1. Project Overview

The objective of this project is to analyze Uber ride booking data to uncover meaningful insights related to ride demand, customer behavior, cancellations, payment preferences, and service performance.
The project demonstrates end-to-end data analytics skills, including data cleaning, feature engineering, SQL-based analysis, and dashboard visualization using real-world ride booking data.

2. Dataset Description

The dataset used for this project is the NCR Ride Booking dataset sourced from Kaggle.
After cleaning and preprocessing in Python, the final dataset was saved as uberdata_clean.csv.

Final Columns in Dataset:

Date
Time
Booking ID
Booking Status
Customer ID
Vehicle Type
Pickup Location
Drop Location
Cancelled Rides by Customer
Cancelled Rides by Driver
Incomplete Rides
Driver Ratings
Customer Rating
Payment Method
Hour
Day
Month

3. Data Cleaning & Feature Engineering

The raw dataset required preprocessing before analysis.

Data Cleaning:
- Removed unnecessary and irrelevant columns
- Handled missing values in ratings, payment methods, and cancellation fields
- Standardized text values
- Verified numerical consistency

Feature Engineering:
- Extracted Hour, Day, and Month
- Created derived fields for trend analysis
- Ensured correct data types for all analytical columns

4. SQL-Based Business Analysis

The cleaned dataset was analyzed using SQL to answer key business questions:

1. Peak ride hours

Hour	total_rides
01 AM	1360
01 PM	5470
02 AM	1339
02 PM	7031
03 AM	1383
03 PM	8202
04 AM	1321
04 PM	9633
05 AM	2786
05 PM	11044
06 AM	4160
06 PM	12397
07 AM	5450
07 PM	11047
08 AM	6861
08 PM	9630
09 AM	8234
09 PM	8103
10 AM	9577
10 PM	5441
11 AM	8390
11 PM	2762
12 AM	1373
12 PM	7006

2. Busiest days

Day	total_rides
Monday	21644
Saturday	21542
Wednesday	21413
Sunday	21398
Friday	21397
Tuesday	21391
Thursday	21215

3. Monthly booking trends

Month	Total_Rides
July	12897
January	12861
May	12778
March	12719
October	12651
August	12636
June	12440
November	12394
December	12250
September	12248
April	12199
February	11927

4. Payment method preference

Payment Method	Total_Rides
NULL	48000
UPI	45909
Debit Card	8239
Cash	25367
Uber Wallet	12276
Credit Card	10209

5. Incomplete ride analysis

Day	Total_Incomplete_Rides
Friday	1259
Monday	1305
Saturday	1287
Sunday	1255
Thursday	1295
Tuesday	1332
Wednesday	1267

6. Vehicle type performance

Vehicle Type	Total_Rides	Avg_Driver_Rating	Avg_Customer_Rating
eBike	10557	4.225614410013734	4.4039535948710125
Go Sedan	27141	4.231812185176269	4.409996402014913
Auto	37419	4.2323688188296105	4.401999568127827
Premier Sedan	18111	4.234864912904311	4.403457163171014
Bike	22517	4.230055579307338	4.403940430383356
Go Mini	29806	4.227694215321552	4.404296727586433
Uber XL	4449	4.238339920948609	4.404850880344944

7. Cancellation trends

Day	Customer_Cancellations	Driver_Cancellations
Friday	1566	3777
Monday	1511	4041
Saturday	1454	3874
Sunday	1508	3835
Thursday	1479	3752
Tuesday	1469	3919
Wednesday	1513	3802

8. Booking status distribution

Booking Status	Total_Rides
No Driver Found	10500
Incomplete	9000
Completed	93000
Cancelled by Driver	27000
Cancelled by Customer	10500

9. High-demand pickup locations

Pickup Location	Total_Rides
Khandsa	949
Barakhamba Road	946
Saket	931
Badarpur	921
Pragati Maidan	920

10. Overall business performance

Total_Rides	Customer_Cancellations	Driver_Cancellations	Total_Incomplete_Rides	Avg_Driver_Rating	Avg_Customer_Rating
150000	10500	27000	9000	4.230992473118317	4.404583870967736

5. Power BI Dashboard

An interactive Power BI dashboard was created to visualize insights such as:

- Total rides, cancellations, and incomplete rides
- Monthly and daily demand trends
- Vehicle usage distribution
- Rating performance

6. Business Insights & Recommendations

- Peak ride hours indicate optimal times for driver allocation
- Certain days experience higher cancellations requiring operational improvement
- Digital payments dominate customer preferences
- Vehicle type usage insights help fleet optimization
- High-demand locations should be prioritized

7. Conclusion

This project demonstrates practical expertise in data cleaning, SQL analysis, and data visualization.
It showcases the ability to convert raw data into meaningful business insights.