Python Mini Project report

Amaan Sadiq and Justin Hatter

aasadiq, hatterj

Project Path 1

Section 1: Dataset

The dataset in question considers the highest and lowest temperatures of a period from April 1st to October 31st as well as the precipitation on each day. The dataset also includes the average “bike traffic” on four separate bridges. The objective is to use the bike traffic data to find the best three bridges for data collection, use the weather data given for each day to predict how often there will be bike traffic on each bridge, and to use the data to determine the date of any random day

Section 2: Methods

To determine the bridge that would be best to use the three sensors to gather data we needed to find the three bridges that had the most traffic. To do so, we took the average traffic of all four bridges and averaged them to find the most traffic. This resulted in us excluding the Brooklyn Bridge in our dataset due to it having the least traffic and allowing for better data to be combined.

In order to determine how often there will be bike traffic, based on the weather, our method of choice was to create a scatter plot for each weather data, by normalizing the weather data as well as the average bike traffic on each day and find the correlation between each one. We expected the results that the bike traffic would decrease when the precipitation increases and to increase bike traffic when the high or low temperature increases. This allows us to predict what date a certain period of time could be based on the correlation of each of the weather data by calculating the average bike traffic for each individual day and graphing it.

Section 3: Analysis

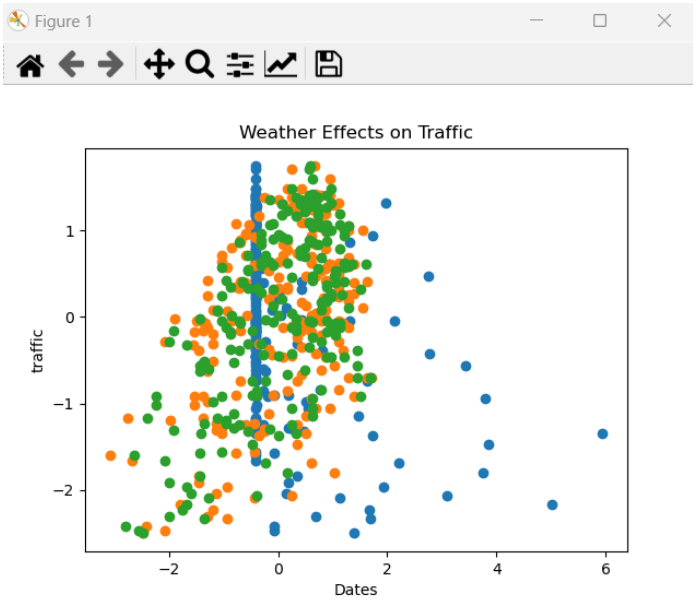

Figure 1: Scatter Plots for each weather data

Blue: Precipitation

Green: High Temperatures

Orange: Low Temperatures

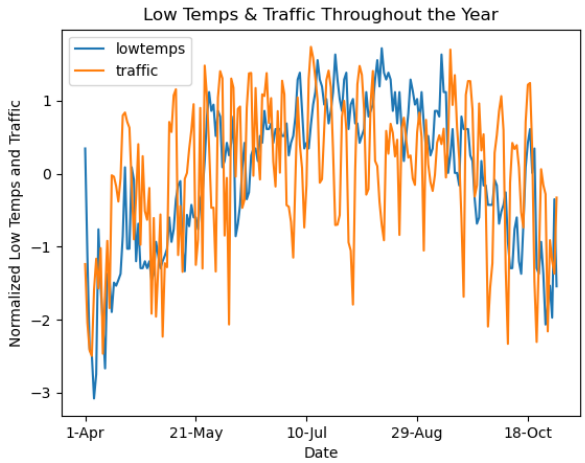

Figure 2: Low Temps and Traffic Correlation (non-normalized)

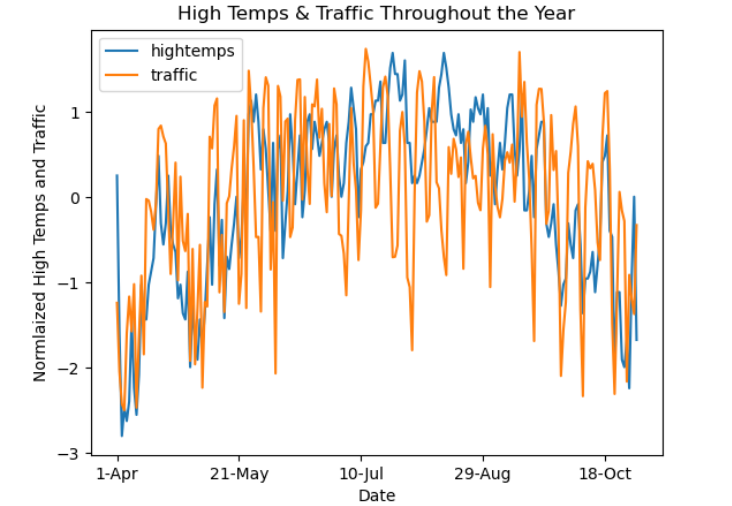

Figure 3: High Temps and Traffic Correlation (non-normalized)

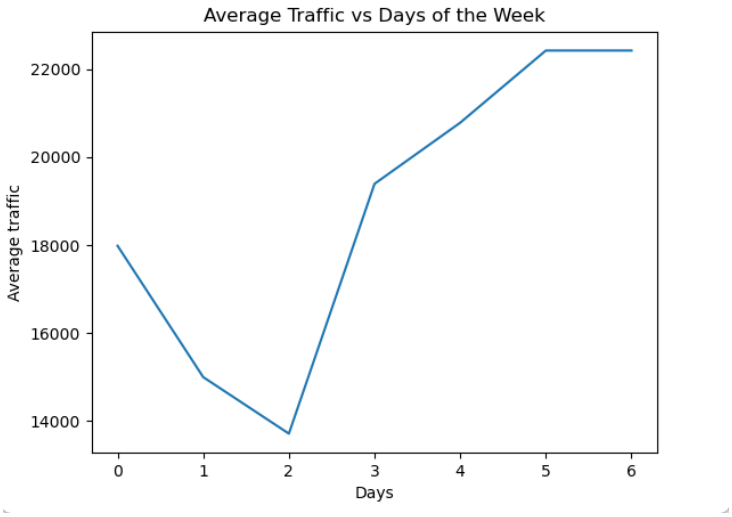

Figure 4: Average Traffic on Each Day

```
C:\Users\amaan\Anaconda3\python.exe C:/Users/amaan/PycharmProjects/project/MiniProjectPath1.py
Avgerages based on days:
Friday:  17984.58064516129
Saturday:  15000.645161290322
Sunday:  13716.387096774193
Monday:  19393.709677419356
Tuesday:  20782.266666666666
Wednesday:  22422.266666666666
Thursday:  20781.3
============================================================
Calculated Correlation Values:
Correlation between weather and bridge traffic:  -0.42071130875526336
Correlation between low temps and bridge traffic:  0.44214931454129536
Correlation between high temps and bridge traffic:  0.5741789145258157
Correlation between days of the week and bridge traffic:  0.37939512898141925

Process finished with exit code 0
```

Figure 5: Calculated Averages and Correlation Values

In Figure 1, the results indicate that there is a positive correlation when both the high and low temperatures increase the amount of traffic displays a positive correlation. This matches the data because typically it is more logical to take a bike ride when the weather is warmer, further proved by our calculated slope in Figure 5. Although, when we normalized the precipitation against the traffic we kept receiving the straight line that could be seen in Figure 1 but when we calculated the correlation values we found a negative correlation which we assumed to be correct. Generally, we assumed a negative correlation because the amount of bike traffic would be lower on a rainy day.

For the first question, we believe placing the three sensors on the Queensboro, Williamsburg, and Manhattan Bridge would result in the best data. When calculating the average bike traffic for each bridge the Brooklyn Bridge had the lowest average bike traffic creating an outlier so we decided to exclude it when calculating our correlations for the weather data.

To answer question 2, yes it is possible to predict the total number of bikers by comparing the correlation values of each scatter plot, given in Figure 5, by graphing the

predicted high and low temperatures as well as the average precipitation for tomorrow it would be possible to predict the number of cyclists total, not for each individual bridge.

For question 3, it would be difficult to predict what day it is based on the bike traffic because the weather is unpredictable and affects the amount of bike traffic. Based on calculating the average bike traffic for each individual day, Figure 4, it is possible to predict what day it is based on the average bike traffic of a certain day.