🚖 Ride Booking Dataset – Project Report

1**. Introduction**

**T**his report provides a detailed analysis of the Ride Booking Dataset, which contains 150,000 records and 21 columns. The dataset captures a wide range of ride-related information, including:  - Booking status - Customer IDs - Vehicle types - Pickup and drop locations - Cancellation details (by both customers and drivers) - Ride distances and booking values - Customer and driver ratings - Payment methods  The goal of this report is to uncover operational insights, customer behavior trends, driver performance indicators, and cancellation patterns that can help improve business decisions and customer satisfaction.

2**. Data Cleaning & Preprocessing**

**S**everal steps were undertaken to ensure data quality and consistency:  • Date/Time Conversion  - Combined Date and Time columns into a single Date\_Time field.  - Extracted Hour and Day of Week for time-based analysis.  • Standardization of Categorical Data  - Converted text fields (e.g., Booking Status, Vehicle Type, Payment Method) into lowercase and trimmed spaces.  • Handling Missing Values  - Numerical fields without outliers (e.g., Avg VTAT, Avg CTAT, Ride Distance) → filled with mean.  - Numerical fields with outliers (e.g., Booking Value, Ratings) → filled with median.  - Categorical fields → missing values replaced with 'NA'.  - Binary flag fields (Cancelled Rides, Incomplete Rides) → missing values replaced with 0.  • Duplicates  - Verified and confirmed no duplicates remained.  Final dataset contained 150,000 clean rows across 24 processed columns.

3**. Descriptive Analysis**

**3.1 Vehicle Type Popularity**

**M**ost popular: Auto, Bike, Sedan Least popular: Premium services (Premier Sedan, Uber XL)  This indicates that the affordable and fast commute options dominate ride demand.

3**.2 Average Ride Distance & Booking Value**

**A**verage Ride Distance: ~ 24.64 KM Average Booking Value: ~ ₹478.12  This shows that typical rides are medium-range urban trips, priced in the mid-segment range.

3**.3 Rating Distribution**

**C**ustomer Ratings: Skewed towards 4–5 stars → reflects high customer satisfaction. Driver Ratings: Also skewed towards higher values, indicating positive customer perception of drivers.

3**.4 Cancellation Reasons**

**B**y Customers: - Wrong address - Change of plans - Driver not moving towards pickup - Driver asked customer to cancel - AC not working  By Drivers: - Customer-related issues - Customer coughing/sick - Personal or vehicle issues - More than permitted passengers

4**. Customer Behaviour Insights**

**F**requent Cancellers: A small set of customers cancelled multiple rides (IDs highlighted in dataset).  Peak Cancellation Times: - Morning around 10 AM - Evening between 6–7 PM  Day of Week Trend: - Higher cancellations during weekdays compared to weekends.  This suggests that office commute timings drive most cancellations.

5**. Correlation Analysis**

**A** correlation heatmap was created between Ride Distance, Booking Value, and Customer Rating.  Very weak correlations were found.  Implication: Longer rides or higher fares do not directly influence customer satisfaction.  Thus, service quality factors (driver behavior, vehicle condition, punctuality) matter more than price or distance.

6**. Driver Performance**

**T**otal Driver Cancellations: ~ 27,000 rides  Top Reasons: Customer-related issues, illness concerns, and overloading.  Due to lack of unique driver identifiers, highest and lowest rated drivers could not be determined.

7**. Operational Metrics**

**7.1 VTAT & CTAT (Turnaround Times)**

**A**verage Vehicle Turnaround Time (VTAT): ~8–10 minutes. Average Customer Turnaround Time (CTAT): ~28–30 minutes.  Variation was observed across locations and vehicle types.

7**.2 Peak Demand Analysis**

**T**op Pickup Locations: - Metro stations - Commercial hubs - High-traffic areas  Peak Ride Hours: - Morning (9–10 AM) - Evening (6–8 PM)  This aligns with commute hours.

8**. Conclusion**

**T**he Ride Booking Dataset highlights several important insights:  • Affordable vehicle types dominate ride demand. • Customers and drivers generally report high satisfaction ratings. • Cancellations remain a major issue, driven by both logistics (wrong address, driver delays) and personal reasons. • Demand strongly peaks during commute times and in urban hotspots. • Price and distance are not strongly correlated with satisfaction—suggesting that service quality matters most.  Recommendations: - Improve pickup accuracy using better GPS and address verification. - Introduce penalties for frequent cancellers (both customers and drivers). - Deploy more drivers in peak-demand zones and times. - Focus on maintaining service quality rather than lowering prices, since satisfaction is not price-dependent.