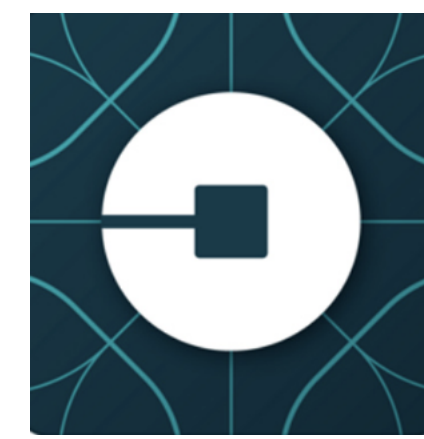


Uber Pick Up in NYC

Omar Sagga | Miranda Rivera



Motivation

Find a more efficient way of carpooling with Uber during rush-hour

Rush-hour occurs from 7am-10am and 5pm-8pm

Carpooling with others is cheap, but inefficient | the pick up points provided by Uber are not optimal

Attempt to relieve commuters from walking long distances to their ubers or experiencing detours that delay the ride

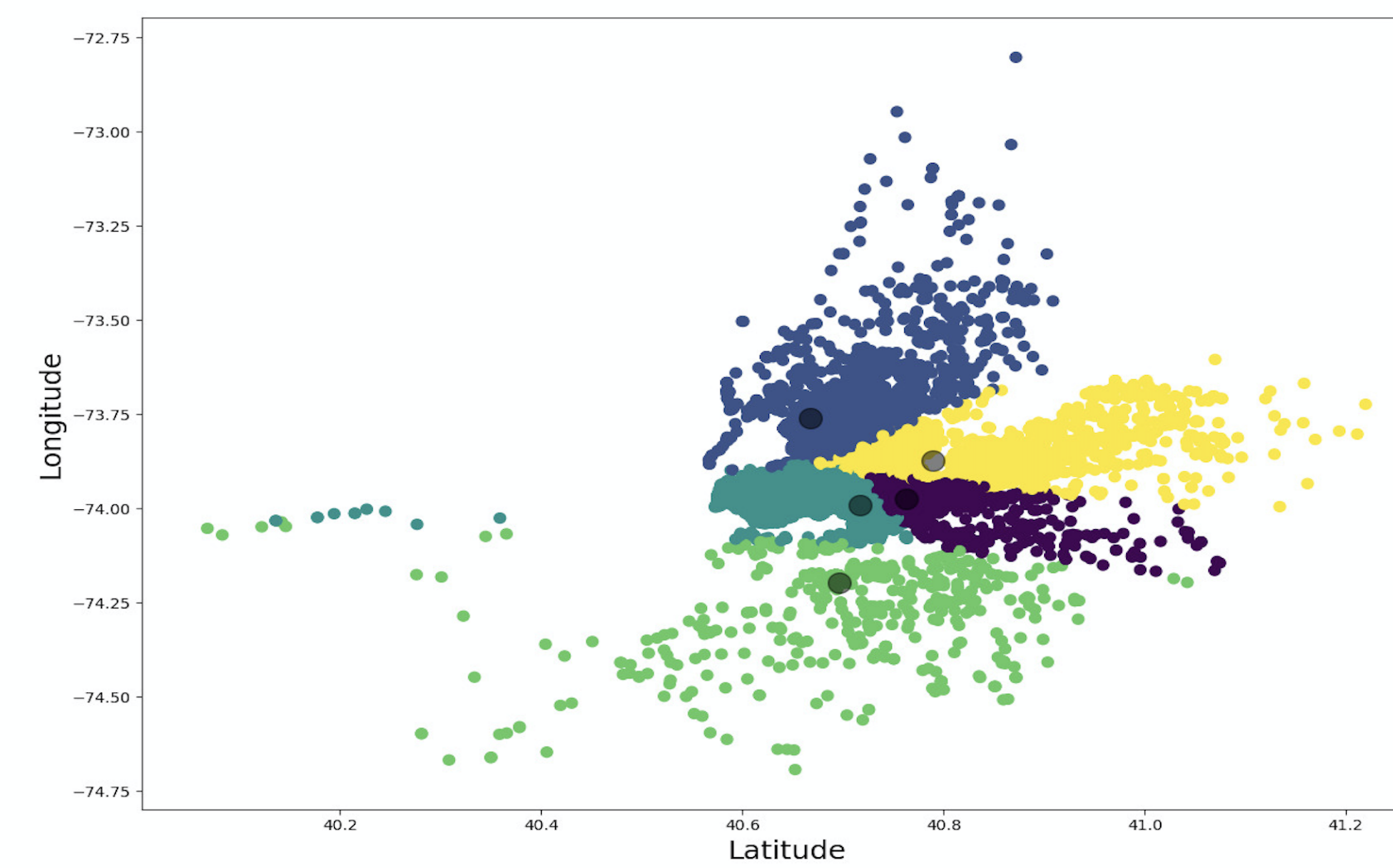
Data Set

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

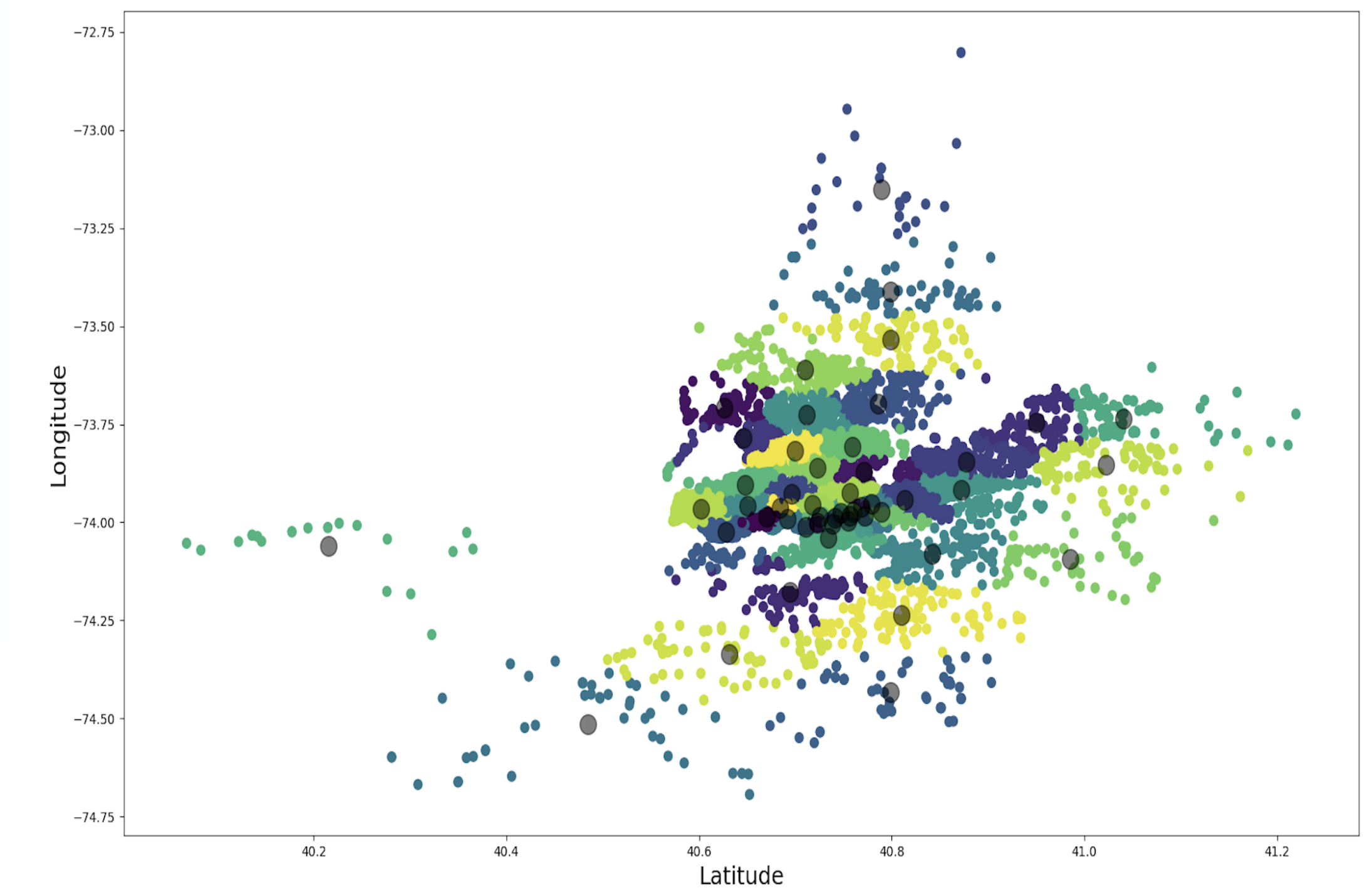
Constraints

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

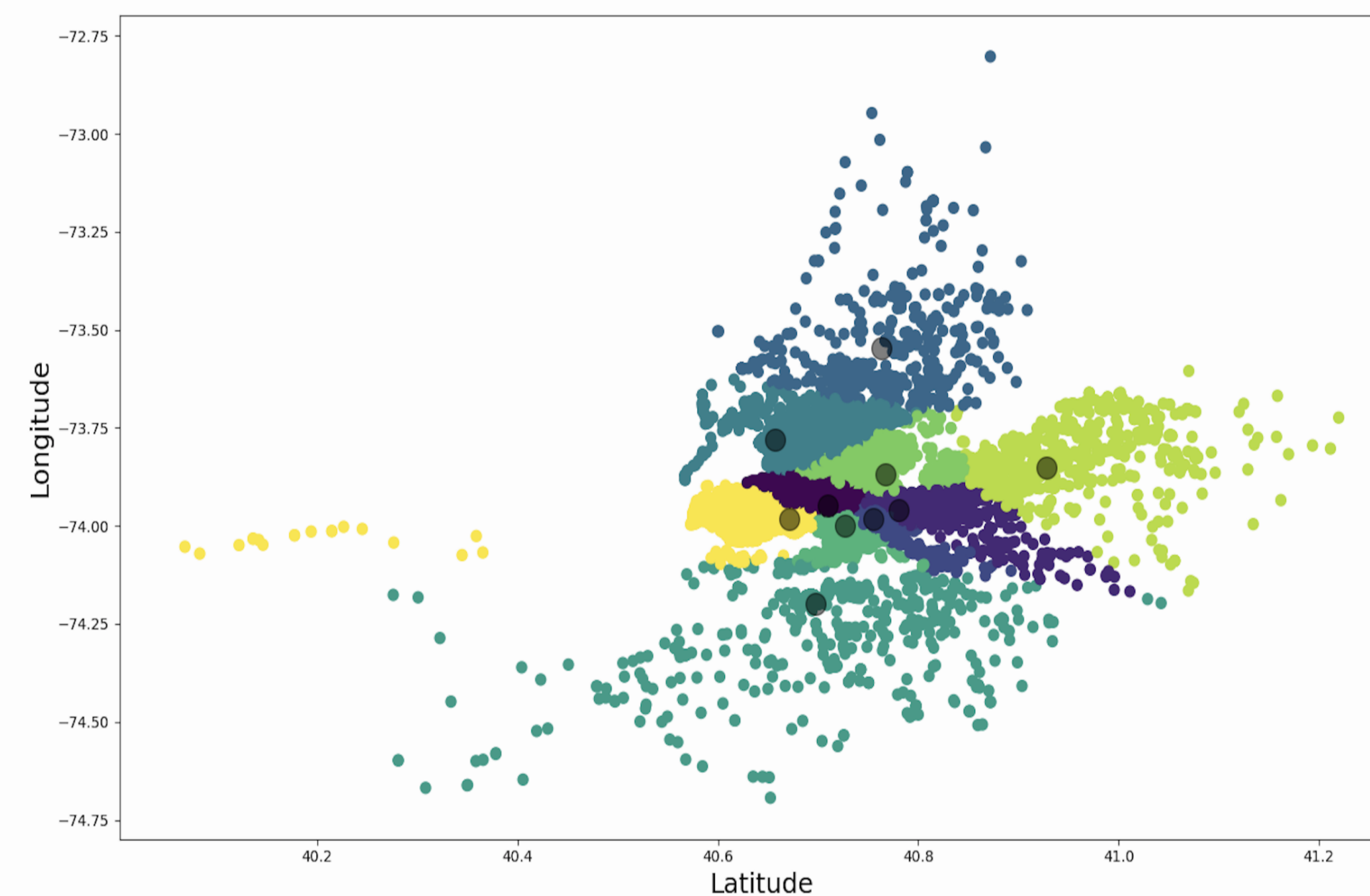
The dataset we used has a time range of about a month and makes assumptions about popular pick up points in general



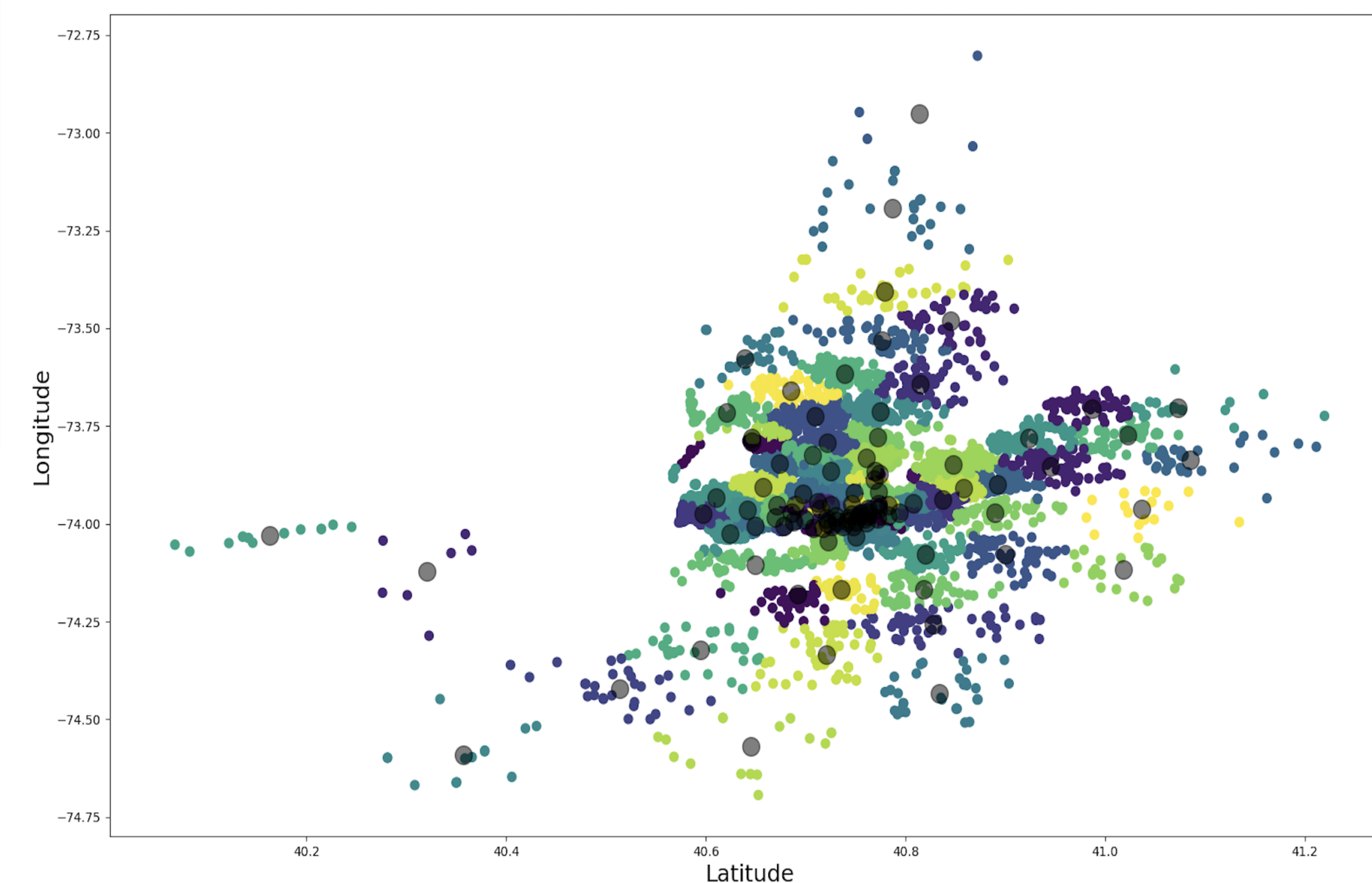
K-means algorithm with K = 5



K-means algorithm with K = 50



K-means algorithm with K=10



K-means algorithm with K = 100

Method

First, our application filters the dataset to get the coordinates of the pick up location and the time stamps 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

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

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