An Exploratory Data Analysis of Uber: Data-Driven Insights on Demand,Revenue and Cancellations

Background

This Case Study analyzes **150,000 ride-hailing trips** to uncover trends in demand, cancellation, revenue and customer behavior. The goal is to provide data driven insights that can help optimize operations, improve customer satisfaction and revenue, and enhance driver efficiency.

In a ride-hailing service company, finding and handling cancellation trends and behavior is one the major factors that helps optimize the operations and the usage of resources. In this dataset We found that booking status of **62% rides are completed**, the rest 38% are not completed, which includes **6% Incomplete**, **7.1% No Driver Found**, **7.1% Cancelled by Customer** and **18.1% Cancelled by Driver**.

Moreover, Uber (or an Uber-like e-hailing service) operates in a dynamic environment where understanding rider behavior, pricing, and cancellations is essential to maintaining profitability and service levels.

This case study aims to analyze the provided ride dataset to:

- Identify demand patterns by time of day, day of week,month and geographically (pickup/drop locations).
- Examine cancellation trends (when, where, and why rides are cancelled by drivers or customers).
- Look closely into Incomplete booking status trends and reasons.
- Analyze revenue drivers, especially how ride distance, vehicle type, and fare correlate.
- Perform Customer behavior analysis(ride frequency,average CTAT,spend analysis, average ratings etc.)
- Propose actionable strategies to improve operations, pricing, and customer satisfaction.

Data Description & Structure

The dataset consists of detailed ride-hailing data from Uber Operation for the year 2024 covering the NCR region of India.The data spans multiple months and captures rich insights into booking patterns,revenue flow, both customer and driver behaviors across different vehicle types.

Dataset Source: [Uber Ride Dataset 2024]

Structure: This dataset contains 150,000 rows of data with 21 columns, each row corresponds to one booking made by a customer. It includes variables of ride bookings: their start/end locations, ride duration/distance, fare amounts, cancellation tags , and time metadata.

Exploratory Analysis & Findings

As we see in **Fig.1**, **62% of Booking Status is Completed** and the rest 38% includes Cancelled by Customer,Cancelled by Driver,Completed,Incomplete and No Driver Found. **Cancelled by Driver** is a total of **18.1%**. In **Fig.2.a & Fig.2.b** we deep dive through the reasons for Cancellation by Customer and Driver respectively.

In **Customer Cancellation Reasons**, Change of plans,Driver asked to cancel,Driver is not moving towards pickup location and Wrong Address, each of the reasons contributes **22%(approx.)** and AC not working is **11%.** Whereas in **Driver Cancellation Reasons**, Customer related issues,More than permitted people in there,Personal & Car related issues,The customer was coughing/sick are the reasons category and each of the categories constitute **25%(approx.)**.

Fig.3 draws the hourly Cancellation trend of **Cancelled by Customer** and **Cancelled by Driver** based on **booked rides booking count,** meaning on a subset of overall booking data with Booking Status as Completed,Cancelled by Customer and Cancelled by Driver because 'No Driver Found' and Incomplete booking necessarily doesn't mean booking was successful due to vagueness of data. Whereas for a ride to be cancelled it needs to be booked first. Hence we included Cancelled by Customer and Cancelled by Driver in booked subset data along with Completed. C**ancelled by Customer** and **Cancelled by Driver both** peak the cancellations during 9-11 a.m. and 6-8 p.m. hour with highest number of cancellation **889** and **2257** respectively at **6 p.m.** To further dig into the **peak hour(6 p.m.)** cancellation reasons for both Cancelled by Customer and Cancelled by Driver, we have a pie chart(**Fig.4**) representing the reasons for both types of cancellations. **Peak hour Customer Cancellation Reasons** shows that **Wrong Address** contributes **24.6%** to cancellation reasons, which is a bit higher than the other reasons. The percentage of other three reasons: Change of plans,Driver asked to cancel and Driver not moving towards pickup location is 22-23%(Approx.) and AC is not working is 9.3%. The Wrong Address reason of Cancellation by Customer on booked rides is also higher than the **Wrong Address(22.5%)** reason on overall rides. Whereas, the reasons for Cancellation by Driver on booked data shows the same pattern of reason distribution to the reasons for Cancellation by Driver on overall data which is **25%(approx.).**

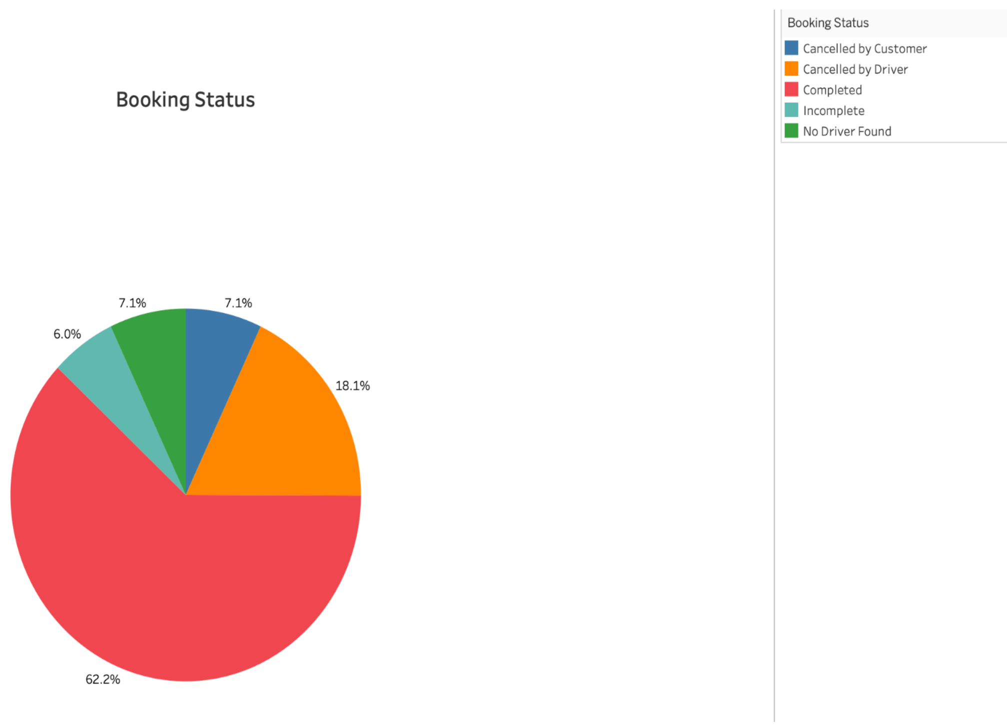


Fig.1 Booking Status

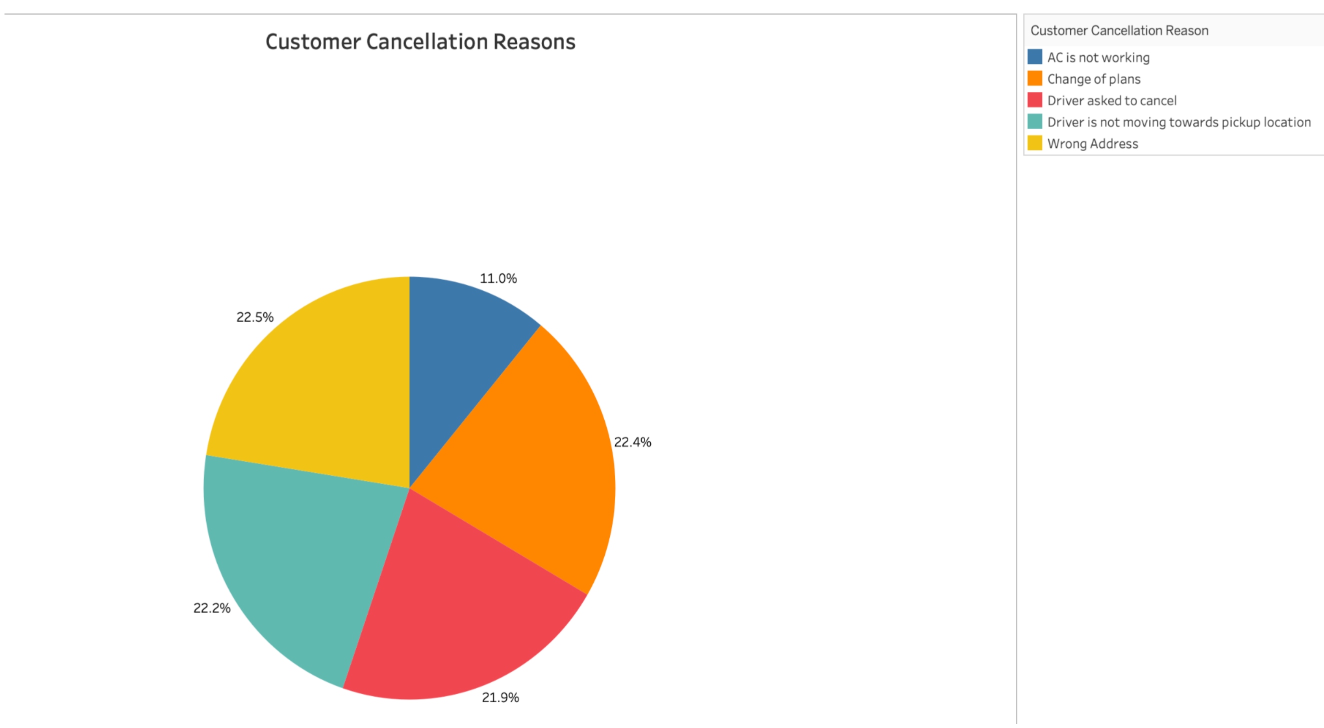


Fig.2.a Customer Cancellation Reasons

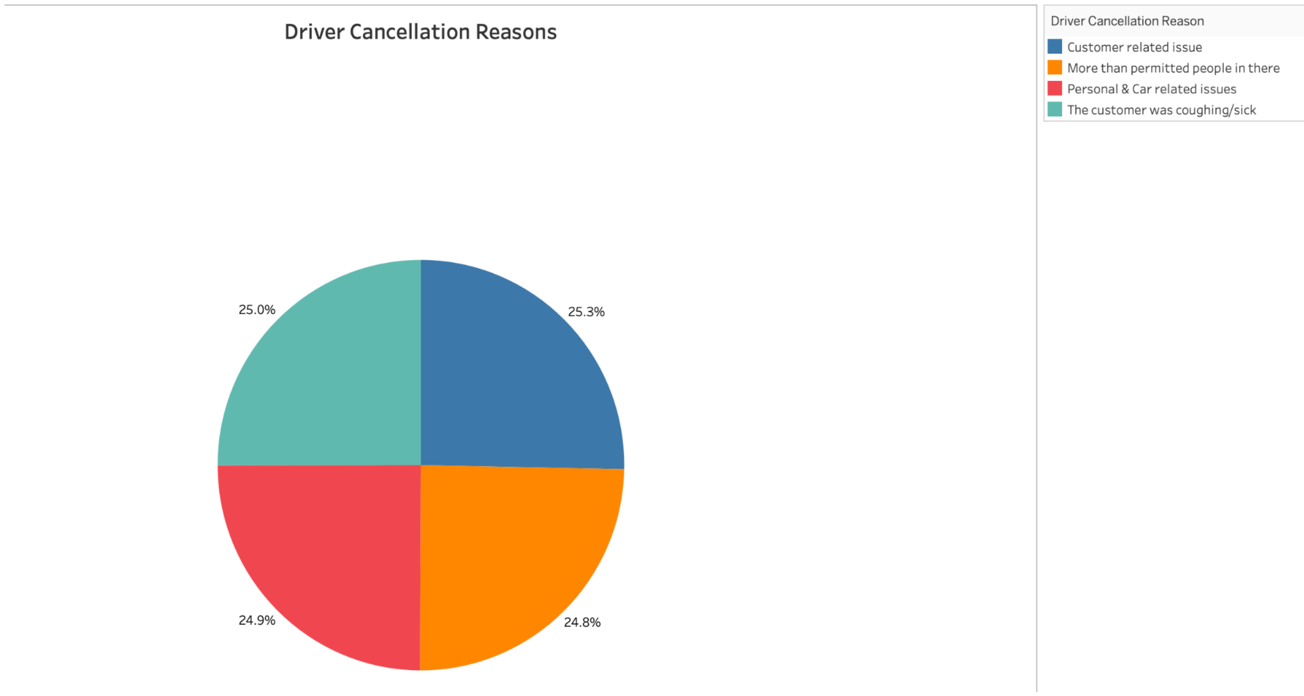


Fig.2.b Driver Cancellation Reasons

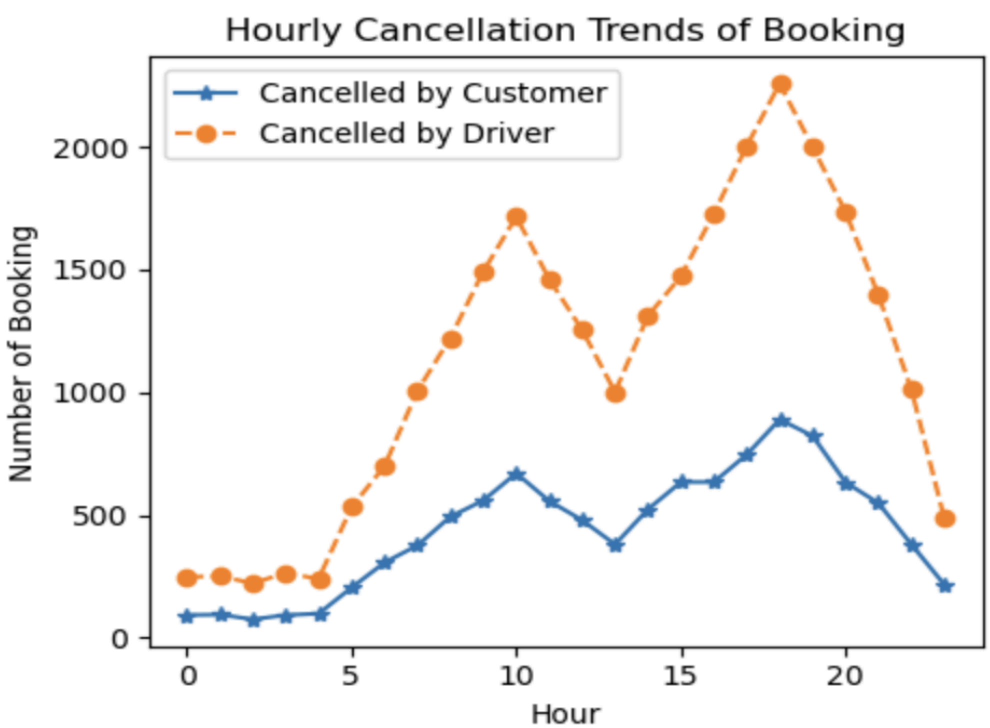


Fig.3 Hourly Cancellation Trends of Booking on Booked Rides

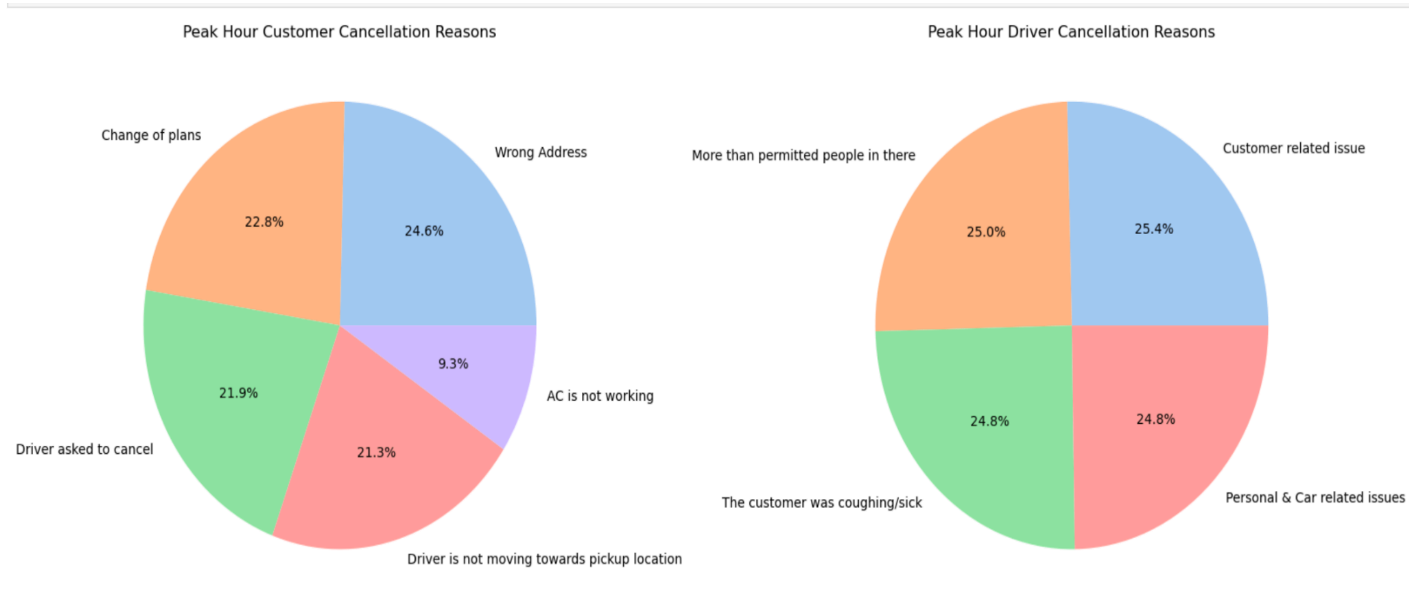


Fig.4 Peak Hour Cancellation Reason by Customer and Driver

As we find the **top ten pickup and drop locations(Fig.5)** based on the **overall booking count**, we see that **Khandsa** is the top pickup location with 900+ bookings. **Ashram** is the top drop location with 900+ bookings based on the overall booking count.

We find the top ten pickup locations based on **booked rides booking count**, and also on the data with booking status as **Completed** i.e. **completed rides booking count.**

In all the three bar charts of the top ten pickup locations, **Khandsa** is the top pickup location based on booking counts along with some other overlapping pickup locations such as **Barakhamba Road, Badarpur, Madipur, Mehrauli,Dwarka Sector 21** etc.

Saket and AIIMS are in top ten in overall and booked rides booking but not present in top ten in completed rides booking chart. Similarly, **Shivaji Park and IGI Airport** are in top in booked booking but not in completed booking. In **Fig.6** we evaluate the **cancellation rate of the top ten pickup locations of booked booking** and find **Saket, Shivaji Park, AIIMS and IGI Airport** as top four cancellation locations with highest percentage of cancellation rate on booked booking i.e. **30.89%, 30.22%, 29.57%, 29.56%** respectively. which justifies why these four pick up locations were not present in **Fig.5.c chart**. To further analyse the reasons for cancellation in these four locations we check for Customer/Driver cancellation reasons respectively.

In **Customer Cancellation Reason** at **Saket**, **Driver** is **not moving towards pickup location** is **28.6%,** whereas at **Shivaji Park**, **Change of plans** is **25.9%** & **AC** is **not working is 17.2%**. At **AIIMS**, **Wrong Address** is **25%** & **AC is not working is 15%**. And at **IGI Airport**, **Change of plans** is **23.9%** & **Driver is not moving towards pickup location is 26.8%**. These percentage of reasons are higher compared to Customer Cancellation Reasons on overall booking.

In **Driver Cancellation Reasons** at **Saket, Customer related issues** is **30.7%** & **More than permitted people in there** is **29.6%,** at **AIIMS, The customer was coughing/sick** is **28.4% & More than permitted people in there** is **27.8%,** at **IGI Airport, Customer related issues** is **29.9%**. Whereas the other factors of reason for cancellation by driver in Saket,Shivaji Park,AIIMS & IGI Airport do not seem much different, either they are similar in number or less compared to driver cancellation reasons on overall booking. And above mentioned highlighed factors are higher in percentage compared to overall booking.

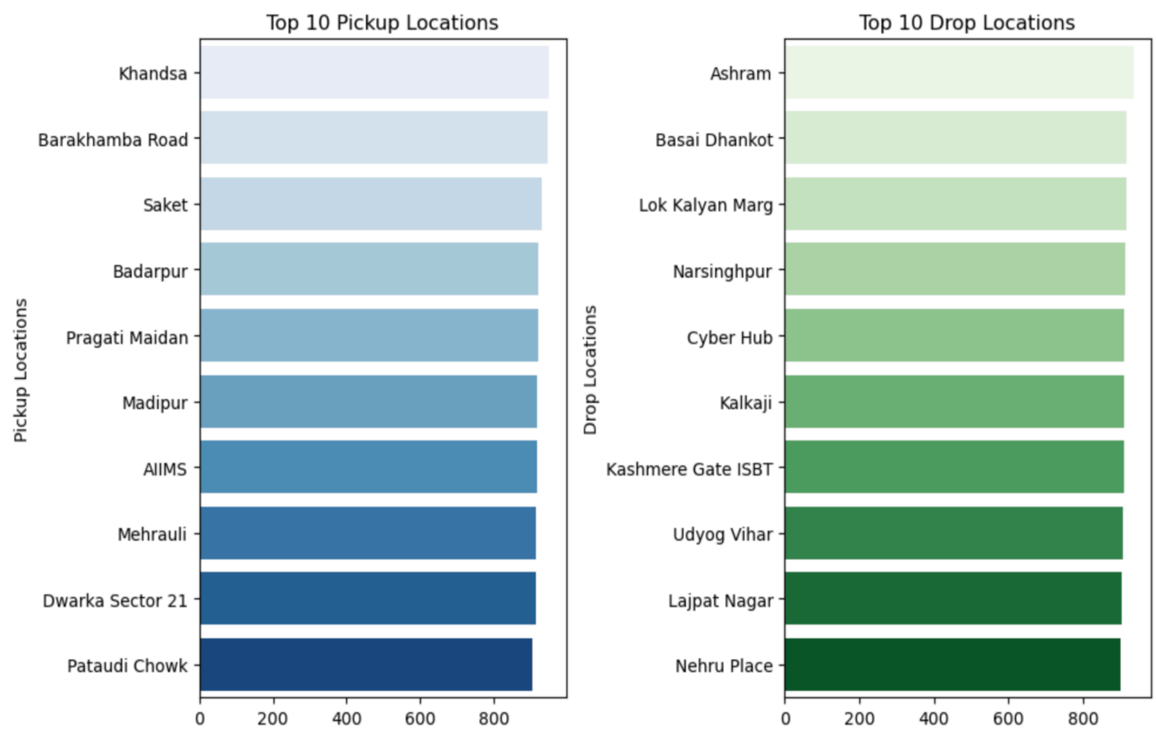


Fig.5.a Top 10 Pickup Locations and Top 10 Drop Locations on overall Booking

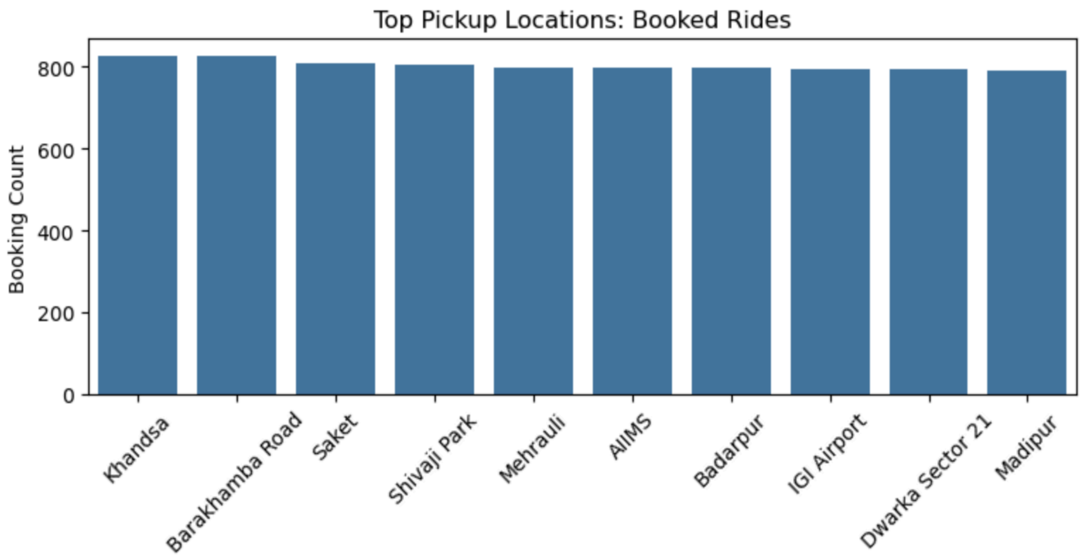


Fig.5.b Top 10 Pickup Locations: Booked Rides

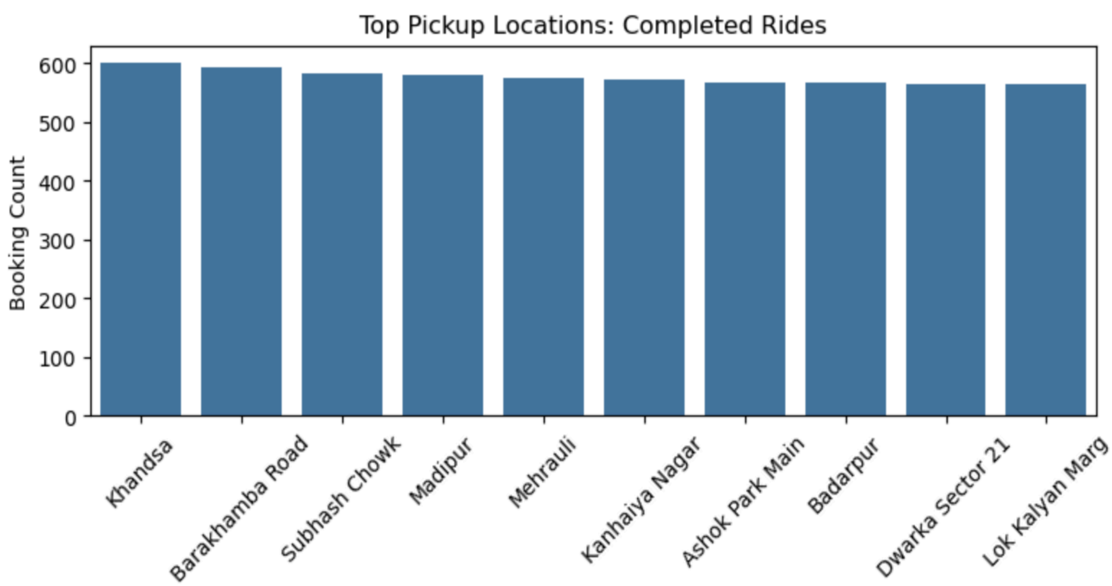


Fig.5.c Top 10 Pickup Locations: Completed Rides

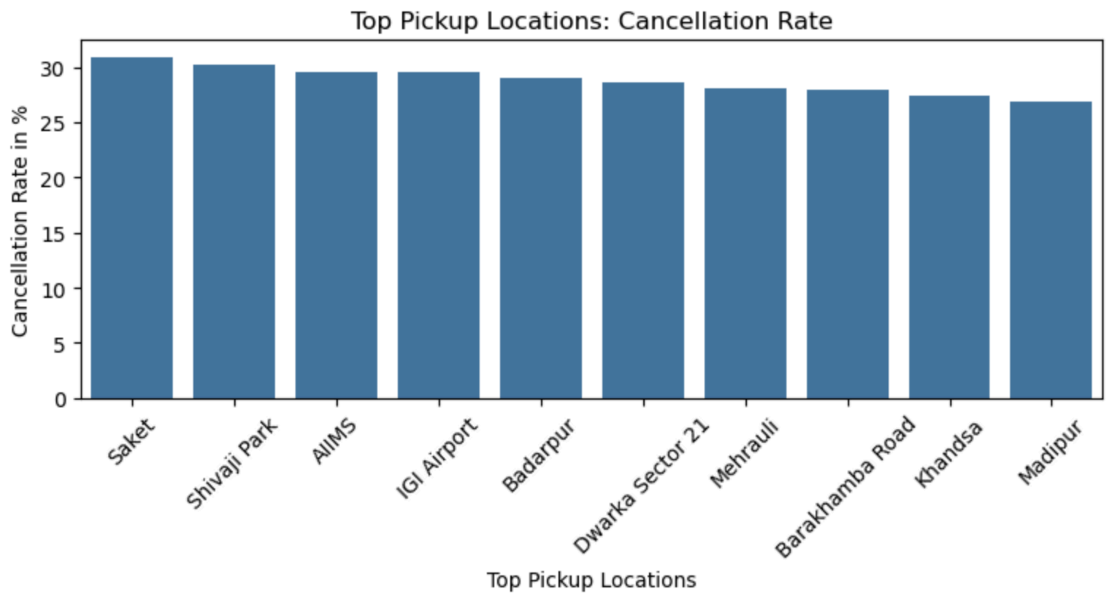
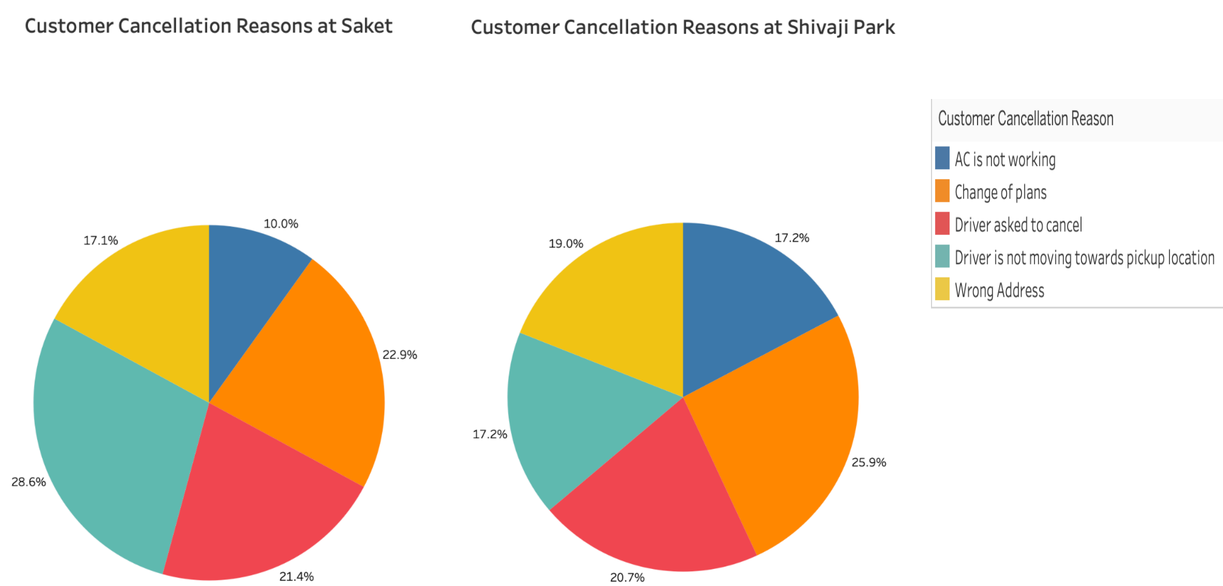


Fig.6 Top Pickup Locations based on Cancellation Rate of booked rides



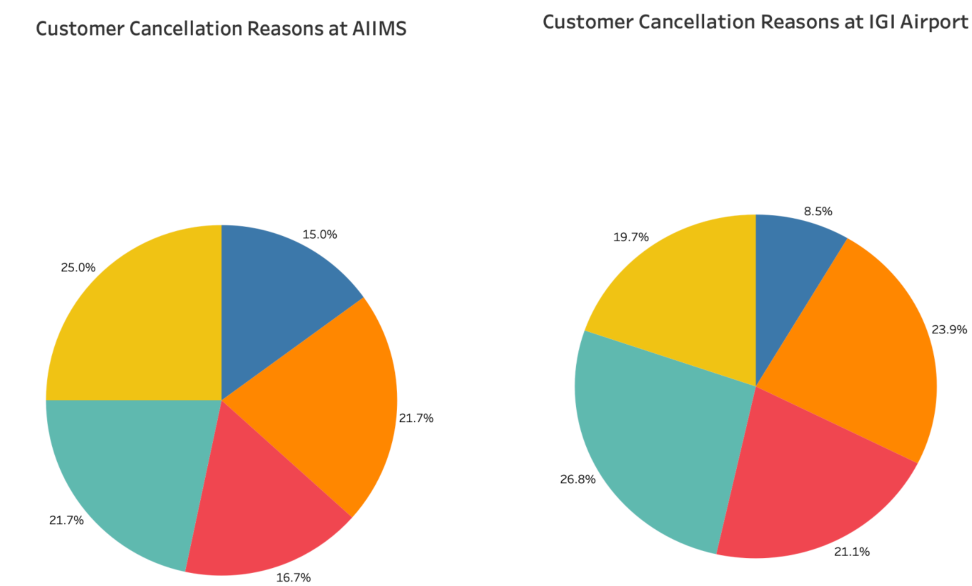
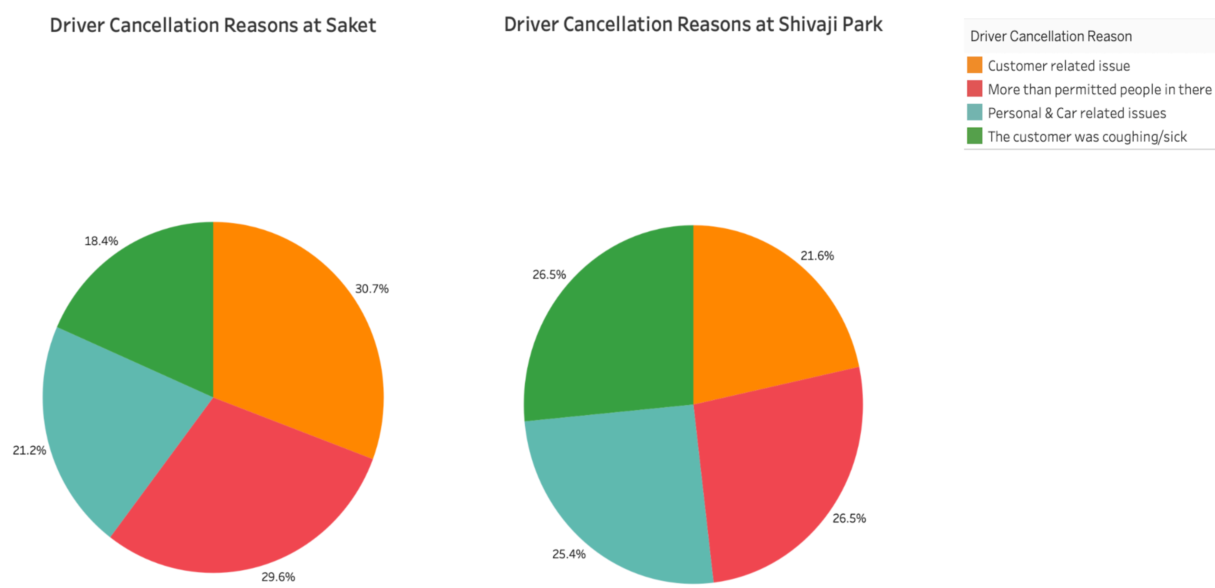


Fig.7.a Customer Cancellation Reasons at Saket,Shivaji Park,AIIMS & IGI Airport



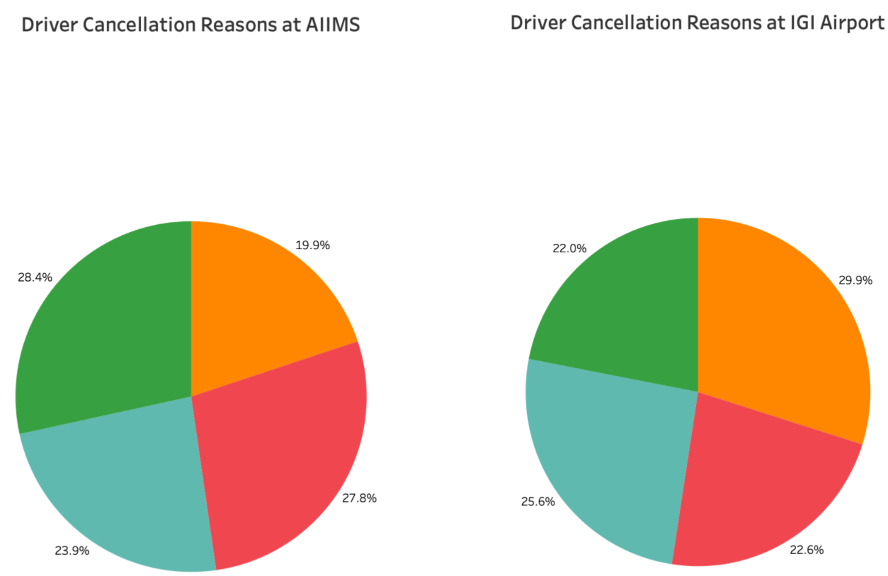


Fig.7.b Driver Cancellation Reasons at Saket,Shivaji Park,AIIMS & IGI Airport

As we analyse the **Rides Booking by Hour** of the day on overall booking, we see peak booking hour in the morning is between **9-11 a.m.** and in the evening between **3-9 p.m.** with higher number of rides bookings. Whereas, **Rides Booking by Day of Week** shows a flat trend throughout the day of the week with **20k+ booking** everyday.

In **Fig.9. Total Monthly Rides** ranges between **12,000 and 13,000 rides per month**. The highest number of bookings are done in the months of January, March, May & July. **Completed Monthly Rides** ranges between **7000 and 8000 rides per month**. There is slight dip in February and September month in both total and completed rides, then recovery in later months, basically the completed rides are mirroring the total rides. There is a large and **consistent gap** between the total monthly rides and completed monthly rides around **4000-5000 rides per month** which are not completed. But in the months of **April and June** we see a sharp **dip in total booking** but a **slight dip in completed booking,** not proportional to the drop in total rides suggesting a narrower gap than usual between total and completed rides booking. In **Fig.10.(Monthly Cancellation Rate)**, the month of **April has the lowest cancellation rate 24.59%** and the month of **June** also shows a low cancellation rate of **24.83%**. March & October also have low cancellation rate 24.73% & 24.65% respectively. March & October, both months are having high counts of total rides and completed rides, hence the narrower gap and low cancellation rate.

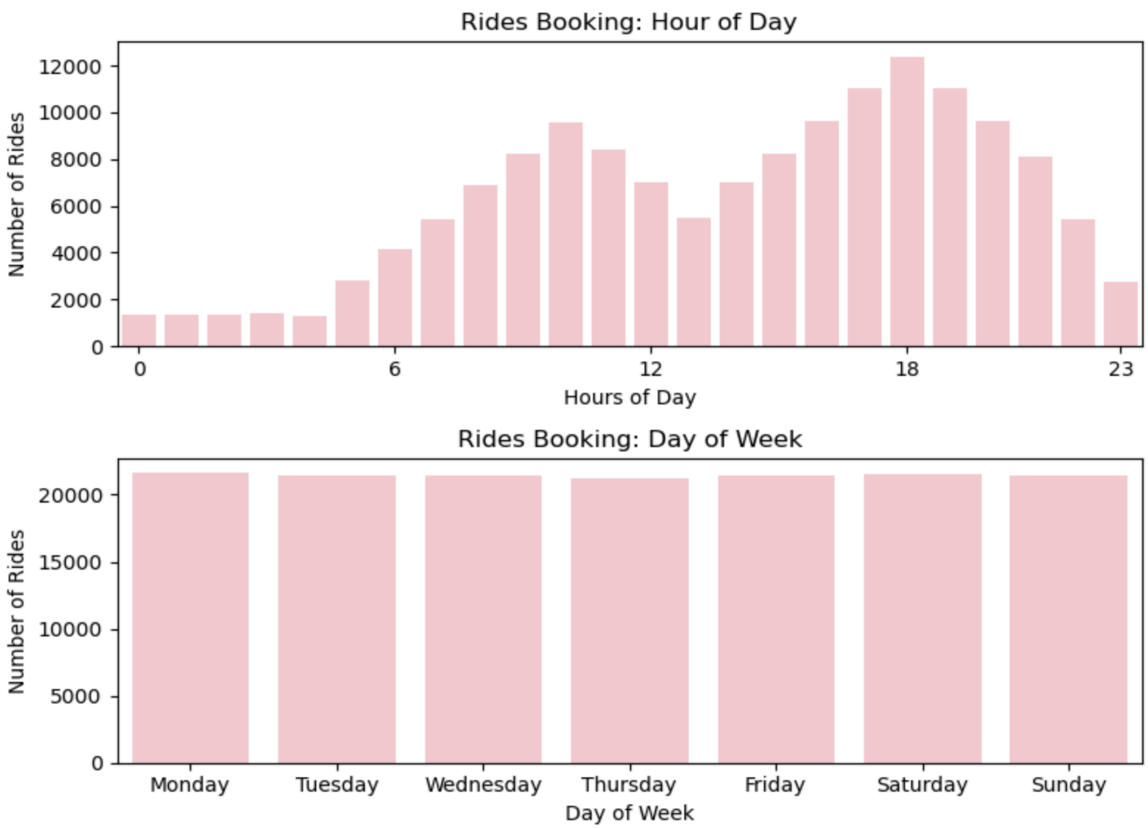


Fig.8. Rides Booking: Hour of Day and Day of Week Comparison

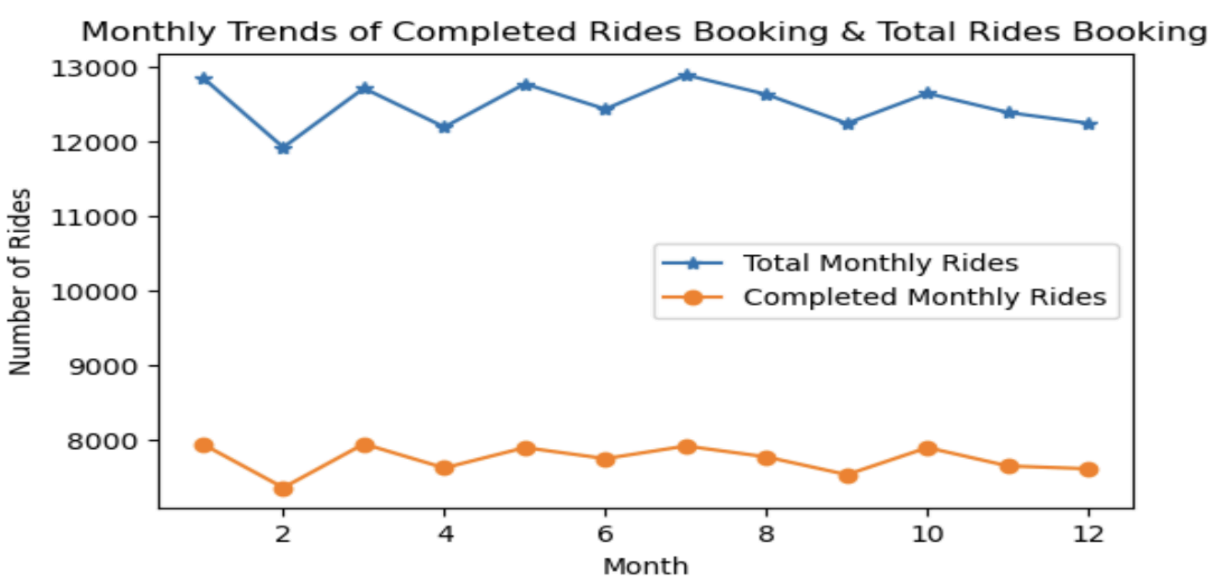


Fig.9. Monthly Trends of Completed Rides & Total Rides

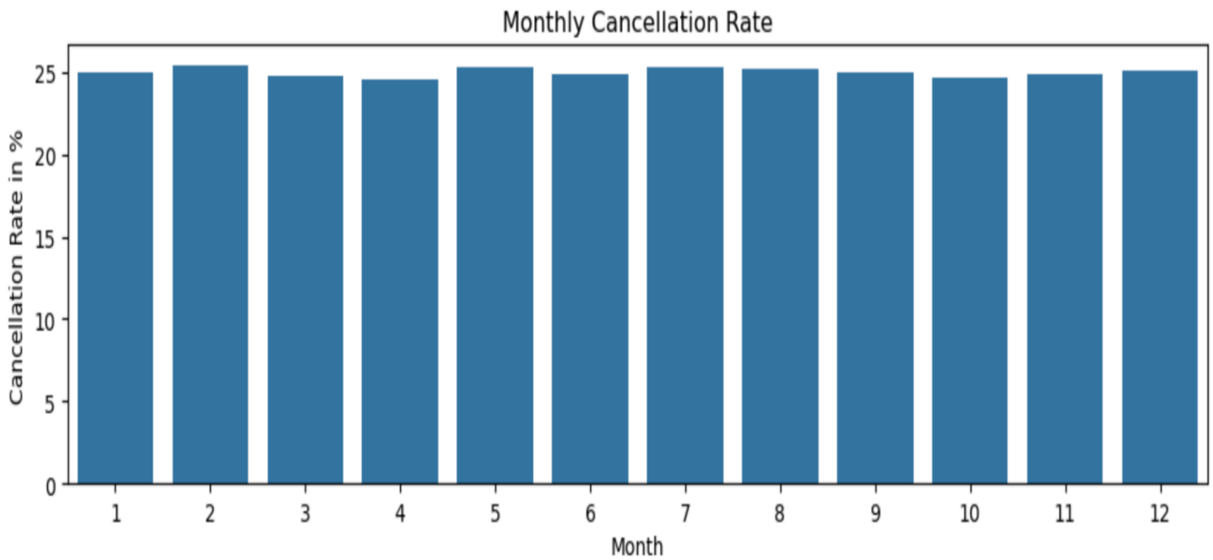


Fig.10. Monthly Cancellation on Overall Rides Booking

In **Fig.11** we have a Monthly Incompletion rate and No Driver Found rate on total booking. Month of **November** sees the **highest number of Incompletion rate i.e, 6.37%** and month of August has the highest number of No Driver Found rate i.e, 7.41%.

As we further analyse **Hourly Trends of Incomplete Bookings** and the reasons for incomplete booking we see that the highest number of incomplete bookings is at 7 p.m with 750+ number of incomplete bookings. Vehicle breakdown is the top reason for peak hour incomplete bookings.

In the **Day of the Week trend of incomplete booking,** Tuesday has the peak number of incomplete booking with 1320+ number of incomplete bookings and Customer Demand is the top reason for booking incompletion.

In **Monthly trend of incomplete booking**, January has 800+ number incomplete bookings and Vehicle Breakdown, Customer Demand and Other issues have almost equal contribution towards the incompletion of Rides.

In **Fig.13** as we compare the average number of completed rides on weekdays and weekends, we find a similar trend, no sharp weekdays drop or no weekends spike. In terms of completed rides by vehicle type, **Auto with 20k+ number of completed rides** is in the top vehicle type followed by **Go Mini & Go Sedan with 15k number of completed rides** followed by Bike, Premier Sedan, eBike and Uber XL.

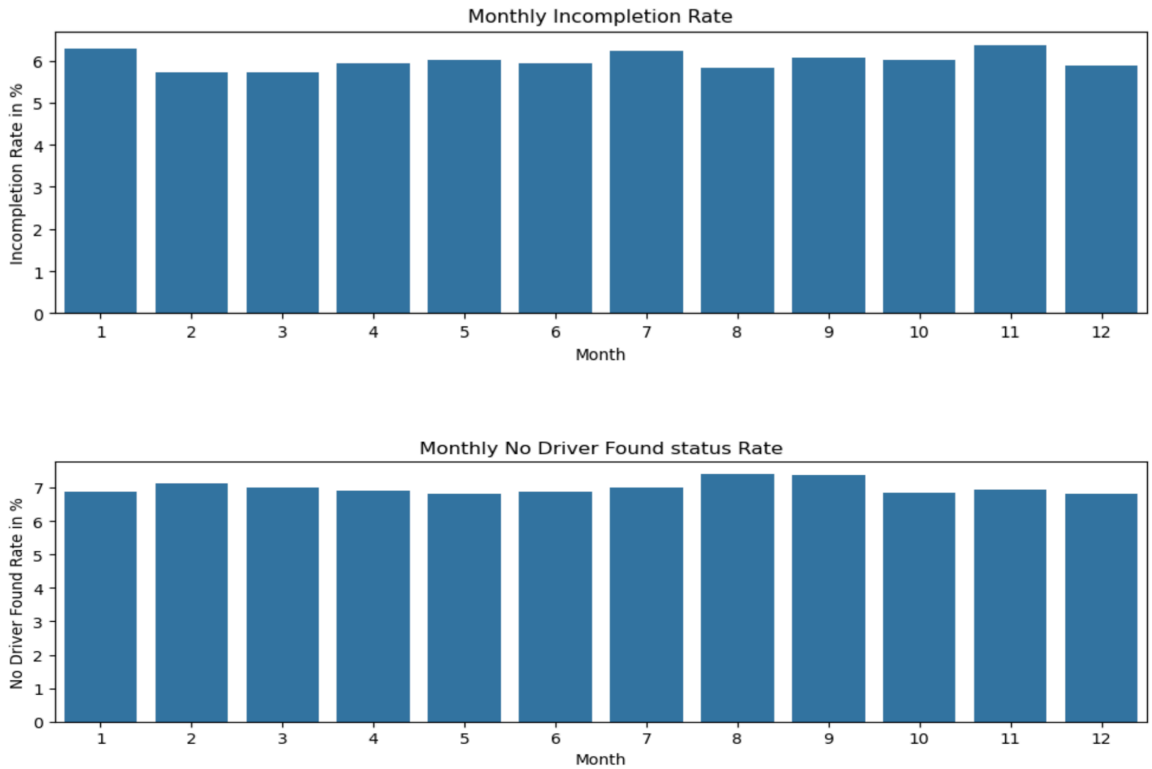


Fig.11. Monthly Incompletion/No Driver Found rate on Overall Rides Booking

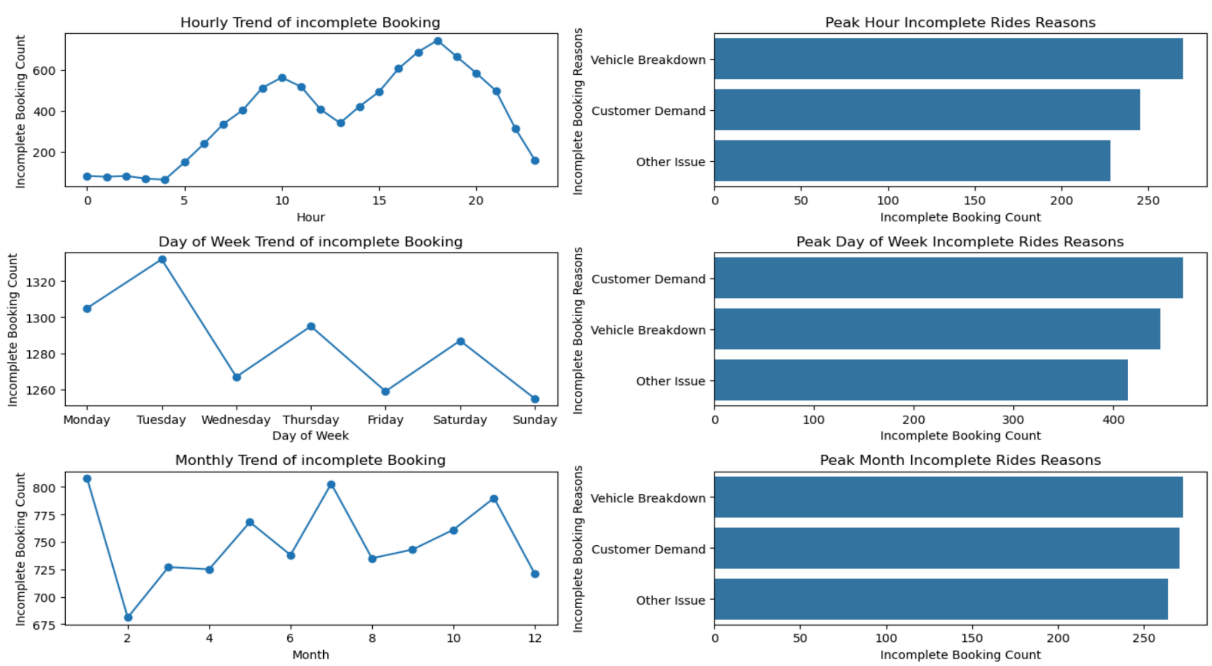


Fig.12. Hourly/Day of Week/Monthly Trend and reasons for Incomplete Rides

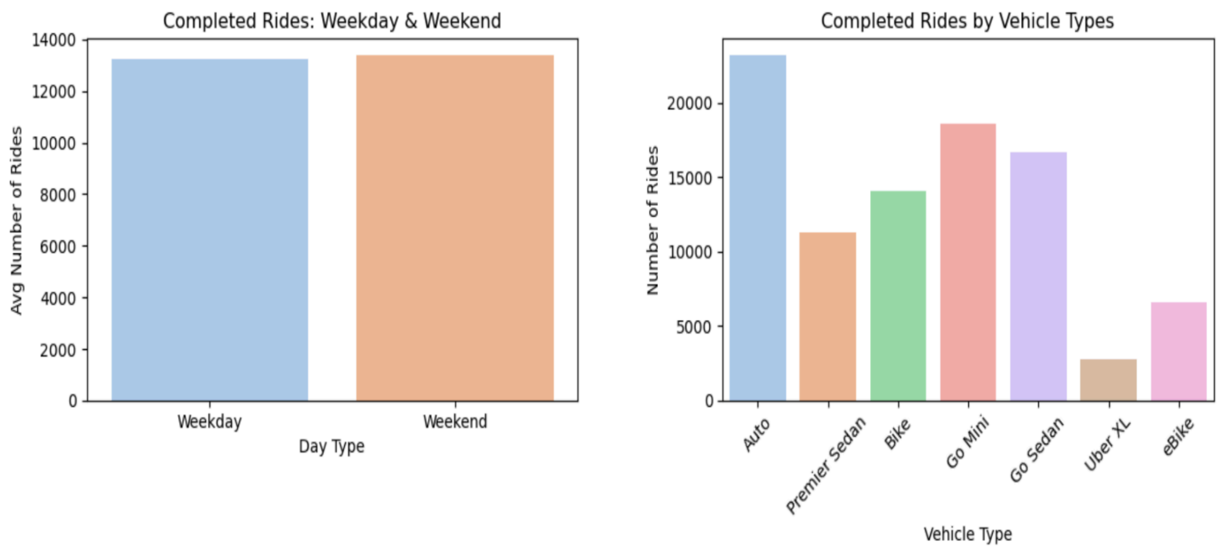


Fig.13. Completed Rides Trends: Weekday vs Weekend and Vehicle Type

As we analyse the total revenue we consider **overall booking** cause even the incomplete rides have a contribution towards the revenue generation. In **Fig.14** we see **Auto contributes 12.8M+ revenue** for the Uber rides, whereas **Go Mini and Go Sedan contribution is 10.3M+ & 9.3M+ respectively.** Bike contribution is 7.8M+, Premier Sedan 6.2M+, eBike 3.6M+ and being the least contributor, Uber XL revenue contribution is around 1.5M+. Whereas all the vehicle types' average contribution towards revenue generation is almost similar around 300K+.

In **Fig.15** we see the total distance and average distance covered by vehicle types. **Auto, Go Mini** and **Go Sedan** are in top three of vehicle types with the highest total distance covered followed by Bike, Premier Sedan, eBike and Uber XL respectively. Whereas the average distance covered by all types of vehicle are similar in trend.

To further dig into revenue generation analysis based on the distance covered, we create five distance ranges viz. 0-5km, 5-10km, 10-20km, 20-50km, 50km+ and find the number of rides booked by distance range and revenue generated by each distance range on overall rides and completed rides respectively. In **Fig.16** we can observe that the maximum number of booking on both overall rides booking and completed rides booking is done in the **distance range 20-50km** followed by 10-20km, 5-10km & 0-5km respectively. We further observe that **average revenue** generated by distance range on both **overall rides booking** and completed rides booking shows that all the ranges show a similar contribution towards average revenue.

In **Fig.17** we find **average revenue per km** by distance range of completed rides which imply that **0-5km distance range** generates the highest average revenue per km followed by 5-10km, 10-20km and 20-50km distance ranges respectively. As we analyse the revenue per km drop between Short Trips(0-5km) and Long Trips(20-50km) on completed rides booking, we see a **more than 90% drop** in the revenue per km.

In the 0-5km distance range, **Go Sedan** has a significantly high contribution towards **average revenue** and **average revenue per km**. Rest vehicle types show almost a similar amount of contribution in all distance ranges for both average revenue and average revenue per km.

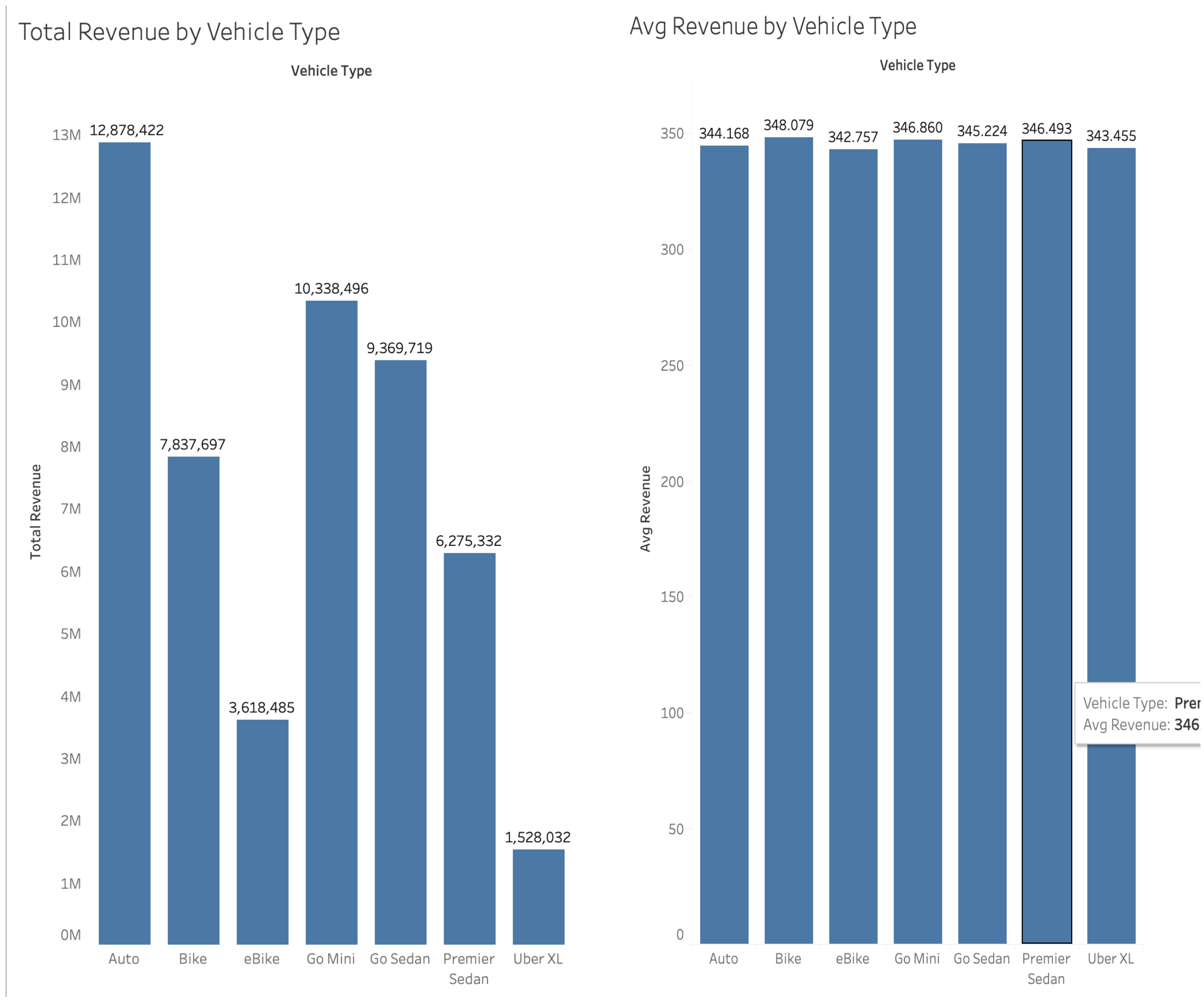


Fig.14. Total Revenue/Avg Revenue by Vehicle Type on Overall Booking

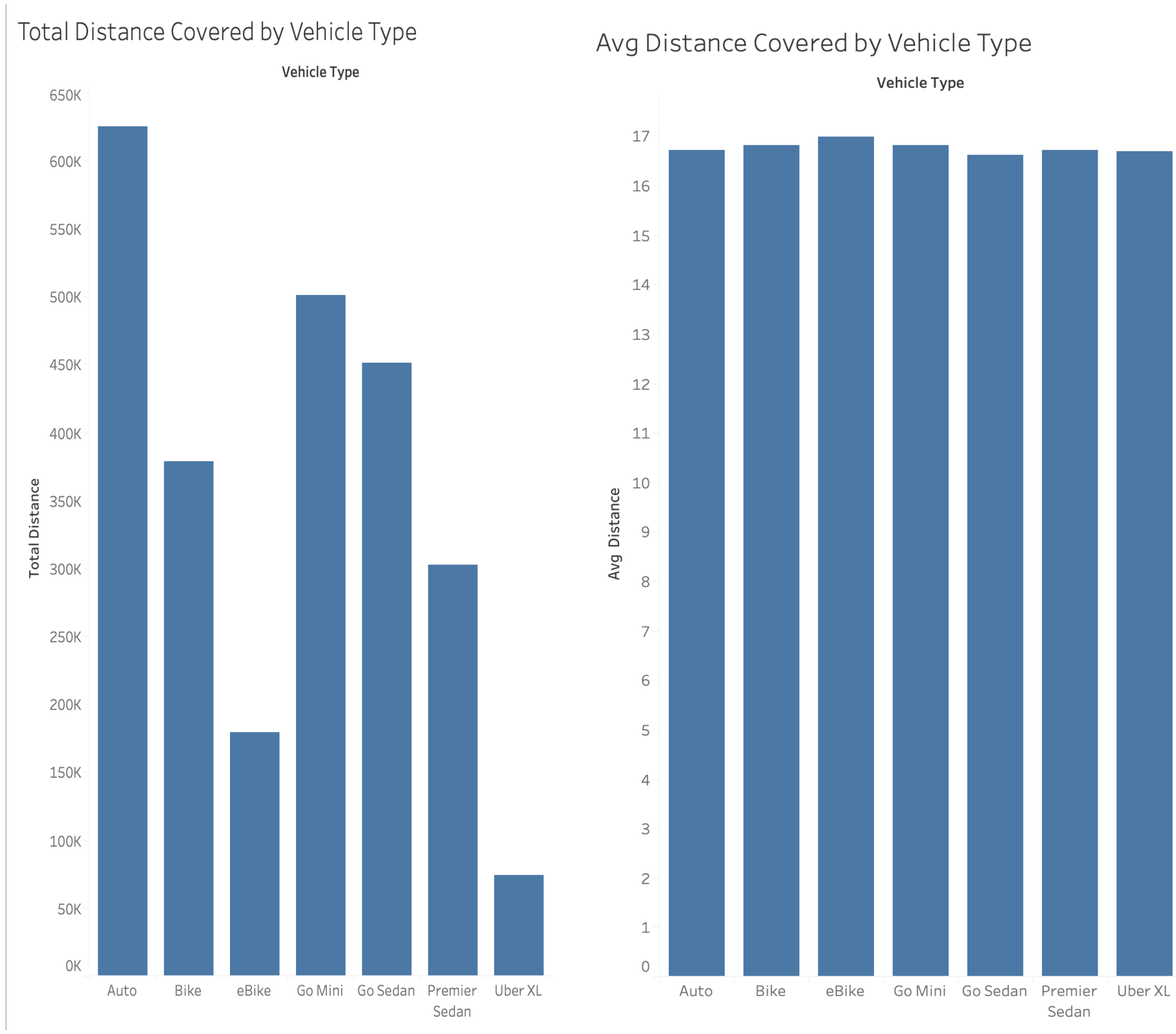


Fig.15. Total Distance/Avg Distance covered by Vehicle Type on overall booking

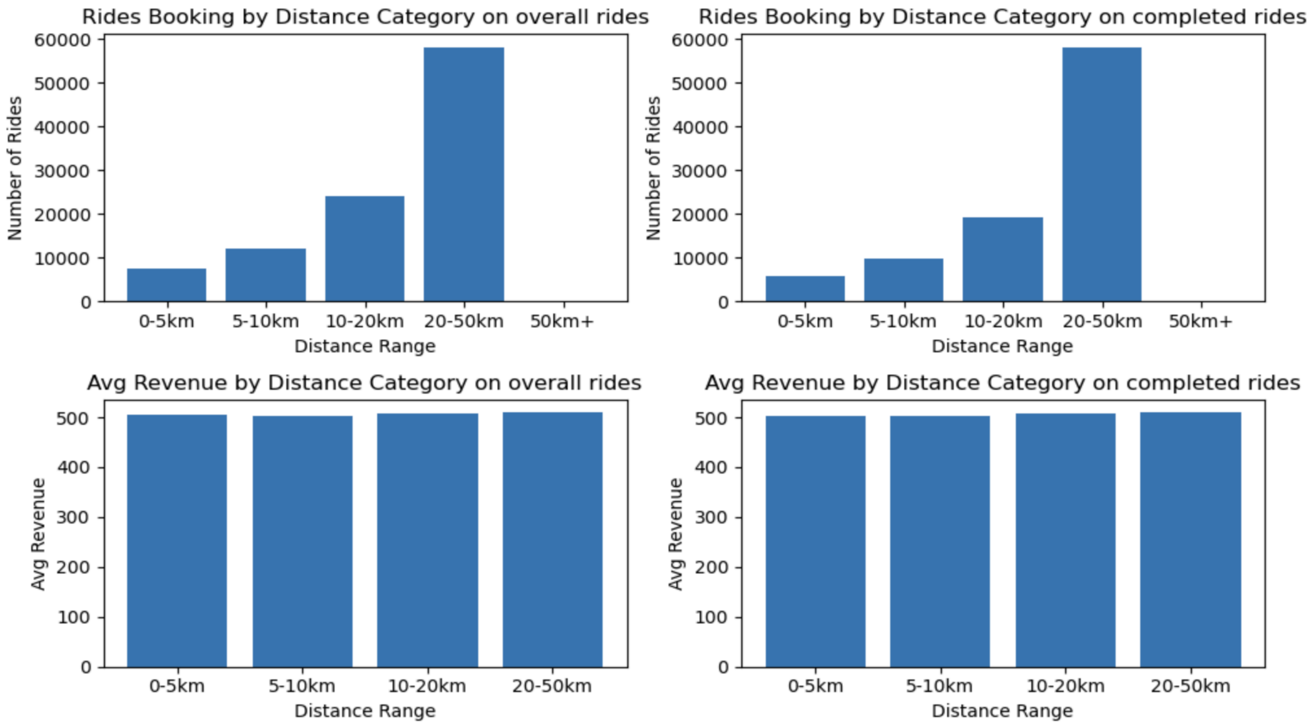


Fig.16. Rides Booking on overall rides booking vs Rides Booking on Completed Rides booking and Average Revenue by Distance Range on overall & Completed Rides

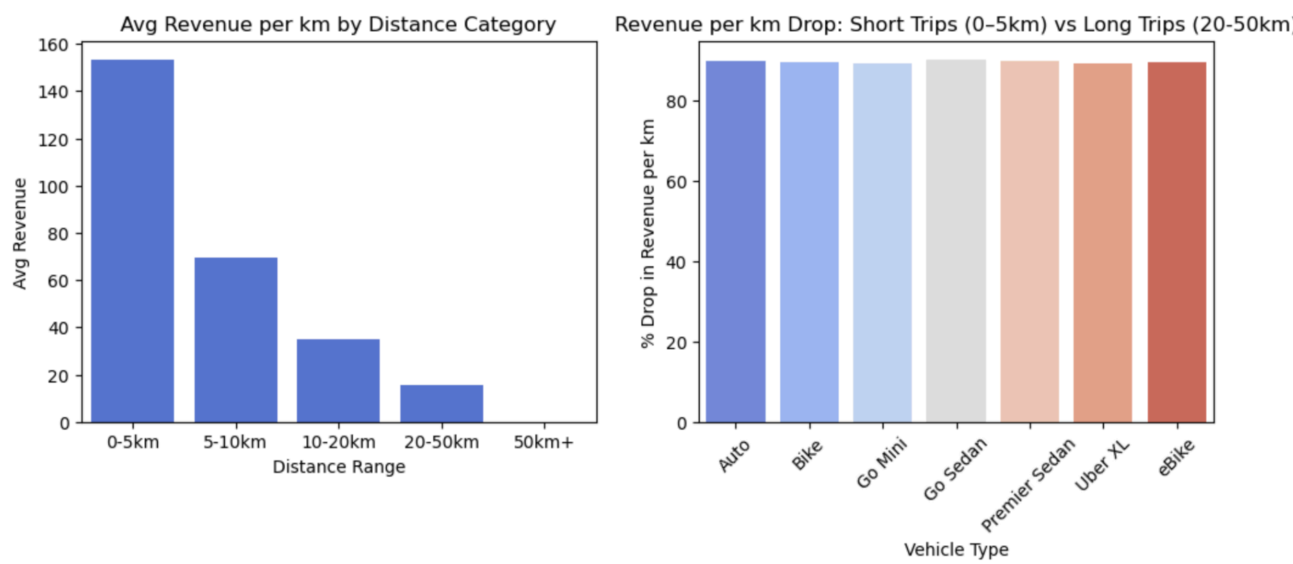
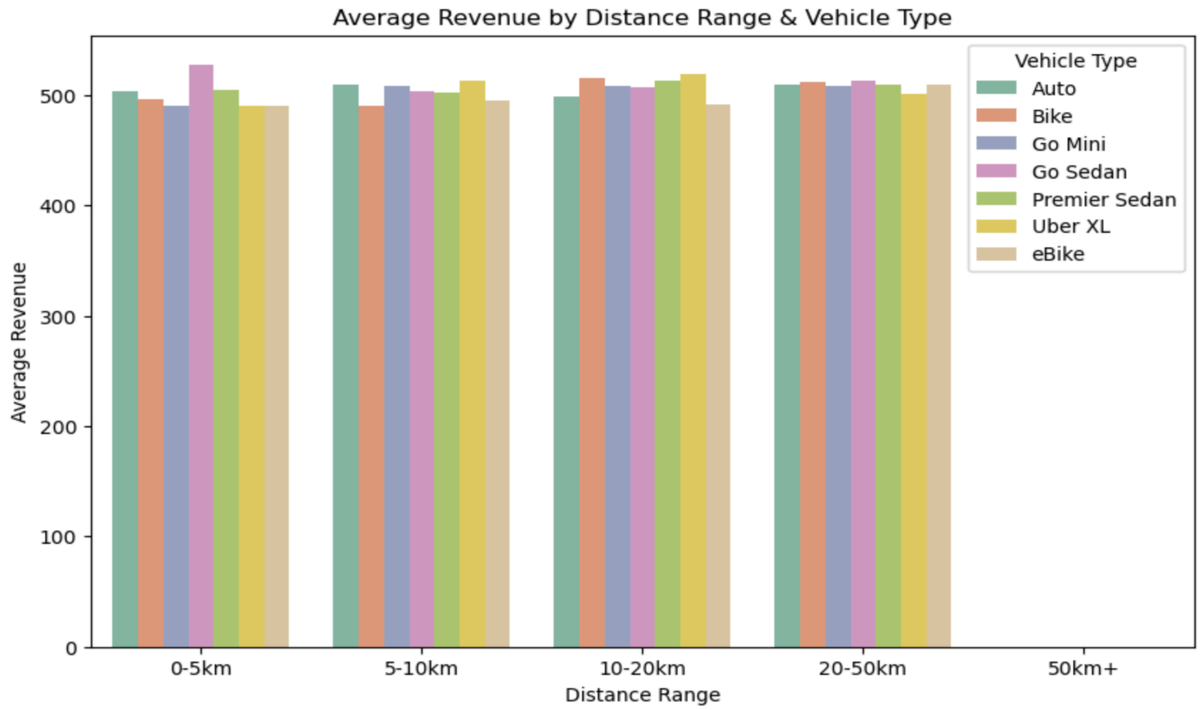


Fig.17. Average Revenue per km by Distance Range of Completed Rides & Revenue per km drop between Short Trips(0-5km) & Long Trips(20-50km) on Completed Rides Booking



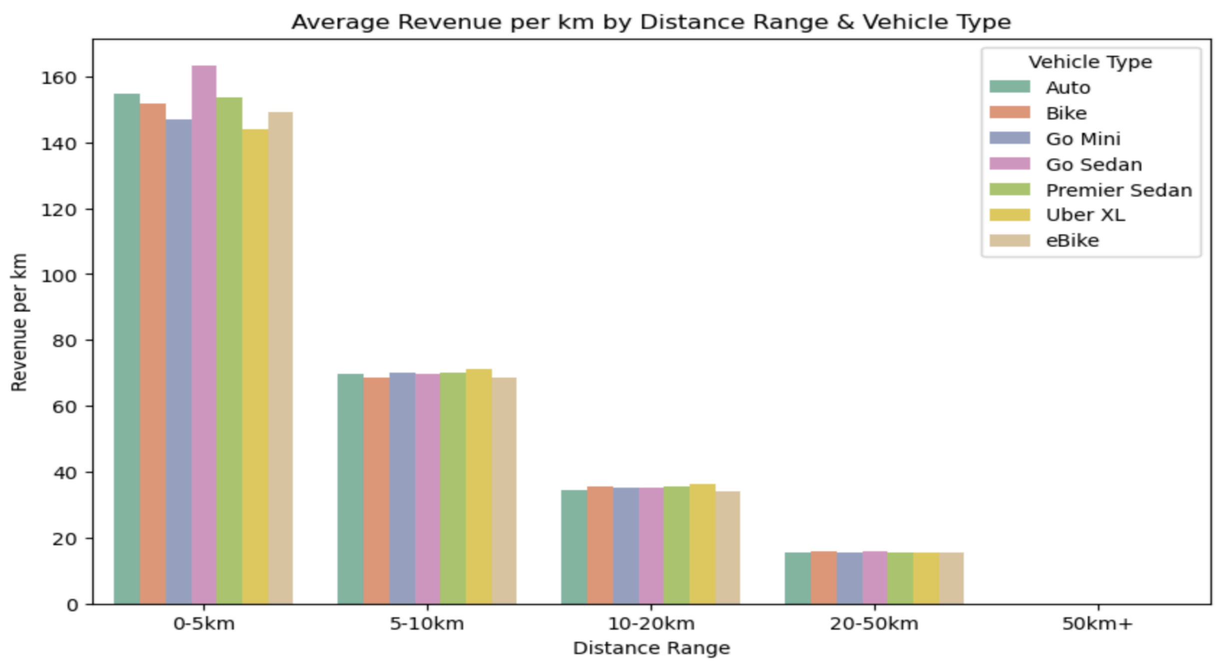


Fig.18. Average Revenue & Average Revenue Per Km by Distance & Vehicle Type of Completed Rides

Ride Frequency Is extremely low as the majority of customers have booked only one ride and very few have taken two or three rides. This confirms **low repeat engagement**, suggesting most users are either one-time or infrequent customers.

Total Spend Distribution Is Heavily Right-Skewed. Most customers have **low total spend**, concentrated below ₹1000. A **small minority of high spenders** contribute more revenue around ₹4000-₹5000.

Cancellations Are Rare for Most Users. The majority of customers have zero cancellations, while a smaller group cancelled one or two rides.

Customer Ratings Are Generally Positive. The average ratings are skewed toward 4–5 ratings, few customers rated below 2. A large number of customers give **no reviews**.

Fig. 19. Customer Behavior Analysis

Highlights & Recommendations

Insight Highlights

In our findings we see that uber ride hailing service is having **62% of ride completion rate**

- Only **about 6 in 10 bookings** are successfully completed - a moderate success rate.
- This indicates **room for improvement in operational reliability** and **customer satisfaction**.

Driver cancellation rate is 18.1% which means

- Driver cancellations alone make up approximately **47.4% of all failed rides** indicating it being the **major contributor towards failed rides**.
- This is a strong signal of potential **driver-side issues** such as poor incentive alignment, potential issues in driver availability, unprofitable trips, or long pickup distances etc.

Customer Cancellations , No Driver Found & Incomplete rides form the rest booking (~20%), suggesting:

- **Customer impatience or change of plans** (e.g., long wait times, high surge pricing, or ETA unreliability etc).
- **Operational gaps** (e.g., inefficient driver distribution or high mismatch between supply and demand).
- **Platform or communication issues** (e.g., ETA mismatches, app glitches, trip disputes, or network/connectivity issues).

The top four reasons for **Customer Cancellation Reasons** – Change of plans, Driver asked to cancel, Driver not moving toward pickup, and Wrong address represent **nearly 90% of all customer cancellations**, which suggest **recurring behavioral and operational issues**. The most striking part is “Driver asked to cancel” and “Driver not moving toward pickup” — both point to **driver behavior issues**, not customer intent. AC not working (11%) signals **vehicle quality concerns** that could affect customer satisfaction.

Four main reasons for **Driver Cancellation Reasons –** Customer-related issues, Too many passengers, Personal/Car issues, and Customer sick/coughing, each make up **~25%** of total driver cancellations showing a **balanced distribution** of issues. However, Customer-related issues and More than permitted people highlight **policy non-compliance or mismatch between driver expectations and customer behavior**. Personal/Car issues suggest a need for **vehicle maintenance or driver availability management**, while Customer sickness shows **health-safety caution** post-pandemic.

Both **Customer** and **Driver cancellations** peak during **commute rush hours** meaning the system experiences the most cancellations when both riders and drivers are the busiest – suggesting operational strain during peak hours. The **top Customer Cancellation Reasons** during peak rush hours**– Wrong Address (24.6%)** is slightly higher than the same reason in overall rides (22.5%) suggesting location inaccuracy, pickup confusion, or dense-traffic area mismatches between map pins and real pickup points. **Customers are more likely to cancel if they can't locate the driver or if the driver struggles to reach the pickup spot during traffic-heavy hours.** The distribution of driver cancellation reasons remains consistent with the overall trend – each around 25%. This suggests **driver behavior and motivation are stable across hours**, but the **volume of cancellations spikes** due to **increased ride requests and urban congestion** at 6 PM.

Analysis of booking data reveals **Khandsa** as the leading pickup location across all ride categories(overall,booked & completed), while **Ashram** is the top drop location. High booking but low completion rates at **Saket, AIIMS, Shivaji Park, and IGI Airport** (**high cancellation rates ≈30%**) are driven by customer and driver cancellations — mainly due to driver delays, incorrect pickup details, and comfort or capacity issues.

Morning(9-11 a.m.) and Evening(3-9 p.m.) time slots show **highest booking activity**, indicating the strongest operational and driver allocation needs. Flat trend across all days of the week indicates **steady demand throughout the week**, with **no strong weekday vs. weekend variation**. A large gap of 4,000–5,000 rides/month not completed (≈33–38% loss potential). Monthly completion efficiency peaks in **March, April, June, and October** due to lower cancellations. The months of **April & June** show lower total bookings but **proportionally higher completion rates**, indicating **operational stability.**

Vehicle breakdown is the top reason for peak incomplete booking hours suggesting fleet maintenance. **Customer demand** is the top reason for peak incomplete booking day. In the monthly trend of incomplete booking we see Vehicle Breakdown, Customer Demand, and Other Issues – all **contributing almost equally.**

The Auto fleet with the highest number of completed rides booking is the backbone of ride completions, showing reliability and high adaptability followed by Go Mini and Go Sedan. Despite being lower-fare rides, **Autos dominate total revenue generation** due to their **high volume of booking.** Go Mini and Go Sedan follow. No single type of vehicle shows extreme deviation in terms of average revenue generation, implying **balanced performance** among vehicle classes. Auto, Go Mini & Go Sedan top the total distance covered by a vehicle followed by others. Average distance is fairly **uniform across all vehicle types**, showing **consistent trip lengths** regardless of category. Distance coverage and revenue generation trends **align closely**, showing that **trip volume (frequency)** rather than **trip length** is the main revenue driver — especially for Autos.

Both overall and completed rides show the **highest ride volume in the 20–50 km range**, indicating strong demand for **medium-to-long-distance trips**. Other distance range rides show stable performance across both datasets. There is **very low customer demand** for extremely long trips(50km+) via this platform, likely due to cost or service preference for intercity alternatives.

Short-distance trips are **far more profitable per km**, whereas long trips yield **lower per-km efficiency.** More than 90% drop in average revenue per km between short and long trips indicates a massive efficiency drop. **A drop in revenue per km is normal** because - 1. Each ride includes a **base fare** (fixed cost) that doesn't scale with distance. 2. **Per km rates** are often **discounted for longer rides**. But a **90%+ drop** is larger than typical and **could indicate an imbalance** in **Uber's current distance-based fare model**. For example – if a 2 km ride earns ₹120 (₹60/km), a 40 km ride may earn ₹800 (₹20/km), which indicates that while long trips contribute more to total revenue, they are less profitable per kilometer. **Short trips are overpriced per km**, while **long trips are underpriced per km.** For drivers, **long trips yield lower returns per km**, which may discourage them from accepting long-distance bookings. While this pricing encourages customers to take longer rides, it can **reduce overall revenue efficiency for Uber.**

Within the 0–5 km segment, **Go Sedan stands out** with a significantly higher contribution towards both average revenue and average revenue per km, reflecting customers' willingness to pay a premium for comfort and reliability even on short rides.

Recommendations

1. Reduce Driver Cancellations –
    - **Driver Incentive Optimization:** Offer performance-based bonuses for maintaining high acceptance and low cancellation rates.
    - **Smart Ride Assignment:** Prioritize assigning rides that are closer to the driver's current location or match preferred trip lengths.
    - **Penalty System:** Introduce gentle penalties or ranking impacts for repeated cancellations.
2. Improve Supply Availability –
    - **Demand Forecasting:** Use historical data to anticipate high-demand zones and proactively alert nearby drivers.
    - **Driver Onboarding:** Recruit and retain more drivers in under-served areas or high-demand time slots.
3. Address Driver-Linked Customer Cancellations by tracking drivers frequently reported for "Driver asked to cancel" or " Driver not moving toward pickup" and retrain or penalize those drivers. Improve pickup algorithms by using real-time driver location and ETA optimization to ensure drivers closer to the customer are assigned.
4. Improve Vehicle Quality Issues by introducing **Vehicle inspection programs** to regularly check AC and basic comfort features – customer comfort affects completion and ratings.
5. Introduce a **penalty or dynamic cancellation fee** for repeated customer-initiated cancellations to encourage commitment and reduce booking drop-offs.
6. Fix Location Accuracy Issues –
    - **Improve Map and GPS Accuracy** – Integrate better geolocation APIs, **Auto-correct pickup pins** by snapping them to the nearest valid road or known pickup point and **Use building or landmark-level geocoding** instead of just street names.

- To reduce confusion and waste less time in hard-to-reach zones, introduce **“Recommended Pickup Points”** – safe, easy-to-find landmarks near the user’s current location. The app can prompt: “Your location may be hard to reach. Would you like to meet at ‘XYZ point’’ (2 mins walk) instead?. This option would be very useful specially during rush hours where the cancellation occurs due to urban congestion.
- **Customer Confirmation & Validation:** Ask customers to **double-check the pickup address** before confirming the ride. App can prompt: “Is this the correct pickup location? Tap to adjust if not.”. If GPS accuracy is poor, ask for **manual address confirmation or landmark input** to reduce location mismatches.
- **Data-Driven Optimization**: Identify locations where “Wrong Address” cancellations occur frequently. Flag these zones. Uber can use this data to correct mapping errors.

7. Handling Rush Hours Cancellation **–**
    - Address Wrong Address issues during rush hours which is likely to occur due to urban congestions.
    - **Driver Navigation Assistance:** Provide alternate route suggestions or “smart pickup routes” to avoid traffic deadlocks during the rush hours.
    - **Dynamic Driver Allocation:** Anticipate the 6 PM(peak cancellation hour) demand surge and **pre-position and incentivize drivers** near high-demand zones before rush hour to manage surge effectively.
    - **Enhance in-app communication** between drivers and customers for smoother coordination at congested pickup points.
8. Improvement of service reliability in **Saket, AIIMS, Shivaji Park, and IGI Airport** through faster driver dispatch, clearer pickup point mapping, vehicle condition checks (AC, cleanliness), and driver training for customer handling.
9. **Driver Incentives**(Completion-based bonus, Low-cancellation reward etc.) to Improve Ride Completion and **Customer Incentives**(Dynamic cancellation fee, grace cancellations, loyalty rewards for low cancellation customers etc) to Reduce Cancellations.
10. Since demand is steady throughout the day of week, driver deployment and marketing campaigns can remain **uniform across days**, focusing instead on **hourly variations** rather than day-based differences. Monthly Cancellation reduction may correlate with improved completion efficiency – operational and environmental factors such as weather, driver availability, seasonal demand etc may influence these patterns. Hence we can increase driver availability during **9–11 a.m.** and **3–9 p.m.** to match peak demand, reducing wait time and cancellations while replicating operational conditions from low-cancellation months.
11. Handling Incomplete Booking:
    - Fleet & Vehicle Management – Schedule **routine checks and servicing** for vehicles. Introduce an **in-app driver vehicle health checklist** (AC, brakes, tires, battery). Auto-flag vehicles with repeated breakdown-related cancellations for **temporary suspension and inspection**.
    - Driver Supply Optimization

- Customer Experience & Demand Management – When demand exceeds supply show **accurate wait times** or **alternate transport options**.

12. Strategic Fleet Focus:
    - Scale Auto Fleet Capacity – Expand the **Auto fleet**, given its strong completion performance, top contribution towards revenue & distance. Prioritize **vehicle maintenance** and **peak-hour allocation** for Autos to sustain performance.
    - Leverage Go Mini & Go Sedan Efficiency – Improve maintenance & availability support for **Go Mini** and **Go Sedan**, which also have stable completion performance and good contribution towards revenue and distance coverage. This will have the **strongest positive impact** on completion and reliability. Offer **priority allocation** in high-demand zones to maximize utilization.
    - Boost eBike & Bike Utilization – Promote **eco-friendly options** for **short-distance urban trips** through customer incentives (e.g., lower fares, reward points) to improve fleet balance. It can reduce "No Driver Found" cases during peak hours while diversifying fleet mix.
    - Revitalize Uber XL – As the **lowest revenue contributor**, Uber XL utilization should be **targeted for specific segments** (airport transfers, group rides etc). Consider **dynamic pricing** or **event-based promotions such as special fare discounts, offers, or incentives tied to specific events, holidays, or local occasions to encourage more bookings** & increase ride frequency.
13. Handing average revenue per km drop magnitude:
    - Re-evaluate long-distance fare scaling and fine-tune base fare and per km rates.
    - Create tiered per km pricing and increase per km rate moderately(e.g., higher rate for first 5 km, moderate for 5–20 km, slightly reduced for 20+ km and so-on).
    - Provide **long-trip completion bonuses** or **distance-based rewards** to drivers which may help sustain driver motivation and prevent "long trip cancellations".
14. Go Sedan demonstrates good revenue generation in short-distance travel compared to other vehicle types. This suggests that customers view it as a comfort-oriented, value-added option. By strategically optimizing pricing, marketing, and driver incentives, Uber can **amplify Go Sedan's profitability.**

⚠️ Limitations & Assumptions

The analysis of Uber ride data is based on certain assumptions regarding data accuracy, pricing consistency, and rider–driver behavior. While every effort was made to ensure accuracy, the study is limited by few constraints. Hence, findings and recommendations should be interpreted within these constraints.

Incomplete Ride Status Ambiguity:

The “Incomplete” booking status in the dataset lacks clarity regarding when the trip was stopped and how fare amounts were determined. As these rides contribute to overall revenue figures, the absence of detailed fare computation or termination reasons may lead to an overestimation of revenue. Incomplete rides reason “Others” is not clear but it equally contributes towards incomplete rides reasons, so we may have problems fixing the incomplete rides with limited clarity.

Absence of Driver/Vehicle Identification:

The dataset lacks unique driver identifiers,limiting the ability to analyze driver-level performance, efficiency, and behavioral patterns. As a result, key operational metrics such as turnaround time, driver utilization rate, and revenue per driver cannot be computed. This also constrains the ability to identify patterns in driver-related cancellations & delays. Consequently, making it impossible to analyze **ride allocation** or detect **uneven workload distribution** among drivers. Due to the absence of the vehicle identifiers we are unable to analyse vehicle related behavior.

Low Customer Ride Frequency:

Uber’s customer base in this dataset is **broad but shallow** — many users, but minimal ride frequency. The dataset contains customers with very few ride bookings — most users have ridden once, and even the top ten customers booked only 2–3 rides. This limits the ability to derive meaningfully accurate insights into customer loyalty, retention, and behavioral trends(e.g., preferred ride time, location patterns, cancellation habits, or loyalty indicators). As a result, customer-level analyses (such as repeat booking patterns or high-value user identification) may not fully represent actual rider behavior or engagement over time.

To ensure interpretability of the analysis, several assumptions were made regarding data completeness & accuracy and pricing structure. It is assumed that the provided dataset accurately represents the total population of Uber rides during the analysis period. In the absence of fare breakdowns (base fare, surge, waiting time), total fare is assumed to be a **single consistent revenue indicator** across all ride types. Vehicle types are assumed to operate under **standardized Uber pricing logic** (e.g., Go Sedan always has higher base fare than Auto). Driver incentive structure is not considered in analysis. Despite many customers having only one or two rides, it is assumed that the dataset still provides a **valid cross-section** of Uber’s urban customer base.