Jack Huckenpoehler *Path 2* *https://github.com/Huckenpoehler/ECE-20875-Mini-Project*

Abstract

The dataset under analysis describes the bicycle usage of four bridges in New York City: Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge, and Queensboro bridge. This analysis intends to answer a selection of statistically-grounded questions about the dataset and draw intrinsic conclusions.

Analysis

There are three essential questions to be answered from this dataset:

1. If we are to draw conclusions from the four bridges by only analyzing three, which three bridges should be selected?
2. Can weather forecasts be used to predict the total number of bicyclists on a given day?
3. Can the day of the week be predicted based on the number of bicyclists on the bridges?

To answer Question 1, the total sample size for each bridge is to be calculated to determine which bridges see the most traffic. This is done to preserve as much of the original sample size as possible to as to maximize the accuracy of the data analysis' results.

To answer Question 2, the total mean bicyclist quantity will be calculated across two criteria, being the temperature and the precipitation. 'Clear' and 'Rainy' days will be split by whether precipitation was measured on that day. 'Warm' and 'Cool' days are divided by a temperature of 60°F, where warm days are those which feature a high temperature greater than 60°F, and cool days are defined by a high temperature less than 60°F. This question can be answered if a significant difference can be ascertained between the categories.

To answer Question 3, a multi-line graph is to be drawn across every week of the dataset, comparing the totals of all the bicyclists present across the four bridges with the day of the week. From this, a determination can be made if there exists a correlation between the day of the week and the quantity of bicyclists present which can allow predictions to be made.

Addendum

The sample sizes across the three bridges are as follows:

Brooklyn Bridge:	648570
Manhattan Bridge:	1081178
Williamsburg Bridge:	1318427
Queensboro Bridge:	920355

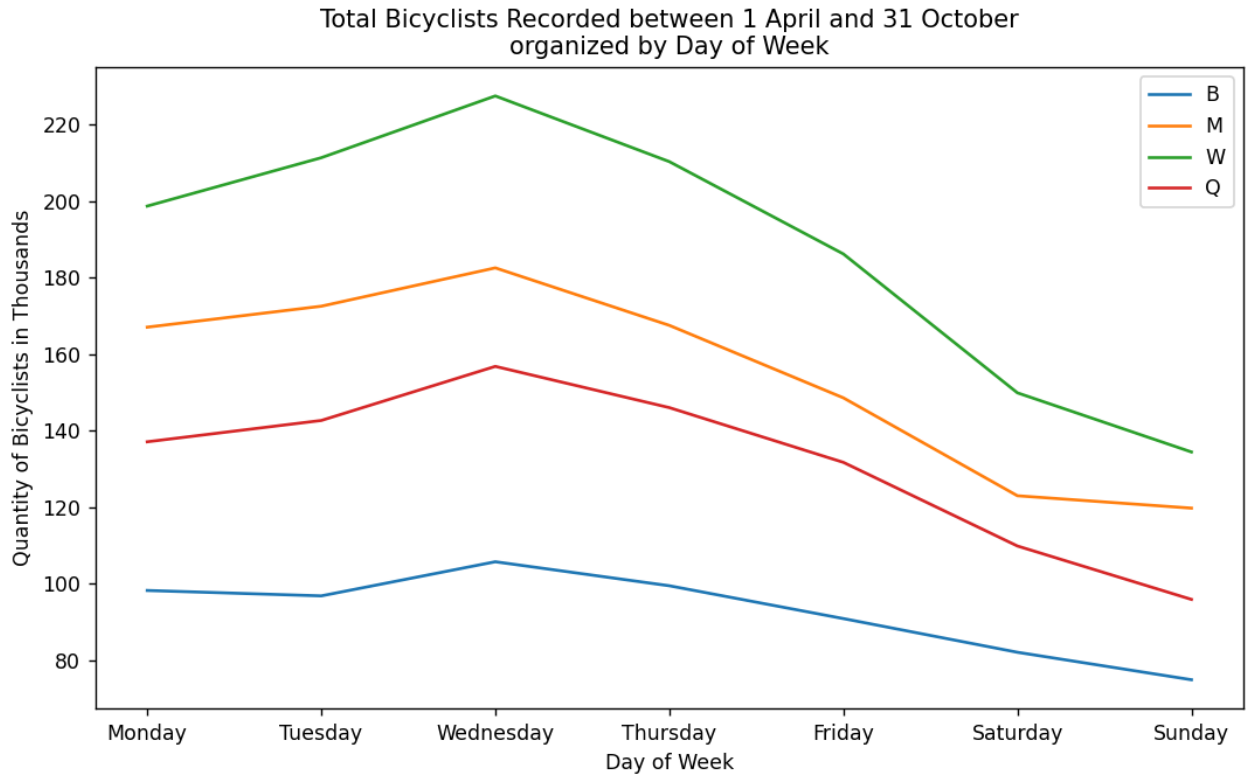
Based on these results, the Brooklyn Bridge is selected for omission provided that it has the smallest sample size. Additionally, the Brooklyn Bridge and Manhattan Bridge are geographically nearby, only separated by a distance of about a mile. Given this, it can be said that the Manhattan Bridge, Williamsburg Bridge, and Queensboro Bridge can be used exclusively for data analysis, with the potential inaccuracies in the data analysis being mitigated by the proximity of the Manhattan Bridge from the excluded Brooklyn Bridge.

Average daily bicyclists organized by weather conditions are as follows:

Clear Day:	20511.71
Rainy Day:	14320.88
Warm Day:	19810.59
Cool Day:	10467.93

These averages reveal that, with rain present, a decrease in bicyclist activity of nearly 50% can be expected. The same decrease can be seen between cool and warm days, with near 50% of the average bicyclist activity favoring temperatures greater than 60°F. These results can support predictions that, with cooler and rainier weather, significantly fewer quantities of bicyclists can be expected.

Plotting the quantities of bicyclists across the four bridges organized by the day of the week presents in the following figure:



It can be seen that the quantity of bicyclists recorded across the four bridges increases towards Wednesday, and experiences a steep dropoff for the remainder of the week. The more popular Williamsburg Bridge sees a decrease of over 30% in its bicyclist activity between Wednesday and Sunday, whereas the less-frequented Brooklyn Bridge loses about 20% of its bicyclist population. While the exact reason that Wednesday is favored by bicyclists over other days lies outside of the abilities of the dataset provided, this analysis does reveal that it is possible to predict the day of the week based on the bicyclist activity by comparing the current activity to the previous day. Taking Wednesday as the maximum for the week, the week progresses with a decrease in activity until it hits the minimum on Sunday, before rising again as the new week begins.