Bike Traffic Data Analysis

Ian Denzer (ipdenzer) and Evan Cresta (echesta)

ECE 20875-005 Python for Data Science

12/4/2022

Path #1

Introduction

We have been given data related to the bike traffic across a number of bridges in New York City. The data is collected from 124 days between April 1 to October 31. Data in this time period consists of the date, weekday, temperature high, temperature low, precipitation, total bikers crossed over the Brooklyn, Manhattan, Williamsburg, and Queensboro Bridge, and the total amount of bikers that crossed over all four bridges.

Data Analysis Strategy

With this data we are challenged to try and answer three questions:

1. You want to install sensors on the bridges to estimate overall traffic across all the bridges. But you only have enough budget to install sensors on three of the four bridges. Which bridges should you install the sensors on to get the best prediction of overall traffic?

2. The city administration is cracking down on helmet laws and wants to deploy police officers on days with high traffic to hand out citations. Can they use the next day's weather forecast (low/high temperature and precipitation) to predict the total number of bicyclists that day?

3. Can you use this data to predict what day (Monday to Sunday) is today based on the number of bicyclists on the bridges?

Since the city is looking to represent the overall traffic, the representation would consist of the three bridges with the most traffic. That way the least amount of data is lost. So, for question one the strategy is to plot the individual day traffic for each bridge and compare each bridge to each other. A trendline for each bridge will also be placed on this graph to help see any patterns. On this graph we can look to see if there is any obvious answer to which bridges consist of the mostly daily traffic. Another useful piece of data will be to calculate the total and average amount of traffic over the time given. Both numbers will help with our decision.

For question two, one graph showing the relationship between the total precipitation and the traffic for that day and a second graph showing the relationship between average temperature and total traffic will be the way of approaching this challenge. With the temperature vs traffic graph the average between the high and low of the day is chosen since most of the day the temperature falls between these values. The high and low represent the extremes of the temperature which happens for the least amount of the day. A trendline within these graphs could also help depending on if there is a recognizable pattern in the data.

The final question asks about predicting what day it is based on the amount of people traveling. Gathering the sum of all traffic across all bridges and the day the traffic happens we can plot this data on a scatter plot. This scatter plot with trendlines could show a recognizable pattern and help guess which day of the week it is.

Analysis Results

After following the strategy for analysis one the results were very conclusive. Figure 1 below shows the traffic across all four bridges each day. Looking at this graph it is clear that the picking order for most traffic to least is the William Bridge, Manhattan Bridge, Queensboro Bridge, and then the Brooklyn Bridge.

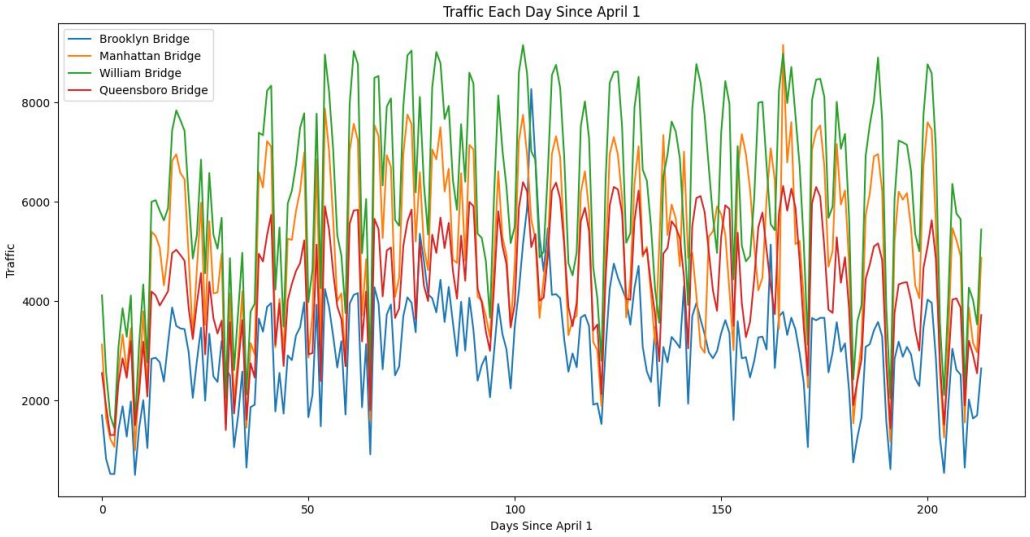

Figure 1: Traffic Each Day Since April 1

If this visual representation of the traffic was not a clear enough indicator more calculations were done to help support this claim. Total and average traffic is show in Table 1 below.

	Total Traffic	Average Traffic
Brooklyn Bridge	648570	3030
Manhattan Bridge	1081178	5052
William Bridge	1318427	6160
Queensboro Bridge	920355	4300

Table 1: Total and Average Traffic for Each Bridge (rounded to the nearest number)

The data shown in Table 1 follows the same pattern as Figure 1. The order for most traffic as mentioned above also applies to the total traffic and average traffic. Using this data, it is suggested that the State of New York put traffic monitoring systems on the William Bridge, Manhattan Bridge, and Queensboro Bridge to best represent traffic in New York. The Brooklyn Bridge contributes the least amount to overall traffic between all the bridges which is only 16% and would lead to the least amount of data lost if not monitored.

Analysis two resulted in there being no way to accurately predict the traffic based on temperature or precipitation. Below is Figure 2. This figure is a scatter plot that shows the total traffic at the average temperature for all 124 days. Figure 3 is just below Figure 2 and shows the total amount of traffic compared to the precipitation for that day.

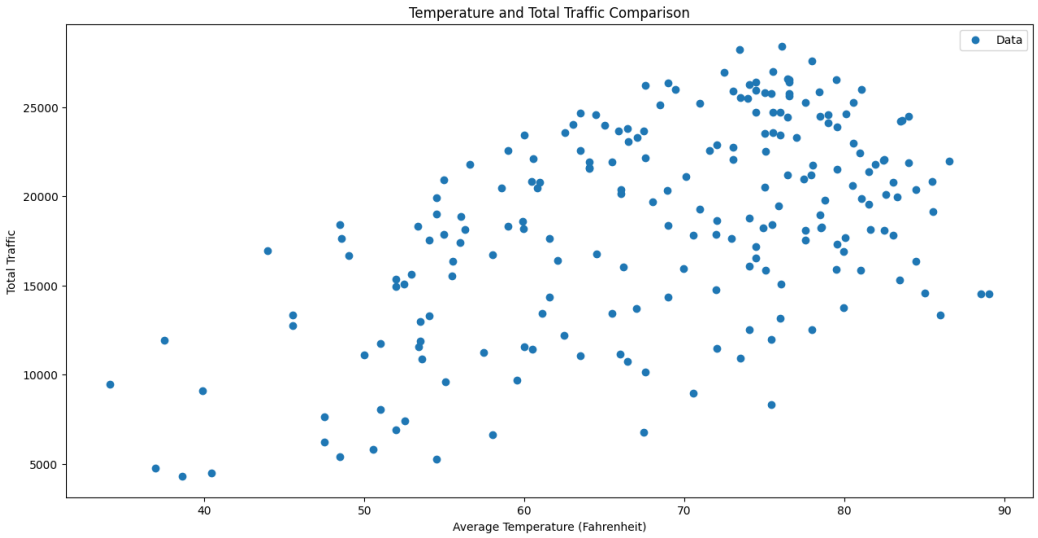

Figure 2 Temperature vs Total Traffic

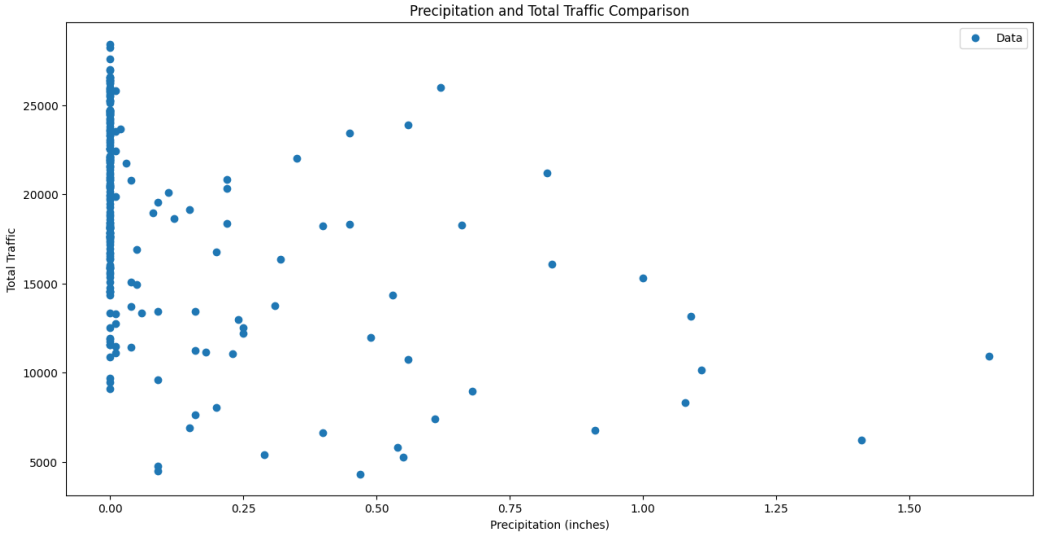

Figure 3 Precipitation vs Total Traffic

In Figure 2 there is a noticeable peak that the data appears to follow but overall, the range of values are considered to be too great to make an accurate prediction. Using a trendline to model the data we do get a r-squared value of around 0.8 but the range of total traffic values is

too great. The data has very large ranges that can cause error of around 5000 people if not more. A general change of traffic based on temperature can be made but not accurately.

Figure 3 does not help with accurate predictions either. 75% of all the traffic is done when there is zero precipitation, and the standard deviation of those values is around 4430 bikers. This is once again too large of a range of error for good predictions. Like temperature a general conclusion about the amount of traffic can be made based upon the precipitation but not accurately. Precipitation is no better of an estimation tool than temperature. Figure 4 can show us a guess can be made seeing that there will be more bikers if the temperature is around 70 degrees on average and 0 precipitation, but actual results can vary by around 4000 bikers. Looking at both data sets and their calculations there is not an accurate way to predict the number of bikers using these data sources.

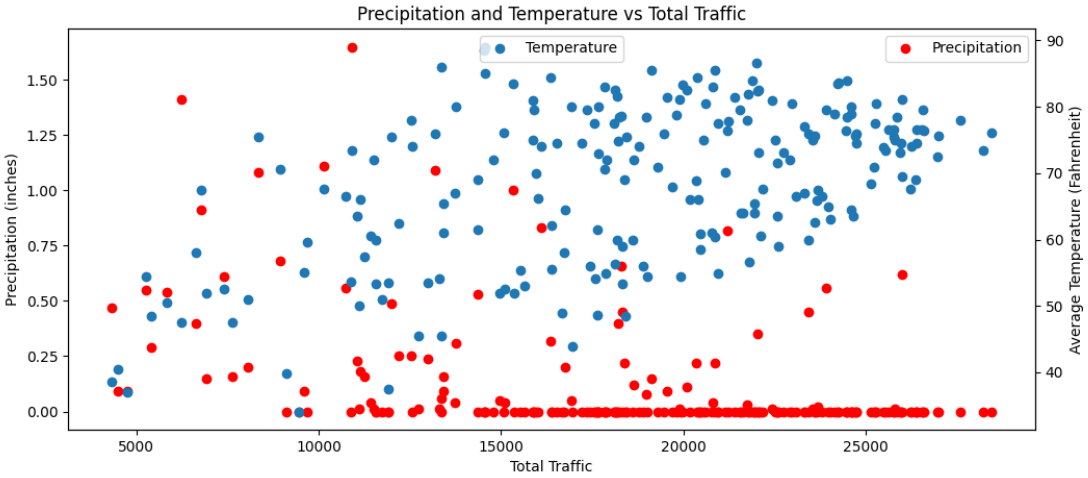

Figure 4 Precipitation and Temperature vs Total Traffic

The final data analysis wants to see if the day of the week can be predicted using the number of bikers crossing the bridges. Below is Figure 5 which shows the average amount of traffic on each day of the week. Below that is Figure 6 which is all the data points that contribute to the totals on each day.

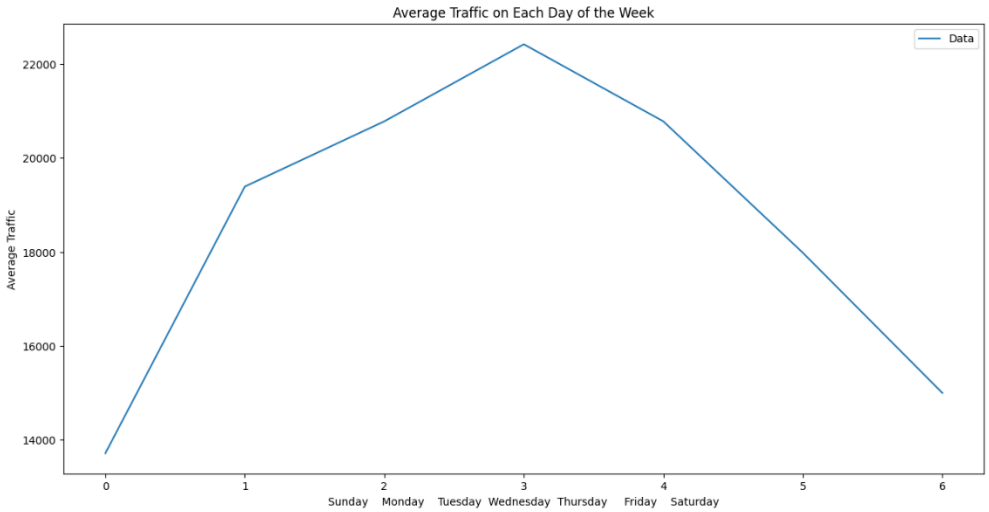

Figure 5 Average Traffic on Each Day of The Week

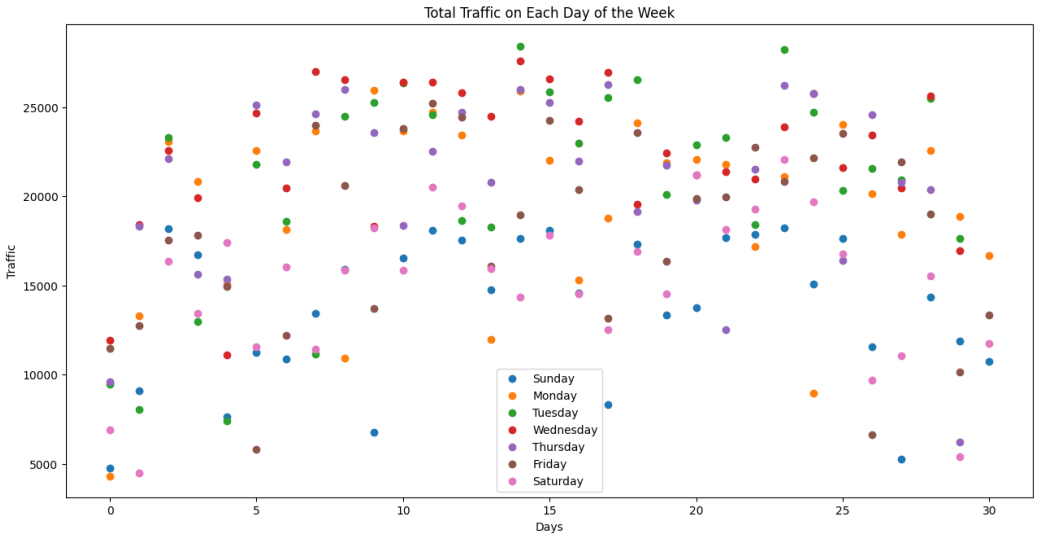

Figure 6 Total Traffic on Each Day of The Week

Observing these graphs, a trend can be seen in Figure 5 and one in Figure 6. Figure 5 shows a promising trend that could be used to make a trendline, but accuracy is this in question. There is not much of a difference between some days of the week on opposite sides of its peak. Figure 6 shows the similarities between all the data set. Most data points are within hundreds of each other. The arch shaped trend can be seen in Figure 7 and 8 but the large ranges of values can also be seen.

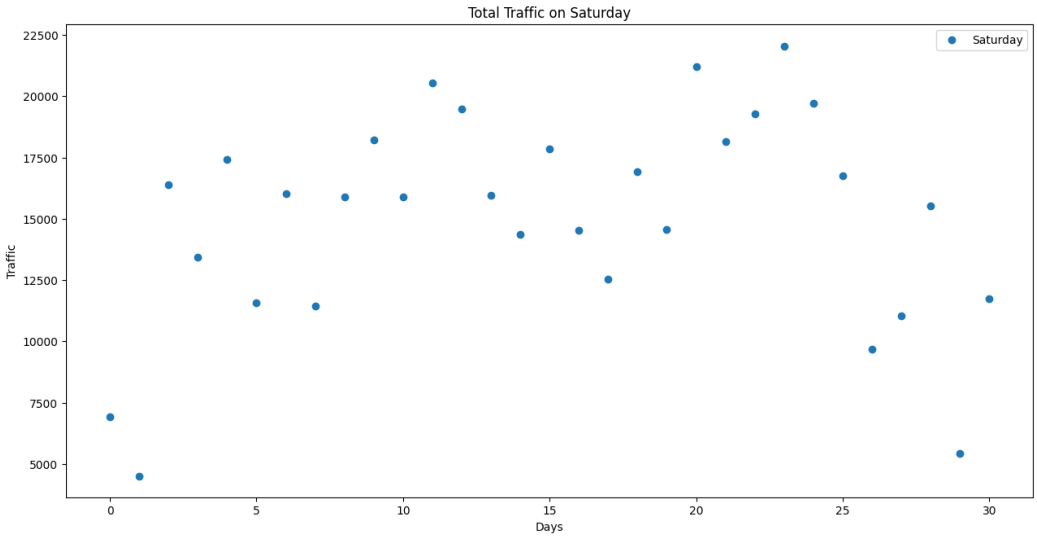

Figure 7 Total Traffic on Saturday

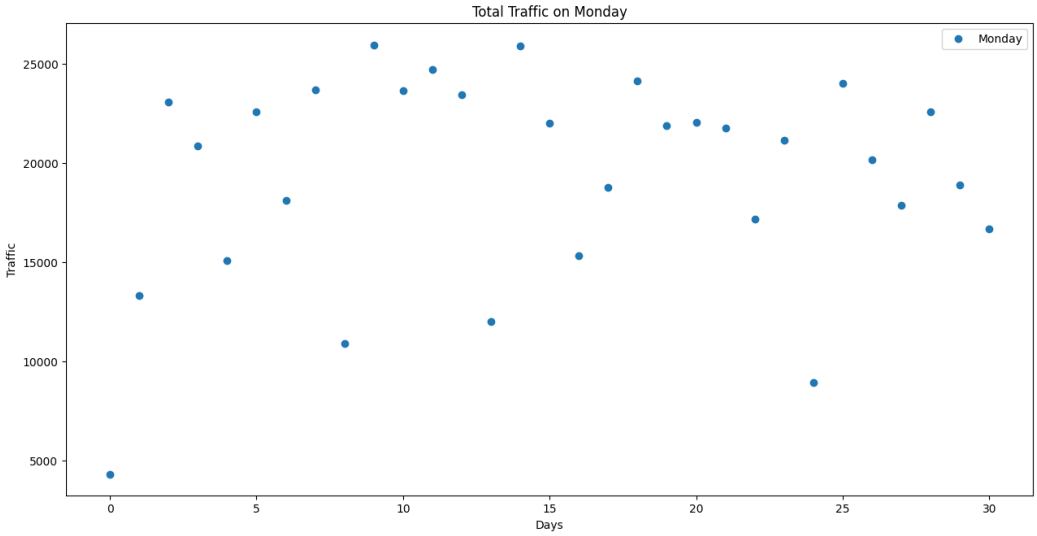

Figure 8 Total Traffic on Monday

More data concluding that some days are too similar can be seen in Table 2. Looking at average traffic for Tuesday and Thursday we can see that the difference between them is only 1 biker on average. In Figure 9 we can see just how similar the data points are. Overall, there is no way to accurately predict what day it is based upon the amount of traffic. A guess can be made as to whether it's a weekday or not but that is all that can be done.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Average Traffic	13716	19393	20782	22422	20781	17984	15000
Maximum Traffic	18249	25965	28437	27579	26263	25228	22055
Minimum Traffic	4759	4335	7400	11105	6248	5387	4402

Table 2 Calculated Data for Each Day

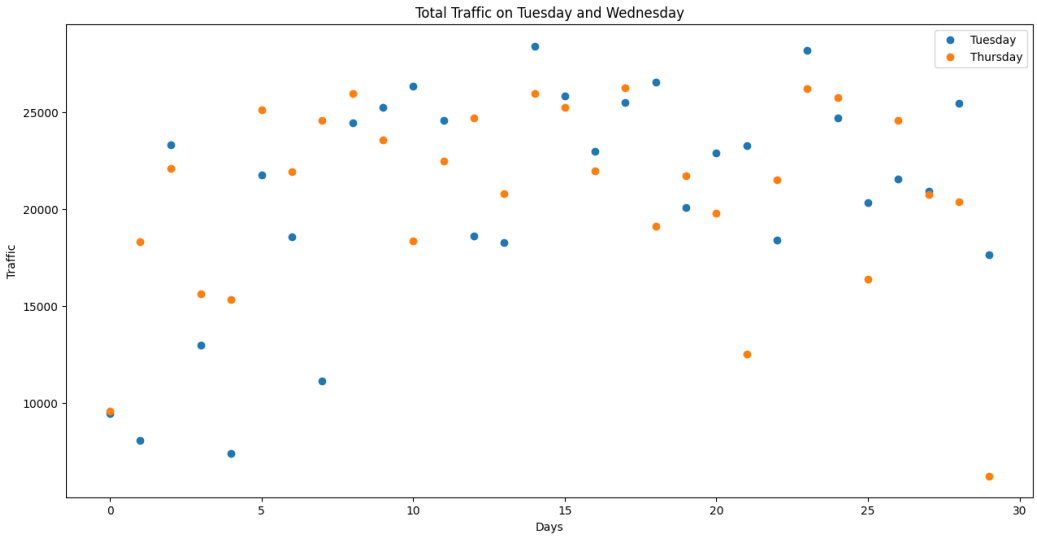

Figure 9 Total Traffic comparison between Tuesday and Thursday

In conclusion analysis one determined the Brooklyn Bridge should not be monitored since it contributes the least amount to overall traffic between all the bridges which is only 16% of total traffic. Analysis two concluded that the data for precipitation and temperature shows no reliable way of predicting traffic accurately. A guess can be made seeing that there will be more bikers if the temperature is around 70 degrees on average and 0 precipitation, but actual results can vary by around 4000 bikers. This is determined to be too big a margin for error. In the end it is up to the State of New York if this range is too great. Analysis three concluded that the average values for the traffic are too similar across all the days of the week. Once again someone could predict how far away from Wednesday the day is based on the data, but they could not know which half of the week it is.