RIDING THROUGH NYC: INSIGHTS FROM 2024

1. Introduction

New York City's Yellow Taxi system generates millions of trips and an enormous amount of data each year. By analyzing the 2024 trip records, this project explores travel patterns, fare trends, and rider behavior across the city's boroughs. The findings support data-driven decision-making for transportation planners, taxi operators, and other mobility stakeholders.

2. Problem Statement

How can we use New York City's millions of yearly yellow-taxi trip records to uncover meaningful patterns in demand, pickup hotspots, rider behavior, and pricing, so that planners and transportation authorities can better understand city movement and make data-driven decisions around mobility and congestion?

3. Goal

The goal of this project is to:

- Analyze how yellow-taxi demand changes by hour, day, and month to understand when New York is busiest.
- Identify which pickup zones, vendors, and payment methods influence trip patterns, fares, and tips.
- Examine how external factors, like rush hours, weekends, and airport travel impact overall ride behavior and revenue.
- Build a visual presentation summarizing these insights.
- Provide actionable recommendations for stakeholders who rely on NYC taxi mobility insights.

4. Target Audience

This analysis is relevant for:

- City planners and transportation authorities who need insights to improve mobility and manage congestion.
- Government and policy teams who rely on travel patterns to guide infrastructure and transportation decisions.
- Taxi operators, mobility companies, and urban analysts who depend on understanding rider behavior, pricing, and demand.

5. Dataset

Source: NYC Taxi & Limousine Commission (TLC) – Yellow Taxi Trip Records for 2024.

Format: Parquet files (monthly).

6. Data Cleaning

Data was cleaned and prepared by:

- Combined monthly parquet files.
- Dropped nulls from congestion surcharge (4,091,232) Airport fee (4,091,232), RatecodeID (4,091,232) store and fwd flag (4,091,232) and passenger count (4,091,232)
- Converted datetime fields.
- Added derived columns such as pickup hour and trip duration.
- Filtered unrealistic trips and outliers.
- Creating aggregated tables for rates, vendors, airports, and weekdays.

7. Analysis and Findings

- **Busiest Month for Taxi Trips:** *May and October are peak travel months, likely due to tourism and event season.*
- **Daily Ride Demand in NYC:** *Across 2024, Thursdays showed the highest travel volume, reaching 5.8 million trips.*
- **Hourly Ride Patterns Across the City:** *Throughout 2024, 6 PM was the busiest hour, with 2.65 million recorded trips.*
- **Top Pick-up Locations:** *The data was analyzed and the top 10 pickup locations were visualized. Upper East Side (North and South) emerged as major hotspots due to their high residential density, while Times Square showed heavy pickup activity because of constant tourist movement, shopping areas, and overall traffic flow.*
- **Top Drop-off locations:** *The top drop-off locations identified in the analysis were Upper West Side South, Upper East Side South, and Times Square. Similar to the pickup trends, Times Square appears again due to its heavy foot traffic, shopping areas, and overall city activity.*
- **Taxi Usage Across NYC Boroughs:** *Manhattan generates the highest taxi revenue at $676.5 million, followed by Queens at $225.8 million and Brooklyn at $17.2 million. This indicates that Manhattan relies on taxi services far more than the other NYC boroughs.*
- **Neighborhood Payment Preferences:** *Manhattan relies heavily on credit card payments, with around 85% of trips paid digitally and only 15% in cash. In contrast,*

Staten Island shows the opposite pattern, with just 36% credit card usage and 63.6% paid in cash. This contrast suggests that Manhattan riders are more likely to be commuters, tourists, or people moving through commercial areas where digital payments are the norm, while Staten Island riders may rely more on local trips where cash payments are still common.

- **Tipping Behavior by Borough:** *Manhattan has the highest tipping rate at 19%, and the 'Unknown' zone in the dataset also shows a similar average tip of around 19%. In comparison, Queens has a lower tipping rate of about 15%. This suggests that areas with higher tourist activity or higher-income riders tend to tip more consistently.*
- **Tipping rate changes by hour:** *Tipping peaks at 7–8 PM, reaching about 20%, likely reflecting commuters returning home or people heading out for evening activities. Tipping rates fluctuate across the 24-hour cycle—dropping in the early morning hours and gradually rising throughout the day before hitting their highest point in the early evening.*

8. Recommendations

- Optimize traffic flow in high-demand zones during peak hours.
- Allocate transportation resources based on borough-level taxi usage.
- Adjust fleet distribution to match hourly and monthly demand patterns.
- Enhance forecasting models by combining taxi data with weather, events, and flight information.

9. Limitations

- The dataset includes only yellow taxis, not Uber, Lyft, or green cabs so it doesn't represent all NYC transportation.
- Pickup and drop-off zones are broad areas, not precise GPS locations, limiting location accuracy.
- The dataset doesn't include external factors like weather, events, or holidays that strongly affect taxi demand.
- Payment and tipping behavior can be misleading.

10. References

- NYC TLC Trip Record Data
- TLC Data Dictionary
- GA Data Analytics Bootcamp Study Materials
- Python libraries used: Pandas, NumPy, Matplotlib, Polars, PyArrow, Python datetime