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. 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 the sources below that includes NYC MTA tunnels and bridges and metro card usage, NYC traffic accidents (which can also be considered as traffic congestion) and, carbon emissions over time.

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

<http://web.mta.info/developers/data/bandt/trafficdata.html>

<https://data.ny.gov/Transportation/Fare-Card-History-for-Metropolitan-Transportation-/v7qc-gwpn>

<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions-Crashes/h9gi-nx95>

[https://nyc-ghg-inventory.cusp.nyu.edu/#data](https://nyc-ghg-inventory.cusp.nyu.edu/" \l "data)

Hypothesis and Model Simulation:

The hypothesis will be finding a optimization model to simulate toll prices that maximize revenue and mass transit volume while minimizing traffic congestion, accidents and/or carbon emissions. This model can motivate NYC residents and Metro area commuters to use mass transit versus using vehicles. Doing this can help to reduce the city’s use of carbon emissions and footprint and reduce traffic congestion.

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.