EAS 503 PROJECT

Chicago Taxi Trip Analysis

Group 4

NAMES	UBIT NAME
ADVAIT KULKARNI	advaitan
SUHIT DATTA	suhitdat
VARAD TUPE	varadsha

Abstract

The space of public transport has exploded in the recent years with the advent of many a ride-hailing services. A major concern for users has been about the pricing of rides dependent on timings, areas and surges. The project aims to perform an exploratory analysis on Chicago Taxi Trips Data.

Introduction

The data comprises of all the Chicago taxi records between April-July 2017 extracted from the Chicago Data Portal. It has a total of 113 million records containing features like trip duration, miles, area codes and pricing.

Data

Taxi Trip Data

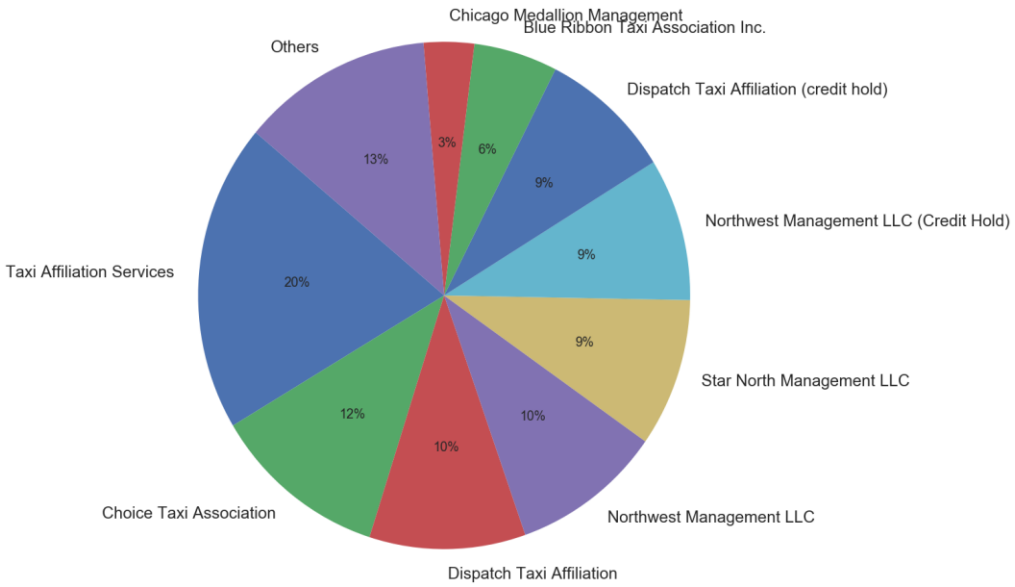
Column Name	Description	Type
Trip ID	A unique identifier for the trip.	Plain Text
Taxi ID	A unique identifier for the taxi.	Plain Text
Trip Start Timestamp	When the trip started, rounded to the nearest 15 minutes.	Date & Time
Trip End Timestamp	When the trip ended, rounded to the nearest 15 minutes.	Date & Time
Trip Seconds	Time of the trip in seconds.	Number
Trip Miles	Distance of the trip in miles.	Number
Pickup Census Tract	The Census Tract where the trip began. For privacy, this Census Tract is not shown for some trips.	Plain Text
Dropoff Census Tract	The Census Tract where the trip ended. For privacy, this Census Tract is not shown for some trips.	Plain Text
Pickup Community Area	The Community Area where the trip began.	Number
Dropoff Community Area	The Community Area where the trip ended.	Number
Fare	The fare for the trip.	Money
Tips	The tip for the trip. Cash tips generally will not be recorded.	Money

Column Name	Description	Type
Tolls	The tolls for the trip.	Money
Extras	Extra charges for the trip.	Money
Trip Total	Total cost of the trip, the total of the previous columns.	Money
Payment Type	Type of payment for the trip.	Plain Text
Company	The taxi company.	Plain Text
Pickup Centroid Latitude	The latitude of the center of the pickup census tract or the community area if the census tract has been hidden for privacy.	Number
Pickup Centroid Longitude	The longitude of the center of the pickup census tract or the community area if the census tract has been hidden for privacy.	Number
Dropoff Centroid Latitude	The latitude of the center of the dropoff census tract or the community area if the census tract has been hidden for privacy.	Number
Dropoff Centroid Longitude	The longitude of the center of the dropoff census tract or the community area if the census tract has been hidden for privacy.	Number

Analysis

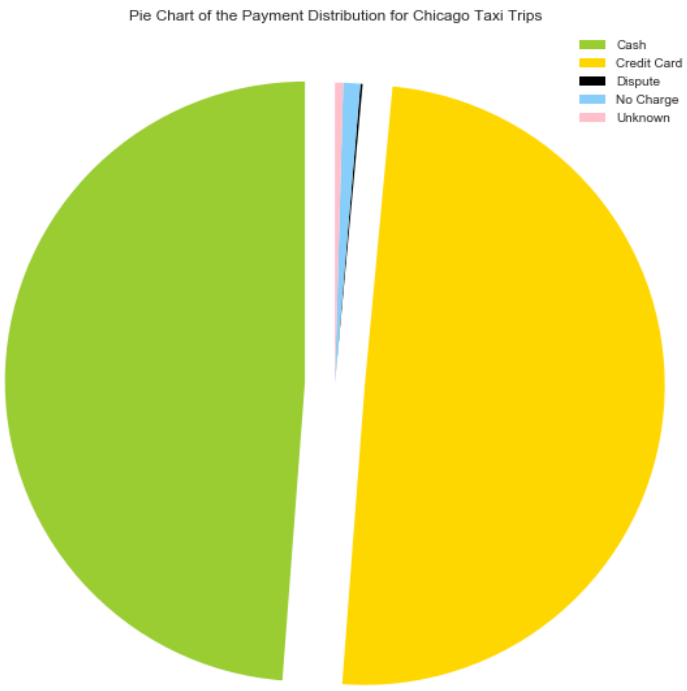
Distribution of Taxi ownership by each company

The following pie chart represents the share of each company in the Chicago Taxi space. Taxi Affiliation Services has the maximum share in Chicago.



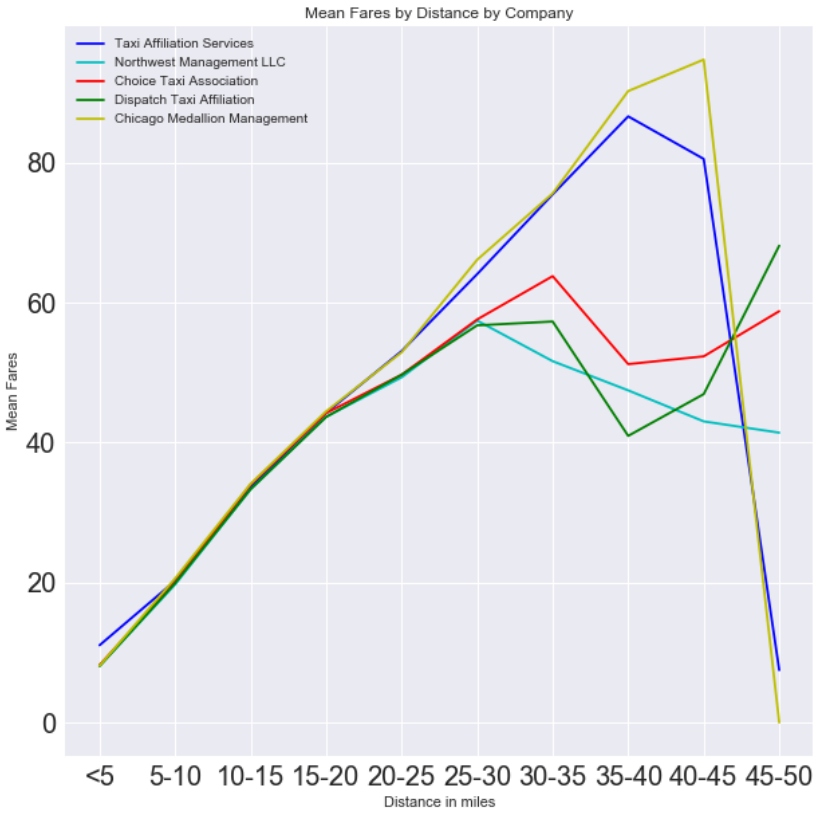
Payment distribution for taxi trips in Chicago

The distribution of payment types is as follows :-



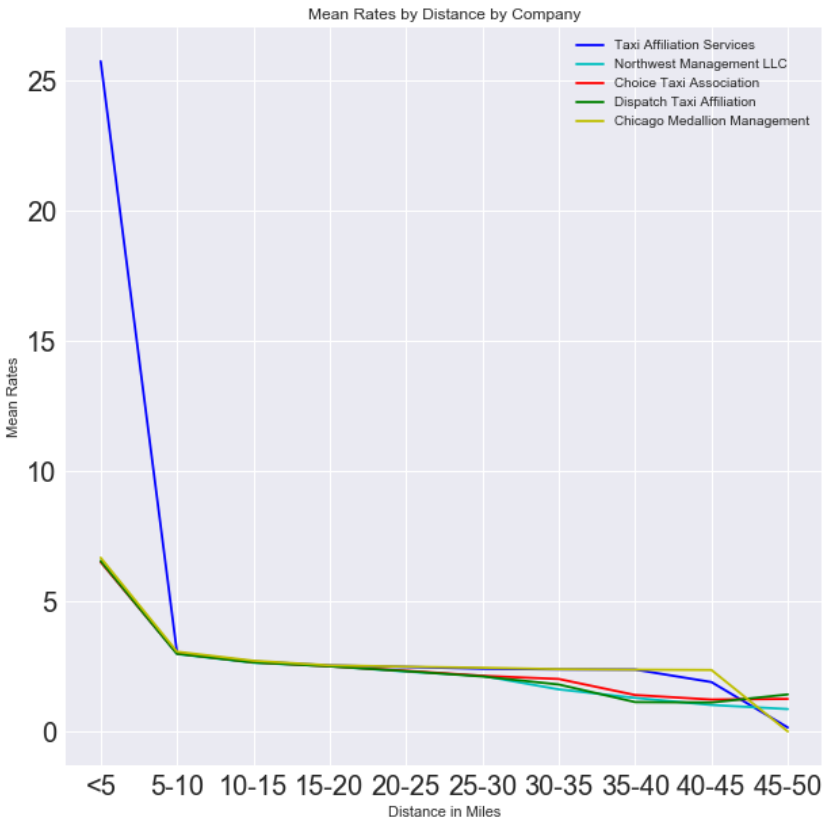
Fare vs Trip Miles of Top 5 taxi companies

The following plots display the variation of fares with distance for the Top 5 Taxi companies based on the volume of Trips.



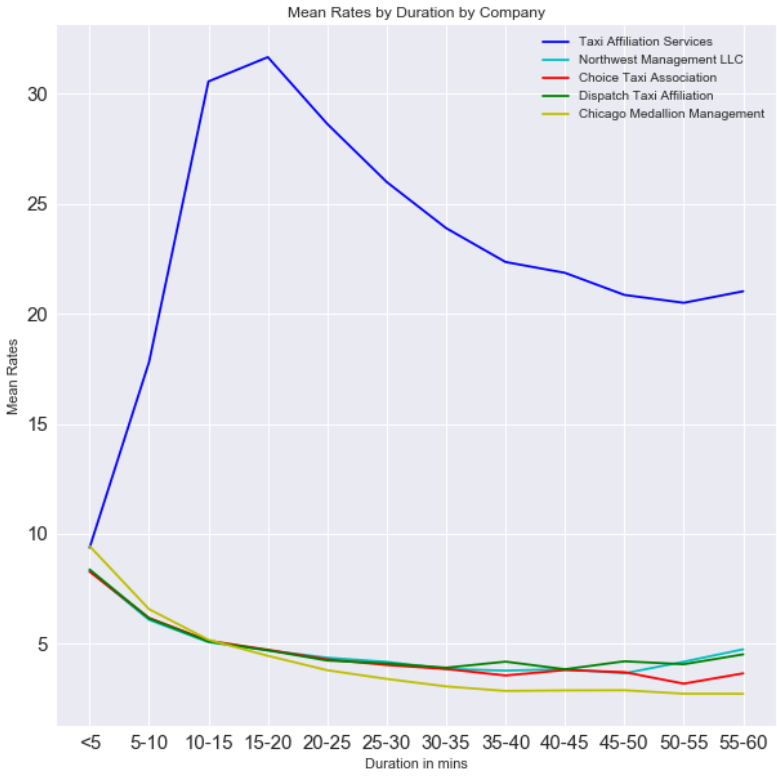
Fare Rate vs Trip Miles of Top 5 taxi companies

The following plots display the variation of Rates with distance for the Top 5 Taxi companies based on the volume of Trips :



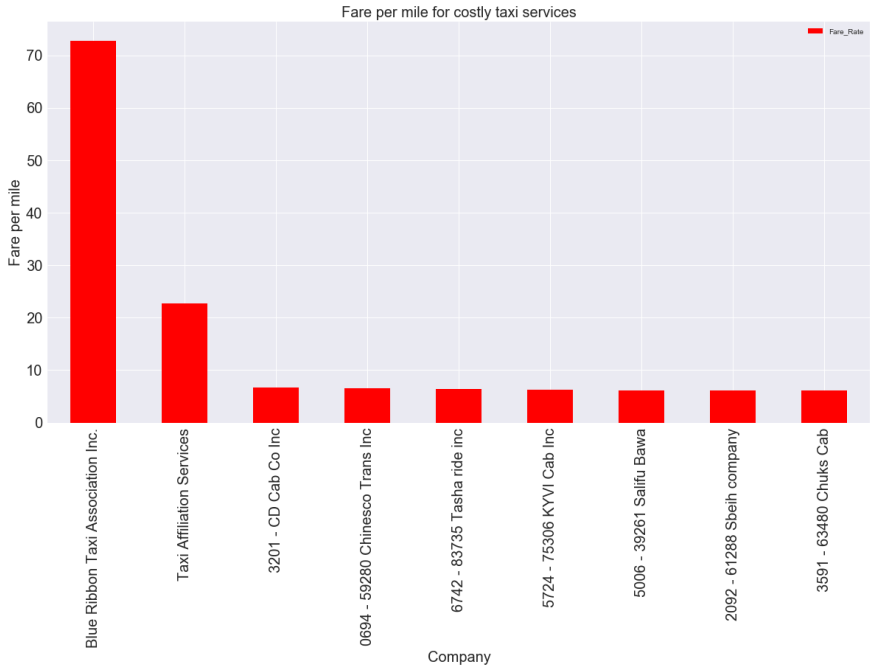
Fare Rate vs Trip Duration of Top 5 taxi companies

The following plots show the variation of fares and rates with duration :

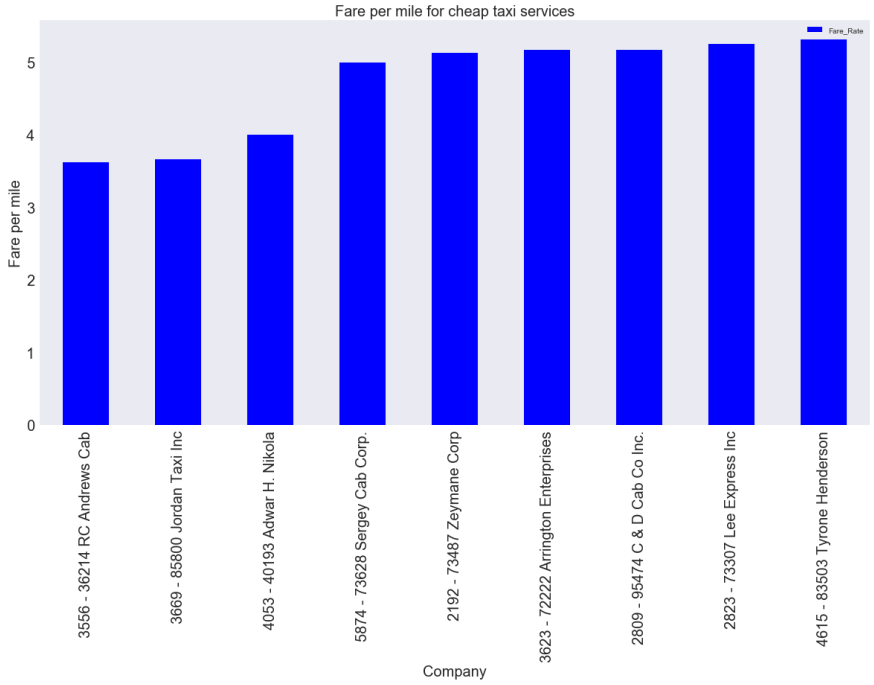


Top 10 expensive fare taxi companies

These plots show us the relative costs of the costliest and the cheapest Taxi services.

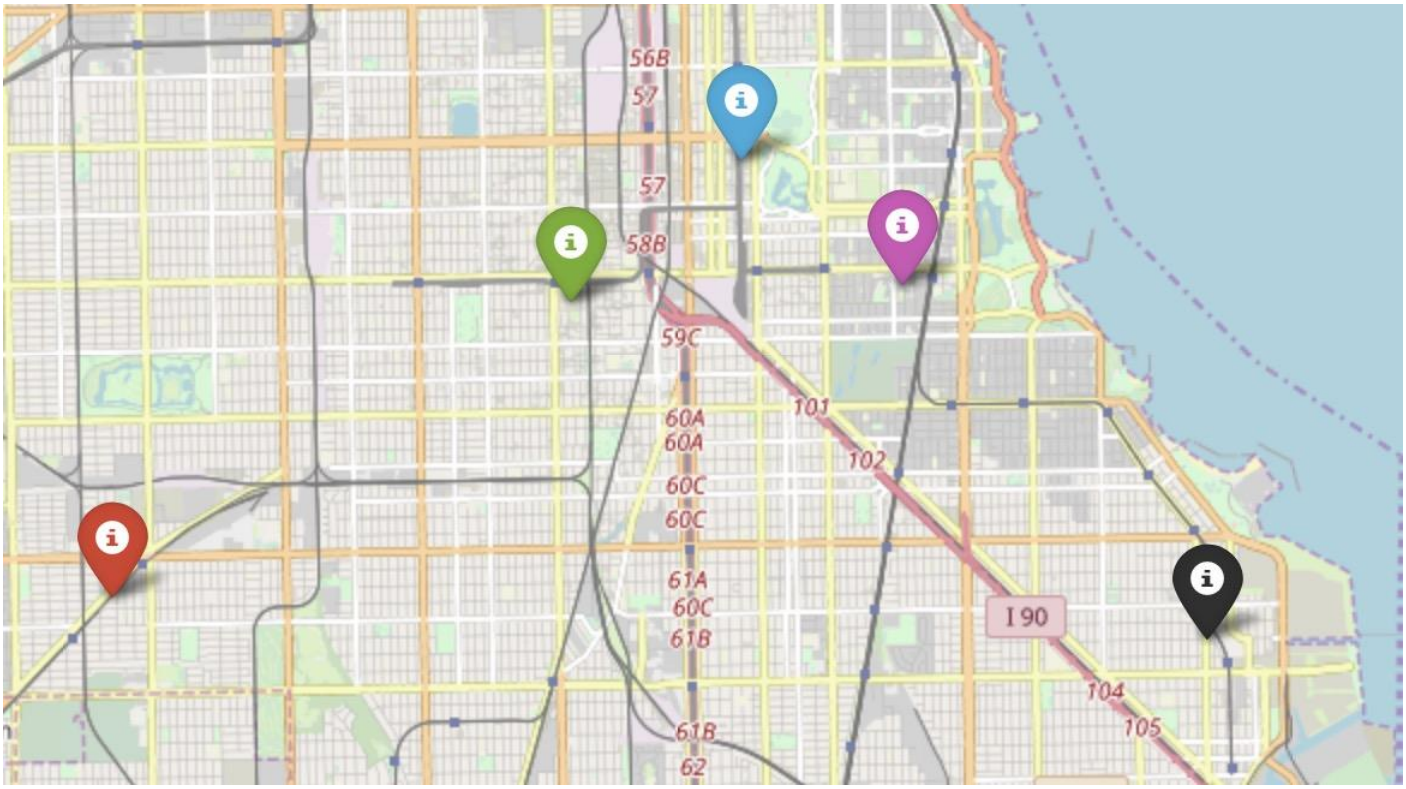


Top 10 cheapest fare taxi companies



Top 5 Community area with highest fare rate

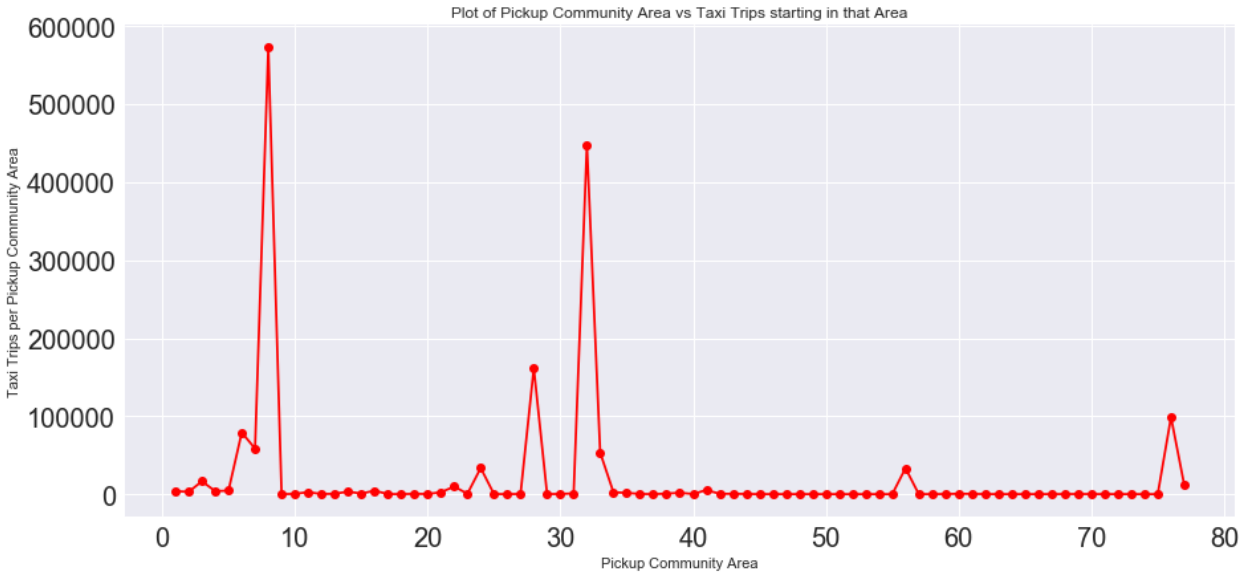
The image below shows the top community areas with highest average fare rate:

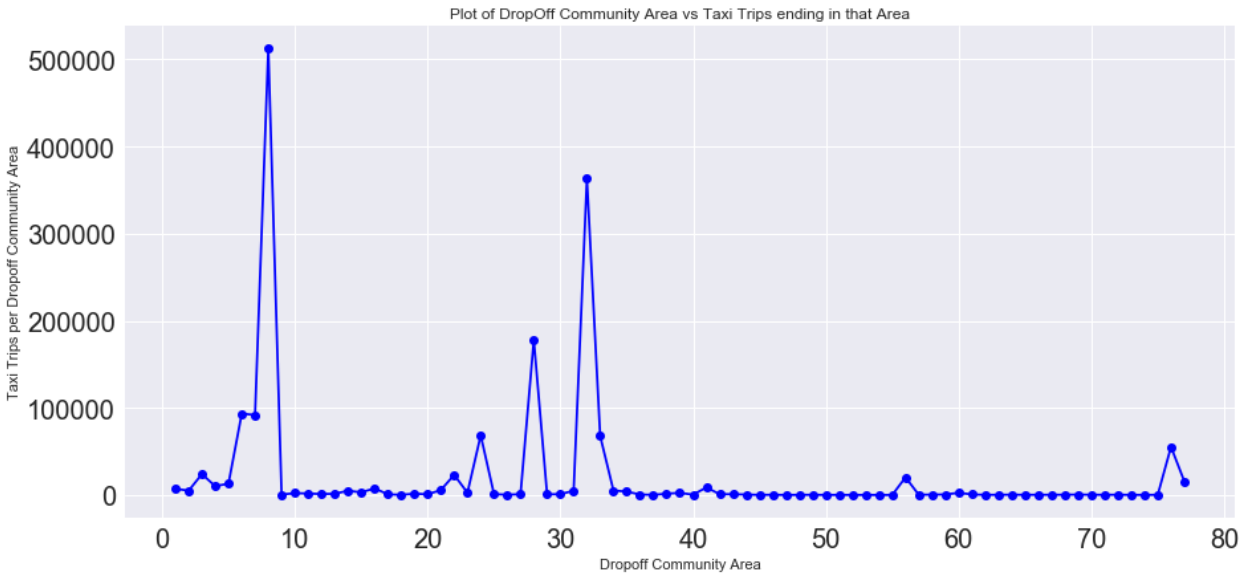


Area	Fare per Mile	Fare per Minute
South Chicago	39.824209	1.031072
Ashburn	37.744285	0.920967
Woodlawn	36.182858	1.081462
Washington Park	34.433751	1.03916
Englewood	32.776405	1.157654

Frequency of trips by pickup and dropoff community

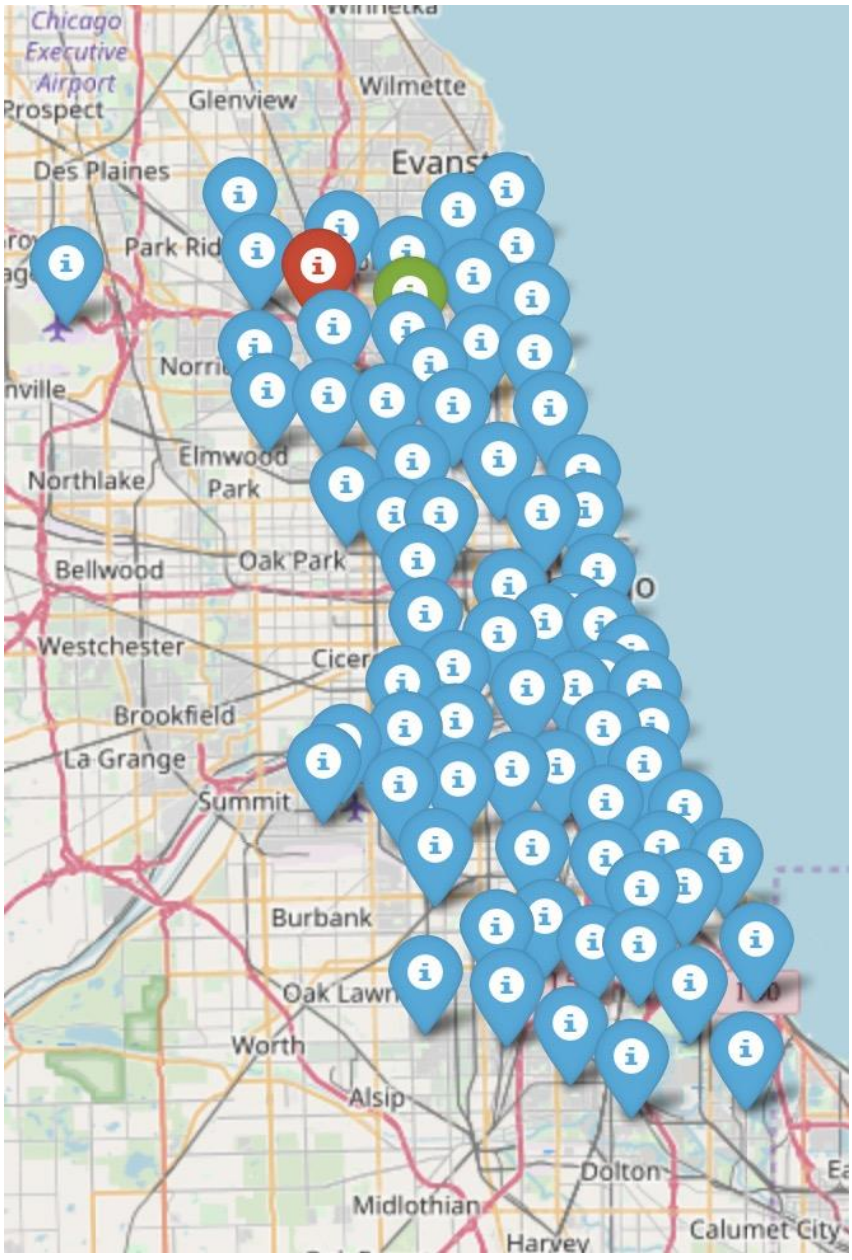
Plots of Pickup/Dropoff Community Area vs Taxi Trips starting/ending in that Area





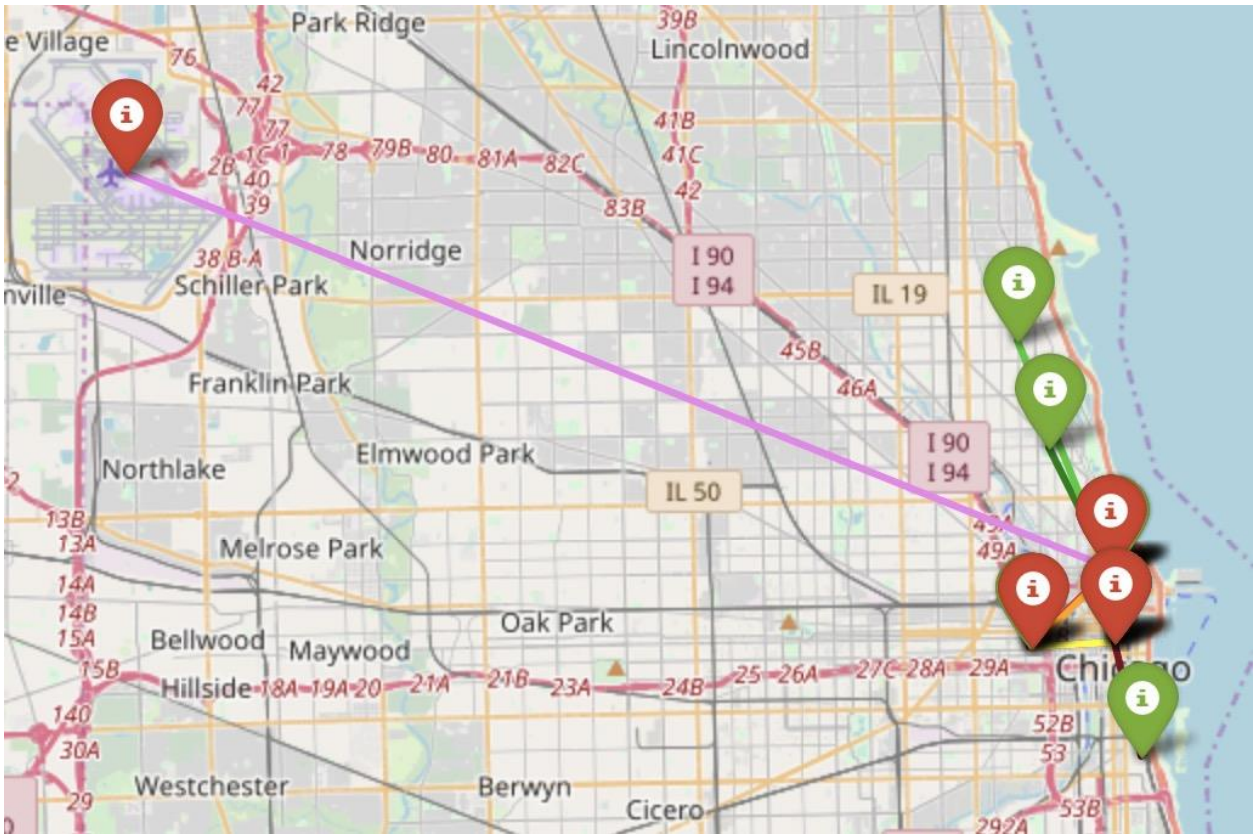
It can be seen that Community Area 8 and 32 have the maximum pickups and drop-offs.

Taxi Company dominance by Community Area



Taxi Affiliation Services dominates Chicago in most of the community areas except Dispatch Taxi Affiliation which dominates Jefferson Park and Northwest Management LLC which dominates at Albany Park

Top 10 busiest Taxi Routes in Chicago :



NO_OF_TRIPS	PICKUP_COM	DROPOFF_COM
152656	The Loop	Near North Side
140459	Near North Side	The Loop
67040	The Loop	Near West Side
63313	Near North Side	Near West Side
54009	Near West Side	Near North Side
45810	Near West Side	The Loop
36919	Near North Side	Lincoln Park
32351	O Hare	Near North Side
28767	Near North Side	Lake View
27502	The Loop	Near South Side

Conclusion

The project provided an elaborate understanding of the Chicago taxi business. It can be useful for newer players like Uber to move into bigger markets. Conversely, it can also be useful for traditional companies to maintain their businesses in their territories. It provides an idea of how competitive pricing can be used both at the user and the company's end.

Github Link : https://github.com/varadtupe/ChicagoTaxiTrip/

NBViewer Link : http://nbviewer.jupyter.org/github/varadtupe/ChicagoTaxiTrip/blob/master/Python%20Notebooks/Final_Chicago.ipynb