Cecilia Bell 4.21.2021

Project Proposal

Liveboard Advertising Pricing

The Question:
How can OutFront Media optimize the pricing of their liveboard advertising spaces in MTA stations based on the foot traffic of the stations?

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. Given that liveboard advertisement spaces are available for purchase at different times, on different days, and at different stations, how can the pricing of these advertisement spaces be optimized based on the foot traffic of the stations? Assuming that higher foot traffic should correlate with a higher price, I aim to answer the following questions to offer actionable insights for OutFront Media:

1. What stations have the highest foot traffic?
2. What days and times have the highest foot traffic by station?

New York is a multilingual city and some companies seeking advertising space from OutFront may also be interested in knowing whether their content should feature languages other than English, such as Spanish or Chinese. For OutFront to offer these companies recommendations on including non-English languages in their content, I aim to answer the following questions:

3. What are the most common household languages/most commonly spoken languages of the populations of the neighbourhoods where the stations are located?
4. What neighborhoods (where the stations are located) have populations with the highest percentage of limited english proficiency?

The Data
To answer the first two questions above, I plan to use the New York MTA turnstile data from January to March, 2019 and January to March, 2021. I plan to use pre-Covid and during-Covid data 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 it can be accessed here:

- http://web.mta.info/developers/turnstile.html

An individual sample of analysis might be the cumulative foot traffic in a week for the top five busiest stations. I plan to start by focusing on finding the cumulative foot traffic in a week for each station and then focus on finding the foot traffic in an hour in a day for each station. I predict that the stations with the highest foot traffic will be the stations with the largest number

of train lines and be located in Manhattan, and that the highest traffic days and times for each station will be weekday mornings and evenings.

To answer the the third and fourth questions, I will use the US Census Bureau's 2016 data on language usage by New York City neighborhood, compiled and slightly altered by Jill Hubley (I cannot access the original), which can be accessed here:

- http://www.jillhubley.com/project/nyclanguages/#about

The Tools

I plan to ingest the raw data into an SQL database and then query from that data into Python via SQLAlchemy, cleaning and aggregating the data. I will then perform exploratory data analysis in Pandas. Finally, I will use Matplotlib or Seaborn to create plots based on my analysis. Alternatively, I might use Tableau to visualize the data.

The MVP Goal

My MVP goal is to create a bar chart depicting the five stations with the highest foot traffic. In order to do this, I will need to clean and organize my chosen MTA turnstile data first.