Derek Huang (huan1093)/ Christina Lo(lo67)
ECE 20875 Project: Path 1

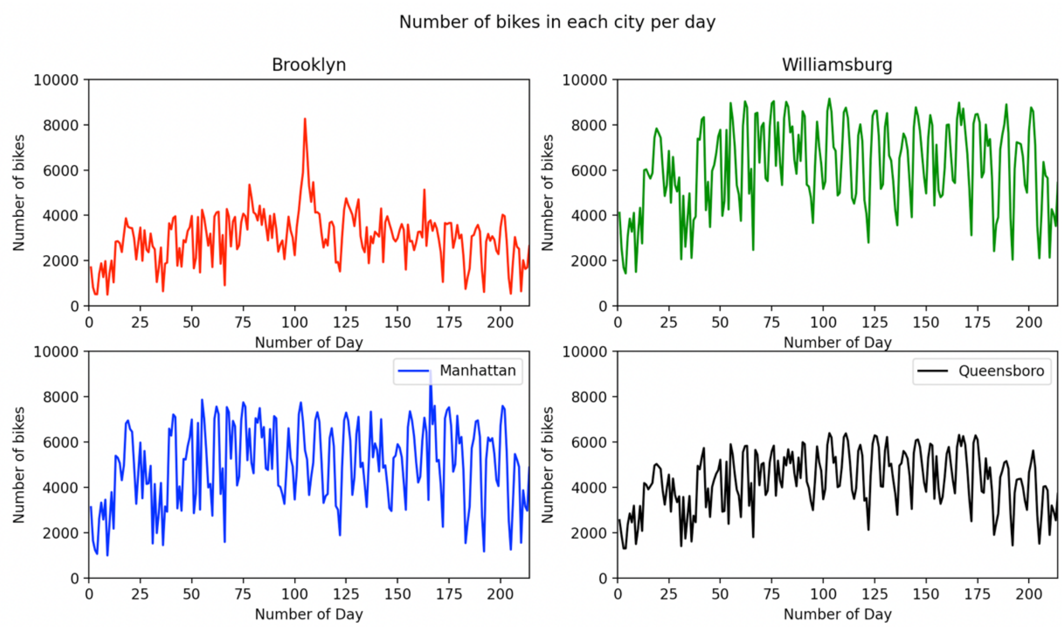
Data Sets:

We have chosen Path 1, the Bike traffic problem, and therefore we're using the 'NYC_Bicycle_Counts_2016_Corrected.csv'. This is an excel document that shows the date, day, the highest and lowest temperature on that specific day, the precipitation rate on a specific day, and the number of bikers observed in each city of New York during a specific day.

Problem1:

The first problem asks to install sensors on the bridges to estimate the traffic across all the bridges, however there's only enough budget for 3 sensors. We are supposed to determine which 3 bridges we are supposed to use install the sensors on. One way to do this is to look at the distribution of the number of bikers per day on each of the 4 bridges and observe if there's a similar pattern on 2 of the bridges. If there is a similar pattern on the bridges, then we should install the sensors on only one of them, because we want to cover all of the different distributions. For this problem, we plotted the number of bikers on each day at each of the four bridges. This shows the distribution of the bikers and we can therefore visualize how the number of bikers differ between each city. Figure 1 shows the results of the distribution.

Figure 1.



From looking at the distribution, one can tell that the distribution of bikers in Williamsburg and Manhattan are somewhat similar. Therefore, only one of the two cities, Williamsburg and Manhattan, needs a sensor installed. Due to the number of bikers per day, it would be more beneficial to place the sensor on Williamsburg because there are more samples from that city, as observed from the distribution. In conclusion, we suggest installing the 3 sensors on Brooklyn, Williamsburg and Queensboro to get the prediction of the overall traffic.

Problem2:

We are asked to see if the police officers could use the next day's weather forecast to predict the number of bicyclists that day. We decide to look at the relationship between the high temperature and the number of bicyclists, and the relationship between low temperature and number of bicyclists, independently. From figure 2 we can see the relationship between high temperature and the number of bicyclists. From figure 3 we can see the relationship between low temperature and the number of bicyclists.

Figure 2.

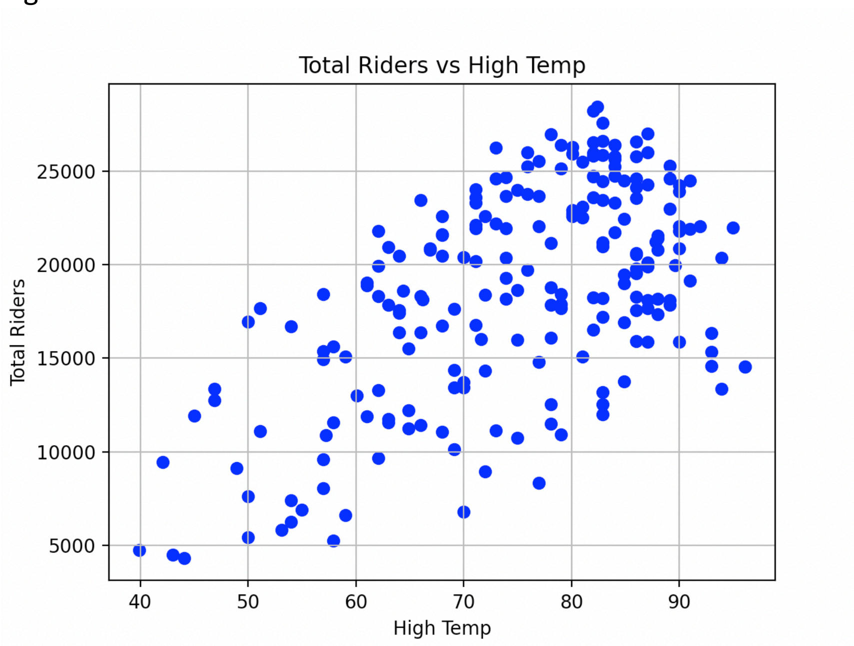


Figure 3.

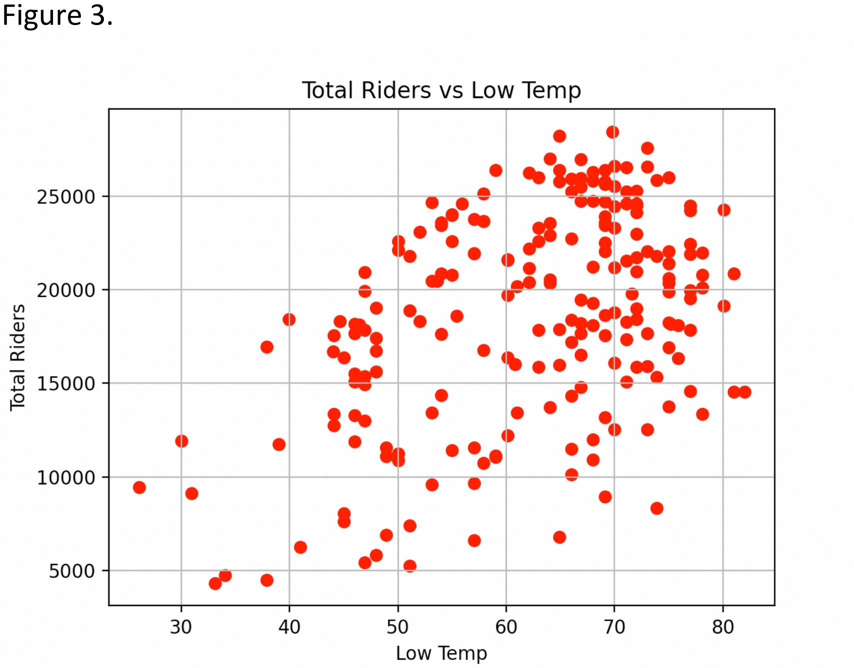


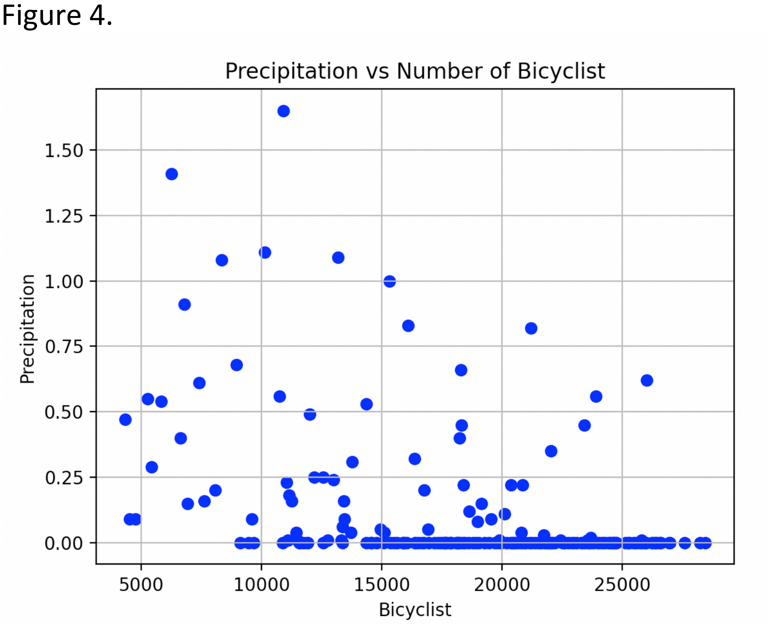
Figure 2 and Figure 3 both have a linear looking relationship between the number of riders and the temperature. We can predict that on warmer days, there are usually more bicyclists. This could be due to the factor of a nice weather encourages people to bike more. The police officers could assume there would be more bicyclists on a day where the low and high temperature lies between 70-90 degrees Fahrenheit.

However, if we were to predict the number of cyclists on a day given the high and low temperature, we would need to construct a regression analysis to estimate the number of riders. We used the linear regression analysis and calculated the coefficients and intercepts for the two-variable equation. The coefficient for the high temperature is 483.61, the coefficient for the low temperature is -260.88 and the intercept is -1526.68. The equation is $y = 483.61x1 - 260.88x2 - 1526.68$, where x1 is the high temperature and x2 is the low temperature for the day. Furthermore, we constructed an R squared analysis to see if our model could accurately predict the number of bikers given the high and low temperatures. The R squared value was 0.375, which is fairly low and suggests that the equation might not be very suitable for predicting the number of bicyclists given the temperature. This would go against our assumption, since from figure 2 and 3 we could observe there is a relationship between the temperature and number of bicyclists. However, this could be due to the fact that the precipitation rate was not considered. People tend to not ride bicycles on a rainy weather due to the inconveniences, and thus even though we can see a pattern between warmer weathers and the number of bicyclists, the precipitation rate also affects the number of bicyclists on a specific day.

Problem3:

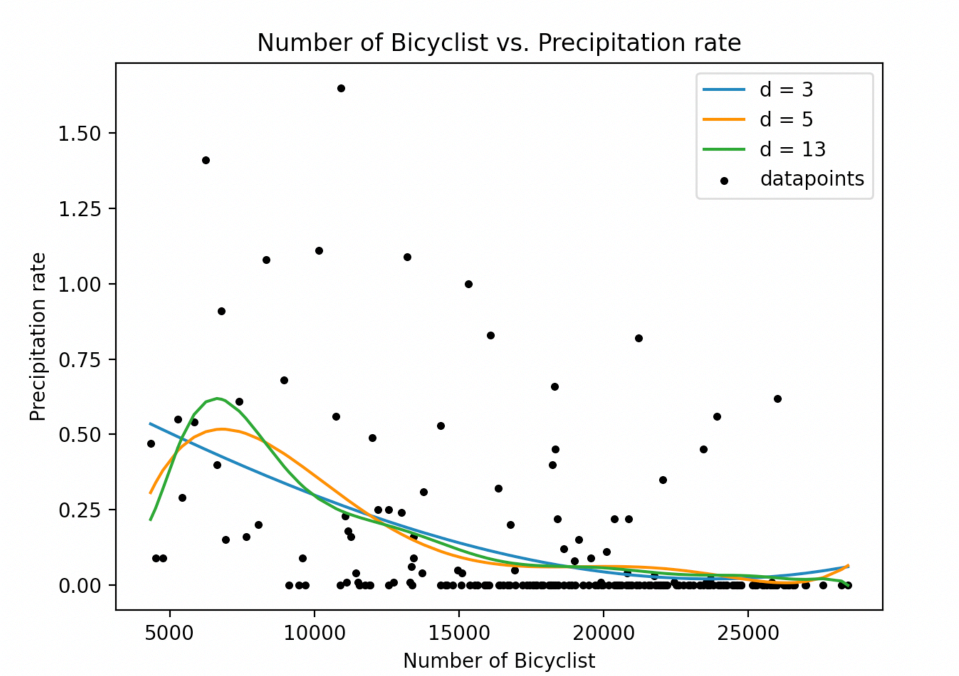
This problem asks if we could use the number of bicyclists on the bridges to predict whether it is raining. The independent variable here would be the number of bicyclists, and the dependent variable would be the precipitation, since we are using the number of bicyclists to predict the precipitation rate. To show the relationship between precipitation and the number of bicyclists, we first created a scatterplot to observe the overall relationship shown in figure 4.

Figure 4.



It can be observed that there is an association between the number of bicyclist and the precipitation rate. There tends to be more bicyclists on days when it is not raining, and the vice versa, according to figure 4. To find the equation to predict the precipitation rate, we constructed a regression with different degrees to find out which equation best estimates precipitation. The graph in figure 5 shows a use of different degrees in attempt to find the best fit line and therefore, the best prediction for the precipitation on a specific day, using the number of bicyclists.

Figure 5.



As seen in figure 5, the higher the degree, the better the line fits and therefore the more accurate the equation predicts. However, the line of fit looks similar after a large degree, such as 13, and therefore it is not significant to keep increasing the degree because it is substantial to use a lower degree. Using the 5^{th} degree, we get an equation of y = 3.305x^5e-21 – 2.954x^4e-16 + 1x^3e-11 – 1.573x^2e-7 + 0.001x – 2.167. We simply just plug in the total number of bicyclists on the bridges to get the precipitation rate. The equation might not be an accurate prediction of the precipitation, but from Figure 4 only, one could assume that there will be less bicyclists on a rainy day, and there will be a lot more bicyclist on a sunny day.