

MTA Project Write-Up

Liveboard Advertising Pricing

Abstract

Across New York MTA stations, OutFront Media has a network of liveboards, digital advertising boards that are often arranged as a triptych and allow full motion or static video content. The goal of this project was to explore foot traffic levels across MTA stations to determine ways in which OutFront Media could optimize liveboard advertising space pricing in the year ahead. Using MTA turnstile data, I found the highest traffic stations (large commuter stations), the highest traffic days of the week (weekdays), and the highest traffic times of the day (evenings), recognizing interesting patterns emerge in foot traffic before and during the pandemic.

Design

Given that liveboard advertising spaces are available for purchase at different stations, on different days, and at different times, the project's guiding goal was to find the highest traffic stations, highest traffic days of the week, and highest traffic times, and compare data from before and during the pandemic to explore trends in the likely changes in foot traffic (see more below).

Data

MTA turnstile data from January to March, 2019 and January to March, 2021 was used. The raw dataset consisted of 5,359,570 observations. Pre-Covid and during-Covid data was used to offer both a current and normalized picture of subway usage which, when looked at together, could inform predictions for subway usage in the year ahead. The MTA turnstile data is open source and can be accessed here:

<http://web.mta.info/developers/turnstile.html>

Algorithms

The project began with data cleaning, which included identifying and removing duplicate rows of data and identifying and reconciling anomalous data. Notably, the entries and exits data was found to be cumulative, so the differences between entries and exits and previous day/hour entries and exits was taken to compute daily and hourly foot traffic. Entries and exits data was also aggregated to give total station foot traffic. Visualizations (bar charts, line graphs, and scatter plots) were then selected to showcase the findings.

Tools

Raw data was ingested into an SQL database. SQL queries were then run with SQLAlchemy to explore the data and find duplicate records. Exploratory data analysis was then performed in Pandas. Finally, Matplotlib was used to create plots visualizing findings.

Communication

Matplotlib plots visualizing the findings are displayed in a PowerPoint, which will be presented to the Metis class and shared on GitHub.