Uber: Maximising Revenue per Driver Hour (RDH) through Demand Optimisation

Bangalore, known for its **heavy traffic congestion, fluctuating ride demand, and diverse commuting patterns**, presents significant challenges for Uber, including **inefficient driver deployment, high idle times, and unpredictable trip profitability**. Uber aims to optimize the demand-supply balance, minimize idle time, and enhance earnings per trip. Addressing these challenges will improve driver earnings while ensuring better operational efficiency.

Problem Objective

The goal of this project is to provide suggestions for optimizing **Revenue per Driver Hour (RDH)**, improving **fleet efficiency**, and enhancing **overall revenue strategies**. The analysis will focus on:

- Demand trends across Bangalore to understand ride patterns.
- Driver earnings per trip/hour and factors influencing RDH.
- Fleet efficiency, focusing on idle time and driver deployment.
- Trends and insights that can contribute to improving RDH.
- Assessing factors influencing RDH beyond predefined ones to uncover further insights.

Methodology:

The analysis follows these structured steps:

- **Define Metrics and Hypotheses**: Identify metrics like Total Amount, Distance, Trip Duration and Payment Type. Formulate hypotheses, e.g., "Vendors who operate during weekdays generate higher revenue compared to weekends."
- **Data Cleaning**: Handle missing, duplicate, or inconsistent data. Standardize formats and remove outliers.
- **Exploratory Data Analysis (EDA)-** Cleaning, preprocessing, and identifying patterns.
- **Feature Engineering:** Creating new variables such as Trip Duration, Idle Time, Pickup Location, Drop-off Location, Payment Type Category.
- **Summarize Insights**: Identify insights that can help optimize **RDH, fleet efficiency, and overall revenue strategies**.

Dataset:

This dataset contains information on 6.5 million rows of Uber driver data, providing detailed insights into trips, including trip distance, pickup and drop-off timestamps, pickup location IDs, and more. The objective is to leverage this data to optimize **Revenue per Driver Hour (RDH)**, enhance **fleet efficiency**, and improve overall **revenue strategies**.

The dataset includes the following columns: VendorID, tpep_pickup_datetime, tpep_dropoff_datetime, passenger_count, trip_distance, PULocationID, DOLocationID, payment_type, fare_amount, extra, gst, tip_amount, tolls_amount, improvement_surcharge, total_amount, congestion_surcharge.

Data Cleaning and Preparation:

1. **Handling Missing Values :**

The Uber dataset contains missing values in the passenger_count and payment_type columns. Missing values in passenger_count were imputed using the mean, as it showed minimal correlation with other columns. Meanwhile, payment_type is a categorical ID column representing the payment method, where missing values were replaced with 0 and mapped to the "Unknown" payment category.

2. **Handling Outlier Values:**

The Uber dataset contains extremely high or unrealistic outlier values, such as a fare amount of 20K for a single trip. Most of these outliers column values were capped at the 99th percentile, and any negative values were replaced with zero.

3. **Converting Data Types:**

tpep_pickup_datetime and tpep_dropoff_datetime columns were converted from object to datetime format, while payment_type and passenger_count were converted from float to int for better data consistency and understanding.

4. **Adding New Columns:**

PULocationID and DOLocationID were mapped to their respective locations, creating new columns pickup_location and dropoff_location. Similarly, payment_type IDs were mapped to relevant payment_type_categories, resulting in a new column named payment_type_category.

Exploratory Data Analysis (EDA)

Key Metrics:

- Total Amount
- Trip Distance
- Trip Duration
- Passenger Count
- Payment Type

Divided the analysis into four key segments to identify patterns and formulate actionable hypotheses.

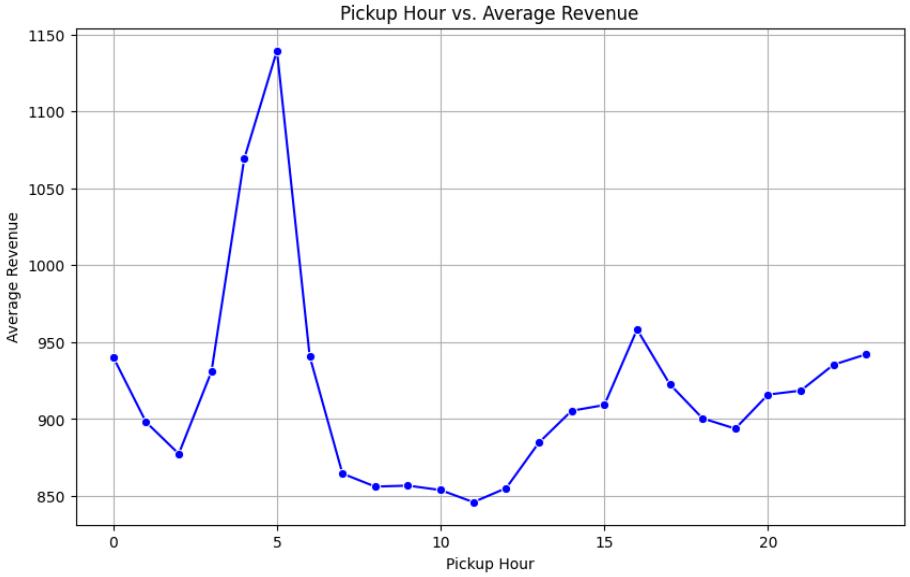
1. **Time Based Analysis**
2. **Location Based Analysis**
3. **Operational Analysis**
4. **Revenue & Payment Analysis**

Time Based Analysis: Analyses hourly/daily/monthly revenue peaks, cancellations, and demand patterns and formulating relevant hypothesis

Columns: - tpep_pickup_datetime, total_amount, tip_amount, congestion_surcharge

Hypothesis

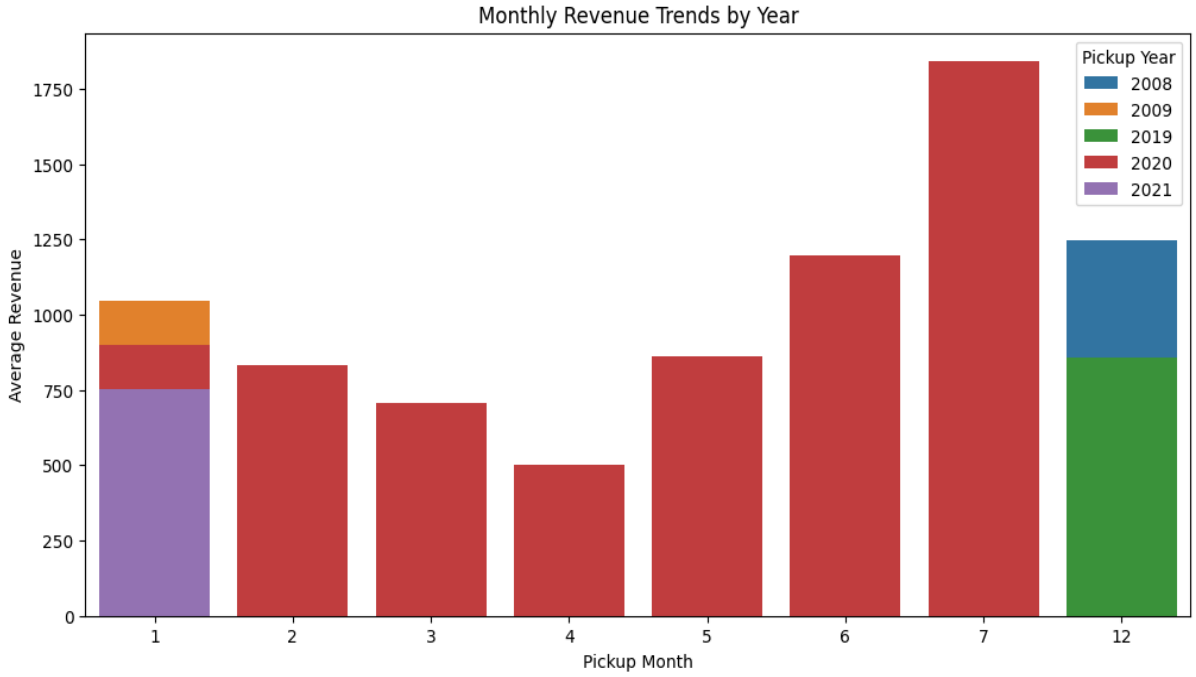
H1:- Vendors who operate during early morning (5-6 AM) and evening (4-6 PM) tend to generate higher total revenue compared to other times of the day.



Insights:

1. **Peak Revenue** occurs around the 5th hour, indicating high demand and increased earnings during early morning hours.
2. **Lowest Revenue** is observed between 7 AM - 10 AM, followed by a gradual increase during afternoon and evening hours.

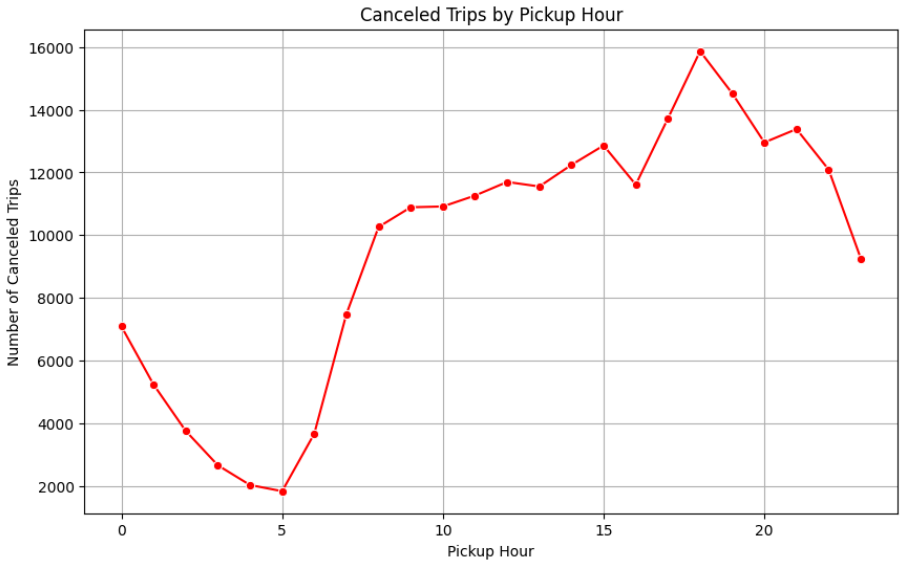
Hypothesis 2:- Vendors generate the highest revenue in July, with notable increases in January and December, while revenue remains lower from April to June. Revenue trends improved in 2020 and 2021, showing higher variability across months.



Inshights

- **Revenue trends improved in 2020 and 2021, with higher variability across months, suggesting changing demand patterns in recent years.**
- Revenue spikes in **July**, indicating higher demand during this month, possibly due to seasonal events or increased travel.
- **January and December** show relatively higher revenue, suggesting festive or holiday season boosts.

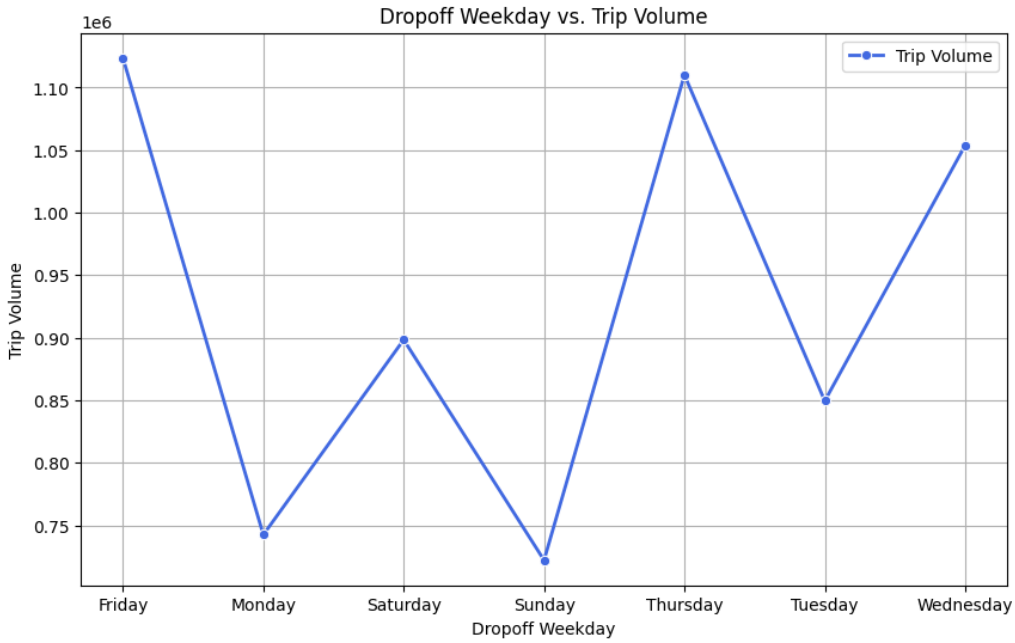
Hypothesis3:- Cancellation rates peak during late afternoon and evening hours (5 PM – 7 PM), suggesting that high traffic and increased ride demand during these periods contribute to higher cancellation probabilities.



Insights

- Cancelled trips are lowest between 4 AM - 6 AM, while they peak around 5 PM - 7 PM, indicating higher cancellations during evening hours.

Hypothesis5:- Trip volume peaks on Monday and Friday, with notable increases on Sundays. Midweek and early weekdays show lower demand compared to the beginning and end of the week.



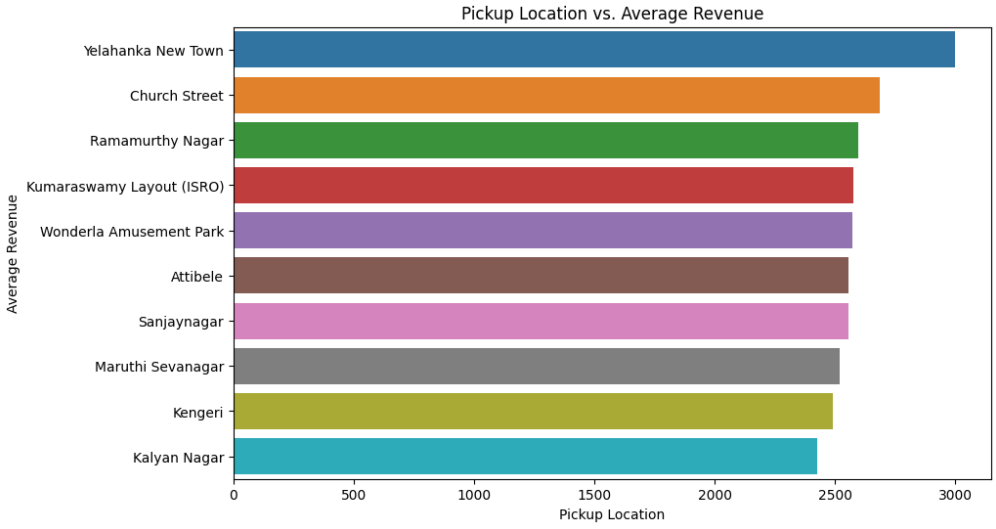
Insights

- Trip volume peaks on Friday and Thursday, while it drops significantly on Sunday and Monday, indicating higher demand towards the end of the workweeks

Location Based Analysis : Analyses pickup location and drop-off location, cancellations, and demand patterns and formulating relevant hypothesis

Hypothesis:

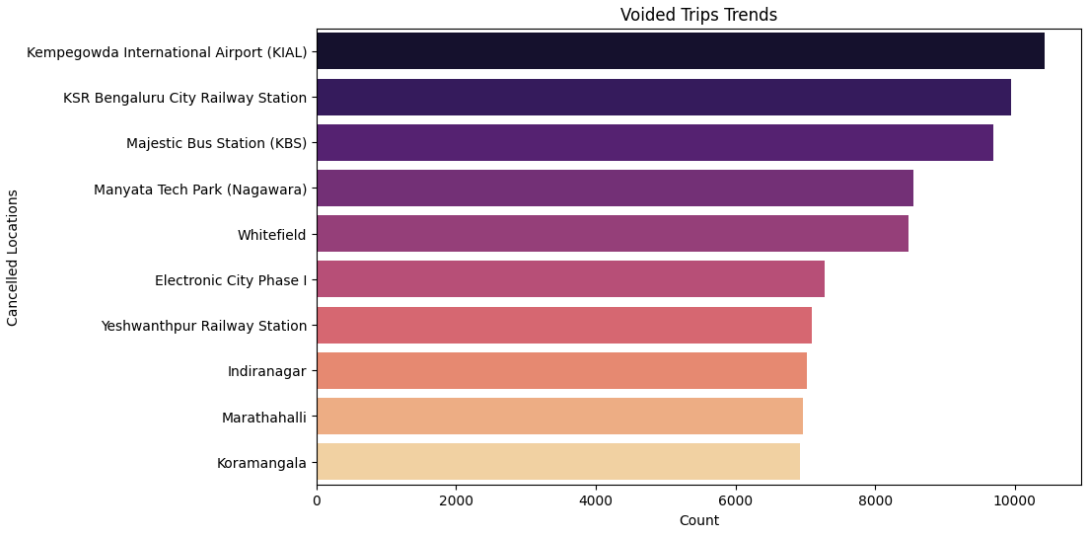
H1:- Yelahanka New Town has the highest average revenue for vendors



Insights

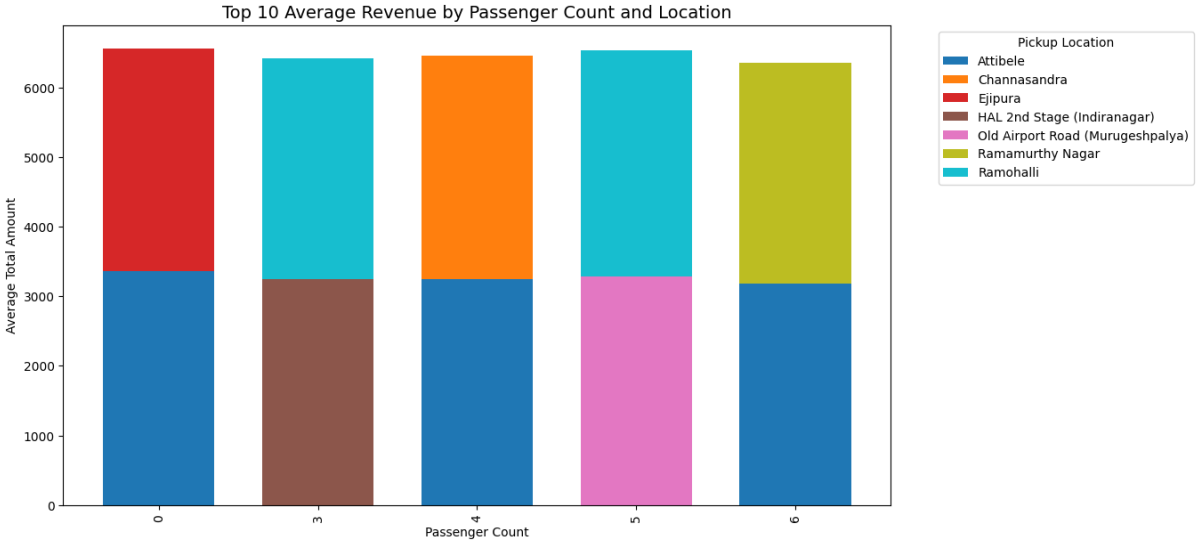
- Yelahanka New Town generates the highest average revenue, followed by Church Street and Ramamurthy Nagar, indicating that these areas contribute significantly to revenue growth.

Hypothesis2:- Kempegowda International Airport (KIAL) is the most frequently canceled pickup location by vendors.



Insights:- Kempegowda International Airport (KIAL) has the highest number of voided trips, followed by KSR Bengaluru City Railway Station and Majestic Bus Station, indicating frequent cancellations at major transportation hubs.

Hypothesis3: Trips with 3-4 passengers, mainly from locations like Ramamurthy Nagar and Ejipura, tend to generate higher average revenue.



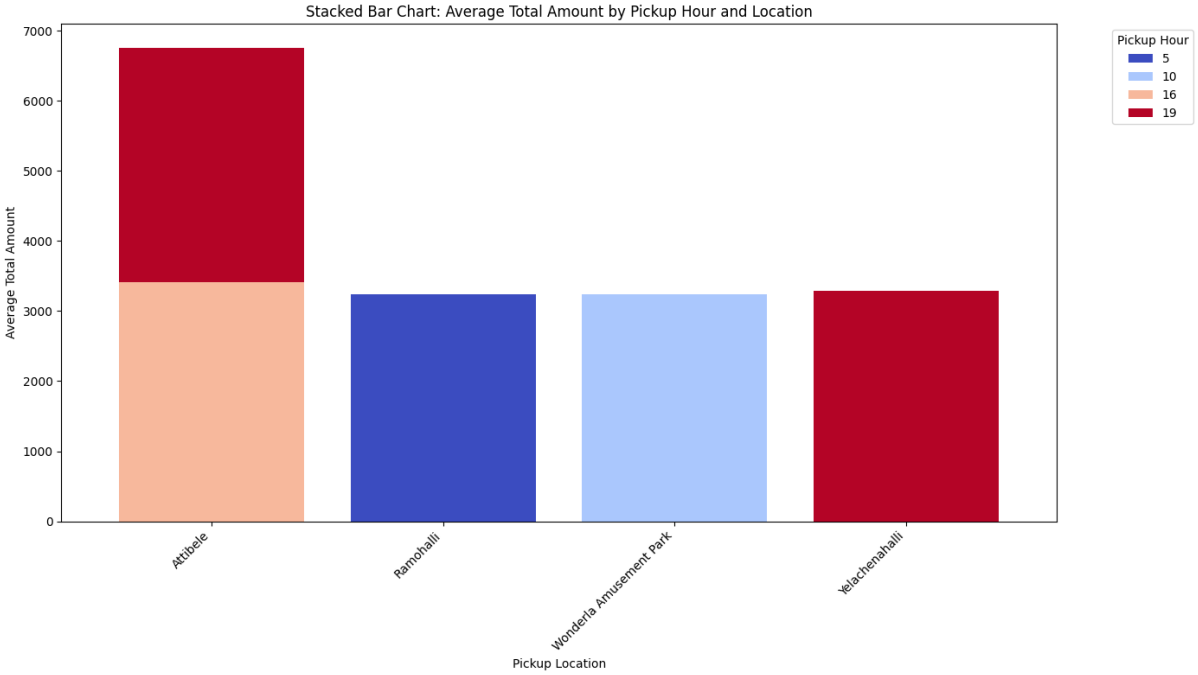
Insights :

Kempegowda International Airport (KIAL) experiences the highest number of voided trips, indicating frequent cancellations due to long wait times or driver dissatisfaction.

Operational Analysis:- Analyzing pickup and drop-off locations, passenger count, trip duration, and cancellations and formulating revelant hypothesis

Hypothesis

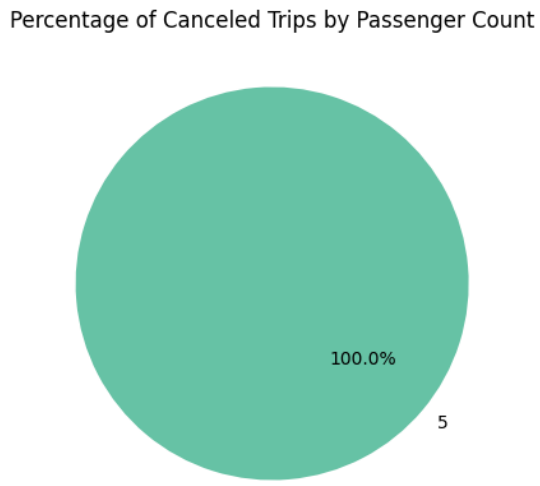
Hypothesis1:- Pickup locations such as Attibele generate the highest average revenue between 10 AM and 8 PM, while Ramohalli, Wonderla Amusement Park, and Yelachenahalli show consistent revenue across multiple pickup hours.



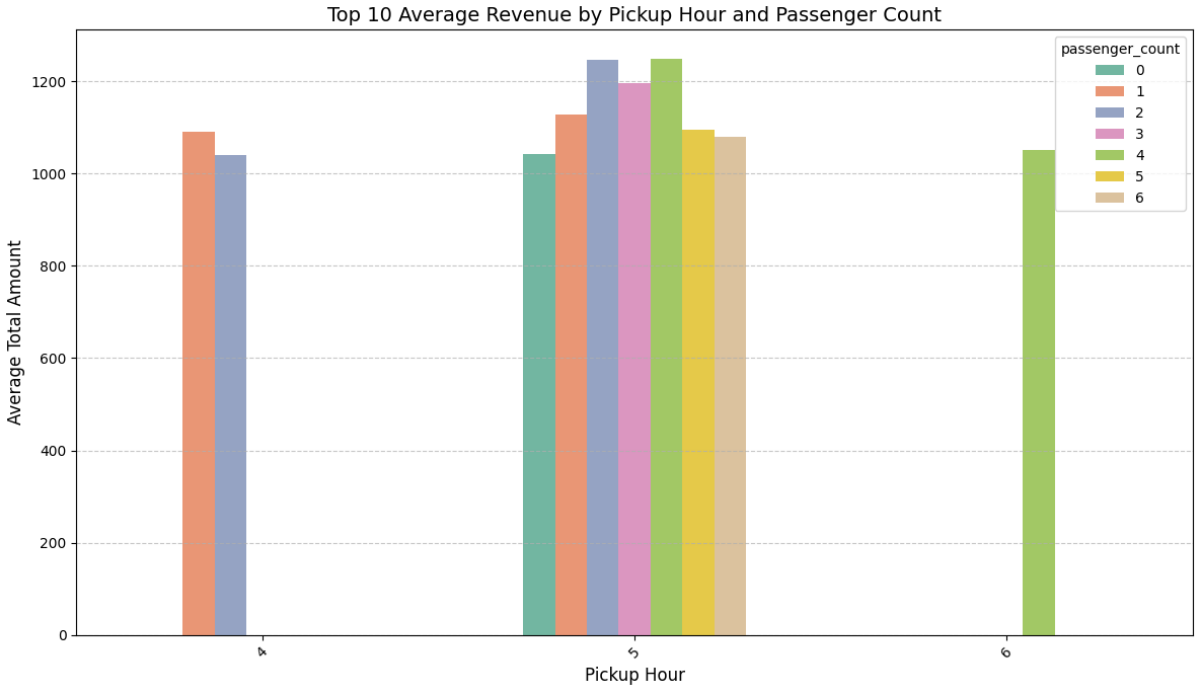
Insights:

Attibele generates the highest revenue between 10 AM and 8 PM, while Ramohalli and Wonderla show consistent revenue across multiple hours, with Yelachenahalli experiencing a surge around 7 PM, indicating that peak revenue hours vary across locations.

Hypothesis2: Trips with 5 passengers account for 100% of the canceled trips, suggesting that group rides with maximum capacity are more likely to be canceled.



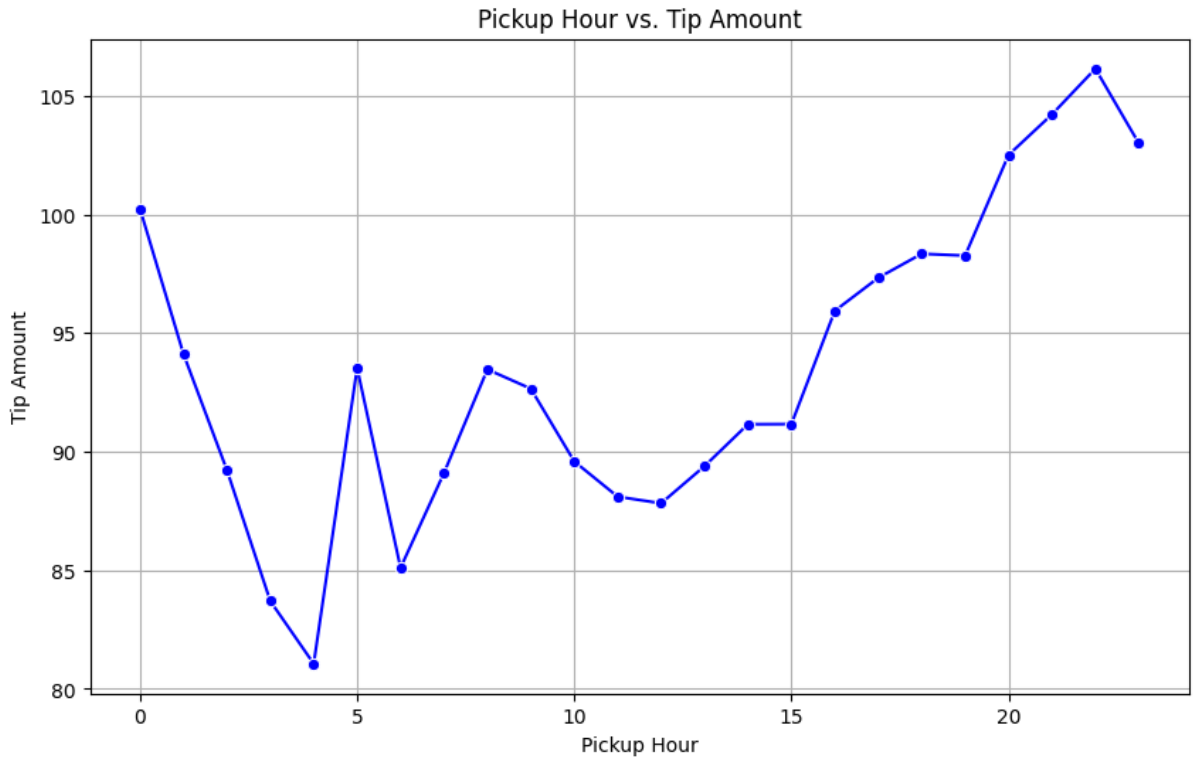
Hypothesis3: Higher passenger counts during peak pickup hours (5-6 AM) generate the highest average total revenue, with passenger counts between 3 to 5 contributing the most. Beyond this range, additional passengers have minimal impact on increasing total fare.



Revenue and Payment Based Analaysis:- Analysising revenue, payment patterns, and tipping behavior across different time periods and payment types, covering columns like total_amount, payment_type, tip_amount, pickup_datetime, and congestion_surcharge.

Hypothesis

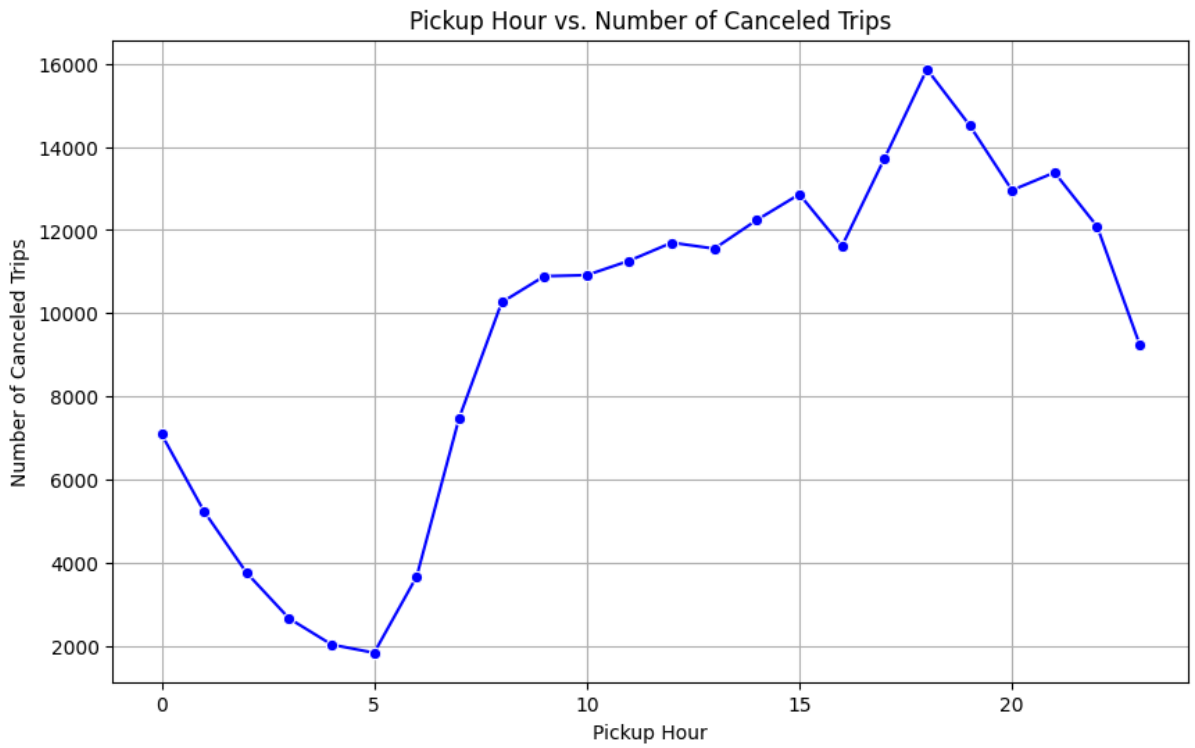
Hypothesis1:- Tip amounts are generally higher during evening hours (16:00 - 22:00), indicating passengers tip more during peak evening periods.



Insights:

1. Tip amounts are highest around midnight (12:00 AM) and peak again after 9:00 PM.
2. Tips tend to drop significantly between 3:00 AM – 5:00 AM, indicating lower tipping behavior during early morning hours.
3. A gradual increase in tip amounts is observed between 4:00 PM – 10:00 PM, with consistent higher tips during peak evening periods.

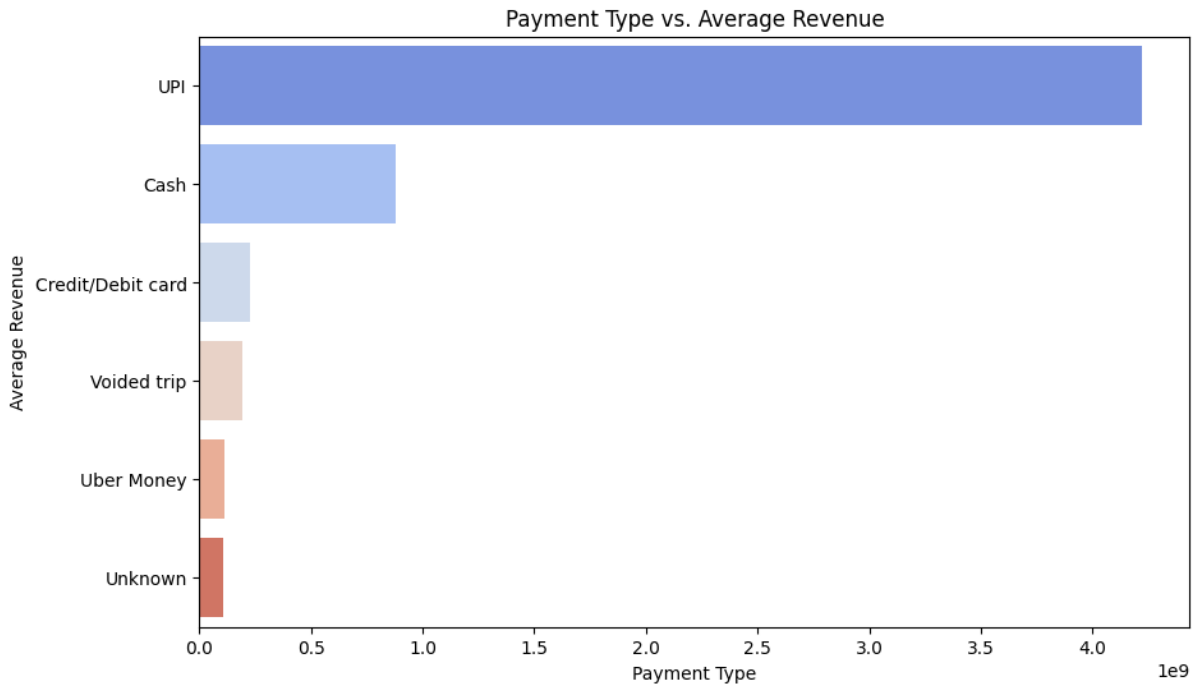
Hypothesis2:- The number of canceled trips peaks between 18:00 and 20:00, suggesting that high demand during these hours leads to increased trip cancellations.



Insights:

1. Tip amounts are highest around midnight (12:00 AM) and peak again after 9:00 PM.
2. Tips tend to drop significantly between 3:00 AM – 5:00 AM, indicating lower tipping behavior during early morning hours.
3. A gradual increase in tip amounts is observed between 4:00 PM – 10:00 PM, with consistent higher tips during peak evening periods.

Hypothesis3:- : Rides paid in UPI yield the highest average revenue compared to all other payment types, highlighting a preference for digital payments.



Business Insights and Recommendations

Revenue per Driver Hour (RDH)

Insights

Time-Based Revenue Patterns

- Drivers earn the most during early morning (5-6 AM) and evening (4-6 PM) due to commuter demand.
- Weekdays outperform weekends, with **July, January, and December** being the top revenue-generating months.

Location-Based Revenue Disparities

- Yelahanka New Town has the highest average revenue per driver, while Kalyan Nagar consistently underperforms.

Trip Type Impact on Revenue

- Trips with 3-4 passengers generate higher revenue, especially during peak hours.
- Short-distance trips (0-5 km) and 5-passenger trips suffer from high cancellation rates, with the latter showing a 100% cancellation rate.

Recommendations

Aggressive Peak-Time Incentives

- Offer tiered surge bonuses (e.g., +20% for 5 trips, +30% for 10 trips) during 5-6 AM and 4-6 PM to maximize driver availability.
- Launch a "Weekday Warrior" program with a 50-100rs bonus for completing 20+ trips on Mondays and Fridays.

Location-Optimized Deployment

- Implement a geofenced surge multiplier (1.5x fares) in Yelahanka New Town during peak hours.
- Introduce a "Revenue Recovery Plan" in Kalyan Nagar with a minimum hourly rate guarantee (e.g., 200/hour) to maintain service levels.

Trip Type Overhaul

- Prioritize 3-4 passenger trips with a 10% revenue kicker per ride.
- Impose a 50-100 rs cancellation fee on short-distance and 5-passenger trips, with 50% redistributed to drivers.

Fleet Efficiency

Insights

Trip Volume and Duration Trends

- Trip volume peaks on Mondays, Fridays, and Sundays, with a trend toward shorter, more efficient trips.
- Bidarhalli has the longest average trip duration (>60 minutes), limiting trip frequency.

Capacity Utilization

- Trips with 3-4 passengers during peak hours optimize vehicle use, while 5-passenger trips have high cancellation rates.

Recommendations

Trip Frequency Optimization

- Introduce a "Quick Trip Bonus" of $1 per trip for completing 5+ short rides (<20 minutes) per hour.
- For long trips in Bidarhalli, offer a 100 rs long-trip premium and cap assignments at 1 per hour.

Capacity Optimization

- Launch "Pool Priority Zones" with a 25% fare discount for passengers and a 50rs bonus per pooled trip for drivers during peak hours.

Overall Revenue Strategies

Insights

Payment and Surcharge Leverage

- UPI payments yield the highest average revenue per trip.
- Congestion surcharges peak at 8-10 AM, and tips are higher during evening hours (4-10 PM).

Cancellation Revenue Drain

- Cancellations spike during 5 PM – 7 PM and on Fridays and Thursdays, especially for short-distance and 5-passenger trips.

Seasonal and Location Revenue Peaks

- **July, January, and December** are the highest-revenue months.
- Attibele generates strong revenue between 10 AM – 8 PM, and Yelahanka New Town leads in location-based earnings.

Recommendations

Precision Pricing

- Deploy block-by-block surge pricing during peak hours (5-6 AM, 4-6 PM) for a 20-30% revenue lift.
- Increase congestion surcharges by 50% during 8-10 AM in high-traffic zones.

Cancellation Revenue Recovery

- Implement a tiered cancellation penalty (50rs for short trips, 100rs for 5-passenger trips), with 75% redistributed to drivers.
- Offer a 150rs resilience bonus per completed trip during high-cancellation periods (5 PM – 7 PM, Fridays, Thursdays).

Payment and Tipping Boost

- Provide a 1% payout bonus to drivers for UPI-paid trips and a 5% fare discount to passengers using UPI.
- Run a "Double Tip Match" campaign during 4-10 PM, with Uber matching tips up to 100rs per ride.

Seasonal and Location Revenue Amplification

- Activate a "Seasonal Surge Lock" with a 15% fare increase in **July, January, and December**.
- Introduce a 1.3x fare multiplier in Attibele from 10 AM – 8 PM and enhance marketing in Yelahanka New Town