# 

# 

# 

# 

# 

# 

# Ride Booking Dataset — Detailed Analysis Report

#### Power BI Case Study

# Contents

1. Executive Summary

2. 1. Project Background & Objectives

3. 2. Data & Methods

4. 3. Descriptive Findings

5. 4. Customer Behaviour Insights

6. 5. Driver Performance Evaluation

7. 6. Operational Metrics

8. 7. Correlation & Statistical Notes

9. 8. Recommendations & Next Steps (Assistant additions)

10. 9. Limitations

11. 10. Appendix: DAX & Power Query snippets, Glossary

# Summary

This detailed report consolidates findings from a Power BI analysis of a ride booking dataset containing 150,000 records and 21 columns. It covers data quality, cleaning, descriptive statistics, customer and driver behaviour, operational metrics (VTAT/CTAT), cancellation patterns, peak demand analysis, and recommendations for operational and analytical improvements.

Key highlights:  
• Strong representation of 'Auto' bookings (37,419) and lowest volume for 'Uber XL' (4,449).  
• Average booking value is ₹508.30 and average ride distance is 24.64 km.  
• Cancellation is material: drivers cancelled ~27,000 rides (dataset-level number provided by source).  
• Cancellation peaks at ~10:00 AM and ~6:00 PM, with Monday being the highest-cancellation weekday.  
• No obvious correlation observed between booking value, distance and customer satisfaction in the provided analysis.

# 1. Project Background & Objectives

Project: Analysis of Ride Booking Dataset (Power BI)  
  
Primary objectives:  
• Clean and standardize the dataset for reliable reporting.  
• Produce descriptive analysis for vehicle types, distances, booking value and ratings.  
• Surface customer cancellation behaviour and driver cancellations.  
• Evaluate operational KPIs: VTAT (vehicle turnaround time) and CTAT (customer turnaround time).  
• Identify peak demand locations and time slots to inform capacity planning.

# 2. Data & Methods

**2.1 Data source and schema**

The dataset contains 150,000 booking records and 21 columns including Date, Time, Booking ID, Booking Status, Customer ID, Vehicle Type, Pickup/Drop Location, VTAT, CTAT, cancellation flags + reasons, booking value, ride distance, ratings, and payment method.

**2.2 Data cleaning & transformation (Power Query steps)**

• Removed blank rows, removed rows with Power Query errors, and removed duplicate bookings.

• Replaced nulls in binary/flag columns (Cancelled Rides by Customer, Cancelled Rides by Driver, Incomplete Rides) with 0 to create clean binary flags.

• Converted numeric types where needed (e.g., changed decimal numbers to whole numbers for flags).

• Trimmed whitespace on text columns and standardized case for IDs (uppercased Booking ID & Customer ID) so 'ABC1234' and 'abc1234' are treated the same.

• Merged Date + Time into a single DateTime column and changed its type to Date/Time.

• Verified and corrected data types across all columns (dates, numbers, text).

**2.3 DAX measures and derived columns (examples used):**

• Day Name = FORMAT('Table'[DateTime], "dddd") -- weekday names (Monday..Sunday)

• HourOfDay = HOUR('Table'[DateTime]) -- integer 0..23

• Total Cancellation (measure) = SUM('Table'[Cancelled Rides by Customer]) + SUM('Table'[Cancelled Rides by Driver]) + SUM('Table'[Incomplete Rides])

**2.4 Notes on methodology and assumptions**

• The analysis uses the cleaned dataset exported from Power BI. All aggregates and charts referenced were produced inside the Power BI model.  
• When the same location appears with slightly different text (e.g., spelling/casing), it may create duplicate location buckets; standardization is recommended.  
• Driver-level performance metrics cannot be computed without a driver identifier in the source dataset — this is a key limitation.

# 3. Descriptive Findings

## 3.1 Vehicle type popularity

Most and least popular vehicle types (absolute counts provided by source analysis):

| Vehicle Type | Bookings (count) | Notes |
| --- | --- | --- |
| Auto | 37,419 | Most popular |
| Uber XL | 4,449 | Least popular |
| Premier Sedan | — | volume not provided in source summary |

Visualization used: Bar chart (ranked by booking counts).

## 3.2 Average ride metrics

• Average booking value: ₹508.3  
• Average ride distance: 24.64 km  
Visualization: KPI cards in Power BI.

## 3.3 Rating distributions

Rating bins (counts provided):

| Rating Bin | Count (Driver & Customer combined where noted) |
| --- | --- |
| 4.0 | 43,901 |
| 4.5 | 25,713 |
| 5.0 | 2,365 |
| 3.5 | 14,327 |
| 3.0 | 6,694 |

Visualization used: Clustered column chart with rating bins. Note: ensure bins match rating granularity (0.5 steps).

## 3.4 Cancellation reasons (top categories)

| By | Top Reason | Count |
| --- | --- | --- |
| Customer | Wrong address | 2,362 |
| Driver | Customer Related Issue | 6,837 |

Visualization used: Two separate bar charts for customer vs driver cancellation reasons.

# 4. Customer Behaviour Insights

4.1 Frequent cancellers (by Customer ID)

• CID5481002 — 3 cancellations  
• CID6715450 — 3 cancellations  
Visualization used: Table showing Customer ID, total bookings, cancellations, cancellation rate.

4.2 Time-of-day & day-of-week cancellation patterns

• Peak cancellation times: ~10:00 AM and ~6:00 PM.  
• Peak cancellation day: Monday.  
Visualization used: Line chart (hourly cancellations) and bar chart (weekday totals).

4.3 Behavioural insights and operational recommendations (assistant additions)

• Implement a pre-ride confirmation SMS/push for bookings in the 9–11 AM and 5–7 PM windows to reduce no-shows and address-related cancellations.

• Flag customers with repeated cancellations and route them to a manual review or temporary hold with a short message explaining business rules.

• Add cancellation reason standardized categories and enforce structured reason reporting (dropdown) in driver and customer apps to reduce free-text variability.

• Introduce lightweight friction (e.g., small in-app deposit, penalty credit) only for repeat offenders — use A/B testing before wide rollout.

# 5. Driver Performance Evaluation

5.1 Per-driver ratings:

• Not available: driver identifiers were not present in the dataset, therefore individual driver ratings and acceptance rates cannot be computed. Collecting a unique Driver ID is recommended for future driver performance analysis.

5.2 Driver cancellations

• Total rides cancelled by drivers (from source): 27,000  
• Most common driver cancellation reason: Customer Related Issue — 6,837

Operational recommendations (assistant additions):

• Track driver-level metrics: acceptance rate, cancellations per 100 rides, time-to-accept, and average on-trip time.

• Introduce targeted incentives in low-supply windows/locations to reduce 'No driver found' events (e.g., guaranteed hourly earnings for 5–7 PM shifts).

• Provide drivers with clearer structured cancellation reason options to improve root-cause analysis.

# 6. Operational Metrics

6.1 VTAT & CTAT by vehicle type (averages provided)

| Vehicle Type | Avg VTAT (mins) | Avg CTAT (mins) |
| --- | --- | --- |
| Premier Sedan | 8.66 | 29.83 |
| Go Mini | 8.16 | 29.72 |
| Uber XL | 9.50 | 29.60 |
| eBike | 8.47 | 28.98 |
| Bike | 8.40 | 28.97 |
| Go Sedan | 8.69 | 28.87 |
| Auto | 8.43 | 28.72 |

Top locations (source): Vishwavidyalaya, Vishwavidyalaya (duplicate entry), Yamuna Bank.  
Note: duplicate naming indicates the need for location standardization.

## 6.2 Peak demand locations & time slots

• Peak locations: Adarsh Nagar, AIIMS, Akshardham  
• Peak time slots: morning around 10:00 AM; evening 5:00–7:00 PM  
Visualization used: Matrix (Pickup Location × Hour).

## 6.3 Booking status snapshots (morning vs evening)

| Metric (approx time) | Morning (~10:00 AM) | Evening (~6:00 PM) |
| --- | --- | --- |
| Completed | 5,983 | 7,617 |
| Incomplete | 562 | 743 |
| No driver found | 647 | 891 |
| Cancelled by driver | 1,715 | 2,257 |
| Cancelled by customer | 670 | 889 |

# 7. Correlation & Statistical Notes

The existing analysis reports 'no correlation' between ride value, ride distance and customer satisfaction. To validate and quantify this properly, the recommended approach is:  
• Compute Pearson correlation for linear relationships and Spearman rank correlation to capture monotonic relationships.  
• Visualize with scatter plots and add LOESS/smoothed trend lines to detect non-linear patterns.  
• Control for confounders by segmenting (e.g., by vehicle type, location, and time of day) before correlational analysis.

Assistant addition: When performing correlation tests, report both correlation coefficients and p-values, and consider effect sizes — statistically insignificant correlations can still be operationally meaningful in large samples.

# 8. Recommendations & Next Steps

Below are prioritized opportunities split into Quick Wins, Medium-term, and Strategic initiatives. (No time estimates provided; prioritize based on business needs.)

## Quick Wins

• Standardize location naming using a reference table / fuzzy matching to collapse duplicate location entries.

• Enforce structured cancellation reasons (select lists) in driver/customer apps and backfill mappings for existing free-text reasons.

• Create KPI tiles in Power BI for: Daily cancellations, No-driver-found rate, Completed rides, and Active driver supply.

• Add a small 'Confirm pickup address' step in the app flow for bookings during peak cancellation hours.

## Medium-term

• Instrument driver IDs in the dataset and create a driver performance dashboard (acceptance, cancellations, rating trends).

• Build a cancellation risk model (binary classification) using features such as customer history, time of day, pickup location, booking value, and vehicle type.

• Set up an alerting pipeline (Power BI data alerts or Power Automate) for sudden spikes in cancellations or 'no driver found' events.

• Segment customers by lifetime cancellations & booking frequency, and design retention tactics for high-value customers.

## Strategic initiatives

• Develop a supply-demand forecasting model (time series + external features) to proactively incentivize drivers.

• Run experiments (A/B tests) for interventions: pre-ride confirmations, micro-deposits, or targeted driver incentives to measure impact on cancellations.

• Implement a real-time driver dispatch optimization layer to minimize 'no-driver-found' and reduce pickup delays.

KPIs to monitor (suggested):  
• Cancellation rate (by driver/customer)  
• No-driver-found rate  
• Completed ride conversion rate  
• Average VTAT/CTAT  
• Driver acceptance rate  
• Customer repeat booking rate  
• Average booking value by segment

# 9. Limitations

• Driver identifiers are missing — limits ability to compute per-driver KPIs and attribute cancellations to drivers specifically.

• Potential duplicate/noisy location labels — affects location-level aggregates and peak-location identification.

• Free-text cancellation reasons increase classification noise; structured reasons reduce ambiguity.

• Reported 'no correlation' statement should be validated statistically with correlation coefficients and significance testing.

# 10. Appendix

## 10.1 Example DAX measures (reference)

Day Name: Day Name = FORMAT('Bookings'[DateTime], "dddd")

HourOfDay: HourOfDay = HOUR('Bookings'[DateTime])

Total Cancellation: Total Cancellation = SUM('Bookings'[Cancelled Rides by Customer]) + SUM('Bookings'[Cancelled Rides by Driver]) + SUM('Bookings'[Incomplete Rides])

Cancellation Rate: Cancellation Rate = DIVIDE([Total Cancellation], COUNTROWS('Bookings'), 0)

## 10.2 Power Query (Transform) snippets (conceptual)

• 1) Remove blank rows: Table.SelectRows(source, each List.NonNullCount(Record.FieldValues(\_)) > 0)

• 2) Replace nulls for flags: Table.ReplaceValue(..., null, 0, Replacer.ReplaceValue, {"Cancelled Rides by Customer","Cancelled Rides by Driver","Incomplete Rides"})

• 3) Trim text: Table.TransformColumns(...,{{"Pickup Location", Text.Trim}, {"Drop Location", Text.Trim}})

• 4) Uppercase IDs: Table.TransformColumns(..., {{"Customer ID", Text.Upper}, {"Booking ID", Text.Upper}})

## 10.3 Glossary

• VTAT: Vehicle Turnaround Time — time for vehicle readiness between rides (min).

• CTAT: Customer Turnaround Time — time related to customer readiness or handover (min).