MTA Traffic Analysis for Electronic Advertisement Placement for E-sport Team Cheetah

By: Crystal Han

Abstract

The goal of this project was to use the New York MTA Turnstiles to analyze traffic at certain stations targeting our demographics of gamers, students, and working professionals in order to promote a gaming event hosted by the esports team Cheetah.I worked with data provided by the Metropolitan Transport Authority (MTA). Traffic is analyzed between June and November to predict optimal placements of electronic advertisements at stations during specific hours. After pinpointing stations that will have most of our target demographic going through, I provide the traffic of the top five stations with the most traffic and communicate my results using matplotlib and seaborn.

Design

This data for this project is provided by Metropolitan Transport Authority (MTA). This presents weekly data of traffic for each C/A, Unit, Line, Station at specific dates and times. Grouping the average of total traffic per station would enable the Cheetah team to place electronic advertisements at certain stations for 3 months before the event starts. The Cheetah team could allocate resources and energy to advertising at select locations and times that will get word of their event out to people who are interested and our target demographics within these 3 months.

Data

The dataset contains traffic data from June 1, 2021 to November 30, 2021. The data is categorized by C/A, Unit, SCP, Station, Linename, Division, Date, Time, Entries, and Exits. Some of these features can be grouped by station or date to observe the total

number of entries. The entries data can be used to calculate the average number of daily entries per date or station. An in-depth analysis of five stations and the first day of every week was undertaken to help visualize traffic data through plots.

Algorithms

Feature Engineering

1. Plotting average daily entries against station and date to observe the best times and stations to determine best advertisement placement.
2. Converting date and time variables into datetime type
3. Selecting subset of stations with the most average daily entries
4. Fixing outliers that do not present regular behavior of traffic
5. Combining daily entries and taking the average to plot to highlight strong traffic and illogical behavior

Models

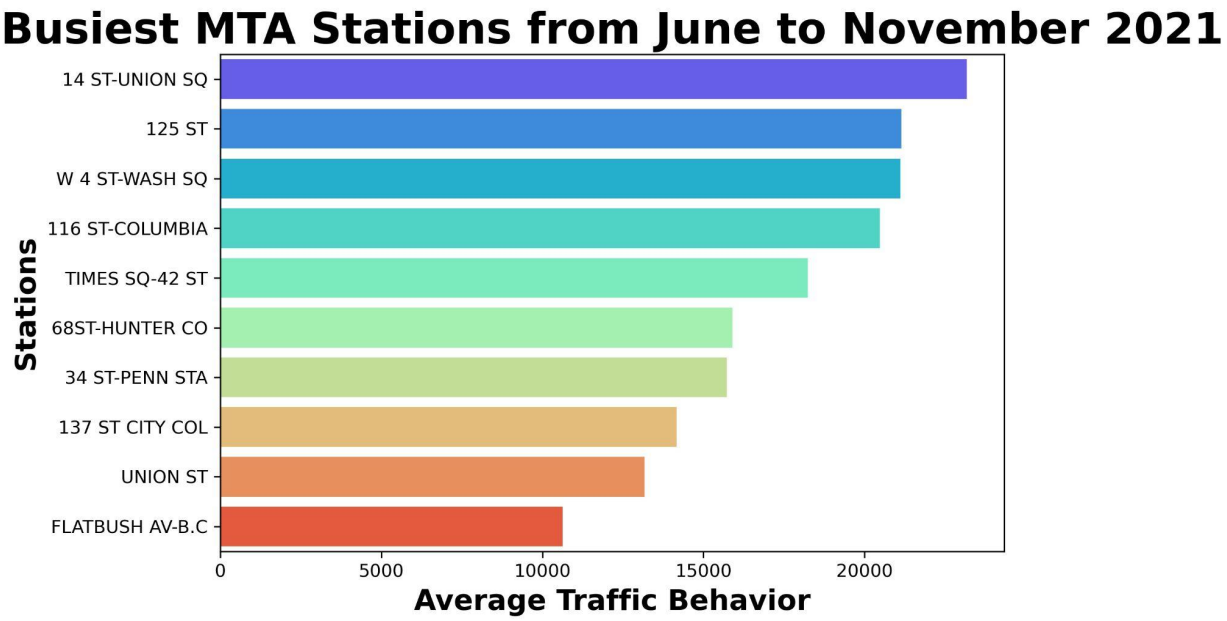
Histograms, box plots, and bar plots were used to observe traffic. Bar plots were used in the end to visualize traffic data against stations and dates to help identify the best location for advertisement placement towards our target demographics. It is important that the traffic was ordered in a descending way.

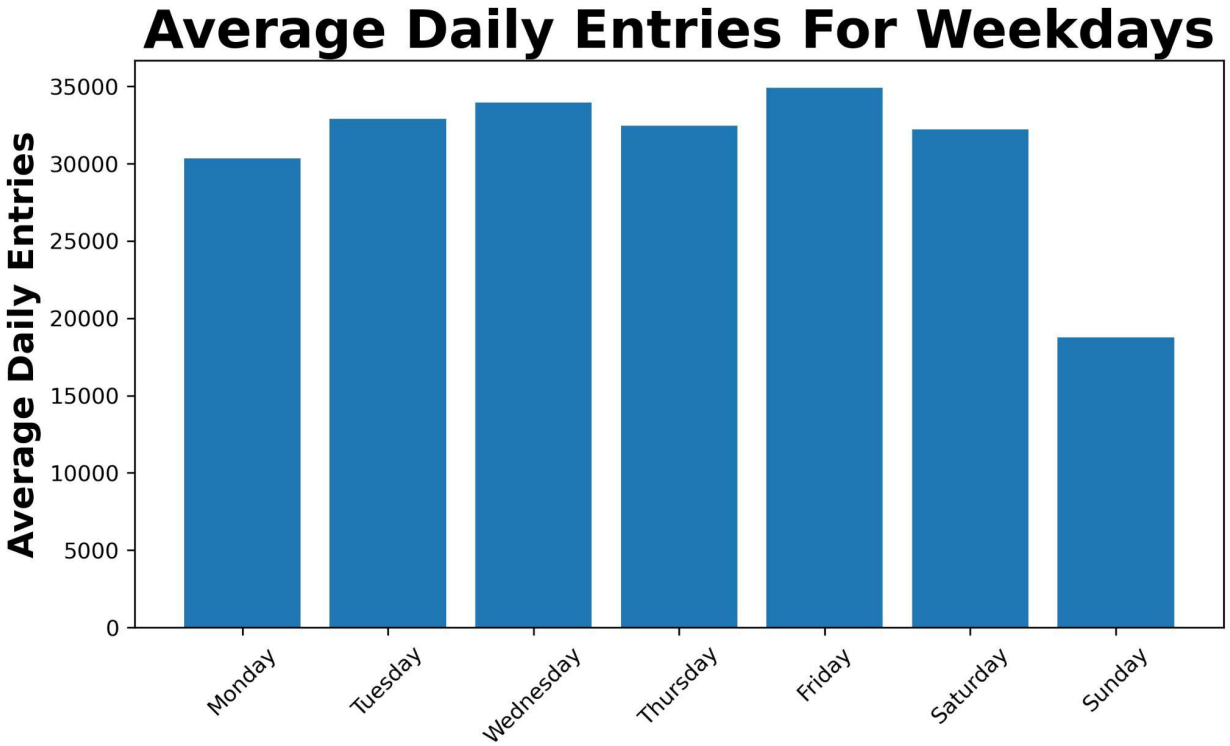
The dataset of June to November was grouped into stations and dates against average daily entries. The average daily entries can help provide visualization of traffic behavior without overwhelming the visualization of it.

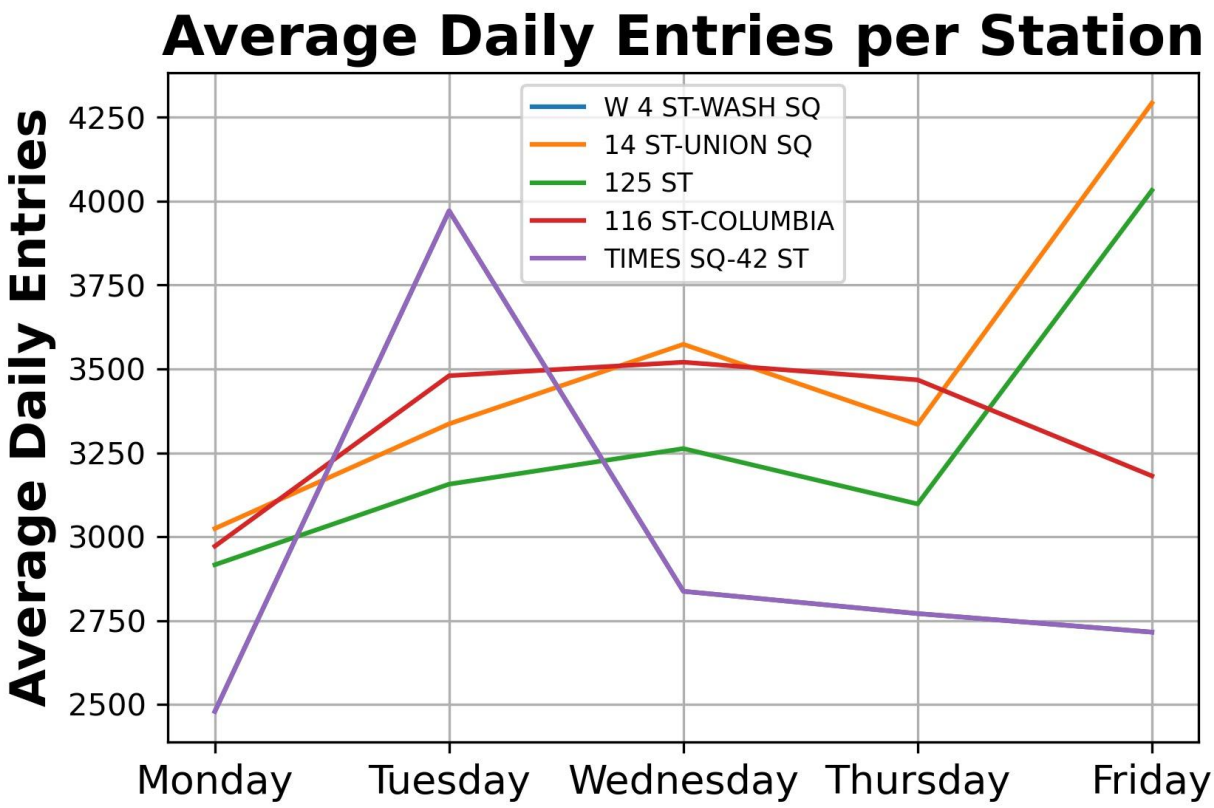
Tools

- NumPy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting and visualizations
- SQLAlchemy to write the python dataframe into a database and read from database

Communication






Average Daily Entries per Station
Average Daily Entries
4250
4000
3750
3500
3250
3000
2750
2500
W 4 ST-WASH SQ
14 ST-UNION SQ
125 ST
116 ST-COLUMBIA
TIMES SQ-42 ST
Monday
Tuesday
Wednesday
Thursday
Friday