

WomenTechWomenYes (WTWY) Member Acquisition Strategy - New York City

1. Abstract

Before their annual gala, WomenTechWomenYes (WTWY) wants to reach out to as many individuals as possible 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 and Census Tract Data which includes total population, racial/ethnic demographic, income, employment, commuting characteristics. The following assumptions are used to drive conclusions

- 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.
- females will be more interested in WTWY
- people from higher income per capita districts will be more likely to donate.

2. Data

MTA turnstiles data from February, 2021 to June, 2021 is used to answer the following questions:

- the busiest subway stations across the city with respect to daily cumulative entries
- the busiest days of a week and time of the day .

The Census Data from more than 2000 census tracts across the five boroughs is used to answer the following questions:

- the borough with the highest average income per capita

- the borough with the highest women percentage
- the borough with the highest employed women percentage
i.e. $\text{WomenPop} * \text{EmployedPop} / \text{TotalPop}$.

3. Approach

To start with, I focus on data cleaning - removing duplicate records, understanding outliers and other inconsistencies, fixing data columns related issues like variable types.

Next, Exploratory Data Analysis is conducted on both datasets to identify the trends and gain insight on the variables of interest.

Finally, conclusions and recommendations are made from results gained from EDA.

4. Tools

- DB Browser for querying the data
- SQLAlchemy for initial data analysis
- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting

5. Communication

The results are presented in a presentation. All data manipulation and analysis can be found in Jupyter notebooks.