

Section 1

Guanxi Zhou, zhou1139, path 1

Yilong Peng, peng280, path 1

Section 2

The highest temperature is on July 23, the lowest temperature is April 3. Over half of date is 0 precipitation, and the highest precipitation is 1.65 on May 30. The lowest number of the bicycle on Brooklyn bridge is on Apr 9 and the highest is on July 14. The lowest number of the bicycle on Manhattan bridge is on Apr 9 and the highest is on Sep 13. The lowest number of the bicycle on Williamsburg bridge is on Apr 4 and the highest is on July 12. The lowest number of the bicycle on Queensboro bridge is on Apr 3 and the highest is on July 12.

This graph shows the number of bicycles of different bridge on different date.

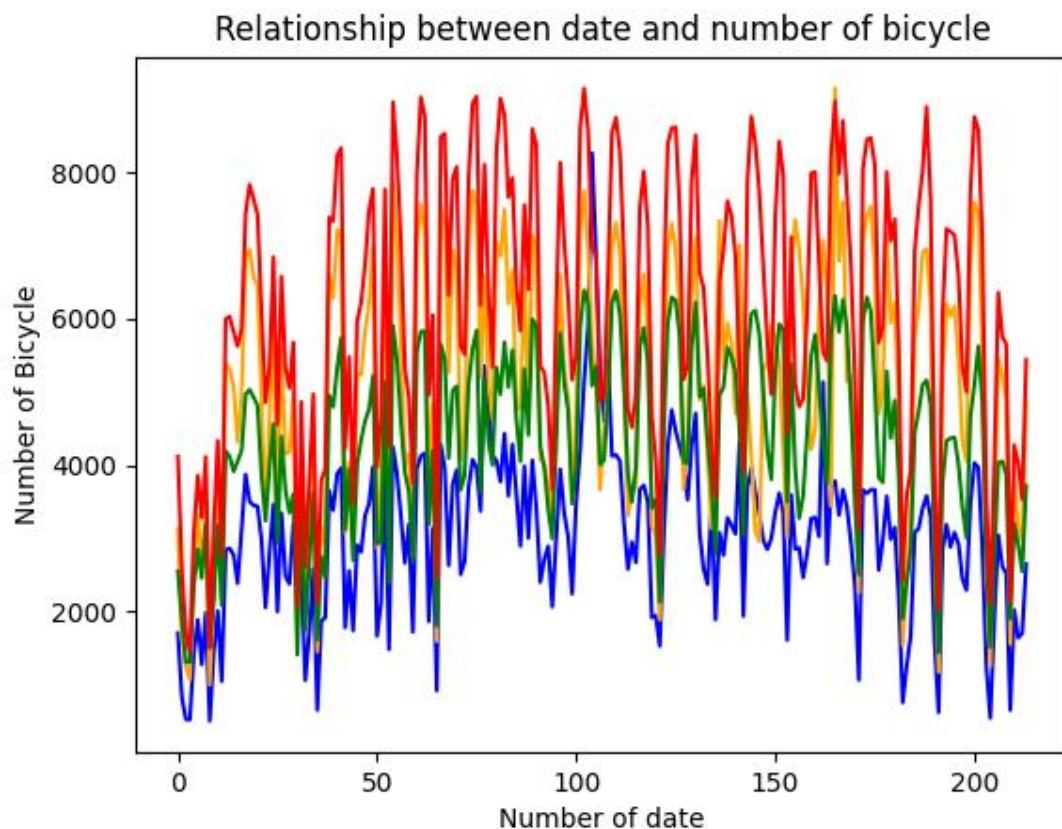
The red line is Williamsburg bridge.

The orange line is Manhattan Bridge.

The green line is Queensboro Bridge.

The blue line is Brooklyn Bridge.

We can see that the Williamsburg bridge is the highest, and Brooklyn bridge is lowest.



Section 3

Problem 1 analysis:

For problem 1, we will choose to use linear regression model. We will use three bridges' traffic data to determine which three bridges are the solution. Since we have 4 different bridges, so we will have different combination.

- i. Brooklyn, Manhattan, Queensboro.
- ii. Brooklyn, Manhattan, Williamsburg.
- iii. Brooklyn, Williamsburg, Queensboro.
- iv. Manhattan, Williamsburg, Queensboro.

We will build the linear regression model for these four different groups of bridges. We will use the model to find the best fits the total bike traffic data. When we are getting the dataset, we can use it to predict the overall traffic. We will find the r squared for each group, r square is ranges from 0 to 1, 0 is mean that the model does not represent the dataset at all and 1 is mean the model is perfectly representing the dataset, so if the r square is closer to 1, that is mean that group will have a better prediction of overall traffic.

Problem 2 analysis:

We can take high temperature, low temperature, and precipitation as input (independent variables). The number of cyclists in four different locations was taken as the output (dependent variable) to set up the model. After setting up the model, we can input different data to predict the number of cyclists that day.

Problem 3 analysis:

We will find the average number of bicyclists on each of the Bridges per day in the week. Depending on the number of cyclists on each bridge, the average number of days on different Bridges obtained in step 1 is similar (close to which day in the week) to determine the corresponding days of the week.