Zhi Wen Huang

Metis: Data Science Bootcamp

EDA Project: WTWY Summer Gala

Abstract:

In this project, we want to get more people to join the Women Tech Women Yes organization summer gala. To do this, I use the MTA turnstile data to get insight of population flow of each subway stations. Using this insight, we can set street teams to optimize the possibility to reach people to register for the summer event.

Design:

I created multiple graphs to visual the population flow of each station in NYC. I’ve grab insights by comparing stations population, stations population flow throughout the week, and average population during the morning to night time interval.

Data:

I used the MTA turnstile data during from January to May 2021. The reason I use this time frame is because the event is starting from early summer. We want to have people ready and register before the event started. Some features of the data like, Control Area, SCP, Unit, Station, Entries, Date, and Time are important to retrieve insights of the population flow.

Tools and Algorithm:

Using SQLite to download the MTA turnstile data. Then I use SQLAlchemy to call queries in Python. I’ve aggregate data by finding mean of entries and sum of total population for each station. Then I use seaborn to create graphs on data I aggregated. Some graphs I created are barplots, lineplot, and heatmap. I include the matplotlib in python to configure my plots to make the visuals more readable and cleaner to understand.

Communications:

I’ve created many graphs and each graphs gives different insights of the data. The bargraph are mainly to compare each stations to see who has more populations. The linegraph is to see the trends and changes throughout the time interval of each day of the week. This is helpful when we want to plan during what time to setup the street teams. To find compare traffics with other stations all together, I created a heatmap to show the population of all stations with their populations during each day of the week. Using this insights we can decide rather we should setup street teams for specific stations and during certain time of the day.