Gulay Samatli-Pac

WomenTechWomenYes (WTWY) Member Acquisition Strategy - New York City

1. Introduction

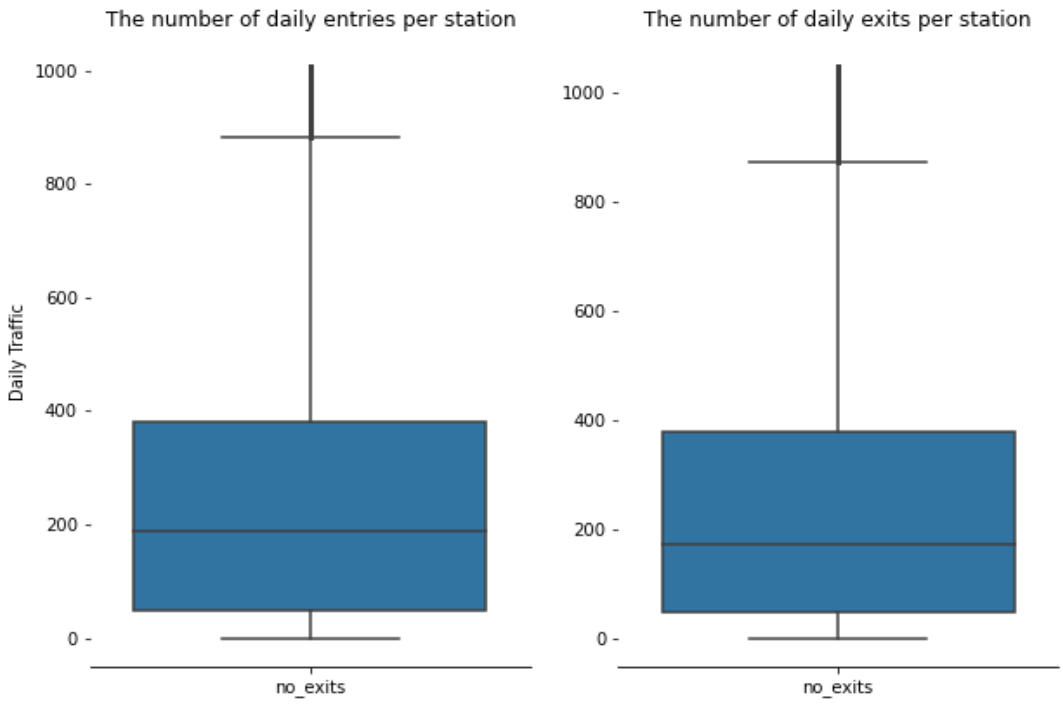
Before their annual gala, WomenTechWomenYes (WTWY) wants to reach out as many as individuals who are passionate about women in technology and simultaneously increase the organization's awareness. The organization plans to collect email addresses at the entrances of the subway stations and send invitations to the gala those who sign up. What I am asked to do is to optimise the placement of their street teams. To do this I will analyze the individual turnstile data from subway stations across New York City: MTA turnstile data. The main purpose of this project is to find the busiest stations across the city, identify the peak days of a week and times to place street teams at these stations.

2. Approach

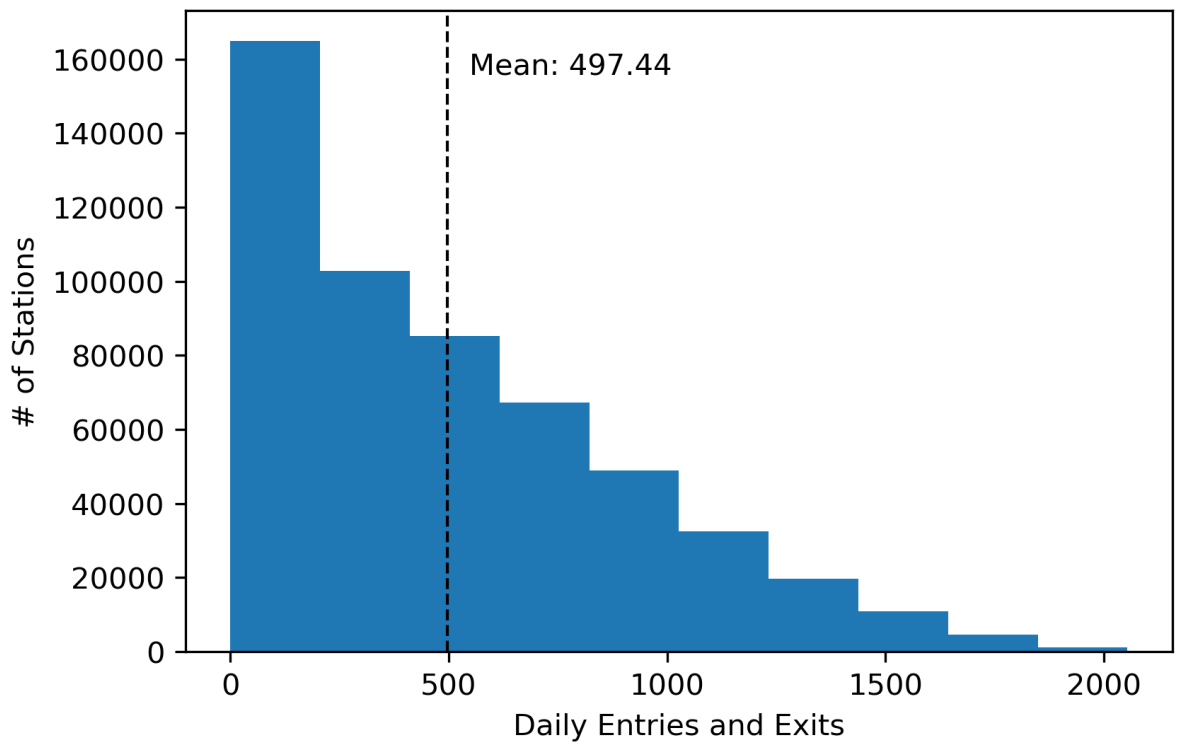
I use MTA turnstile data from February, 2021 to June, 2021. My initial assumptions while analyzing the dataset are:

- the busiest station will provide more signups
- the total of entries and exits from a turnstile is a reasonable estimate of the total foot traffic around the turnstile and so does the traffic around the corresponding station.

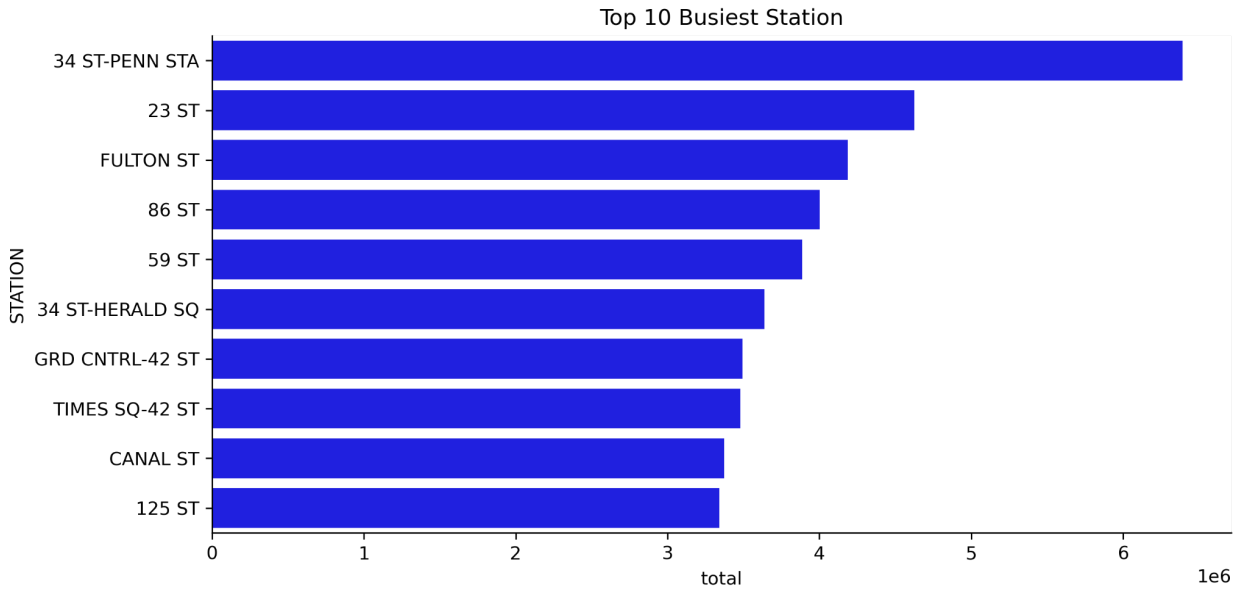
To start with, I focus a relatively good amount of my time on data cleaning - removing duplicate records, understanding outliers, or other abnormal records, fixing data columns related issues like variable types. Then, I produce the following daily entries and exits boxplots. As seen from the graph, the distribution of daily entries and exits looks similar in the sense that they are both right skewed, the majority of the data is towards the lower side.



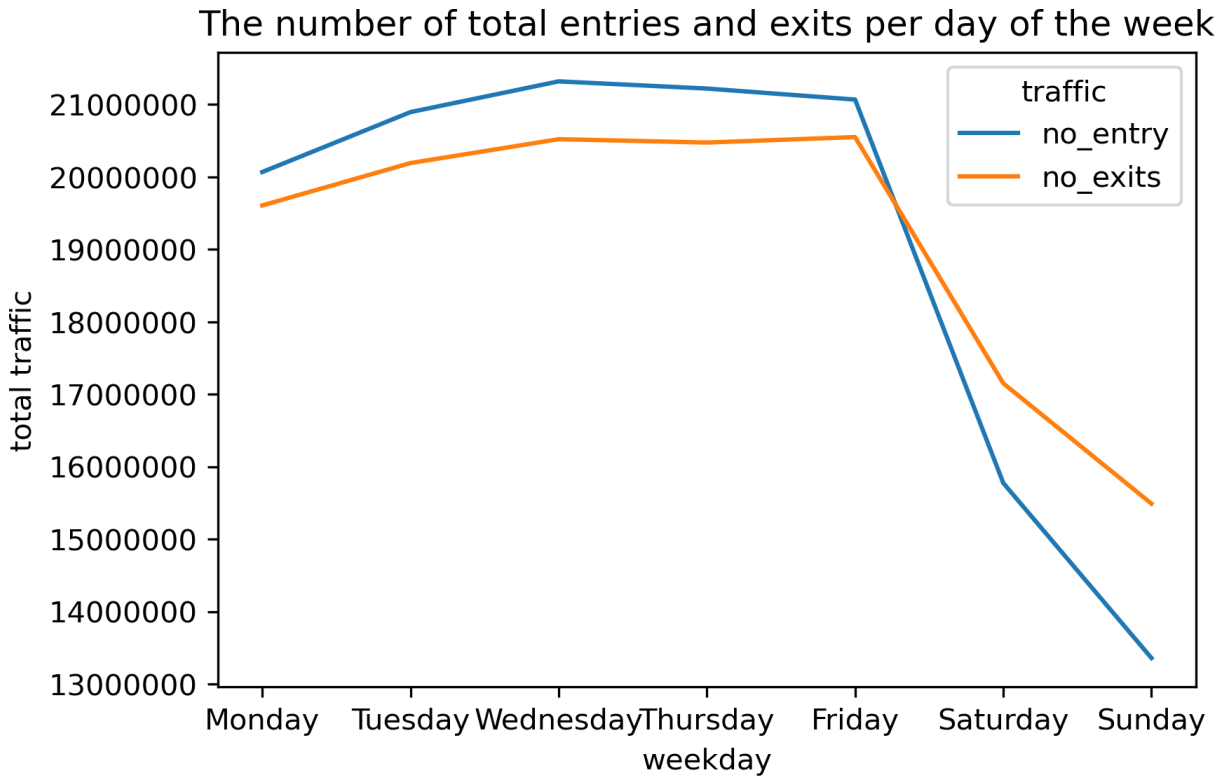
The next graph is the distribution of the total foot traffic which is the combination of daily entries and exits. The histogram of the total traffic shows the same pattern as individual entries and exits, the distribution is skewed right.



The top 10 busiest stations are shown in the graph below.



And finally, the traffic per day of the week is shown below. The foot traffic has reached its peak in the middle of the week and drops significantly toward the weekend.



3. Further Study

Next, I will further analyze the top 3 stations and assess the busiest turnstile within the station and busiest time of the day so I can recommend the optimal schedule for the street time.

Furthermore, I am planning to analyze NYC demographic data as a complement to MTA data to get insights about the sample of population who can be more willing to be part of WTWY e.g. womens age between 30 and 40, middle class. So, the gender, income and age distribution of NYC districts are metrics that I will focus on