

Chicago Taxi Ride Analysis

Team 7:

Prateek Naharia
Tyler Christoforo
Linting Wang
Sifan Zhu
Jiadai Yu

Table of Contents

01

Project
Objective

02

Dataset
Description

03

Key Findings
& Analysis

04

Conclusion

Problem Objective

After Uber and other platforms took over the taxi market, it has left traditional taxi services quite vulnerable

Focus on traditional taxi services

- Analyze trends and patterns in the industry
- Determine potential opportunities
- Develop business decisions relating to taxi pricing, operation, and marketing

Dataset Description

**Data source: Recorded by
City of Chicago's regulatory agency**

<https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew>



- **Total Rows: 206,021,251 entries**
 - 81 traditional taxi companies
 - From 2013 to the current date
 - Updated regularly

- **Total Columns: 22 columns**
 - Fare
 - Tips
 - Duration
 - Tolls
 - Pick-up & Drop-off Location
 - ...



Data Cleaning

Problem Approach

- Hard to manipulate due to data size
 - A significant amount of null values
 - Long run times
- Partition the full data into sets
 - Standard imputing, drop if fail
 - Create problem specific dataframes

Note:

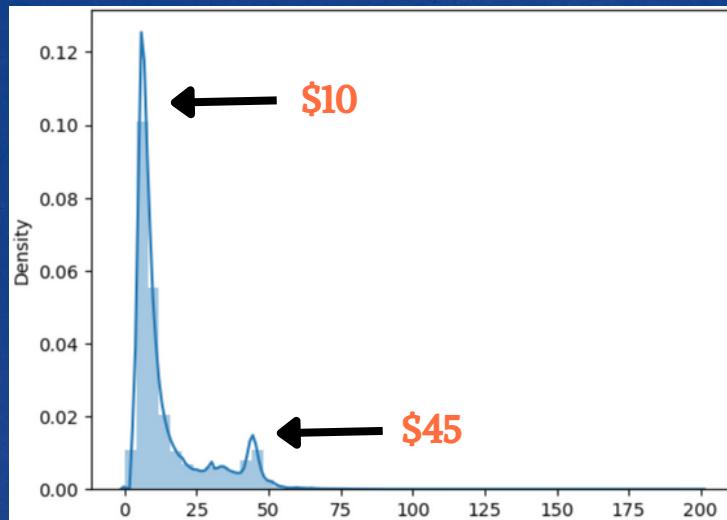
Filter down earlier years with incomplete documentation and missing entries

	# of Existing high null count variables	Max number of null counts
2013	11	11,221,939
2014	10	12,508,340
2015	10	11,458,239
2016	10	11,130,667
2017	9	8,134,193

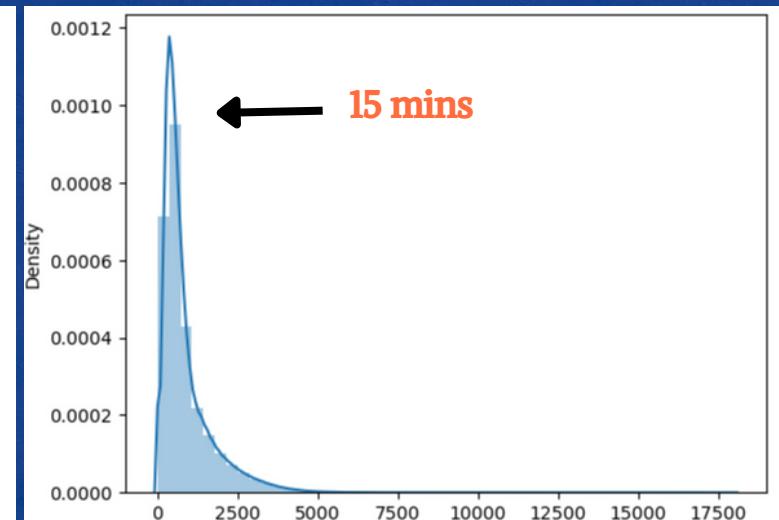
The Current Market

-- CONSUMER BEHAVIOR AT A GLANCE

Price points customers are more likely to take a taxi



Trip Fare Distribution



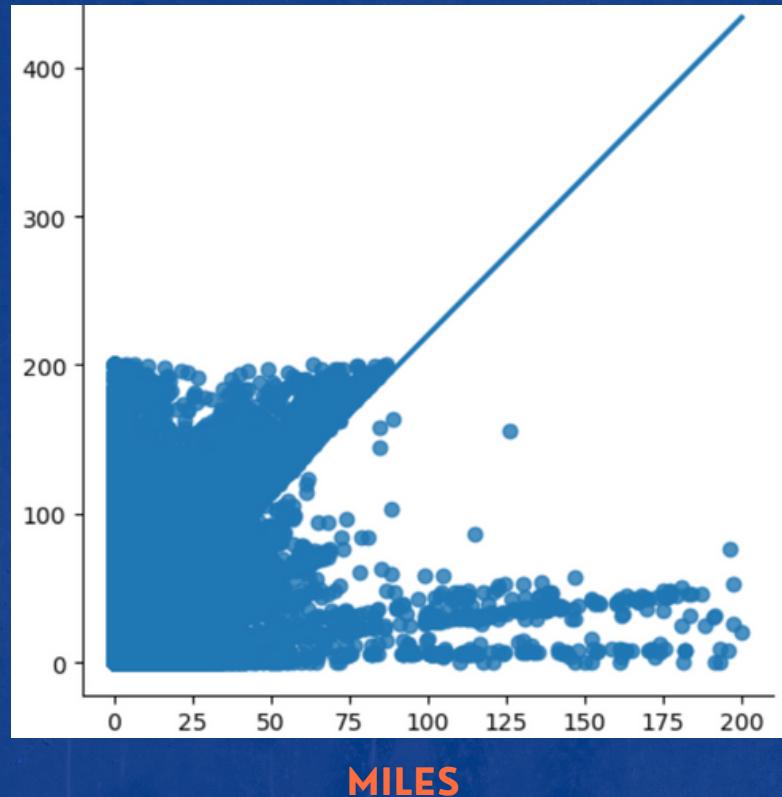
Trip Total Seconds Distribution



The Current Market

-- CORRELATIONSHIP BETWEEN FARE AND MILES

Different pricing strategies

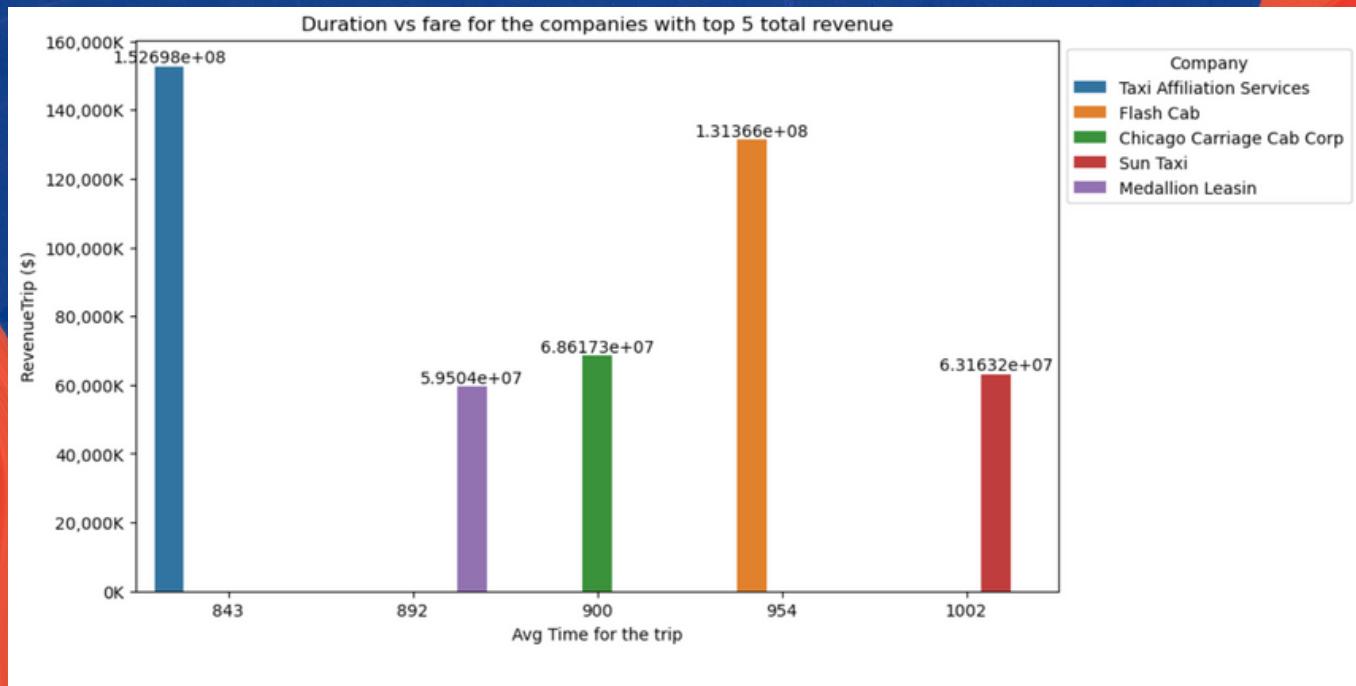


- For trips under 60 miles, fares vary widely, ranging from 0 to 200 dollars (traffic and tolls)
- For trips longer than 60 miles, distance is the key factor



Top Performing Companies

-- THE COMPANIES CAPITALIZING WELL ON THE MARKET

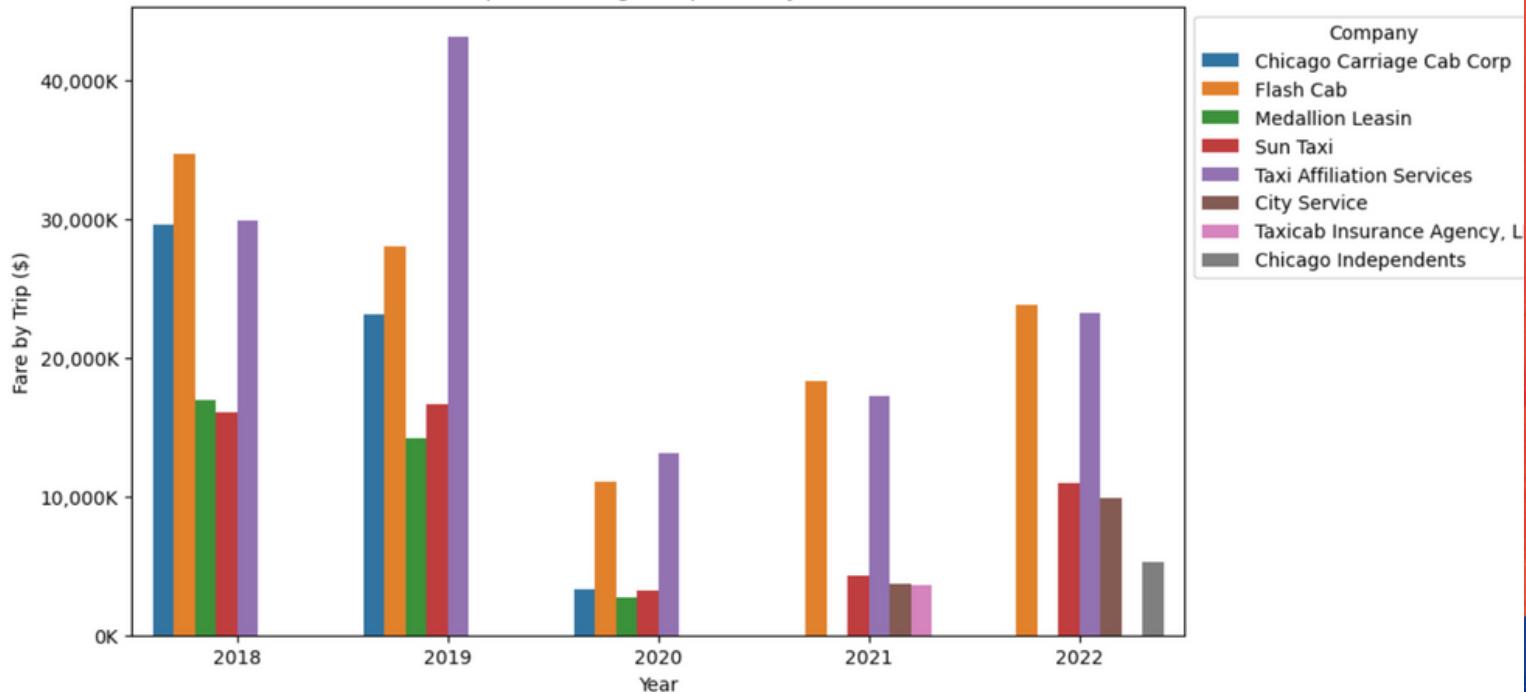


- Companies with top5 revenue have no relationship with trip duration

Top Performing Companies

-- THE COMPANIES CAPITALIZING WELL ON THE MARKET

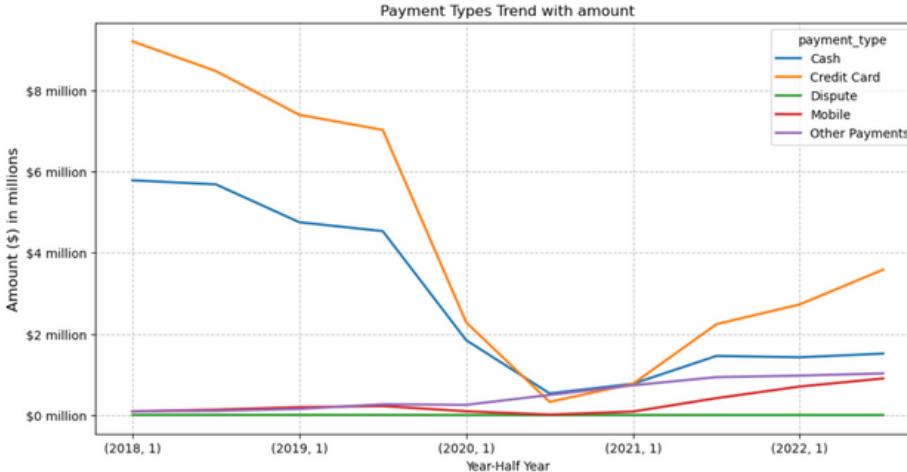
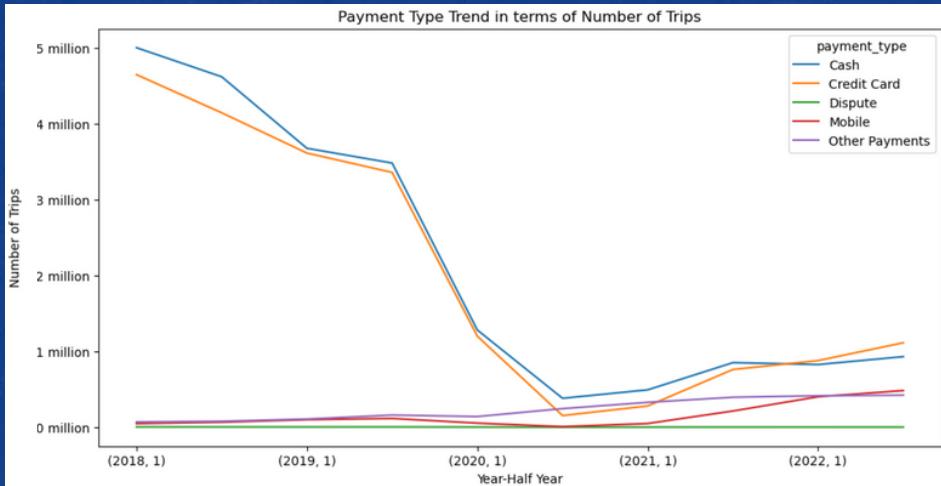
Top Performing Companies by Year



- Bigger players tend to stay in the market lead
- Some companies exit the market or get acquired

Holistic Trends

-- TRENDS OF PAYMENT METHODS CHANGING

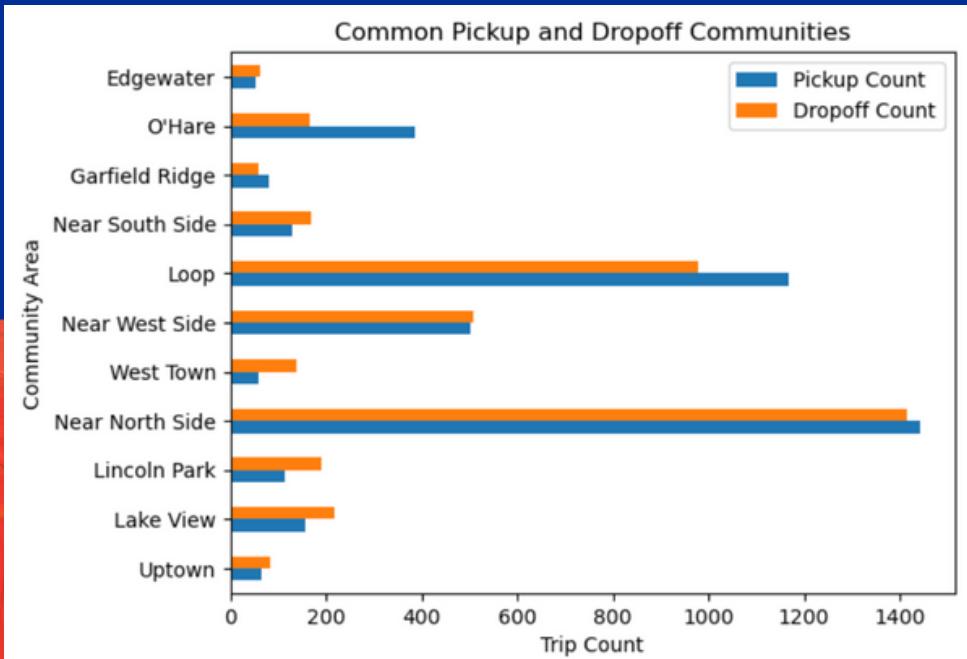


- Total trips dropped dramatically after Covid, increased again after Jan 2021
- Credit card total payment amount increased faster than others



Pickup and Dropoff Locations

WHERE ARE THE HOT SPOTS FOR TAXIS

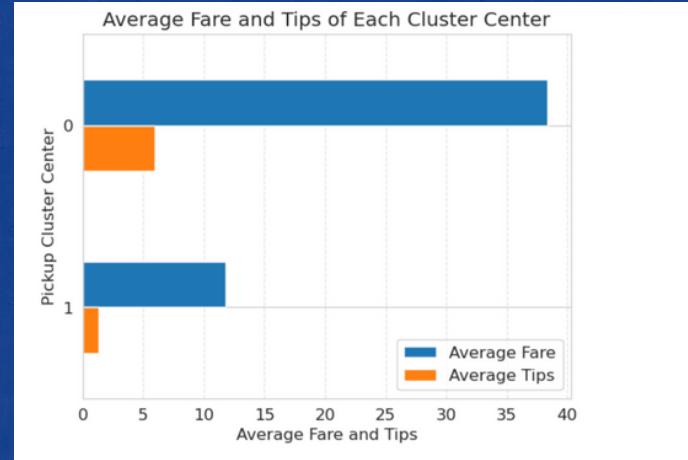
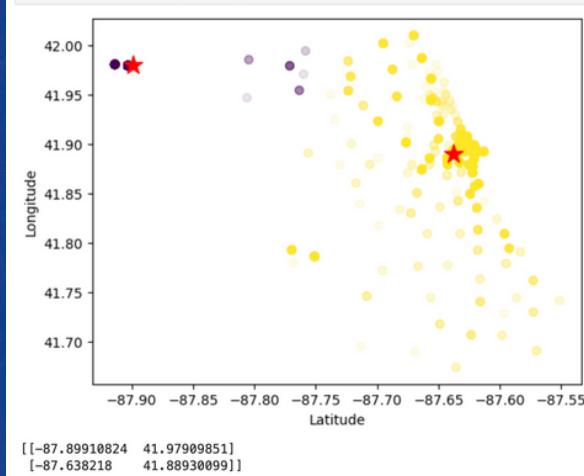


- Locations with higher traffic flow have both more pickups and dropoffs
- Passengers are departing Chicago O'Hare Airport, Loop and Near North Side
- Taking ride to tourists attractions e.g. Lincoln Park and Lake View

Additional Opportunity

MORE TRIPS VS HIGHER TIPS

```
print(kmeans.cluster_centers_)
```



- Pickup Center 0: Chicago O'Hare Airport
Average 13.3 miles per trip, average fare \$38.39, 15.5% tips.
- Pickup Center 1: CBD
10 times more trips, average trip mile 2.7, average fare \$11.77, 11.4% tips.

Conclusions

- No strong variable correlation
- Payment trends
 - Provide more convenient payment methods
- Fare pricing strategies
 - prioritize miles and location, observe actions from top companies

Problem	Potential solution
Diminishing customer size	Create or merge platforms, exhibit DiDi
Weak correlation between variables	Categorize drivers and customers
Two centroids that are very saturated	Design and optimize routes



Thank you!