Uber Operations in Mumbai – Analysis Report

Executive Summary

This project comprehensively analyses Uber ride data in Mumbai, India. The goal was to uncover operational insights from one month of trip records using a combination of SQL, Excel, and Power BI. Key findings include identification of peak demand hours (morning rush and late evening), the busiest pickup areas in the city, patterns in trip cancellations, and revenue contributions by different ride categories. The analysis highlights important business insights, such as a high overall cancellation rate (38% of requests did not result in a completed ride) and the dominance of a few neighbourhoods in ride demand. By integrating data processing and visualization, the project demonstrates how data-driven decisions can be made to improve ride-hailing services, driver allocation, and customer experience.

Tools Used

- **SQL:** Employed for data cleaning and transformation. Large raw datasets were loaded into a SQL environment where queries cleaned inconsistencies (e.g., replacing placeholder values like ""No"" with nulls or 0), combined date and time fields for time-series analysis, and aggregated trip data (by hour, location, etc.) for further study.
- **Excel:** Used for initial data inspection and minor preprocessing. The dataset (an Excel file) was examined using Excel to understand its structure and contents. Excel formulas and filtering helped catch obvious data issues (such as blank rows or misspelt location names) and ensure the data types (dates, numbers) were correctly formatted before importing into Power BI.
- **Power BI:** Utilized for visualization and interactive analysis. The cleaned data was loaded into Power BI to create dashboards showing trends and patterns. Power BI's features (like slicers and drill-downs) allowed for data exploration by different dimensions (date, time, location, ride type). Visual charts (line graphs, bar charts, pie charts, maps) were created to communicate insights clearly to stakeholders.

Data Overview

- **Dataset:** The analysis is based on an Excel dataset of 40,539 Uber bookings in Mumbai spanning one month (July 1–30, 2024). Each record represents a ride booking request and its outcome.
- **Fields:** There are 19 columns capturing details of each booking. Key fields include date and time of the request, unique booking ID, booking status (e.g. *Success*, *Cancelled by Driver*, *Cancelled by Customer*, *Driver Not Found*), customer ID, vehicle type (e.g. Uber X, Uber XL, Comfort, Black), pickup and drop-off locations, travel times, cancellation reasons, fare amount (Booking_Value), payment method, ride distance, and driver/customer ratings.
- **Data Integrity:** The data required some cleaning. For instance, some numeric fields used "No" as a placeholder when a trip did not occur (such as no travel time or fare for cancelled trips). These were interpreted as zero or null values. A few blank rows were present and removed. Overall, the dataset contained rich information for both completed trips and those that were not completed, enabling analysis of demand and supply-side issues.

- **Summary Statistics:** In the month of data, about 25,205 trips (62%) were completed successfully, while the rest were not completed for several reasons (driver or rider cancellations or no driver available). The total recorded revenue from completed trips was approximately ₹9.45 million for the month, from ~25k successful rides (average fare ~₹375 per trip). The rides covered an average distance of ~21 km. Four service categories were available (Uber X, Uber XL, Comfort, and Uber Black), each comprising roughly 10k ride requests in the dataset.

Data Processing and SQL Logic

Data processing was an essential step to prepare for analysis. The raw data from Excel was first cleaned and structured using SQL logic as follows:

- **Cleaning and Filtering:** Using SQL, we filtered out invalid or empty records. For example, blank entries (where most fields were null due to spreadsheet artefacts) were removed via a DELETE or filtering in a SELECT query. Fields containing the text "No" (for cases where data was not applicable) were cleaned by converting them to null or zero. This applied to travel time (V_TAT, C_TAT), cancellation counts, and payment method.

- **Type Conversion:** Date and time fields were combined to form a proper datetime. In SQL, functions like STR_TO_DATE or direct concatenation (depending on the SQL dialect) merge the separate Date and Time columns into a single timestamp. This allowed the extraction of components like hours of day or days of week directly in queries (e.g., using HOUR(timestamp) to get booking hours for peak hour analysis).

- **Derived Columns:** Additional fields were computed to facilitate analysis. For instance, a field for trip outcome was derived to indicate whether a request resulted in a completed ride or a specific cancellation type, making it easier to aggregate outcomes. Similarly, a day-of-week field was added (using SQL date functions) to analyze weekly patterns.

- **Aggregation and Joins:** The dataset was mostly a single table of bookings, so explicit joins were not needed for primary analysis. However, SQL was used to create aggregated views. For example, we wrote queries to get the number of bookings per hour, day, and location. These summary tables (or directly the query outputs) were later used in Power BI to create visuals. In cases where location names had minor inconsistencies (e.g., abbreviations or misspellings), we used SQL (and Excel) to standardize them so that, for example, *""Indira Nagar"* would be treated uniformly.

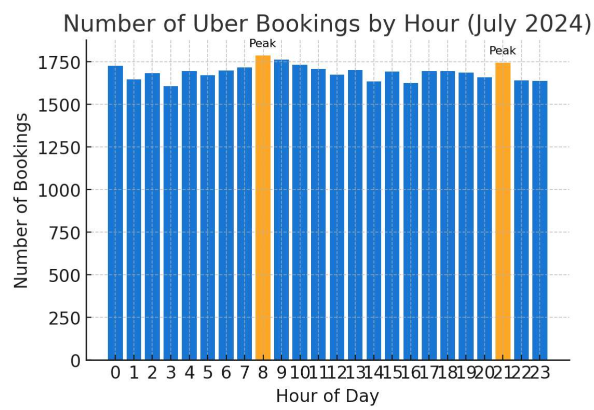
- **Quality Assurance:** Throughout the SQL processing, checks were performed to ensure data consistency. We compared the total counts before and after cleaning to ensure no valid data was lost, and we verified key statistics (such as total successful rides and total revenue) remained logical after transformation. This ensured that the subsequent analysis in Power BI was based on accurate data.

Visual Analysis

After cleaning the data, we used Power BI to create visualizations that reveal important patterns in Uber's Mumbai operations. The following visual insights were derived:

Booking Trends & Demand Patterns: The volume of Uber bookings in Mumbai remained steady day-to-day during the month. On average, there were about 1,350 ride requests daily, with minor fluctuations. Weekdays showed slightly higher demand compared to weekends. For instance, Tuesdays and Mondays saw more rides (since there were 5 Tuesdays/Mondays in the month), while Sundays slightly declined requests. Overall, there was not a strong upward or downward trend over the month – indicating a stable demand period without large events or seasonal effects in July.

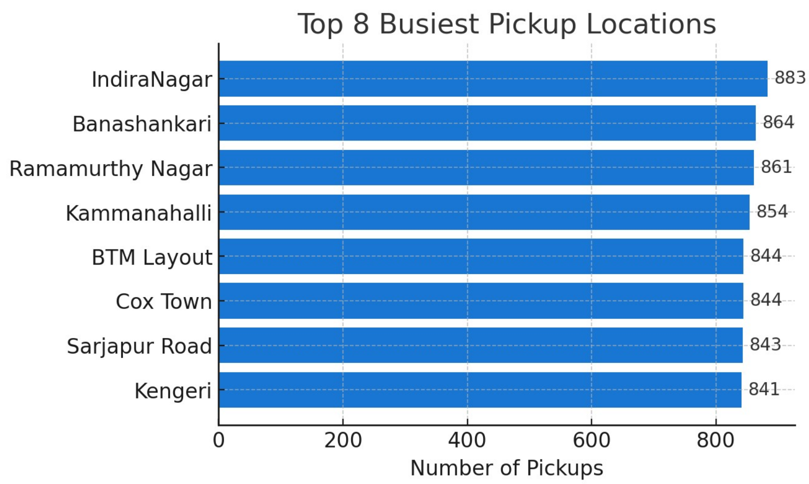
Peak Hours:



Hourly distribution of Uber ride requests in Mumbai. The chart highlights peak demand hours around 8:00 AM and around 9:00 PM.

Analyzing rides by time of day showed distinct peak periods. As depicted above, ride requests start picking up early and **peak around 8 AM**, coinciding with the morning rush hour when people commute for work or education. Demand stays relatively high through the late morning and afternoon, with a mild mid-day dip. In the evening, there is another **surge in rides around 8–9 PM**, reflecting commuters returning home and people going out for dinner or leisure. Late-night hours (after midnight) see lower activity, bottoming out around 3–4 AM when the city is quiet. However, even the lowest hour (around 3 AM) still had substantial requests (over 1,600), indicating that Mumbai's Uber usage continues through the night at a moderate level. The consistent demand throughout the day suggests Uber is a relied-upon mode of transport at all hours, with only relatively modest variation between off-peak and peak times.

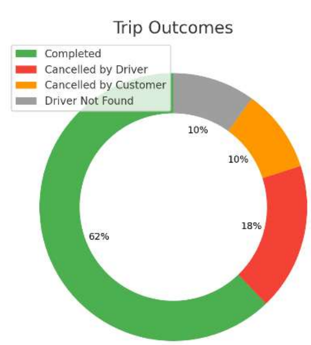
Location Analysis:



Top 8 busiest pickup locations (neighbourhoods) for Uber in the city by the number of pickups in the month.

Certain neighbourhoods emerge as Uber **hotspots**. The bar chart above shows the top pick up locations: **Indira Nagar, Banashankari, Ramamurthy Nagar,** and **Kammanahalli** ranked highest in ride request pickups (each with around 840–880 pickups in the month). These localities are known to be major residential and commercial hubs, which explains the high demand for rides. Other areas in the top eight include **BTM Layout, Cox Town, Sarjapur Road,** and **Kengeri**, all significant parts of the city's geography. This location analysis suggests that Uber usage is concentrated in key neighbourhoods, potentially where population density and commute needs are highest. In terms of planning, Uber could focus driver positioning around these hotspots during peak times to reduce wait times and cancellations. Notably, many of these top areas are well-known for traffic congestion (e.g., Sarjapur Road), which could tie into why so many rides are requested there (people prefer to ride rather than drive themselves in heavy traffic).

Booking Outcomes:



Breakdown of trip outcomes for all booking requests (displayed as a percentage of total requests). Only 62% of requests led to completed rides.

Not every Uber booking request was fulfilled. A sizeable portion of ride requests did not convert into completed trips. The doughnut chart shows 62% **of bookings ended in a successful ride**, while the remaining **38% were unfulfilled**. The unfulfilled requests comprise several scenarios: **18% of all requests were cancelled by drivers** (often last-minute driver cancellations), **10% were cancelled by customers** (riders cancelling, perhaps due to long wait times or change of plans), and about **10% resulted in "Driver Not Found"**, meaning no driver was available to accept the ride. This high cancellation and no-show rate is an important operational insight – it indicates a potential supply-demand mismatch, especially during certain times or locations. Further analysis (not fully shown in this summary) revealed that many "Driver Not Found" incidents occurred during peak hours in outlying areas, suggesting that driver supply was stretched thin. Reducing this 38% failure rate could improve customer satisfaction and increase Uber's revenue; strategies might include incentivizing drivers to serve high-demand time slots or improving the dispatch algorithm to cover underserved neighbourhoods.

Revenue and Service Category Metrics: Different Uber service types show varying performance in terms of revenue. While each service (Uber X, Uber XL, Comfort, and Uber Black) had a similar share of trip requests (25% each), their contribution to revenue differed. **Uber Comfort** rides, for example, generated the highest total revenue (over 3.3 million INR in the month) despite a similar number of trips as other categories. This is because Comfort rides had a much higher average fare per trip (₹500+ on average, due to longer distances or higher base fares). Uber Black, another premium category, also earned a significant revenue share (₹2.76M).

In contrast, Uber X's budget category had the lowest average fare (₹212 per trip) and, consequently, the smallest revenue share (₹1.34M), even though its trip count was comparable. This insight highlights how premium services, while fewer in number, can drive a disproportionate share of income. Uber may benefit from promoting higher-end services in affluent areas or during peak times to maximize revenue. Additionally, we observed that **payment methods** favoured cash and UPI (digital payments). Around half of the completed trips were paid in **cash**, 40% via **UPI**, with only a small minority using credit/debit cards. This mirrors typical consumer behaviour in Indian cities and suggests that any promotions or loyalty programs might be more effective if integrated with UPI or cash-back offers, considering the usage patterns.

Key Findings

- **High Cancellation Rate:** Only about **62% of ride requests led to completed trips**, indicating that more than a third of all Uber bookings did not materialize into actual rides. Driver-initiated cancellations (18% of all requests) were a major factor, pointing to driver availability or willingness issues. Customer cancellations (10%) and lack of available drivers (10% no-shows) also contributed significantly. This highlights a reliability challenge in the service – a gap between user demand and Uber's ability to fulfil it consistently.
- **Consistent Demand with Peak Hours:** Uber rides were in demand throughout the day, with the city averaging ~1.3–1.4k ride requests daily without huge fluctuations. However, **peak usage hours** were observed in the **morning around 8 AM** and **the evening around 8-9 PM**. These peaks coincide with typical commute times, confirming that work/school rush hours drive the highest demand. Late nights (2-4 AM) saw the lowest demand, but the service was used even then. The even distribution of rides across hours suggests a continuous need for transport beyond rush hours, likely due to the city's vibrant 24/7 economy.
- **Busiest Areas:** The analysis identified **key hotspots in the city** where ride requests were highest. Neighbourhoods like Indira Nagar, Banashankari, Ramamurthy Nagar, and Kammanahalli topped the list of pickup locations. These densely populated areas have many workplaces, eateries, and shopping centres, making them central to Uber's operations. High demand in these zones implies that ensuring a strong supply of drivers there (especially during rush hours) is critical. It also suggests potential focus areas for Uber marketing campaigns or partner programs (for example, targeted driver incentives or rider promotions in those localities).
- **Revenue Drivers:** Premium ride categories contributed a large share of revenue. In particular, **Uber Comfort** stood out, contributing the highest total revenue among the four categories, thanks to its higher fare per ride. This indicates a segment of customers willing to pay more for comfort or luxury or take longer trips. Standard Uber X rides, while numerous, brought in less revenue per trip. This insight could guide Uber to balance its fleet, catering to the volume of Uber X users while not neglecting the higher-end market that boosts profitability. Additionally, the **payment method split** showed that most riders paid in cash or through UPI, with very few using cards, reflecting local user preferences.
- **City-specific Behaviour:** Some patterns observed are particular to the local context of Mumbai (or similar urban Indian environments). The prevalence of last-mile connectivity needs in the suburbs (e.g., Kengeri and Sarjapur Road areas) drives Uber usage. The high cancellation rate by drivers might be influenced by traffic congestion (drivers may cancel if a pickup is too far or stuck in traffic) or low fares for long distances (making some trips less attractive to drivers). The strong preference for cash/UPI payments underscores the importance of accommodating local payment habits. Understanding these behaviours is crucial for Uber to tailor its strategies (such as adjusting surge pricing, adding incentives for drivers in traffic-prone areas, or ensuring the app experience remains smooth for cash payments).

Technical Highlights

- **Data Cleaning Challenges:** The raw dataset had inconsistent placeholder values (e.g., "No"") for numerical fields (like travel time or fare when a trip did not occur). This required careful cleaning using Excel and SQL. We addressed it by replacing "No" with 0 for numeric analyses (assuming "No ride = 0 distance/fare"") or treating them as null where appropriate. Excel's find-and-replace, and filtering features helped identify these anomalies, and SQL CASE expressions were used to clean them in bulk systematically.
- **Date-Time Processing:** Separating date and time into two columns made time-series analysis cumbersome. We overcame this by combining them into a single **datetime** field. In SQL, functions were used to parse and merge the fields so that we could easily extract hour-of-day and day-of-week. We also created a calendar table in Power BI to facilitate time-based slicing (e.g., weekday vs weekend performance).
- **Location Data Standardization:** There were minor spelling variants in location names (for example, "IndiraNogar" vs "Indira Nagar"). This could have affected aggregation and map visualizations. It was resolved by standardizing names in Excel (using simple find/replace operations) and verifying the uniqueness of locations. This ensured that all rides from a given locality were counted together. It was a straightforward but essential step to avoid fragmented data insights.
- **Volume and Performance:** With over 40k records, the data volume was moderate. Power BI handled this size well, but we optimized performance using SQL to pre-aggregate some data. For instance, instead of importing all raw data and doing heavy calculations in DAX, we ran SQL queries to get daily ride counts, hourly distributions, etc., using those as the basis for visuals. This approach reduced the load on Power BI and sped up the creation of interactive dashboards.
- **Visualization and DAX:** Crafting the right visuals in Power BI sometimes requires creating custom measures. For example, to compute **the completion rate** or **cancellation percentage**, we wrote DAX measures that calculated these values on the fly, allowing them to be displayed in cards or tooltips. A challenge was ensuring these measures responded correctly to filters (e.g., showing cancellation rate by day or vehicle type). We tested the dashboard extensively, applying filters for dates, locations, and vehicle types to verify that all measures and charts were updated accurately and intuitively.
- **Integration of Tools:** A key learning was the seamless integration of tools – using Excel for a quick preview, SQL for heavy lifting in data prep, and Power BI for visualization. Each tool had strengths that were leveraged: Excel for its ease of manual data correction, SQL for reliable and repeatable transformations, and Power BI for its ability to bring end-user data to life. Ensuring consistency across these (for example, providing the data cleaned in SQL was the same version used in Power BI) was a technical focus area. We maintained version control on the dataset to avoid confusion as adjustments were made.

Conclusion and Learnings

This Uber operations analysis project provided valuable insights into the ride-hailing business and the data analysis process. From a **business perspective**, we learned that Mumbai's Uber demand is robust and fairly steady, with identifiable peaks and hotspots that can inform operational decisions. High cancellation rates signal an area for improvement – for instance, better driver allocation or policy changes to reduce last-minute cancellations. Identifying top-demand areas and times means Uber can proactively position drivers or run promotions during those crucial windows. We also saw the importance of catering to local preferences, such as popular payment methods and the mix of service types that rider's favour.

From a **technical and personal development perspective**, the project was a comprehensive exercise using multiple tools. The user (analyst) gained experience cleaning data using SQL and Excel, ensuring the dataset was reliable for analysis. They learned how to create calculated fields and use SQL queries to answer specific questions (like ""when are rides most likely to be cancelled?"" or ""Which areas generate the most revenue?""). In Power BI, the user honed skills in designing intuitive dashboards — choosing the right chart types and crafting an engaging narrative through visuals. Significant learning outcomes were overcoming challenges such as data inconsistencies and making the visuals interact dynamically.

In **conclusion**, the project yielded actionable insights for Uber's operations (like focusing on driver availability during peak hours in Indira Nagar or improving the success rate of bookings) and enhanced the user's analytical capabilities. The learnings from this analysis lay a solid foundation for future projects. The next steps include expanding the analysis to multiple months to observe trends or incorporating external factors (weather, holidays, major events) to see their impact on Uber usage. Furthermore, these insights could be used to develop predictive models – for example, forecasting demand in various areas or predicting the likelihood of cancellation for a given trip request. Overall, the project demonstrates the power of data-driven decision-making in improving urban mobility services and the importance of a well-rounded skill set to execute such an analysis from end to end.