Jonathan Hernandez

DATA698 Project Proposal Draft

NYC Traffic Fare Levels and Volume

Introduction:

New York City’s Metropolitan Transit Authority (MTA) has been around for many years and heavily controls the majority of transportation of NYC. Anyone who has lived in NYC or even visited the city that never sleeps knows that besides taxis, MTA transportation is always available if needed to move around. It is also perhaps the most popular and busiest transportation system in the country. While many of New Yorkers are dependent on the MTA for commuting, I’ve decided to examine some trends regarding the MTA and see how data can help in learning how traffic fares affect volume and revenue around NYC. Perhaps someone at the MTA can use these findings to see how much subway fare hikes will affect commuters or can focus on more activity on the less frequent tunnels and bridges.

Data Acquisition:

Data will be acquired from Kaggle (website for data science competitions and datasets) and various sources for fare prices over time for public transportation and tunnels and bridges.

<https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data>

<https://mashable.com/2015/03/22/new-york-city-subway-fare/>

Hypothesis:

My hypothesis is that increasing NYC fare hikes over time reduces volume and thus NYC transit revenue. One reason to believe this hypothesis is that as prices increase, people may not want to use the transit system as much as when it was cheaper. Rather than use NYC public transporation, one could use other cheaper (or faster methods) of getting around like bikes or scooters. By fare hikes, I am referring to hikes in not just buses and trains, but in MTA tunnels and bridges. One way to look at this would be to append the fare prices throughout history in the data and run simulations on how transit revenue is affected as a function of fare prices and traffic volume.

Tools and Approaches

The language of choice will be in Python. Various libraries such as Numpy, scipy, matplotlib, pandas, seaborn will most likely be used. Other packages or repositories may be needed for creating visualizations of the data for various simulations.

For the approach of how to create simulations and model, different algorithms like linear/non-linear regression, random forests, neural networks, support vector machines will be looked into.

Looking at data from metro card swipes as well as tunnel and bridge toll transactions we can see how revenue is affected.