CS 504- Final Report

Uber Pick Up in NYC

Motivation:

New York City's rush-hour traffic is one of the worst in the United States. Finding a taxi from 7 to 10 am or 5 to 8 pm in Manhattan is incredibly difficult so most commuters rely on the subway or Uber. Uber, however, is difficult to use because most of the streets are one-way so finding your driver requires walking a few blocks. Along with the difficulty of finding your driver, Uber in NYC has increased traffic because ubers drive without a destination until they are called , worsening congestion throughout the city's streets. This is especially true during rush-hour, the time of the day when most ubers are working.

To solve this problem, we propose a bus-stop-like system for Uber. Our idea consists of express pick up locations for Uber where ubers will wait for passengers in these designated locations. In this way, riders save money, time, and difficulties finding their ubers, and ubers will be off the streets while their idle.

Data Set: Uber Pickups in New York City
https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city

Constraints:
Large city would need a large scale of K in order to present realistic pick-up points, hence K-means algorithm will be slow and inefficient

The dataset we used has a time range of about a month and makes assumptions about popular pick up points in general. For determine permanent pick up points, the dataset used would have to be much larger in order to get accurate information in terms of where are the most accessible pick up points.

Method:

-First, our application filters the dataset to get the coordinates of the pick up location and the timestamps of the typical Uber rides during rush hour in NYC.

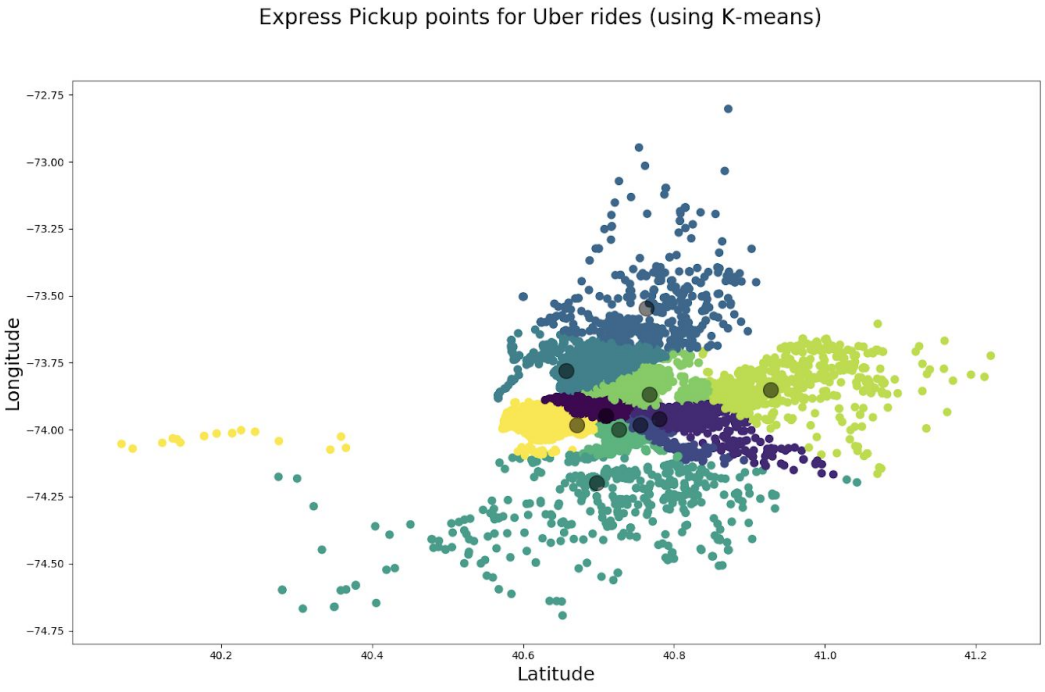
-It aggregates all of the information separated by dates into one data set with all the rides

-Uses K-means to find the K number (defined by the users input) of pick up points that optimize walking time to pick up point and riding time to destination

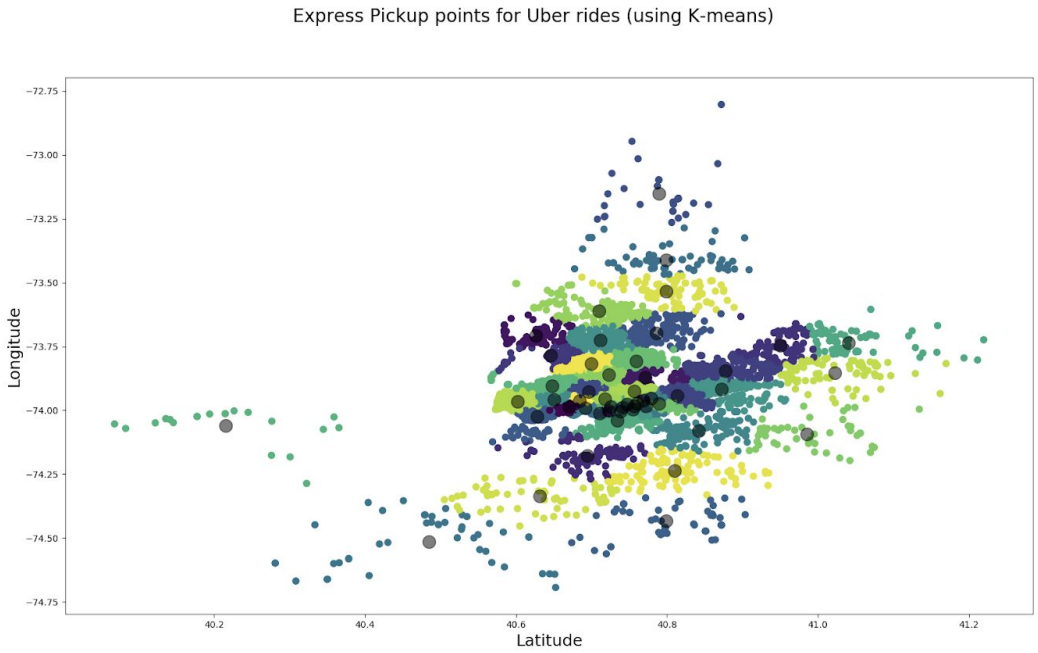
-The set of coordinates found by our algorithm are the distributed means of all the pick up points in the data set. These means represent the appropriate pick up points for all passengers in an area based on the pickup points that are in demand.

-This determines the exact location commuters should walk to, to avoid long pick up waits, enjoy an efficient commute to work and pay cheaper fares while riding with others on Uber.

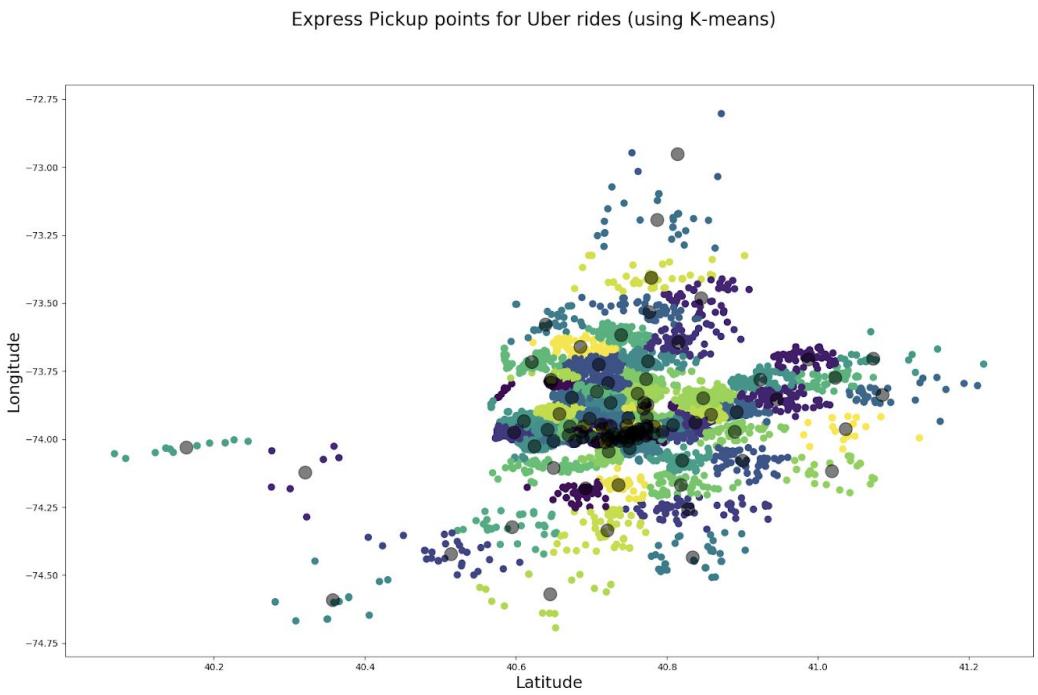
Here are the graphs the K-means algorithm generates:



K = 10


Express Pickup points for Uber rides (using K-means)
Longitude
Latitude
−72.75
−73.00
−73.25
−73.50
−73.75
−74.00
−74.25
−74.50
−74.75
40.2
40.4
40.6
40.8
41.0
41.2

K = 50



K = 100

Statistical Analysis:

-Our statistical analysis of the average distance from original pickup point to the optimal pick up point generated by K-means, helps us score the new pick up point

-The bigger K is the better the score will be, that is the smaller the distance the commuter will have to walk

-If K increases, the score will decrease, meaning a smaller distance from the desired pick up point to the mean pick up point

Web Application:

Our web application allows users to manipulate K, that is the number of pick up stations they would like to generate from the data. The web application will then display a graph of how New York City's Uber system will look like and where, exactly, the optimal pick up points will be.