Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-18/0-6927-P6		

4. Title and Subtitle
GUIDE FOR SEASONAL ADJUSTMENT AND CROWDSOURCED DATA SCALING

5. Report Date
Resubmitted: September 2018

6. Performing Organization Code

7. Author(s)
Bahar Dadashova, Greg Griffin, Subasish Das, Shawn Turner, and Madison Graham

8. Performing Organization Report No.
Product 0-6927-P6

9. Performing Organization Name and Address
Texas A&M Transportation Institute
The Texas A&M University System
College Station, Texas 77843-3135

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
Project 0-6927

12. Sponsoring Agency Name and Address
Texas Department of Transportation
Research and Technology Implementation Office
125 E. 11th Street
Austin, Texas 78701-2483

13. Type of Report and Period Covered
Product
February 2018–August 2018

14. Sponsoring Agency Code

15. Supplementary Notes
Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.
Project Title: Evaluation of Bicycle and Pedestrian Monitoring Equipment to Establish Collection Database and Methodologies for Estimating Non-Motorized Transportation
URL: http://tti.tamu.edu/documents/0-6927-P6.pdf

16. Abstract

This guide describes two different adjustment processes that can be used with pedestrian and bicyclist data. The first adjustment process is seasonal adjustment, which is applied to short-duration counts that are collected during a specific month of the year. Seasonal adjustment annualizes the short-duration counts, such that the resulting adjusted count value is a better estimate of the annual average daily traffic. This guide provides monthly adjustment factors for both pedestrian and bicyclist count data, which is recommended for use with all short-duration count data that include at least seven days of data.

The second adjustment process described in this guide is crowdsourced data scaling, which is applied to crowdsourced bicyclist data samples that are collected from GPS-enabled smartphones. Because the crowdsourced data represent only a sample of the total bicyclists, the number of samples must be scaled or expanded to estimate the total number of bicyclists. This guide describes a simple scaling process that estimates average annual daily bicyclists using the number of crowdsourced data samples, the functional class of the bicyclist travel facility, and the density of high-income households near the bicyclist travel facility.

17. Key Words
Pedestrian and Bicyclist Count Data, Seasonal Adjustment Factors, Crowdsourced Data Scaling

18. Distribution Statement
No restrictions. This document is available to the public through NTIS:
National Technical Information Service
Alexandria, Virginia
http://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	67	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

GUIDE FOR SEASONAL ADJUSTMENT AND CROWDSOURCED DATA SCALING

by

Bahar Dadashova
Associate Transportation Researcher
Texas A&M Transportation Institute

Greg Griffin
Assistant Research Scientist
Texas A&M Transportation Institute

Subasish Das
Associate Transportation Researcher
Texas A&M Transportation Institute

Shawn Turner
Senior Research Engineer
Texas A&M Transportation Institute

and

Madison Graham
Graduate Research Assistant
Texas A&M Transportation Institute

Product 0-6927-P6
Project 0-6927
Project Title: Evaluation of Bicycle and Pedestrian Monitoring Equipment to Establish Collection Database and Methodologies for Estimating Non-Motorized Transportation

Performed in cooperation with the
Texas Department of Transportation
and the
Federal Highway Administration

Resubmitted: September 2018

TEXAS A&M TRANSPORTATION INSTITUTE
College Station, Texas 77843-3135

DISCLAIMER

ACKNOWLEDGMENTS

This project was conducted in cooperation with TxDOT and FHWA. The authors thank the following members of the project monitoring committee:

- Project Manager: Chris Glancy, TxDOT.
- Lead: Teri Kaplan, TxDOT.
- Co-Lead: Bonnie Sherman, TxDOT.
- Mark Wooldridge, TxDOT.
- Bill Knowles, TxDOT.
- Greg Goldman, TxDOT.
- Darren McDaniel, TxDOT.
- Adam Chodkiewicz, TxDOT.
- Shelly Harris, TxDOT.
- Benjamin Miller, TxDOT.
- Michael Flaming, TxDOT.
- Ana Ramirez Huerta, TxDOT.
- Mahendran Thivakaran, TxDOT.
- Diane Dohm, Houston-Galveston Area Council.
- Kelly Porter, Capital Area Metropolitan Planning Organization.
- Karla Weaver, North Central Texas Council of Governments.
- Alexandria Carroll, Alamo Area Metropolitan Planning Organization.
- Jeffrey Pollack, Corpus Christi Metropolitan Planning Organization.
- Francis Reilly, City of Austin.
- Anita Holman, City of Houston.
- Carl Seifert, Jacobs Engineering.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

This guide describes two different adjustment processes that can be used with pedestrian and bicyclist data:

- Seasonal adjustment.
- Crowdsourced data scaling.

Both adjustment processes and their purpose are introduced in the following paragraphs.

SEASONAL ADJUSTMENT

The first adjustment process is seasonal adjustment, which is applied to short-duration counts that are collected during a specific month of the year. Seasonal adjustment annualizes the short-duration counts, such that the resulting adjusted count value is a better estimate of the annual average daily traffic (AADT). A similar adjustment process is also used for short-duration motor vehicle counts that are collected on specific days and specific months.

Chapter 2 in this guide provides monthly adjustment factors for both pedestrian and bicyclist count data, which is recommended for use with all short-duration count data that include at least seven days of data. The monthly adjustment factors are based continuous count data collected from 19 different permanent counter locations in Austin, Dallas, Houston, Plano, and San Antonio. Appendix A includes numerous charts that illustrate the month-of-year, time-of-day, and day-of-week traffic patterns at these permanent counter locations.

CROWDSOURCED DATA SCALING

The second adjustment process described in Chapter 3 is crowdsourced data scaling, which is applied to crowdsourced bicyclist data samples that are collected from GPS-enabled smartphones. Because the crowdsourced data represent only a sample of the total bicyclists, the number of samples must be scaled or expanded to estimate the total number of bicyclists.

Researchers developed the crowdsourced data scaling process by comparing crowdsourced data samples to actual total ground counts at 100 locations throughout Texas. The sample rates varied considerably among the locations, and explanatory variables were tested to determine what variables had the strongest influence on the crowdsourced data sample rate. Researchers then developed simplified equations that included the most influential variables.

Chapter 3 describes the resulting crowdsourced data scaling process that estimates average annual daily bicyclists (AADB) using the number of crowdsourced data samples, the functional class of the bicyclist travel facility, and the density of high-income households near the bicyclist travel facility.

CHAPTER 2. SEASONAL ADJUSTMENT FACTORS

Seasonal adjustment factors are used to process short-duration traffic counts to more accurately estimate AADT, one of the most common traffic count statistics. For example, if bicyclist counts are collected during a month when fewer bicyclists are riding, the collected bicyclist counts should be adjusted up to better represent annual average bicycling levels. Similarly, if pedestrian counts are collected during a month when more people are walking, these collected pedestrian counts should be adjusted down to better represent annual average walking levels. Traffic count analysts routinely use seasonal adjustment factors to annualize motor vehicle counts, as recommended in the Federal Highway Administration's Traffic Monitoring Guide (TMG) (FHWA 2016).

Researchers developed pedestrian and bicyclist seasonal adjustment factors using the methods outlined in the 2016 edition of the TMG. For non-motorized traffic, these methods are detailed in pages 4-25 through 4-32 (Section 4.4). The factor development methods for non-motorized traffic are very similar to those for motorized traffic detailed on pages 3-16 through 3-30 (Section 3.2.1). In general, the method is outlined as follows:

1. **Create a summary of traffic count patterns from continuous counters:** Develop month-of-year, day-of-week, and time-of-day summary charts.
2. **Identify distinct traffic patterns:** Examine charts to identify which continuous counters are most similar or dissimilar.
3. **Classify continuous counters into unique factor groups:** Combine continuous counter locations into unique factor groups.
4. **Calculate average adjustment factors from each factor group:** Calculate average adjustment factors that can be applied to short-duration counts.

In Step 1, researchers created numerous charts to display pedestrian and bicyclist count patterns separately by time-of-day, day-of-week, and month-of-year (see Appendix A). These charts were created for all permanent counters that had at least one full calendar year of complete and valid count data.

In Steps 2 and 3, researchers examined the pedestrian and bicyclist count patterns for each available count location, and classified each location into one of these factor groups as listed in the 2016 TMG:

- Commuter and work/school-based trips: typically have the highest peaks in the morning and evening.
- Recreation/utilitarian: may peak only once daily or be evenly distributed throughout the day.
- Mixed trip purposes (both commuter and recreation/utilitarian): has varying levels of these two different trip purposes, or may include other miscellaneous trip purposes.

TTI's preliminary analysis identified the following number of permanent counter locations in each factor group:

- Commuter and work/school-based trips: one location for pedestrians, one different location for bicyclists.
- Recreation/utilitarian: five locations for pedestrians, two locations for bicyclists.
- Mixed trip purposes: 14 locations for pedestrians, 10 locations for bicyclists.

Since the statewide pedestrian and bicyclist count database currently includes only short-duration counts of at least seven days (including one day of each day of the week), the seasonal adjustment would only need to account for the month of year and not the day of week. Therefore, researchers further analyzed the preliminary factor groups by examining the month-of-year patterns. In looking at these seasonal patterns, researchers concluded that the month-of-year patterns were quite similar, even among different factor groups. To simplify the seasonal adjustment process, TTI combined all analyzed permanent count locations in the three factor groups to create month-of-year count adjustment factors (Figure 1).

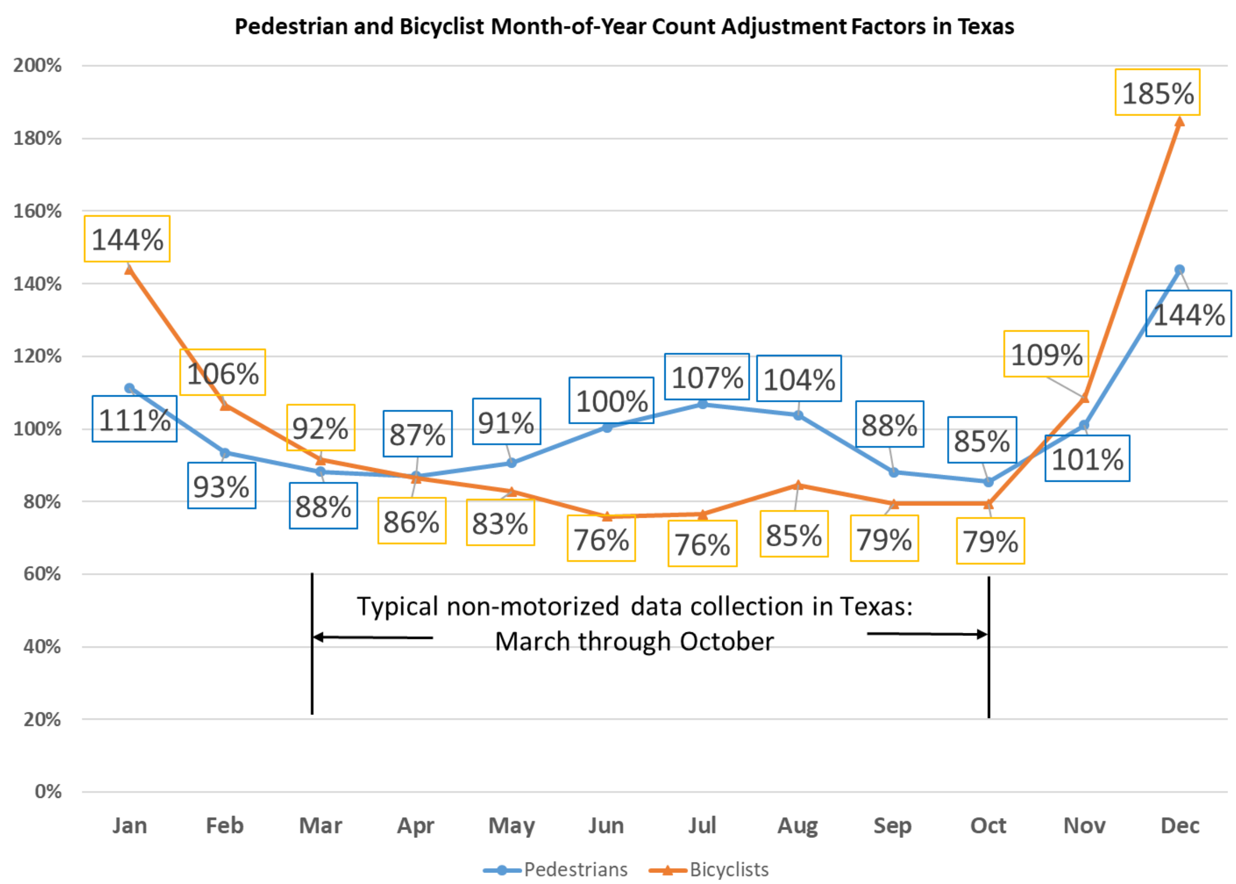


Figure 1. Month-of-Year Count Adjustment Factors for Short-Duration Counts.

To apply these adjustment factors, the seven-day average daily traffic (ADT) volume is multiplied by the factor corresponding to the travel mode and month of short-duration counts. For example, if a seven-day ADT in July for pedestrians is 100 persons, then the annualized ADT (or AADT) is 100×107 percent, or 107 pedestrians. Similarly, if a seven-day ADT in April for bicyclists is 50, then the AADT is 50×86 percent, or 43 bicyclists.

CHAPTER 3. CROWDSOURCED DATA SCALING

INTRODUCTION

Traffic volumes are fundamental for evaluating transportation systems, regardless of travel mode. A lack of counts for non-motorized modes poses a challenge for practitioners developing and managing multimodal transportation facilities, whether they want to evaluate transportation safety, potential need for infrastructure changes, or to answer other questions about how and where people bicycle and walk. This chapter shows how to take advantage of new data sources that provide a limited and biased sample of trips, and combine them with traditional counts to estimate bicycle travel volumes across most of the state of Texas.

Crowdsourcing is an approach to develop insights from a broad pool of individuals through an online platform that aggregates and formats the information for a specific use. In this case, bicycle travel volumes are crowdsourced through a smartphone-based app called Strava, used by bicyclists who want to record and compare their trips. The company aggregates these trips onto a transportation system network, processes them for privacy, and then re-sells the information as a crowdsourced traffic data product, available in many places around the globe. However, only a small portion of all bicyclists use the app, and this proportion varies across time and space. For instance, researchers found 3–9 percent of bicycle trips counted on trails in Austin used Strava at the time of the count (Griffin and Jiao 2015a), but this proportion varies in different contexts and over time (Jestico et al. 2016; Conrow et al. 2018).

Researchers developed a method to scale crowdsourced bicycle volumes by using limited on-ground count data and other factors, resulting in a relatively simple process to estimate bicycle travel using crowdsourced data, combined with the functional class of a network segment from Open Street Map data, and household income from American Community Survey data.

OVERVIEW: DEVELOPING FACTORS TO SCALE CROWDSOURCED BICYCLE VOLUMES

Researchers explored several different approaches to leverage crowdsourced data from Strava Metro to estimate bicycle volumes across the state, focusing on data that practitioners can regularly obtain outside of this research project, and implement their own estimates following this guide. Therefore, researchers limited the data used to Strava Metro's standard data product, the Texas Department of Transportation's (TxDOT's) Roadway Inventory, and American Community Survey data. Researchers also kept to standard statistical analysis methods, focusing on linear regression. The result is a relatively simple travel demand model, using crowdsourced bicycle volumes as a main input, along with functional classification of a transportation segment, and nearby high-income residential areas.

Researchers found that functional classification, or the type of roadway or trail segment, is a key factor for estimating total use with crowdsourced data. This makes sense because Strava is marketed toward a recreation/fitness-oriented user base, and researchers expected these users to more often choose off-street paths based on previous research (Griffin and Jiao 2015b). Therefore, researchers expected Strava to represent a relatively smaller proportion of users on urban arterial streets, where bicyclists may ride more often for work or shopping, rather than recreational trips logged using Strava. Researchers included functional classification (called CLAZZ in Open Street Map or FUN-SYS in TxDOT's Road-Highway Inventory Network [RHiNO] data) to characterize the type of infrastructure on a given segment in the models, and found the model using the Open Street Map classification used in the Strava Metro product had a lower mean absolute percentage error (29 percent versus 38 percent for RHiNO).

Income plays a role in the proportion of bicyclists logging trips on Strava, though it is less important than Strava counts or functional classification. Smartphones more available to higher-income users and the fitness-oriented nature of Strava users may further limit use to those with more disposable income and time (Leao et al. 2017). Preliminary model testing showed the number of households with income more than $200,000 a year was positively associated to the number of bicycle trips recorded on Strava.

Functional class of infrastructure, Strava trip counts, and household income form the basis of the model to estimate total bicycle trip volumes. Refer to project report 0-6927-R1 for additional description of the study methodology.

PRACTITIONER'S GUIDE TO ESTIMATING BICYCLE TRAFFIC WITH CROWDSOURCED DATA

This section describes how to estimate total bicycle traffic, by combining crowdsourced counts from Strava Metro with functional classification and nearby household income. To illustrate the process, this section includes data from the Walnut Creek Trail North of Jain Lane in Austin, Texas. The input data for the estimate includes the Strava activity count in both directions (22 and 23), the number of households nearby with more than $200,000 income (0), and the functional classification (Cycleway).

Step 1 – Record Average Annual Daily Strava Bicycle Counts

TxDOT has access to Strava Metro data starting in summer 2016, and later, subject to annual contract review, viewable on a web-based interface,[1] or with geospatial datasets for analysis in geographic information system (GIS) software. Review the desired Strava count duration, most commonly available as annual roll-ups, whether through Strava's viewer (e.g., http://metro-static.strava.com/dataView/TEXAS/201607_201706/RIDE/#5/31.215/-101.239), or the GIS data products in a desktop GIS software.

[1] July 2016–June 2017 Strava Metro data viewable at http://metro-static.strava.com/dataView/TEXAS/201607_201706/RIDE/#5/31.215/-101.239

Review counts on nearby links to check for accuracy problems. Previous research showed that Strava data "had some routes that were double- or triple-counted because of GPS assignment errors" (Wang et al. 2017). If adjacent segments inexplicably change volumes, use the volume that most closely matches the other nearby links.

If the data area already annual, divide by 365 to estimate average daily Strava bicycle traffic (AADB Strava). If monthly, divide by 30 or the actual number of days in the recorded month, if known. If weekly, divide the total by 7 to estimate daily traffic. Finally, round to the nearest integer. In this case, 8,213 Strava trips were found on our example segment of the Southern Walnut Creek Trail in Austin, resulting in an average annual daily bicyclist estimate of 23.

$$AADB\ Strava_{Walnut\ Creek} = \frac{Annual\ ActivityCount_{Walnut\ Creek}}{365} = 22.5 = 23$$

Step 2 – Identify Segment Functional Classification and Select Equation

Each of the seven functional classifications in Open Street Map has a different relationship to total use, given Strava counts and nearby households with annual income over $200,000.

<u>Functional Classification (CLAZZ in Strava Metro's network dat from Open Street Map)</u>

- Highway, primary (15) $AADB_i = 63 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Highway, secondary (21) $AADB_i = 13 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Highway, tertiary (31) $AADB_i = 22 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Highway, residential (32) $AADB_i = 17 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Highway, path (72) $AADB_i = 72 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Cycleway (81) $AADB_i = 62 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$
- Footway (91) $AADB_i = 28 \times (\exp(AADB\ Strava_i))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,i}))^{0.002}$

Since the Walnut Creek example is a cycleway, researchers choose the following equation for plugging in the other values in Microsoft Excel software:

$$AADB_{Walnut\ Creek} = 62 \times (\exp(AADB\ Strava_{Walnut\ Creek}))^{0.038}(\exp(\text{Household} >\ 200\text{K}_{\,Walnut\ Creek}))^{0.002}$$

Step 3 – Plug in Values to Excel

Insert the daily count of Strava trips (23), and the number of high-income households (0), and the equation becomes:

$$AADB_{Walnut\ Creek} = 62 \times (\exp(23))^{0.038}(\exp(0))^{0.002}$$

To write this equation in Excel, enter the following in a spreadsheet cell:

=62*(EXP(23)^0.038)*(EXP(0)^0.002)

Average Annual Daily Bicyclist traffic at Walnut Creek = 149

The results show that the predicted number of bicycles on this segment is equal to 149. Note that the observed counts are 152, showing this prediction is highly accurate (Appendix B).

Step 4 – Review Results

Finally, review these results against local knowledge and reasonableness. There are several reasons why this model might over-or-under predict bicycle traffic. Strava use itself may be particularly high or low in a certain area. It might over-estimate such if a major event was routed there during the Strava sampling period; or under-estimate if Strava use is particularly low. Researchers expect higher fluctuations in rural areas with lower overall Strava use, as compared with urban areas.

Changes in segment classification over time, such as upgrading a street from a tertiary to secondary segment, could significantly impact bicycle traffic estimation values. Similarly, any errors in the classification will expand error of the traffic estimate. High-income households have a relatively minor, yet statistically significant, role in scaling Strava traffic to estimate totals. However, there may be areas that do not respond to residential income in an average manner, such as bicycling loops in large parks. Use of the route in the park may be rather homogenous, but nearby residential income could skew traffic estimates when they do not, in practice, impact bicycling rates.

This traffic estimation technique is designed to work even with zero Strava bicycle trip counts, by using minimal values observed with manual counts throughout the state. Table 1 can be used to review against estimates with low Strava sample counts.

Table 1. Estimated Number of Bicycle Counts Given Strava Sample and Roadway Class.

Strava Sample Counts	Highway, primary (15)	Highway, secondary (21)	Highway, tertiary (31)	Highway, residential (32)	Highway, path (72)	Cycleway (81)	Footway (91)
0	63	13	22	17	72	63	28
5	76	16	26	21	87	76	34
10	92	19	32	26	105	92	41
20	134	29	46	37	153	135	59

REFERENCES

Conrow, L., E. Wentz, T. Nelson, and C. Pettit. Comparing Spatial Patterns of Crowdsourced and Conventional Bicycling Datasets. *Applied Geography*, Vol. 92, No. December 2017, 2018, pp. 21–30. https://doi.org/10.1016/j.apgeog.2018.01.009.

Federal Highway Administration. *Traffic Monitoring Guide*. Online. November 2016. Available at https://www.fhwa.dot.gov/policyinformation/tmguide/.

Griffin, G. P., and J. Jiao. "Crowdsourcing Bicycle Volumes: Exploring the Role of Volunteered Geographic Information and Established Monitoring Methods." *URISA Journal*, Vol. 27, No. 1, 2015a, pp. 57–66.

Griffin, G. P., and J. Jiao. "Where Does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus." *Journal of Transport & Health*, Vol. 2, No. 2, 2015b, pp. 238–247. https://doi.org/10.1016/j.jth.2014.12.001.

Jestico, B., T. Nelson, and M. Winters. "Mapping Ridership Using Crowdsourced Cycling Data." *Journal of Transport Geography*, Vol. 52, 2016, pp. 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006.

Leao, S., S. Lieske, L. Conrow, J. Doig, V. Mann, and C. Pettit. "Building a National-Longitudinal Geospatial Bicycling Data Collection from Crowdsourcing." *Urban Science*, Vol. 1, No. 3, 2017, p. 23. https://doi.org/10.3390/urbansci1030023.

Wang, H., C. Chen, Y. Wang, Z. Pu, and M. B. Lowry. *Bicycle Safety Analysis: Crowdsourcing Bicycle Travel Data to Estimate Risk Exposure and Create Safety Performance Functions*. Seattle, WA, Pacific Northwest Transportation Consortium, 2017.

APPENDIX A. PEDESTRIAN AND BICYCLIST TRAFFIC PATTERNS AT PERMANENT COUNTER LOCATIONS IN TEXAS

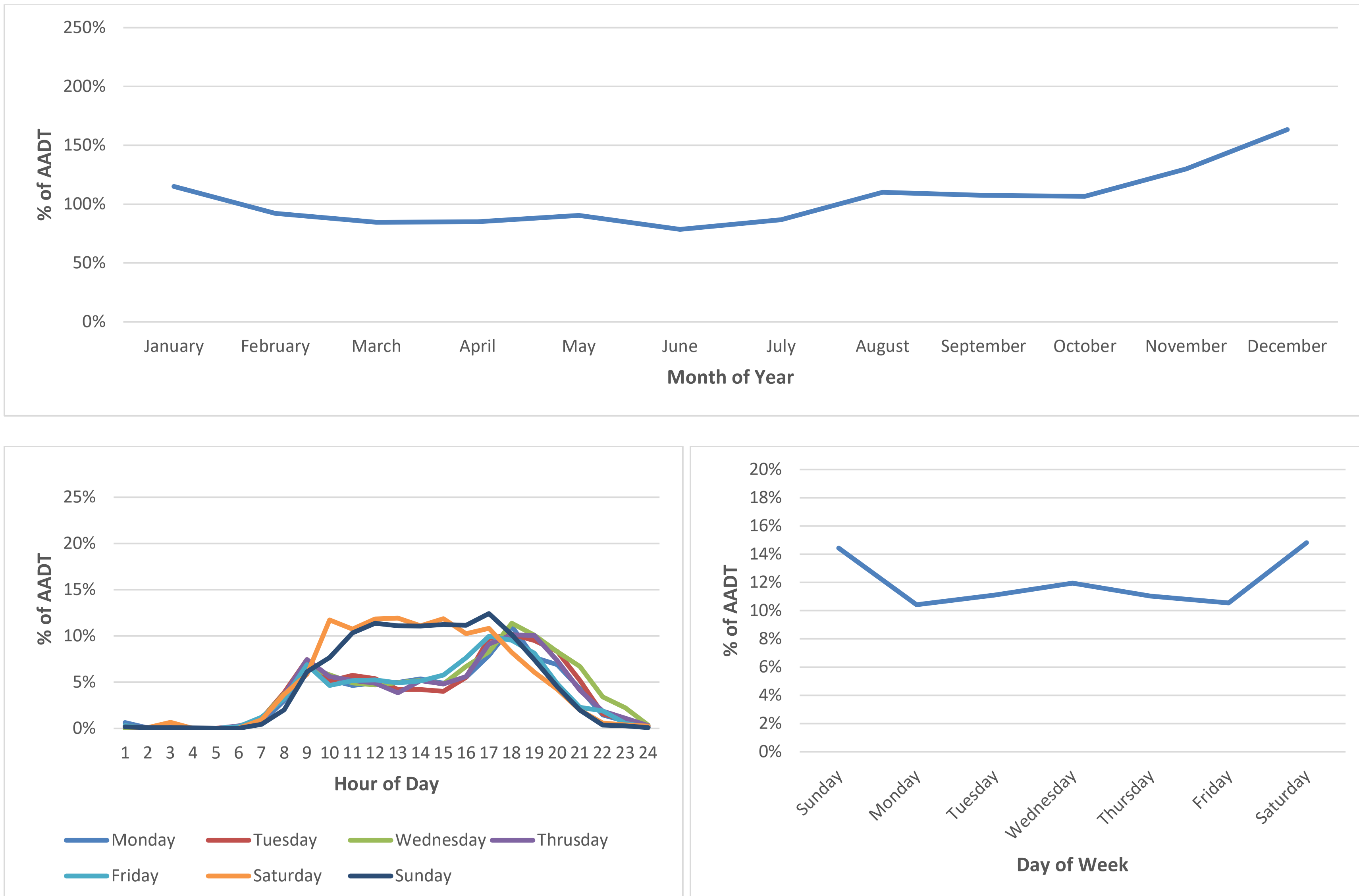


Figure 2. Bicyclist Traffic Count Patterns: Johnson Creek Trail @ MoPac/W 5th St./W 6th St. Interchange (Austin).

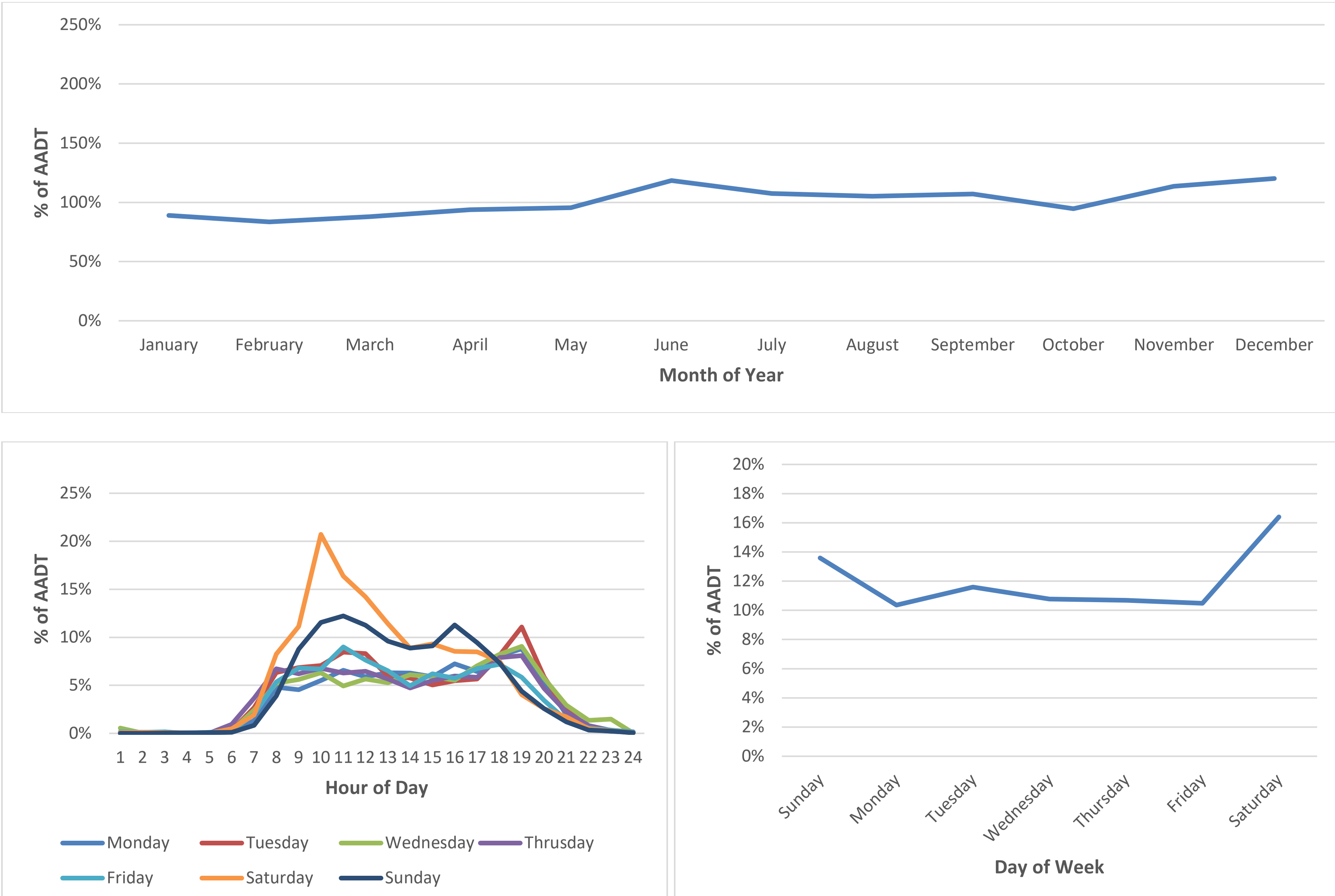


Figure 3. Pedestrian Traffic Count Patterns: Johnson Creek Trail @ MoPac/W 5th St./W 6th St. Interchange (Austin).

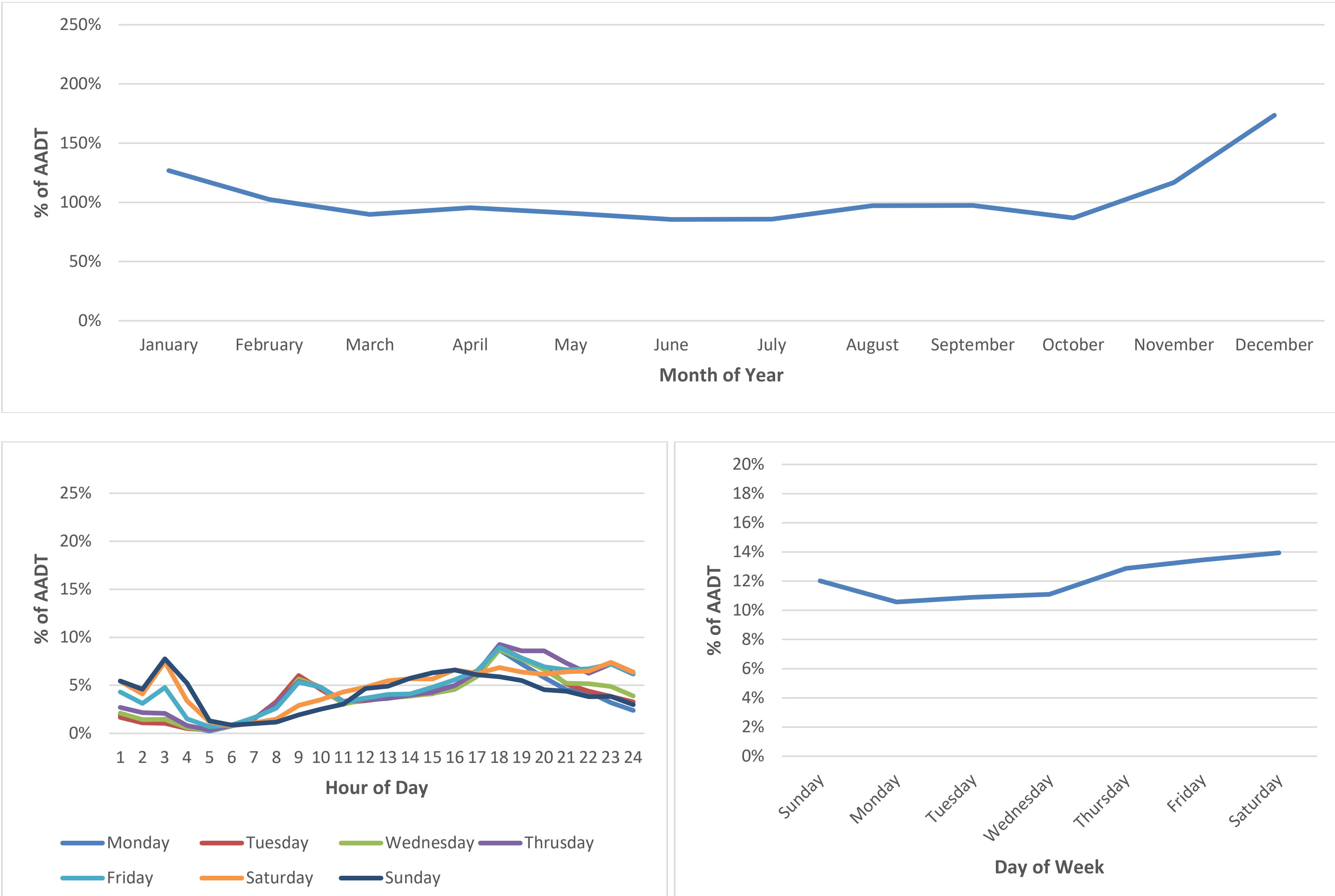


Figure 4. Bicyclist Traffic Count Patterns: Lance Armstrong Bikeway @ Waller Creek (Austin).

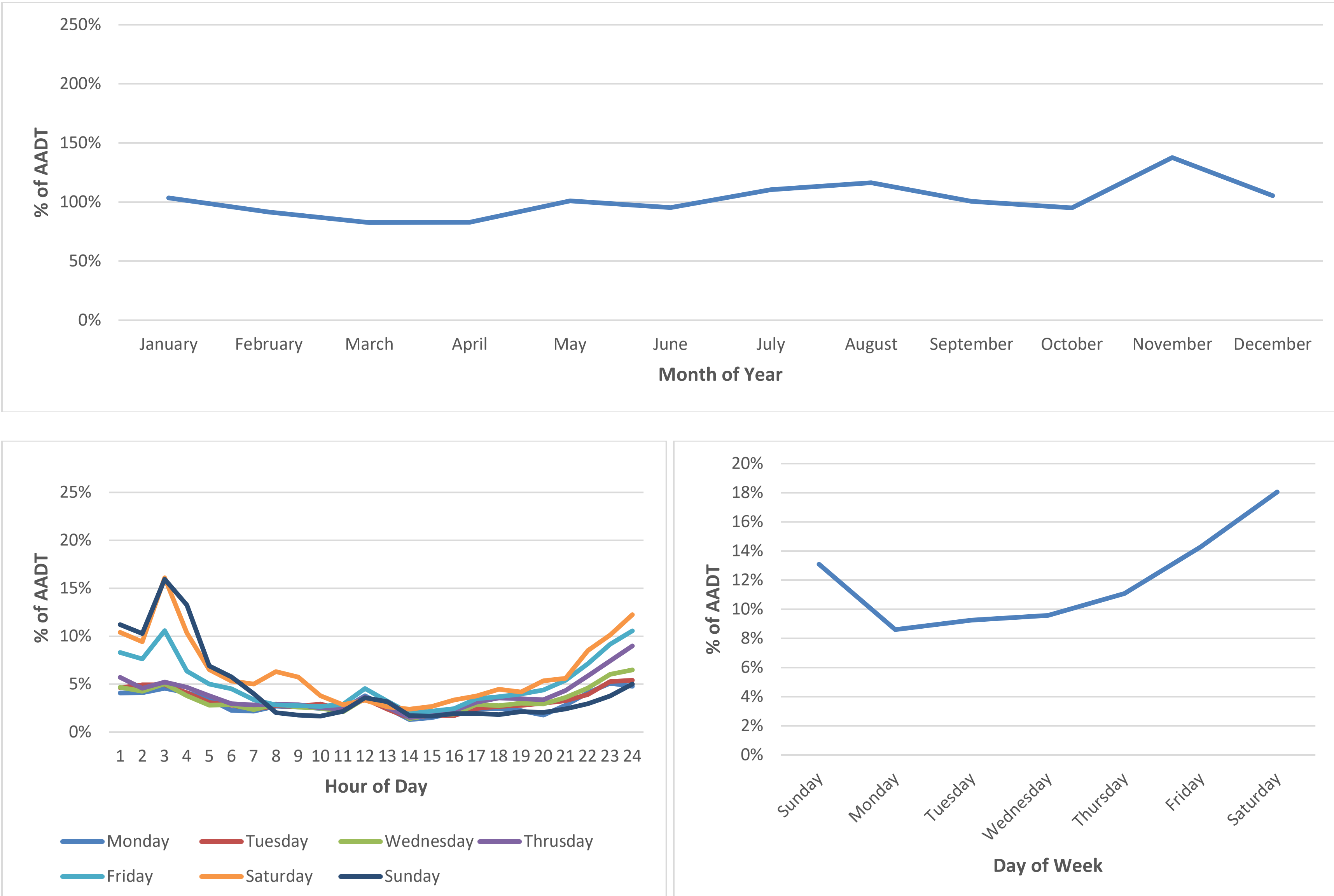


Figure 5. Pedestrian Traffic Count Patterns: Lance Armstrong Bikeway @ Waller Creek (Austin).

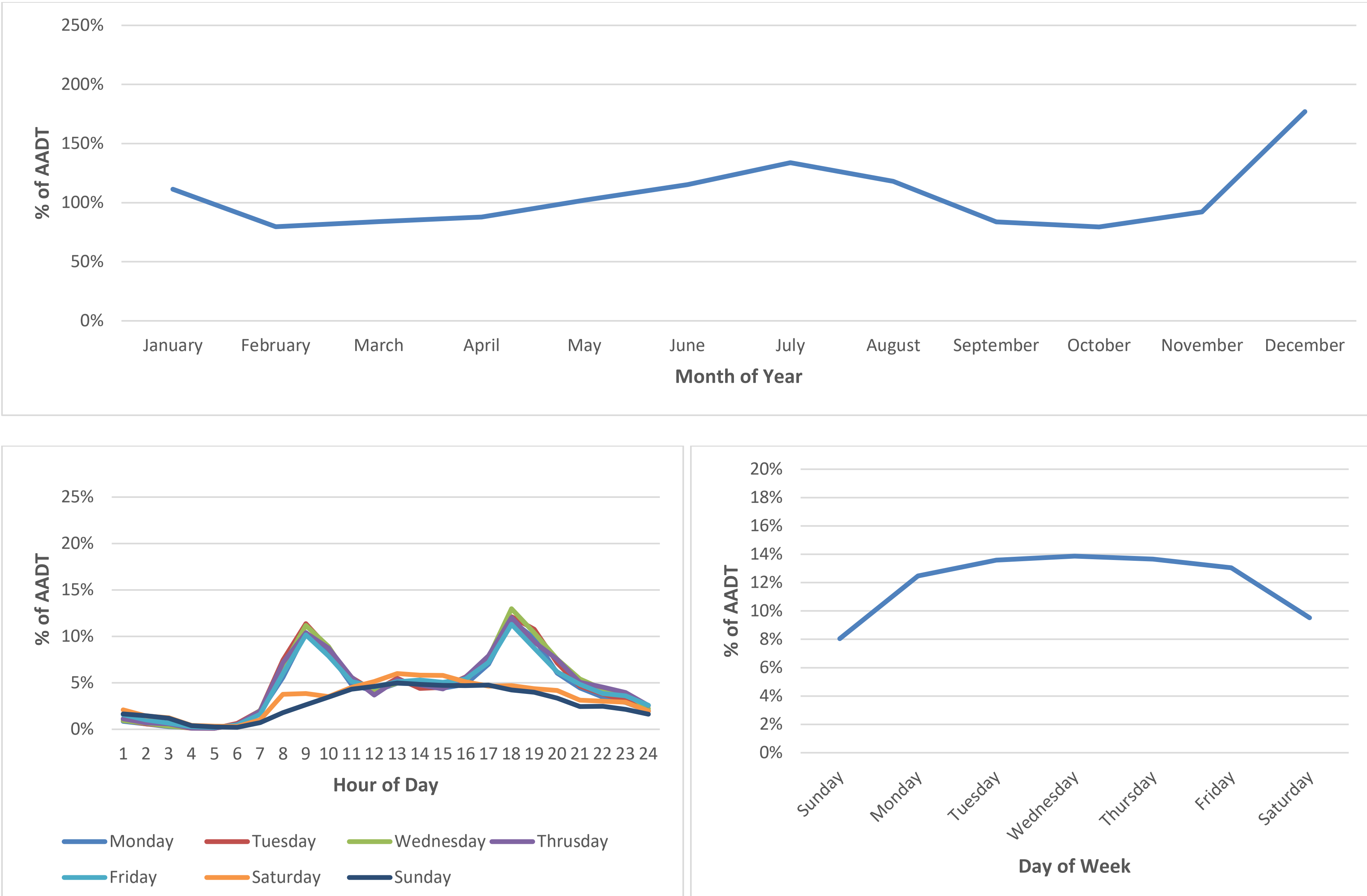


Figure 6. Bicyclist Traffic Count Patterns: Manor Rd. @ Alamo St. (Austin).

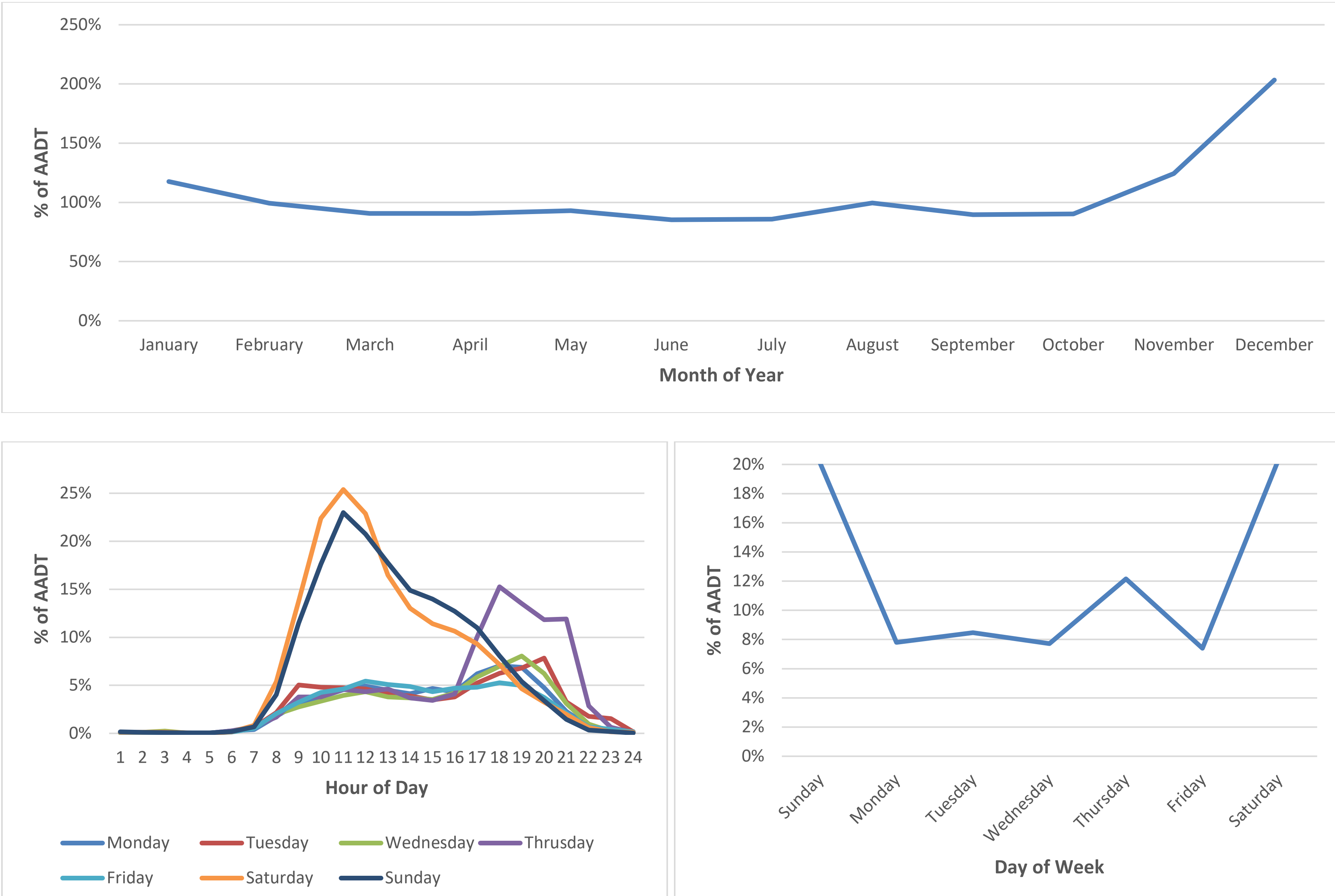
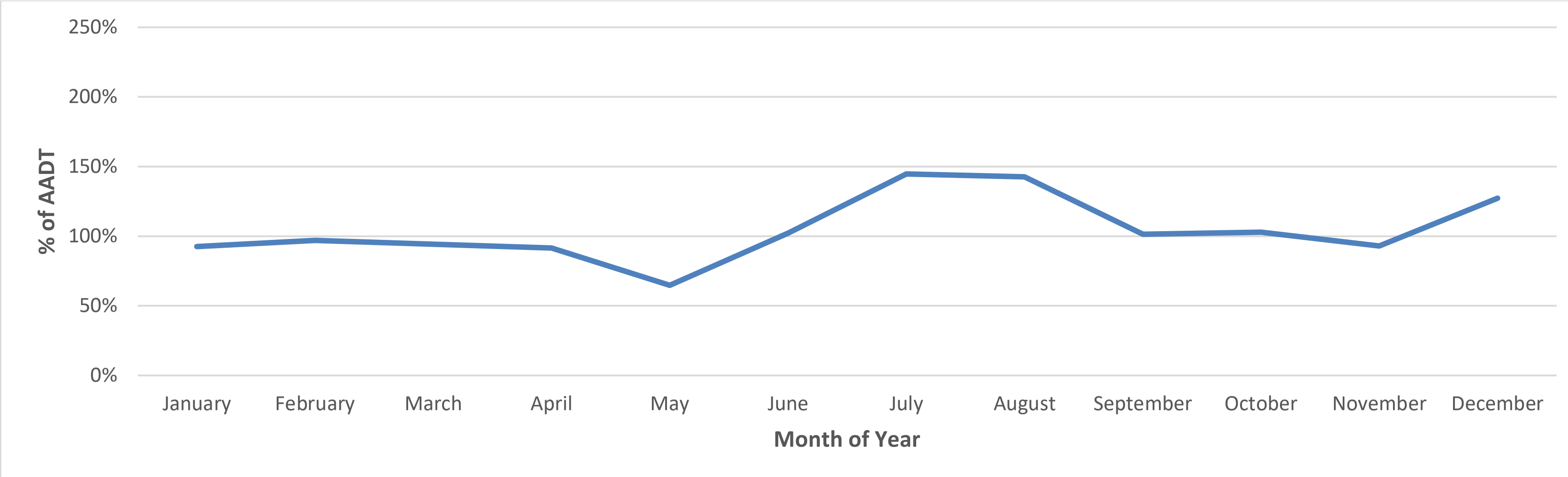
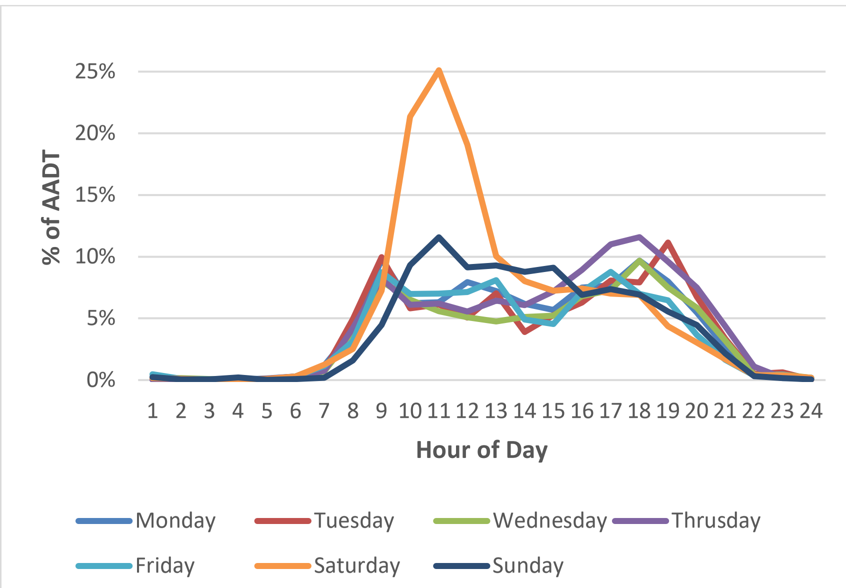


Figure 7. Bicyclist Traffic Count Patterns: Walnut Creek Trail N of Jain Ln. (Austin).





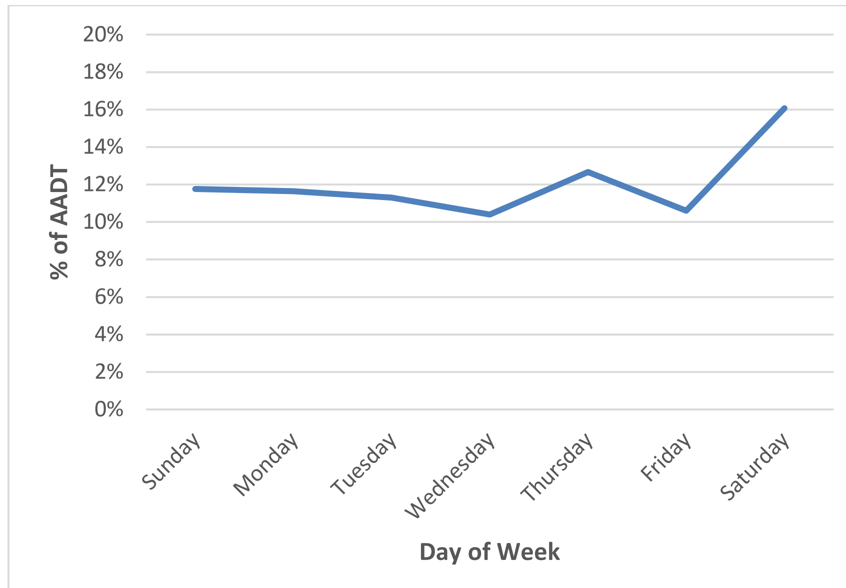
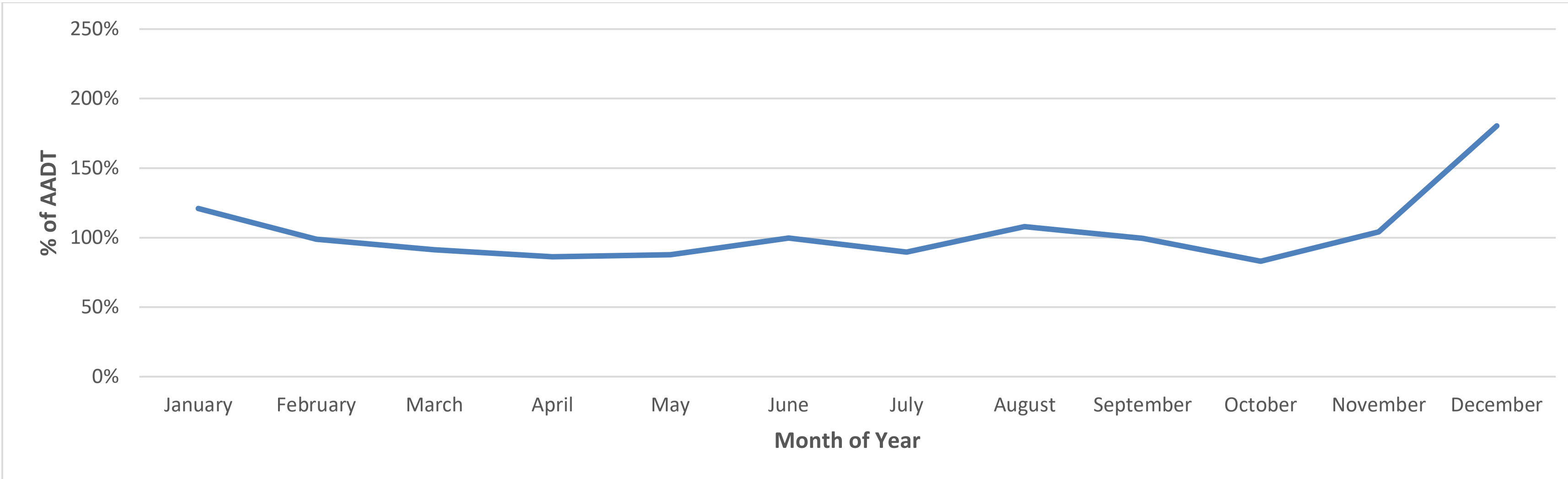
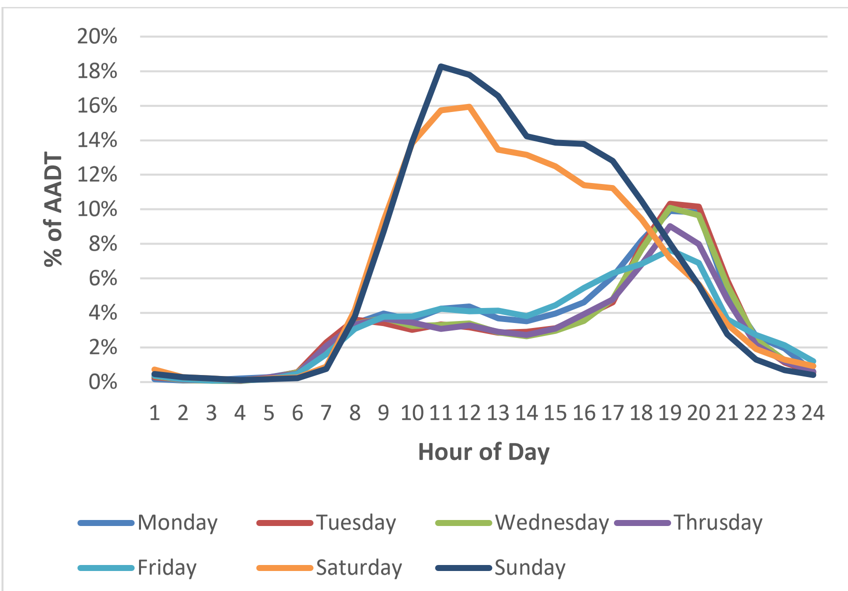


Figure 8. Pedestrian Traffic Count Patterns: Walnut Creek Trail N. of Jain Ln. (Austin).





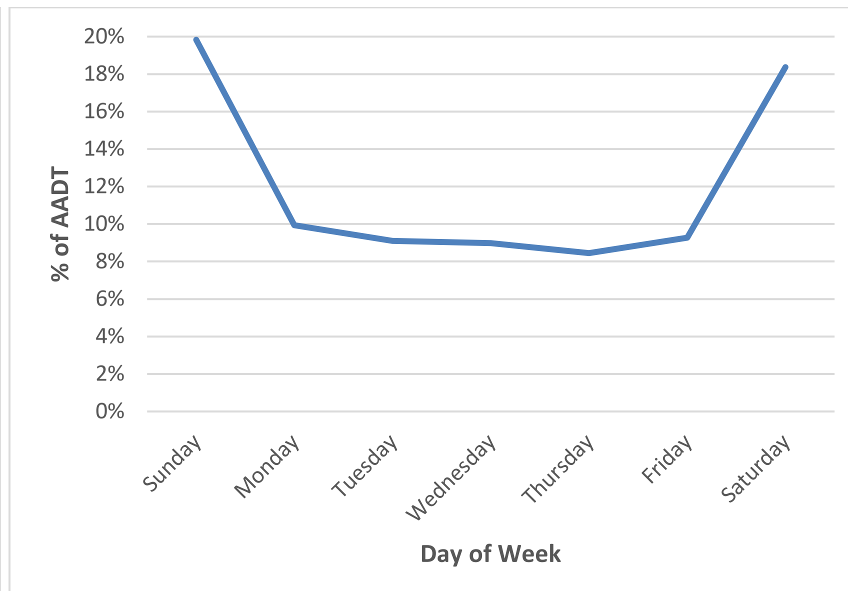


Figure 9. Bicyclist Traffic Count Patterns: Heights Trail @ 5 ½ Street (Houston).

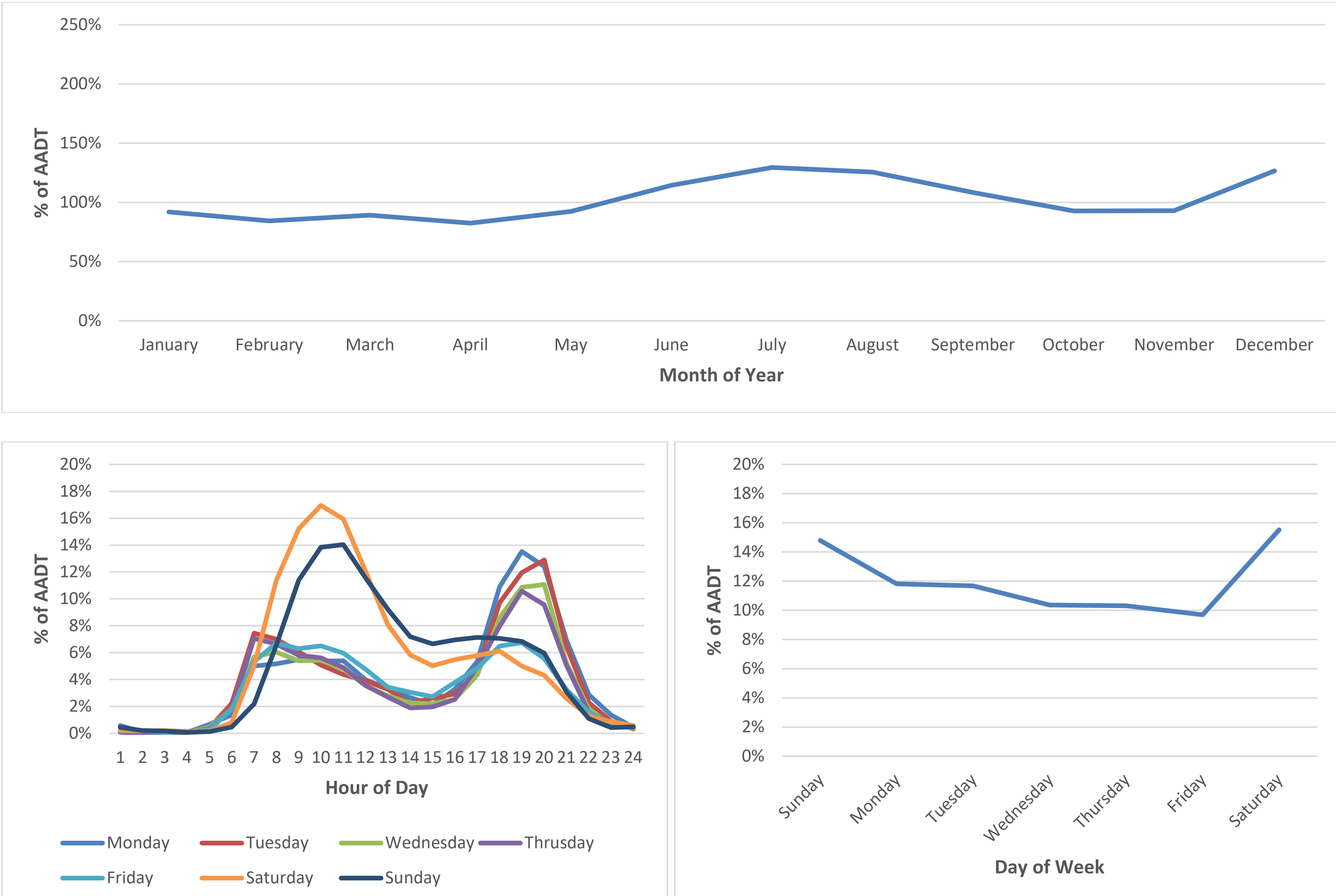


Figure 10. Pedestrian Traffic Count Patterns: Heights Trail @ 5 ½ Street (Houston).

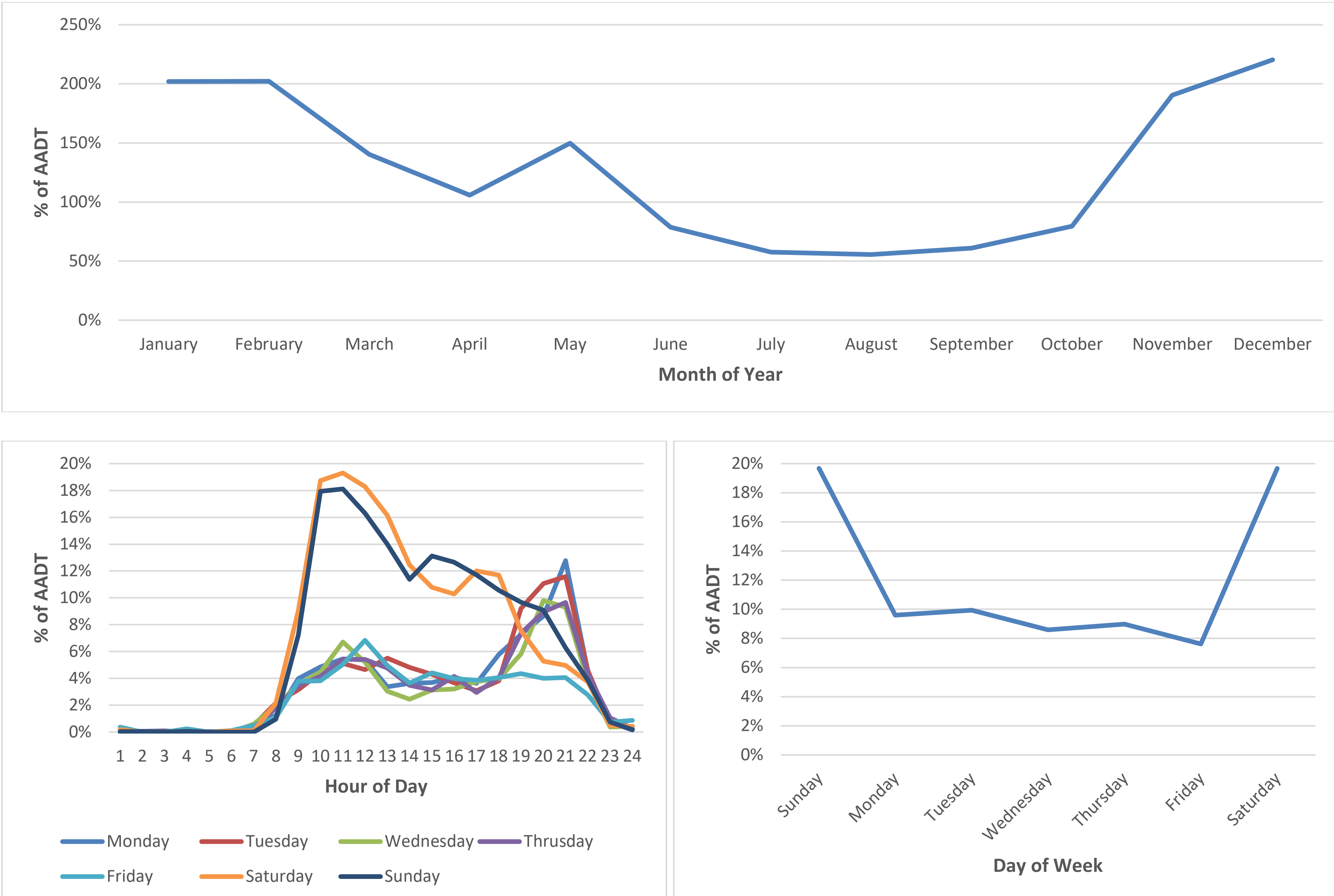


Figure 11. Bicyclist Traffic Count Patterns: Bluebonnet Trail at US 75 (Plano).

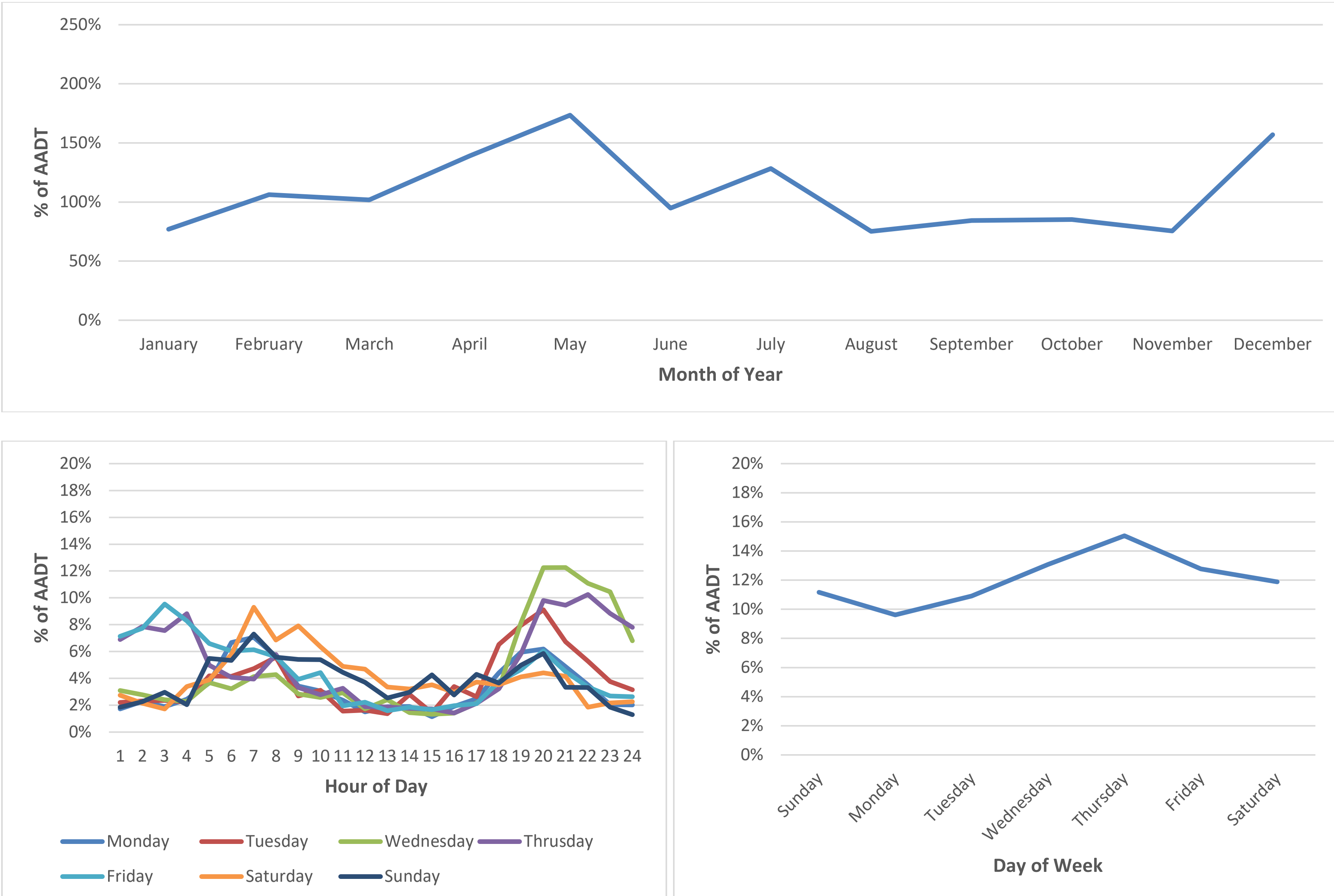


Figure 12. Pedestrian Traffic Count Patterns: Bluebonnet Trail at US 75 (Plano).

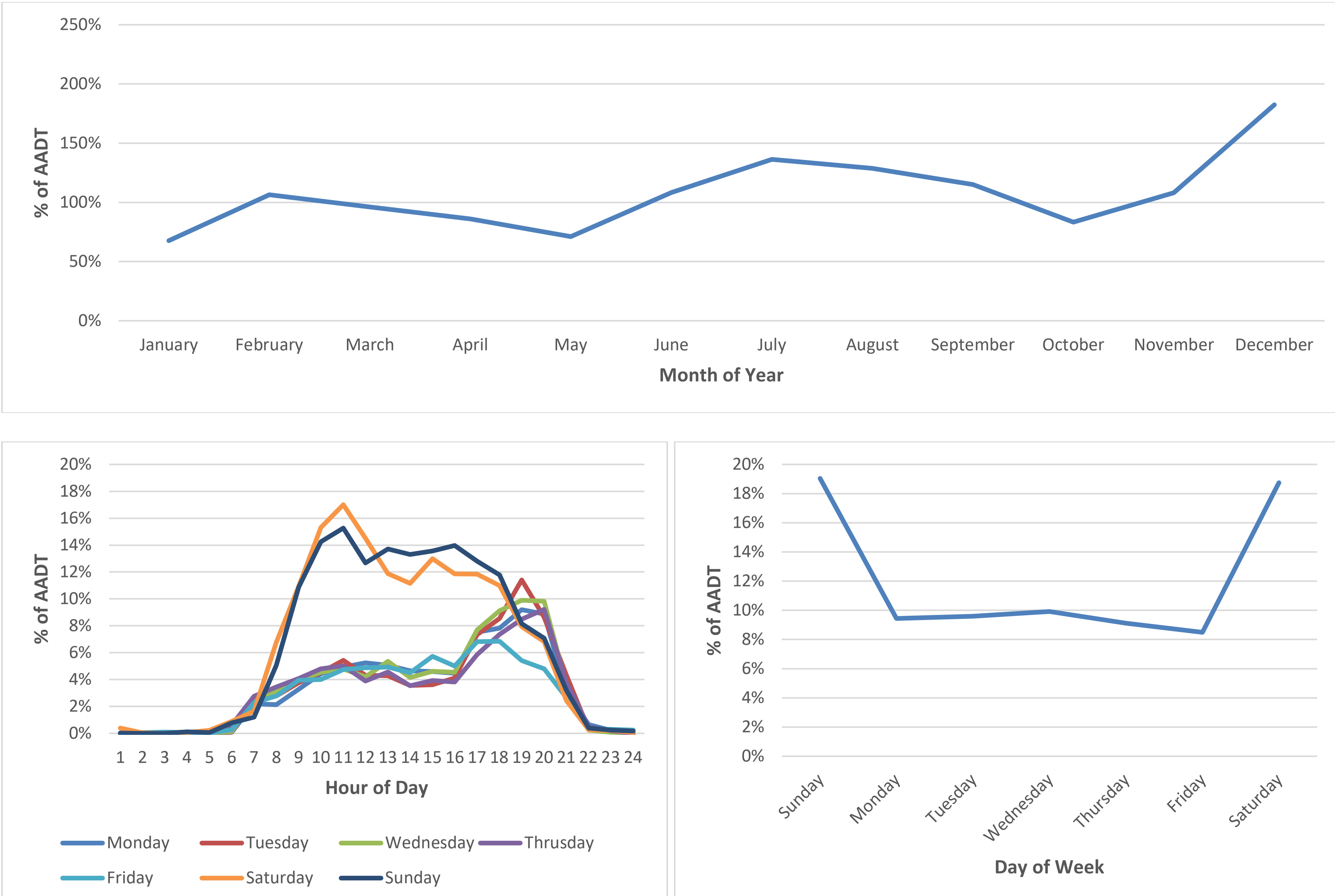


Figure 13. Bicyclist Traffic Count Patterns: OPP and NP Trail (Plano).

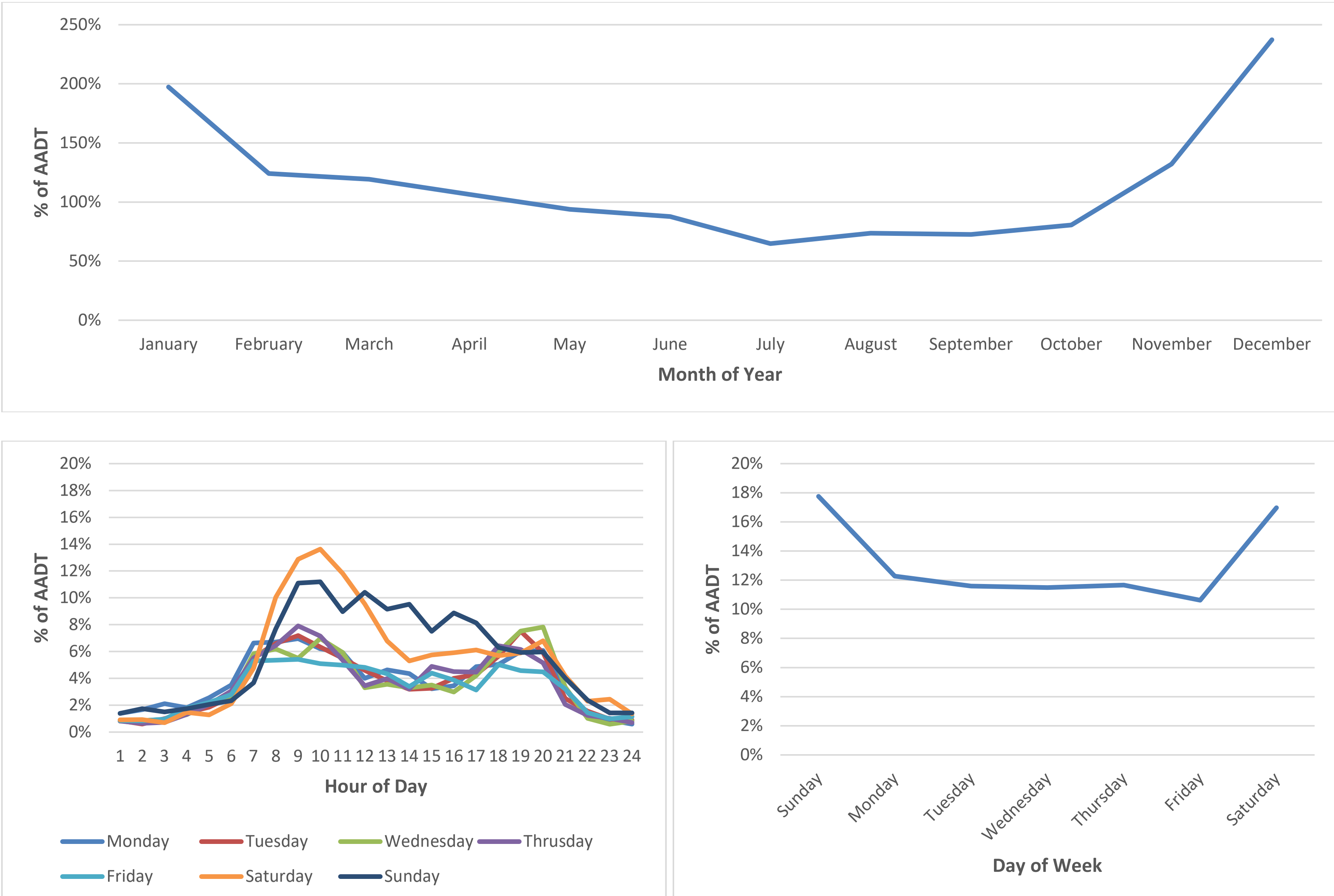


Figure 14. Pedestrian Traffic Count Patterns: OPP and NP Trail (Plano).

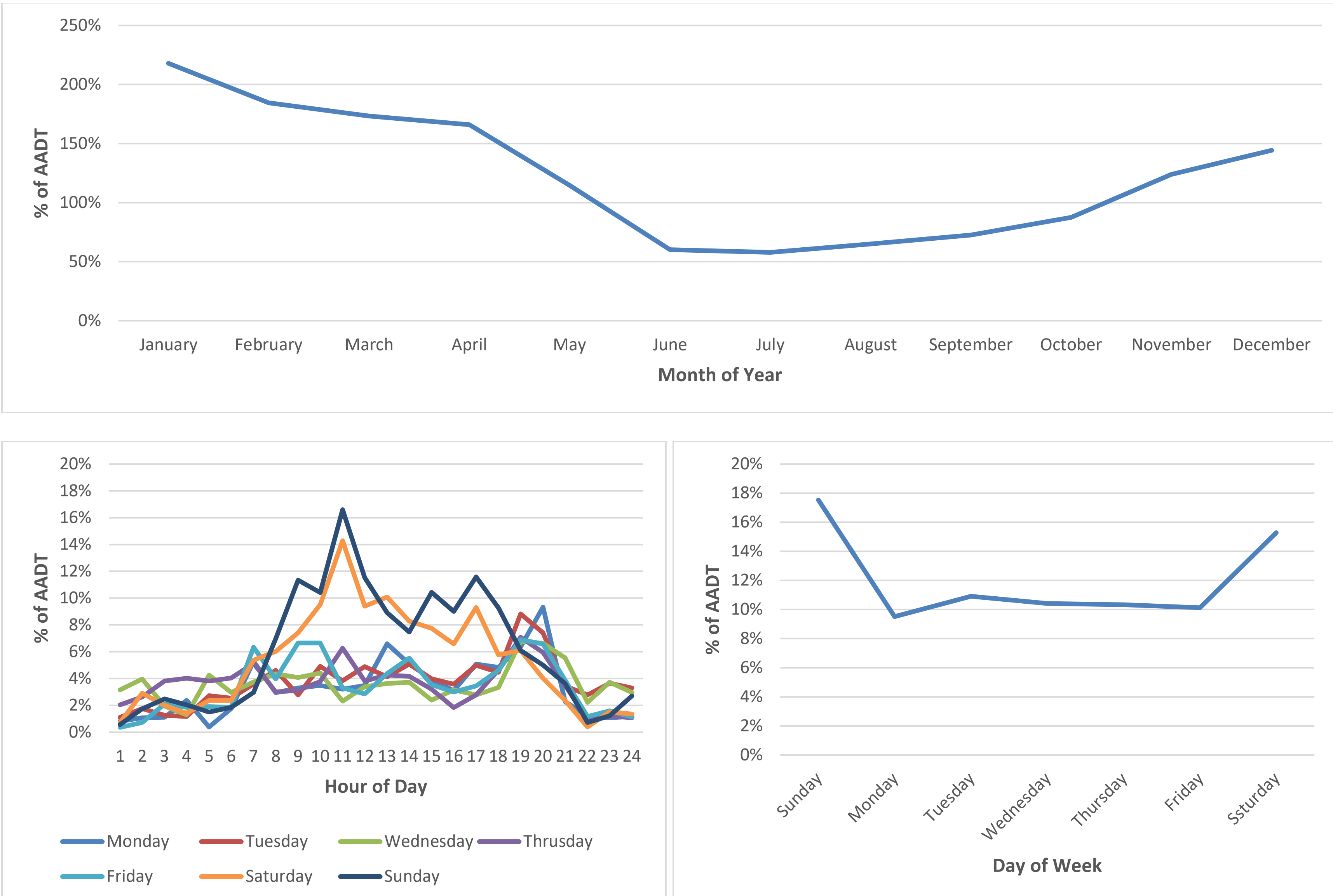


Figure 15. Bicyclist Traffic Count Patterns: Legacy Trail (Plano).

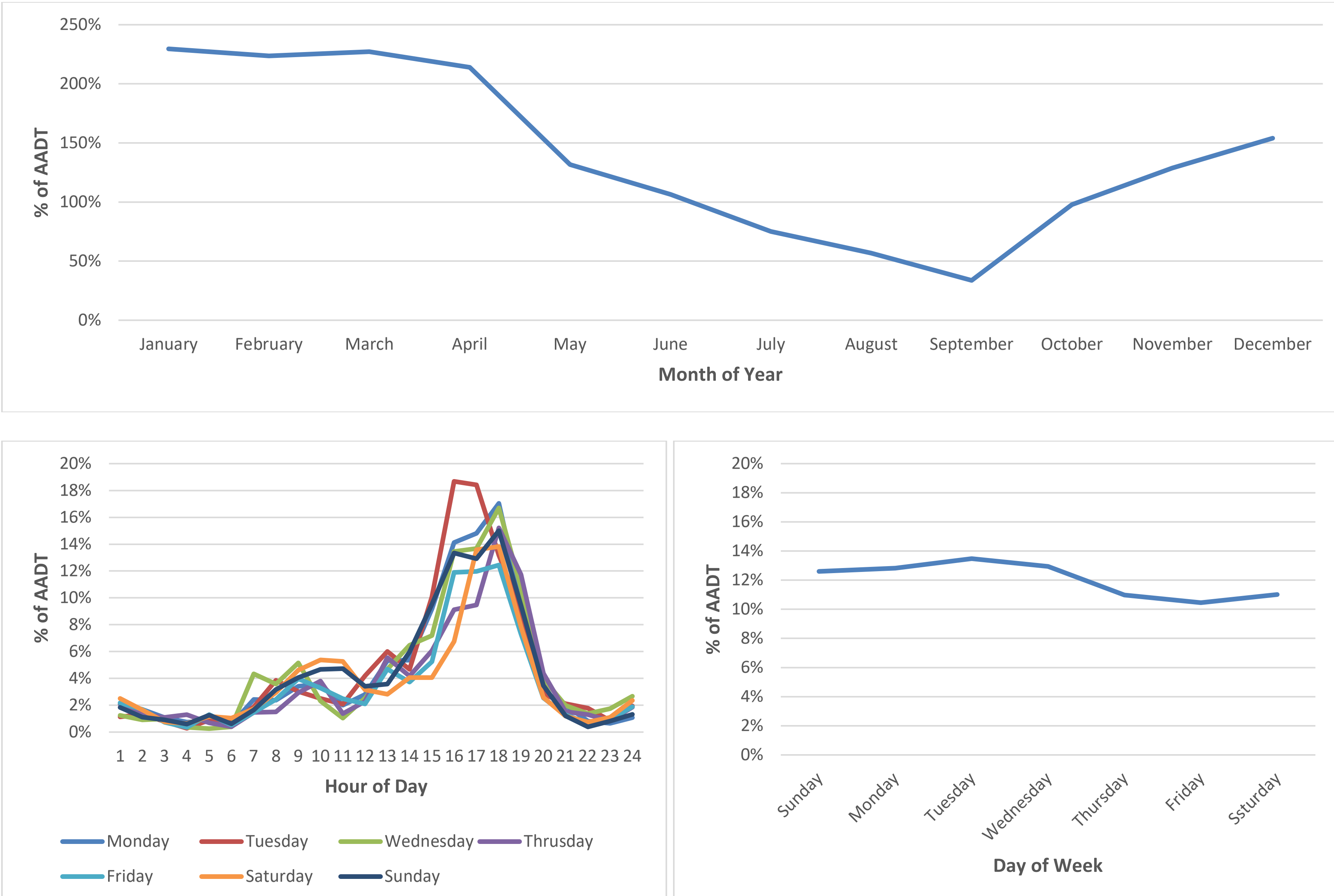


Figure 16. Pedestrian Traffic Count Patterns: Legacy Trail (Plano).

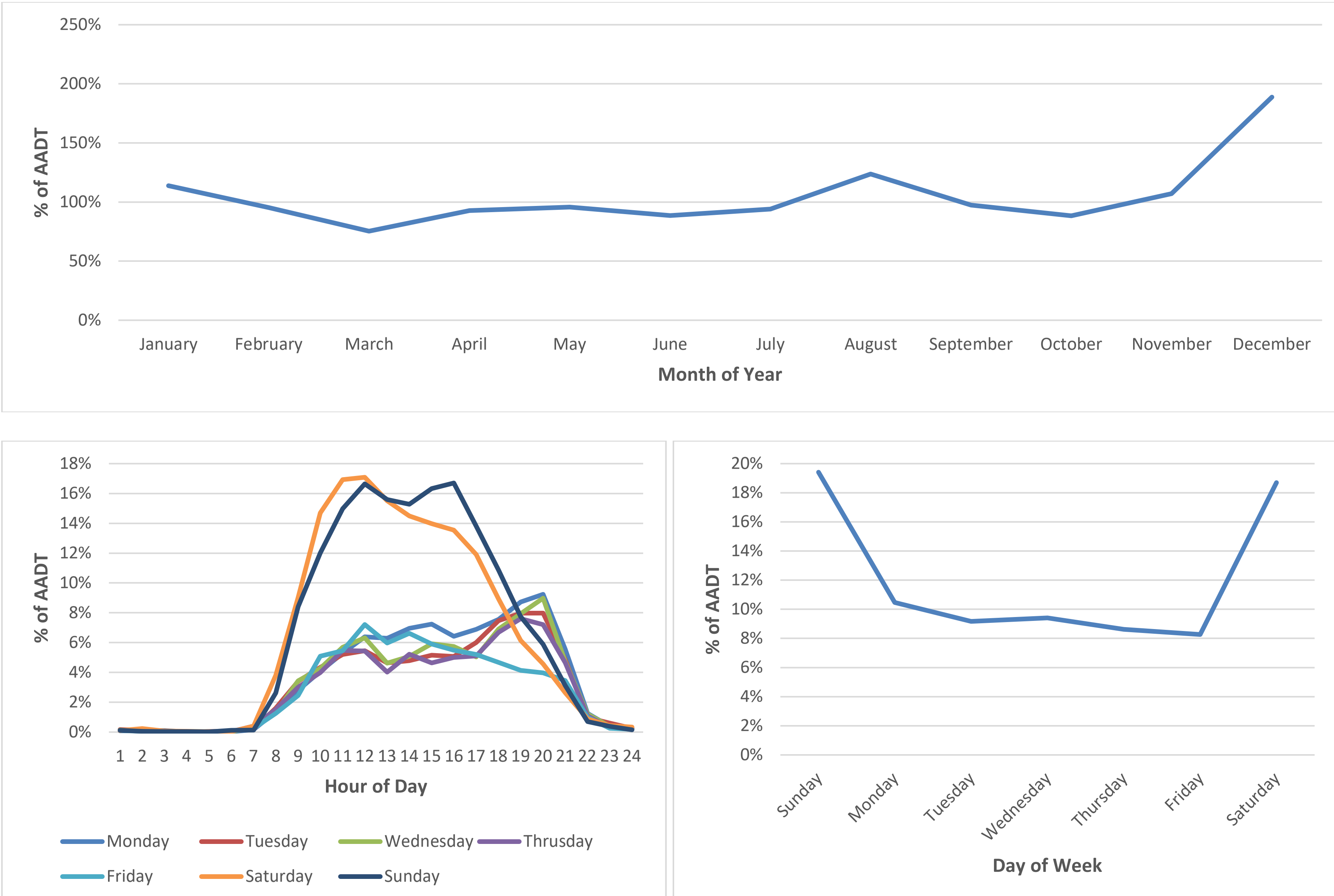


Figure 17. Bicyclist Traffic Count Patterns: Mission Reach Trail South of Theo Ave. (San Antonio).

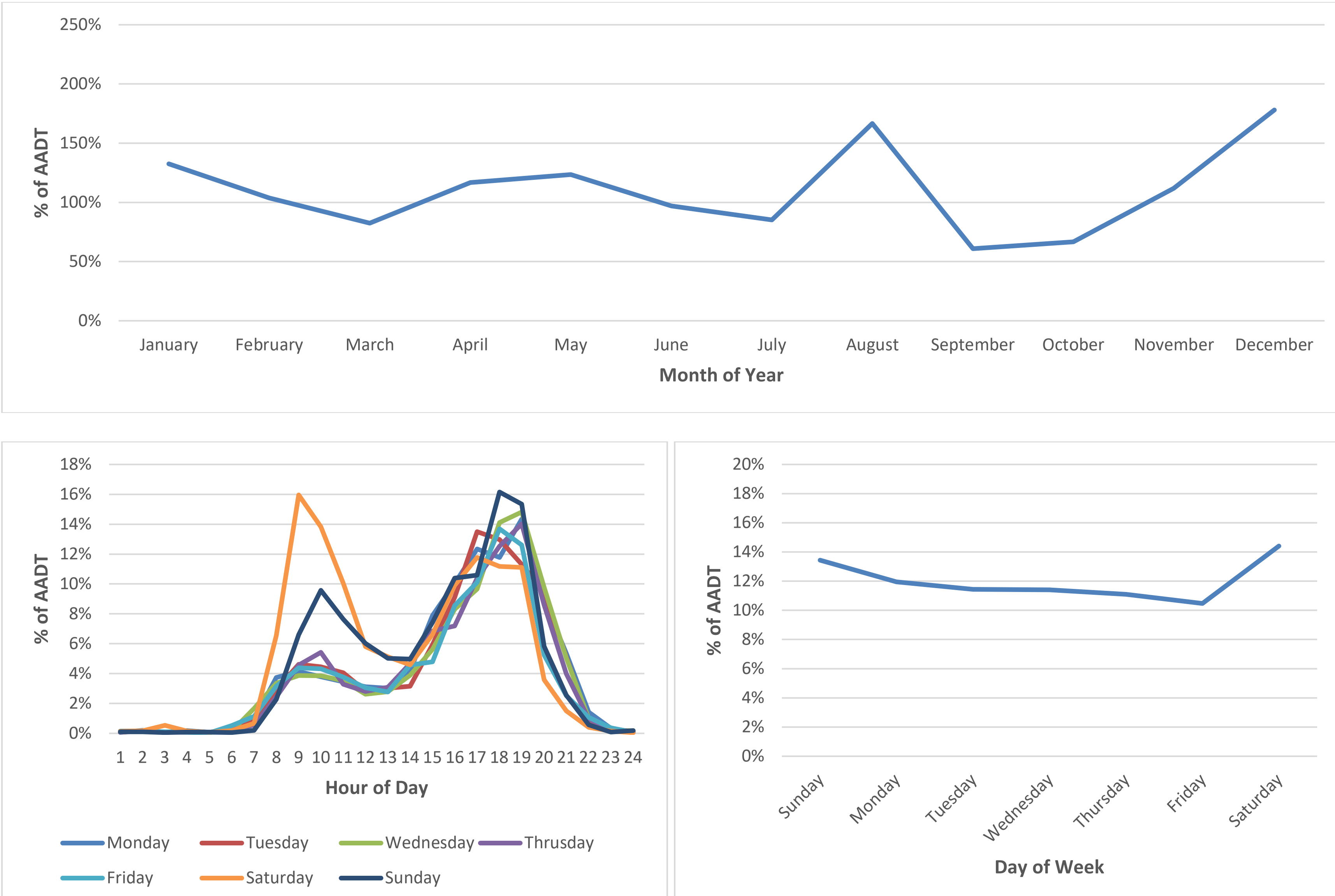
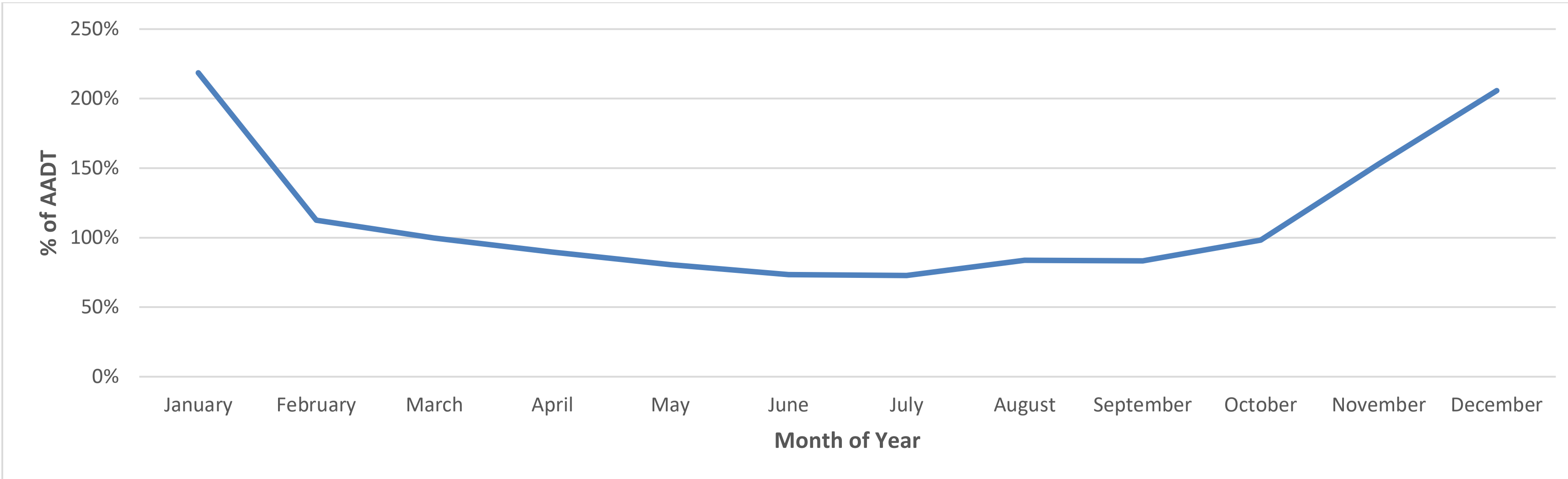
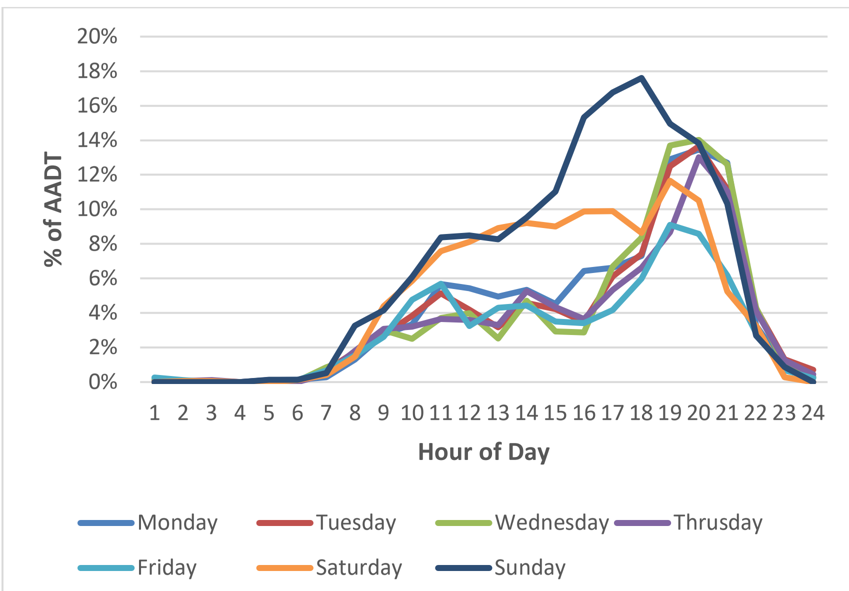


Figure 18. Pedestrian Traffic Count Patterns: Mission Reach Trail South of Theo Ave. (San Antonio).





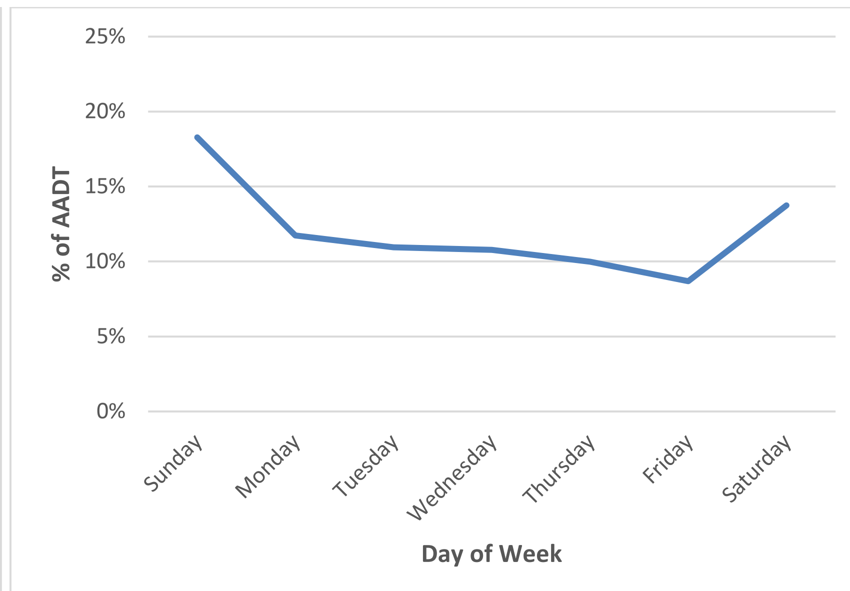


Figure 19. Bicyclist Traffic Count Patterns: Bachman Lake/W North West Highway (Dallas).

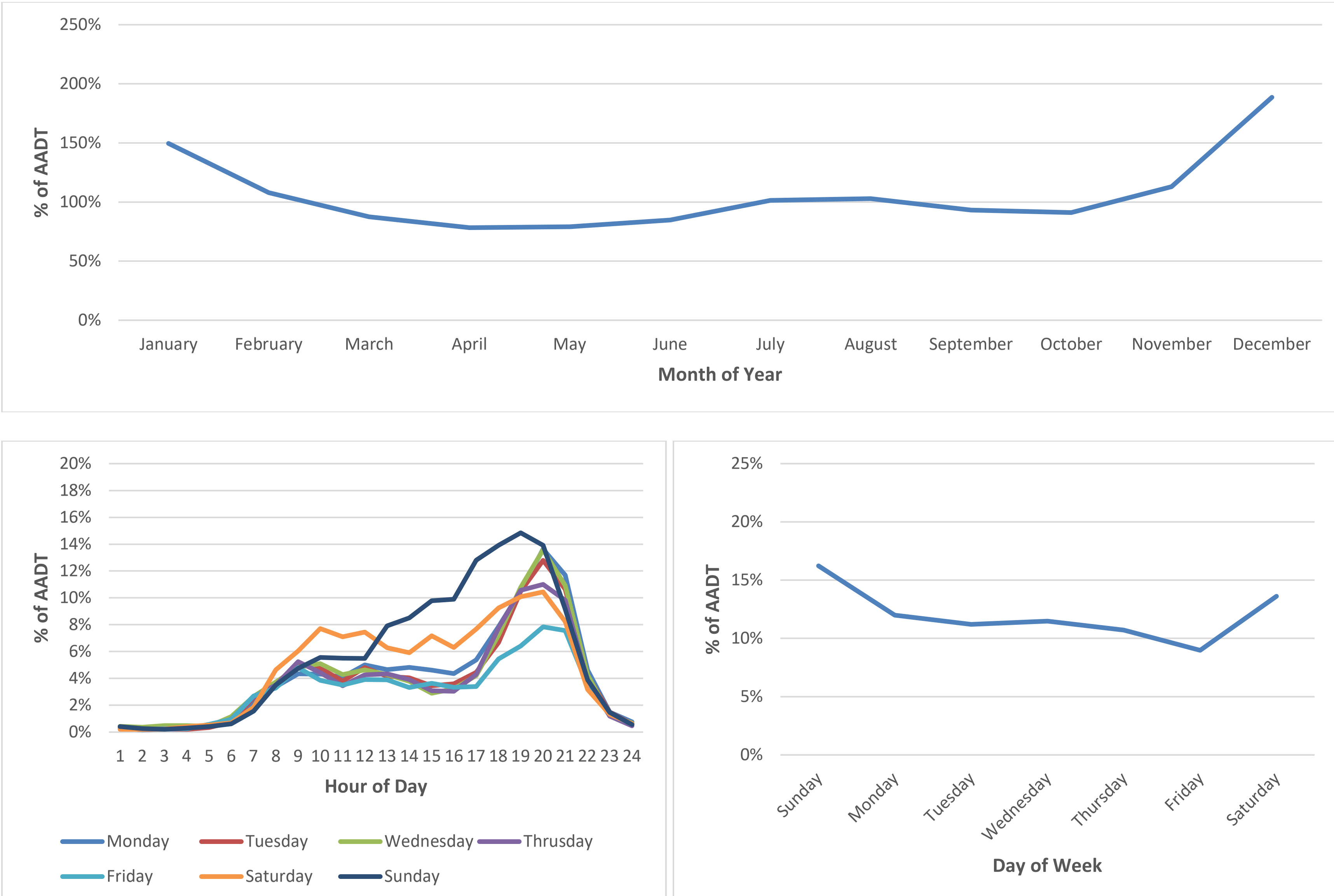
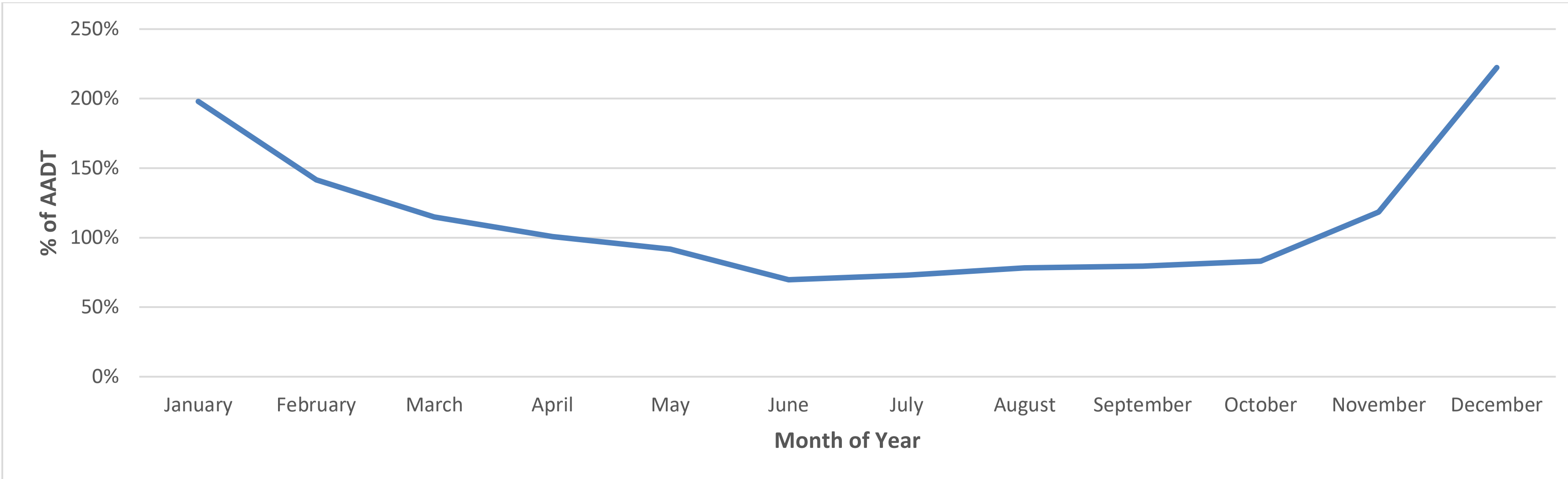
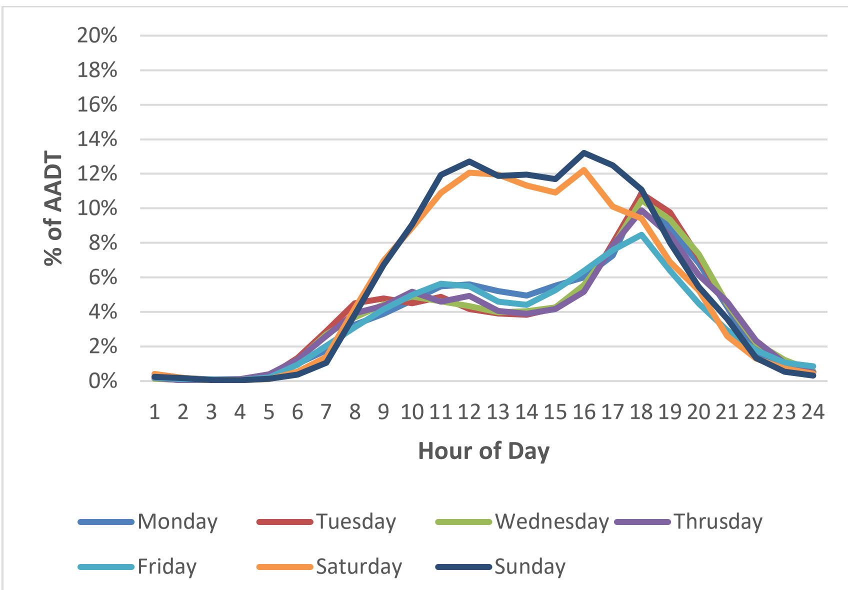


Figure 20. Pedestrian Traffic Count Patterns: Bachman Lake/W North West Highway (Dallas).





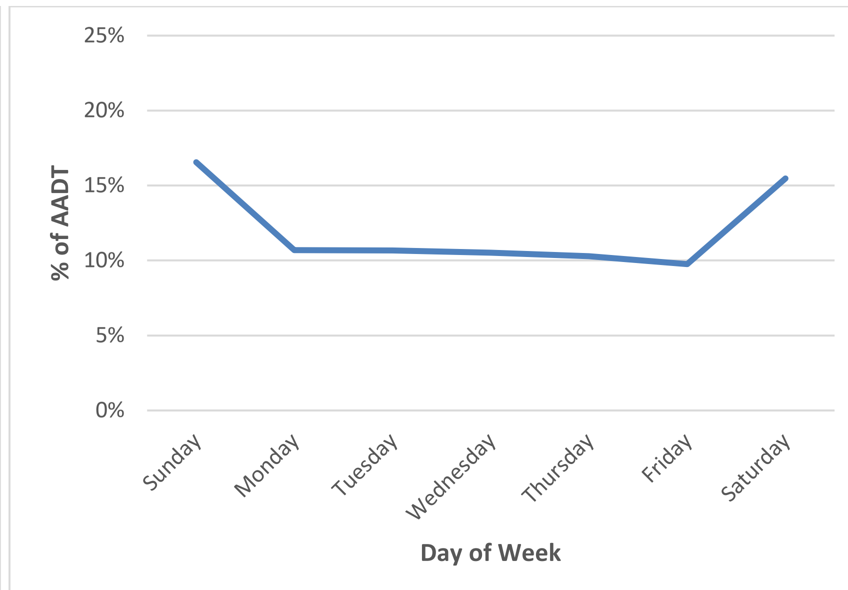


Figure 21. Bicyclist Traffic Count Patterns: Katy Trail at Cedar Springs Rd. (Dallas).

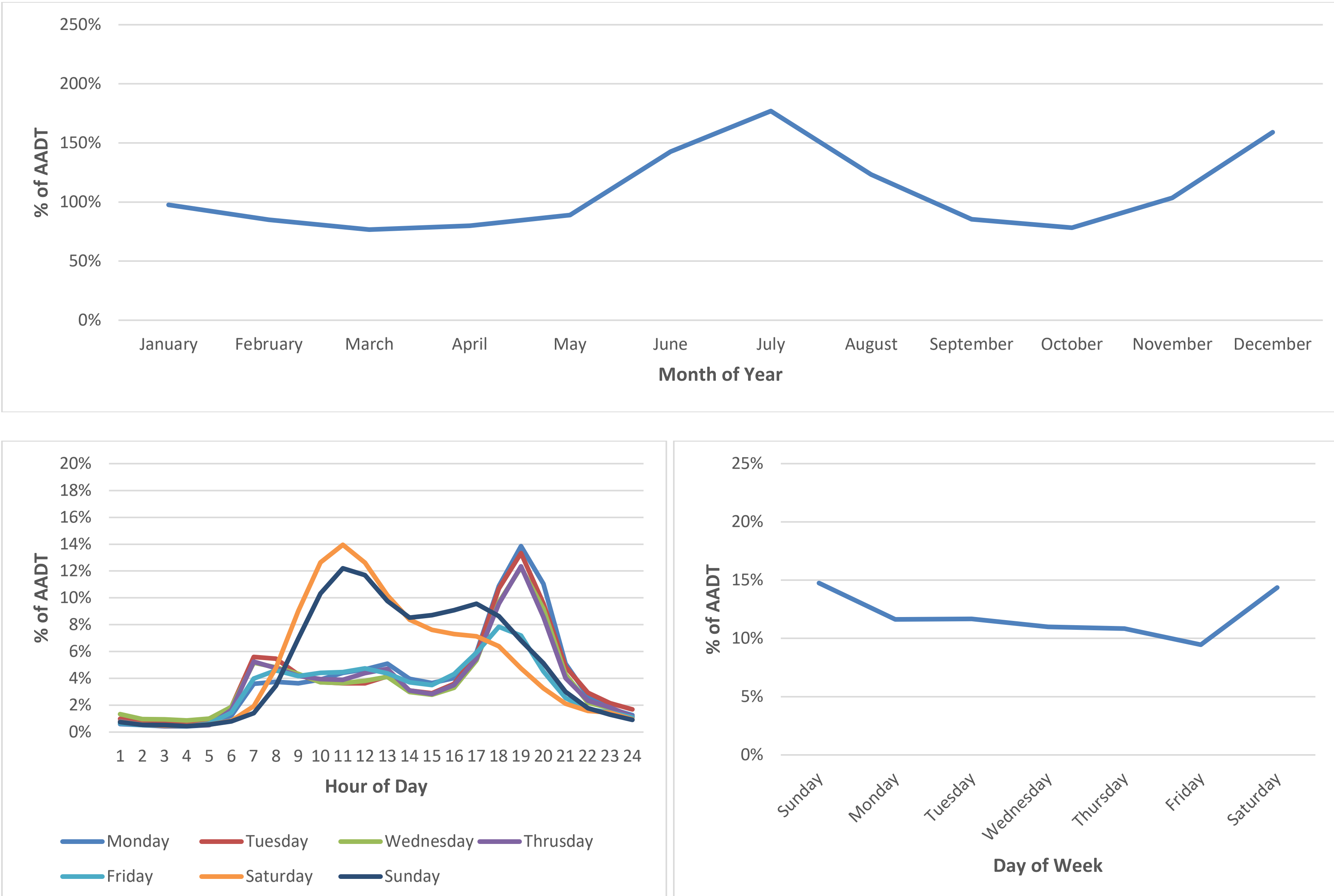


Figure 22. Pedestrian Traffic Count Patterns: Katy Trail at Cedar Springs Rd. (Dallas).

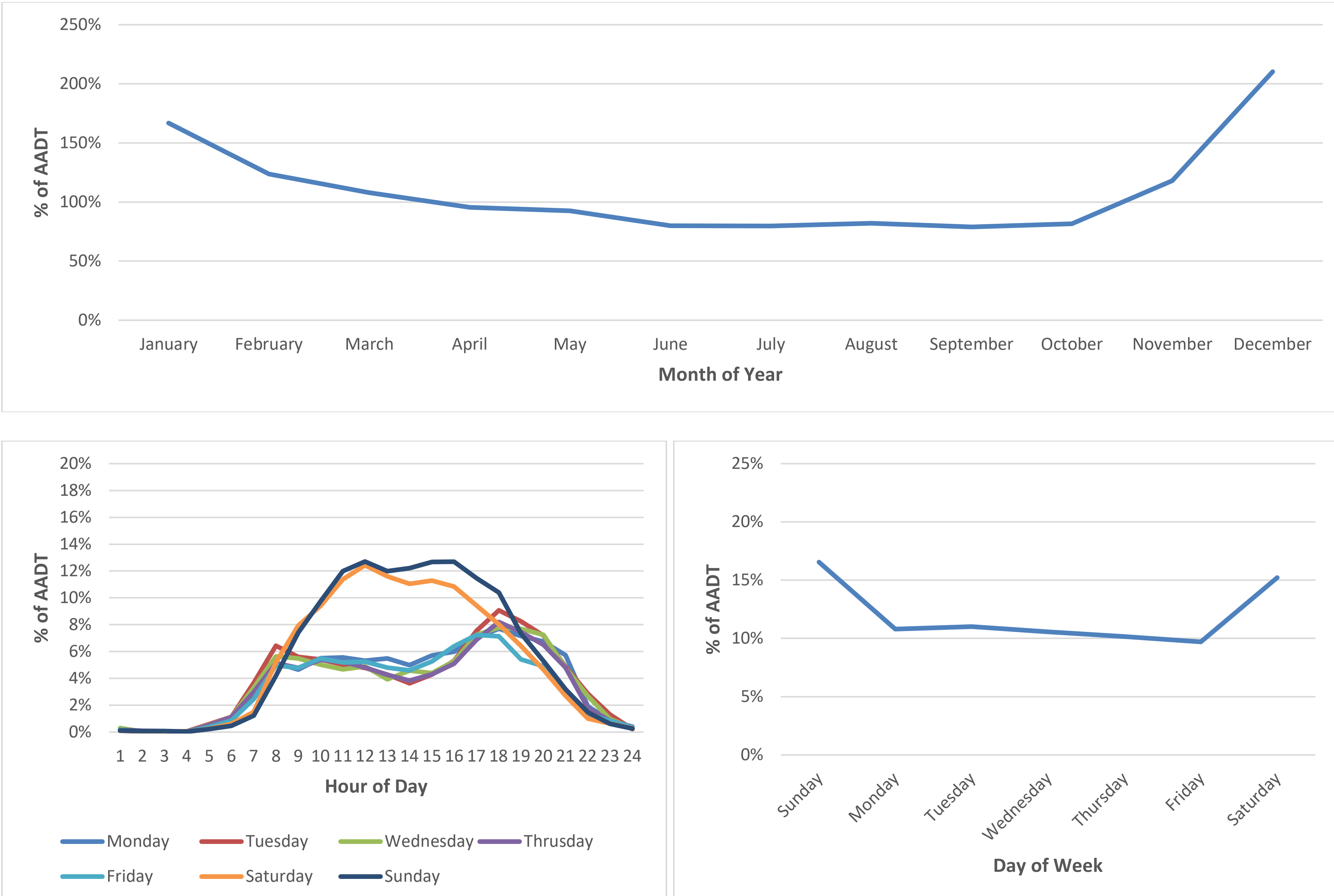


Figure 23. Bicyclist Traffic Count Patterns: Katy Trail at Fitzhugh (Dallas).

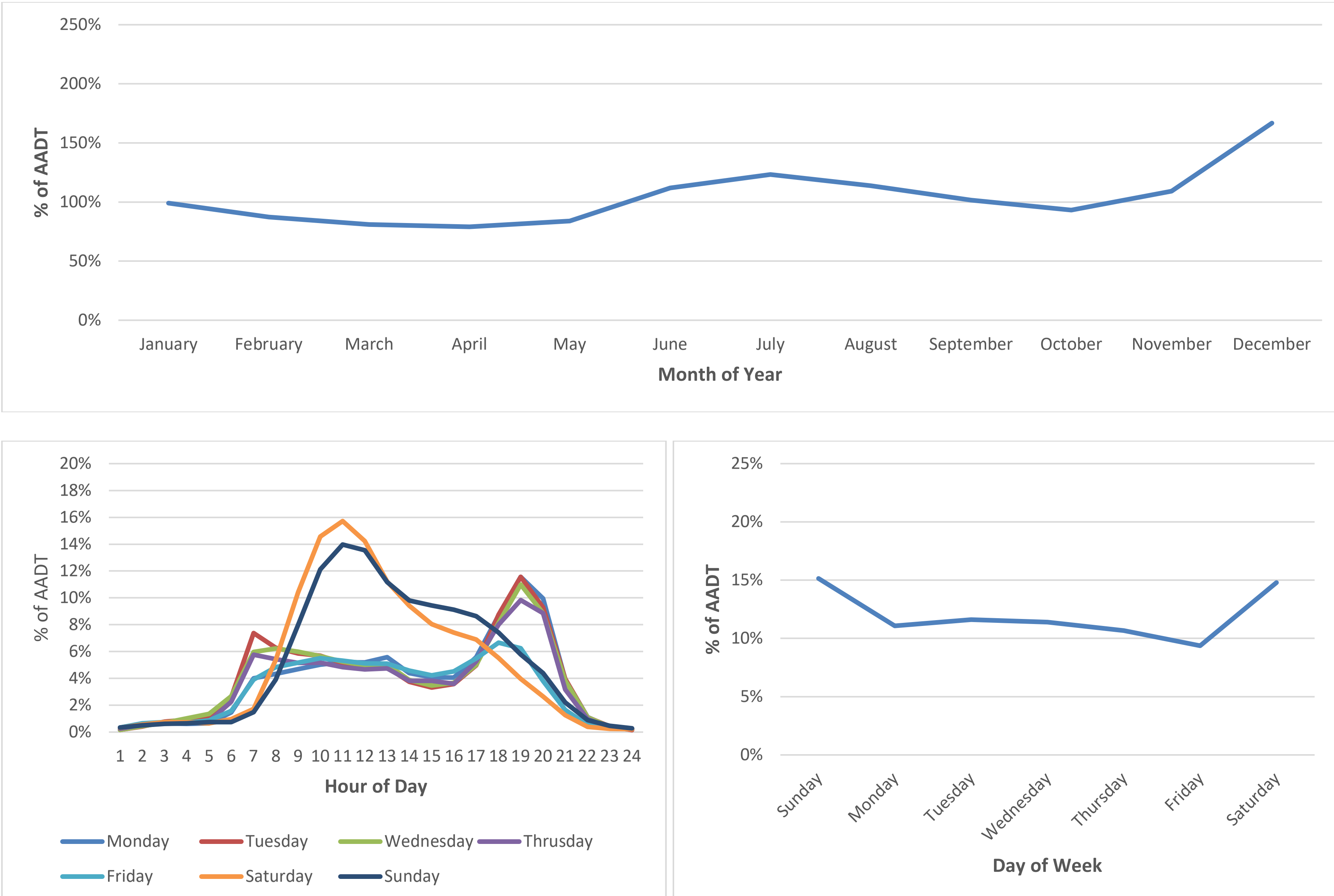
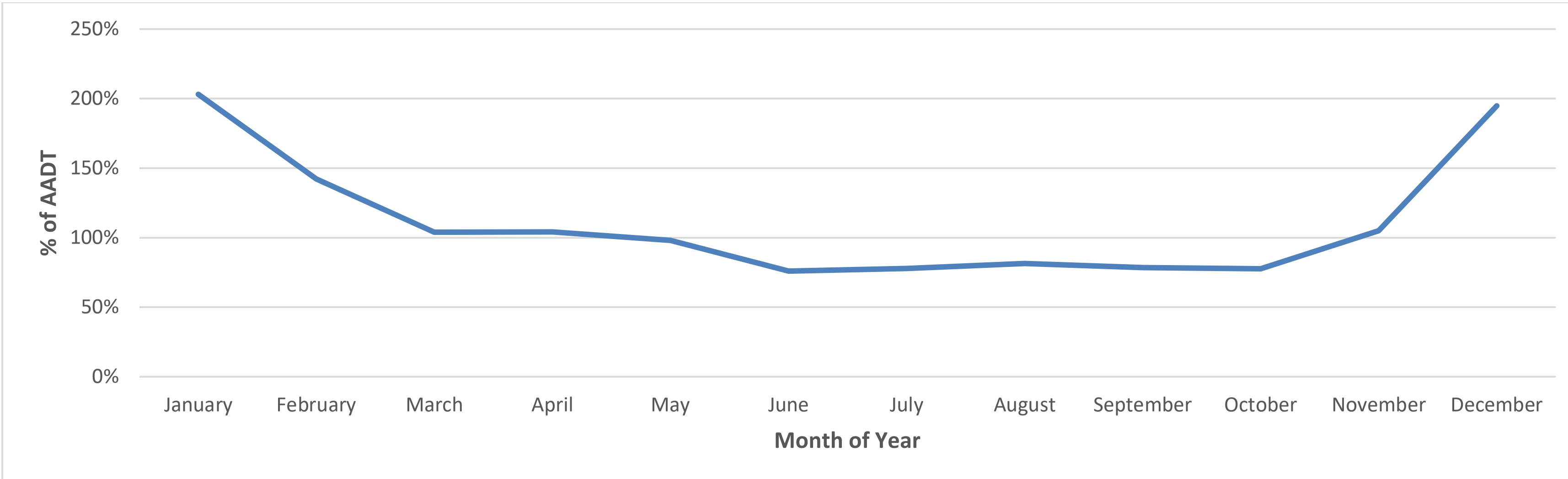
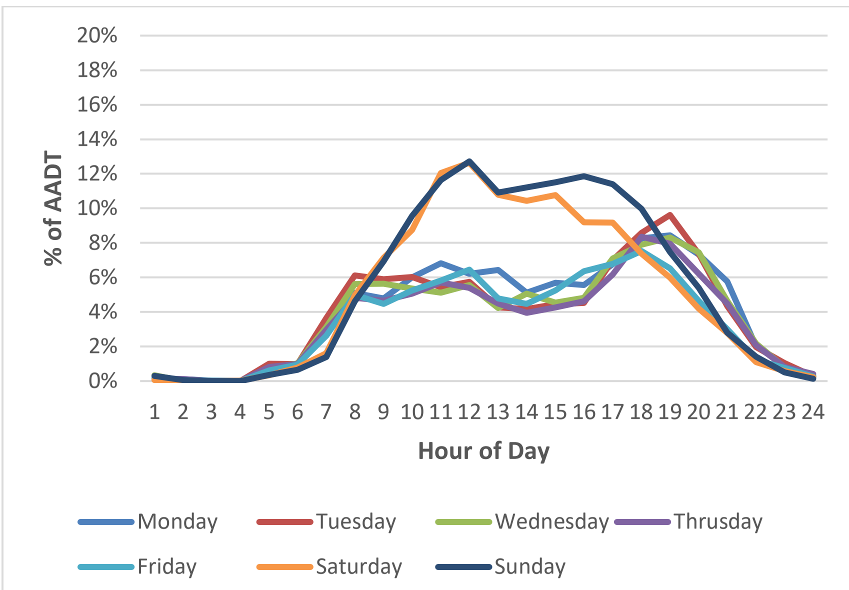


Figure 24. Pedestrian Traffic Count Patterns: Katy Trail at Fitzhugh (Dallas).





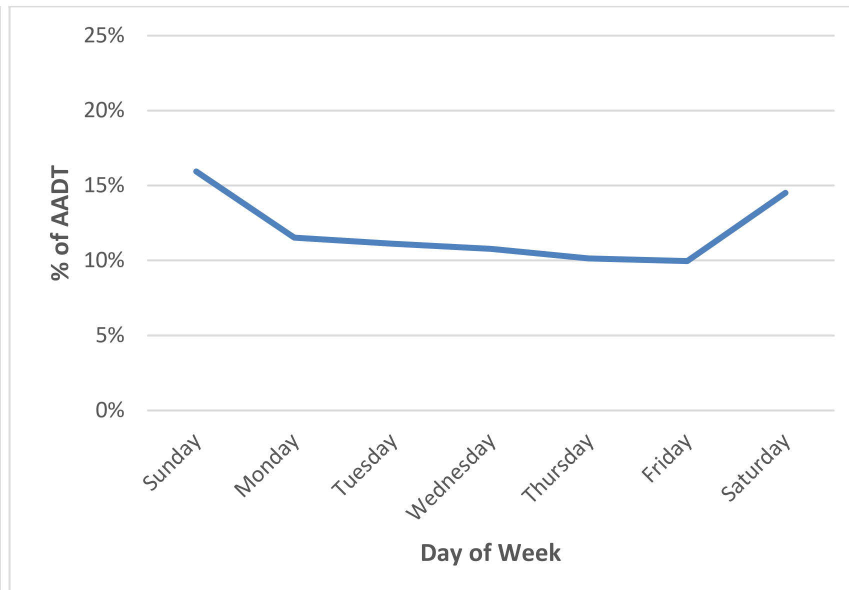


Figure 25. Bicyclist Traffic Count Patterns: Katy Trail at Harvard Avenue (Dallas).

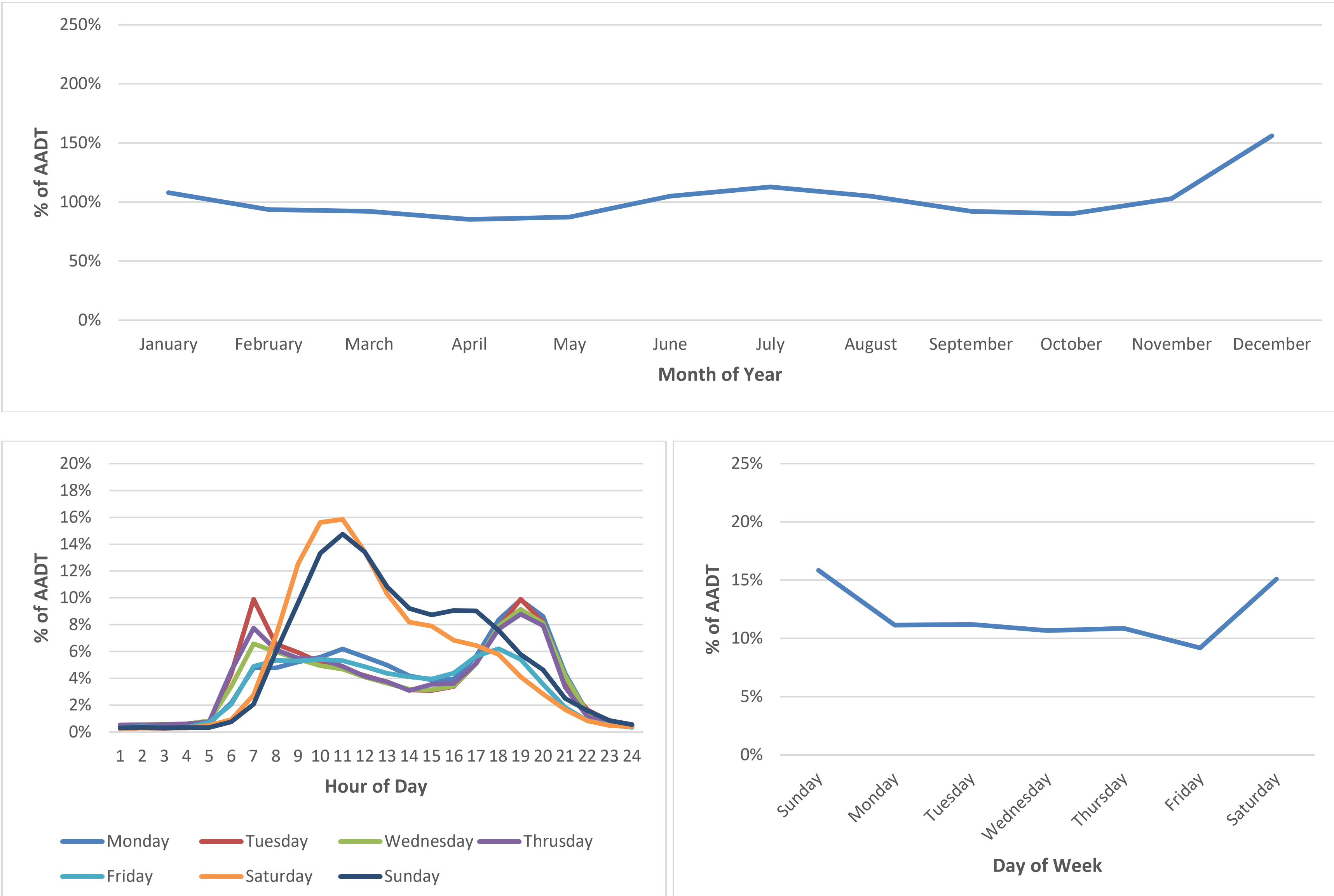
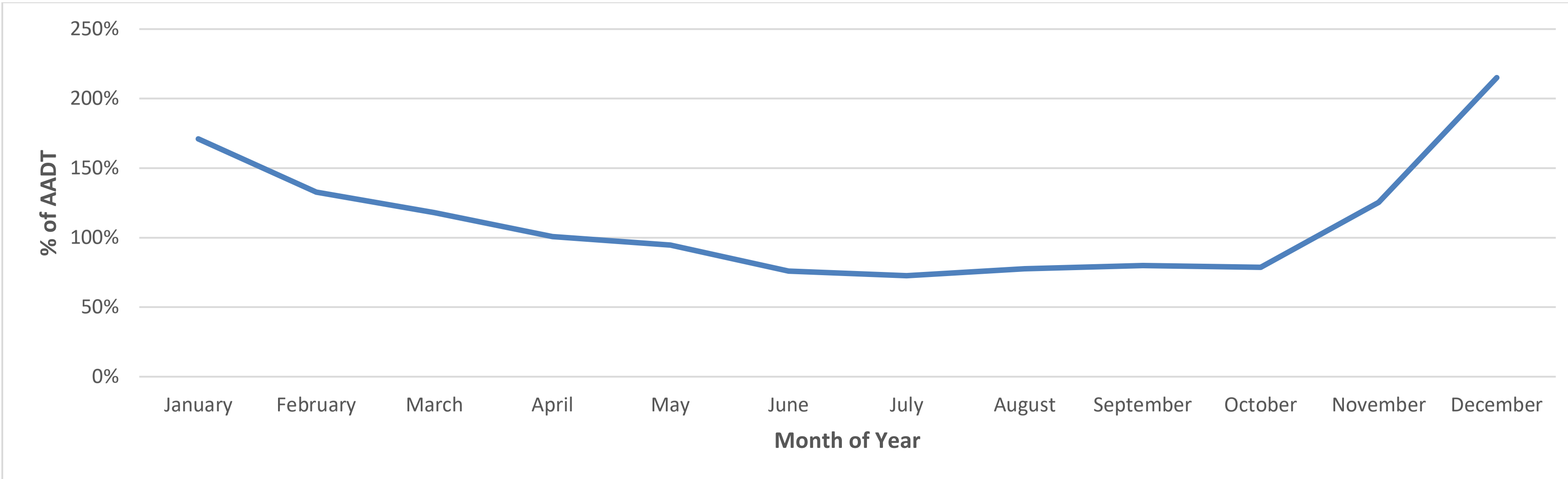
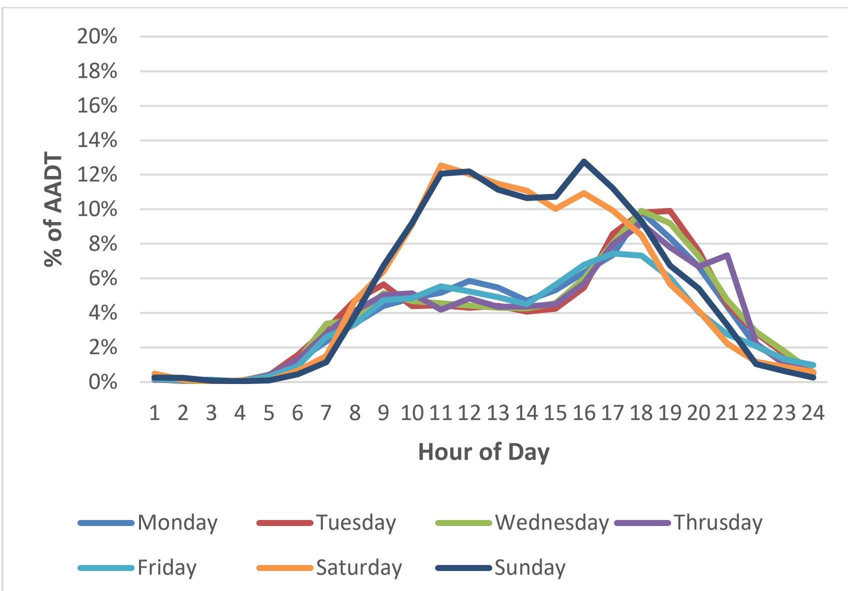


Figure 26. Pedestrian Traffic Count Patterns: Katy Trail at Harvard Avenue (Dallas).





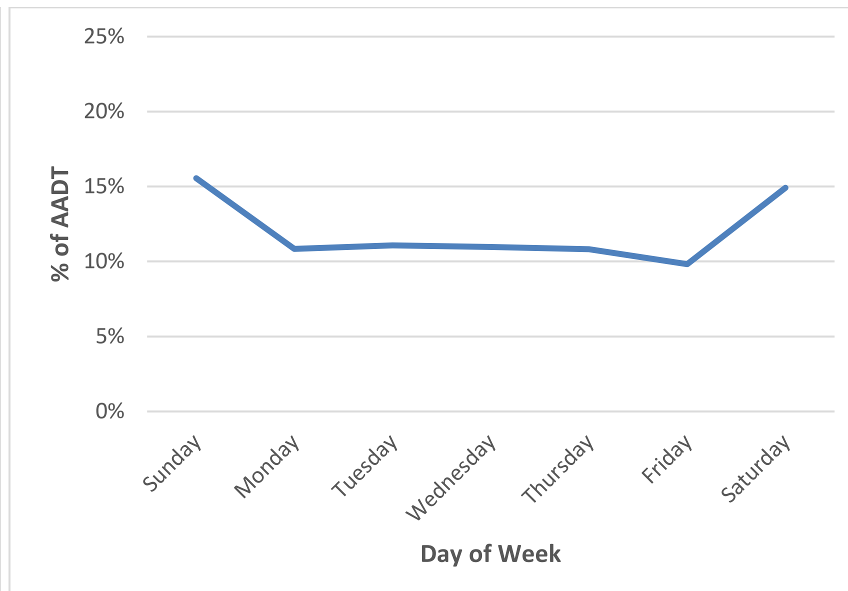
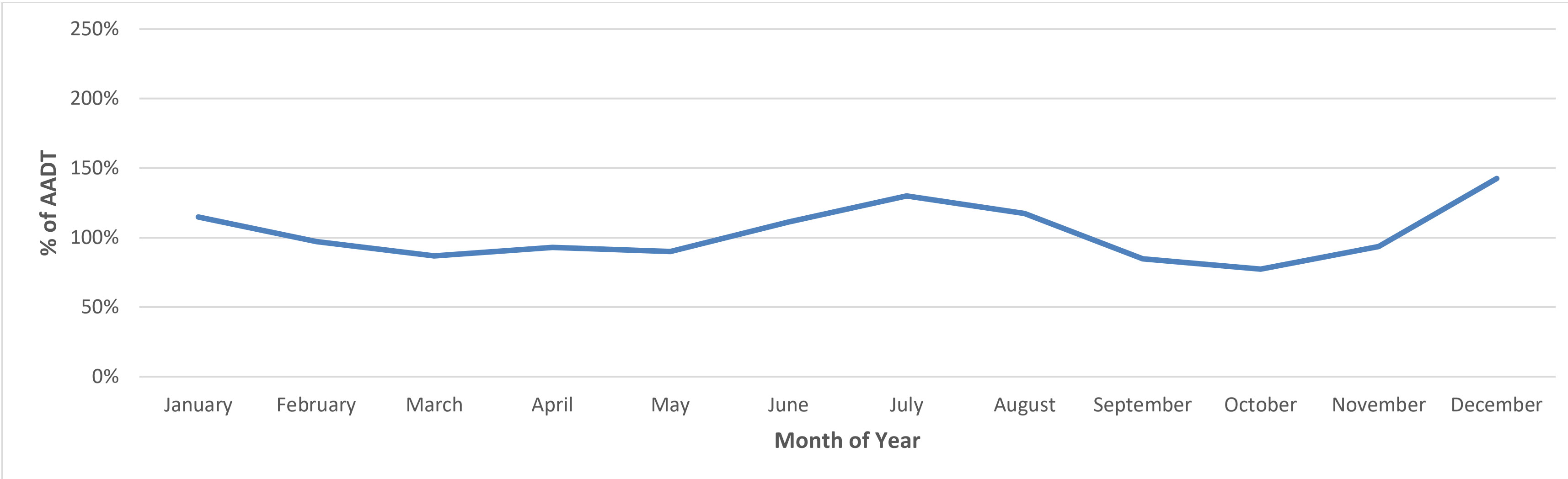
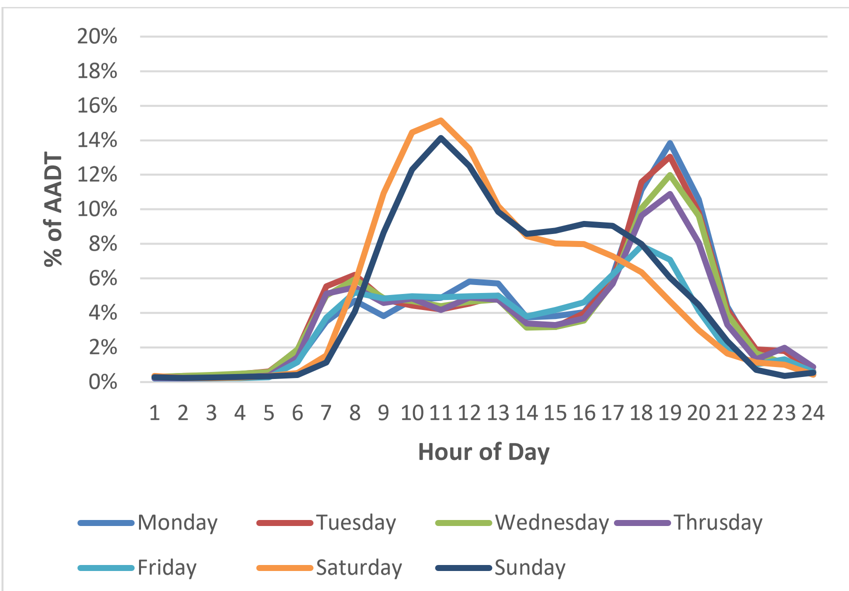


Figure 27. Bicyclist Traffic Count Patterns: Katy Trail (Houston Street/AA Center) (Dallas).





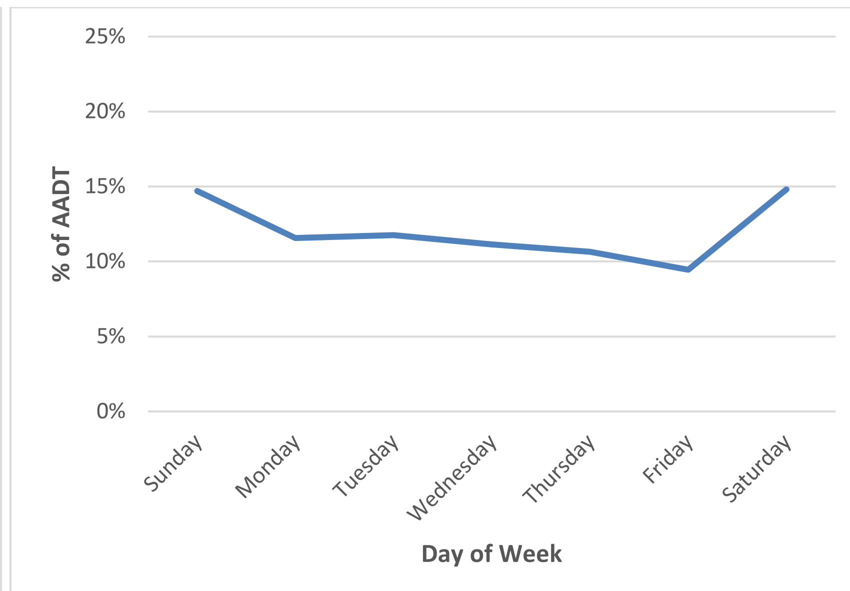


Figure 28. Pedestrian Traffic Count Patterns: Katy Trail (Houston Street/AA Center) (Dallas).

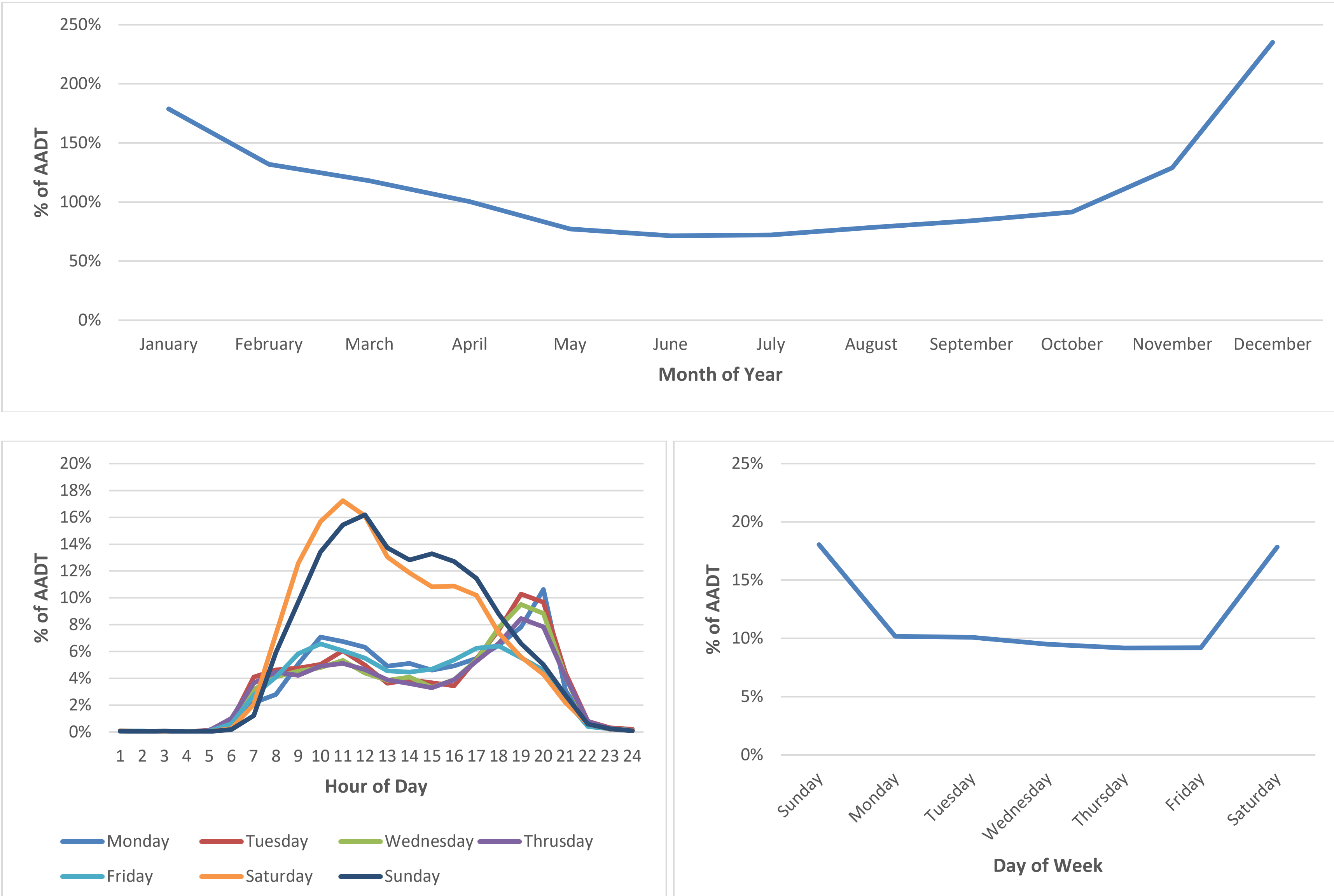


Figure 29. Bicyclist Traffic Count Patterns: White Rock Trail at Big Thicket (Dallas).

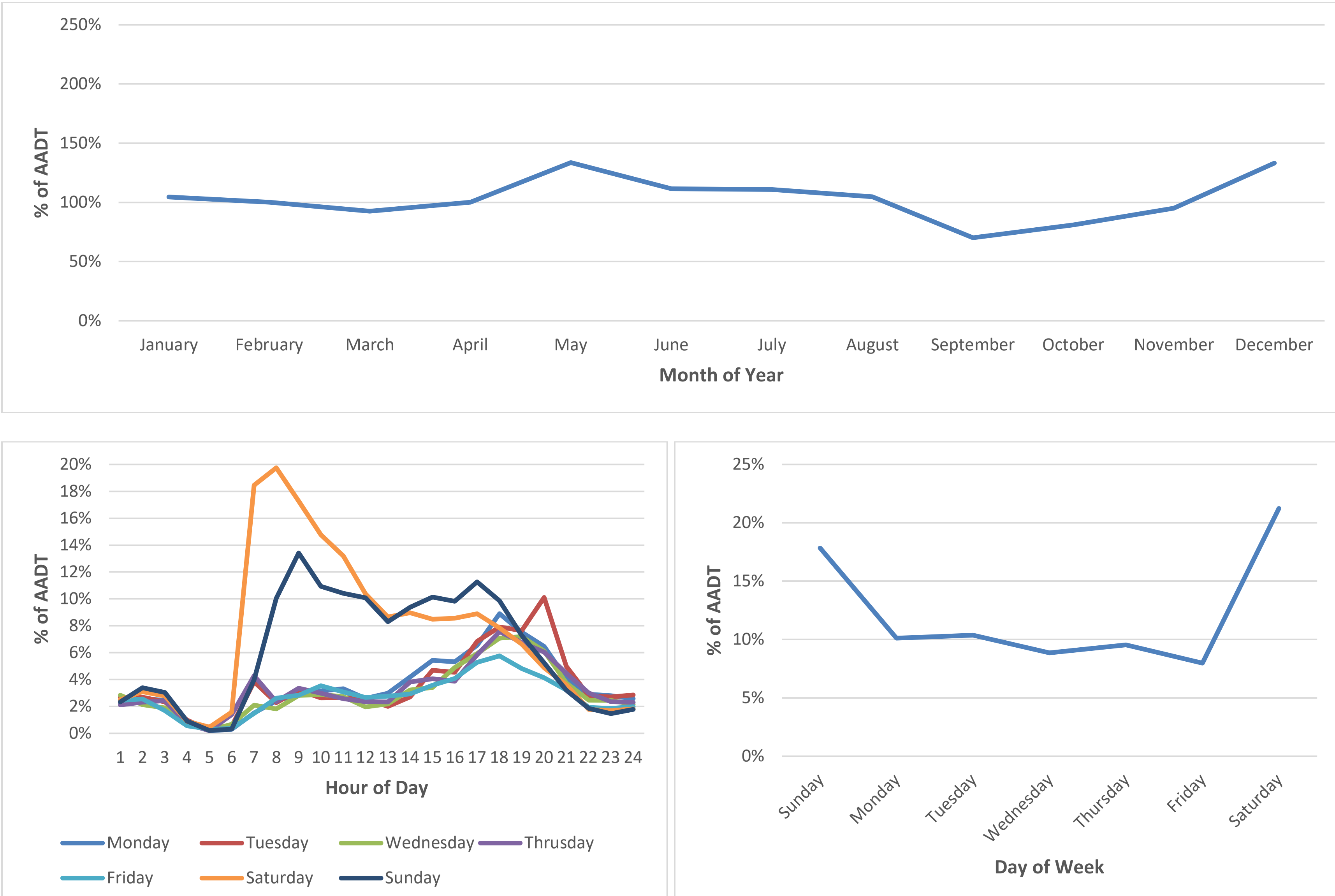
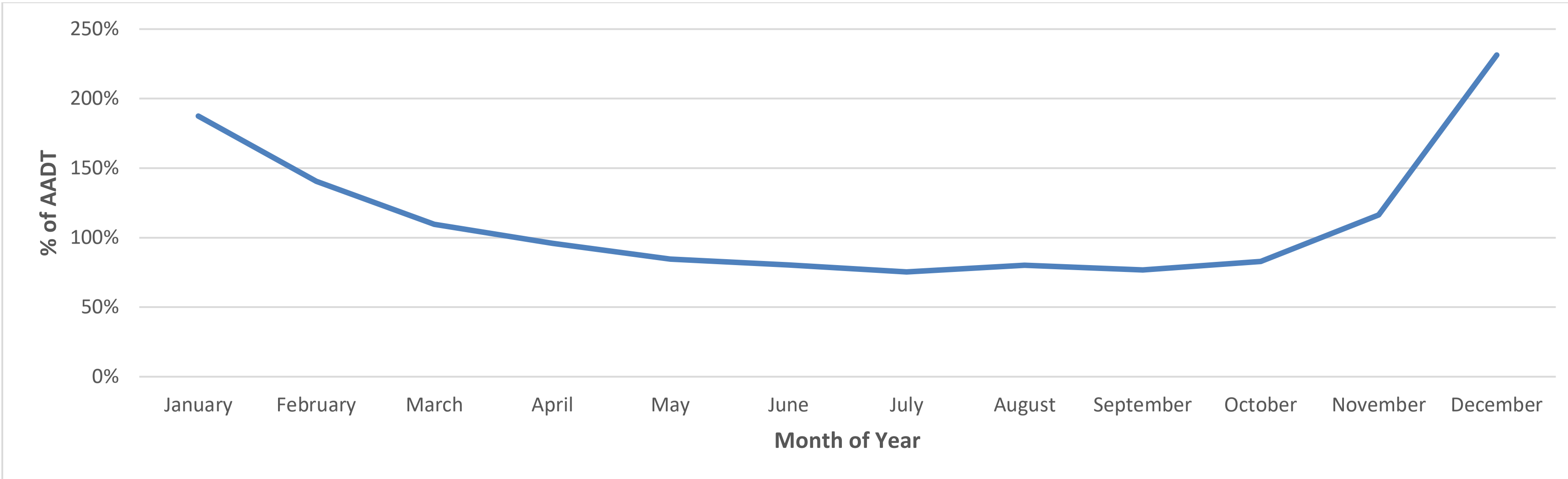
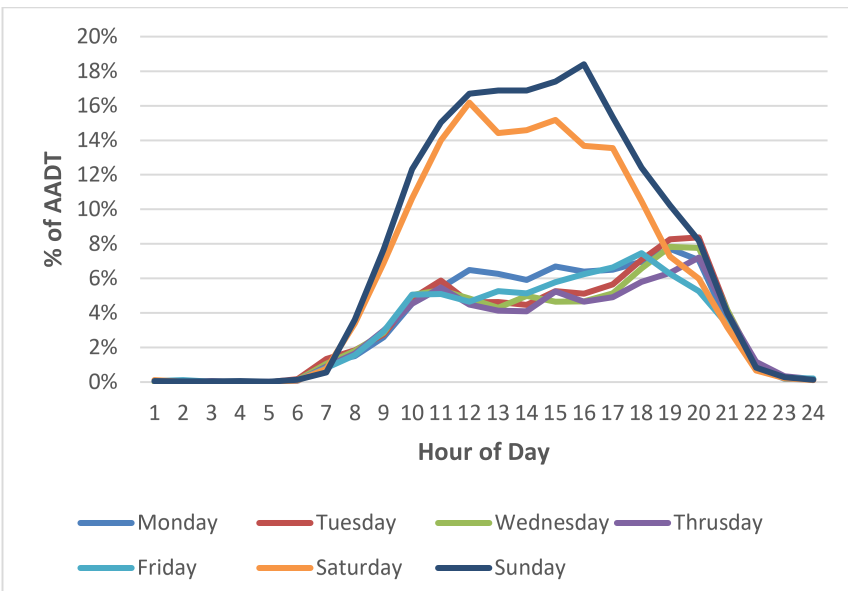


Figure 30. Pedestrian Traffic Count Patterns: White Rock Trail at Big Thicket (Dallas).





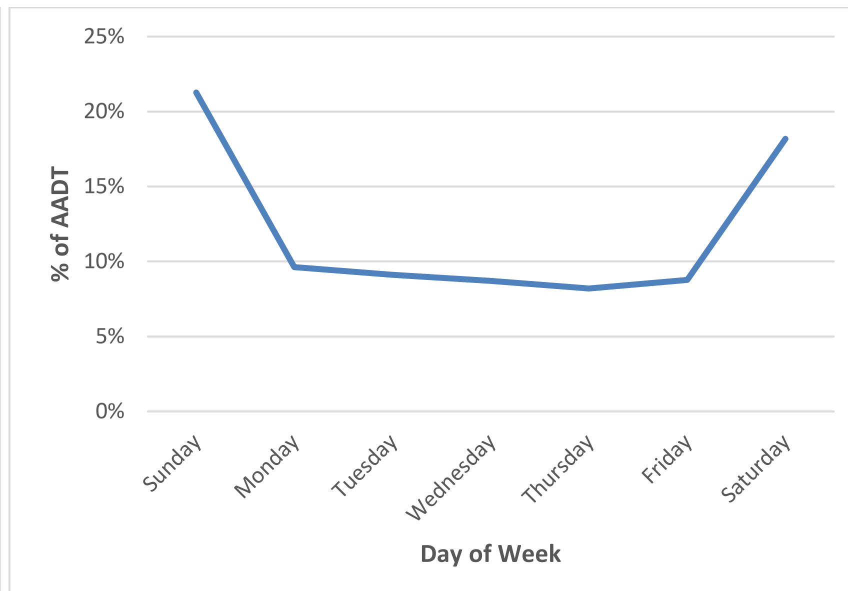


Figure 31. Bicyclist Traffic Count Patterns: White Rock Lake Trail (at Fisher) (Dallas).

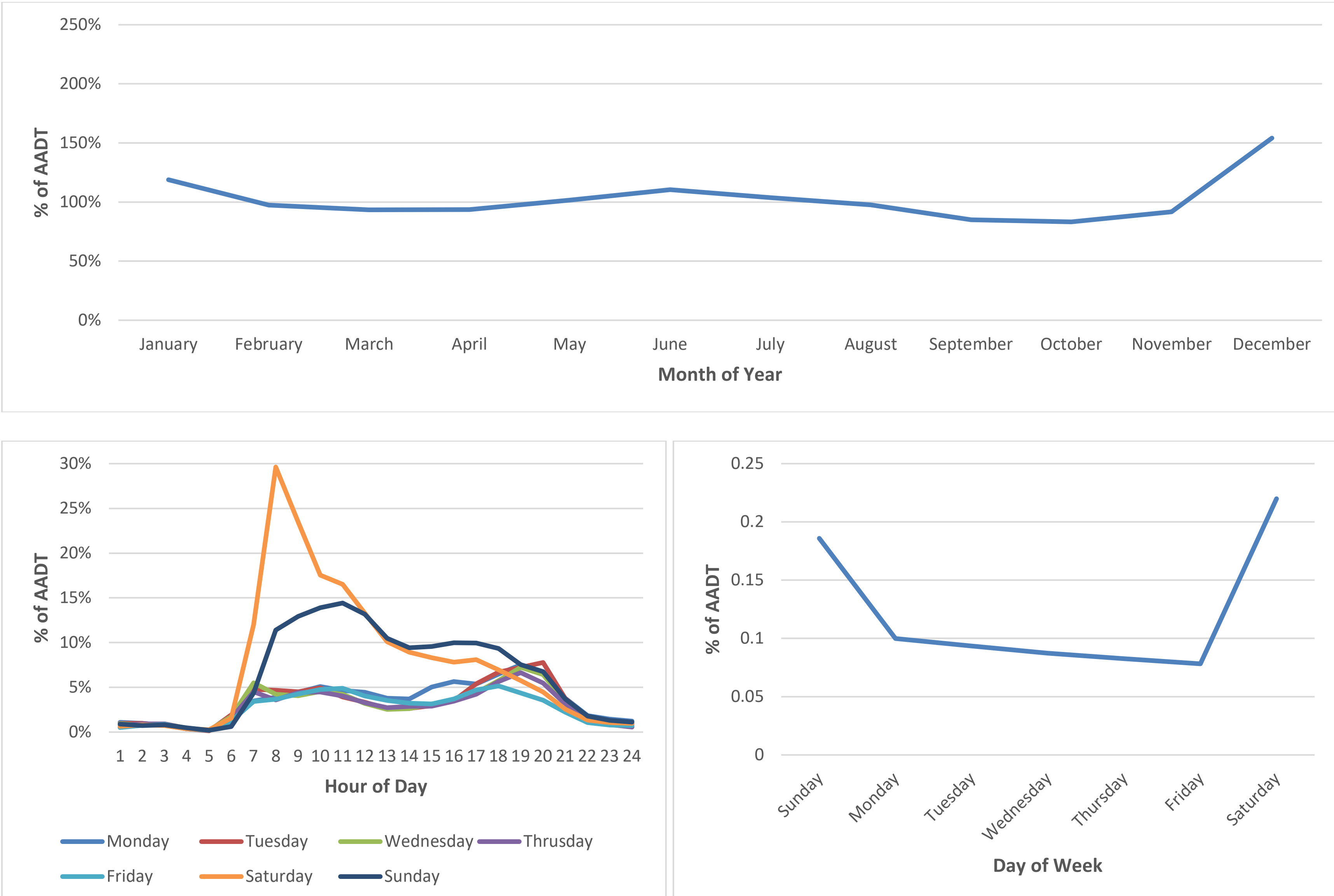
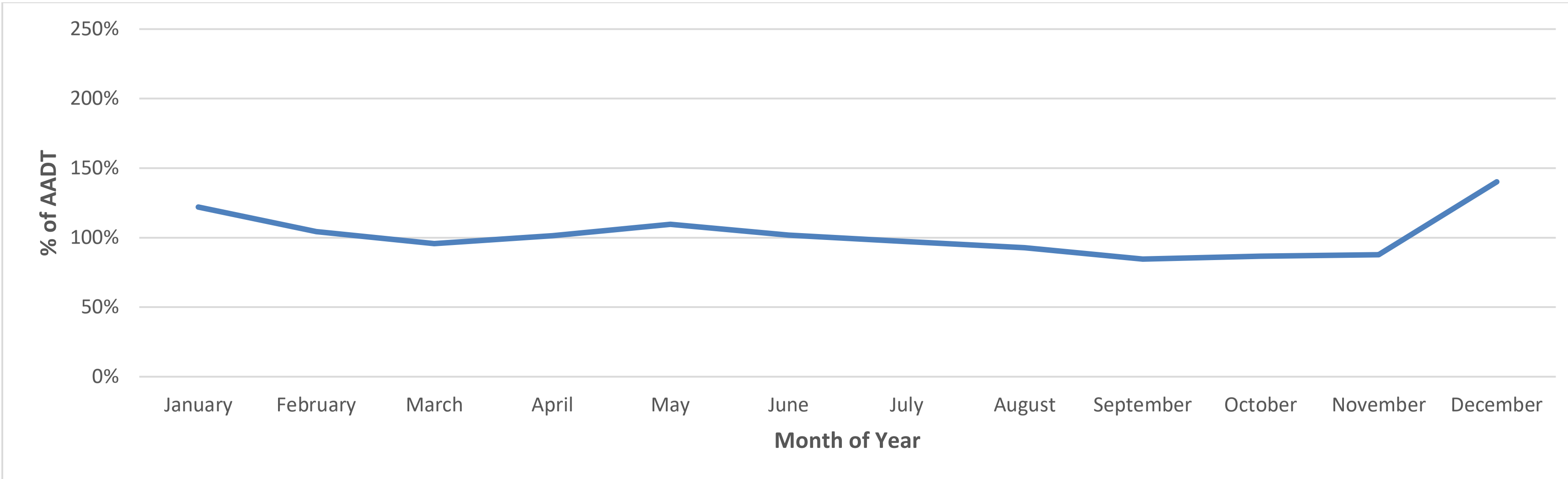
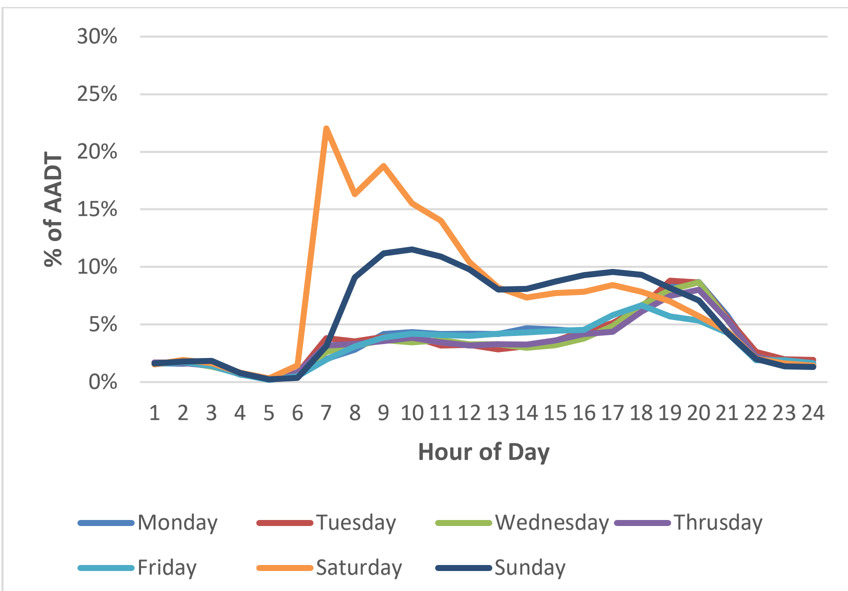


Figure 32. Pedestrian Traffic Count Patterns: White Rock Lake Trail (at Fisher) (Dallas).

Figure 33. Bicyclist Traffic Count Patterns: White Rock Lake Trail at Winfrey Point (Dallas).





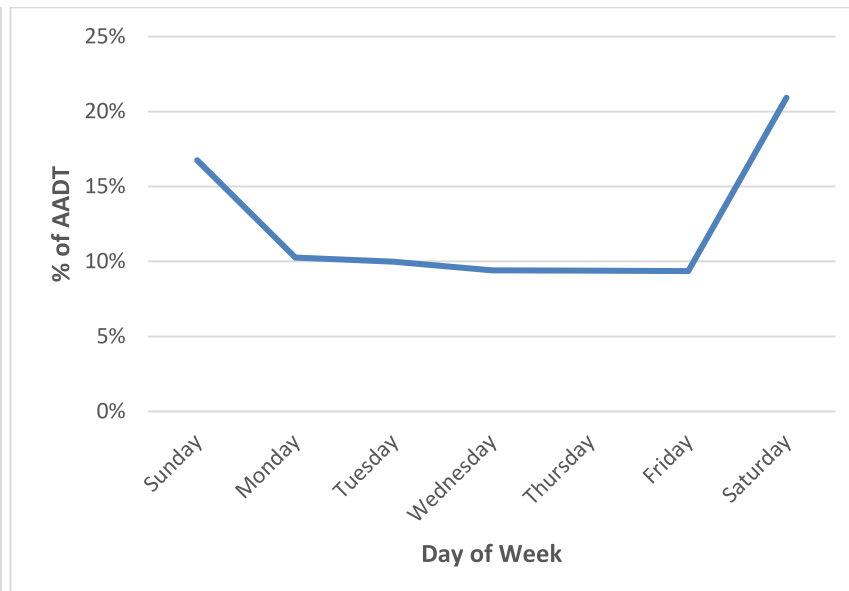


Figure 34. Pedestrian Traffic Count Patterns: White Rock Lake Trail at Winfrey Point (Dallas).

APPENDIX B. DEVELOPMENT OF PROCEDURES FOR CROWDSOURCED DATA SCALING

LIST OF VARIABLES CONSIDERED FOR THE ANALYSIS

Strava	Manual	American Community Survey	RHiNO
Edge ID	Location ID	Total Population	Street Name
Functional Class (CLAZZ)	City	Male Population	Highway Class
Activity	Station Name	Male Population, in Various Age Groups	Functional System
Reverse Activity	Latitude	Female Population	Rural / Urban
Weekend Ratio (both directions of travel)	Longitude	Female Population, in Various Age Groups	Current ADT
Morning Ratio (both directions of travel)	Station ID (both directions of travel)	Total: Households	K-Factor (Peak Hour)
Year	University (School) Present (0.5 miles radius)	Less than $10,000: Households	ADT Combined
Day	Functional Classification	$10,000 to $14,999: Households	Left Shoulder Width
Hour	Facility Type	$15,000 to $19,999: Households	Left Shoulder Type
	Posted Speed Limit	$20,000 to $24,999: Households	Right Shoulder Width
	National Highway System	$25,000 to $29,999: Households	Right Shoulder Type
	Nonmotorized Facility Width	$30,000 to $34,999: Households	Median Width
	Nonmotorized Facility Buffer Width	$35,000 to $39,999: Households	Median Type
	Street Width	$40,000 to $44,999: Households	Number of Lanes
	Parking	$45,000 to $49,999: Households	Surface Width
	Pavement Type	$50,000 to $59,999: Households	Left Curb
	Pavement Condition	$60,000 to $74,999: Households	Right Curb
	ADA Ramps	$75,000 to $99,999: Households	
	Street Lighting	$100,000 to $124,999: Households	
	Street Traffic Volume (ADT)	$125,000 to $149,999: Households	
	Transit	$150,000 to $199,999: Households	
	Shade	$200,000 or more: Households	
		Taxicab: Workers 16 years and over	
		Motorcycle: Workers 16 years and over	
		Bicycle: Workers 16 years and over	

STRAVA SAMPLE PERCENTILE GROUPS

Strava sample percentages refer to the ratio of Strava sample to observed bicycle counts:

$$Strava\ Sample\ Percentage_i = \frac{Strava\ Sample_i}{Bicycle\ Counts_i} \times 100\%$$

Table 2 shows the descriptive statistics of Strava sample percentages for all locations.

Table 2. Descriptive Statistics of Strava Sample Percentages.

Percentage Strava Sample	Min	Max	Mean	St. D.
Travel Direction 1	0%	63%	5%	9%
Travel Direction 2	0%	55%	4%	8%
Average Sample Percentage	0%	59%	5%	8%

Researchers categorized the Strava percentages into five groups, based on the percentiles. Table 3 reports the number of locations per percentile groups.

Table 3. Number of Locations per Percentile Groups.

Strava Percentile Group		Number of Locations
Group 1	Less than 5%	73
Group 2	Equal and more than 5% and less than 10%	15
Group 3	Equal to or more than 10% and less than 15%	4
Group 4	Equal to or more than 15% and less than 20%	2
Group 5	Equal to or more than 20%	6
Total Number of Locations		153

SELECTION OF MOST INFLUENTIAL FACTORS

To select the most influential factors affecting the relationship between Strava counts and ground counts researchers used the Strava percentile groups and conducted the data mining analysis. For this purpose, researchers used Random Forests tool that helps to determine the most important variables based on two criteria:

- Mean Decrease Accuracy.
- Mean Decrease Gini.

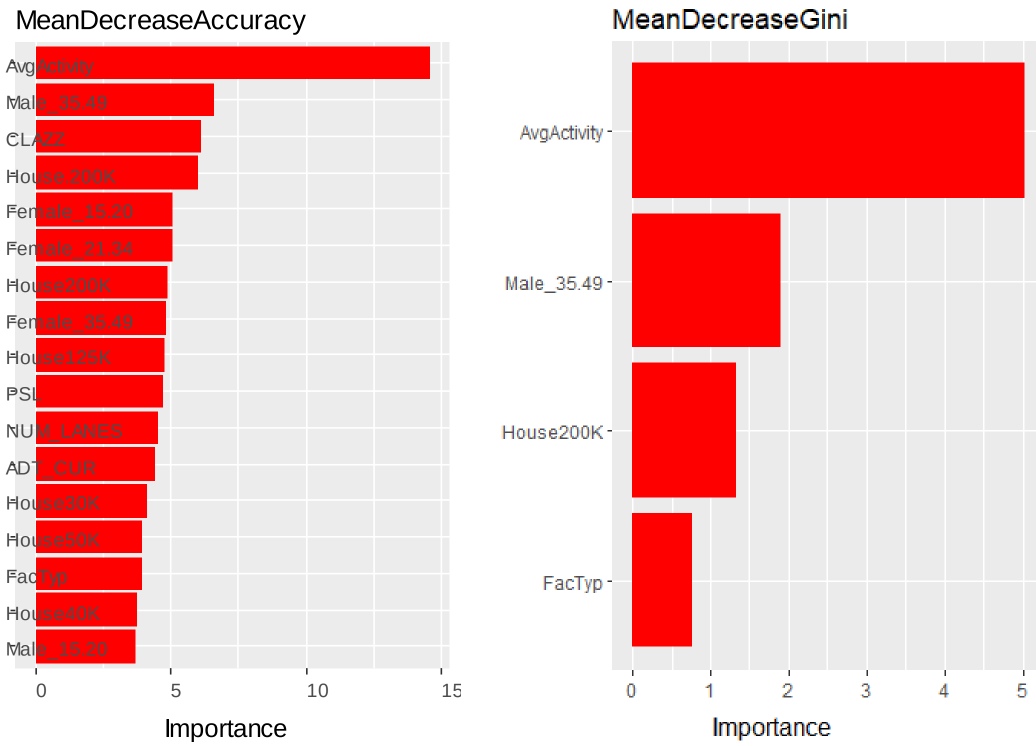


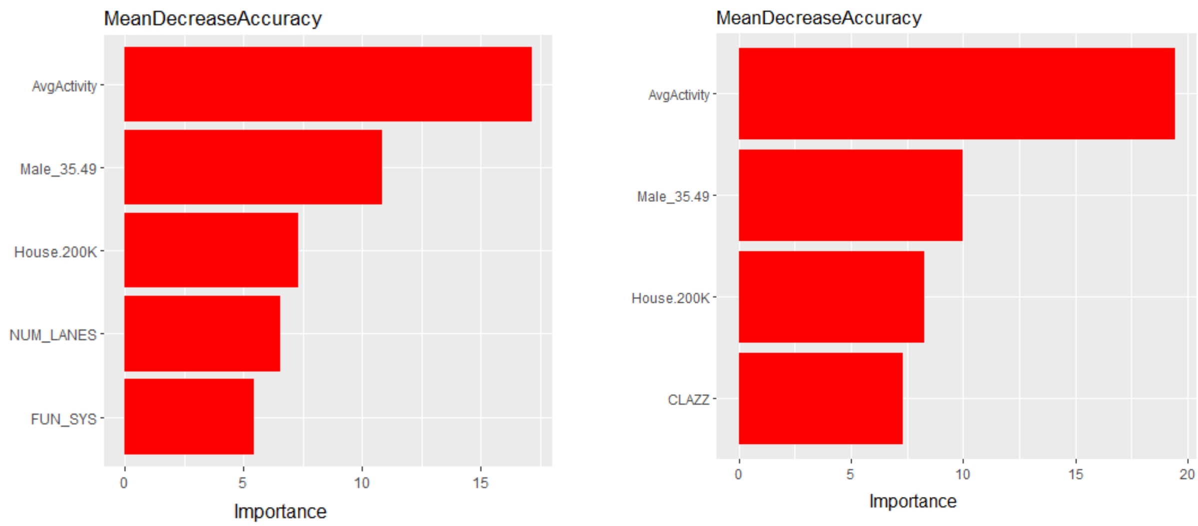
Figure 35. Initial List of Important Variables.

The initial analysis results indicate that the household income and demographic variables are very influential (Figure 35). However, since there are too many variables included in this category, researchers decided to include only the most important variables. These are:

- Household income > $200K.
- Males 35–49.
- Females 21–34.

After conducting the data mining analysis for the second time, by keeping only the most important American Community Survey (ACS) variables, researchers identified the list of most influential variables (Figure 36):

- Strava sample (Strava).
- Male 35–49 (ACS).
- Household income of $200K (ACS).
- Functional System, CLAZZ (Strava).
- Number of Lanes (RHiNO).
- Facility Type (Manual).



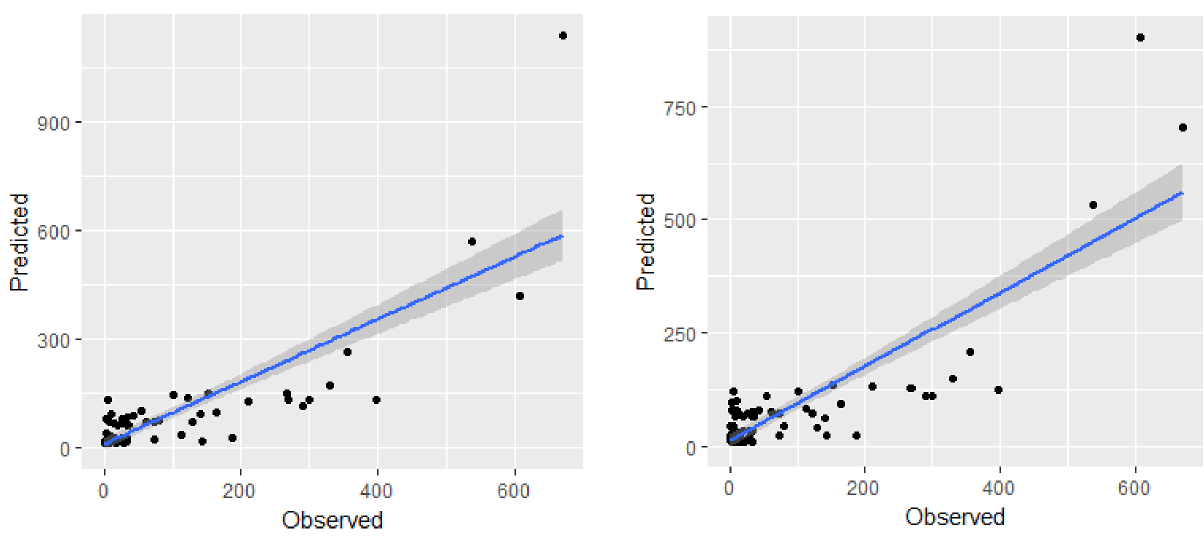
a) If both CLAZZ and FUN_SYS included

b) If only CLAZZ included

Figure 36. Final List of Important Variables.

PREDICTION ERROR MEASURES

Figure 37 indicates the prediction intervals. Table 4 reports the observed and predicted counts and the lower and upper prediction intervals.



a) Model 1

b) Model 2

Figure 37. Prediction Intervals.

Table 4. Strava Sample, Observed and Predicted Counts and 95 Percent Prediction Intervals.

Station Name	City	Strava Sample	Observed Counts	Predicted Counts	Prediction Interval (95%)	
					LL	UL
Ann and Roy Butler Trail @ E Bouldin Creek	Austin	18	396	134	64	134
Ann and Roy Butler Trail @ MoPac Expy	Austin	26	330	170	100	170
Duval St N of E 32nd St	Austin	6	188	28	0	70
Guadalupe St N of W 21st St	Austin	10	141	92	22	92
Johnson Creek Trail @ MoPac	Austin	1	54	102	32	102
Lance Armstrong Bikeway @ Waller Creek	Austin	28	536	567	497	567
Manor Rd @ Alamo St	Austin	9	144	19	0	70
S 1st St @ S Bank Colorado River	Austin	18	121	136	66	136
Shoal Creek Blvd N of W 24th St	Austin	22	81	73	3	73
Shoal Creek Blvd N of W 38th St	Austin	4	10	29	0	70
Walnut Creek Trail N of Jain Ln	Austin	23	152	151	81	151
18th: Fig to McKenzie	Corpus Christi	0	10	18	0	70
Alameda: Atlantic to Naples	Corpus Christi	0	1	14	0	70
Alameda: Mussett to Caldwell	Corpus Christi	0	6	18	0	70
Alameda: Naples to Atlantic	Corpus Christi	0	3	14	0	70
Amber: J to Don Patricio	Corpus Christi	0	6	18	0	70
Antelope: Artesian to Carrizo	Corpus Christi	0	6	14	0	70
Aquarius: Doubloon to Gunwale	Corpus Christi	0	14	20	0	70
Ayers: 2nd to 3rd	Corpus Christi	0	4	14	0	70
Ayers: 3rd to 2nd	Corpus Christi	0	6	15	0	70

Station Name	City	Strava Sample	Observed Counts	Predicted Counts	Prediction Interval (95%)	
					LL	UL
Ayers: 6th to 7th	Corpus Christi	0	5	14	0	70
Ayers: 7th to 6th	Corpus Christi	0	3	14	0	70
Beauvais: Grenade to Beau Terre	Corpus Christi	1	8	23	0	70
Bernice: Rickey to Susan	Corpus Christi	1	6	18	0	70
Blevins: Ormond to Melbourne	Corpus Christi	0	22	18	0	70
Brockhampton: Dunbarton Oak to La Rochelle Way	Corpus Christi	1	7	23	0	70
Buford: 23rd to 22nd	Corpus Christi	0	11	18	0	70
Buford: Santa Fe to 3rd	Corpus Christi	0	5	18	0	70
Carroll: Carrolleton to Gollihar	Corpus Christi	0	7	22	0	70
Carroll: Gollihar to Carrolleton	Corpus Christi	0	4	22	0	70
Carroll: Lamont to Harold	Corpus Christi	0	6	22	0	70
Carroll: Lamont to Marion (#1)	Corpus Christi	0	8	22	0	70
Carroll: Lamont to Marion (#2)	Corpus Christi	0	7	22	0	70
Carroll: Pasadena to Peerman	Corpus Christi	1	15	22	0	70
Cheyenne: Aztec to Washington	Corpus Christi	0	5	18	0	70
Cliff Crenshaw: Guess to Turkey Creek	Corpus Christi	1	33	18	0	70
Corona: Embassy to Flynn Parkway	Corpus Christi	0	3	22	0	70
Corona: Flynn Parkway to Embassy	Corpus Christi	0	3	22	0	70

Station Name	City	Strava Sample	Observed Counts	Predicted Counts	Prediction Interval (95%)	
					LL	UL
Gollihar: Dody to Weber	Corpus Christi	0	4	14	0	70
Gollihar: Driftwood to Dolphin	Corpus Christi	0	2	13	0	70
Gollihar: Kasper to Weber	Corpus Christi	0	1	13	0	70
Gollihar: Laura to Kirkwood	Corpus Christi	0	6	13	0	70
Gollihar: Laura to Randall	Corpus Christi	0	17	13	0	70
Gollihar: Sequoia to Willow	Corpus Christi	0	2	13	0	70
Gypsy: Hawksnest Bay to Whaler	Corpus Christi	0	8	25	0	70
Holly: SH 286 to Martin	Corpus Christi	0	28	13	0	70
Holly: Victor Lara Ortegon to Greenwood	Corpus Christi	0	9	13	0	70
Lawnview: Sunset to Melrose	Corpus Christi	0	10	18	0	70
Matlock: Greenbay to Mciver	Corpus Christi	0	23	18	0	70
Mesquite: Fitzgerald to Resaca	Corpus Christi	0	13	18	0	70
Peoples: Chapparal to Mesquite	Corpus Christi	0	9	18	0	70
Pontchartrain: Kaw Lake to Lake Travis	Corpus Christi	1	5	21	0	70
Retta: Purdue to Selkirk	Corpus Christi	0	14	25	0	70
River Canyon: Teague to Rolling Ridge	Corpus Christi	0	15	18	0	70
Rodd Field: Saratoga to Brooke	Corpus Christi	0	30	14	0	70
Roosevelt: Clark to Vanderbilt	Corpus Christi	0	9	18	0	70

Station Name	City	Strava Sample	Observed Counts	Predicted Counts	Prediction Interval (95%)	
					LL	UL
Sabinas: Dunbar to Tarlton	Corpus Christi	0	4	18	0	70
Wood River: Red River to Rapids	Corpus Christi	1	26	28	0	70
Yorktown: Great Lakes to Oso Parkway	Corpus Christi	1	6	16	0	70
AT&T Trail at Trinity Forest Trail	Dallas	0	8	72	2	72
Coombs Creek Trail	Dallas	2	10	93	23	93
Gateway Park Loop Trail	Dallas	0	34	28	0	70
Glendale Park Loop Trail	Dallas	0	10	28	0	70
Glendale Park South Loop Trail	Dallas	0	15	28	0	70
Katy Trail (Houston street/ AA Center-	Dallas	8	210	127	57	127
Katy Trail at Cedar Springs Rd	Dallas	14	269	133	63	133
Katy Trail at Fitzhugh	Dallas	13	289	115	45	115
Katy Trail at Harvard Avenue	Dallas	10	164	95	25	95
Kiest Park Loop Trail at Conservation	Dallas	0	37	63	0	70
Preston Ridge Trail at La Cosa Dr.	Dallas	3	4	77	7	77
Preston Ridge Trial at Debbe	Dallas	3	42	86	16	86
Trinity strand Trail at Hi Line Dr	Dallas	0	10	32	0	70
W NorthWest Highway	Dallas	0	33	63	0	70
White Rock Lake Trail	Dallas	12	266	149	79	149
White Rock Lake Trail at Winfrey Point	Dallas	76	669	1140	1070	1140
White Rock Trail at Big Thicket	Dallas	81	607	418	348	418
White Rock Trail at Dog Park	Dallas	27	355	264	194	264
Brays Bayou Greenway Trail @ Spur 5	Houston	2	33	82	12	82
Columbia Tap Trail @ Blodgett Street	Houston	2	61	71	1	71
Heights Trail @5 1/2 Street	Houston	13	300	130	60	130

Station Name	City	Strava Sample	Observed Counts	Predicted Counts	Prediction Interval (95%)	
					LL	UL
White Oak Bayou Trail@ 34th Street	Houston	17	102	147	77	147
SH 3 N @ S of Walter Hall Park	League City	2	129	69	0	70
Plano Bluebonnet Trail at US75	Plano	2	3	40	0	70
Plano Legacy Trail	Plano	5	5	130	60	130
Plano OPP & NP Trail	Plano	1	27	80	10	80
Mission Reach Trail at north end of San Juan Ditch	San Antonio	1	15	66	0	70
Mission Reach Trail south of Theo Ave	San Antonio	3	74	71	1	71
Mission Reach Trail south of VFW Blvd (Mission County Park)	San Antonio	2	31	68	0	70
Riverwalk Trail between 8th and 9th Street	San Antonio	6	113	36	0	70
Old Alice Road at N of Belvedere	Brownsville	5	74	21	0	70
FM 802 at W of Habana	Brownsville	1	29	65	0	70
FM 802 at W of Habana	Brownsville	0	19	63	0	70
Sports Park Blvd at W of Brownsville Sports Park	Brownsville	0	4	23	0	70
Mid College Chaparral Creek Bike Lane South Side	Midland	0	8	19	0	70
Oakwood Dr at W of Sunnygrove Dr	Odessa	0	1	18	0	70
Maple Ave at S of E 14th St	Odessa	0	1	18	0	70
W 22nd St at Ventura Ave	Odessa	0	1	18	0	70
Wichita River Trail at E of Broad St/IH-44	Wichita Falls	0	20	63	0	70
FM 369 at N of US 287	Wichita Falls	0	1	14	0	70
Burkburnett Rd (SH 240) at N of Airport Dr	Wichita Falls	1	2	14	0	70

In addition to the prediction intervals, researchers calculated three prediction error measures to test the prediction accuracy of each model:

- Mean Absolute Percentage Error.
- Mean Absolute Error.
- Mean Squared Error.

Mean Absolute Percentage Error

$$MAPE = \frac{1}{n}\sum_{n} 100 \times \left|\frac{\hat{Y}_{PM_p,t} - Y_{PM_p,t}}{Y_{PM_p,t}}\right|$$

Mean Accuracy Error

$$MAE = \frac{1}{n}\sum_{n}\left|\hat{Y}_{PM_p,t} - Y_{PM_p,t}\right|$$

Mean Squared Error

$$MSE = \frac{1}{n}\sum_{n}\left(\hat{Y}_{PM_p,t} - Y_{PM_p,t}\right)^2$$

Where,

- $n -$ is the size of the validation data (four months).
- $Y_{PM,t} -$ is the performance measures vector for period t, where $t = 1, T$.
- $\hat{Y}_{PM_p,t} -$ is the predicted value of the performance measures.

Table 5 reports the error measures for two models while Table 6 shows the error measures Strava percentile groups. Note that in this table the percentages refer to the error rate and not the accuracy. Hence, the higher the value, the higher the prediction error. According to the error rates, the prediction power of both models are quite similar. Moreover, the accuracy is better for the roadway segments where the Strava sample represents 5–15 percent of bicycle counts (Groups 2 and 3).

Table 5. Relative Accuracy per Strava Percentage Categories.

Prediction Error Measure	Model 1 – Average Counts (CLAZZ)	Model 2 – Average Counts (RHiNO)
Mean Absolute Percentage Error	29%	38%
Mean Squared Error	5855	4836
Mean Absolute Error	41	42

Table 6. Prediction Error per Strava Percentile Groups.

A. Mean Absolute Percentage/Prediction Error

Strava Percentile Group		Model 1 – Average Counts (CLAZZ)	Model 2 – Average Counts (RHiNO)
Group 1	Less than 5%	33%	44%
Group 2	Equal and more than 5% and less than 10%	9.8%	12.5%
Group 3	Equal to or more than 10% and less than 15%	9.9%	10.1%
Group 4	Equal to or more than 15% and less than 20%	42%	46%
Group 5	Equal to or more than 20%	38%	39.7%
	Average MAPE	29%	38%

B. Mean Squared Error

Strava Percentile Group		Model 1 – Average Counts (CLAZZ)	Model 2 – Average Counts (RHiNO)
Group 1	Less than 5%	3066	3500
Group 2	Equal and more than 5% and less than 10%	5686	7501
Group 3	Equal to or more than 10% and less than 15%	9894	2390
Group 4	Equal to or more than 15% and less than 20%	3557	5154
Group 5	Equal to or more than 20%	38286	1606
	Average MSE	5855	4836

C. Mean Absolute Error

Strava Percentile Group		Model 1 – Average Counts (CLAZZ)	Model 2 – Average Counts (RHiNO)
Group 1	Less than 5%	30	33
Group 2	Equal and more than 5% and less than 10%	59	67
Group 3	Equal to or more than 10% and less than 15%	66	103
Group 4	Equal to or more than 15% and less than 20%	49	68
Group 5	Equal to or more than 20%	104	34
	Average MAE	41	42