**Uber Ride Data Analysis Project Report**

**Introduction:**

This project presents an end-to-end analysis of Uber ride data using SQL for backend data handling and Power BI for visual reporting. The aim is to uncover trends, identify inefficiencies, and generate insights about ride demand, cancellations, trip performance, and external influences like weather and events.

**Objectives:**

* Clean and structure raw Uber ride data for analysis.
* Calculate KPIs such as total rides, revenue, average fare, trip duration, and cancellation rates.
* Study customer behavior across different zones and time periods.
* Examine the impact of weather and events on ride volume and fare.
* Visualize findings in an interactive dashboard using Power BI**.**

**Tools & Technologies Used:**

* SQL (MySQL): Data cleaning, transformation, and aggregation.
* Power BI: Dashboard development and visualization.
* ChatGPT: Data Source
* MS Word: Documentation

**Data Sources:**

* rides\_data – Ride-level details (fare, zone, time, ratings, etc.).
* weather\_data – Weather data like temperature and rain by date.
* events\_data – Event-specific information like location, impact zone, and date.
* enriched\_rides – A SQL view combining all data sources with engineered features.

**Data Cleaning Steps:**

* Fixed null or incorrect RideStatus values.
* Standardized CancellationReason.
* Converted string-based dates to proper DATE format.
* Removed invalid or duplicate RideID entries and rows with logical inconsistencies.
* Cleaned up rides with missing distances or invalid fares.

**Data Quality Issues & Fixes:**

* Dropped rides with Dropoff < Pickup
* Nullified fare for cancelled rides
* Removed RideStatus = NULL entries
* Converted event\_data.date to DATE format using STR\_TO\_DATE

**Feature Engineering:**

* fare\_per\_km: FareAmount / DistanceKM
* average\_speed\_kmh: DistanceKM / (ActualTripDuration / 60)
* duration\_deviation\_minutes: Actual - Estimated Duration
* is\_rush\_hour: TRUE for rides between 8–10 AM or 5–8 PM
* is\_event\_day: TRUE if ride date and zone match an event
* day\_type: Weekday or Weekend (based on day of the week)

**Dashboard Insights:**

* Total rides, revenue, and completion rate shown using top KPI cards.
* Ride distribution by hour and day visualized using line and bar charts.
* Cancellation reasons broken down by zone and weekday.
* Surge pricing, fare-per-km, and rush hour effects captured.
* Weather (rain) and event day impacts analyzed.
* Customer and driver satisfaction metrics included.

**Business Insights:**

**High Cancellation Rate (50.69%)**

* More than half of all rides are cancelled, with Gachibowli being the top cancellation zone.
* Indicates potential issues like long wait times, driver unavailability, or inaccurate ETAs.
* **Action:** Investigate service quality and customer experience in Gachibowli.

**Hitec City Has Highest Surge Activity**

* Hitec City shows the most frequent surge pricing — likely due to high demand and limited supply.
* **Action:** Allocate more drivers to this zone during peak hours or use incentive-based driver positioning**.**

**Fare Insights**

* Average Fare: ₹74.78
* Fare per KM: ₹13.56
* Fare per Minute: ₹3.33
* These values vary by hour, peaking during early mornings and late evenings (shown in the line chart).

**Peak Ride Hours and Rush Impact**

* Most rides occur during rush hours (8–10 AM and 5–8 PM).
* Rush hour rides: 13.9K, Non-rush: 5.7K
* **Action:** Consider optimizing pricing and driver availability for these windows.

**Weekday vs Weekend Demand**

* Majority of rides occur on weekdays (86.24%), while only 13.77% are on weekends.
* **Action:** Focus marketing or promotions on weekend usage to balance demand.

**Top Event Impact Zones**

* Most rides during events occurred in Citywide zones (666 rides).
* Events in Banjara Hills and Hitec City had minimal influence.
* **Action:** Consider further validating which events impact demand and align driver deployments accordingly.

**Cancellation by Day of Week**

* Cancellation is fairly consistent across days, with Monday having slightly more cancellations.
* **Action:** Consider adjusting weekday scheduling or offering time-based discounts.

**On-Time Completion Rate**

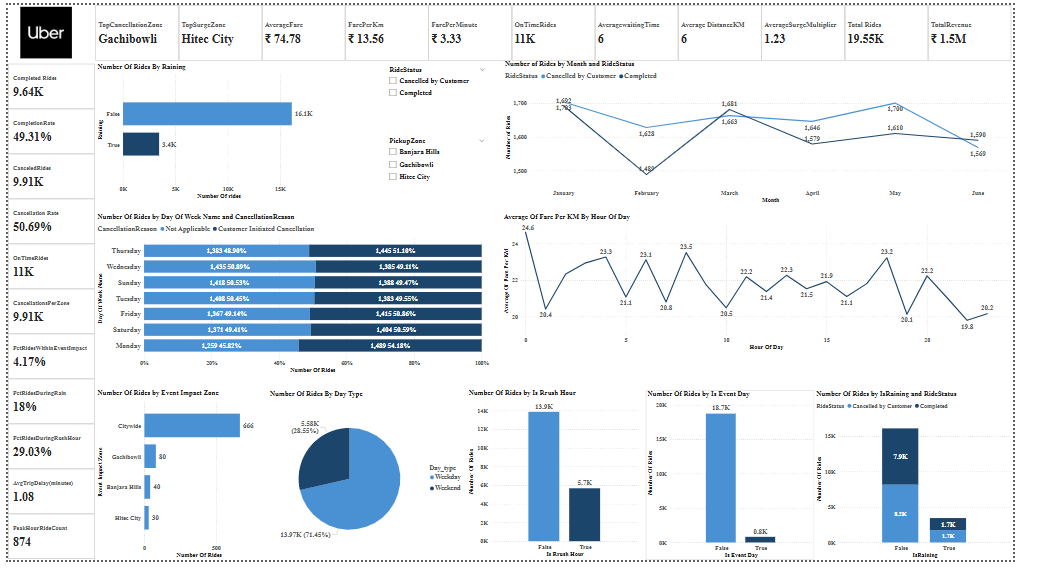
* 11K rides were completed on time, indicating a moderate performance level.
* **Action:** Analyze delays in other zones to further improve performance.

**Customer Rating Trends**

* 7.5K rides were rated by customers, indicating decent engagement.
* Cancellations had fewer ratings, possibly due to incomplete ride experience.
* **Action:** Encourage post-ride feedback, even on partial or cancelled trips, to gather service insights.

**Low Event-Based Ride Ratio**

* Only 874 rides (18%) occurred during event days.
* This suggests that either events were not major influencers or weren't well captured in the data.
* **Action:** Ensure event data coverage is accurate and up to date; promote ride services during local events.



**About Me**

I’m Ajay, a B.Tech graduate in Electronics and Communication Engineering, currently building my skills in data analytics.  
This Uber data analysis project reflects my hands-on practice with SQL, Power BI, and real-world business logic.  
My goal is to transition into a career as a data analyst, where I can solve practical problems using data, insights, and storytelling.  
I’m passionate about creating clear, actionable dashboards and continuously learning through real-world projects.

**Ajay Sammeta**

**Email:** ajaysammeta03@gmail.com

**LinkedIn:** [**linkedin.com/in/ajay-sammeta-6b1369237**](https://www.linkedin.com/in/ajay-sammeta-6b1369237)

**Portfolio:** https://github.com/ajaysammeta03/Data-Analytics-Portfolio