Chicago Taxi Trips

Trends, Insights, and Operational Efficiency

Syamsul Rizal Fany
saemfany@gmail.com
September 2024

Introduction

The **Chicago Taxi Trips** dataset provides detailed information about taxi journeys in the city of Chicago, capturing key attributes such as pickup and dropoff locations, trip durations, distances, fares, tips, and payment methods. This dataset offers valuable insights into the patterns of taxi usage, customer behavior, and operational efficiency within the city's transportation network. By analyzing this data, we aim to uncover trends related to popular routes, fare structures, peak hours, and the performance of different taxi companies. The insights gained can inform decisions for optimizing services and improving customer experiences.

Dataset Overview

The **Chicago Taxi Trips** dataset contains detailed information about taxi trips in Chicago, covering several years of trip data. The dataset includes over **200 million records**, each representing a single taxi trip, with information on **trip start and end times**, **pickup and dropoff locations**, **trip duration**, **fare**, and other important trip characteristics.

Time Period:
The dataset includes trips from **various years**, allowing for the analysis of trends over time. In this report, we focus on specific time periods, such as **2013** and **2023**, to explore **annual patterns** and **seasonal trends**.

Data Usage:
This dataset provides insights into taxi operations in Chicago, including **route popularity**, **customer behavior**, and **fare patterns**, making it a valuable resource for **business optimization** and **policy-making**.

Dataset Overview

Key Columns:

- **unique_key**: A unique identifier for each taxi trip.
- **taxi_id**: An identifier for the taxi that performed the trip.
- **trip_start_timestamp**: The exact time when the trip started.
- **trip_end_timestamp**: The exact time when the trip ended.
- **trip_seconds**: The total duration of the trip in seconds.
- **trip_miles**: The distance traveled during the trip in miles.
- **pickup_community_area**: The community area where the trip started.
- **dropoff_community_area**: The community area where the trip ended.
- **fare**: The fare charged for the trip (excluding tips and tolls).
- **tips**: The tips given by the passenger.
- **tolls**: The amount charged for tolls during the trip.
- **trip_total**: The total amount charged for the trip, including fare, tips, tolls, and extras.
- **payment_type**: The method of payment used by the customer (e.g., cash, credit card).
- **company**: The taxi company that operated the trip.

Tools and Technologies

SQL

Queries were run in **Google BigQuery** to extract, aggregate, and manipulate large amounts of data

Python

Pandas and **NumPy** were used for additional data wrangling, and **Matplotlib** and **Seaborn** for visualization

Google Colab

The project was executed in **Google Colab** for interactive analysis and visualization

Analysis 1

Calculate the average, median, and standard deviation of trip duration (trip_seconds) for trips made on **Monday** and **Saturday**. Compare the results for both days.

Row	weekday	avg_seconds	median_seconds	stddev_seconds
1	Saturday	737.41055150983675	552	1146.3358880111678
2	Monday	839.47978351934137	540	1346.0602084326094

Monday trips generally take longer on average, but they also have greater variability, with a larger spread of trip durations.

Saturday trips tend to be more consistent in duration but have a slightly higher median, which suggests that while most trips are similar in length, they may be a bit longer than those on Monday.

These findings could indicate different travel patterns or congestion factors between weekdays and weekends.

Analysis 2

Find the top five routes (from the pickup_community_area to dropoff_community_area) with the most trips in the year 2023.

The most popular route is **within Community Area 8**, with **464,844 trips** starting and ending in this area. This suggests that **Community Area 8** is a central hub, possibly a business district or a high-traffic area where people frequently travel short distances within the same area.

Row	pickup_community_area	dropoff_community_area	num_trips
1	8	8	464844
2	32	8	291722
3	76	8	274747
4	8	32	267673
5	32	32	241596

Community Area 8 also stands out as a central destination, receiving the highest volume of trips from other areas, especially from **Community Areas 32 and 76**.

Community Area 32 is another key location, serving as both an origin and destination for a large number of trips, with significant movement within its own area as well.

Analysis 3

Compare the average taxi trip cost (fare, tips, and taxes) based on payment methods in the year 2019.

Credit Card and Mobile payment methods are the most lucrative for drivers, yielding higher fares and tips.

Prepaid fares are higher but offer no tips, while **Cash** and **Disputed** payments show lower fares and minimal tips.

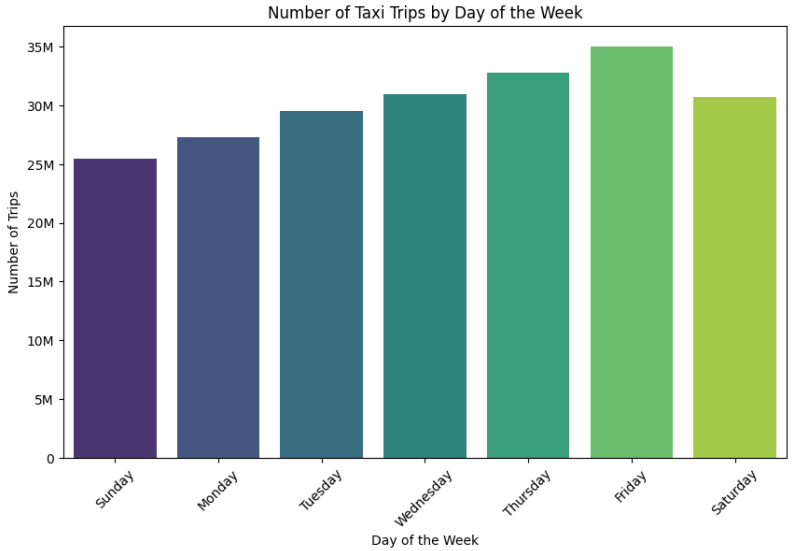
Row	payment_type	average_fare	average_tips	average_tolls
1	Prepaid	19.46141479...	0.0	0.0
2	Credit Card	16.81413780...	3.774544696...	0.002266943...
3	Prcard	16.13000271...	0.203527591...	0.001529141...
4	Mobile	15.97317403...	3.112698023...	4.968012969...
5	Unknown	15.87237993...	0.082129280...	0.000147177...
6	No Charge	15.75350963...	0.249531863...	0.019141699...
7	Dispute	15.65584022...	0.001445658...	0.078991435...
8	Cash	12.97830635...	0.002387440...	0.001790926...
9	Pcard	11.32747148...	0.038022813...	0.0

Different payment methods may appeal to different customer segments, with credit card and mobile users showing more willingness to tip compared to cash or prepaid users.

Analysis 4

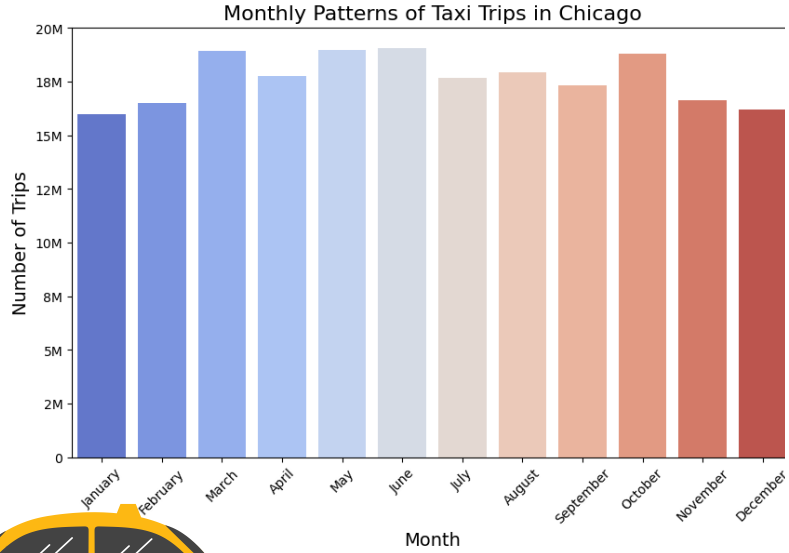
Are there any seasonal patterns in taxi trips?
Daily Patterns

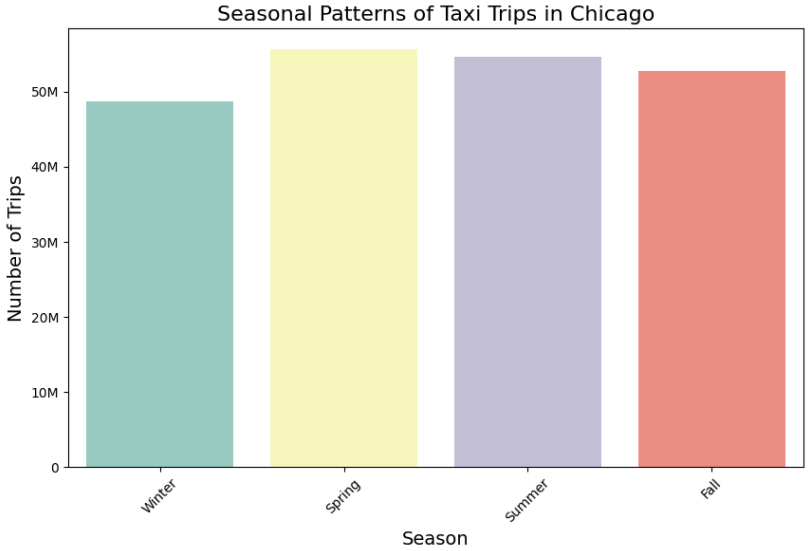
Taxi demand is highest on **Fridays** and gradually builds throughout the week, peaking before the weekend.
Sunday experiences the lowest demand, reflecting reduced activity and commuting.
This pattern suggests that taxi services can optimize their fleet and staff resources based on the expected increase in demand from Monday to Friday, especially focusing on peak times on Friday and Saturday.



Analysis 4

Are there any seasonal patterns in taxi trips?
Monthly and Seasonal Patterns





Analysis 4

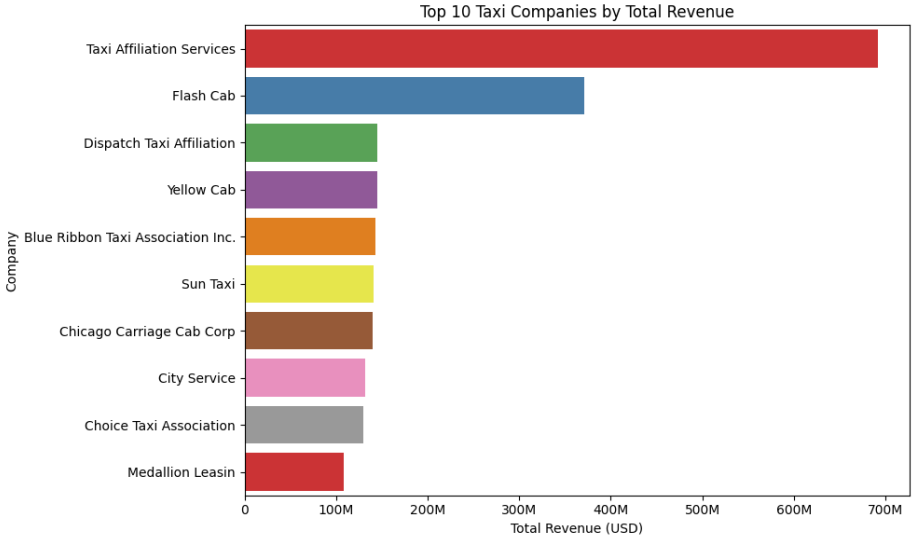
Are there any seasonal patterns in taxi trips?
Monthly and Seasonal Patterns

Spring is the busiest season, with **May** and **June** leading the way, while **winter** is the slowest, driven by low demand in **January** and **February**. Taxi operators can anticipate higher demand in spring and summer and adjust their services accordingly, while preparing for a slowdown during the winter months.

Analysis 5

Which taxi company has the highest revenue?

Taxi Affiliation Services and **Flash Cab** dominate the Chicago taxi market in terms of revenue, controlling a large portion of the industry's income. Mid-tier companies like **Dispatch Taxi Affiliation** and **Yellow Cab** are still key players but operate on a smaller scale. The concentration of revenue among a few companies suggests a competitive landscape where a few large players lead the market, while smaller companies cater to more localized or specialized segments.

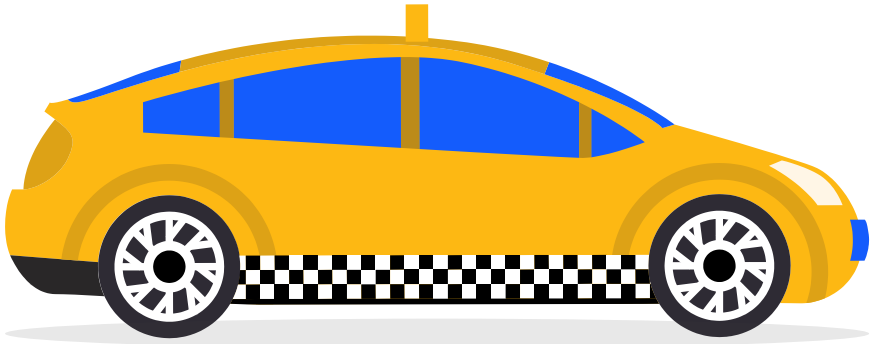


Recommendation and Conclusion

1. Top Companies and Revenue Distribution

Recommendation: Taxi operators, especially smaller companies, should analyze the strategies of market leaders like **Taxi Affiliation Services** and **Flash Cab** to identify potential areas for growth. These companies might be leveraging technology, customer loyalty programs, or more extensive networks that can be emulated or adapted by smaller companies.

Conclusion: The taxi market in Chicago is highly concentrated, with a few key players dominating revenue. Smaller companies must differentiate themselves, perhaps through niche services, better customer experience, or innovative payment solutions to increase their market share.

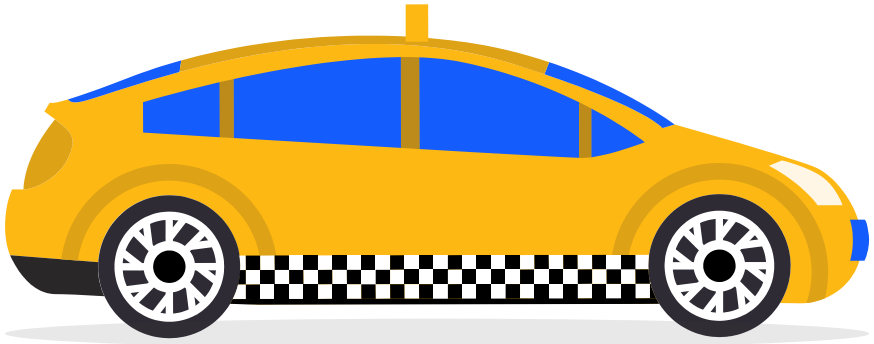


Recommendation and Conclusion

2. Seasonal Patterns

Recommendation: Taxi companies should adjust their operations based on seasonal demand. For example, fleet availability and driver hours can be increased in the **spring** and **summer** when demand is highest, while the **winter** months should focus on optimizing operations and reducing costs due to the lower volume of trips.

Conclusion: **Spring** and **summer** present the highest demand for taxi services, while **winter** is the slowest. Adjusting fleet sizes and operational hours to match seasonal patterns can optimize profitability.

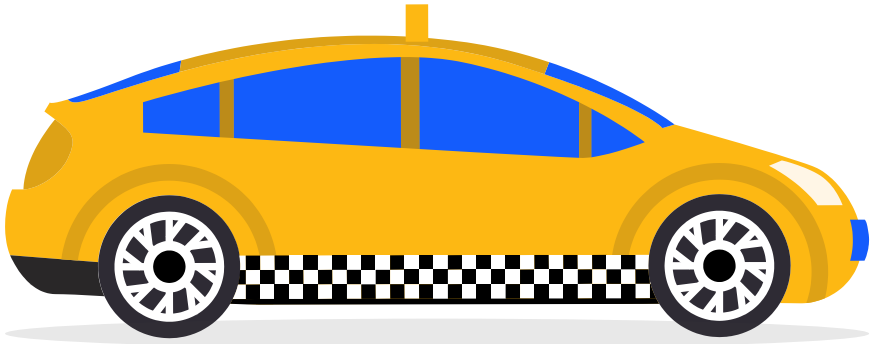


Recommendation and Conclusion

3. Daily Trip Patterns

Recommendation: Taxis should focus on peak days such as **Friday** and **Thursday**, ensuring higher driver availability and more vehicles on the road. Promotions or incentives for drivers working on these peak days could help meet the increased demand.

Conclusion: **Fridays** are the busiest days for taxi trips, suggesting increased transportation needs towards the end of the workweek. Companies should maximize operational efficiency and service availability during these peak periods to capture more business.

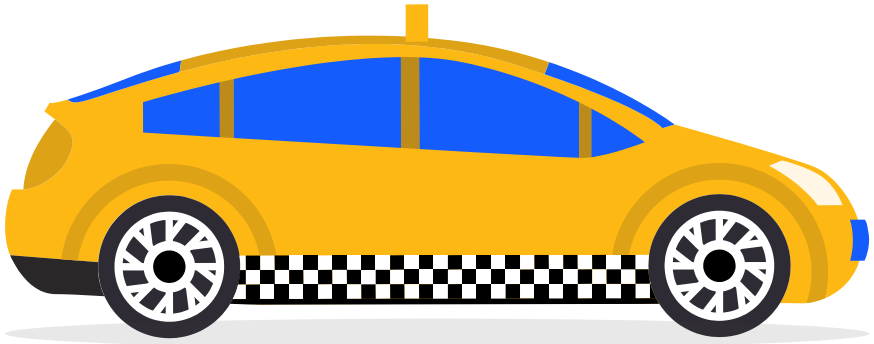


Recommendation and Conclusion

4. Impact of Payment Methods on Revenue

Recommendation: Companies should focus on promoting **credit card** and **mobile** payment methods, as they are associated with higher average fares and tips. Developing partnerships with mobile payment platforms or offering discounts for cashless payments might encourage more usage of these methods.

Conclusion: **Credit card** and **mobile** payments are more profitable compared to **cash** transactions, both in terms of fares and tips. Emphasizing and encouraging the use of these methods can increase revenue per trip.

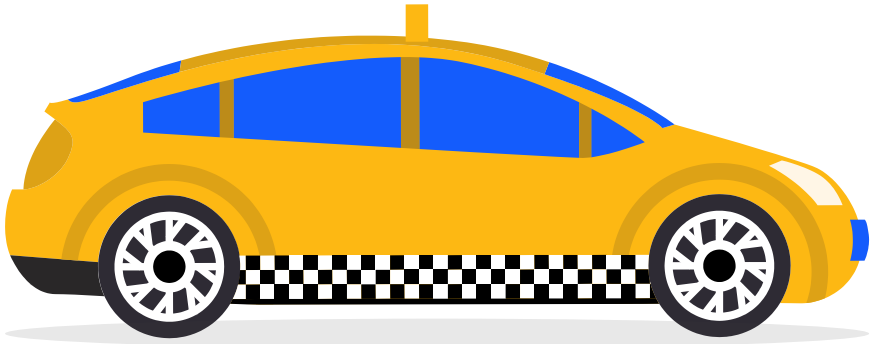


Recommendation and Conclusion

5. Popular Routes

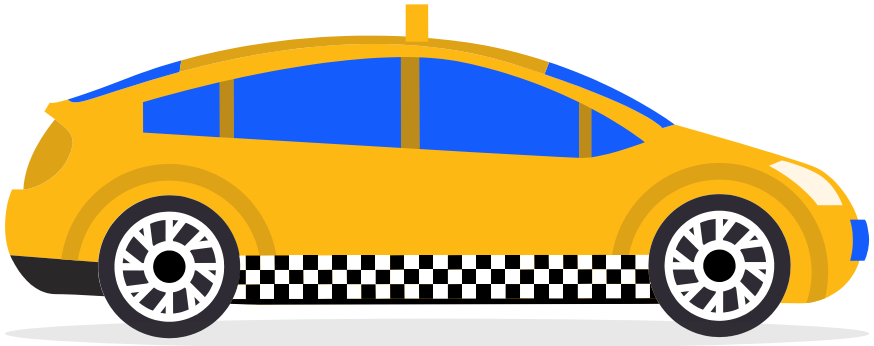
Recommendation: Taxi companies should optimize their service in areas like **community area 8 (Near North Side)** and **32 (Loop)**, which are popular pickup and drop-off locations. Additional marketing, fleet deployment, and targeted services in these areas could boost the number of trips and revenue.

Conclusion: The **Near North Side** and **Loop** are key areas with the most frequent trips. Taxi companies should focus on these high-demand regions to improve efficiency and increase profitability.



General Conclusion

The analysis of the **Chicago Taxi Trips dataset** reveals that a few major companies dominate the market, seasonal and daily patterns significantly affect demand, and modern payment methods result in higher fares and tips. By focusing on optimizing operations based on these insights—such as increasing fleet availability during peak seasons and days, promoting cashless payments, and targeting popular routes—taxi companies can enhance profitability and customer satisfaction in a competitive industry.



THANK YOU