

NYC Yellow Cab Analysis

2/6/2019

Team Members

Monica Chaturvedi
Simeon Markind
Julia Marshall
Alexandra McKeon
George Posner
Shiv Ganesh Sankara Bagam
Trace Sauter

Dataset

The NYC Yellow Cab dataset is available at <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>. The data is available as a series of CSV files, each corresponding to one month of rides. Together, these files contain records of every yellow cab fare from January 2009 through June 2015. Each record in the dataset corresponds to one trip and includes the following information:

- Date and time for each of pickup and dropoff
- Latitude and longitude for each of pickup and dropoff
- Trip distance
- Fare amount, tip amount, toll amount, and surcharge amounts

Since each CSV is 1.4 GB in size, we intend to take a random sample in order to make the analysis feasible. Separately, we are also exploring incorporating other datasets (e.g., residential rents from Zillow Research and crime data from the NYPD).

Project Idea

We would like to explore the New York City taxicab data. Specifically, we're interested in understanding the relationship between taxicab trips and the following:

- City-mandated cab fares: In September 2012, the NYC Taxi and Limousine Commission announced a 17% hike in fares. We would like to examine how this impacted the volume, duration, and lengths of trips and if any identifiable subset of trips was differentially impacted.
- Commuting patterns: We would like to determine if it is possible to identify blocks or neighborhoods with increased residential or commercial activity.
- Inventory planning: We can hence propose to build a forecasting model that can clearly estimate how the demand for taxis would be for the next month/year across multiple localities and come up with some kind of inventory planning insights.
- Major events in New York. Examples include the New York Marathon, the opening of the 2nd Avenue Subway, and criminal or terrorist activity.