Women Tech Women Yes - Maximizing Outreach

Mitchell Seiter

Abstract

The leaders of Women Tech Women Yes (WTWY) reached out to see if Metis could help predict how they could be most successful in reaching out to as many people as possible who might be supportive of their cause. Using NYC MTA data that records the use of every subway turnstile, I was able to find the most active subway stations around the college campuses with the most enrollment. After cleaning and refining this data I was able to provide several graphically pleasing charts to support WTWY

Design

To tackle this project, I collaborated with the WTWY team to figure out who their ideal demographics were. After finding out they were mainly hoping to target a female, younger, and/or more socially active crowd, I began to look for various university and college campuses around NYC, and to isolate the subway stations that were in the nearby vicinities, and figure out which ones were the most heavily trafficked.

In addition to that, I have narrowed down the busiest days of the week and the time slots when the subway stations are most frequently used as well.

Data

The MTA dataset consisted of a little over three months worth of subway data, which came out to about 2.5 million rows of data overall. A lot of the data set was in disarray and needed some work to be done on it before it could be usable.

Algorithms

Algorithms used to clean up this data set included finding a way to first make the data usable. This involved going through the data frame and first getting rid of all NaN values.

We were then able to find the daily use of each turnstile and continue to get rid of any erroneous data. A lot of the values were well in the range of being an outlier. However, because the range of the data set was so large I only got rid of the most extreme outliers, and used an IQR(Interquartile Range) * 100 instead of the standard 1.5.

From there it was just a matter of grouping the data in various subtables to get the visualizations and information we were looking for.

Tools

- Pandas and SQL for data manipulation
- Matplotlib and Seaborn for plotting
- Input functions for future use by WTWY

Communication

I have assembled a slide deck that will be presented to WTWY and sent over to them for their future use as well as provide them with a tool that will allow them to compare subway stops on their own.