# Guide for Seasonal Adjustment and Crowdsourced Data Scaling

Technical Report 0-6927-P6

Cooperative Research Program

TEXAS A&M TRANSPORTATION INSTITUTE
COLLEGE STATION, TEXAS

in cooperation with the
Federal Highway Administration and the
Texas Department of Transportation
http://tti.tamu.edu/documents/0-6927-P6.pdf

**Technical Report Documentation Page**

| 1. Report No.<br>FHWA/TX-18/0-6927-P6 | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| 4. Title and Subtitle<br>GUIDE FOR SEASONAL ADJUSTMENT AND CROWDSOURCED DATA SCALING | | 5. Report Date<br>Published: December 2018<br>6. Performing Organization Code |
| 7. Author(s)<br>Bahar Dadashova, Greg Griffin, Subasish Das, Shawn Turner, and Madison Graham | | 8. Performing Organization Report No.<br>Product 0-6927-P6 |
| 9. Performing Organization Name and Address<br>Texas A&M Transportation Institute<br>The Texas A&M University System<br>College Station, Texas 77843-3135 | | 10. Work Unit No. (TRAIS)<br>11. Contract or Grant No.<br>Project 0-6927 |
| 12. Sponsoring Agency Name and Address<br>Texas Department of Transportation<br>Research and Technology Implementation Office<br>125 E. 11th Street<br>Austin, Texas 78701-2483 | | 13. Type of Report and Period Covered<br>Product<br>February 2018–August 2018<br>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<br>Pedestrian and Bicyclist Count Data, Seasonal Adjustment Factors, Crowdsourced Data Scaling | 18. Distribution Statement<br>No restrictions. This document is available to the public through NTIS:<br>National Technical Information Service<br>Alexandria, Virginia<br>http://www.ntis.gov | | |
|---|---|---|---|
| 19. Security Classif. (of this report)<br>Unclassified | 20. Security Classif. (of this page)<br>Unclassified | 21. No. of Pages<br>76 | 22. Price |

**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

Published: December 2018

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

# DISCLAIMER

This research was performed in cooperation with the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the FHWA or TxDOT. This report does not constitute a standard, specification, or regulation.

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

# 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 on continuous count data collected from 17 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 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 17 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): have 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: 11 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 at least 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 (Figure 1). 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 2).

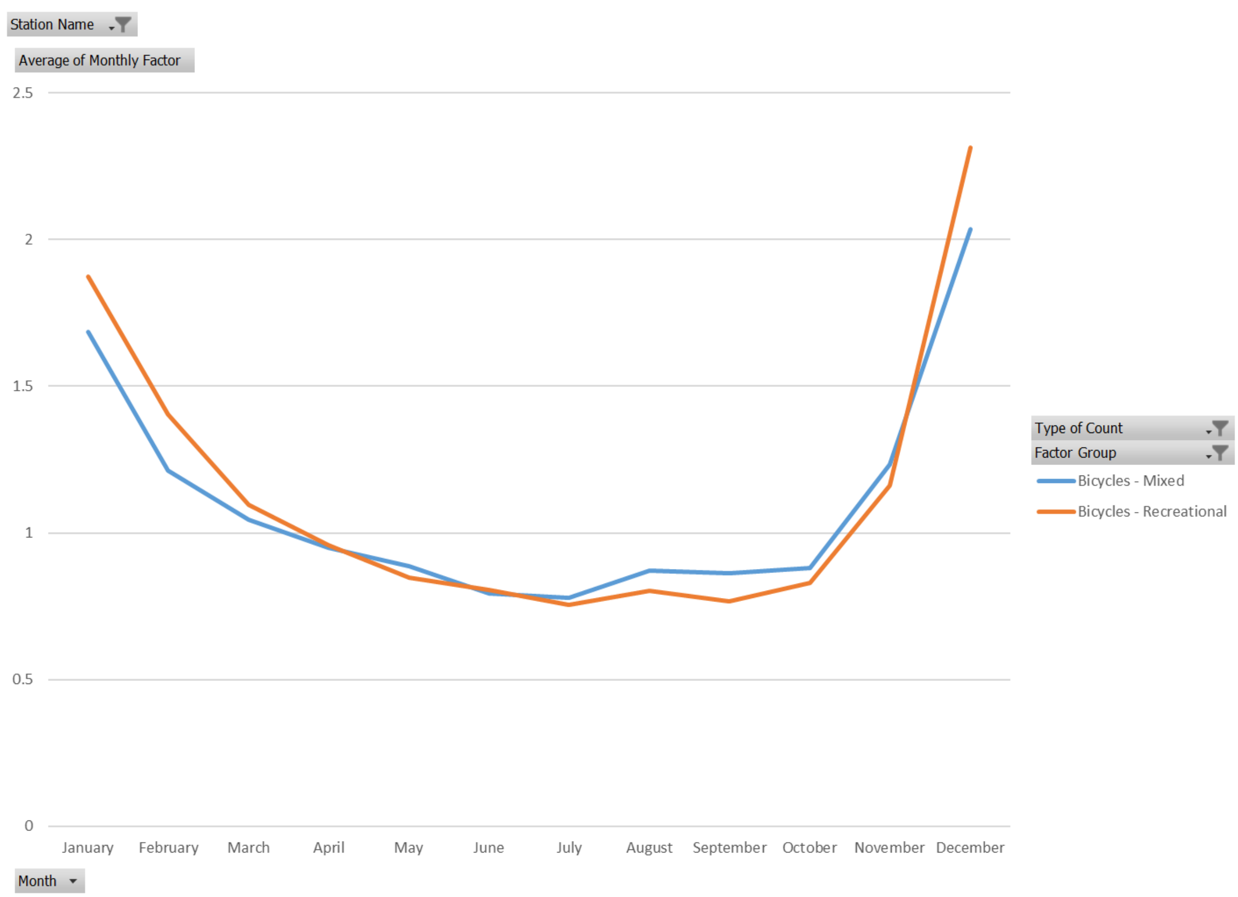



**Figure 1. Chart Illustrating Similar Seasonal Patterns for Different Factor Groups.**

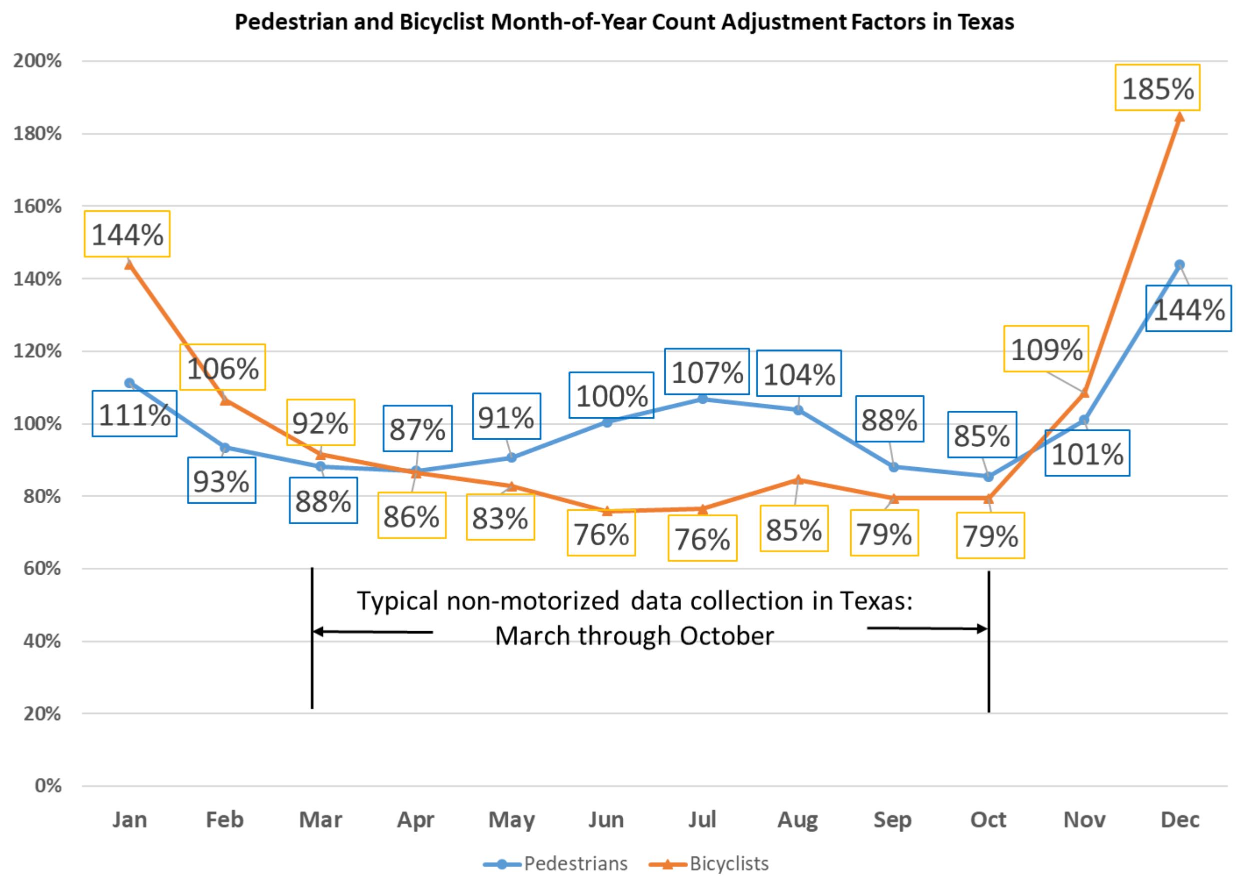



**Figure 2. Month-of-Year Count Adjustment Factors for Short-Duration Counts.**

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, then combine them with traditional counts to estimate bicycle travel volumes across most of the state of Texas.

Crowdsourcing uses a broad pool of individuals through an online platform that aggregates and formats the information for a specific use. In this case, bicyclist travel is 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 trips 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 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 model, using crowdsourced bicyclist trips as a main input, along with functional classification of a transportation segment, and nearby high-income residential areas (see Appendix B for details about model development).

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 data 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. Researchers found that the model using the Open Street Map classification (also used in the Strava Metro product) had a lower mean absolute percentage error (29 percent versus 38 percent for RHiNO). Therefore, researchers decided to use the CLAZZ variable instead of FUN-SYS as the roadway functional classification variable.

Income plays a role in the proportion of bicyclists logging trips on Strava, though it is less important in the model than Strava activity or functional classification. Smartphones may be more available for higher-income users, and the fitness-oriented nature of Strava users may further result in higher use among 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 activity, and household income form the basis of the model to estimate total bicycle trip volumes. Refer to project report 0-6927-R1 and Appendix B for additional description of the study methodology.

## STEPS TO ESTIMATE 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 annual number of bicyclist activities logged via Strava in both directions (TACTCNT = 16,271), the density of households with more than $200,000 income in the given block group ($\text{Household Density}_i = 0$), and the functional classification ($CLAZZ_i$= Cycleway).

### Step 1 – Record Annual Daily Strava Bicyclist Activities

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. Strava activity data are available through Strava's Dataviewer (http://metro-static.strava.com/dataView/TEXAS/201607_201706/RIDE/#5/31.215/-101.239), and in GIS

[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

shapefiles. In the Strava Dataviewer, data are displayed as annual roll-ups of activities, i.e. the total Strava activities are pooled for a given point or linear segment during the entire year. In GIS shapefiles, Strava provides activity count data at different geographies (streets, intersections, areas), and time periods (i.e. annual, monthly, hourly), described further in the current Strava Metro Comprehensive User Guide that is provided with the company's data deliveries. In Strava data, segments are referred to as edges and are assigned a unique identifier. In our example, the edge ID of Walnut Creek Trail is 1644966. The bicyclist activity for the Strava edges can be found in the following files:

- Annual roll-up: texas_201607_201706_ride_rollup_total.csv
- Monthly roll-up: texas_201607_201706_ride_rollup_month_2016_7_total.shp
- Weekday of Month roll-up: texas_201607_201706_ride_rollup_month_2016_7_weekday.shp
- Weekend of Month roll-up: texas_201607_201706_ride_rollup_month_2016_7_weekend.shp

Strava activity are available for both directions of travel (total activity count, TACTCNT), for default direction of travel (activity count, ACTCNT) and for reverse direction of travel (reverse activity count, RACTCNT). Strava does not report the name of the travel direction. The default direction can be identified by using the arrow symbols in ArcGIS.

After selecting the Strava segment (edge) for analysis, review Strava activities 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 using the annual roll-up data, divide the total activity counts (TACTCNT) by 365 to estimate average daily Strava bicycle traffic (AADB Strava). If monthly, divide TACTCNT by 30 or the actual number of days in the recorded month. If weekly, divide TACTCNT by 7 to estimate daily traffic. Finally, round to the nearest integer.

In this case, 16,271 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 45.

$$AADB\ Strava_{Walnut\ Creek} = \frac{Annual\ TACTCNT_{Walnut\ Creek}}{365} = 44.57 = 45$$

## 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 activities and the number of nearby households with annual income over $200,000.

| *Functional Classification (CLAZZ in Strava Metro's network data from Open Street Map)* | |
|---|---|
| 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 chose the following equation:

$$AADB_{Walnut\ Creek} = 62 \times (\exp(AADB\ Strava_{Walnut\ Creek}))^{0.038} \times$$

$$(\exp(\text{Household} > \ 200\text{K}_{\,Walnut\ Creek}))^{0.002}$$

## Step 3 – Plug in Values to Excel

Insert the daily count of Strava trips (45), and the number of high-income households (0), and the equation becomes:

$$AADB_{Walnut\ Creek} = 62 \times (\exp(45))^{0.038}(\exp(0))^{0.002}$$

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

=62*(EXP(45)^0.038)*(EXP(0)^0.002)

Average Annual Daily Bicyclist traffic at Walnut Creek = 343

The results show that the predicted number of bicycles on this segment is equal to 343. Calculation of lower and upper prediction intervals for AADB are 272 and 412 respectively. Additional detail on prediction interval calculation is provided in Appendix B. Note that the observed counts are 304 at this trail, indicating that the AADB model predicted the ground count at this location relatively accurately.

### 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 through the area 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 activities 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 activities, since the input data used counts at some low-activity-bicycling locations throughout the state. Table 1 can be used to review against estimates with low Strava activity levels.

**Table 1. Estimated Annual Bicycle Traffic Given Strava Activity 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 |

## SUMMARY AND CAVEATS OF USING AADB ESTIMATION MODELS

To develop the AADB models, researchers have used the ground counts collected from 100 count stations. The ground counts were mainly collected from urban areas and shared use paths. Moreover, as indicated earlier, Strava uses Open Street Map (OSM) as the basemap. OSM classifies the roadways into 22 categories or CLAZZ (Appendix C). The sites used in this study only represent 7 CLAZZ categories. Although the model goodness of fit measures are within acceptable range (i.e. 29% error margin, and 70% accuracy level), the researchers suggest that the practitioners take caution when implementing these models to estimate the bicycle counts on rural segments and CLAZZs that are not included in this study. Appendix C provides further guidelines how the AADB model can be used to estimate the AADB counts for the roadway functional classes that were not included in the modelling process.

# 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.

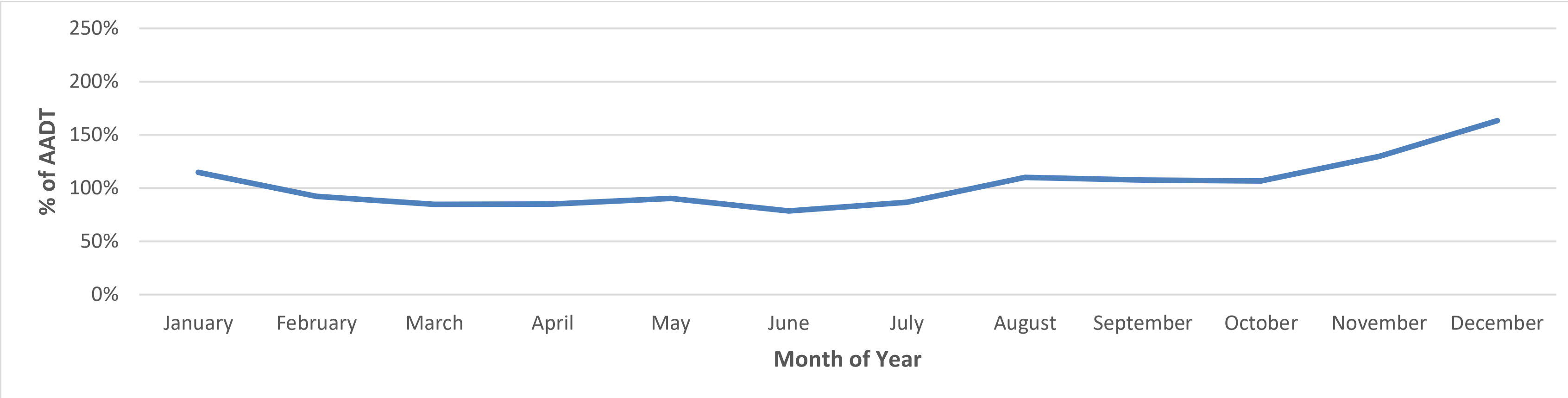
Griffin, G. P., and J. Jiao. "Where Does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus." *Journal of Transport & Health*, Vol. 2, No. 2, 2015b, pp. 238–247. https://doi.org/10.1016/j.jth.2014.12.001.

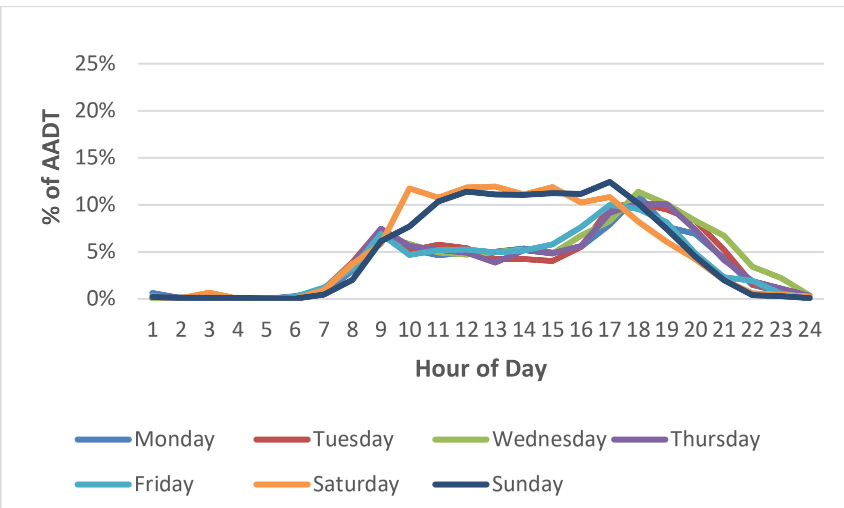
Jestico, B., T. Nelson, and M. Winters. "Mapping Ridership Using Crowdsourced Cycling Data." *Journal of Transport Geography*, Vol. 52, 2016, pp. 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006.

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



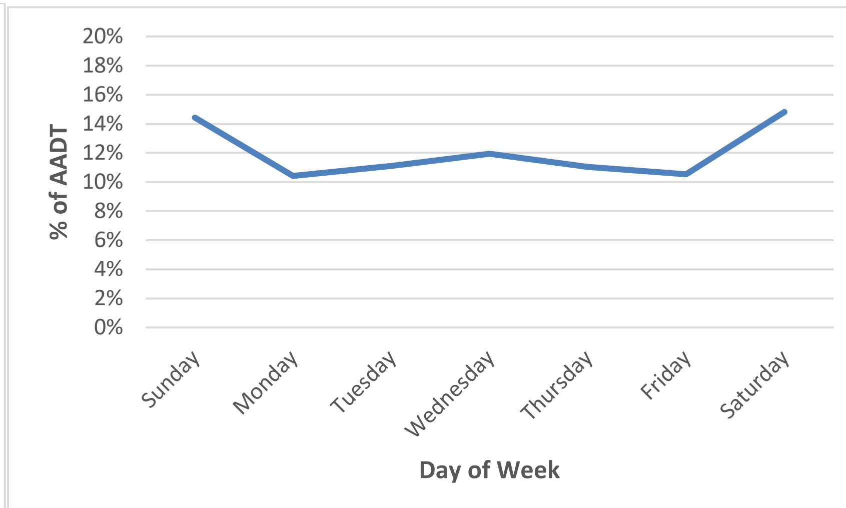




**Figure 3. Bicyclist Traffic Patterns: Johnson Creek Trail @ MoPac/W 5th St./W 6th St. Interchange (Austin).**

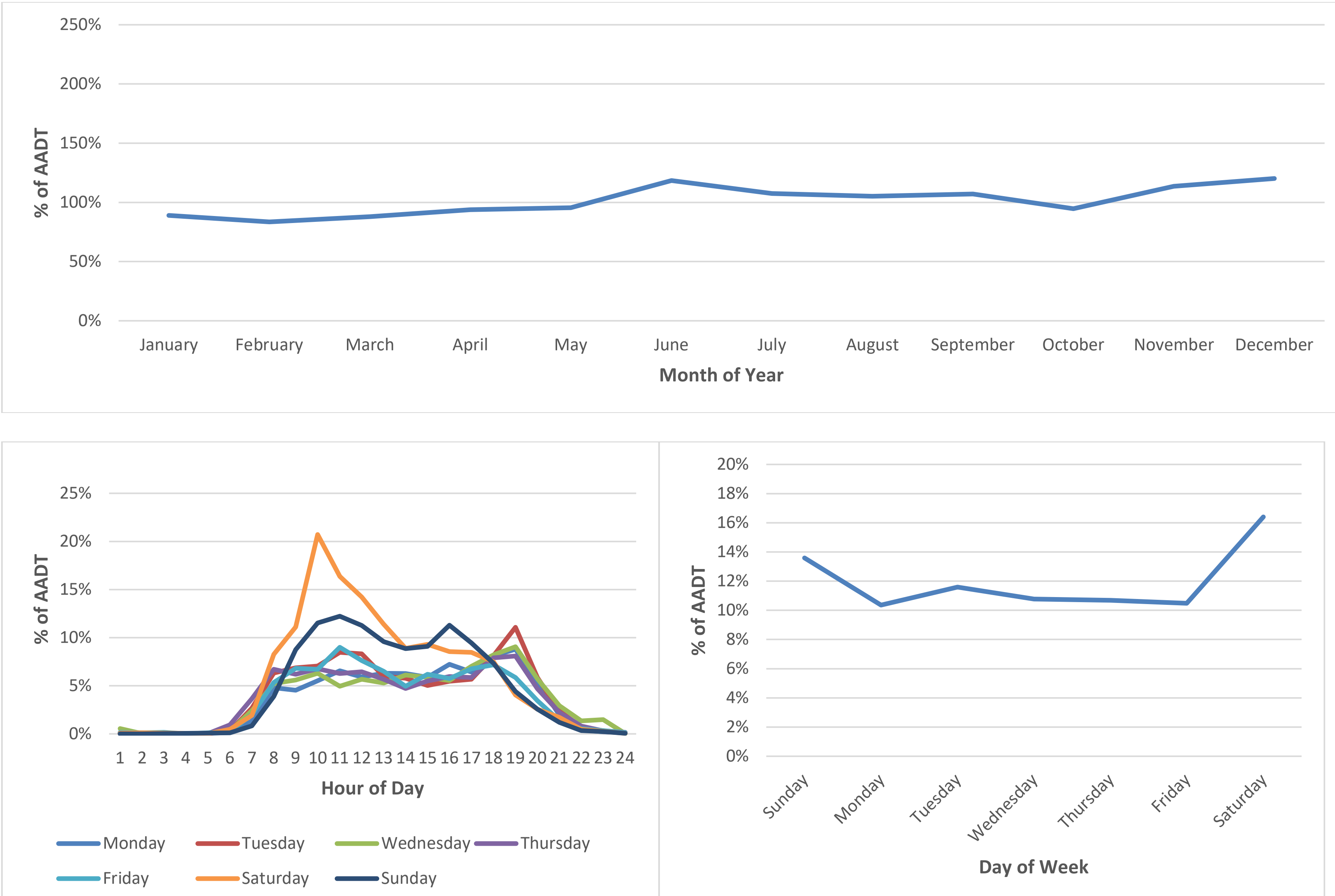



**Figure 4. Pedestrian Traffic Patterns: Johnson Creek Trail @ MoPac/W 5th St./W 6th St. Interchange (Austin).**

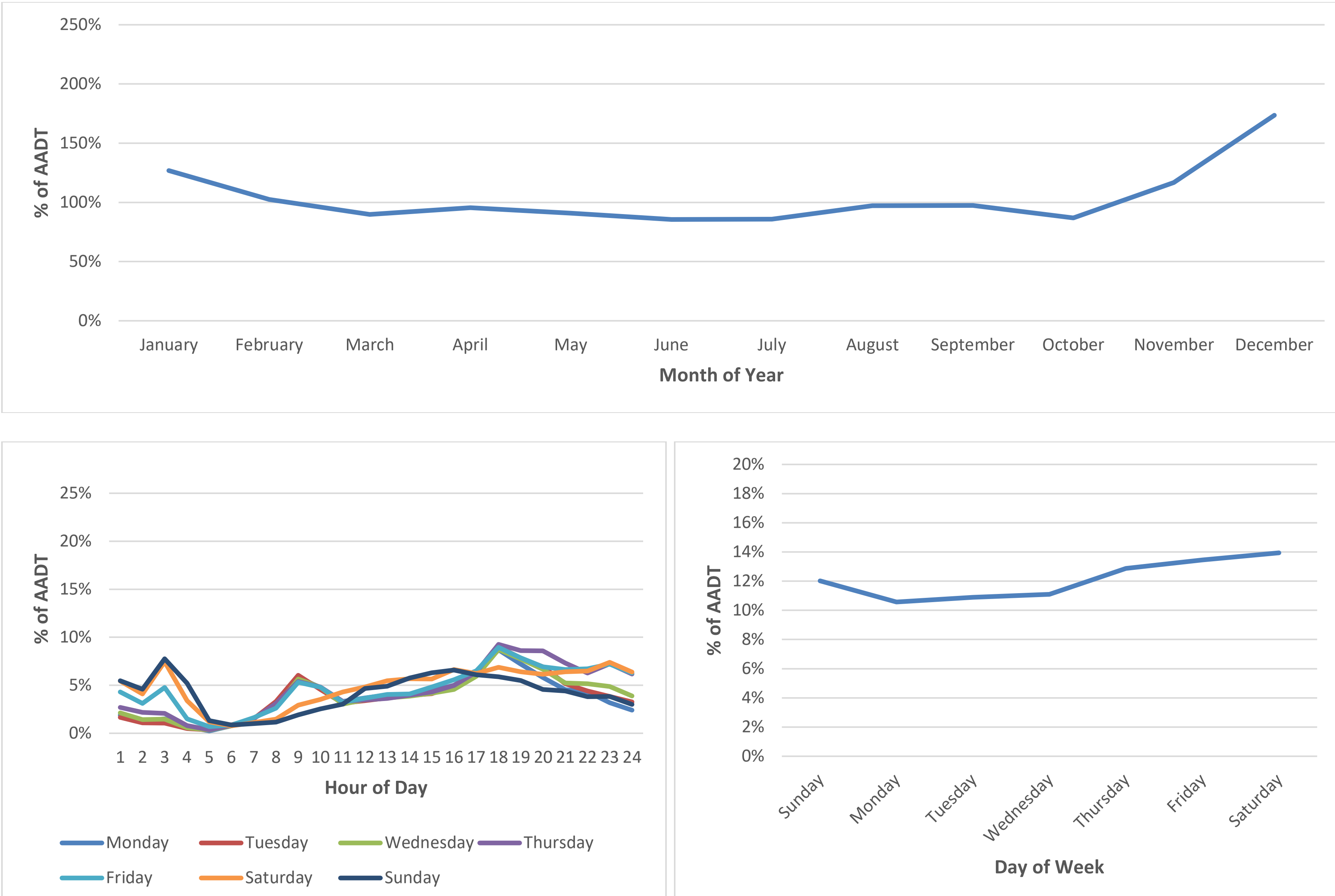



**Figure 5. Bicyclist Traffic Patterns: Lance Armstrong Bikeway @ Waller Creek (Austin).**

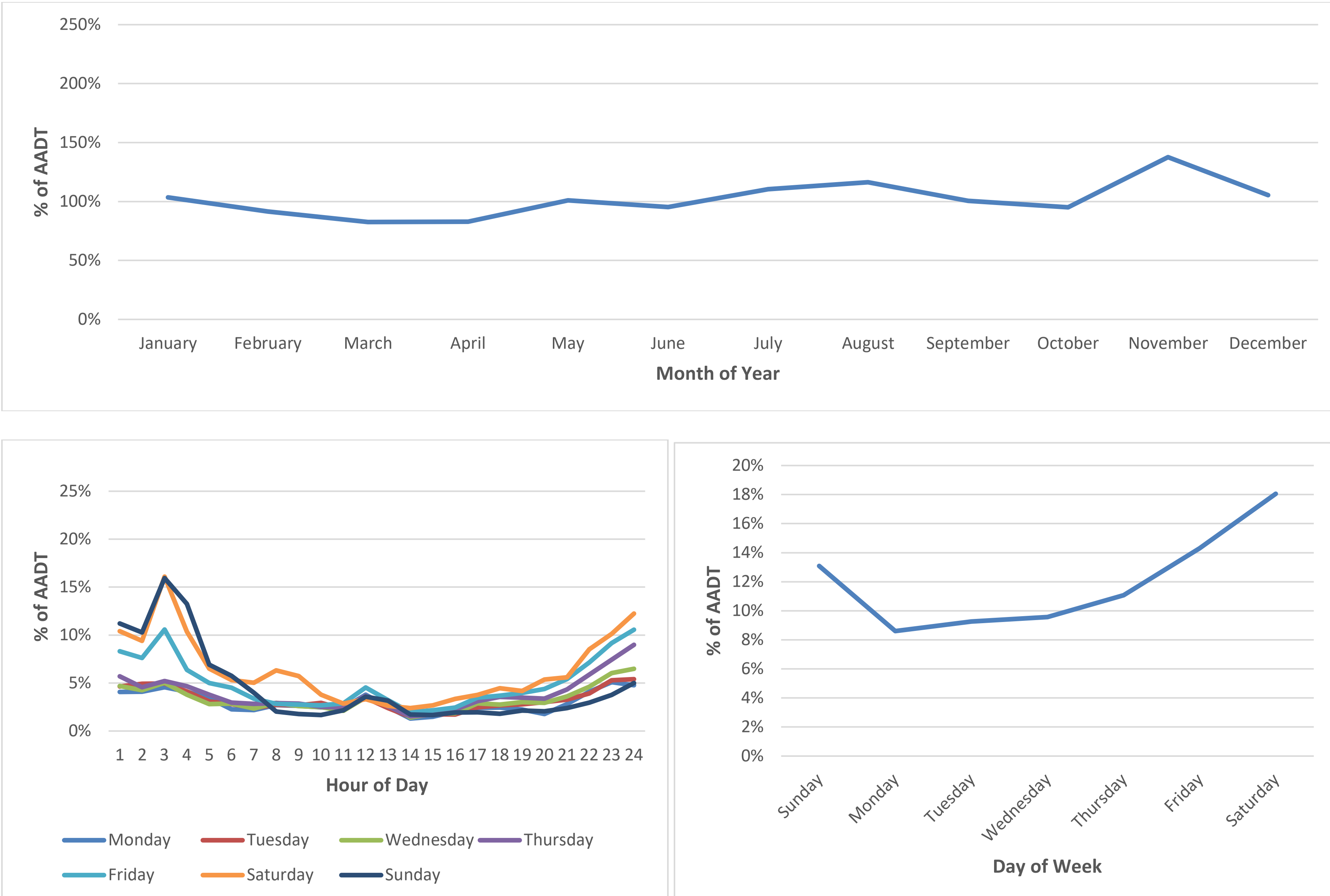



**Figure 6. Pedestrian Traffic Patterns: Lance Armstrong Bikeway @ Waller Creek (Austin).**

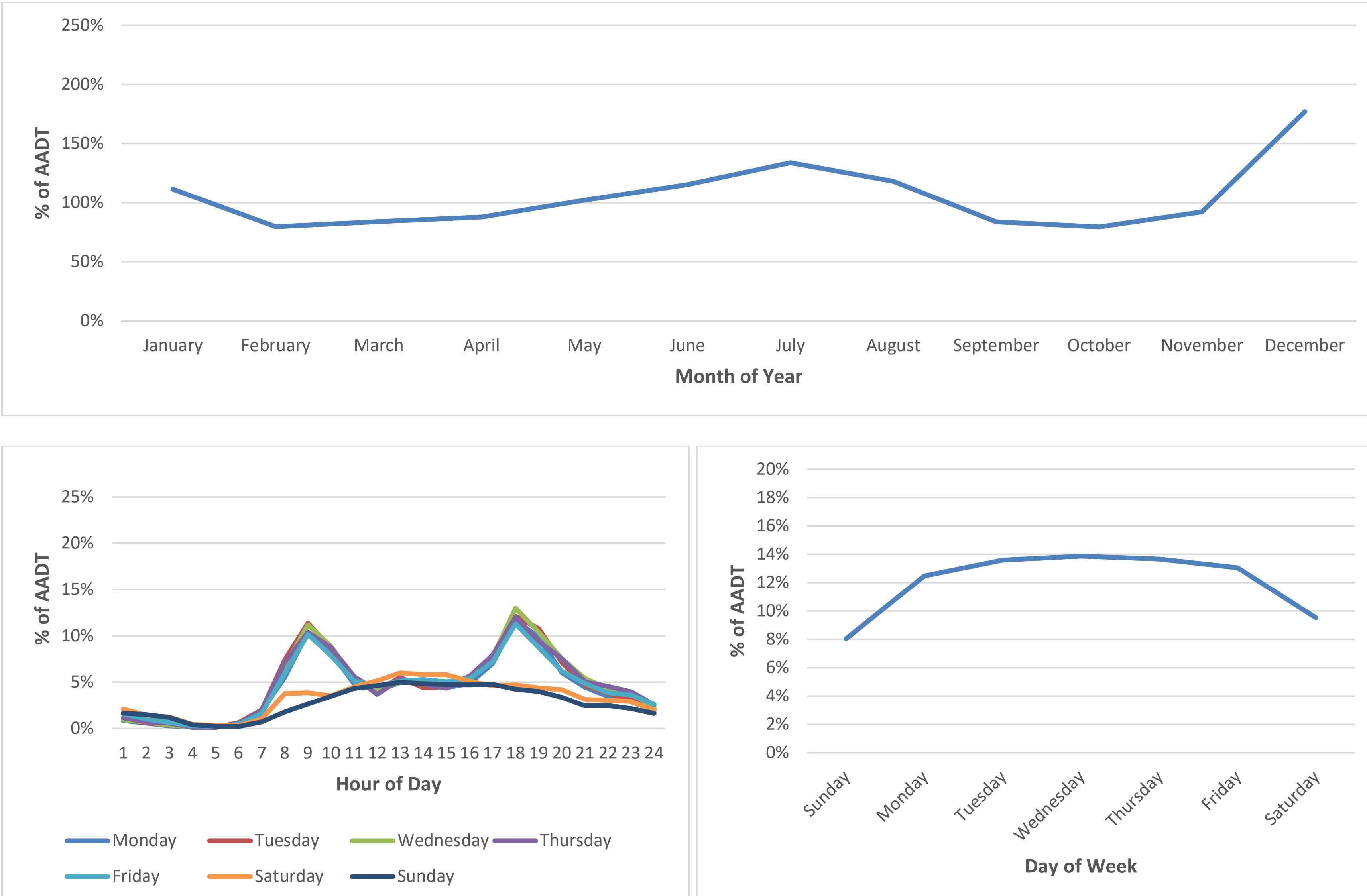
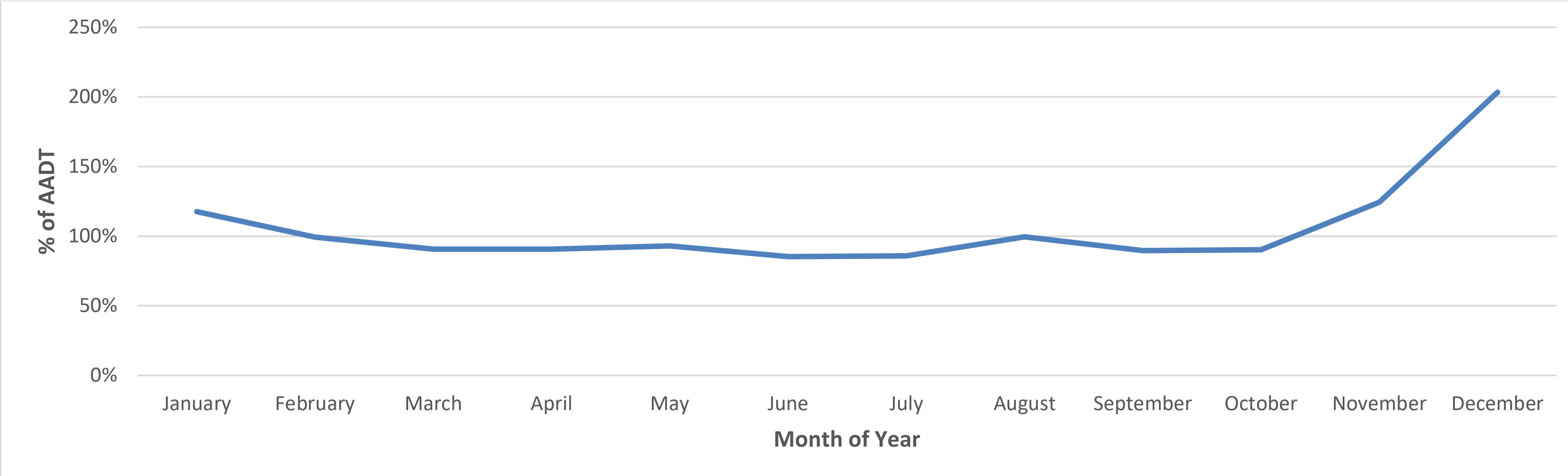
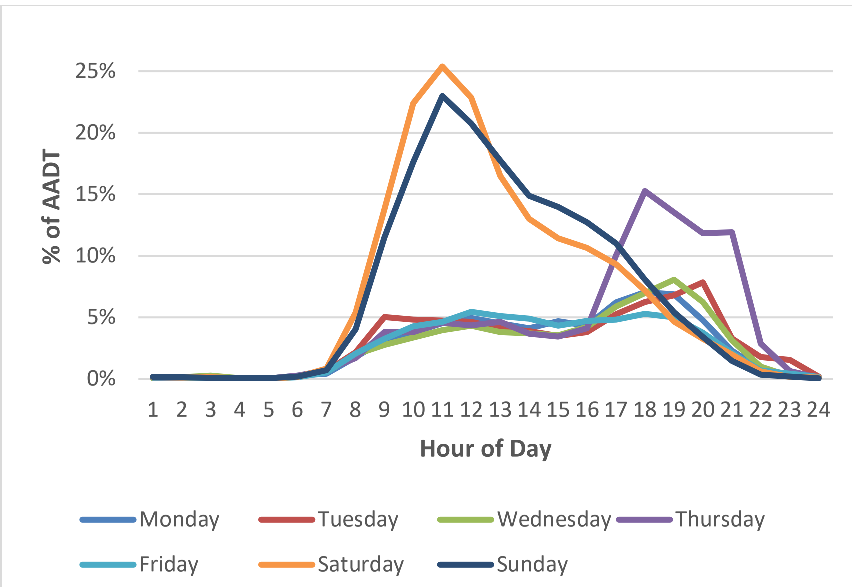



**Figure 7. Bicyclist Traffic Patterns: Manor Rd. @ Alamo St. (Austin).**



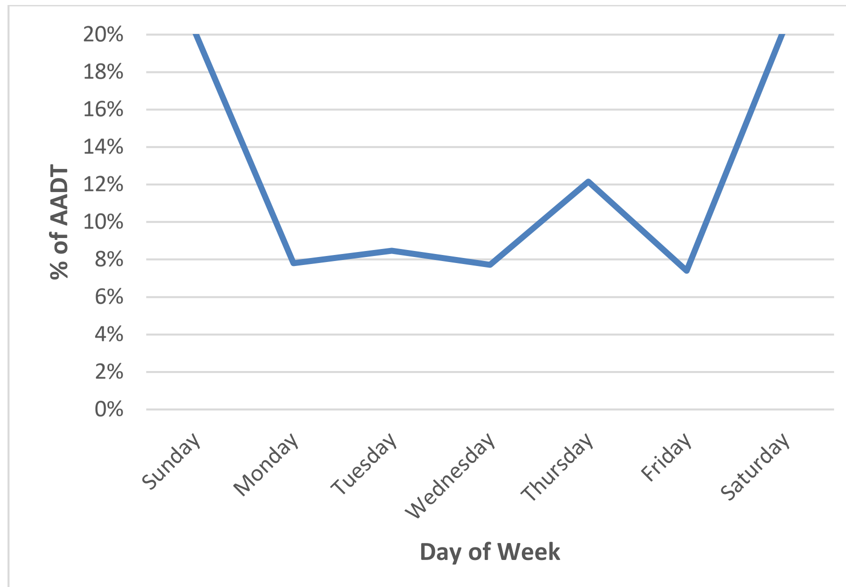




**Figure 8. Bicyclist Traffic Patterns: Walnut Creek Trail N of Jain Ln. (Austin).**

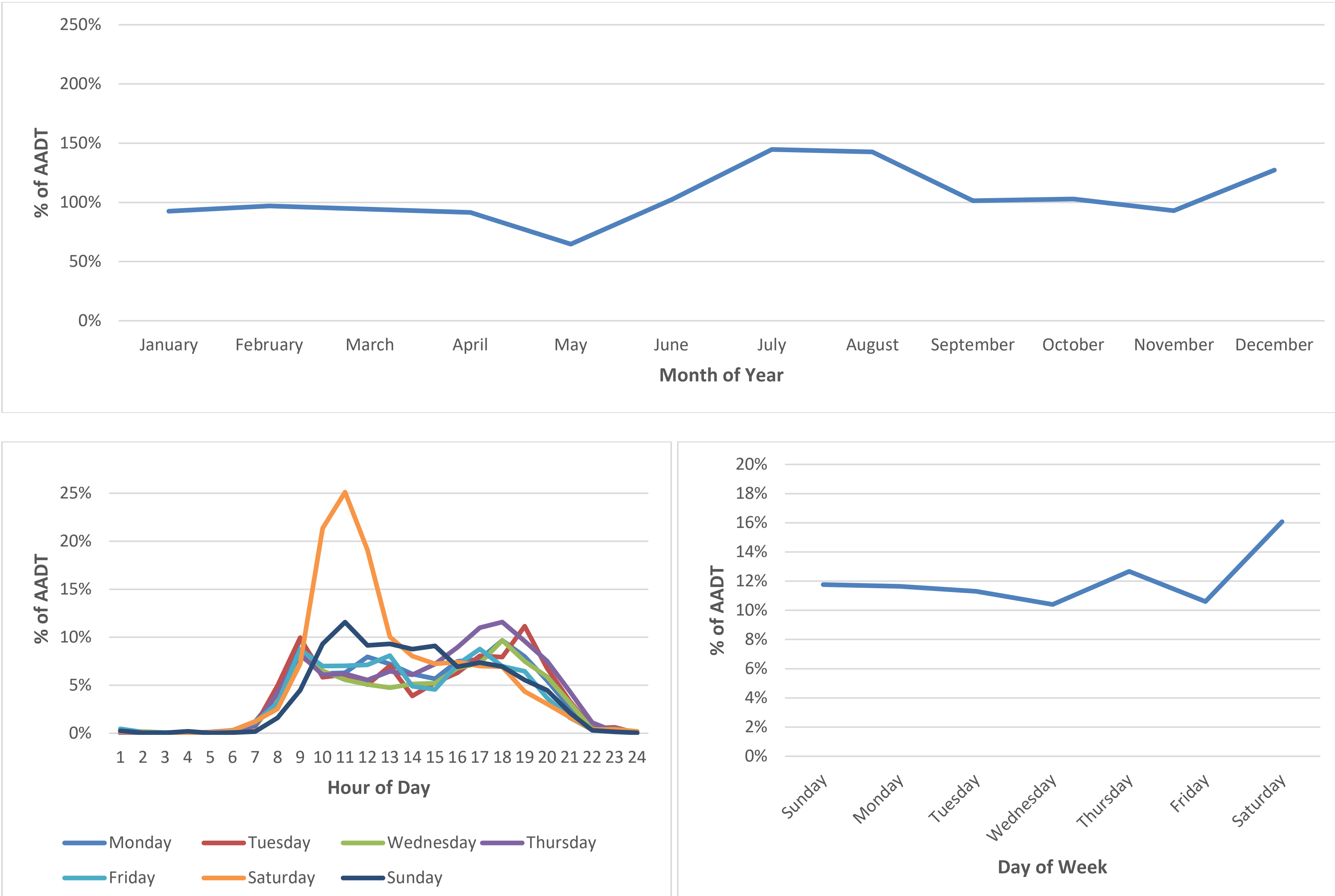



**Figure 9. Pedestrian Traffic Patterns: Walnut Creek Trail N. of Jain Ln. (Austin).**

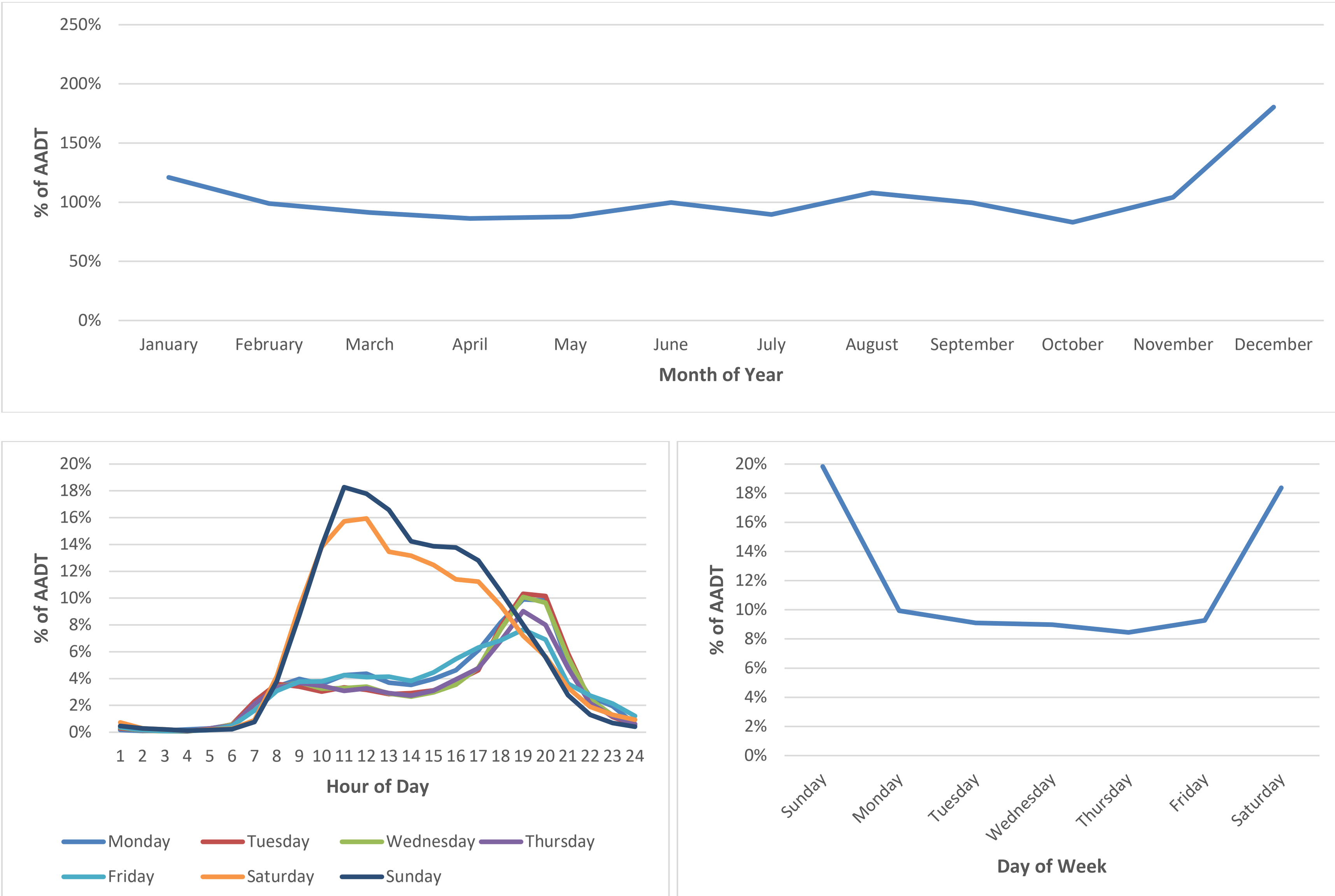



**Figure 10. Bicyclist Traffic Patterns: Heights Trail @ 5 ½ Street (Houston).**

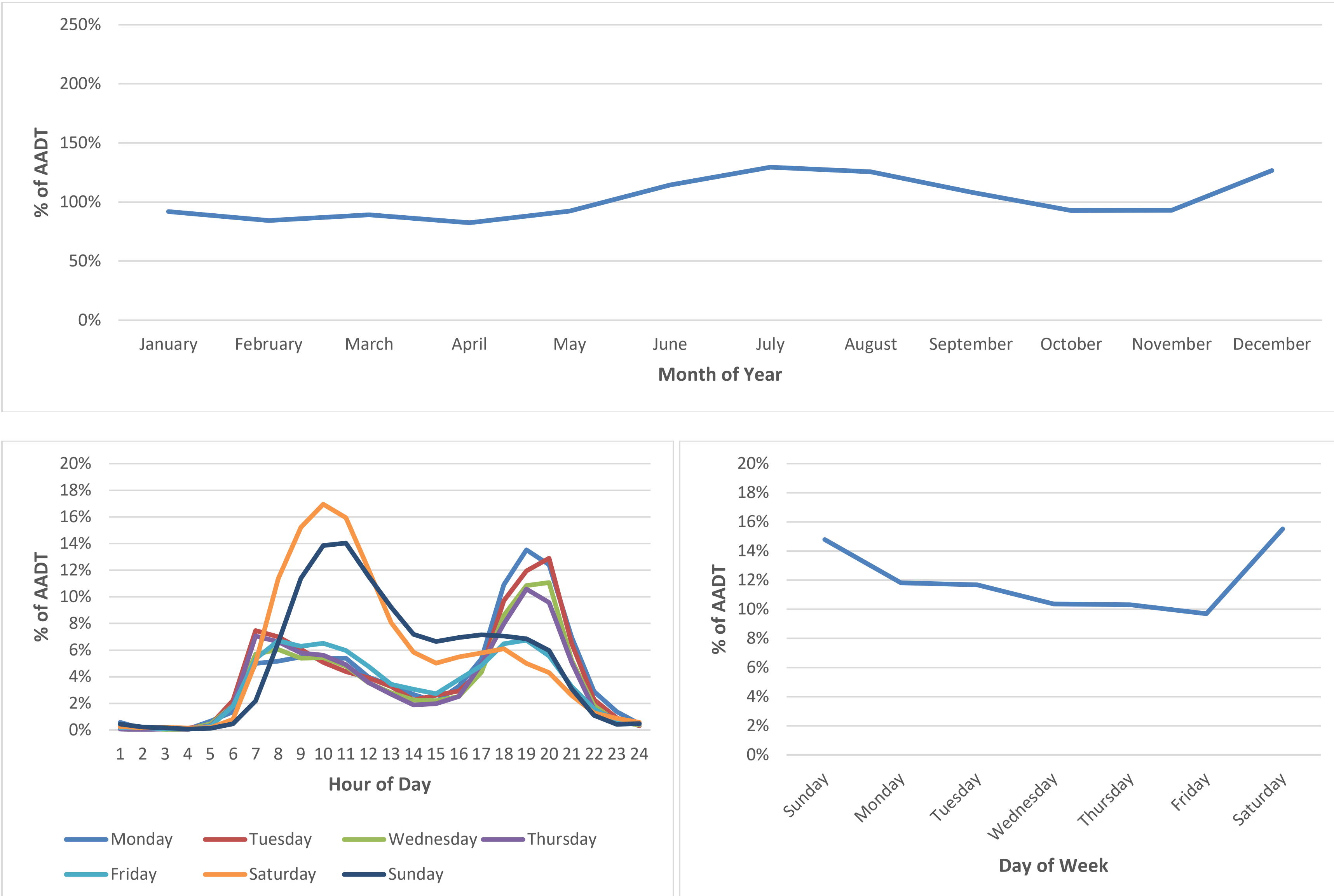
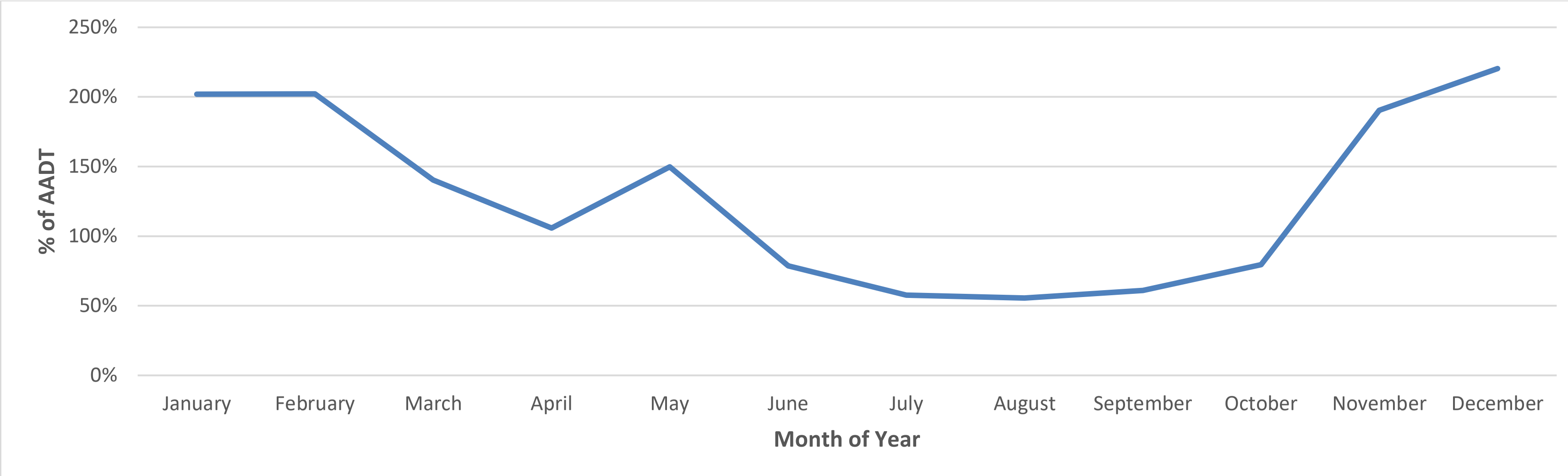
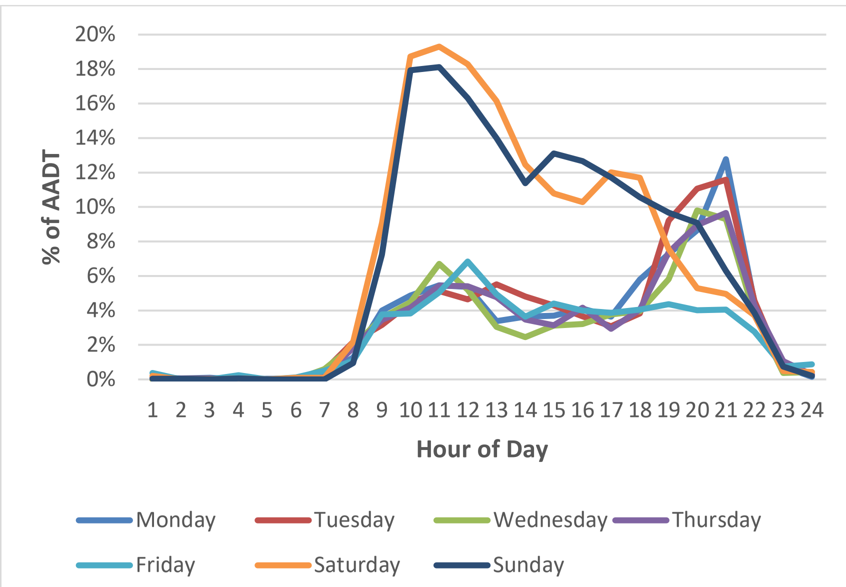



**Figure 11. Pedestrian Traffic Patterns: Heights Trail @ 5 ½ Street (Houston).**



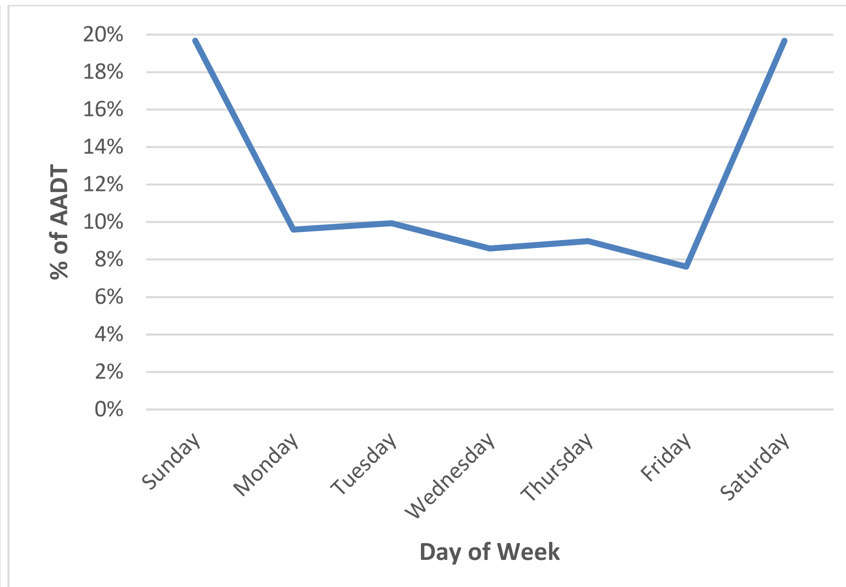




**Figure 12. Bicyclist Traffic Patterns: Bluebonnet Trail at US 75 (Plano).**

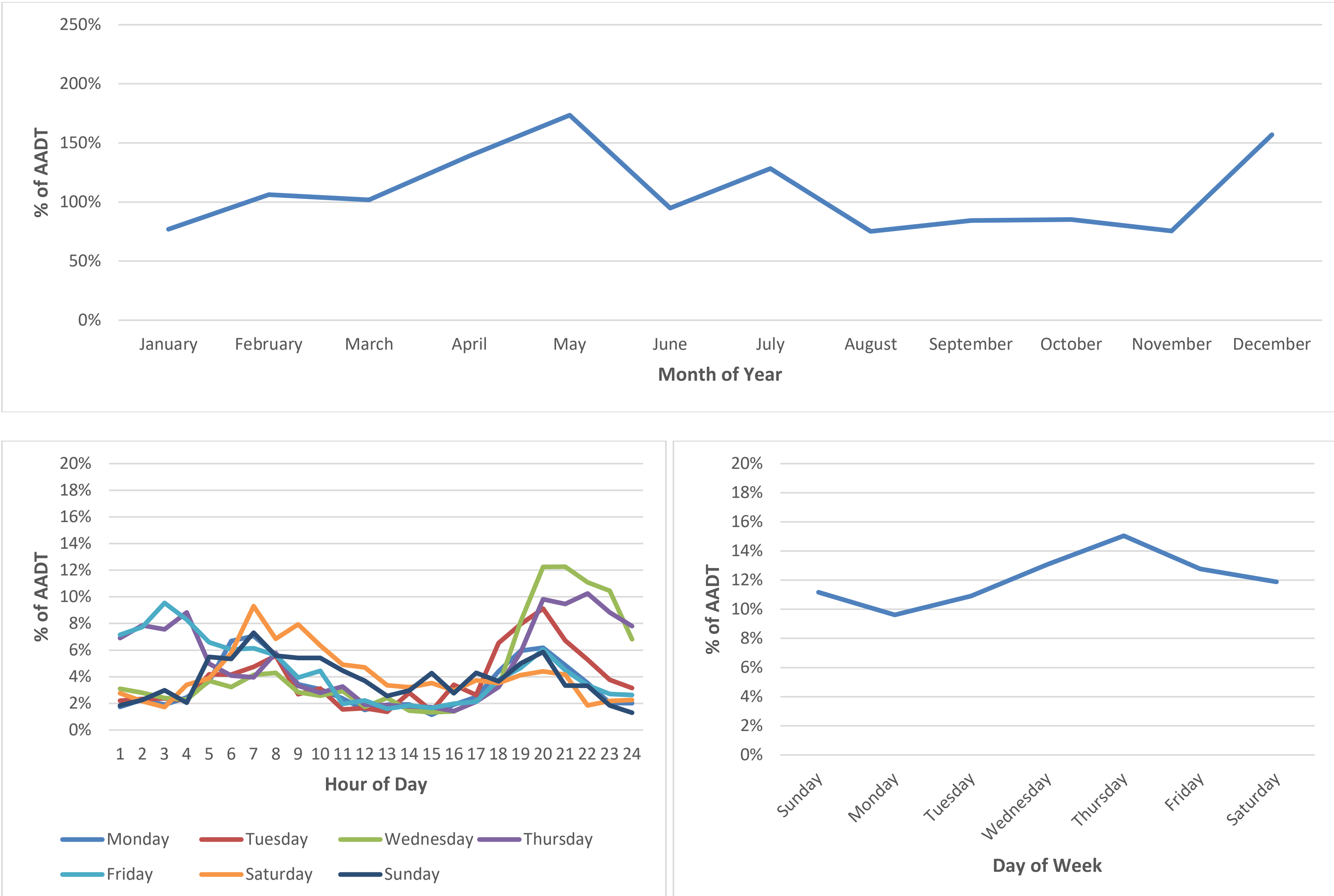
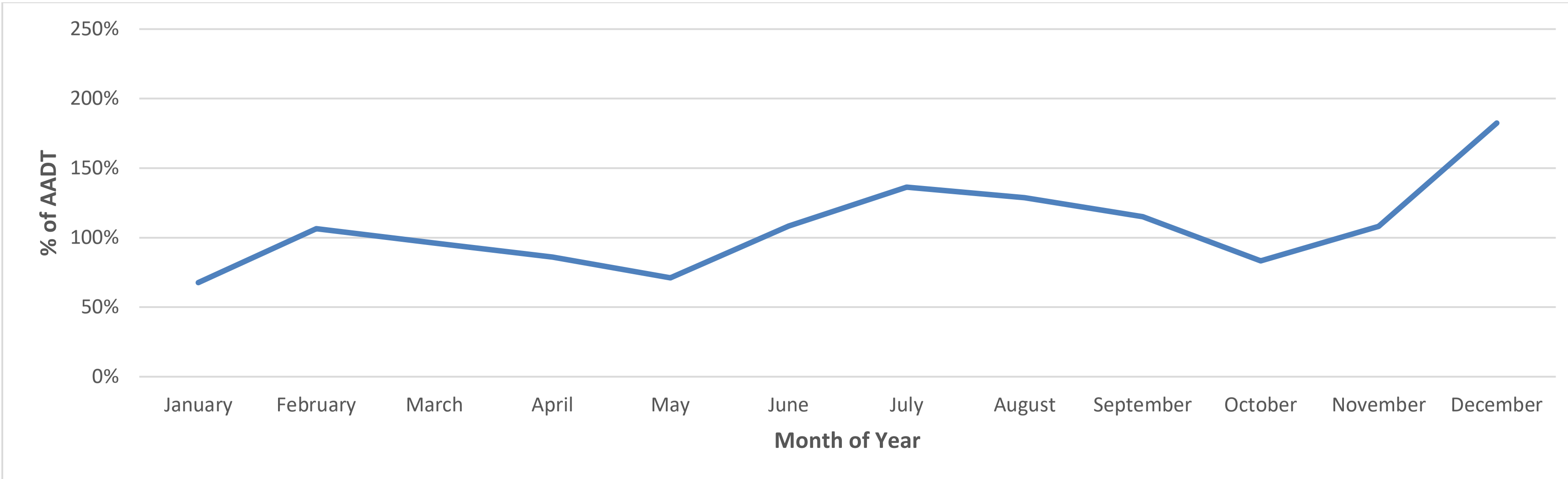
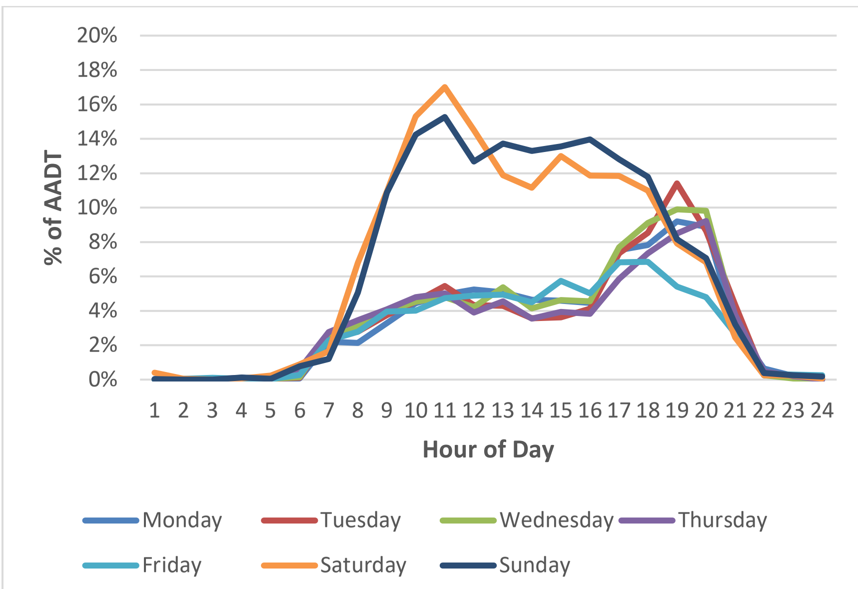



Figure 13. Pedestrian Traffic Patterns: Bluebonnet Trail at US 75 (Plano).



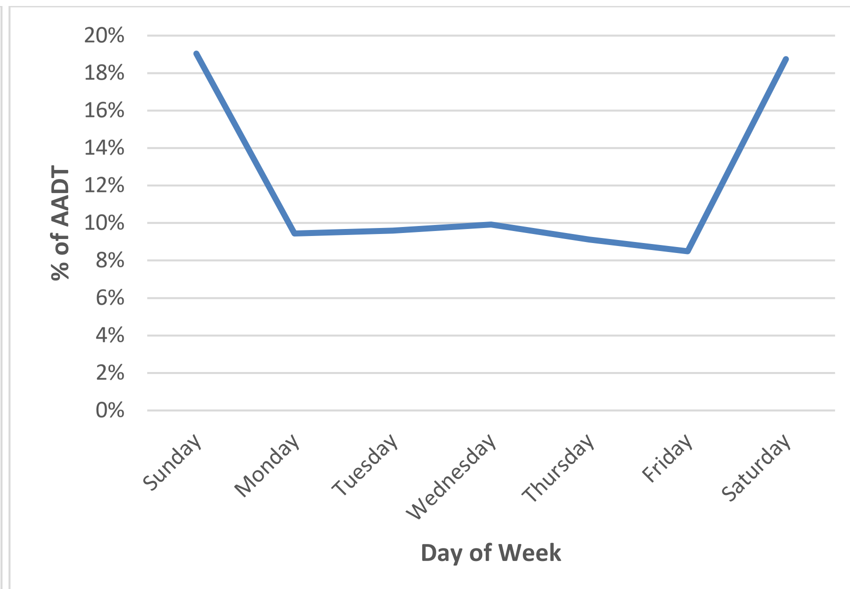
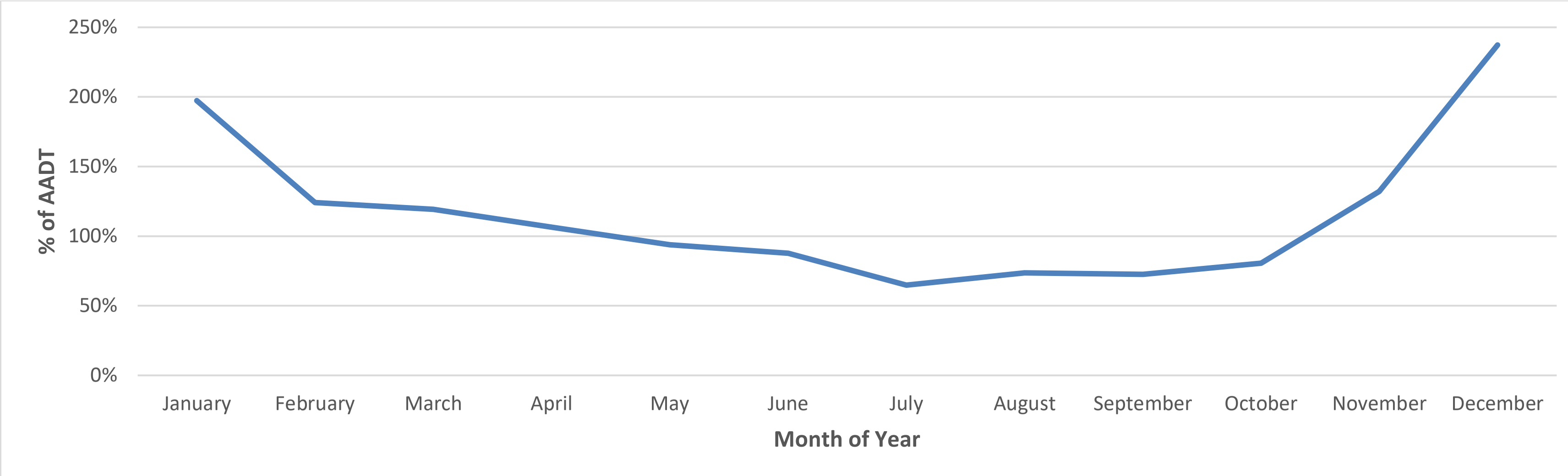
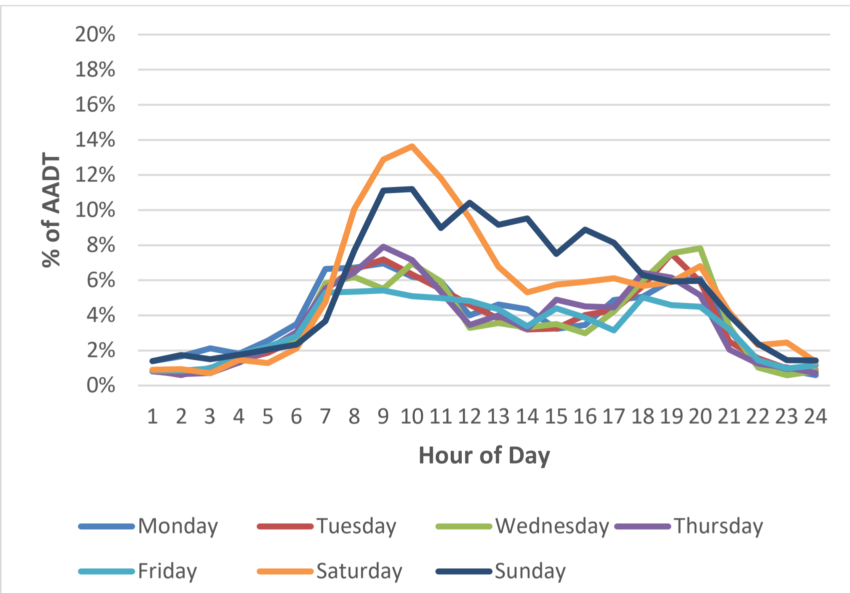




**Figure 14. Bicyclist Traffic Patterns: OPP and NP Trail (Plano).**



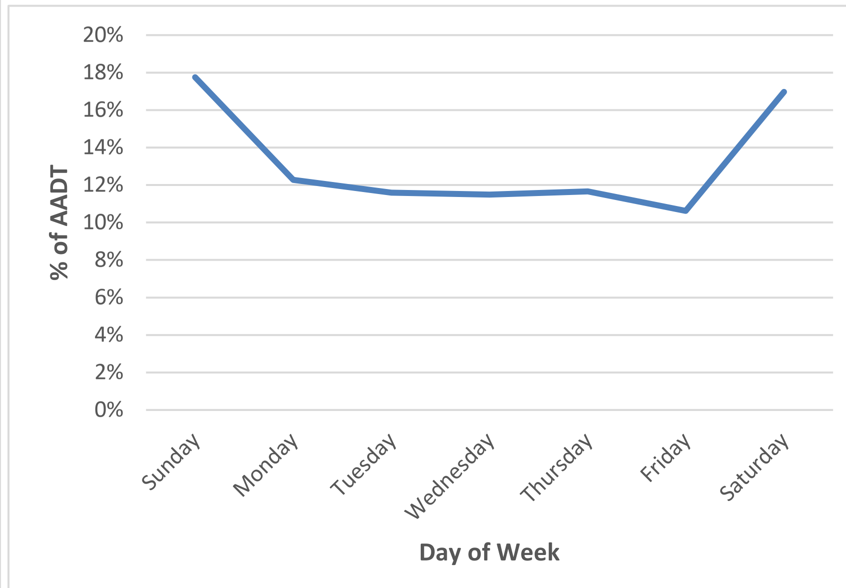




**Figure 15. Pedestrian Traffic Patterns: OPP and NP Trail (Plano).**

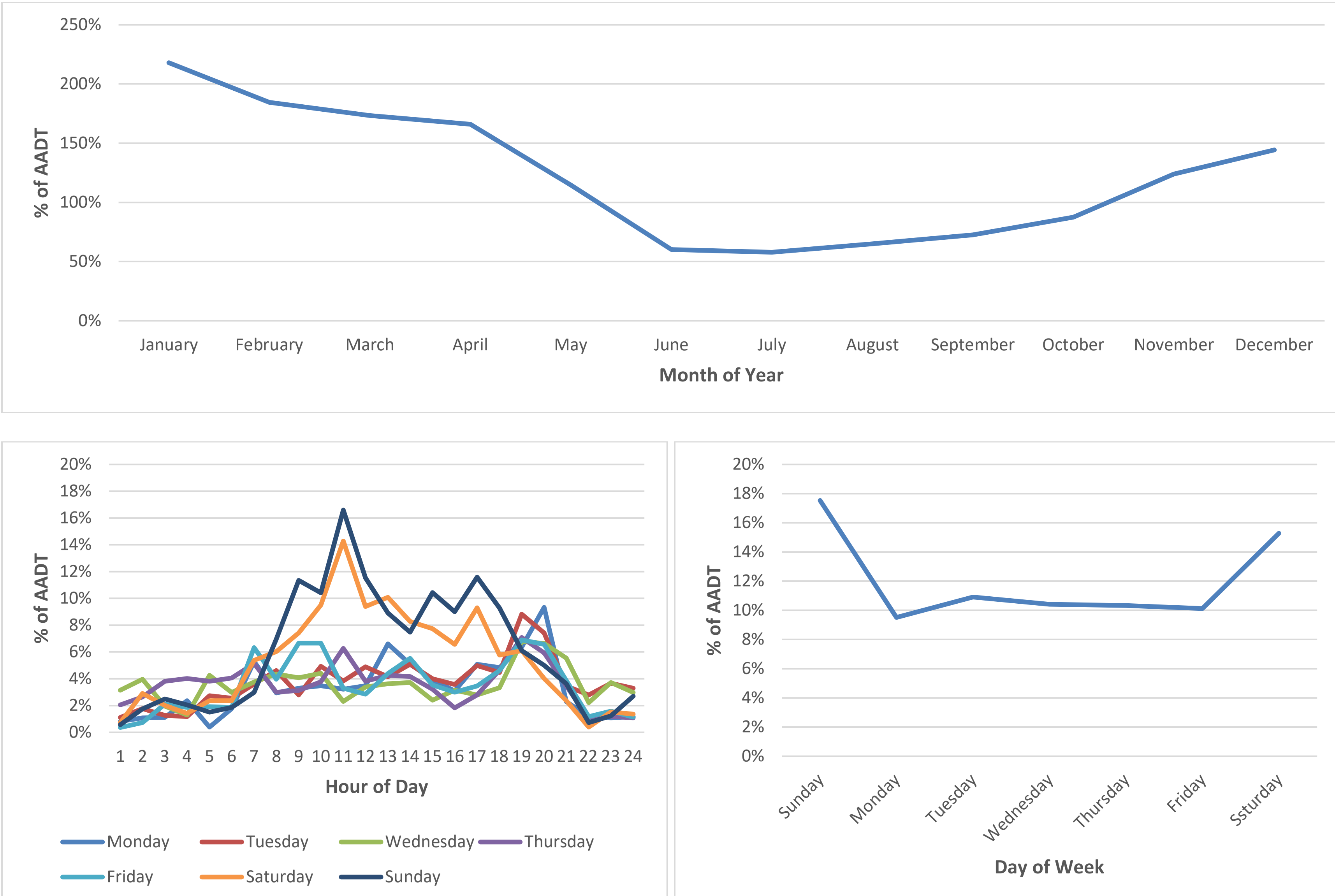



Figure 16. Bicyclist Traffic Patterns: Legacy Trail (Plano).

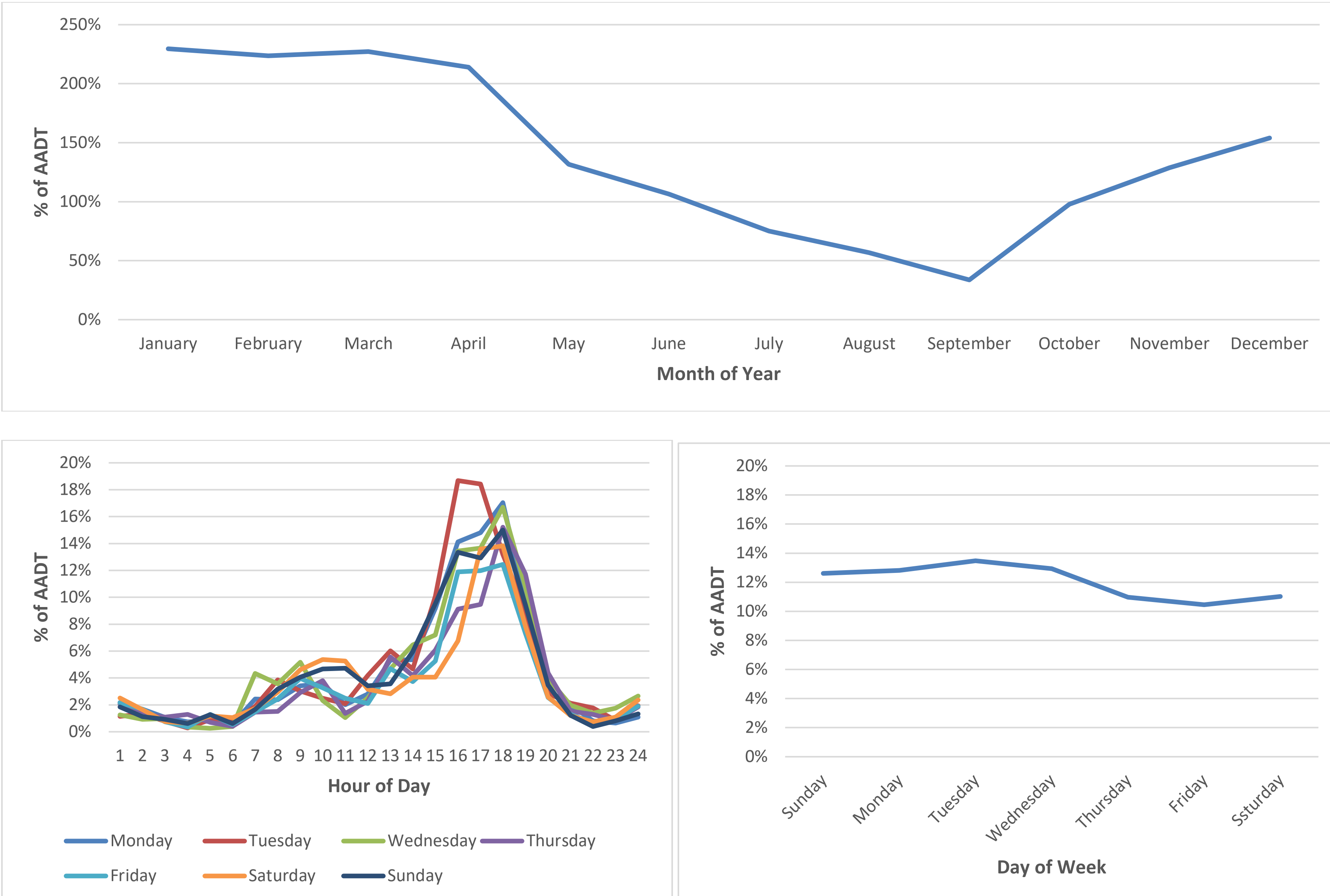



Figure 17. Pedestrian Traffic Patterns: Legacy Trail (Plano).

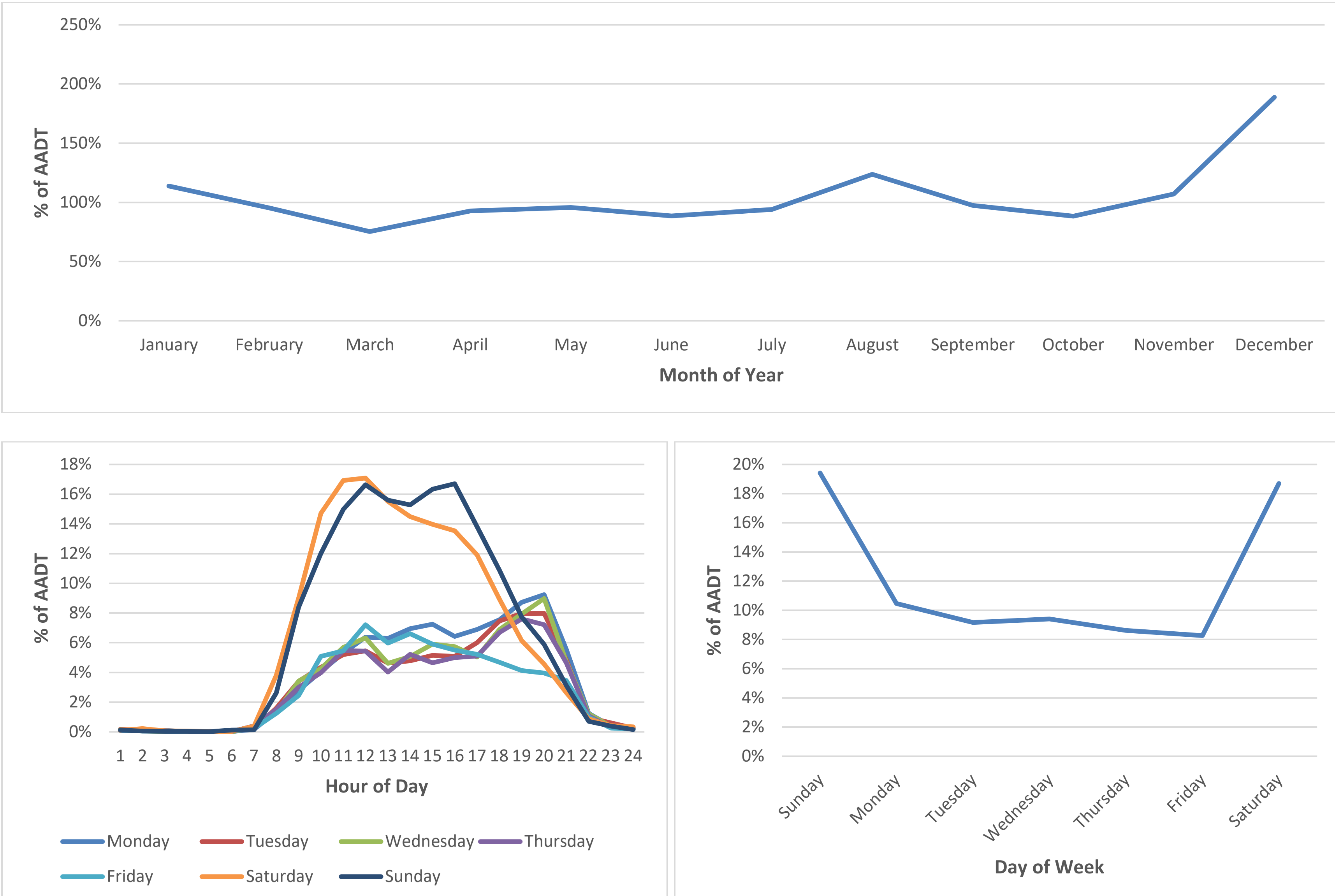



**Figure 18. Bicyclist Traffic Patterns: Mission Reach Trail South of Theo Ave. (San Antonio).**

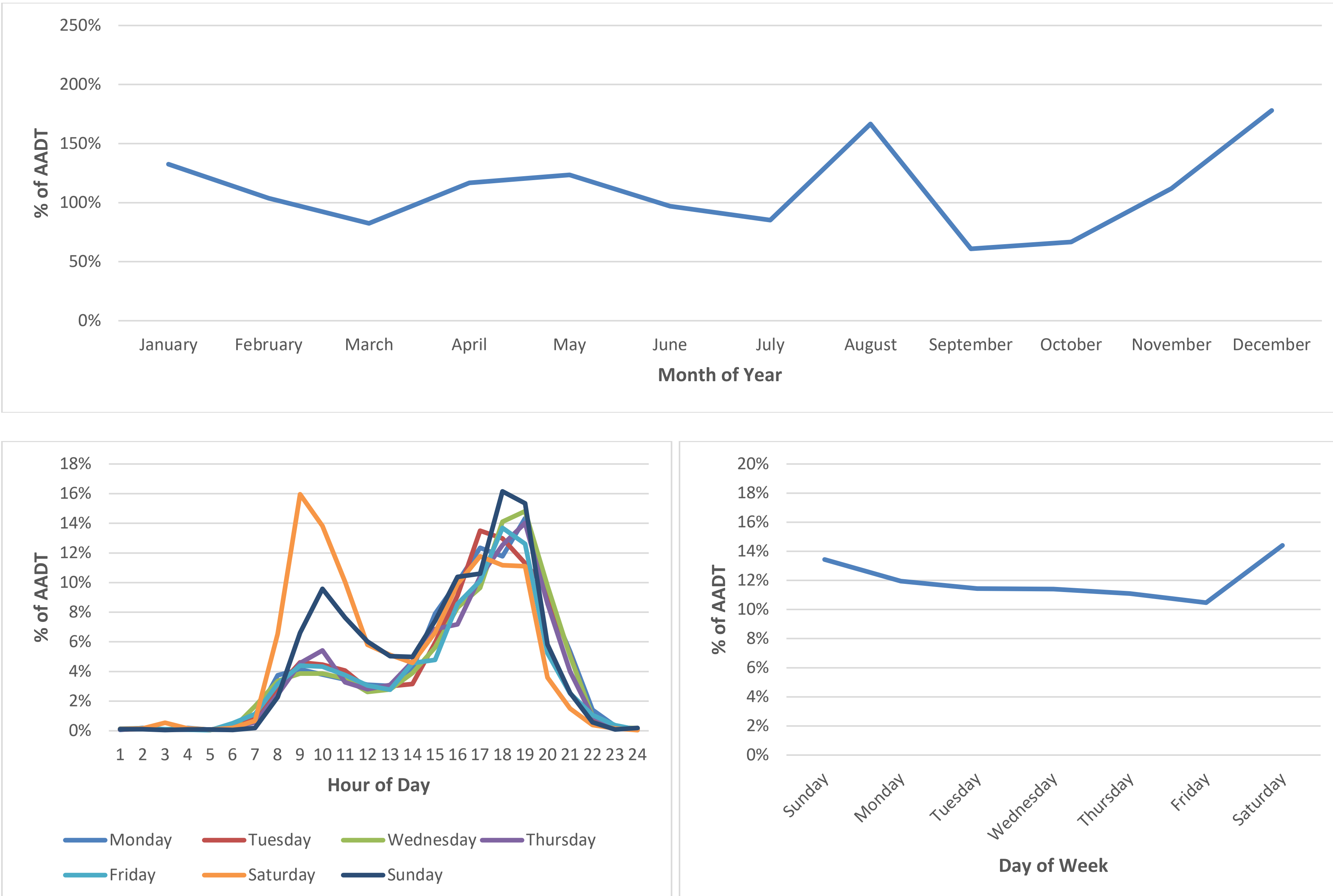



Figure 19. Pedestrian Traffic Patterns: Mission Reach Trail South of Theo Ave. (San Antonio).

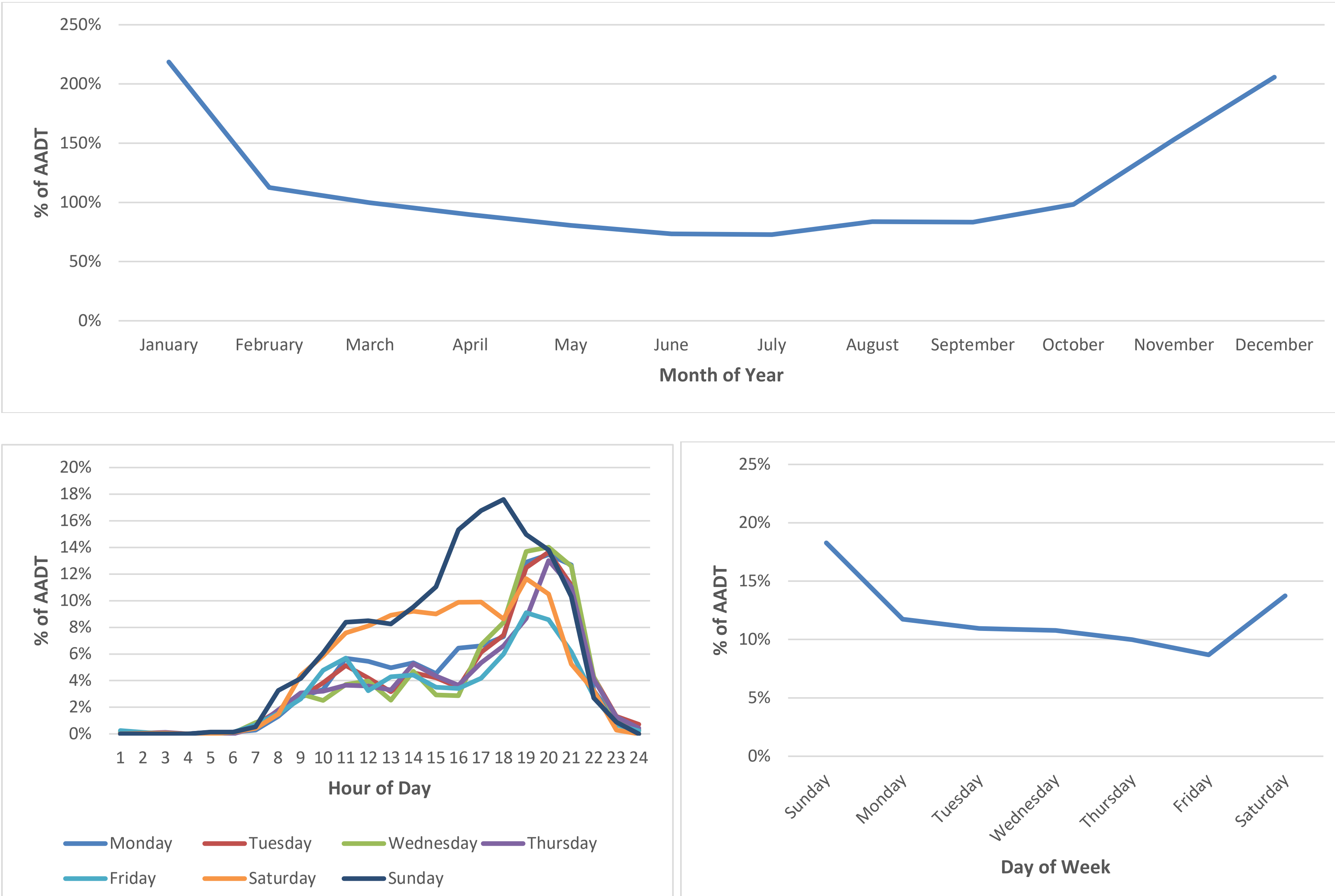



**Figure 20. Bicyclist Traffic Patterns: Bachman Lake/W North West Highway (Dallas).**

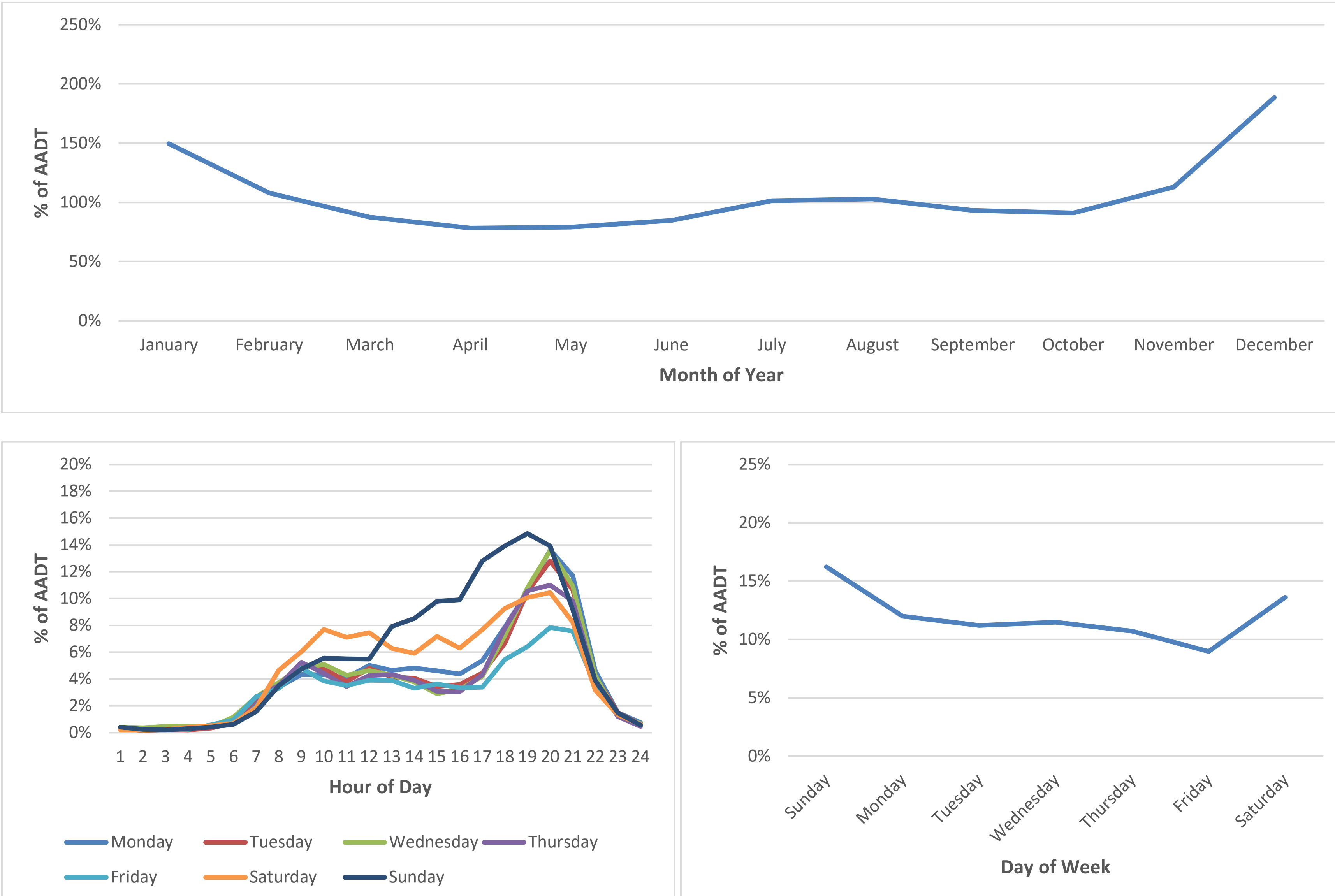



**Figure 21. Pedestrian Traffic Patterns: Bachman Lake/W North West Highway (Dallas).**

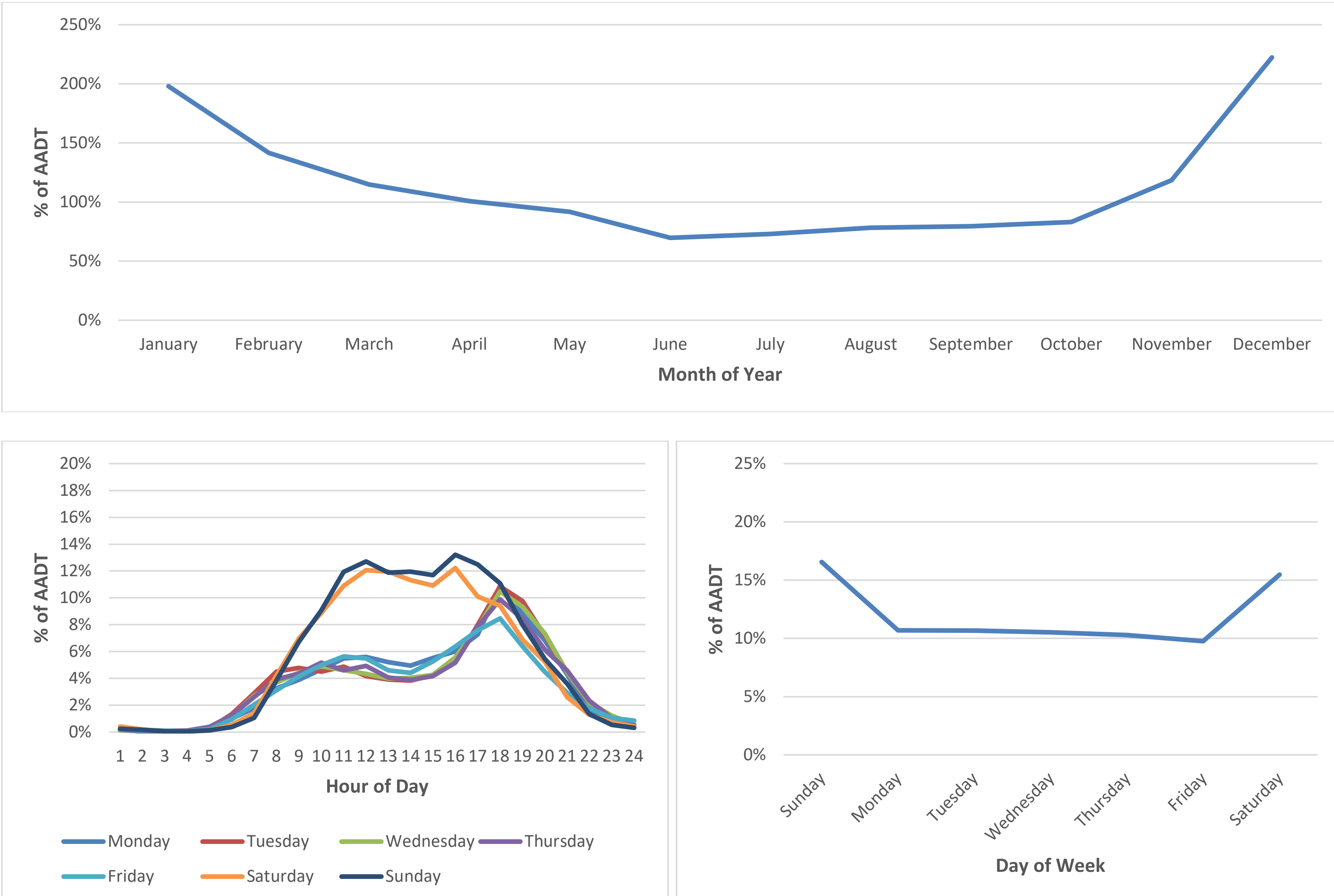



**Figure 22. Bicyclist Traffic Patterns: Katy Trail at Cedar Springs Rd. (Dallas).**

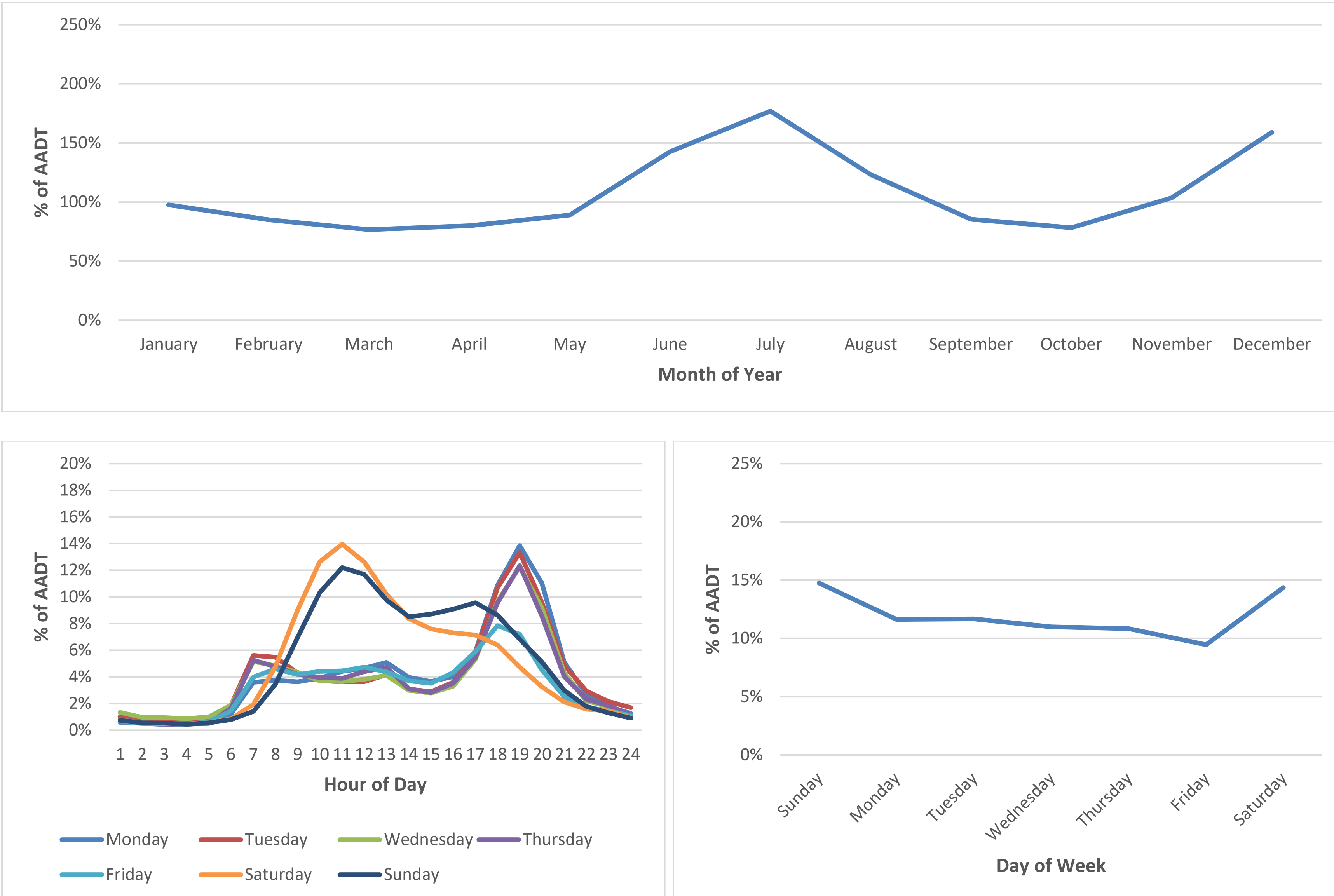



**Figure 23. Pedestrian Traffic Patterns: Katy Trail at Cedar Springs Rd. (Dallas).**

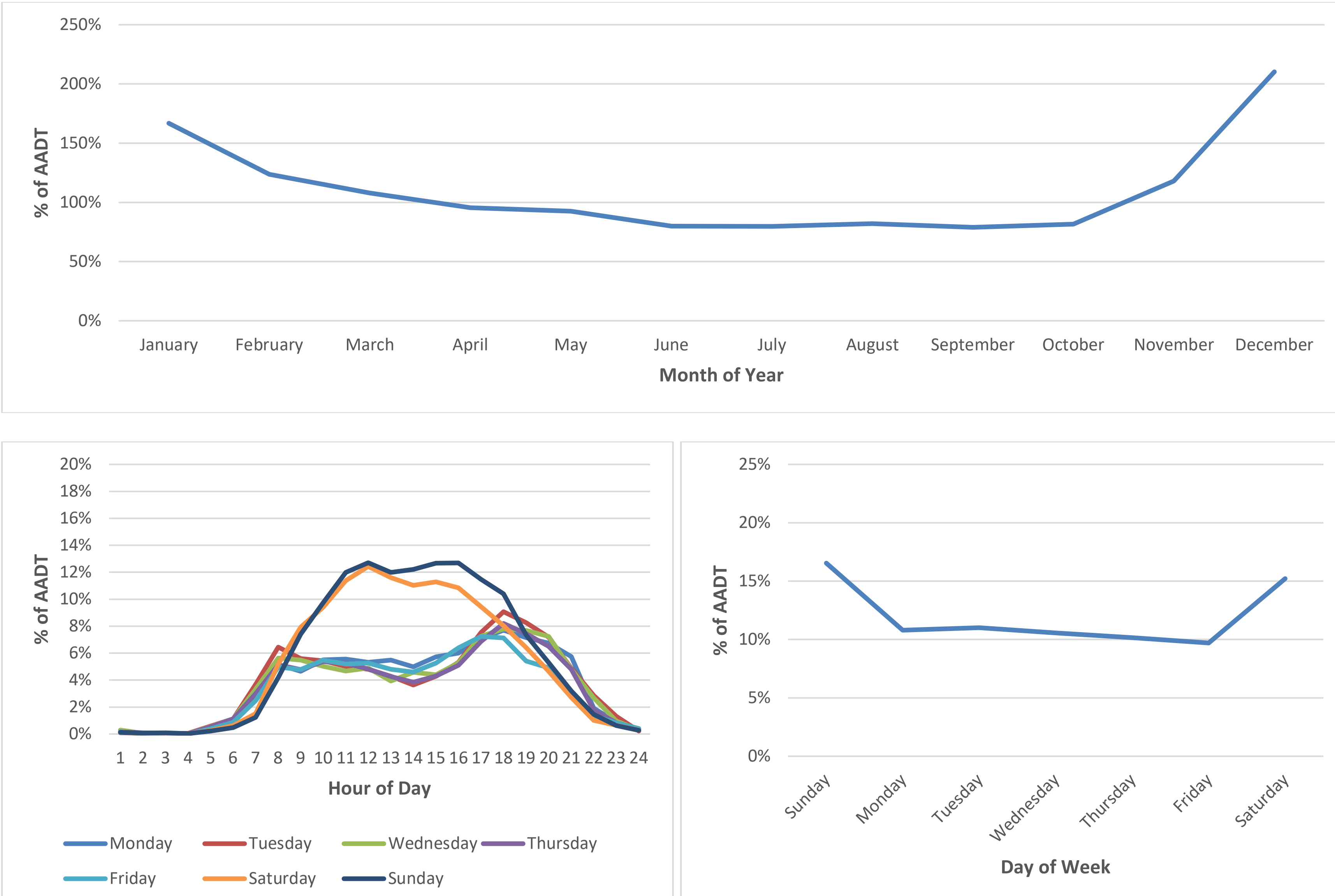



**Figure 24. Bicyclist Traffic Patterns: Katy Trail at Fitzhugh (Dallas).**

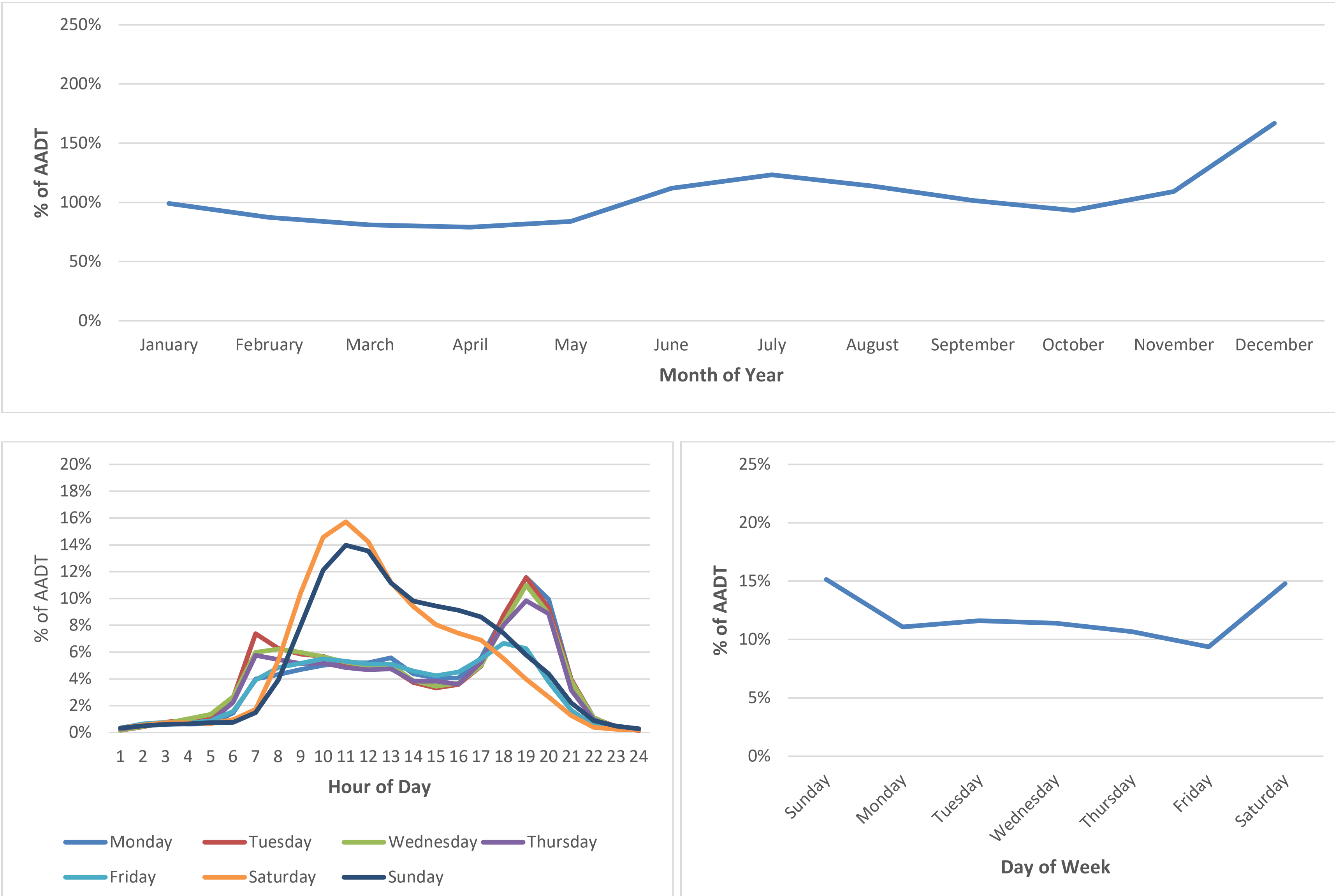



**Figure 25. Pedestrian Traffic Patterns: Katy Trail at Fitzhugh (Dallas).**

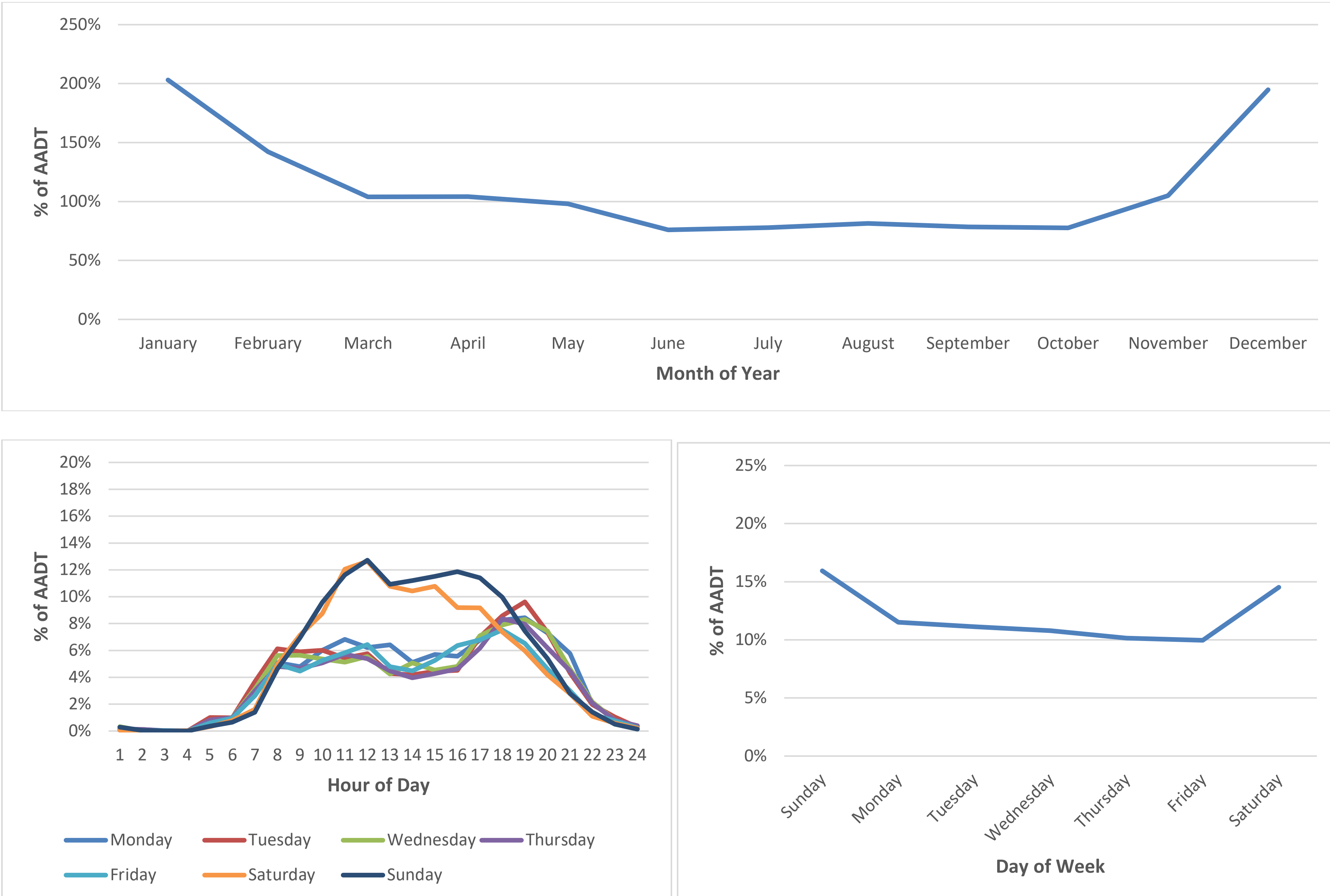



**Figure 26. Bicyclist Traffic Patterns: Katy Trail at Harvard Avenue (Dallas).**

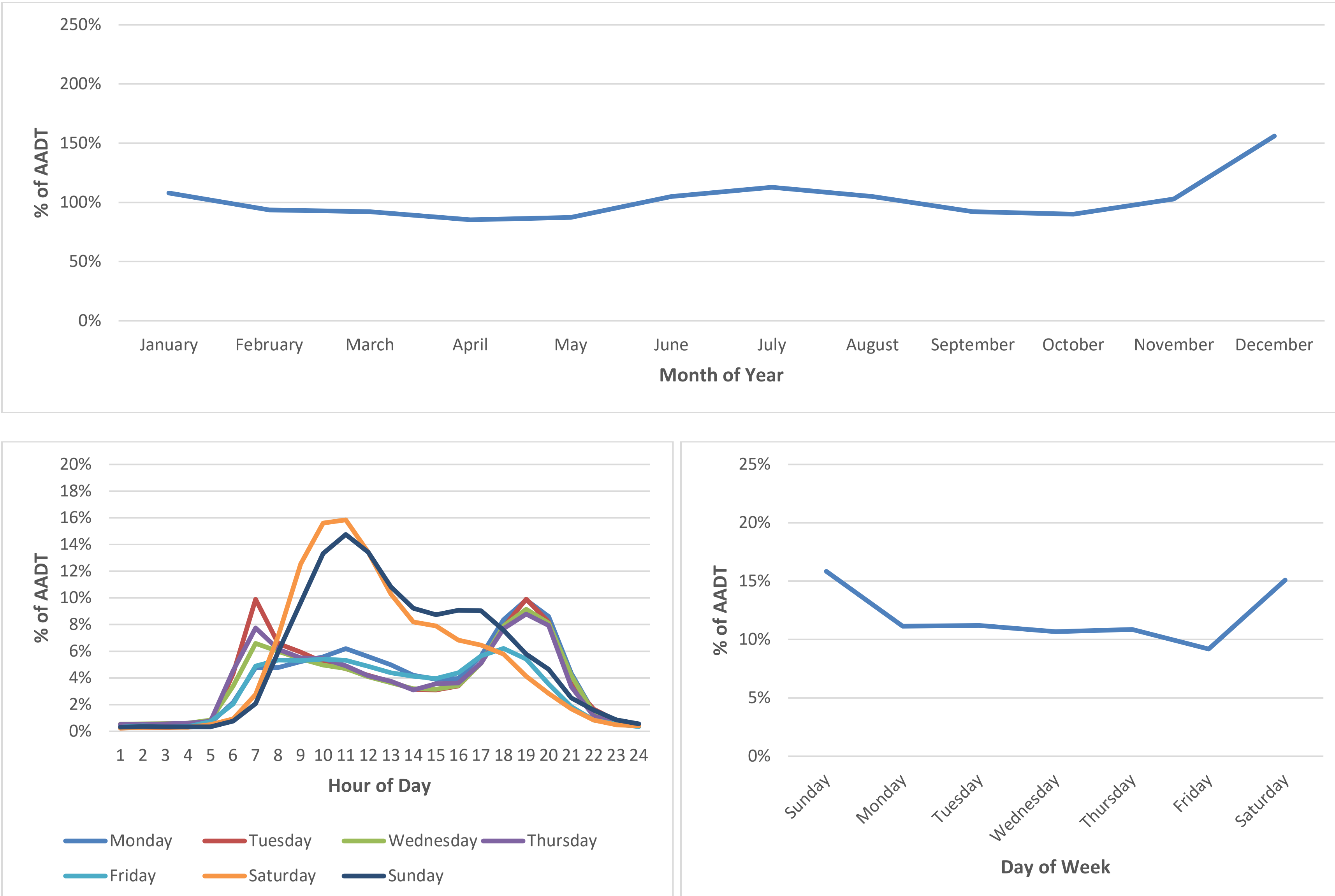
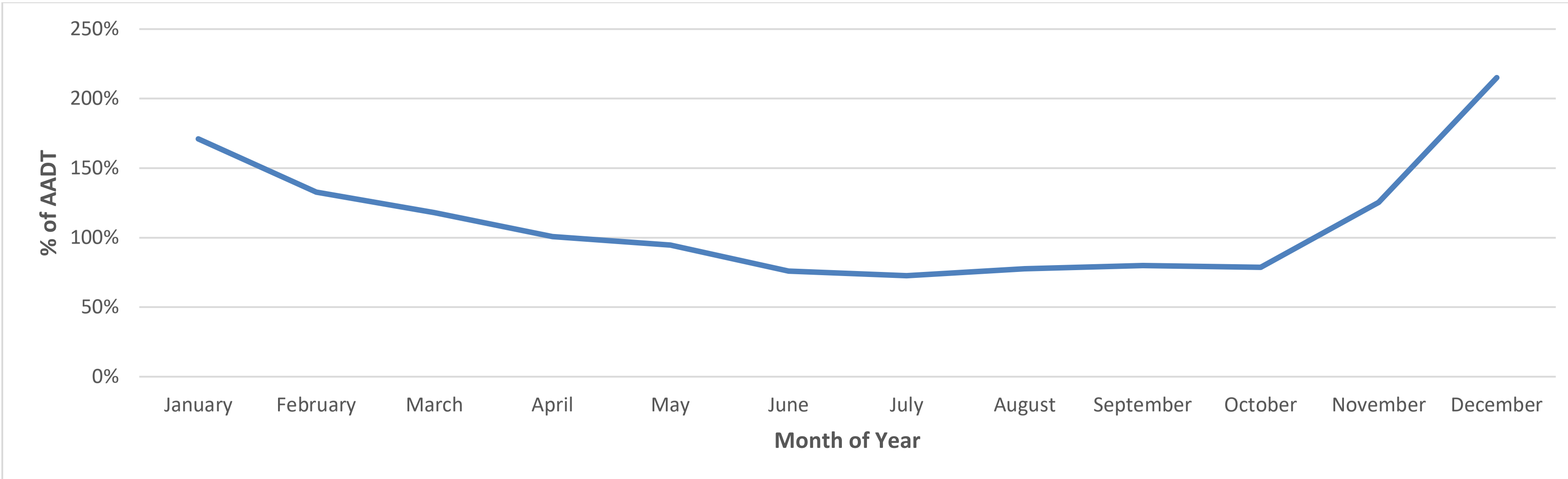
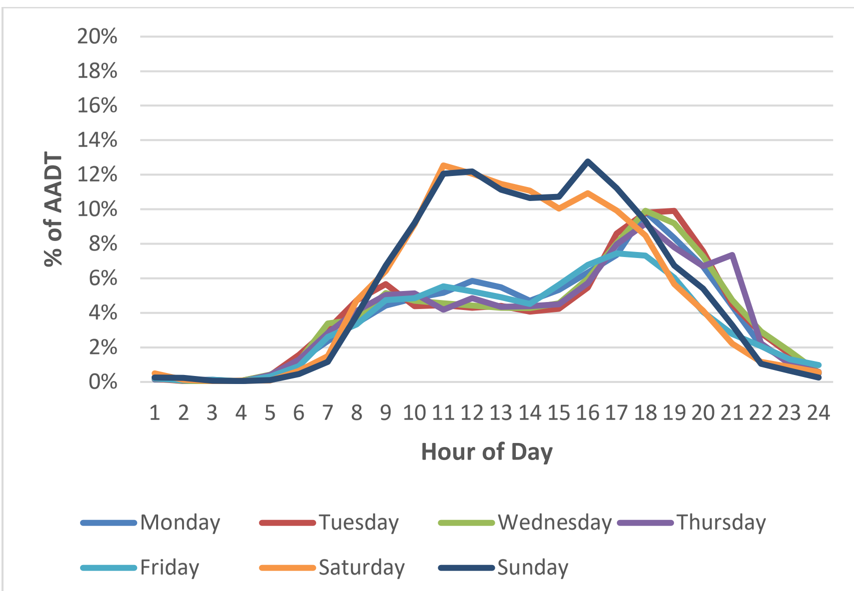



Figure 27. Pedestrian Traffic Patterns: Katy Trail at Harvard Avenue (Dallas).



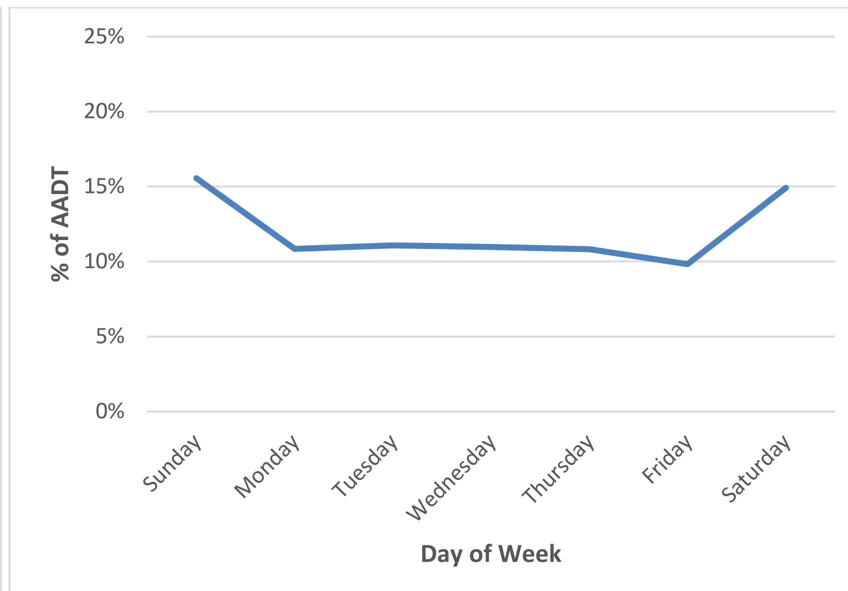
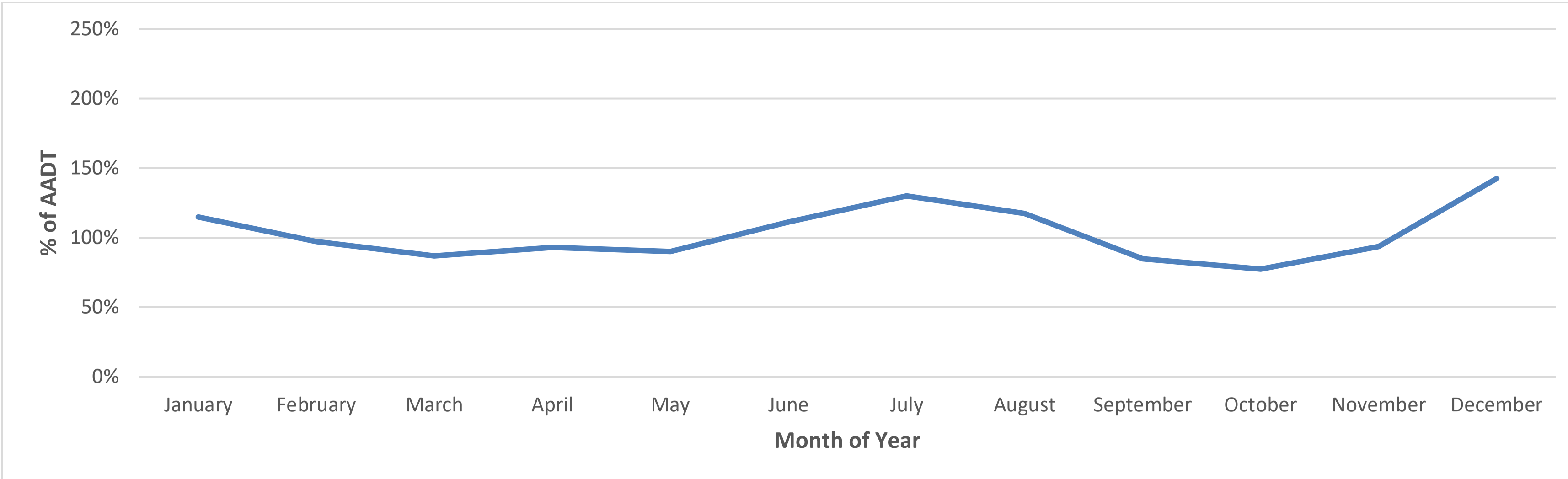
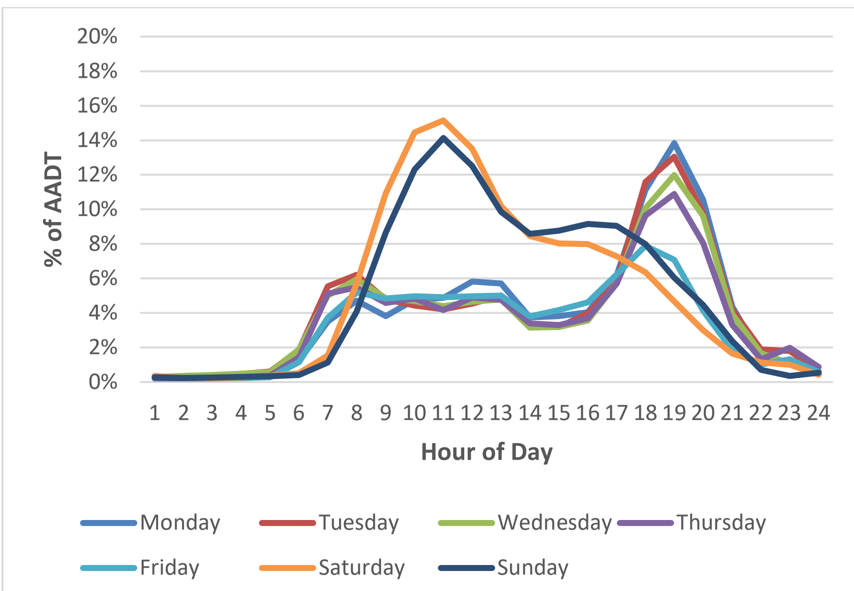




**Figure 28. Bicyclist Traffic Patterns: Katy Trail (Houston Street/AA Center) (Dallas).**



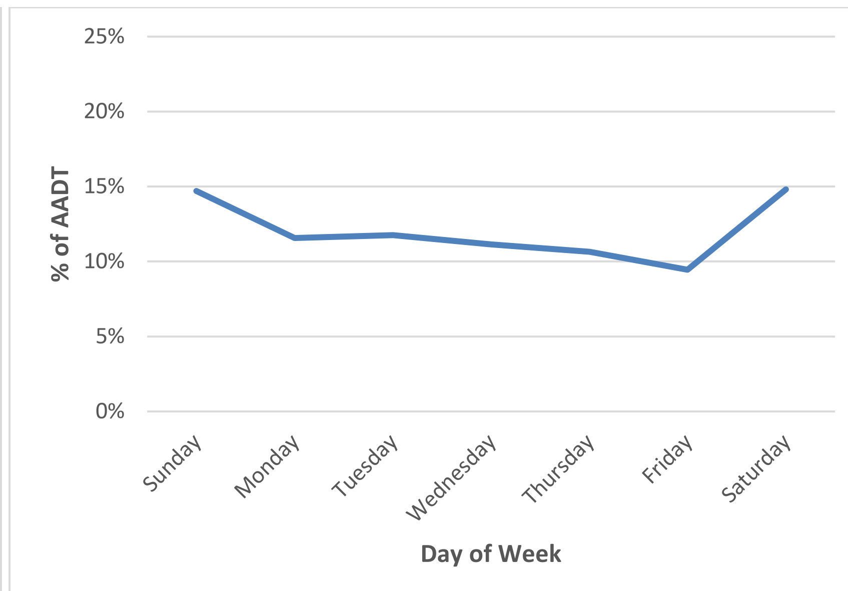




**Figure 29. Pedestrian Traffic Patterns: Katy Trail (Houston Street/AA Center) (Dallas).**

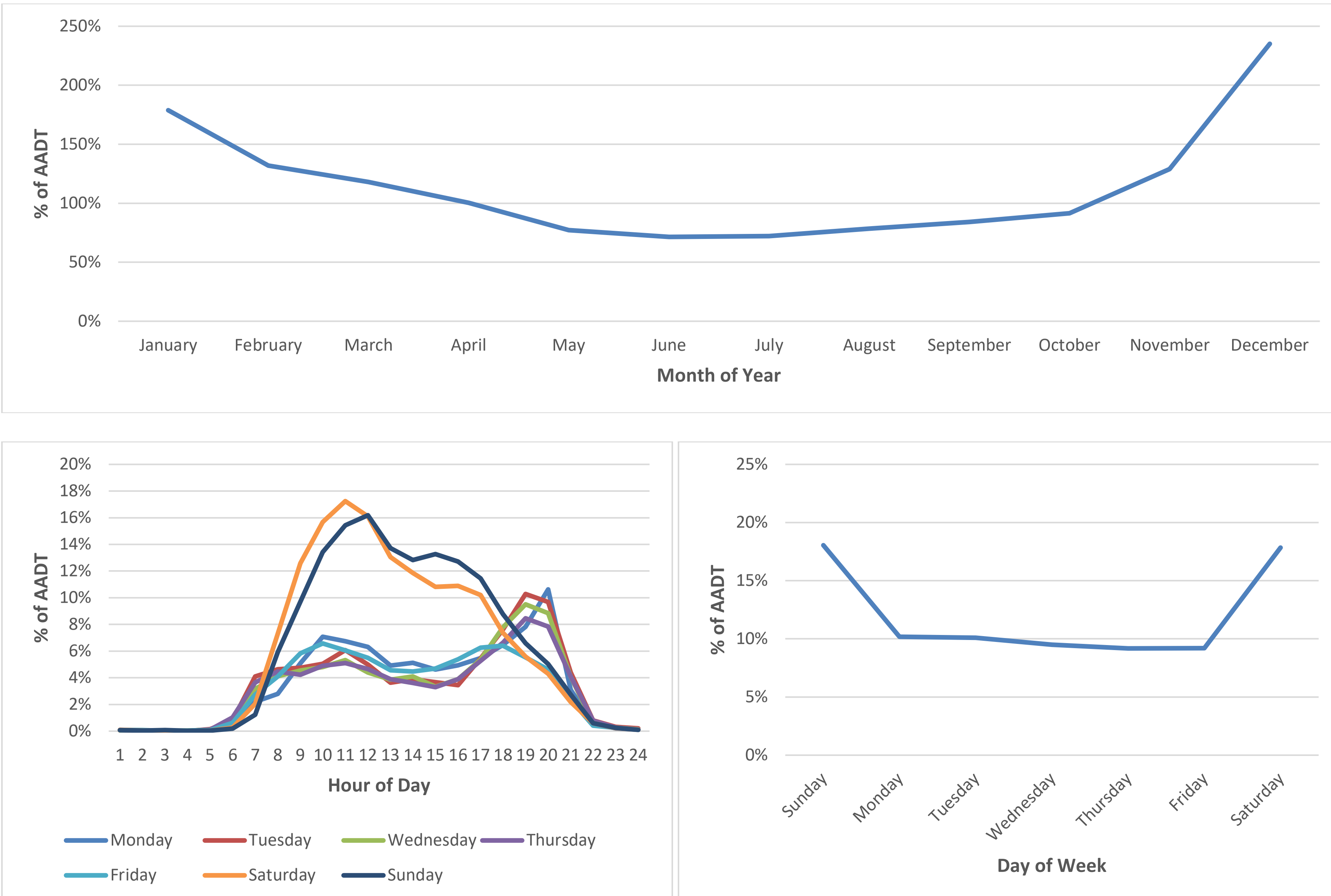



**Figure 30. Bicyclist Traffic Patterns: White Rock Trail at Big Thicket (Dallas).**

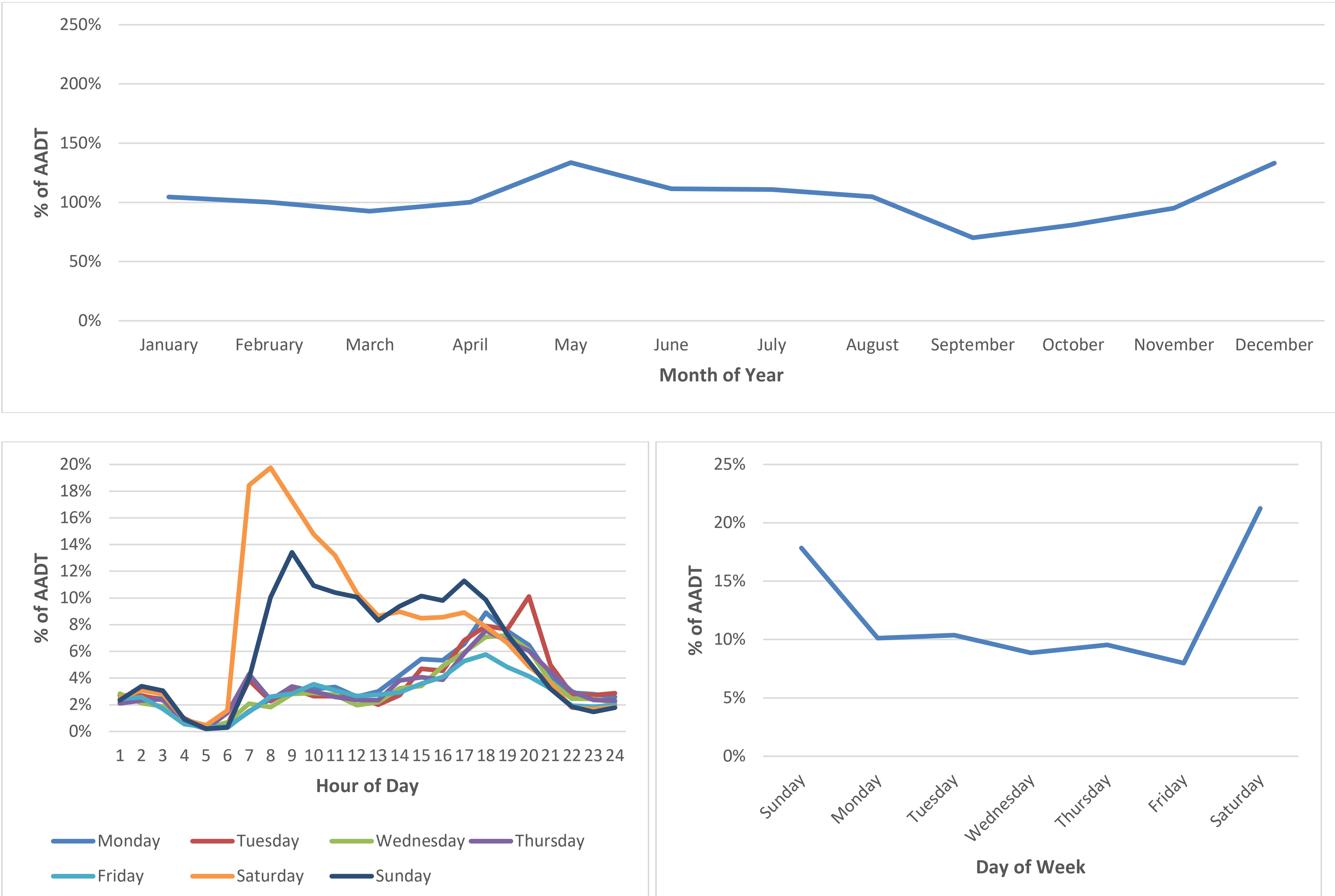



**Figure 31. Pedestrian Traffic Patterns: White Rock Trail at Big Thicket (Dallas).**

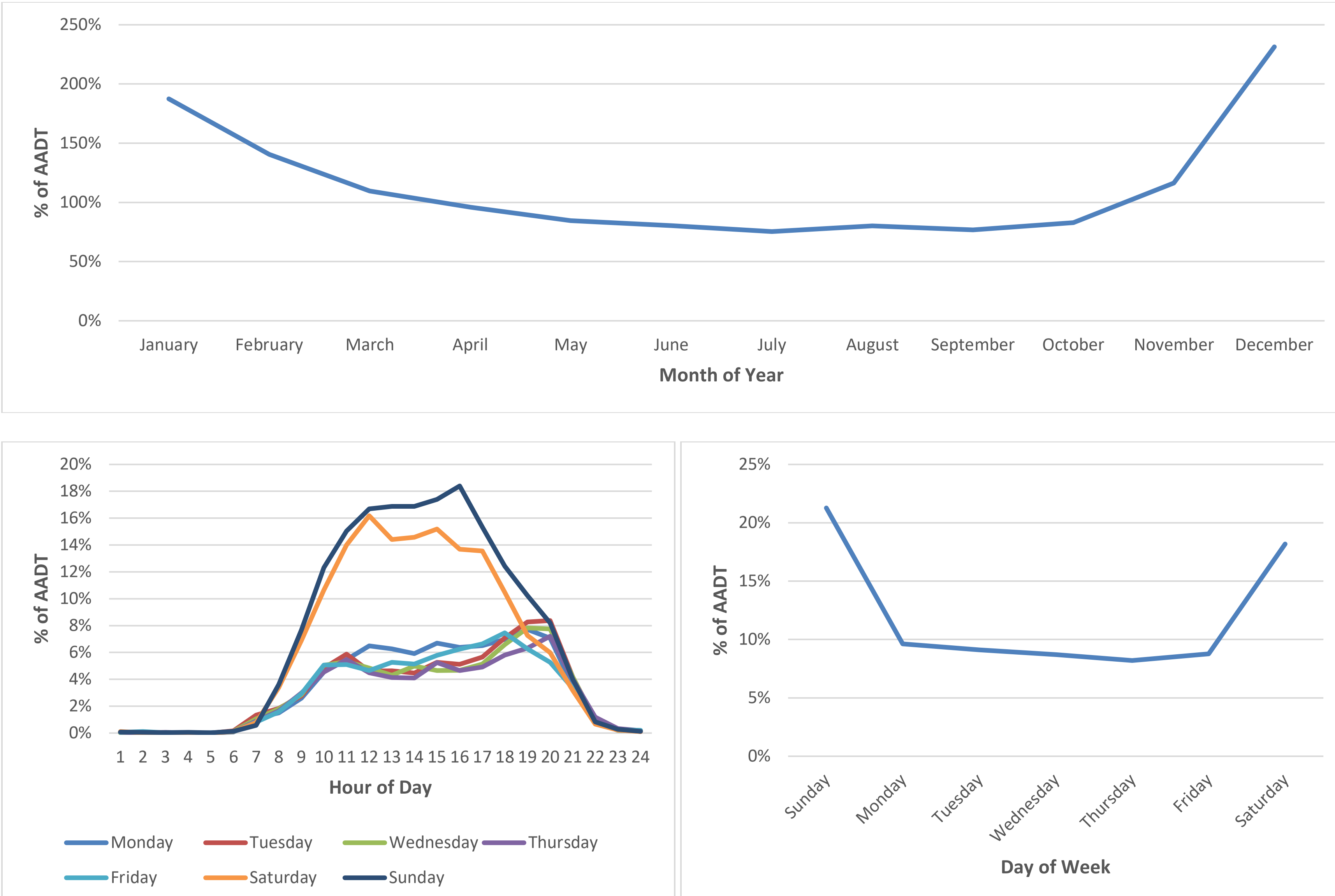



**Figure 32. Bicyclist Traffic Patterns: White Rock Lake Trail (at Fisher) (Dallas).**

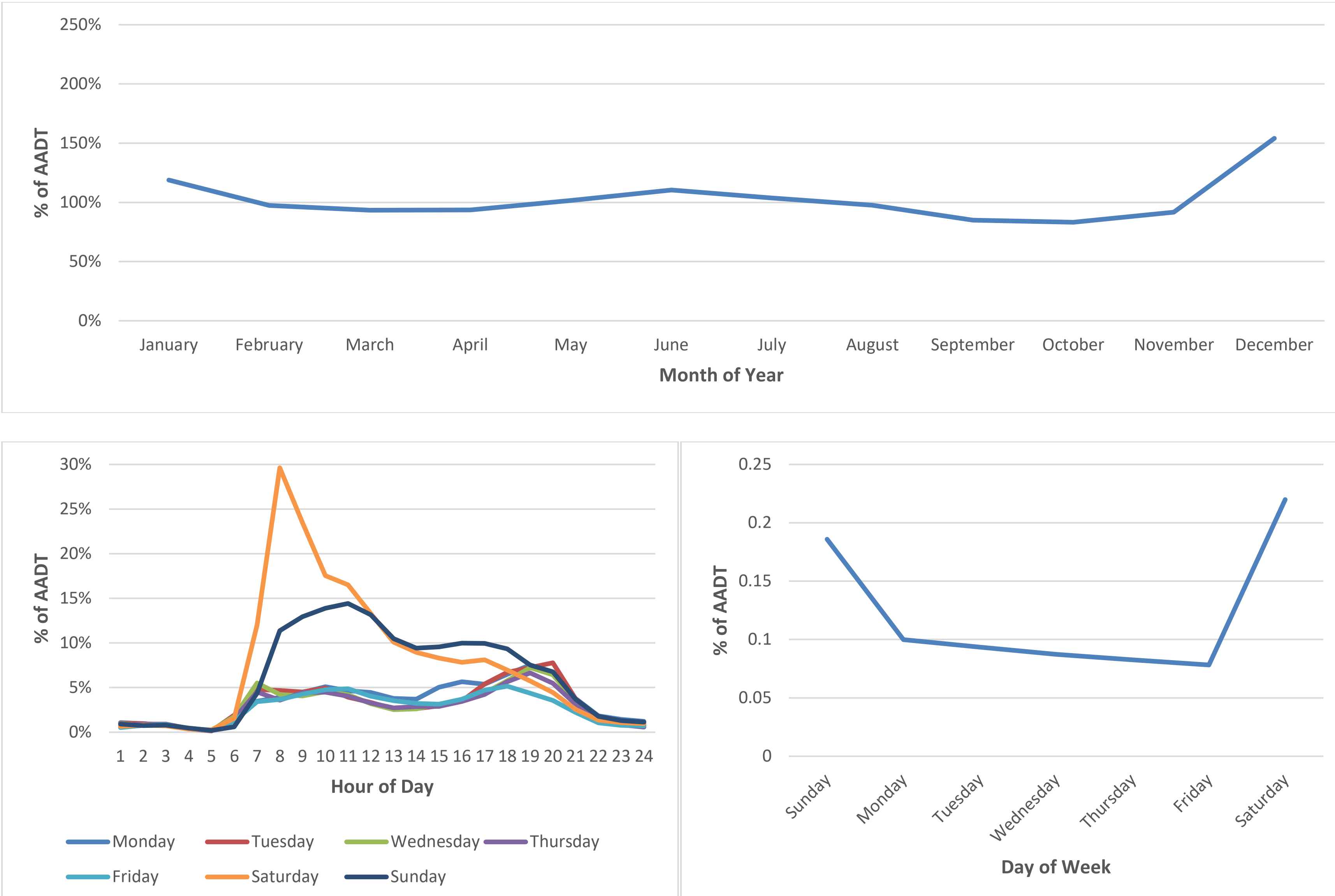



**Figure 33. Pedestrian Traffic Patterns: White Rock Lake Trail (at Fisher) (Dallas).**

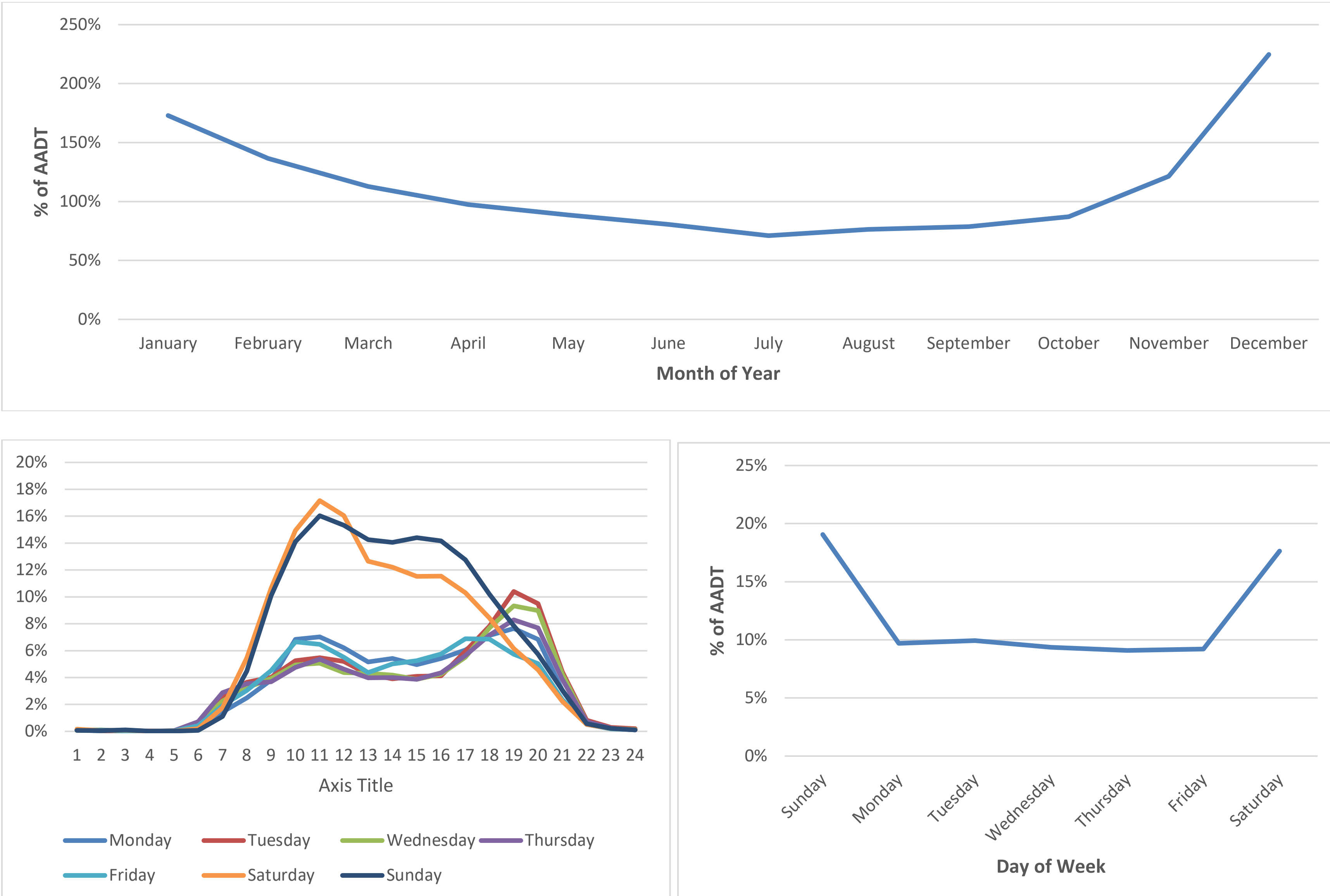
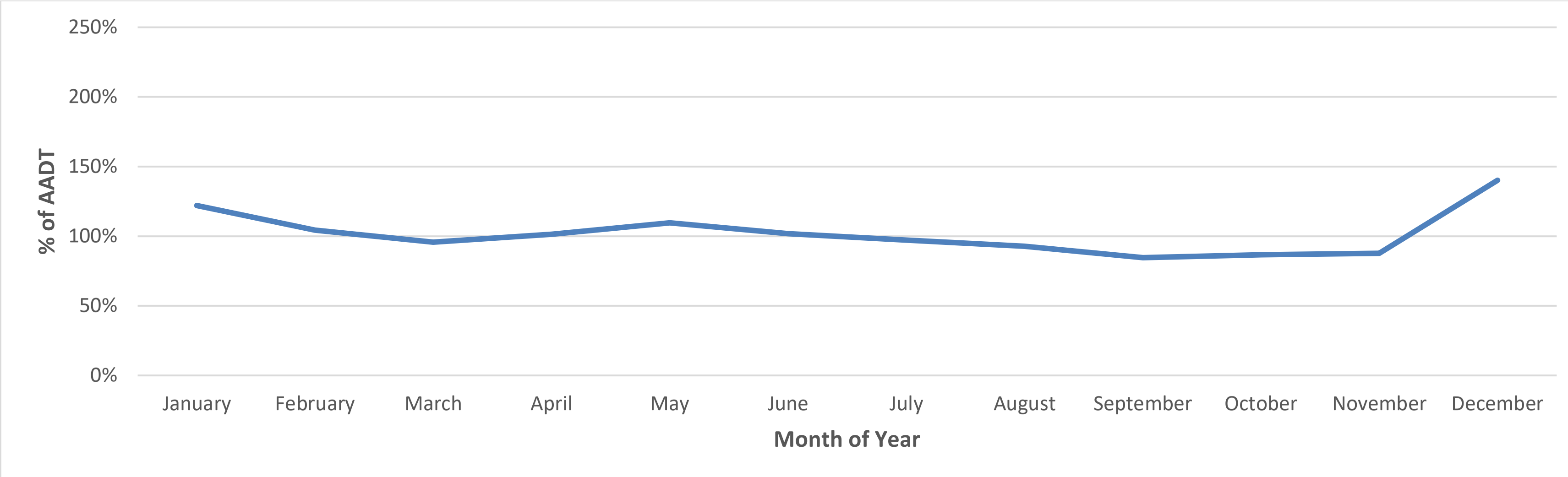
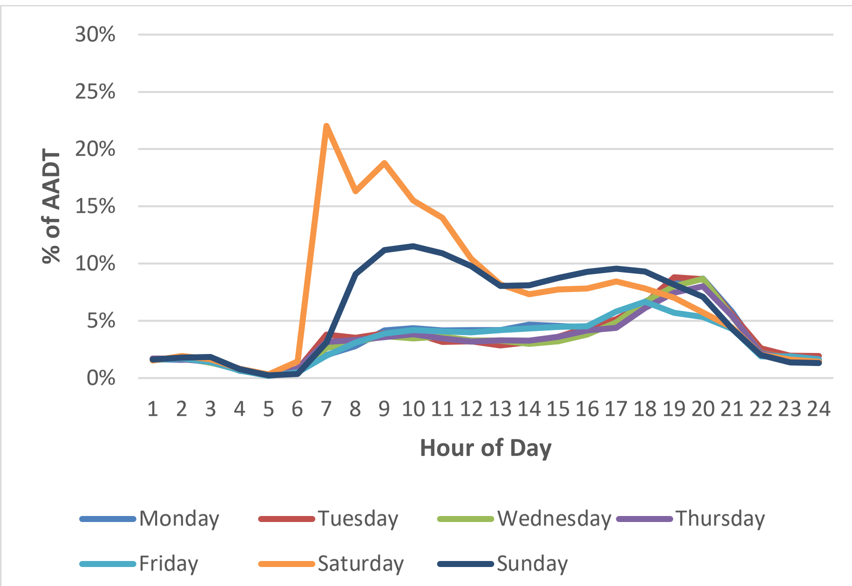



**Figure 34. Bicyclist Traffic Patterns: White Rock Lake Trail at Winfrey Point (Dallas).**



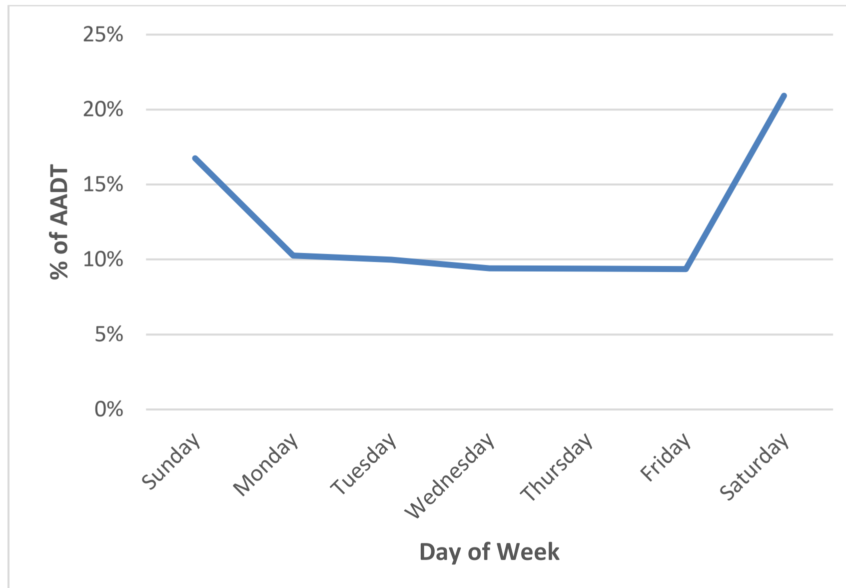




Figure 35. Pedestrian Traffic Count Patterns: White Rock Lake Trail at Winfrey Point (Dallas).

# APPENDIX B. DEVELOPMENT OF PROCEDURES FOR CROWDSOURCED DATA SCALING

## SELECTION OF MOST INFLUENTIAL VARIABLES

Researchers compiled a list of potentially important variables that can help to explain the relationships between the observed bicycle counts and Strava activity. Table 2 depicts the list of variables and data sources considered for the analysis.

**Table 2. List of Variables Considered for the Analysis.**

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

### 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 3 shows the descriptive statistics of Strava sample percentages for all locations.

**Table 3. 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 4 reports the number of locations per percentile groups.

**Table 4. 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 activity and ground counts researchers used the Strava percentile groups and conducted the data mining analysis. For this purpose, researchers used Random Forest tool that helps to determine the most important variables based on two criteria:

- Mean Decrease Accuracy.
- Mean Decrease Gini.[2]

The initial analysis results indicate that the household income and demographic variables are very influential (Figure 36). 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.

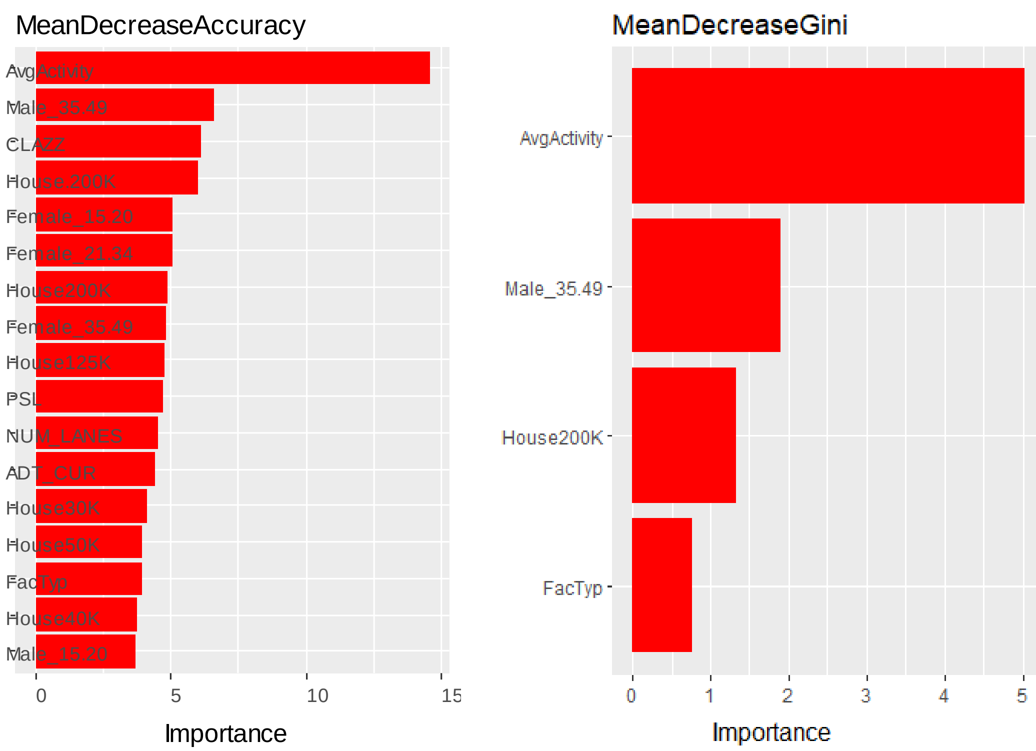


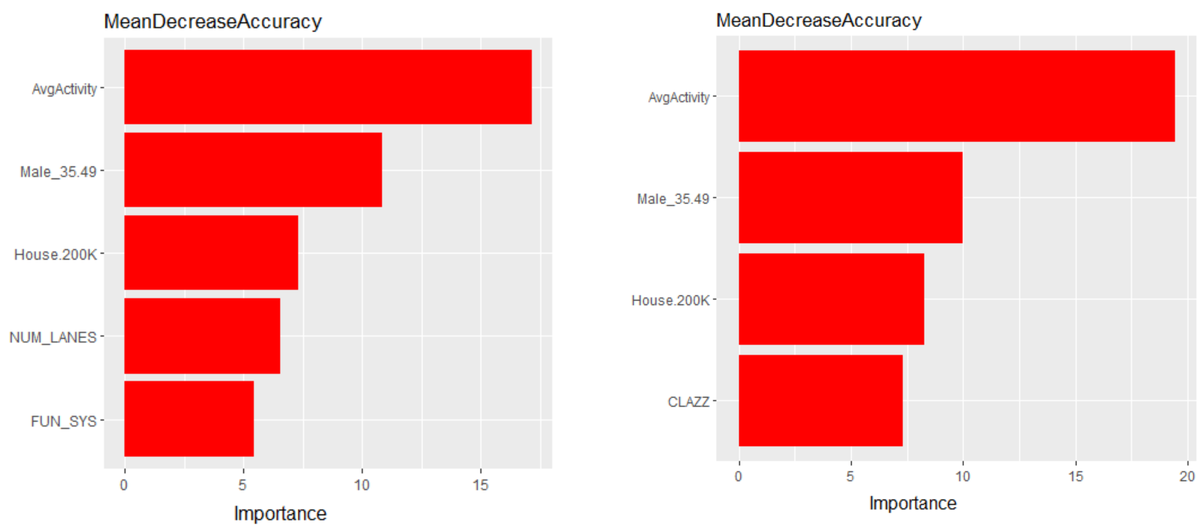

**Figure 36. Initial List of Important Variables.**

After conducting the data mining analysis for the second time, by keeping only the most important American Community Survey (ACS) variables in the Random Forest Tool, researchers identified the list of most influential variables as follows (Figure 37):

- Strava sample (Strava).
- Male 35–49 (ACS).

[2] Mean Decrease Gini (MDG) refers to a process to select variables in large datasets based on a Random Forest method. A higher importance factor for a variable indicates greater influence on an outcome variable, in this case, bicycle counts.

- Household income of $200K (ACS).
- Functional System, CLAZZ (Strava/OSM).
- Number of Lanes (RHiNO).
- Facility Type (Manual).




a) If both CLAZZ and FUN_SYS included

b) If only CLAZZ included

**Figure 37. Final List of Important Variables.**

The CLAZZ variable indicates the roadway class. In this study, the researchers have used the ground counts collected from some of the roadway class types. However, Texas Strava shapefiles include other roadway classes that were not included in the study. Table 5 indicates the list of the roadway classes included in the Texas Strava (OSM) shapefiles, their definitions, default speeds and allowed transportation types (Source: Strava User Guide). The table also shows which of these roadway class types were included in the data analysis. The last column shows the corresponding roadway class types that can be used as a surrogate for the missing roadway classes. Note that the _link tags refer to the ramps and channelization. Since this roadway class was not included in the data analysis the coefficient for this category is zero. Moreover, the CLAZZ codes 43 and 74 have not been defined in the User Guide although they were included in the Texas Strava map. The coefficient for these types will also be equal to zero.

**Table 5. CLAZZ Definitions.**

| CLAZZ Codes in Texas OSM | Definition | Default Speed (KMH) | Approximate Speed in MPH | Allowed Transportation Type | Included in the Model | Compatible CLAZZ Code |
|---|---|---|---|---|---|---|
| 11 | Highway, motorway | 120 | 75 | Car | No | 15 |
| 12 | Highway, motorway_link* | 30 | 20 | Car | No | NA** |
| 13 | Highway, trunk | 90 | 60 | Car | No | 15 |
| 14 | Highway, trunk_link* | 30 | 20 | Car | No | NA** |
| 15 | Highway, primary | 70 | 45 | Car | Yes | 15 |
| 16 | Highway, primary_link* | 30 | 20 | Car | No | NA** |
| 21 | Highway, secondary | 60 | 40 | Car | Yes | 21 |
| 22 | Highway, secondary_link* | 30 | 20 | Car | No | NA** |
| 31 | Highway, tertiary | 40 | 25 | Car, Bike | Yes | 31 |
| 32 | Highway, residential | 50 | 30 | Car, Bike | Yes | 32 |
| 41 | Highway, road | 30 | 20 | Car, Bike | No | 31 |
| 42 | Highway, unclassified, | 30 | 20 | Car, Bike | No | 31 |
| 43 | Not Defined | | 0 | | No | |
| 51 | Highway, service | 5 | 5 | Car, Bike | No | 91/81 |
| 62 | Highway, pedestrian | 5 | 5 | Bike, Foot | No | 91 |
| 63 | Highway, living_street | 7 | 5 | Car, Bike, Foot | No | 91 |
| 71 | Highway, track | 10 | 10 | Bike, Foot | No | 72 |
| 72 | Highway, path | 10 | 10 | Bike, Foot | Yes | 72 |
| 73 | Highway, bridleway | 10 | 10 | Bike, Foot | No | 72 |
| 74 | Not Defined | | 0 | | No | |
| 81 | Highway, cycleway | 15 | 10 | Bike | Yes | 81 |
| 91 | Highway, footway | 5 | 5 | Foot | Yes | 91 |

*The **_link** tags are used to identify slip roads/ramps and "channelized" (physically separated) at-grade turning lanes connecting the through carriageways/through lanes of highways to other roadways of all types. **_link** tags should also be used for physical channelization of turning traffic lanes at traffic signal junctions and in roundabout designs that physically separate a specific turn from the main roundabout (Source: https://wiki.openstreetmap.org/wiki/Highway_link).

## MODEL ESTIMATION RESULTS

After identifying the most important variables, researchers developed travel demand models to estimate the impact of these variables on the observed bicycle counts. Researchers developed two sets of models using:

- Strava and ACS databases (Model 1).
- Strava, ACS and RHiNO databases (Model 2).

Each set of models consists of three models:

- Total Counts.
- Default Direction Counts.
- Reverse Direction Counts.

Hence a total of six models were developed to estimate the annual average daily bicycle (AADB) counts. Variable Males 35–49 was not included in the final models, since the sign of the variable kept changing based on the model. Researchers used the Strava Activity (AADB Strava), number of households with > 200K income (Households > 200K) and CLAZZ variables to develop Model 1.

*Model 1 (CLAZZ)*

$$AADB_i = \exp(\beta_0 + \beta_1 \times AADB\ Strava_i + \beta_2 \times \text{Household} > \ 200\text{K}_{\ i} + \beta_3 \times CLAZZ_i)$$

Where

- $AADB_i -$ represents the estimated Annual Average Daily Bicycles at segments/edge $i$.
- $AADB\ Strava_i -$ represents the Strava sample activity at location $i$ for the given time period.
- $\text{Household} > \ 200\text{K}_{\ i} -$ represents the number of households with 200K income
- $CLAZZ_i -$ represents the functional system according to Strava/OSM.
- $\beta_k -$ are the coefficient estimates.

Researchers used Strava Activity, number of households with > 200K income, FUN_SYS and number of lanes to develop Model 2.

Model 2 (FUN_SYS)

$$AADB_i = \exp(\beta_0 + \beta_1 \times AADB\ Strava_i + \beta_2 \times \text{Household} > \ 200\text{K}_{\ i} + \beta_3 \times FUN\_SYS_i + +\beta_4 \times NUM\_LANES_i)$$

Where

- $AADB_i -$ represents the estimated Annual Average Daily Bicycles at segments/edge $i$.
- $AADB\ Strava_i -$ represents the Strava sample activity at location $i$ for the given time period.
- $\text{Household} > \ 200\text{K}_{\ i} -$ represents the number of households with 200K income
- $FUN\_SYS_i -$ represents the roadway functional system according to RHiNO
- $NUM\_LANES_i -$ represents the number of lanes on the roadway segment.
- $\beta_k -$ are the coefficient estimates.

Table 6 and Table 7 show the estimation results for both models, as well as the goodness of fit measures (overdispersion parameter and $R^2$).

## Table 6. Estimation Results, Model 1.

**Model 1**: $AADB_i = \exp(\beta_0 + \beta_1 \times AADB\ Strava_i + \beta_2 \times \text{Household} > \$200\text{K}_{\ i} + \beta_3 \times CLAZZ_i + \varepsilon_i)$

| Variables Included in the Models | | Total Counts | | | Default Direction Counts | | | Reverse Direction Counts | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ***Estimate*** | ***St.D.*** | ***p-value*** | ***Estimate*** | ***St.D.*** | ***p-value*** | ***Estimate*** | ***St.D.*** | ***p-value*** |
| Intercept (CLAZZ) | Highway, primary | 4.138 | 0.053 | < 0.001 | 2.782 | 0.109 | < 0.001 | 4.343 | 0.041 | < 0.001 |
| | Highway, secondary | 2.590 | 0.060 | < 0.001 | 2.598 | 0.060 | < 0.001 | 2.521 | 0.061 | < 0.001 |
| | Highway, tertiary | 3.078 | 0.062 | < 0.001 | 3.227 | 0.058 | < 0.001 | 2.830 | 0.068 | < 0.001 |
| | Highway, residential | 2.862 | 0.037 | < 0.001 | 3.184 | 0.032 | < 0.001 | 2.460 | 0.045 | < 0.001 |
| | Highway, path | 4.271 | 0.031 | < 0.001 | 4.551 | 0.029 | < 0.001 | 3.468 | 0.039 | < 0.001 |
| | Cycleway | 4.144 | 0.027 | < 0.001 | 4.134 | 0.027 | < 0.001 | 3.754 | 0.032 | < 0.001 |
| | Footway | 3.323 | 0.062 | < 0.001 | 3.468 | 0.060 | < 0.001 | 2.135 | 0.071 | < 0.001 |
| AADB Strava | | 0.038 | 0.000 | < 0.001 | 0.022 | 0.000 | < 0.001 | 0.100 | 0.001 | < 0.001 |
| Number of Households with >200K income | | 0.002 | 0.000 | < 0.001 | 0.002 | 0.000 | < 0.001 | 0.000 | 0.000 | < 0.01 |
| $R^2$ (Model Accuracy) | | 75% | | | 85% | | | 83% | | |
| Overdispersion | | 1.172 | | | 0.822 | | | 1.338 | | |

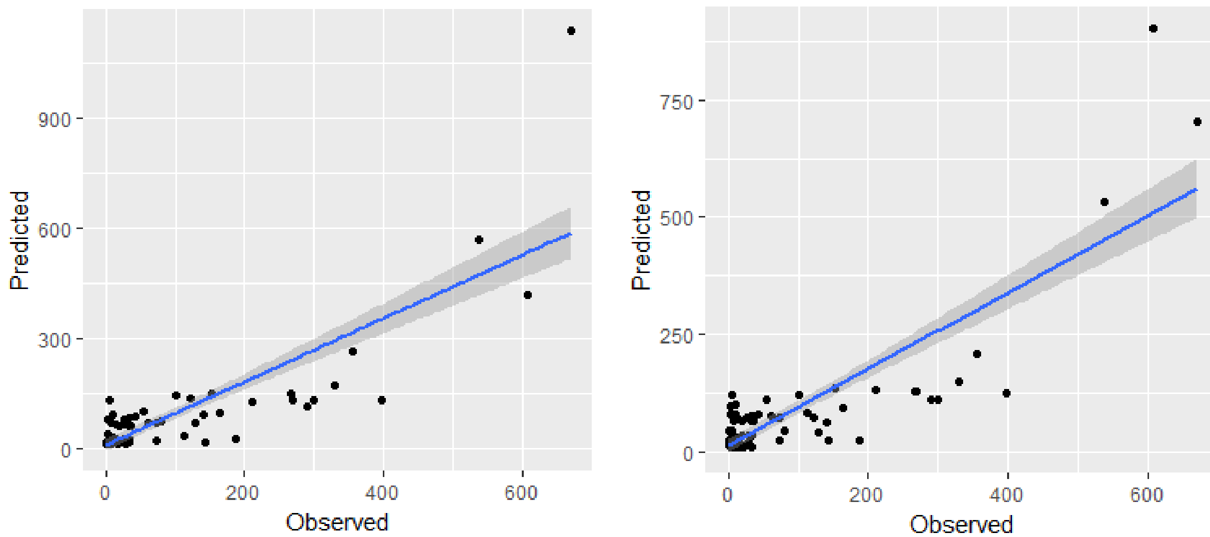
### Table 7. Estimation Results, Model 2.

**Model 2**: $AADB_i = \exp(\beta_0 + \beta_1 \times AADB\ Strava_i + \beta_2 \times \text{Household} >\ 200\text{K}_{\ i} + \beta_3 \times FUN\_SYS_i + +\beta_4 \times NUM\_LANES_i + \varepsilon_i)$

| Variables Included in the Models | | Total Counts | | | Default Direction Counts | | | Reverse Direction Counts | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ***Estimate*** | ***St.D.*** | ***p-value*** | ***Estimate*** | ***St.D.*** | ***p-value*** | ***Estimate*** | ***St.D.*** | ***p-value*** |
| Intercept (FUN_SYS) | Collector (Minor) | 3.211 | 0.078 | < 0.001 | 2.491 | 0.094 | < 0.001 | 3.078 | 0.082 | < 0.001 |
| | Local Road | 2.506 | 0.083 | < 0.001 | 1.631 | 0.092 | < 0.001 | 3.036 | 0.077 | < 0.001 |
| | Minor Arterial | 2.987 | 0.118 | < 0.001 | 1.548 | 0.139 | < 0.001 | 3.662 | 0.102 | < 0.001 |
| | Principal Arterial | 3.929 | 0.116 | < 0.001 | 1.539 | 0.162 | < 0.001 | 4.741 | 0.088 | < 0.001 |
| | Trail | 4.270 | 0.035 | < 0.001 | 4.047 | 0.038 | < 0.001 | 3.969 | 0.035 | < 0.001 |
| AADB Strava | | 0.031 | 0.000 | < 0.001 | 0.017 | 0.000 | < 0.001 | 0.091 | 0.001 | < 0.001 |
| Density of Households with >200K income | | 0.002 | 0.000 | < 0.001 | 0.002 | 0.000 | < 0.001 | 0.000 | 0.000 | < 0.001 |
| Number of Lane | | -0.066 | 0.027 | < 0.05 | 0.310 | 0.032 | < 0.001 | -0.271 | 0.023 | < 0.001 |
| $R^2$ (Model Accuracy) | | 70% | | | 64% | | | 70% | | |
| Overdispersion | | 0.967 | | | 0.734 | | | 1.129 | | |

**PREDICTION ERROR MEASURES**

Figure 38 indicates the prediction intervals.




a) Model 1 b) Model 2

**Figure 38. Prediction Intervals.**

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

<u>*Mean Absolute Percentage Error*</u>

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

<u>*Mean Accuracy Error*</u>

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

<u>*Mean Squared Error*</u>

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

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 measure.

Table 8 reports the error measures for the two models while Table 9 shows the error measures for the Strava percentile groups listed in Table 3. 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 8. Relative Accuracy per Strava Percentage Categories.**

| Prediction Error Measure | Model 1 (CLAZZ) | Model 2 (RHiNO) |
|---|---|---|
| **Mean Absolute Percentage Error** | 29% | 38% |
| **Mean Squared Error** | 5855 | 4836 |
| **Mean Absolute Error** | 41 | 42 |

**Table 9. Prediction Error per Strava Percentile Groups.**

| Strava Percentile Groups | | Mean Absolute Prediction Error | | Mean Squared Error | | Mean Absolute Error | |
|---|---|---|---|---|---|---|---|
| | | *Model 1 (CLAZZ)* | *Model 2 (RHiNO)* | *Model 1 (CLAZZ)* | *Model 2 (RHiNO)* | *Model 1 (CLAZZ)* | *Model 2 (RHiNO)* |
| Group 1 | Strava Ratio < 5% | 33% | 44% | 3066 | 3500 | 30 | 33 |
| Group 2 | 5% ≤ Strava Ratio < 10% | 9.80% | 12.50% | 5686 | 7501 | 59 | 67 |
| Group 3 | 10% ≤ Strava Ratio < 15% | 9.90% | 10.10% | 9894 | 2390 | 66 | 103 |
| Group 4 | 15% ≤ Strava Ratio < 20% | 42% | 46% | 3557 | 5154 | 49 | 68 |
| Group 5 | 20% ≤ Strava Ratio | 38% | 39.70% | 38286 | 1606 | 104 | 34 |
| | Average | 29% | 38% | 5855 | 4836 | 41 | 42 |

Table 10 reports the observed and predicted counts and the lower and upper prediction intervals based on Model 1 (which uses CLAZZ to represent functional class).

**Table 10. Prediction Intervals Based on Model 1 (CLAZZ functional class).**

| Station Name | City | CLAZZ | Strava AADB | Observed AADB | Predicted AADB | Lower Prediction Interval (95%) | Upper Prediction Interval (95%) |
|---|---|---|---|---|---|---|---|
| Ann and Roy Butler Trail @ E Bouldin Creek | Austin | 81 | 35 | 791 | 238 | 168 | 308 |
| Ann and Roy Butler Trail @ MoPac Expy | Austin | 81 | 52 | 660 | 455 | 385 | 525 |
| Duval St N of E 32nd St | Austin | 31 | 11 | 376 | 33 | 0 | 103 |
| Guadalupe St N of W 21st St | Austin | 15 | 20 | 281 | 134 | 64 | 204 |
| Johnson Creek Trail @ MoPac/W 5th St/W 6th St Interchange | Austin | 81 | 1 | 108 | 65 | 0 | 135 |
| Lance Armstrong Bikeway @ Waller Creek | Austin | 81 | 55 | 1071 | 510 | 440 | 580 |
| Manor Rd @ Alamo St | Austin | 21 | 17 | 288 | 25 | 0 | 95 |
| S 1st St @ S Bank Colorado River | Austin | 81 | 36 | 241 | 248 | 178 | 318 |
| Shoal Creek Blvd N of W 24th St | Austin | 91 | 43 | 162 | 130 | 60 | 200 |
| Shoal Creek Blvd N of W 38th St | Austin | 31 | 8 | 20 | 29 | 0 | 99 |
| Walnut Creek Trail N of Jain Ln | Austin | 81 | 45 | 304 | 349 | 279 | 419 |
| Old Alice Road at N of Belvedere | Brownsville | 32 | 10 | 148 | 26 | 0 | 96 |
| FM 802 at W of Habana | Brownsville | 15 | 2 | 57 | 68 | 0 | 138 |
| FM 802 at W of Habana | Brownsville | 15 | 0 | 38 | 63 | 0 | 133 |
| Sports Park Blvd at W of Brownsville Sports Park | Brownsville | 32 | 0 | 7 | 17 | 0 | 87 |
| 18th: Fig to McKenzie | Corpus Christi | 32 | 0 | 19 | 17 | 0 | 87 |
| Alameda: Atlantic to Naples | Corpus Christi | 21 | 0 | 2 | 13 | 0 | 83 |

| Station Name | City | CLAZZ | Strava AADB | Observed AADB | Predicted AADB | Lower Prediction Interval (95%) | Upper Prediction Interval (95%) |
|---|---|---|---|---|---|---|---|
| Alameda: Mussett to Caldwell | Corpus Christi | 32 | 0 | 11 | 17 | 0 | 87 |
| Alameda: Naples to Atlantic | Corpus Christi | 21 | 0 | 6 | 13 | 0 | 83 |
| Amber: J to Don Patricio | Corpus Christi | 32 | 0 | 11 | 17 | 0 | 87 |
| Antelope: Artesian to Carrizo | Corpus Christi | 21 | 0 | 11 | 13 | 0 | 83 |
| Aquarius: Doubloon to Gunwale | Corpus Christi | 32 | 0 | 27 | 17 | 0 | 87 |
| Ayers: 2nd to 3rd | Corpus Christi | 21 | 0 | 8 | 13 | 0 | 83 |
| Ayers: 3rd to 2nd | Corpus Christi | 21 | 0 | 11 | 13 | 0 | 83 |
| Ayers: 6th to 7th | Corpus Christi | 21 | 0 | 9 | 13 | 0 | 83 |
| Ayers: 7th to 6th | Corpus Christi | 21 | 0 | 6 | 13 | 0 | 83 |
| Beauvais: Grenade to Beau Terre | Corpus Christi | 32 | 1 | 16 | 18 | 0 | 88 |
| Bernice: Rickey to Susan | Corpus Christi | 32 | 1 | 12 | 18 | 0 | 88 |
| Blevins: Ormond to Melbourne | Corpus Christi | 32 | 0 | 44 | 17 | 0 | 87 |
| Brockhampton: Dunbarton Oak to La Rochelle Way | Corpus Christi | 31 | 1 | 13 | 23 | 0 | 93 |
| Buford: 23rd to 22nd | Corpus Christi | 32 | 0 | 21 | 17 | 0 | 87 |
| Buford: Santa Fe to 3rd | Corpus Christi | 32 | 0 | 10 | 17 | 0 | 87 |

| Station Name | City | CLAZZ | Strava AADB | Observed AADB | Predicted AADB | Lower Prediction Interval (95%) | Upper Prediction Interval (95%) |
|---|---|---|---|---|---|---|---|
| Carroll: Carrolleton to Gollihar | Corpus Christi | 31 | 0 | 13 | 22 | 0 | 92 |
| Carroll: Gollihar to Carrolleton | Corpus Christi | 31 | 0 | 8 | 22 | 0 | 92 |
| Carroll: Lamont to Harold | Corpus Christi | 31 | 0 | 12 | 22 | 0 | 92 |
| Carroll: Lamont to Marion (#1) | Corpus Christi | 31 | 0 | 15 | 22 | 0 | 92 |
| Carroll: Lamont to Marion (#2) | Corpus Christi | 31 | 0 | 13 | 22 | 0 | 92 |
| Center: Pasadena to Peerman | Corpus Christi | 32 | 1 | 30 | 18 | 0 | 88 |
| Cheyenne: Aztec to Washington | Corpus Christi | 32 | 0 | 10 | 17 | 0 | 87 |
| Cliff Crenshaw: Guess to Turkey Creek | Corpus Christi | 32 | 1 | 65 | 18 | 0 | 88 |
| Corona: Embassy to Flynn Parkway | Corpus Christi | 31 | 0 | 6 | 22 | 0 | 92 |
| Corona: Flynn Parkway to Embassy | Corpus Christi | 31 | 0 | 6 | 22 | 0 | 92 |
| Gollihar: Dody to Weber | Corpus Christi | 21 | 0 | 8 | 13 | 0 | 83 |
| Gollihar: Driftwood to Dolphin | Corpus Christi | 21 | 0 | 4 | 13 | 0 | 83 |
| Gollihar: Kasper to Weber | Corpus Christi | 21 | 0 | 2 | 13 | 0 | 83 |
| Gollihar: Laura to Kirkwood | Corpus Christi | 21 | 0 | 11 | 13 | 0 | 83 |
| Gollihar: Laura to Randall | Corpus Christi | 21 | 0 | 33 | 13 | 0 | 83 |

| Station Name | City | CLAZZ | Strava AADB | Observed AADB | Predicted AADB | Lower Prediction Interval (95%) | Upper Prediction Interval (95%) |
|---|---|---|---|---|---|---|---|
| Gollihar: Sequoia to Willow | Corpus Christi | 21 | 0 | 4 | 13 | 0 | 83 |
| Gypsy: Hawksnest Bay to Whaler | Corpus Christi | 32 | 0 | 15 | 17 | 0 | 87 |
| Holly: SH 286 to Martin | Corpus Christi | 21 | 0 | 55 | 13 | 0 | 83 |
| Holly: Victor Lara Ortegon to Greenwood | Corpus Christi | 21 | 0 | 17 | 13 | 0 | 83 |
| Lawnview: Sunset to Melrose | Corpus Christi | 32 | 0 | 20 | 17 | 0 | 87 |
| Matlock: Greenbay to Mciver | Corpus Christi | 32 | 0 | 46 | 17 | 0 | 87 |
| Mesquite: Fitzgerald to Resaca | Corpus Christi | 32 | 0 | 26 | 17 | 0 | 87 |
| Peoples: Chapparal to Mesquite | Corpus Christi | 32 | 0 | 17 | 17 | 0 | 87 |
| Pontchartrain: Kaw Lake to Lake Travis | Corpus Christi | 32 | 1 | 9 | 18 | 0 | 88 |
| Retta: Purdue to Selkirk | Corpus Christi | 31 | 0 | 27 | 22 | 0 | 92 |
| River Canyon: Teague to Rolling Ridge | Corpus Christi | 32 | 0 | 30 | 17 | 0 | 87 |
| Rodd Field: Saratoga to Brooke | Corpus Christi | 21 | 0 | 59 | 13 | 0 | 83 |
| Roosevelt: Clark to Vanderbilt | Corpus Christi | 32 | 0 | 17 | 17 | 0 | 87 |
| Sabinas: Dunbar to Tarlton | Corpus Christi | 32 | 0 | 7 | 17 | 0 | 87 |
| Wood River: Red River to Rapids | Corpus Christi | 32 | 1 | 52 | 18 | 0 | 88 |

| Station Name | City | CLAZZ | Strava AADB | Observed AADB | Predicted AADB | Lower Prediction Interval (95%) | Upper Prediction Interval (95%) |
|---|---|---|---|---|---|---|---|
| Yorktown: Great Lakes to Oso Parkway | Corpus Christi | 21 | 2 | 12 | 14 | 0 | 84 |
| AT&T Trail at Trinity Forest Trail | Dallas | 72 | 0 | 16 | 72 | 2 | 142 |
| Coombs Creek Trail | Dallas | 81 | 4 | 19 | 73 | 3 | 143 |
| Gateway Park Loop Trail | Dallas | 91 | 0 | 68 | 25 | 0 | 95 |
| Glendale Park Loop Trail | Dallas | 91 | 0 | 20 | 25 | 0 | 95 |
| Glendale Park South Loop Trail | Dallas | 91 | 0 | 29 | 25 | 0 | 95 |
| Katy Trail (Houston street/ AA Center- | Dallas | 81 | 15 | 419 | 111 | 41 | 181 |
| Katy Trail at Cedar Springs Rd | Dallas | 81 | 28 | 537 | 183 | 113 | 253 |
| Katy Trail at Fitzhugh | Dallas | 81 | 26 | 577 | 169 | 99 | 239 |
| Katy Trail at Harvard Avenue | Dallas | 81 | 19 | 327 | 130 | 60 | 200 |
| Kiest Park Loop Trail at Conservation | Dallas | 81 | 0 | 73 | 63 | 0 | 133 |
| Preston Ridge Trail at La Cosa Dr. | Dallas | 81 | 6 | 7 | 79 | 9 | 149 |
| Preston Ridge Trial at Debbe | Dallas | 72 | 6 | 83 | 90 | 20 | 160 |
| Trinity strand Trail at Hi Line Dr | Dallas | 91 | 0 | 20 | 25 | 0 | 95 |
| W NorthWest Highway | Dallas | 15 | 0 | 66 | 63 | 0 | 133 |
| White Rock Lake Trail | Dallas | 72 | 23 | 532 | 172 | 102 | 242 |
| White Rock Lake Trail at Winfrey Point | Dallas | 81 | 151 | 1337 | 19575 | 19505 | 19645 |
| White Rock Trail at Big Thicket | Dallas | 32 | 161 | 1214 | 7943 | 7873 | 8013 |
| White Rock Trail at Dog Park | Dallas | 72 | 53 | 709 | 536 | 466 | 606 |
| Brays Bayou Greenway Trail @ Spur 5 | Houston | 72 | 4 | 65 | 83 | 13 | 153 |
| Columbia Tap Trail @ Blodgett Street | Houston | 81 | 4 | 121 | 73 | 3 | 143 |

| **Station Name** | **City** | **CLAZZ** | **Strava AADB** | **Observed AADB** | **Predicted AADB** | **Lower Prediction Interval (95%)** | **Upper Prediction Interval (95%)** |
|---|---|---|---|---|---|---|---|
| Heights Trail @5 1/2 Street | Houston | 72 | 26 | 599 | 192 | 122 | 262 |
| White Oak Bayou Trail@ 34th Street | Houston | 72 | 33 | 204 | 251 | 181 | 321 |
| SH 3 N @ S of Walter Hall Park | League City | 15 | 4 | 258 | 73 | 3 | 143 |
| Mid College Chaparral Creek Bike Lane South Side | Midland | 32 | 0 | 15 | 17 | 0 | 87 |
| Oakwood Dr at W of Sunnygrove Dr | Odessa | 32 | 0 | 2 | 17 | 0 | 87 |
| Maple Ave at S of E 14th St | Odessa | 32 | 0 | 2 | 17 | 0 | 87 |
| W 22nd St at Ventura Ave | Odessa | 32 | 0 | 2 | 17 | 0 | 87 |
| Plano Bluebonnet Trail at US75 | Plano | 91 | 4 | 5 | 29 | 0 | 99 |
| Plano Legacy Trail | Plano | 72 | 9 | 10 | 101 | 31 | 171 |
| Plano OPP & NP Trail | Plano | 72 | 1 | 53 | 74 | 4 | 144 |
| Mission Reach Trail at north end of San Juan Ditch | San Antonio | 81 | 1 | 30 | 65 | 0 | 135 |
| Mission Reach Trail south of Theo Ave | San Antonio | 81 | 6 | 148 | 79 | 9 | 149 |
| Mission Reach Trail south of VFW Blvd (Mission County Park) | San Antonio | 81 | 3 | 61 | 71 | 1 | 141 |
| Riverwalk Trail between 8th and 9th Street | San Antonio | 91 | 11 | 225 | 38 | 0 | 108 |
| Wichita River Trail at E of Broad St/IH-44 | Wichita Falls | 81 | 0 | 40 | 63 | 0 | 133 |
| FM 369 at N of US 287 | Wichita Falls | 21 | 0 | 2 | 13 | 0 | 83 |
| Burkburnett Rd (SH 240) at N of Airport Dr | Wichita Falls | 21 | 2 | 3 | 14 | 0 | 84 |