A Comprehensive Analysis of NYC Yellow Taxi Rides

Sagar Dudhat

www.linkedin.com/in/sagar-dudhat-505072213

Table of Contents

1. Introduction

This project centers on a comprehensive analysis of the **NYC Yellow Taxi Rides** dataset using **Power BI** to uncover critical insights into urban mobility, taxi utilization, and revenue generation in New York City. By examining millions of taxi ride records, our objective was to reveal significant patterns and trends in passenger demand, trip frequency, and profitability. The primary goal of this analysis was to enhance operational efficiency and boost revenue by strategically optimizing driver allocation and improving service delivery based on data-driven insights. Through this project, we aimed to propose actionable recommendations for improving the overall performance of NYC's yellow taxi operations.

2. Goals and Expectations

Our primary goal was to optimize driver allocation and increase revenue by 10% through data-driven insights. We anticipated identifying clear patterns in taxi demand, varying across different times of day, days of the week, and seasons. By uncovering these trends, we aimed to improve resource planning, enhance operational efficiency, and ensure drivers are deployed in high-demand areas at the most optimal times, ultimately leading to better service and increased profitability.

3. Information on the Open-Source Data

We selected the **NYC Yellow Taxi Rides** dataset, which contains detailed trip information including pick-up and drop-off locations, trip times, fare amounts, and additional metrics. This data is publicly available from the **New York City Taxi & Limousine Commission (TLC)**. It provides comprehensive insights into urban mobility patterns, revenue, and taxi utilization in NYC.

Dataset Overview

- **Source**: NYC Taxi & Limousine Commission
- **Content**: Pick-up and drop-off times, locations, fare amounts, passenger counts, and trip distances.
- **Why We Chose It**: The dataset offers a rich set of variables that can be used to explore operational efficiency, revenue optimization, and service demand across different geographical and temporal dimensions.

4. Key Questions

The analysis aimed to answer three key questions:

1. What are the top locations where most pick-ups occur?

2. How do taxi trip patterns and demand vary by time of day, day of the week, and season?
3. Which routes generate the highest revenue, and how can driver allocation be optimized for these routes?

5. Our Objectives

These objectives are crucial for ensuring the sustainability and profitability of NYC's yellow taxi services. By understanding demand patterns and route profitability, we can make informed decisions that maximize both driver efficiency and customer satisfaction. This, in turn, helps increase revenue, reduce idle time for drivers, and ensure that services are available when and where they are needed most.

Our key objectives were to:

A. **Analyze taxi demand patterns** to identify high-demand areas and times, enabling better driver deployment.
B. **Investigate the profitability of different routes**, ensuring drivers are assigned to the most lucrative trips.
C. **Propose strategic changes** to improve driver efficiency, reduce downtime, and increase overall revenue for the taxi operations.

These objectives provide the foundation for data-driven decision-making, which is essential for operational success and revenue growth.

6. Key Performance Indicators (KPIs)

Defining and tracking KPIs is critical for understanding the performance and efficiency of taxi operations. These metrics allow us to measure how well the service meets demand, identify potential areas for improvement, and ensure that resources are being used optimally. By focusing on key indicators such as pick-up locations, passenger counts, and revenue per route, we can make data-driven decisions that improve overall operational efficiency and maximize profitability.

The following KPIs were selected to monitor and assess taxi operations:

1. **Pick-ups per Location**: Tracking demand at different locations helps us identify high-demand areas where taxis are most needed.
2. **Total Number of Passengers Served**: This KPI monitors overall taxi utilization, ensuring that enough taxis are available to meet passenger demand.
3. **Pick-ups per Hour/Minute**: Analyzing real-time demand fluctuations allows for more dynamic scheduling, ensuring taxis are available when demand spikes.
4. **Total Revenue per Route**: By determining which routes generate the most revenue, we can optimize driver allocation to prioritize the most profitable trips.

These KPIs are essential for guiding operational decisions and improving the overall effectiveness of taxi services.

7. Data Model in Power BI

In Power BI, our data model plays a vital role in structuring the dataset for accurate analysis and meaningful visualizations. By creating logical relationships between various tables, we can explore the dataset from multiple dimensions and derive deeper insights. A well-constructed data model ensures that the relationships between key entities, such as trip details, locations, payments, and revenue, are clearly defined.

In this data model Power BI is structured around the **Fact Table**, which contains core trip information such as:

a) **Pick-up and drop-off locations** (location IDs).
b) **Fare amount** and additional charges (airport fees, congestion surcharges, etc.).
c) **Passenger count** and payment types.

The **Fact Table** is linked to various dimension tables to enhance our analysis:

1. **PUDO Locations Table**: This table contains information about the pick-up and drop-off locations, linked to the Fact Table by location IDs. It helps us analyze demand based on zones and identify high-demand areas.
2. **Rate Code Table**: This table defines the rate types (e.g., standard, premium) and is linked to the Fact Table, enabling analysis of how fare types of influence revenue.
3. **Vendor Table**: Linked by the vendor ID, this table allows us to examine taxi trips based on service providers, enabling vendor-specific analysis.
4. **Payments Table**: This table stores payment types and custom payment information, providing insight into the modes of payment used by passengers.
5. **Pickups per Zone Table**: This table tracks the number of pick-ups in each zone, allowing us to identify high-traffic areas across the city.
6. **Measures Table**: This includes calculated measures such as **Approximate Total Revenue**, which provides a crucial KPI for assessing overall financial performance.

By connecting these tables through key fields like **locationID**, **ratecodeID**, and **vendorID**, we can dynamically explore the data across different dimensions. This allows us to create powerful visualizations that reveal important trends in taxi operations, including high-demand zones, top revenue-generating routes, and the impact of fare types on revenue.

Here's a revised version with an explanation of how each dashboard component helps achieve the project's goals. **[Refer Appendix A]**

8. Power BI Dashboard

Our Power BI dashboard provides key insights into NYC Yellow Taxi operations, helping us achieve our goal of optimizing driver allocation and increasing revenue by 10%. Each visualization on the dashboard serves a specific purpose in meeting these objectives by offering real-time tracking of demand, revenue, and other performance indicators. Here's how each component contributes to our goal:

1. **Key Metrics**:
   a. **Most Pickups: East Village (62)** – This metric identifies the area with the highest demand for taxi services, allowing us to allocate more drivers to the East Village during peak hours. By ensuring that we meet demand in high-traffic areas, we can reduce idle time for drivers and increase revenue.

   b. **Total Passengers Served: 1517** – Monitoring the total number of passengers served helps track overall taxi utilization. A higher number of passengers indicates better coverage of demand, which directly influences our goal of optimizing operations and increasing profitability.

   c. **Approximate Total Revenue in 2 Hours: $25.19K** – This provides a snapshot of revenue generated over a short period, helping us evaluate whether we are on track to meet our revenue goals. Real-time tracking of revenue allows us to make quick adjustments to driver deployment or pricing strategies if necessary.

   d. **Average Fare: $16.95** – Knowing the average fare per ride helps us assess whether the current fare structure is maximizing profitability. If fares are too low or high in comparison to demand, adjustments can be made to balance affordability for passengers with profitability for the service.

2. **Line Graph – Pickups Every Minute from 12 AM**:
   This graph visualizes the fluctuation of taxi demand throughout the early hours. By identifying the minutes when demand is highest, we can schedule more drivers during those peak periods, ensuring optimal coverage. This will reduce wait times for passengers and help drivers serve more customers, directly contributing to increased revenue.
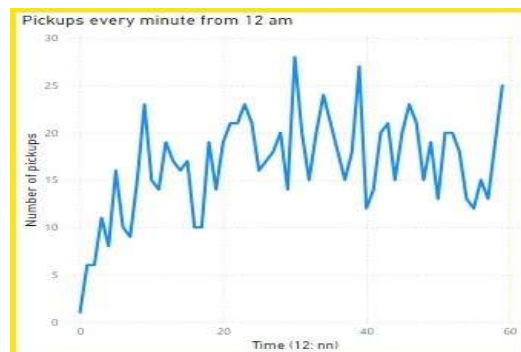

3. **Gauge – Average Tip per Ride (3.25)**:
   This gauge shows the average tip amount given by passengers. A higher tip average can be an indicator of customer satisfaction, which is linked to better service. By maintaining or increasing this number, we can ensure that passengers have a positive experience, encouraging repeat usage and contributing to overall revenue growth.

4. **Pie Chart – Passenger Count by Payment Type**:
The pie chart shows that most passengers prefer to pay by credit card (77.46%), followed by cash (20.76%). Understanding these payment preferences allows us to streamline the payment process and offer more convenient options for passengers. Making payment methods easier and faster improves the overall customer experience, which in turn can help increase customer loyalty and ride frequency.

Sum of passenger_count by PaymentType
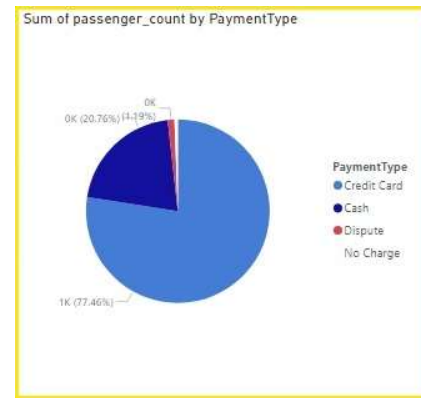
How These Components Help Achieve Our Goal

Each of these metrics and visualizations is critical in driving informed decision-making:

1. **Optimizing Driver Allocation**: By identifying high-demand locations (East Village) and peak times (from the line graph), we can efficiently deploy drivers where and when they are needed most. This reduces wasted time and maximizes the number of passengers served, directly boosting revenue.
2. **Revenue Growth**: Real-time monitoring of total revenue, fare averages, and high-revenue areas ensures that we stay on track to achieve our 10% revenue increase. If revenue is falling short, we can immediately adjust strategies, such as reallocating drivers or changing pricing.
3. **Improving Customer Satisfaction**: The average tip per ride and the insights from payment preferences allow us to gauge customer satisfaction and make improvements where necessary. A satisfied customer base is more likely to use the service again, contributing to long-term revenue growth.

Overall, this dashboard empowers us to track and measure key performance indicators that align directly with our goal of improving operational efficiency and increasing revenue. By leveraging these insights, we can make data-driven decisions to optimize taxi services in New York City.
[Refer Appendix B]

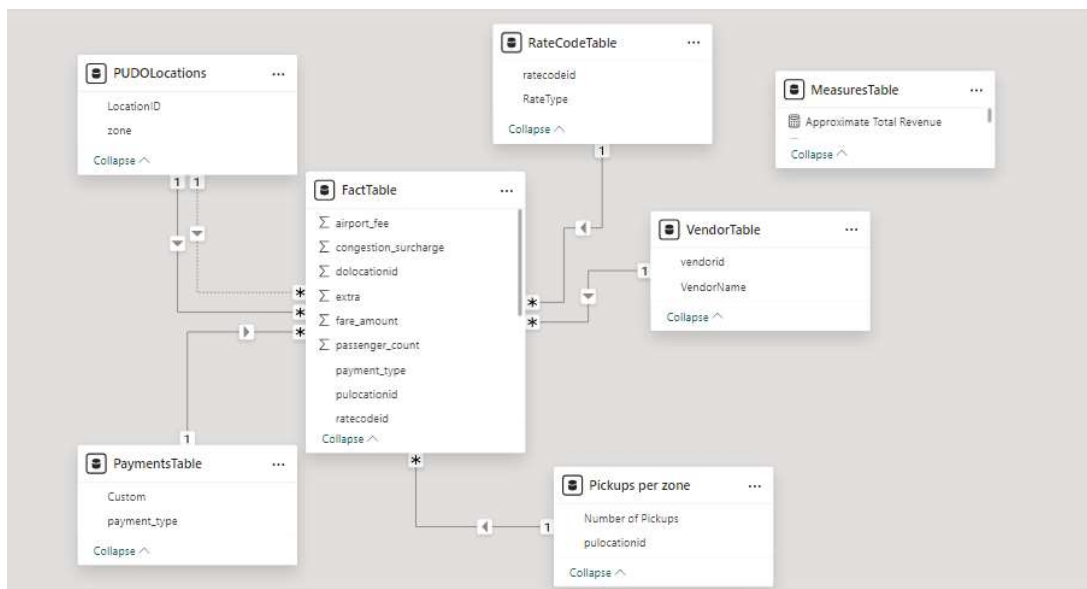# 9. Power BI Analysis Benefits/Results

The Power BI analysis revealed several valuable insights:

1. **High-Demand Areas**: Midtown Manhattan had the highest taxi pick-up activity, especially during peak rush hours.
2. **Time-Based Demand**: Taxi demand peaks in the mornings and evenings on weekdays, with lower demand during the night and early afternoon.
3. **Profitable Routes**: Routes from central Manhattan to the city's airports (JFK, LaGuardia) proved to be the most profitable.

10. Conclusion

Using Power BI to analyze the NYC Yellow Taxi dataset, we uncovered key insights to improve operational efficiency and boost profitability. The analysis identified high-demand areas like East Village and peak times for optimal driver allocation, reducing idle time and enhancing service. We also highlighted the most profitable routes and tracked metrics such as average fare and passenger count, enabling real-time adjustments. Additionally, insights into customer payment preferences and tipping patterns can improve the passenger experience. By leveraging these data-driven strategies, we expect to achieve a 10% revenue increase while enhancing overall taxi operations and service quality.

Appendix A: Data model in PowerBI

Appendix B: PowerBI Dashboard