The Effects of Bike Traffic on Police Responses in Seattle

Group 9:
Nguyen Hau
Benjamin Nguyen
Luke Sweeney

CSE 4334
Prof. Zaruba
Fall 2023

Table of Contents

Introduction

This project delves into the relationship between bike traffic and crime in an urban setting, leveraging datasets from the Seattle Open Data portal. The project specifically focuses on police responses near automated bike counters to unveil insights into how urban mobility and public safety may be connected through the use of advanced analytics, statistical models, and visualization tools.

Problem Statement

The specific topic addressed by this project is to investigate whether or not areas with higher bike traffic exhibit corresponding higher crime rates. By analyzing datasets from the Seattle Open Data portal, we aim to discern patterns in the data that will tell us information about the dynamics of these two variables. The goal of the project is to produce valuable insights that contribute to an understanding of the relationship.

Analysis and Results

Data Description

Below are the data sets used and some descriptions of what kind of information they contain. For the source of the data, please see the bibliography.

Police Call Data

- **CAD Event Number**: Unique ID
- **Event Clearance Description**: How the call was resolved, as reported by the primary officer.
- **Call Type**: How the call was received by the Communications Center.
- **Priority**: Priority of the call, as assigned by the CAD system.
- **Initial Call Type**: How the call was classified, initially by the Communication Center.
- **Final Call Type**: How the call was classified, finally by the primary officer.
- **Original Time Queued**: Time queued in the CAD system.
- **Arrived Time**: Time the first officer arrived on the call.
- **Precinct**: Precinct where the call originated.
- **Sector**: Sector where the call originated. All Sectors roll up to one of five Precincts.
- **Beat**: Beat where the call originated. All Beats roll up to Sectors.
- **Blurred_Longitude**: Longitude, blurred to protect privacy.
- **Blurred_Latitude**: Latitude, blurred to protect privacy.

Fremont Bridge Bicycle Data

- **Date**: The date and hour of day object(s) are detected by the sensor.

- **Fremont Bridge Sidewalks**, south of N 34th St: Total of both sidewalks
- **Fremont Bridge Sidewalks, south of N 34th St Cyclist East Sidewalk**: The total number of bicyclists traveling on the East sidewalk in one hour as recorded by the sensor.
- **Fremont Bridge Sidewalks, south of N 34th St Cyclist West Sidewalk**: The total number of bicyclists traveling on the West sidewalk in one hour as recorded by the sensor.

NW 58th St Greenway Bike Traffic

- **Date**: The date and hour of day object(s) are detected by the sensor.
- **NW 58th St Greenway st 22nd Ave NW Total**: The total number of bicyclists recorded, in both directions, in one hour. This field adds together data from the East and West columns.
- **East**: The total number of bicyclists traveling east in one hour as recorded by the sensor.
- **West**: The total number of bicyclists traveling west in one hour as recorded by the sensor.

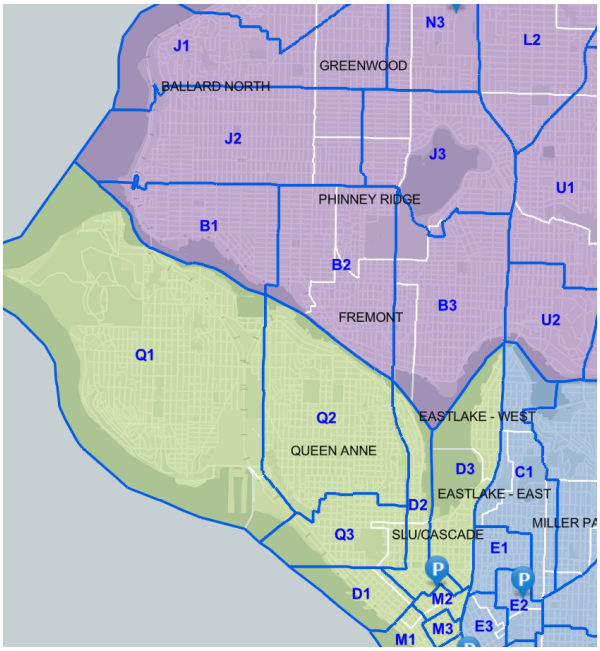
Data Cleaning

The first task was to explore and clean our data. The datasets provided were very well kept. There were few missing values in the relevant police call data, and the values that were missing were not critical. Most of the data is non-scalar values, so outliers were not really a problem.

We began by filtering the police call data. There were over 5.5 million rows, but not all of this was relevant. We only wanted to keep police call data that could likely be related to bike traffic, meaning we only want to look at police calls in the time frame represented by the bike counter data, and that are nearby the bike counters.

To achieve this, we found the date ranges represented by the bike counters and only kept the police call data for those dates. The Fremont counter contained data from 2012-10-3 to 2023-8-31, and the Greenway counter contained data from 2014-1-1 to 2022-7-31.

We then used the Seattle “Find your Precinct” tool[1] pictured here to find the police beat in which each bike counter is located. We used those beats, as well as neighboring beats that are easily accessible from the bike counters[2]. For the Fremont counter, we chose beats Q2, B2, B3, and D2, and for the Greenway counter we chose B1, B2, B3, J2, J3. Only police calls from within these beats were kept. After discarding police calls outside of the relevant time frames and outside of the selected beats, the resulting police call data was split into two datasets, one for each traffic data set. After this, we have two pairs of data sets: Fremont traffic and calls, and Greenway traffic and calls. There were 314k calls in the Fremont data set, and 246k calls in the Greenway data set.



Resampling and Aggregation

Resampling was necessary to analyze this data. The traffic data was already aggregated by hour when we found it, so it made sense to put the police call data in the same format. We resampled both of the call datasets into hourly totals. In the resampled data, we performed the following operations on the call data:

- Call Type, Initial Call Type: kept every unique value
- Clearance Description: kept every unique value
- Priority, Arrival Delay: averaged

The resampled data can now be aggregated with the bike traffic by hour and we can process this combined data.

[1] https://www.seattle.gov/police/about-us/police-locations/precinct-locator

[2] We could have used all police beats that lie within a certain radius. However, in the case of the Fremont Bridge counter, beat D3 is technically close but lies on the other side of Lake Union, making it practically far to travel to.

EDA and Data Visualization

Data Visualization

Listed below in the appendix as figure 1.1 and figure 1.2 is a scatter plot and a time series plot made to visualize all of our data points related to bike traffic and crime rate. In the case of the time series plot, it is intended to show data over a period of time from around 2014 to present time to see if there are some consistent trends and patterns that allow us to uncover any insights about the relationship. The scatter plot suggests that contrary to our initial expectations, crime rate is actually lower whenever there is higher bike traffic. Furthermore the time series plot fails to show any consistent trends or patterns between bike traffic and crime rate over time. Despite the seasonal fluctuations in bike traffic, the crime rate tends to stay at a fairly similar level with seemingly random spikes further emphasizing the lack of a statistical relationship between the two variables. Therefore, the conclusion drawn from these two plots is limited to the observation that a relationship existing between bike traffic and crime rate is highly unlikely.

Correlation

Correlation tables showing any correlation between bike traffic and average priority, average arrival delay, and amount of calls in the hour are listed in the appendix as figures 4.1 and 4.2.

Based on these correlation tables, there is very low correlation between these three variables and any of the traffic data. This would suggest that bike traffic does not have a significant impact on any of those three variables. Based on this very low correlation, it's very unlikely that any regression model will be a good fit for this data. However, because of the nature of this project, we will attempt to fit some models anyway.

Regression Models

Using the same merged data as the correlation section above, we attempted to fit three models anyway to see if there was a good fit. We primarily looked at the R^2 value for each model. For all the models fit, the R^2 value was very low. This indicated that very little change in the response variable was caused by the independent variables. All three of these models were very poor fits for this data, so no further analysis was done. This is evidence that there is no evident causation between bike traffic and police response.

Independent variables	Linear R^2	Polynomial R^2	Random Forest R^2
average_priority, average_arrived_delay, call_count	0.036	0.05	-0.16
call_count	0.025	0.032	0.031
average_priority	0.01	0.015	0.022
average_arrived_delay	0.006	0.00	-0.29

Conclusions

Despite the initial expectations and assumptions, our analysis of the relationship between bike traffic and crime in the city of Seattle yielded some insightful results. Our findings indicate a lack of discernible and statistically significant correlations or relationships between bike traffic and rate in the studied areas. However it is crucial to note that the absence of a clear relationship might be influenced by contextual factors not accounted for in our analysis as they are out of the project's scope. While our study did not show a clear connection, it does prompt for further exploration to identify additional factors influencing bike traffic and crime rates in these settings.

Lessons Learned

I learned a lot about the data analysis process through this project. I think what surprised me the most was that more time was spent on EDA and getting the data ready to be analyzed than actual analysis. If I want to make my future projects successful, I will spend more time on EDA in the beginning of the project, which can help to inform the questions I ask about the data. I also learned that communicating the results of our project can be tricky, so a bit of care is needed there.

-Luke

I learned that data mining takes a lot more time, and effort than I had expected. Following methodology makes things easier

-Benjamin

Clearly understanding the data and having a clear goal in mind is critical to good data mining.

-Nguyen

Bibliography

- Seattle Data Portal: https://data.seattle.gov/
- Police Call Data: https://data.seattle.gov/Public-Safety/Call-Data/33kz-ixgy
- Fremont Bridge Bicycle Data: https://data.seattle.gov/Transportation/Fremont-Bridge-Bicycle-Counter/65db-xm6k
- NW 58th St Greenway Bicycle data: https://data.seattle.gov/Transportation/NW-58th-St-Greenway-at-22nd-Ave-NW-Bicycle-Counter/47yq-6ugv

Appendix

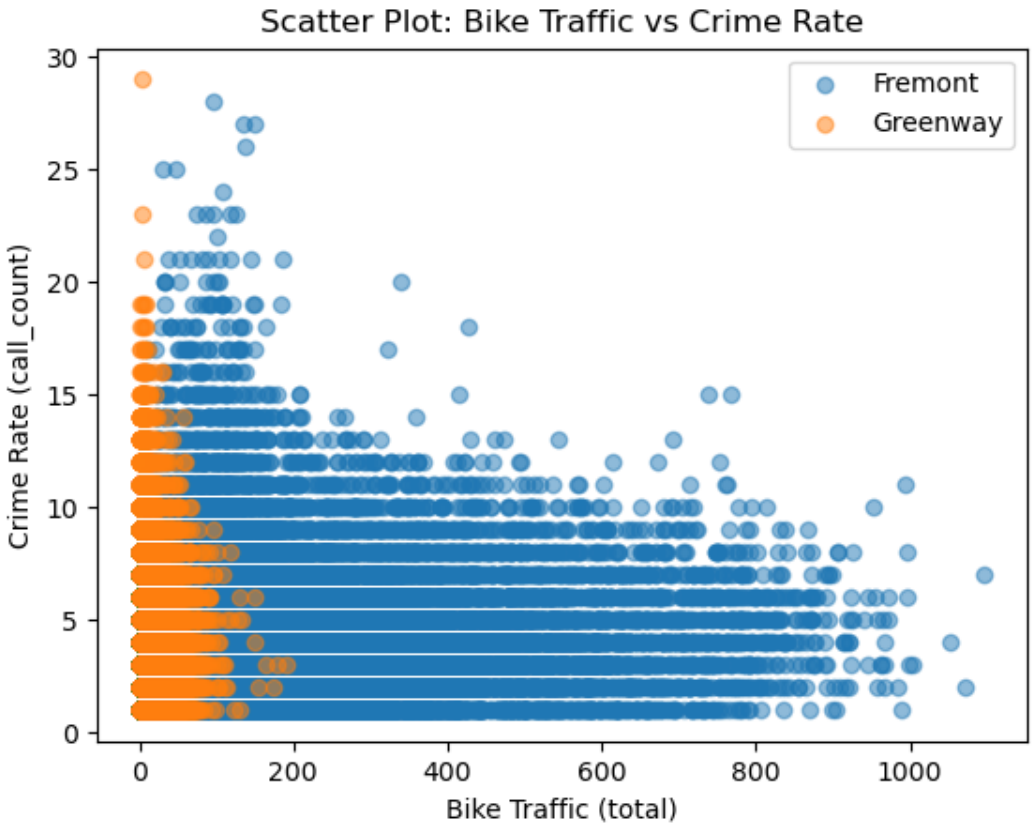


Figure 1.1

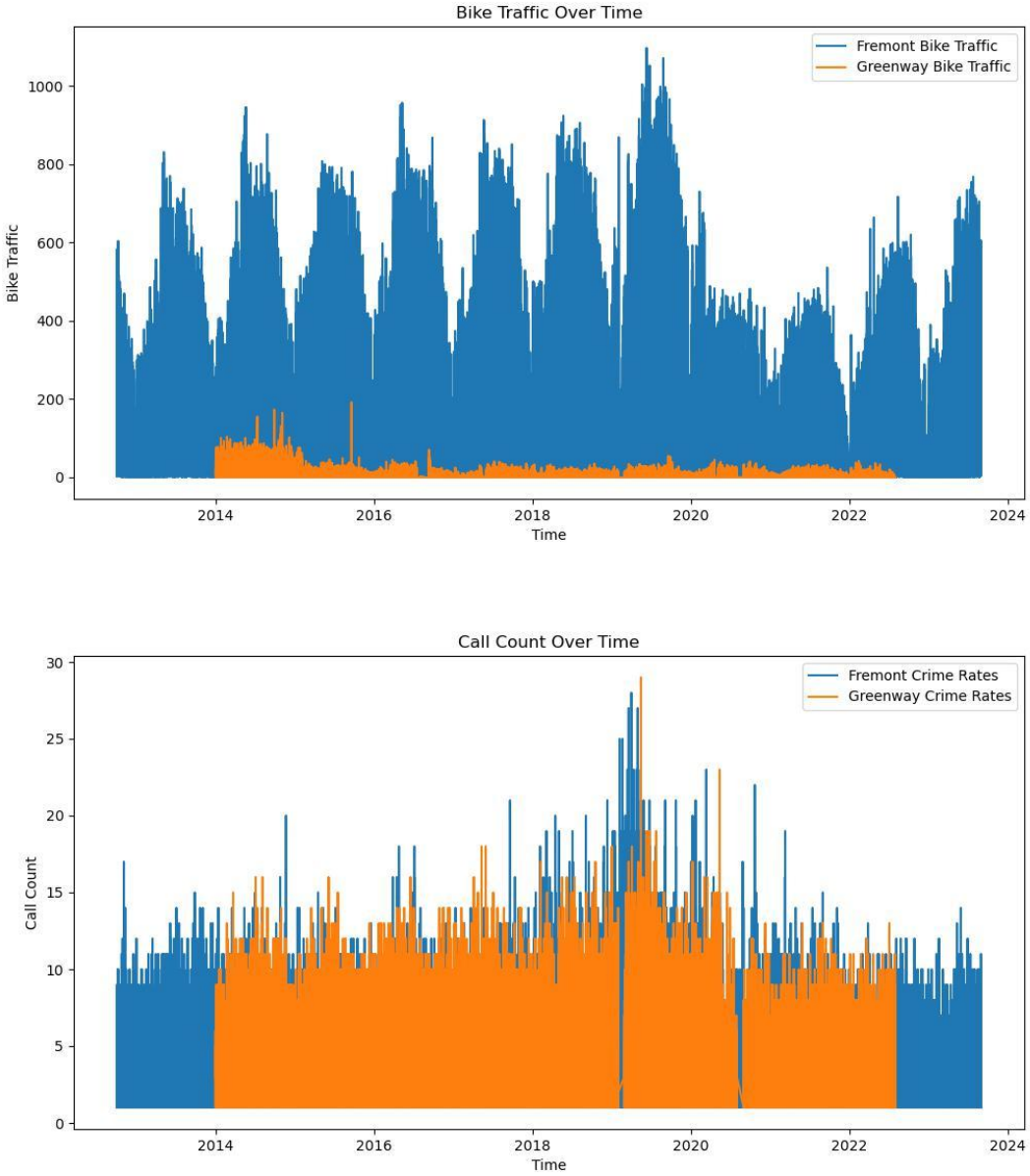

Bike Traffic Over Time
Fremont Bike Traffic
Greenway Bike Traffic
Bike Traffic
0
200
400
600
800
1000
2014
2016
2018
2020
2022
2024
Time
Call Count Over Time
Fremont Crime Rates
Greenway Crime Rates
Call Count
0
5
10
15
20
25
30
2014
2016
2018
2020
2022
2024
Time

Figure 1.2

	bank name	address	predicted response time (mins)	lat	long
0	Washington Federal	1200 NE 45th St Seattle, WA 98105	9.225388	47.661480	-122.315210
1	US Bank	10700 Meridian Ave N Ste G10 Seattle, WA 98133	9.580675	47.707370	-122.333090
2	HSBC	523 Union St, Seattle WA 98101	8.681955	47.609890	-122.333680
3	Key Bank	2105 Queen Anne Ave N Seattle, WA 98109	8.848104	47.637480	-122.357140
4	M&T Bank	4522 45th Ave SW Seattle, WA 98116	8.028426	47.561460	-122.389420
5	BNY Mellon	601 Union St Ste 520 Seattle, WA 98101	8.690951	47.610320	-122.332330
6	Chase Bank	1201 3rd Ave Seattle, WA 98101	8.645289	47.607040	-122.336480
7	Silicon Vallley Bank	920 5th Ave Ste 300 Seattle, WA 98104	8.659463	47.606470	-122.331280
8	Wells Fargo Bank	999 3rd Ave Seattle, WA 98104	8.635259	47.605040	-122.334240
9	Bank of America	401 Union St Fl 24 Seattle, WA 98101	8.669595	47.609220	-122.335340
10	WaFd Bank	425 Pike St, Seattle, WA 98101	8.678403	47.610321	-122.335693

Figure 2 - A Toy Linear Regression

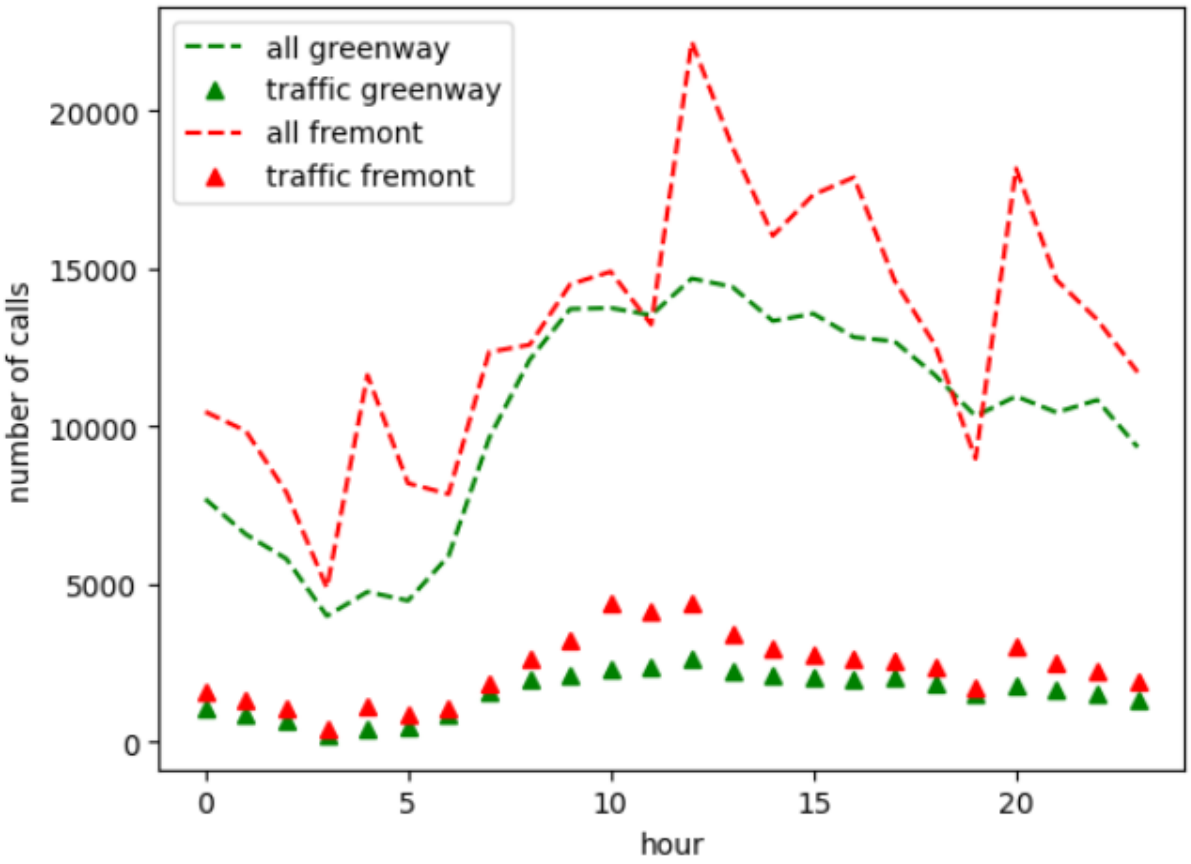


Figure 3: Number of Calls by the Hour

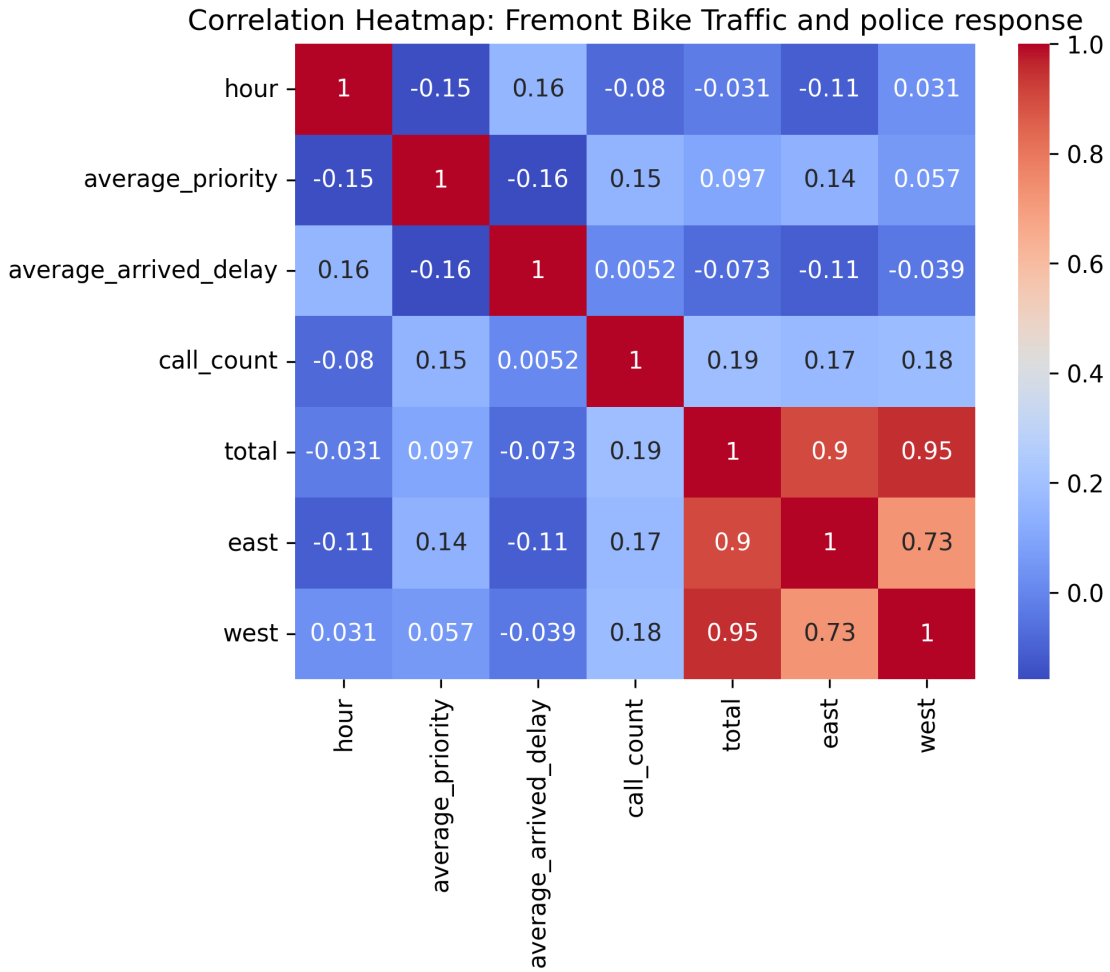


Figure 4.1: Fremont Traffic Correlation Table

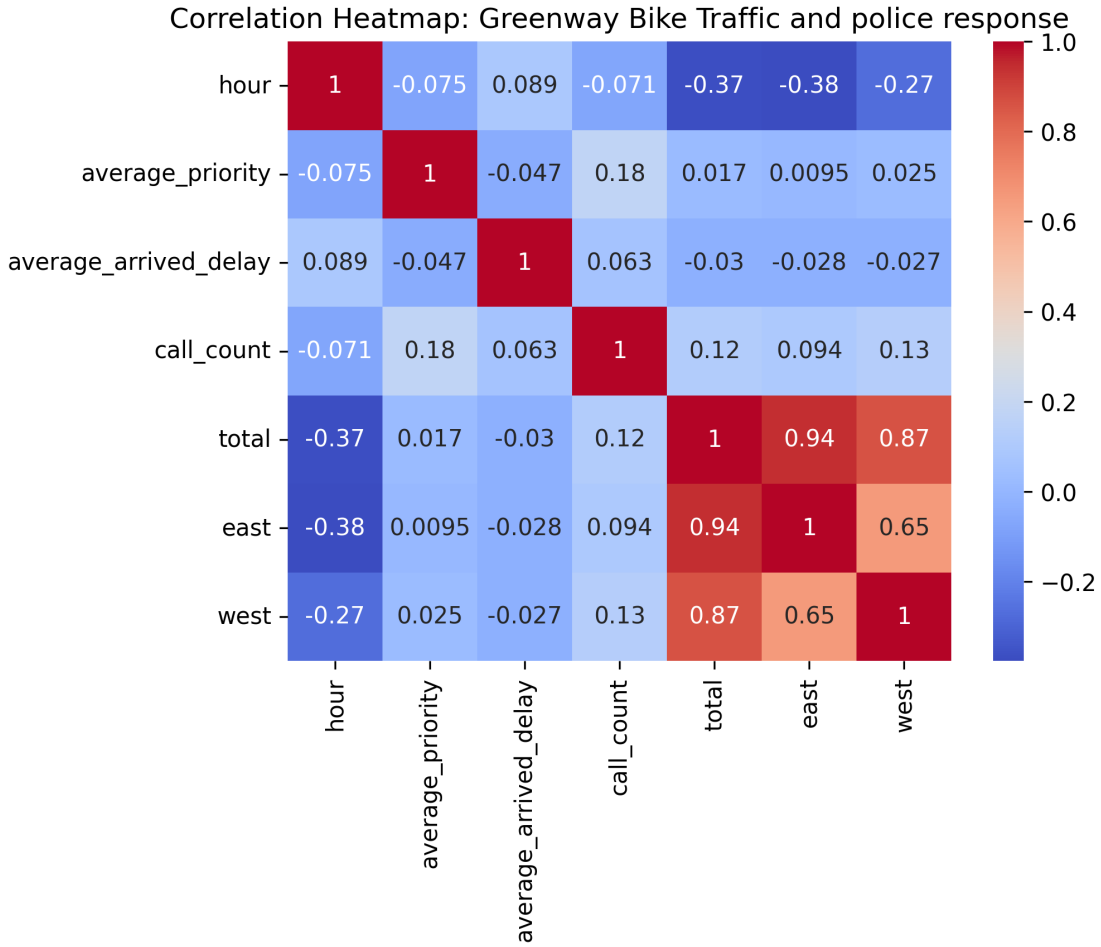


Figure 4.2: Greenway Traffic Correlation Table