Adam Dulloo 995318

Distribution of trips in New York Yellow Taxi Trips

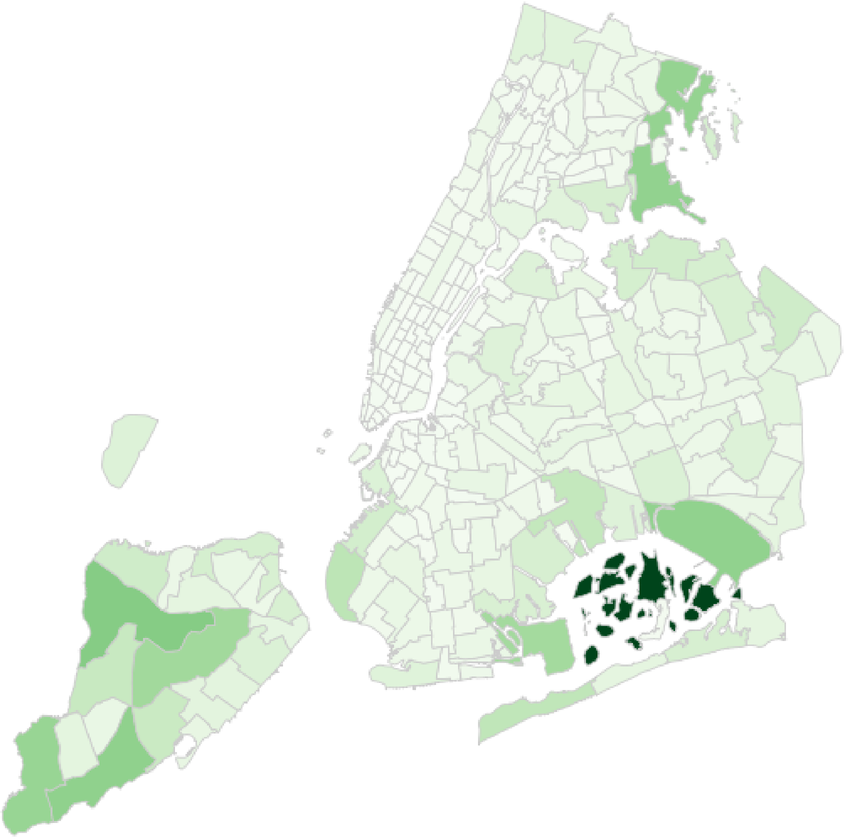
Introduction

This report aims to discover factors that correlate the distribution of trips (ie pickups and dropoffs) with time. The New York Taxi and Limousine Trip Record Dataset (TLC) was used to produce geospatial visualisations of the aforementioned attributes.

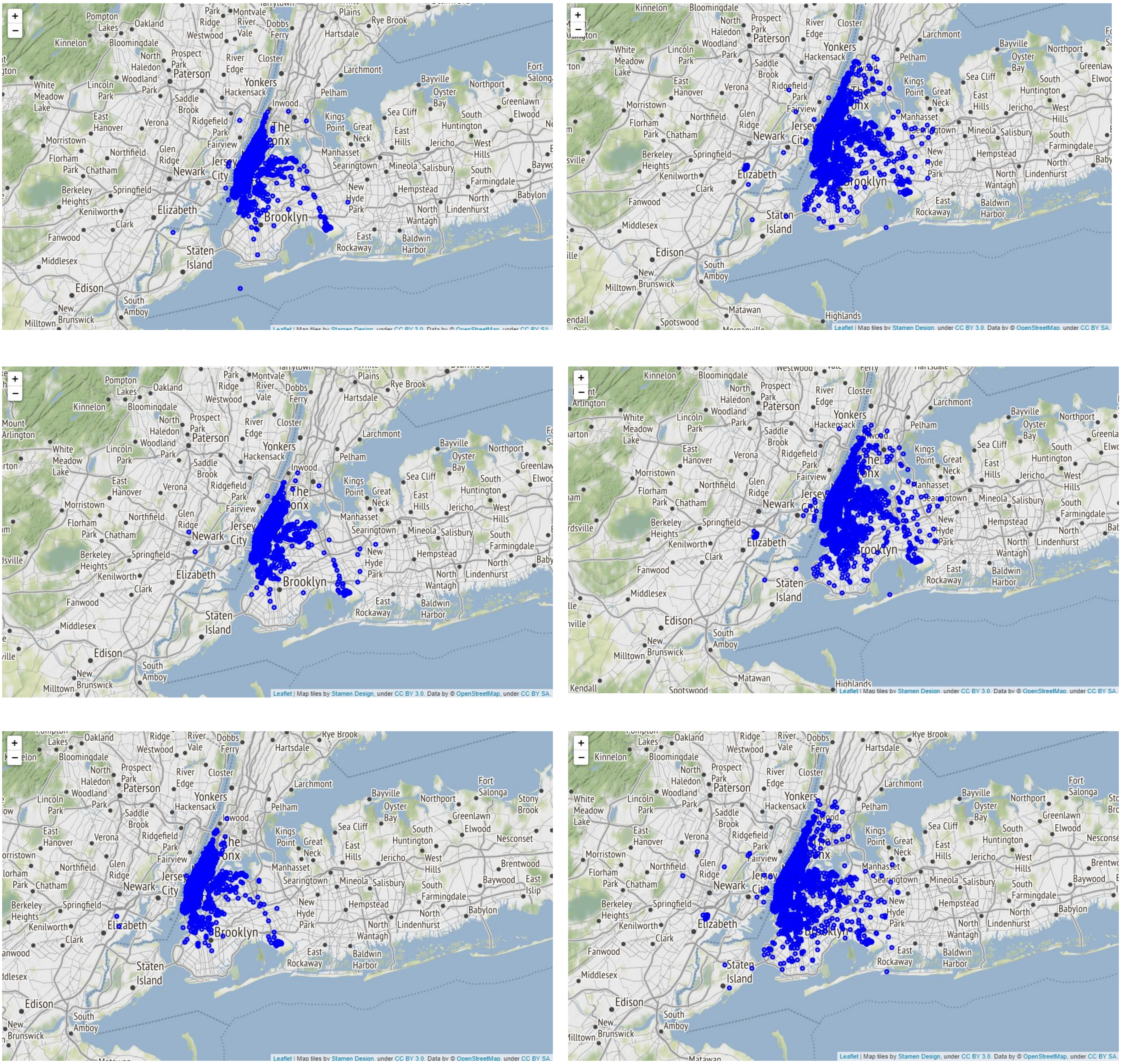
Pre-processing and cleaning

Since there was a change in the TLC data in 2016 due to privacy concerns, pickup and dropoff longitudes and latitudes were omitted in data post 2016. Because of this, Yellow Taxi Trip Records of the three months of October, November and December of 2015 was chosen as it is more geographically precise than the TLC data post 2016.
The data cleansing steps that were undertaken were removing any records that had pickups or dropoffs that were not in New York City, rows with null values in the longitudes and latitudes along with any in the time column were dropped. As some mapping functions are incredibly computationally taxing, only the scatter plots used a random sample of 10,000 records. Outliers were calculated by making sure all records had points within New York City's minimum and maximum longitude and latitude.

Analysis

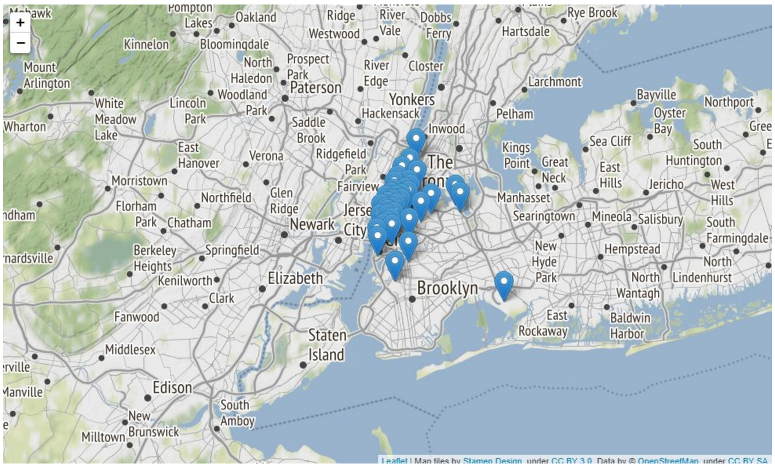


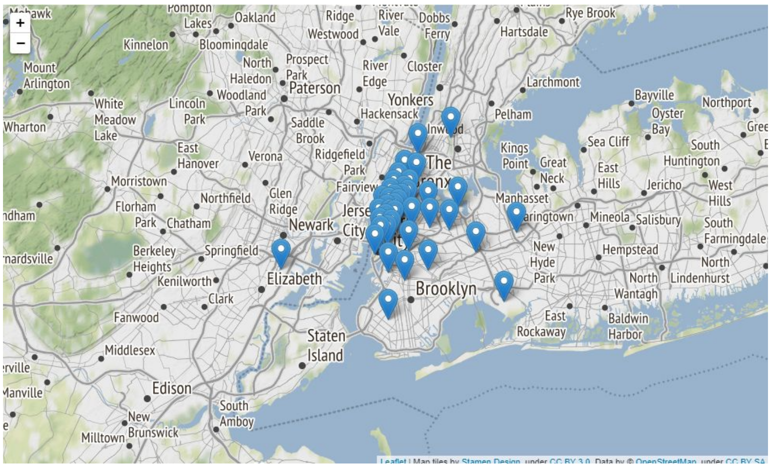
The figure above shows the taxi zones that the yellow taxi cab operates in New York City. The pickup (left) and dropoff (right) distributions are shown for October, November, December 2015 respectively.

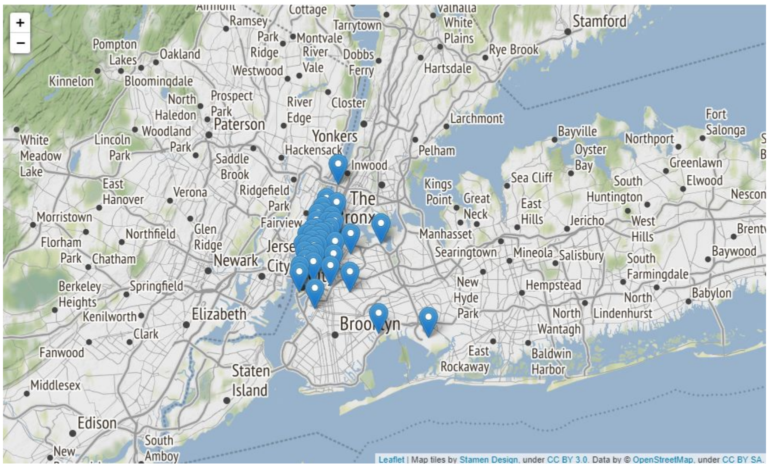

+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Elizabeth
Yonkers
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.

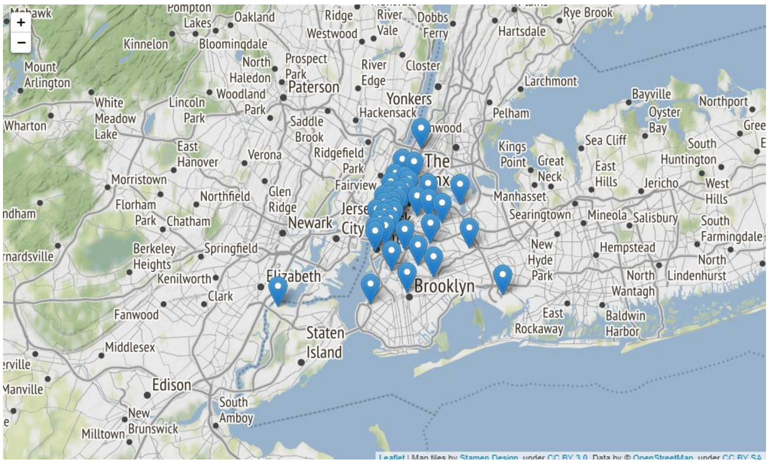
Adam Dulloo 995318

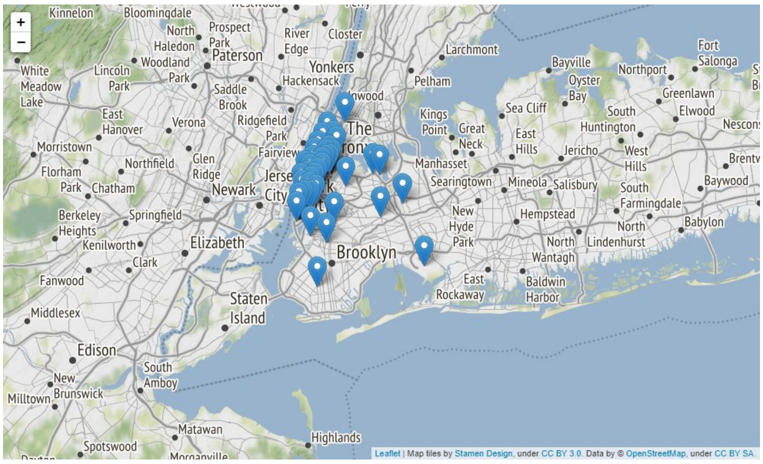
As there are so many points, k means clustering with k=40 was used. K=40 was found from trial and error of various different k.

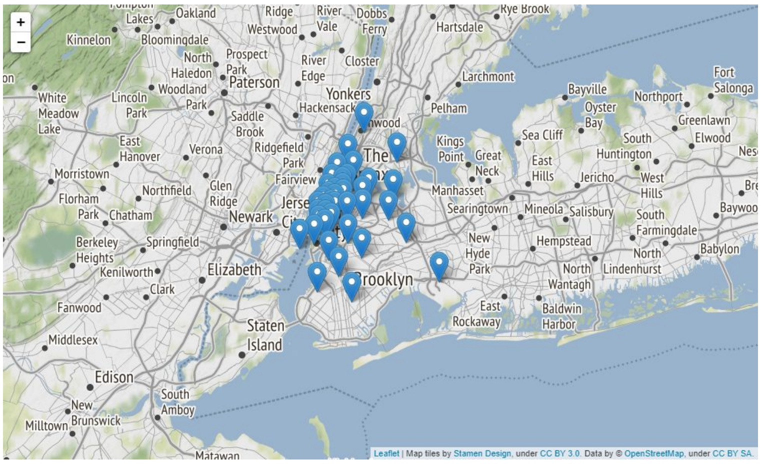

+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.


+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.


+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Stamford
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.

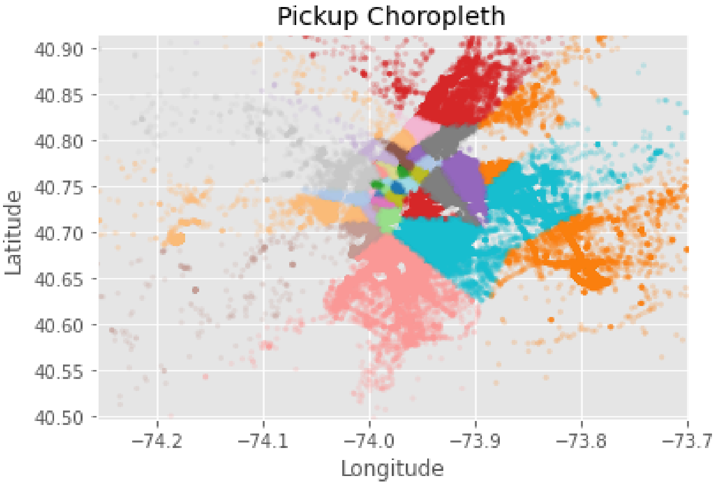

+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.

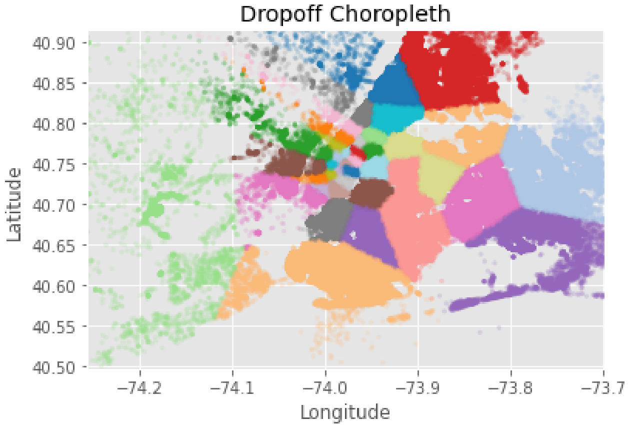

+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.

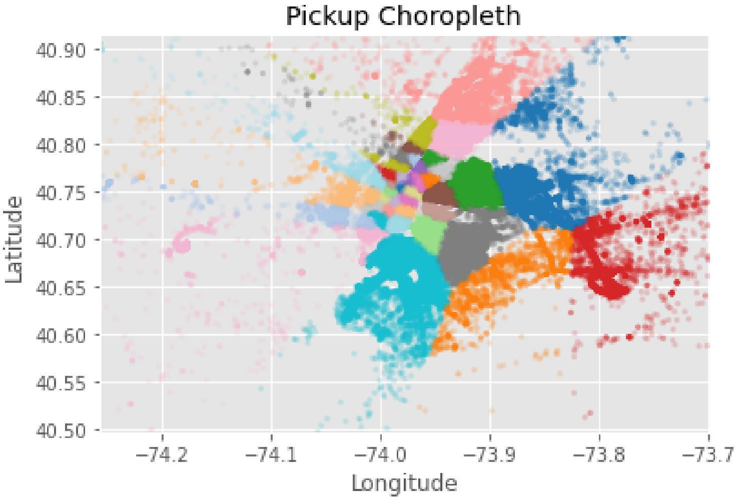

+
−
The Bronx
Brooklyn
Staten Island
Newark
Jersey City
Yonkers
Elizabeth
Leaflet | Map tiles by Stamen Design, under CC BY 3.0. Data by © OpenStreetMap, under CC BY SA.

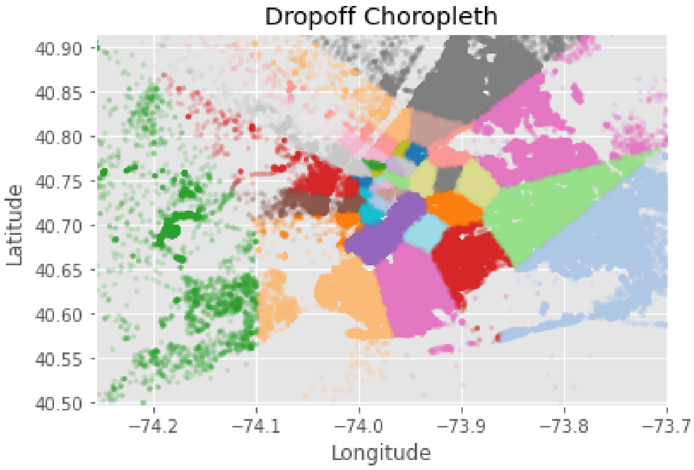
It is now clearer to see lots of clusters especially in the Manhattan borough. The pickup locations looked to be concentrated in the Manhattan borough, but the dropoff locations look to be a bit more spread out over Queens and Brooklyn.

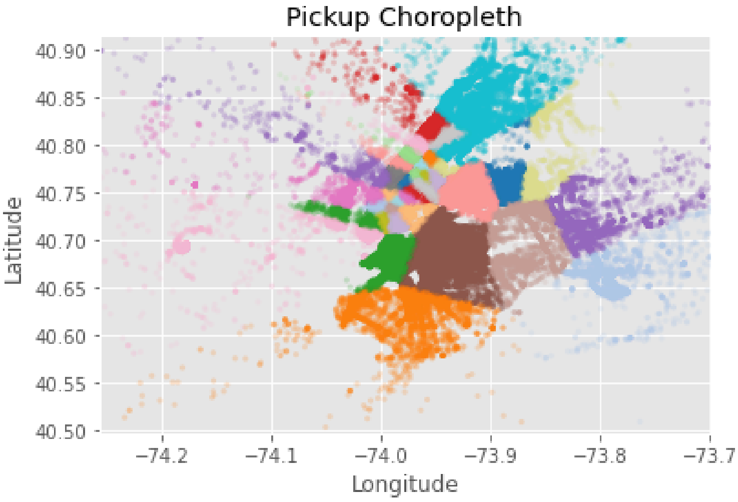
Adam Dulloo 995318

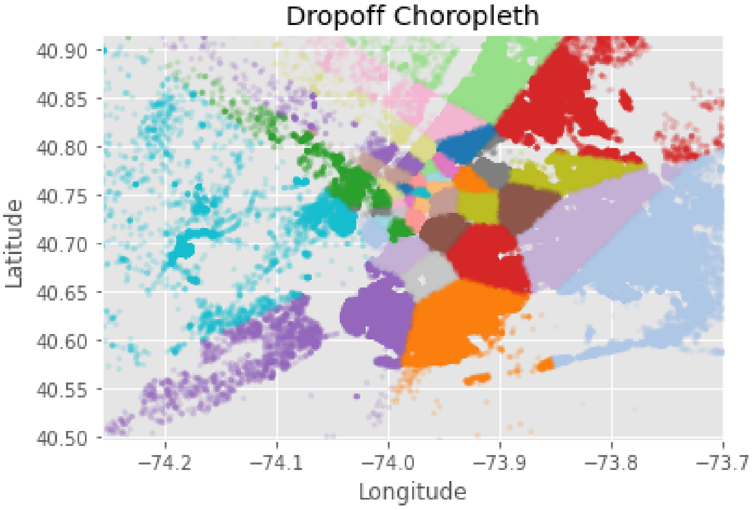






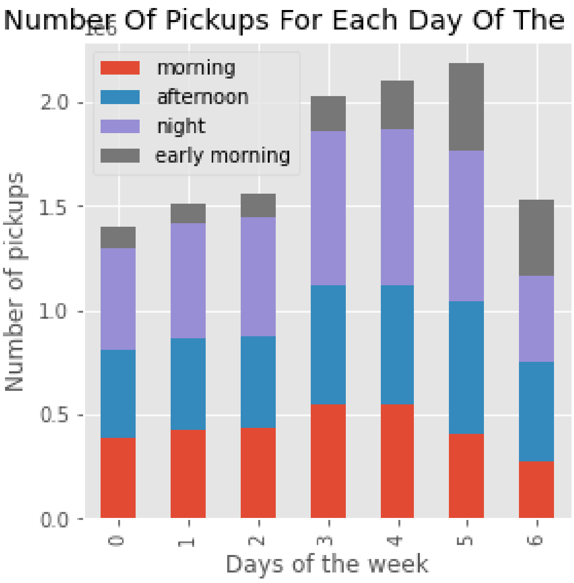


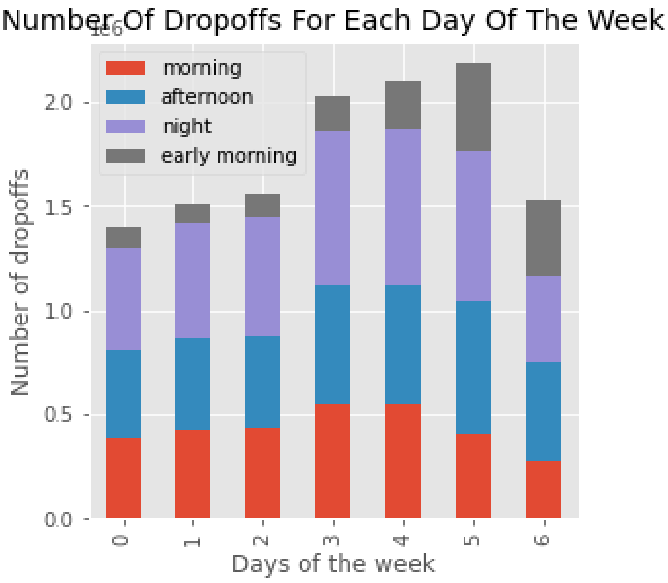


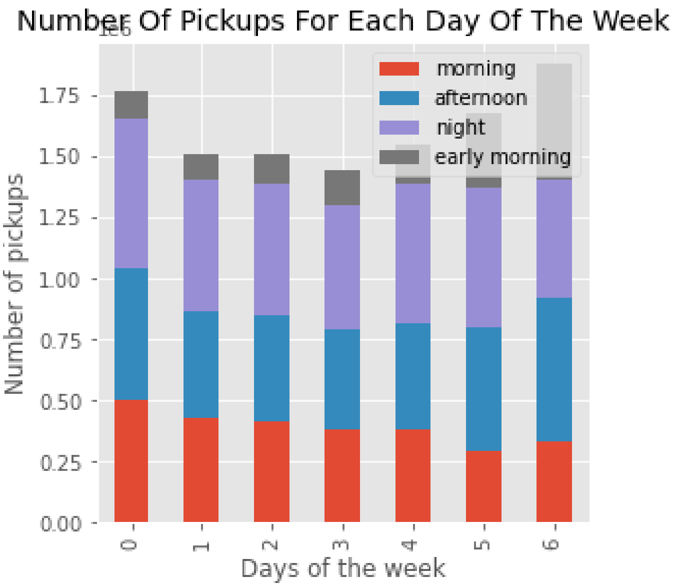


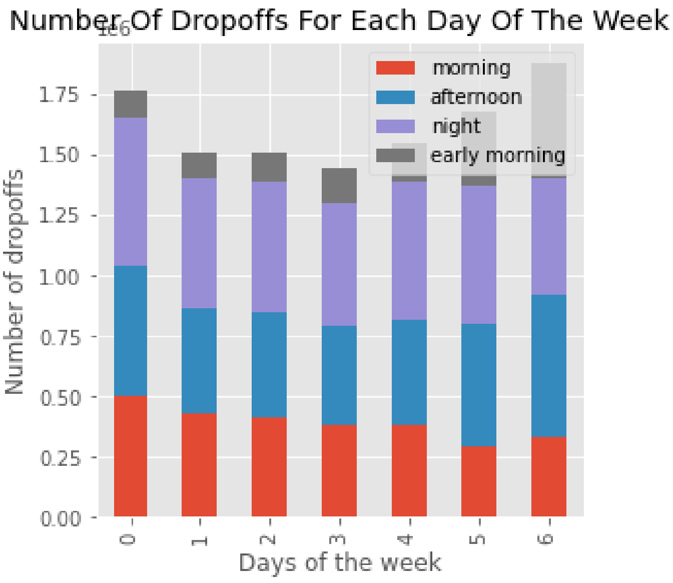
To confirm this, choropleth maps were made using the clusters of k-means. It is clear to see that pickups are heavily concentrated in the Manhattan borough and dropoffs whilst still concentrated in the Manhattan borough are definitely more dispersed throughout New York City. This means that most people live in the Manhattan borough and commute to work in the other boroughs, especially Queens and Brooklyn are more concentrated in the dropoff Choropleth maps. Comparing the three months, there aren't a lot of differences which makes sense as there weren't any significant reasons to leave the Manhattan borough for residence or any major job cuts that would drastically change the dropoff locations.

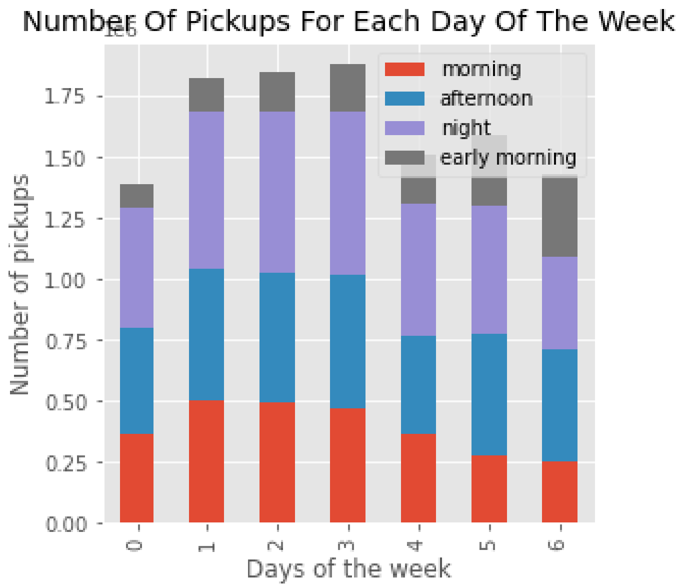
We can try and find a relationship with time by first graphing the number of pickups and dropoffs for each day of the week distributed by early morning, morning, afternoon and night. Graphs are shown in October, November and December 2015 respectively.

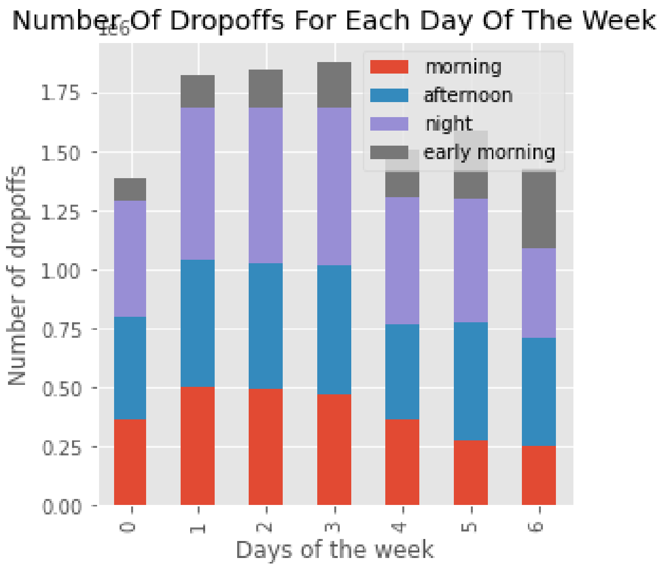




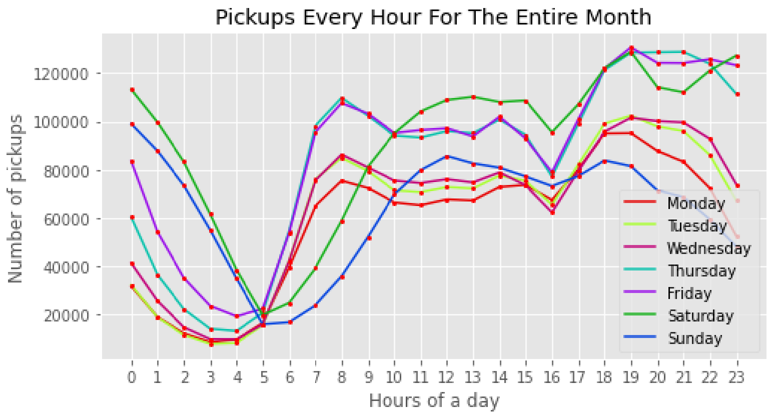


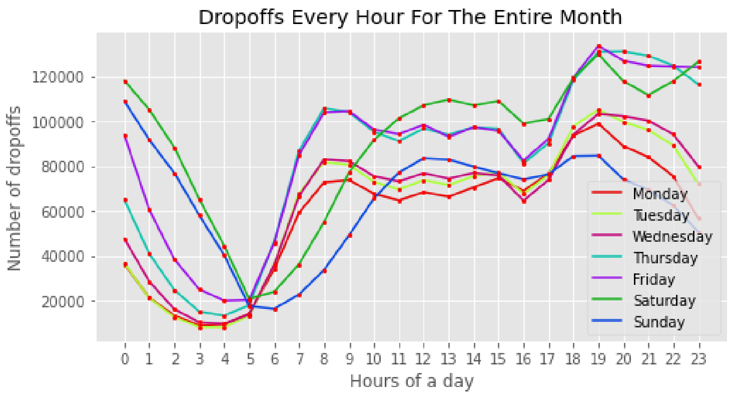




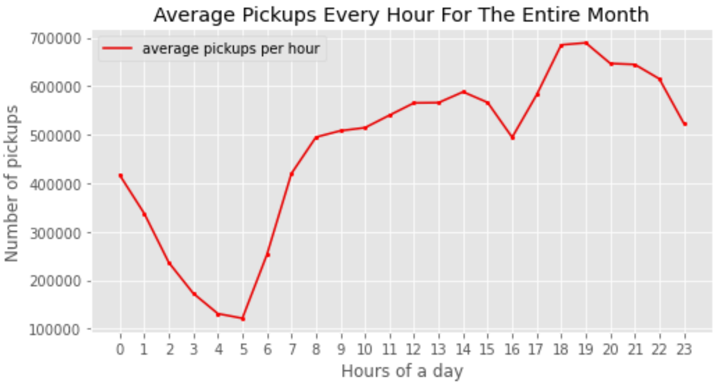


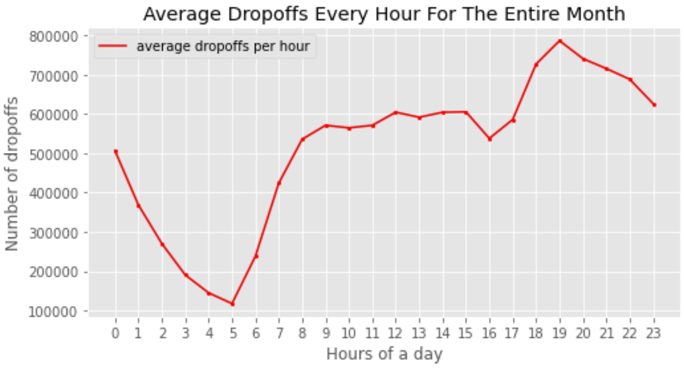
Days of the week are Monday (0), Tuesday (1), Wednesday (2), Thursday (3), Friday (4), Saturday (5), Sunday (6) and the time frames are early morning (23h to 6h), morning (6h to 12h), afternoon (12h to 18h), and night (18h to 23h). Interestingly, there is no one day that has the most pickups/dropoffs throughout the months as Saturday had the most for October, Sunday for November and Thursday for December. Saturday and Sunday makes sense as it is the weekend where most people have free time to go out. Thursday being the most popular day in December suggests that a lot of people use taxis to get to work from home and vice versa. Naturally the “early morning” section is high during the weekends and low during the first few days of the work week. In general, this aligns with what was expected, with the data showing that people probably preferred to use taxis to commute to work from home during Tuesdays, Wednesdays and Thursdays of December.

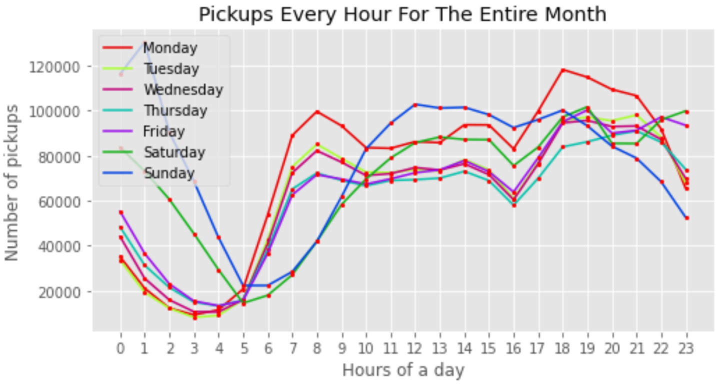


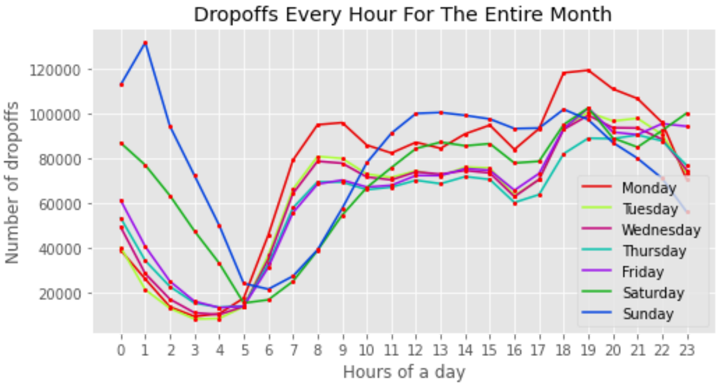


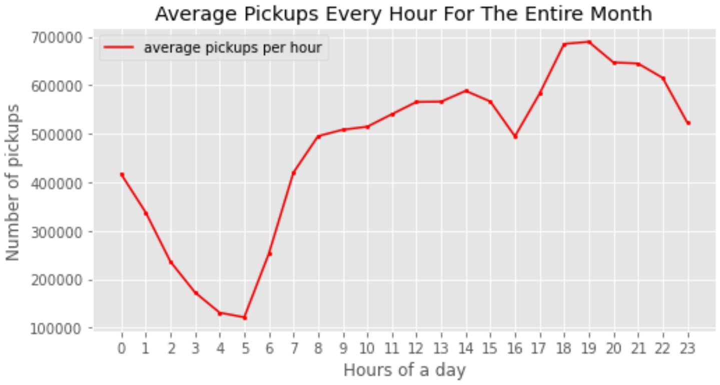
Adam Dulloo 995318

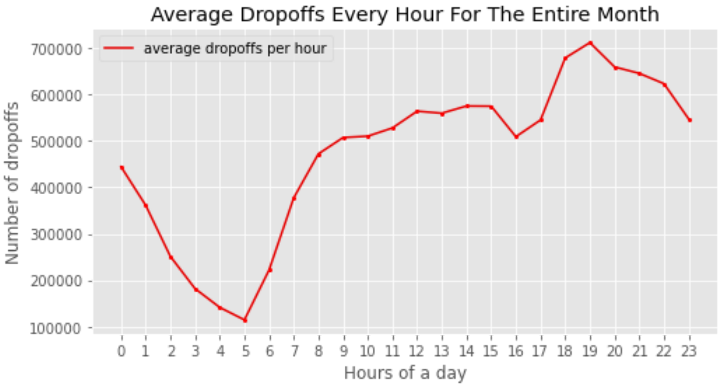


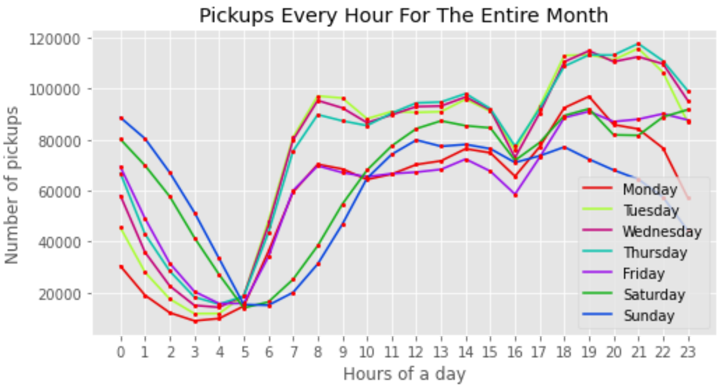


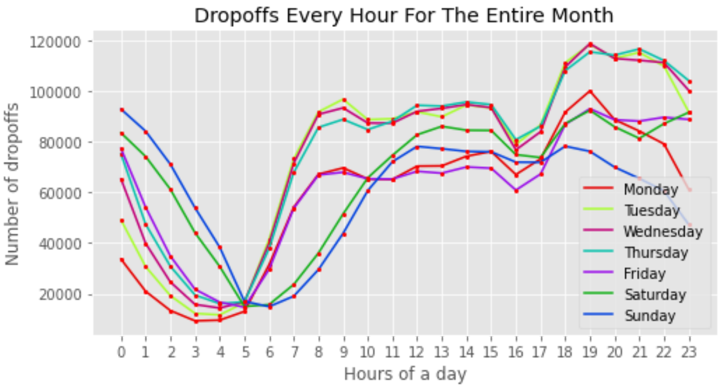


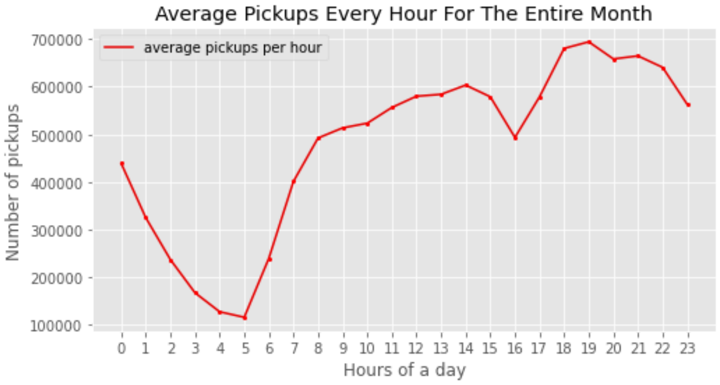


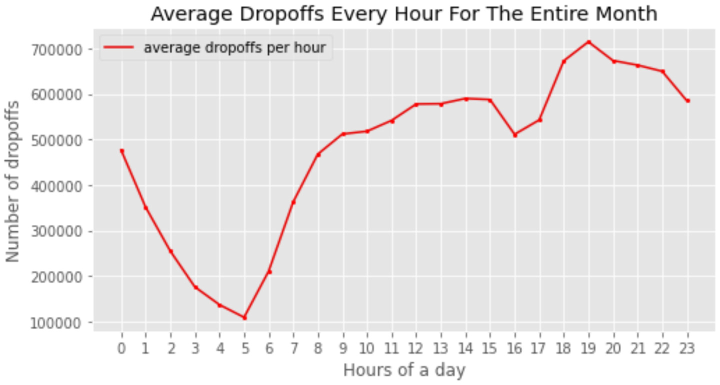












The graphs represent the pickups, dropoffs, average pickups and average dropoffs for the months of October, November, December 2015 respectively. Naturally, the pickups and dropoffs are similar as you would expect every pickup to have a dropoff roughly in the same hour or two. However, there is a significant trend throughout these months in that there is a dip from midnight to 5 a.m.Then a significant spike to around 9 a.m. followed by a steady rise to approximately 2 p.m. Next there's a dip in pickups/dropoffs from 2 p.m. to 4 p.m. before another spike to around 7/8 p.m and finally a gradual decrease in pickups/dropoffs till midnight. Interestingly enough, this is true for most of the days of the week as well with Saturdays having an increase from 9 p.m. to midnight and Fridays having a flat line from 9 p.m. to midnight. This makes sense as more people tend to go out and stay late on the week-ends. However, throughout the day from midnight to 9 p.m. all days of the week follow the same pattern which suggests that many people follow the same routines on the week-ends as on the week-days.

Conclusion

There is a definite correlation between the pickups/dropoffs and the time of day. The actual day of the week affects the rate of pickups/dropoffs less than expected. The only difference being an increase in midnight rates and a decrease in morning rates on the week-ends which is to be expected. If there is one day that commutes are the highest, more data would be needed to analyse. Furthermore, there is no one day that has a higher commute than others which is interesting as it would be expected that Friday commutes would be the highest since people need to go to work in the morning and come back home whilst also going out in the nighttime.

Resources

TLC Trip Record Data - Yellow Taxi Trip Records (October 2015, November 2015, December 2015) [ONLINE]
https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

TLC Trip Record Data - Yellow Trips Data Dictionary [ONLINE]
https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

TLC Trip Record Data - Taxi Zone Shapefile [ONLINE]
https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

New York City Department of City Planning 2020 New York City Borough Boundary [ONLINE]
https://www1.nyc.gov/assets/planning/download/pdf/data-maps/open-data/nybb_metadata.pdf?ver=18c