**Path: 1 – Bicycle Traffic**

**Group:**

1. **Miguel Castilho Oliveira**
   1. **Purdue Username: mcastilh**
   2. **GitHub Username: MigCast9**
2. **Pedro Paulo Pinto**
   1. **Purdue Username: pintop**
   2. **GitHub Username: pepabrdp**

***Dataset:***

The dataset chosen gives information of bike traffic across four different bridges in New York City: Manhattan Bridge, Brooklyn Bridge, Queensboro Bridge, and Williamsburg Bridge. Additionally, it also provides information regarding the dates of when each data sample was collected, as well as weather characteristics, such as precipitation and temperature.

***Analyses Methods:***

**1 – Sensor Data**

The first analyses question inquired about the bridges which sensors should be installed on to get a better prediction of overall traffic. Considering there were enough sensors for only three of the four bridges, it was necessary to devise a method to disregard the bridge that least represented the data. To that end, the group decided that the bridge that presented the average daily bike traffic furthest from the average of all average daily bike traffic combined to be the one that would be eliminated.

For this, the group should calculate the average daily bike traffic for each bridge by summing their daily bike traffic and dividing it by the number of days. After calculating that number for each of the four bridges, it is necessary to calculate the total average daily traffic by summing up the numbers just found and dividing it by four. In sequence, all that is left is to check which average daily bike traffic is furthest from the total average daily bike traffic.

**2 -** **Weather forecast to predict number of bicyclists**

In this problem, it is necessary to consider the possibility of predicting the number of cyclists on the following day based on the weather forecast. To that goal, what the group decided to do was a linear Ridge regression in which it will be necessary to use the values of Low Temperature, High Temperature, and Precipitation as the X parameters in order to try to predict the corresponding values for total traffic. In sequence, it is possible to compare the predicted total bike traffic to the actual values and calculate the Mean Squared Error to have some diagnostic of the model as well as the coefficient of correlation to see how effective the model is in predicting the correct values.

The team does not believe the weather forecast will be a good indicator of the number of bikers.

***Results:***

**1 – Sensor Data**

Through the implementation of the described method, the average daily bike traffic for the Manhattan Bridge was of 5052 bicycles, for the Brooklyn Bridge was of 3030 bicycles, for Queensboro Bridge was of 4283, and for the Williamsburg Bridge the average daily bike traffic was of 6160. After obtaining those values, the total average daily bike traffic obtained was of 4636 bikes, and the bridge that deviated the most from such value was the Brooklyn Bridge. Therefore, the bridges chosen to have the sensors installed are the Queensboro, Manhattan and Williamsburg Bridges to get the best sense of overall traffic.

**2 -** **Weather forecast to predict number of bicyclists**

After running through the described process of separating the data into training and test data in order to create a model and test it with the test data, it was possible to arrive at the follow model:

Total Traffic = ﻿-1942.78367 \* x1 -1543.50895 \* x2 + 4347.27326 \* x3 + 18412.6

As seen by the graph below, this model is obtained by the lambda value that creates the smallest Mean Squared Error, in which the Lambda value is 0.1 and the MSE is ﻿13961.

Chart

Description automatically generated

As observed, the MSE value obtained is extremely high, and moreover, the correlation coefficient yielded is at a low of 0.159, meaning that the weather cannot be used to accurately predict the number of bikers on a certain day.