MTA DATA ANALYSIS

Yonga Lhamo

Abstract
MTA map is complicated even to just look at and diving into the data within, was lets just say a little bit more complicated. WTWY needs help in analyzing the MTA dataset to optimize the stationing of their street team members in order to gain maximum number of signatures and attendees for their summer Gala.The recommendations will be done by filtering stations and turnstiles based on traffic and neighborhood profiling.

Design: The objective of the process was to find locations based on traffic and context. So the flow of the process was to first grab all of the turnstiles, grouped and ordered by the day . And then get the grouped traffic for stations. Getting stations with highest average traffic give WTWY more exposure. And to further filter that since WTWY is a small organization and may not have that many resources, I took the top 30 stations by traffic and did a neighborhood profile filter selecting just five based on it being a tech hub or on financial grounds . so neighborhoods within the tech industry would have the probability of higher interest in the cause and richer neighborhoods could possibly have more money to donate.

Data:

- MTA dataset, March-May 2019 because these are pre-summer months and gala is beginning of summer. 2019 because the years after have been different from norm due to covid and hopefully by next summer things would have normalised.
- Zip code dataset based on subway station to locate stations.
- Neighborhood dataset based on zip code to research profiles on neighborhood.
- Turnstile cluster addresses to get more specific coordinates than a station name

Tools:Pandas, SQL,Tableau

Communication: After following the design process, we have top turnstile geo locations that WTWY can focus their resources on. Amongst the top stations they also have specific cross streets with busiest traffic (turnstile cluster locations). The findings can be further improved if given more time to find better datasets to correlate with the primary data set .