAILABILITY RATE	164
8.2.3	E-SCOOTER GEOGRAPHIC VEHICLE COVERAGE RATE	166
8.3.	TUGO BIKE-SHARE SYSTEM PERFORMANCE MEASURES	171
8.3.1	TUGO BIKE-SHARE SYSTEM LEVEL VEHICLE UTILIZATION RATE	171
8.3.2	TUGO BIKE-SHARE SYSTEM GEOGRAPHIC VEHICLE UTILIZATION RATE	173
8.3.3	TUGO BIKE-SHARE SYSTEM GEOGRAPHIC TIME AVAILABILITY RATE.....	176
8.3.4	TUGO BIKE-SHARE SYSTEM GEOGRAPHIC VEHICLE COVERAGE RATE.....	179
CHAPTER 9.	DATA QA/QC AND PERFORMANCE MEASURE TARGET DEVELOPMENT	181
9.1.	DELIVERABLES.....	181
9.2.	TRANSIT QA/QC AND VALIDATION	183
9.2.1	TRIP UPDATE AND VEHICLE POSITIONS COMPARISON	183
9.2.2	TRANSIT PERFORMANCE MEASURES RESULT VALIDATION	187
9.2.3	TRANSIT RECOMMENDED PERFORMANCE TARGETS	189
9.3.	PEDESTRIAN PERFORMANCE TARGETS.....	192
9.3.1	PEDESTRIAN VOLUME ESTIMATION LITERATURE REVIEW	192
9.3.2	PEDESTRIAN VOLUME ESTIMATION MODEL RESULTS.....	192
9.3.3	COMPARISON TO OTHER MODELS	194
9.3.4	PEDESTRIAN RECOMMENDED ACCEPTANCE CRITERIA.....	199
9.4.	MICROMOBILITY DATA QA/QC	200
9.4.1	TUGO BIKE SHARE GBFS DATA ACCURACY REVIEW	200
9.4.2	E-SCOOTER GBFS DATA ACCURACY REVIEW	202
9.4.3	MICROMOBILITY RECOMMENDED PERFORMANCE MEASURE TARGET	208
CHAPTER 10.	PLATFORM DEVELOPMENT FOR REGION-WIDE MULTI-MODAL PERFORMANCE 209	
10.1.	TRANSIT FREQUENCY AND PEDESTRIAN ACTIVITY	211

10.2. FEATURE IMPORTANCE ANALYSIS USING SHAP	212
10.2.1 INTRODUCTION TO SHAP	212
10.2.2 RESULTS OF SHAP	215
10.2.3 PLATFORM DEVELOPMENT	223

LIST OF TABLES

1.1. BACKGROUND	1
1.2. ORGANIZATION OF THE REPORT	1
TABLE 1.1. TIMELINE.....	3
2.1. LITERATURE REVIEW.....	4
TABLE 2.1. SUMMARY OF STUDIES ON TRANSIT PERFORMANCE MEASURES IDENTIFICATION AND COLLECTION	5
TABLE 2.2. SUMMARY OF STUDIES ON PEDESTRIAN PERFORMANCE MEASURES IDENTIFICATION AND COLLECTION.....	9
TABLE 2.3. SUMMARY OF STUDIES ON E-SCOOTER PERFORMANCE MEASURES IDENTIFICATION AND COLLECTION	12
TABLE 2.4. SUMMARY OF STUDIES ON BIKE PERFORMANCE MEASURES IDENTIFICATION AND COLLECTION.....	14
TABLE 2.5. SUMMARY OF STUDIES ON AIR QUALITY DUE TO TRANSPORTATION OPERATION MEASURES IDENTIFICATION AND COLLECTION	16
TABLE 2.6. SUMMARY OF MULTI-MODAL TRANSPORTATION PERFORMANCE MEASURES APPLICATION EXAMPLES	18
TABLE 2.7. SUMMARY OF TRANSPORTATION OPERATION IMPACT ON AIR QUALITY PERFORMANCE MEASURES APPLICATION EXAMPLES	22
3.1. DELIVERABLES.....	23
TABLE 3.1. SUMMARY OF TRANSIT DATA SOURCES AND USE CASES CONSIDERED.....	24
3.2. GENERAL TRANSIT FEED SPECIFICATION (GTFS)	24
TABLE 3.2. SUN TRAN GTFS FEEDS SUMMARY	25
TABLE 3.3. STOPS PARAMETERS DEFINITIONS.....	28
TABLE 3.4. ROUTES PARAMETERS DEFINITIONS.....	29
TABLE 3.5. TRIPS PARAMETERS DEFINITIONS	30
TABLE 3.6. SHAPES PARAMETERS DEFINITIONS	30
TABLE 3.7. STOP TIMES FILE PARAMETERS DEFINITIONS.....	32
4.1. DELIVERABLES.....	43
TABLE 4.1. PEDESTRIAN DATA SOURCES IN PIMA COUNTY	44
TABLE 4.2. AVAILABLE PEDESTRIAN RELATED MAXVIEW DATA CODES.....	47
TABLE 4.3. AVAILABLE MAXVIEW DEVICE COUNT EVERY MONTH.....	48
TABLE 4.4. MIOVISION SENSOR (ARCHIVED IN CATS SERVER) LOCATION DISTRIBUTION.....	69
TABLE 4.5. OUTLIERS IN BOTTOM RIGHT OF FIGURE 4.25	76
5.1. DELIVERABLES.....	77
TABLE 5.1. MICROMOBILITY DATA SOURCES IN PIMA COUNTY	78
5.2. E-SCOOTER DATA	79
TABLE 5.2. AVAILABLE E-SCOOTER GBFS DATA SUMMARY	81
5.3. BIKE SHARING DATA	88
TABLE 5.3. AVAILABLE BIKE SHARE GBFS DATA SUMMARY	90
TABLE 5.4. TUGO BIKE SHARE PRICING PLAN SUMMARY	92

TABLE 5.5.	STREETLIGHT DATA ANALYSIS AVAILABILITY FOR EACH TRANSPORTATION MODE	97
6.1. DELIVERABLES		102
TABLE 6.1.	TRANSIT PERFORMANCE MEASURES SUMMARY	103
6.2. SPEED		106
6.3. IN-VEHICLE TRAVEL TIME RELIABILITY		114
6.4. HEADWAYS		122
6.5. HEADWAY IRREGULARITY		128
6.6. HEADWAY RELIABILITY		132
7.1. DELIVERABLES		139
TABLE 7.1.	PEDESTRIAN PERFORMANCE MEASURES SUMMARY	140
7.2. PEDESTRIAN VOLUME		142
TABLE 7.2.	SAMPLE LOCATIONS OF INTERSECTIONS EQUIPPED WITH PEDESTRIAN SIGNAL DEVICES	143
TABLE 7.3.	RAW DATA SAMPLE	144
TABLE 7.4.	DESCRIPTION OF MODEL VARIABLES	144
TABLE 7.5.	DESCRIPTION, STRENGTHS AND LIMITATIONS OF METRICS	151
7.3. PEDESTRIAN DELAY		152
TABLE 7.6.	MAXVIEW EVENT CODE TABLE	153
TABLE 7.7.	MIOVISION EVENT CODE TABLE	153
7.4. PEDESTRIAN LEVEL OF SERVICE		158
TABLE 7.8.	RECOMMENDED HCM PLOS CRITERIA FOR SIGNALIZED CROSSING DELAY (ROUPHAIL, NAGUI M., ET AL.) 159	
TABLE 7.9.	RECOMMENDED HCM PLOS CRITERIA FOR UNSIGNALIZED CROSSING DELAY (ROUPHAIL, NAGUI M., ET AL.) 159	
8.1. DELIVERABLES		161
TABLE 8.1.	MICROMOBILITY PERFORMANCE MEASURES SUMMARY	162
8.2. E-SCOOTER PERFORMANCE MEASURES		162
8.3. TUGO BIKE-SHARE SYSTEM PERFORMANCE MEASURES		171
9.1. DELIVERABLES		181
TABLE 9.1.	DATA QA/QC AND PERFORMANCE MEASURE VALIDATION SUMMARY	182
9.2. TRANSIT QA/QC AND VALIDATION		183
TABLE 9.2.	TRANSIT PERFORMANCE MEASURES LITERATURE REVIEW SUMMARY	188
TABLE 9.3.	SUN TRAN TRANSIT PERFORMANCE MEASURES SUMMARY	189
TABLE 9.4.	TRANSIT PERFORMANCE TARGETS	190
9.3. PEDESTRIAN PERFORMANCE TARGETS		192
TABLE 9.5.	PEDESTRIAN VOLUME ESTIMATION LITERATURE REVIEW SUMMARY	192
TABLE 9.6.	DESCRIPTION, STRENGTHS AND LIMITATIONS OF EACH METRIC	193

TABLE 9.7.	PERFORMANCE METRICS FOR ALTERNATIVE MODELS	195
TABLE 9.8.	PEDESTRIAN PERFORMANCE MEASURE TARGET	199
9.4.	MICROMOBILITY DATA QA/QC	200
TABLE 9.9.	TUGO BIKE SHARE OBSERVATIONS RECORD	201
TABLE 9.10.	MICROMOBILITY PERFORMANCE MEASURE TARGET.....	208
TABLE 10.1.	MULTI-MODAL PERFORMANCE MEASURES SUMMARY	210
10.1.	TRANSIT FREQUENCY AND PEDESTRIAN ACTIVITY	211
10.2.	FEATURE IMPORTANCE ANALYSIS USING SHAP.....	212
TABLE 10.2.	INTERPRETATION OF DIFFERENT COMBINATIONS OF FEATURE AND SHAP VALUES	215
TABLE A.1	MIOVISION CROSSWALK PEDESTRIAN DATA	229
TABLE A.2	THE TWO INTERSECTIONS INSIDE CITY OF TUCSON.....	230

LIST OF FIGURES

1.1. BACKGROUND	1
1.2. ORGANIZATION OF THE REPORT	1
2.1. LITERATURE REVIEW	4
3.1. DELIVERABLES	23
3.2. GENERAL TRANSIT FEED SPECIFICATION (GTFS)	24
FIGURE 3.1. SUN TRAN GTFS DATA STRUCTURE	26
FIGURE 3.2. GTFS ROUTE SYSTEM.....	27
FIGURE 3.3. GTFS TRIP UPDATE DATA STRUCTURE.....	28
FIGURE 3.4. GTFS SHAPES STATIC DATA	31
FIGURE 3.5. TRANSIT VEHICLE LOCATIONS ANALYSIS	34
FIGURE 3.6. WEEKDAY AND WEEKEND AVERAGE SPEED PATTERN	36
FIGURE 3.7. SCHEDULED AND ACTUAL HEADWAYS COMPARISONS BY ROUTE.....	40
FIGURE 3.8. “ACTUAL HEADWAY – SCHEDULED HEADWAY” BOXPLOT	41
FIGURE 3.9. WEEKDAYS AND WEEKENDS HEADWAY PATTERN.....	42
4.1. DELIVERABLES	43
FIGURE 4.1. METHODOLOGY FLOWCHART	45
FIGURE 4.2. MAXVIEW EVENT-BASED DATA LOCATIONS	46
FIGURE 4.3. AVAILABLE MAXVIEW DEVICE COUNT EVERY MONTH.....	49
FIGURE 4.4. HOURLY AVERAGE PUSH BUTTON COUNTS BY DAY TYPE AND DEVICE TYPE FROM JANUARY 1 TO JANUARY 31, 2023 (HAWK DEVICES 362 AND 307 EXCLUDED)	50
FIGURE 4.5. HOURLY AVERAGE PUSH COUNTS IN JANUARY (WEEKDAYS).....	51
FIGURE 4.6. HOURLY AVERAGE PUSH COUNTS IN JANUARY (WEEKENDS)	53
FIGURE 4.7. HOURLY PEDESTRIAN COUNTS FOR DEVICE 214 IN JANUARY 2023	54
FIGURE 4.8. HOURLY COUNTS FOR DEVICE 362 IN JULY 2023	55
FIGURE 4.9. HOURLY AVERAGE PUSH BUTTON COUNTS BY DAY TYPE AND DEVICE TYPE FROM JULY 1 TO JULY 31, 2023 (EXCLUDING HAWK DEVICE 362).....	56
FIGURE 4.10. HOURLY AVERAGE PUSH COUNTS IN JULY (WEEKDAYS)	57
FIGURE 4.11. HOURLY AVERAGE PUSH COUNTS IN JULY (WEEKENDS).....	59
FIGURE 4.12. HOURLY AVERAGE WAITING TIME BY DAY TYPE AND DEVICE TYPE OF JANUARY 2023	61
FIGURE 4.13. HOURLY AVERAGE WAITING TIME IN JANUARY (WEEKDAYS).....	62
FIGURE 4.14. HOURLY AVERAGE WAITING TIME IN JANUARY (WEEKENDS)	64
FIGURE 4.15. HOURLY AVERAGE WAITING TIME BY DAY AND DEVICE TYPE IN, JULY 1 – JULY 31, 2023	65
FIGURE 4.16. HOURLY AVERAGE WAITING TIME IN JULY (WEEKDAYS)	66
FIGURE 4.17. HOURLY AVERAGE WAITING TIME IN JULY (WEEKENDS).....	67
FIGURE 4.18. MIOVISION SENSOR (ARCHIVED IN CATS SERVER) LOCATION DISTRIBUTION	68
FIGURE 4.19. NUMBER OF MONTHLY ACTIVE MIOVISION SENSORS.....	69
FIGURE 4.20. MONTHLY AVERAGE PUSH-BUTTON COUNTS ACROSS ALL DETECTORS.....	70
FIGURE 4.21. HOURLY AVERAGE PEDESTRIAN COUNT BY DAY IN JANUARY AND JULY	71
FIGURE 4.22. MAXVIEW AND MIOVISION LOCATIONS OF JANUARY 2023	73

FIGURE 4.23. DISTINCT INTERSECTION NUMBER FROM THE SURVEY DATA 74

FIGURE 4.24. RELATIONSHIP BETWEEN HOURLY PUSH BUTTON COUNTS (MAXVIEW) AND HOURLY PEDESTRIAN COUNTS (PAG_TMC) OF SEPTEMBER 2023 75

FIGURE 4.25. RELATIONSHIP BETWEEN HOURLY PUSH BUTTON COUNTS (MAXVIEW) AND HOURLY PEDESTRIAN COUNTS (PAG_TMC) OF OCTOBER 2023 75

FIGURE 4.26. HOURLY PEDESTRIAN COUNTS FOR DEVICE 582 IN OCTOBER 2023 76

5.1. DELIVERABLES 77

FIGURE 5.1. MICROMOBILITY METHODOLOGY FLOWCHART 79

5.2. E-SCOOTER DATA 79

FIGURE 5.2. E-SCOOTER GBFS GEOFENCING MAP 80

FIGURE 5.3. E-SCOOTER AVAILABILITY MAP ON JULY 31ST (WEEKDAY) AND AUGUST 3RD (WEEKEND), 2024 (CONTINUED) 85

FIGURE 5.4. E-SCOOTER AVAILABILITY PLOT FROM SEPTEMBER 16TH TO SEPTEMBER 22ND, 2024 86

FIGURE 5.5. E-SCOOTER USAGE PLOT 87

FIGURE 5.6. DISABLED E-SCOOTER PLOT FROM SEPTEMBER 16TH TO SEPTEMBER 22ND, 2024 88

5.3. BIKE SHARING DATA 88

FIGURE 5.7. BIKESHARE GBFS STATION LOCATIONS MAP 89

FIGURE 5.8. TUGO BIKESHARE AVAILABILITY MAP ON JULY 31ST (WEEKDAY) AND AUGUST 3RD (WEEKEND), 2024 94

FIGURE 5.9. TUGO BIKE SHARE TOTAL VEHICLE AVAILABILITY FROM SEPTEMBER 30TH TO OCTOBER 6TH, 2024 95

FIGURE 5.10. TUGO BIKE SHARE TOTAL VEHICLE USAGE FROM SEPTEMBER 30TH TO OCTOBER 6TH, 2024 96

FIGURE 5.11. TUCSON BIKE ACTIVITY ANALYSIS IN JANUARY 2022 99

FIGURE 5.12. E-SCOOTER AND BIKE SHARE AVAILABILITY AND DEMAND MAP 101

6.1. DELIVERABLES 102

FIGURE 6.1. GTFS-ST PREPROCESSING FLOW CHART 104

FIGURE 6.2. GTFS SEGMENT LIST DATA EXAMPLE 104

FIGURE 6.3. FIGURE 3. SPEED PERFORMANCE MEASURE FLOW CHART 105

6.2. SPEED 106

FIGURE 6.4. SUN TRAN SYSTEM LEVEL TRANSIT VEHICLE SPEED BOXPLOT 107

FIGURE 6.5. SUN TRAN ROUTE LEVEL SPEED PERFORMANCE HEATMAP 109

FIGURE 6.6. SUN EXPRESS SYSTEM LEVEL TRANSIT VEHICLE SPEED BOXPLOT 110

FIGURE 6.7. SEGMENT-WEIGHTED SUN EXPRESS ROUTE LEVEL TRANSIT VEHICLE SPEED HEATMAP 111

FIGURE 6.8. SUN LINK SYSTEM LEVEL TRANSIT VEHICLE SPEED BOXPLOT 112

FIGURE 6.9. SUN LINK STOP LEVEL TRANSIT VEHICLE SPEED BOXPLOT 113

6.3. IN-VEHICLE TRAVEL TIME RELIABILITY 114

FIGURE 6.10. TRAVEL TIME RELIABILITY PERFORMANCE MEASURE FLOW CHART 115

FIGURE 6.11. SUN TRAN SYSTEM LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY BOXPLOT 116

FIGURE 6.12. NORMALIZED SUN TRAN ROUTE LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY HEATMAP 117

FIGURE 6.13. SUN EXPRESS SYSTEM LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY BOXPLOT 118

FIGURE 6.14. SUN EXPRESS ROUTE LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY HEATMAP 119

FIGURE 6.15. SUN LINK SYSTEM LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY BOXPLOT 120

FIGURE 6.16. SUN LINK STOP LEVEL IN-VEHICLE TRAVEL TIME RELIABILITY HEATMAP 121

FIGURE 8.3.	WEEKEND TIME AVAILABILITY RATE.....	166
FIGURE 8.4.	WEEKDAY VEHICLE COVERAGE RATE.....	168
FIGURE 8.5.	WEEKEND VEHICLE COVERAGE RATE.....	170
8.3.	TUGO BIKE-SHARE SYSTEM PERFORMANCE MEASURES.....	171
FIGURE 8.6.	TUGO BIKE-SHARE SYSTEM LEVEL VEHICLE UTILIZATION.....	172
FIGURE 8.7.	WEEKDAY UTILIZATION RATE.....	174
FIGURE 8.8.	WEEKEND TIME AVAILABILITY RATE.....	175
FIGURE 8.9.	WEEKDAY TIME AVAILABILITY RATE.....	177
FIGURE 8.10.	WEEKEND TIME AVAILABILITY RATE.....	178
FIGURE 8.11.	VEHICLE COVERAGE RATE.....	180
9.1.	DELIVERABLES.....	181
9.2.	TRANSIT QA/QC AND VALIDATION.....	183
FIGURE 9.1.	TRIP UPDATE DATA AND VEHICLE POSITIONS DATA COMPARISON METRIC EXAMPLE.....	183
FIGURE 9.2.	VEHICLE POSITIONS RECORD SEGMENT ASSIGNMENT.....	184
FIGURE 9.3.	TIME DIFFERENCE OF EACH ROUTE ON WEDNESDAY, MARCH 26, 2025.....	185
FIGURE 9.4.	TIME DIFFERENCE OF EACH ROUTE ON SATURDAY, MARCH 29, 2025.....	186
FIGURE 9.5.	SUN TRAN TRANSIT PERFORMANCE MEASURES AND RIDERSHIP SCATTER PLOTS.....	189
9.3.	PEDESTRIAN PERFORMANCE TARGETS.....	192
FIGURE 9.6.	PARITY PLOT OF THE FINAL XGBOOST MODEL.....	194
FIGURE 9.7.	PARITY PLOT COMPARISON OF ALTERNATIVE MODELS.....	196
FIGURE 9.8.	RMSE COMPARISON BETWEEN LINEAR AND XGBOOST MODELS IN DIFFERENT RANGES.....	196
FIGURE 9.9.	MAPE COMPARISON BETWEEN LINEAR AND XGBOOST MODELS IN DIFFERENT RANGES.....	197
FIGURE 9.10.	MAE COMPARISON BETWEEN LINEAR AND XGBOOST MODELS IN DIFFERENT RANGES.....	197
FIGURE 9.11.	PERCENT RMSE VALUES FOR THE XGBOOST MODEL IN DIFFERENT RANGES.....	198
9.4.	MICROMOBILITY DATA QA/QC.....	200
FIGURE 9.12.	STATION 27 BIKE AVAILABILITY STATUS.....	201
FIGURE 9.13.	STATION 30 BIKE AVAILABILITY STATUS.....	202
FIGURE 9.14.	E-SCOOTER OBSERVATION AT THE INTERSECTION OF E UNIVERSITY BLVD AND N 3 RD AVE.....	203
FIGURE 9.15.	E-SCOOTER OBSERVATION AT E 8 TH ST AND N 4 TH AVE.....	204
FIGURE 9.16.	E-SCOOTER OBSERVATION AT NORTHEAST OF E 8 TH ST AND N 4 TH AVE.....	205
FIGURE 9.17.	E-SCOOTER OBSERVATION AT E 7 TH ST AND N 4 TH AVE.....	206
FIGURE 9.18.	E-SCOOTER RETURN ACTIVITY OBSERVATION.....	207
10.1.	TRANSIT FREQUENCY AND PEDESTRIAN ACTIVITY.....	211
FIGURE 10.1.	TRANSIT FREQUENCY AND PEDESTRIAN.....	212
10.2.	FEATURE IMPORTANCE ANALYSIS USING SHAP.....	212
FIGURE 10.2.	BREAKDOWN OF FEATURE CONTRIBUTIONS TO A SAMPLE PREDICTION USING SHAP VALUES.....	214
FIGURE 10.3.	GLOBAL FEATURE IMPORTANCE OF FEATURES.....	216
FIGURE 10.4.	SHAP BEESWARM PLOT OF THE TOP FEATURES OF THE PEDESTRIAN MODEL.....	217
FIGURE 10.5.	SHAP BEESWARM PLOTS FOR TOTALSIGNALS AND LEISURE.....	218
FIGURE 10.6.	SHAP BEESWARM PLOT FOR EDUCATION.....	218

FIGURE 10.7.	SCATTER PLOT OF SHAP VALUES FOR DIFFERENT NUMBER OF PUSH BUTTON EVENTS	219
FIGURE 10.8.	SCATTER PLOT OF SHAP VALUES FOR DIFFERENT RATIOS OF ZERO-CAR HOUSEHOLDS.....	220
FIGURE 10.9.	SCATTER PLOT OF SHAP VALUES FOR DIFFERENT VALUES OF NATIONAL WALKING INDEX	221
FIGURE 10.10.	SCATTER PLOT OF SHAP VALUES FOR DIFFERENT VALUES OF D4A	222
FIGURE 10.11.	PLATFORM PAGE.....	223
FIGURE 10.12.	READ-ME FILE SAMPLE	224
FIGURE 10.13.	TASK 3 CODE PLATFORM FOLDER	224
FIGURE 10.14.	TASK 4 CODE PLATFORM FOLDER	225
FIGURE 10.15.	TASK 5 CODE PLATFORM FOLDER	226
FIGURE 10.16.	TASK 6 CODE PLATFORM FOLDER	226
FIGURE 10.17.	TASK 7 CODE PLATFORM FOLDER	227
FIGURE 10.18.	TASK 8 CODE PLATFORM FOLDER	227
FIGURE 10.19.	TASK 9 PLATFORM FOLDER	228
FIGURE 10.20.	TASK 10 PLATFORM FOLDER	228

CHAPTER 1. INTRODUCTION

1.1. Background

The growing need for sustainable, equitable, and efficient transportation systems has prompted regional planning agencies to shift from auto-centric strategies toward multi-modal transportation planning. In this context, the Pima Association of Governments (PAG) has initiated the development of the Regional Mobility and Accessibility Plan (RMAP), a long-range, performance-based transportation plan aimed at enhancing safety, mobility, accessibility, and environmental outcomes across the Tucson metropolitan area. RMAP emphasizes an integrated approach that includes all transportation modes, motorized vehicles, public transit, bicycles, and pedestrians, aligning with broader goals of reducing greenhouse gas emissions and promoting active transportation.

To support these objectives, accurate and mode-specific performance measures are essential. Building on previous efforts focused primarily on motorized vehicle performance using event-based data, this project extends the scope by targeting active and alternative transportation modes. By utilizing a variety of existing data sources, including General Transit Feed Specification (GTFS), General Bikeshare Feed Specification (GBFS), and local sensor data, this study establishes a robust framework for assessing the performance of the regional multi-modal transportation system. The outcome includes not only the identification and estimation of key performance indicators but also the development of data collection methodologies, QA/QC procedures, and analytical methods, thereby laying the groundwork for improved mobility, air quality, and long-term planning across the PAG region.

1.2. Organization of the Report

This report is organized for each task results and data summary as shown in 3. Chapter 2 focuses on the literature reviews of different performance measures for different modes of transportation, transit, pedestrian, and micromobility. The multimodal transportation performance measures and emission modeling are reviewed in the section. Chapter 3 explored transit data and discussed data analysis. Chapter 4 explored existing pedestrian data and conducted pedestrian data analysis. Chapter 5 reviewed micromobility data for each e-scooter and TUGO bike sharing system and conducted analysis. Chapter 6 used the identified data and conducted different performance measures including speed, travel time reliability, headway, headway reliability, irregularity, and arrival delay for each transit modes, Sun Tran, Sun Express, and Sun Link. Chapter 7 used pedestrian data to estimate the pedestrian volume on different intersections, calculated pedestrian delay and measured Pedestrian Level of Service (PLOS). Chapter 8 measures performance of e-scooters and TUGO bike sharing system including system level vehicle utilization rate, geographic vehicle utilization rate, geographic time availability rate, and geographic vehicle coverage rate.

Based on the data, findings, and methods discussed in Chapter 3 through Chapter 8, acceptance criteria are discussed in Chapter 9. Chapter 10 discussed multimodal performance measures including different modes of transportation explored and explained the developed Jupyter note platform.

CHAPTER 2. LITERATURE REVIEW

2.1. Literature Review

The objective of the project is to improve regional mobility and air quality. The project focused on obtaining the performance measures of the regional multi-modal transportation system by leveraging existing data sources in the Pima Association of Governments (PAG) region. As a final output, methods of data collection and multi-modal performance measure estimation as well as QA/QC procedures were developed.

This section summarizes the existing performance measures for multi-modal transportation systems including multimodal data resource identification, performance measures identification and collection, and applications of multimodal performance measurements.

2.1.1 Mode-specific Performance Indicators

Multi-modal transportation systems performance measures are important to understand and evaluate the current system's efficiency, effectiveness, and reliability for further operation planning. However, there may be limited data sources available for certain modes. A challenge is obtaining appropriate performance measures for each mode of transportation with existing data sources. Some studies show how to evaluate the performance of various transportation modes, including transit, pedestrian, and micromobility with existing data sources.

2.1.1.1 Transit

For transit performance measures, appropriate usage of accurate and detailed information is critical. Some common sources of data for operation and service performance analysis are Automatic Vehicle Location (AVL), Automated Passenger Counts (APC), and electronic fareboxes (Liao and Liu, 2010; Carrasco, N., 2012; Handley et al., 2019). AVL systems track real-time locations and stores bus arrival and departure times at specified locations; APC systems collect passenger boarding automatically; and electronic fareboxes provide route-level and system-level ridership data (Ryus, P., 2003).

Table 2.1. Summary of studies on transit performance measures identification and collection

Reference	Performance Measures	Metric/Methods	Data Source(s)	Key Findings
Ryus, P. (2003)	<ul style="list-style-type: none"> • Availability measures • Service delivery measures • Community measures • Travel time measures • Safety and security measures • Economic measures • Etc. 	<ul style="list-style-type: none"> • Descriptive statistics 	<ul style="list-style-type: none"> • In-House • National Transit Database (NTD) • Demographic data • Traffic data • GIS data • AVL, APC and Farebox Data 	<ul style="list-style-type: none"> • Given 12 case studies of successful performance-measure programs. • Provided resources for the process of developing performance-measurement programs including Data sources, data collection techniques, and data applications.
Rodier, C. J., & Issac, E. (2016).	<ul style="list-style-type: none"> • Service Coverage • Travel time • Frequency • Stop Accessibility • On-time performance • Passenger load • Mobility service equitability 	<ul style="list-style-type: none"> • Descriptive statistics 	<ul style="list-style-type: none"> • In-house data • NTD • Census data • Travel demand models • Surveys 	<ul style="list-style-type: none"> • Identified key transit performance measures, including service availability, service delivery, community impact and ETC. • Examined the use of performance measures in recent reports and publications by the four major California metropolitan planning organizations.
Liao, C. F., & Liu, H. X. (2010)	Route level performance <ul style="list-style-type: none"> • Running Time • Dwell Time at Stop • Delay at Signal • Transfer Activity 	<ul style="list-style-type: none"> • TP time model • Link travel time model 	<ul style="list-style-type: none"> • Archived transit data from Metro Transit in the Twin Cities of Minnesota including Schedule Data, AVL APC, and Farecard data • Arterial Traffic Data • Environmental Factors 	<ul style="list-style-type: none"> • Developed an integrated data-processing framework to support transit performance analysis and research.
Wong, J. C. (2013).	<ul style="list-style-type: none"> • Average Headway • Hours of service • Passenger Load • On-time performance • Travel Time Difference 	Descriptive statistics	<ul style="list-style-type: none"> • GTFS data • NTD data 	<ul style="list-style-type: none"> • Developed a tool that can calculate performance measures from GTFS data • Compared metrics calculated from raw GTFS feeds to those reported in the NTD.
Carrasco, N. (2012).	Reliability measures <ul style="list-style-type: none"> • Travel time • Speed • Punctuality • Regularity 	Descriptive statistics	<ul style="list-style-type: none"> • AVL data • A set of spreadsheets containing aggregate information on schedule deviation, vehicle speed, and travel time 	<ul style="list-style-type: none"> • Quantifying reliability of transit service in Zurich, Switzerland

	<ul style="list-style-type: none"> • Passage waiting time 			
Liu, J., He, M., Schonfeld, P. M., Kato, H., & Li, A. (2022).	<ul style="list-style-type: none"> • Accessibility (destination-reachability metrics) • Time reliability 	<ul style="list-style-type: none"> • Monte Carlo simulation. • Linear Regression 	<ul style="list-style-type: none"> • Attributes of lines • Average Tavel Time, Time reliability. • Walking distance 	<ul style="list-style-type: none"> • Measured accessibility by incorporating time reliability in Wuhan's rail transit system • The accessibility of Wuhan's URTN during weekday morning peak hours indicates that the metrics that do not consider time reliability overestimate a URTN's accessibility
Al Mamun, M. S., & Lownes, N. E. (2011)	<ul style="list-style-type: none"> • Composite index of public transit accessibility 	<ul style="list-style-type: none"> • The Local Index of Transit Availability • TCQSM • Time-of-day Tool 	<ul style="list-style-type: none"> • Transit data and census data • Travel demand for the specific time period. 	<ul style="list-style-type: none"> • Examine how consistently the three methods rated transit accessibility for each tract of the study area. • Proposed weighting factors for individual methods to formulate a composite measure.
Handley, J. C., Fu, L., & Tupper, L. L. (2019)	<ul style="list-style-type: none"> • Connectivity measure for transit systems 	<ul style="list-style-type: none"> • Spatial statistical model • Simultaneous autoregressive model 	<ul style="list-style-type: none"> • GTFS data • Automated passenger count (APC) data • Tables on schedule adherence 	<ul style="list-style-type: none"> • Stop level connectivity is defined as the total area reachable from the bus stop of interest within a certain travel time budget. • Network connectivity measure in terms of the proportion of city area and population having walking access to the core network. • A case study with operational data from Rochester, New York illustrates connectivity measure at different levels.
Liu, J., Schonfeld, P. M., F.ASCE, Peng, Q., Yin, Y. (2020)	<ul style="list-style-type: none"> • Travel reliability; connectivity reliability, travel time reliability, capacity reliability 	<ul style="list-style-type: none"> • Weighted average number of tolerable travel paths 	<ul style="list-style-type: none"> • Automatic Fare Collection (AFC) data (OD flow) 	<ul style="list-style-type: none"> • Travel reliability of urban rail transit is related to tolerable coefficient • A station with better connectivity reliability tends to have higher capacity reliability
Park, Y., Mount, J., Liu, L., Xiao, N., Miller, H. J. (2019)	<ul style="list-style-type: none"> • Transit route-based network distance • Total propagating delay size 	<ul style="list-style-type: none"> • Compare scheduled and observed vehicle data • Determine spatial boundaries 	<ul style="list-style-type: none"> • GTFS data 	<ul style="list-style-type: none"> • Systematic approach to analyze the empirical delay propagation patterns based on schedule and real-time GTFS data

	<ul style="list-style-type: none"> • Distance decay 			
Ma, X., M.ASCE, Wang, Y. (2014)	<ul style="list-style-type: none"> • Network-level Commercial Speed • Route-level transit travel time reliability • Stop-level ridership • Stop-level headway variance 	<ul style="list-style-type: none"> • Transit path-finding algorithm • 95th percentile travel time • buffer index 	<ul style="list-style-type: none"> • AFC • AVL 	<ul style="list-style-type: none"> • Demonstrated feasibility of establishing a web-based e-science system for transit performance measures
Liu, L., Porr, A., Miller, H. J. (2022)	<ul style="list-style-type: none"> • Accessibility 	<ul style="list-style-type: none"> • Space-time prismgeography method 	<ul style="list-style-type: none"> • GTFS data 	<ul style="list-style-type: none"> • Introduced time geography approach (realizable real-time space) in the calculation of accessibility
Morri, N., Hadouaj, S., Said, L. B. (2021)	<ul style="list-style-type: none"> • Punctuality index • Regularity index • Correspondence index 	<ul style="list-style-type: none"> • Measure difference between the scheduled and observed vehicle data 	<ul style="list-style-type: none"> • GTFS data • AnyLogic 	<ul style="list-style-type: none"> • Developed a transit control system that detects the traffic disturbance of itineraries and generate regulation action in real time

Note: AVL, Automatic Vehicle Location; APC, Automatic Passenger Count; GTFS, General Transit Feed Specification Transit Data; NTD, National Transit Database data; TCQSM: Transit Capacity and Quality of Service Manual

For instance, Liao and Liu (2010) developed an integrated data-processing framework with raw AVL-APC and electronic fare collection data enabling the creation of systematic transit performance measures and Liu et al. (2020) evaluated travel reliability by identifying three reliability indicators, connectivity reliability, travel time reliability, and capacity reliability, with OD station pair data from fare collection data. General Transit Feed Specification (GTFS) data, provides information on agency services' stops, trips, routes, and fares (Wong, J. C., 2013) and was used by Handley et al. (2019) and Wong, J. C. (2013) to evaluate service connectivity and deficiencies in the transit system. National Transit Data (NTD) offers information about the agency, the financial status of the transit operation, and the transit service (Ryus, P., 2003). Rodier and Issac (2016) suggested performance measures addressing safety, mobility, spare ratio, cost-effectiveness, and energy consumption using the NTD and survey data. Other analyses may use additional data to supplement the frequently used transit data sources. For instance, Al Mamun and Lownes (2011) used Census and transit data to develop tract-based transit accessibility scores, and Liu et al. (2022) measured Wuhan's rail transit system using attributes of routes and walking distance data. Moreover, real-time transit data can be used to generate real-time transit performance measures. By comparing the GTFS schedules and real-time data, Park et al (2019) measured delays; Liu et al (2022) measured accessibility by introducing a new time geography approach, called a realizable real-time space-time prism; and Morri et al (2021) developed three indices, punctuality index, regularity index, and correspondence. Alternatively, using AFC and AVL data, various real-time performance measures such as speed, travel time, ridership, and headways were measured at the network, route, and stop levels (Ma, X. et al., 2014).

2.1.1.2 Pedestrians

Walking is increasingly promoted to citizens and communities, but there have been limited studies addressing pedestrians compared to motorized vehicles (Landis et al., 2016). An examination of literature found studies that introduced performance measurement methods for pedestrian mode. Most of the reviewed studies developed a Pedestrian Level of Service (PLOS) metric using different regression models (Petritsch, T. A., 2005; Daniel et al., 2016; Landis et al., 2016). The models include Pearson correlation analysis and regression model (Petritsch, T. A., 2005), multiple linear regression (Daniel et al., 2016), and a stepwise multivariable regression (Landis et al., 2016). Field collected data, video, and survey were used for most of the developed PLOS metrics. Choobchian et al. (2024) evaluated walkability using a data-driven approach with structural equation modeling using three different walkability indicators. Additionally, Hubbard et al. (2008) introduced real-time pedestrian performance measures, integrating real-time performance measures with existing controller and vehicle detection technology.

Table 2.2. Summary of studies on pedestrian performance measures identification and collection

Reference	Performance Measures	Methods	Data Source(s)	Key Findings
Petritsch, T. A. (2005)	<ul style="list-style-type: none"> • Pedestrian LOS 	<ul style="list-style-type: none"> • Pearson correlation analyses and regression model 	<ul style="list-style-type: none"> • Field data collection and video simulations 	<ul style="list-style-type: none"> • Developed a LOS model for pedestrians at signalized intersections. • Identified six primary factors in the Pedestrian LOS model for intersections, such as right-turn-on-red volumes, permissive left turns from the street parallel to the crosswalk, vehicle volume, midblock 85 percentile speed of the vehicles on the street being crosses.
Hubbard, S. M., Bullock, D. M., & Day, C. M. (2008)	<ul style="list-style-type: none"> • Real-time pedestrian performance measures at signal intersections 	<ul style="list-style-type: none"> • HCM pedestrian delay method • Pedestrian space method • Right-turn flow rate during pedestrian interval 	<ul style="list-style-type: none"> • Loop or video detection data including vehicle occupancy, vehicle gaps, vehicle speeds, and V/C ratio. • Signal controller event-based logging data 	<ul style="list-style-type: none"> • Provided real-time pedestrian performance measures using existing controllers and vehicle detection technology.
Daniel et al. (2016)	<ul style="list-style-type: none"> • Pedestrian Footpath LOS 	<ul style="list-style-type: none"> • Multiple linear regression 	<ul style="list-style-type: none"> • On-site measurement, video & visual walkthrough surveys 	<ul style="list-style-type: none"> • The factors considered fall into three categories: physical, location, and user attributes. • Pedestrian's perception of comfort and safety is influenced not only by the walking environment and conditions but also by the adjacent road and traffic characteristics
Landis, B. W., Vattikuti, V. R., Ottenberg, R. M., McLeod, D. S., & Guttenplan, M. (2001).	<ul style="list-style-type: none"> • Pedestrian LOS within a Roadside environment 	<ul style="list-style-type: none"> • A stepwise multivariable regression 	<ul style="list-style-type: none"> • Surveys 	<ul style="list-style-type: none"> • Lateral separation factors, outside lane traffic volume, speed, and driveway access frequency are the main factors related to pedestrian LOS. • The model was developed through a stepwise multivariable regression analysis of 1,250 observations from an event that placed 75 people on a roadway walking course in the Pensacola, Florida, metropolitan area.

Choobchian, P., Mohammadi, A., Zou, B., Hair, J. F., Valinejad, M., Shin, J., Siriraj, P. S. (2024)	<ul style="list-style-type: none">• Walkability	<ul style="list-style-type: none">• Data-driven approach using structural equation modeling, Partial Least Square Structural Equation model	<ul style="list-style-type: none">• American Community Survey (ACS)	<ul style="list-style-type: none">• Walkability affects the share of commute walk trips the most.• Transit- and job density-related indicators are the most critical.• Socio-economic factors influence the tendency of walking to commute
---	---	---	---	--

2.1.1.3 Micromobility

Micromobility has emerged in recent years as a popular mode of transportation for short-distance travel. One common micromobility mode is the e-scooter, for which performance measures derived from trajectory data is most commonly used. Using the e-scooter trajectory data, Zuniga-Garcia et al. (2021) conducted a spatial-temporal analysis, and Caspi et al. (2020) evaluated travel behavior patterns of e-scooter in Austin, Texas by the descriptive statistical analysis and geographically weighted regression model. Alternatively, Yan et al. (2021) correlated transit and station-based bike share system data with the e-scooter performance with correlation analysis.

Table 2.3. Summary of studies on e-scooter performance measures identification and collection

E-Scooter				
Reference	Performance Measures	Methods	Data Sources (s)	Key Findings
Zuniga-Garcia, N., Juri, N. R., Perrine, K. A., & Machemehl, R. B. (2021).	<ul style="list-style-type: none"> • VMT and speeds across different urban infrastructure types 	<ul style="list-style-type: none"> • Spatial-temporal analysis 	<ul style="list-style-type: none"> • E-Scooter trajectory data 	<ul style="list-style-type: none"> • An average e-scooter trip distance is split between sidewalks (18 percent), bike lanes (11 percent), and roadways (33 percent), with 38 percent across other unclassified areas. • E-scooter riders have an average of 15.2 km/h on weekdays and 13.7 km/h on weekends.
Caspi, O., Smart, M. J., & Noland, R. B. (2020).	<ul style="list-style-type: none"> • Travel behavior patterns of e-scooter use in Austin 	<ul style="list-style-type: none"> • Descriptive statistical analysis. • Geographically Weighted Regression • 	<ul style="list-style-type: none"> • E-scooter trajectory data, land use, and street network data 	<ul style="list-style-type: none"> • Commuting is not the main trip purpose for e-scooter riders. Also, most trips are in residential, commercial, and industrial areas. • The usage of e-scooter is associated with areas with high employment rates, and in areas with bicycle infrastructure. • People use e-scooter regardless of the affluence of the neighborhood. And college towns are a ready market for e-scooter sharing services.
Yan, X., Yang, W., Zhang, X., Xu, Y., Bejleri, I., & Zhao, X. (2021).	<ul style="list-style-type: none"> • The relationships with transit and station-based bike sharing system 	<ul style="list-style-type: none"> • Correlation analysis 	<ul style="list-style-type: none"> • General Bikeshare Feed Specification data and real-time bike-share system data 	<ul style="list-style-type: none"> • E-scooters have both complementary and substitution effects on bike share and public transit. • Travelers pay a price premium and save time when choosing e-scooters over transit.

Note: VMT: Vehicle miles traveled

Bike-sharing system performance measures frequently used bike-sharing GPS data. For example, using the bike-sharing GPS data, Kabra et al. (2020) evaluated the accessibility and availability of the bike-sharing system in Paris, France with origin-destination model and user-choice model, and Pashkevich et al. (2021) assessed the bike-sharing system in Krakow, Poland using static analysis method. Additionally, Berke et al. (2024) used bike-share network data and census data to evaluate the U.S. cities' bike-share network's equity by quantifying spatial equity with the Gini index.

Table 2.4. Summary of studies on bike performance measures identification and collection

Bikeshare system				
Reference	Performance Measures	Methods	Data Sources	Key Findings
Kabra, A., Belavina, E., & Girotra, K. (2020).	<ul style="list-style-type: none"> • Accessibility (how far the user walks to reach stations) and availability (likelihood of finding a bicycle) 	<ul style="list-style-type: none"> • Origin-Density Model • User-Choice Model 	<ul style="list-style-type: none"> • Bike sharing data in Paris • POI from Google • Metro Ridership and tourist volume • Weather data 	<ul style="list-style-type: none"> • Every additional meter of walking to a station decreases a user's likelihood of using a bike from that station by 0.194%. • 80% of bike-share usage comes from areas within 300 m of stations.
Berke, A., Truitt, W., & Larson, K. (2024).	<ul style="list-style-type: none"> • Transportation equity 	<ul style="list-style-type: none"> • Quantifying spatial equity with the Gini index • Quantifying equitable access by demographic inclusion 	<ul style="list-style-type: none"> • bike-share data and census data 	<ul style="list-style-type: none"> • Evaluated how well five large U.S. city bike-share networks have addressed equity as they expanded. • Higher-income and White populations were consistently included in bike-share service areas at higher rates.
Pashkevich, A., Kłos, M. J., Jaremski, R., & Aristombayeva, M. (2021)	<ul style="list-style-type: none"> • Load of station as an origin (or destination) of trips • Travel time and Trip distance • Attractiveness points of each station 	<ul style="list-style-type: none"> • Statistical analysis 	<ul style="list-style-type: none"> • Bike-sharing GPS data 	<ul style="list-style-type: none"> • Evaluated the station-based bike-sharing system in Krakow, Poland. • criteria of stations attractiveness were introduced based on everyday activities and public transport accessibility

2.1.1.4 Air Quality

Sustainability has been one of the primary goals of transportation systems and planning. Measuring the air quality according to the transportation performance allows the evaluation of environmental impact of the current transportation system. According to some studies regarding the emission due to transportation, the most important air pollutants include volatile organic compounds (VOCs), carbon monoxide (CO), oxides of nitrogen (NO_x), oxides of sulfur (SO_x), and particulate matter (PM) (Delucchi, M., 1996; Clark et al., 2017; Aminzadegan et al., 2022). The air quality and emission were evaluated by different emission models such as Emission Factor (EMFAC) model, greenhouse-gas emissions model, and regression model with emission data from agencies, studies, and United States Environmental Protection Agency (U.S. EPA) (Delucchi, M., 1996; Aminzadegan et al., 2022).

Table 2.5. Summary of studies on air quality due to transportation operation measures identification and collection

Transportation Impact to Air Quality Measures				
Reference	Criteria Air pollutants	Methods	Data Sources (s)	• <i>Key Findings</i>
Delucchi, M. A., McCubbin, D. R., (1996)	<ul style="list-style-type: none"> • volatile organic compounds (VOCs) • carbon monoxide (CO) • oxides of nitrogen (NO_x) • oxides of sulfur (SO_x) • particulate matter (PM; including small-diameter PM₁₀, in some cases) 	<ul style="list-style-type: none"> • CARB's EMFAC emission model, CEC's Elfin model, greenhouse-gas emissions model 	<ul style="list-style-type: none"> • CARB's EMFAC raw data, toxic-emission inventory data 	<ul style="list-style-type: none"> • social cost of a transportation mode considering the health effects of air pollution
Clark, L. P., Millet, D. B., Marshall, J. D. (2017)	<ul style="list-style-type: none"> • NO₂ 	<ul style="list-style-type: none"> • Regression model with Census Demographic data 	<ul style="list-style-type: none"> • Air pollution data from published research (Bechle et al., 2015) 	<ul style="list-style-type: none"> • Estimated annual average NO₂ concentrations decreased from 2000 to 2010 for all of the race-ethnicity following the declining transportation-related air pollutant emissions
Aminzadegan, S., Shahriari, A., Mehranfar, F., Abramović, B.(2022)	<ul style="list-style-type: none"> • SO₂ • PM • CO • NO_x • N₂O • CH₄ • CO₂ 	<ul style="list-style-type: none"> • Analysis based on the investigation of literature 	<ul style="list-style-type: none"> • Literature Review 	<ul style="list-style-type: none"> • Effect of greenhouse gases from transportation, impact of pollutants for each transportation mode

2.1.2 Examples of Multimodal Performance Indicator Applications

Agencies and researchers have been applying different performance measures of multimodal transportation to system and network planning, Transportation Systems Management & Operations (TSMO) strategies, corridor or project planning, street design, etc.

The Florida Department of Transportation (FDOT) has two applications of multimodal transportation measures: the Complete Streets Implementation Plan (FDOT and Smart Growth America, 2015) and the TSMO strategic plan (FDOT, 2017). FDOT and Smart Growth America (2015) used seven types of complete streets goals: safety, accessibility, economy, environmental, public health, social equity, and life quality, to evaluate the Complete Streets Implementation Plan. To evaluate the quality of experience, FDOT and Smart Growth America (2015) also applied LOS to the transit, bicycle, and pedestrian modes. For the TSMO strategic plan, FDOT (2017) used Regional Integrated Traffic Information System (RITIS) for performance metrics calculation, tracking the performance, and identifying optimal set of TSMO. The Oregon Department of Transportation (ODOT) (2021) used RITIS/INRIX probe data, internal data warehouse/business intelligence system, central signal system, and PORTAL system developed by Portland State University to enhance the availability and use of mobility performance measures. Semler et al. (2016) developed performance measures for pedestrian and bike systems based on community goals of connectivity, economic, environment, equity, health, livability, and safety, aligning with the performance measures such as accessibility, demand, reliability, mobility, and infrastructure. Additionally, multimodal level of service (MMLOS) was applied to assess auto, bus, bicycle, and pedestrian level of service (Washington State Transportation Commission, 2018; Khedri et al., 2022). The studies highlighted the needs of different MMLOS standards according to the project-specific goals.

Table 2.6. Summary of multi-modal transportation performance measures application examples

References	Measures	Applications	Study Locations	Findings
FDOT and Smart Growth America. (2015).	<ul style="list-style-type: none"> Seven types of complete street goals and each goal consider several different performance measures. 	<ul style="list-style-type: none"> Complete Streets Implementation Plan 	Florida	<ul style="list-style-type: none"> Successfully integrating a complete street approach into FDOT's practice requires measures and criteria used to evaluate proposed future investments, the performance of individual transportation facilities, the performance of the network as a whole, and the general effectiveness of FDOT's programs
Washington State Transportation Commission. (2018).	<ul style="list-style-type: none"> Multimodal LOS 	<ul style="list-style-type: none"> Statewide Transportation Policy Plan 	Washington	<ul style="list-style-type: none"> The statewide policy plan provides the overarching framework of statewide goals. The plan requires local and state agencies to establish multimodal LOS standards based on their goals and use concurrency to maintain them since there is no one-size-fits-all measure of transportation system performance.
FDOT, 2017.	<ul style="list-style-type: none"> Three types of program goals (Mobility, Safety, and ITS/communication networks maintenance) have several performance metrics, respectively. 	<ul style="list-style-type: none"> TSMO Strategic Plan 	Florida	<ul style="list-style-type: none"> Performance metrics can track the project performance enhancement and identify the optimal set of TSMO strategies. Archived in RITIS includes FDOT traffic detector data, privately collected vehicle probe data, and SunGuide® event data. These data are used to calculate the metrics.
ODOT, 2021	<ul style="list-style-type: none"> Mobility-related performance measures 	<ul style="list-style-type: none"> Operations program performance management plan 	Oregon	<ul style="list-style-type: none"> Create actionable performance measures that directly support the stewardship of Oregon's transportation system. Core performance measures are used to understand ODOT's ability to plan for, implement, operate, and maintain facilities. Also, these measures can be used to validate whether a project is producing the intended outcome.

References	Measures	Applications	Study Locations	Findings
Khedri, B., Malarkey, D., & MacKenzie, D. (2022).	<ul style="list-style-type: none"> Multimodal level of service measures 	<ul style="list-style-type: none"> Project Evaluation 	NA	<ul style="list-style-type: none"> Summarized practice for multimodal design and the use of multimodal performance measures in transportation projects. Current practices have not yet standardized on a single approach or list of performance measures to fit all projects.
ODOT, 2021	<ul style="list-style-type: none"> Mobility-related performance measures 	<ul style="list-style-type: none"> Operations program performance management plan 	Oregon	<ul style="list-style-type: none"> Created actionable performance measures that directly support the stewardship of Oregon's transportation system. Core performance measures are used to understand ODOT's ability to plan for, implement, operate, and maintain facilities. Also, these measures can be used to validate whether a project is producing the intended outcome.
Semler, C., Vest, A., Kingsley, K., Mah, S., Kittelson, W., Sundstrom, C., & Brookshire, K. (2016).	<ul style="list-style-type: none"> Pedestrian and bike system performance measures 	<p>Local Jurisdiction (county, city);</p> <ul style="list-style-type: none"> Street Design, System/Network Planning Corridor or project planning Development review <p>Regional Planning Agency (MPO);</p> <ul style="list-style-type: none"> System/network planning Regional Policy Development Funding Allocation <p>State Agency (DOT)</p>	NA	<ul style="list-style-type: none"> Helped communities develop performance measures that can fully integrate pedestrian and bicycle planning in ongoing performance management activities. Discussed how the measures can be tracked and what data required.

References	Measures	Applications	Study Locations	Findings
		<ul style="list-style-type: none">• Statewide system/network planning• Statewide Policy Development• Code Compliance		

Some studies were conducted to apply air quality measures to multimodal transportation systems. Moon et al. (2022) used the relationship between the traffic volume, simulated with Google maps according to the 2012 California Household Travel Survey (CHTS), and emission data, obtained from the Air Quality System (AQS) from U.S. EPA, to evaluate the air quality in California. Another approach to the air quality measure was using simulation tools. Chen et al. (2020) evaluated link-based energy consumption rate and emission rate for transit buses during peak-hour traffic in South Carolina with the traffic simulation using Simulation of Urban Mobility (SUMO), emission simulation using MOVES, and deep learning-based surrogate model.

Table 2.7. Summary of transportation operation impact on air quality performance measures application examples

References	Measures	Method	Data Source	Study Locations	Findings
Moon, et al., (2022)	<ul style="list-style-type: none"> Correlation of traffic volume with traffic-related air pollutant 	<ul style="list-style-type: none"> Calculate correlation of monitored traffic volume in the local area with air quality data 	Traffic data: Google Maps Air quality data: Air Quality System (AQS) from EPA	California	<ul style="list-style-type: none"> Correlation between traffic volume and two common vehicular pollutants is more significant when traffic volume is high. Potential interference of other determining factors (weather, industrial activities)
Chen et al., (2020)	<ul style="list-style-type: none"> Statistical correlations of the energy and air quality impacts with changes in traffic activities 	<ul style="list-style-type: none"> Applied traffic simulation's output to EPA's Motor Vehicle Emissions Simulator 	Reference for traffic data: iPeMs Air quality data: EPA's Motor Vehicle Emissions Simulator, MOVES	South Carolina	<ul style="list-style-type: none"> Developed neural network models by applying the traffic simulation of city of Columbia to the EPA's Motor Vehicle Emissions Simulator Achieved 15% mean absolute percentage error relative to MOVES results

CHAPTER 3. TRANSIT DATA COLLECTION AND ANALYSIS

In order to understand the data before defining performance measures, various available data sources were collected and explored to determine whether they would be further considered for the transit performance measurements in **CHAPTER 6**.



3.1. Deliverables

Table 3.1 summarizes the available data sources, collection methods, and use cases. This study comprehensively examines the GTFS data and is structured as outlined in the following. **Section 3.2.1** provides an overview of the GTFS data, summarizing its components and general applications. **Sections 3.2.2** and **3.2.3** delve deeper into two particular GTFS data entities, namely GTFS Trip Updates and GTFS Static data, respectively. These sections cover the details which were further used in developing performance measures in **CHAPTER 6**. Subsequently, **Section 3.2.4** focuses on data analysis derived from GTFS, highlighting insights into bus locations, transit vehicle speeds, and headways across the system. The appendix collects overviews of all other sources that have been considered throughout the tasks but tabled according to various conversations with PAG.

Table 3.1. Summary of Transit Data Sources and Use Cases Considered

Data		Source	Collection Method	Use Cases
GTFS*	Vehicles	Sun Tran	API provided by Sun Tran	GTFS* data accuracy measure
	Trip Update	Sun Tran	API provided by Sun Tran	Performance measures: <ul style="list-style-type: none"> • <i>Speed</i> • <i>Travel time reliability</i> • <i>Headway</i> • <i>Headway reliability</i> • <i>Irregularity</i> • <i>On-time performance</i>
	GTFS* Static Data	Sun Tran	API provided by Sun Tran	Performance measures: <ul style="list-style-type: none"> • <i>Speed</i> • <i>Travel time reliability</i> • <i>Headway</i> • <i>Headway reliability</i> • <i>Irregularity</i> On-time performance
GTFS*: General Transit Feed Specification				

3.2. General Transit Feed Specification (GTFS)

This section is dedicated to a comprehensive overview of the General Transit Feed Specification (GTFS) data and its analytical applications. It delves into the specifics of each GTFS component, detailing the type of information provided by each dataset. The analysis conducted using this data encompasses bus locations, speeds, and headways, providing a thorough understanding of transit operations and performance metrics.

3.2.1 GTFS General Overview

The project's goal is to establish performance measures that measure high-resolution performance of multi-modal transportation system in the PAG region. To achieve this goal, obtaining real-time data for each transportation mode is critical. GTFS provides information on Sun Tran's operation in real-time. As depicted in [Table 3.2](#), there are two types of data, GTFS Real Time (hereafter, GTFS-RT) and GTFS Schedule (hereafter, GTFS-ST). The GTFS-RT provides real time updated information of transit system about trip updates, vehicle positions, and alerts. The GTFS-ST provides static and scheduled transit system information such as scheduled time of departure and transit system networks.

Table 3.2. Sun Tran GTFS feeds summary

GTFS Data	Feed Entity	Description
RT	Trip Update	Realtime update for departure events for a given stop on a trip
	Vehicle	Realtime positioning information for a given vehicle
	Alert	Indicating some sort of incident in the public transit network
ST	agency	Transit agencies with service represented in this data
	calendar	Service dates specified using a weekly schedule with start and end dates
	calendar_dates	Exceptions for the services defined in the calendar
	feed_info	Dataset metadata, including publisher, version, and expiration information
	stops	Stops where vehicles pick up or drop off riders
	transfers	Rules for making connections at transfer points between routes
	trips	Trip for each route. A trip is a sequence of two or more stops that occur during a specific time period
	routes	Transit routes. A route is a group of trips that are displayed to riders as a single service
	shapes	Rules for mapping vehicle travel paths (route alignments)
stop_times	Times that a vehicle arrives at and departs from stops for each trip	

Note: Ref.: Google Inc. MobilityData (Oct 3, 2024)

As shown in [Figure 3.1](#), GTFS has different types and entities. Under the upper branch, the “Sun Tran” (including the Sun Tran modes of Sun Tran bus, Express Bus, and Sun Link), there are GTFS-RT and GTFS-ST. The Sun Tran GTFS-RT data includes Vehicle, Trip Update, and Alert entities. The Sun Tran GTFS-ST includes agency, calendar, calendar_dates, feed_info, routes, shapes, stop_times, stops, transfers, and trips entities.

The lower branch, the “Sun Shuttle”, is a GTFS data entity provided separately from the upper branch. Sun Shuttle only has the GTFS-RT which encompasses Vehicle and Trip Update entities.

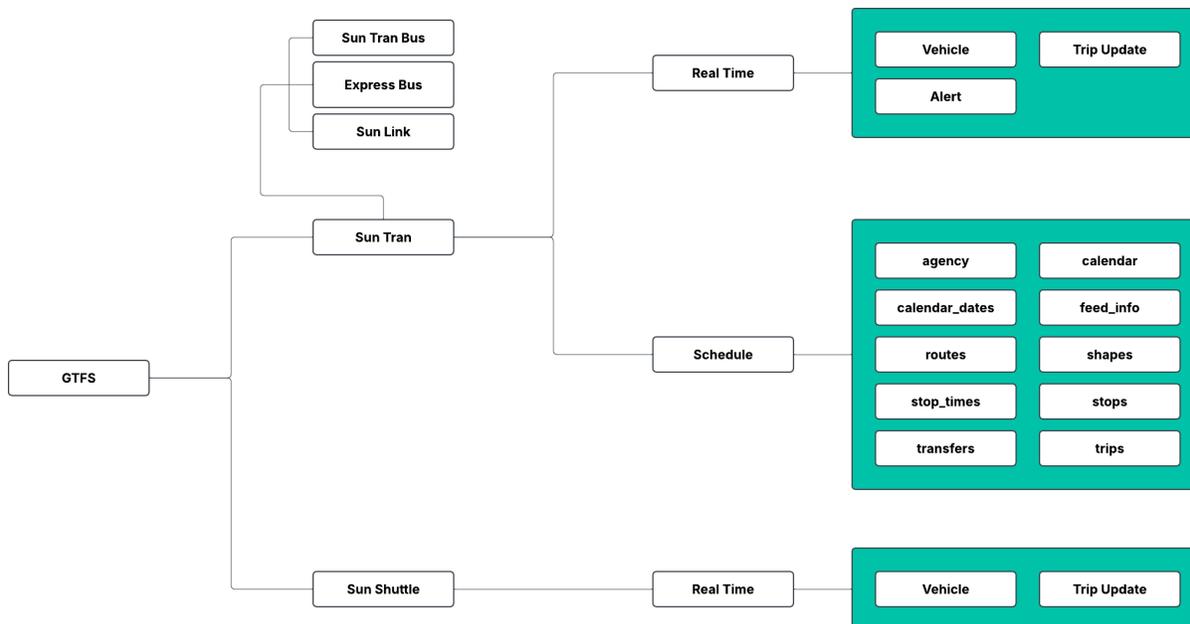


Figure 3.1. Sun Tran GTFS Data Structure

3.2.2 *GTFS Trip Update*

GTFS Trip Update provides real-time updates for departure events for a given stop on a trip. As illustrated in [Figure 3.2](#), the Sun Tran system is composed of different elements: vehicle, trip, stop, and route. The vehicle is the vehicle serving each route and running for a trip. The trip refers to a set of stop sequences in a route that the vehicle traveled. Stops are physical transit stops, which can lie on multiple routes. The route is a group of trips that are displayed to riders as a single service. The path of each vehicle should follow the designated route. The Trip Update data is a predicted departure time for stops along the route considering the operation and road conditions. According to Google Inc. MobilityData (2023), Trip Update reflects deviations from the scheduled timetable by providing predicted arrival and departure times for stops along a route.

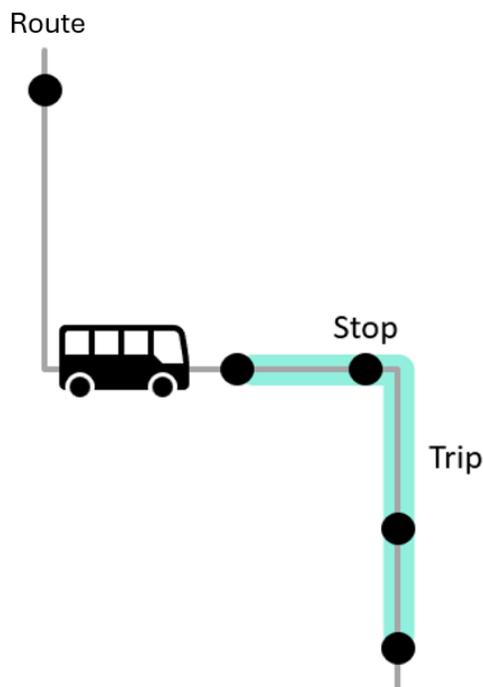


Figure 3.2. GTFS Route System

To understand the GTFS Trip Update data information availability, raw data was collected without specified data structure formatting. The *gtfs-realtime-bindings* module, providing Python classes generated from the GTFS-realtime Protocol Buffer specification (Google Inc. MobilityData, 2023), and the *requests* library were installed and used with Python 3.10.0 to obtain the GTFS data from the Protocol Buffer specification. The raw data provided for each vehicle's trip sequence information as described in [Figure 3.3](#).

To organize the data at the route level, a data structure was defined with a dictionary format containing the route ID as the first dimension, the vehicle ID as the second dimension, and the stop updates information as the third dimension. With this data structure, the GTFS Trip Update data was archived at 1 second time intervals. The observed data updating behavior indicates that when a trip sequence is updated, the newly retrieved data includes the complete set of updated records for the trip.

The Trip Update data was used to measure transit's travel time reliability, headway, irregularity, headway reliability, speed, and on-time performance. With the departure time of each route's stops, the performance could be quantified and measured.

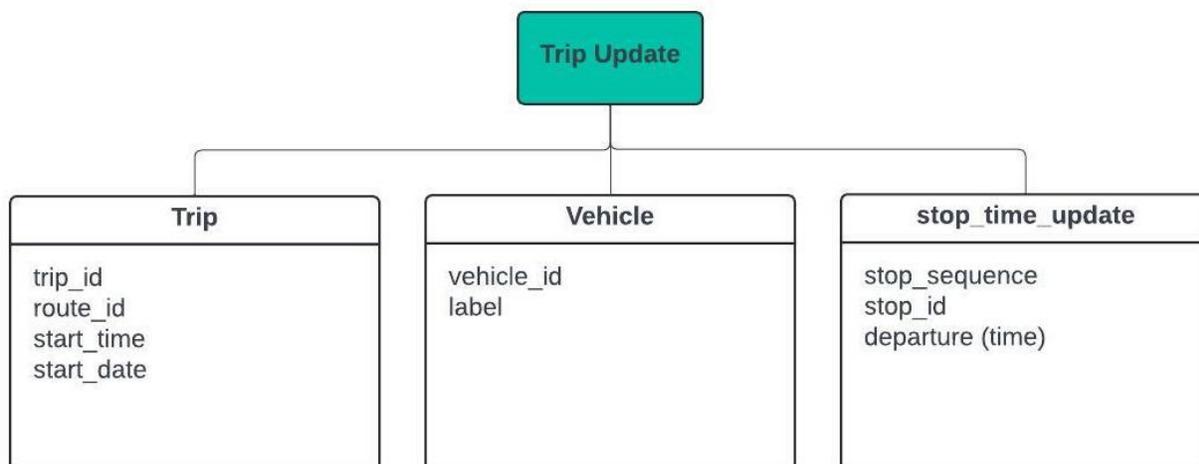


Figure 3.3. GTFS Trip Update Data Structure

3.2.3 GTFS Static Data

GTFS static data was provided along with the GTFS API links. The static data offers an understanding of how each ID of the transit system elements, such as route_id, trip_id, shape_id, and stop_id, were coded and what they represent. The route_id in GTFS data refers to the identification number of each route. This data consists of nine files in comma-separated values (CSV) format, including the file names of *agency*, *calendar*, *calendar_dates*, *feed_info*, *routes*, *shapes*, *stop_times*, *stops*, and *trips*. For transit performance measures, understanding the geographic information of each GTFS data is crucial. The geographic data from the GTFS static data was referenced with the API data to ensure analysis comprehensiveness.

3.2.3.1 Stops

As indicated in [Table 3.3](#), the *stops* file provides information about each stop and assigns a unique identifier to each stop. It describes where the stop is located with the coordinates (longitudes and latitudes) and provides the stop's name and code.

Table 3.3. Stops parameters definitions

Parameter	Description
stop_id	Identifies a location: stop/platform, station, entrance/exit, generic node, boarding area. <ul style="list-style-type: none"> ID must be unique across all stops.stop_id Multiple routes may use the same stop_id
stop_code	Short text or a number that identifies the location for riders
stop_name	Name of the location
stop_lat /stop_lon	Latitude and longitude of the location

Note: Ref.: Google Inc. MobilityData (May 22, 2024)

3.2.3.2 Routes

As outlined in [Table 3.4](#), the *routes* file elaborates on each transit route, clarifying how route identifiers and color codes correlate to specific routes operated by Sun Tran allowing different transit routes to be identified and distinguished.

Table 3.4. Routes parameters definitions

Parameters	Description	Options Definition
route_id	Identifies a route	NA
route_short_name	Short name of a route (often a short, abstract identifier)	NA
route_long_name	Full name of a route (generally more descriptive than the route_short_name)	NA
route_type	Indicates the type of transportation used on a route	0: Tram, Streetcar, Light rail 1: Subway, Metro 3: Bus 4: Ferry 5: Cable tram 6: Aerial lift 7: Funicular 11: Trolleybus 12: Monorail

Note: Ref.: Google Inc. MobilityData (May 22, 2024)

3.2.3.3 Trips

The *trips* file lists all potential trips for each route within the transit system. A trip is a sequence of two or more stops that occur during different operation days and times (Google Inc. MobilityData, Retrieved May 24, 2024). The list includes the information of each trip's identifier, route and block it aligns with, and shape identifier associated with its route as can be seen from [Table 3.5](#). This file is critical for understanding the scheduling and operations of each trip.

Table 3.5. Trips parameters definitions

Parameters	Description	Options Definition
route_id	Identifies a route	NA
service_id	Identifies a set of dates when service is available for one or more routes	1: Monday – Friday 2: Saturday 3: Sunday 10201, 10401, 7701, 1901: N/A
trip_id	Identifies a trip	NA
trip_headsign	Text that appears on signage identifying the trip's destination to riders	NA
Trip_short_name	Uniquely identifies a trip within a service day	NA
direction_id	Indicates the direction of travel for a trip	0: Travel in one direction 1: Travel in the opposite direction
block_id	Identifies the block to which the trip belongs	NA
shape_id	Identifies a geospatial shape describing the vehicle travel path for a trip	NA

Note: Ref.: Google Inc. MobilityData (May 22, 2024)

3.2.3.4 Shapes

The *shapes* file provides the geographical information of each trip ID. As shown in [Table 3.6](#), the longitudinal and latitudinal coordinates of each trip ID and the traveled distance from the beginning point of the sequence to the current sequence point were defined. [Figure 3.4](#) demonstrates each route's geographical visualization from the *shapes* file data. The *shapes* file was enhanced to include route identifiers for the route-level geographical analysis.

Table 3.6. Shapes parameters definitions

Parameters	Description
shape_id	Identifies a shape
shape_pt_lat	Latitude of a shape point
shape_pt_lon	Longitude of a shape point
shape_pt_sequence	Sequence in which the shape points connect to form the shape (increase along the trip but do not need to be consecutive)
shape_dist_traveled	Actual distance traveled along the shape from the first shape point to the point specified in the record.

Note: Ref.: Google Inc. MobilityData (May 22, 2024)

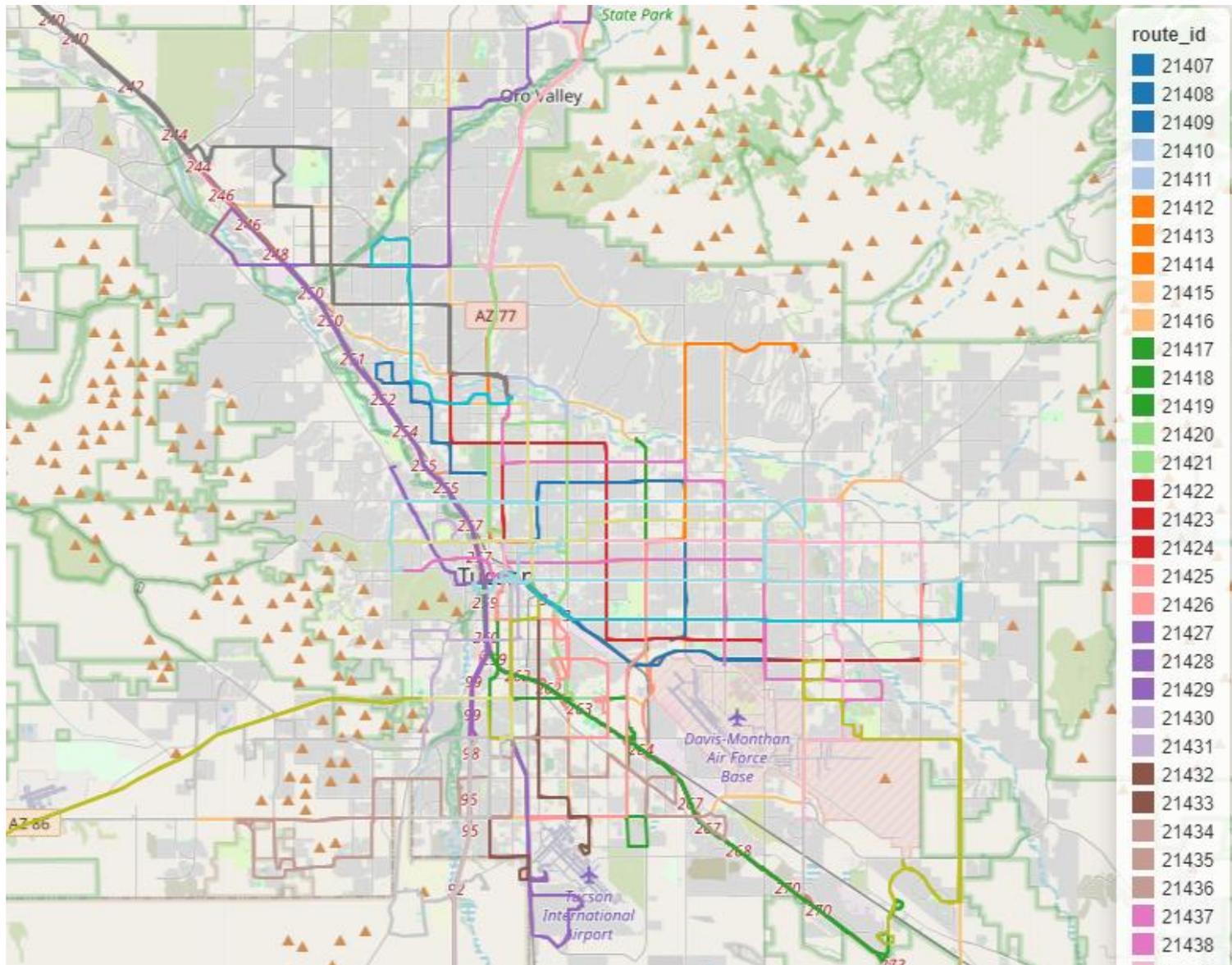


Figure 3.4. GTFS Shapes Static Data

3.2.3.5 Stop Times

The *stop_times* file provides the sequences, locations, and arrival and departure times of each stop in each trip as shown in [Table 3.7](#).

Table 3.7. Stop times file parameters definitions

Parameters	Description	Options Definition
trip_id	Identifies a trip	NA
arrival_time/ departure_time	Arrival / departure time at the stop for a specific trip	NA
stop_id	Identifies serviced stop	NA
stop_sequence	Order of stops, location groups, or GeoJSON locations for a particular trip	NA
pickup_type	Indicates pickup method	0 / empty: Regularly scheduled pickup 1: No pickup available 2: Must phone agency to arrange pickup 3: Must coordinate with driver to arrange pickup
drop_off_type	Indicates drop off method	0 / empty: Regularly scheduled drop off 1: No drop off available 2: Must phone agency to arrange drop off 3: Must coordinate with driver to arrange drop off
shape_dist_traveled	Actual distance traveled along the associated shape, from the first stop to the stop specified in this record	NA
timepoint	Indicates if arrival and departure times for a stop are strictly adhered to by the vehicle or if they are instead approximate and /or interpolated times	NA

Note: Ref.: Google Inc. MobilityData (May 22, 2024)

The GTFS Static data provided critical information on how the GTFS data can be interpreted and how the geographical visualization can be designed with the given data.

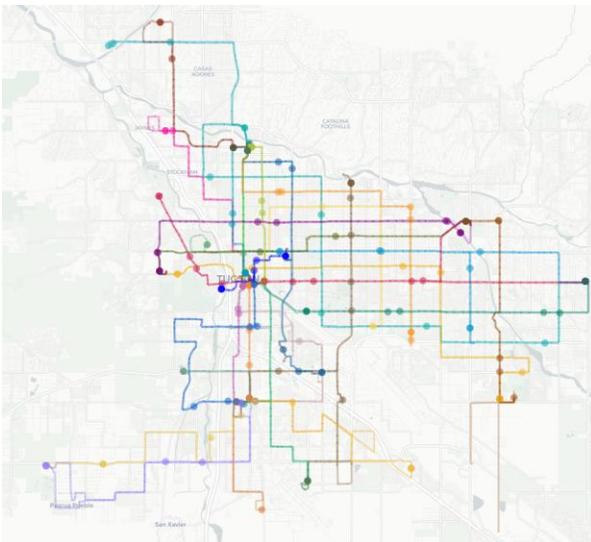
3.2.4 GTFIS Data Analysis

To effectively assess the Sun Tran transit system's operation, data from both the GTFIS-RT and GTFIS-ST were utilized. This approach allowed for comprehensive data analysis, including hour-by-hour system reviews and geospatial explorations. The geospatial analyses were facilitated by integrating geographic information from GTFIS-ST with real-time trip updates from GTFIS-RT. For this analysis, archived GTFIS-RT Trip Update data spanning from July 19, 2024, to August 27, 2024, were examined. The analysis periods were defined as follows: AM peak from 6:30 AM to 8:30 AM, midday from 8:30 AM to 4:00 PM, and PM peak from 4:00 PM to 6:00 PM. This structured timeframe is used for capturing the transit system's performance across different parts of the day, thereby providing insights into peak and off-peak operational efficiencies.

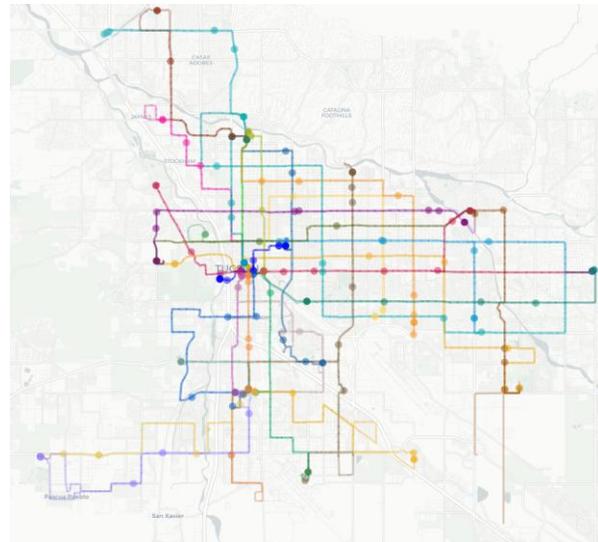
3.2.4.1 Bus Locations

The distribution of bus locations represents the density of transit service across different times of the day. [Figure 3.5](#) illustrates the bus location during selected hours for each time period, which were chosen to represent typical operational patterns: 7:00 AM to 8:00 AM for the AM peak, 12:00 PM to 1:00 PM for midday, and 5:00 PM to 6:00 PM for the PM peak. The bus location is displayed for each unique vehicle with its last observed location. If a vehicle completes its route within a given time frame, the location of its final stop is depicted as the vehicle's position on the map. According to the GTFIS-ST defined schedule, more vehicles operate during daytime, from AM peak to PM peak, and less vehicles are operated during outside of daytime as the ridership decrease. This analysis utilized GTFIS-RT Trip Update data, focusing on departure updates within the specified time period of the AM peak, midday, and PM peak intervals.

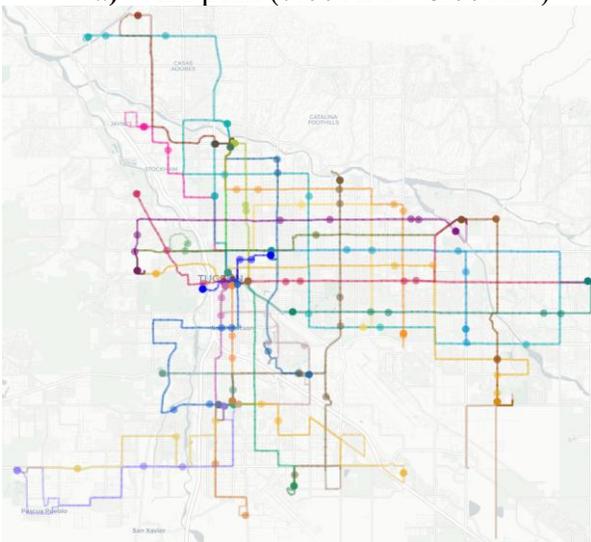
The findings from different time periods show a consistent pattern: a high concentration of vehicles in downtown areas, contrasting with fewer buses and likely longer headways in more remote regions.



a) AM peak (7:00 AM – 8:00 AM)



b) Midday (12:00 PM – 1:00 PM)



c) PM peak (5:00 PM – 6:00 PM)

Figure 3.5. Transit Vehicle Locations Analysis

3.2.4.2 Transit Vehicle Speed

Analyzing transit speed provides valuable insights into service operational performance. It helps identify delayed service routes and time. To assess the efficiency of the transit service, segment-level speeds between stops were analyzed. Using GTFS-RT Trip Update data, departure times at each stop were extracted. Travel time for each segment was calculated by subtracting the departure time at stop i from the departure time at stop $i + 1$. Corresponding segment geometries and travel distances were obtained from the GTFS-ST files. With these values, the speed for each segment was calculated as below (Federal Highway Administration (FHWA), 1997).

$$\frac{t_{i+1} - t_i}{d_{i, i+1}}$$

In which:

t_{i+1} : departure time at stop $i + 1$

t_i : departure time at stop i

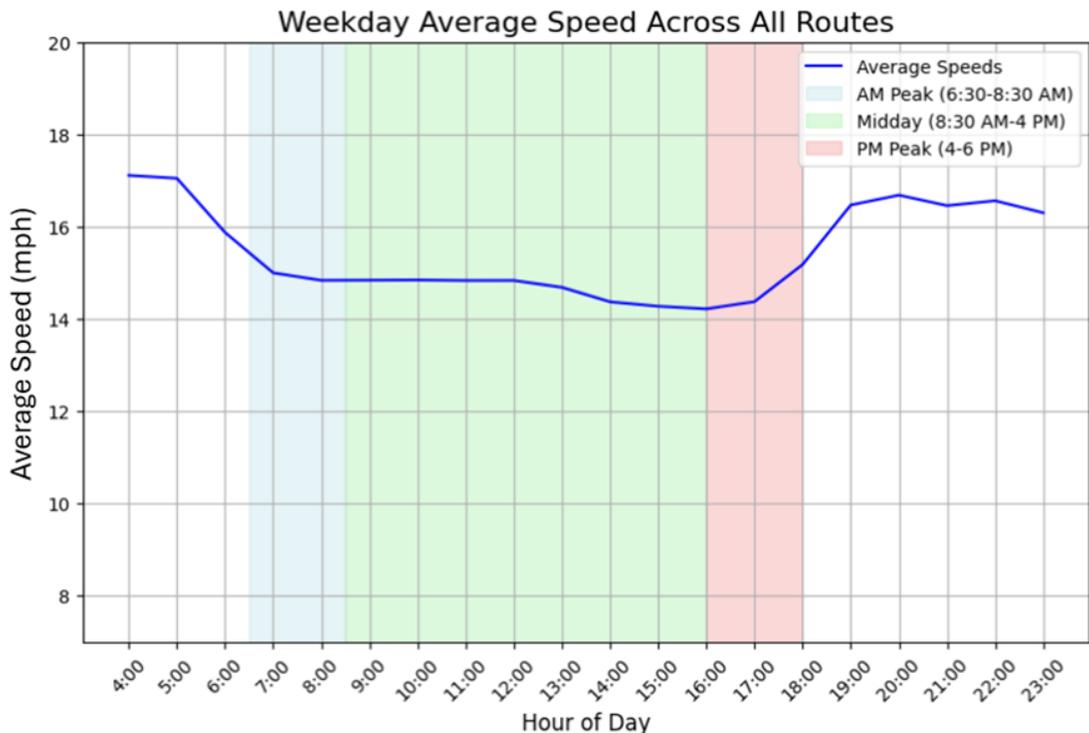
$d_{i, i+1}$: stop-to-stop segment distance from stop i to stop $i + 1$

Weekday and Weekend Average Speed Pattern:

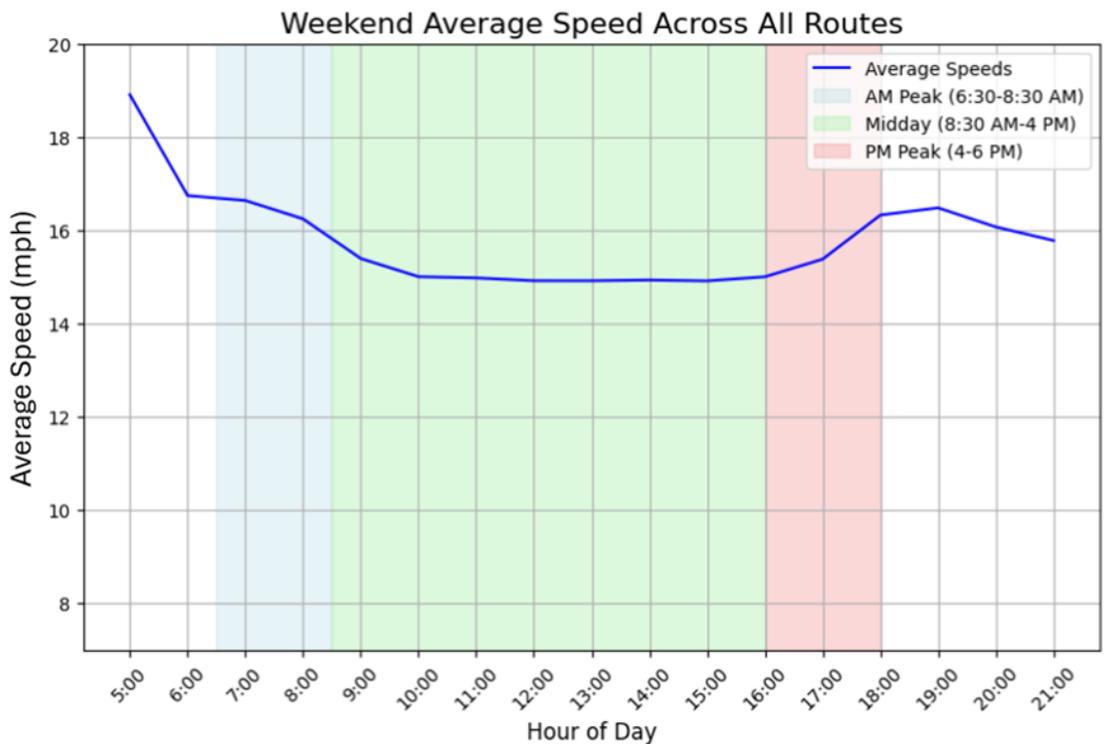
To assess variations in transit speed between weekdays and weekends, we analyzed the traffic patterns, acknowledging that these periods exhibit different characteristics. The weekdays include five weekdays, from Monday to Friday, and weekends include two weekend days, Saturday and Sunday. **Figure 3.6** provides a comparative visualization of the average transit speeds for these two distinct time frames.

During weekdays, speeds are higher before the AM peak and after the PM peak, averaging between 16 mph and 17 mph. However, there is a noticeable decrease during the AM to PM peak period, with speeds ranging between 14 mph and 15 mph. This pattern is likely influenced by increased traffic congestion and higher passenger volumes during these peak hours. There is also a decrease beginning at 2PM and continuing through the start of the PM peak period. The weekend speed pattern shows higher averages before the AM peak, ranging between 17 mph and 19 mph, and a similar pattern to weekdays during the midday period.

This analysis highlights the influence of differing traffic volumes and passenger behavior on transit speed across weekdays and weekends, emphasizing the need for tailored service strategies for each time frame to optimize efficiency and rider experience.



a) Weekday average speed



b) Weekend average speed

Figure 3.6. Weekday and Weekend Average Speed Pattern

3.2.4.3 Transit Vehicle Headways

To comprehensively understand the operations of the transit system, we analyzed and calculated the headways for both scheduled and actual vehicle arrivals at each stop. The vehicle's arrival time was subtracted from the next consecutive vehicle arrival time for each transit station to calculate headways.

According to the Sun Tran website at the time of writing this report, the typical headway is approximately 30 minutes during midday hours. This confirms the range of calculated headway using GTFS-ST and GTFS-RT to be consistent with the operator's defined schedule by ranging 30 – 40 minutes headway during the midday hours as demonstrated in [Figure 3.7](#). The headway calculation was also verified through manual examination of the raw data (Vogel, 2003).

$$t_{v+1} - t_v$$

In which:

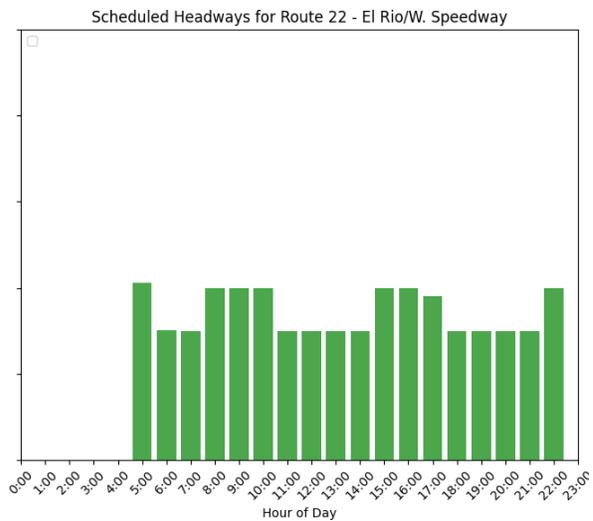
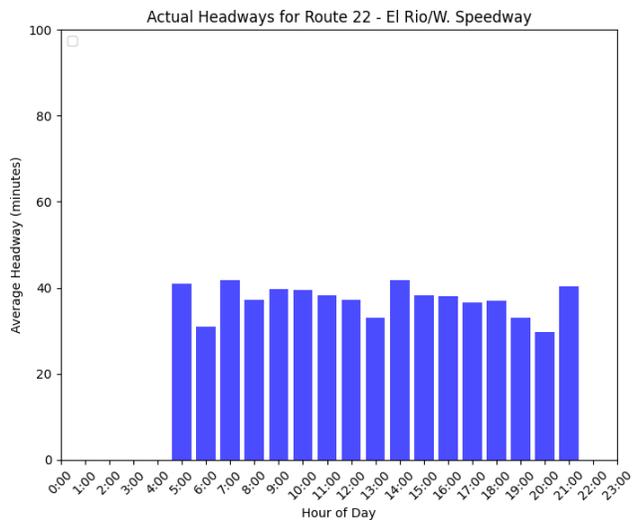
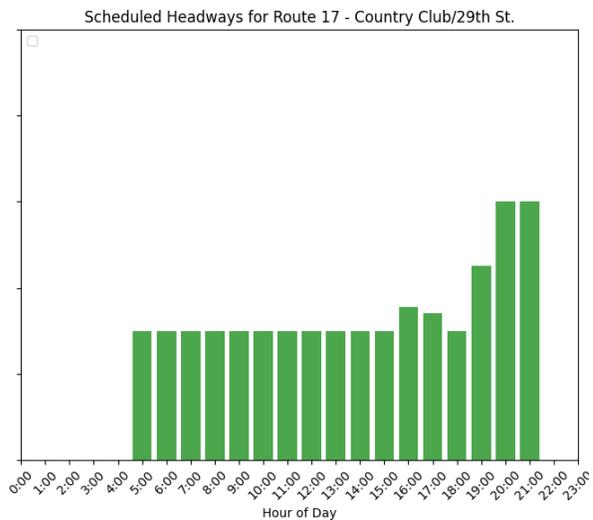
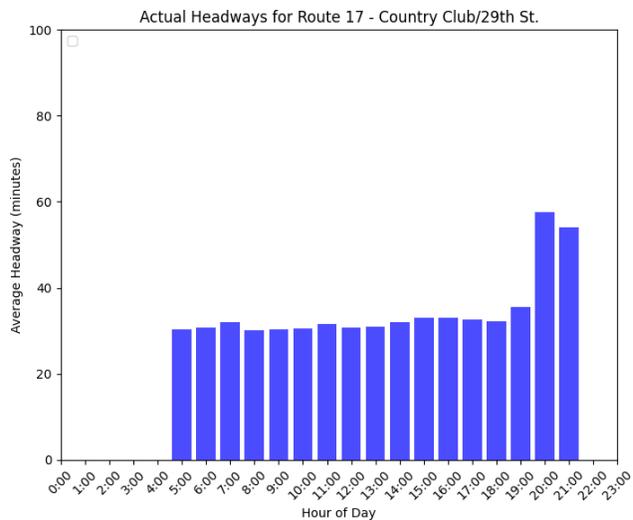
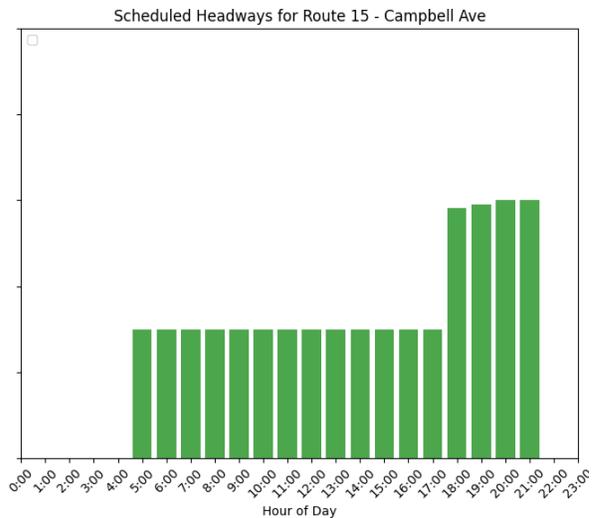
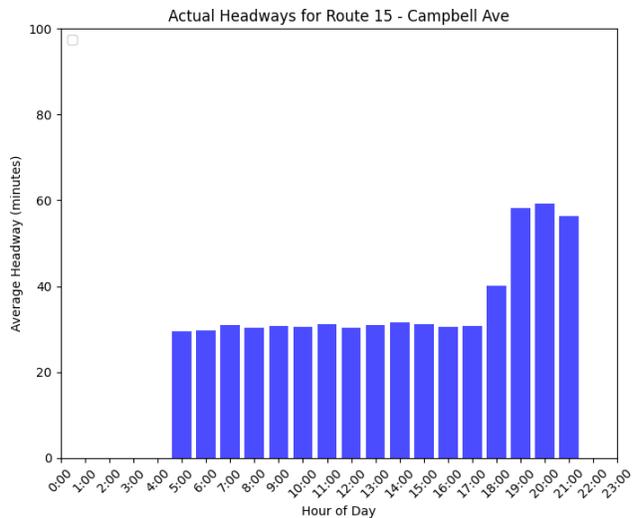
t_v : vehicle's arrival time

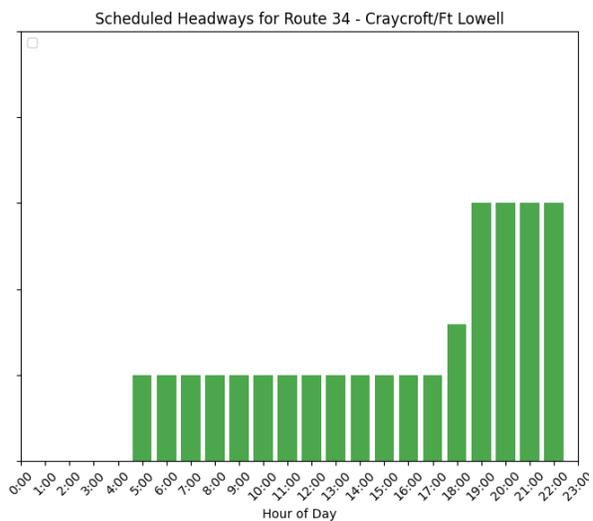
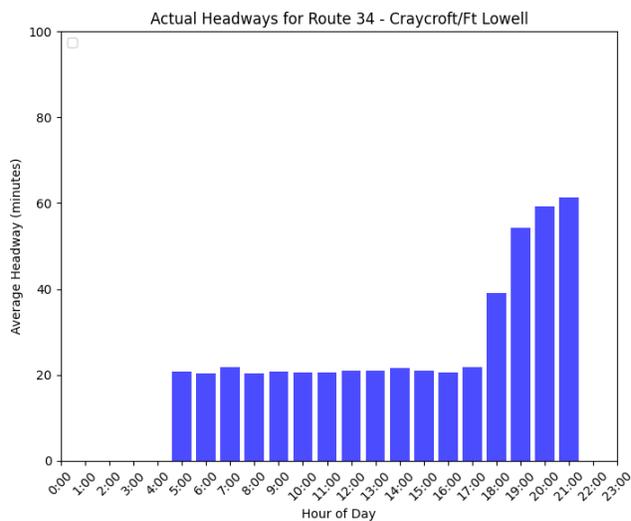
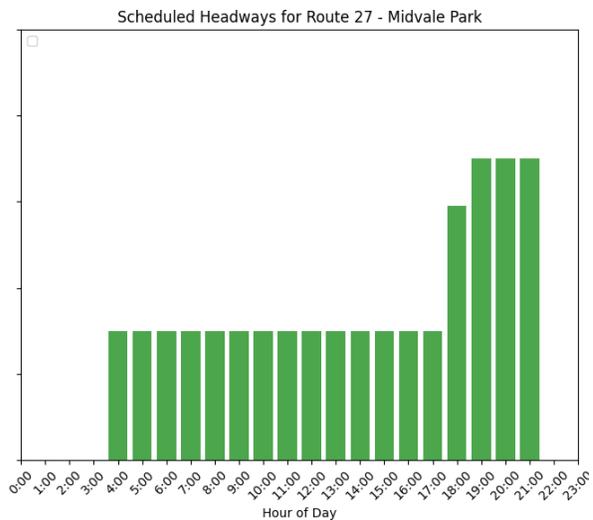
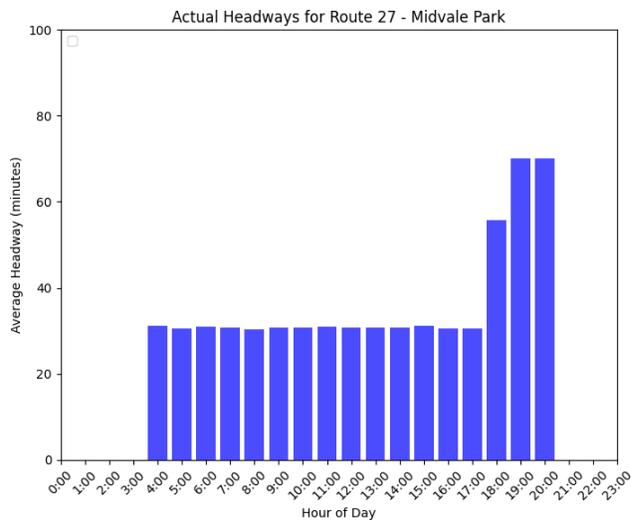
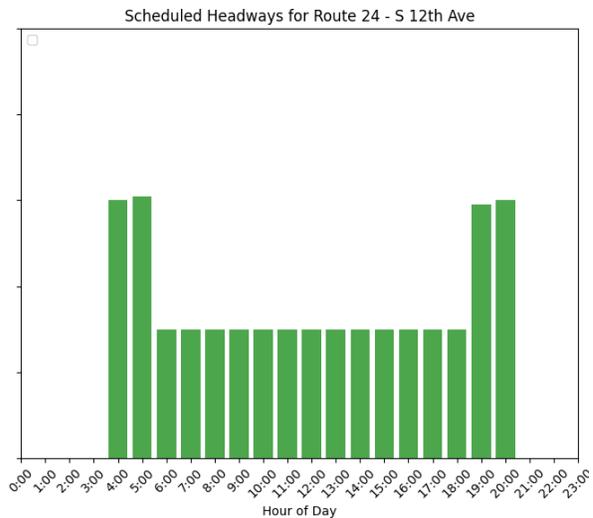
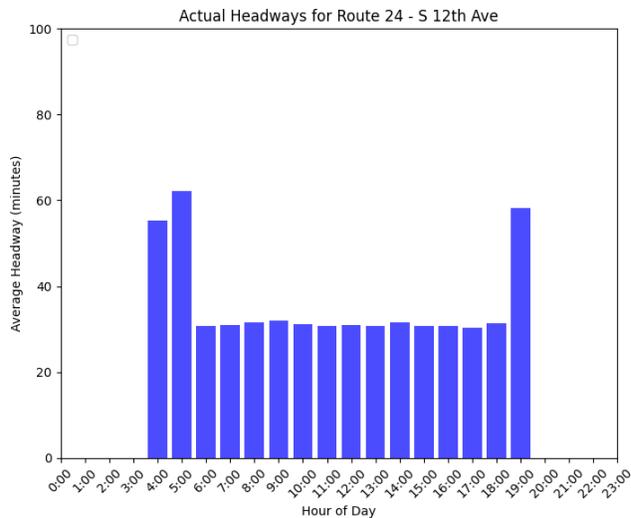
t_{v+1} : next consecutive vehicle arrival time

Headways were calculated using both GTFS-RT and GTFS-ST data. GTFS-RT Trip Update data provided actual departure times for each trip at each stop, enabling calculation of real-time headways. Similarly, scheduled departure times from the GTFS-ST stop_times file allowed for the estimation of scheduled headways based on the operator-defined timetable.

Scheduled and Actual Headways Comparison:

In our comprehensive analysis of the differences between scheduled and actual headways, we examined different transit routes individually to identify operational variances and patterns across the network. For both GTFS-RT and GTFS-ST data, the median headway for each route was calculated hourly by aggregating stop-level headways. As shown in [Figure 3.7](#), the differences between actual and scheduled headways are generally minimal, indicating the Sun Tran system operates consistently aligning with its service plan. Some discrepancies and lags between the actual and scheduled headways, such as the dropped last operating hour for Routes 22 and 24, are due to methodological constraints around calculating headways for the last vehicle on a route.





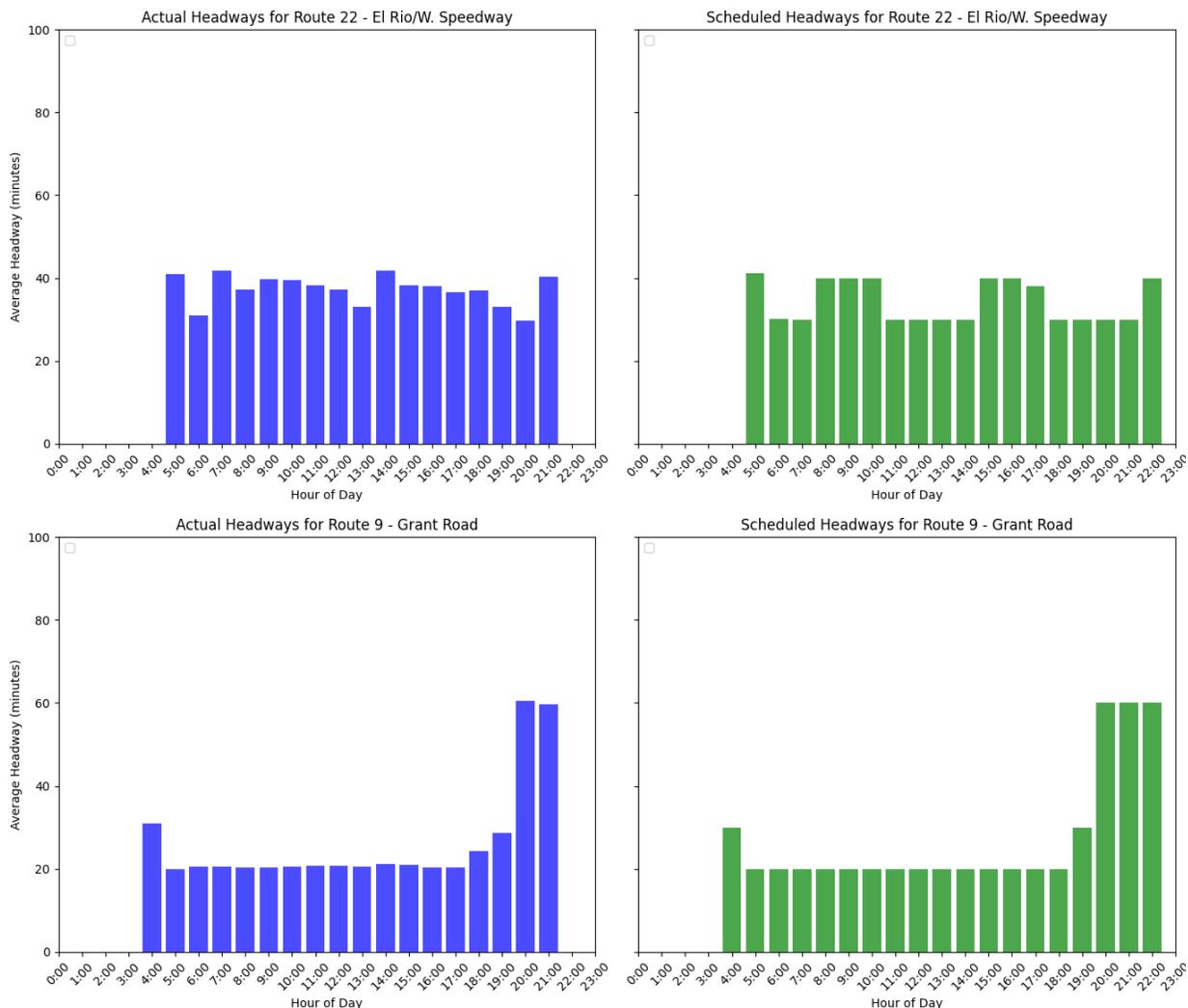


Figure 3.7. Scheduled and Actual Headways Comparisons by Route

“Actual Headway – Scheduled Headway” System Level Analysis

In our system level headway analysis, the focus extends beyond individual transit routes to evaluate the entire network, assessing difference between actual and scheduled headways across all hours of operation. As illustrated in [Figure 3.8](#), the median difference between actual and scheduled headways generally range from -5 to +5 minutes, indicating that actual service closely follows the planned schedule throughout most of the day.

Negative differences, where actual headways are shorter than scheduled, are observed from the early morning through the PM peak period. After the PM peak, the difference shifts to positive values, indicating that actual headways exceed scheduled intervals during late evening hours. This transition from negative to positive values suggests that the system maintains relatively shorter headways during the day and longer headways in the late evening.

Notably, variance in headway differences is more pronounced during certain hours, as shown by the interquartile range (IQR) boxes. Specifically, the hours around 18:00 and 21:00 demonstrate higher variability. The increased variance at 18:00 may be attributed to changes in route headways following the PM peak period, while the larger variance at 21:00 likely reflects the end of transit service, as also observed in the route-level headway patterns shown in [Figure 3.7](#).

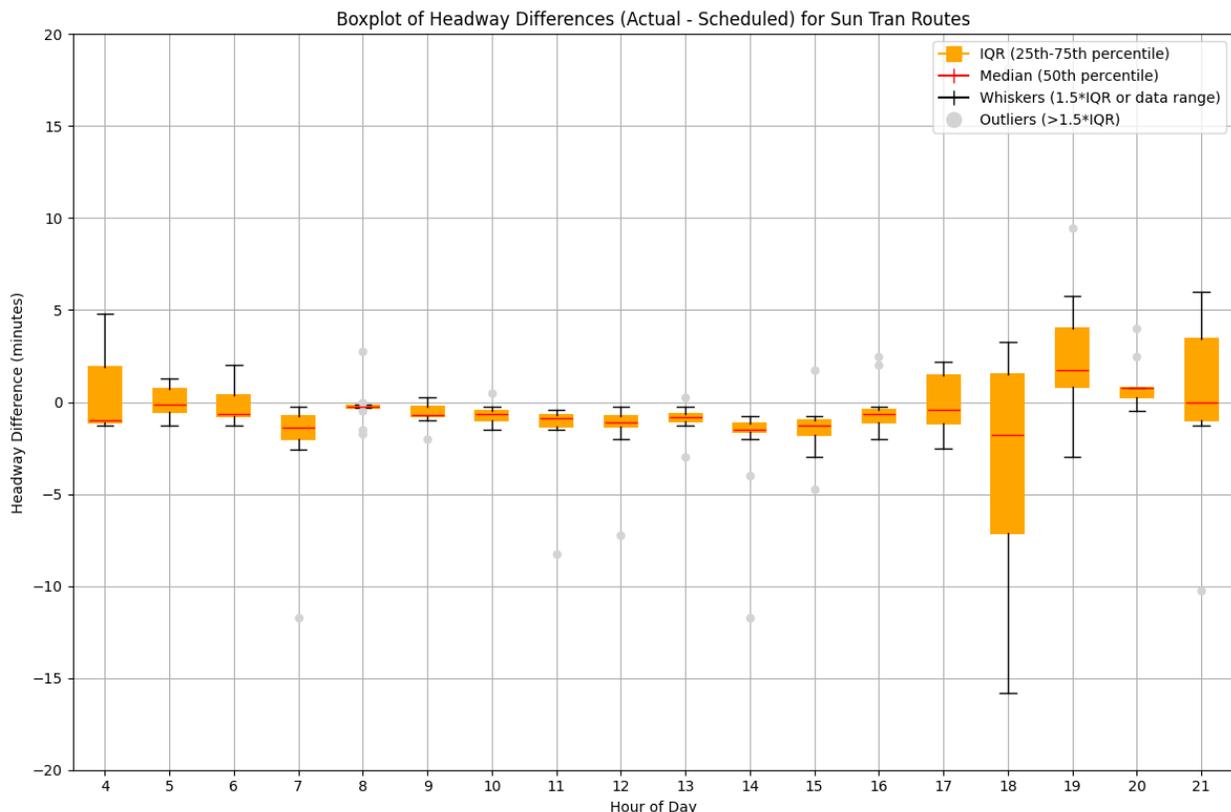


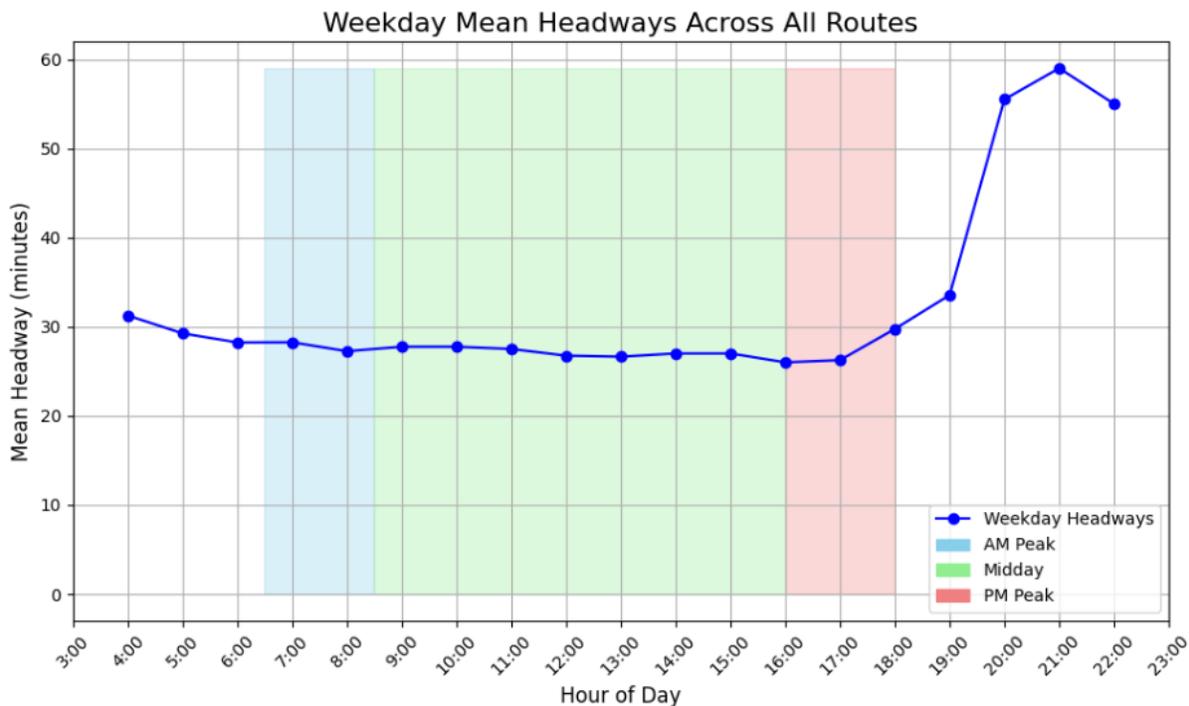
Figure 3.8. “Actual Headway – Scheduled Headway” Boxplot

Headway Weekdays and Weekends Pattern Analysis

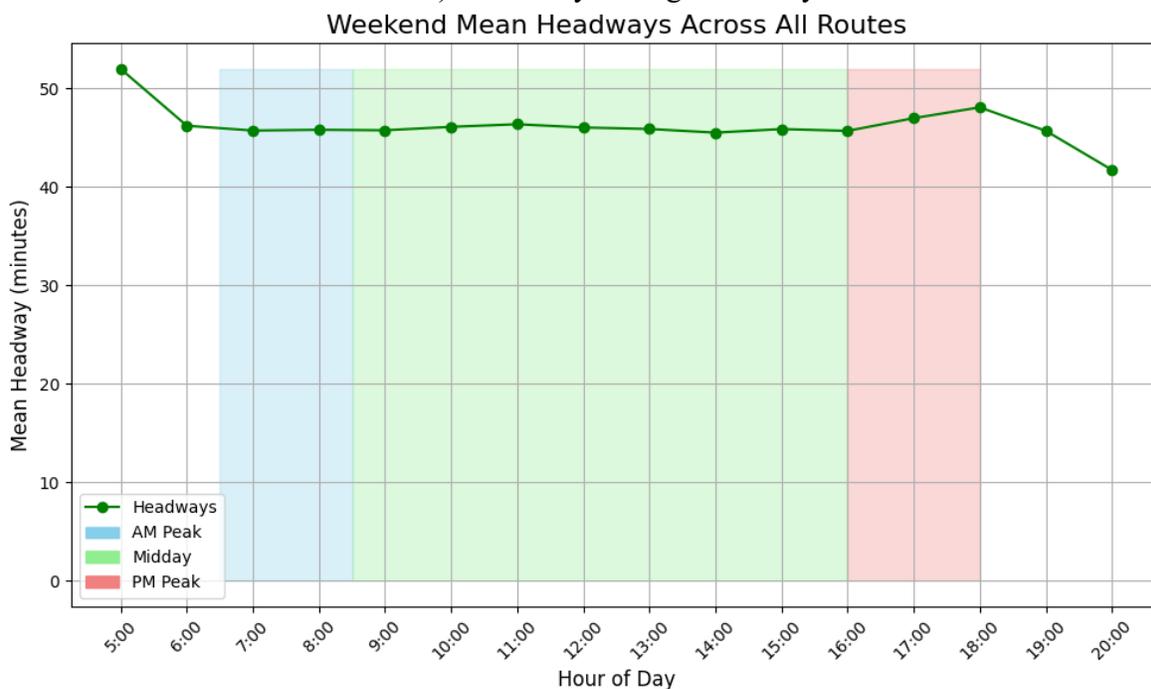
To evaluate differences in transit operation patterns between weekdays and weekends, an average headway analysis was conducted across the entire transit system using GTFS-RT Trip Update data, as shown in [Figure 3.9](#). The results highlight notable variations in the timing a duration of headway across the two periods.

On weekdays, longer headways are observed after PM peak hours, 6:00 PM to 10:00 PM, reflecting lower frequency of the transit service. The highest average headway occurs around 9:00 PM and drops off in the late evening coinciding with service ending for routes operating with extended headways.

Weekend headways remain relatively uniform throughout the day, typically ranging from 45 to 50 minutes. The overall mean headway on weekends is higher than that on weekdays, indicating a generally lower service frequency during weekends.



a) Weekday average headway



b) Weekend average headway

Figure 3.9. Weekdays and Weekends Headway Pattern

CHAPTER 4. Pedestrian Data Collection and Analysis

Walking, as reviewed by Distefano and Leonardi (2023), Gallo and Marinelli (2020), and Allen and Nolmark (2022), is widely recognized in literature as a sustainable and healthy mode of transportation. From an environmental perspective, promoting walking can significantly improve air quality and reduce urban noise. From a public health standpoint, walking can notably increase physical activity levels among less active populations, which is linked to lower rates of obesity and cardiovascular diseases, as well as improved mental health.



4.1. Deliverables

This chapter focuses on documenting various data sources available for deriving high-resolution pedestrian performance indicators in PAG region that support walking. [Table 4.1](#) lists these data sources, summarizing their collection methods, use cases, and current collection status. Subsequent sections provide detailed descriptions and analysis of each pedestrian data source collected.

Table 4.1. Pedestrian data sources in Pima County

Data	Source	Collection Method	Use Cases	Number of Intersections (Year 2023)
MaxView	MaxView event-based data	CATS* SQL** server	Performance measures: <ul style="list-style-type: none"> • <i>Pedestrian volume</i> • <i>Pedestrian Level of Service</i> • <i>Pedestrian Delays</i> 	306
PAG TMC Pedestrian	Yearly manual count by PAG	PAG MS2 Website (pag.ms2soft.com)	Performance measures: <ul style="list-style-type: none"> • <i>Pedestrian volume</i> 	87
Miovision	Video detection	API*** token	Performance measures: <ul style="list-style-type: none"> • <i>Pedestrian volume</i> 	123
Miovision	Miovision event-based data****	CATS* SQL** server	Performance measures: <ul style="list-style-type: none"> • <i>Pedestrian volume</i> • <i>Pedestrian Level of Service</i> • <i>Pedestrian Delays</i> 	105
CATS*: Center for Applied Transportation Sciences SQL**: Structured Query Language API***: Application Programming Interfaces Miovision event-based data****: Miovision event-based data share the same format as MaxView event-based data				

Data from Miovision (video detection and event-based) and MaxView (event-based) are processed separately, with Miovision focusing on pedestrian counts and MaxView providing pedestrian waiting times and push button counts after processing with the event-based data. PAG TMC manually collected pedestrian counts, which share some overlapping locations, spatially and temporally, with MaxView, are processed to obtain pedestrian counts for building a model of pedestrian count estimation from the event-based push button count data. Also, a waiting time estimation model was developed through the count data and the waiting time data. In this case, for intersections with available MaxView data, the waiting time of an intersection is directly calculated using the event-based data, and an estimation model was developed to estimate the pedestrian count using event-based data. The intersection LOS for pedestrian traffic was calculated from the average delay (waiting time). For intersections with Miovision data, the pedestrian count of an intersection is directly from the Miovision API, the pedestrian waiting time was estimated through Miovision event-based data, and the LOS of pedestrian was calculated from the average delay. **Figure 4.1** below shows the methodology flowchart.

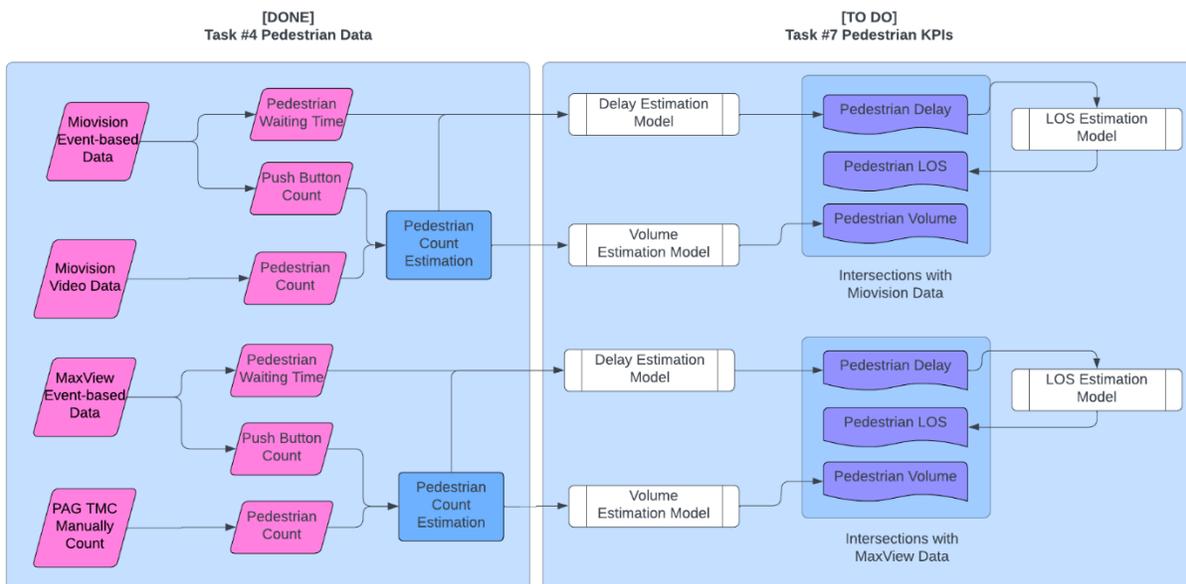


Figure 4.1. Methodology flowchart

In the following section, we demonstrate the event-based data (using MaxView as an example), the video detection data (using MaxView Video Detection as an example), and the PAG TMC pedestrian count data which was manually collected.

This section elaborates on the data sources under consideration to be further explored and used in the project.

4.1.2 MaxView Data

4.1.2.1 Data Elaboration

The most critical measure of evaluating the performance of walking infrastructure in the region is the pedestrian counts. The MaxView system, an advanced traffic management system (ATMS), enables this measurement at the level of signalized intersections and High-intensity Activated crossWalK (HAWK) beacons, via providing push button event records. **Figure 4.2** illustrates, respectively, the locations of push button devices across the region, and in selected intersection and HAWK locations. As can be seen, the push button devices are available across the City of Tucson and the Pima County area. There is a total of 676 sensors, in which 567 are intersections (red dots) and 109 are HAWKs (blue dots).

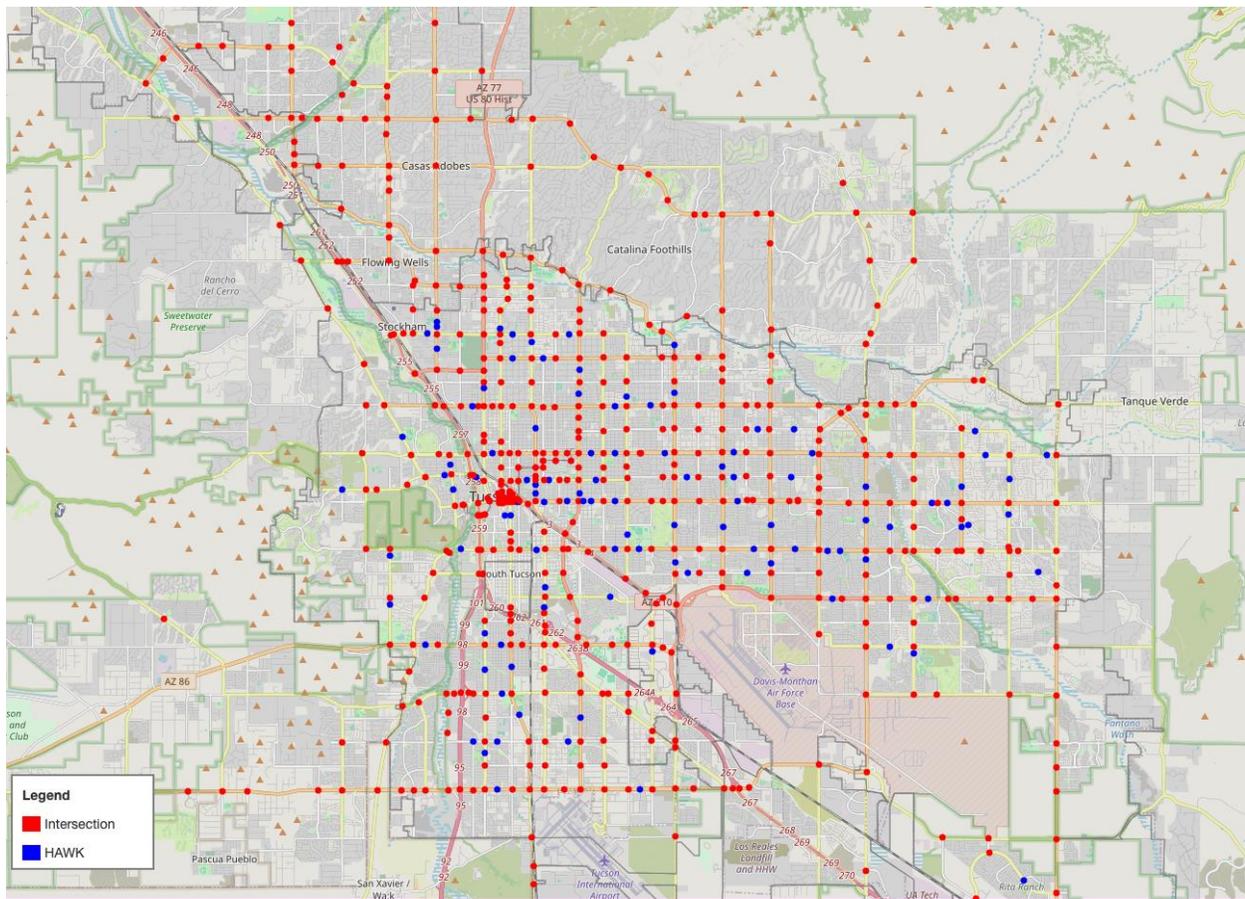


Figure 4.2. Maxview event-based data locations

The MaxView system records the events of the pedestrian's push buttons at the signalized intersections or HAWAKs. [Table 4.2](#) lists the obtainable event-based data from MaxView and each defined event identifiers. The two main events for pedestrians are: Event ID' = 45 when the pedestrian pushes the 'request to walk' button and 'Event ID' = 21 when the pedestrian starts to walk. By counting the number of 'Event ID' = 45, the pedestrian push counts can be extracted. The count of hourly push is the number of events where 'Event ID' = 45. The waiting time is the time gap between the event of '45' and the following event of '21'. Multiple pedestrians could arrive at the intersection. Thus, the maximum waiting time is the waiting time of the pedestrian who arrives first at the intersection in each cycle and the average waiting time is the mean of the waiting time for the pedestrians in each cycle.

Table 4.2. Available pedestrian related MaxView data codes

	Event IDs	Event Descriptor	Description *
Pedestrian Phase Events	21	Pedestrian begin walk	Set when walk indication becomes active.
	22	Pedestrian begin Clearance	Set when flashing don't walk indication becomes active.
	23	Pedestrian begin solid don't walk	Set when don't walk indication becomes solid (non-flashing) from either termination of ped clearance, or head illumination after a ped dark interval.
Detector Events	89	Pedestrian detector off	Ped detector events shall be triggered post any detector delay/extension processing and may be set multiple times for a single pedestrian call. (With future intent to eventually support ped presence and volume)
	90	Pedestrian detector on	N/A
Phase Control Events	45	Pedestrian call registered	Call to service on a phase is registered by pedestrian demand. This event will not be set if a recall exists on the phase.

Notes: * From the code table of "ControllerEventTypes" of MaxView Event-based data archived in CATS

4.1.2.2 Data Analysis

MaxView operates multiple sensors within the city of Tucson area. Currently, the data from these sensors is archived in the CATS SQL database as event-based data. **Table 4.3** provides a summary of the number of sensors that are actively updating data. A sensor is considered available for the month if it has recorded at least one data entry during that period.

Table 4.3. Available MaxView device count every month

[Year]	[Month]	MaxViewUniqueDeviceCount
2022	1	337
2022	2	332
2022	3	338
2022	4	329
2022	5	330
2022	6	325
2022	7	319
2022	8	324
2022	9	323
2022	10	325
2022	11	325
2022	12	324
2023	1	306
2023	2	305
2023	3	303
2023	4	309
2023	5	305
2023	6	301
2023	7	306
2023	8	301
2023	9	294
2023	10	297
2023	11	293
2023	12	291
2024	1	286
2024	2	299
2024	3	290

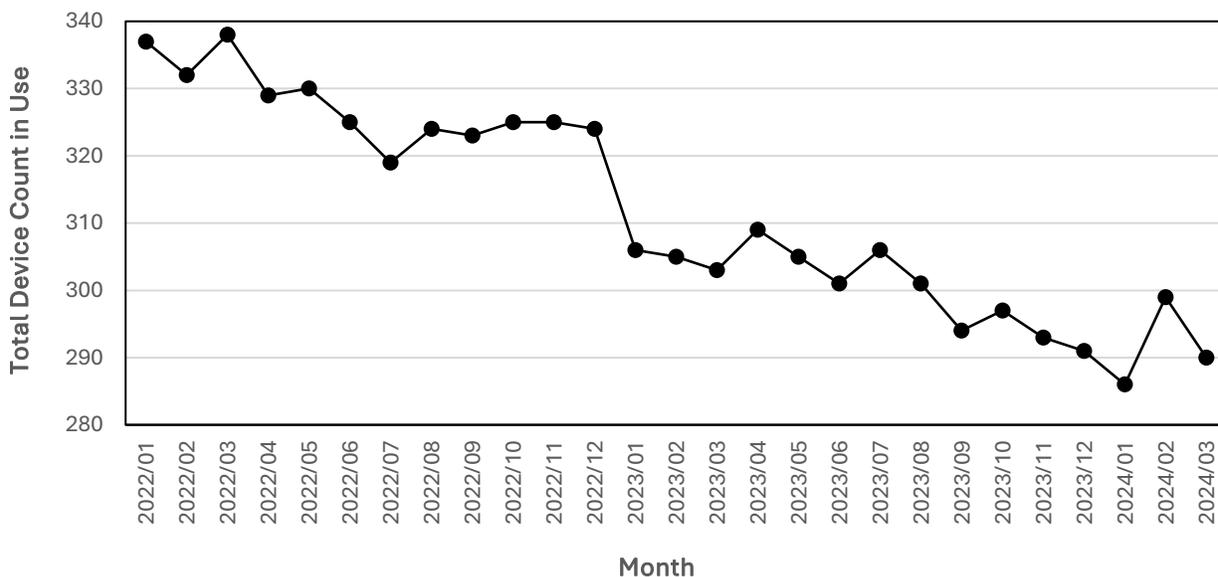


Figure 4.3. Available MaxView device count every month

To understand the daily trends of pedestrian activity at intersection levels, the pedestrian push button count can be used either directly (Humagain and Singleton, 2021), or after being fed to a Bayesian Additive Regression Trees (BART) model that is trained to derive pedestrian volumes more accurately (Li and Wu, 2021). Humagain and Singleton (2021) used pedestrian push-button data from Utah traffic signals to understand the pedestrian activity patterns. Li and Wu (2021) presented a method to estimate pedestrian volume using button-pushing and signal timing events data from MaxView system by modeling pedestrian arrival as a Poisson process. Below we just look at the data of the push button count, not the actual volume count of pedestrians. To analyze the data, we define the time periods as AM_PEAK (06:30:00–08:30:00), MID_DAY (08:30:00–16:00:00), and PM_PEAK (16:00:00–18:00:00). The City of Tucson has swapped the MaxView AMTS with a new Kinetic Signals ATMS by Spring 2024.

4.1.2.2.2 Push count Analysis

Data from January 1 to January 31, 2023.

We began the analysis by selecting January 2023, focusing on the wintertime. **Figure 4.4** displays the hourly average push button counts for both intersections and HAWK crossings, distinguishing between weekdays (i.e., Mondays, Tuesdays, Wednesdays, Thursdays, and Fridays) and weekends (i.e., Saturdays and Sundays).

The graphs below show the average push button count by day type and device type in January 2023.

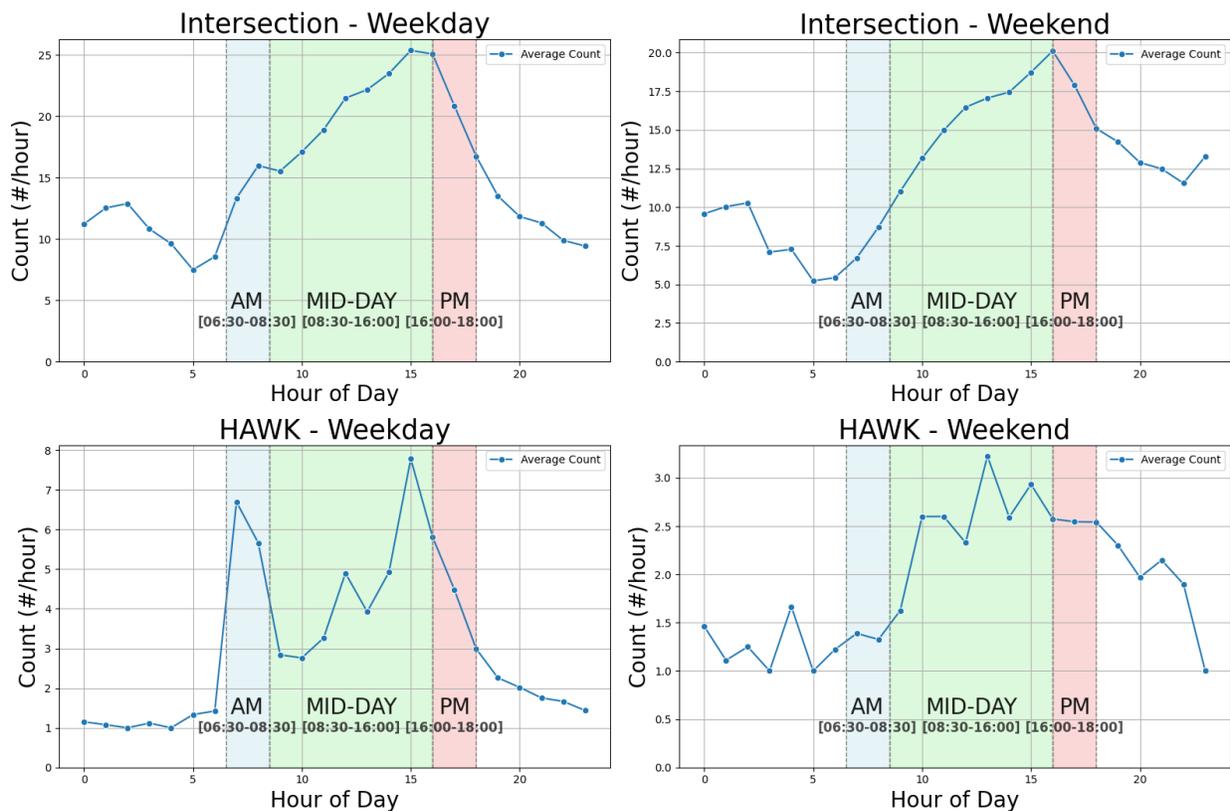
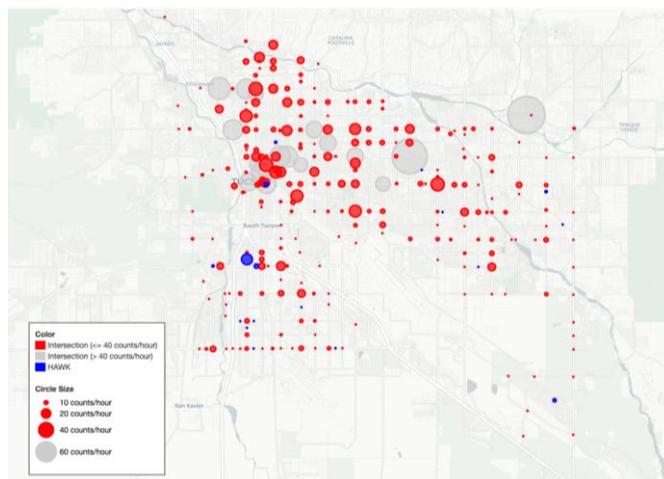


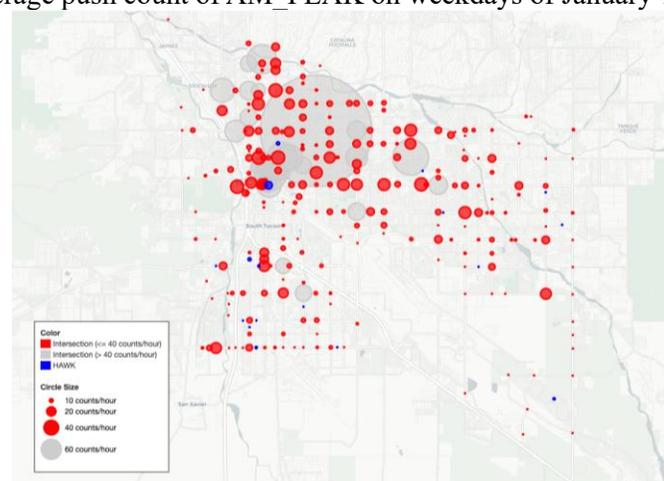
Figure 4.4. Hourly average push button counts by day type and device type from January 1 to January 31, 2023 (HAWK devices 362 and 307 excluded)

On weekdays, push counts at intersections increase steadily throughout the morning, peaking around 3 PM during the Mid-Day period, reaching up to 25 per hour before tapering off during the PM peak (4-6 PM). Weekend push counts follow a similar pattern but peak later in the day at up to 20 per hour. HAWK crossings show a different trend, with high push counts during the AM peak on weekdays, followed by fluctuations throughout the day and a peak of 8 counts per hour in the Mid-Day period. On weekends, HAWK push counts are lower across the day, peaking around noon with around 3 counts per hour.

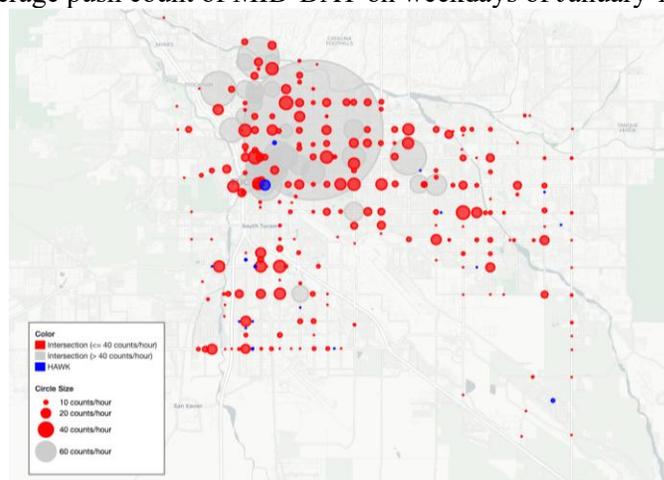
To take a closer look at the details, the following figures demonstrate the push counts of different intersections of AM peak, Mid-day, and PM peak.



(a) Hourly average push count of AM_PEAk on weekdays of January 1 – January 30, 2023



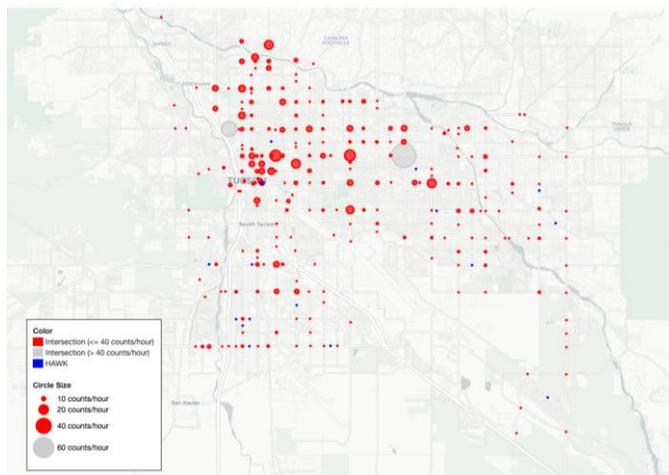
(b) Hourly average push count of MID-DAY on weekdays of January 1 – January 30, 2023



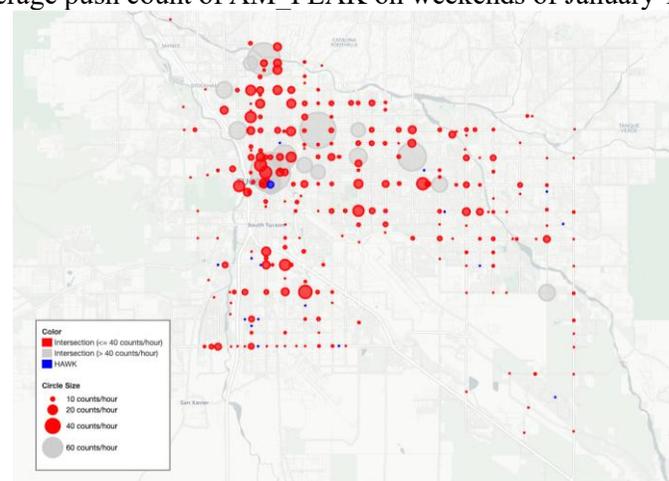
(c) Hourly average push count of PM_PEAk on weekdays of January 1 – January 30, 2023

Figure 4.5. Hourly average push counts in January (weekdays)

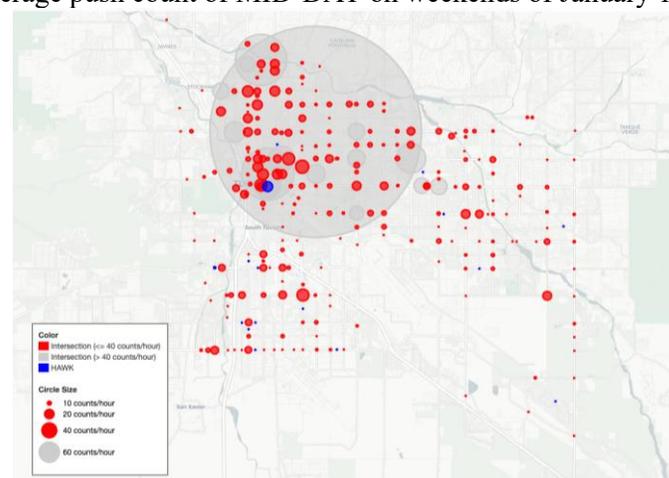
Figure 4.5 depicts the push button counts during AM Peak, Mid-day, and PM Peak periods in Tucson on weekdays. Across all time periods, intersections (in red) dominate pedestrian activity compared to HAWK signals (in blue), particularly in central Tucson. The light grey shaded circles are the intersections with hourly count larger than 40 per hour. Locations around downtown Tucson, the University of Arizona, and 4th Ave demonstrate consistently high activation counts, and the intersections of major arterials, which tend to have shopping plazas or other major attractors, show increased activity in the midday and PM peak periods.



(d) Hourly average push count of AM_PEAk on weekends of January 1 – January 30, 2023



(e) Hourly average push count of MID-DAY on weekends of January 1 – January 30, 2023



(f) Hourly average push count of PM_PEAk on weekends of January 1 – January 30, 2023

Figure 4.6. Hourly average push counts in January (weekends)

Figure 4.6 depicts the average push counts during different time periods (AM Peak, Mid-Day, and PM Peak) on weekends. The AM Peak figure shows smaller and more dispersed push button activity across Tucson, with most intersections experiencing relatively low counts (10–20 counts/hour). Observed push button counts are heaviest during the afternoon and evening, particularly around downtown Tucson, 4th Ave, and the area around the Tucson Mall.

On both weekdays and weekends, the extremely large marker at the intersection of Grant Road & Tucson Boulevard is due to extremely high push button counts reported by Device 214, shown in **Figure 4.7**. While the majority of data points remain within normal ranges, there are notable outliers, with the highest points exceeding 5,000 counts per hour. These outliers may indicate irregularities, such as technical glitches or events that caused anomalous data spikes.

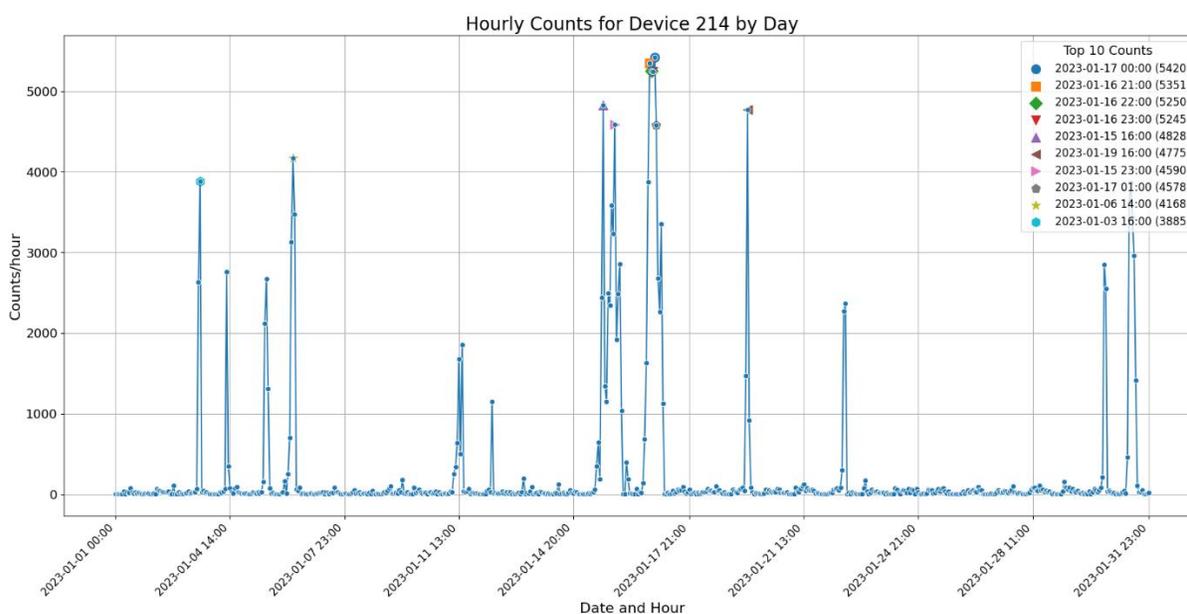


Figure 4.7. Hourly pedestrian counts for Device 214 in January 2023

To address these outliers, one approach is to set thresholds for identifying abnormal data points. This can be done by examining historical averages to determine typical pedestrian counts for specific times of day or locations, particularly during peak hours. The standard deviation method can also be applied, where any count falling beyond three standard deviations from the mean is flagged as an outlier, capturing the natural variability in the data. Additionally, using a percentile threshold (e.g., the 95th or 99th percentile) allows us to flag values that are unusually high in comparison to the majority of the data, marking them for further review. These approaches will help refine the process of identifying and addressing outliers in pedestrian count data more effectively.

Data of July 1 to July 31, 2023

Figure 4.9 and **Figure 4.10** display the hourly average push button counts for both intersections and HAWK crossings, distinguishing between weekdays (Mondays, Tuesdays, Wednesdays, Thursdays, and Fridays) and weekends (Saturdays and Sundays).

The average push button count at HAWK intersections on weekends is unusually high around midnight, because Device 362, located at Broadway Blvd & 4th Ave, has unusually high counts. The duration between pushes looks fine (several seconds or minutes) most of the time, but there are a few occasions when multiple counts are registered in 1 second.

Figure 4.8 shows the hourly push counts at HAWK Device 362 in July 2023.

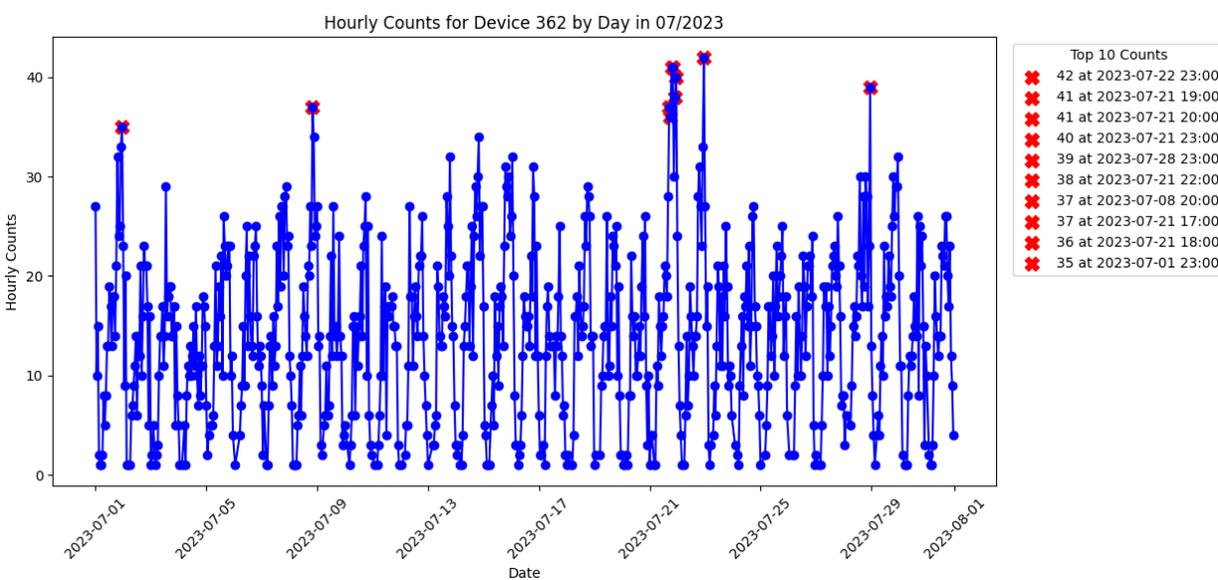


Figure 4.8. Hourly Counts for Device 362 in July 2023

The graphs below show the average push button count by day type and device type in July 2023, after removing HAWK Device 362.

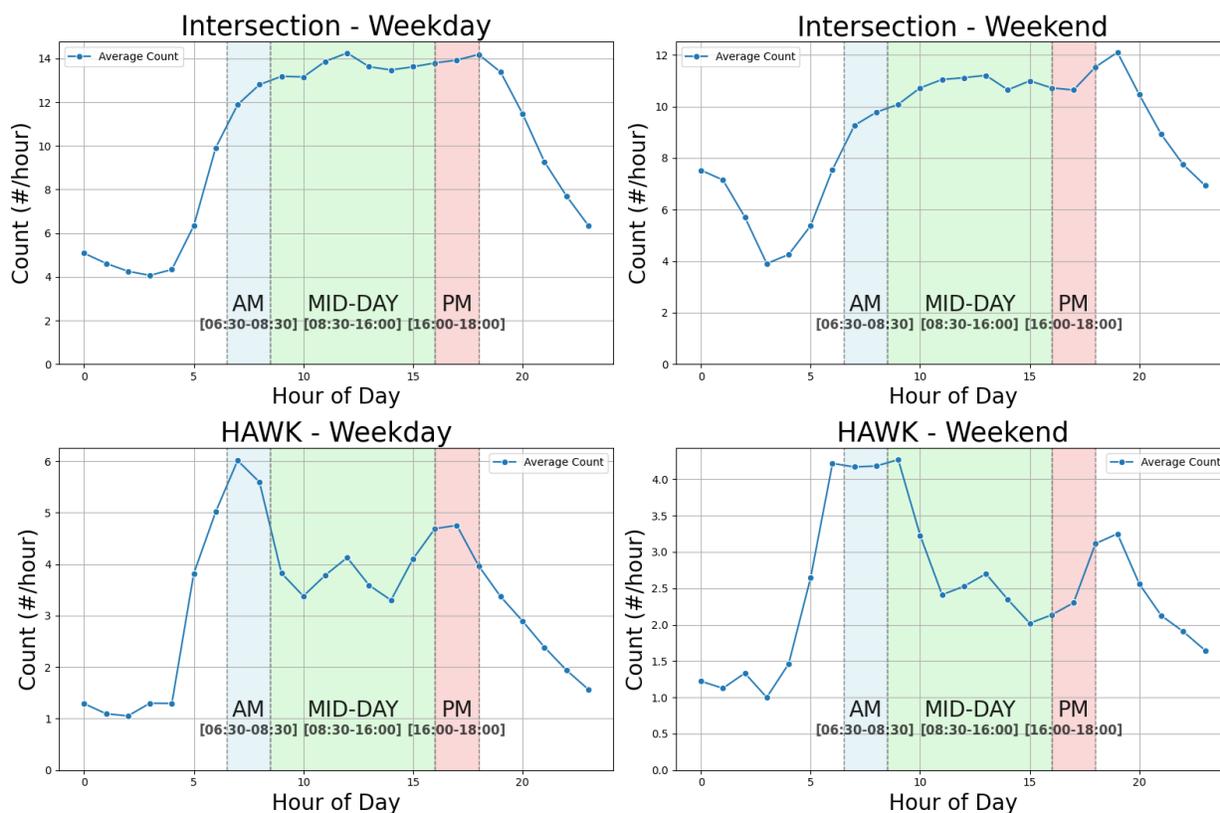
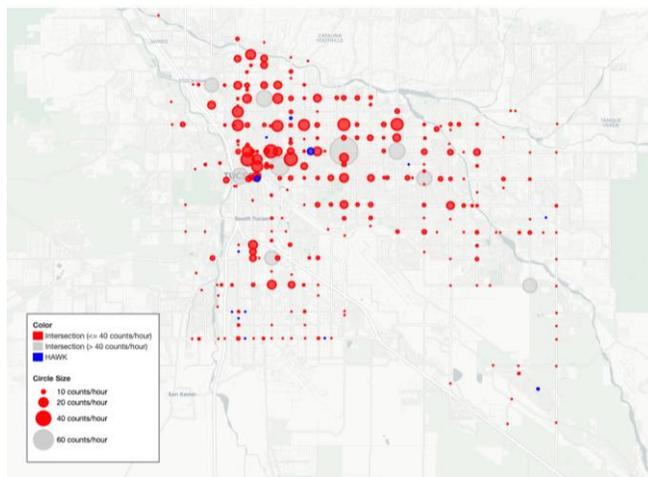


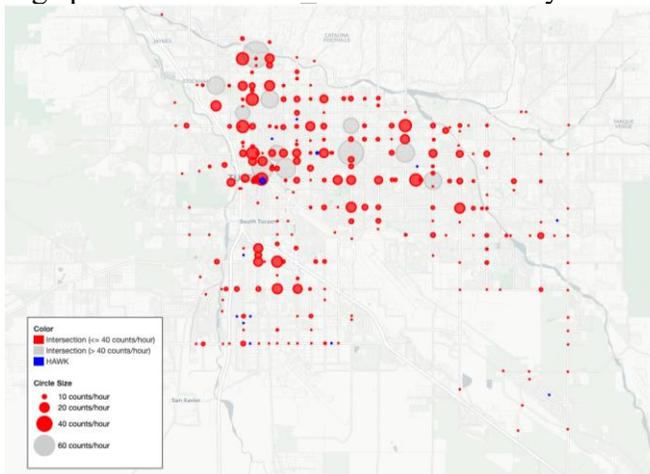
Figure 4.9. Hourly average push button counts by day type and device type from July 1 to July 31, 2023 (excluding HAWK device 362)

This figure illustrates the hourly average push counts for intersection and HAWK devices during weekdays and weekends, across the AM, Mid-Day, and PM time periods in July 2023. On weekdays, intersection push button counts show a typical triple peak pattern, with counts rising during the AM peak period, a midday peak during the noon lunch period, and a post-work peak in the evening. A similar pattern occurs at intersections on weekends, but with slightly lower average counts overall. On both weekdays and weekends, HAWKS show a unique pattern, with a daily high in the AM period, and lower period peaks around the noon lunch period, and in the early afternoon on weekdays and early evening on weekends.

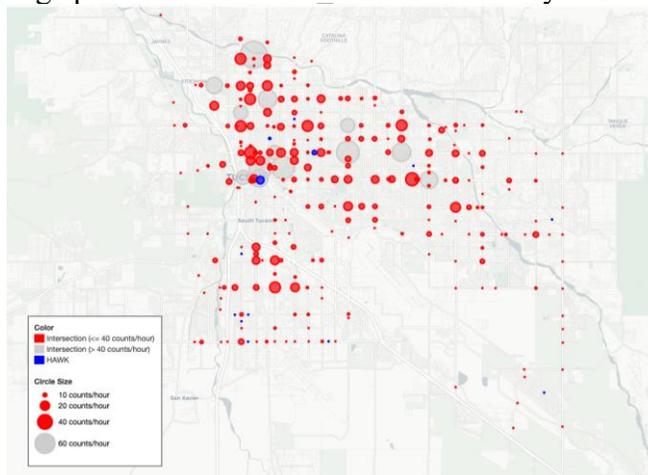
Comparing the figures from January and July 2023, push counts are lower for all days and device types except for weekend HAWKS, which have an AM peak just above 4 activations per hour, as compared to the midday peak of just above 3 activations per hour in January.



(a) Hourly average push count of AM PEAK on weekdays of July 1 to July 31, 2023



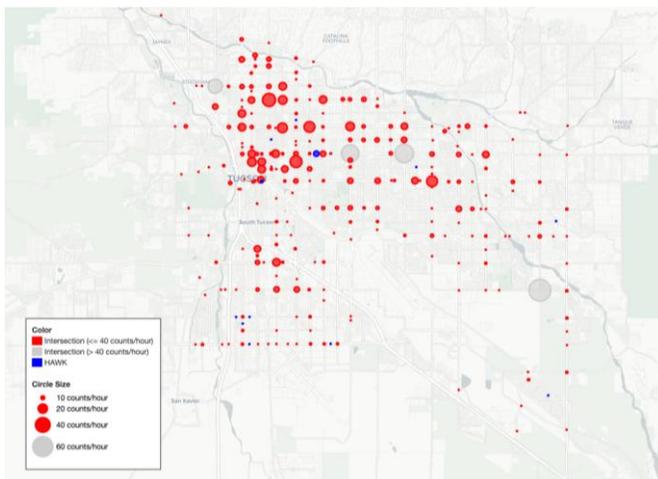
(b) Hourly average push count of MID DAY on weekdays of July 1 to July 31, 2023



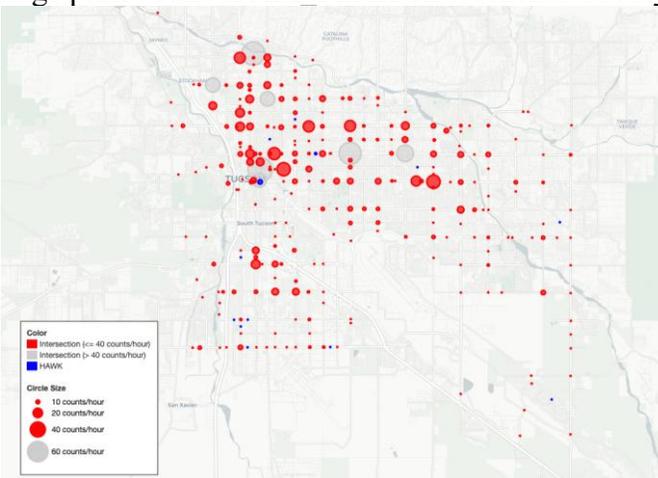
(c) Hourly average push count of PM-PEAK on weekdays of July 1 to July 31, 2023

Figure 4.10. Hourly average push counts in July (weekdays)

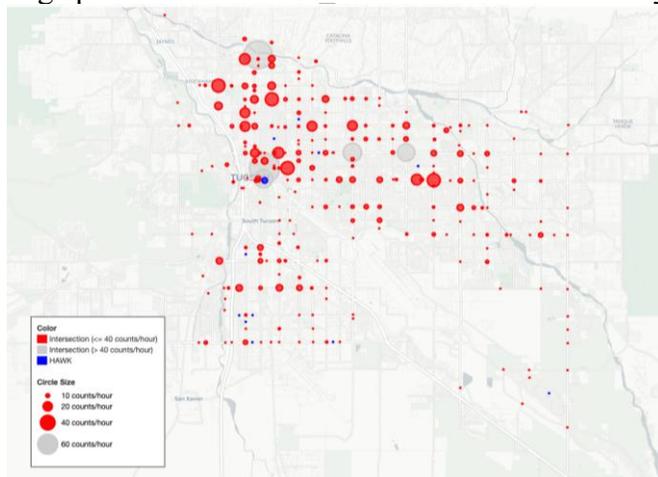
Figure 4.10 shows the average hourly push counts at intersections and HAWK signals by period for weekdays in July 2023. While push button counts are lower overall than in the winter evaluation period, many of the same high-count intersections identified earlier remain apparent, showing consistent year-round activity at key attractors and areas, or the prevalence of non-discretionary trips.



(a) Hourly average push count of AM PEAK on weekends of July 1 to July 31, 2023



(b) Hourly average push count of MID DAY on weekends of July 1 to July 31, 2023



(c) Hourly average push count of PM PEAK on weekends of July 1 to July 31, 2023

Figure 4.11. Hourly average push counts in July (weekends)

Figure 4.11 shows the average hourly push counts at intersections and HAWK signals by period for weekends in July 2023. As with the weekday counts, there is a significant decrease in recorded activity, especially in the PM Peak period, but high counts remain around the University of Arizona and along the Oracle Rd and Stone Ave corridors,

4.1.2.2.3 Pedestrian waiting time

Next, we analyze the waiting time of pedestrians of the same time in January 2023 and July 2023.

$$AWT_p = \frac{\sum_{i=1}^{N_p} WT_i}{N_p}$$

$$WT_i = T_{i,1} - T_{i,0}$$

In which:

WT_i : Waiting time of each pushing event i

$T_{i,0}$: Time of pushing event happens for i^{th} event_code = 45

$T_{i,1}$: Time of next light turns green (event code = 21) after i^{th} event_code = 45

N_p : Number of pushing events in the time period p (e.g. one hour)

AWT_p : Average waiting time of pedestrians in this time period p

Data of January 1 to January 31, 2023

Figure 4.12 illustrates hourly maximum and average pedestrian waiting times at intersections and HAWK signals on weekdays and weekends of January 2023. The maximum waiting time refers to the time the first pedestrian arriving at the intersection waits during each cycle, while the average waiting time represents the mean waiting time for all pedestrians within the same cycle.

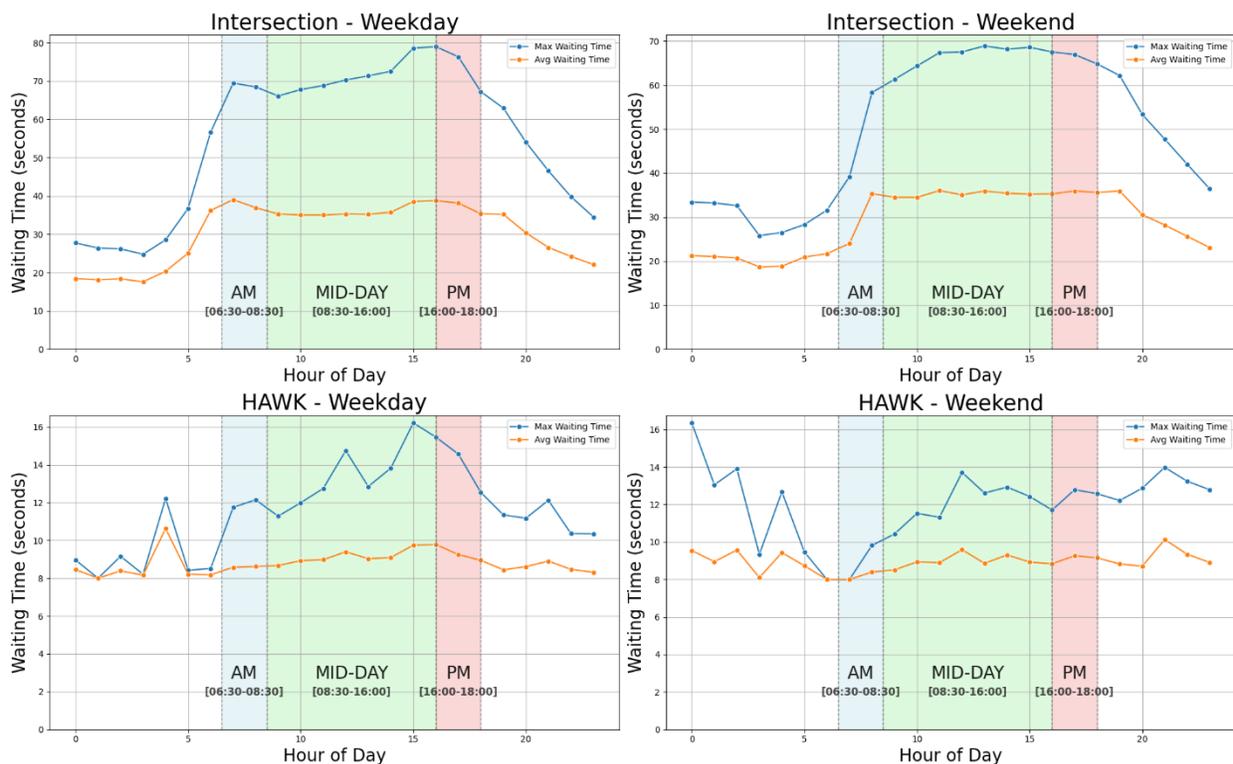
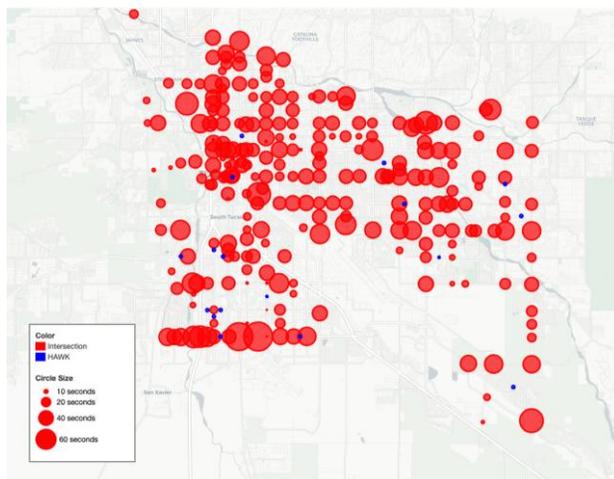


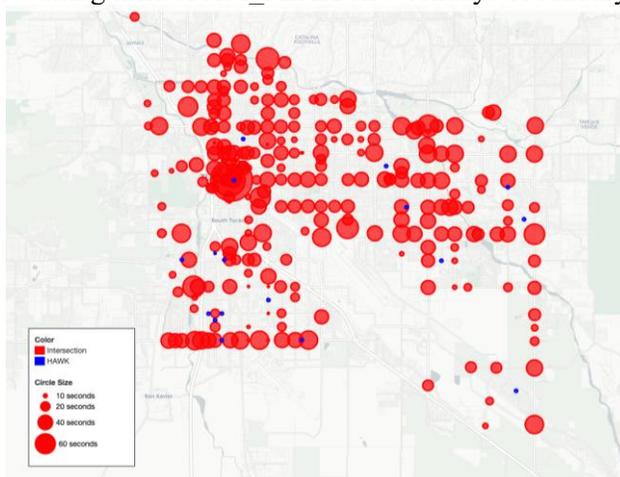
Figure 4.12. Hourly average waiting time by day type and device type of January 2023

The data reveals that intersections have considerably higher waiting times compared to HAWK signals, with weekday intersections experiencing an average waiting time of 40 seconds during peak midday hours and maximum times peaking over 80 seconds. Weekends show similar trends at intersections, with slightly lower waiting times during AM and PM peaks. HAWK signals, on the other hand, maintain much lower waiting times, averaging around 12 seconds, with slight variations throughout the day, and the maximum waiting time rarely exceeds 17 seconds. These findings are consistent with the intended functions of the two types of devices, with HAWK signals prioritizing serving pedestrian crossings as soon as possible, while traffic intersections may need to progress through an entire cycle serving vehicle movements before serving a pedestrian movement.

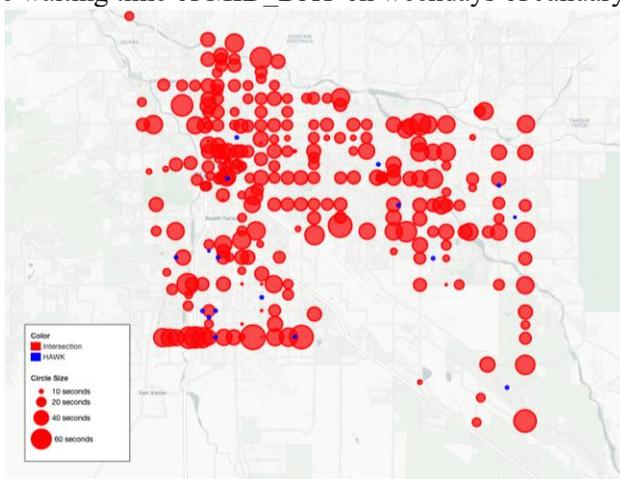
To take a closer look at the details, [Figure 4.13](#) demonstrates the average waiting time of different intersections during the AM peak, Mid-day, and PM peak.



(a) Hourly average waiting time of AM_PEAK on weekdays of January 1–January 31, 2023



(b) Hourly average waiting time of MID_DAY on weekdays of January 1–January 31, 2023

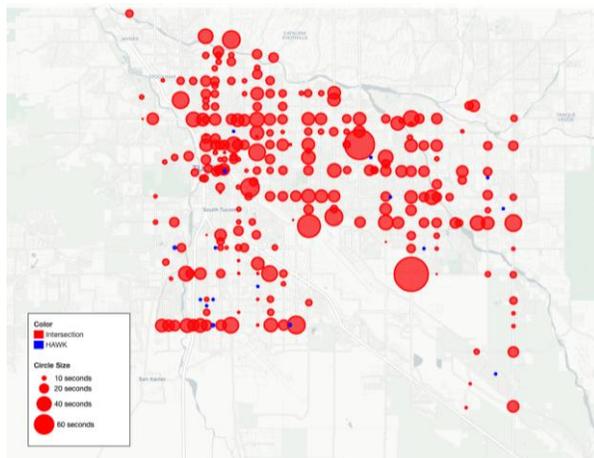


(c) Hourly average waiting time of PM_PEAK on weekdays of January 1–January 31, 2023

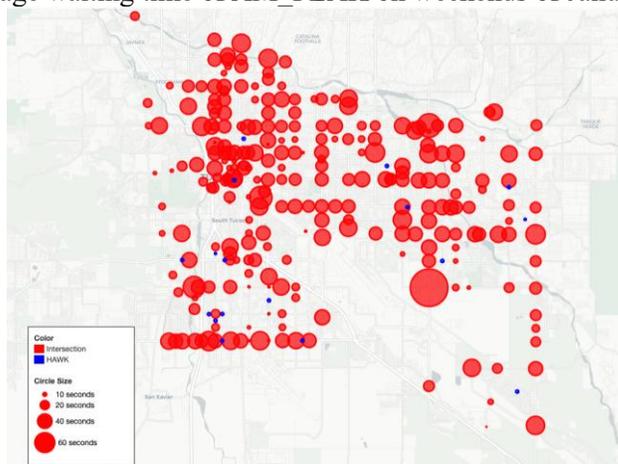
Figure 4.13. Hourly average waiting time in January (weekdays)

Figure 4.14 depicts the average waiting times at intersections and HAWK crossings during the AM Peak, Mid-Day, and PM Peak periods on weekdays in January 2023. The intersections of principal arterials stand out in all three time periods, due to their longer cycle times, intersections along principal arterials have moderate waiting times, while HAWK signals show consistently low waiting times.

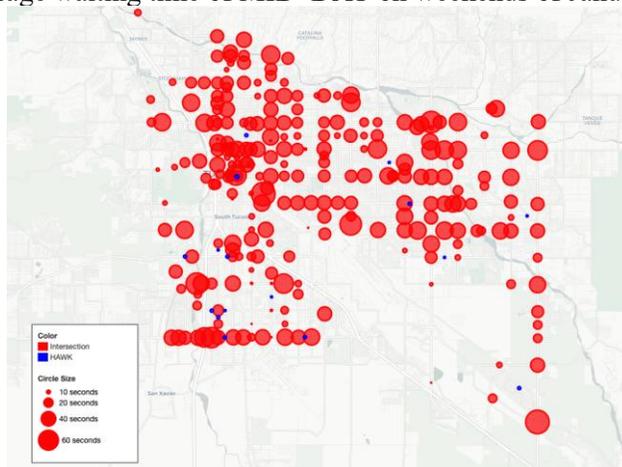
Across all periods, intersections dominate the map, with larger red circles indicating longer waiting times, particularly in the city center and northern areas. The average waiting times are shorter during the Mid-day period, with smaller circles overall, while the AM Peak and PM Peak periods exhibit longer waiting times with larger circles, especially in central intersections. HAWK crossings generally show shorter waiting times compared to intersections, represented by smaller blue circles, with some variation across different periods. Overall, waiting times increase in more urbanized areas, with noticeable consistency in trends across different times of day.



(a) Hourly average waiting time of AM_PEAk on weekends of January 1–January 31, 2023



(b) Hourly average waiting time of MID_DAY on weekends of January 1–January 31, 2023



(c) Hourly average waiting time of PM_PEAk on weekends of January 1–January 31, 2023

Figure 4.14. Hourly average waiting time in January (weekends)

Data of July 1 to July 31, 2023

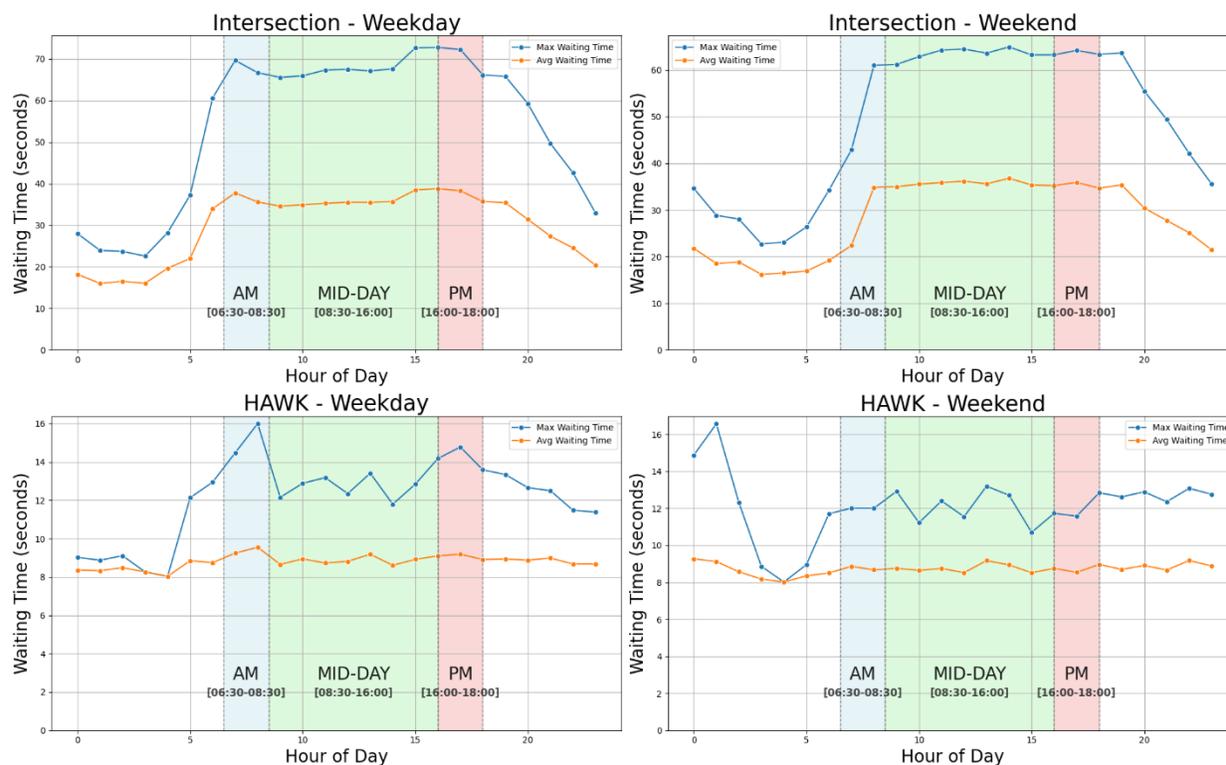
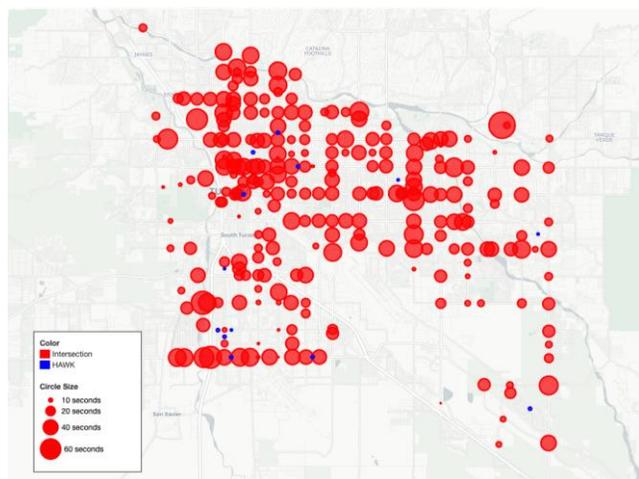
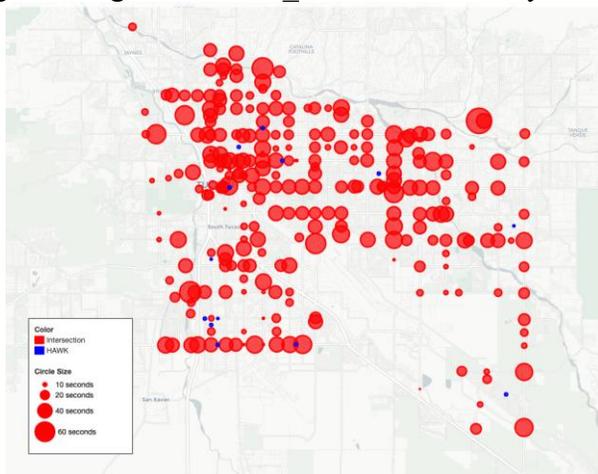


Figure 4.15. Hourly average waiting time by day and device type in, July 1 – July 31, 2023

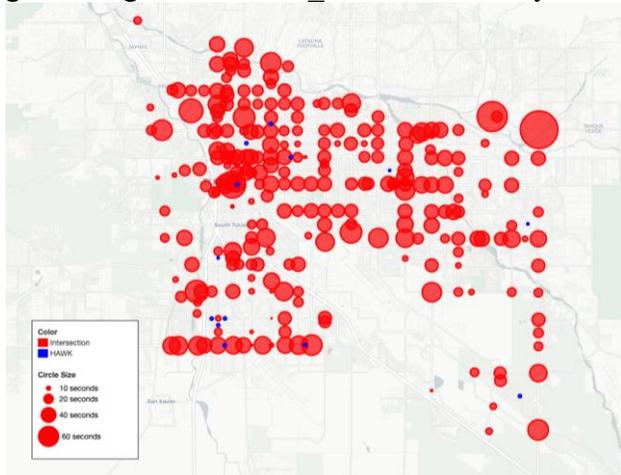
Figure 4.15 presents hourly waiting times for pedestrian crossings, comparing weekdays and weekends across two device types: intersections and HAWKs. The graphs show both average and maximum waiting times in seconds. On weekdays, waiting times peak during the midday and PM periods for both intersections and HAWKs. During weekends, waiting times tend to be lower overall, with a similar pattern of midday peaks. As in the January evaluation period, HAWK signals have lower average and maximum waiting times, due to the differences in operational characteristics of the two signal types.



(a) Hourly average waiting time of AM_PEAK on weekdays of July 1 – July 31, 2023

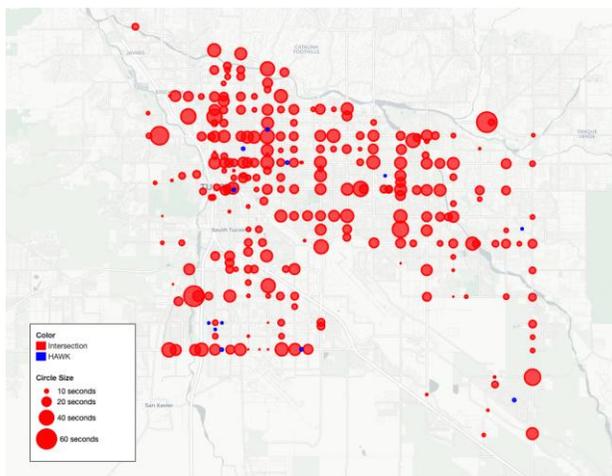


(b) Hourly average waiting time of MID_DAY on weekdays of July 1 – July 31, 2023

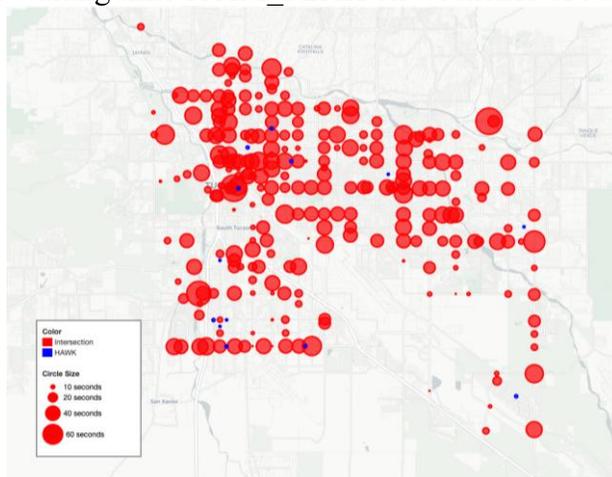


(c) Hourly average waiting time of PM_PEAK on weekdays of July 1 – July 31, 2023

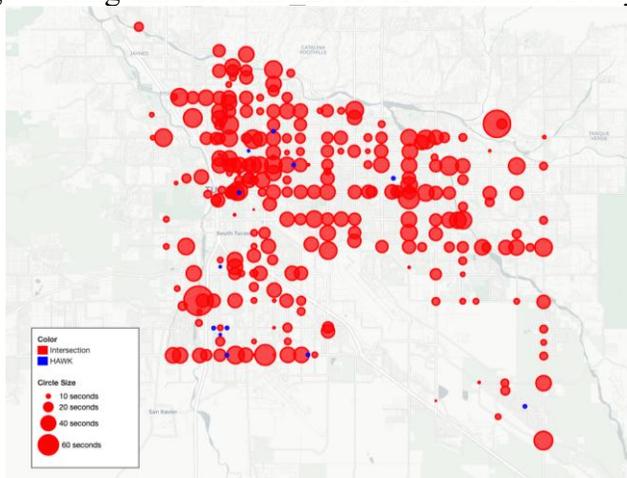
Figure 4.16. Hourly average waiting time in July (weekdays)



(a) Hourly average waiting time of AM_PEAK on weekends of July 1 – July 31, 2023



(b) Hourly average waiting time of MID_DAY on weekends of July 1 – July 31, 2023



(c) Hourly average waiting time of PM_PEAK on weekends of July 1 – July 31, 2023

Figure 4.17. Hourly average waiting time in July (weekends)

Figure 4.16 and **Figure 4.17** visualize the average waiting times for pedestrian crossings during different times of day (AM peak, mid-day, PM peak) on weekdays and weekends. The trends are largely similar to those of the January evaluation period, with the largest waiting times at the intersections of principal arterials, moderate waiting times along principal arterials, and the lower waiting times at HAWK signals.

4.1.3 *Miovision data*

4.1.3.1 Data Elaboration

4.1.3.1.1 Miovision sensor locations

Miovision sensors can provide several traffic parameters, including turning movement count data in one-minute intervals. 126 intersections in the PAG region have been configured with Miovision sensors as of March 2024. The U of Arizona team used computer code developed earlier with Pima County Department of Transportation (PCDOT) that uses the API keys (provided by PCDOT and the Town of Marana). The code collects data via the Miovision API. Most intersections use a single Miovision camera, but there are large intersections where two cameras are used to accurately collect traffic data, located at La Cholla Boulevard & Ruthrauff Road, La Cholla Boulevard & River Road, Alvernon Way & Brandi Fenton Driveway & River Road, and La Cholla Boulevard & Orange Grove Road in Pima County, and Cortaro Road & Arizona Pavilions, Cortaro Road & Silverbell Road, and Thornydale Road & Orange Grove Road in the Town of Marana. **Table 4.4** summarizes the number of Miovision intersections.

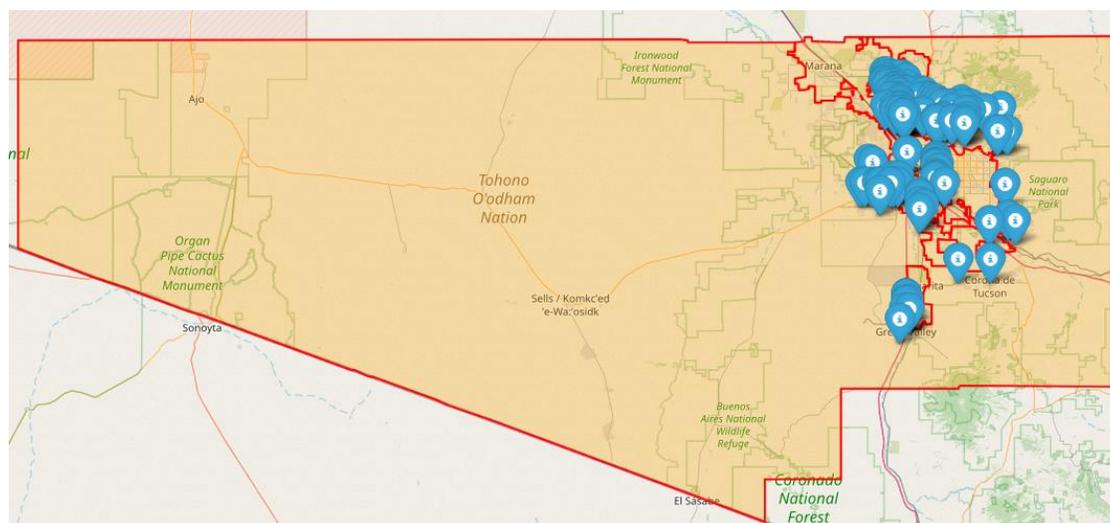


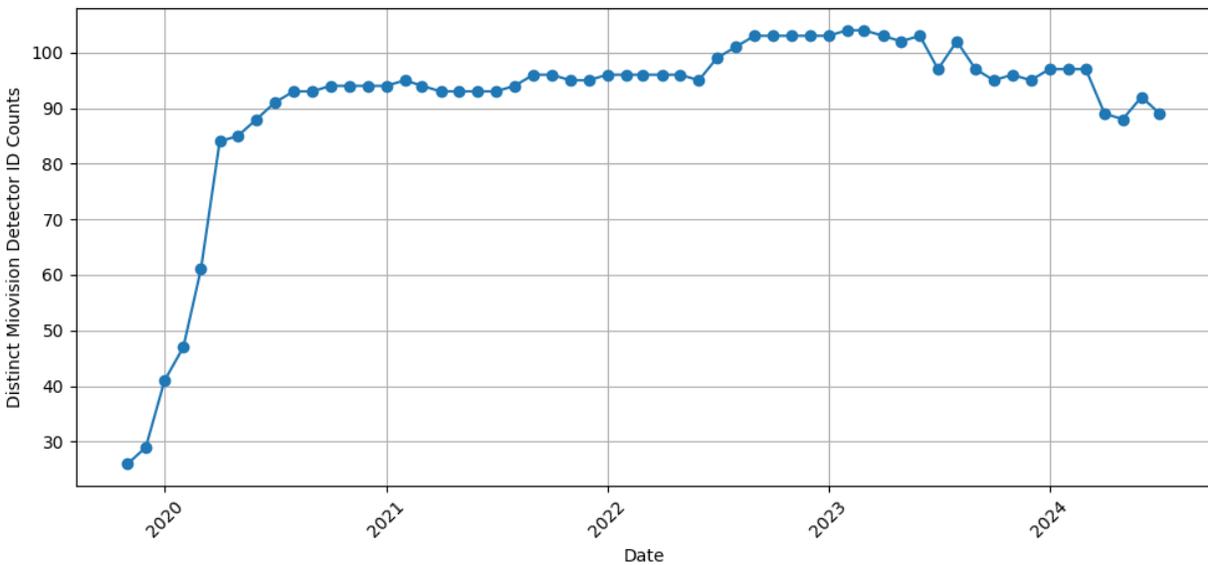
Figure 4.18. Miovision sensor (archived in CATS server) location distribution

Table 4.4. Miovision sensor (archived in CATS server) location distribution

NAME	COUNT
MARANA	0
UNINCORPORATED PIMA COUNTY	123
SAHUARITA	0
ORO VALLEY	0
TUCSON	0
SOUTH TUCSON	0

4.1.3.1.2 *Miovision sensor count*

The **Figure 4.19** below displays the monthly trend of active Miovision detector ID counts from 2020 through 2024. In 2020, the number of detectors increased rapidly, starting from around 30 and reaching nearly 90 active detectors within a few months. After this initial growth phase, the number of detectors stabilizes at around 90-100 between late 2020 and mid-2023, indicating consistent detector deployment and usage. However, towards the end of 2023 and into 2024, the count starts to slightly decline, dipping below 90 by mid-2024, suggesting a possible reduction in the number of active or distinct detectors during this period.



4.1.3.2 Data Analysis

4.1.3.2.1 Miovision volume data

We picked the data from January 2021 to June 2024 to do the analysis because the sensor counts are steadier. **Figure 4.20** illustrates the monthly average push-button counts across all intersections from 2021 to mid-2024, revealing clear seasonal patterns. Push-button activity tends to peak during the cooler months, particularly in early 2022 and 2023, which had peaks at 1000 push-button counts per day. However, there are significant drops mid-year, especially around the summer months, where counts dip as low as 500. This pattern repeats each year, suggesting that external factors such as weather conditions or seasonal behavior heavily influence pedestrian movement, with reduced activity during warmer months and higher activity in cooler periods. As can be seen, the overall pattern also indicates a decreasing trend after 2022. This trend may or may not have been due to decreasing number of pedestrians. Other possibilities, which could be very likely, are reduced sensor accuracy and reduced efficiency of the controller communications causing more data losses.

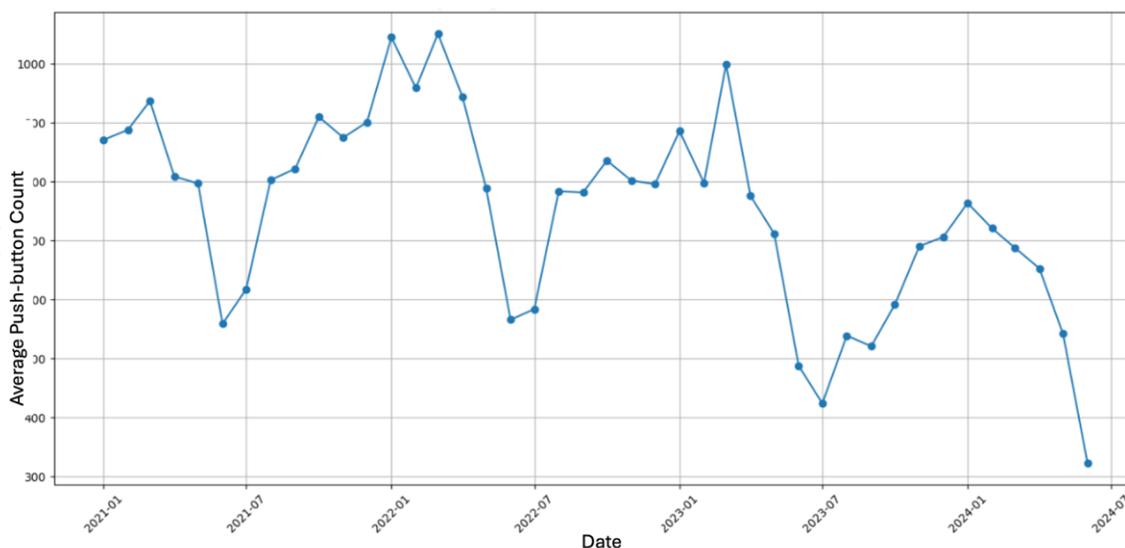
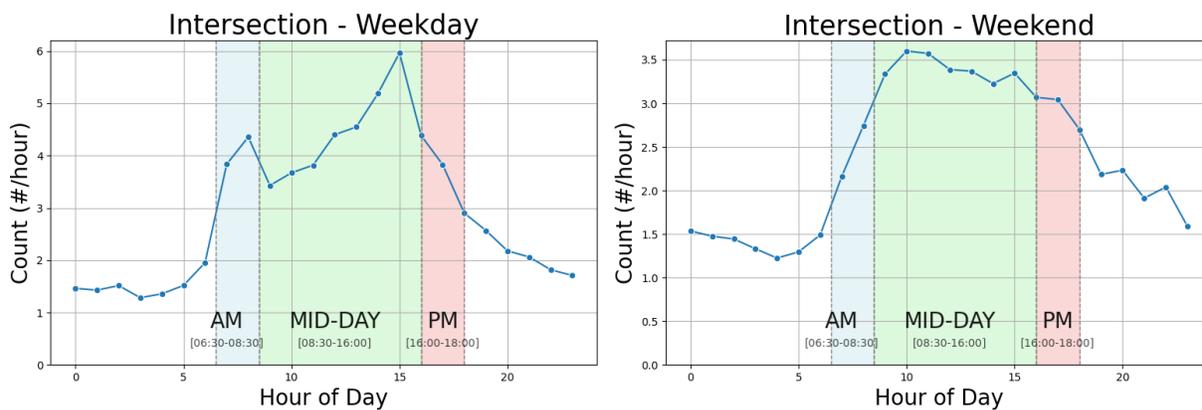


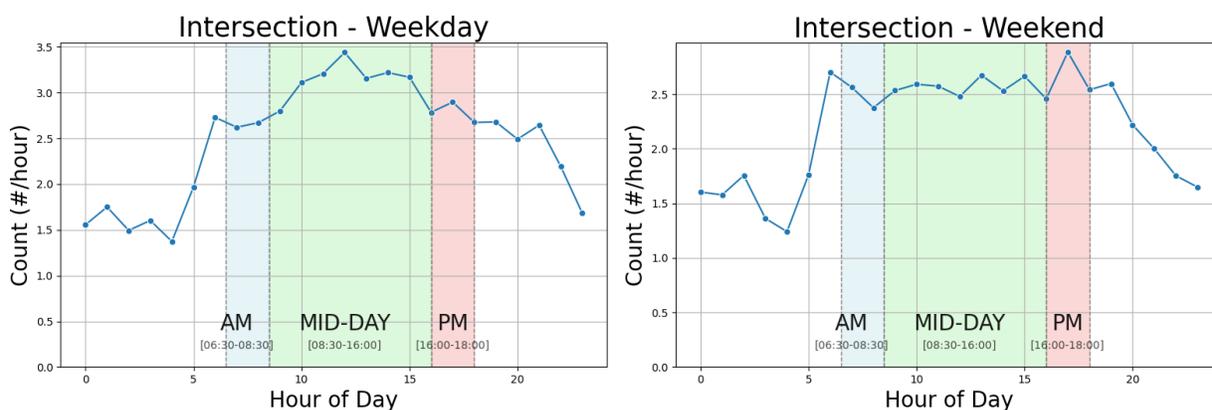
Figure 4.20. Monthly Average Push-button Counts Across All Detectors

4.1.3.2.2 Miovision average pedestrian count

Figure 4.21 compares hourly average pedestrian counts at intersections for two distinct periods, January 2023 and July 2023, focusing separately on weekdays and weekends. The data reveals a strong standard daily temporal pattern, a slight seasonal pattern, but it has low counts overall, likely due to the Miovision-equipped intersections being located in suburban areas.



(a) Hourly average pedestrian count by day of January 2023



(b) Hourly average pedestrian count by day of July 2023

Figure 4.21. Hourly average pedestrian count by day in January and July

On weekdays in January 2023, the pedestrian count peaks during the Mid-day period, with activity reaching approximately 5 counts/hour between 10 AM and 2 PM. Pedestrian counts are low overall due to the suburban and rural nature of the Miovision-equipped intersections but still show a pattern of regular daytime activity. July 2023 shows a similar daytime activity pattern but with reduced pedestrian activity, likely due to the extreme summer heat, which discourages outdoor activity, particularly during the day.

On weekends, the daytime temporal patterns are largely the same, but with lower average counts than weekdays, which then decline further during the summer.

4.1.3.2.3 Overall Trends

The comparison shows clear seasonal variations, with January 2023 having higher pedestrian counts across both weekdays and weekends than July 2023. The drop in July, particularly during

mid-day, is likely influenced by higher temperatures, deterring outdoor movement. July also shows a slight decrease in activity towards the late afternoon, likely in response to the summer heat. These patterns highlight how weather and seasons affect pedestrian activity, which could be crucial for transportation planning and infrastructure development aimed at meeting pedestrian needs throughout the year.

4.1.3.3 Miovision and MaxView sensor location overlap

To investigate the relationship between push button counts and actual pedestrian counts, we require event-based push button data along with ground-truth pedestrian count data that are both temporally and spatially aligned. Accordingly, we examined the locations of the MaxView and Miovision data in January 2023.

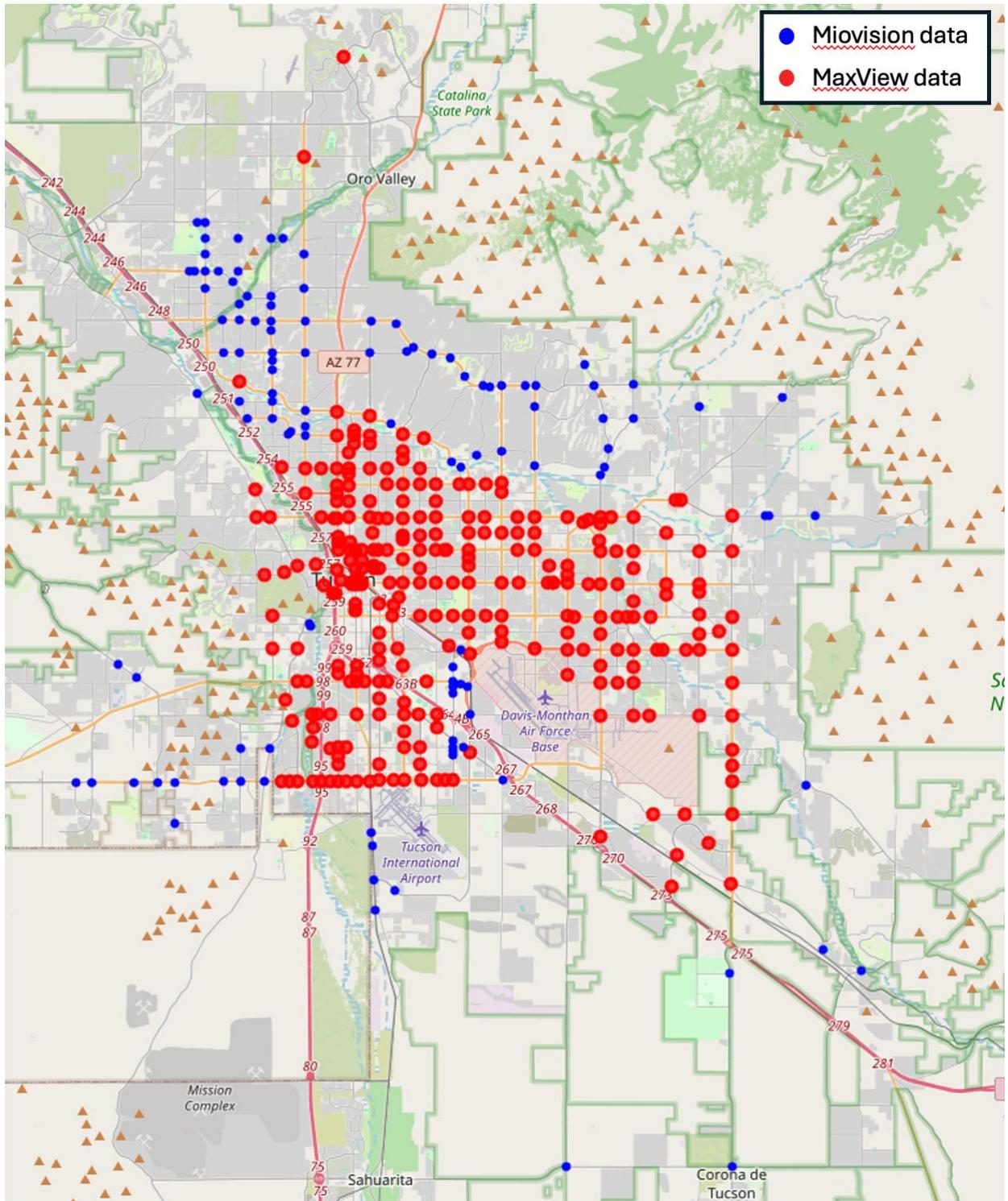


Figure 4.22. MaxView and Miovision locations of January 2023

Given the lack of location overlap between the available datasets, we have decided to use the PAG TMC pedestrian count data as our source of ground-truth information for actual pedestrian counts. This data was collected at major intersections throughout the PAG region, ensuring a higher level of accuracy and reliability. By utilizing this dataset, we can more effectively validate the push button counts and establish a meaningful comparison across locations where automatic pedestrian detection data is unavailable or insufficient. This approach allows us to maintain consistency in our analysis while accounting for the spatial limitations of the original datasets.

4.1.4 PAG Turning Movement Count (TMC) Pedestrian Data

4.1.4.1 Data Elaboration

Pedestrian survey data are collected as part of PAG's TMC data collection efforts, typically for durations of 4, 16, or 24 hours, depending on the specific needs and funding availability of the program. These counts are conducted either in the fall (September to November) or the spring (January to April). [Figure 4.23](#) below shows the counts collected per month from 2013 to 2023, typically averaging around 150 locations per year.

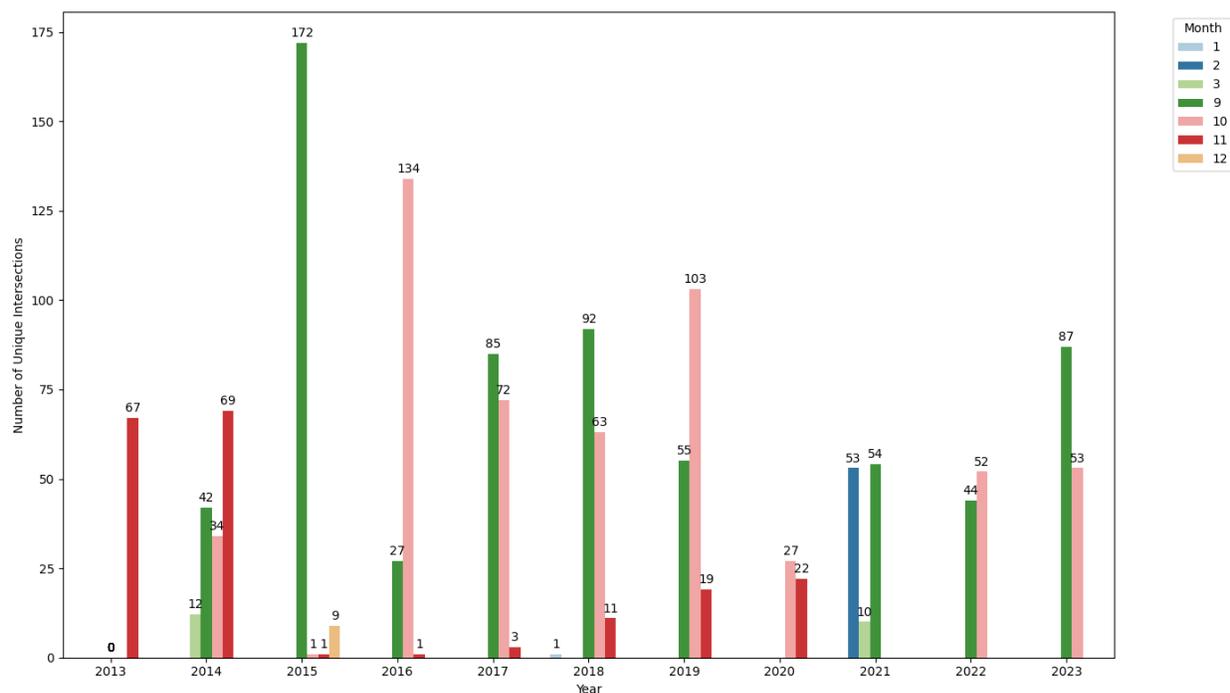


Figure 4.23. Distinct intersection number from the survey data

We choose the months of September and October of the year 2023 to compare the push button count data with the actual pedestrian count data.

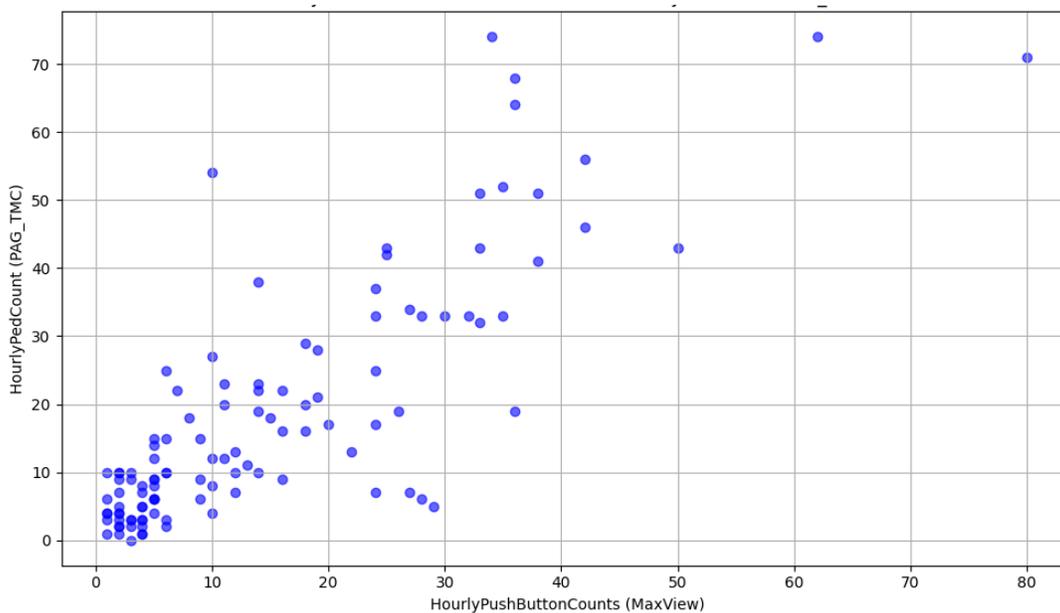


Figure 4.24. Relationship Between Hourly Push Button Counts (MaxView) and Hourly Pedestrian Counts (PAG_TMC) of September 2023

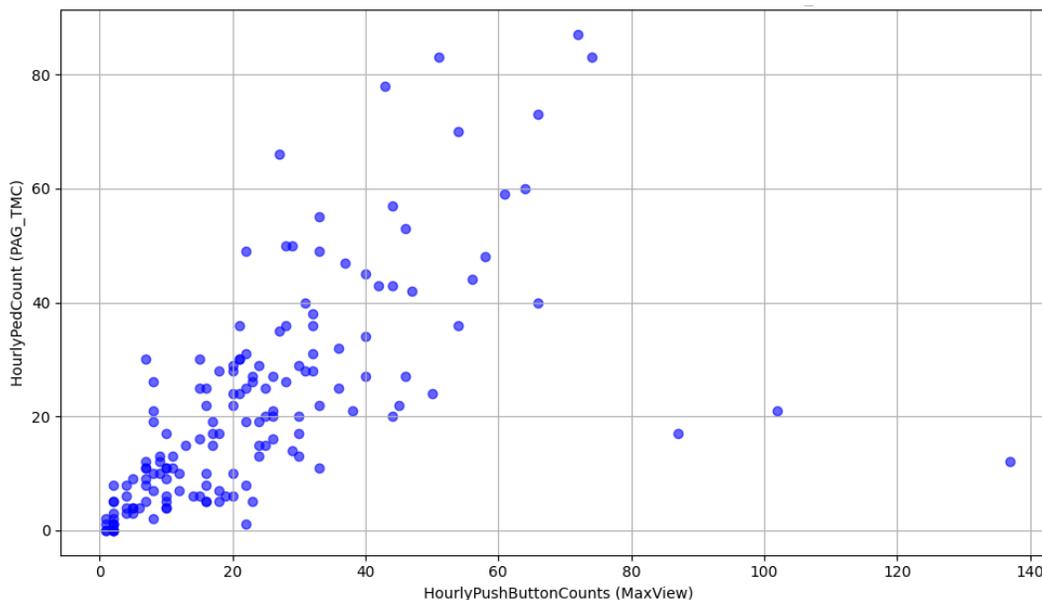


Figure 4.25. Relationship Between Hourly Push Button Counts (MaxView) and Hourly Pedestrian Counts (PAG_TMC) of October 2023

The scatter plots for September 2023 and October 2023 show overall strong correlations between the pedestrian counts and push counts, with pedestrian counts generally in line with push button counts. The outliers in the bottom right area of [Figure 4.25](#) are listed in [Table 4.5](#).

Table 4.5. Outliers in bottom right of Figure 4.25

Index	DeviceID	Date	Hour	HourlyPedCount	HourlyPushButtonCounts
137	582	10/4/23	16:00:00	12	137
138	582	10/4/23	17:00:00	21	102

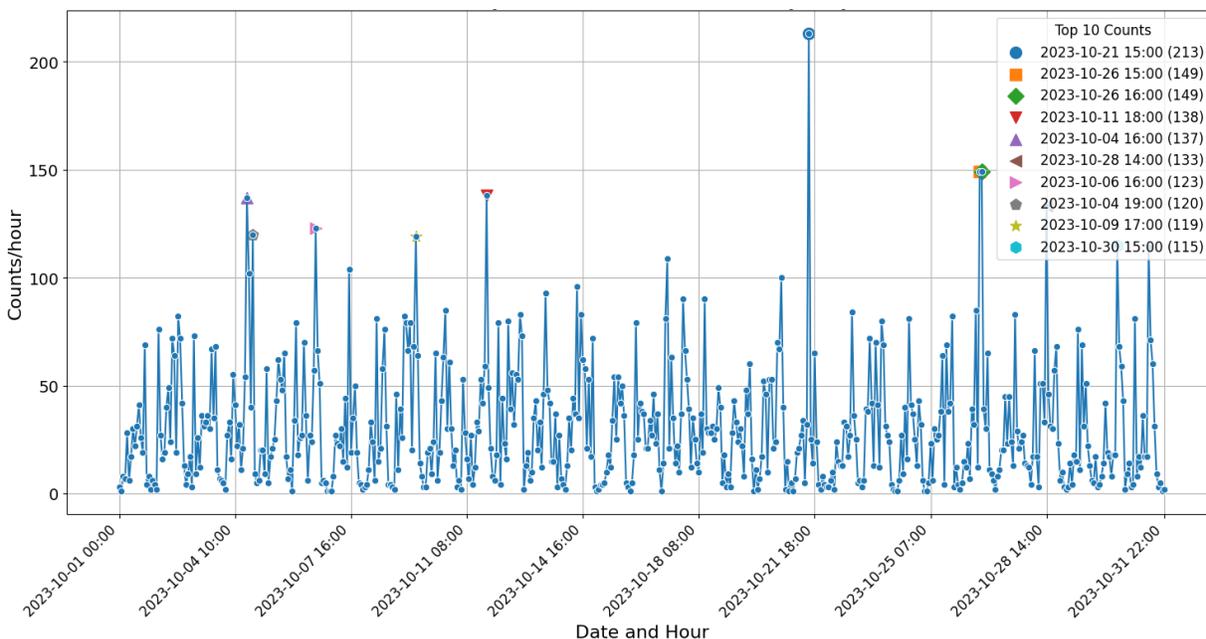


Figure 4.26. Hourly pedestrian counts for Device 582 in October 2023

This appears to be an anomaly in the push button data, where multiple pushes were recorded between 16:43:00 and 16:44:00 on October 4, 2023. In practice, we aggregate these consecutive pushes to minimize the impact of such outliers on the estimation model. After eliminating the outliers, the data shows a strong linear relationship between push button counts and actual pedestrian counts in both months.

CHAPTER 5. Micromobility (Bike sharing and E-scooter) Data Collection and Analysis

Micromobility is increasingly recognized as a viable alternative for short-distance travel and as a complement to other modes of transportation, such as transit, according to Zuniga-Garcia, N. et al. (2021). Berke, A. et al. (2024) and Kabra, A. et al. (2020) highlight the diverse benefits of integrating micromobility into urban transport systems. Expanding



micromobility options not only provides alternatives to traditional transportation modes but also significantly reduces vehicle congestion on roads. This reduction in road traffic contributes to improved traffic flow and advances in transportation sustainability.

This chapter involves a thorough review and analysis of existing micromobility data in Pima County to inform the development of performance measures of micromobility. Each dataset was carefully examined and evaluated to determine its suitability for use in performance measure development.

5.1. Deliverables

Table 5.1 provides an overview of the available micromobility data, including the used cases considered for each dataset. In subsequent sections, each data source will be detailed, with a focus on providing comprehensive information and analysis.

Table 5.1. Micromobility data sources in Pima county

Data		Source	Collection method	Use Cases
GBFS* Real time	E-scooter GBFS	Spin	API**	Performance measure: <ul style="list-style-type: none"> • <i>System level vehicle utilization rate</i> • <i>Geographic time availability rate</i> • <i>Geographic vehicle coverage rate</i>
	Bike Share GBFS	TUGO Bike	API**	Performance measure: <ul style="list-style-type: none"> • <i>System level vehicle utilization rate</i> • <i>Geographic vehicle utilization rate</i> • <i>Geographic time availability rate</i> • <i>Geographic vehicle coverage rate</i>
StreetLight Data***		Location Based Service	Account provided by Pima County	Micromobility trips attribute analysis

GBFS*: General Bikeshare Feed Specification

API**: Application Programming Interfaces

StreetLight Data***: We used 2022 StreetLight data, which is the latest time available

Data from both the e-scooter and TUGO bike share systems are sourced from the General Bikeshare Feed Specification (GBFS) real-time API, providing detailed information on the availability of each micromobility mode. The API offers real-time data on the location of each e-scooter and the number of bikes available at each TUGO bike share station. Estimated demand for the micromobility mode of transportation is derived from StreetLight data, which utilizes location-based services (LBS) to estimate traffic patterns. As illustrated in [Figure 5.1](#), the GBFS data is primarily utilized for analyzing the availability and usage patterns of the e-scooter and TUGO bike share systems, encompassing both geographic and system-level analyses. Conversely, Streetlight data informs demand analysis by estimating the micromobility traffic at Pima County areas. The integration of these datasets will be used for further development of various performance measures for micromobility services. Detailed methodologies pertaining to data processing, description, and analysis are elaborated in subsequent sections of this report.

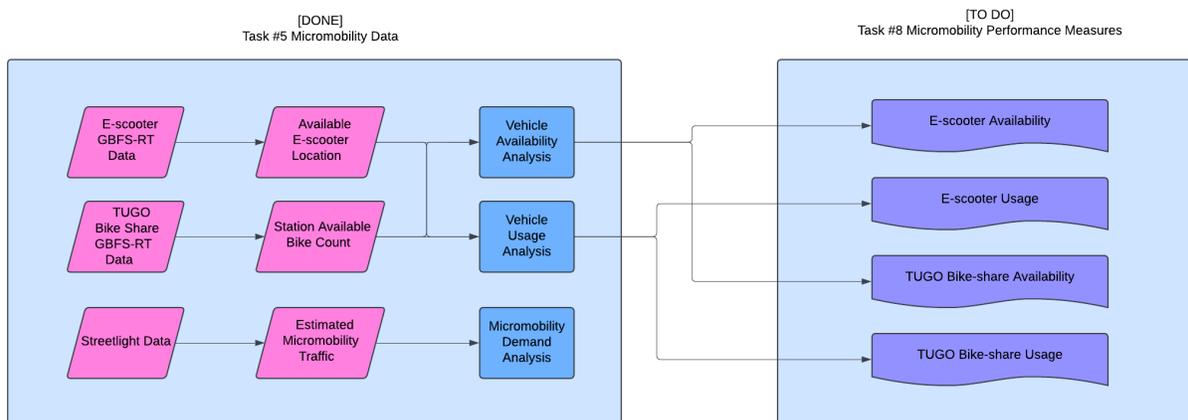


Figure 5.1. Micromobility methodology flowchart

5.2. E-scooter Data

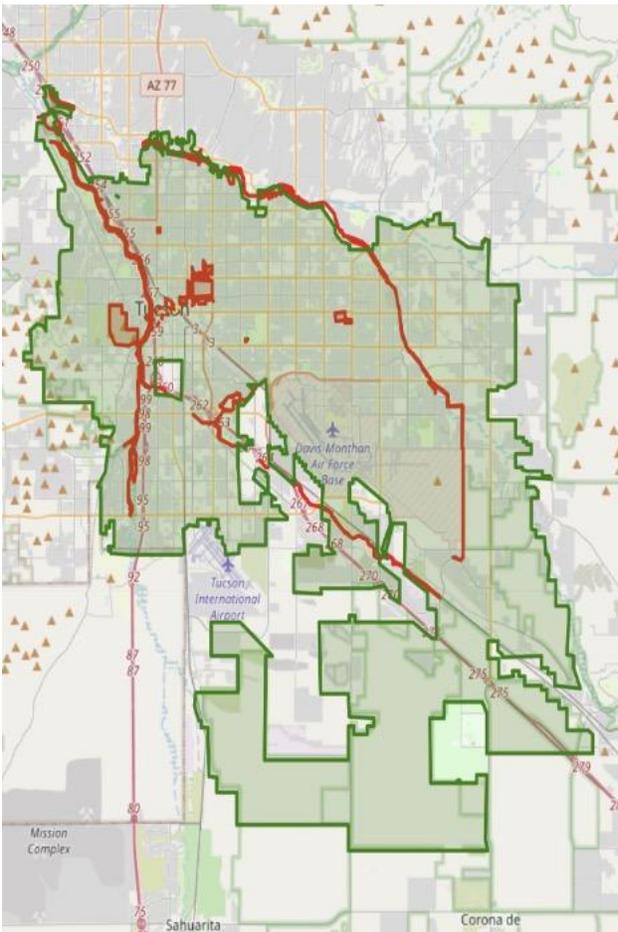
This section explores the e-scooter data that was used for the development of performance measures. We collect data daily through the GBFS real-time API provided by the City of Tucson, recording and archiving it for further processing. Subsequently, we clean and organize the archived data to calculate metrics for micromobility.

5.2.1 E-scooter GBFS (General Bikeshare Feed Specification)

5.2.1.1 Data Elaboration

In early 2018, the City of Tucson was approached by several dockless scooter companies, prompting the formation of a Shared Mobility Working Group. This group engaged with stakeholders such as the Fourth Avenue Merchants Association, Downtown Tucson Partnership, Mercado District, the University of Arizona, and Pima Community College to gauge the potential impacts of a Shared Mobility Pilot Program.

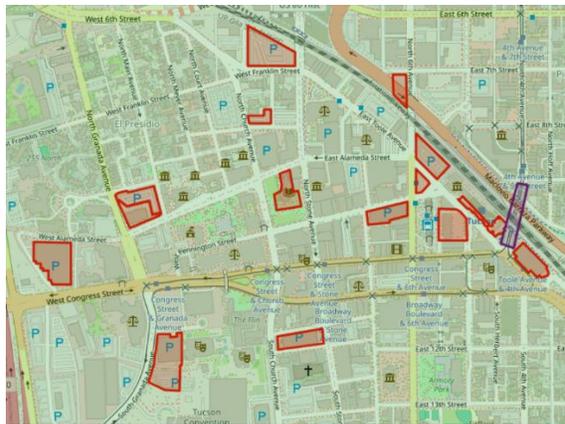
This pilot assessed e-scooters as a mode of transportation to enhance access to public transit and included e-scooter GBFS data that provides shared e-scooter information in real time, such as the location, vehicle status, battery status. **Figure 5.2** illustrates the geofencing area of the e-scooter operation with different colors indicating different regional regulations. Green areas permit e-scooter use, purple areas such as the 4th Ave underpass in **Figure 5.2(c)** allow riding but not starting or ending rides, and red areas prohibit e-scooter use. As depicted in **Figure 5.2(a)**, e-scooter operation is allowed in the City of Tucson area. As illustrated in **Figure 5.2(b), (c), and (d)**, the areas that are not allowed for the operation are schools, parking lots, and some bike paths.



(a) E-scooter GBFS geofencing area



(b) E-scooter not allowed: School



(c) E-scooter not allowed: Parking lots



(d) E-scooter not allowed: Bike path

Figure 5.2. E-scooter GBFS geofencing map

5.2.1.2 Data Processing

Data was obtained from the GBFS realtime API. GBFS data is recorded in real time, but the frequency and timing of the data update varies for each GBFS file. For example, the “*free_bike_status*” is updated every 2.5 to 5 minutes, while most of the other system and operation data are less frequently updated. Thus, the primary limitation of the data lies in the inability to determine the data update interval based on the available observations, as well as the fact that the *bike_id* changes with each data request. Consequently, it is not possible to obtain complete trajectory data for individual e-scooters. Therefore, the focus of data processing should be on accurately tracking the real-time count of e-scooters in different zones/areas. This information will be instrumental in analyzing the availability of e-scooters across various regions, particularly in conjunction with public transit stations or in densely populated areas such as downtown Tucson.

The GBFS data is collected at one-minute intervals through a script that operates continuously, which is designed with robust error handling and memory optimization strategies. Once retrieved, the data is stored in a Pandas DataFrame, where a ‘*query_time*’ column is appended to capture the exact time of each data retrieval, in Mountain Standard Time (MST). This data is then systematically saved into daily CSV files, with records being appended every minute. Each file is a compilation of all daily data. This methodical collection process supports the effective and consistent gathering of real-time micromobility data, which is crucial for detailed analysis and strategic planning in micromobility management.

The e-scooter GBFS data is detailed in [Table 5.2](#). To assess e-scooter accessibility and availability, the most critical information pertains to the locations and status of available e-scooters. The “*free_bike_status*” GBFS file provides information on vehicles that are not currently rented. By analyzing the geographic location and corresponding timestamp of these available vehicles, the availability and accessibility of e-scooters can be effectively measured. The fields highlighted are the attributes that were used.

Table 5.2. Available E-scooter GBFS data summary

Files	Description	Field Name	Type
“ <i>free_bike_status</i> ”	Describes all vehicles that are not currently active rental	<i>bike_id</i>	ID
		Lat	Latitude
		lon	Longitude
		<i>is_disabled</i>	Boolean
		<i>is_reserved</i>	Boolean
		<i>vehicle_type_id</i>	ID
		<i>last_reported</i>	Timestamp
		<i>current_range_meters</i>	Non-negative float
		<i>current_fuel_percent</i>	Non-negative float
		<i>pricing_plan_id</i>	ID
“ <i>gbfs_versions</i> ”		versions	Array<Object>

Files	Description	Field Name	Type
	Describes all of the versions that are available	versions[].version	String
		versions[].url	URL
"geofencing_zones"	Describes geofencing zones and their associated rules and attributes	geofencing_zones	GeoJSON FeatureCollection
		geofencing_zones[].type	String
		geofencing_zones[].features	Array<Object>
		geofencing_zones[].features[].type	String
		geofencing_zones[].features[].geometry	GeoJSON MultiPolygon
	geofencing_zones[].features[].properties.start	Array<Rule>	
"system_information"	Describes system operation information	system_id	ID
		languages	Array<Language>
		name	Array<Localized String>
		opening_hours	String
"system_pricing_plans"	Describes e-scooter system pricing for the users	plans	Array<Object>
		plans[].plan_id	ID
		plans[].currency	String
		plans[].price	Non-Negative Float
		plans[].is_taxable	Boolean
		plans[].description	Array<Localized String>
"system_regions"	Describes the regions for a system.	regions	Array<Object>
		regions[].region_id	ID
		regions[].name	Array<Localized String>
"vehicle_types"	Describes the vehicle types that are available for rental	vehicle_types	Array<Object>
		vehicle_types[].vehicle_type_id	ID
		vehicle_types[].form_factor	Enum

: Data used for Performance Measures

5.2.1.3 System Information Data

The system information is provided by the GBFS data's "system_information" and "system_pricing_plan" files as described in [Table 5.2](#). The "system_information" file includes system time zone, license, operator, operation rental application, and the updated time. According to the GBFS data that was archived on August 11, 2024, the e-scooter system is operated by Bird

Rides, Inc, an operator of SPIN shared e-scooter service. The system time zone is set as US/Arizona, and the rental application is available for both Android and IOS. The “*system_pricing_plan*” file includes the pricing plans’ name, price, price description, and the currency of pricing. One pricing plan was available according to the GBFS pricing plan data that was archived on August 11, 2024. The e-scooter has “Bird Pricing Plan” with USD as currency, pricing \$1.00 to unlock the vehicle and \$0.39 per minute since the unlock. The payment is taxable.

5.2.2 *Data Analysis*

5.2.2.1 *E-scooter availability and accessibility*

To assess the availability of e-scooters in the City of Tucson, the operational areas and specific locations of e-scooters are visualized using heatmaps for distinct time periods: 7:00 AM – 8:00 AM, 12:00 PM – 1:00 PM, and 5:00 PM – 6:00 PM, which represent AM peak, midday, and PM peak, respectively. These periods were selected to illustrate typical usage patterns, while other hourly intervals analysis results are also available.

For a comprehensive comparison between weekdays and weekends, July 31st (a weekday) and August 3rd (a weekend) in 2024 were chosen for further analysis. The proximity of e-scooters to public transit was analyzed by overlaying transit station locations on the e-scooter heatmap. As shown in [Figure 5.3](#), each blue circle marker indicates the presence of an e-scooter. The opacity of these markers varies, with denser areas indicating a higher availability of e-scooters.

These maps show consistent patterns of e-scooter availability from AM to PM peak across both weekdays and weekends. A significant concentration of e-scooters is consistently observed in downtown Tucson and near the University of Arizona campus. On July 31st, e-scooters were more dispersed during the AM peak, were more concentrated in the 4th Avenue area during the midday and were more concentrated in both the 4th Avenue and downtown areas during the PM peak. Similarly, on the weekend, a broader dispersion of e-scooters was shown during the AM peak compared to other time periods.

The maps also highlight accessibility, with e-scooters frequently located near Sun Link streetcar stations in the downtown area, suggesting a possibility of their use as a link between residential or workplace locations and the Streetcar network.

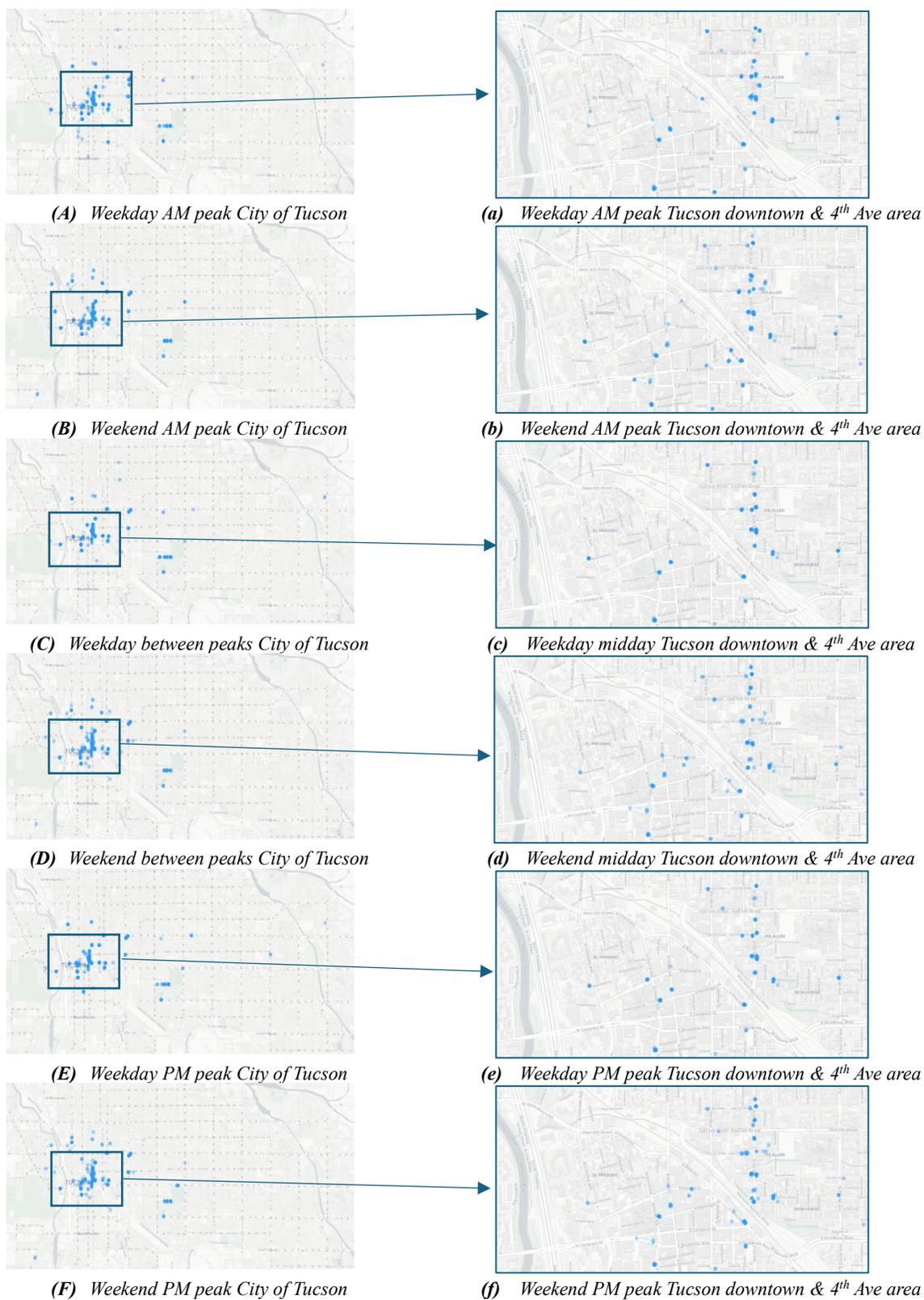


Figure 5.3. E-scooter availability map on July 31st (weekday) and August 3rd (weekend), 2024(continued)

5.2.2.2 System level e-scooter availability

To assess the pattern of e-scooter availability, an analysis was conducted over a one-week period from September 16th, 2024, to September 22nd, 2024, during a school semester. This analysis provides insights into fluctuations in e-scooter availability as shown in **Figure 5.4**. The findings indicate that Monday had relatively low availability, with figures ranging from 210 to 220 e-scooters, suggesting higher usage rates at the start of the week. Availability level persisted throughout the weekdays. On Friday, however, availability dipped slightly to a range of 225 to 235 e-scooters. Over the weekend, availability further decreased, fluctuating between 210 and 230 e-scooters. The daily pattern showed a consistent trend, with higher availability in the early morning hours. As the AM peak hours commenced, availability diminished, continuing to decline through midday. The lowest availability was observed during the PM peak hours, followed by a recovery in availability after these peak periods. This pattern of fluctuating availability was consistent throughout the different days, illustrating a weekly cycle in e-scooter usage.



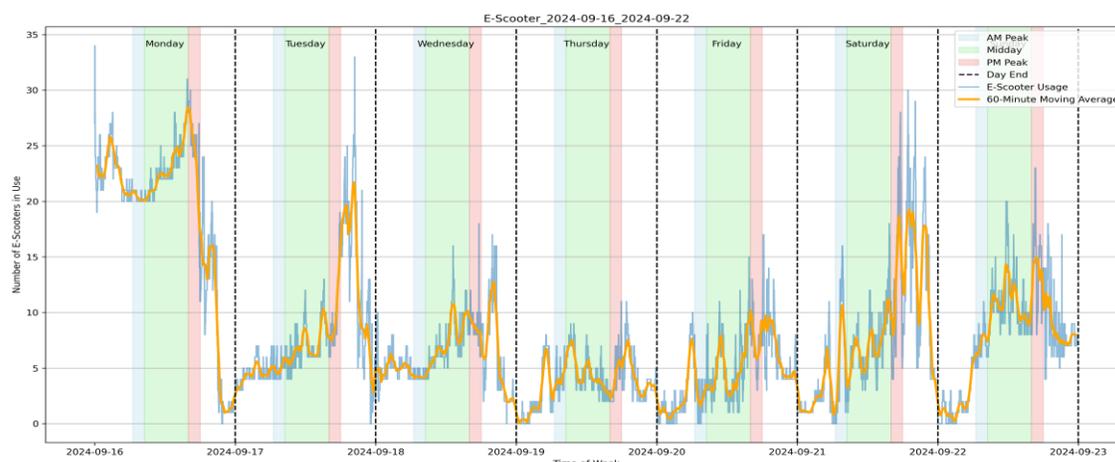
Figure 5.4. E-scooter availability plot from September 16th to September 22nd, 2024

5.2.2.3 System level usage analysis

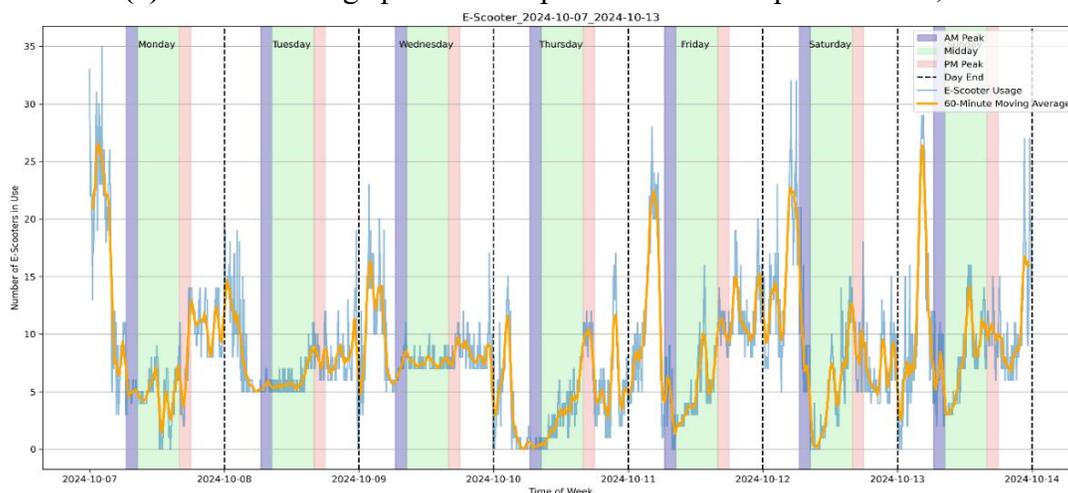
Based on the e-scooter availability analysis, system-level usage patterns are identified using the assumption that the highest number of vehicles available during any point in a day represents the system's maximum capacity for that day as illustrated in **Figure 5.4** with the red line indicating the daily maximum available vehicles. **Figure 5.5** illustrates actual e-scooter usage over time, where the blue line shows the number of scooters actively in use and the yellow line represents a 60-minute moving average.

Figure 5.5(a) and **Figure 5.5(b)** show the number of E-scooters in use during the week of September 16 to 22 and the week of October 7 to 13, respectively. With the exception of Monday September 16, weekdays have 10 or fewer active E-scooters until the post-PM-peak and early AM periods, where it repeatedly spikes to over 20 active E-scooters. Weekends show similar overall patterns, but with increased daytime activity, peaking at 10-15 active E-scooters and similar post-PM-peak activity.

These findings highlight the dynamic nature of micromobility usage. While certain patterns such as peak-hour activity and weekend surges are consistent, variability between weeks suggests external factors like weather, events, or academic schedules may influence demand. To capture a comprehensive picture, CHAPTER 8 incorporates an average usage profile across multiple weeks to derive robust performance measures for each day of the week.



(a) E-scooter usage plot from September 16th to September 23rd, 2024



(b) E-scooter usage plot from October 7th to October 13th, 2024

Figure 5.5. E-scooter usage plot

5.2.2.4 System level disabled vehicle analysis

According to MobilityData (2024), “disabled vehicles” refer to those that, although present in the field, are not available for rental due to mechanical issues, low battery, or other similar reasons. The GBFS data includes disability status for each available vehicle, enabling a detailed system-level analysis of vehicle disabilities as illustrated in [Figure 5.6](#).

The data reveals that throughout the week, the number of disabled e-scooters remains relatively low. On weekdays, the count of disabled vehicles typically ranges from 0 to 2. However, during the weekend, and particularly on Saturday, there is a noticeable increase in disabled vehicles, with counts ranging from 0 to 4. This trend is particularly striking given that Monday shows a higher estimated usage of vehicles compared to Saturday yet displays fewer disabled vehicles. This analysis suggests a need for targeted maintenance and battery management strategies to enhance vehicle availability, particularly on weekends.

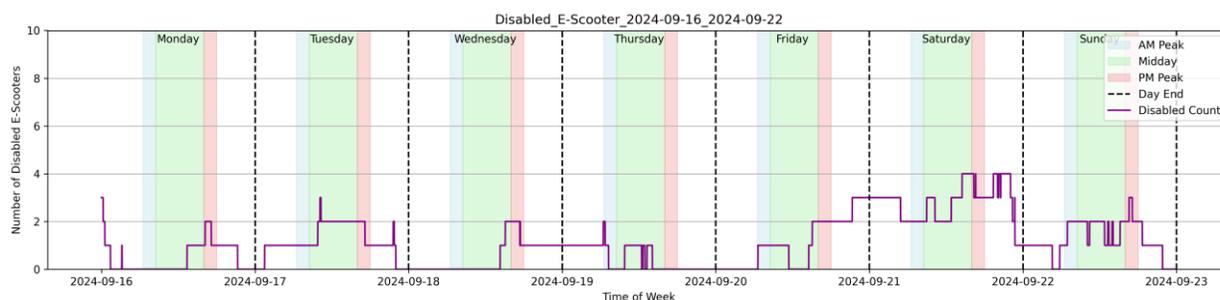


Figure 5.6. Disabled e-scooter plot from September 16th to September 22nd, 2024

5.3. Bike sharing data

This section explores the bike-share data that was used for the development of performance measures. As with the e-scooter data, data was collected continuously through the General Bikeshare Feed Specification (GBFS) real-time API provided by the City of Tucson.

5.3.1 *TUGO Bike Share GBFS*

5.3.1.1 Data Elaboration

TUGO Bike Share, available 24/7 year-round, offers 330 bikes across 40 stations strategically placed in 13 Tucson neighborhoods. This initiative results from a collaboration between the City of Tucson, Shift Transit as the system operator, and PBSC Urban Solutions as the equipment provider, with station locations chosen to enhance access to community resources, employment centers, and established bike infrastructure.

Bike share systems have been adopted by cities to improve traffic congestion and sustainability. Evaluating the system’s operational performance level is important to understand its impact on the city’s transportation system. Kabra, A. et al (2020) estimated the impact of the bikeshare system’s

performance, accessibility and bike-availability on the ridership. To assess the bikeshare system's performance, the most important data is bike-share station information and availability at each bike-share station spatially and temporally. TUGO bike-share GBFS data provides station information including geographic information and real-time vehicle availability of each station. **Figure 5.7** illustrates the locations of the TUGO bike-share station locations. The stations are in the City of Tucson, more densely located in the Tucson downtown and the University of Arizona campus areas.

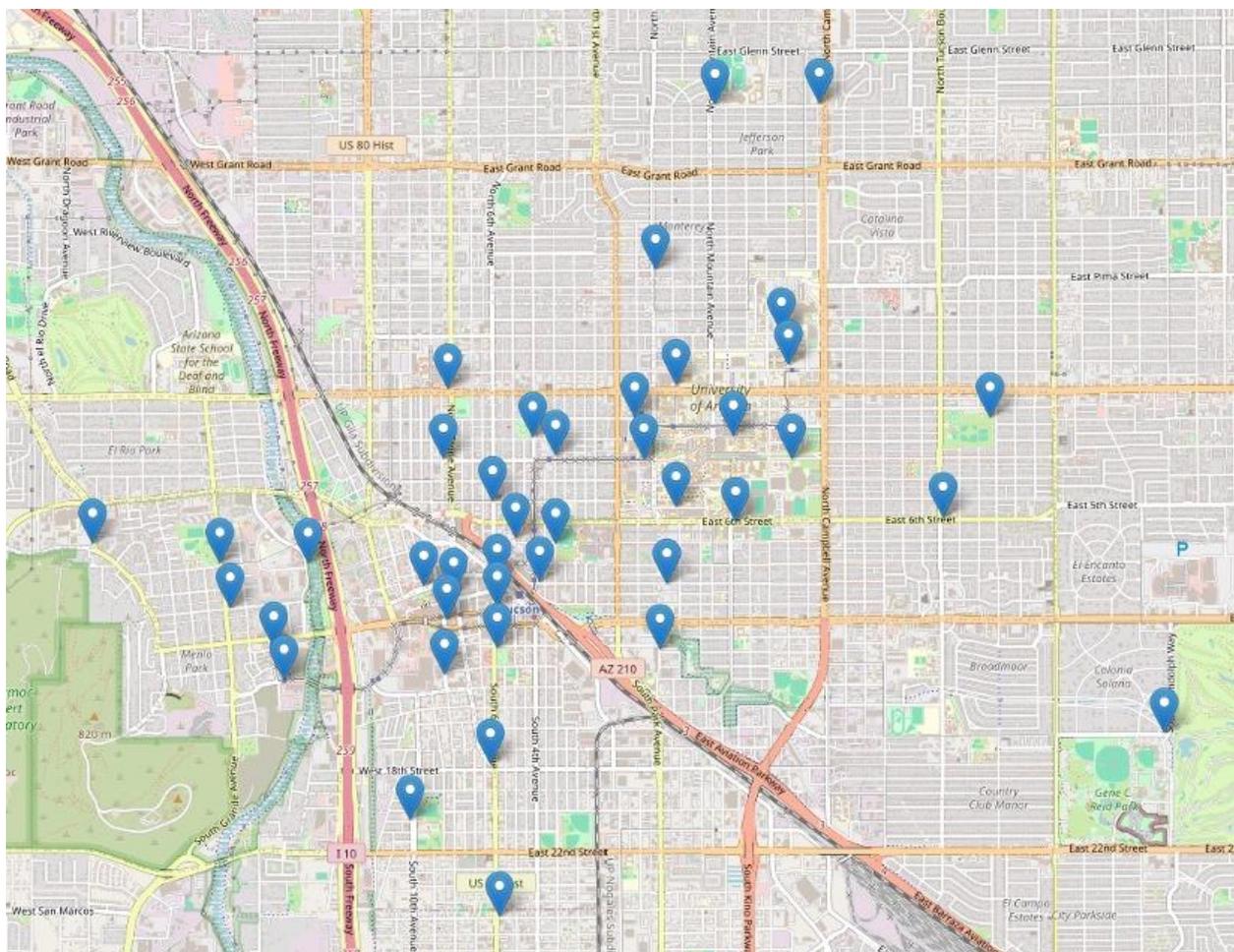


Figure 5.7. Bikeshare GBFS station locations map

5.3.1.2 Data Processing

The bike share data is collected from the GBFS API, similar to the e-scooter data updates. The TUGO bike-share system utilizes real-time data from the GBFS, with varying update intervals. Among the GBFS files, the “station_status” file is updated most frequently. It provides comprehensive information about the status of each station, including vehicle availability and operational status of the station. Updates for this file are applied collectively for all bike-share stations. As with the e-scooter GBFS data, the TUGO data is collected every minute using a

continuously running script that is optimized for error handling and memory efficiency. The data is stored in a Pandas DataFrame, and a ‘query_time’ column is appended to record the exact time of each query, synchronized to Mountain Standard Time (MST). This data is then saved into daily CSV files, with new entries appended every minute. These files are organized chronologically by the query time, ensuring systematic recording and easy retrieval of data. This structured approach to data archiving facilitates efficient analysis and monitoring of bike-share system usage.

Table 5.3 below summarizes bike share GBFS data, each GBFS file’s available data and descriptions. The most critical data files for assessing the Bike Share are “*station_information*” and “*station_status*”. The “*station_status*” GBFS file provides the availability information at the station level. The “*station_information*” provides the geographic information of each bike-share station which can be interpreted as accessibility metrics to the public transit.

Table 5.3. Available bike share GBFS data summary

Files	Description	Field Name	Type
“ <i>gbfs_versions</i> ”	Describes all of the versions that are available	versions	Array<Object>
		versions[].version	String
		versions[].url	URL
“ <i>system_information</i> ”	Describes system operation information	system_id	ID
		languages	Array<Language>
		name	Array<Localized String>
		opening_hours	String
“ <i>system_pricing_plans</i> ”	Describes e-scooter system pricing for the users	plans	Array<Object>
		plans[].plan_id	ID
		plans[].currency	String
		plans[].price	Non-Negative Float
		plans[].is_taxable	Boolean
		plans[].description	Array<Localized String>
“ <i>system_regions</i> ”	Describes the regions for a system.	regions	Array<Object>
		regions[].region_id	ID
		regions[].name	Array<Localized String>
“ <i>vehicle_types</i> ”	Describes the vehicle types that are available for rental	vehicle_types	Array<Object>
		vehicle_types[].vehicle_type_id	ID
		vehicle_types[].form_factor	Enum
“ <i>station_information</i> ”	Describes the bike-share stations	stations	Array<Object>
		Stations[].station_id	ID

Files	Description	Field Name	Type
	operation information	Stations[].name	Array<Localized String>
		Stations[].lat	Latitude
		Stations[].lon	Longitude
"station_status"	Describes the capacity and availability of each bike-share station	stations	Array<Object>
		stations[].station_id	ID
		stations[].num_vehicles_available	Non-negative integer
		stations[].vehicle_types_available[].vehicle_type_id	ID
		stations[].vehicle_types_available[].count	Non-negative integer
		stations[].is_installed	Boolean
		Stations[].is_renting	Boolean
		Stations[].is_returning	Boolean
Stations[].is_reported	Timestamp		

: Data used for Performance Indicators

System Information Data:

The system information is provided by the GBFS data's "system_information" and "system_pricing_plan" files as outlined in [Table 5.3](#). Similar to the e-scooter system information data, the "system_information" file contains system time zone, license, operator, total vehicle's mechanical count, station count, and the updated time. According to the GBFS data that was archived on August 11, 2024, the bike-share system is operated by TUGO bike share. The system time zone is set as US/Arizona. The vehicle count and station count are defined as a total of 329 vehicles and 41 stations. The "system_pricing_plan" file includes the pricing plans' name, price, price description, and the currency of pricing. A total of eleven different pricing plans were defined according to the GBFS data archived on August 11, 2024. Different types of passes including Annual Pass, Monthly Pass, Senior Citizen Pass, TUGO For All, Corporate Annual Pass, etc., are described in [Table 5.4](#) with USD as currency. The payment is not taxable, and the payment will only be charged once at the beginning of trip.

Table 5.4. TUGO bike share pricing plan summary

Pricing Plan Name	Pass Type	Pass Price	Vehicle Unlock Fee (\$)	Price(\$)/ Minute
Annual Pass	Annual	80.0	NA	NA
Annual Pass (monthly installment)	Annual	96.0	NA	NA
Monthly Pass	Monthly	18.0	NA	NA
Senior Citizen Pass	Annual	65.0	NA	NA
TUGO For All	Daily	5.0	NA	NA
Corporate Annual Pass	Annual	80.0	NA	NA
Pay As You Go	Daily	NA	1.0	0.25

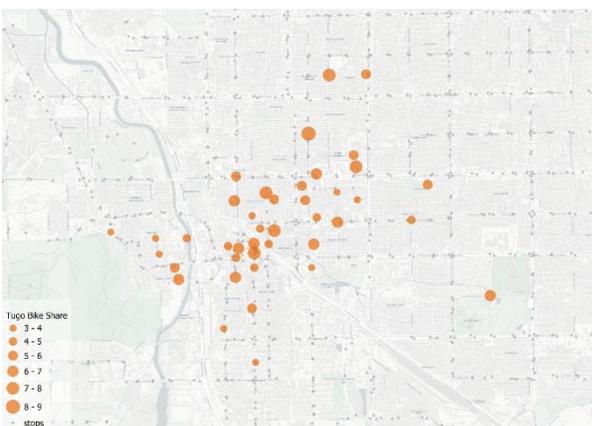
Note: TUGO For All pricing plan is for members of Supplemental Nutrition Assistance Program (SNAP), Medicaid and SunTran Economy Fare plan.

5.3.2 Data Analysis

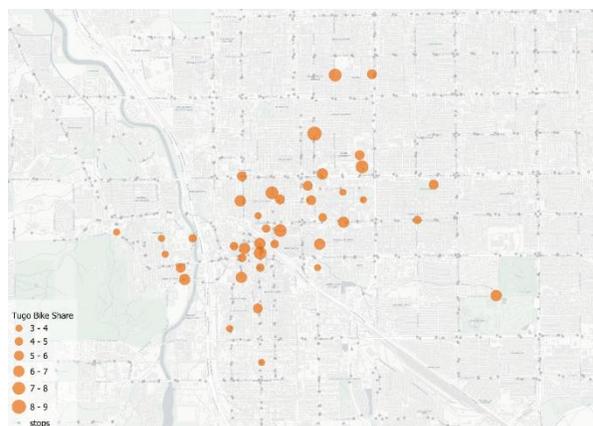
5.3.2.1 Bikeshare availability and accessibility

To provide insight into weekday and weekend usage of the TUGO bike-share system, data from two specific dates, July 31st and August 3rd, 2024, was analyzed. The availability of bikes at each station during selected time periods, 7:00 AM to 8:00 AM, 12:00 PM to 1:00 PM, and 5:00 PM to 6:00 PM, corresponds to AM peak, midday, and PM peak times. These intervals were chosen to showcase the availability patterns while maintaining the option for more detailed hourly analysis if needed. The proximity of bike stations to public transit routes was also analyzed to assess the integration of the bike-share system with other transportation modes. Stations are notably concentrated around downtown and near the University of Arizona campus area, which are significant hubs of activity. In **Figure 5.8**, the availability of bikes at each TUGO bike-share station is visually represented with orange circles, where larger circles indicate a higher number of available bikes.

Figure 5.8(a) through **5.8(f)** illustrate the availability patterns across the day from AM to PM peak, as well as between the selected weekday and weekend. Notably, there is no significant variation in vehicle availability across different times and dates. This consistent availability pattern suggests low usage during the observation period, possibly influenced by the extreme heat typical in Tucson during late July to early August. Furthermore, considering that one of the primary user groups of the bike-share system is University of Arizona students, the observed period coinciding with the school vacation may also explain the lack of fluctuation in bike availability. This analysis highlights the importance of contextual factors, such as weather and academic schedules, in influencing micromobility usage patterns.



(a) Weekday AM peak City of Tucson



(d) Weekend AM peak City of Tucson

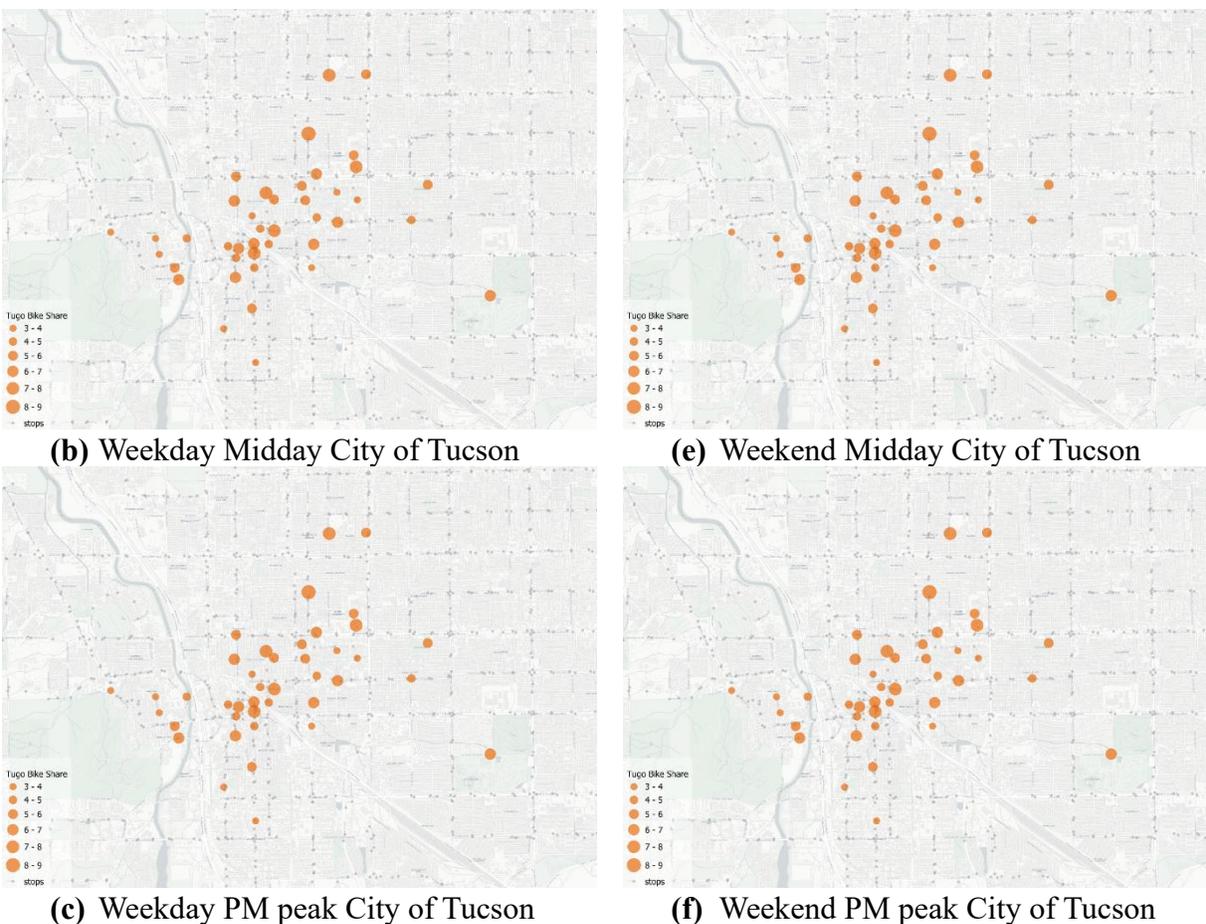


Figure 5.8. TUGO bikeshare availability map on July 31st (weekday) and August 3rd (weekend), 2024

5.3.2.2 TUGO bike sharing total vehicle availability analysis

To assess the pattern of bike availability, the TUGO bike-share system was analyzed over a week-long period from September 30, 2024, to October 6, 2024. The total availability of bikes was calculated by aggregating the number of available bikes at each station using GBFS data recorded at specific times throughout each day.

The detailed daily and hourly breakdown, depicted in [Figure 5.9](#), provides insights into the usage patterns and potential operational challenges faced by the bike-share system, indicating the times and days when additional resources might be needed to ensure optimal service levels.

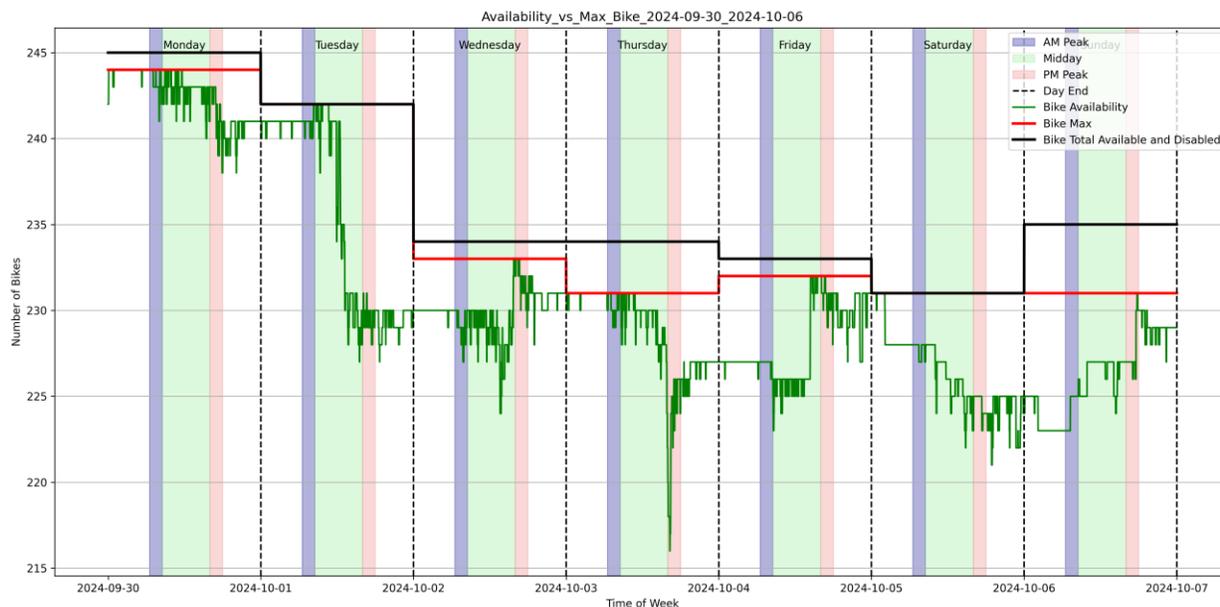


Figure 5.9. TUGO bike share total vehicle availability from September 30th to October 6th, 2024

5.3.2.3 TUGO bike share total vehicle usage analysis

To analyze the daily usage pattern of the TUGO bike share system, usage was calculated based on bike availability data from September 30, 2024, to October 6, 2024. The total bike usage was derived by subtracting the number of available bikes from the daily total of bikes, excluding any disabled vehicles, as depicted with a red line in [Figure 5.9](#).

[Figure 5.10](#) illustrates the varying patterns of bike usage throughout the week. Monday displayed the lowest usage, with bike utilization peaking at six bikes. In contrast, Tuesday and Thursday showed the highest usage, peaking at fifteen bikes. For the rest of the week, bike usage peaks fluctuated between seven and ten bikes.

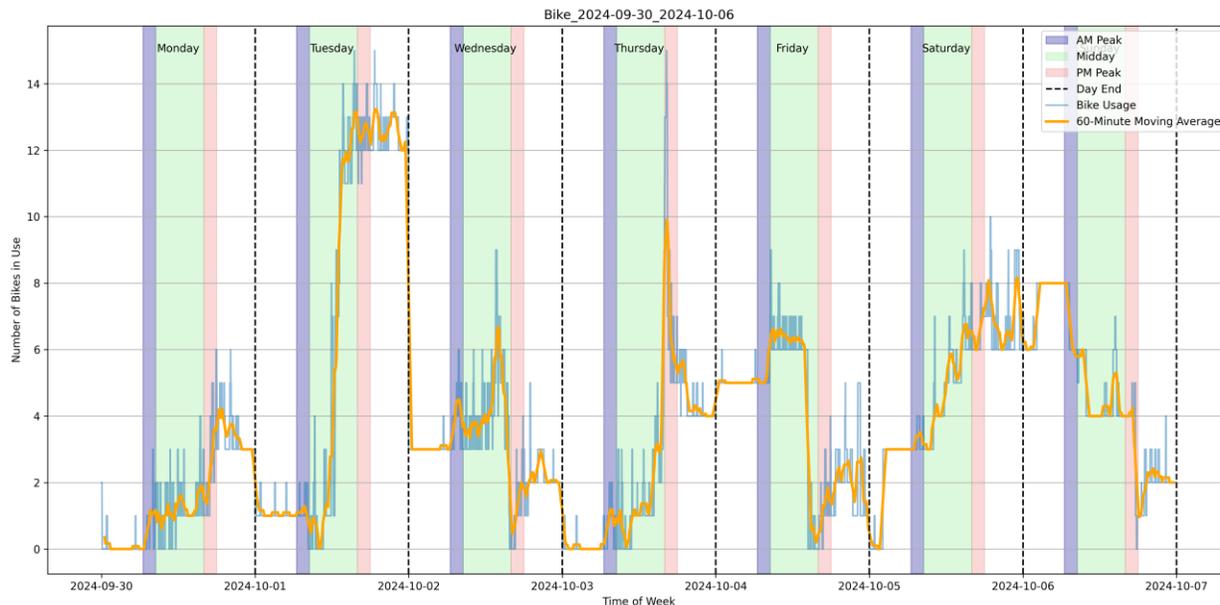


Figure 5.10. TUGO bike share total vehicle usage from September 30th to October 6th, 2024

5.3.3 *StreetLight Data*

5.3.3.1 Data Elaboration

StreetLight Data is a software as a service (SaaS) provider that offers data online and centrally managed, via their InSight platform. InSight provides various types of transportation analyses for different modes of transportation: trucks, bicycles, pedestrians, bus, and rail. Each mode of transportation has different available analyses and data sources, summarized in [Table 5.5](#). For example, StreetLight’s bicycle mode analyses include zone activity, origin-destination, origin-destination through middle filter, and trips to/from pre-set geography.

Zone activity analyzes all travel in a zone set, regardless of origins or destinations. Origin-destination analyzes trip and traveler attributes of user-designated trips’ origin and destination zone sets. O-D through middle filter analyzes origin, destination, and also the segments trips pass through. Trips to/from pre-set geography provides the trip starting and ending zone sets with pre-set geographies, such as US Census Block Groups, ZIP codes, and Traffic Analysis Zones (TAZ). Existing TAZ data available on the InSight platform was developed and updated by PAG. The UA team considered and applied the OD data from StreetLight InSight platform to analyze the demand of bikeshare in PAG area.

Table 5.5. StreetLight Data analysis availability for each transportation mode

Type of Analysis	Transportation Mode					
	All Vehicle	Truck	Bicycle	Pedestrian	Bus	Rail
Zone Activity						
Origin-Destination						
O-D Through Middle Filter						
Trips to/from Pre-Set Geography						
Top Routes for Zones						
Top Routes Between OD						
AADT						
Segment Analysis						
Turning Movement Counts						

(Reference: StreetLight Data (June 19, 2024))

In StreetLight Data, the bicycle data used in this analysis is sourced from LBS data collected through mobile apps. Using this LBS data, the volume of bicycle trips is calculated by normalizing the trip samples with a weighted population factor to account for regional variations in device usage. Additionally, a bicycle penetration rate is estimated based on the trip samples, considering the estimated vehicle penetration rate in the zones under analysis. These estimates are further refined by scaling them against nationwide permanent bicycle counters (StreetLight Data, retrieved June 19, 2024).

5.3.3.2 Data Limitation

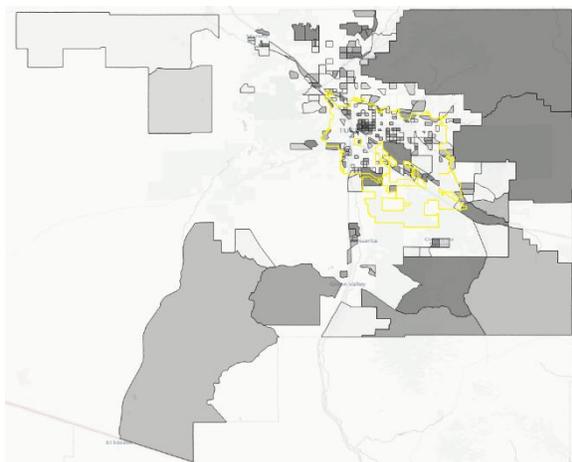
The StreetLight data utilizes location-based service estimates for pedestrian metrics, as previously described. Consequently, the accuracy and resolution of this data are lower compared to event-based and sensor-based data. Additionally, StreetLight data does not support performance analyses at the intersection level. More importantly, the most recent StreetLight data available to us through the shared account with PAG is from April 2022, which does not align with the GBFS data timeline. Therefore, for demonstration purposes, we have utilized StreetLight data from January 2022. This analysis can be easily extended to other time periods if additional data becomes available.

5.3.3.3 Data Analysis

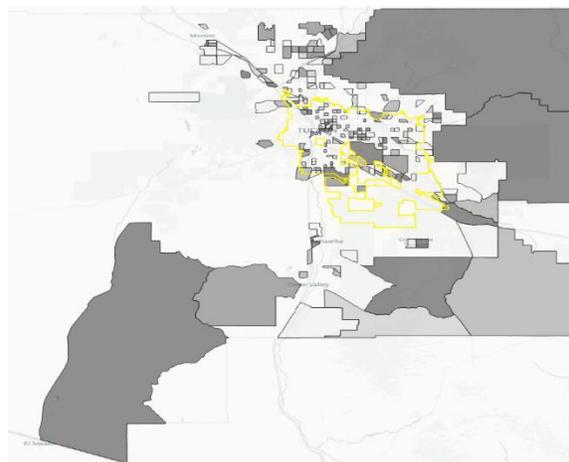
The StreetLight data provides the region level bicycle demand analysis, focusing on origins and destinations. To explore the bicycle demand across the region, the area in Tucson was analyzed at block group level. Along the geographic demand analysis, understanding the periodic trend is also important. The illustration of bicycle demand for each time period, AM peak, between peaks, and

PM peak are provided in **Figure 5.11**. StreetLight data has the definitions of time periods as: AM peak: 6:00 AM – 10:00 AM; Midday: 10:00 AM – 3:00 PM; PM peak: 3:00 PM – 7:00 PM. January 2022 was selected for the data period. In this visualization, darker colors indicate a higher volume of trips within the specified block groups. The analysis considers trips starting or ending in predefined geographic areas.

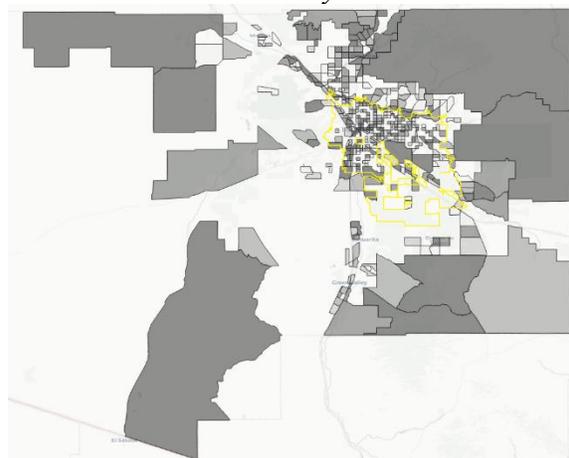
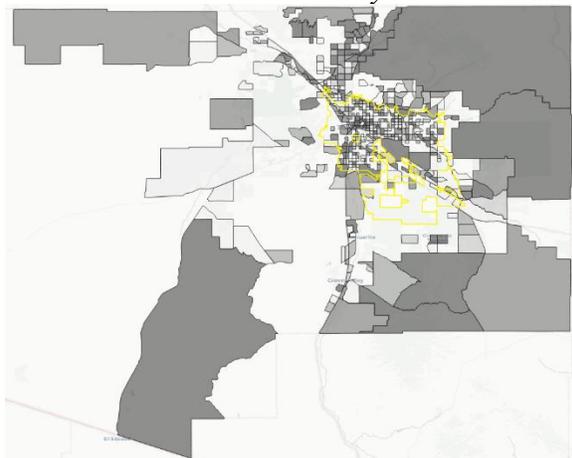
According to the geographic and periodic analysis of bike activity in block group level, **Figures 5.11(A)** and **11(a)** shows low bike traffic in eastern Pima County during the AM peak. The highest bike traffic was observed at the university, downtown, and airport areas. **Figures 5.11(B)** and **5.11(b)** represent between AM peak and PM peak and **Figures 5.11(C)** and **5.11(c)** represent PM peak bike traffic in the City of Tucson. The between peaks and PM peak time periods showed a similar trend with high bike traffic in the city. Compared to the AM peak, the high traffic area was observed at west Tucson area in between peaks and PM peak. The high traffic is not only focused on the downtown, university, and airport areas (block-level total traffic), but also more broadly observed in the city. The origin and destination showed similar trends for each of the time period.



(A) AM peak City of Tucson bike origin activity



(a) AM peak City of Tucson bike destination activity



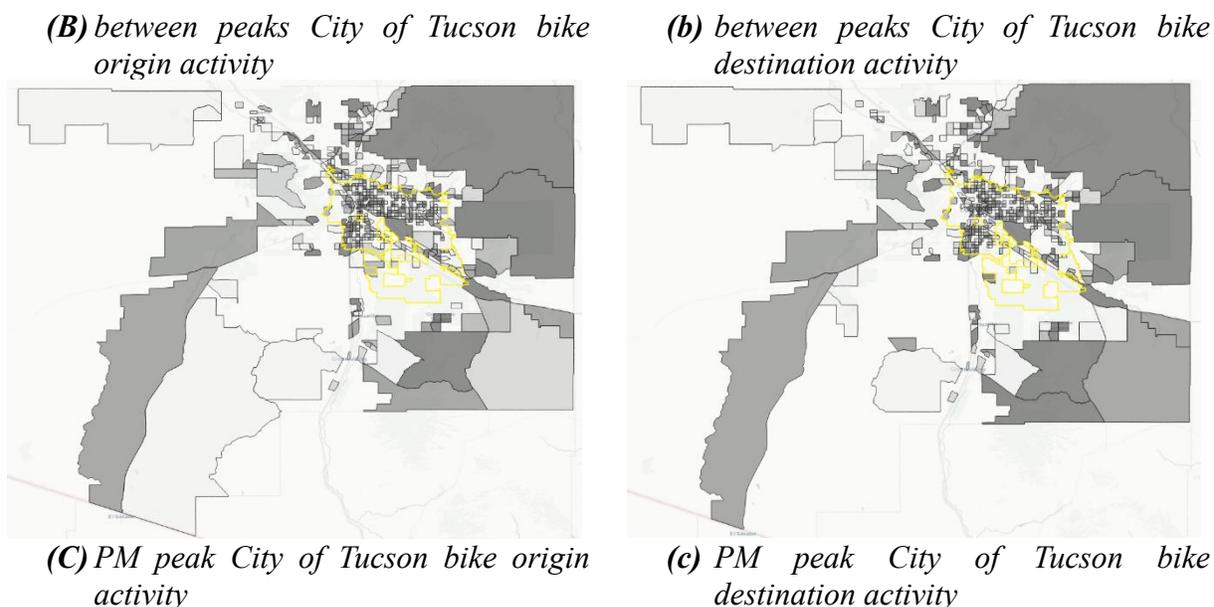


Figure 5.11. Tucson bike activity analysis in January 2022

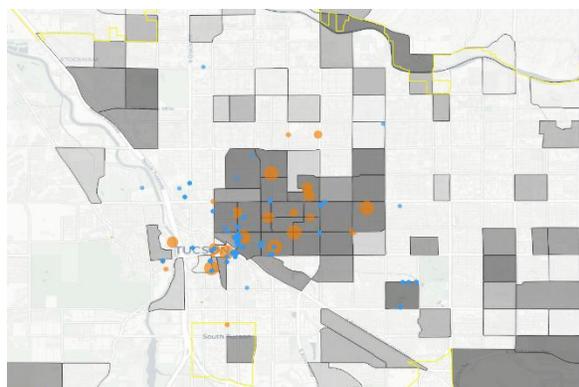
5.3.3.4 Demand, accessibility, and availability analysis

Given that the analysis of micromobility availability, accessibility, and demand has been conducted using the available data, aggregating these results will offer a more comprehensive perspective on the relationship between estimated user demand and the actual operational status.

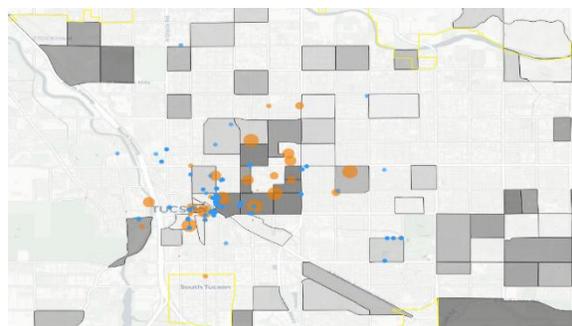
Figure 5.12 synthesizes the analyses conducted on e-scooter and bike share availability, transit station locations as accessibility indicators, and estimated bike demands. It utilizes various visual markers to represent different data points clearly: blue markers indicate the locations where e-scooters are available, while orange circle markers vary in size to represent the availability at each TUGO bike share station. The origin and destination demand, sourced from StreetLight data, is depicted using varying shades of darkness across different Traffic Analysis Zones (TAZs). In this visual representation, darker blocks indicate higher bike origin or destination traffic, as detailed in the figure legend.

The analysis reveals that the highest concentrations of micromobility availability are consistently found in downtown and university areas across all time periods. These areas also exhibit higher estimated origin and destination traffic, suggesting a strong correlation between micromobility supply and demand. Despite the steady availability of bike share stations, there is a noticeable density of e-scooters in the downtown area along 4th Ave., as depicted in **Figures 5.12**. **Figure 5.12** underscores a significant alignment between demand, as indicated by origin-destination data from Streetlight data, and supply, measured by the presence of e-scooters and TUGO bike-share units. This correlation is particularly pronounced in the downtown area, where both demand and supply peaks suggest effective placement and utilization of micromobility services to accommodate user needs. This comprehensive visualization aids in understanding the dynamics of

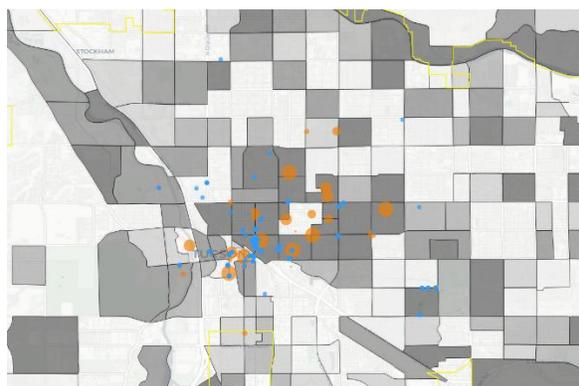
micromobility use and its integration with other transit options, highlighting areas of high activity and potential focal points for further infrastructure development or service adjustments.



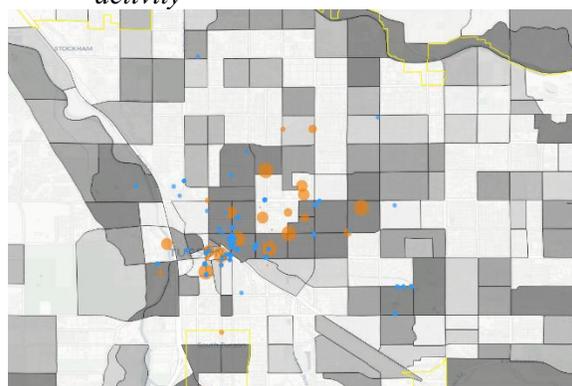
(A) AM peak City of Tucson bike origin activity



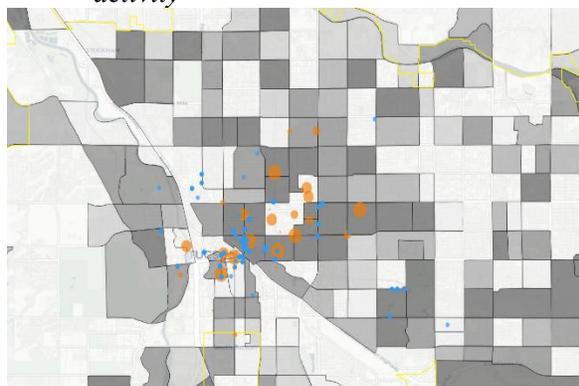
(a) AM peak City of Tucson bike destination activity



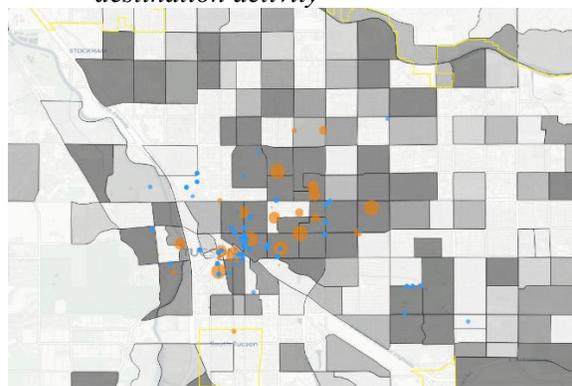
(B) between peaks City of Tucson bike origin activity



(b) between peaks City of Tucson bike destination activity



(C) PM peak City of Tucson bike origin activity



(c) PM peak City of Tucson bike destination activity

E-scooter
Tugo bike share availability

- 2 - 4
- 4 - 5
- 5 - 6
- 6 - 7.2
- 7.2 - 13



Origin / Destination traffic

- 1 - 9
- 9 - 12
- 12 - 18
- 18 - 30
- 30 - 310



Figure 5.12. E-scooter and bike share availability and demand map

CHAPTER 6. Transit Performance Measure Extraction

This chapter is directed to identify and quantify key performance measures for the transit system in the region. As discussed in the report of CHAPTER 3, the General Transit Feed Specification (GTFS) data, provided by Sun Tran, can be used to measure major transit system performance indicators of interest, pertaining to speed, in-vehicle travel time reliability, on-time performance, and headway.



Our methodology leverages both the GTFS Real Time (GTFS-RT) and GTFS Schedule (GTFS-ST) to evaluate performance in two levels: (i) system-wide, encompassing the entire transit network, and (ii) route-specific, focusing on individual transit routes.

6.1. Deliverables

The performance measures that have been explored under this task are outlined in [Table 6.1](#). The table of contents, list of tables and list of figures are outlined in the following page. GTFS-RT Trip Updates are employed for performance measurement, as also used in previous studies by Newmark (2024), Arias et al (2021), Elliott and Lumley (2020), and Wu et al (2023).

For the analysis of performance measures reported in this document, GTFS-RT data archived from January 1st, 2025, to February 8th, 2025, was utilized. A preprocessing was conducted using the Python “*gtfs_segment*” package. This package facilitated the creation of lists of transit system segments, incorporating static data provided by the GTFS-ST, as illustrated in [Figure 6.1](#).

Table 6.1. Transit Performance Measures Summary

Performance Measure	Location in the Report	Description	Used Data
Speed	Section 6.2	Evaluate the overall stop-to-stop segment speed for each hour and route	GTFS-RT: <ul style="list-style-type: none"> • <i>Trip Update</i> GTFS-ST: <ul style="list-style-type: none"> • <i>stops</i> • <i>routes</i> • <i>shapes</i> • <i>trips</i>
In-vehicle Travel Time Reliability	Section 6.3	Evaluate the reliability of travel time once the passenger is on the transit vehicle for each route	
Average Headway	Section 6.4	Measure average headways for each hour and route and across the system	
Headway Irregularity	Section 6.5	Measure the irregularity of headway for each hour and route across the system	
Headway Reliability	Section 6.6	Measure the headway reliability for each hour and route across the system	
On-Time Performance	Section 6.7	Measure the time gap between scheduled and actual arrival time for each route	

The output includes the traveled distance and geospatial information of the transit routes for each stop-to-stop segment as depicted in [Figure 6.2](#). Each stop-to-stop segment is generated using the sequence of consecutive stops for each route. It also identifies the consecutive stop pairs that form each segment for every route in the defined transit system. This segment data is then imported for extracting various performance measures. For the consistency of data analysis time period definition, AM peak is defined from 6:30 AM to 8:30 AM and PM peak is defined from 4:00 PM to 6:00 PM. Detailed descriptions of each performance measure are provided in subsequent sections of this report. Due to the distinct operational characteristics of each transit mode, regular bus, express bus, and light rail streetcar, the performance of each mode was evaluated separately. Sun Tran operates as a frequent transit service from early morning to late evening hours. Sun Express, by contrast, provides a service with limited frequency, operating primarily during morning and evening peak hours, 4:00 AM to 8:00 AM, and 3:00 PM to 7:00 PM, respectively. Sun Link functions as a frequency-based transit service. Rather than adhering to a detailed timetable, Sun Link maintains consistent headways based on time of day and day type, allowing for regular and predictable service intervals.

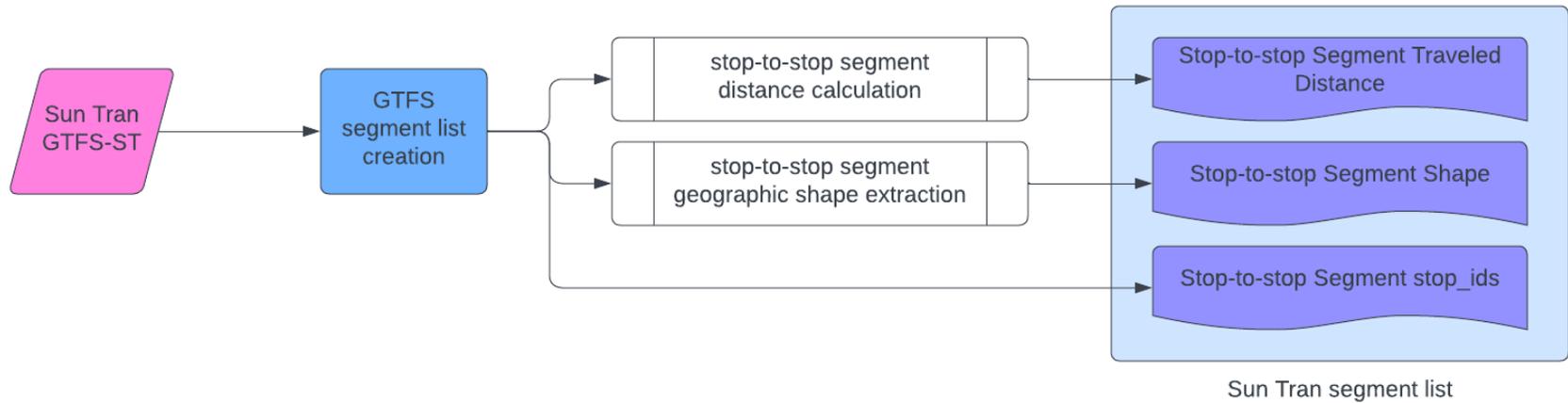


Figure 6.1. GTFS-ST preprocessing flow chart

	segment_id	route_id	trip_id	distance	stop_id1	stop_id2	geometry
0	10858-10859-1	21407	2503341	323.82	10858	10859	LINestring (-110.97085333333334 32.22433233333333,
1	10859-10861-1	21407	2503341	830.74	10859	10861	LINestring (-110.97260193333334 32.22278433333334,
2	10861-13604-1	21407	2503341	730.75	10861	13604	LINestring (-110.96568975 32.221199, -110.9656395000
3	10925-14204-1	21407	2503357	1030.44	10925	14204	LINestring (-110.95988529411764 32.22145429411765,
4	11263-11264-1	21407	2503341	347.26	11263	11264	LINestring (-110.95676416666667 32.239659, -110.9567
5	11264-11265-1	21407	2503341	399.91	11264	11265	LINestring (-110.956799 32.24279153846154, -110.9567
6	11265-11266-1	21407	2503341	480.92	11265	11266	LINestring (-110.95682 32.246399, -110.956821111111
7	11266-11267-1	21407	2503341	382.39	11266	11267	LINestring (-110.956859 32.25073714285714, -110.9568
8	11267-11268-1	21407	2503341	349.43	11267	11268	LINestring (-110.956749 32.25418, -110.9567484705882
9	11268-16169-1	21407	2503341	465.76	11268	16169	LINestring (-110.95673238095237 32.25733233333333,
10	11272-11273-1	21407	2503341	399.1	11272	11273	LINestring (-110.9393278 32.25758, -110.93928 32.2575
11	11273-11274-1	21407	2503341	407.48	11273	11274	LINestring (-110.93509119512196 32.257629, -110.9350
12	11274-11275-1	21407	2503341	417.59	11274	11275	LINestring (-110.93076522058824 32.25765127941177,
13	11275-11276-1	21407	2503341	330.8	11275	11276	LINestring (-110.9263325 32.25769475, -110.92628075 3

Figure 6.2. GTFS segment list data example

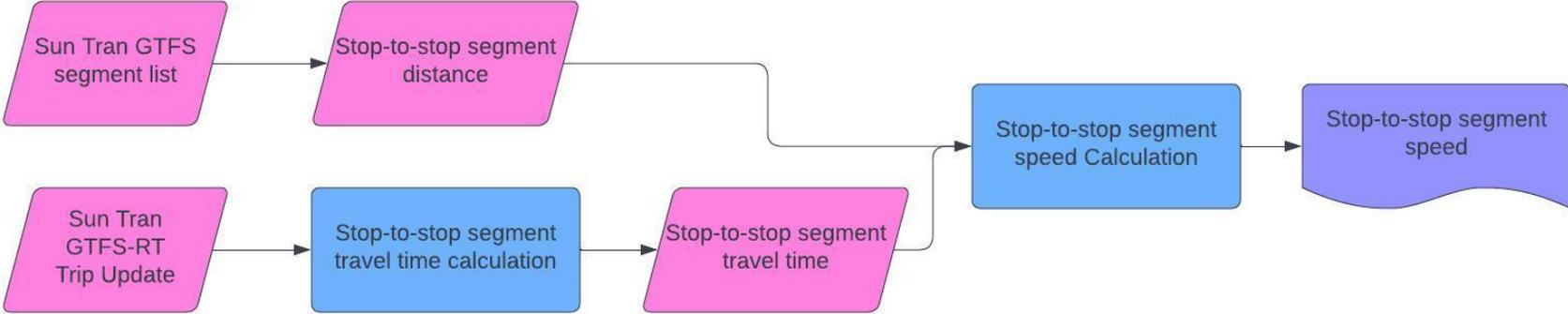


Figure 6.3. Figure 3. Speed performance measure flow chart

6.2. Speed

To evaluate the efficiency of the transit system from the perspective of the passengers, speed performance of transit vehicles was analyzed for each stop-to-stop segment (Carrasco, 2012). This method involves calculating the travel time by subtracting the departure time at a stop from the departure time at the subsequent stop using GTFS-RT Trip Update data. This calculated travel time includes the dwelling time of the vehicle at each stop. The distance between consecutive stops was derived from a preprocessed segment list. The speed for each segment was then calculated by dividing the distance by the corresponding travel time. The methodology for these calculations is depicted in [Figure 6.3](#).

As Sun Tran's regular service, express service, and Sun Link operate on different schedules and scheduling methods, each mode of transit system's performance was measured separately. As compared to Sun Tran's regular service, Sun Express is a less frequent service which operates only during AM and PM peak periods. Sun Express' AM operation hours are mostly between 4:00 AM to 8:00 AM and PM operation hours are mostly between 3:00 PM to 7:00 PM. Differing from both, Sun Link is a frequency-based service operating with fixed headways for each time and day types.

6.2.1 *Sun Tran System level speed performance measure*

To provide a comprehensive understanding of high-level transit operations, a system-level vehicle speed boxplot was utilized for evaluation. This visualization effectively summarizes the distribution of vehicle speeds and highlights outliers, enabling comparison across different times of day. As shown in [Figure 6.4](#), the upper panel presents the distribution of overall vehicle speeds in the Sun Tran system by hour, while the lower panel displays the number of observations per hour, providing context on data availability. The boxplot reveals that median vehicle speeds generally range between 13 mph and 18 mph throughout the day. Speeds tend to be higher before the AM peak and after the PM peak, with medians around 16–17 mph, in contrast to the slightly lower values of 14–15 mph observed during daytime hours (approximately 6:30 AM to 6:00 PM).

The mean speeds, indicated by the blue lines in the boxplot, follow a similar pattern but are consistently slightly lower than the medians across most hours. This difference implies the presence of lower-speed outliers, which pull the mean down while having less influence on the median. The greatest divergence between mean and median is visible during the midday and PM peak hours, indicating increased variability in operating conditions during those times. The accompanying bar chart reinforces these observations by showing lower data volumes before 6:00 AM and after 6:00 PM, consistent with reduced service levels.

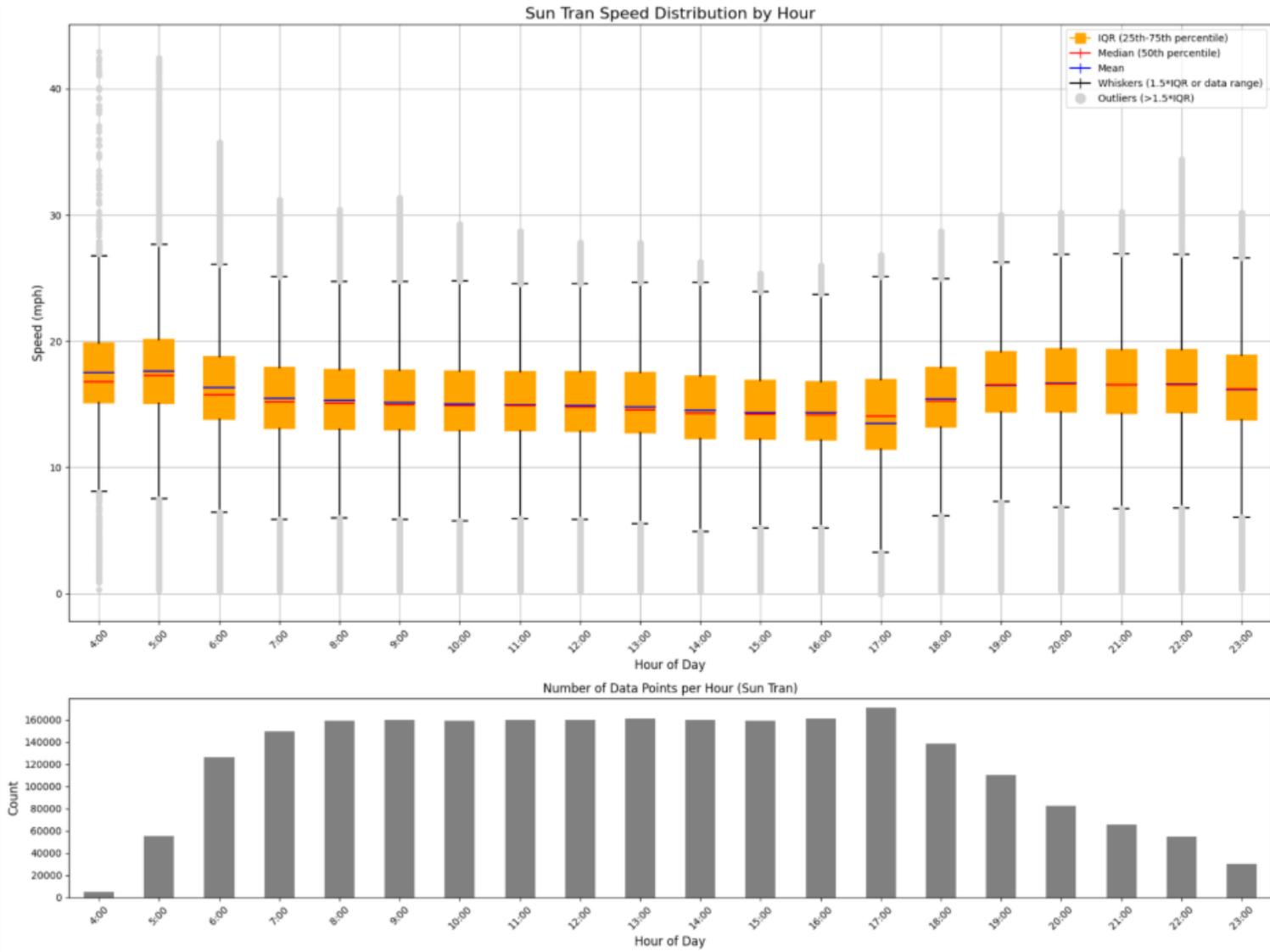


Figure 6.4. Sun Tran System level transit vehicle speed boxplot

6.2.2 *Sun Tran route level speed performance measure*

To further investigate transit operating performance, a route-level analysis of vehicle speed was conducted. **Figure 6.5** presents a heatmap visualizing the mean speed for each Sun Tran route by hour of the day. Darker green shades represent higher speeds, while lighter green indicates lower speeds. Gray cells denote missing or unavailable data for specific route-hour combinations. The x-axis displays operating hours from 4:00 AM to 11:00 PM, and the y-axis lists the 29 Sun Tran routes by their short and long names, consistent with GTFS-ST route designations.

Overall, most routes exhibit mean speeds ranging from 10 to 15 mph during daytime operations, particularly between the AM and PM peak periods. In contrast, speeds tend to be 15–25 mph before the AM peak (prior to 6:30 AM) and after the PM peak (after 6:00 PM), indicating improved travel conditions during off-peak hours.

The heatmap reveals two major speed patterns. Most routes show a U-shaped curve, with higher speeds in the early morning and evening, and lower speeds from the AM peak through the PM peak, due to increased activity on the road during the day, while a few routes, such as Route 6 and Route 18, experience consistent speeds across all hours of operations.

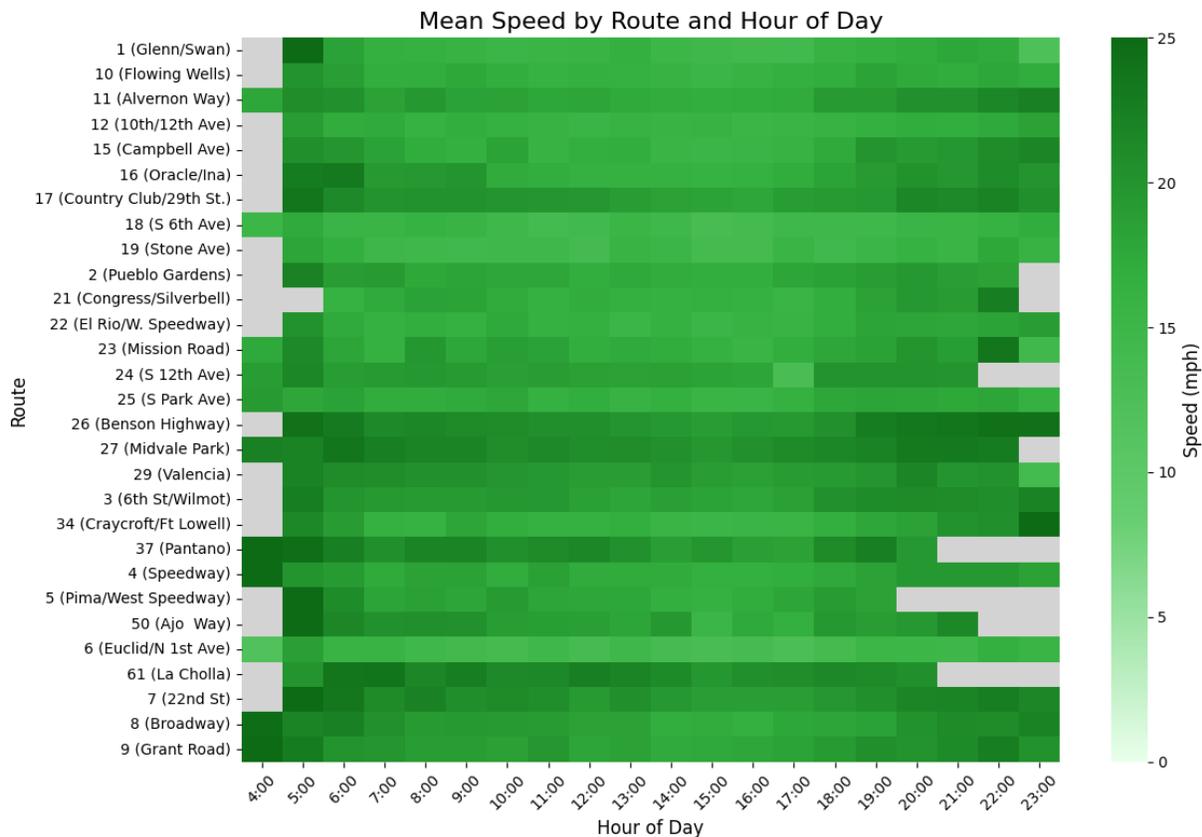


Figure 6.5. Sun Tran Route level speed performance heatmap

6.2.3 Sun Express System level speed performance measure

Figure 6.6 presents the system-level speed distribution of the Sun Express transit service using a segment-based speed calculation method. The boxplot illustrates the variability in bus speeds across different operational hours.

During AM operation hours, the analysis shows that the median speed generally exceeds the mean speed, indicating potential skewness in the data due to lower-speed trip records. Over the AM period, speeds steadily decline, likely reflecting increased roadway congestion during this time.

During the PM peak hours, the opposite pattern is observed, with slower speeds in the early peak, and improving over time. At 7:00 PM, the speed drops, due to express routes reaching the slower endpoints of their routes and ending service, leading to fewer datapoints in general, and a larger fraction of low-speed datapoints.

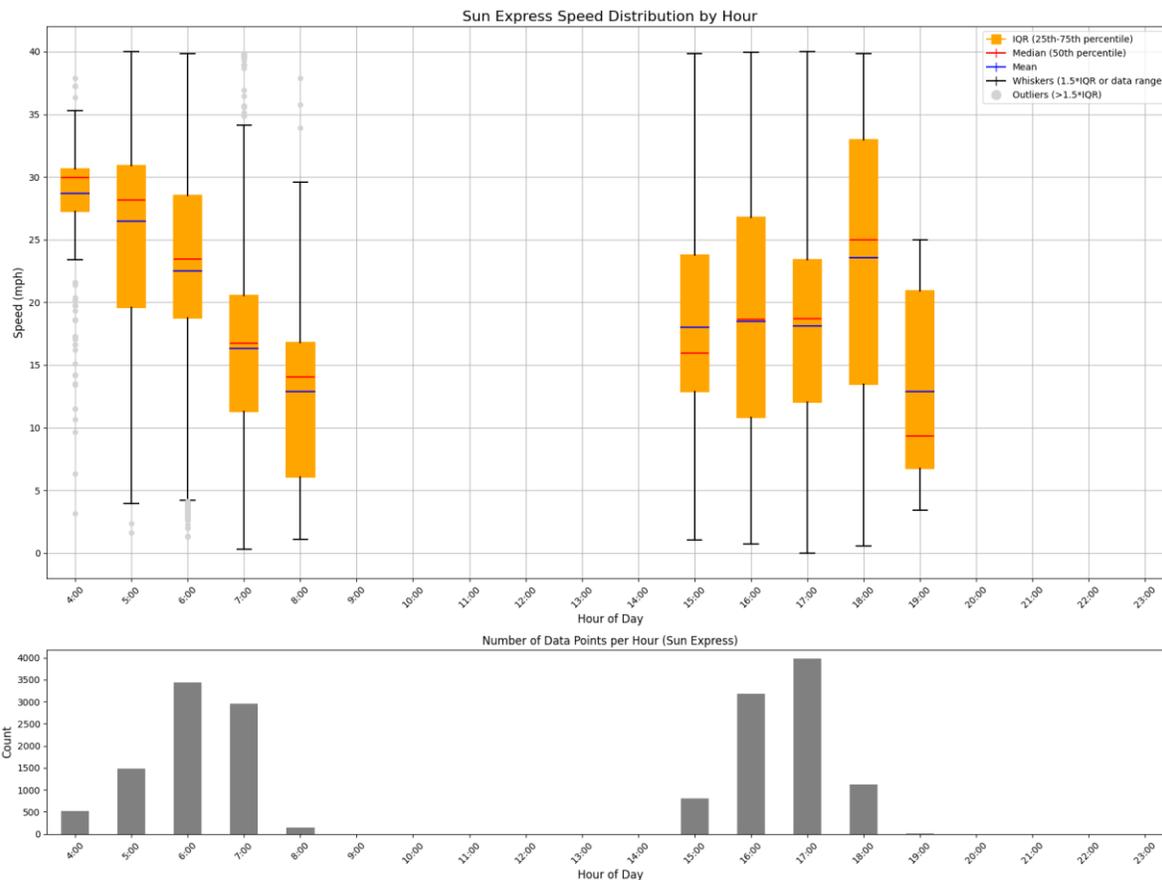


Figure 6.6. Sun Express System level transit vehicle speed boxplot

6.2.4 Sun Express route level speed performance measure

To further delve into express bus speeds, analysis was conducted at the individual Sun Express route level. [Figure 6.7](#) presents a heatmap of route level median speeds derived from segment-level speed calculations. The sharp decreases in the median speed generally align with portions of routes in downtown or denser portions of the route, where there are higher numbers of shorter and slower segments

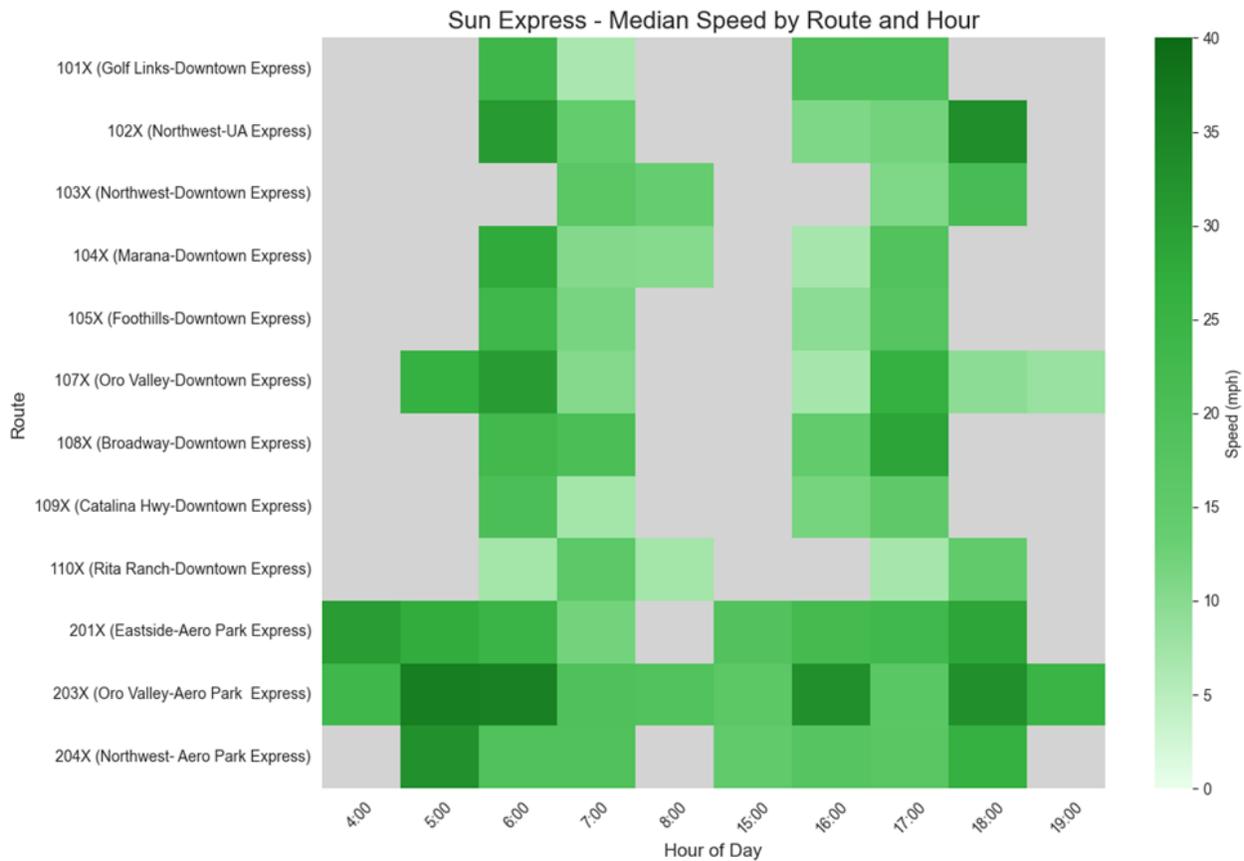


Figure 6.7. Segment-weighted Sun Express route level transit vehicle speed heatmap

6.2.5 Sun Link System level speed performance measure

Figure 6.8 presents the system-level speed distribution of the Sun Link streetcar service, derived from segment-level speed calculations. The boxplot illustrates how vehicle speeds vary by hour of the day, covering Sun Link’s operating period from 6:00 AM to 11:00 PM. A corresponding bar chart below the boxplot shows the number of data points collected per hour, providing context on data availability and volume throughout the day. Sun Link is extremely consistent, with an average speed of about 7 mph and a median speed of about 10 mph throughout its operating period.

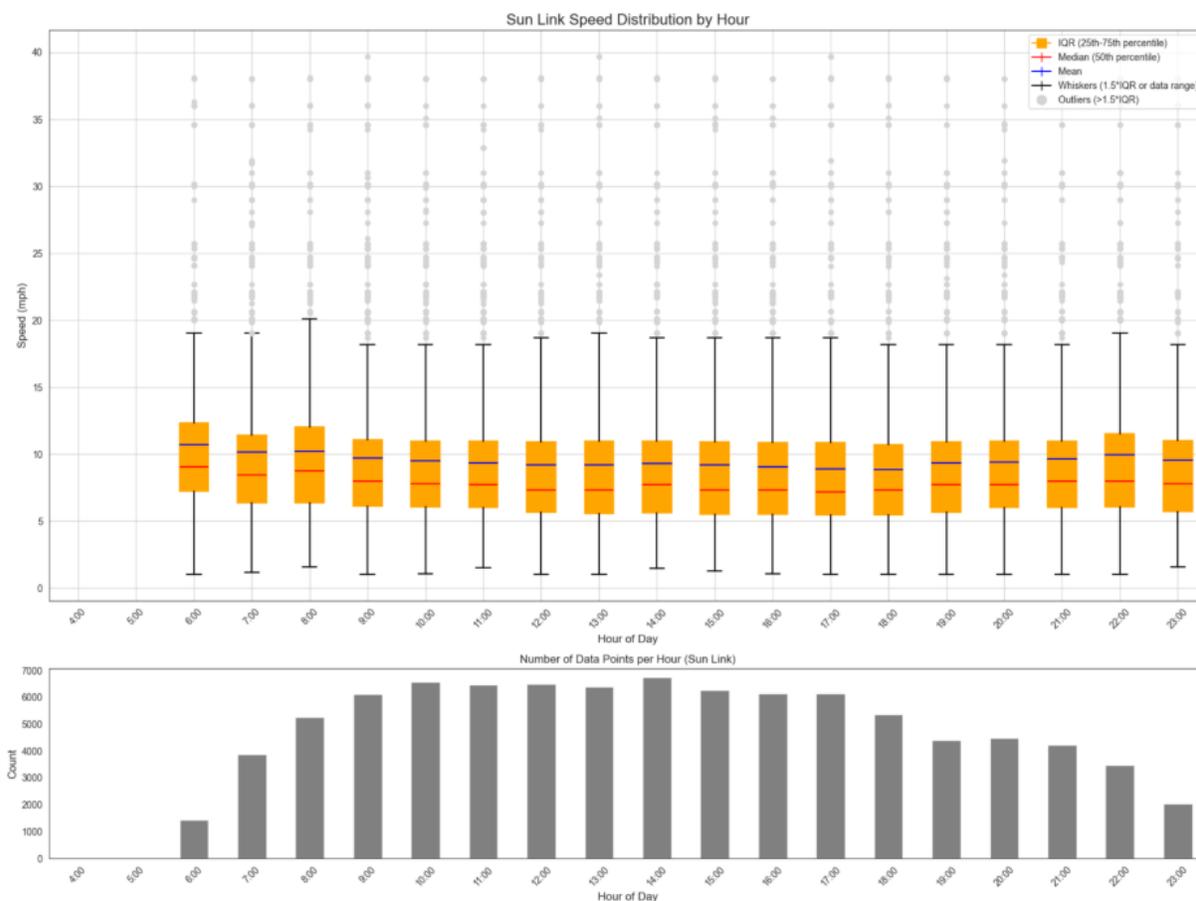


Figure 6.8. Sun Link system level transit vehicle speed boxplot

6.2.6 Sun Link stop level speed performance measure

To further examine the operational characteristics of streetcar service, a stop-level speed analysis was conducted for the Sun Link stop. **Figure 6.9** presents a detailed heatmap visualization of stop-level speeds across different stops. Since Sun Link operates a single route, this granular analysis provides valuable insight into how speed patterns vary by segment and time of day.

Across the Sun Link route, most segments maintain fairly consistent median speeds across all operating hours, with only minor decreases in median speeds during busier daytime or peak periods.

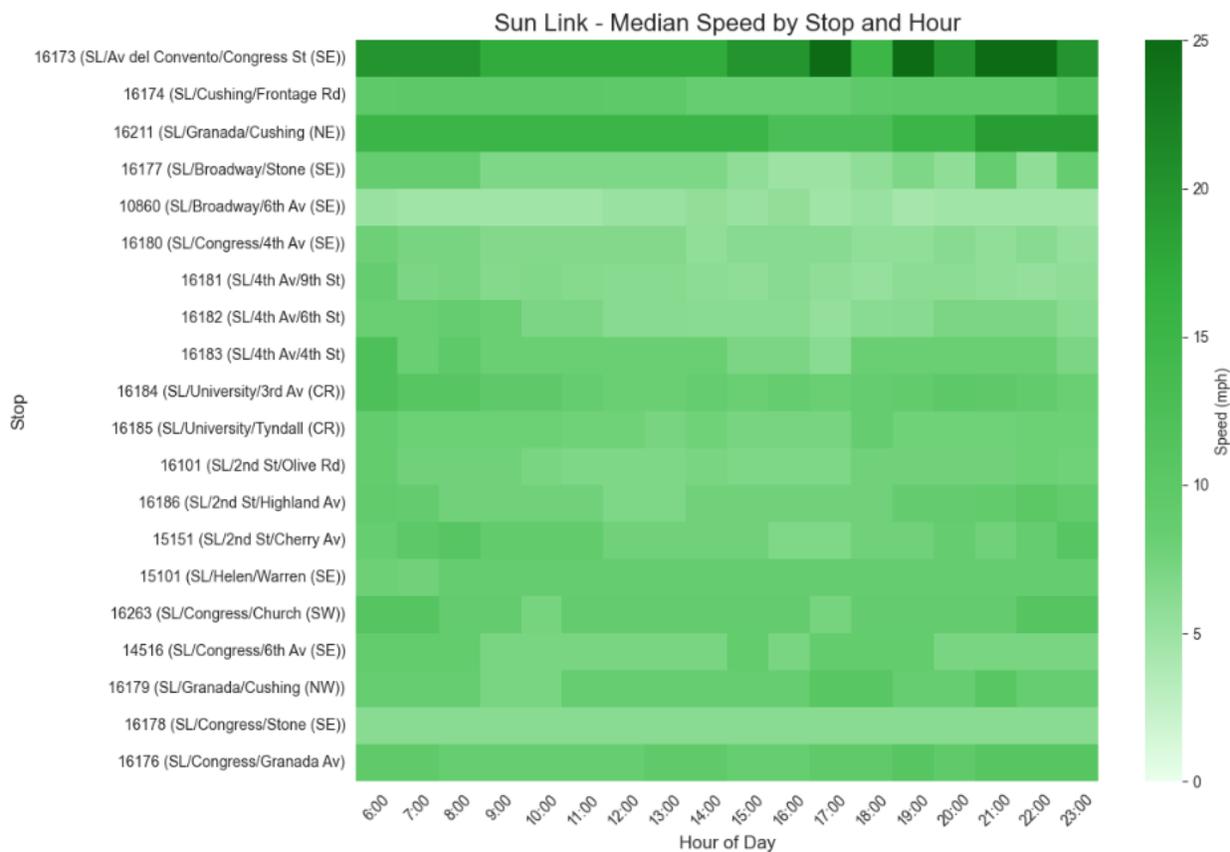


Figure 6.9. Sun Link stop level transit vehicle speed boxplot

6.3. In-vehicle Travel Time Reliability

To assess travel time reliability from the passengers' perspective once onboard, we measure the consistency of travel times between stops. Travel time is calculated by subtracting the departure time at one stop from the departure time at the next, including the vehicle's dwelling time at each stop. Travel time reliability reflects the difference between extreme, 95th percentile, and average travel time that the passengers experienced. To quantify this reliability, the Buffer Time Index (BTI) is referred (Harsha, M. M. & Mulangi, R. H., 2021).

$$R = 1 - \frac{tt_{95} - tt_{mean}}{tt_{mean}} = \begin{cases} \geq 0 \text{ and } < 1, \text{ if } tt_{mean} < tt_{95} \leq 2tt_{mean} \\ < 0, \text{ if } tt_{95} > 2tt_{mean} \end{cases}$$

In which:

tt_{95} : 95th percentile of stop-to-stop segment travel time for each hour

tt_{mean} : mean of stop-to-stop segment travel time for each hour

This metric evaluates the disparity between the 95th percentile and the average travel time, illustrating the extent of variability passengers may experience. The methodology for data collection and the calculation process from importing segment list data and GTFS-RT Trip Update data to calculate travel time reliability separately for the system level and individual route level is illustrated in [Figure 6.10](#). For the system level analysis, the 95th percentile and mean travel times are calculated hourly across the entire Sun Tran system. For the route level analysis, these same metrics are computed hourly for each individual route. When R equals 1, the 95th percentile and the mean travel time are identical, indicating highest reliable service. When R equals 0, the 95th percentile is twice the mean value, reflecting low reliability. When R is less than 0, the 95th percentile exceeds twice the mean, showing low reliability in travel time.

To make the reliability measure easier to interpret, R is then normalized between 0 and 1 as below, with 0 corresponding to the cases where tt_{95} is larger than or equal to $4 \times tt_{mean}$ and 1 corresponding to the case of tt_{95} being equal to the tt_{mean} (i.e., being most reliable).

$$R^{normalized} = \frac{R - R_{min}}{R_{max} - R_{min}}$$

Where:

$R^{normalized}$: Normalized travel time reliability index

R_{min} : Minimum travel time reliability ($R_{min} = -3$ for all $tt_{95} \geq 4 \times tt_{mean}$)

R_{max} : Maximum travel time reliability ($R_{max} = 1$)

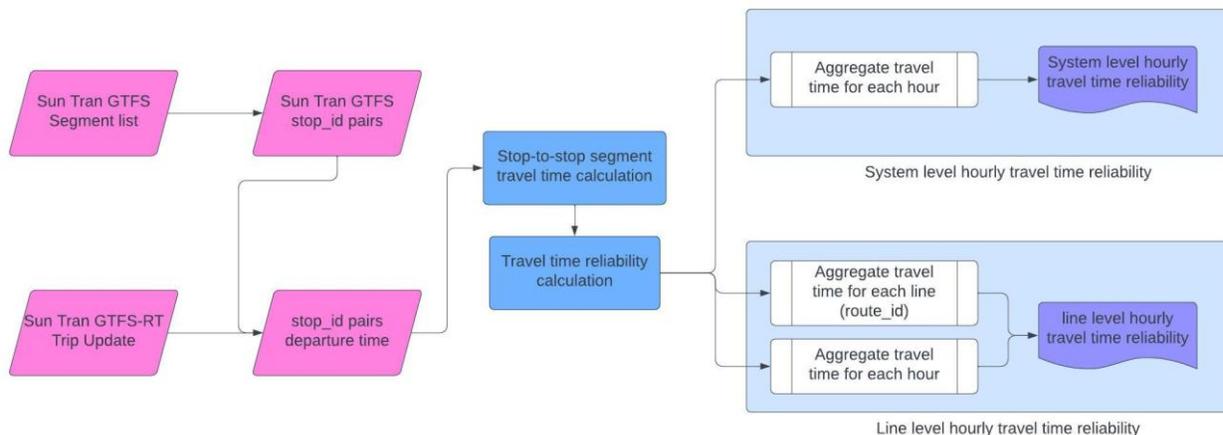


Figure 6.10. Travel time reliability performance measure flow chart

6.3.1 Sun Tran system-level in-vehicle travel time reliability performance measure

Figure 6.11 shows the normalized travel time reliability ($R^{normalized}$) of Sun Tran buses against each hour of the day. The $R^{normalized}$ values are represented by Box-and-Whisker plots in each hour of the day. Across all plots within the chart, all $R^{normalized}$ mean and median values are noted to be higher than 0.9 throughout all hours. This suggests that the Sun Tran system provides consistent travel times, having high reliability in travel times. Another observation is that the $R^{normalized}$ mean values across all hours are lower than $R^{normalized}$ median values, thus indicating the influence of lower reliability of incidents which appears as outliers in the plot.

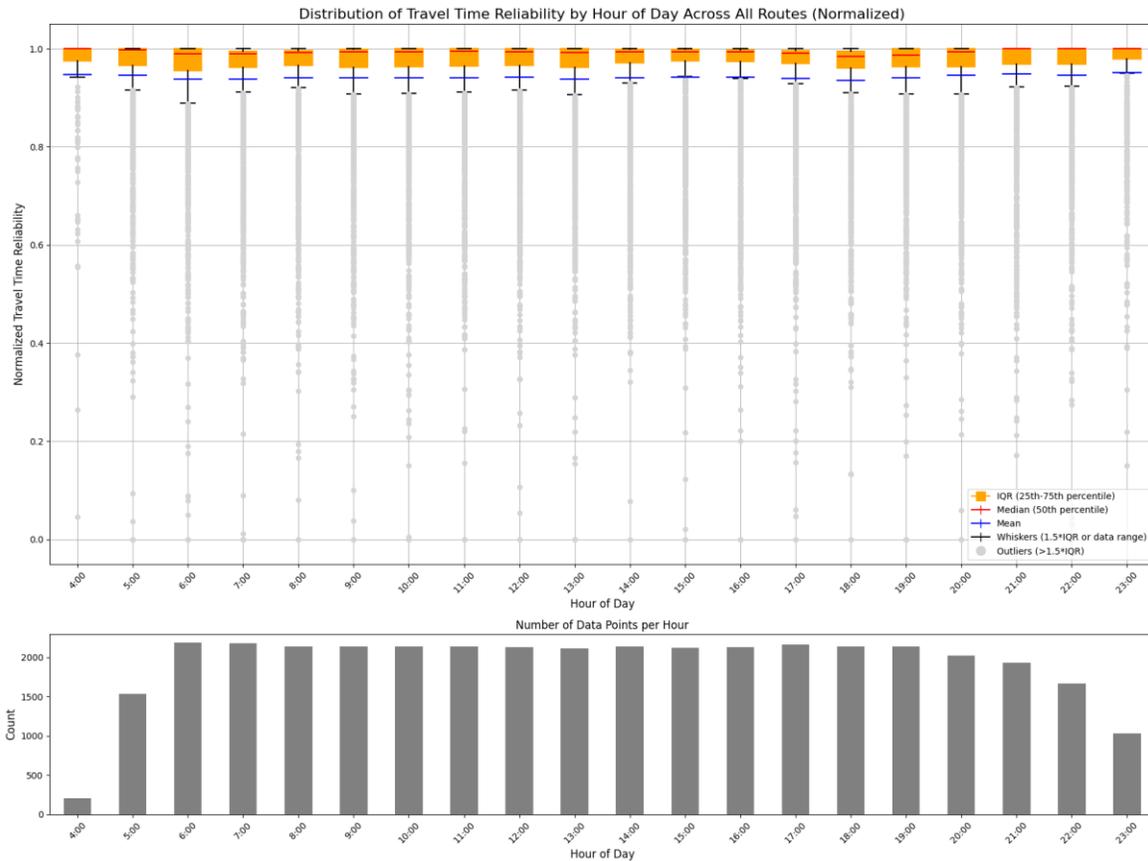


Figure 6.11. Sun Tran System level in-vehicle travel time reliability boxplot

6.3.2 Sun Tran route level in-vehicle travel time reliability performance measure

To further examine travel time reliability patterns, the analysis was extended to the route level, with normalized results visualized in the heatmap shown in [Figure 6.12](#). The analysis shows that reliability is consistently high across all routes and hours of operation.

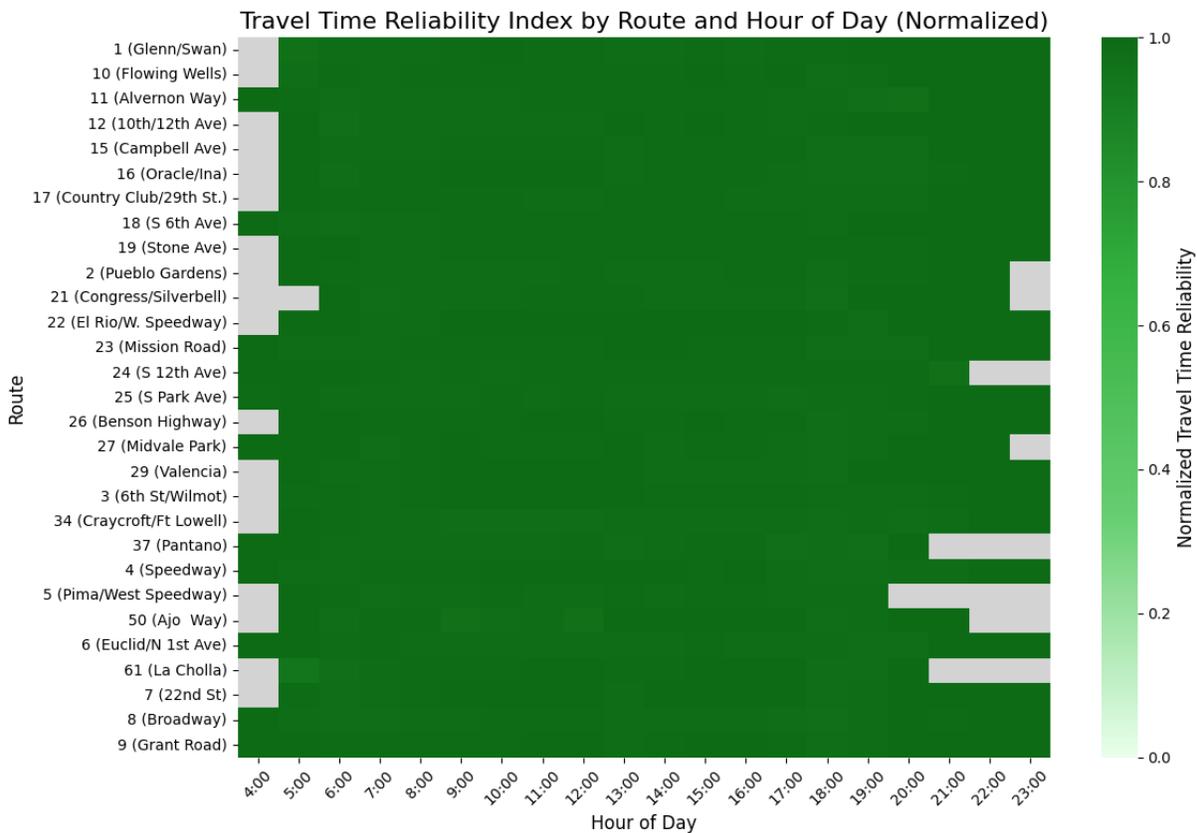


Figure 6.12. Normalized Sun Tran route level in-vehicle travel time reliability heatmap

6.3.3 Sun Express System level travel time reliability performance measure

Figure 6.13 shows a boxplot of the normalized travel time reliability ($R^{normalized}$) of Sun Express routes by hour. To complement this, a corresponding bar chart illustrates the volume of data available for each hour, offering context regarding the number of vehicle observations analyzed.

Sun Express routes are shown to have consistently high travel time reliability, with median values above 0.90 and mean values between 0.80 and 0.95 across all operating hours. The lower mean values and long tails on the boxplot indicate some low-reliability outlier points, which may be due to atypical operating conditions at certain points and times in the analysis period.

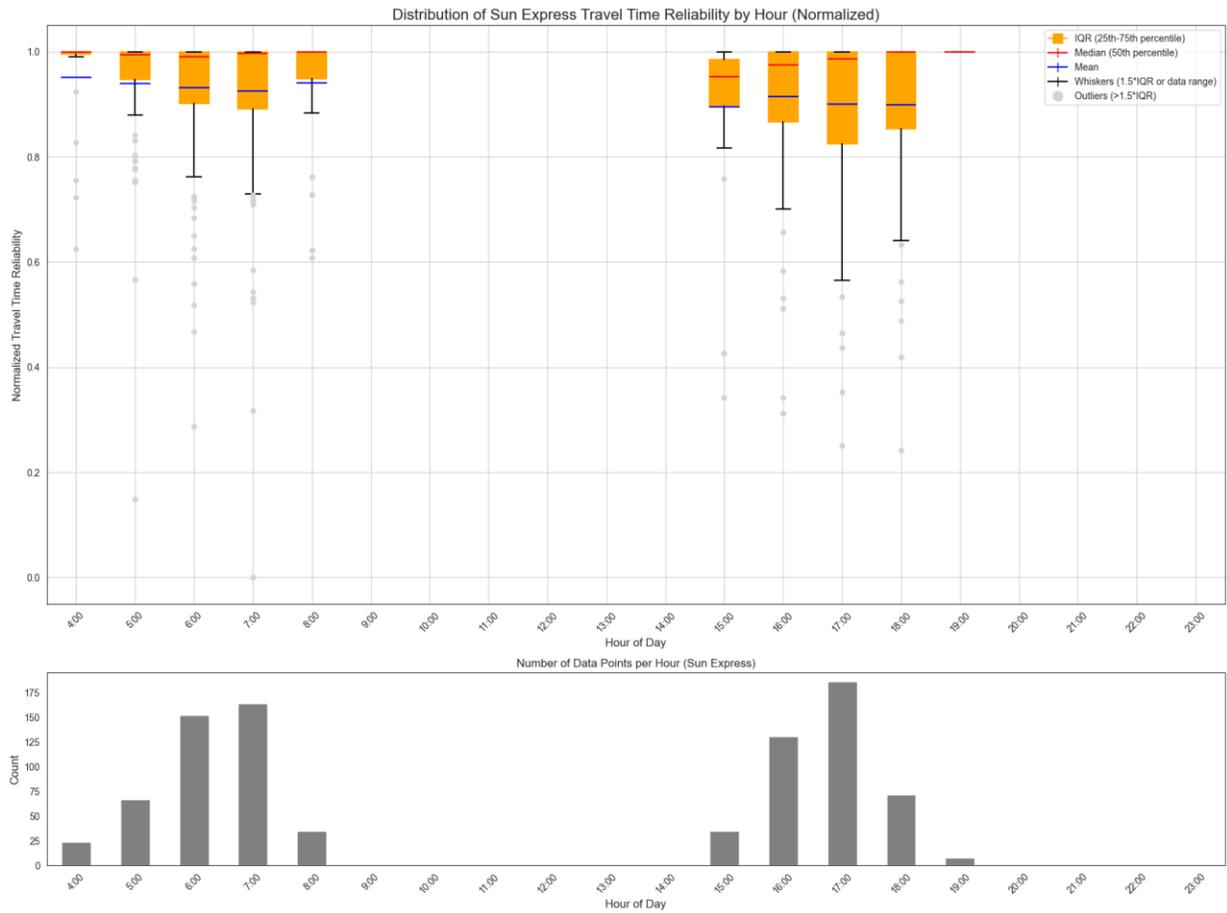


Figure 6.13. Sun Express System level in-vehicle travel time reliability boxplot

6.3.4 Sun Express route level travel time reliability performance measure

To further examine travel time reliability within the Sun Express system, a detailed route level analysis was conducted. **Figure 6.14** presents the distribution of normalized travel time reliability for each express route across different times of day. Using this metric, all routes show consistently high performance during their operating periods.

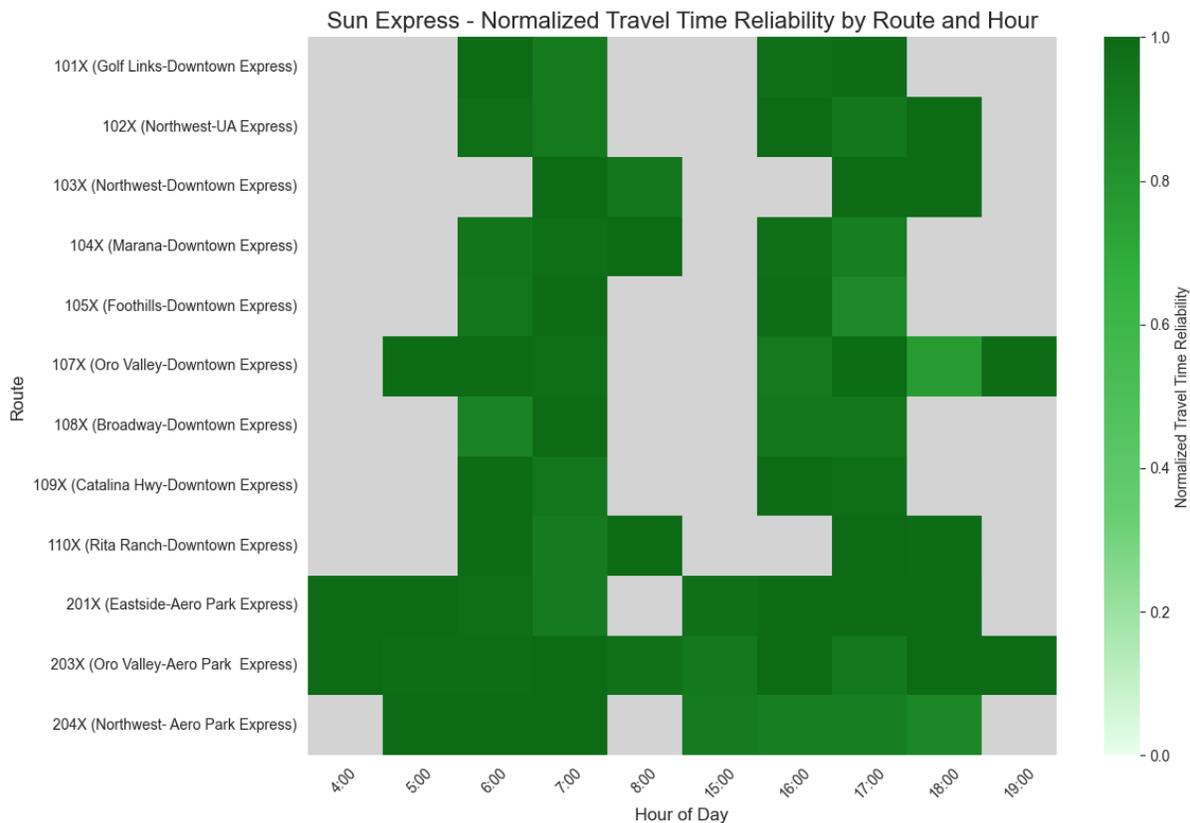


Figure 6.14. Sun Express route level in-vehicle travel time reliability heatmap

6.3.5 Sun Link System level travel time reliability performance measure

Figure 6.15 shows a boxplot of the normalized system level travel time reliability within the SunLink system across its operating period. Accompanying the boxplot, a bar chart details the volume of data available for each hour, offering insights into the number of vehicles analyzed during these intervals.

To assess the overall reliability of travel time within the Sun Link system, a normalized system level travel time reliability was analyzed. Both the mean and median values generally range between 0.8 and 0.9, indicating overall high reliability. At 4:00 PM and 10:00 PM, there is increased variability in normalized reliability, the former likely due to increased congestion in and the latter possibly due to changes in route service in the late evening.

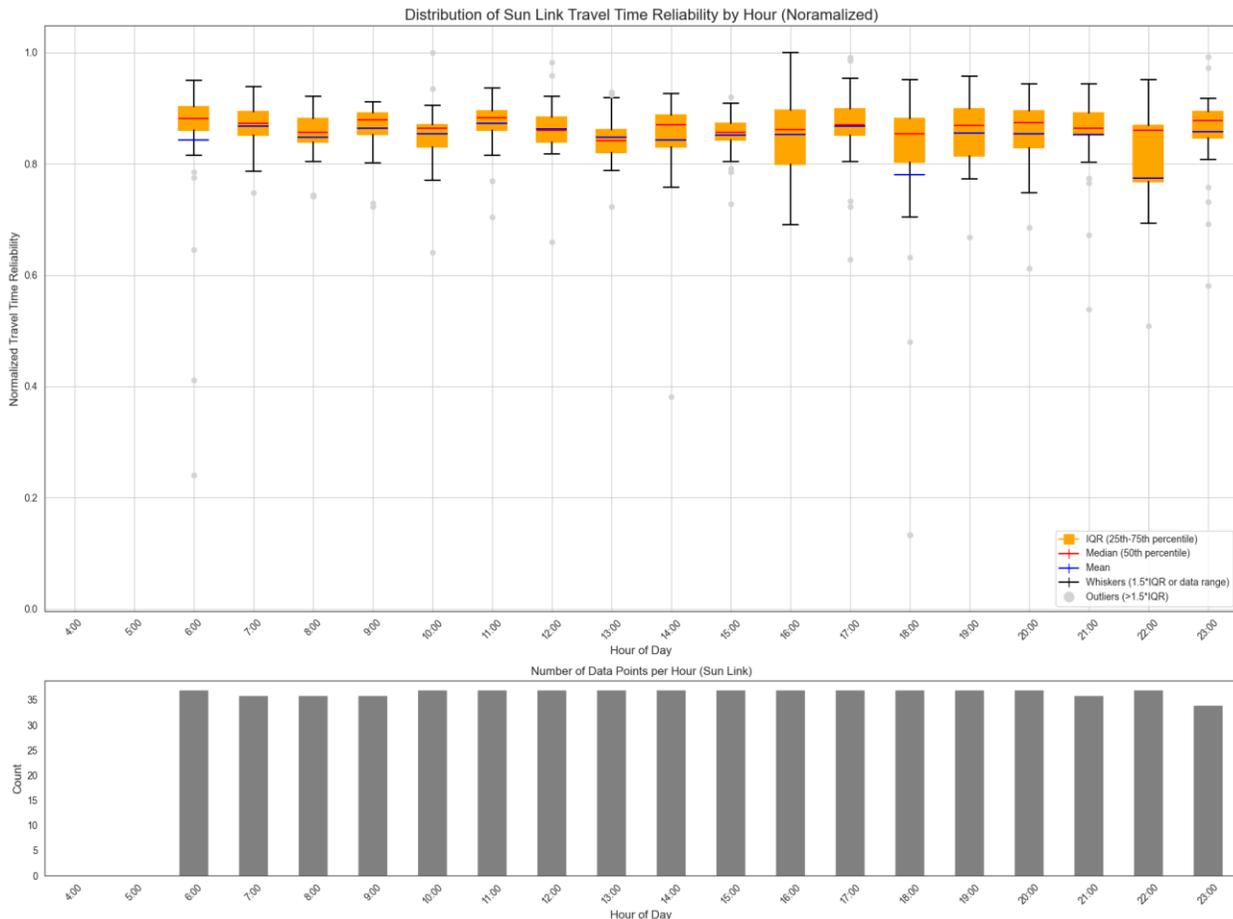


Figure 6.15. Sun Link system level in-vehicle travel time reliability boxplot

6.3.6 Sun Link stop level travel time reliability performance measure

To further examine the Sun link system, a stop-level analysis of travel time reliability was conducted. **Figure 6.16** presents a detailed heatmap visualization of normalized travel time reliability by stop and hour. The overall observed range of $R^{normalized}$ is 0.8 to 1.0, showing high reliability overall operating hours, though the downtown portion of the Sun Link routes have slightly lower reliability at most hours of the day, possibly indicating the ongoing impact of traffic in that area.

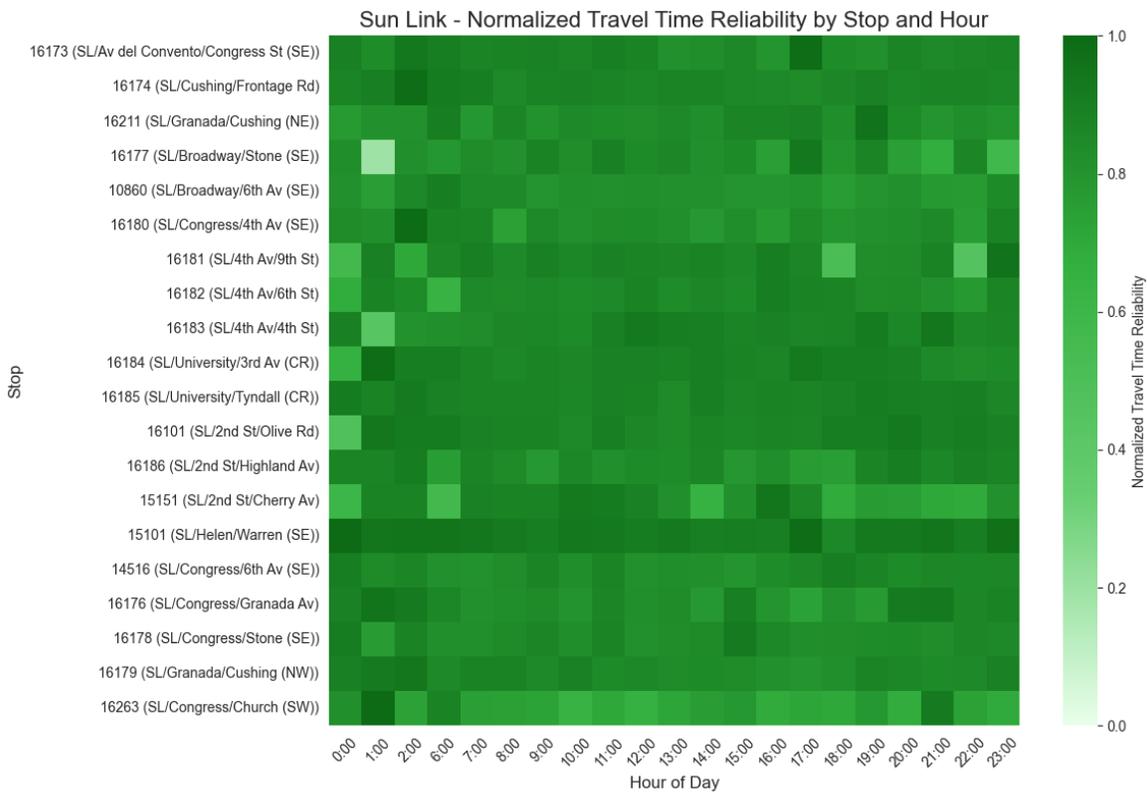


Figure 6.16. Sun Link stop level in-vehicle travel time reliability heatmap

6.4. Headways

Transit headways indicate [Figure 6.17](#) the expected waiting time for the passengers who randomly come to a transit stop. As detailed in CHAPTER 3, headways were calculated by measuring the time difference between two consecutive vehicles at each stop. For a more accurate analysis, data from vehicles operating outside of scheduled service hours and the last vehicle of each operational day were excluded. This approach ensures that the analysis reflects typical service conditions and provides a reliable representation of average waiting times.

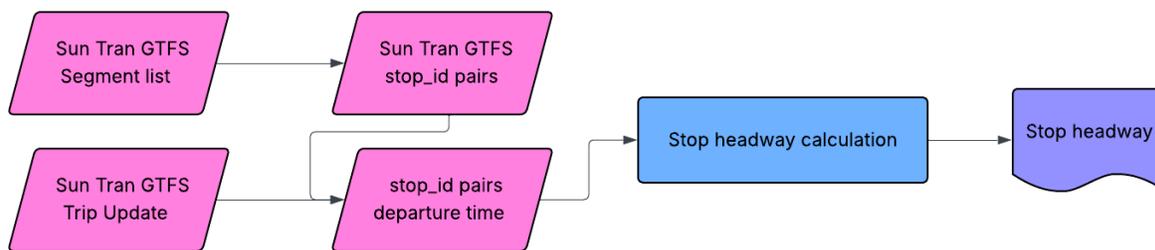


Figure 6.17. Headways performance measure flow chart

6.4.2 Sun Tran System level headways performance measure

To assess the distribution of headways across the Sun Tran transit system, actual headways were analyzed for each operational hour and visualized using a boxplot. This figure captures the variability in service frequency through interquartile ranges (IQR), whiskers, and outliers, providing insight into the consistency headway consistencies throughout the day.

As depicted in [Figure 6.18](#), median headways generally range between 25 and 35 minutes for most hours, reflecting the daytime frequency of most Sun Tran routes. In the early morning and early evening hours, median headways are higher, reflecting the lower frequency of some routes during those time periods. In the late evening, the lower frequency routes end service while higher frequency routes continue to operate, which causes a reduction in the median headway distribution.

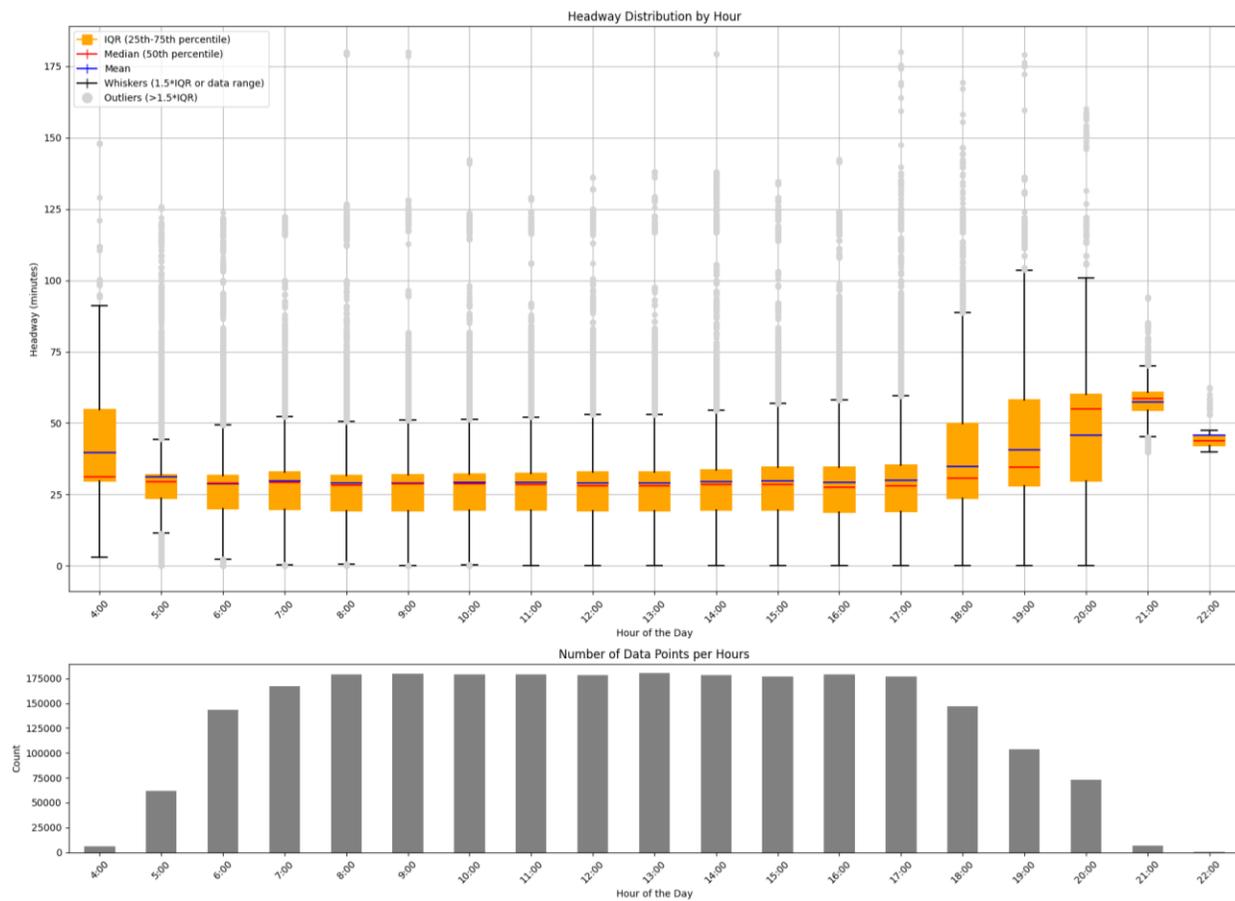


Figure 6.18. Sun Tran system level headways boxplot

6.4.3 Sun Tran route level headways performance measure

To enhance understanding of Sun Tran’s headway performance, a detailed route level analysis of median headways was conducted across all operational hours. **Figure 6.19** displays the results in the form of a heatmap, where darker red indicates longer median headways and lighter red denotes shorter intervals. The use of median values better reflects typical service frequencies for each route and hour.

The analysis shows that the median headways on all routes have only minor variation throughout the day, only shifting significantly at route frequency changes, such as early morning shifts to daytime frequencies, or the shift from daytime frequencies to evening or late-evening frequencies.

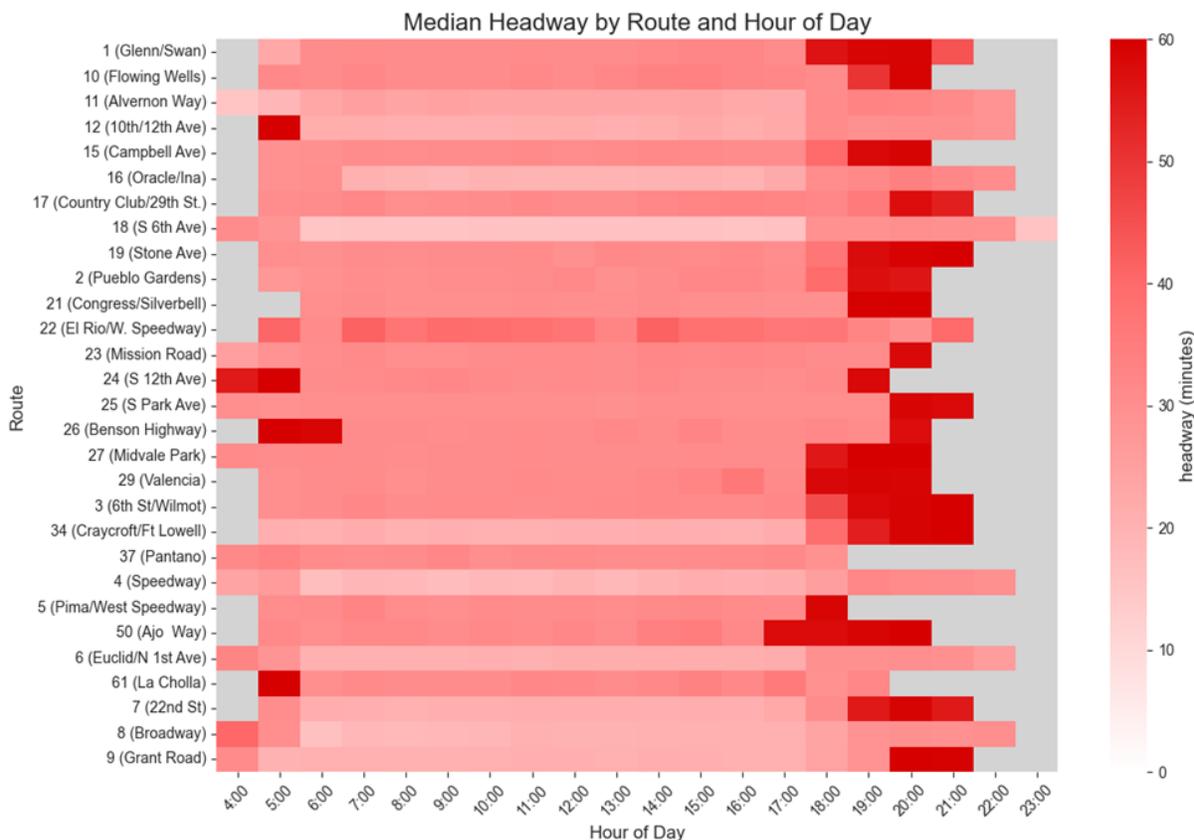


Figure 6.19. Sun Tran route level headways heatmap

6.4.4 Sun Link System level travel time reliability performance measure

Figure 6.20 shows a boxplot representation of system-level headways for Sun Link across its hours of operation. The data indicates that Sun Link has regular headways of 10-15 minutes, consistent with its scheduled service frequency. Some extreme outliers in the boxplot could be due to roadway incidents, such as parked cars blocking the streetcar track

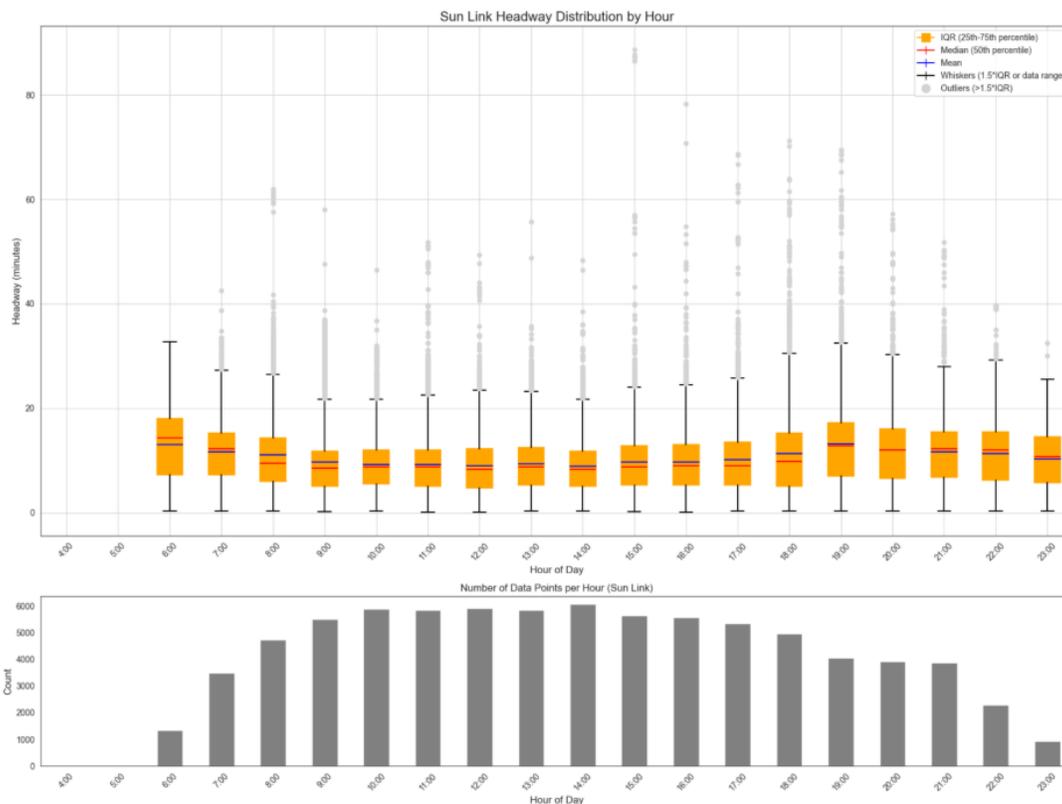


Figure 6.20. Sun Link system level headways boxplot

6.4.5 Sun Link stop level headway performance measure

To gain deeper insights into the operational performance of the Sun Link streetcar, a stop-level headway analysis was conducted for both directions of travel. Figure 6.21 and Figure 6.22 present heatmaps that illustrate the hourly headways at each stop for Westbound and Eastbound, respectively.

For both directions, headways largely reflect the Sun Link scheduled frequencies of 10, 15, or 30 minutes, depending on time of day, with only minor variations across stops. Exceptions to this trend are most notable in the westbound direction at 12PM, 4PM, and 8PM, and localized in the downtown area at 2PM and 5PM. Eastbound exceptions occur at 4PM and 7PM and localized in the 4th Ave/University area at 12PM and 8PM. At these times and segments, headways are longer than scheduled, indicating slower streetcar speeds, possibly due to increased vehicle or pedestrian traffic in the area.

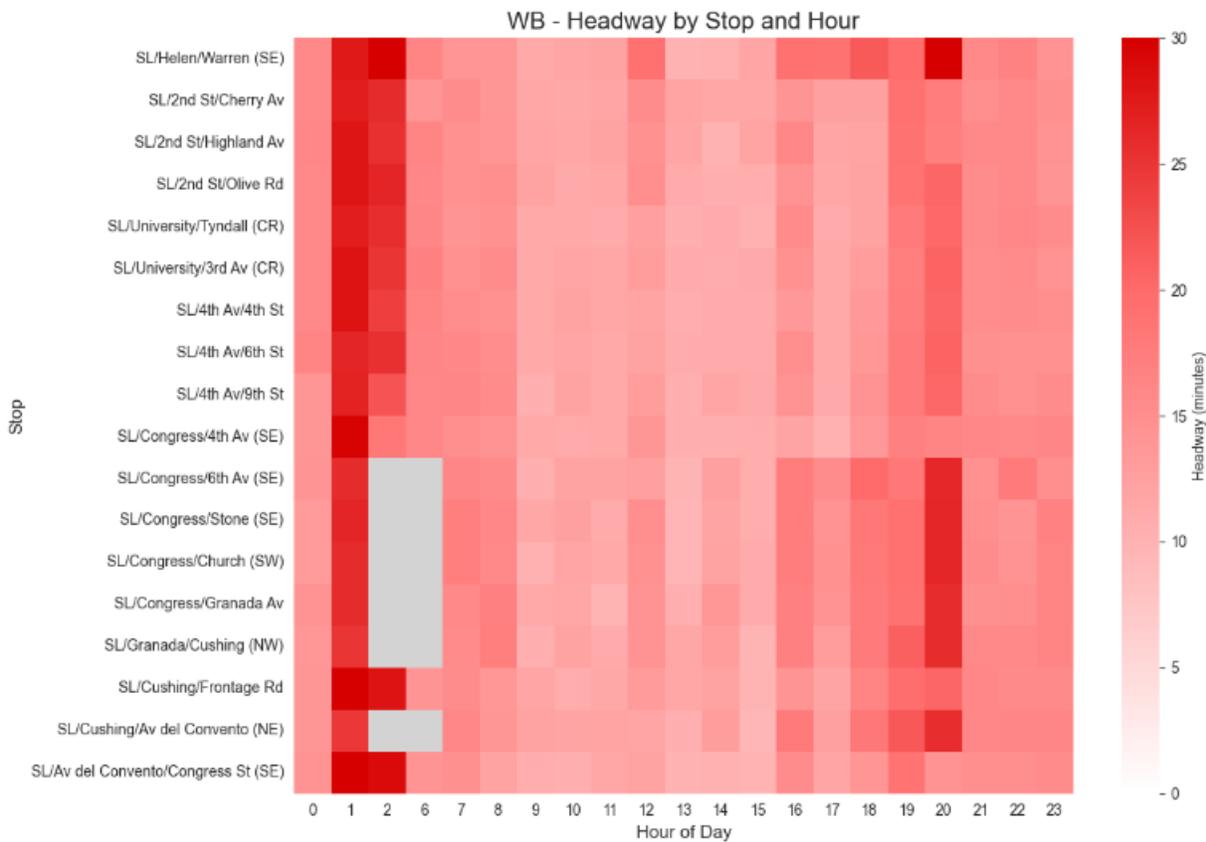


Figure 6.21. Sun Link stop level headways heatmap – Westbound

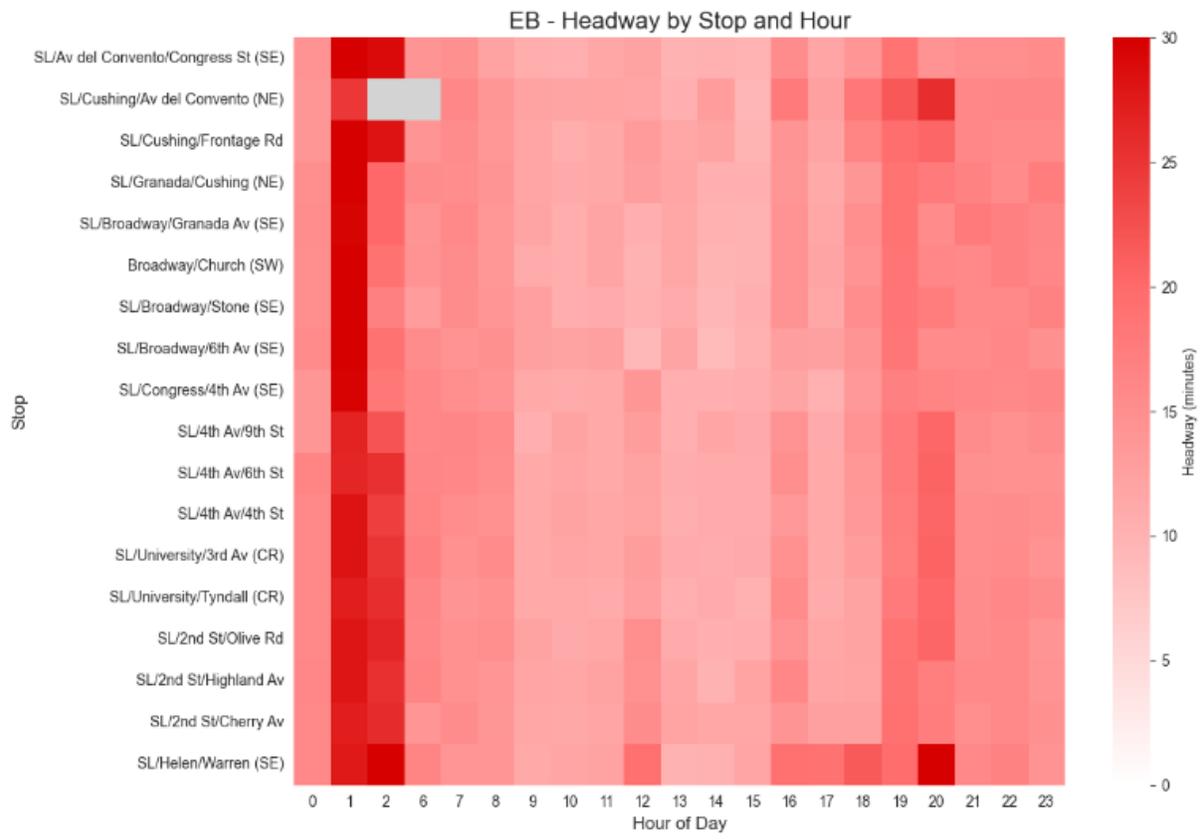


Figure 6.22. Sun Link stop level headways heatmap – Eastbound

6.5. Headway Irregularity

The analysis of headway performance, conducted at both system-wide and individual route levels, reveals variations in headway times across different transit routes. At the system level, this shows the range of headways at a given time of day, at the route level, this shows the variation in headways due to roadway or rider factors.

To quantify the irregularity of headways, we measure the standard deviation of headways for each hour, separately for the system as a whole and for each specific route. The standard deviation, denoted as σ , is calculated using the difference in arrival times between consecutive vehicles.

$$\sigma_h = \sqrt{\frac{1}{N} \sum_{i=1}^N (h_i - \mu_h)^2}$$

In which:

N : number of calculated headways record each hour

h_i : each individual headway value

μ_h : mean headway of the hour

σ_h : standard deviation of headways for each hour

6.5.1 *Sun Tran System level headway standard deviation performance measure*

To show the range of irregularity in the Sun Tran system, the standard deviation was calculated for each hour of each transit route over the study period, and each route-hour was visualized in [Figure 6.23](#). Each boxplot illustrates the interquartile range (IQR), representing the middle 50% of the data, while whiskers extend to show the spread beyond the IQR.

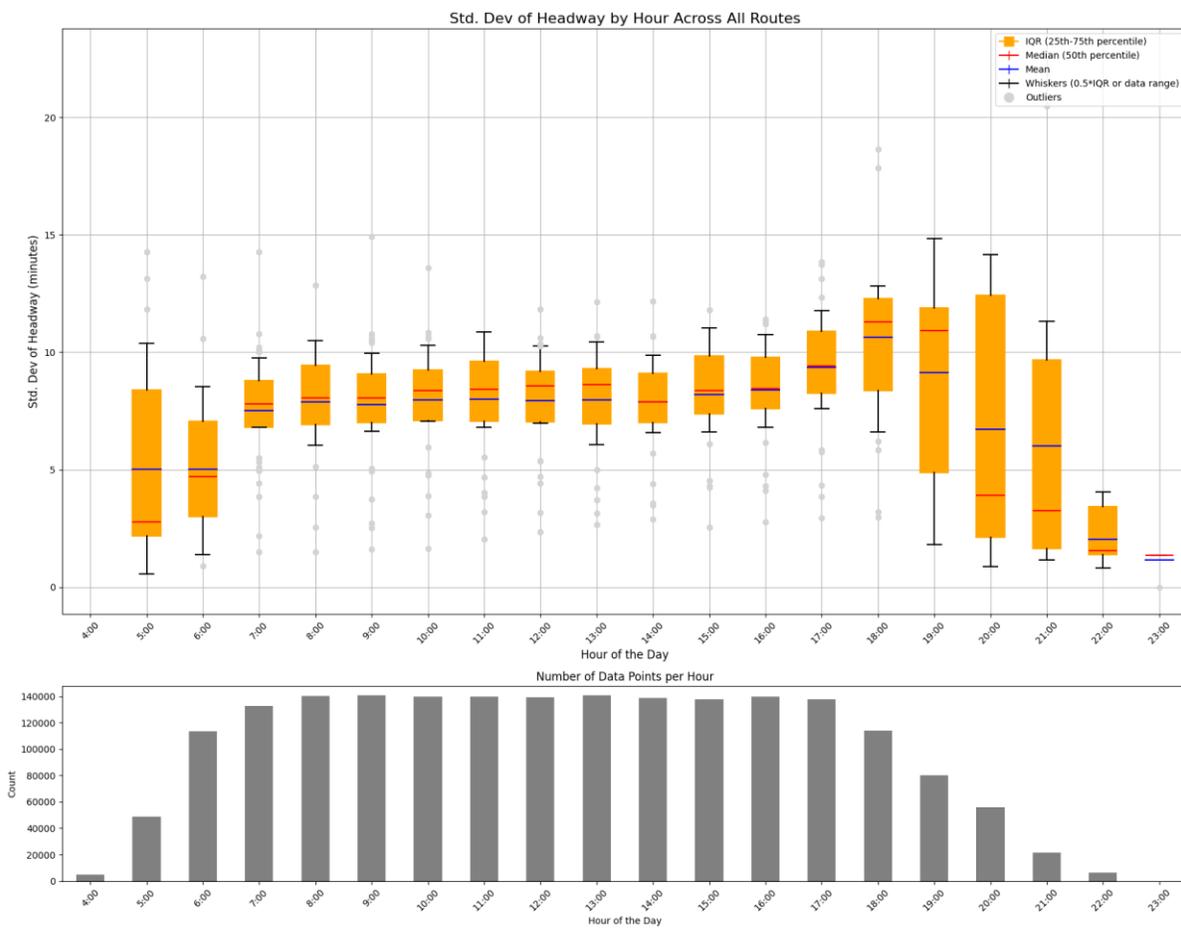


Figure 6.23. Sun Tran system level headway standard deviation boxplot

6.5.2 Sun Tran route level headways irregularity performance measure

To assess the regularity of service across individual transit routes, headway irregularity, measured as the standard deviation of actual headways, was analyzed at the route level, and visualized through the heatmap in **Figure 6.24**. In this visualization, darker red shades represent higher irregularity, indicating less consistent intervals between vehicles, while lighter reds reflect lower irregularity, suggesting more uniform service. Gray areas denote hours for which data was unavailable. This heatmap allows for an intuitive identification of when and where irregular service patterns emerge, helping to inform targeted operational improvements.

Overall, the transit system maintains relatively consistent service before the AM peak and after the PM peak, with notable increases in irregularity during core daytime hours. Based on observed patterns, most transit routes exhibit higher headway irregularity during daytime hours, compared to the more consistent service seen in early morning and late evening periods, showing the impacts of traffic and increased ridership.

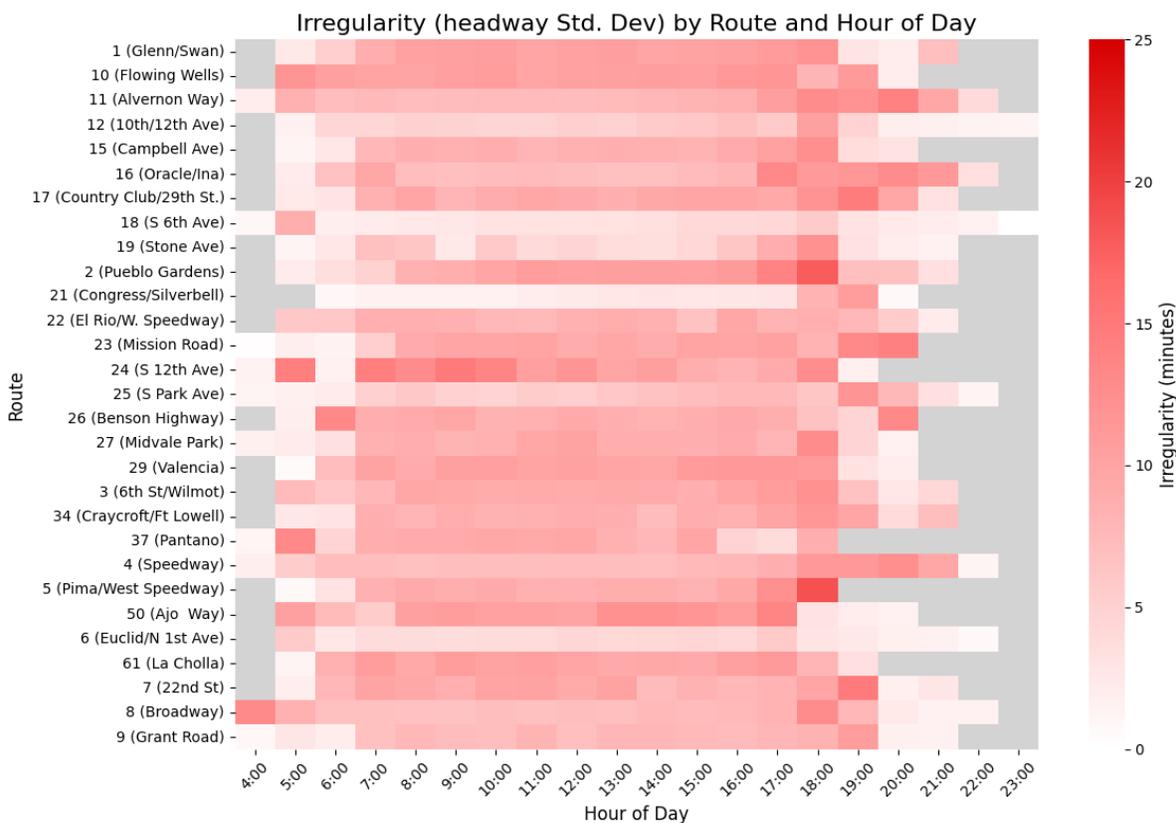


Figure 6.24. Sun Tran route level irregularity heatmap

6.5.3 Sun Link stop level Irregularity performance measure

To assess the consistency of service for the Sun Link streetcar, headway irregularity was analyzed at the stop level for each direction of travel. The heatmaps in **Figure 6.25** and **Figure 6.26** illustrate headway irregularity by stop and hour for the westbound and eastbound directions, respectively.

The analysis shows that headway irregularity tends to be more pronounced during midday and early evening hours. Both directions show increases in irregularity at 8-9AM and beginning around 12:00 PM and remaining elevated into the evening hours, with westbound headways being more irregular in the downtown area, and eastbound headways being more irregular in the University area.

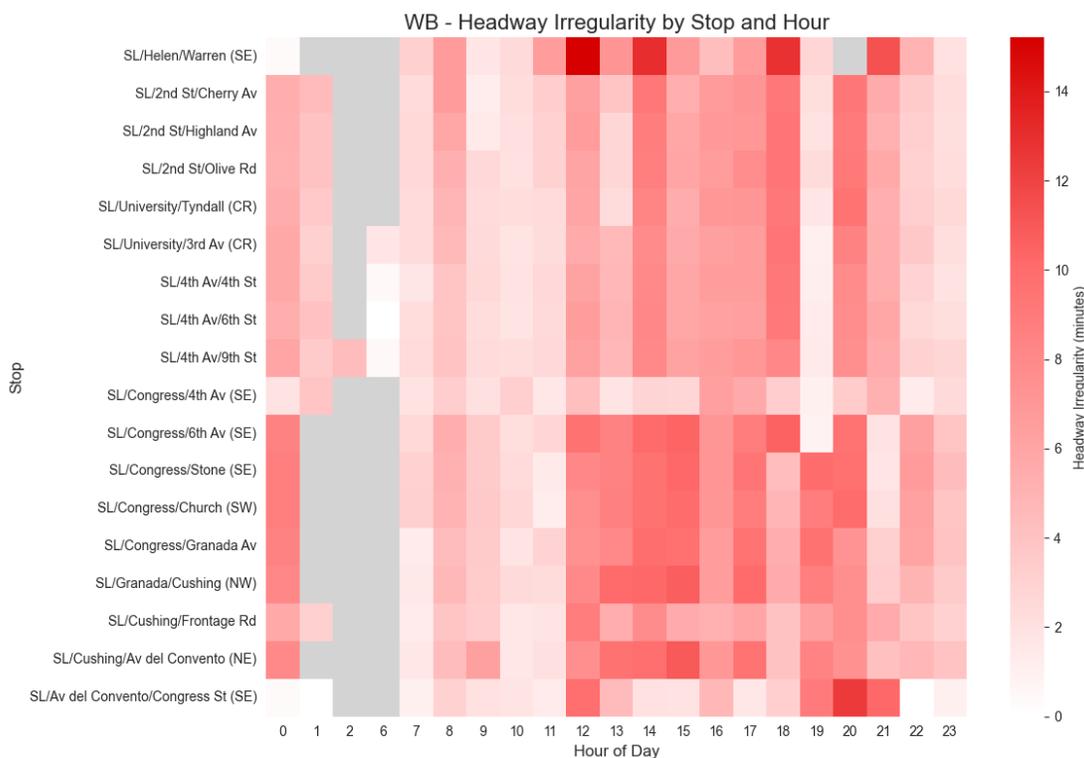


Figure 6.25. Sun Link stop level irregularity heatmap – Westbound

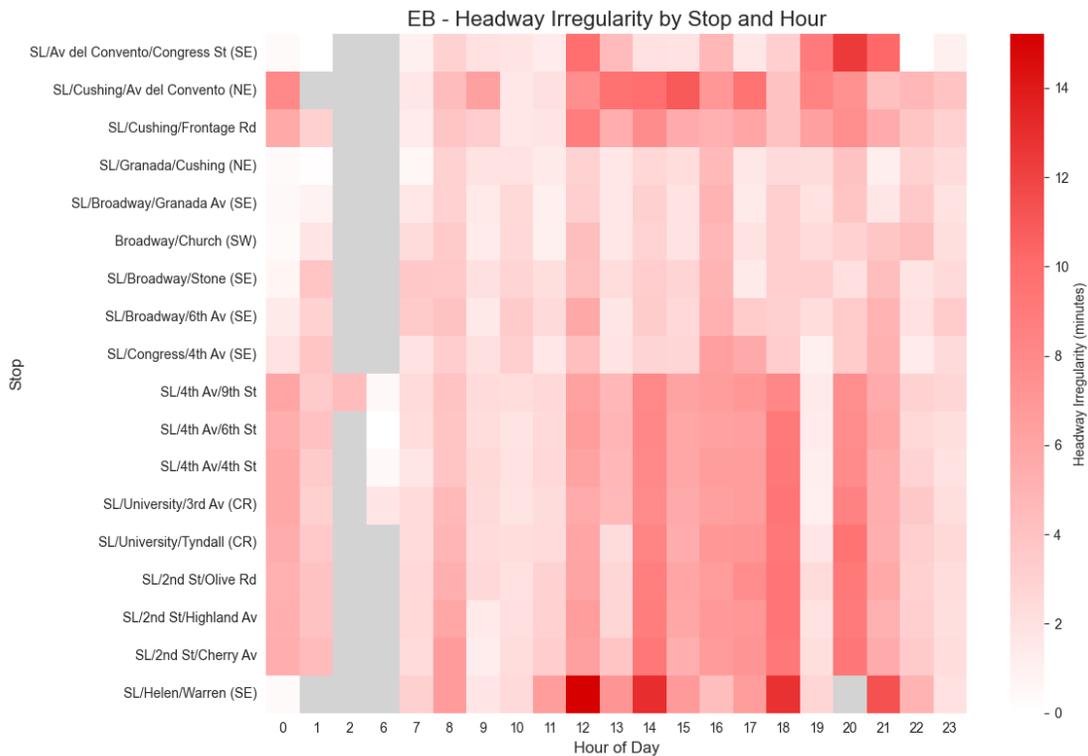


Figure 6.26. Sun Link stop level irregularity heatmap – Eastbound

6.6. Headway Reliability

To understand how reliable the headway is for passengers, headway reliability was measured using a method adapted from the travel time reliability performance measure. Knowing the reliability of headways helps transit agencies identify inconsistencies in vehicle arrivals. For passengers, improved headway reliability translates to reduced waiting time uncertainty, better trip predictability. To calculate Headway reliability, the method referenced for travel time reliability is applied (Harsha, M. M. & Mulangi, R. H., 2021).

$$R = 1 - \frac{h_{95} - h_{mean}}{h_{mean}} = \begin{cases} \geq 0 \text{ and } < 1, & \text{if } h_{mean} < h_{95} \leq 2h_{mean} \\ < 0, & \text{if } h_{95} > 2h_{mean} \end{cases}$$

In which:

h_{95} : 95th percentile of stop-to-stop segment headway for each hour

h_{mean} : mean of stop-to-stop segment headway for each hour

This formula quantifies the deviation of the less reliable headway from the average, providing reliability rate, R , that measures the passenger's ability to predict the headway based on a typical case a they would keep in mind. Measuring headway reliability in this way helps identify times and routes with potential issues in service consistency. The range of R can be referred to the travel

time reliability section. The analysis offers clear insights into the predictability of service, which is crucial for passenger satisfaction and operational planning. Understanding where and when headways deviate significantly from the norm allows transit agencies to make targeted improvements to reduce variability, thereby enhancing the overall reliability of the transit system.

As with the travel time reliability, the headway reliability rate (R) is also normalized between 0 and 1 for ease of interpretation. Here, 0 corresponds to the cases where h_{95} is larger than or equal to $4 \times h_{mean}$ and 1 corresponding to the hypothetical case of h_{95} being equal to the h_{mean} (i.e., being most reliable).

$$R^{normalized} = \frac{R - R_{min}}{R_{max} - R_{min}}$$

Where:

$R^{normalized}$: the normalized headway reliability index

R_{min} : the minimum headway reliability ($R_{min} = -3$ for all $h_{95} \geq 4 \times h_{mean}$)

R_{max} : maximum travel time reliability ($R_{min} = 1$)

6.6.1 Sun Tran system level headway reliability performance measure

Figure 6.27 shows the normalized headway reliability ($R^{normalized}$) of Sun Tran buses against each hour of the day. The $R^{normalized}$ values are represented by Box-and-Whisker plots in each hour of the day. Between 7:00 AM to 5:00 PM, with $R^{normalized}$ values range from 0.8 to 0.9, while the early morning and evening periods $R^{normalized}$ values range from 0.85 to 0.98, indicating consistently reliable service even during times with more vehicle and rider activity.

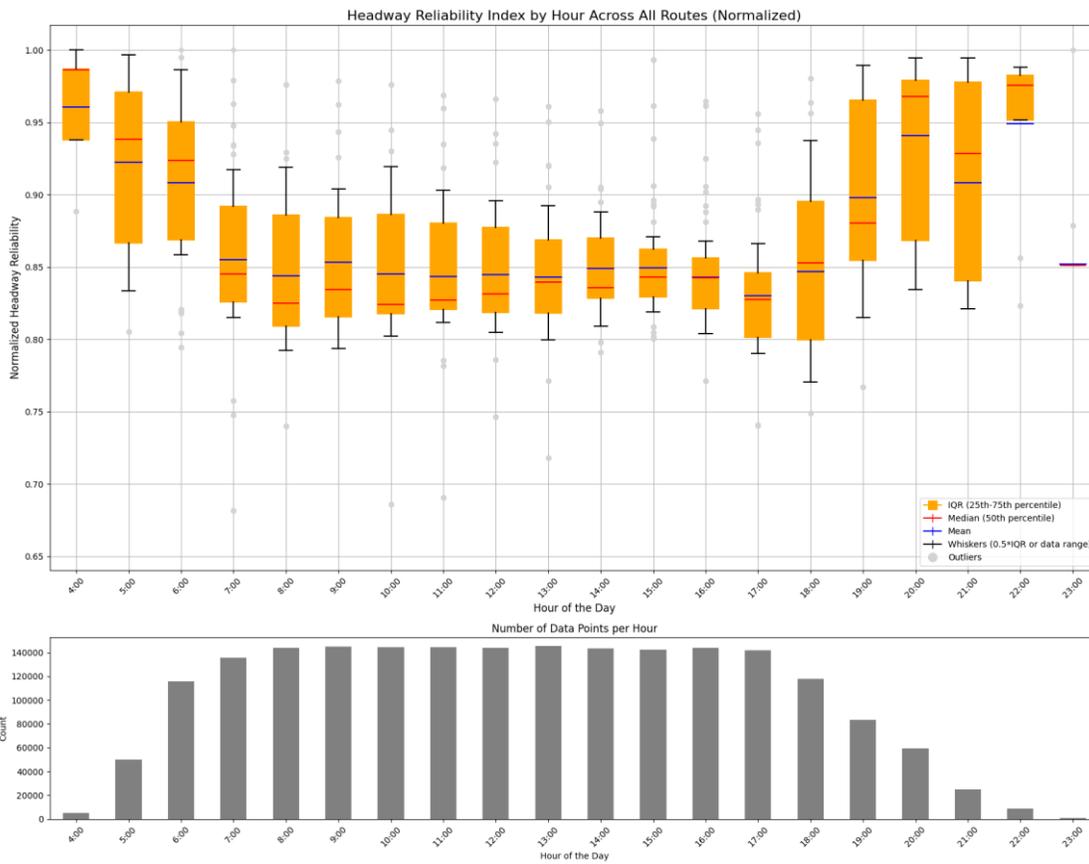


Figure 6.27. Sun Tran system level headway reliability boxplot

6.6.2 Sun Tran route level headway reliability performance measure

Figure 6.28 shows normalized travel time reliability for all SunTran routes, which shows the same overall headway reliability pattern, with most routes’ highest in the early morning and evening, and lower during daytime hours but still high overall.

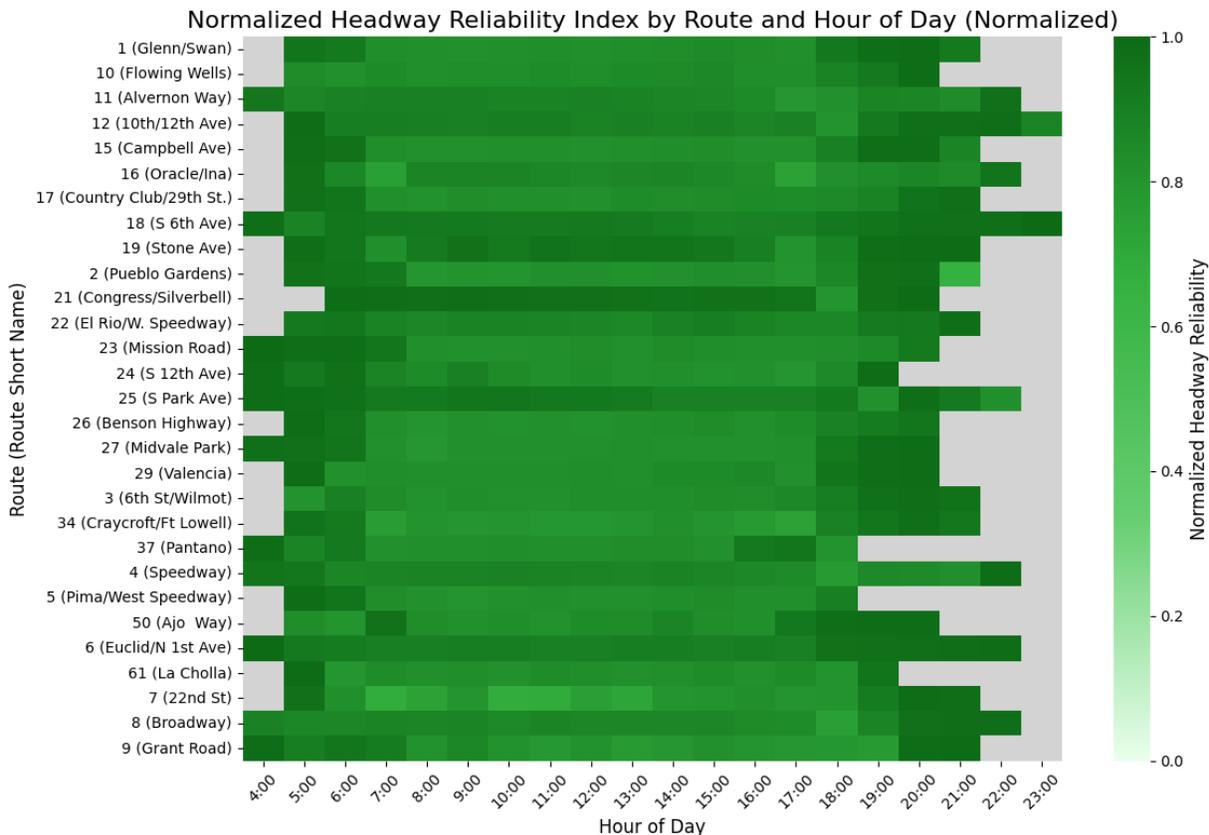


Figure 6.28. Sun Tran route level headway reliability heatmap

6.6.3 Sun Link stop level headway reliability performance measure

As shown in [Figure 6.29](#) and [Figure 6.30](#), $R^{normalized}$ is high in both directions over most of the day, though the westbound direction experiences slightly lower reliability from 12PM to 4PM along the entire route, and lower reliability from 4PM to 10PM along the University of Arizona and 4th Ave segments, and the eastbound direction experiences slightly lower reliability from 2PM to 7PM along 4th Ave and in the university area.

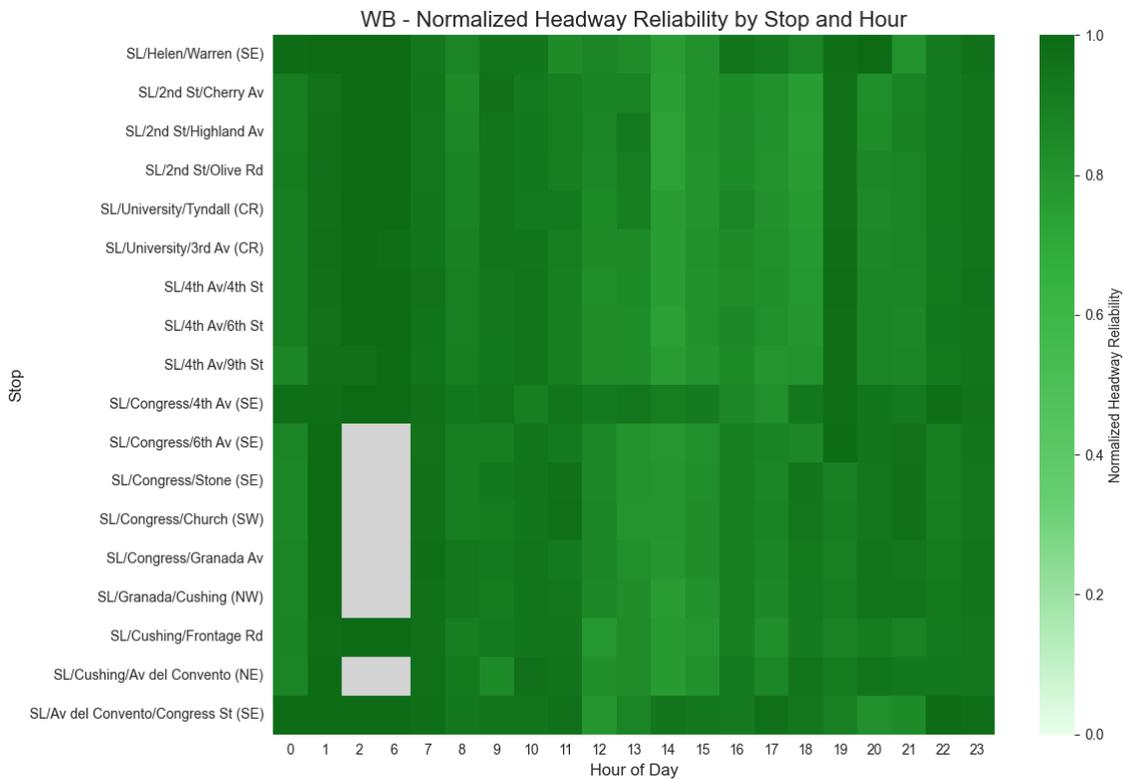


Figure 6.29. Sun Link stop level headway reliability heatmap – Westbound

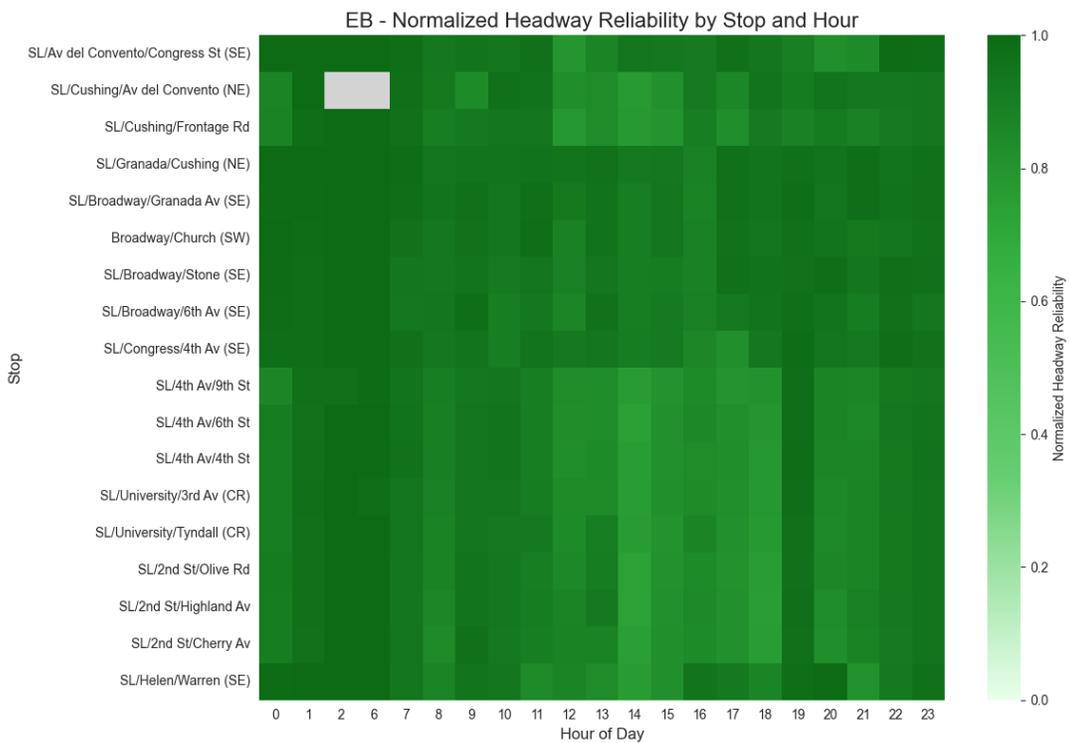


Figure 6.30. Sun Link stop level headway reliability heatmap – Eastbound

6.6.4 On-Time Performance

To evaluate how well Sun Tran service adheres to its schedule, On Time performance is measured as the percentage of trips that arrive within an acceptable time window. According to the Sun Tran website, passengers are advised to arrive at the stop five minutes earlier than the scheduled bus arrival time. Based on this assumption, arrival up to five minutes early is considered acceptable. In addition, referring to the literature (Strathman and Hopper, 1993), an arrival up to five minutes late is also defined on time.

$$P_{On-Time} = \frac{n_{On-Time}}{N}$$

In which:

$P_{On-Time}$: percentage of On Time arrival

n_1 : number of trips arriving on time

N : total number of total trips

As shown in **Figure 6.31**, On-Time percentages are observed to have high mean and median values, with both above 90% throughout the day. On-time percentages are slightly lower during the afternoon and PM peak, corresponding to periods of higher passenger and roadway volumes.

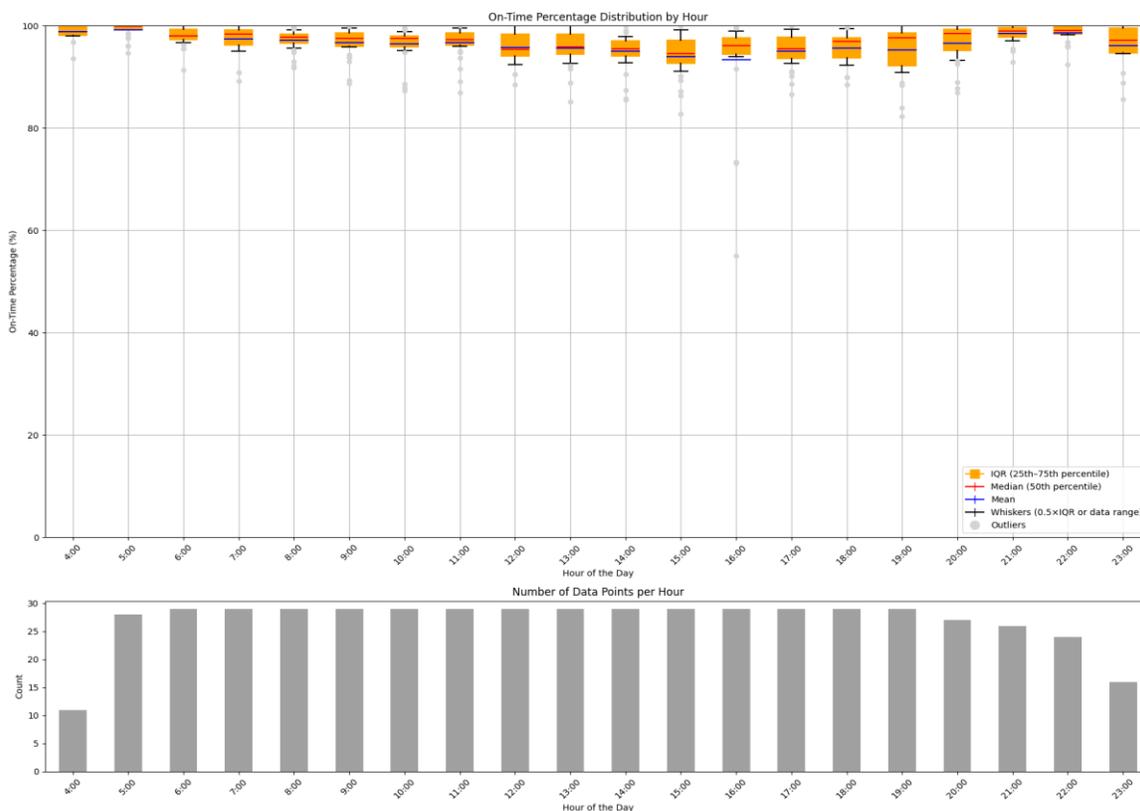


Figure 6.31. Sun Tran On Time percentage per hour of the day

Figure 6.32, below, show on-time percentage values close to 100% for almost all routes, though some, such as Route 18 showed lower than average on-time rates for most operating hours, and others showed isolated drops in on-time rates, such as Route 15, 21, and 37, at 4PM, likely attributable to increased traffic in the PM peak period.

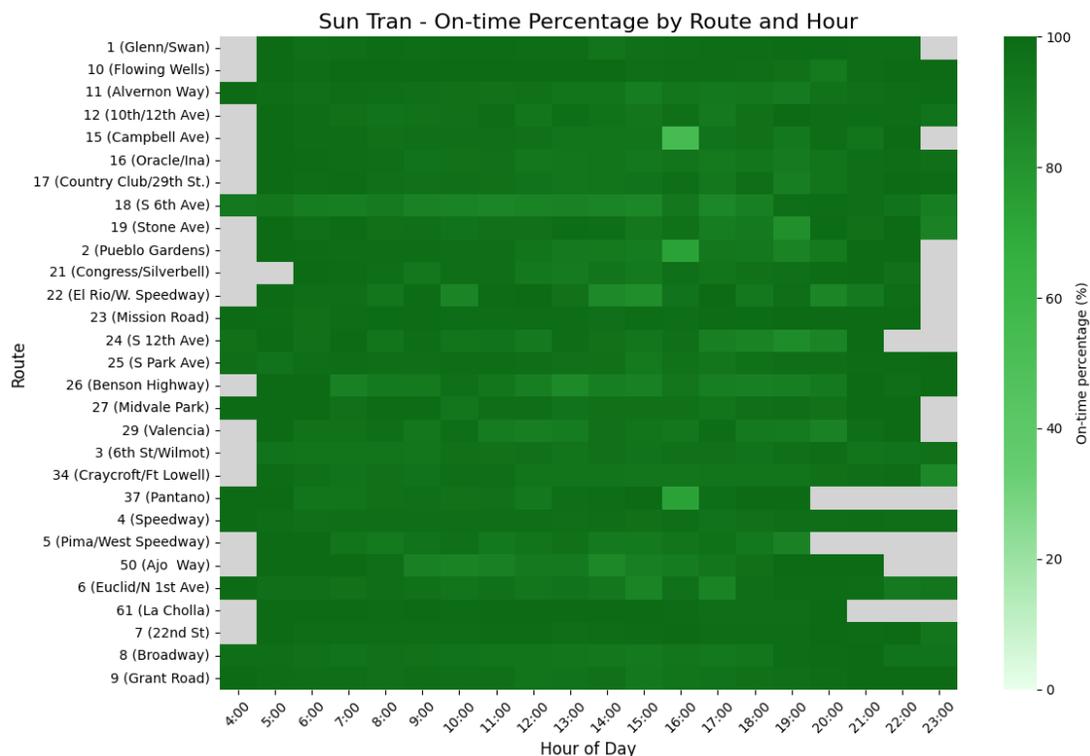


Figure 6.32. On-Time Percentage Reliability Index by Route

CHAPTER 7. Pedestrian Performance Measure Extraction

This chapter focuses on identifying and quantifying key performance measures for the pedestrian system in PAG region. As outlined in **CHAPTER 4**, Miovision (video detection and event-based) and MaxView (event-based) data are processed independently. To develop models for estimating pedestrian counts, PAG turning movement counts (TMC's) manually collected pedestrian counts,



which align with some MaxView locations, are utilized as ground-truth data. Furthermore, waiting times will be estimated using event-based data, with pedestrian delay represented by the hourly pedestrian waiting time. The pedestrian level of service (PLOS) will be evaluated following the HCM model based on the estimated pedestrian delay. For intersections with MaxView data, pedestrian counts will be estimated with a machine learning model, such as XGBoost, based on event-based data, while pedestrian delays will be directly calculated as an average waiting time. For intersections covered by Miovision data, pedestrian counts will be derived from Miovision's video detection data, delays will be estimated from event-based data, and the PLOS will be determined based on pedestrian delays.

7.1. Deliverables

Table 7.1 below presents the three key performance measures—pedestrian volume, pedestrian delay, and PLOS—along with their detailed descriptions.

Table 7.1. Pedestrian performance measures summary

Performance Measure	Location in the Report	Description	Used Data
Pedestrian volume	Section 7.2	Hourly average volume of pedestrians in a crossing walk	<ul style="list-style-type: none"> • <i>PAG TMC pedestrian data</i> • <i>MaxView event-based data</i> • <i>Miovision event-based data</i> • <i>Miovision video detection data</i>
Pedestrian delay	Section 7.3	Calculated as hourly average waiting time of pedestrians in a crossing walk	
PLOS	Section 7.4	Hourly PLOS of a crossing walk	

To explore the relationship between push button counts and actual pedestrian counts, event-based push button data was aligned, spatially and temporally, with pedestrian count data. Upon reviewing the MaxView and Miovision data of 2023, we found no overlapping locations between the two datasets. As a result, we opted to use PAG TMC pedestrian count data, which was manually collected at key intersections across Tucson, providing a reliable ground-truth reference. **Figure 7.1** shows the locations of both the MaxView and Miovision data which is also shown in CHAPTER 4. The PAG TMC pedestrian count data enables us to validate push button counts effectively and ensures meaningful comparisons, especially in areas where automated pedestrian detection data is unavailable or insufficient, ensuring analytical consistency while addressing the spatial gaps in the MaxView and Miovision datasets.

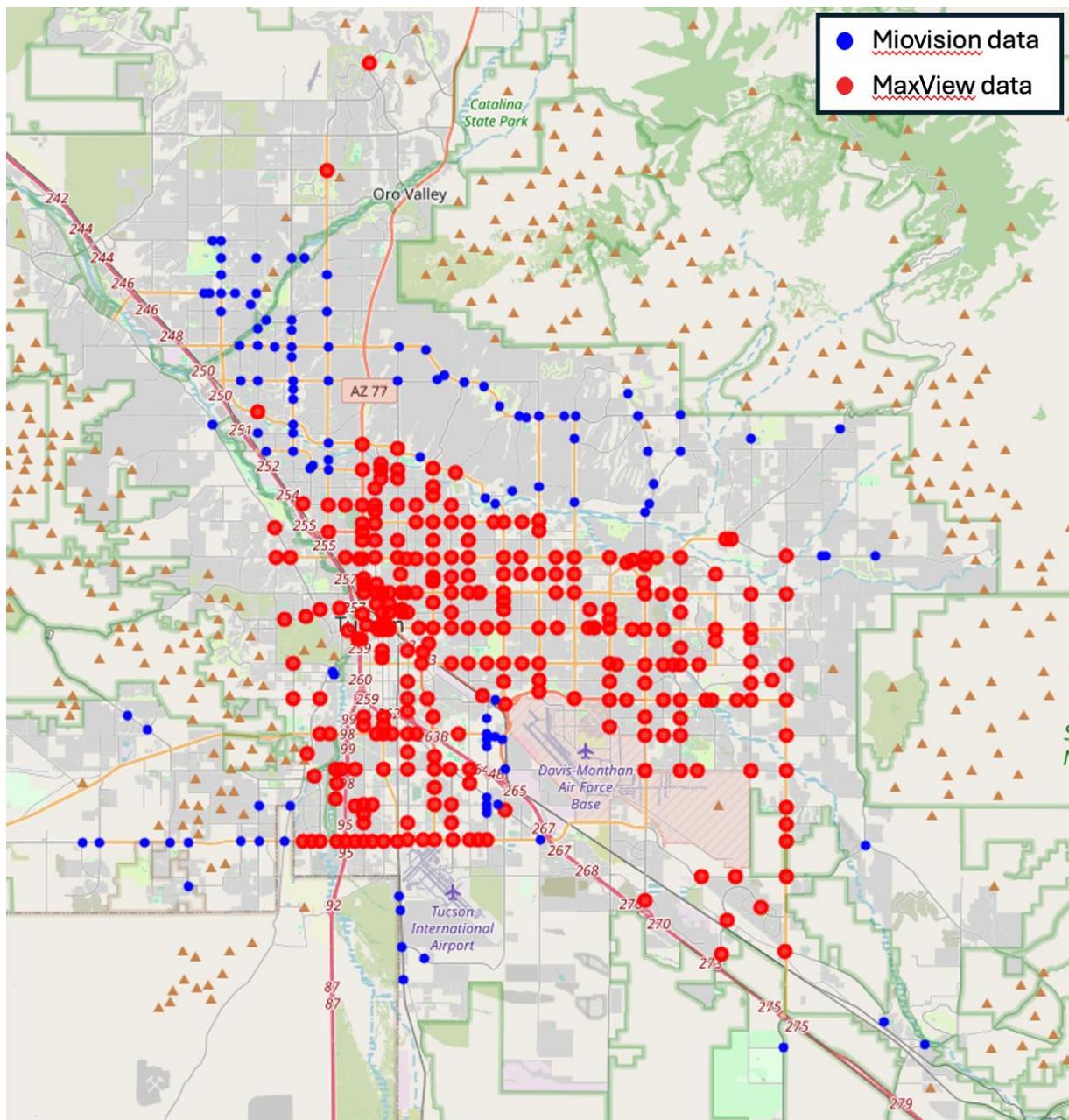


Figure 7.1. MaxView and Miovision locations of 2023

7.2. Pedestrian volume

7.2.1 Introduction

This section focuses on estimating pedestrian volume using hourly pedestrian push button data from the MaxView system in the City of Tucson and the ground truth data obtained from PAG TMC manually collected pedestrian volume data. The MaxView devices provide data on pedestrian push button activities, which can be used to estimate pedestrian traffic patterns. By using advanced machine learning models, this analysis captures time-dependent features and incorporates additional relevant data (i.e., spatial and time features) for high-accuracy estimations. This section does not use Miovision event-based data as many entries showed no recorded pedestrian volumes despite push button activity being detected. Due to these inconsistencies, the data was excluded from this analysis.

Previous work by Li and Wu (2021) developed a method to estimate pedestrian volumes at midblock crosswalks by analyzing pushbutton activity at a single junction. Their approach focused on predicting pedestrian counts at isolated midblock locations and was conducted at six locations. Li et al. (2022) expanded the approach using a Bayesian Additive Regression Trees (BART) model over 70 intersections in Pima County. This study estimates pedestrian volume across entire intersections, involving more complex patterns of pedestrian movement and utilizing all available MaxView data, covering 237 intersections across multiple months. Due to the broader scope and larger dataset, previous methods were not suitable for our analysis. Therefore, a fresh deployment ensured better representativity.

7.2.2 Methodology

The dataset used for this model is based on MaxView's pedestrian push button data and PAG TMC's actual pedestrian volumes collected across scattered months between 2017 and 2024. However, data is only available for the months of September through March; no data was recorded for April through August. It contains 45,478 push button events recorded over 1708 intersection-hours at 237 intersections. An intersection-hour refers to a single hour of data collected at a specific intersection. For example, if data is recorded at one intersection from 2 to 3 PM, that constitutes one intersection-hour.

To get a better understanding of where the intersections are, [Figure 7.2](#) presents a map displaying the geographic locations of a sample of intersections equipped with pedestrian signal devices throughout Tucson that were used for the model.

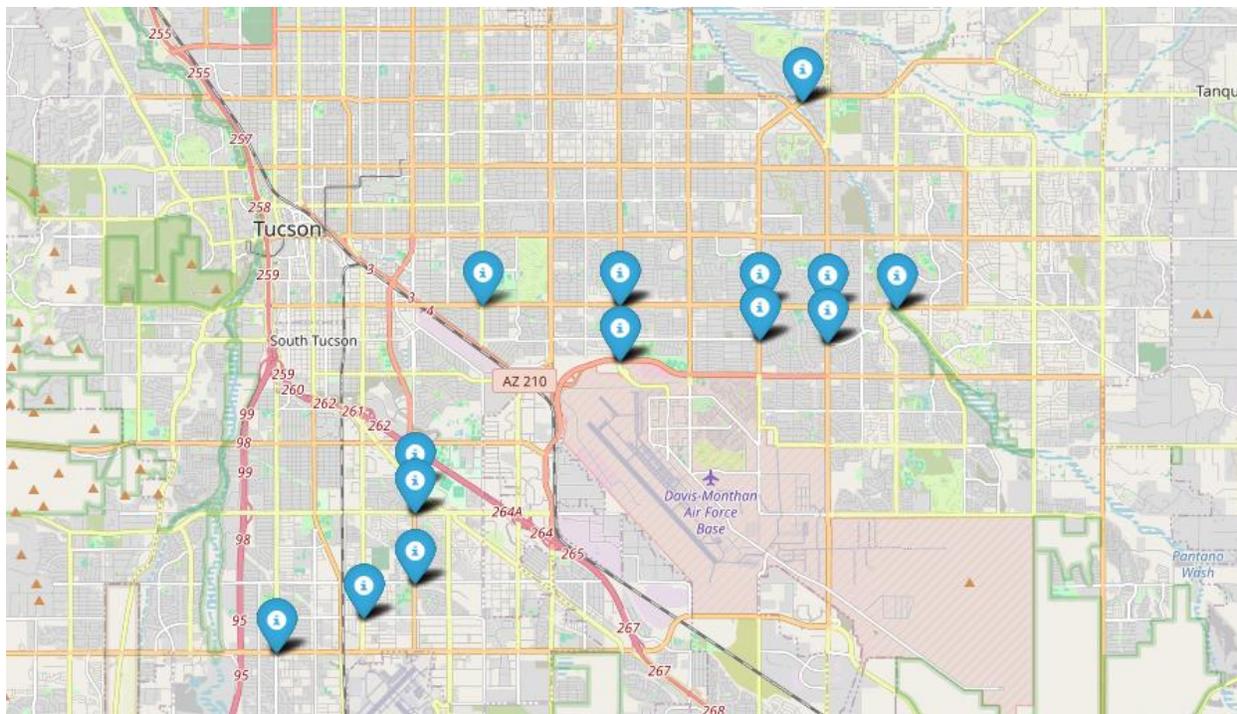


Figure 7.2. Sample locations of intersections that are equipped with pedestrian signal devices

Table 7.2 lists the sample intersections equipped with pedestrian signal devices throughout Tucson, including the intersection names and their corresponding latitude and longitude coordinates.

Table 7.2. Sample locations of intersections equipped with pedestrian signal devices

Name	Latitude	Longitude
22nd St & Country Club Rd	32.20692	-110.9267
22nd St & Swan Rd	32.20697	-110.8924
22nd St & Wilmot Rd	32.20676	-110.858
22nd St & Kolb Rd	32.20651	-110.841
22nd Street & Pantano Pkwy	32.20645	-110.8238
Wilmot Rd & 29th St	32.19969	-110.858
Kolb Rd & 29th St	32.19902	-110.841
Golf Links Rd & Swan Rd	32.19531	-110.8926
Kino Pkwy & Benson Hwy	32.16887	-110.9434
Irvington Rd & Campbell Ave	32.1633	-110.9434
Valencia Rd & 12th Ave	32.13399	-110.9778
Campbell Ave & Drexel Rd	32.14868	-110.9435
Park Ave & Bilby Rd	32.14134	-110.9562
Tanque Verde Rd & Camino Principal	32.24966	-110.8473

Below is a detailed overview of the new data preparation process:

7.2.2.1 Raw data

The raw data includes attributes such as *TimeStamp*, *DeviceID*, *EventId*, *Parameter*, and *DateHour*. Only *EventIds* of 45, corresponding to pedestrian push button events, were considered for the model. The *Parameter* attribute shows the direction in which the signal was pressed. Since *Parameter* is categorical, one-hot encoding was applied to preserve its categorical nature. For the purpose of the model, *DeviceID* and *DateHour*, were used to sort the data, an example of which is shown in [Table 7.3](#).

Table 7.3. Raw data sample

TimeStamp	DeviceID	EventId	Parameter	DateHour
2023-09-13 07:00:41.600*	722	45	4	2023-09-13 07:00:00
2023-10-03 07:45:18.100	217	45	2	2023-10-03 07:00:00
2023-10-04 08:57:44.000	486	45	6	2023-10-04 08:00:00

Note: *TimeStamp format is YYYY-MM-DD HH:mm:ss:fff

Additional sources of data were incorporated as additional model training variables. [Table 7.4](#) below describes the variables used for training the model as well as their origin.

Table 7.4. Description of model variables

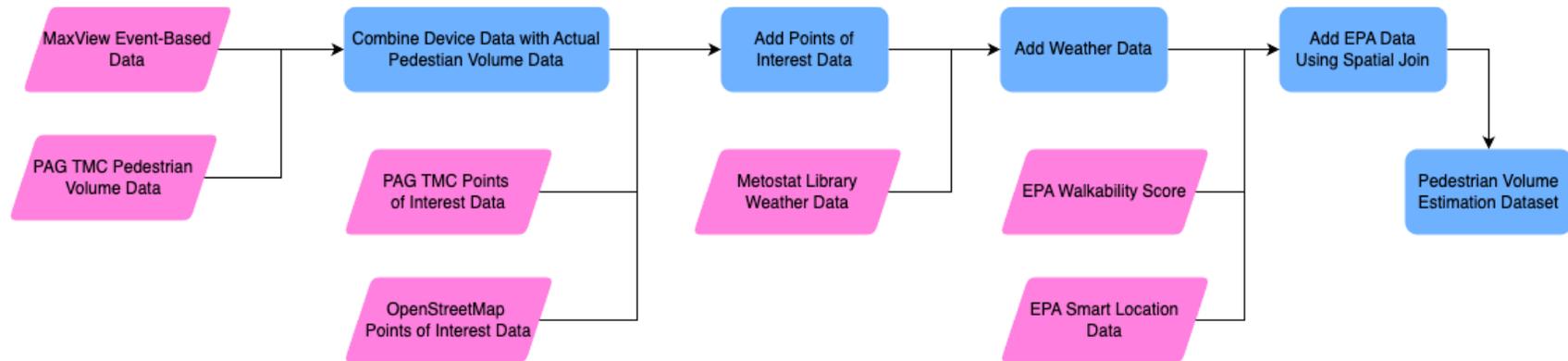
Variable name	Source	Description
TotalSignals	MaxView devices	Total number of push button events in an hour
Fast_food	OpenStreetMap points of interest (POI)	Number of fast foods within a quarter-mile radius of device
Parking		Number of parkings within a quarter-mile radius of device
Theatre		Number of theatres within a quarter-mile radius of device
School		Number of schools within a quarter-mile radius of device
Fuel		Number of gas stations within a quarter-mile radius of device
Shops		Number of shops within a quarter-mile radius of device (e.g., convenience stores, supermarkets, clothing stores)
Leisure		Number of leisure activities within a quarter-mile radius of device (e.g., park, playground, pitch)
Tourism		Number of tourism attractions within a quarter-mile radius of device (e.g., museum, hotel, general tourist attractions)
TotPop		Census block population around the device's location

Variable name	Source	Description
Pct_AO0	U.S. Environmental Protection Agency (EPA) Smart Location attributes	Percentage of zero-car households in the census block population around the device's location
D4A		Distance to the nearest transit stop in the census block population around the device's location
Temperature	Meteostat library attributes	Temperature at the location of the device at a given date and hour
Resource-Dependent Activities	PAG POI attributes	Count of each type of location within a quarter-mile radius of the device's location
Consumer Svcs		
Finance, Insurance, & Real Estate (FIRE)		
Government, Social, & Advocacy Svcs		
Business Svcs		
High Tech Manufacturing & Development		
Health Care		
Hospitality, Tourism, & Recreation		
Consumer Goods Manufacturing		
Construction		
Media, Publishing, & Entertainment		
Retail		
Transportation & Distribution		
Education		
Metal Inputs & Transportation-Related Manufacturing		
Telecommunications		
Non-Metallic Manufacturing		
NatWalkInd	EPA Walkability Score (based on census blocks)	National walkability score near the device's location ranging from 0 to 20

Figure 7.3 shows the multi-step process used to compile the dataset by integrating data from seven different sources. Pedestrian volume measurements are combined with corresponding MaxView event-based data, which includes the numbers of push-button events and device-specific information. In the next step, points of interest (POI) data from OpenStreetMap (e.g., fast-food, parking, theatre, etc.) and PAG (e.g., business services, health care, telecommunications, etc.) are included by identifying POIs within a quarter-mile radius of each device in order to capture the potential influence of surrounding facilities on pedestrian volume at each intersection.

Next, temperatures are added to each data point to account for the potential impact of weather on walking behavior. Finally, EPA attributes (e.g., NatWalkInd, TotPop, Pct_AO0, and D4A) are integrated through a spatial join, where each device location is mapped to the corresponding census

block to extract relevant EPA attributes. These attributes provide insights into the built environment and transportation-related features of the surrounding population, helping to assess their potential influence on pedestrian activity.



* Refer to description of model variables for more information.

Figure 7.3. Methodology flow chart for calculation of pedestrian volume

7.2.2.2 Model

The predictive model used in this analysis is an XGBoost Regressor. Hyperparameters for the model were optimized using grid search with a 5-fold cross-validation process, sampling 50 parameter combinations. The parameter grid explored various settings for tree depth, learning rate, subsample ratios, and the number of estimators. The best parameters identified were:

- subsample: 0.6
- n_estimators: 300
- min_child_weight: 5
- max_depth: 9
- learning_rate: 0.05
- colsample_bytree: 0.6

Based on the top model, outlier removal was performed to improve the predictive power of the model by removing the noise. In particular, since push button events are a key attribute for predicting hourly pedestrian volume, inaccuracies in these measurements could significantly impact the model's predictive ability. Such inaccuracies may result from device malfunctions or repeated button presses within a short period. To address this, data points with potentially inaccurate push button events were removed from the data.

For each device, push button events exceeding the threshold (defined as the 75th percentile plus 1.5 times the interquartile range) for each device were excluded. This process resulted in the removal of 68 data points from a total of 1,708. Alternative noise thresholds were also tested but did not show improvements in model accuracy. [Figure 7.4](#) shows parity plots of pedestrian volume vs. number of push buttons with outliers marked as red.

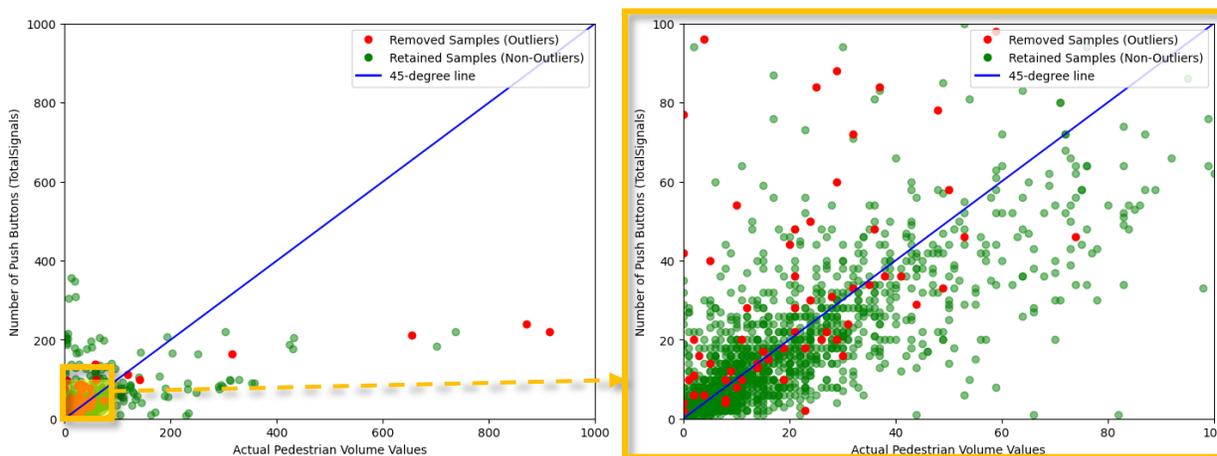


Figure 7.4. Parity plots of pedestrian volume vs. number of push buttons (TotalSignals)

Figure 7.4 highlights that the data points are widely dispersed, indicating that push-button events alone are not sufficient for accurately predicting pedestrian volume. The right plot is a zoomed-in version of the left plot, focusing on lower pedestrian volumes, which further emphasizes this dispersion. Additionally, by applying this outlier detection strategy, it can be observed that many of the points labeled as outliers are positioned far from the 45-degree line, reinforcing this strategy.

Figure 7.5 shows parity plots of pedestrian volume vs. XGBoost model's predictions with outliers marked as red.

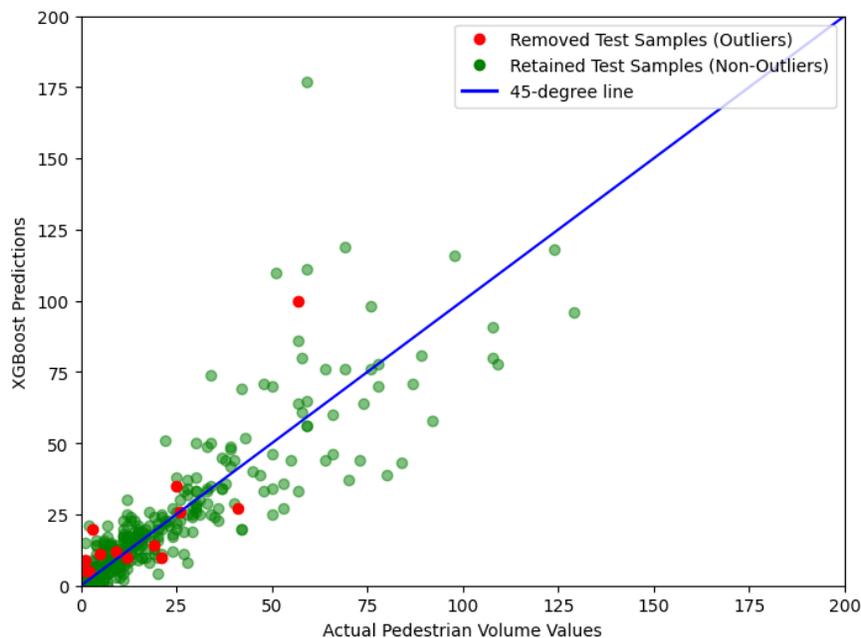


Figure 7.5. Parity plot of pedestrian volume vs. XGBoost predictions on test data before outlier removal

The plot shows that although test outliers are few, they deviate noticeably from the 45-degree line, indicating that the model struggles to predict their values accurately. Removing these outliers helps improve the model's overall performance by reducing the influence of extreme or irregular cases. **Figure 7.6** shows the distribution of data points after outlier removal, with 80% allocated to the training set and 20% to the test set.

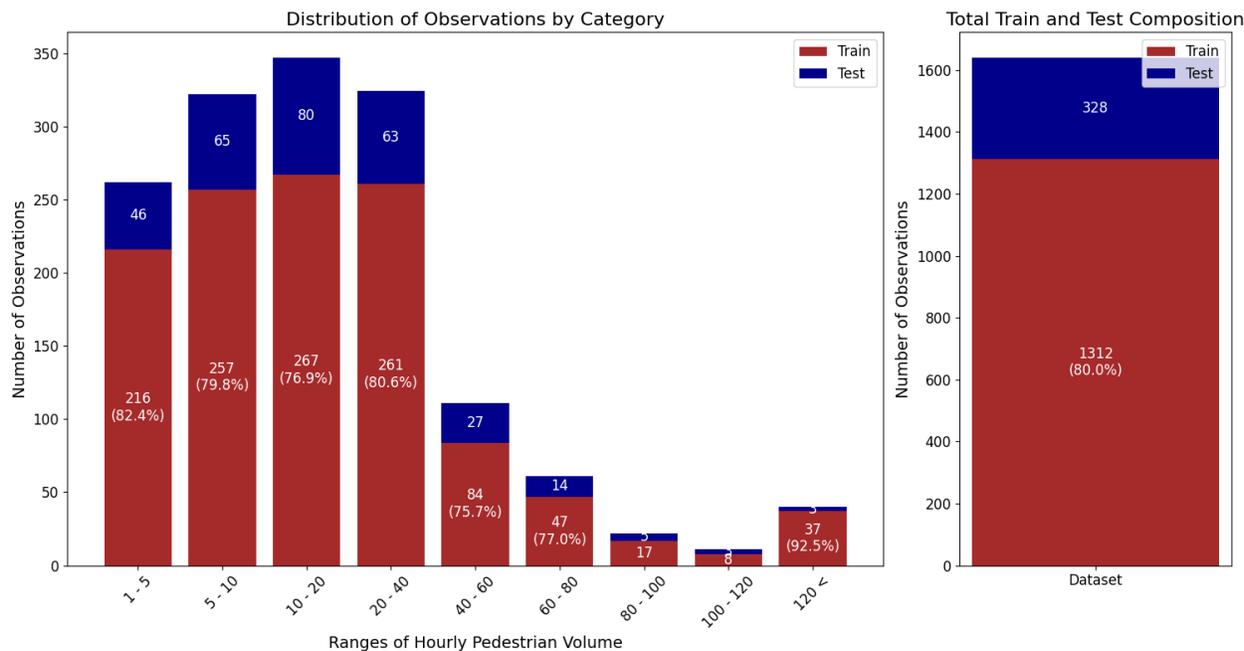


Figure 7.6. Distribution of data points by train and test

Figure 7.6 shows that most of the samples have lower pedestrian volume ranges (i.e. less than 40 pedestrians per hour), with a balanced representation between train and test splits. As pedestrian volume increases, the number of available observations significantly decreases, leading to a data imbalance in higher volume ranges. This suggests that model performance may be more reliable for lower pedestrian counts but could struggle with generalization in higher-volume scenarios due to limited data.

7.2.3 Results

The final data contains 31 features and 1,640 data points after merging the event-based data and PAG TMC's actual pedestrian volumes. Testing the XGBoost model on the test samples resulted in the following metrics:

- Root mean squared error (RMSE): 12.61
- Mean absolute percentage error (MAPE): 43.48%
- R-squared (R^2): 0.819
- Mean absolute error (MAE): 7.03
- Average 5-fold R-squared (R^2): 0.828

To evaluate the performance of the models effectively, multiple metrics are necessary to capture different aspects of prediction accuracy. A single metric may not fully reflect model performance, especially in cases where errors vary across different ranges of pedestrian volume. By using multiple evaluation criteria, we can ensure a more comprehensive assessment of how well the

model generalizes to new data. [Table 7.5](#) outlines the description, strengths, and limitations of each metric and provides a basis on why each metric was used.

Table 7.5. Description, strengths and limitations of metrics

Metric	Description	Strengths	Limitations
Root mean squared error (RMSE)	Average magnitude of error, penalizing large errors	Highlights large errors; unit-based interpretation	Sensitive to outliers; scale-dependent
Mean absolute percentage error (MAPE)	Average percentage error between predictions and actual values	Easy to interpret; scale-invariant	Unreliable for values near zero
R-Squared (R^2)	Proportion of variance explained by the model	Intuitive fit measure; good for model comparison	Misleading for non-linear or overfitted models
Mean absolute error (MAE)	Average of the absolute differences between predicted and actual values	Easy to interpret; not as sensitive to outliers as RMSE	Does not penalize large errors as heavily as RMSE

[Figure 7.7](#) displays the parity plot of the final model after outlier removal. The plot shows that the model generally follows the expected trend, with predictions aligning more closely to the 45-degree line for lower pedestrian volumes

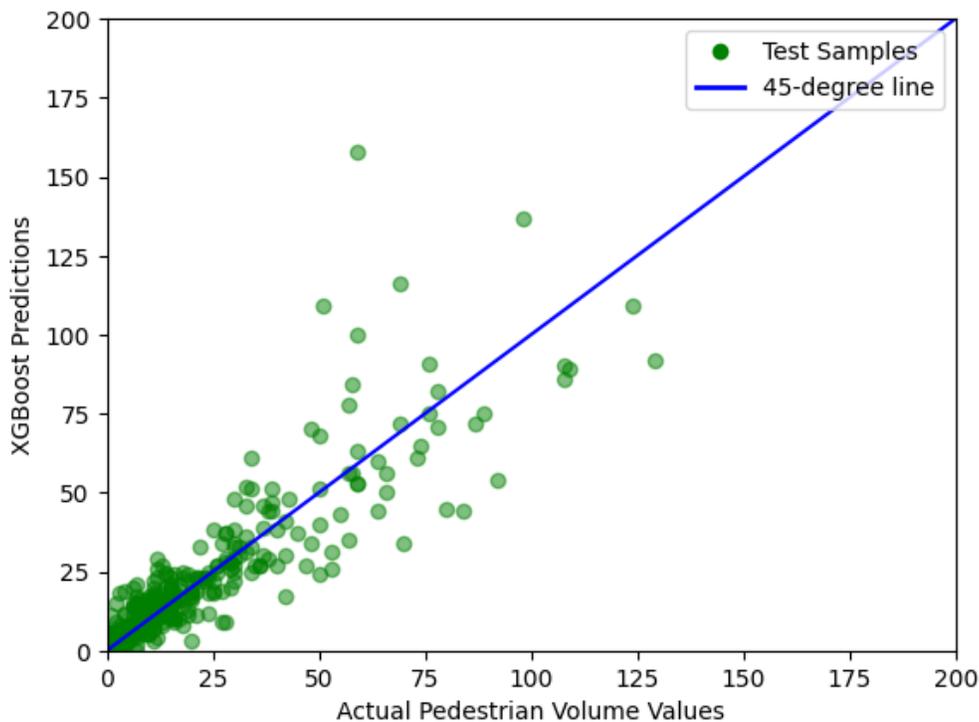


Figure 7.7. Parity plot of the final XGBoost model

7.3. Pedestrian delay

7.3.1 Introduction

Research on directly utilizing push-button data to determine pedestrian delay remains limited. Pedestrian delay is generally measured as the time difference between when a pedestrian requests a crossing by pressing the push button (push-button event) and when the crossing signal is granted (Karimpour et al., 2021). However, several challenges arise in accurately calculating delay using push-button data. A single pedestrian may press the button multiple times or not press it at all, and subsequent pedestrians arriving at the intersection may or may not activate the button. Additionally, those arriving at an intersection with a walk signal already active will not engage the button (Li and Wu, 2020). These scenarios can lead to either underreporting or overreporting pedestrian activity, complicating the delay calculations.

To address these challenges, some methodologies (Karimpour et al., 2022; Marisamynathan & Vedagiri, 2018) have been applied to estimate pedestrian delay. One such approach (Karimpour et al., 2022) utilizes a finite mixture model based on event-based data, with factors such as traffic flow, cycle length, and pedestrian green duration playing a critical role. This model was applied to four signalized intersections along major arterial corridors in Pima County and demonstrated superior performance compared to traditional methods, including the Highway Capacity Manual (HCM), Virkler and Dunn methods (Karimpour et al., 2021). Similar techniques could be extended to other intersections to enhance delay estimation accuracy and support level of service (LOS) assessments.

In the following sections, pedestrian delay for each intersection/HAWK will be calculated as the average waiting time for each push-button activation within a given cycle length. For the original event-based data, events exceeding the threshold (defined as the 75th percentile plus 1.5 times the interquartile range) for each device were excluded.

7.3.2 Methodology

Now, both the MaxView and Miovision event-based data are accessible through PAG data server or CATS SQL server. The two data share similar formats, but not the same. In MaxView data it has:

Table 7.6. MaxView event code table

	Event ID	Event descriptor	Description
Pedestrian phase events	21	Pedestrian begin walk	Set when walk indication becomes active.
Phase control events	45	Pedestrian call registered	Call to service on a phase is registered by pedestrian demand. This event will not be set if a recall exists on the phase.

However, in the Miovision event-based data, there does not exist the event ID of 45, thus we use the event ID of 90 to represent the time of actual push button event. The event with event ID of 45 is the time when the pedestrian pushes the button, and the event with event ID of 90 is the time when the signal controller receives the push button signal. Thus, the time gaps between them are relatively low, usually within 0.1 second.

Table 7.7. Miovision event code table

	Event ID	Event descriptor	Description
Pedestrian phase events	21	Pedestrian begin walk	Set when walk indication becomes active.
Phase control events	90	PedDetector on	Ped detector events shall be triggered post any detector delay/extension processing and may be set multiple times for a single pedestrian call.

The actual waiting time is calculated through the following (Li & Wu, 2021; Xu et al., 2023):

$$AWT_p = \frac{\sum_{i=1}^{N_p} WT_i}{N_p}$$

$$WT_i = T_{i,1} - T_{i,0}$$

In which:

WT_i : Waiting time of each push button i

$T_{i,0}$: Time of push button initially activated for i^{th} event_code = 45 or 90

$T_{i,1}$: Time of next light turns green (event code = 21) after i^{th} event_code = 45 or 90

N_p : Number of pushes of push buttons in the time period p (e.g. one hour)

AWT_p : Average waiting time of pedestrians in this time period p

Figure 7.8 below illustrates the process of calculating pedestrian delays using event-based data from both MaxView and Miovision. First, pedestrian waiting times are calculated individually for each crosswalk using data from both sources. These calculated waiting times are then averaged to determine the overall pedestrian delay for each crosswalk.

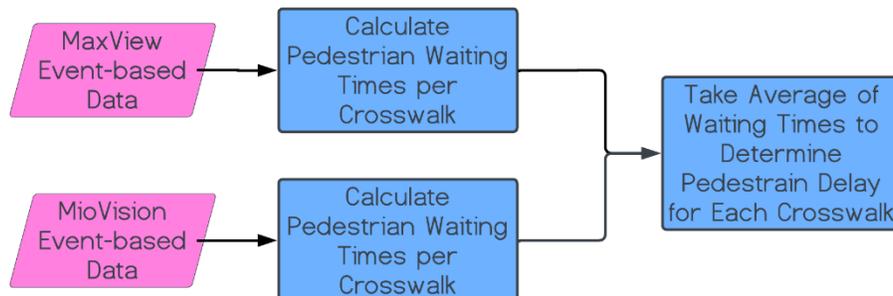
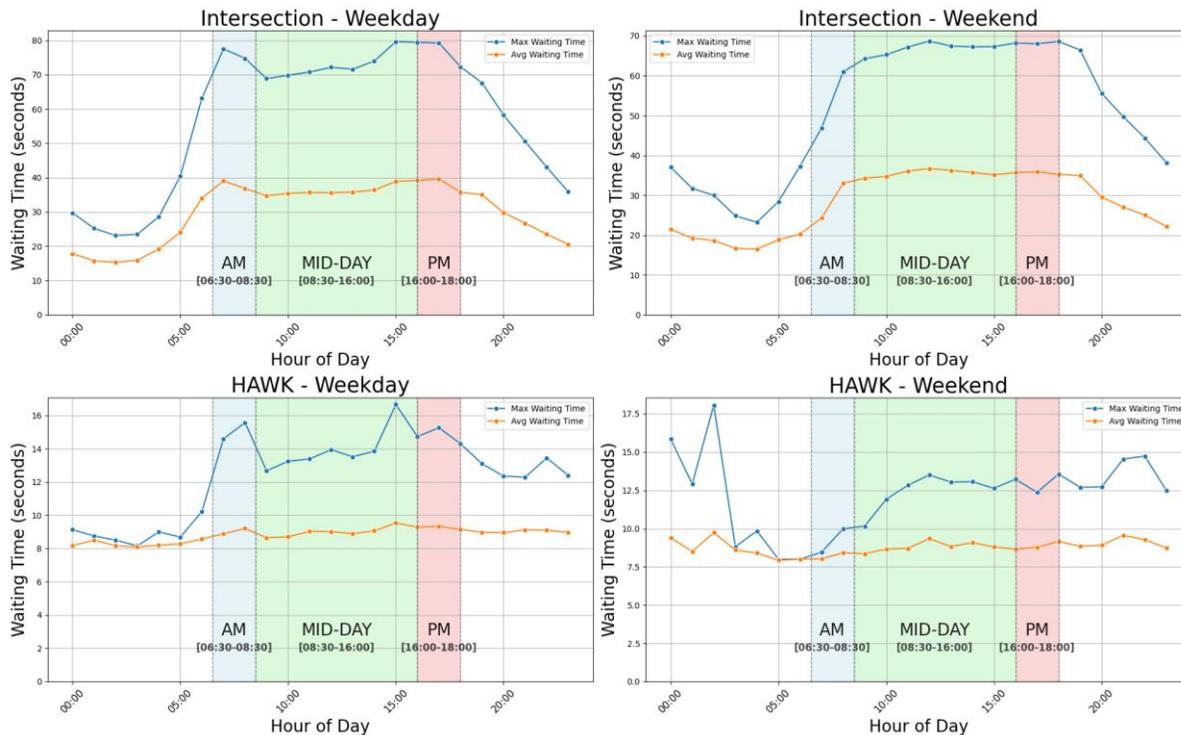


Figure 7.8. Methodology flow chart for calculation of pedestrian delay

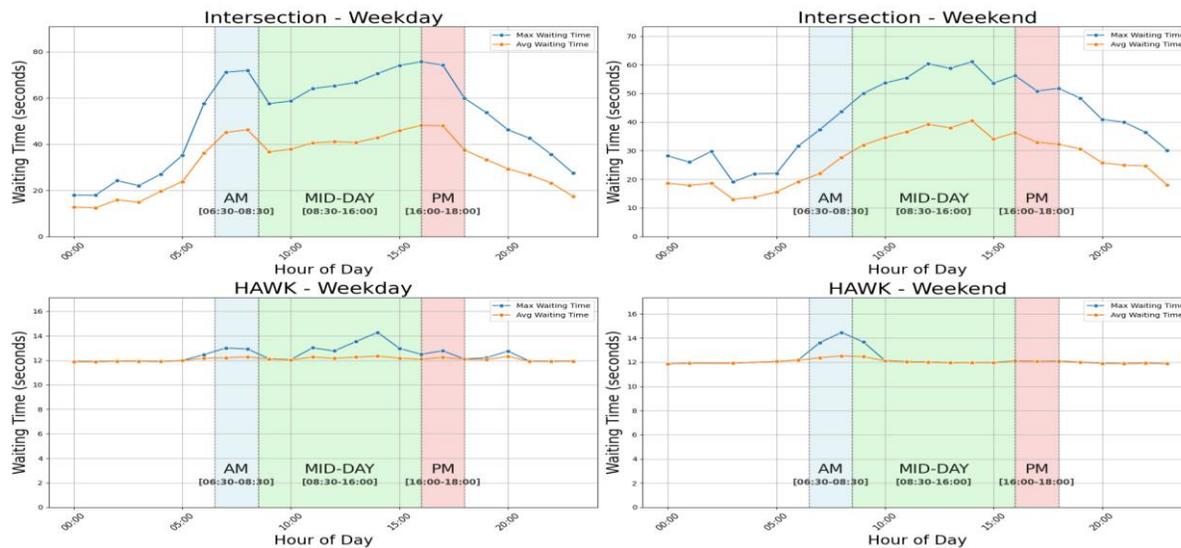
7.3.3 Results

7.3.3.1 Average waiting time and maximum waiting time

Figure 7.9 compares the differences in pedestrian waiting times captured by MaxView and Miovision. There are 256 intersections and 12 HAWKs in the MaxView dataset, and 100 intersections and 2 HAWKs in the Miovision dataset. Both datasets display similar overall patterns across intersections and HAWKs, with higher waiting times during AM peak, mid-day, and PM peak hours, especially at intersections. However, MaxView data shows slightly more pronounced peaks, with waiting times reaching higher maximum values, particularly during PM hours. Conversely, Miovision data presents smoother trends with less variability, particularly at HAWKs, where waiting times remain relatively stable throughout the day.



(a) MaxView intersections

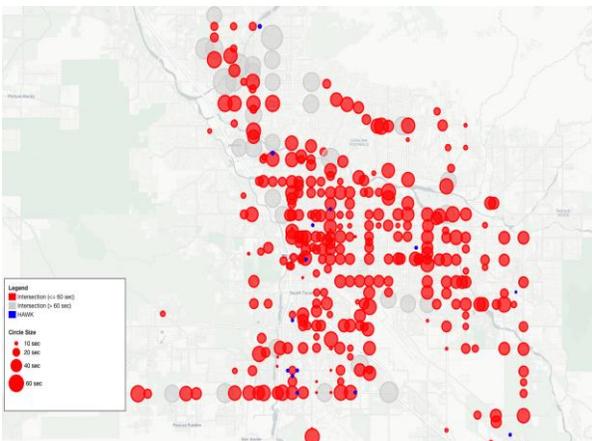


(b) Miovision intersections

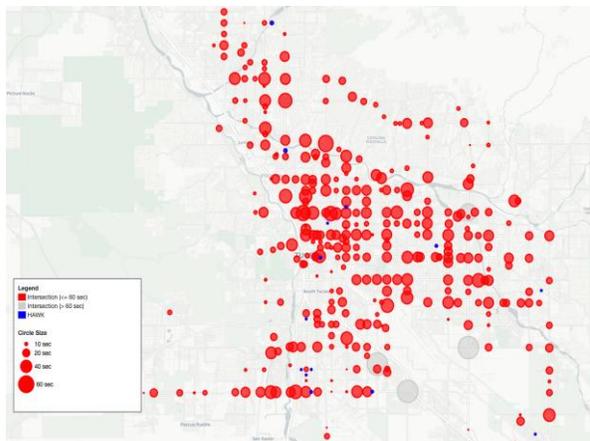
Figure 7.9. Hourly average waiting time by day type and device type at September 2023

7.3.3.2 Average waiting time on map

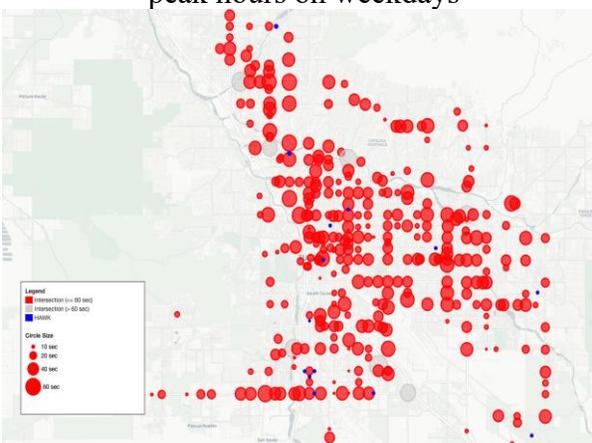
To better check the average waiting time of each crosswalk, the following maps from **Figure 7.10** depict pedestrian waiting times across intersections and HAWK signals in PAG region. Each figure represents different time periods: AM peak hours (6:30 - 8:30 AM), mid-day hours (8:30 AM - 4:00 PM), and PM peak hours (4:00 - 6:00 PM) for both weekdays (Monday to Friday) and weekends (Saturday and Sunday). Across all time slots, intersections with longer waiting times (over 60 seconds) are marked in grey, while those with shorter waiting times are red. The blue circles represent HAWK signals, mostly showing modest waiting times. Tucson's downtown area has the highest density of intersections with varying waiting times, especially during weekday peak hours. Meanwhile, the waiting times across the northwest remain relatively varying as well, with longer waiting times during the AM peak and PM peak hours on weekdays. The data indicates that pedestrian delays peak during PM peak hours on weekdays, with a general trend of increased congestion in central Tucson.



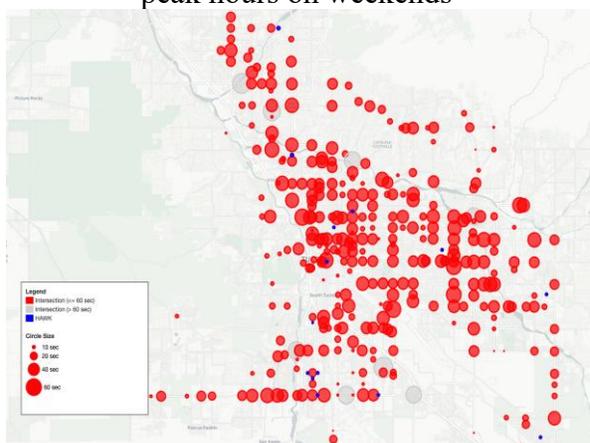
(a) Average pedestrian waiting time in AM peak hours on weekdays



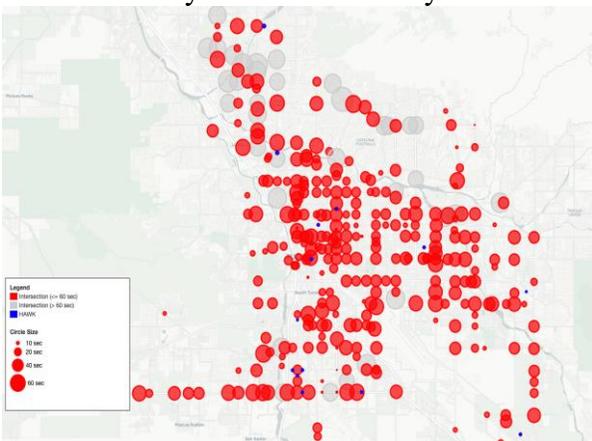
(d) Average pedestrian waiting time in AM peak hours on weekends



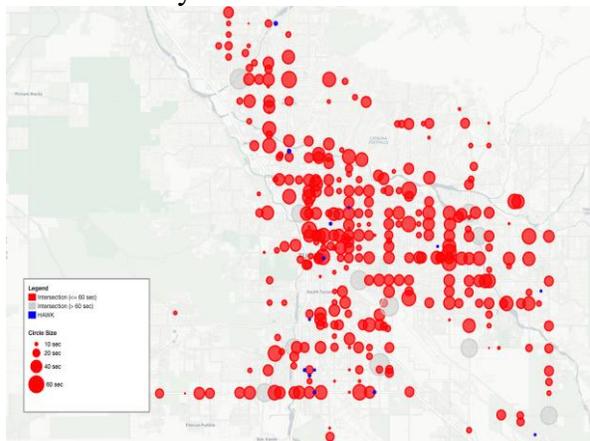
(b) Average pedestrian waiting time in mid-day hours on weekdays



(e) Average pedestrian waiting time in mid-day hours on weekends



(c) Average pedestrian waiting time in PM peak hours on weekdays



(f) Average pedestrian waiting time in PM peak hours on weekends

Figure 7.10. Average pedestrian waiting times in different hours of day and days of week

7.4. Pedestrian level of service

7.4.1 Introduction

Pedestrian level of service (PLOS) evaluates walking conditions along routes, paths, or facilities, reflecting pedestrian perceptions of mobility, comfort, and safety (Daniel et al., 2016). The Highway Capacity Manual (HCM) provides a method for determining LOS based on pedestrian delay, while other studies explore additional factors, both at signalized and unsignalized intersections. Research on PLOS falls into two categories: those using survey data (Campisi et al., 2022) and those relying solely on analytic data (Marisamynathan and Vedagiri, 2019; Ahmed et al., 2021). Some studies focus on intersections, while others examine entire roadway segments (Jensen, 2007).

A study in Gainesville, Florida, used analytic data to score roadways by evaluating pedestrian facilities, conflict points, amenities, multimodal opportunities, and vehicle LOS, correlating scores to LOS. The study emphasized the need for standardized bike and pedestrian LOS measures to predict volume (Dixon, 1996). Hubbard et al. (2008) used real-time data from inductive loop detectors and video to calculate PLOS by integrating pedestrian and vehicle data, combining HCM metrics with new measures such as compromised signal crossings.

Other research incorporates perceived safety through survey data. Petritsch et al. (2005) analyzed perceived conflicts, exposure, and delays, finding that video simulations aligned closely with real-world conditions. Similarly, Landis et al. (2001) a PLOS model based on survey data with roadway conditions, such as separation widths, traffic volume, and vehicle speed. However, Landis et al.'s study did not include pedestrian delay, unlike the others.

While these studies offer comprehensive methodologies for calculating PLOS, they require extensive roadway data. The HCM's delay-based method offers a simplified approach where detailed data is unavailable. The following research will adopt this HCM method, introducing a streamlined PLOS model that draws on key elements from these studies.

7.4.2 Methodology

7.4.2.1 HCM method

Once pedestrian delay is calculated, [Table 7.8](#) and [Table 7.9](#) below from the HCM are used to determine the PLOS grade from delay data for each intersection based on intersection type (signalized or unsignalized).

Table 7.8. Recommended HCM PLOS criteria for signalized crossing delay (Rouphail, Nagui M., et al.)

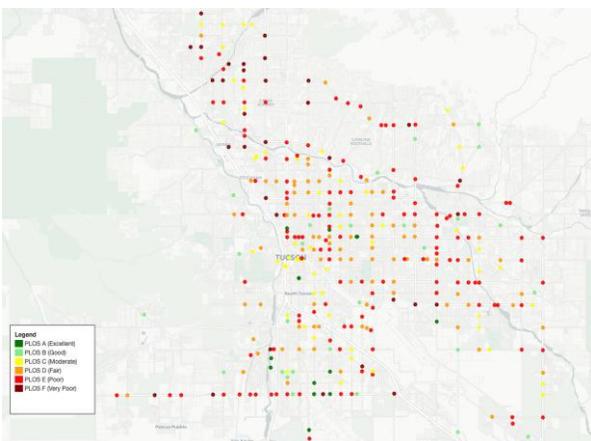
LOS	Average delay per pedestrian (s)	Likelihood of pedestrian noncompliance
A	<10	Low
B	10-20	
C	20-30	Moderate
D	30-40	
E	40-60	High
F	≥60	Very high

Table 7.9. Recommended HCM PLOS criteria for unsignalized crossing delay (Rouphail, Nagui M., et al.)

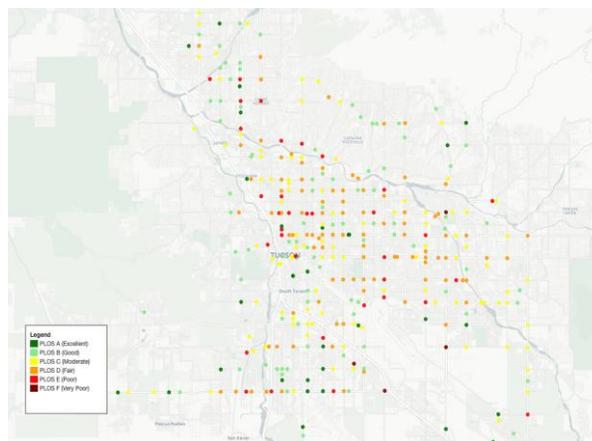
LOS	Average delay per pedestrian (s)	Likelihood of risk-taking behavior by pedestrians
A	<5	Low
B	5-10	
C	10-20	Moderate
D	20-30	
E	30-45	High
F	>45	Very high

7.4.3 Results

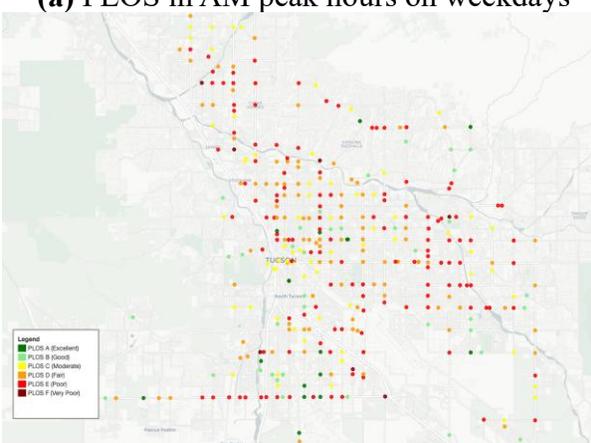
Figure 7.11 shows a series of maps indicating the PLOS across intersections in Tucson and surrounding areas, segmented into different times of day—AM peak hours (6:30 - 8:30 AM), mid-day hours (8:30 AM - 4:00 PM), and PM peak hours (4:00 - 6:00 PM)—during both weekdays and weekends. The color-coded points represent different PLOS grades, ranging from green (PLOS A - Excellent) to red (PLOS F - Very Poor). The analysis reveals a predominance of lower PLOS grades (D, E & F) in central Tucson. Notably, the patterns suggest that pedestrian service levels tend to deteriorate during PM peak hours, while mid-day and AM peak times show more variability. This analysis helps identify critical areas where pedestrian infrastructure improvements are needed to enhance service quality.



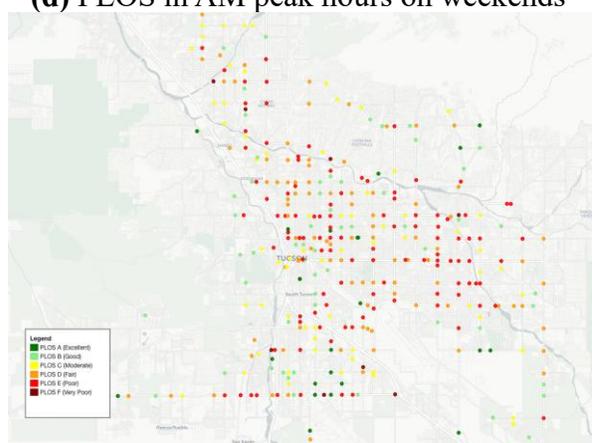
(a) PLOS in AM peak hours on weekdays



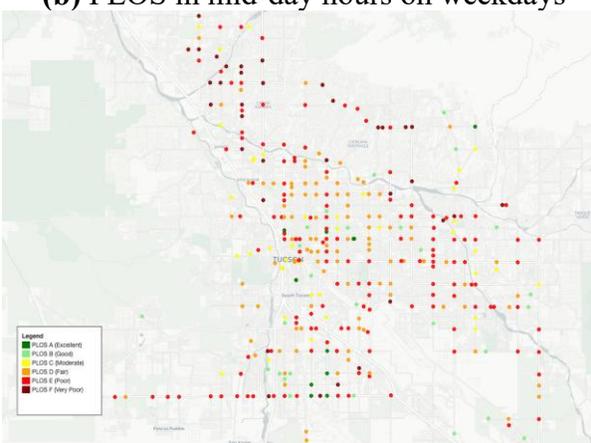
(d) PLOS in AM peak hours on weekends



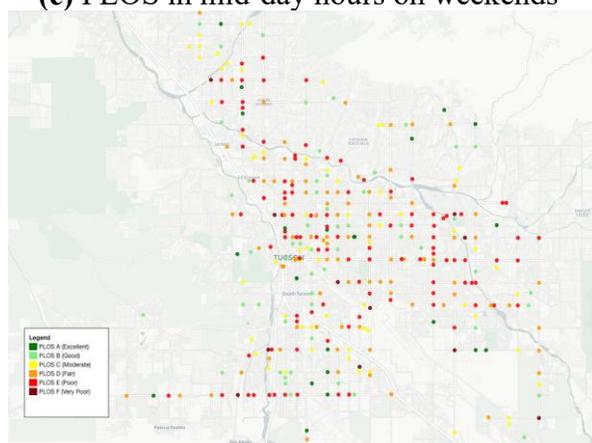
(b) PLOS in mid-day hours on weekdays



(e) PLOS in mid-day hours on weekends



(c) PLOS in PM peak hours on weekdays



(f) PLOS in PM peak hours on weekends

Figure 7.11. PLOS in different times of day and days of week

CHAPTER 8. Micromobility Performance Measure Extraction

This chapter focuses on developing comprehensive performance measures for e-scooters and the TUGO bike sharing system in the region by utilizing General Bikeshare Feed Specification (GBFS) data. As previously discussed in **CHAPTER 5**, GBFS data for e-scooters provides the locations of available vehicles, and TUGO bike share system data details



bike availability at each station. The archived GBFS Real Time (GBFS-RT) data from September 13, 2024, to November 30, 2024, was used for e-scooters, while data from September 23, 2024, to November 30, 2024, was used for the TUGO bike-share system. The “*station_information*” file from the GBFS Schedule (GBFS-ST) provided geographic information about bike share stations.

8.1. Deliverables

The performance measures analysis including availability, utilization rate, and coverage rate, evaluated from both a geographic and system-level perspective. These metrics are outlined in **Table 8.1**. Availability measures the accessibility of e-scooters and bikes across stations, utilization rate quantifies the frequency of use per available unit, and coverage rate evaluates the distribution and reach of service areas within the county. The analysis incorporated time period segments defined as Before AM Peak (12:00 AM to 6:30 AM), AM Peak (6:30 AM to 8:30 AM), Midday (8:30 AM to 4:00 PM), PM Peak (4:00 PM to 6:00 PM), and After PM Peak (6:00 PM to 12:00 AM). The results are visually represented through geographic displays and plots to provide a clear visual assessment of performance across different times of the day.

Table 8.1. Micromobility Performance Measures Summary

Performance Measure	Mode	Location in the Report	Description	Used Data
System level vehicle utilization rate	e-scooter	Section 8.2.1	Measure percentage of vehicles in-use in system level	GBFS-RT: <ul style="list-style-type: none"> • <i>E-scooter</i> • <i>TUGO bike share</i> GBFS-ST: <ul style="list-style-type: none"> • <i>Station information</i>
	Bike sharing	Section 8.3.1		
Geographic vehicle utilization rate	Bike sharing	Section 8.3.2	Measure percentage vehicles in use for each station	
Geographic time availability rate	e-scooter	Section 8.2.2	Measure percentage of time that vehicle(s) are available for users at service areas	
	Bike sharing	Section 8.3.3		
Geographic vehicle coverage rate	e-scooter	Section 8.2.3	Measure the area that can be reached in 5-minute walk from the vehicle over the entire operation area. City of Tucson	
	Bike sharing	Section 8.3.4		

8.2. E-scooter Performance Measures

A comprehensive evaluation was conducted to understand the performance of e-scooters in the region, focusing on various metrics that reflect usage, availability, and accessibility. This analysis delves into system-level utilization rates, which indicate how frequently e-scooters are used within the system, alongside geographic time availability rates, highlighting the temporal availability of these scooters across different areas. Additionally, the geographic vehicle coverage rate was assessed to gauge the spatial distribution and accessibility of e-scooters throughout Pima County. Collectively, these performance measures shed light on the effectiveness of the e-scooter program, providing valuable insights into how these vehicles are integrated into the daily transport dynamics of the region.

8.2.1 *E-scooter system level vehicle utilization rate*

Identifying the usage pattern of e-scooters allows operators to manage resources and meet customer demands effectively. To assess the overall utilization rate, we estimate the system-level utilization rate based on certain assumptions. The total number of e-scooters available throughout the day is assumed to represent the maximum number of vehicles available daily. This figure accounts for the highest number of e-scooters present in the system at any point during the day, excluding any vehicles that are temporarily disabled or out of service.

To calculate the utilization rate, we track the number of available vehicles at each one-minute interval throughout the day. This figure is then subtracted from the maximum daily available vehicle count. The result is divided by the maximum daily available vehicle count and multiplied by 100 to convert this figure into a percentage. This percentage reflects the proportion of e-scooters in use at any given time compared to the total available. The formula used to calculate the utilization rate is outlined below.

$$\text{Vehicle Utilization Rate} = \frac{\text{Maximum daily available vehicle} - \text{available vehicle}}{\text{Maximum daily available vehicle}} \times 100(\%)$$

Using this calculation method for system-level utilization, the e-scooter usage pattern was analyzed using data from September 30 to November 30, 2024. The daily usage patterns of e-scooters were averaged for each day, as depicted in [Figure 8.1](#), to elucidate the typical usage behavior across different days of the week. From Monday to Thursday, the pattern shows a peak in usage at midnight, which then decreases throughout the before AM peak hours. The utilization rate sees a slight increase from the AM peak to the PM peak. However, after the PM peak, there is a surge in usage, with the highest utilization occurring again at midnight. This suggests that e-scooters are frequently used for late evening activities during weekdays. Conversely, from Friday to Sunday, the highest daily utilization rates are observed during the before AM peak hours, indicating a shift in usage patterns on weekends. After reaching the peak during these early morning hours, there is a sharp decline in utilization. Post-PM peak hours on weekends show a different trend compared to weekdays; there is a slight drop in the utilization rate immediately after the PM peak, followed by an increase during the later after PM peak hours. The result shows a similar pattern with the e-scooter usage from October 7th to October 13th, 2024, in [Figure 5.5](#).

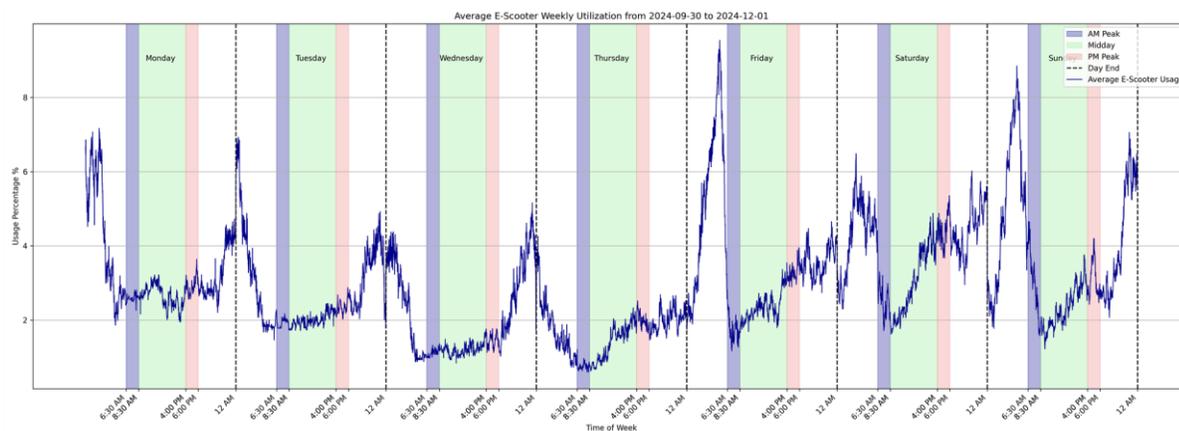


Figure 8.1. E-scooter system level vehicle utilization

8.2.2 *E-scooter geographic Time availability rate*

Ensuring e-scooters' availability to prospective users contributes to effective system operations. To enhance our understanding of e-scooter availability, an evaluation of the time availability of the e-scooter has been conducted.

To calculate time availability, a grid with cells measuring 364.0 feet by 308.2 feet was overlaid over the Tucson metro area, and the percentage of time during each hour that at least one e-scooter was available within each cell was evaluated. This metric provides a direct indication of how readily accessible e-scooters are to users throughout different times of the day. By monitoring this availability, service operators can identify patterns of e-scooter usage and respond to changes in demand (Netes, 2018).

$$\text{Time Availability Rate (\%)} = \frac{t_{va}}{60} \times 100$$

t_{va} : time at least one vehicle available (minutes)

For this performance evaluation, the GBFS data from September 13, 2024, to November 30, 2024, was analyzed to calculate the time availability. The day types of analysis are weekdays, from Monday to Friday, and weekends, Saturday and Sunday. The average time availability for each hour was calculated for these day types, illustrating the typical e-scooter availability during different periods of each day type.

8.2.2.1 *Weekday e-scooter geographic time availability rate*

The analysis of e-scooter time availability on weekdays, as illustrated in [Figure 8.2](#), shows that there are consistent patterns of e-scooter availability and activity across various parts of the day and different locations in Tucson.

Throughout the day, availability is highest in and around the areas including downtown Tucson, 4th Ave, the University of Arizona, and Reid Park. Regular clusters of activity can be seen in other parts of the city, such as along Oracle Rd between Miracle Mile and Wetmore Rd in the 5-6PM period, the San Juan Plaza, and Broadway Campbell Plaza, indicating regular use of e-scooters for medium-length trips. Additionally, occasional activity is seen as far west as Mission Rd/36th St, and as far east as Harrison Rd, suggesting some use even for longer trips.

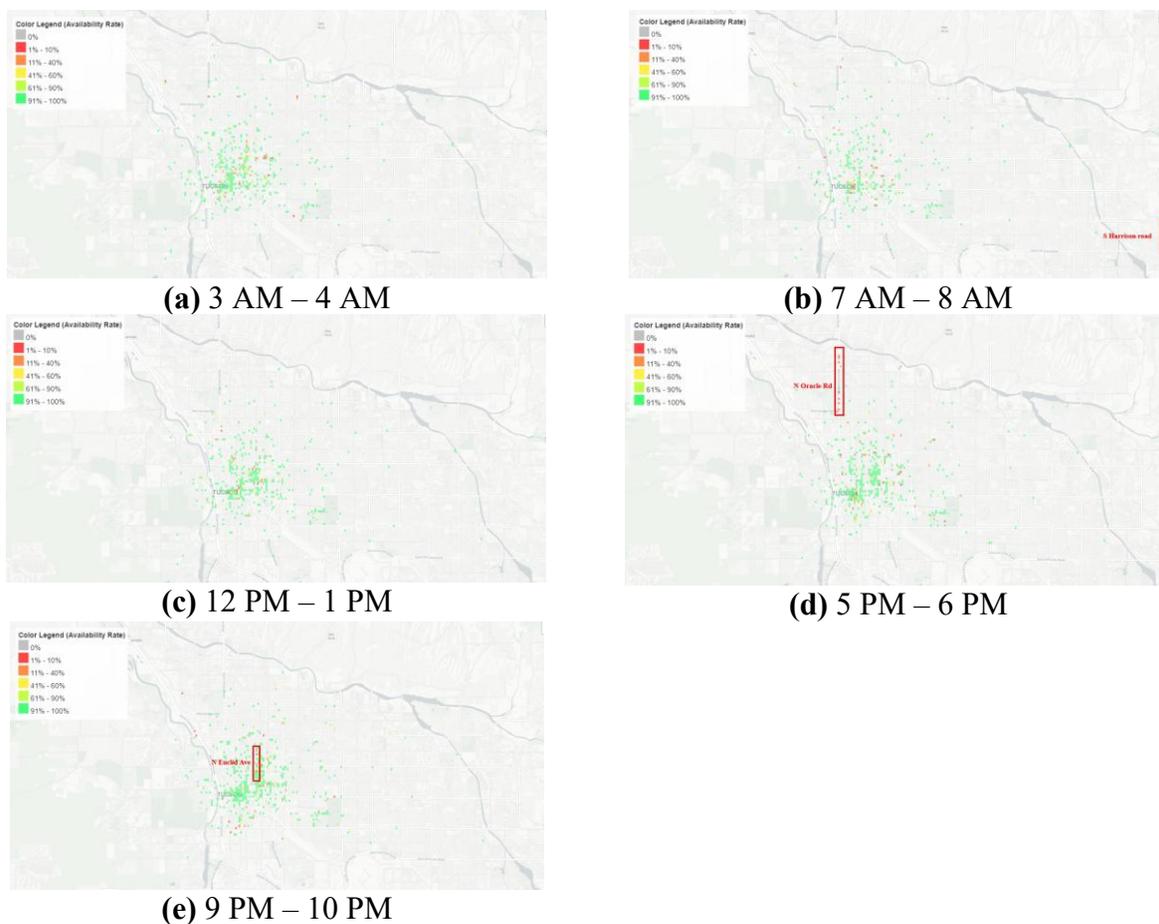


Figure 8.2. Weekday time availability rate

Figure 8.3 shows analysis of e-scooter geographic time availability rate on weekends, within the same data period previously mentioned.

The patterns are largely similar to weekdays, with activity clustered in downtown Tucson, 4th Ave, University of Arizona, Reid Park, and nearby shopping plaza areas. These patterns indicate that the primary users of e-scooters are based in these areas and are using them primarily for short and medium length trips and could be more likely to use other modes for longer trips.

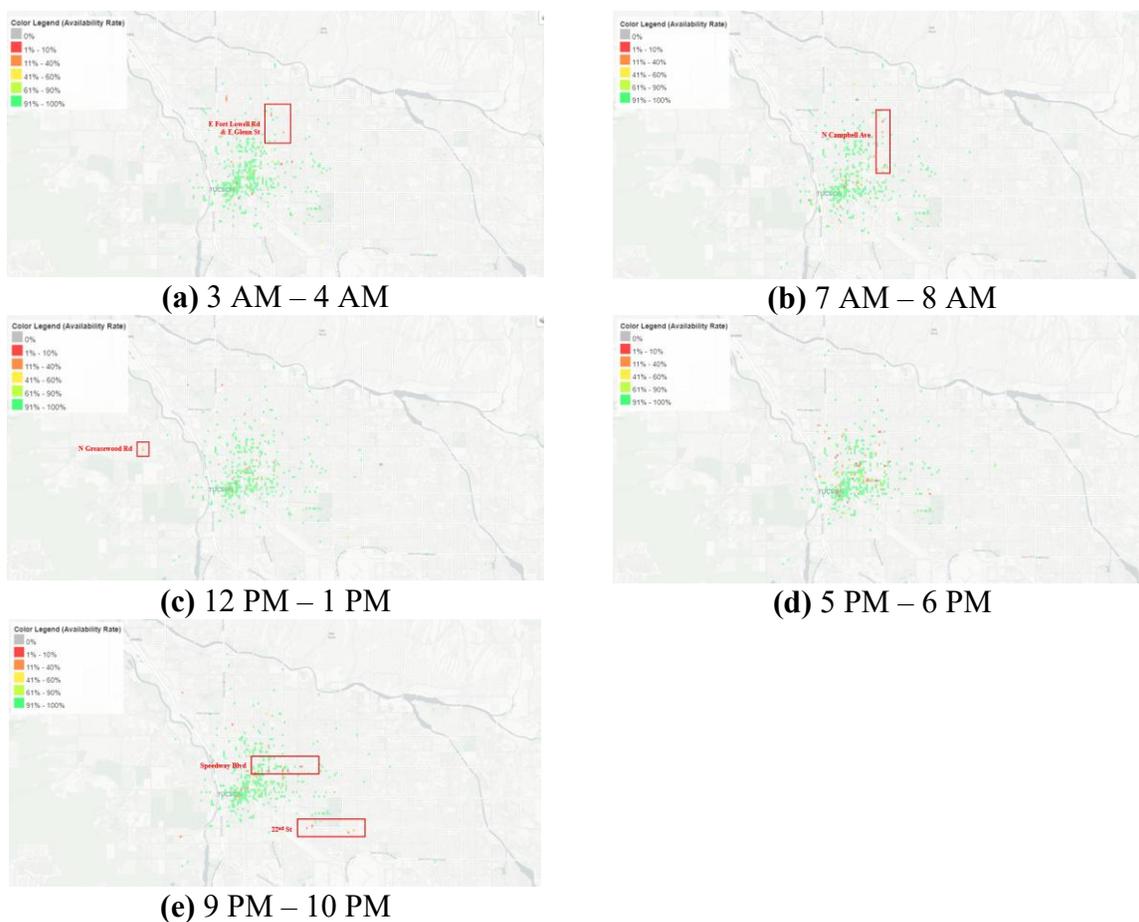


Figure 8.3. Weekend Time availability rate

8.2.3 *E-scooter geographic vehicle coverage rate*

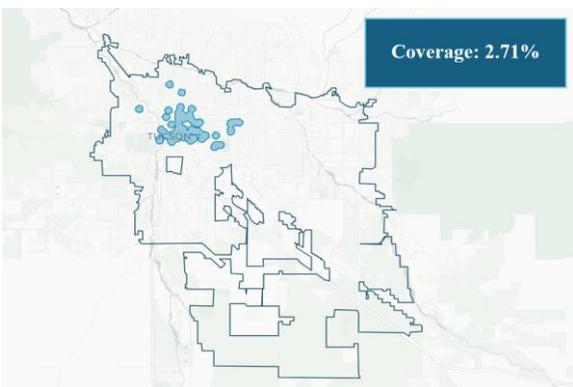
To evaluate e-scooter accessibility within Tucson, vehicle coverage is quantified using catchment areas, defined as the regions reachable within a 5-minute walk from any given e-scooter's location. These individual catchment zones were aggregated to determine the total area covered. The final metric is expressed as the proportion of Tucson's total area that falls within these combined catchment zones. This calculation is performed for each hour and across various dates, and for each period, the observation with the maximum coverage value was selected and added to a map of the Tucson operation area.

This visual and quantitative analysis of vehicle coverage provides a clear view of the areas being served by the e-scooter service.

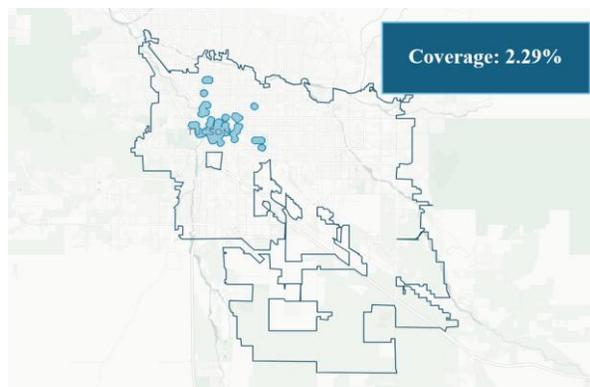
8.2.3.1 Weekday e-scooter geographic vehicle coverage rate

The e-scooter coverage analysis for weekdays reveals distinct patterns of geographic distribution and coverage rates across various time periods, as detailed in [Figure 8.4](#).

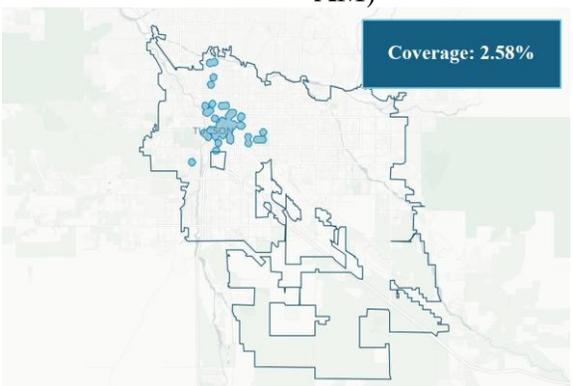
During the analyzed weekdays, maximum coverage was at its lowest during the AM peak period, at 2.29%, and at its highest in the post-PM peak, at 2.88%, supporting the findings of the previous section, e-scooter usage and availability is concentrated in and around downtown Tucson, 4th Ave, the University of Arizona, and other central attractors.



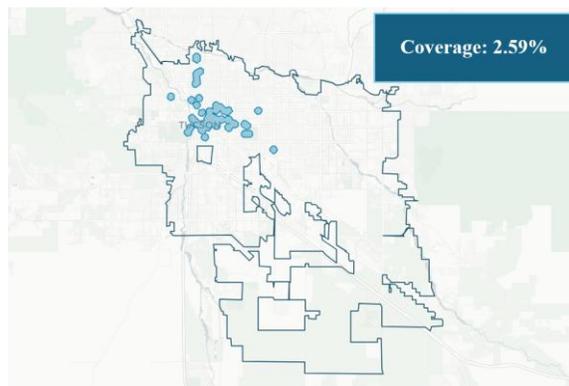
(a) Before AM Peak (12:00 AM to 6:30 AM)



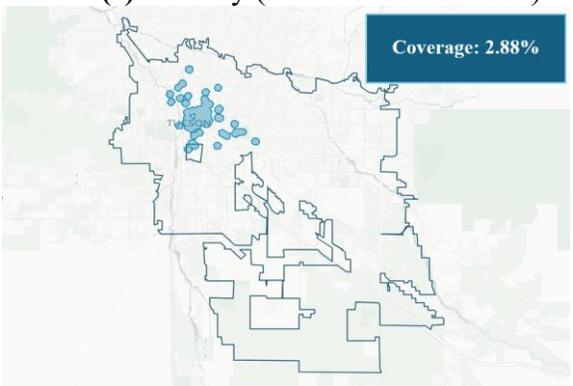
(b) AM Peak (6:30 AM to 8:30 AM)



(c) Midday (8:30 AM to 4:00 PM)



(d) PM peak (4:00 PM to 6:00 PM)



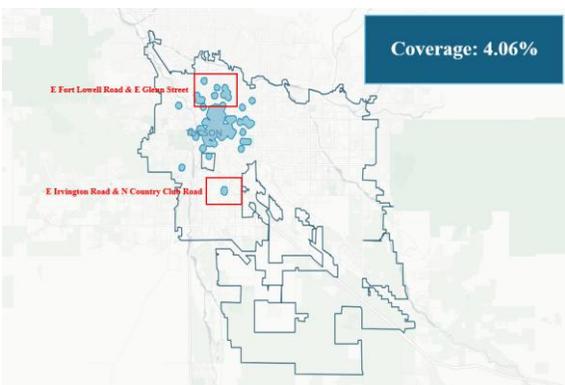
(e) After PM peak (6:00 PM to 12:00 AM)

Figure 8.4. Weekday vehicle coverage rate

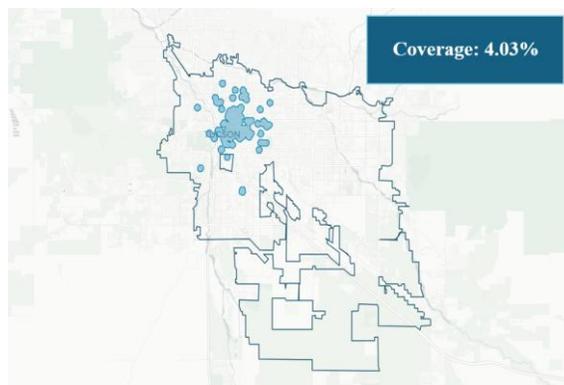
8.2.3.2 Weekend e-scooter geographic vehicle coverage rate

During the weekends, the e-scooter coverage pattern across Tucson exhibits distinctive features and generally higher rates compared to weekdays, as highlighted in [Figure 8.5](#). The average maximum coverage on weekends is around 4%, indicating a more extensive use and distribution of e-scooters during these days.

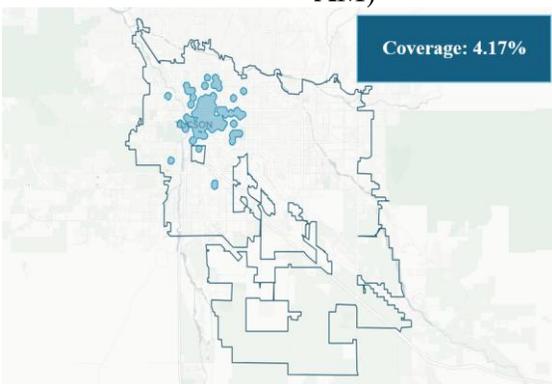
As compared to weekdays, greater activity is seen in commercial areas, such as the strip malls and restaurants east of the University of Arizona along Speedway Blvd from Campbell Ave to Alvernon Way, as well as those north of campus, along Glenn St and Ft. Lowell Rd along and west of Campbell Ave; additional activity also occurs in recreational areas, such as the Sam Lena South Tucson Library and Silverlake Park.



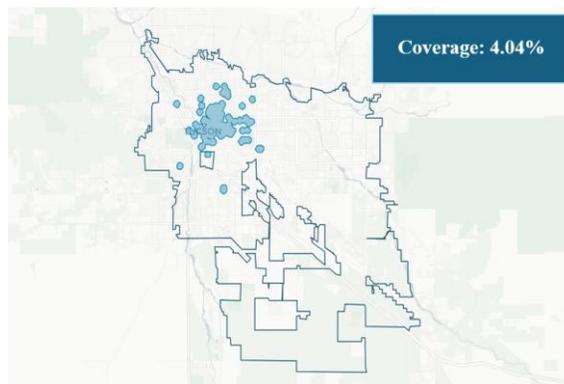
(A) Before AM Peak (12:00 AM to 6:30 AM)



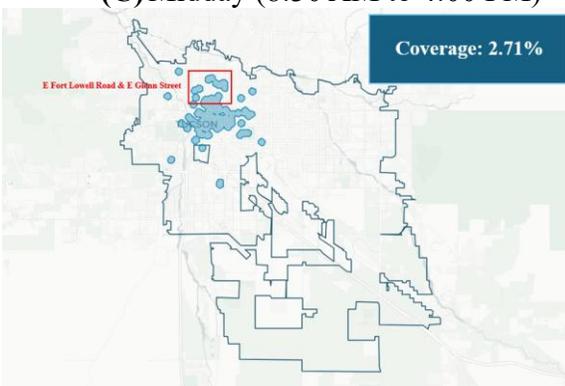
(B) AM Peak (6:30 AM to 8:30 AM)



(C) Midday (8:30 AM to 4:00 PM)



(D) PM peak (4:00 PM to 6:00 PM)



(E) After PM peak (6:00 PM to 12:00 AM)

Figure 8.5. Weekend vehicle coverage rate

8.3. TUGO Bike-share system Performance Measures

To assess the performance of the bike-sharing system in Pima County, operated by TUGO, a thorough examination of various performance measures is conducted. This analysis includes exploring the system-level utilization rate, which provides insights into the frequency of bike usage across all stations, highlighting the overall activity level within the system. Additionally, the geographic vehicle usage rate is evaluated to pinpoint specific locations within the system that experience higher demand.

Further, the geographic time availability rate is measured to determine the consistency of bike availability at each station over time. This metric is crucial for assessing the reliability of the bike sharing system from the perspective of users, ensuring that bikes are accessible when and where they are needed. Lastly, the geographic vehicle coverage rate is analyzed to gauge the extent of the service area covered by the bike sharing stations. This involves studying the catchment areas of each station to see how much of the city is accessible to potential users within reasonable walking distance.

Together, these evaluations provide comprehensive information about how the bikes are used and the accessibility and availability of the bike sharing system in Pima County. By understanding these aspects, stakeholders can make informed decisions to improve the system's efficiency, expand its reach, and enhance user satisfaction, all of which contribute to the bike-sharing system's success and sustainability.

8.3.1 TUGO Bike-share system level vehicle utilization rate

Identifying the usage patterns of the bike sharing system is crucial for operators to effectively manage resources and meet customer needs. Understanding the overall utilization rate is a key aspect of this analysis. The methodology used to estimate the system-level utilization rate for bikes is similar to that applied to e-scooters. The process begins by assuming the total number of bikes available in the system each day as the maximum daily available vehicle numbers. This figure represents the highest count of bikes that are operational and present in the system, excluding any that are disabled or out of service.

To calculate the utilization rate, the number of bikes available at each minute throughout the day is recorded. This count is then subtracted from the maximum daily available vehicle count. The result of this subtraction is divided by the maximum daily available vehicle count and then multiplied by 100 to convert this figure into a percentage. This percentage effectively quantifies the proportion of time that bikes are actively in use as compared to being available within the system.

This utilization rate calculation provides a dynamic view of how frequently the bikes are rented and returned throughout the day, offering valuable insights into user behavior and system performance. This metric is critical for operators to optimize bike allocations, maintain sufficient availability, and ensure the system meets the mobility needs of the community efficiently. The formula used can be summarized as follows.

$$\text{Vehicle Utilization Rate} = \frac{\text{Maximum daily available vehicle} - \text{available vehicle}}{\text{Maximum daily available vehicle}} \times 100(\%)$$

With the established assumptions, a system-level analysis of the bike-sharing usage patterns was conducted using data from September 30 to November 30, 2024. The results, depicted in **Figure 8.6**, illustrate distinct daily utilization patterns for the bike-sharing system that differ between each day.

From Monday to Thursday, the data shows that the daily utilization rate of the bike-sharing system is at its lowest before AM Peak (6:30 AM to 8:30 AM) hours. As the day progresses, there is a noticeable increase in bike usage from the AM peak to the midday hours, indicating a higher demand during the morning commute and lunchtime. During the PM peak hours, there is a slight dip in utilization early on, followed by an increase as the evening commute begins. After the PM peak hours, the utilization rate gradually decreases, reflecting a reduction in bike-sharing usage as the day ends. Conversely, from Friday to Sunday, the pattern shifts slightly. The utilization rate remains lowest from before AM peak hours to AM peak hours, suggesting a later start to bike usage on weekends. The rate then climbs during the early midday hours as more users engage with the system for leisure or errands, peaking before gradually decreasing throughout the latter part of the midday. This peak is followed by a consistent decrease in bike usage, which stabilizes from the PM peak hour through to after PM peak hours, indicating a more subdued evening activity level on weekends compared to weekdays.

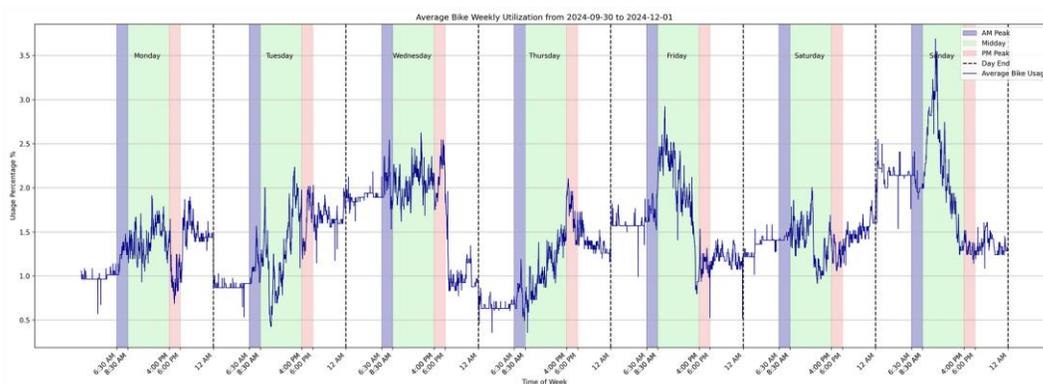


Figure 8.6. TUGO bike-share system level vehicle utilization

8.3.2 *TUGO Bike-share system geographic vehicle utilization rate*

To understand how the bike share system usage is distributed across the TUGO bike-sharing stations throughout Pima County, the station-level utilization rate was assessed. This specific metric identifies which stations are most used and which may require adjustments in bike availability or station capacity.

The station-level utilization rate was calculated by subtracting the number of available bikes from the total station capacity and divided by the station capacity to provide a percentage, indicating the utilization rate of that particular station. The formula can be expressed mathematically as follows.

$$\text{Vehicle Utilization rate} = \frac{\text{station capacity} - \text{available vehicle}}{\text{station capacity}} * 100$$

The station capacity for the TUGO bike-sharing system is sourced from the GBFS-ST data, specifically from the station information component. Concurrently, the number of available vehicles at each station is determined using the GBFS real-time data, which provides updates on the availability of bikes in real time. The utilization rate calculation is based on the assumption that the number of empty docking stations at each location serves as an indicator of the station's usage. The utilization rate is recorded every minute, and then averaged for each day type during the data time period.

8.3.2.1 Weekday TUGO Bike-share system geographic vehicle utilization rate

The utilization rates of the TUGO bike-sharing system's stations are calculated based on the previously mentioned assumption that the number of empty docking stations indicates station usage. As illustrated in [Figure 8.7](#), stations with lower utilization rates are predominantly located around the campus and downtown areas. In contrast, most stations exhibit medium utilization rates, with rates between 61% to 90% particularly noted at stations positioned outside the downtown and campus areas, specifically around 6th and 9th Streets. High utilization rates are observed at stations near the north and south of the university campus, and notably at a station on N Ash Avenue.

This pattern of utilization rate distribution remains consistent across various time periods. The consistency in the available number of bikes at each station throughout different times indicates that the utilization of the bike-sharing system is relatively unchanged over time. This could be due to a mix of factors, such as a large number of two-way trips self-balancing the stations, relatively low overall usage, keeping the stations stable at their target availability rates, or active rebalancing of the number of available bikes at each station.

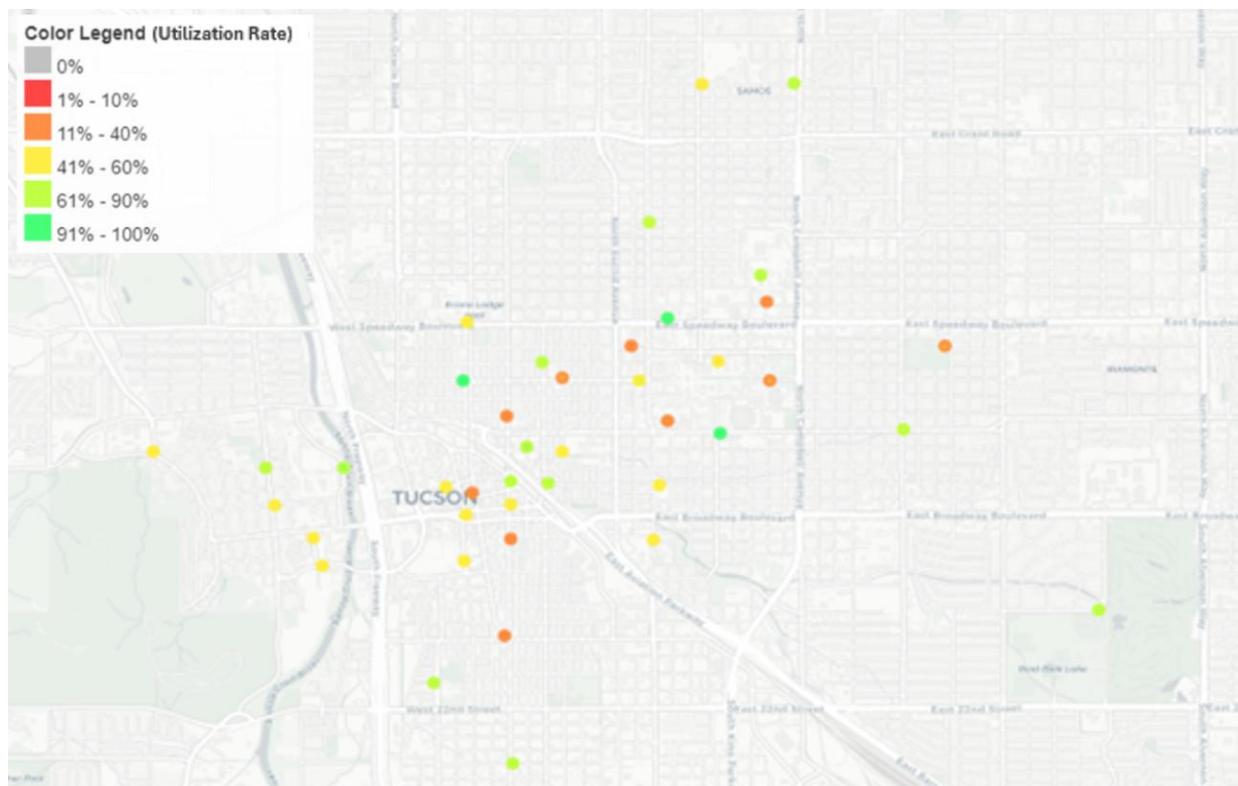
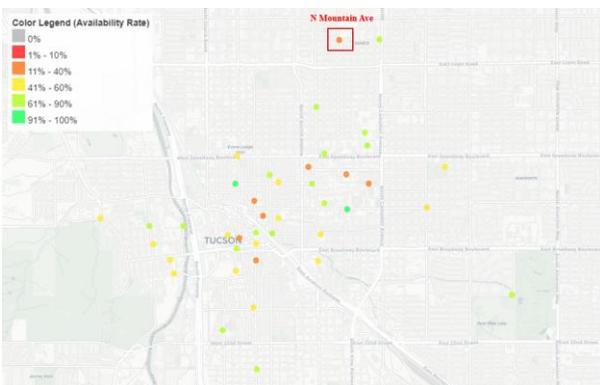


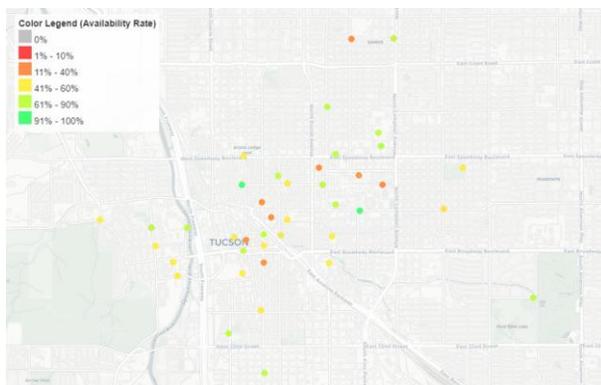
Figure 8.7. Weekday utilization rate

8.3.2.2 Weekend TUGO Bike-share system geographic vehicle utilization rate

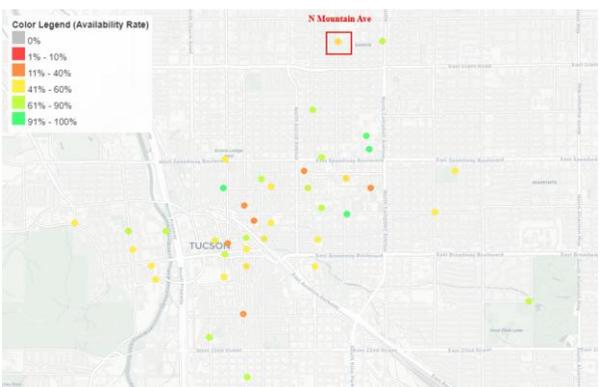
During the weekends, the TUGO bike-sharing system exhibits a utilization pattern that distinctly varies from weekdays, as detailed in [Figure 8.8](#), which maps the vehicle utilization rates at each station across different time periods. In the early morning before the AM peak at 3:00 AM, stations in the downtown and university campus areas, along with a station on N Mountain Avenue, show lower utilization rates, ranging from 11% to 40%. Conversely, higher utilization rates are observed at other stations within the campus areas and at central downtown locations. As the day progresses to the AM peak hour at 7:00 AM, this pattern of utilization continues with stations outside of the downtown and campus core displaying medium utilization rates, typically between 41% and 60%. This indicates a more moderate level of bike usage in areas peripheral to major activity centers. By midday, an increase in utilization rates is particularly notable at northern campus stations and again at the station on N Mountain Avenue, suggesting a concentration of bike-sharing activity in these areas. The rest of the stations maintain similar utilization rates as earlier in the day. This trend of increased utilization at specific stations persists in the afternoon and after the PM peak hours. Remarkably, the overall utilization rate across the downtown and campus areas during weekends is higher compared to weekdays, pointing to a shift in bike usage patterns.



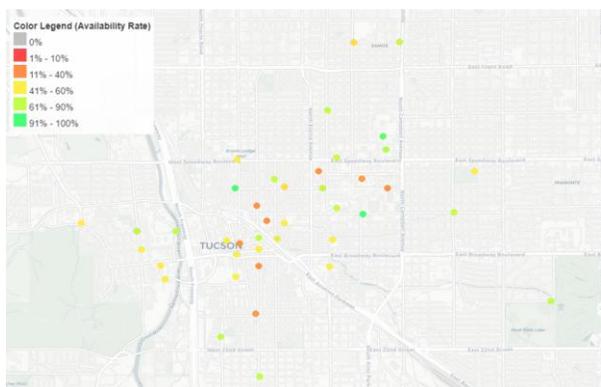
(a) 3 AM – 4 AM



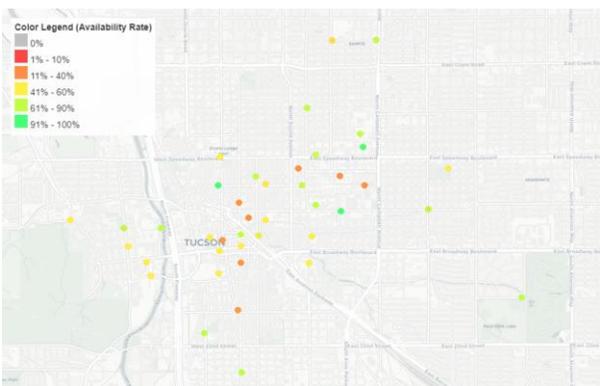
(b) 7 AM – 8 AM



(c) 12 PM – 1 PM



(d) 5 PM – 6 PM



(e) 9 PM – 10 PM

Figure 8.8. Weekend time availability rate

8.3.3 *TUGO Bike-share system geographic time availability rate*

To effectively understand how bike availability is distributed across the operating stations of the TUGO bike-sharing system, the station-level availability rate was assessed using a method analogous to that used for evaluating e-scooter availability. However, unlike the grid-based analysis used for e-scooters, the availability rate for the bike-sharing system is calculated specifically at the station level. This approach reflects a key difference in the geographic analysis definition level, tailoring it to the distinct infrastructure and operational dynamics of bike-sharing.

The station-level availability rate is determined by the presence of at least one bike at each station. If a station has at least one bike available for use, it is considered available. This binary measure provides a straightforward indicator of whether a station can serve users at any given time, focusing on the essential function of providing at least one bike. The availability rate is then averaged for each day type over different days during the data time period.

The necessary data for this analysis is sourced from the GBFS system. The station geographic information, including location and capacity, is pulled from the GBFS static data under "station information." Concurrently, the number of bikes available at each station is updated in real time and obtained from the GBFS real-time data. This setup ensures that the analysis reflects current conditions and can respond dynamically to changes in bike availability throughout the day.

By employing this methodology, stakeholders can obtain a clear and actionable understanding of bike availability across the TUGO network. This information is crucial for managing the system efficiently, ensuring that bikes are available where and when they are needed, thus enhancing the overall user experience and system reliability.

8.3.3.1 *Weekday TUGO Bike-share system geographic time availability rate*

As depicted in **Figure 8.9**, the availability of the TUGO bike-sharing system remains largely consistent across various times of the day. Under the criterion that a station is considered available if at least one bike is present, all stations consistently meet this threshold throughout the day. This high level of availability underscores the system's reliability and the effectiveness of its operational management.

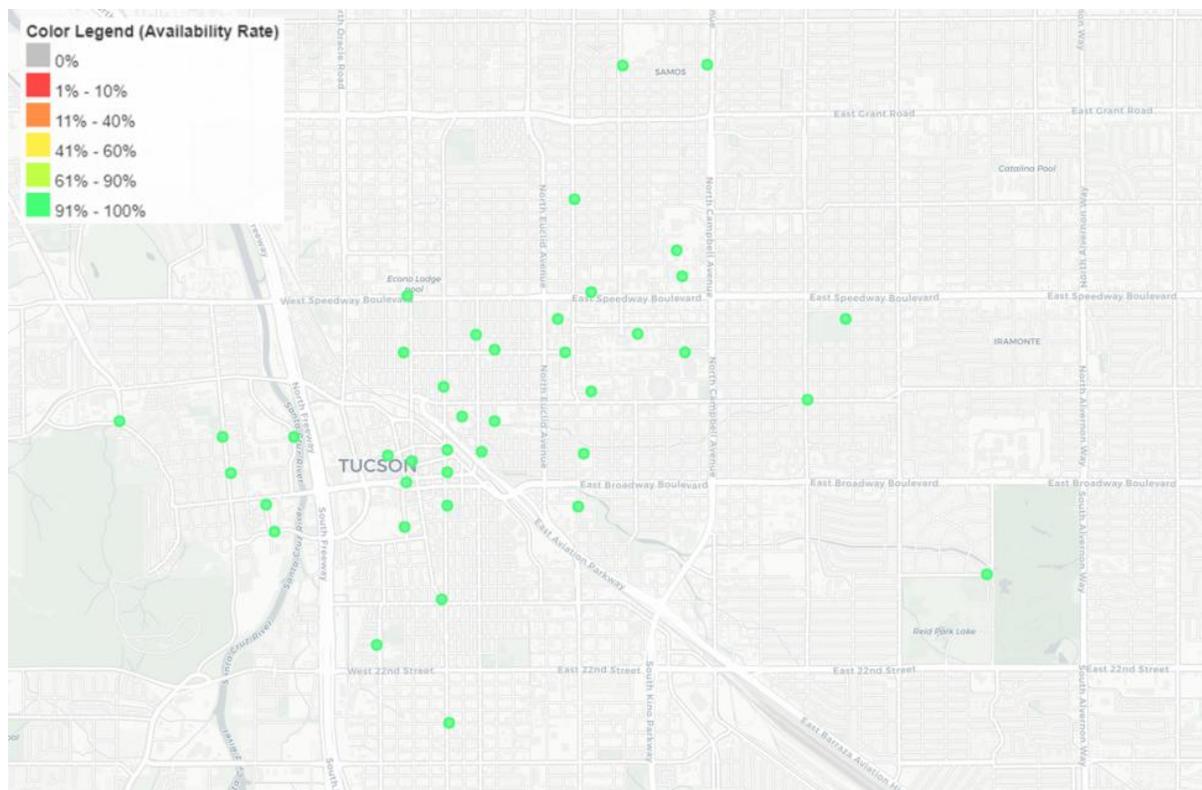


Figure 8.9. Weekday time availability rate

8.3.3.2 Weekend TUGO Bike-share system geographic time availability rate

During the weekends, the station-level time availability rate for the TUGO bike-sharing system mirrors the pattern observed during weekdays, as illustrated in [Figure 8.10](#). Bikes are available for all the stations throughout all time periods.

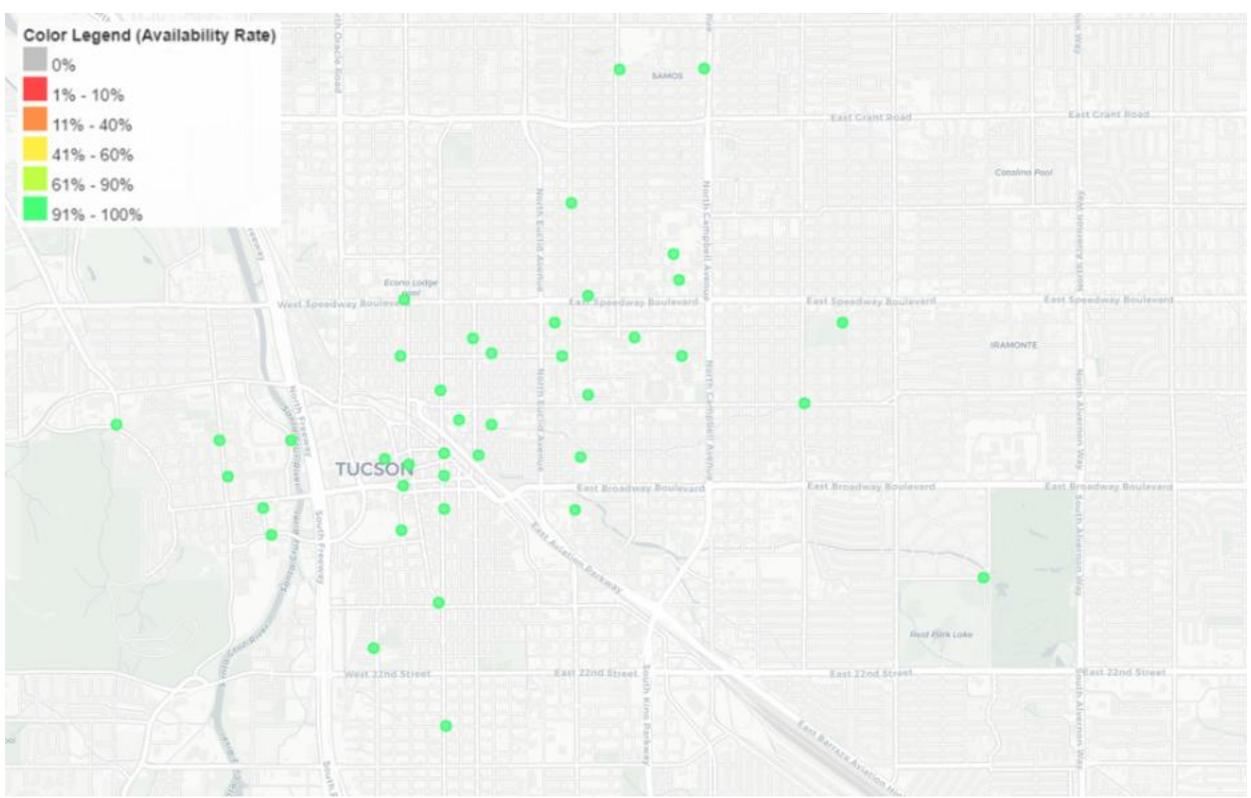


Figure 8.10. Weekend time availability rate

8.3.4 *TUGO Bike-share system geographic vehicle coverage rate*

To assess the extent of the area served by the available bikes in the TUGO bike-sharing system, the vehicle coverage rate is analyzed using a methodology similar to that employed for the e-scooter coverage rate. This approach involves determining the catchment area, which is defined as the walkable distance reachable from each bike station within a five-minute walk. The premise is that if a station has at least one bike available, the area around the station counts as covered; if no bikes are available at a station, that station's surrounding area is not included in the covered area.

The coverage rate is calculated by dividing the total area covered by bike stations with available bikes by the total operational area of the bike-sharing system. This ratio provides a quantifiable measure of how much of the operational area is accessible to users via the bike-sharing service. Consistent with the method used for calculating the e-scooter coverage rate, the maximum coverage rate observed during each day type's specific time period is selected to represent the service's availability.

As depicted in [Figure 8.11](#), the coverage rate during weekdays remains stable consistently measuring at 1.94%. This is consistent with the findings in the previous section, as all stations were found to have, on average, at least one bike during all evaluated periods. While low as a percentage of the Tucson area, the concentration of Tugo bike stations around the University of Arizona, 4th Ave, downtown Tucson, and nearby attractors such as Tumamoc Hill, Himmel Park, Reid Park, and the Banner-University Medical Center highlights its use for traditionally car-less populations.

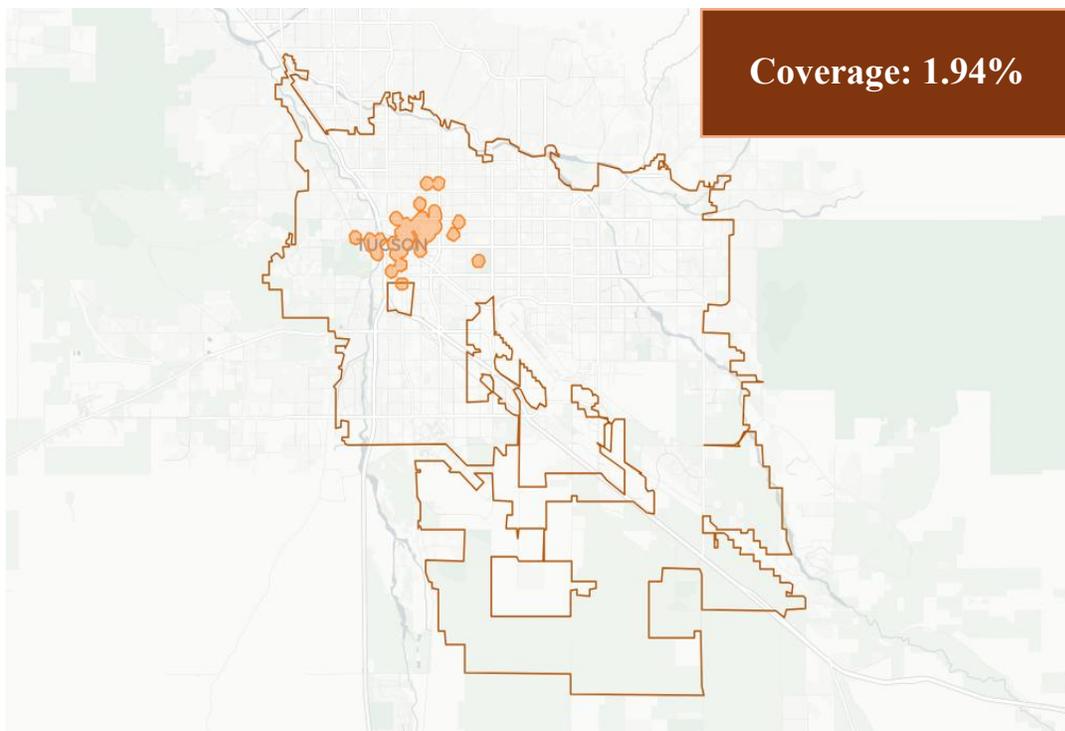


Figure 8.11. Vehicle coverage rate

CHAPTER 9. DATA QA/QC AND PERFORMANCE MEASURE TARGET DEVELOPMENT

This chapter is designed to ensure the accuracy, consistency, and reliability of output data generated through the proposed performance measurement methodologies. Spanning CHAPTER 3 through 8 focuses on validating the integrity of input data and assessing the performance of analytical models used



throughout each step of the workflow. As each transportation mode, transit, pedestrian, and micromobility, relies on different data sources and employs unique methodologies, the quality assurance process is tailored to address the specific requirements and challenges of each mode. This chapter includes systematic checks for data completeness, temporal and spatial consistency, and alignment with expected values based on transportation theory or historical trends. These efforts help ensure the credibility and robustness of all subsequent performance analyses.

9.1. Deliverables

As summarized in [Table 9.1](#), the quality checks for transit and micromobility primarily focus on verifying the integrity of and completeness of archived GTFS and GBFS data. In contrast, pedestrian performance relies on a volume estimation model, necessitating a model-focused quality review to assess its predictive accuracy and reliability. Details regarding the data periods, quality assurance methodologies, and findings are elaborated in the following sections.

Table 9.1. Data QA/QC and Performance Measure Validation Summary

Mode	QA/QC or Validation	Description	Data
Transit	GTFS Trip Update data quality control	Compare sample of GTFS Trip Update to sample of GTFS Vehicle Positions data for vehicle trip data accuracy check	<ul style="list-style-type: none"> • <i>GTFS Trip Update data</i> • <i>GTFS Vehicle Positions data</i>
	Transit performance measures result validation	Compare resulted performance measure average to the literature reviewed values	<ul style="list-style-type: none"> • <i>GTFS Trip Update data</i>
Pedestrian	Model performance metrics	Measure performance metrics, R ² , RMSE, and MAPE, for estimation models with sample data	<ul style="list-style-type: none"> • <i>MaxView event-based data</i> • <i>PAG TMC pedestrian data</i> • <i>Miovision event-based data</i> • <i>Miovision video detection data</i>
Micromobility	GBFS data quality control	Compare sample GBFS data to observed e-scooters and Tugo bike sharing status	<ul style="list-style-type: none"> • <i>E-scooter GBFS vehicle status data</i>

9.2. Transit QA/QC and Validation

To review how GTFS Trip Update data is accurate and appropriate for the performance measures, the data is reviewed with Trip Update data comparison with Vehicle Positions data.

9.2.1 Trip Update and Vehicle Positions comparison

As mentioned in **CHAPTER 3**, GTFS Trip Update uses predicted vehicle arrival and departure times for stops along a route (Mobility Data, 2023). To understand how well the estimation is predicted in the GTFS Trip Update data, this data is compared to the GTFS Vehicle Positions data which provides the real time vehicle locations with vehicle ID, trip ID, latitude, and longitude.

As shown in **Figure 9.1**, Vehicle Positions data from a single vehicle can be reported with different timestamps (11:52 am at location V1, 11:53 am at location V2, and 11:55 am at location V3) along a specified route. And independently, Trip Update estimates arrival time of the bus, 11:57 am, at the stop. So, there is a 2-minute gap in terms of data reporting between the latest vehicle position data and trip update data.



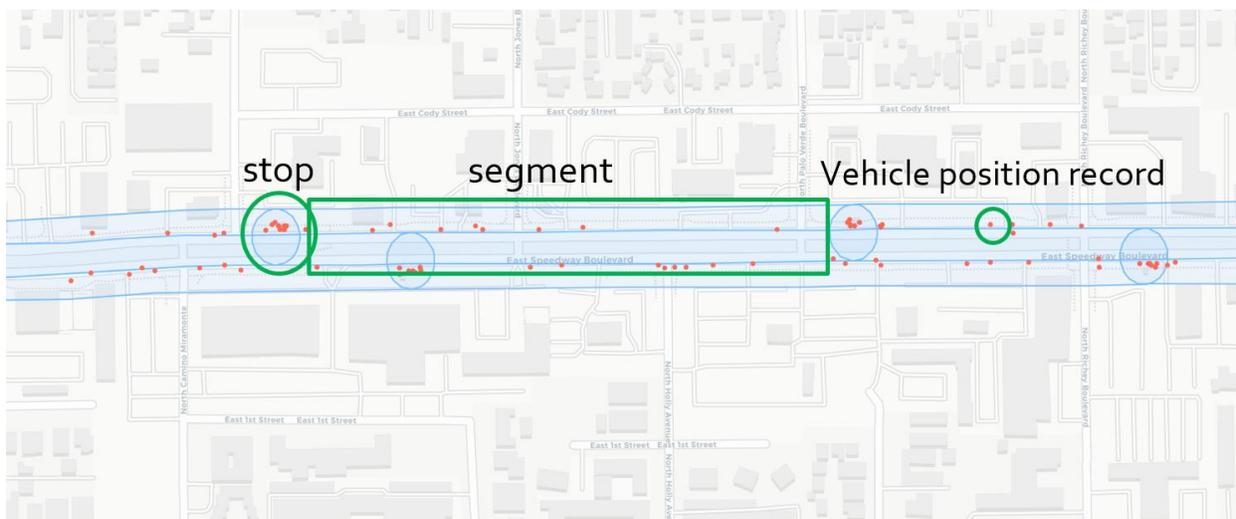
Figure 9.1. Trip Update data and Vehicle Positions data comparison metric example

Figure 9.2 displays the Vehicle Positions record assigned to each segment, matched to pre-processed segment geometry. Each red point shows the location of a bus at a specific time, based on its GPS location. The blue lines represent street segments where buses travel between stops. These segments were derived from GTFS-ST stops file with two consecutive stops along a trip. A buffer area (approximately 65 ft wide) was added around each segment to check which buses passed close to the path. When two segments overlap, blue circle in **Figure 9.2 (b)**, it represents the location of a bus stop. Each Vehicle Positions GPS point is linked to the closest street segment it falls within. Then, for each Trip Update departure time record, a Vehicle Positions GPS point is assigned by finding the one within the stop buffer that occurred closest in time. Any matches where the GPS point and Trip Update record are more than 20 minutes apart are excluded.

We note that past stop updates may be dropped out of the GTFS real-time feed if a bus pass a stop ahead of schedule. For example, if a bus is predicted to arrive at a stop two minutes earlier than the schedule, that prediction must remain in the feed until after the scheduled time passes. But it could remove the record earlier than the feed update.



(a) Vehicle Positions record segment assignment visualization



(b) Vehicle Positions record segment assignment (zoomed in)

Figure 9.2. Vehicle Positions record segment assignment

Figure 9.3 illustrates the distribution of time differences, in minutes, between the Trip Update departure time and the closest corresponding timestamp from the archived Vehicle Positions dataset for each Sun Tran route, based on data collected on a weekday, March 26th, 2025. The x-axis represents individual transit routes identified by both route number and route name, while the y-axis shows the time difference in minutes. Each boxplot presents a summary of the distribution for each route. The orange boxes represent the interquartile range (IQR), red lines indicate the median values, black whiskers show the $1.5 \times$ IQR range, and black dots mark the outliers.

Routes such as 1 (Glenn/Swan) and 37 (Pantano) stand out with higher median time differences, approximately 4-5 minutes, and noticeably wider IQRs. Conversely, several routes, 50 (Ajo Way), 6 (Euclid/1st Ave), and 63 (La Cholla), show relatively low median time differences, less than 2 minutes, and narrow IQRs. The limited number of outliers also supports the conclusion that these routes experience more reliable and synchronized data reporting.

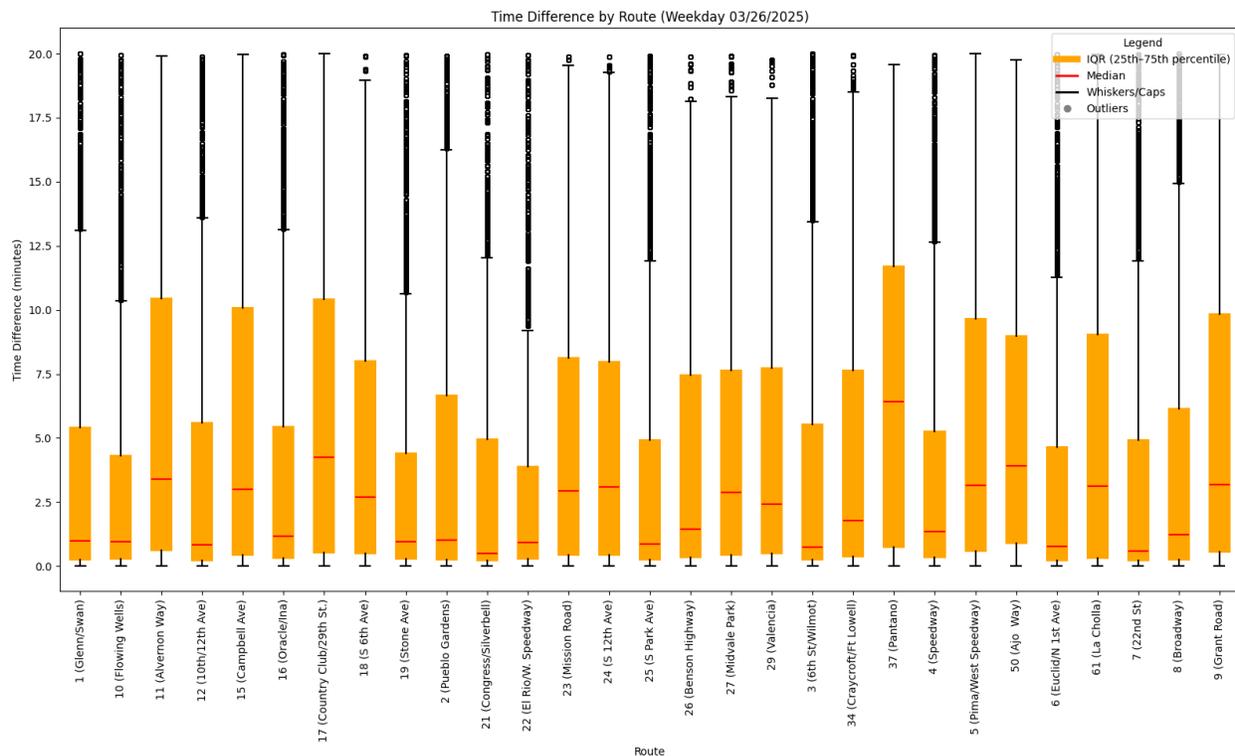


Figure 9.3. Time difference of each route on Wednesday, March 26, 2025

Figure 9.4 presents the distribution of time differences between Trip Update departure times and the nearest corresponding time stamps from Vehicle Positions on a weekend day, Saturday, March 29, 2025. Routes including 37 (Pantano), 24 (Craycroft/Ft. Lowell), and 3 (6th St/Wilmot) exhibit larger IQRs and higher medians, approximately 5 minutes.

More stable patterns are observed for routes such as 2 (Pueblo Gardens), 19 (Stone Ave), and 18 (5 6th Ave), which maintain lower medians, around 2–3 minutes, and narrower IQRs.

When comparing the weekend data to the weekday results, a general trend of increased median values and broader IQRs is observed across most routes on the weekend. For example, routes like 1 (Glenn/Swan) and 22 (E. Roger/Speedway) exhibit wider spreads and higher medians on the weekend, implying reduced synchronization accuracy.

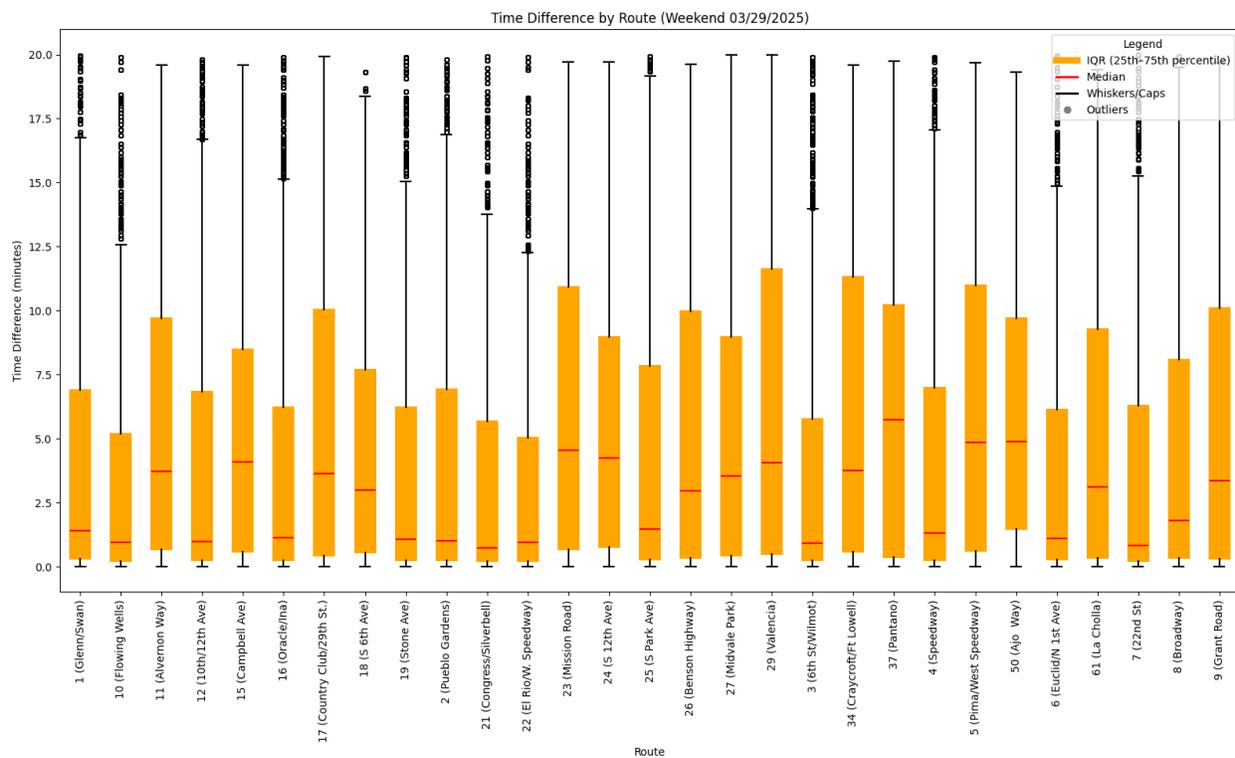


Figure 9.4. Time difference of each route on Saturday, March 29, 2025

9.2.2 *Transit performance measures result validation*

9.2.2.1 *Transit performance measures literature review*

A literature review was conducted to validate the performance measures results observed in Sun Tran system by comparing the calculated metrics for the Sun Tran system to typical value ranges found in studies and transit agency reports.

As shown in [Table 9.2](#), Recent literature highlights the impacts of different performance measures on bus ridership. Arrival delay, commonly defined as the deviation from scheduled arrival time, has a consistently negative relationship with ridership. Studies such as Pulugurtha & Jayanthi (2021) and WMATA evaluations (Mineta Institute, 2021) indicate that average delays of 0 to 5 minutes can significantly deter rider satisfaction and trust in the system. Similarly, median headway, a measure of service frequency, has been shown to negatively affect ridership when headways extend beyond 10–15 minutes, particularly in urban areas (Zhang & Zhang, 2023; UITP, 2023). Beyond average frequency, headway irregularity—captured through the standard deviation of headways—also correlates negatively with ridership, as variability between bus arrivals (ranging from 2 to 5 minutes) undermines reliability and increases perceived waiting time (Zhang & Zhang, 2023; Strathman et al., 2013).

In contrast, the headway reliability, which quantifies the consistency of bus intervals by comparing the 95th percentile headway to the mean, shows a positive relationship with ridership. Riders are more likely to rely on services with higher reliability, and R values above 0.875 are generally preferred (FHWA, 2019; Jayanthi et al., 2023). Likewise, mean operating speed is positively associated with ridership, with systems with average speeds of 8–12 mph, typical for local buses, tending to attract more users, as these speeds indicate that services are efficient and not delayed by congestion (APTA, 2022; Cato Institute, 2018). Finally, travel time reliability, often expressed as on-time performance, is also positively associated with ridership, as systems achieving 75–95% on-time performance are generally seen as dependable, a conclusion supported by multiple sources including FHWA (2019), Pulugurtha & Jayanthi (2021), MBTA audits, and WMATA reports.

Table 9.2. Transit performance measures literature review summary

Performance Measure	Relationship to Ridership	Reference
Arrival Delay	Negative	Pulugurtha & Jayanthi (2021); WMATA (Mineta Institute, 2021)
Median Headway	Negative	Zhang & Zhang (2023); UITP (2023)
Headway Irregularity	Negative	Zhang & Zhang (2023); Strathman et al. (2013)
Headway Reliability	Positive	FHWA (2019); Jayanthi et al. (2023)
Mean Speed	Positive	APTA (2022); Cato Institute (2018)
Travel Time Reliability	Positive	FHWA (2019); Pulugurtha & Jayanthi (2021); MBTA Audit (2018); WMATA reports

9.2.2.2 Transit performance measures result validation

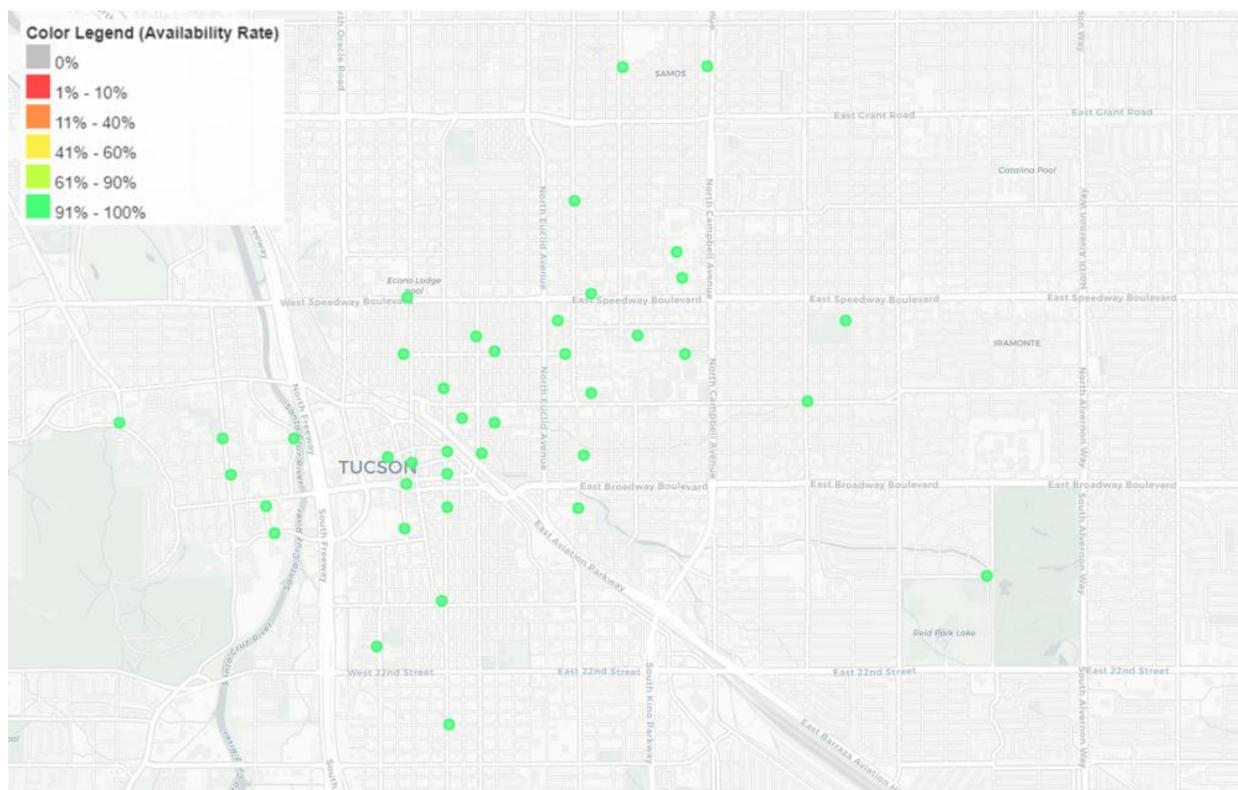
To assess the performance of the Sun Tran system, each key performance measure was compared to typical value ranges reported in the literature. This comparison provides an evidence-based perspective on how Sun Tran’s current operational characteristics align with or differ from peer systems based on the result shown in [Table 9.3](#) and [Figure 9.9](#).

Starting with arrival delay, the Sun Tran system reported an average of 10.76 minutes, which exceeds the typical 0–5 minute range observed in U.S. systems such as WMATA and others studied by Pulugurtha and Jayanthi (2021). In terms of median headway, Sun Tran averages 28.27 minutes, which is higher than the 10–15 minutes commonly cited as acceptable for urban routes (Zhang & Zhang, 2023; UITP, 2023). The negative correlation observed between headway and ridership in both the literature and Sun Tran’s data. Regarding headway irregularity, Sun Tran’s average standard deviation of 10.03 minutes is higher than the benchmark range of 2–5 minutes reported by Strathman et al. (2013) and Zhang & Zhang (2023). The headway reliability index, which compares the 95th percentile headway to the mean, averaged 50.35% for Sun Tran, positioned at the upper edge of the 20–50% range reported by FHWA (2019) and Jayanthi et al. (2023). Interestingly, Sun Tran’s mean operating speed was 18.22 mph, which exceeds the 8–12 mph range commonly reported for local services (APTA, 2022; Cato Institute, 2018). This higher speed may reflect more suburban or express-style service patterns or relatively uncongested roadways. Lastly, travel time reliability, often assessed through on-time performance, averaged 78.15% for Sun Tran. This result aligns with the literature range of 75–95% (FHWA, 2019; Pulugurtha & Jayanthi, 2021).

In summary, while several of Sun Tran’s performance measures fall within the ranges reported in national studies, headway length and irregularity are relatively high compared to peer systems.

Table 9.3. Sun Tran transit performance measures summary

Performance Measure	Relationship to Ridership	Average Value
On-time Percentage	Positive	96.41%
Median Headway	Negative	28.27 minutes
Headway Irregularity	Negative	10.03 minutes
Headway Reliability Index	Positive	50.35%
Mean Speed	Positive	18.22 mph
Travel Time Reliability	Positive	78.15 %

**Figure 9.5. Sun Tran transit performance measures and ridership scatter plots**

9.2.3 Transit recommended performance targets

To ensure high-quality transit data and performance, clear targets with threshold criteria for different performance measures can be considered. These targets can serve quality assurance and quality control (QA/QC) practices for GTFS-derived datasets. Each measure has a defined relationship with ridership and plays a role in operational planning and monitoring.

Table 9.4 summarizes the key performance measures in the transit mode. For instance, arrival delay serves as a direct indicator of reliability at the stop level and is frequently used to identify delay hotspots or improve on-time performance, with a recommended range of 0-5 minutes late on average (Pulugurtha & Jayanthi, 2021; WMATA, 2021). Headway-related metrics, including

median headway, irregularity, and the headway reliability index, are widely used in high-frequency service evaluation. For example, the median headway for urban routes is ideally between 10-15 minutes, helping determine if a route meets frequency standards (Zhang & Zhang, 2023; UITP, 2023).

Operational instability and bunching can be identified through headway irregularities with a threshold of 2-5 minutes standard deviation (Strathman et al., 2013). A reliability index of 20-50% is considered moderate to good for frequent transit routes (FHWA, 2019; Jayanthi et al., 2023). Mean operating speeds, ideally 8 – 12 mph for local routes, are essential for prioritizing corridor treatments and assessing network efficiency (APTA, 2022; Cato Institute, 2018). Lastly, travel time reliability, typically ranging from 75 - 95 %, is commonly used in system-wide performance audits (Pulugurtha & Jayanthi, 2021; MBTA, 2018)

Table 9.4. Transit Performance Targets

Performance measures	Use cases	Performance Targets	Reference
Arrival Delay	<ul style="list-style-type: none"> • Identify delay hotspots • Improve on-time performance 	0 – 5 minutes late (avg)	Pulugurtha & Jayanthi (2021); WMATA (Mineta Institute, 2021)
Median Headway	<ul style="list-style-type: none"> • Evaluate frequency adequacy • Determine whether routes meet high-frequency standards in network design 	10 – 15 minutes for urban routes	Zhang & Zhang (2023); UITP (2023)
Headway Irregularity	<ul style="list-style-type: none"> • Detect operational instability • Justify implementation of active dispatching strategies or bus lanes 	2-5 minutes std dev	Zhang & Zhang (2023); Strathman et al. (2013)
Headway Reliability Index	<ul style="list-style-type: none"> • Schedule validation and frequent service monitoring • Evaluate real-time operations performance 	20 – 50 %	FHWA (2019); Jayanthi et al. (2023)
Mean Speed	<ul style="list-style-type: none"> • Identify low-speed corridors for priority treatments • Assess corridor productivity 	8 – 12 mph for local buses	APTA (2022); Cato Institute (2018)

Travel Time Reliability	<ul style="list-style-type: none">• Assess service dependability for riders• Trigger schedule modifications• Measure success of priority projects	75 – 95 %	FHWA (2019); Pulugurtha & Jayanthi (2021); MBTA Audit (2018); WMATA reports
-------------------------	---	-----------	---

9.3. Pedestrian Performance Targets

9.3.1 *Pedestrian volume estimation literature review*

Recent research demonstrates the viability of pedestrian push button event data as a cost-effective and scalable proxy for estimating pedestrian volumes, as summarized in [Table 9.5](#). While the literature of pedestrian volume estimation using push button events is limited, these studies have shown the practicality of using this data.

Singleton and Runa (2021) explored signal-level button press data to develop piecewise regression models, achieving strong correlation with observed volumes across 20 intersections in Utah. Building on this, Kothuri et al. (2024) validated similar methods across a broader network of 65 intersections in Oregon, employing simplified regression to achieve average errors below 2.5 pedestrians per hour. Li and Wu (2021) introduced a more granular modeling approach, applying a Poisson arrival process and Maximum Likelihood Estimation (MLE) to event-based data from midblock crossings. Finally, Li et al. (2022) advanced the use of machine learning by implementing Bayesian Additive Regression Trees (BART) to model crossing volumes at signalized intersections, achieving R^2 values up to 0.83.

Together, these foundational studies illustrate both the methodological evolution of push button data in estimating pedestrian activity, particularly when paired with contextual data. However, they also highlight the dependence on high-resolution, continuous data availability, which may limit applicability in resource-constrained settings.

Table 9.5. Pedestrian volume estimation literature review summary

Study Name	Methodology	Performance Targets	Data Size
Singleton & Runa (2021)	Piecewise linear & quadratic regression	MAE = 3.0	20 intersections, ~1,100 hourly observations
Kothuri et al. (2024)	Quadratic regression	MAE = 2.4	12 sites, thousands of hours of signal data
Li & Wu (2021)	Maximum Likelihood Estimation (MLE) at midblock crosswalks	MAE = 2.27	7 sites, ~1,200 hourly samples
Li et al. (2022)	Bayesian Additive Regression Trees (BART)	$R^2 = 0.83$	4 intersections, 1,258 hourly records

9.3.2 *Pedestrian volume estimation model results*

To evaluate the performance of the models effectively, multiple metrics are necessary to capture different aspects of prediction accuracy. A single metric may not fully reflect model performance, especially in cases where errors vary across different ranges of pedestrian volume. By using multiple evaluation criteria, we can ensure a more comprehensive assessment of how well the

model generalizes to new data. **Table 9.6** outlines the description, strengths, and limitations of each metric and provides a basis on why each metric was used.

Table 9.6. Description, Strengths and Limitations of Each Metric

Metric	Description	Strengths	Limitations
Root Mean Squared Error (RMSE)	Average magnitude of error, penalizing large errors.	Highlights large errors; unit-based interpretation.	Sensitive to outliers; scale-dependent.
Mean Absolute Percentage Error (MAPE)	Average percentage error between predictions and actual values.	Easy to interpret; scale-invariant.	Unreliable for values near zero.
R-Squared (R^2)	Proportion of variance explained by the model.	Intuitive fit measure; good for model comparison.	Misleading for non-linear or overfitted models.
Mean absolute error (MAE)	Average of the absolute differences between predicted and actual values	Easy to interpret; not as sensitive to outliers as RMSE	Does not penalize large errors as heavily as RMSE

The final data contains 31 features and 1,640 data points, collected across 237 intersections. Testing the XGBoost model on the test samples resulted in the following metrics:

- *Root Mean Squared Error (RMSE): 12.61*
- *Mean Absolute Percentage Error (MAPE): 43.48%*
- *R-squared (R^2): 0.819*
- *Mean Absolute Error (MAE): 7.03*
- *Average 5-fold R-squared (R^2): 0.828*

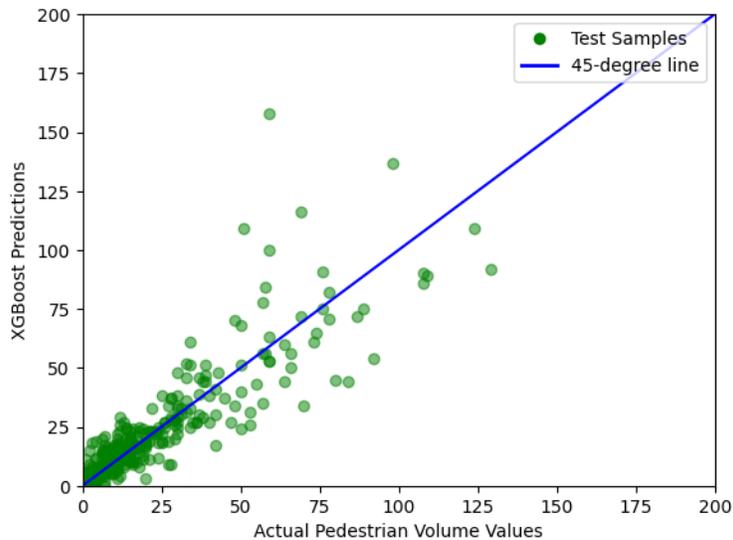


Figure 9.6. Parity Plot of the Final XGBoost Model

Figure 9.6 shows that the model generally follows the expected trend, with predictions aligning more closely to the 45-degree line for lower pedestrian volumes.

The model achieved strong performance given the constraints of the available dataset, with an R^2 of 0.819 and a 5-fold cross-validated average R^2 of 0.828. While error values such as MAE are somewhat higher than those reported in recent literature, it is important to note the differences in study conditions. Unlike prior studies that typically focused on a small number of intersections with high-resolution and continuous button press data, which often span thousands of hourly observations, our model was developed using 1,640 scattered data points distributed across 237 intersections. This broader spatial scope, combined with limited sample sizes per location and instances of missing or incomplete signal data, presents a more challenging modeling environment. Despite these constraints, the model demonstrates robust predictive accuracy and generalizability across a diverse urban network.

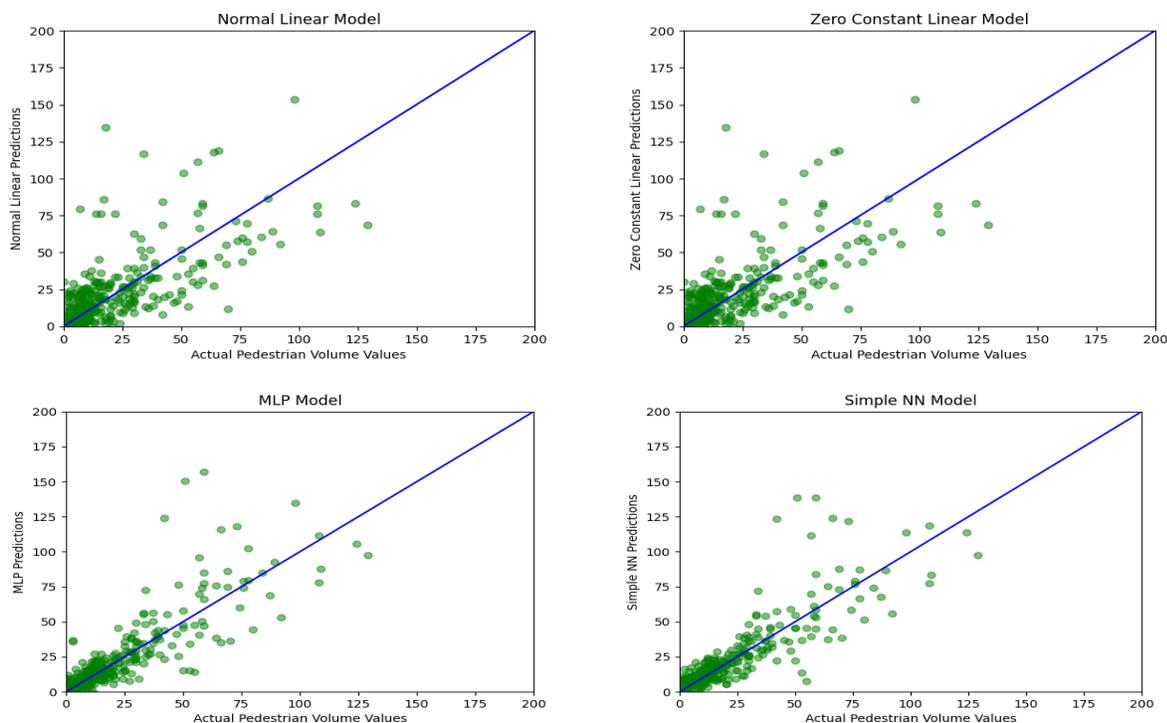
9.3.3 Comparison to other models

The XGBoost model was evaluated against other machine learning models in terms of performance and consistently outperformed other machine learning models tested, including classic linear regression, zero-constant linear regression, simple neural network (NN), and Multi-Layer Perceptron (MLP). Performance metrics for each of the models are shown in Table 9.6.

Table 9.7. Performance Metrics for Alternative Models

Metric	Linear Regression	Zero Constant Linear Regression	Simple NN	MLP	XGBoost
RMSE	28.86	29.19	14.88	15.41	12.61
MAPE ¹	126.19%	123.47%	58.75%	57.82%	43.48%
R ²	0.621	0.698	0.749	0.731	0.819
MAE	15.24	15.15	7.87	9.52	7.03

Table 9.7 highlights that XGBoost outperforms all other models across all three metrics, achieving the lowest RMSE (12.61), lowest MAPE (43.48%), and highest R² (0.819), indicating better overall fit and predictive capability. The neural network-based models (Simple NN and MLP) show noticeable improvement over linear models, but they still underperform compared to XGBoost. XGBoost performed the best on this dataset due to its ability to handle a wide variety of feature relationships—linear, nonlinear, and interactions—using its tree-based architecture. **Figure 9.11** shows parity plots of alternative models tested.



¹ Values below 1 were masked out to ensure sensible calculations for MAPE, as the metric becomes unreliable when actual values are close to zero.

Figure 9.7. Parity Plot Comparison of Alternative Models

Figure 9.7 shows that normal linear regression and zero-constant linear regression models show significant scatter and a weaker fit, particularly for higher pedestrian volumes. The simple NN and MLP models show improved alignment with the 45-degree line and better fit, though some dispersion remains for larger values. Overall, while the neural networks perform better than the linear models, they still show variability, highlighting the need for a more robust model such as XGBoost.

The performance of both linear regression and XGBoost models varies across different ranges of the target variable. **Figure 9.8**, **Figure 9.9**, and **Figure 9.10** show the performance of zero-constant linear regression vs XGBoost model across different categories, RMSE, MAPE, and MAE respectively. In each figure, the plots show the side-to-side comparison of zero-constant linear regression and XGBoost.

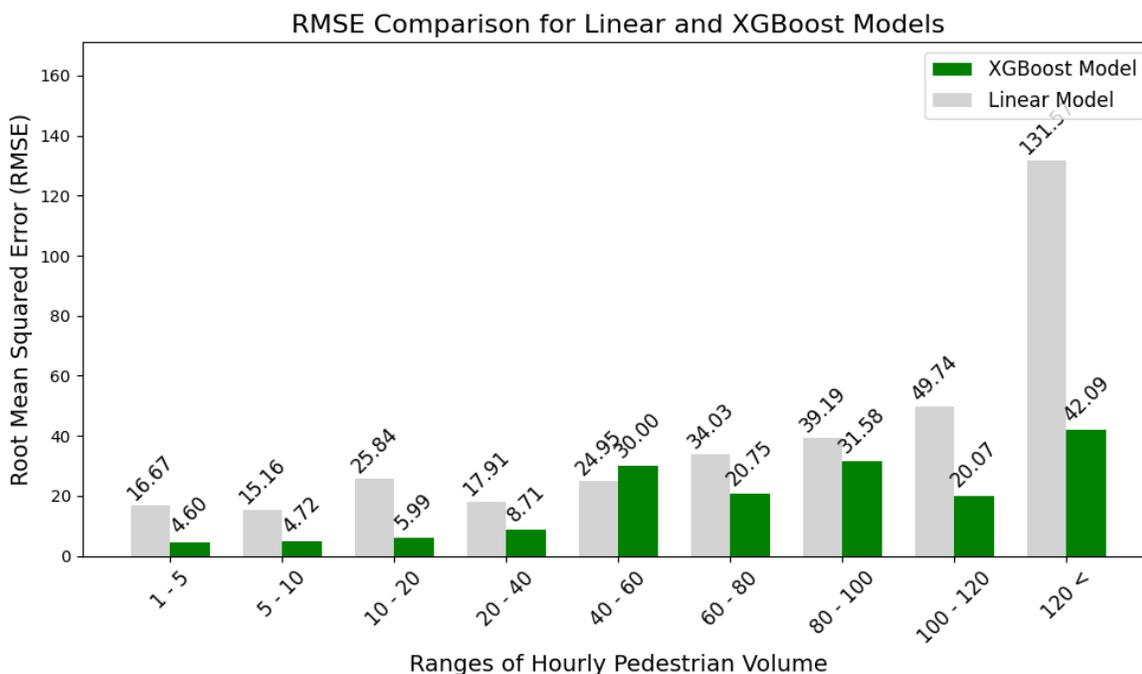


Figure 9.8. RMSE Comparison Between Linear and XGBoost Models in Different Ranges

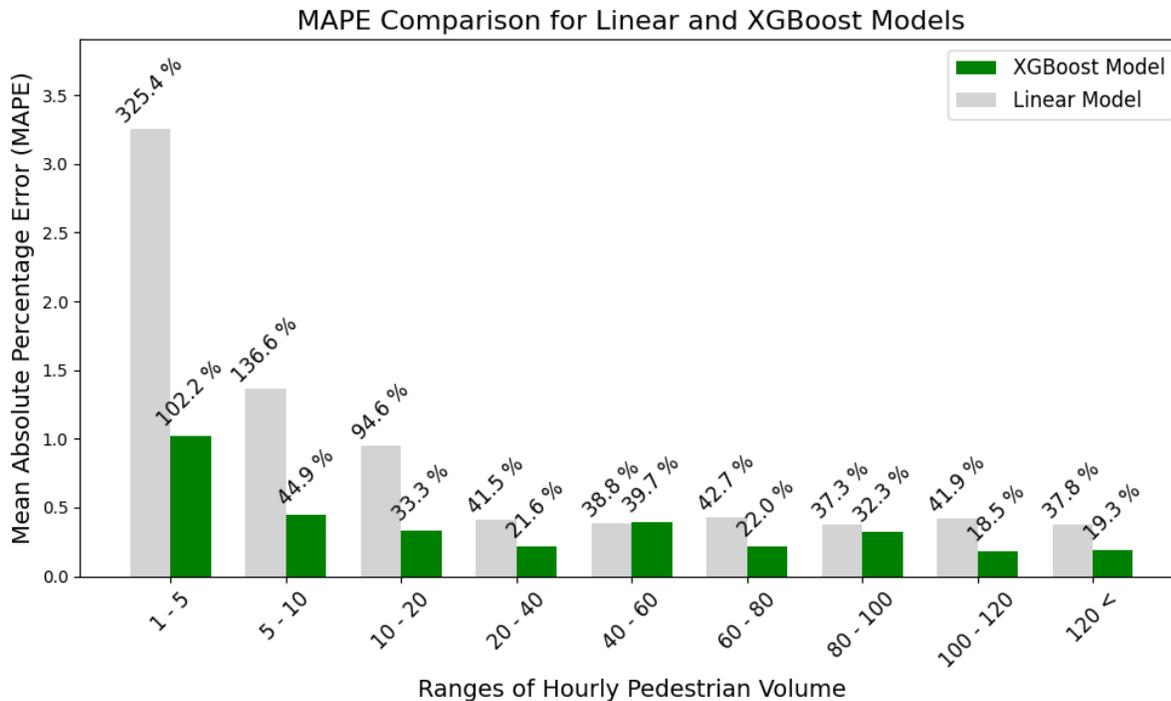


Figure 9.9. MAPE Comparison Between Linear and XGBoost Models in Different Ranges

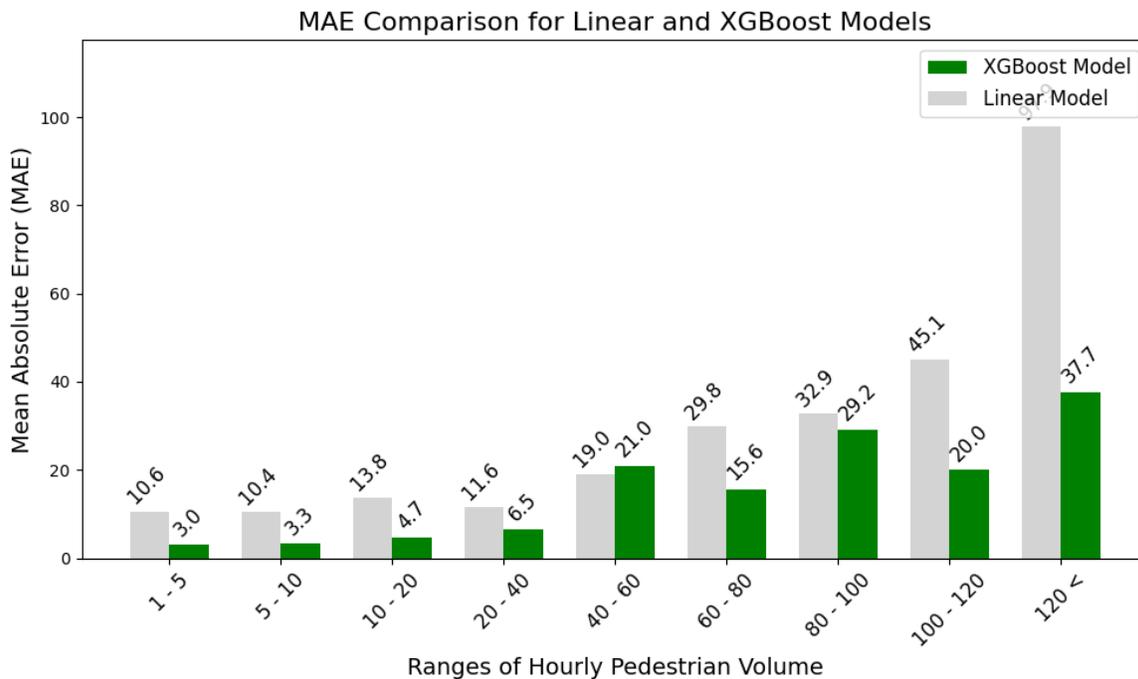


Figure 9.10. MAE Comparison Between Linear and XGBoost Models in Different Ranges

The figures show that XGBoost outperforms over the zero-constant linear model in almost all pedestrian volume groups, except for the 40-60 range. XGBoost is particularly effective for lower pedestrian volume values in the 0-20 range.

In addition, the % RMSE values were obtained by calculating the RMSE for each category and dividing it by the average pedestrian volume for that category. This allows for the comparison of model's performance across different pedestrian volume ranges. **Figure 9.11** shows these values. The metric is computed using the following formula (TF-Resource, n.d.):

$$\%RMSE = \frac{RMSE}{Average\ Pedestrian\ Volume} \times 100$$

Figure 9.11 shows that, based on % RMSE values, relative errors are highest in low-volume ranges, particularly in the 1-5 range, where values are small. As pedestrian volume increases, the % RMSE decreases, indicating that the model's predictions become more stable relative to actual values. However, a slight increase was observed in the 40-60 range. This is a range where the model struggles, even compared to the zero-constant linear model. Its underperformance in the 40-60 range is likely due to the relationship between features and pedestrian volume being linear in this range.

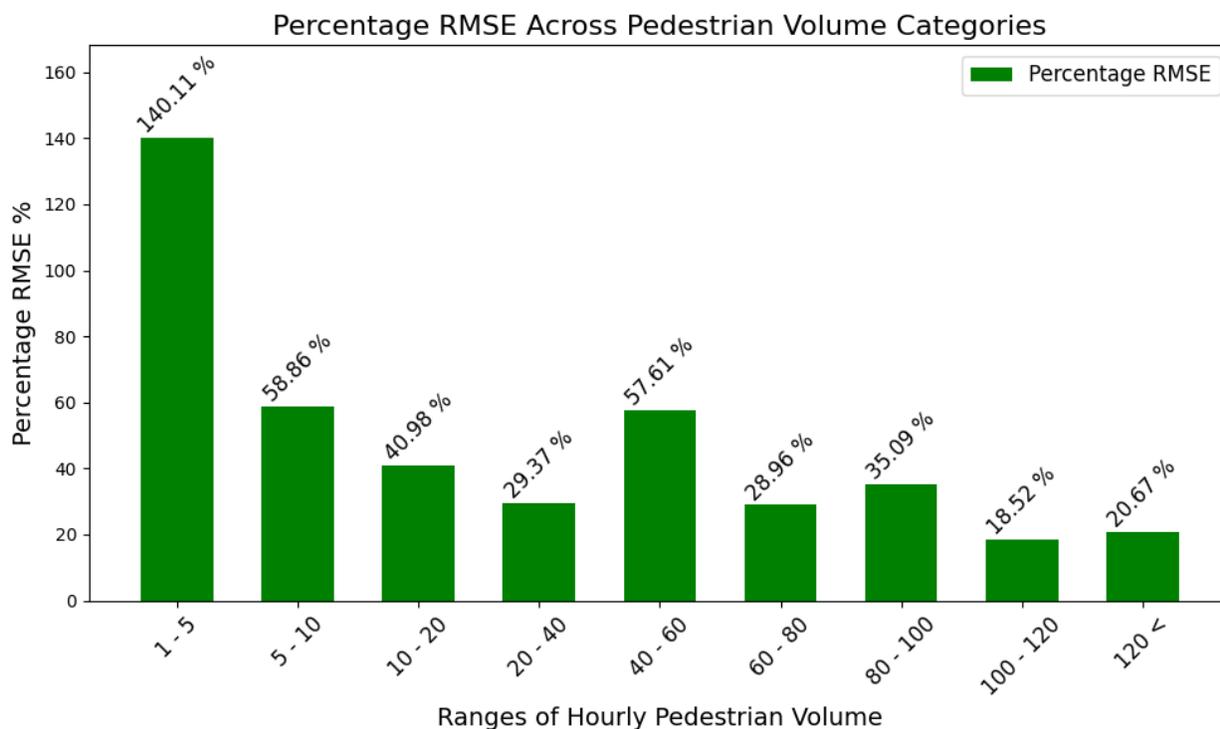


Figure 9.11. Percent RMSE Values for The XGBoost Model in Different Ranges

9.3.4 Pedestrian recommended acceptance criteria

To evaluate pedestrian infrastructure and performance, delay, LOS, and volume estimation accuracy measures offer quantifiable standards for assessing and planning walking environments. **Table 9.8** outlines the key pedestrian performance measure targets. Pedestrian delay, with a preferred average delay of less than 30 seconds, is critical for identifying signal design deficiencies and safety risks (Rouphail et al., 2005; NCHRP, 2022). Pedestrian LOS, typically requiring LOS C or better, at least 15 ft² per person, is widely used to assess comfort and mobility in crowded urban corridors. These metrics have been applied to prioritize sidewalk expansion projects and crossing redesigns in cities like New York (Landis et al., 2001; NYC DOT, 2015). Pedestrian volume estimation models using signal push-button data or short counts have proven to be accurate ($R^2 > 0.70$, $MAE \leq 3$) and are used in infrastructure planning and QA/QC validation where continuous counts are unavailable (Singleton and Runa, 2021; Choobchian et al., 2024).

While our model's MAE value exceeds the benchmarks reported in these studies, this variation is expected given the different study contexts. As addressed earlier, our model was developed using a relatively sparse and heterogeneous dataset, covering 237 intersections with incomplete or missing sensor inputs and limited data points per location. These conditions present challenges not encountered in studies with high-quality structured data. Nonetheless, the model still delivers a high explanatory capacity with an R^2 of 0.819 and provides a valuable tool for estimating pedestrian volumes at a network scale where traditional count methods are impractical.

Table 9.8. Pedestrian performance measure target

Performance measures	Use cases	Target	Reference
Pedestrian Delay	<ul style="list-style-type: none"> • <i>Intersection signal retiming</i> • <i>Safety analysis</i> 	< 30 seconds average delay	Rouphail et al. (2005); NCHRP 969 (2022)
Pedestrian Level of Service	<ul style="list-style-type: none"> • <i>Sidewalk widening</i> • <i>Intersection redesign</i> 	\geq LOS C; > 15sq ft/person	Landis et al. (2001); NYC DOT (2015)
Pedestrian Volume Estimation	<ul style="list-style-type: none"> • <i>Equity audits</i> • <i>Infrastructure planning</i> 	$R^2 > 0.70$; $MAE \leq 3$	Singleton and Runa (2021); Choobchian et al. (2024)

9.4. Micromobility Data QA/QC

To ensure the reliability of the General Bikeshare Feed Specification (GBFS) data used in this study, a comprehensive data accuracy review was conducted by comparing archived data against direct field observations collected during a site visit. This validation process aimed to assess the consistency and accuracy of real-time bikeshare system status records—such as vehicle availability and station occupancy—by cross-referencing the recorded GBFS outputs with ground-truth conditions observed in the field. This comparison serves as a critical step in verifying the integrity of the data, identifying any systematic discrepancies, and confirming the suitability of the GBFS feed for performance analysis and decision-making in micromobility system evaluation.

9.4.1 *Tugo Bike Share GBFS data accuracy review*

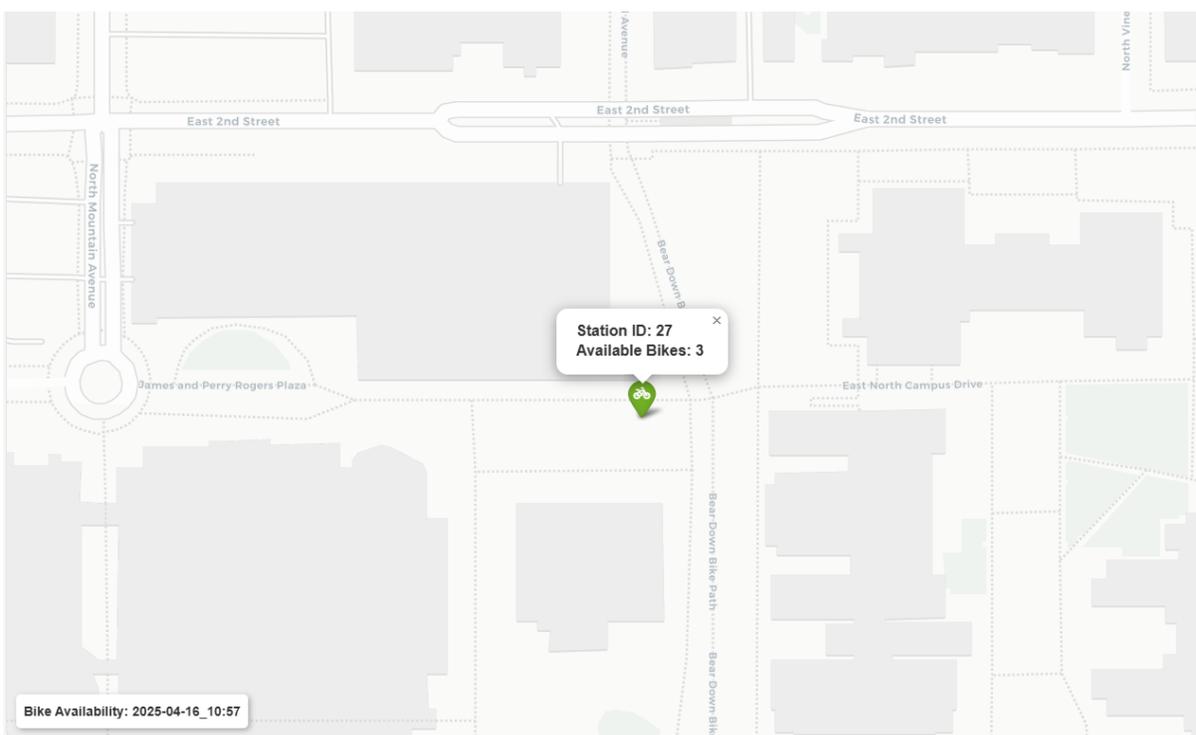
As part of the GBFS data quality validation process, the Tugo Bike Share system’s archived GBFS data was systematically compared with observed bike station statuses collected during an in-person site visit. A total of nine stations were visited and manually inspected on April 16, 2025, with the number of available bikes at each station recorded along with the corresponding observation time. These observations were then cross-referenced with the GBFS data captured at the same timestamps.

Table 9.9 summarizes the station IDs, observed bike availability, and corresponding timestamps used in the validation. Visual comparisons for each station examples are illustrated in **Figures 9.12** and **9.13**, representing station 27 and 30 each, which display the real-time GBFS data overlays with the observation timestamps noted in the legend at the bottom left corner of each figure. Concurrent observations of the bike docks were also made to confirm the reported number of docked bikes.

The validation results confirm a complete match between the observed station statuses and the GBFS-recorded data, indicating a high level of accuracy and reliability in the Tugo GBFS feed. This consistency validates the use of the archived GBFS data for performance monitoring, operational analysis, and broader micromobility system evaluations.

Table 9.9. Tugo bike share observations record

Station ID	Bike Status	Time	Date
27	3	10:57 AM	04/16/2025
30	5	11:01 AM	04/16/2025
31	8	11:05 AM	04/16/2025
32	12	11:06 AM	04/16/2025
28	2	11:13 AM	04/16/2025
25	2	11:18 AM	04/16/2025
24	8	11:21 AM	04/16/2025
21	8	11:24 AM	04/16/2025
23	11	11:31 AM	04/16/2025

**Figure 9.12. Station 27 bike availability status**

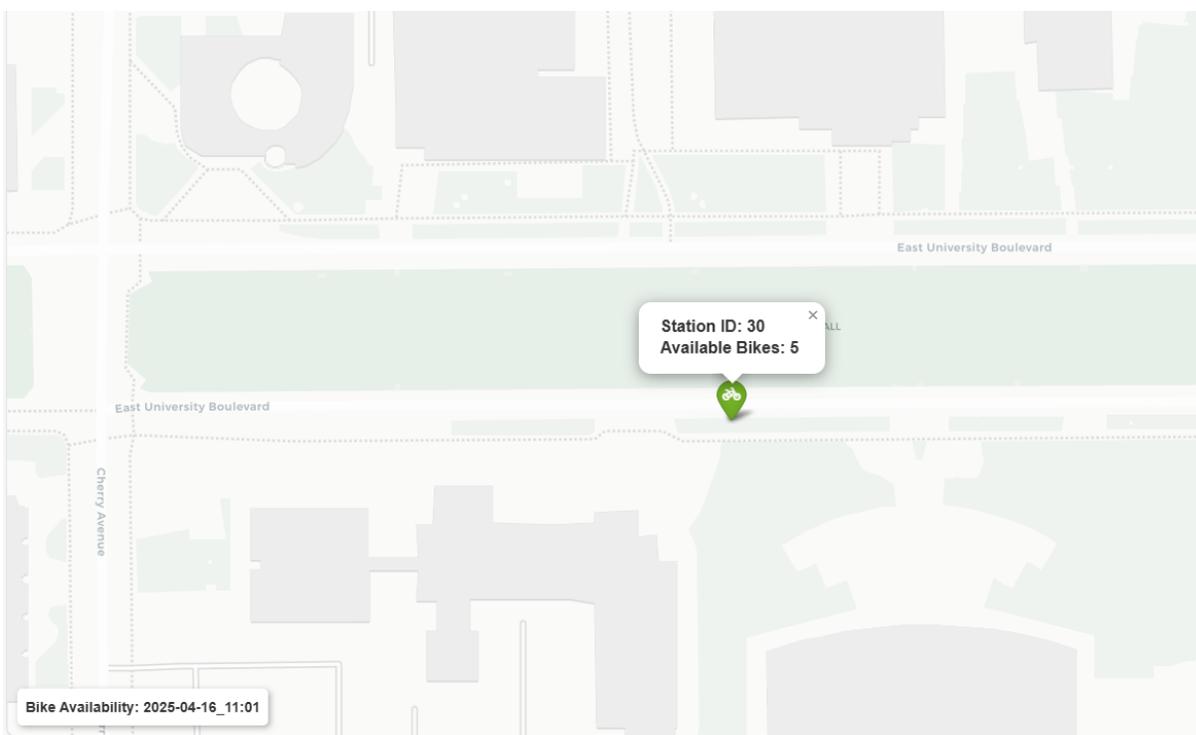


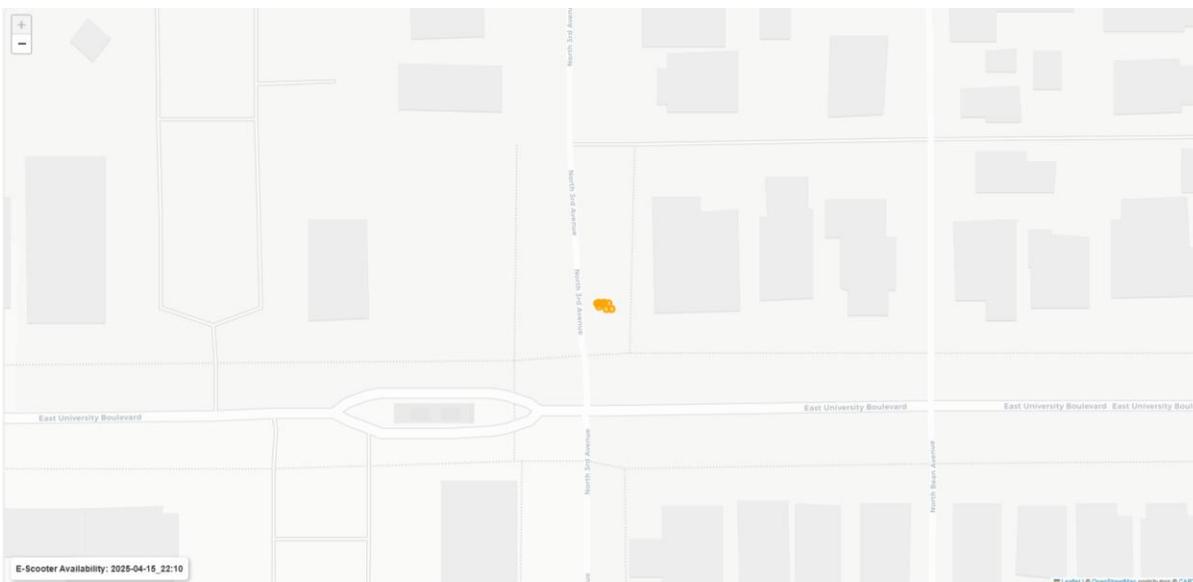
Figure 9.13. Station 30 bike availability status

9.4.2 E-scooter GBFS data accuracy review

To validate the accuracy of the E-scooter GBFS data, a field GBFS data, a field verification was conducted on April 15, 2025, at five different locations, comparing observed e-scooter availability with archived GBFS records.

At a temporary e-scooter station located at the intersection of E University Blvd and N 3rd Ave, a total of eight e-scooters were observed at 10:00 PM, which matched the number recorded in the archived GBFS data as displayed in [Figure 9.14](#). Additional validations were performed around E 8th St and N 4th Ave. On the southeast corner, two e-scooters were observed and confirmed in the data as shown in [Figure 9.15](#), while on the northeast corner, four e-scooters were recorded both in-person and in the GBFS feed as displayed in [Figure 9.16](#). At the southwest corner of E 7th St and N 4th Ave, five e-scooters were present during observation, and this count was consistently reflected in the archived data as in [Figure 9.17](#).

In [Figure 9.18](#), a user return activity was captured at the moment of observation. The image clearly shows two e-scooters with lights on, indicating recent returns, and this activity was also reflected in the GBFS records. This alignment between the observed behavior and the archived state further supports the accuracy of the GBFS feed.

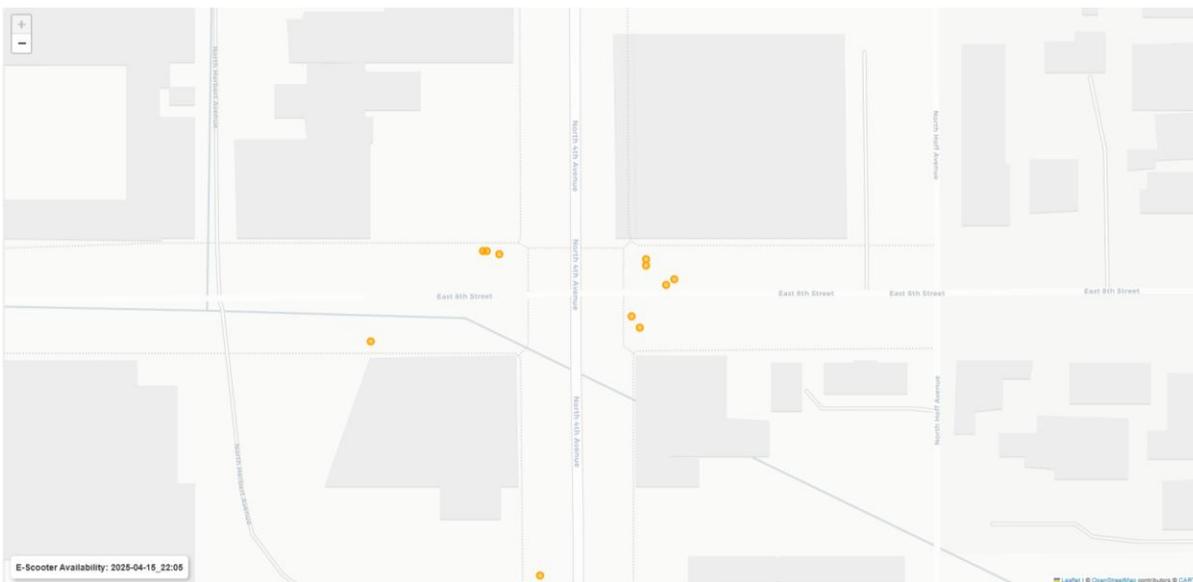


(a) GBFS data e-scooter availability map



(b) Observed e-scooter availability

Figure 9.14. E-scooter observation at the intersection of E University Blvd and N 3rd Ave

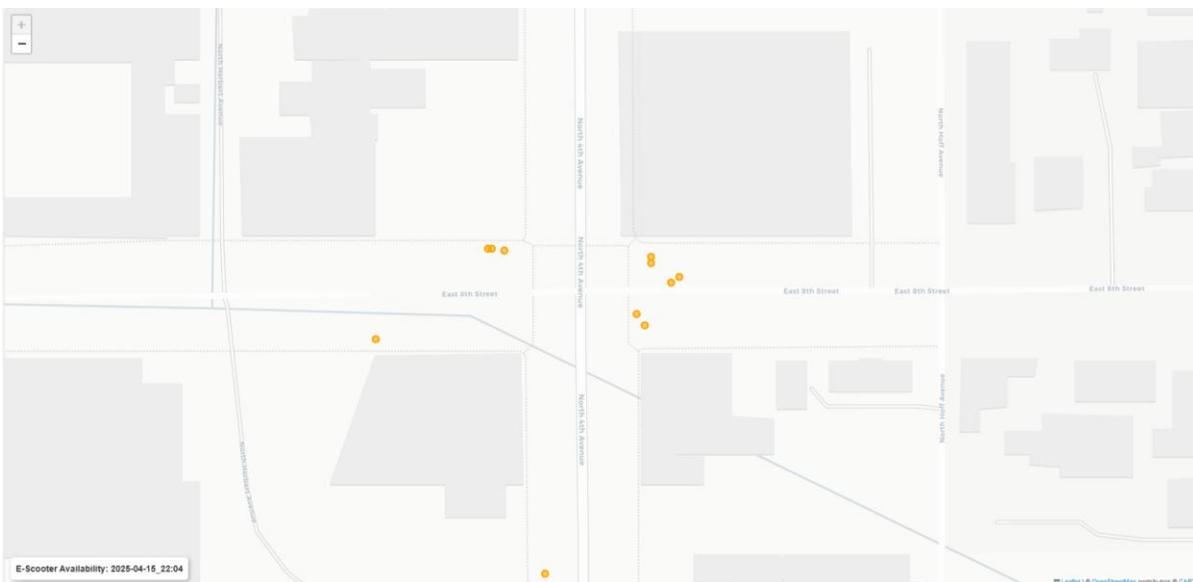


(a) E-scooter GBFS data e-scooter availability map



(b) Observed e-scooter availability

Figure 9.15. E-scooter observation at E 8th St and N 4th Ave

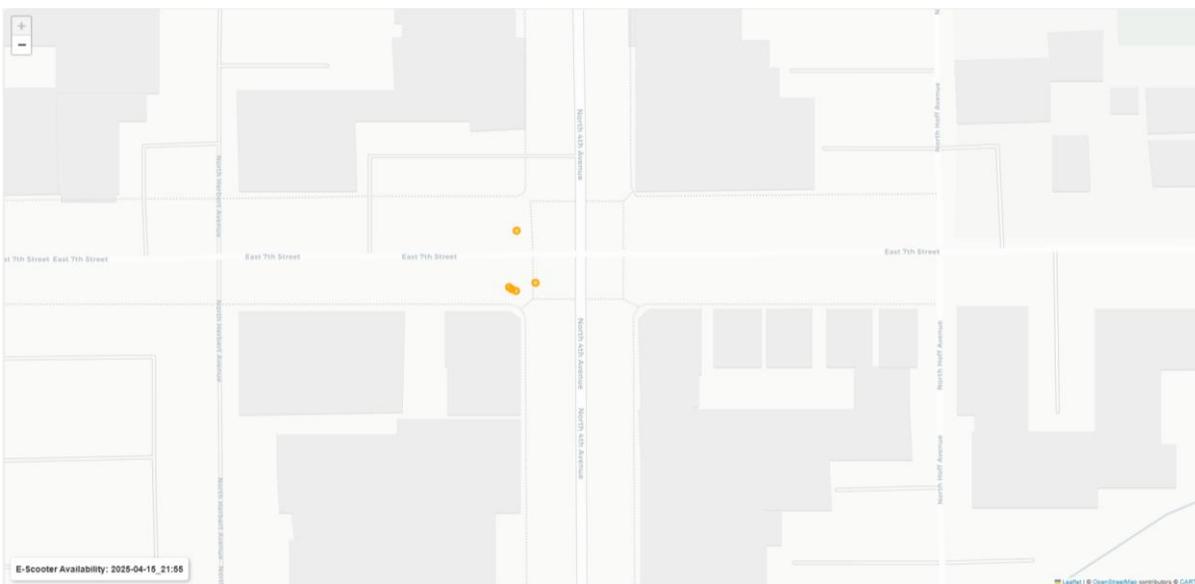


(a) E-scooter GBFS data e-scooter availability map



(b) Observed e-scooter availability

Figure 9.16. E-scooter observation at Northeast of E 8th St and N 4th Ave

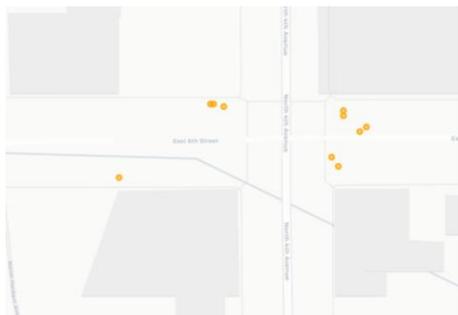


(a) E-scooter GBFS data e-scooter availability map



(b) Observed e-scooter availability

Figure 9.17. E-scooter observation at E 7th St and N 4th Ave



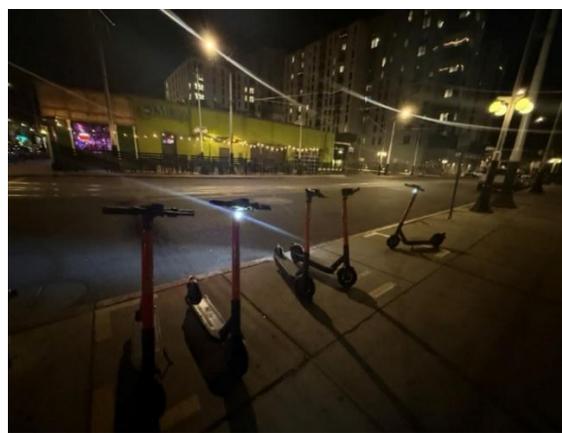
(a) E-scooter GBFS data e-scooter availability map before return



(b) E-scooter GBFS data e-scooter availability map after return



(c) Observed e-scooter return activity



(d) Observed e-scooter availability

Figure 9.18. E-scooter return activity observation

9.4.3 *Micromobility recommended performance measure target*

To evaluate micromobility performance measures, such as availability, utilization, and spatial coverage can be used to understand user access and operational efficiency. These indicators help cities manage equity goals, rebalance fleets, and plan infrastructure support for shared mobility systems.

Table 9.10 summarizes performance measures with defined thresholds for assessing micromobility services. Availability, measured as the percentage of time a scooter or bike is accessible in designated areas, should ideally exceed 95% during operating hours (Kabra et al., 2020; Caspi et al., 2020). This metric is crucial for rebalancing enforcement and ensuring geographic equity in access.

Table 9.10. Micromobility performance measure target

Performance measures	Use cases	Target	Reference
Micromobility availability	<ul style="list-style-type: none"> • <i>Fleet rebalancing</i> • <i>Equity policy enforcement</i> 	≥ 95% availability during operating hours	Kabra et al. (2020); Caspi et al. (2020)

CHAPTER 10. Platform Development for Region-wide Multi-modal Performance

This chapter focuses on estimating and analyzing region-wide multi-modal performance measures using existing data sources within the PAG region. Building on the methodologies developed in CHAPTER 6 to 8, the analysis incorporates data from various transportation modes, including public transit

(GTFS), micromobility (GBFS), and roadway traffic systems. Key performance measures include travel time reliability, headway consistency, vehicle availability, and accessibility. A Jupyter Notebook-based platform is developed to visualize these performance measures in an interactive and spatially detailed manner. This work supports a more integrated and data-driven approach to transportation planning in the PAG region. It provides a flexible foundation for future analysis, scenario testing, and policy evaluation aimed at improving mobility, accessibility, and equity across the entire multimodal network.

As summarized in [Table 10.1](#), the quality checks for transit and micromobility primarily focus on verifying the integrity of and completeness of archived GTFS and GBFS data. In contrast, pedestrian performance relies on a volume estimation model, necessitating a model-focused quality review to assess its predictive accuracy and reliability. Details regarding the data periods, quality assurance methodologies, and findings are elaborated in the following sections.



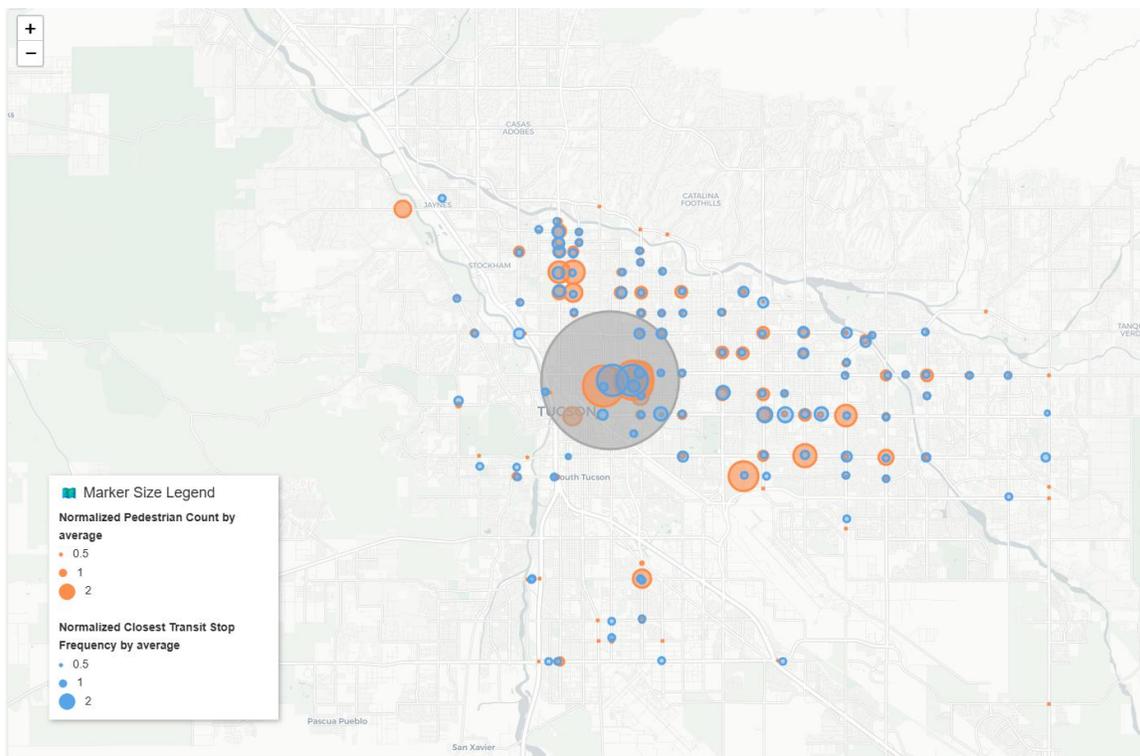
Table 10.1. Multi-modal performance measures summary

Performance Measure	Location in Report	Description	Mode			
			Transit	Pedestrian	Bike	E-scooter
Transit frequency and pedestrian activity	Section 10.2	Evaluate the relationship between close transit headway to the pedestrian activity				
Feature Importance Analysis Using SHAP	Section 10.3	Perform SHAP matrix for each feature				
Closest transit station distance to e-scooter	Section 10.4	Evaluate the distance to the closest transit station for each e-scooter				
Closest transit station distance to bike sharing station	Section 10.5	Evaluate the distance to the closest transit station for each bike sharing station				

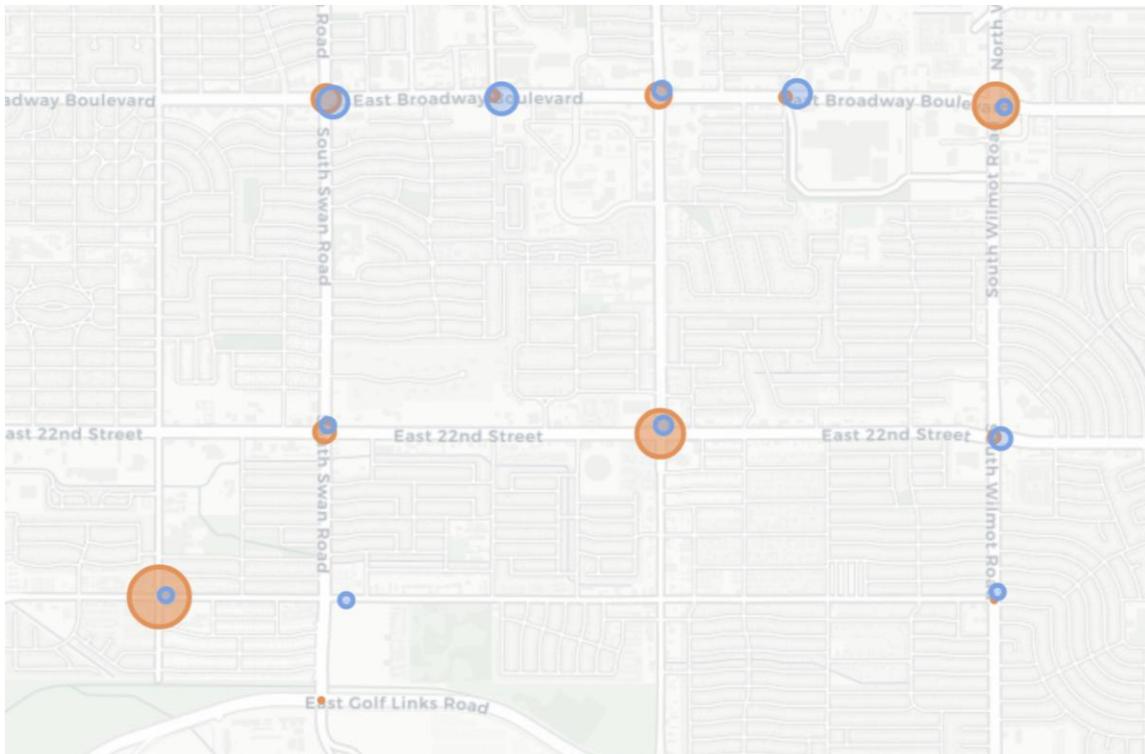
10.1. Transit frequency and pedestrian activity

The pedestrian-transit overlay map shown in **Figure 10.1** visualizes the spatial relationship between pedestrian activity and transit service frequency throughout the study area. Pedestrian counts, sourced from TMC data, were scaled based on the average pedestrian count of 22 in the 4 peak hours covered by TMC data, and are represented by orange proportional markers at intersection locations. A gray marker highlights the intersection at E 2nd St & N Park Ave, located on the University of Arizona campus, where a high pedestrian volume was recorded due to heavy student pedestrian traffic. Nearby transit stops are shown in blue, with marker sizes corresponding to the average service frequency during weekday peak periods, 6:00–8:00 AM and 4:00–6:00 PM. Transit frequency was calculated as the inverse of headway using performance results from **CHAPTER 6**. This combined visualization supports the identification of locations where transit access aligns with pedestrian demand and highlights areas with potential for improved multimodal integration.

Several intersections exhibit high scaled pedestrian volumes, ranging from 2 to 3, but are served by nearby transit stops with relatively low service frequencies, as illustrated in **Figure 10.1(b)**. With assumption that pedestrians are potential transit users, these locations may present opportunities to improve transit frequency, to enhance accessibility and potentially encouraging greater multimodal use.



(a) Transit headway and pedestrian count plotting map



(b) High pedestrian count and low transit frequency location examples

Figure 10.1. Transit frequency and pedestrian

10.2. Feature Importance Analysis Using SHAP

10.2.1 Introduction to SHAP

SHapley Additive exPlanations (SHAP) is a widely used tool for interpreting machine learning models, including tree-based models (e.g., XGBoost) and deep learning (e.g., neural networks). It is based on Shapley values, a concept from the cooperative game theory literature, ensuring a fair allocation of contributions among features (Lundberg & Lee, 2017). SHAP outputs show how much each input affects a prediction by comparing results with and without that input, across all possible combinations. This reveals which inputs matter most and their levels of significance.

Let's show the machine learning model of interest as $\hat{Y} = f(X)$, where:

\hat{Y} denotes the set of predictions of the target variable (i.e., Y) for all data instances;

X denotes the input data set consisting of I data instances ($i = 1, 2, \dots, I$) and J features ($j = 1, 2, \dots, J$); and

$f(\cdot)$ denotes the machine learning model of interest.

Let's also denote the model prediction for each i by y_i , and the set of all instances i for each feature j by X_j . Suppose for a given data instance i , which represents an intersection-hour, we are predicting the hourly pedestrian volume y_i . Each column j corresponds to a feature; for example, TotalSignals, Temperature, and Shops can be denoted by X_1 , X_2 , and X_3 , respectively.

The raw output of SHAP is a matrix where each row represents a single data instance i and each column represents a feature (or variable) j . Each element of this matrix is the “SHAP Value” for the corresponding pair of (i, j) and represents the extent to which j pushes the model's prediction for i up (i.e., towards higher y_i) or down (i.e., towards lower y_i).

The SHAP value of each j for i is calculated by averaging its marginal contributions across all possible combinations of features in which j appears. These values are calculated relative to a “Base Value.” The Base Value is the statistical expected value of the model's predictions across all instances i . That is, Base Value = $E[f(X)]$. For a regression model, this Base Value is calculated as the statistical average of the model predictions across all i .

In other words, the summation of the Base Value and the SHAP Value for each data instance i yields the predicted value for that i . **Figure 10.2** is a standard SHAP output (referred to as SHAP Waterfall plot) that depicts this relationship for each data instance². The plot represents a random sample from the dataset, with feature values listed on the left. The base value is 24.959 which is the average model output (prediction) across all samples in the dataset. The final prediction is 13.605, determined by adding or subtracting each feature's SHAP value. For example, TotalSignals has a value of 46 and a SHAP value of +8.41, indicating an increase in the prediction by 8.41 units. Conversely, Parking, with a value of 6, decreases the prediction by 4.16 units. The unit of measurement in this case is the hourly pedestrian volume. This example shows how SHAP values help break down a model's decision into understandable feature contributions.

² In this case, i corresponds to a data instance at Grant Road & Country Club Road intersection.

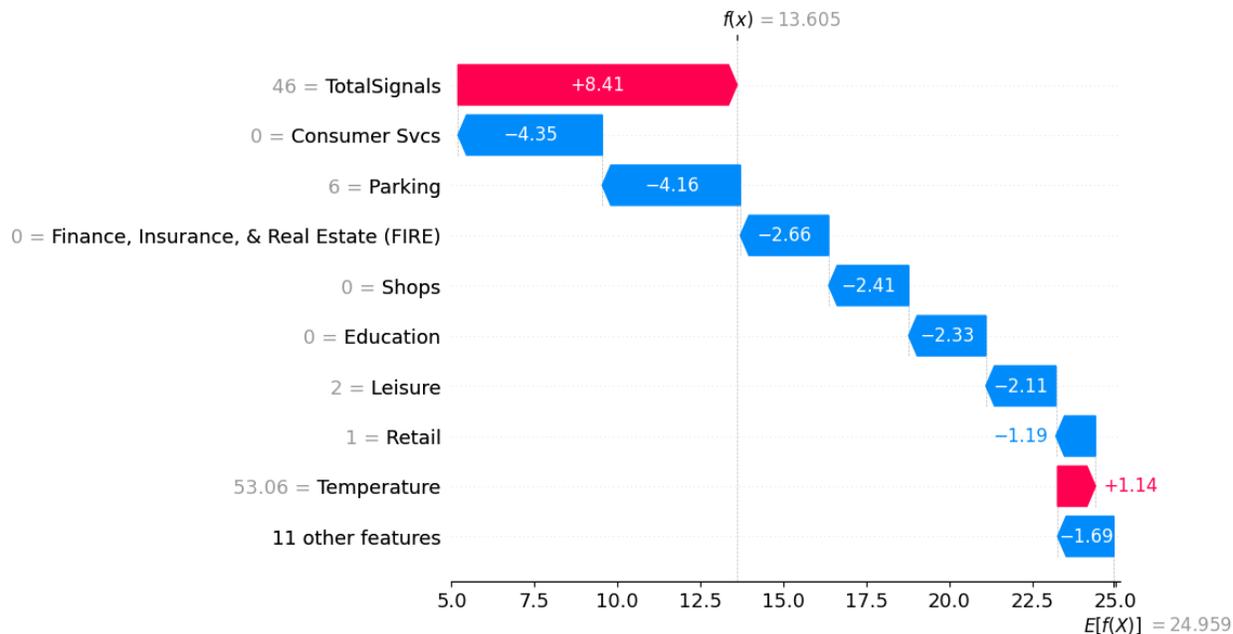


Figure 10.2. Breakdown of Feature Contributions to a Sample Prediction Using SHAP Values

Thus, when interpreting SHAP results, it is very critical not to confuse the Base Value with the average of the target variable (i.e., $Mean(Y)$), the average of each feature (i.e., $Mean(X_j) \forall j$), the minimum of each feature (i.e., $Min(X_j) \forall j$), or zero.

In addition, a “Global Feature Importance” (or “Feature Importance” for short) of each feature j is calculated by averaging the absolute SHAP Values of j across all i . This value helps assess the overall significance of each feature in predicting the model’s output.

SHAP provides a wide range of visualizations, among which Beeswarm Plots and Scatter Plots stand out for applications similar to the pedestrian volume estimation model of interest in this project. These visualizations are explained below, before deepening further into presenting our model results.

A SHAP Beeswarm Plot visualizes the overall impact of features on the model’s predictions by showing their SHAP values across all data points. It helps identify which features are most important and whether they increase or decrease predictions. The plot also reveals patterns in how feature values influence the model’s output. Below is a list of important notes about the plot:

- *Feature Ordering: Features are arranged from top to bottom based on their global feature importance, with the most important features at the top. As an example, TotalSignals is the most important feature for the model.*
- *Color of Dots: The color represents the feature’s actual value for that instance, ranging from light blue for low values to red for high values. As an example, national walking index*

(*NatWalkInd*) ranges from 3.16 to 18.83. In the Beeswarm visualization, 3.16 would be light blue and 18.83 would be red.

- *X-Axis*: Represents SHAP values, indicating how much a feature pushed the prediction higher or lower. Dots to the right increase the prediction value, while dots to the left decrease it.
- *Each Dot Represents*: A single data point where the position along the x-axis shows that instance's SHAP value for the feature.
- *Spread of Dots*: A wide horizontal spread means the feature has a varying impact across different predictions, while a narrow spread indicates a consistent effect.

Table 10.2 shows interpretation of different combinations of feature and SHAP values.

Table 10.2. Interpretation of Different Combinations of Feature and SHAP values

Feature Value & SHAP Impact	Interpretation
High feature value, positive SHAP (Red dots, right side)	As the feature value increases, the prediction value also increases, indicating a positive correlation.
High feature value, negative SHAP (Red dots, left side)	As the feature value increases, the prediction value decreases, indicating a negative correlation.
Low feature value, positive SHAP (Blue dots, right side)	As the feature value decreases, the prediction value increases, indicating a negative correlation.
Low feature value, negative SHAP (Blue dots, left side)	As the feature value decreases, the prediction also decreases, indicating a positive correlation.

A SHAP Scatter Plot shows the impact of a single feature on model predictions by plotting SHAP values against the feature's actual values. It helps reveal patterns and relationships between the feature and the target variable. The x-axis represents the actual feature values, and the y-axis shows the SHAP values, which indicate the feature's impact on the model's prediction—higher SHAP values increase the prediction, while lower SHAP values decrease it.

10.2.2 Results of SHAP

The SHAP “TreeExplainer” was used on the final XGBoost model to calculate SHAP values for each feature and sample. TreeExplainer is a class in the SHAP library that is specifically designed for tree-based models like XGBoost, which utilizes the model's internal structure to efficiently determine how much each feature contributes to a prediction. This enables a detailed breakdown of individual predictions, helping to understand the role of each feature in shaping the model's output. Additionally, to gain a general understanding of each feature's impact in model's predictive

power, global SHAP values have been calculated. **Figure 10.3** shows the global SHAP values computed using TreeExplainer, ranked by their importance.

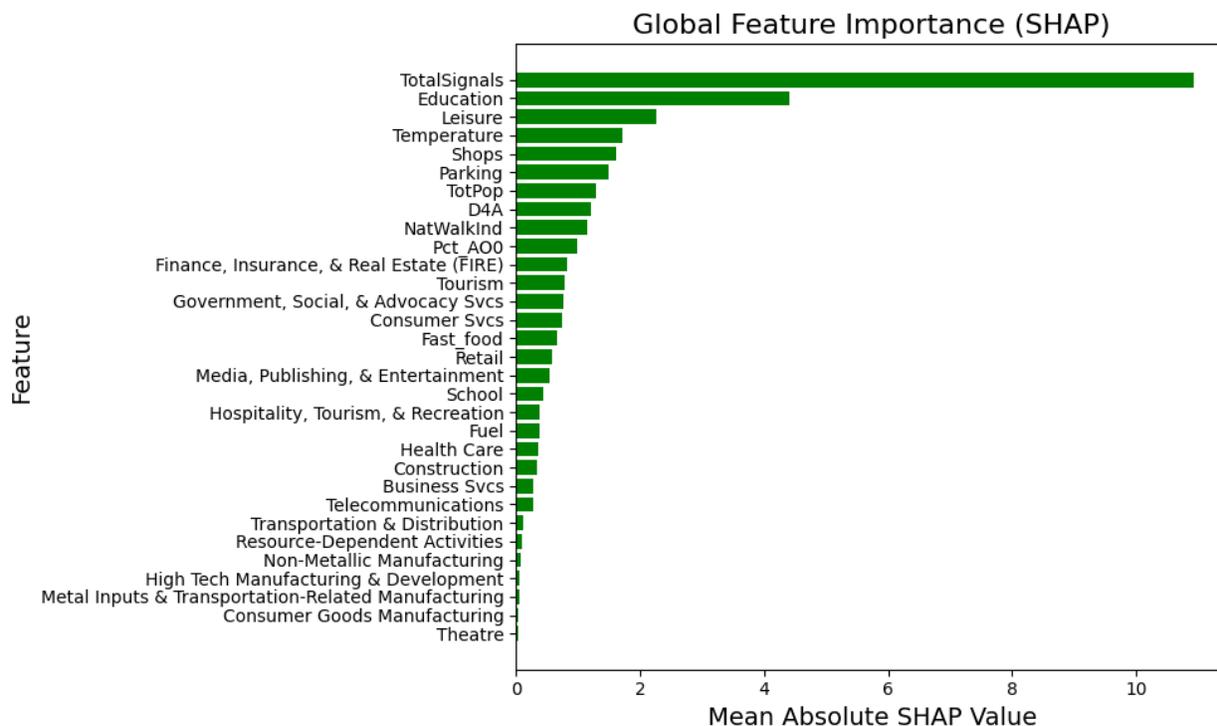


Figure 10.3. Global Feature Importance of Features

The chart shows that TotalSignals is the most significant feature, followed by Education, Leisure and Temperature, while Theatre, Consumer Goods Manufacturing, and Metal Inputs have the least influence on the model's predictions. TotalSignals has a mean absolute SHAP value of approximately 11, indicating that, on average, it shifts the model's predictions by 11 units or 11 pedestrian volumes. Despite being the most impactful feature, previously demonstrated that it is not sufficient on its own to accurately predict pedestrian volumes. While all features are included in the model, only the top 20 is displayed, as they account for the majority of the cumulative mean absolute SHAP values. **Figure 10.4** presents a Beeswarm plot to illustrate not only the importance of each feature but also how their values influence individual predictions across the dataset.

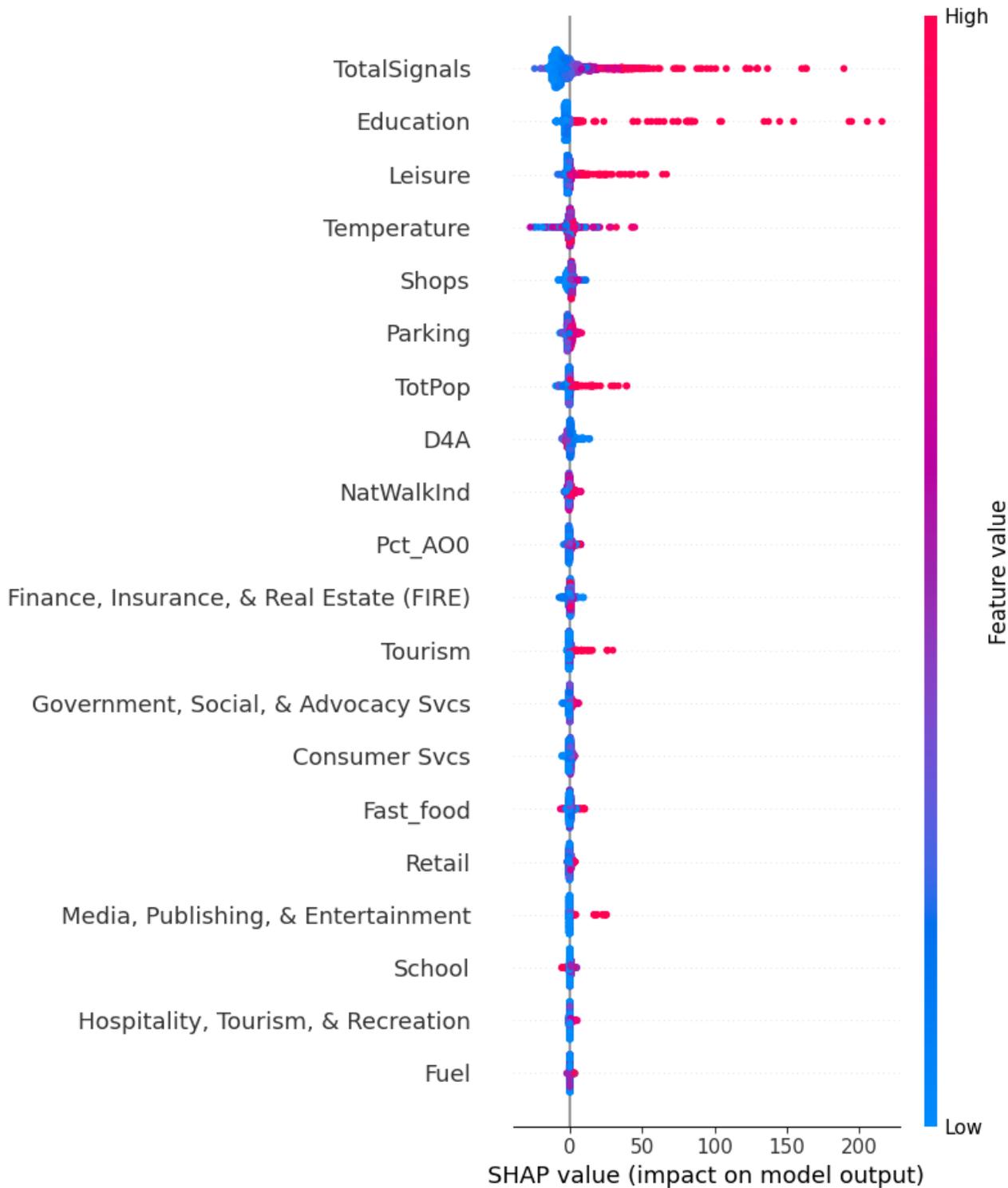


Figure 10.4. SHAP Beeswarm Plot of the Top Features of the Pedestrian Model

As shown, Push Button Events (TotalSignals) and Leisure both have positive correlation with pedestrian hourly volume. The high clustering of blue dots on the left indicates that low push-

button activity and low leisure availability generally results in low pedestrian volumes with minimal variability. In contrast, the wide spread of red dots on the right suggests that high push-button activity and high leisure availability can lead to significant but variable increases in pedestrian volume. **Figure 10.5** shows the Beeswarm plots for TotalSignals and Leisure individually.

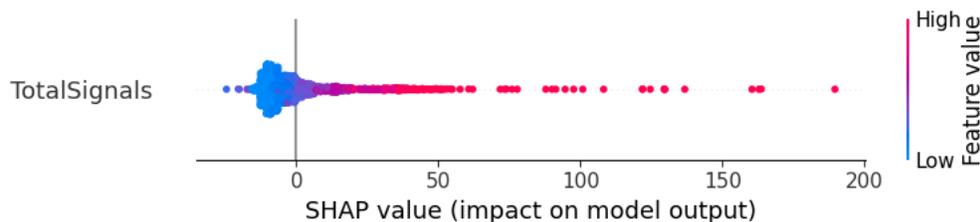


Figure 10.5. SHAP Beeswarm Plots for TotalSignals and Leisure

Education has a positive correlation with hourly pedestrian volume. Higher education values lead to a significant increase in pedestrian volume, while lower values have minimal impact and do not strongly reduce pedestrian volume. **Figure 10.6** shows the individual Beeswarm plot for Education.

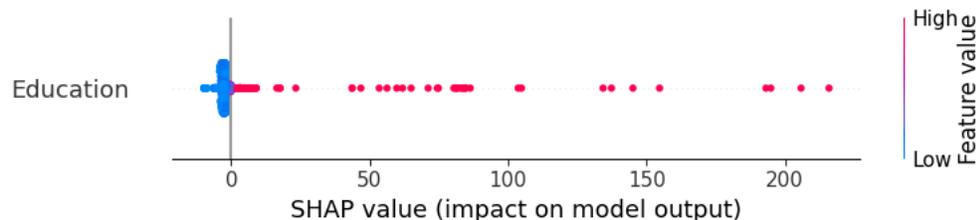


Figure 10.6. SHAP Beeswarm Plot for Education

While a Beeswarm plot provides an overview of feature importance and distribution, a Scatter plot offers more detailed look at individual feature relationships. They help visualize how specific feature values influence predictions, revealing patterns and potential threshold effects that may not be as apparent in the Beeswarm plot. This section presents SHAP scatter plots highlighting the most notable findings from the feature set. **Figure 10.7** displays the scatter plot for TotalSignals.

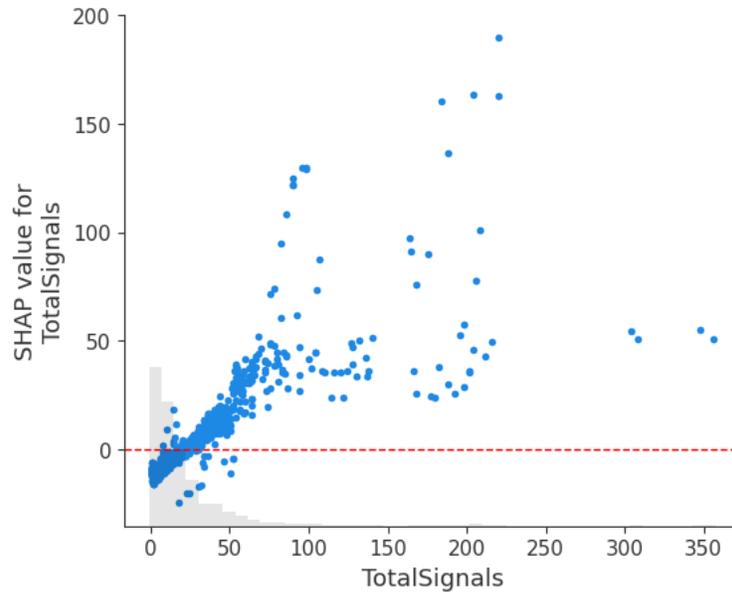


Figure 10.7. Scatter Plot of SHAP Values for Different Number of Push Button Events

Figure 10.7 reveals that the number of push-button events (TotalSignals) has a positive relationship with hourly pedestrian volume. This relationship appears nearly linear up to approximately 70 push-button events, after which it becomes more complex with greater variability in impact. This suggests that the number of push-button events is a strong predictor of pedestrian volume when the number of push-button events is relatively low, but its influence becomes less predictable at higher values.

This observation aligns with findings from previous research. For instance, Li et al. (2022) utilized Bayesian additive regression trees to estimate pedestrian crossing volumes at signalized intersections and found that traditional linear models were less effective at higher pedestrian volumes, indicating increased variability in the relationship between push-button events and actual pedestrian counts. Similarly, Humagain and Singleton (2021) analyzed pedestrian push-button data from Utah traffic signals and observed that while push-button data is valuable for monitoring pedestrian activity, the relationship between push-button presses and actual pedestrian volumes can vary due to factors such as intersection characteristics and pedestrian behavior, leading to less predictable patterns at higher volumes.

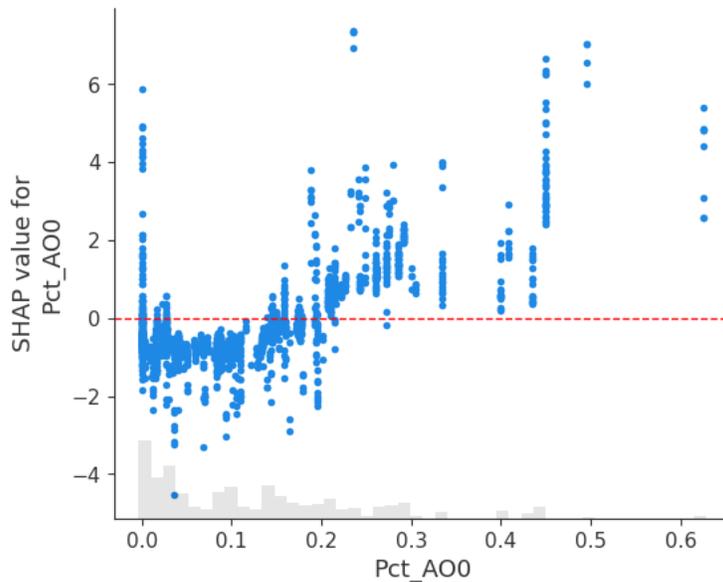


Figure 10.8. Scatter Plot of SHAP Values for Different Ratios of Zero-Car Households

Figure 10.8 highlights that the percentage of zero-car households (Pct_AO0) has a positive relationship with hourly pedestrian volume. As the proportion of zero-car households in a census block increases, pedestrian volume generally rises as well, suggesting greater reliance on walking in areas with fewer personal vehicles.

This pattern aligns with findings from Liu et al. (2024), which observed that areas with higher zero-car household rates also have higher pedestrian traffic but are more exposed to pedestrian crashes. Similarly, Yousefzadeh Barri et al. (2023) demonstrated that low-income zero-car households depend heavily on walking and transit, supporting the idea that pedestrian activity increases in regions where car ownership is low.

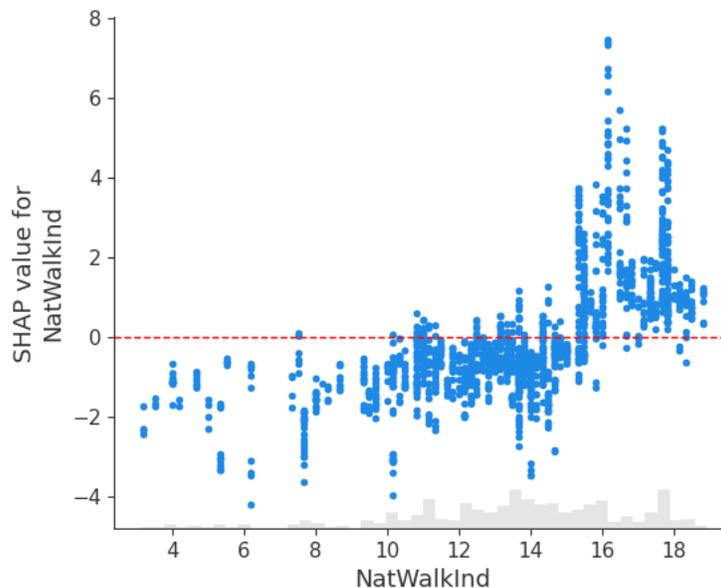


Figure 10.9. Scatter Plot of SHAP Values for Different Values of National Walking Index

Figure 10.9 shows that the National Walking Index has a positive correlation with pedestrian volume and follows a binary relationship with it; however, the exact threshold is less distinct compared to the number of shops. National Walking Index values lower than 15 are mostly irrelevant and, in most cases, correspond to lower than average pedestrian volume, while values above 15 show a clear positive effect on walking activity.

This threshold effect aligns with findings from Wali & Frank (2024), who observed that walkability indices like the National Walkability Index exhibit a nonlinear relationship with pedestrian volume, where lower values fail to significantly encourage walking, but once a sufficient walkability level is reached, pedestrian activity rises notably.

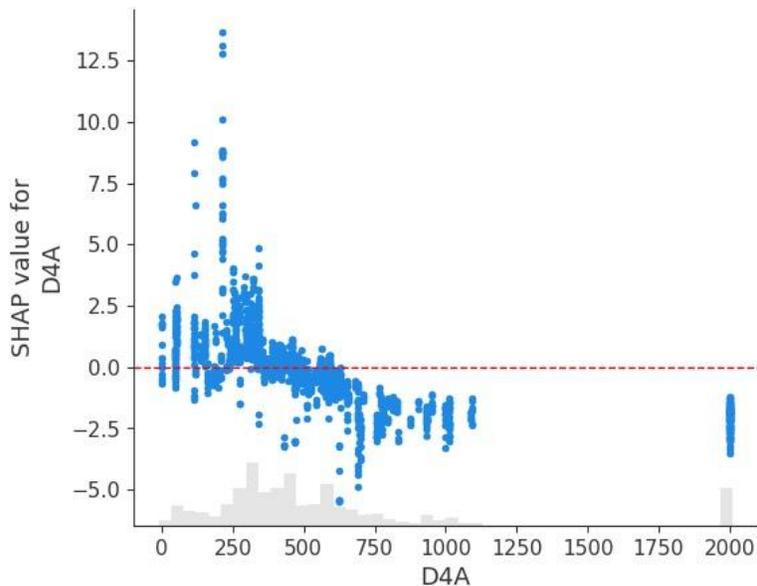
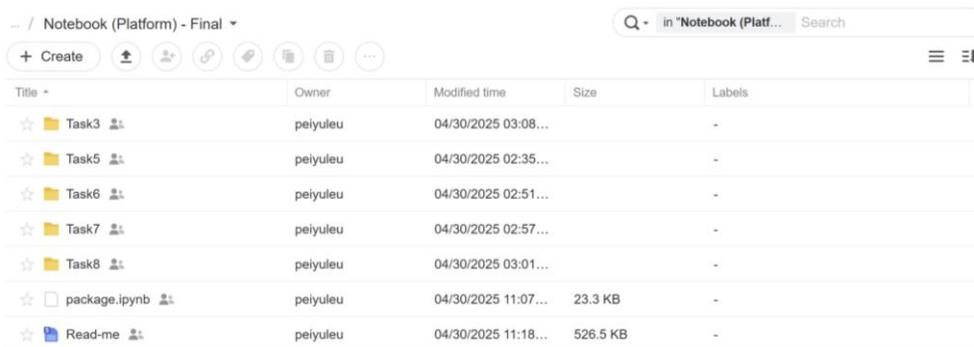


Figure 10.10. Scatter Plot of SHAP Values for Different Values of D4A

Figure 10.10 shows that distance to the nearest transit station (D4A) has a negative relationship with pedestrian volume. As D4A increases, SHAP values generally decrease, indicating that areas farther from transit tend to have lower pedestrian activity. This pattern is especially noticeable at higher distances, where the impact remains consistently negative. Overall, the figure suggests that proximity to transit stations is associated with higher pedestrian volume, while greater distance tends to reduce it.

10.2.3 Platform Development

In this section, the developed analysis platform is presented, and its structure and development process are documented, as illustrated in [Figure 10.11](#). The platform is organized into clearly labeled folders, each corresponding to a specific task, allowing for intuitive and streamlined navigation. The platform was built using Jupyter Notebook, incorporating markdown cells to document input requirements, analytical procedures, and supplementary information as needed.



Title	Owner	Modified time	Size	Labels
Task3	peiyuleu	04/30/2025 03:08...		-
Task5	peiyuleu	04/30/2025 02:35...		-
Task6	peiyuleu	04/30/2025 02:51...		-
Task7	peiyuleu	04/30/2025 02:57...		-
Task8	peiyuleu	04/30/2025 03:01...		-
package.ipynb	peiyuleu	04/30/2025 11:07...	23.3 KB	-
Read-me	peiyuleu	04/30/2025 11:18...	526.5 KB	-

Figure 10.11. Platform page

[Figure 10.12](#) presents a sample of the README file, which outlines the expected inputs and outputs for each script associated with the respective tasks. At the beginning of the document, required packages for the analysis are defined. This documentation serves as a navigation guide, helping users efficiently locate and understand the purpose of each code module within the analysis platform.

```

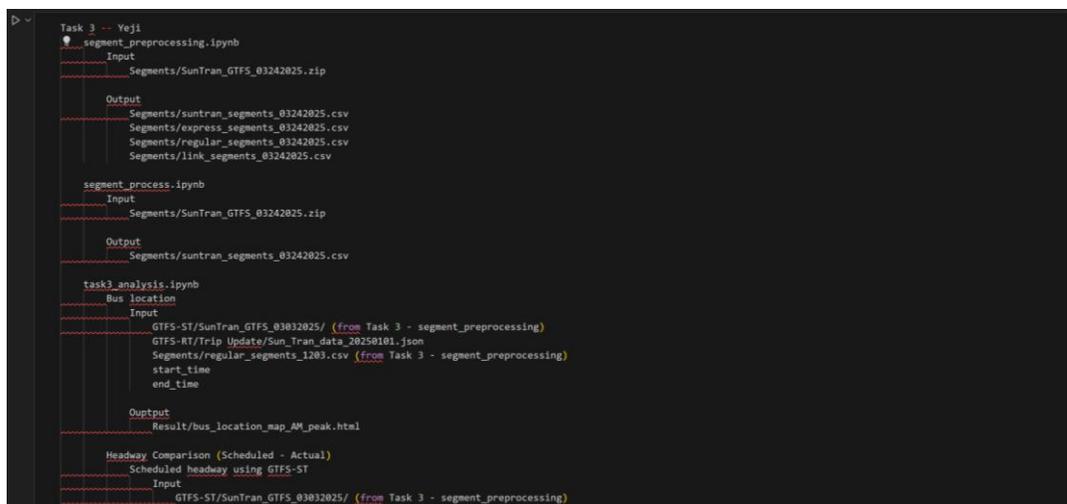
The following packages are used in the code.

Please install all the packages to have a fluent user experience.

pandas - Task 3, 4, 5, 6, 7, 8, 9, 10
pytz - Task 3, 6, 9
folium - Task 3, 8, 9, 10
seaborn - Task 3, 4, 7
matplotlib - Task 3, 4, 5, 6, 7, 8, 9
gtfs_segments - Task 3
shapely - Task 7, 8
geopandas - Task 3, 4, 7, 8
pickle - Task 8
numpy - Task 3, 4, 5, 6, 7, 8
sklearn - Task 4, 7
xgboost - Task 7
statsmodels - Task 7
shap - Task 7
meteostat - Task 7
osmnx - Task 7
fiona - Task 7
openpyxl - Task 6
requests - Task 5
logging - Task 4
pyodbc - Task 4

```

(a) Read-me file package definition



(b) Read-me file task sample

Figure 10.12. Read-me file sample

Each task folder contains two main types of files: code notebooks and sample input/output data. Example input and output files are included for reference, while the corresponding Jupyter Notebook files are clearly titled to reflect their specific functions.

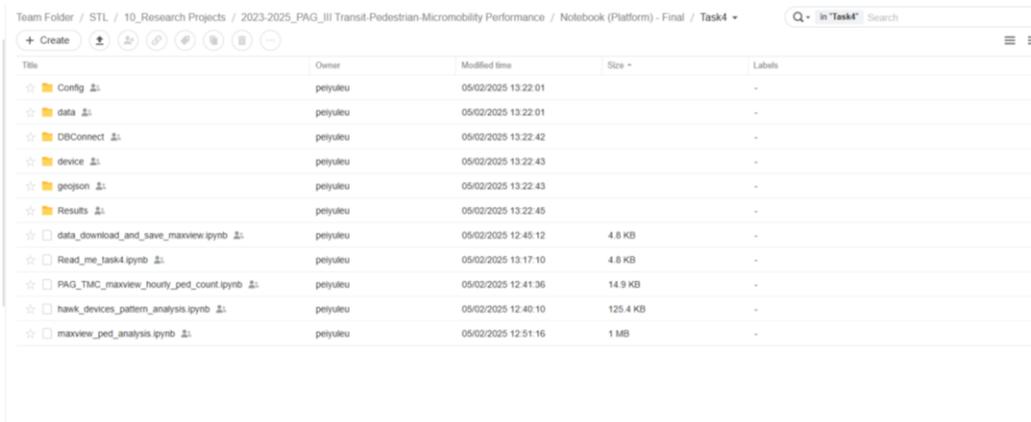
In task 3, as illustrated in [Figure 10.13](#), two Jupyter Notebooks are provided. The *segment_preprocessing* notebook contains the code used to preprocess transit segments in preparation for further analysis, as detailed in the CHAPTER 3. This preprocessing generates a list of segments for each route. The second notebook, *task3_analysis*, utilizes the preprocessed data along with the computed performance measure results to produce visualizations and analytical summaries.

Title	Owner	Modified time	Size	Labels
GTFS-RT	peiyuleu	04/30/2025 03:08...		-
GTFS-ST	peiyuleu	04/30/2025 03:10...		-
Results	peiyuleu	04/30/2025 03:10...		-
segment_based_kpis	peiyuleu	04/30/2025 03:11...		-
Segments	peiyuleu	04/30/2025 03:10...		-
segment_preprocessing.i...	peiyuleu	04/29/2025 14:33...	7 KB	-
task3_analysis.ipynb	peiyuleu	04/29/2025 14:56...	13 MB	-

Figure 10.13. Task 3 code platform folder

In Task 4, as illustrated in [Figure 10.14](#), the project directory is systematically organized. The folder includes subdirectories for code, a readme file, input data, and sample output results. Th

code folder contains several analysis scripts, including *data_download_and_save_maxview*, *PAG_TMC_maxview_hourly_ped_count*, *hawk_devices_pattern_analysis*, *maxview_ped_analysis*, each housed in a separate Jupyter notebook with clearly defined titles. The Read_me file outlines the input and output files associated with each script. Data query scripts are located within the Config and DBConnect folders. Input datasets are stored in data, device, and geojson directories, with their roles and formats also described in the Read_me documentation.



Title	Owner	Modified time	Size	Labels
Config	petyuleu	05/02/2025 13:22:01		-
data	petyuleu	05/02/2025 13:22:01		-
DBConnect	petyuleu	05/02/2025 13:22:42		-
device	petyuleu	05/02/2025 13:22:43		-
geojson	petyuleu	05/02/2025 13:22:43		-
Results	petyuleu	05/02/2025 13:22:45		-
data_download_and_save_maxview.ipynb	petyuleu	05/02/2025 12:45:12	4.8 KB	-
Read_me_task4.ipynb	petyuleu	05/02/2025 13:17:10	4.8 KB	-
PAG_TMC_maxview_hourly_ped_count.ipynb	petyuleu	05/02/2025 12:41:36	14.9 KB	-
hawk_devices_pattern_analysis.ipynb	petyuleu	05/02/2025 12:40:10	125.4 KB	-
maxview_ped_analysis.ipynb	petyuleu	05/02/2025 12:51:16	1 MB	-

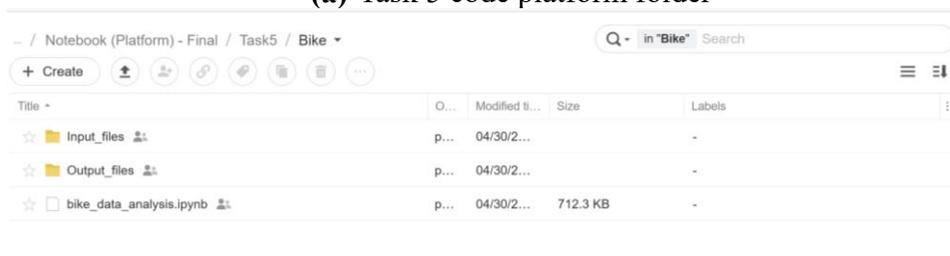
Figure 10.14. Task 4 code platform folder

In Task 5, as shown in [Figure 10.15](#), the directory is organized into two subfolders corresponding to the two micromobility modes explored: e-scooters and bikes. Each subfolder contains an example input data folder, an example output data folder, and a Jupyter Notebook used for analysis. The notebook includes detailed markdown annotations that describe the input requirements, expected outputs, and step-by-step analysis procedures, ensuring clarity and transparency throughout the workflow.



Title	O...	Modified ti...	Size	Labels
Bike	p...	04/30/2...		-
E-Scooter	p...	04/30/2...		-

(a) Task 5 code platform folder



Title	O...	Modified ti...	Size	Labels
Input_files	p...	04/30/2...		-
Output_files	p...	04/30/2...		-
bike_data_analysis.ipynb	p...	04/30/2...	712.3 KB	-

(b) Task 5 code platform folder – Bike

Title	Owner	Modified time	Size	Labels
☆ Input_files	peiyuleu	04/30/2025 02:51...	-	-
☆ Output_files	peiyuleu	04/30/2025 02:53...	-	-
☆ E-Scooter_data_analysis.ipynb	peiyuleu	04/30/2025 02:55...	1 MB	-

(c) Task 5 code platform folder – E-Scooter

Figure 10.15. Task 5 code platform folder

As shown in [Figure 10.16](#), the Task 6 platform folder follows a structure similar to that of the Task 3 platform. It includes sample input and output files, along with three Jupyter Notebooks: *SunTran_Performance*, *SunTran_analysis_visualization*, and *SunLink_SunExpress_Performance_measures*. The *SunTran_Performance* notebook contains the code to compute a variety of performance measures—namely speed, headway, headway irregularity, headway reliability, travel time reliability, and arrival delay—across the entire Sun Tran transit system. Building on these outputs, the *SunTran_analysis_visualization* and *SunLink_SunExpress_Performance_measures* notebooks generate visualizations and analysis results for each of the three transit modes: Sun Tran, Sun Express, and Sun Link.

Title	Owner	Modified time	Size	Labels
☆ GTFS-RT	peiyuleu	04/30/2025 02:51...	-	-
☆ GTFS-ST	peiyuleu	04/30/2025 02:53...	-	-
☆ Max_load_each_line	peiyuleu	04/30/2025 02:53...	-	-
☆ Results	peiyuleu	04/30/2025 02:53...	-	-
☆ segment_based_kpis	peiyuleu	04/30/2025 02:55...	-	-
☆ Segments	peiyuleu	04/30/2025 02:55...	-	-
☆ Max_load_each_line_AP...	peiyuleu	04/24/2025 15:30...	22.7 KB	-
☆ SunLink_SunExpress_Per...	peiyuleu	04/30/2025 02:33...	1.2 MB	-
☆ SunTran_analysis_visuali...	peiyuleu	04/24/2025 15:30...	58.8 KB	-
☆ SunTran_Performance_m...	peiyuleu	04/29/2025 13:25...	49.7 KB	-

Figure 10.16. Task 6 code platform folder

The Task 7 platform folder contains the pedestrian volume modeling code, *Pedestrian_model*, as well as additional performance measure scripts, which are organized within the "*Pedestrian Analysis Code*" subfolder as depicted in [Figure 10.17](#). This folder also includes the necessary input datasets for both the modeling and analysis components. Each Jupyter Notebook provides detailed

explanations of the required inputs and processing steps using markdown annotations, ensuring transparency and ease of use for users navigating the code.

Title	Owner	Modified time	Size	Labels
hpt	peiyuleu	04/30/2025 02:57...		-
Input	peiyuleu	04/30/2025 02:57...		-
Natl_WI.gdb	peiyuleu	04/30/2025 02:57...		-
Pedestrian Analysis code	peiyuleu	04/30/2025 02:57...		-
SmartLocationDatabase.gdb	peiyuleu	04/30/2025 02:59...		-
Data_preparation_final.py	peiyuleu	04/24/2025 15:31...	8.3 KB	-
output_combined_data2.csv	peiyuleu	04/24/2025 15:31...	398.8 KB	-
Pedestrian_model.ipynb	peiyuleu	04/24/2025 15:31...	2.7 MB	-

Figure 10.17. Task 7 code platform folder

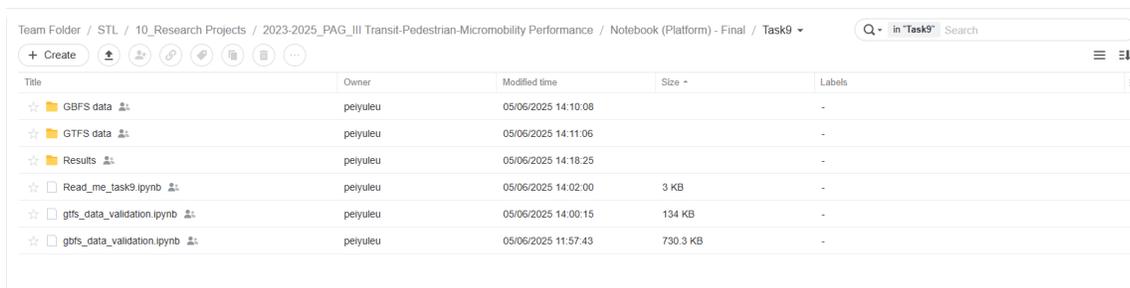
The Task 8 platform folder, as illustrated in [Figure 10.18](#), includes three separate Jupyter Notebooks, input data files, and an output results folder. The *Bike_data_visualization* and *E-Scooter_data_visualization* notebooks provide mode-specific analyses of micromobility usage patterns. Additionally, the *micromobility_data_analysis_visualization* notebook generates a series of performance measures and corresponding visualizations across both modes. Each notebook is fully documented using markdown cells, with clearly specified input file requirements and step-by-step descriptions of the analytical processes.

Title	Owner	Modified time	Size	Labels
GBFS-ST	peiyuleu	04/30/2...		-
GeoJSON	peiyuleu	04/30/2...		-
Results	peiyuleu	04/30/2...		-
Shapefile	peiyuleu	04/30/2...		-
Bike_data_visualization.ipynb	peiyuleu	04/30/2...	386.6 KB	-
E-Scooter_data_visualization.ipynb	peiyuleu	04/30/2...	717.2 KB	-
micromobility_data_analysis_visualization.ipynb	peiyuleu	04/30/2...	304.7 KB	-

Figure 10.18. Task 8 code platform folder

The Task 9 platform is organized to support data validation efforts, as shown in [Figure 10.19](#). It includes two main scripts, a readme file, sample input data, and sample output results. The two scripts, *gfts_data_validation* and *gbfs_data_validation*, perform validation analyses for GTFS and GBFS datasets, respectively. The pedestrian model validation is part of the broader pedestrian

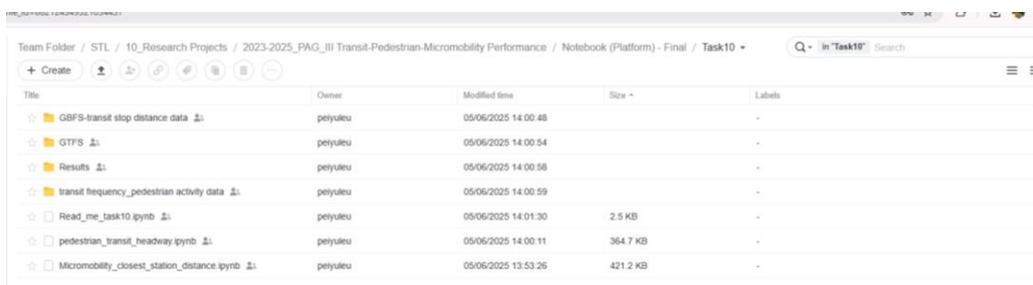
modeling workflow and is therefore located within the Task 7 platform directory. Input data for each validation script is stored in the GTFS_data and GBFS_data folders, while sample results are available in the Results_data directory. The Read_me file provides a clear description of the required inputs and expected outputs for each script to assist users in navigating the platform.



Title	Owner	Modified time	Size	Labels
☆ GBFS data	peiyuleu	05/06/2025 14:10:08	-	-
☆ GTFS data	peiyuleu	05/06/2025 14:11:06	-	-
☆ Results	peiyuleu	05/06/2025 14:18:25	-	-
☆ Read_me_task9.ipynb	peiyuleu	05/06/2025 14:02:00	3 KB	-
☆ gtfs_data_validation.ipynb	peiyuleu	05/06/2025 14:00:15	134 KB	-
☆ gbfs_data_validation.ipynb	peiyuleu	05/06/2025 11:57:43	730.3 KB	-

Figure 10.19. Task 9 platform folder

The Task 10 platform supports multimodal performance analysis and is organized with two analysis scripts, a readme file, input datasets, and sample output results, as illustrated in [Figure 10.20](#). The two scripts, `pedestrian_transit_headway` and `Micromobility_closest_station_distance`, analyze the interactions between transit and pedestrian activity, and transit and micromobility services, respectively. Since SHAP analysis is integrated with pedestrian modeling, its corresponding code is located in the Task 7 platform directory. Input data for each analysis is stored in dedicated folders, `GBFS-transit_stop_distance_data` and `transit_frequency_pedestrian_activity_data`. The `Read_me` file defines the required inputs and expected outputs for each script to guide users. Sample analysis results are available in the results folder for reference.



Title	Owner	Modified time	Size	Labels
☆ GBFS-transit stop distance data	peiyuleu	05/06/2025 14:00:48	-	-
☆ GTFS	peiyuleu	05/06/2025 14:00:54	-	-
☆ Results	peiyuleu	05/06/2025 14:00:58	-	-
☆ transit frequency pedestrian activity data	peiyuleu	05/06/2025 14:00:59	-	-
☆ Read_me_task10.ipynb	peiyuleu	05/06/2025 14:01:30	2.5 KB	-
☆ pedestrian_transit_headway.ipynb	peiyuleu	05/06/2025 14:00:11	364.7 KB	-
☆ Micromobility_closest_station_distance.ipynb	peiyuleu	05/06/2025 13:53:26	421.2 KB	-

Figure 10.20. Task 10 platform folder

APPENDIX A: MIOVISION MICROMOBILITY DATA EXPLORATION

This section details the micromobility data that was collected and initially explored but ultimately excluded from further analysis. The primary reason for this decision is that the Miovision data only captures one intersection with bike data, rendering it insufficient for comprehensive analysis.

Data Elaboration

One of the services provided by Miovision is TrafficLink. This service allows agencies to manage the transportation system remotely. The pedestrian/bike data is collected from video detection using computer vision algorithms. Miovision cameras are located in intersections and provide videos for the detection. The data could be obtained and used using their application programming interface. The Miovision API provides intersection level pedestrian data. For the intersections that the Miovision cameras are located, the information of each crosswalk turning movement counts of pedestrians is provided. **Table A.1** summarizes the available crosswalk pedestrian/bike data. The timestamp, class of detected object, side definition of the intersection, direction the pedestrian crossed at, and number of pedestrian/bike data are available from the API (Miovision, Retrieved June 19, 2024).

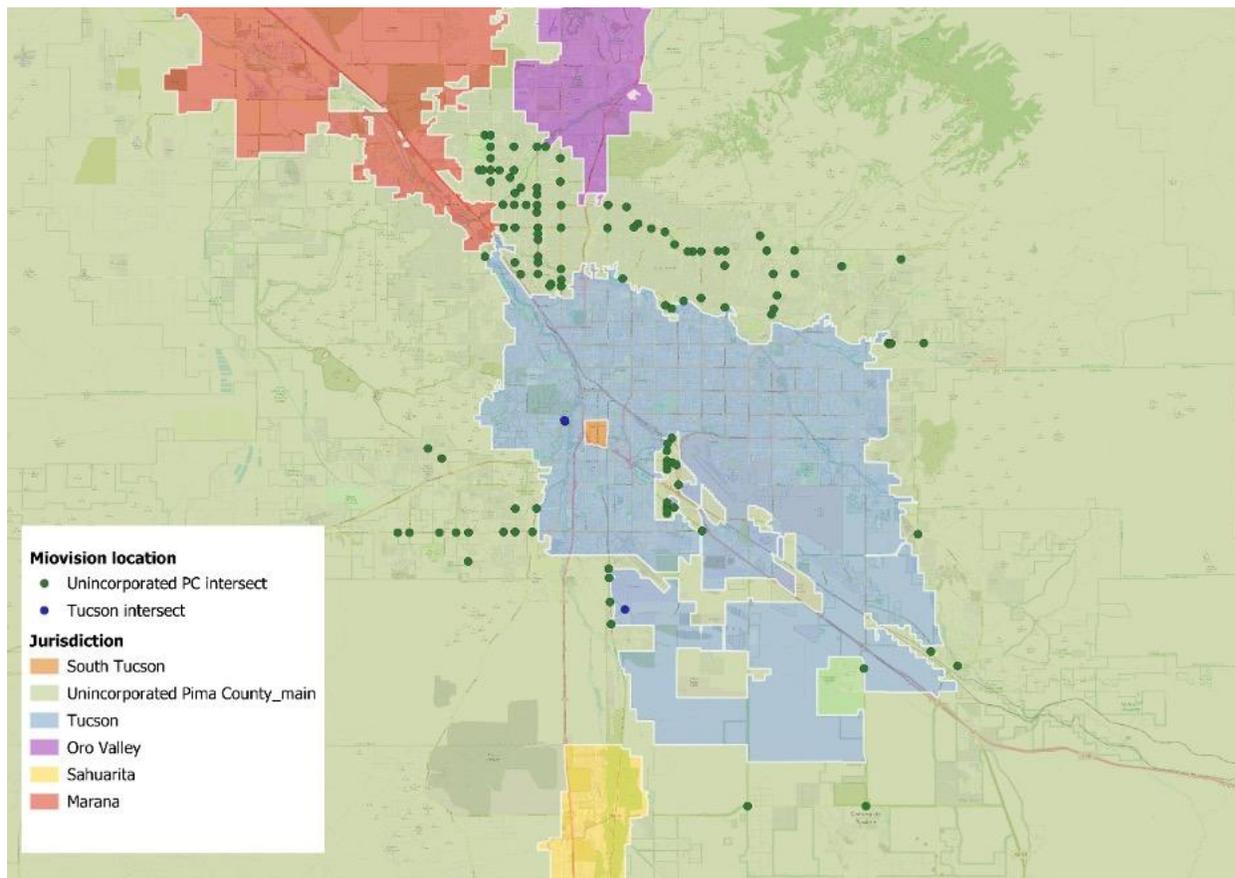
Table A.1 Miovision crosswalk pedestrian data

Data	Data Type	Description	Format
timestamp	string	Minute in time included in the bucket in the organizations time zone	ISO-8601
class	string	Buckets classification	Pedestrian/bicycle
crosswalkSide	string	Side of the intersection the pedestrians crossed at	N/S/E/W/NE/NW/SE/SW
direction	string	Way the pedestrians crossed at	CW/CCW
qty	integer	Number of pedestrians in the bucket	NA

(Reference: Miovision (June 20, 2024))

Reason Behind not Using the Data

The Miovision TrafficLink API provides appropriate data including the pedestrian counts in the intersection level. However, as **Figure A.1** illustrates, only two testing Miovision video detection cameras are installed in Tucson shown in **Table A.2**, limiting the applicability of the data for this project.



Reference: Miovision (June 20, 2024)

Figure A.1 Miovision camera location map

Table A.2 The two intersections inside city of Tucson

IntName	Longitude	Latitude
AAA CESAR TEST	-110.992294	32.2025167
San Marcos / Mission Road (Shop Test Cabinet)	-110.992750	32.203295

Even though we could analyze the data in unincorporated area, however, the original data of Miovision shows only one intersection keep updating the bicycle data, i.e., 36th St / Palo Verde Rd (-110.917179, 32.188828)

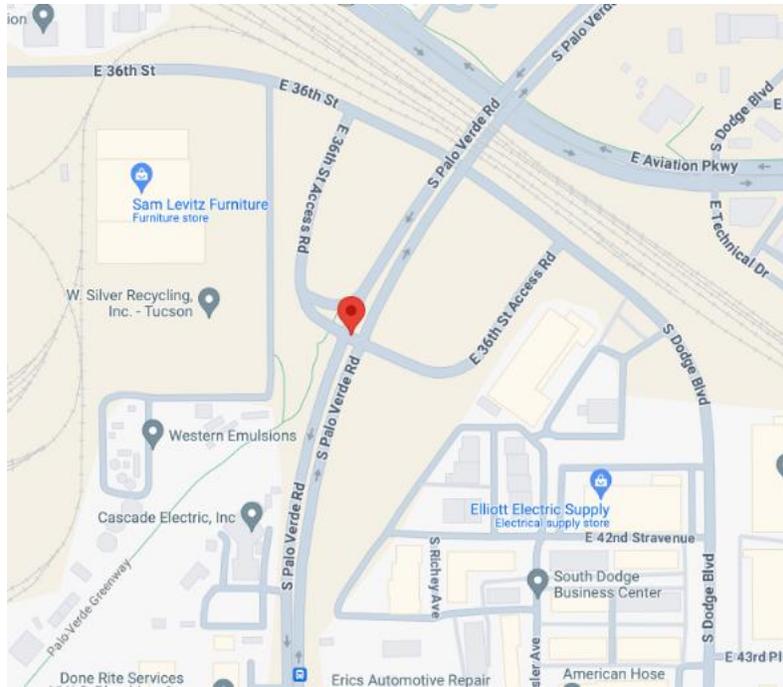


Figure A.2 **Location of 36th St / Palo Verde Rd intersection**

APPENDIX B: LIST OF INTERSECTIONS WITH THE LONGEST WAITING TIMES

Table B.1 Intersections with the longest average waiting times in AM peak hours on weekdays

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	95.0
Miovision	Hardy Rd, La Canada Dr & Overton Rd	32.36638	-110.99575	87.8
Miovision	Magee Rd, Shannon Rd & Tuscany Dr	32.34801	-111.02481	81.7
MaxView	River Rd & Campbell Ave	32.28716	-110.9439	74.9
Miovision	Benson Hwy, Swan Rd & Valencia Rd	32.13439	-110.89201	74.1
Miovision	Cortaro Farms Rd & Oldfather Dr	32.35900	-111.05542	73.3
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	69.9
Miovision	Cortaro Farms Rd & Thornydale Rd	32.35905	-111.0469	69.4
Miovision	Ina Rd & Shannon Rd	32.33739	-111.02946	68.9
Miovision	La Cholla Blvd & Ruthrauff Rd	32.29436	-111.0121	67.7

Table B.2 Intersections with the longest average waiting times in mid-day hours on weekdays

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	73.0
MaxView	River Rd & Campbell Ave	32.28716	-110.9439	69.1
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	68.1
MaxView	Irvington Rd & Calle Santa Cruz	32.16318	-110.9914	67.7
MaxView	Tanque Verde Rd & Sabino Canyon Rd	32.25079	-110.8409	63.9
Miovision	Benson Hwy, Swan Rd & Valencia Rd	32.13438	-110.89201	62.7
Miovision	La Cholla Blvd & Ruthrauff Rd	32.29436	-111.0121	62.3
Miovision	Ina Rd & La Canada Dr	32.33719	-110.99515	59.5
Miovision	Ina Rd & La Cholla Blvd	32.33728	-111.01271	59.4
MaxView	Miracle Mile Strip & Flowing Wells Rd	32.26122	-110.995	58.8

Table B.3 Intersections with the longest average waiting times in PM peak hours on weekdays

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Nogales Hwy & Old Nogales Hwy	32.07647	-110.95858	125.9
Miovision	Orange Grove Rd & Skyline Dr	32.32569	-110.93867	87.2
Miovision	La Cholla Blvd & Magee Rd	32.34797	-111.01277	82.4
Miovision	Magee Rd, Shannon Rd & Tuscany Dr	32.34801	-111.02481	80.5
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	78.9
Miovision	Foothills Mall Dr & La Cholla Blvd	32.34391	-111.01276	75.0
MaxView	Golf Links Rd & Swan Rd	32.19531	-110.8926	73.3
MaxView	Miracle Mile Strip & Flowing Wells Rd	32.26122	-110.995	73.1
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	72.7
Miovision	Cortaro Farms Rd & Oldfather Dr	32.35900	-111.05542	72.6

Table B.4 Intersections with the longest average waiting times in AM peak hours on weekends

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	73.0
MaxView	River Rd & Campbell Ave	32.28716	-110.9439	69.1
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	68.1
MaxView	Irvington Rd & Calle Santa Cruz	32.16318	-110.9914	67.7
MaxView	Tanque Verde Rd & Sabino Canyon Rd	32.25079	-110.8409	63.9
Miovision	Benson Hwy, Swan Rd & Valencia Rd	32.13439	-110.89201	62.7
Miovision	La Cholla Blvd & Ruthrauff Rd	32.29436	-111.0121	62.3
Miovision	Ina Rd & La Canada Dr	32.33719	-110.99515	59.5
Miovision	Ina Rd & La Cholla Blvd	32.33728	-111.01271	59.4
MaxView	Miracle Mile Strip & Flowing Wells Rd	32.26122	-110.995	58.8

Table B.5 Intersections with the longest average waiting times in mid-day hours on weekends

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	74.0
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	70.8
Miovision	Hospital Dr & La Cholla Blvd	32.31937	-111.01216	66.4
MaxView	Irvington Rd & Calle Santa Cruz	32.16318	-110.9914	65.3
MaxView	River Rd & Campbell Ave	32.28716	-110.9439	64.8
MaxView	Valencia Rd & Air National Guard	32.13409	-110.9491	64.4
MaxView	Tanque Verde Rd & Sabino Canyon Rd	32.25079	-110.8409	62.4
Miovision	Benson Hwy, Swan Rd & Valencia Rd	32.13439	-110.89201	61.9
MaxView	Alvernon Way & Aviation Hwy	32.19012	-110.9092	59.5
MaxView	Grant Rd & Fairview Ave	32.2503	-110.9864	58.9

Table B.6 Intersections with the longest average waiting times in PM peak hours on weekends

Source	Intersection	Latitude	Longitude	Average waiting time
Miovision	Benson Hwy & Palo Verde Rd	32.15206	-110.91766	76.1
MaxView	Alvernon Way & Aviation Hwy	32.19012	-110.9092	72.6
Miovision	Camino de la Tierra & Ina Rd	32.33747	-111.03782	71.4
MaxView	22nd St & Tucson Blvd	32.20689	-110.9352	66.3
MaxView	Houghton Rd & Valencia Rd	32.11894	-110.7727	65.6
MaxView	Kolb Rd & Irvington Rd	32.16283	-110.8409	64.1
Miovision	Mission Rd & Valencia Rd	32.13386	-111.01604	63.9
MaxView	Grant Rd & Fairview Ave	32.2503	-110.9864	61.9
MaxView	Irvington Rd & Calle Santa Cruz	32.16318	-110.9914	61.8
MaxView	Tanque Verde Rd & Sabino Canyon Rd	32.25079	-110.8409	60.9

APPENDIX C: GEOGRAPHICAL AIR POLLUTION ANALYSIS

This section presents a geographical analysis of three key air pollutants: fine particulate matter (PM_{2.5}), coarse particulate matter (PM₁₀), and ozone (O₃) in the PAG region sourced from EPA AirNow (www.airnow.gov).

PM_{2.5} refers to airborne particles with a diameter of 2.5 micrometers or less. These fine particles can penetrate deep into the respiratory system and are linked to various health issues, including cardiovascular and respiratory diseases. PM₁₀ includes particles with a diameter of 10 micrometers or less. While larger than PM_{2.5}, these particles can still be inhaled and may cause respiratory irritation and other health problems. O₃ is a harmful air pollutant at ground level and exposure to elevated levels of ground-level ozone can cause breathing difficulties and exacerbate lung diseases such as asthma.

The geographical distribution of monitoring stations for each pollutant is shown in the **Figures C.1 to C.6**. Each site is marked with a circle, where the radius and color of the circle represent the average and 95th percentile concentration of each pollutant at that location, respectively. As displayed in the figures, each pollutant is monitored using distinct sensor networks, meaning the spatial coverage and device types vary by pollutant.

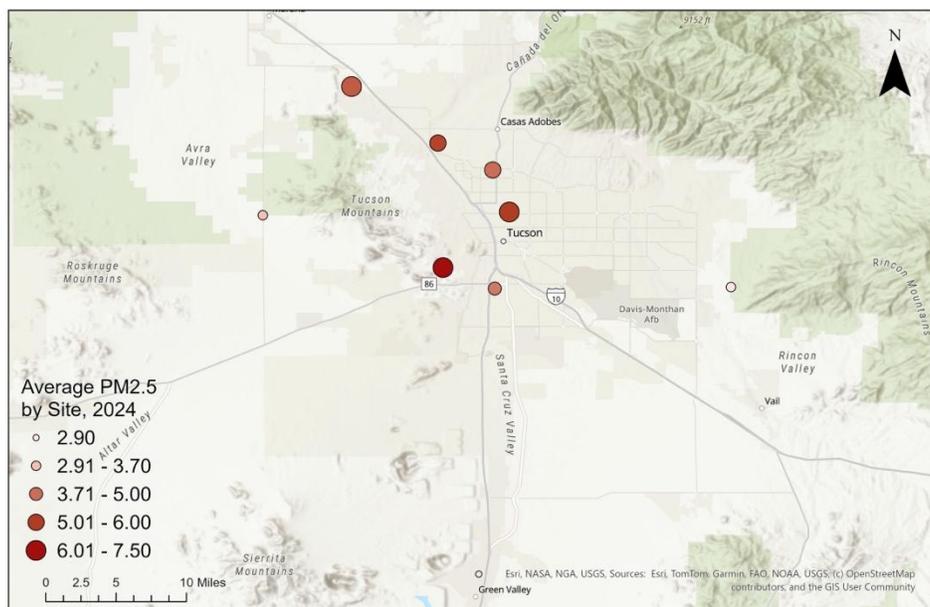


Figure C.1 Average PM_{2.5} at different sensor locations

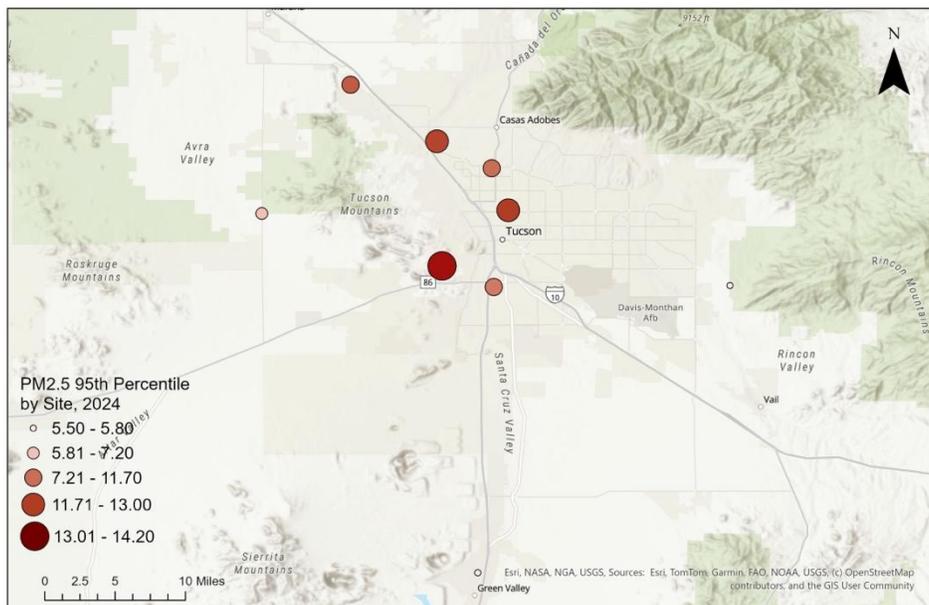


Figure C.2 95th percentile of PM2.5 at different sensor locations

Figure C.1 shows that the average PM2.5 concentrations across various locations in the City of Tucson generally remain below $7.5 \mu\text{g}/\text{m}^3$, indicating consistently low levels of fine particulate pollution. **Figure C.2** further shows that the 95th percentile values that capture higher-end exposure levels do not exceed approximately $12 \mu\text{g}/\text{m}^3$ at any site.

These findings suggest that even during peak conditions, PM2.5 levels in Tucson remain below the U.S. EPA's threshold of $12 \mu\text{g}/\text{m}^3$ for "Good" air quality based on the 24-hour standard. Overall, the figures indicates that fine particulate pollution is not a major health concern in the region under current conditions.

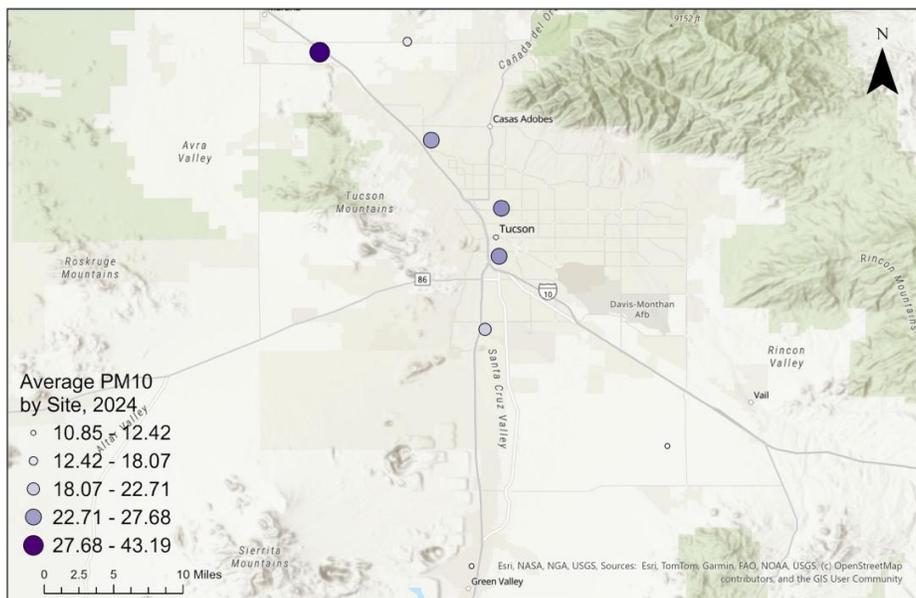


Figure C.3 Average PM10 at different sensor locations

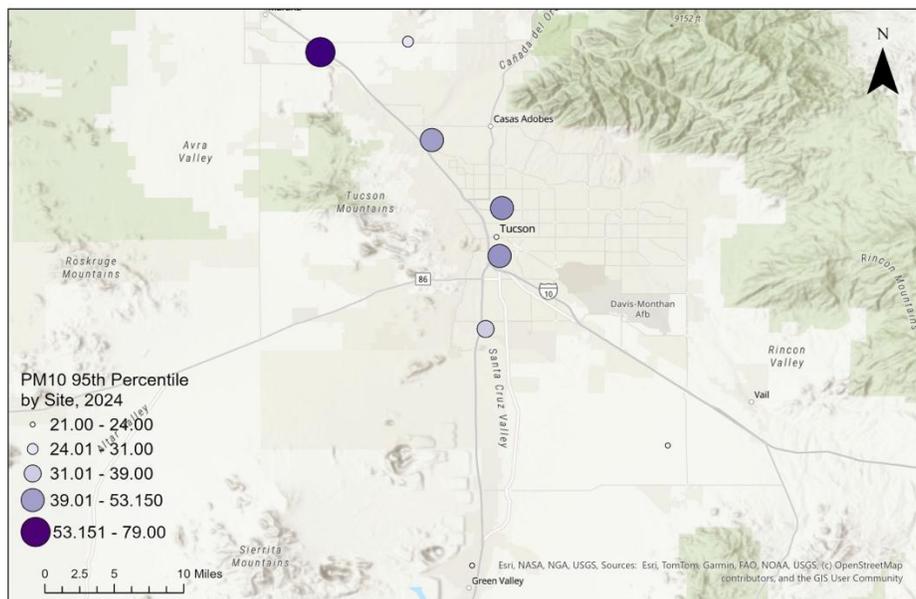


Figure C.4 95th percentile of PM10 at different sensor locations

Figure C.3 illustrates that average PM10 concentrations across Tucson are well below the EPA’s “Good” threshold of 54 $\mu\text{g}/\text{m}^3$, with the highest average at any monitoring site reaching only 43 $\mu\text{g}/\text{m}^3$. This indicates that on most days, coarse particulate matter levels remain within healthy limits.

However, **Figure C.4**, which displays the 95th percentile PM10 values, reveals that one sensor that is located in the northwest area near I-10, records values reaching up to approximately 79 $\mu\text{g}/\text{m}^3$. While this level still falls within the “Moderate” category, it suggests that there are occasional days when PM10 concentrations are elevated in that specific region. Despite these occasional highs, PM10 levels across the city generally do not pose a significant health concern.

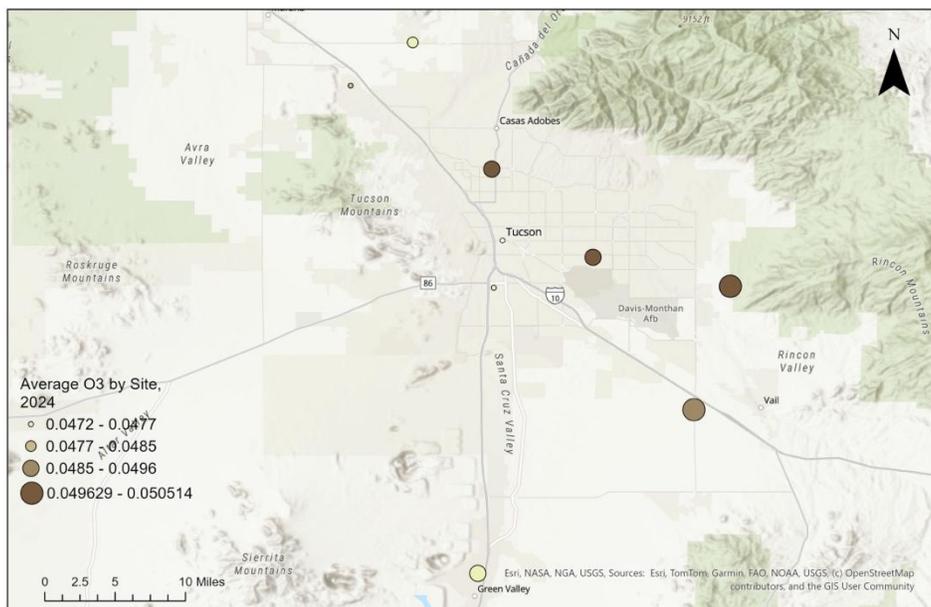


Figure C.5 Average O₃ at different sensor locations

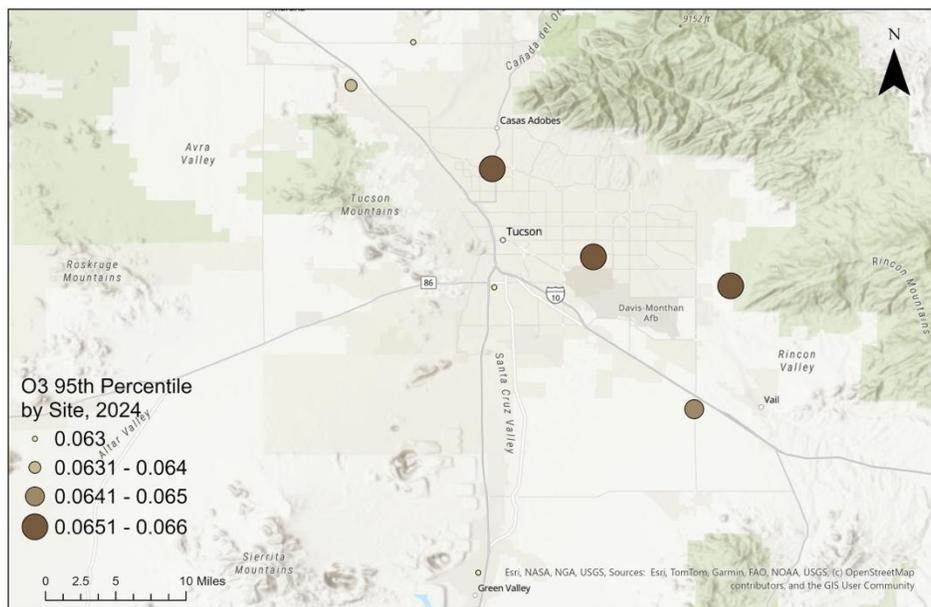


Figure C.6 95th percentile of O₃ at different sensor locations

Figure C.5 shows that the average O₃ concentrations across all monitoring sites in Tucson remain below the EPA's "Good" threshold of 0.054 ppm which indicates healthy air quality. However, it is notable that sensors located on the eastern side of the city tend to report slightly higher average values.

Figure C.6 presents the 95th percentile O₃ concentrations, revealing that all monitoring locations exceed the 0.054 ppm threshold, placing them in the "Moderate" air quality category. This indicates that although typical daily ozone levels are safe, short-term peaks may occur that could affect sensitive individuals, such as those with asthma or other respiratory conditions. Overall, ozone levels in Tucson are generally acceptable, but periodic spikes may warrant caution for sensitive individuals.

APPENDIX D: TEMPORAL AIR POLLUTION ANALYSIS

This section presents a temporal analysis of PM_{2.5}, PM₁₀, and O₃ levels in the City of Tucson. The analysis is based on the daily average concentrations, calculated by averaging readings across all available monitoring sensors for each day through EPA AirNow (www.airnow.gov).

Initially, only data from 2024 was examined. However, early results revealed not only seasonal fluctuations but also an apparent upward trend in pollutant levels over time. To better understand these patterns and isolate true seasonality effects, the analysis window was expanded to include data from January 2023 to May 2025. This broader time frame enabled a clearer view of both short-term variability and long-term seasonal trends. [Figures D.1](#) to [D.6](#) display the daily and monthly average values for each pollutant.

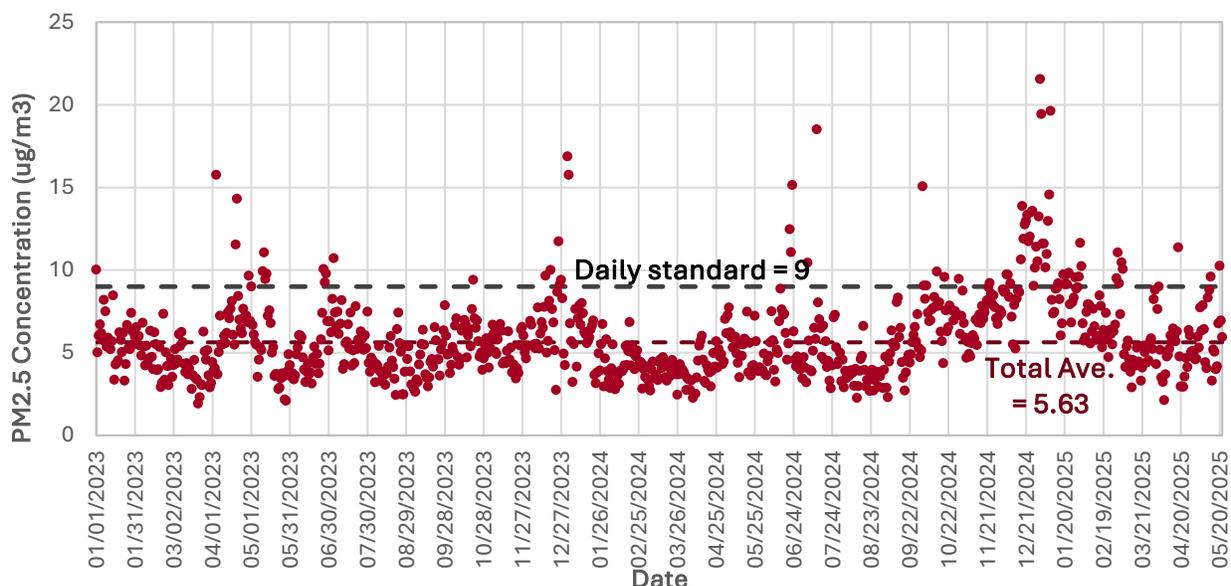


Figure D.1 Average daily PM_{2.5} concentration

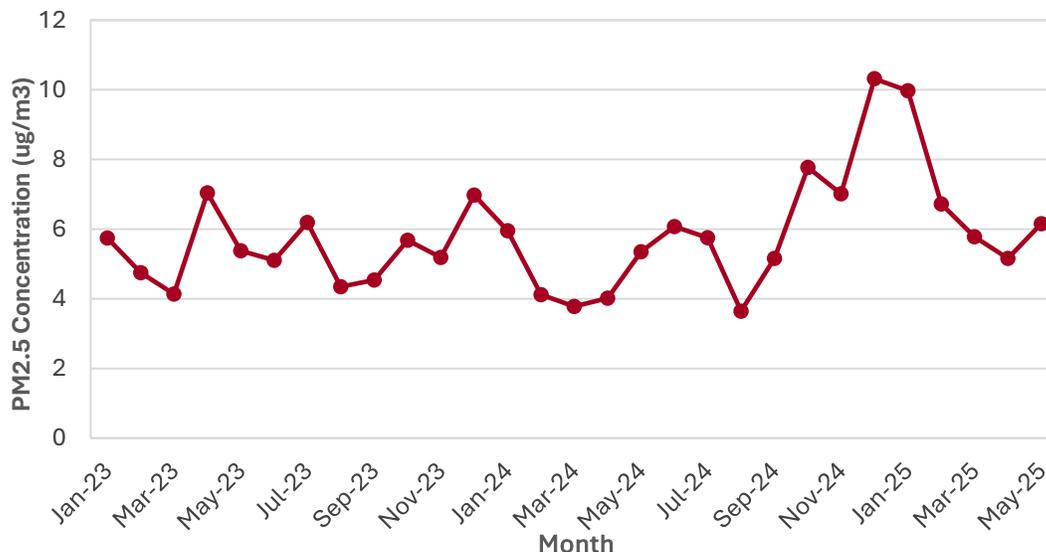


Figure D.2 Average monthly PM2.5 concentration

Figure D.1 and **Figure D.2** illustrate both the seasonal variation and a notable upward trend in PM2.5 concentrations across Tucson. The data shows that March typically experiences the lowest average concentrations, while cooler months show higher levels. Higher levels in cooler months such as December could be attributed to temperature inversion and increased residential heating in winter months. On the other hand, March is characterized by milder weather and better atmospheric mixing and tend to have improved air quality.

Notably, in 2024, average daily PM2.5 concentrations peaked around 10 $\mu\text{g}/\text{m}^3$ in December, nearly 50 percent higher than the levels observed in December 2023. Furthermore, data from the early months of 2025 continues to show elevated PM2.5 levels, suggesting a continuing upward trend.

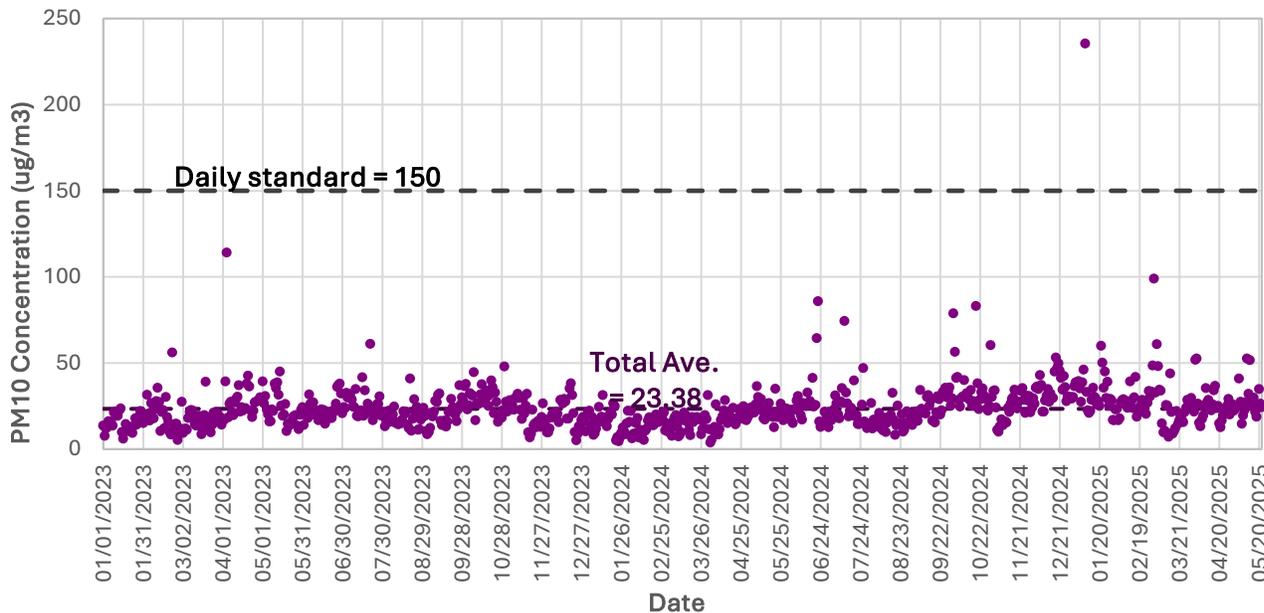


Figure D.3 Average daily PM10 concentration



Figure D.4 Average monthly PM10 concentration

Figure D.3 and **Figure D.4** show that the temporal behavior of PM10 closely resembles that of PM2.5 exhibiting both a seasonal cycle and a gradual upward trend over time. Higher concentrations are observed in cooler months, while March tends to have lowest levels. The data similarly indicates increasing levels in PM10 values and suggests increasing PM10 values in 2025.

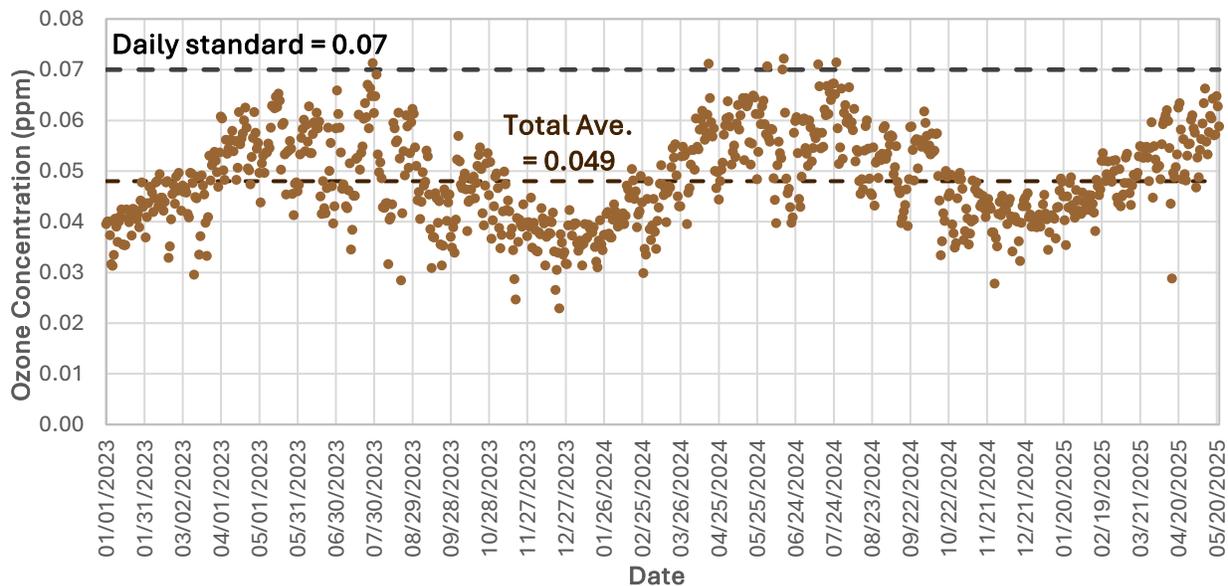


Figure D.5 Average daily O₃ concentration

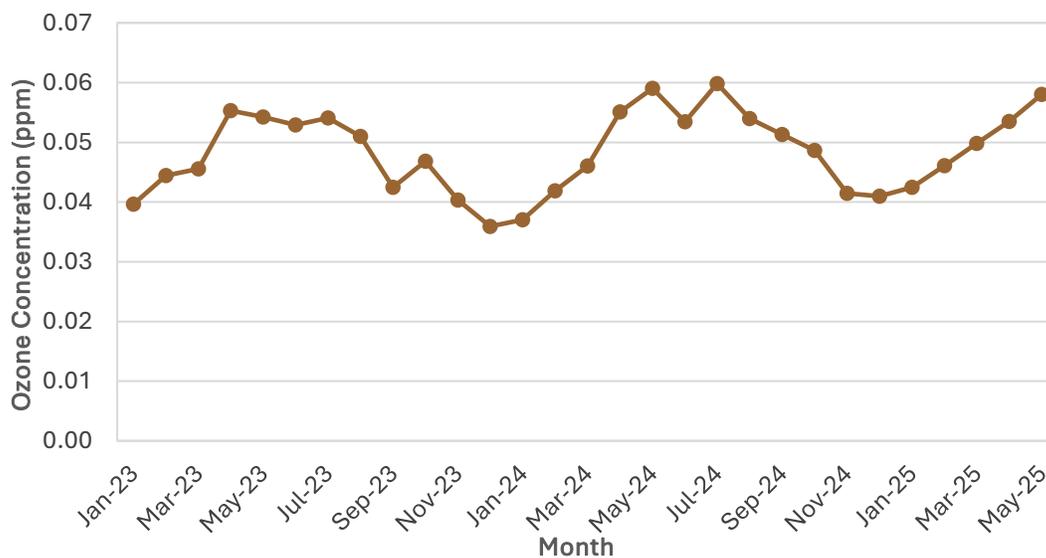


Figure D.6 Average monthly O₃ concentration

Figure D.5 and **Figure D.6** demonstrate that O₃ concentrations exhibit a clear and consistent seasonal pattern. O₃ levels are highest during the warmer months, particularly from March through July, and lowest during the cooler months of November to January. This seasonal pattern aligns with the photochemical nature of ozone formation, which is affected by sunlight and higher temperatures. In addition, the data shows that O₃ values in 2025 are slightly elevated compared to the same months in previous years, suggesting a subtle upward trend.

References

1. Ryus, P., 2003. *A summary of TCRP report 88: A guidebook for developing a transit performance-measurement system*. TCRP Research Results Digest 56.
2. Handley, J. C., Fu, L., and Tupper, L. L., 2019. *A case study in spatial-temporal accessibility for a transit system*. *Journal of Transport Geography* 75, 25-36.
3. Al Mamun, M. S., and Lownes, N. E., 2011. *A composite index of public transit accessibility*. *Journal of Public Transportation* 14.2, 69-87. <https://doi.org/10.5038/2375-0901.14.2.4>
4. Liao, C. F., and Liu, H. X., 2010. *Development of data-processing framework for transit performance analysis*. *Transportation research record* 2143.1, 34-43. <https://doi.org/10.3141/2143-05>
5. Liu, J., He, M., Schonfeld, P. M., Kato, H., Li, A., 2022 *Measures of accessibility incorporating time reliability for an urban rail transit network: A case study in Wuhan, China*. *Transportation Research Part A: Policy and Practice* 165, 471-489. <https://doi.org/10.1016/j.tra.2022.09.011>
6. Carrasco, N. 2012. *Quantifying reliability of transit service in Zurich, Switzerland: case study of bus line 31*. *Transportation research record* 2274.1, 114-125. <https://doi.org/10.3141/2274-13>
7. Rodier, C. J., and Issac, E., 2016. *Transit performance measures in California*. No. CA-MTI-16-1208. Mineta Transportation Institute.
8. Wong, J. C. 2013. *Use of the general transit feed specification (GTFS) in transit performance measurement*. Diss. Georgia Institute of Technology. <https://doi.org/10.3141/2338-02>
9. Hubbard, S. M., Bullock, D. M., Day, C. M., 2008. *Integration of real-time pedestrian performance measures into existing infrastructure of traffic signal system*. *Transportation research record* 2080.1, 37-47. <https://doi.org/10.3141/2080-05>
10. Petritsch, T. A., 2005. *Level-of-service model for pedestrians at signalized intersections*. *Transportation research record* 1939.1, 54-62. <https://doi.org/10.1177/0361198105193900107>
11. Landis, B. W., Vattikuti, V. R., Ottenberg, R. M., McLeod, D. S., Guttenplan, M., 2001. *Modeling the roadside walking environment: pedestrian level of service*. *Transportation research record* 1773.1, 82-88. <https://doi.org/10.3141/1773-10>
12. Daniel, B. D., Mohamad Nor, S. N., Rohani, M. M., Prasetijo, J., Aman, M. Y., Ambak, K., 2016. *Pedestrian footpath level of service (FOOT-LOS) model for Johor Bahru*. MATEC web of conferences. Vol. 47. EDP Sciences. <https://doi.org/10.1051/mateconf/20164703006>
13. Yan, X., Yang, W., Zhang, X., Xu, Y., Bejleri, I., Zhao, X., 2021. *A spatiotemporal analysis of e-scooters' relationships with transit and station-based bikeshare*. *Transportation research part D: transport and environment* 101, 103088. <https://doi.org/10.1016/j.trd.2021.103088>

14. Zuniga-Garcia, N., Juri, N. R., Perrine, K. A., Machemehl, R. B., 2021. *E-scooters in urban infrastructure: Understanding sidewalk, bike lane, and roadway usage from trajectory data. Case studies on transport policy* 9.3, 983-994.
<https://doi.org/10.1016/j.cstp.2021.04.004>
15. Caspi, O., Smart, M. J., Noland, R. B., 2020. *Spatial associations of dockless shared e-scooter usage." Transportation Research Part D: Transport and Environment* 86, 102396. <https://doi.org/10.1016/j.trd.2020.102396>
16. Kabra, A., Belavina, E., Girotra, Karan., 2020. *Bike-share systems: Accessibility and availability. Management Science* 66.9, 3803-3824.
<https://doi.org/10.1287/mnsc.2019.3407>
17. Berke, A., Truitt, W., Larson, K., 2024. *Is access to public bike-share networks equitable? A multiyear spatial analysis across 5 US Cities. Journal of Transport Geography* 114, 103759. <https://doi.org/10.1016/j.jtrangeo.2023.103759>
18. Pashkevich, A., Klos, M. J., Jaremski, R., Aristombayeva, M., 2021. *Decision Support Methods in Modern Transportation Systems and Networks. Springer.*
19. Moon, J., Hong, J. G., Park, T. W., 2022. *Novel Method for Traffic Estimation and Air Quality Assessment in California." Sustainability* 14.15, 9169.
<https://doi.org/10.3390/su14159169>
20. Chen, Y., Huynh, N., Comert, G., Zhang, Y., 2020. *Data-driven Multimodal Transportation Energy Consumption Prediction and Analysis Framework for Sustainable Transit and Transportation Planning.*
21. Choudhary, R., Ratra, S., Agarwal, A., 2022. *Multimodal routing framework for urban environments considering real-time air quality and congestion. Atmospheric Pollution Research* 13.9, 101525. <https://doi.org/10.1016/j.apr.2022.101525>
22. FDOT, Smart Growth America, 2015. *Complete Streets Implementation Plan – M2D2: Multimodal Development and Delivery.*
23. Khedri, B., Malarkey, D., Mackenzie, D., 2022. *Emerging Practices in Multimodal Design and Performance Measurement: Review of Recent Literature and Practical Documents. Transportation Research Record* 2676, 672-684.
<https://doi.org/10.1177/03611981221082545>
24. USDOT, 2016. *Guidebook for Developing Pedestrian & Bicycle Performance Measures.*
25. ODOT, 2021. *Operations Program Performance Management Plan.*
26. FDOT, 2017. *Transportation Systems Management & Operations – Strategic Plan.*
27. Washington State Transportation Commission, 2018. *Washington Transportation Plan – 2040 and Beyond.*
28. Clark, L. P., Millet, D. B., Marshall, J. D., 2017. *Changes in Transportation-Related Air Pollution Exposures by Race-Ethnicity and Socioeconomic Status: Outdoor Nitrogen Dioxide in the United States in 2000 and 2010. Environmental Health Perspectives* 125, 097012. <https://doi.org/10.1289/EHP959>

29. Delucchi, M. A., McCubbin, D. R., 1996. *The Contribution of Motor Vehicles and Other Sources to Ambient Air Pollution*. UC Berkeley: University of California Transportation Center.
30. Aminzadegan, S., Shahriari, A., Mehranfar, F., Abramović, B., 2022. Factors affecting the emission of pollutants in different types of transportation: A literature review. *Energy Reports* 8, 2508-2529. <https://doi.org/10.1016/j.egy.2022.01.161>
31. Liu, J., Schonfeld, P. M., F. ASCE, Peng, Q., Yin, Y., 2020. *Measures of Travel Reliability on an Urban Rail Transit Network*. *Journal of Transportation Engineering, Part A: Systems* 146 (6), 04020037. <https://doi.org/10.1061/JTEPBS.0000361>
32. Choobchian, Pooria., Mohammadi, A., Zou, B., Hair, J. F., Valinejad, M., Shin, J., Sriraj, P. S., 2024. *Calibrating walkability indicators for commute walk trips: A structural equation modeling approach*. *Transportation Research Part A: Policy and Practice* 179, 103896. <https://doi.org/10.1016/j.tra.2023.103896>
33. Kumar, S., Fedujwar, R., Agarwal, A., 2024. *Travel Time Variability of Bus Routes in Delhi using Real-Time GTFS Data*, 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS).IEEE, p.210-215.
34. Huber, S., and Rust, C., 2016. *Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM)*. *The Stata Journal* 16, 416-423.
35. Google Maps, 2024. *Google Maps Platform Documentation – Directions API overview*. Google. <https://developers.google.com/maps/documentation/directions/overview>
36. Google Inc. *MobilityData*, 2023. *Gtfs-realttime-bindings*. 2.0. <https://github.com/MobilityData/gtfs-realttime-bindings/tree/master>
37. Google Inc. *MobilityData*, 2024. *GTFS Schedule Reference*. *General Transit Feed Specification*. Retrieved May 24, 2024, from <https://gtfs.org/schedule/reference/>
38. Distefano, N. and Leonardi, S., 2023. *Fostering urban walking: strategies focused on pedestrian satisfaction*. *Sustainability*, 15(24), p.16649.
39. Gallo, M. and Marinelli, M., 2020. *Sustainable mobility: A review of possible actions and policies*. *Sustainability*, 12(18), p.7499.
40. Allen, H. and Nolmark, H., 2022. *Active Transportation, the Ultimate Low Carbon Way to Travel—A Review of International Research and Education*. *Frontiers in Sustainable Cities*, 4, p.824909.
41. Humagain, P., Singleton, P. A., 2021. *Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from Utah traffic signals*. *The Journal of Transport and Land Use* 14, 1341-1360.
42. Singleton, P. A., Runa, F., Humagain, P., 2020. *Utilizing archiving traffic signal performance measures for pedestrian planning and analysis*
43. Miovision, *Miovision Open Data API*. Retrieved June 20, 2024, from <https://docs.api.miovision.com/>
44. Streetlight, *Streetlight Training*. Retrieved June 19, 2024 from <https://training.streetlightdata.com/>

45. Google Earth 10.56.0.1, 2022. Retrieved July 8, 2024, from <https://earth.google.com/web>
46. Li, X., Wu, Y., 2021. Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. *Transportation Research Part C* 122, 102876.
47. Zuniga-Garcia, N., Juri, N. R., Perrine, K. A., Machemehl, R. B., 2021. E-scooters in urban infrastructure: Understanding sidewalk, bike lane, and roadway usage from trajectory data.
48. Berke, A., Truitt, W., Larson, K., 2024. Is access to public bike-share networks equitable? A multiyear spatial analysis across 5 U.S. Cities
49. Kabra, A., Belavina, E., Girotra, K., 2020. *Bike-Share Systems: Accessibility and Availability*
50. Google Inc. MobilityData, 2024. GBFS Reference. General Bikeshare Feed Specification. Retrieved July 27, 2024, from <https://gbfs.org/specification/reference/>
51. Lime, 2019. Lime 100 million rides, Retrieved December 30, 2019, <https://web.archive.org/web/20191230201613/https://www.li.me/en-us/100mm>
52. Kabra, A., Belavina, E., & Girotra, K. (2020). *Bike-Share Systems: Accessibility and Availability*. *Management Science*, 66(9), 3803–3824. <https://doi.org/10.1287/mnsc.2019.3407>
53. Streetlight, *Streetlight Training*. Retrieved June 19, 2024 from <https://training.streetlightdata.com/>
54. Alarcon, D., Raine, R., Credio, S., Ginn, S., Toothaker, J., Hansen, K., Bemis, A., Durband, D., Kessler, T., Yusufova, M., Lusk, R., Johnson, P., Whiting, J., Longanecker, D., Cervantez, T., Ramirez, M., Carranza, C., Leon, V., Aranda, E., ... Anderson, M. (2020). *E-scooter Pilot Program Evaluation: City of Tucson*.
55. Federal Highway Administration (FHWA). (1997). *Travel Time Data Collection Handbook*. <https://www.fhwa.dot.gov/ohim/handbook/chap1.pdf>
56. Jensen, S. U. (2007). Pedestrian and bicyclist level of service on roadway segments. *Transportation Research Record*, 2031(1), 43–51. <https://doi.org/10.3141/2031-06>
57. Karimpour, A., Anderson, J. C., Kothuri, S., & Wu, Y. J. (2022). Estimating pedestrian delay at signalized intersections using high-resolution event-based data: a finite mixture modeling method. *Journal of Intelligent Transportation Systems*, 26(5), 511–528. <https://doi.org/10.1080/15472450.2020.1863404>
58. Li, X., & Wu, Y. J. (2021). Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. *Transportation Research Part C: Emerging Technologies*, 122, 102876. <https://doi.org/10.1016/j.trc.2020.102876>
59. Marisamynathan, S., & Vedagiri, P. (2018). A new approach to estimate pedestrian delay at signalized intersections. *Transport*, 33(1), 249–259. <https://doi.org/10.3846/16484142.2016.1174881>

60. Netes, V. (2018). *End-to-end availability of cloud services*. 2018 22nd Conference of Open Innovations Association (FRUCT), 198–203. <https://doi.org/10.23919/FRUCT.2018.8468302>
61. TF-Resource. (n.d.). *Model validation and reasonableness checking/assignment*. Travel Forecasting Resource.
62. Vogel, K. (2003). *A comparison of headway and time to collision as safety indicators*. *Accident Analysis & Prevention*, 35(3), 427–433. [https://doi.org/10.1016/S0001-4575\(02\)00022-2](https://doi.org/10.1016/S0001-4575(02)00022-2)
63. Xu, P., Li, X., Wu, Y. J., & Noh, H. (2023). *Network-level turning movement counts estimation using traffic controller event-based data*. *Journal of Intelligent Transportation Systems*, 27(5), 677–691. <https://doi.org/10.1080/15472450.2021.1992335>
64. Newmark, G. L. 2024. *Assessing GTFS Accuracy*.
65. Arias, D., Todd, K., Krieger, J., Maddox, S., Haley, P., Watkins, K. E., Berrebi, S. 2021. *Using GTFS to Calculate Travel Time Savings Potential of Bus Preferential Treatments*, *Transportation Research Record*, 2675.9, p.1643-1654
66. Elliott, T., Lumley, T. 2020. *Modeling the travel time of transit vehicles in real-time through a GTFS-based road network using GPS vehicle locations*. *Australian & New Zealand Journal of Statistics* 62.2, p.153-167
67. Wu, J., Du, B., Gong, Z., Wu, Q., Shen, J., Zhou, L., Cai, C. 2023. *A GTFS data acquisition and processing framework and its application to train delay prediction*. *International Journal of Transportation Science and Technology*, 12.1, p.201-216
68. Carrasco, N. 2012. *Quantifying reliability of transit service in Zurich, Switzerland: case study of bus line 31*. *Transportation research record* 2274.1, p.114-125. <https://doi.org/10.3141/2274-13>
69. Harsha, M. M., Mulangi, R. H., 2021. *Impact of Side Friction on Travel Time Reliability of Urban Public Transit*, *International Journal of Civil Engineering*, p.1221-1237. <https://doi.org/10.1007/s40999-021-00622-y>
70. Ahmed, T., Moeinaddini, M., Almoshaogeh, M., Jamal, A., Nawaz, I., & Alharbi, F. (2021). *A New Pedestrian Crossing Level of Service (PCLOS) Method for Promoting Safe Pedestrian Crossing in Urban Areas*. *International Journal of Environmental Research and Public Health*, 18(16), 8813. <https://doi.org/10.3390/ijerph18168813>.
71. Campisi, T., Tesoriere, G., Skoufas, A., Zeglis, D., Andronis, C., & Basbas, S. (2022). *Perceived pedestrian level of service: The case of Thessaloniki, Greece*. *Transportation Research Procedia*, 60, 124–131. <https://doi.org/10.1016/j.trpro.2021.12.017>.
72. Cooper, C. H., Harvey, I., Orford, S., & Chiaradia, A. J. (2021). *Using multiple hybrid spatial design network analysis to predict longitudinal effect of a major city centre redevelopment on pedestrian flows*. *Transportation*, 643–672.
73. Daniel, B. D., Nor, S. N. M., Md Rohani, M., Prasetijo, J., Aman, M. Y., & Ambak, K. (2016). *Pedestrian footpath level of service (FOOT-LOS) model for Johor Bahru*. *MATEC Web of Conferences*, 47, 3–7. <https://doi.org/10.1051/mateconf/20164703006>

74. de Montigny, L., Ling, R., & Zacharias, J. (2012). *The effects of weather on walking rates in nine cities*. *Environment and Behavior*, 821-840.
75. Dixon, L. B. (1996). *Bicycle and Pedestrian Level-of-Service Performance Measures and Standards for Congestion Management Systems*. *Transportation Research Record*, 1538(1), 1-9. <https://doi.org/10.1177/0361198196153800101>.
76. Hubbard, S. M. L., Bullock, D. M., & Day, C. M. (2008). *Integration of Real-Time Pedestrian Performance Measures into Existing Infrastructure of Traffic Signal System*. *Transportation Research Record*, 2080(1), 37-47. <https://doi.org/10.3141/2080-05>.
77. Humagain, P., & Singleton, P. A. (2021). *Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from Utah traffic signals*. *Journal of Transport and Land Use*, 2112.
78. Karimpour, A., Anderson, J. C., Kothuri, S., & Wu, Y. (2021, May 25). *Estimating pedestrian delay at signalized intersections using high-resolution event-based data: a finite mixture modeling method*. *Journal of Intelligent Transportation Systems*, 26(5), 511-528. <https://doi.org/10.1080/15472450.2021.1926246>.
79. Kim, S., & Woo, A. (2022). *Streetscape and business survival: Examining the impact of walkable environments on the survival of restaurant businesses in commercial areas based on street view images*. *Journal of Transport Geography*, 105.
80. Landis, B. W., Vattikuti, V. R., Ottenberg, R. M., McLeod, D. S., & Guttenplan, M. (2001). *Modeling the Roadside Walking Environment: Pedestrian Level of Service*. *Transportation Research Record*, 1773(1), 82-88. <https://doi.org/10.3141/1773-10>.
81. Li, P., Zhao, P., & Schwanen, T. (2020). *Effect of land use on shopping trips in station areas: Examining sensitivity to scale*. *Transportation Research Part A*, 969-985.
82. Li, X., Xu, P., & Wu, Y. J. (2022). *Pedestrian crossing volume estimation at signalized intersections using Bayesian additive regression trees*. *Journal of Intelligent Transportation Systems*, 26(5), 557-571. <https://doi.org/10.1080/15472450.2021.1933471>
83. Li, X., & Wu, Y. (2021). *Real-time estimation of pedestrian volume at Button-activated midblock crosswalks using traffic controller event-based data*. *Transportation Research Part C: Emerging Technologies*, 122. <https://doi.org/10.1016/j.trc.2020.102876>.
84. Liu, J., Das, S. Z., & Khan, M. N. (2024). *Spatial analysis of geographical disparities in pedestrian safety*. *Transport Policy*, 164-181.
85. Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. *Advances in neural information processing systems*, 30.
86. Marisamynathan, S., & Vedagiri, P. (2016). *A new approach to estimate pedestrian delay at signalized intersections*. *Transport*, 33(1), 249-259. <https://doi.org/10.3846/16484142.2016.1158208>.
87. Petritsch, T. A., Landis, B. W., McLeod, P. S., Huang, H. F., Challa, S., & Guttenplan, M. (2005). *Level-of-Service Model for Pedestrians at Signalized Intersections*. *Transportation Research Record*, 1939(1), 54-62. <https://doi.org/10.1177/0361198105193900107>.

88. Roupail, N. M., Allen, D. P., Milazzo, J. S., & Hummer, J. E. (1998). *Recommended Procedures Chapter 13, "Pedestrians," of the Highway Capacity Manual*. North Carolina State University, Department of Civil Engineering. <https://highways.dot.gov/sites/fhwa.dot.gov/files/FHWA-RD-98-107.pdf>.
89. Wali, B., & Frank, L. D. (2024). *Redefining walkability to capture safety: Investing in pedestrian, bike, and street level design features to make it safe to walk and bike. Transportation research part A: policy and practice, 103814.*
90. Yousefzadeh Barri, E., Farber, S., Jahanshahi, H., Tiznado-Aitken, I., & Beyazit, E. (2023). *Exploring the joint impacts of income, car ownership, and built environment on daily activity patterns: a cluster analysis of trip chains. Transportmetrica A: Transport Science, 1-31.*
91. American Public Transportation Association (2022). *Public Transportation Fact Book*. Available at: <https://www.apta.com>
92. Federal Highway Administration (2019). *Does Travel Time Reliability Matter? FHWA-HOP-19-062*. Available at: <https://ops.fhwa.dot.gov>
93. Jayanthi, L.S., Pulugurtha, S.S. and Mishra, R., 2023. *Examining the association between bus transit reliability and the number of boarding passengers. Public Transport, 15(3), pp.675–696.*
94. Pulugurtha, S.S. and Jayanthi, L.S., 2021. *Does Transit Service Reliability Influence Ridership? Mineta Transportation Institute*. Available at: <https://transweb.sjsu.edu/research/2003-Transit-Reliability>
95. Singleton, P. A., & Runa, F. (2021). *Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. Transportation research record, 2675(6), 429-440.*
96. Kothuri, S., Singleton, P. A., Saheli, M. V., Yates, E., & Broach, J. (2024). *Active Transportation Counts from Existing On-Street Signal and Detection Infrastructure (No. FHWAOR-RD-24-03)*. Oregon. Dept. of Transportation. Research Section.
97. Li, X., & Wu, Y. J. (2021). *Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. Transportation research part C: emerging technologies, 122, 102876.*
98. Li, X., Xu, P., & Wu, Y. J. (2022). *Pedestrian crossing volume estimation at signalized intersections using Bayesian additive regression trees. Journal of Intelligent Transportation Systems, 26(5), 557-571.*
99. Strathman, J.G., Kimpel, T.J. and Callas, S., 2013. *Headway Deviation Effects on Bus Passenger Loads: Analysis of TriMet's Archived AVL/APC Data. Transportation Northwest (TransNow)*. Available at: <https://nacto.org>
100. UITP (2023). *What is bus headway and how it impacts public transport quality*. Available at: <https://www.uitp.org>
101. Zhang, Y. and Zhang, Y., 2023. *The adverse impact of headway variability on bus transit ridership. Transportation Research Part A: Policy and Practice, 170, pp.103–116.*

102. O'Toole, R., 2018. *Transit: The Urban Parasite*. Cato Institute, Policy Analysis No. 845. Available at: <https://www.cato.org>
103. Pessaro, B., Catalá, M., Wang, Z., & Spicer, M. (2017). *Impact of transit stop location on pedestrian safety*.
104. Mandolakani, F. S., Subedi, A., Singleton, P. A., & Mekker, M. (2024). *Pedestrian Safety and Traffic Operations Around Near-Side Versus Far-Side Transit Stops: Emerging Observational Evidence from Utah*. arXiv preprint arXiv:2408.08890.
105. Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. *Advances in neural information processing systems*, 30.
106. Li, X., Xu, P., & Wu, Y. J. (2022). *Pedestrian crossing volume estimation at signalized intersections using Bayesian additive regression trees*. *Journal of Intelligent Transportation Systems*, 26(5), 557-571. <https://doi.org/10.1080/15472450.2021.1933471>
107. Humagain, P., & Singleton, P. A. (2021). *Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from Utah traffic signals*. *Journal of Transport and Land Use*, 2112.
108. Liu, J., Das, S. Z., & Khan, M. N. (2024). *Spatial analysis of geographical disparities in pedestrian safety*. *Transport Policy*, 164-181.
109. Yousefzadeh Barri, E., Farber, S., Jahanshahi, H., Tiznado-Aitken, I., & Beyazit, E. (2023). *Exploring the joint impacts of income, car ownership, and built environment on daily activity patterns: a cluster analysis of trip chains*. *Transportmetrica A: Transport Science*, 1-31.
110. Wali, B., & Frank, L. D. (2024). *Redefining walkability to capture safety: Investing in pedestrian, bike, and street level design features to make it safe to walk and bike*. *Transportation research part A: policy and practice*, 103814.
111. Pulugurtha, S.S. and Jayanthi, L.S., 2021. *Does Transit Service Reliability Influence Ridership?* Mineta Transportation Institute. Available at: <https://transweb.sjsu.edu/research/2003-Transit-Reliability>
112. WMATA (Mineta Institute), 2021. *Performance and Safety Analysis for Washington Metropolitan Area Transit Authority*. San Jose, CA: Mineta Transportation Institute.
113. Zhang, Y. and Zhang, W., 2023. *Urban bus service reliability: A statistical approach to analyze headway variations*. *Transportation Research Part A: Policy and Practice*, 165, pp.278–293.
114. UITP, 2023. *Frequency is Freedom: Measuring and Improving Transit Headways*. Brussels: International Association of Public Transport.
115. Strathman, J.G., Kimpel, T.J., Dueker, K.J., Gerhart, R.L. and Callas, S., 2013. *Evaluation of transit operations: a data envelopment analysis approach*. *Transportation Research Part A: Policy and Practice*, 37(10), pp.911–928.
116. FHWA, 2019. *Transit Performance Monitoring and Evaluation*. Washington, DC: Federal Highway Administration.

117. Jayanthi, J., Pulugurtha, S. and Lentz, J., 2023. *Influence of bus stop features on ridership and on-time performance: A case study in Charlotte, NC. Journal of Public Transportation*, 25, pp.74–88.
118. APTA, 2022. *Public Transportation Fact Book*. Washington, DC: American Public Transportation Association.
119. Cato Institute, 2018. *The High Cost of Slow Buses. Policy Report*. Washington, DC.
120. MBTA, 2018. *Massachusetts Bay Transportation Authority On-Time Performance Audit*. Boston, MA: Office of the Inspector General.
121. Roupail, N.M., Hummer, J.E. and Milazzo II, J.S., 2005. *Level of service for pedestrians at signalized intersections. Transportation Research Record*, 1920(1), pp.52–62.
122. NCHRP, 2022. *Guidebook for Incorporating Pedestrian and Bicycle Performance Measures into the Planning and Design Processes*. NCHRP Report 969. Washington, DC: Transportation Research Board.
123. Landis, B.W., Vattikuti, V.R., Ottenberg, R.M., McLeod, D.S. and Guttenplan, M., 2001. *Modeling the roadside walking environment: Pedestrian level of service. Transportation Research Record*, 1773(1), pp.82–88.
124. NYC DOT, 2015. *Pedestrian Level of Service Study: Sidewalk Congestion in Manhattan*. New York: NYC Department of Transportation.
125. Singleton, P.A. and Runa, B., 2021. *Estimating pedestrian volumes at signalized intersections using push-button data. Transportation Research Record*, 2675(4), pp.262–272.
126. Choobchian, P., Mohammadi, A., Zou, B. and Wang, Y., 2024. *Data-driven approach to evaluating walkability using SEM. Transportation Research Part F: Traffic Psychology and Behaviour*.
127. Kabra, A., Belavina, E. and Girotra, K., 2020. *Bike-share systems: Accessibility and availability. Management Science*, 66(9), pp.3803–3824.
128. Caspi, O., Smart, M.J. and Noland, R.B., 2020. *Spatial associations of dockless shared e-scooter usage. Transportation Research Part D: Transport and Environment*, 86, p.102396.