Company: YourChauffeur (private cars application)
Purpose: Increase overall revenue through usership

Email from client:

Dear data scientists,

YourChauffeur is a ride-sharing service looking to expand our market share in New York City. Our smartphone application connects passengers with drivers of vehicles for hire. We would like to increase the number of users who download our application as well as the frequency with current users utilize our service.

Toward that end, we would like you to use freely available MTA subway data to determine optimal physical and virtual advertising to geo-located smartphone users. Furthermore, we would like to increase usership among current/previous YourChauffeur users through smart push notifications about special deals during bad weather and peak times and locations.

We look forward to hearing your proposal.
Miss D. Daisy
YourChauffeur Chief Executive

Advertising and Discounts

Problem Statement:

YourChauffeur needs to optimize participation by: (1) increasing awareness of YourChauffeur through efficient physical advertising in the most crowded subway stations; (2) increasing awareness of YourChauffeur through efficient virtual advertising by targeting geo-located smartphone users based on time and location; and (3) strengthening usership by offering discounts based on weather and peak travel times/locations to users. Ideally, implementation of these steps will convert to a higher rate of usership and revenue for YourChauffeur.

MTA data obtained from the NYC Data Portal provides rich information about the travel patterns of New Yorkers -- using this information, can we create an advertising and discounting plan that will optimize YourChauffeur's resources toward achieving their goals?

Preliminary Results:

Here are some preliminary results we found based on mta turnstile data only:
We focused our analysis on subway entries and not on exits.

Top five stations in terms of average daily turnstile entrances:

1. 34th St - Penn Station
2. 42nd St - Grand Central Station
3. 34th St - Herald Square
4. 14th St - Union Sq
5. 42nd St - Times Sq

We also tried to estimate which were the most crowded stations by normalizing the number of entries by the number of booth per station.

Top five Most Crowded stations:

1. Path WTC
2. Bedford Avenue
3. Main Street
4. 68ST- Hunter College
5. Roosevelt Avenue

The more crowded the station is, the more people will struggle to get into a train. These people will be particularly receptive to the advertisements YourChauffeur sends them and may consider utilizing YourChauffeur.

We recommend YourChauffeur to focus on targeting users near Midtown stations. YourChauffeur could geo-locate its users and send them push notifications or coupons as they come close to one these subway stations.

What Else We Could Do/Adding more Data Sources:

The most promising area of additional research would be the marrying of YourChauffeur's existing proprietary ad click and utilization information with our subway data. Our subway turnstile research has given us a good starting point. However, yours is an existing business with considerable information about where your existing users work and reside, where they utilize your service, what ad information they respond to and which rides tend to be most profitable. By combining the two, we believe we can help you expand both your customer base and the utilization of your existing customer base.

To further this aim, we plan to integrate real-time MTA scheduling information to infer the most crowded subway stations and lines which are not necessarily the stations with the most throughput. Like the turnstile data, MTA's realtime feed is freely available on the official website located here: http://datamine.mta.info/feed-documentation

We plan to integrate real-time NYC traffic data to compare the utility of surface transportation vs sub-surface transportation at any given time. Users will be more likely to utilize your service if it will save them significant time. There are several freely available traffic APIs, including MapQuest and Yahoo.

We could also integrate real-time weather data as well. Anecdotally, we believe users will be more likely to use a car service like YourChauffeur when the weather is bad. Weather Underground, for example, has a freely available weather API.
http://www.wunderground.com/weather/api/

All of these data sources will help you micro-target your ad coverage and direct it at the most receptive audiences.

Conclusion:

We've learned a lot by analyzing the free MTA data already. We believe we could leverage even more data to make a big impact on YourChauffeur's client base. We have identified three methods to increase YourChauffeur usage in NYC. First, placing advertisements in high traffic subway stations in an effort to target the entire population. Second, targeting likely users who already have smartphones and are located in an area when and where they are likely to be traveling by showing them targeted mobile ads. Finally, we propose a method for sending coupons to subscribed users based on their location and likelihood of needing to travel. We look forward to hearing your opinions about our ideas and improving upon them. Let's set up a meeting to further discuss.