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 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 and Model Simulation:

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. 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.

The model that will be looked into will be a model that computes optimal fares to maximize revenue and mass transit usage while minimizing traffic flow. This will be a optimization model that will include several features such as fare prices, time, location, fare usage. The fare card history and hourly traffic on MTA bridges and tunnels dataset from the Kaggle MTA site will be used in these simulations and hypothesis testing.

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.

GENERAL COMMENTS

Jonathan,

I still don’t think your hypothesis (NYC fare hikes over time reduces volume and thus NYC transit revenue levels) is valid. Yes, subway ridership has declined since its peak in 2015, but according to the MTA, that’s more a function of fare evasion, service disruptions from planned maintenance and repairs, poor service and the expansion of ride-hailing services https://www.osc.state.ny.us/osdc/rpt8-2019-mta-financial-outlook.pdf

If I were you, I’d focus more on bridge and tunnel toll rates, and build an optimization model (based on historical fare rates, revenue volume and any traffic congestion data that you can find (perhaps traffic citations or accident rates as a proxy for heavy motor vehicle traffic, carbon emission release amounts?) to simulate the optimal toll rates that maximizes revenue and mass transit usage, while minimizing traffic congestion, accidents, and/or carbon emissions.

The hypothesis would be to use machine learning to find the optimal (either surge/congestion or straight) pricing to motivate NYC residents and Metro areas commuters to use mass transit versus private motor vehicles – thus achieving city planning goals of reducing the city’s carbon footprint as well as traffic congestion.

A