Module 1 Summary- Snizhana Kurylyuk

Abstract

The goal of this project was to use New York subway data to deliver a preliminary analysis for my client. My client wanted to know where the perfect location in New York was for a fitness studio. I worked with Turnstile Data from September 4 - October 2 to address their request. After refining my raw data, I built a visual presentation to displace the top 3 perfable locations based on ridership on various days of the week and times with peak movement. In addition, to cleaning, exploring, and aggregating the MTA data, I applied my own analysis, future prediction and recommendation for my client.

Design

A client wants to open a fitness studio in New York and reached out to see if I can utilize MTA subway data to predict which proximal location is best for her studio. She would like to place her studio as close to a station as possible. In addition, she would like to know the peak days and times to schedule classes.

Data

The MTA Dataset contains millions of data points collected weekly going back to May 5, 2010. I originally began my project with 3 months of data. However due to changes in COVID regulations and companies returning back to work, I reduced my data set to the 5 most recent weeks to allow for a more accurate analysis. This consisted of 1,049,604 raw data points.

Algorithms

(In this project there were not many algorithm options since we are working with exploratory data analysis)

- Numpy and Pandas
- Matplotlib and Seaborn
- SQL and SQLite
- SQLAlchemy

Tools

- NumPy and Pandas for data manipulation
- Google maps for geographical visualization
- Matplotlib and Seaborn for plotting
- SQLite for data ingestion, storage and extraction
- SQLAlchemy for extraction into python

Communication

My project was completed in line with time requirements and can be viewed here.