DS 241 – Introduction to Data Science Final Project Report

Nathan Collinsworth, Bryce Butler, Thomas Robbins, Eric Fingar

Identification of sub-problem:

We chose to analyze the connection between ridership and bikeshare location's distance to metro stations. We believed that through analysis we would see that ridership increases as the distance to a metro station decreases.

Recommendation to increase ridership:

Our recommendation to increase ridership would be to strategically place more bikeshare locations near metro station entrances. This would attract more users who find the convenience of biking to and from metro stations appealing. This would also increase the total number of bikes available to be used at the more popular bikeshare stations.These strategies allows more people to integrate the use of bikeshare bikes into their daily commute.

Explanation of assumptions and data used:

We assumed that riders of bikes are also riders of the metro, and use bikes to travel shorter distances to and from the metro compared to the distance they travel on the metro.

Figure #1:

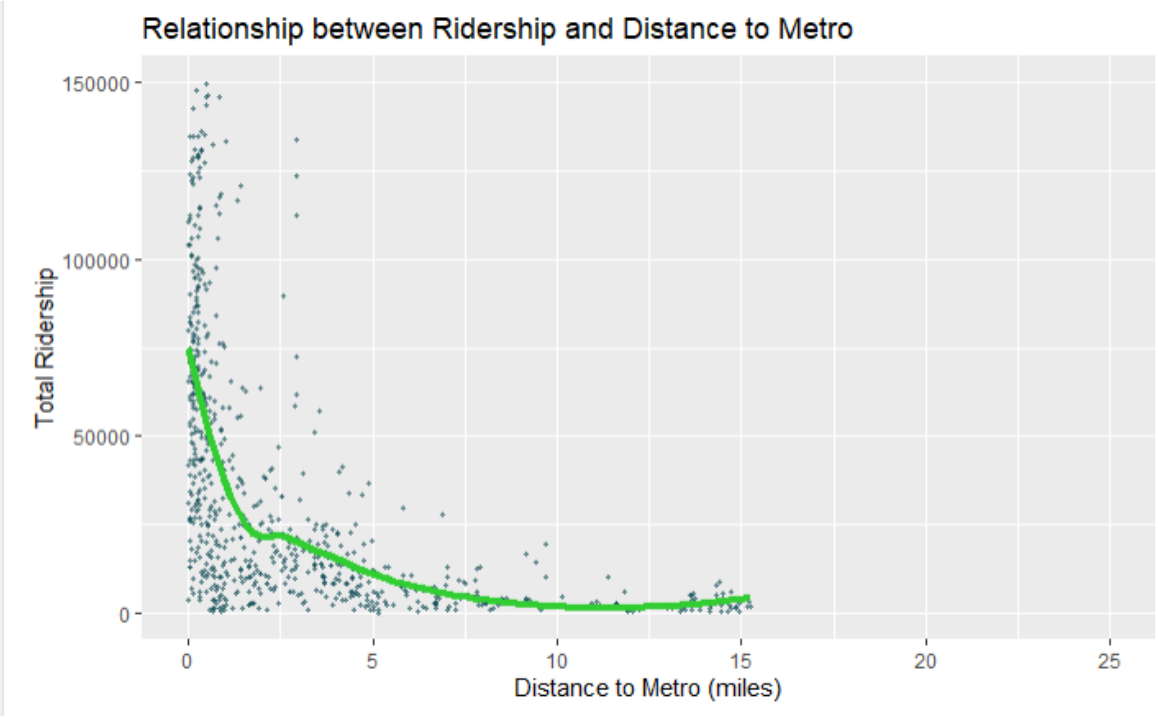

Figure #1 shows the correlation between bikeshare station ridership and their proximity to metro stations. This graph was constructed using the ggplot2 library, the x-axis represents the distance to the nearest metro station and the y-axis depicts the total ridership at each bikeshare station.

Figure #2:

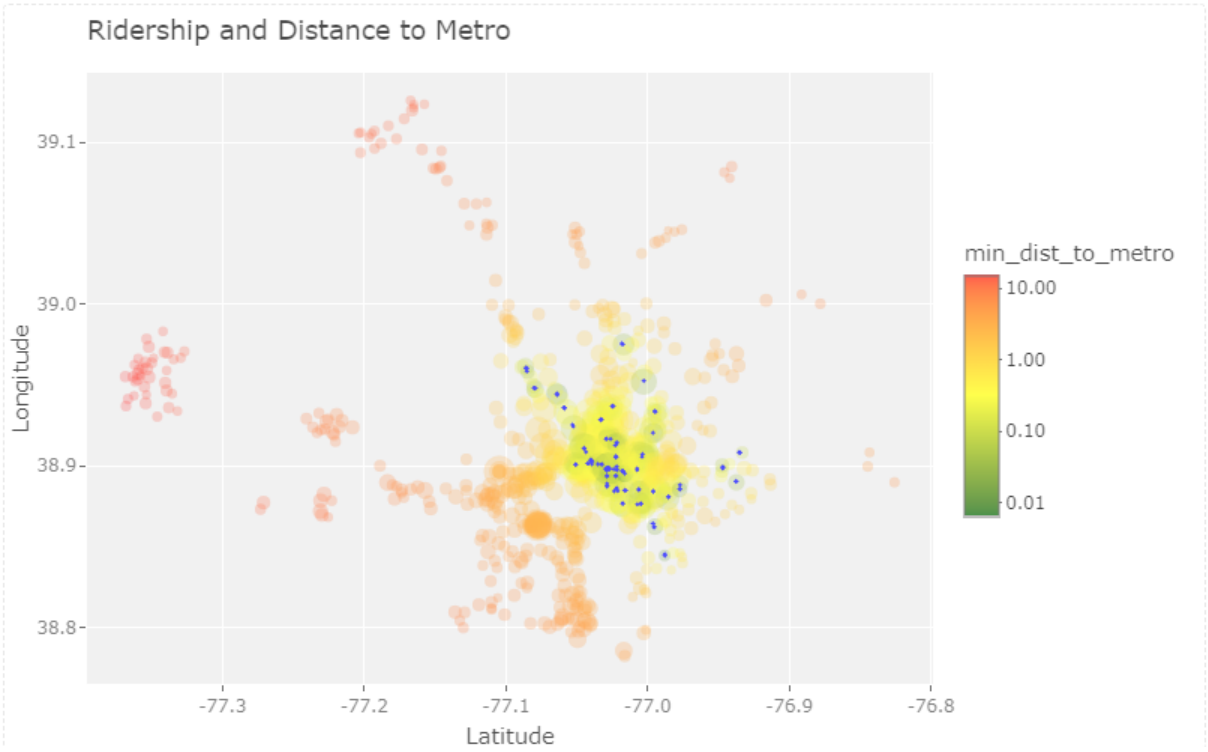

Figure #2 is a geographic plot rendered using Plotly and shows a geographical perspective on bikeshare and metro stations in Washington DC. Using ggplot2 initially, the map displays bikeshare stations as points, with color indicating their distance to the nearest metro station. The blue points denote the locations of metro stations.

Outliers in Data:

The outliers present in our dataset are shown in the map clearly as larger red/orange circles. The was predicted by us and is assumed to be a a result of a combination of communities that do not need use of the metro as frequently. This could include college campuses, local shops and store in a suburban area, and other similar outliers. This deos not impact or takeaway from out analysis as these outliers were assumed to exist and do not take change our overall reccomendation