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

Introduction:

Bangalore, a city renowned 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. To improve driver earnings and overall platform efficiency, Uber aims to increase revenue per driver hour (RDH) by optimising demand-supply balance, minimising idle time, and ensuring that drivers maximize their earnings per trip.

Objective:

The objective of this project is to explore:

1. Uber's ride data to uncover insights that can help optimise **RDH**, **fleet efficiency**, and overall **revenue strategies**
2. Assess demand trends across Bangalore, **analyse driver earnings per trip per hour**, **focusing on RDH** and the factors that influence it and share the findings not limited to predefined factors.

Business Impact

Optimising RDH will not only boost driver satisfaction but also enhance customer experience by reducing waiting times and ensuring better availability of rides.

- **Increased Driver Retention:** Higher earnings motivate drivers to stay on the platform.
- **Reduced Idle Time:** Efficient driver allocation reduces idle time, maximising revenue.
- **Improved Customer Experience:** Faster ride availability during peak hours leads to higher customer satisfaction.
- **Higher Revenue for Uber:** Efficient demand-supply management directly impacts Uber's bottom line.

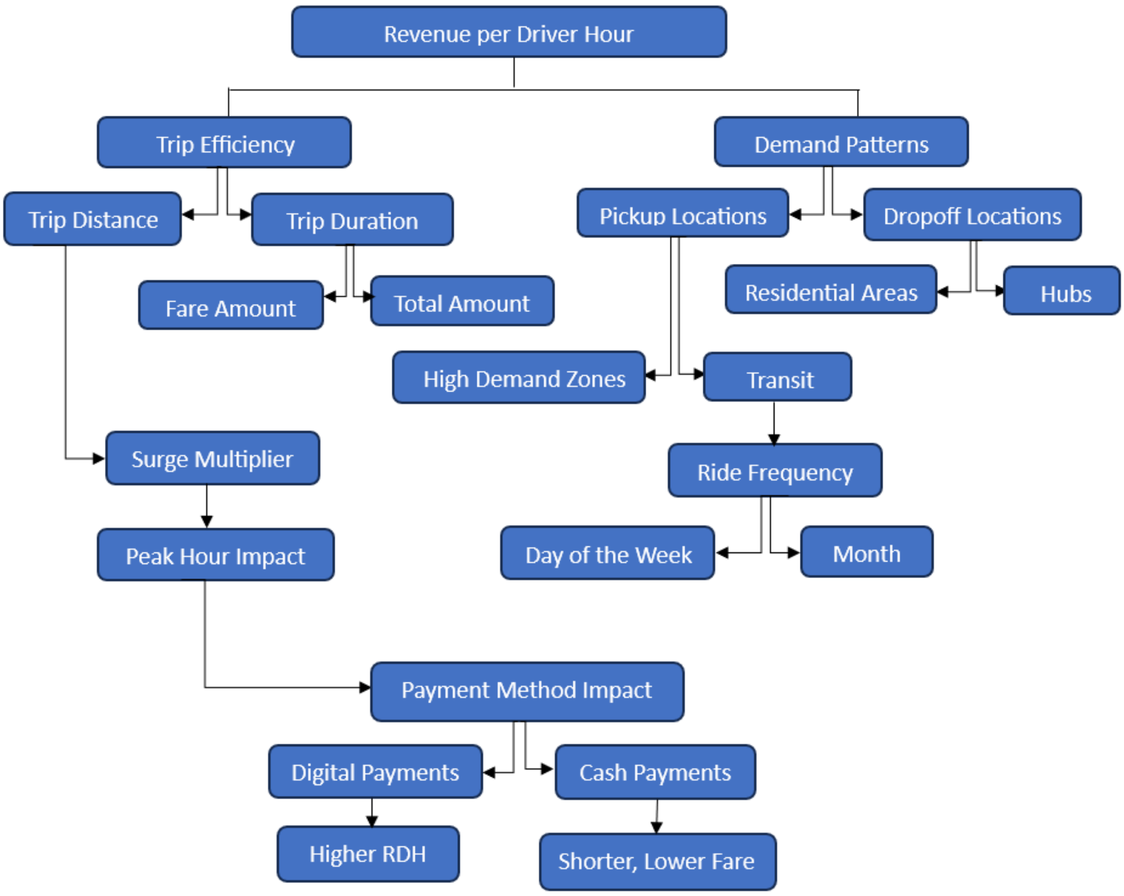
Dataset Overview

- Dataset Name: **Uber Dataset March'25**
- Number of Rows: **6500000**
- Number of Columns: **16**

Column Definitions

- **VendorID**: Identifier for the taxi service provider.
- **tpep_pickup_datetime**: The date and time when the uber trip started.
- **tpep_dropoff_datetime**: The date and time when the uber trip ended.
- **passenger_count**: The total number of passengers in the trip.
- **trip_distance**: The distance of the trip measured in KMs.
- **PULocationID**: The unique location ID where the passenger was picked up.
- **DOLocationID**: The unique location ID where the passenger was dropped off.
- **payment_type**: The method of payment used for the trip.
- **fare_amount**: The base fare for the trip, excluding additional charges.
- **extra**: Additional charges such as late-night surcharges or peak-hour fees.
- **gst**: Goods and Services Tax (GST) applied to the trip.
- **tip_amount**: The tip given to the driver.
- **tolls_amount**: Charges for tolls during the trip.
- **improvement_surcharge**: A fixed surcharge used for Uber service improvements.
- **total_amount**: The total fare amount including base fare, extras, tolls, and taxes.
- **congestion_surcharge**: An additional charge applied for trips in congested areas.

KPI Tree:



Data Cleaning and Preparation:

1. **Data Import:** The CSV file containing Uber trip data which was imported using gdown from Google Drive and loaded into a Pandas DataFrame in Google Colab for further processing.

2. **Finding Null Values:** The dataset contained missing values in: passenger_count & payment_type. Appropriate strategies were applied to handle these missing values.

3. **Checking Datatypes:** Verifying and converting data types was done to ensure proper formatting and efficient processing.

4. **Timestamp Handling:** Datetime columns were parsed and converted to a consistent datetime format.

5. **Missing Data Handling:** Missing values in passenger_count were filled with the rounded mean value of the column and Missing values in payment_type were replaced with 7, indicating **"Unknown"** to avoid inconsistencies.

6. **Handling Duplicates:** A total of **12,949 duplicate rows** were identified and successfully dropped from the dataset.

7. **Handling Outliers:** We handled outliers using capping with the IQR method to ensure extreme values do not distort the analysis.Outliers were clipped to lie within 1.5 times the IQR range to maintain data consistency.

8. **Creating New Columns:** Several new columns were created to facilitate deeper analysis and uncover meaningful insights:
    - **pickup_day –** Extracted the day of the week from tpep_pickup_datetime, enabling the analysis of demand patterns by day.
    - **pickup_hour –** Extracted the hour of the day from tpep_pickup_datetime to evaluate peak and off-peak hours.
    - **trip_duration_min** – Calculated trip duration in minutes by taking the difference between tpep_dropoff_datetime and tpep_pickup_datetime, giving insights into ride lengths.
    - **revenue_per_min** – Derived by dividing total_amount by trip_duration_min, enabling the calculation of earnings per minute.
    - **RDH (Revenue per Driver Hour)** – Calculated by multiplying revenue_per_min by 60 to measure driver earnings per hour effectively.

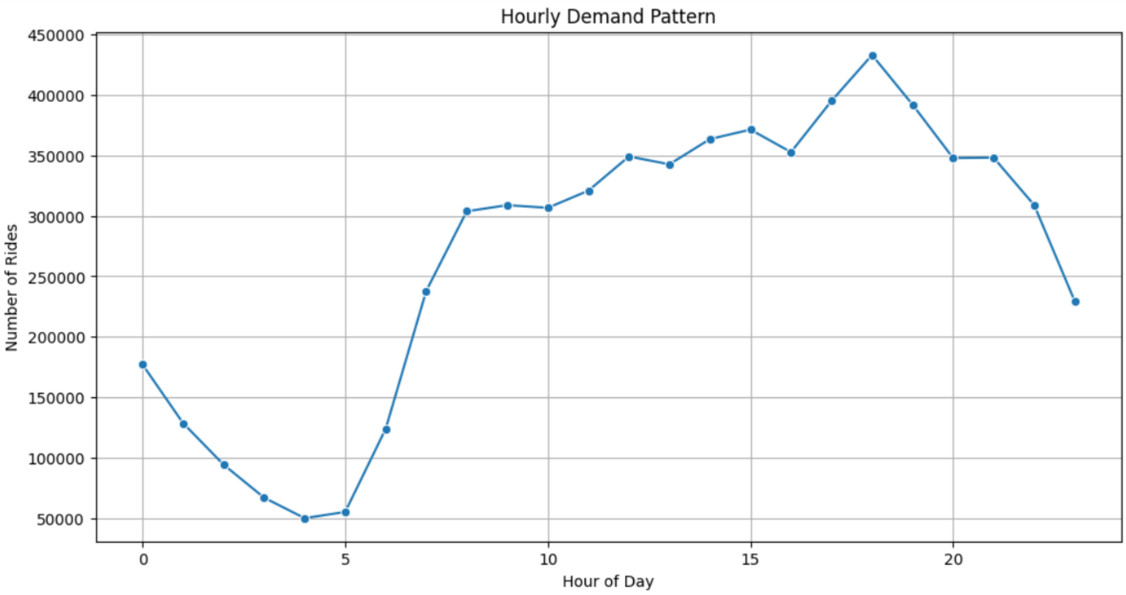
9. **Filling Missing Location Values:** Missing values in Pickup_Location and Dropoff_Location were replaced with **"Unknown"** to ensure data completeness.

10. **Dropping Unnecessary Columns:** After verifying relevance, unnecessary columns that did not contribute to the core analysis were dropped to streamline data for modeling and visualization.

Exploratory Data Analysis(EDA):

Hypothesis 1:

Peak hours contribute significantly to RDH.



Findings:

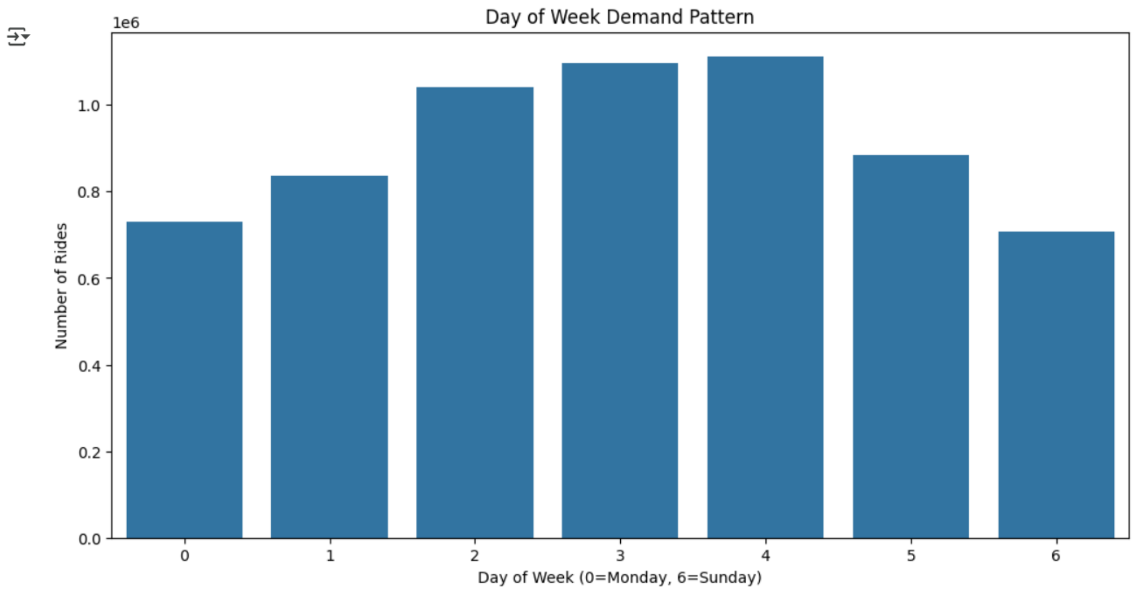
- Peak hours (morning and evening rush hours) show a significant contribution to higher Revenue per Driver Hour (RDH), aligning with increased commuter demand.
- Non-peak hours contribute less to RDH, indicating that demand is highly concentrated during peak travel periods.
- Hypothesis Status: **TRUE.** Peak hours play a crucial role in boosting RDH, validating the hypothesis.

Recommendations:

- Increase Drivers in Peak Hours: Allocate more drivers during 6 AM - 9 AM and 5 PM - 8 PM to reduce wait times.
- Reduce Fleet During Low Demand: Scale down operations between 12 AM - 5 AM to cut costs.
- Promote Off-Peak Rides: Offer discounts during 10 AM - 4 PM to boost demand.
- Ensure Late-Night Safety: Maintain driver availability and introduce secure ride options between 9 PM - 12 AM.

Hypothesis 2:

Ride demand is higher on weekends (Friday-Sunday) due to increased leisure and social travel.



Findings:

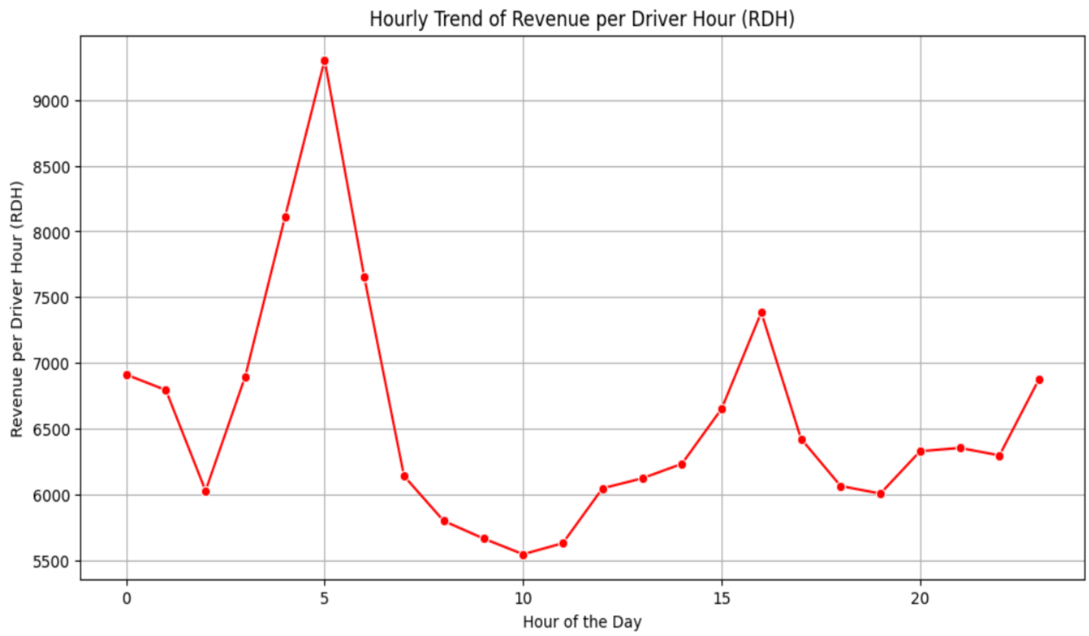
- **Friday records the highest demand,** followed by **Thursday and Wednesday,** indicating increased end-of-week travel.
- **Weekend demand (Saturday-Sunday)** is comparatively lower, suggesting that weekday rides, likely driven by work and commuting, dominate the overall demand.
- Hypothesis Status: **Partially FALSE.** Ride demand is highest on **weekdays (Tuesday-Friday),** with lower demand observed during weekends.

Recommendations:

- Focus on increasing promotions or incentives during weekends to boost ride demand.
- Analyze customer preferences and trip patterns to understand lower weekend demand.
- Optimize driver allocation to match peak demand on weekdays.

Hypothesis 3:

RDH spikes during peak hours due to high demand and surge pricing, while off-peak hours see lower earnings.



Findings:

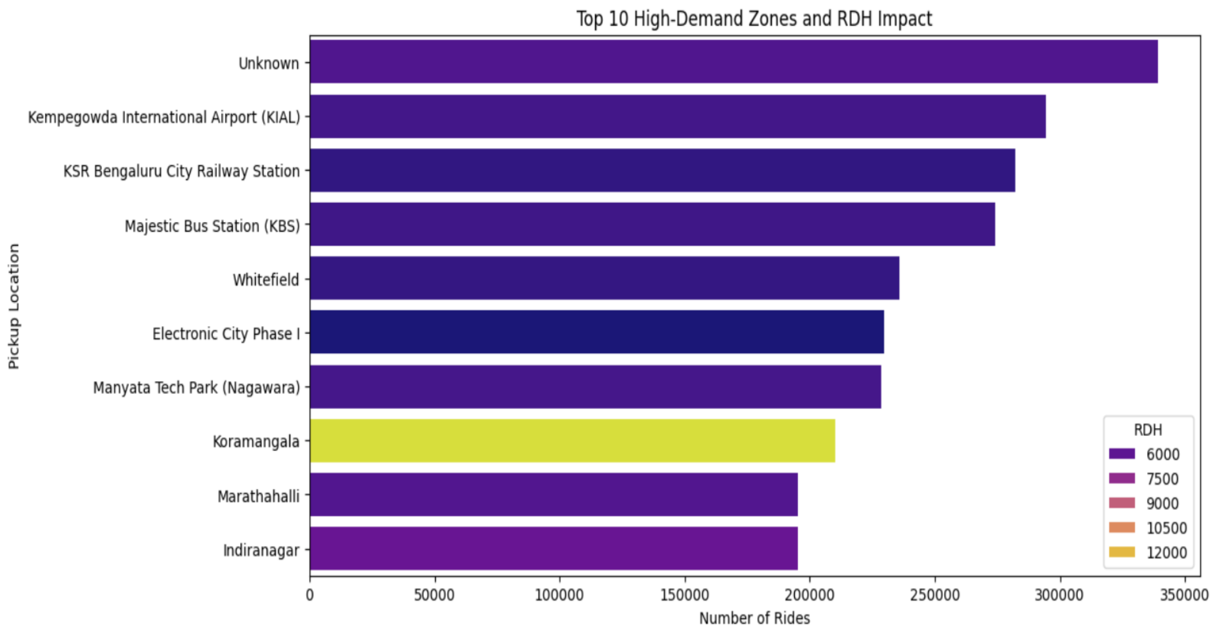
- **RDH peaks during early morning hours (3-5 AM),** likely due to airport or long-distance rides.
- A secondary rise occurs in the evening at **4-6 PM**, likely due to evening commuters, office pickups, and post-work travel.
- Hypothesis Status: **FALSE.** RDH does not peak during traditional rush hours, indicating that demand patterns may be influenced by other factors beyond commuter traffic.

Recommendations:

- **Refine surge pricing models** to capture potential demand during evening hours more effectively.
- **Optimize driver allocation** by analyzing ride patterns to ensure adequate coverage during high-demand periods.
- **Use predictive modeling** to dynamically adjust pricing and driver availability based on real-time demand trends.

Hypothesis 4:

High-demand pickup zones generate higher Revenue per Driver Hour (RDH) due to frequent ride requests.



Findings:

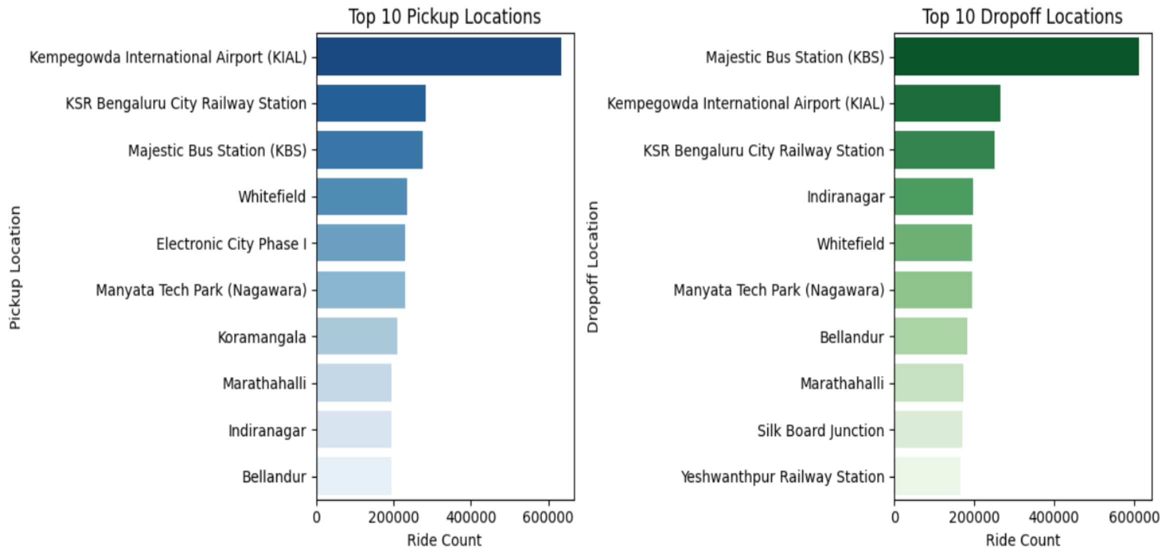
- **Koramangala** offers the highest RDH, making it a lucrative zone for drivers despite lower ride counts.
- **Airport**, **railway**, and **tech zones** show high ride volume but moderate RDH, indicating frequent but shorter trips.
- Hypothesis Status**: FALSE.** High-demand zones do not guarantee higher RDH, highlighting that factors like distance, duration, and pricing influence RDH significantly.

Recommendations:

- Optimize pricing strategies in high-demand zones to balance ride volume and revenue.
- Study Koramangala’s high RDH to identify patterns and apply them in other zones.
- Refine driver deployment by focusing on zones with higher revenue potential, not just high demand.

Hypothesis 5:

High-traffic areas (business hubs, transit points, and nightlife zones) dominate pickups and drop-offs, confirming demand concentration in key urban hotspots.



Findings:

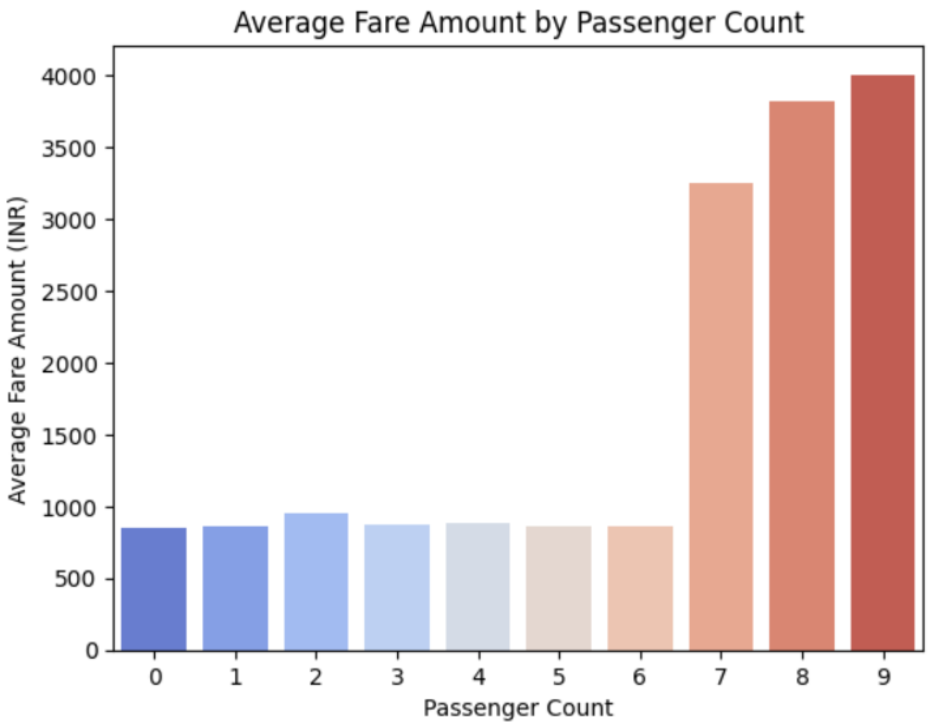
- Top Pickup Locations: Airports, railway stations, and key transit hubs show high demand, indicating a focus on intercity and long-distance travelers.
- Top Dropoff Locations: Major transit hubs and commercial areas dominate, reflecting frequent one-way trips originating from travel points.
- Hypothesis Status: **TRUE.** Demand is concentrated around high-traffic areas, validating the hypothesis that urban hotspots drive consistent ride volume.

Recommendations:

- Prioritize driver availability around peak hours at transit hubs to capture demand surges.
- Implement zone-based dynamic pricing to maximize RDH in high-demand zones.
- Analyze trip patterns from key hubs to optimize route allocation and reduce idle time for drivers.

Hypothesis 6:

Higher passenger count leads to increased average fare, as shared/group rides generate more revenue per trip.



Findings:

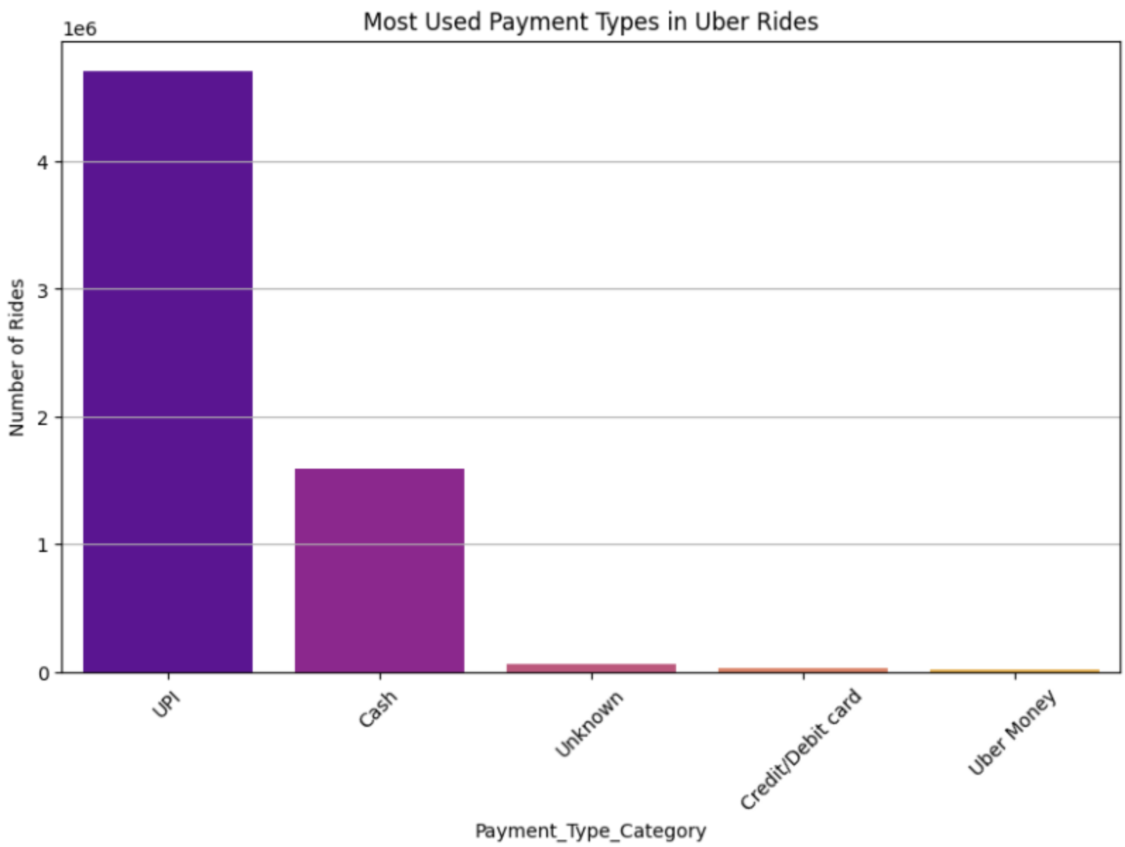
- Group/shared rides result in significantly higher fare amounts, especially for **7+** passengers.
- Larger groups and shared rides drive higher revenue, contributing positively to RDH.
- Hypothesis Status: **TRUE.** Higher passenger count directly correlates with increased average fare, supporting the hypothesis that shared/group rides generate more revenue.

Recommendations:

- Promote group ride options with dynamic pricing to optimize revenue.
- Introduce incentives for shared rides during peak hours to capitalize on increased fare potential.
- 🚘 Optimize vehicle allocation for larger passenger groups to reduce idle time and maximize RDH.

Hypothesis 7:

Riders using digital payments (UPI, Credit/Debit Cards) tend to take longer or more expensive trips compared to cash users.



Findings:

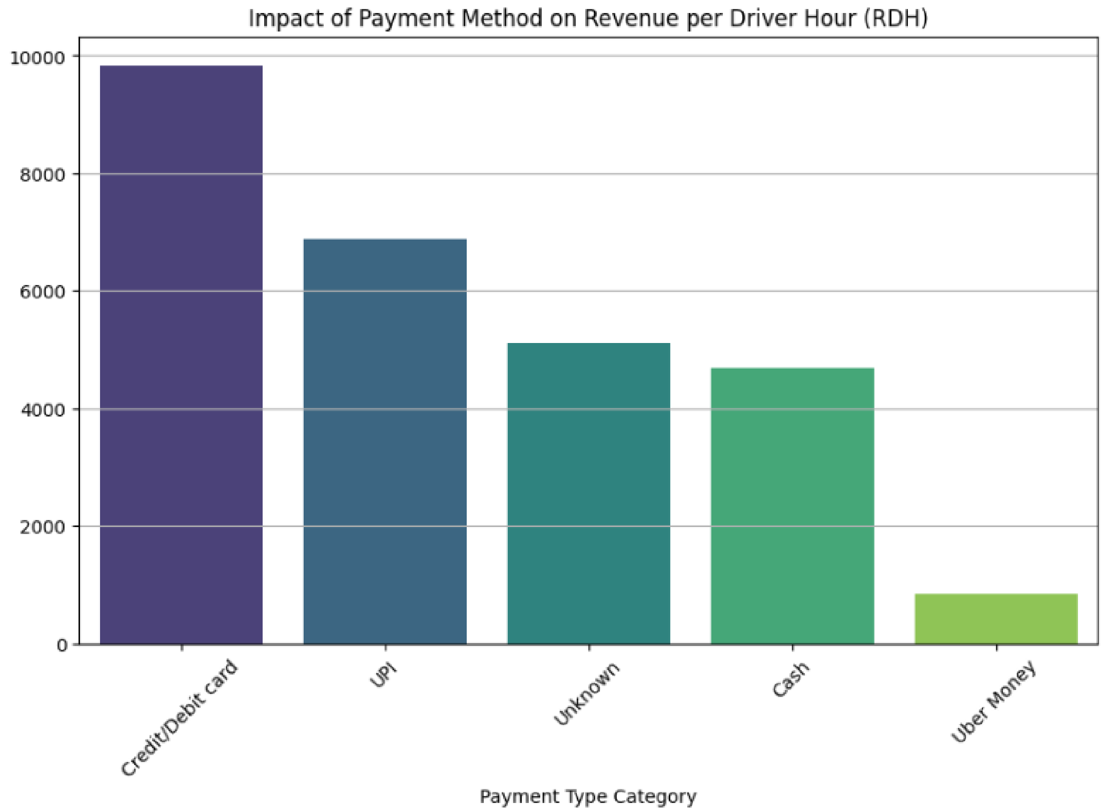
- UPI dominates with the highest number of rides, far exceeding other payment types.
- Credit/Debit cards and Uber Money show minimal adoption, indicating rider preference for UPI.
- Hypothesis Status: **Partially TRUE.** • While digital payments can lead to longer or more expensive trips, the dominance of UPI over credit/debit cards contradicts this trend, as card usage is **least preferred**.

Recommendations:

- Incentivize UPI and digital payments by offering cashback or loyalty rewards.
- Analyze high-value UPI transactions to identify premium users and design targeted offers.

Hypothesis 8:

Riders using digital payments (UPI, Credit/Debit Cards) tend to take longer or more expensive trips compared to cash users.



Findings:

- Though credit/debit cards are the least used, they generate the highest **Revenue per Driver Hour (RDH)**, indicating that card users tend to opt for longer or premium rides, often accompanied by generous tipping.
- UPI, being the most used payment method, contributes significantly but doesn't surpass card payments in terms of RDH.
- Hypothesis Status: **TRUE.** Digital payment users (especially credit card users) generate higher RDH compared to cash users, driven by longer trips and possible tipping behavior.

Recommendations:

- Boost Card Usage: Incentivize credit card payments with reward programs to maximize high-RDH rides.
- Promote Premium Services: Encourage UPI users to explore premium ride options for increased RDH.
- Target High-RDH Segments: Focus promotions on digital payment users to enhance overall revenue growth.

Formula Making:

Based on our hypothesis testing, we identified key factors such as trip duration, payment type, ride demand, and high-demand zones that significantly impact **Revenue per Driver Hour (RDH)** and overall ride efficiency. These factors were integrated to create metrics that ensure optimized ride pricing, better driver deployment, and improved revenue.

Here are the steps of the formula:

1. Trip Duration Calculation:
We calculate the total trip duration in minutes by taking the difference between drop-off and pickup times.

$$\text{trip\_duration\_min} = \frac{\text{tpep\_dropoff\_datetime} - \text{tpep\_pickup\_datetime}}{60}$$

2. Revenue per Minute Calculation:
We calculate the revenue per minute generated from each trip.

$$\text{revenue\_per\_min} = \frac{\text{total\_amount}}{\text{trip\_duration\_min}}$$

3. Revenue per Driver Hour (RDH) Calculation:
To assess the driver's hourly earnings, we multiply the revenue per minute by 60.

$$\text{RDH} = \text{revenue\_per\_min} \times 60$$

4. Handling Outliers Using IQR:
We cap outliers by using the Interquartile Range (IQR) to limit extreme values and maintain data consistency.

$$\text{Outliers Removed} = Q1 - 1.5 \times IQR \text{ and } Q3 + 1.5 \times IQR$$

5. Demand Grouping by Hour:
We group ride counts by hour to analyze hourly demand patterns.

$$\text{hourly\_demand} = \text{uber\_df.groupby('hour').size()}$$

6. Ride Count per Zone Calculation:
We calculate the total number of rides for each zone.

$$\text{zone\_demand} = \text{uber\_df.groupby('Pickup\_Location').size()}$$

7.Average RDH per Zone:
To analyze revenue impact by zone, we calculate the average RDH for each pickup location.

$$\text{zone\_rdh} = \text{uber\_df.groupby('Pickup\_Location')['RDH'].mean()}$$

Overall Analysis:

- **Peak Demand on Fridays:** Ride demand peaks on **Fridays** due to increased travel at the end of the workweek. However, demand gradually decreases over the weekend, highlighting the dominance of weekday commuter rides.
- **Inconsistent RDH During Peak Hours:** Contrary to expectations, RDH does not peak significantly during typical rush hours, suggesting possible pricing inefficiencies or an oversupply of drivers.
- **High-Demand Zones and RDH Impact:** High-traffic pickup zones like **KIAL, KSR Bengaluru City Railway Station, and Majestic** contribute to high ride counts but not necessarily higher RDH, indicating that ride volume does not directly translate to higher revenue.
- **Digital Payment Dominance: UPI** is the dominant payment method, contributing to higher revenue per trip. However, **Credit/Debit Card** usage remains minimal, indicating user preference for faster and more seamless payment options.
- **Longer Trip Duration for Digital Payments:** Riders using **digital payments** (UPI and Credit/Debit Cards) tend to take longer or more expensive trips, contributing positively to RDH, although cash remains a preferred method for shorter rides.
- **Impact of Passenger Count on Fare:** Higher passenger counts result in increased average fare, making shared/group rides more profitable.
- **Ride Demand Distribution:** Demand is highest around **transit hubs, commercial zones, and nightlife areas**, reinforcing the need for better driver deployment in these zones to minimize idle time.

Recommendations:

- **Demand-Driven Surge Pricing:** Optimize **surge pricing** models in high-demand zones during peak hours to balance driver availability and maximize RDH.
- **Zone-Based Driver Deployment:** Refine **driver allocation strategies** to focus not only on high-ride volume zones but also on maximizing RDH from zones with higher ride value.
- **Promote Digital Payment Incentives:** Encourage greater adoption of **Credit/Debit Card** payments through targeted promotions, as higher-value trips and longer durations contribute positively to revenue.
- **Segmented Pricing Strategy for Zones:** Implement **dynamic pricing** models based on zone-level ride patterns to capture both high-demand and high-RDH zones effectively.

- **Increase Ride-Sharing Promotion:** Incentivize **shared/group rides** to capitalize on higher passenger counts that lead to increased trip revenue.

- **Refining Payment Options for Low-Frequency Users:** Introduce payment flexibility (EMI or wallet integrations) to attract low-frequency cash users and shift them towards digital payments for higher RDH.

- **Leverage Data from Koramangala's High RDH:** Analyze Koramangala’s high RDH to identify replicable factors that can be extended to other zones with similar demographics.