Arcenas, Carlos
De Guzman, Carlo
Seechung, Vito
Tiongson, Zen
CS129.1–A
Revised Big Data Problem Proposal

1. Big data problem:

   Given the data of taxi and for-hire vehicle pick ups in New York City, how do we determine the top boroughs in NYC in which passengers are picked up?
   (Results will be visualized on a map diagram)

   Solving this problem can be helpful for drivers to know where he or she is most likely to get passengers, at different times of day. Solving this problem would also be helpful for passengers, as they would know which companies are more present in their area throughout the day, increasing the likelihood that they will find transport.

   Our solution would first make use of the “date/time” key in the dataset to group the pickups into hours of the day, to reflect the changing patterns of traffic as dictated by the rhythms of New York life. From there, we will use the “lat” and “long” keys to form the actual coordinates of the pickups, and then group them according to what borough of New York they actually took place in.

2. Description of data:

   According to the GitHub repository for the Uber response to the New York Taxi and Limousine Commission, as released through a Freedom of Information Law request:

   “The Uber TLC FOIL Response contains data on over 4.5 million Uber pickups in New York City from April to September 2014, and 14.3 million more Uber pickups from January to June 2015. Trip-level data on 10 other for-hire vehicle (FHV) companies, as well as aggregated data for 329 FHV companies, is also included. All the files are as they were received on August 3, Sept. 15 and Sept. 22, 2015.”

   The data is currently available on the Github repo from fivethirtyeight (https://github.com/fivethirtyeight/uber-tlc-foil-response), all provided in CSV form. Data will be obtained by simply downloading the provided files.

   Alongside this dataset, we will be making use of the borough boundaries geographic dataset created by the government of the City of New York (https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm/data) to accurately locate the pick-up points to their respective boroughs.