MTA Authority (New York, New York)

Introduction:

We propose developing an application that gives commuters an estimate of how crowded or congested a given subway line will be at a given point in time at a given station. This congestion on the subway system can be impactful for a number of reasons. Riders become late, unhappy, and dissatisfied with their ride using the Metro Transit lines. This dissatisfaction may also increase the use of taxis and other personalized rider programs. JTAG Analytics hopes to help the Metro Transit Authority inform its users about how to avoid congestion and so optimize their experience and ease their use of the subway system.

From a brief analysis of some of the data you've made available, including MTA turnstile data, we've determined that congestion is predictable. Most obvious, congestion is heightened during rush hour. Trains also tend to get excessively full at off-peak hours when too few trains are running. Possible solutions to reducing this congestion include (1) change the infrastructure to accommodate the flow through the system (but these are large investments, which are costly and take many years for their benefits to take effect), and (2) changing how people flow through the system to alleviate stress on and in the system. Solution 2 is where focused our analysis.

Congestion can be alleviated by riders slightly adjusting their schedule or travel trajectory, but the information necessary to making this adjustment is difficult to individually assess and forecast. We propose to use the data made available by the MTA to report on the level of congestion present in the subway system over time. Using the ridership data along with the frequency of trains at a given point and the capacity of a particular subway car, we can develop metrics of capacity, or how full subway cars will be when they arrive at a subway station. Riders will be able to access this information up to a week in advance in order to better plan their travel. For example, a commuter might access this information the night before a morning commute and in turn choose to depart the next morning earlier than originally planned.

Methodology:

Current methods of determining congestion on other venues often occur by actively sampling cellular data or other real-time information and then reporting the data as it occurs. This is useful for relatively irregular behavior, such as traffic flow on roadways.

The behavior of train riders is starkly different from car traffic since it is governed by the regularity of the train schedule and permanence of train station locations. Because of this regularity, we can construct a forecasting model to predict the congestion that appears at different points of the day at each station. Using a year's worth of data, we can make useful predictions on train use.

With a forecasting model, based on observed seasonality and other variables, we will predict levels of congestion at different times of day at each subway stop. We will compute congestion at each station along a given subway line at a given time point. This will be accomplished using the capacity of each particular train, the frequency of the trains running at that point in time, and the number of people entering the subway station at that point in time.

MTA Data Requests:

In order to enhance the accuracy of our model, we request that MTA fine grain their data output as a part of our collaboration. Current ridership reporting comes from turnstiles which report information every four hours. The train schedule, however, moves at a much faster rate than this information is reported. By logging this data at a more frequent duration (at least every minute, optimistically) we will be able to build out a forecasting model that is more effective for prediction.

In addition, we request real-time data to better serve the MTA riders. Real-time data reported directly from the turnstiles would put less burden on a forecasting model and create the possibility of seeing congestion immediately.

MILESTONES (6 mo - 1 year)

- Visualization model of the expected congestion rate at a given time point for up to next 3 months for one train (N – line).
- Prototype to display forecast of daily expected congestion for one train (N – line).
- Include other lines in MTA. Add functionality to display alternative route options for departures/arrivals (Where from? / Where to? / Arrive by? / Date?). Also display warnings when a particular line and station are near capacity.
- Factor in weather (from the National Climatic Data Center and week's weather forecast) and special circumstances (i.e., special events, construction, and other circumstances anything that causes an Uber surge!).
- We encourage the MTA to install a device (Eurotech's DynaPCN 10-20 Automatic Passenger Counter) in the subway system that counts people entering and exiting each subway car. This tool will allow the information communicated to the commuters to be more precise and reliable.