WTWY street teams

Problem Statement

Women Tech Women Yes (WTWY) is hosting their annual gala in the beginning of the summer, and they would like to maximize attendance of their event and increase awareness of their organization. A designated group of street teams will solicit e-mail addresses from people at various locations, and those people will be contacted with free tickets to the event.

By using publicly available MTA subway data, can we devise a plan to optimize WTWY's deployment of their street teams in order to obtain the most e-mail addresses and maximize attendance of their event while increasing awareness?

Our Approach and Results

We first quantified traffic for each station by computing the total number of turns (adding together entries and exits) over the course of one week. Based on this information, we identified stations that had the highest traffic:

1)   Penn Station
2)   Grand Street
3)   Herald Square
4)   Union Square
5)   86th Street

We found that that was a large variation in the number of turnstiles per station, 100 (Penn Station) vs 3 (Neptune Ave), so we wanted to target bottlenecks and identify the "densest" stations. We quantified density by calculating the total number of turnstiles per station and dividing the total number of turns by this number. The following are the "densest" stations:

1)   Union Square
2)   York Street
3)   1st Ave
4)   23rd St - 6th Ave
5)   Grand - 30th Ave

In addition to measuring density, we identified those stations which were more likely to contain commuters (who are more likely to attend the gala than tourists). We accomplished this by subtracting the traffic for off peak hours from peak hours by the traffic for off-peak hours and comparing the different stations based on this new metric. As before, we calculated density by dividing by total number of turnstiles per station. The top 5 stations based on this criteria were:

1) Main St
2) 59th St
3) 86th St
4) Penn Station
5) Herald Square

Additional Approaches

Next, we will develop an approach for finding the highest traffic time periods (typically 4 hour windows) in the highest traffic "commuter" stations. In addition, we will explore the possibility of comparing traffic between the different entrances into each station.

Lastly, to improve targeting of people who are likely to attend our event, we will use demographic data online in conjunction with insights from MTA data. For example, we can use public data to target areas with higher concentrations of young female professionals.

Conclusion

Based on a quick manipulation of the MTA data and resulting analysis, we came up with a straightforward approach to improve yield of your street teams' efforts. We are excited to continue working with you to discuss different approaches and to add additional layers of complexity to our analysis.

By combining the results of our first two studies we were able to calculate a preliminary determination for which stations you should target:

1. Penn Station
2. 86th Street
3. 59th Street
4. Hudson - 80th Street
5. Roosevelt Ave

Thank you,

Metis Data Science Team
Aaron Mangum
Daniel Shin
Nick Smit
Tohei Yokogawa