NYC Yellow Taxi 2023 Data Analysis Report

By, Shyju S R

Date : 20-05-2025

Introduction

In this case study, I perform exploratory data analysis (EDA) on yellow taxi rides in New York City for the year 2023. The goal is to uncover insights that can help optimize taxi operations, improve service efficiency, maximize revenue, and enhance passenger experience.

Data Overview

The dataset includes yellow taxi trip records from NYC Taxi for the year 2023. The data includes fields such as pick-up and drop-off dates/times, locations, trip distances, itemized fares, rate types, payment types, and passenger counts. Because the data volume is enormous, I took only 5% of the hourly data for each day as sampling for the study.

Data Cleaning

The data was cleaned by removing duplicates, dropping unnecessary columns, fixing negative values in monetary fields, handling missing values using median/mode, and removing invalid datetime and outlier records.

Exploratory Data Analysis Findings

Temporal Trends

Hourly, daily, and monthly pickup patterns were analyzed to identify peak hours and high-demand days. The analysis revealed that taxi demand is highest during morning and evening rush hours, and weekends show increased demand in entertainment zones.

Revenue Analysis

Monthly and quarterly revenue trends were analyzed. The analysis showed that revenue peaks during holiday seasons and weekends. A pie chart was created to show the revenue distribution by quarter.

Correlation Analysis

A strong positive correlation ($r \approx 0.95$) was found between trip distance and fare amount, indicating that longer trips generate higher fares.

Outlier Handling

Outliers were identified and handled by removing trips with distances greater than 50 miles, fares greater than $200, tips greater than $50, and invalid payment types and vendor IDs.

Recommendations

Taxi Demand Observation and Allocation Strategies

1. Observer the Hourly Demand in the report

- **Track taxi demand by hour** throughout the day.
- Allocate taxis dynamically based on hourly demand patterns to ensure optimal availability.

2. Zone and Time-Based Allocation

- Monitor demand across **different zones** (high vs. low demand) and **time periods** (peak vs. off-peak).
- Adjust taxi distribution accordingly to balance supply and demand.

3. Observe the Weekday vs. Weekend Demand Patterns

- Analyze how demand shifts between **weekdays and weekends**.
- Pay special attention to **zones with entertainment venues** (e.g., clubs, theatres) that may see increased demand during weekends.

4. Observe and Prioritize High-Revenue Zones

- Identify and prioritize **high-revenue areas** such as **airports**.
- Ensure consistent taxi availability in these zones to maximize revenue.

5. Long-Distance Trip Planning

- Observe the **number of taxis required for long-distance trips**.
- Allocate a suitable number of vehicles to handle these trips without affecting local availability.

Dynamic Taxi Pricing Strategies

1. Hour-Based Pricing Strategy

- Implement **different pricing tiers** for:
    - **Peak Hours**
    - **Off-Peak Hours**
    - **Night Time**
- Helps manage demand and incentivize travel during less busy hours.

2. Zone-Based Pricing Strategy

- Apply **variable pricing** based on zone demand:
    - **High-Demand Zones**: Higher pricing
    - **Low-Demand Zones**: Standard or discounted pricing
- Encourages better taxi distribution and service availability.

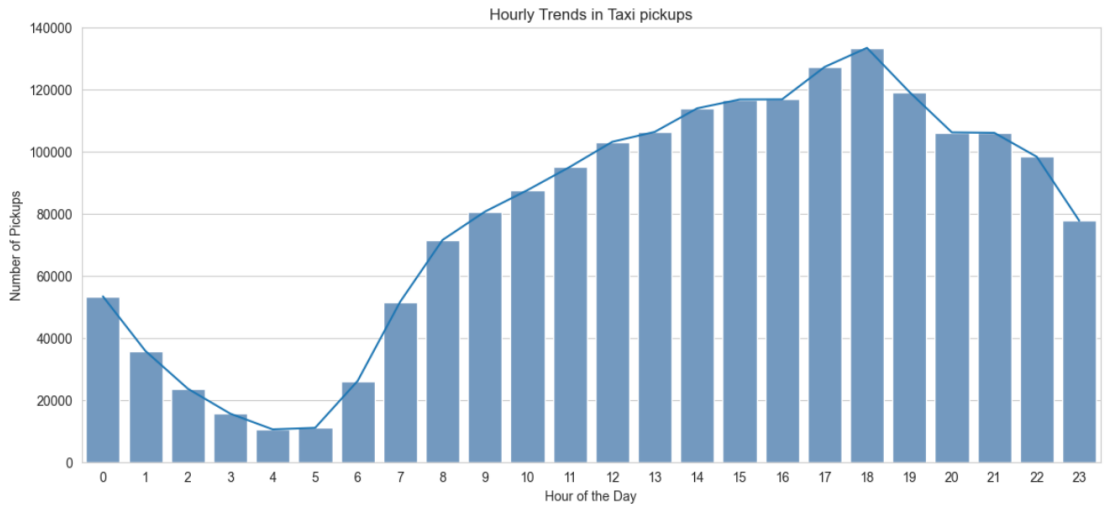
3. Weekend/Entertainment Zone Pricing Strategy

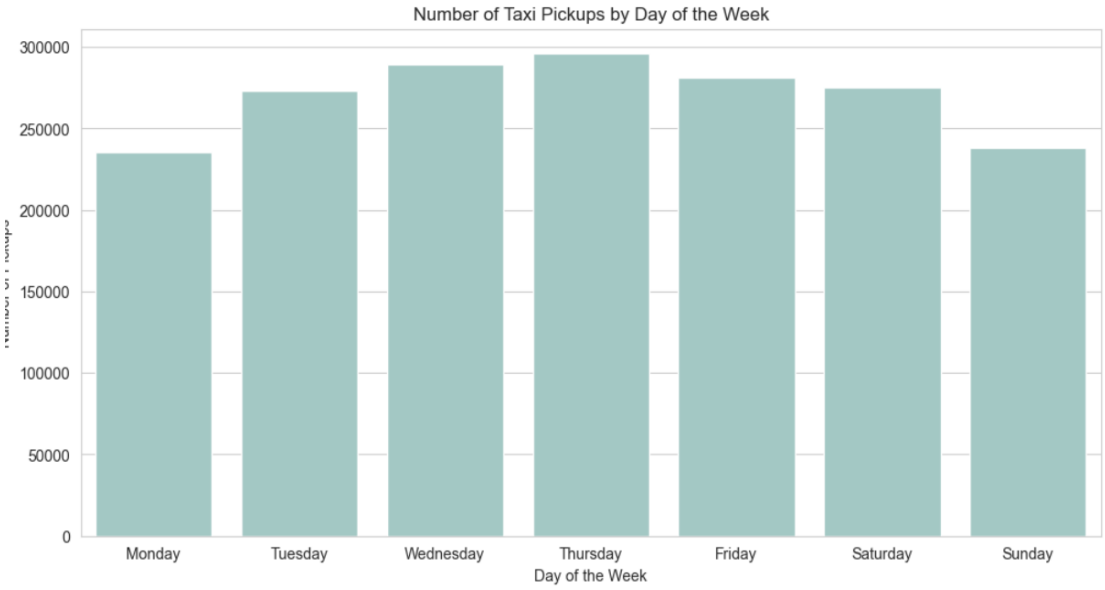
- Introduce **special pricing** for zones with:
    - **Theatres**
    - **Clubs**
    - **Other entertainment venues**
- Adjust pricing especially during **weekends and late evenings** to reflect increased demand.

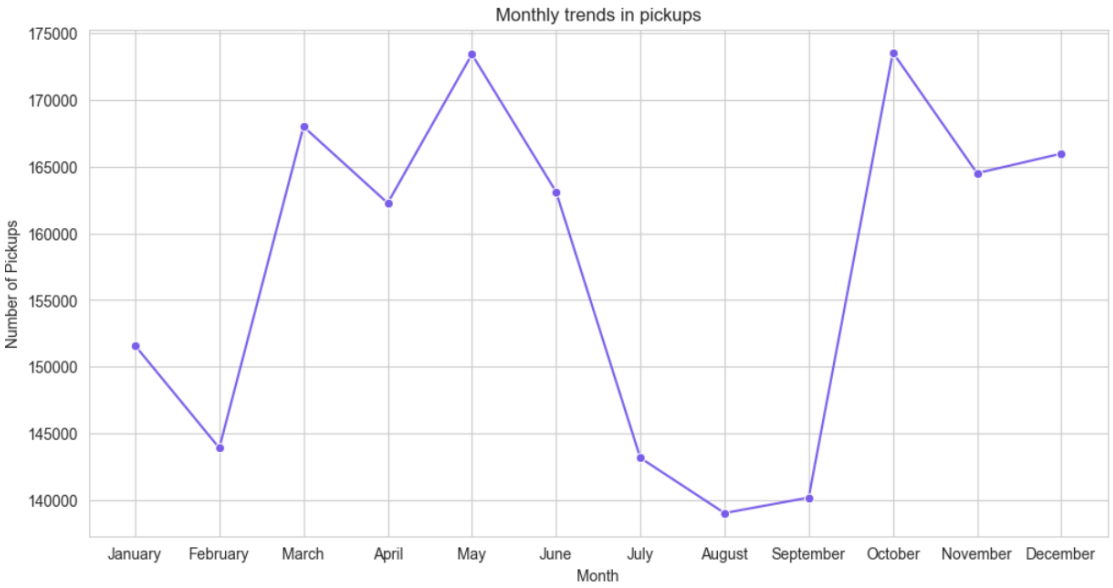
4. Trip Duration-Based Pricing Strategy

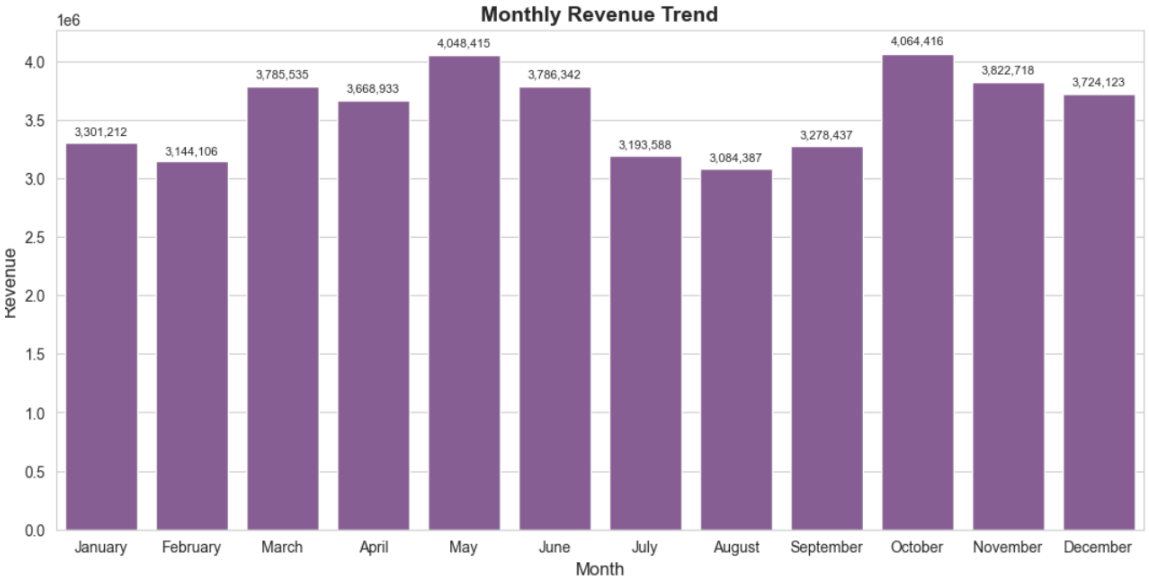
- Define pricing based on **trip length**:
    - **Short Trips**
    - **Medium Trips**
    - **Long Trips**
- Tailor pricing to reflect operational costs and optimize revenue.

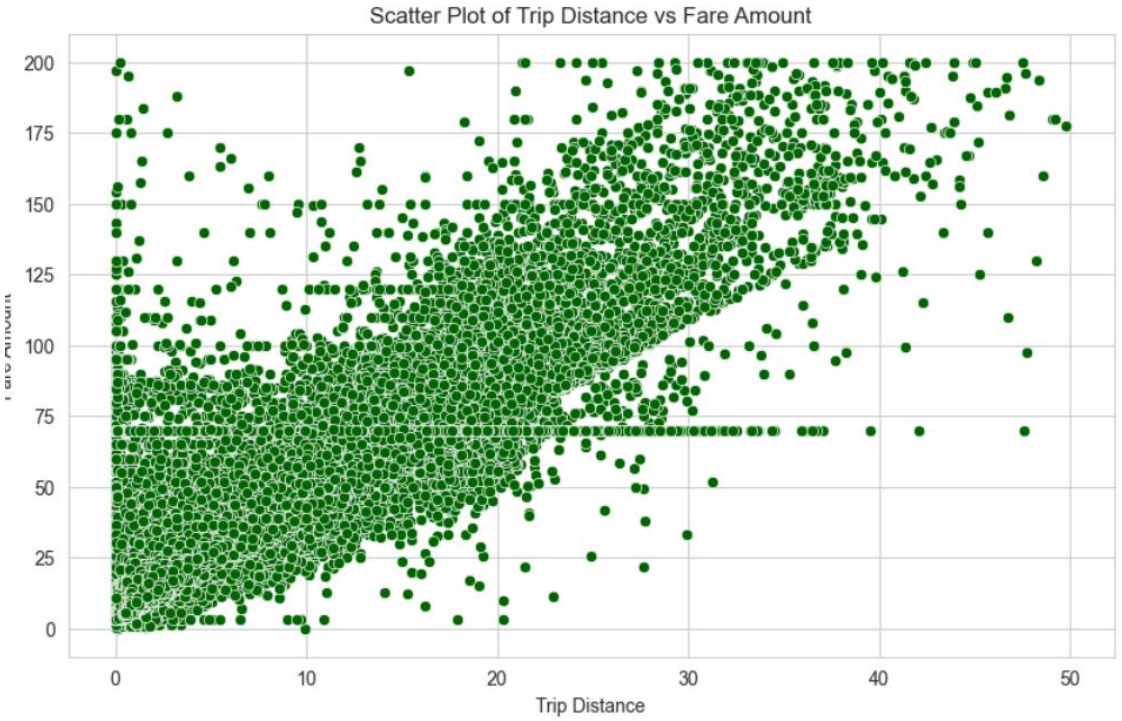
Note: The exact **percentage of surcharges** should be determined by the transportation provider based on **domain knowledge**, local demand patterns, and operational considerations.

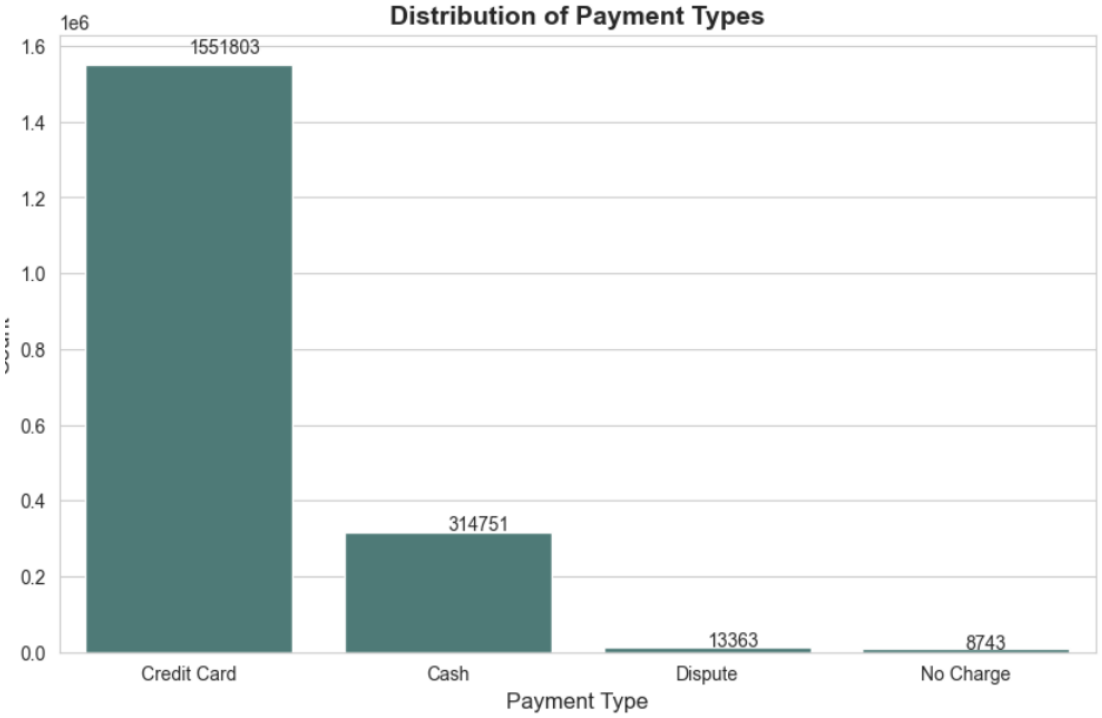


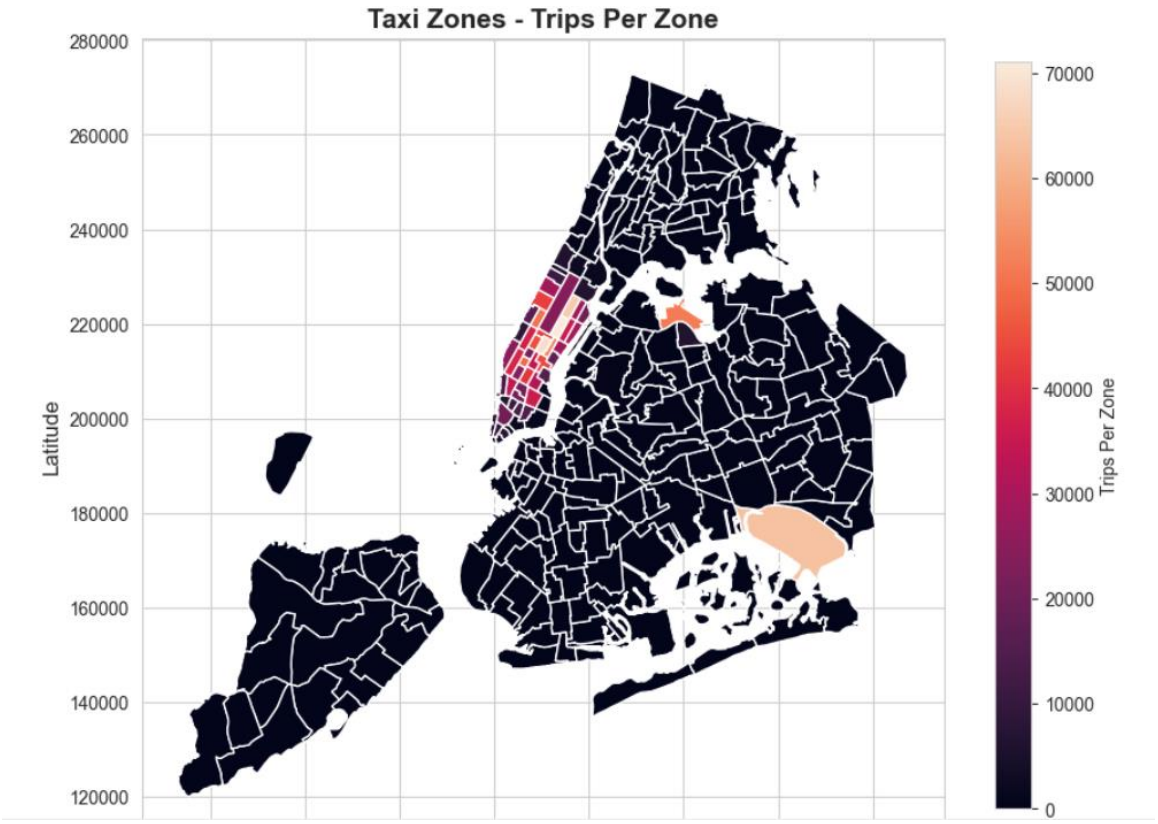

Number of Taxi Pickups by Day of the Week
300000
250000
200000
150000
100000
50000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week

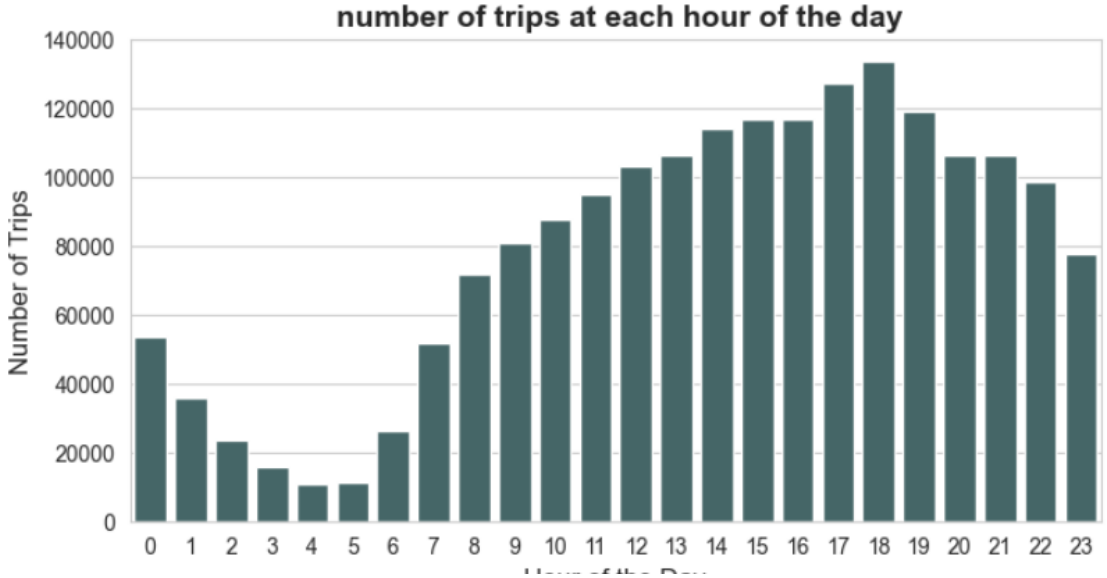

Monthly trends in pickups
175000
170000
165000
160000
155000
150000
145000
140000
Number of Pickups
January
February
March
April
May
June
July
August
September
October
November
December
Month

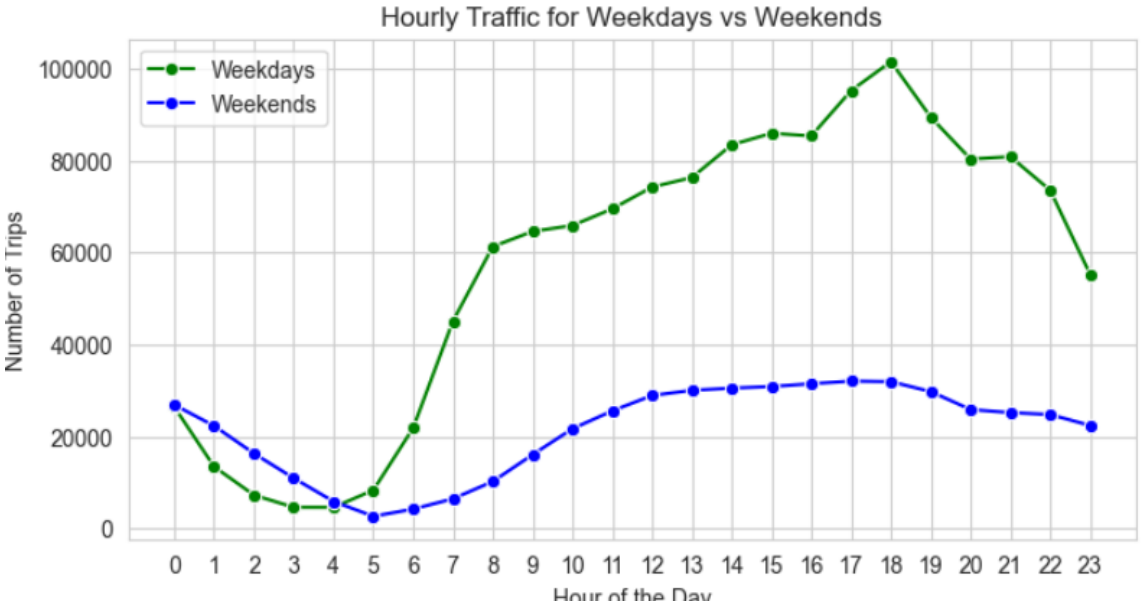

Monthly Revenue Trend
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Revenue
3,301,212
3,144,106
3,785,535
3,668,933
4,048,415
3,786,342
3,193,588
3,084,387
3,278,437
4,064,416
3,822,718
3,724,123
January
February
March
April
May
June
July
August
September
October
November
December
Month

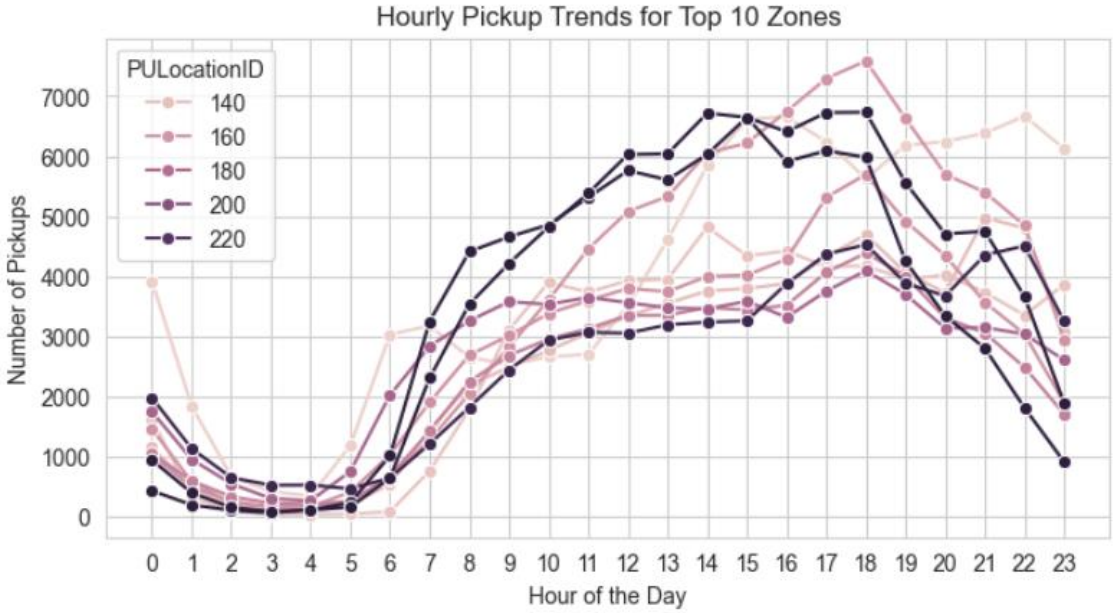

Scatter Plot of Trip Distance vs Fare Amount
200
175
150
125
100
75
50
25
0
0
10
20
30
40
50
Trip Distance

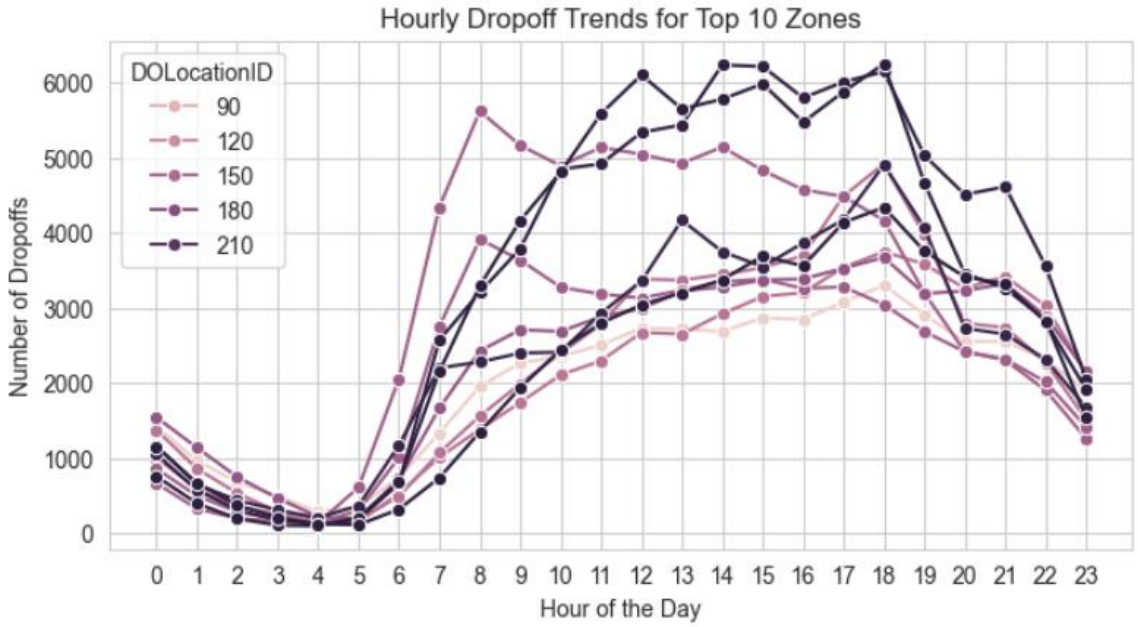

Distribution of Payment Types
1e6
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1551803
314751
13363
8743
Credit Card
Cash
Dispute
No Charge
Payment Type

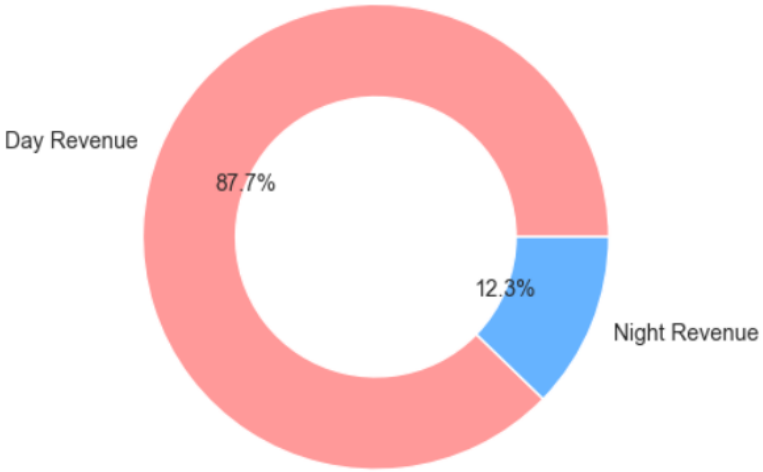

Taxi Zones - Trips Per Zone
Latitude
280000
260000
240000
220000
200000
180000
160000
140000
120000
Trips Per Zone
70000
60000
50000
40000
30000
20000
10000
0

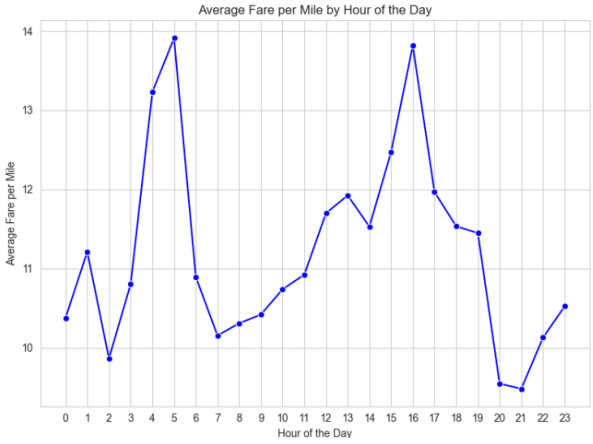

number of trips at each hour of the day
Number of Trips
140000
120000
100000
80000
60000
40000
20000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

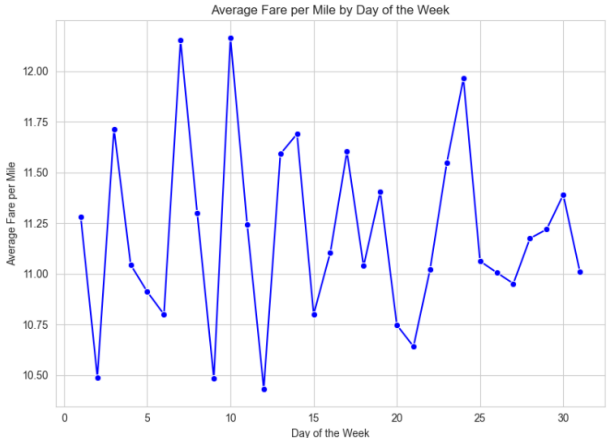

Hourly Traffic for Weekdays vs Weekends
Weekdays
Weekends
Number of Trips
100000
80000
60000
40000
20000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of the Day

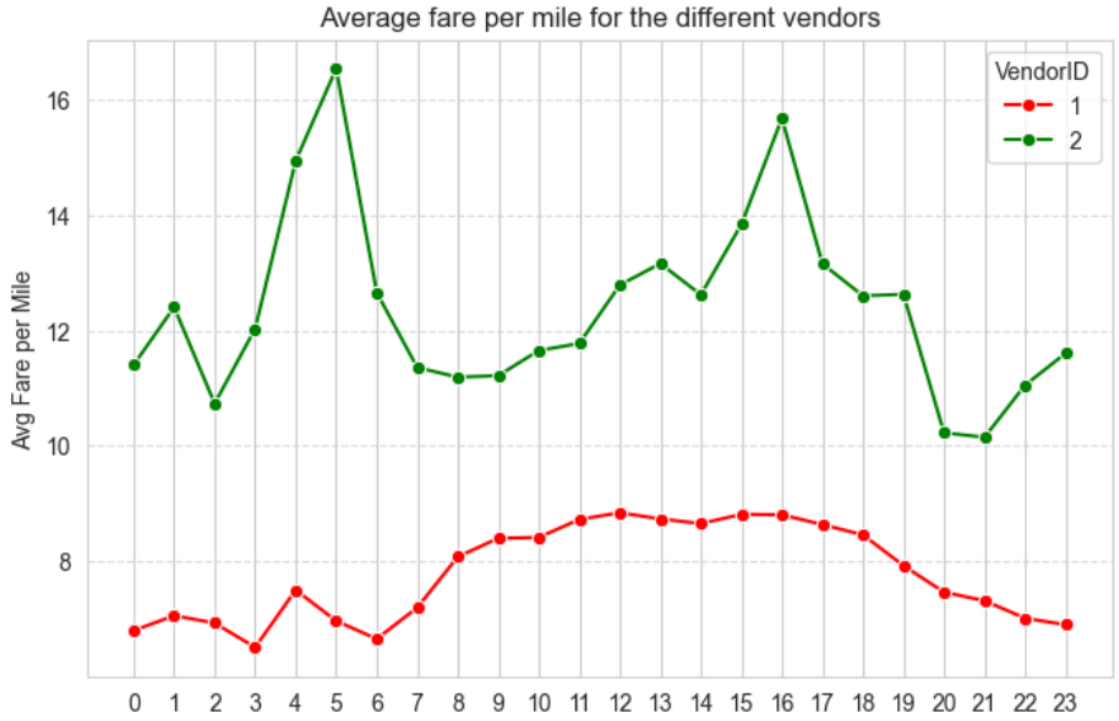

Hourly Pickup Trends for Top 10 Zones
PULocationID
140
160
180
200
220
Number of Pickups
7000
6000
5000
4000
3000
2000
1000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of the Day

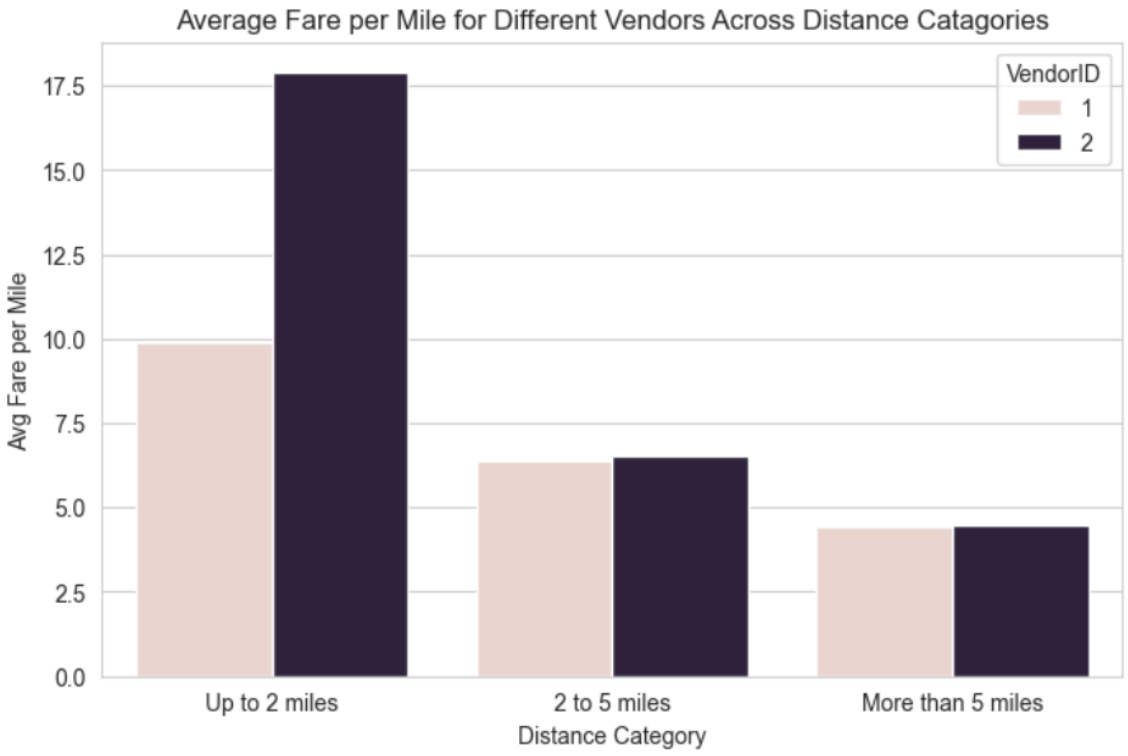

Hourly Dropoff Trends for Top 10 Zones
DOLocationID
90
120
150
180
210
Number of Dropoffs
Hour of the Day

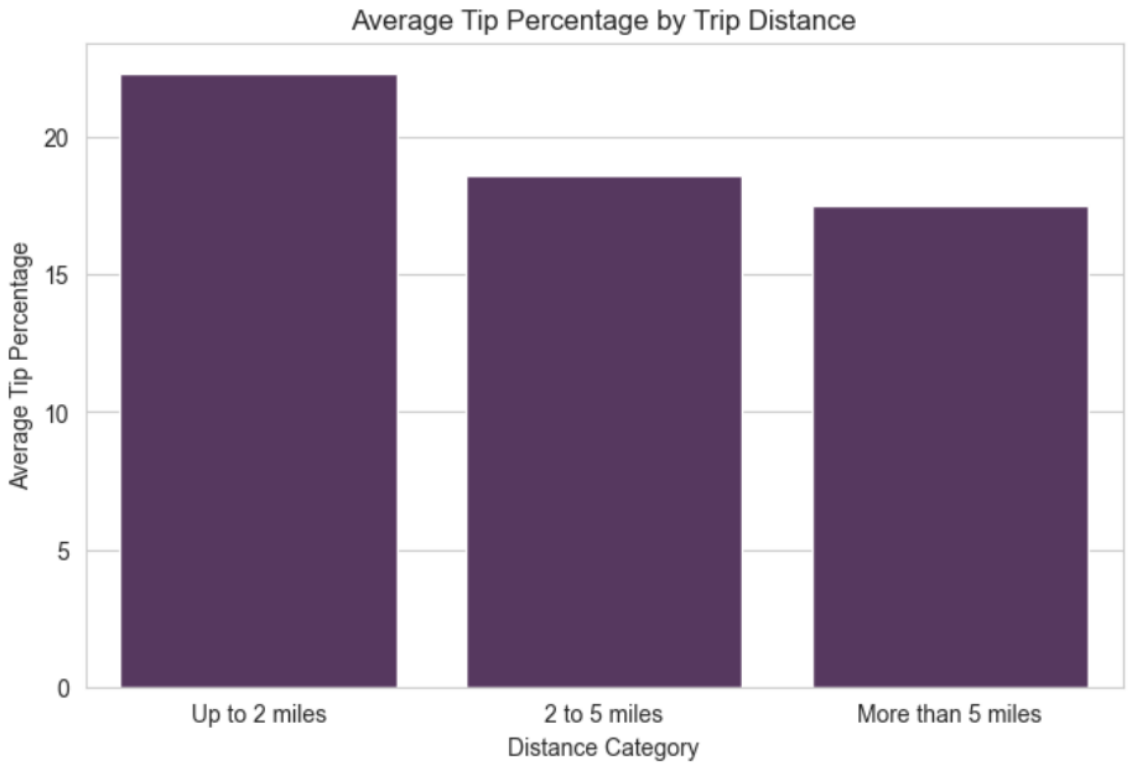

Day Revenue
87.7%
12.3%
Night Revenue

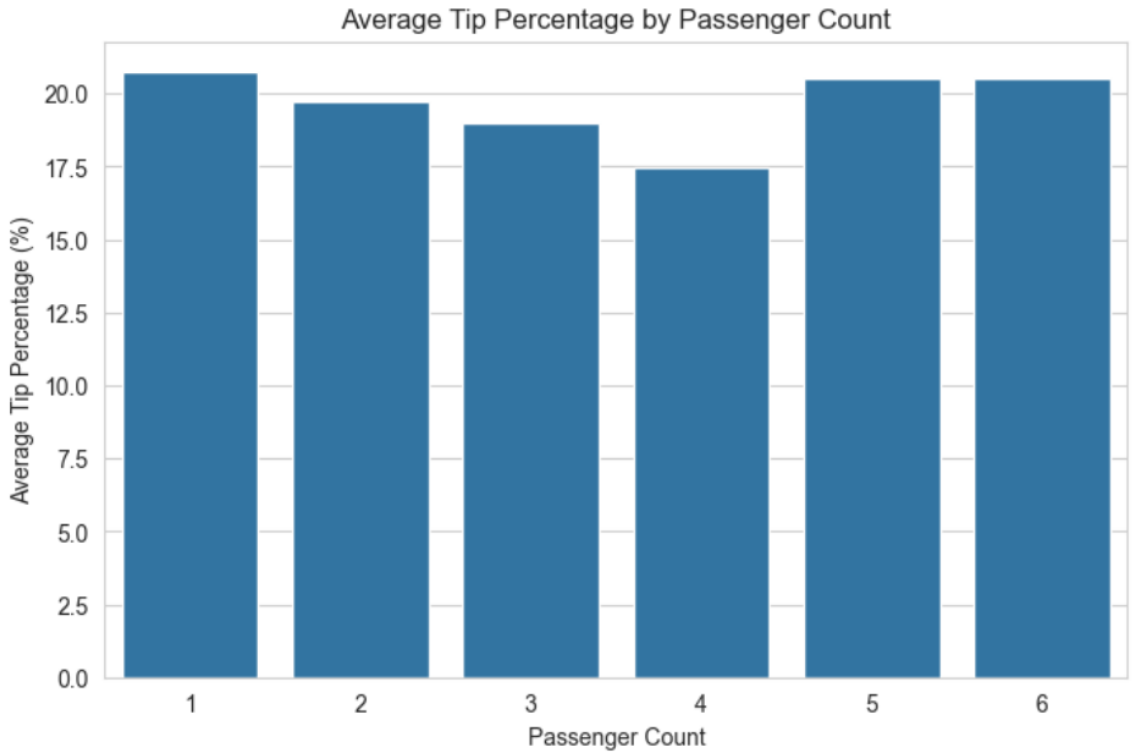

Average Fare per Mile by Hour of the Day
Average Fare per Mile
Hour of the Day

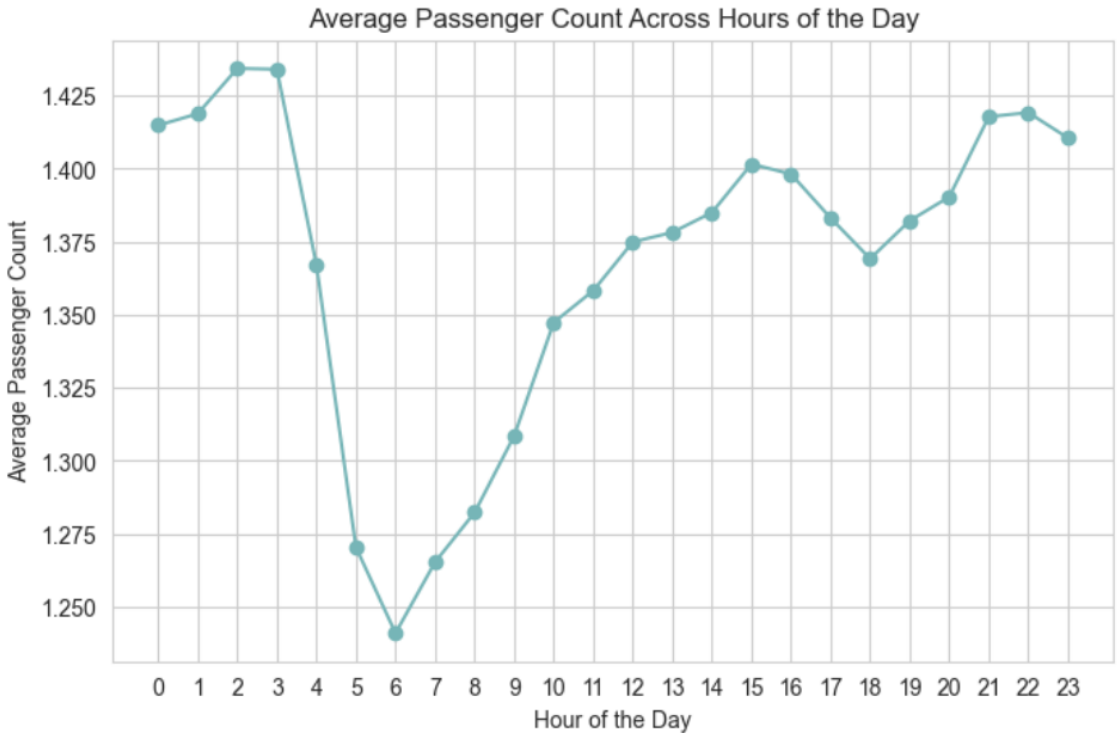

Average Fare per Mile by Day of the Week
Average Fare per Mile
Day of the Week

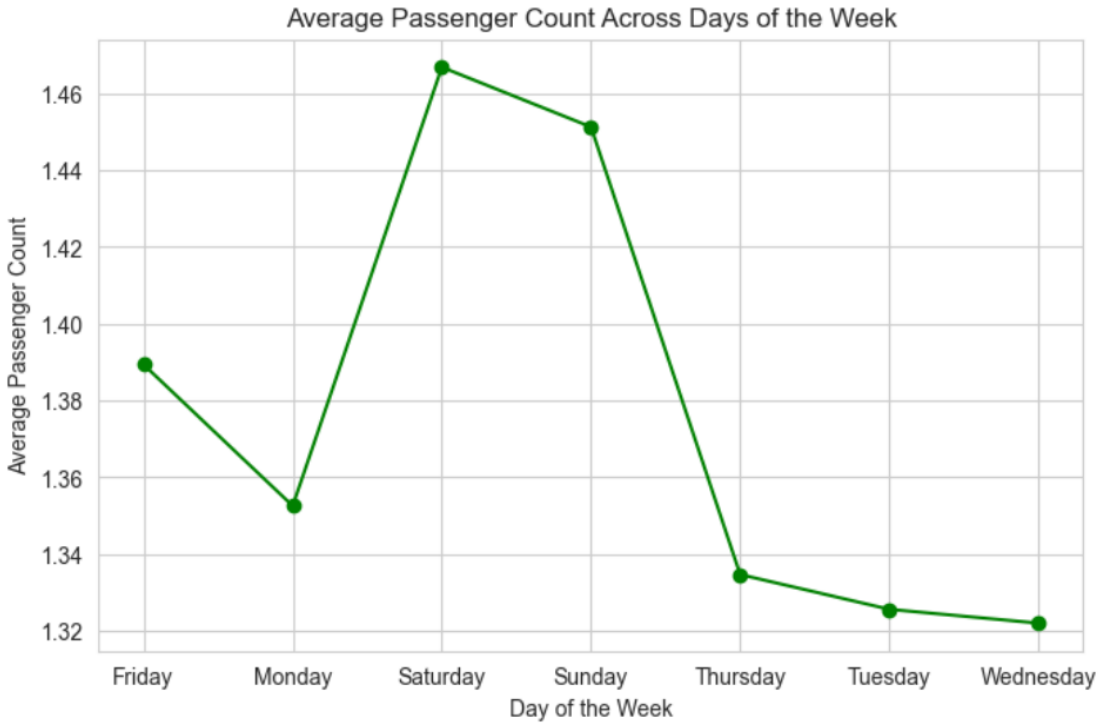

Average fare per mile for the different vendors
VendorID
1
2
Avg Fare per Mile
16
14
12
10
8
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23


Average Fare per Mile for Different Vendors Across Distance Catagories
VendorID
1
2
Avg Fare per Mile
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Up to 2 miles
2 to 5 miles
More than 5 miles
Distance Category


Average Tip Percentage by Trip Distance
Average Tip Percentage
0
5
10
15
20
Up to 2 miles
2 to 5 miles
More than 5 miles
Distance Category

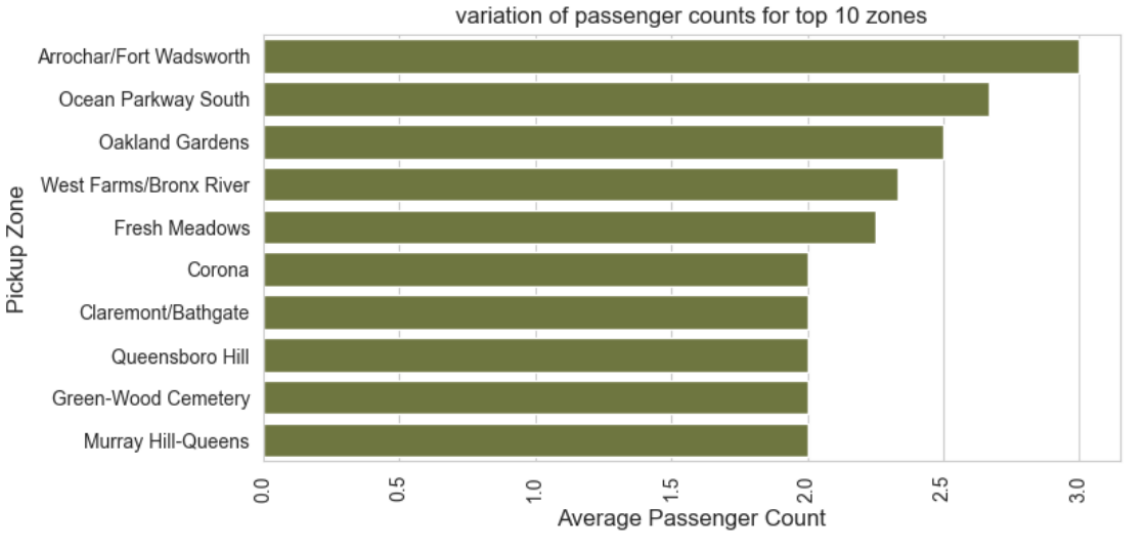

Average Tip Percentage by Passenger Count
Average Tip Percentage (%)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
1
2
3
4
5
6
Passenger Count

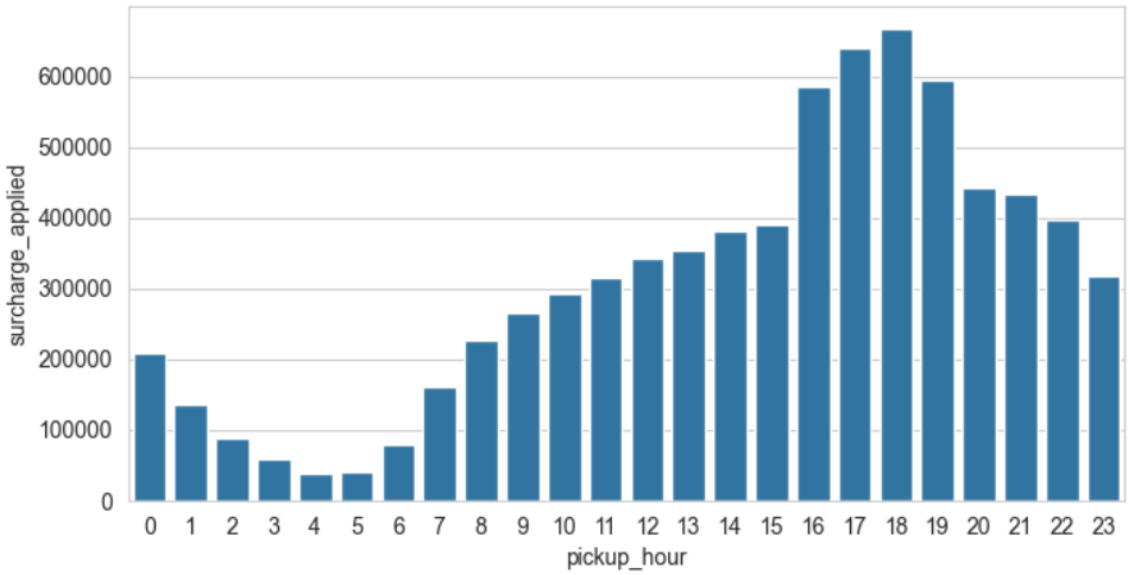




The following images provide some insights based on the analysis.

The detailed report is available in EDA_Assg_NYC_Taxi_Starter_SHYJU_S_R.pdf.


variation of passenger counts for top 10 zones
Arrochar/Fort Wadsworth
Ocean Parkway South
Oakland Gardens
West Farms/Bronx River
Fresh Meadows
Corona
Claremont/Bathgate
Queensboro Hill
Green-Wood Cemetery
Murray Hill-Queens
Pickup Zone
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Average Passenger Count


600000
500000
400000
300000
200000
100000
0
surcharge_applied
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
pickup_hour