Skyler Svanda

ECE 20875

12/9/2021

Python Mini Project

Github Username: ssvanda

Path: Bike Traffic

Skyler Svanda

ECE 20875

12/9/2021

Dataset:

I chose to work with the data set for the New York City cyclist count on the Brooklyn Bridge, Manhattan Bridge, the Williamsburg Bridge, and the Queensboro Bridge. The dataset also included the day, date, high and low temperature, as well as percipitation.

Skyler Svanda

ECE 20875

12/9/2021

Chosen Form of Analysis:

For the first problem, I chose to compare the standard deviations of the total number of people who cross each respective bridge. In doing so, I will be able to determine which bridges have the most variation in population, thus calling for the need for sensors to be installed in order to more accurately predict traffic. The bridge with the lowest standard deviation or spread will not need sensors installed because that bridge is already relatively predictable in how busy it will be.

For the second problem, I administered a two-sample Z test, using the high and low temperatures as conditions for the total number of bikers. By taking the mean of the high and low temperatures, we can compare this to the true mean temperature regardless of high or low. After doing so, we map the values for the number of people who bike during high and values in order to determine the differences in the groups of people who go out in the cold vs the heat. This analysis will tell us if there is a difference in the number of people who go biking in the cold and the heat.

In the third problem, we use the data to predict the number of cyclists. By comparing the mean number of people who bike in the rain versus the mean number of cyclists on the bridges regardless, we can run a two sample Z test on how far the total number of rain cyclists strays from the average number of cyclists. By creating an extra dataset of rain cyclists, we can compare this against the true dataset and find a correlation between whether or not it is raining and the number of bicyclists.

Skyler Svanda

ECE 20875

12/9/2021

Analysis:

For our first question, I found the mean, standard deviations for all the bridges to be 1134.04, 1745.48, 1910.64, and 1260.98 across the Brooklyn Bridge, the Manhattan Bridge, the Williamsburg Bridge, and the Queensboro Bridge respectively. Now that we have this information, the Brooklyn Bridge has the smallest standard deviation, leading us to know that the total number of people who cross over the Brooklyn Bridge has the least amount of variability in comparison with the other bridges who all have a larger standard deviation. The sensors need to be installed on the bridges with the most unpredictability in order to get the most use out of the sensors. By placing this hardware on the remaining bridges, the day to day totals of people can be more accurately predicted and allow for the public to time traffic across bridges with more variability in population.

The second question revolved around predicting the number of bicyclists based on the day's temperature readings. We started by splitting the data set into two different populations around total average temperature across high and low for all days. After obtaining that mean, we created a data set for the high temperatures above that mean and then the low temperatures below. We then performed a two sample Z test and found the difference between the number of people who go out on hot days and cold days, finding that there are about 3022 more people who go out on hot days. With a standard deviation of 657 and a Z score of 4.66, we obtained a P value of 3.06e-06. At a significance level of .05, we can determine that our alternate hypothesis is true, and that the means are different. Since the means are different, we can determine that it is possible to use the weather forecast to predict the number of cyclists on the bridges for any given day.

Skyler Svanda

ECE 20875

12/9/2021

The final question asks if it is possible to use precipitation in order to predict the number of cyclists. In order to do this, another two sample Z test was performed on the initial population, and then the population of people who went out strictly on rainy days. The difference in the two population means is about 4223 people while the standard deviation of this difference is 805. Already we see that the difference between the two means is massive, especially when we look at the standard deviation. From this data it is clear to see why we get a Z score of -5.24 and a p value of 1.598e-07. At a significance level of .05 it can be seen that our alternate hypothesis is true: The mean of the total number of cyclists is different than the mean of the number of cyclists who cross the bridges in the rain.