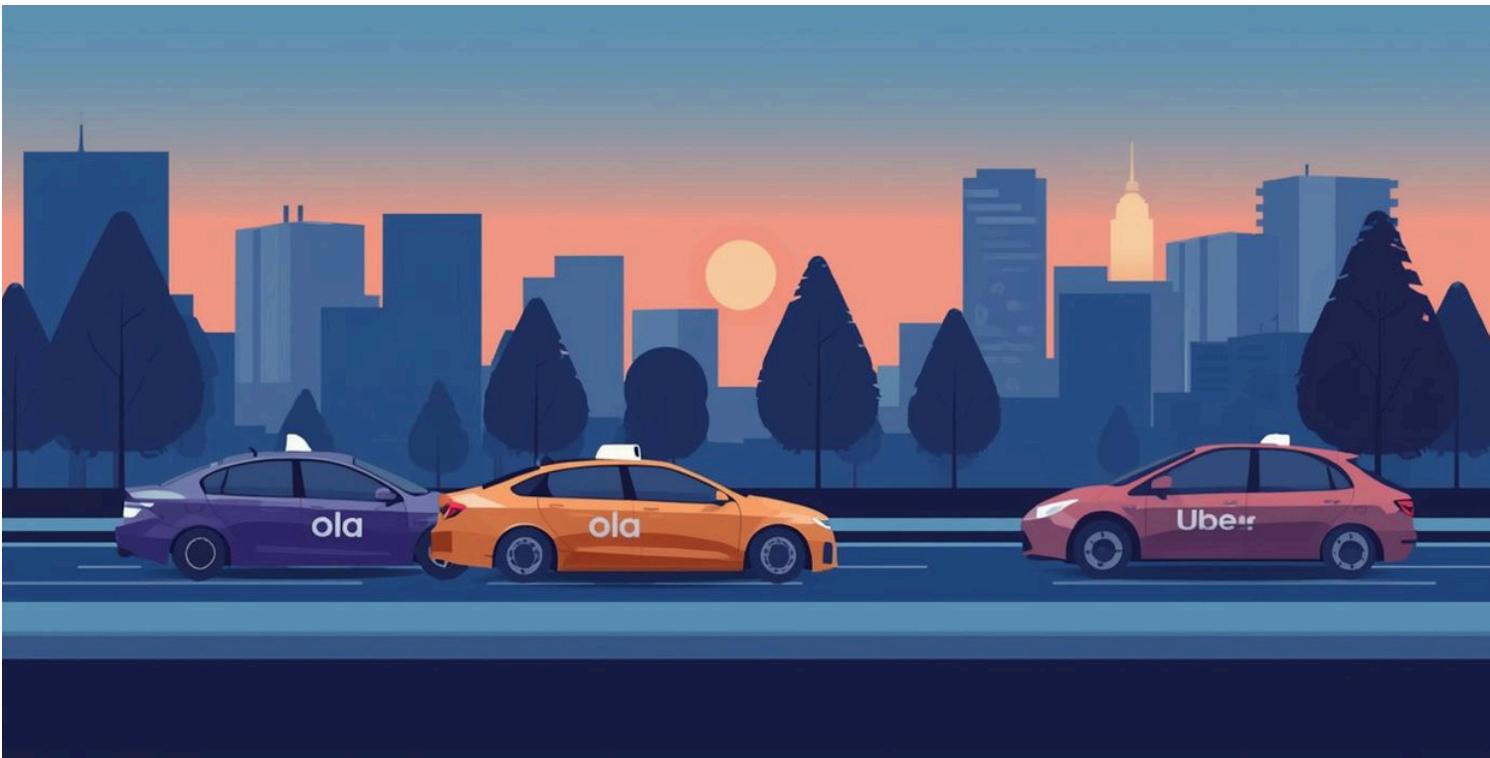


Ola and Uber

Ride Booking & Cancellation Analysis



Presented by: Amey Ghotankar

About Me

- Data Analyst with hands-on experience in SQL-based exploratory and business analysis
- Strong focus on real-world problem solving using large datasets
- This project demonstrates my ability to analyze operational inefficiencies and customer behavior using SQL



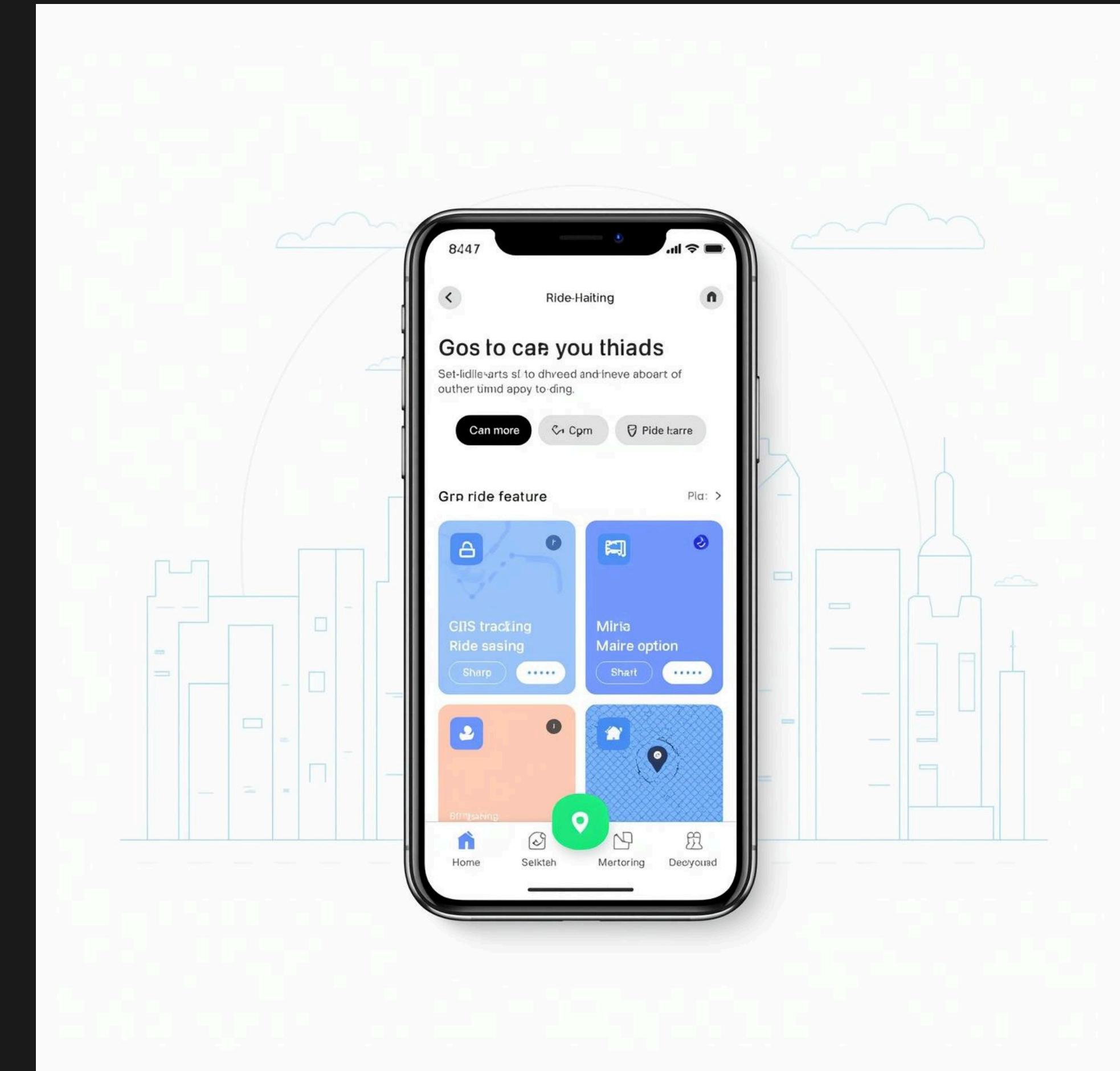
Project Overview

Project Objective

- Analyze ride booking behavior, cancellations, turnaround times, and revenue
- Identify operational bottlenecks affecting ride completion and revenue

Key Business Questions

- Why are rides getting cancelled?
- Who cancels more: customers or drivers?
- Which vehicle types perform best?
- How does turnaround time impact success rates?



Dataset Overview



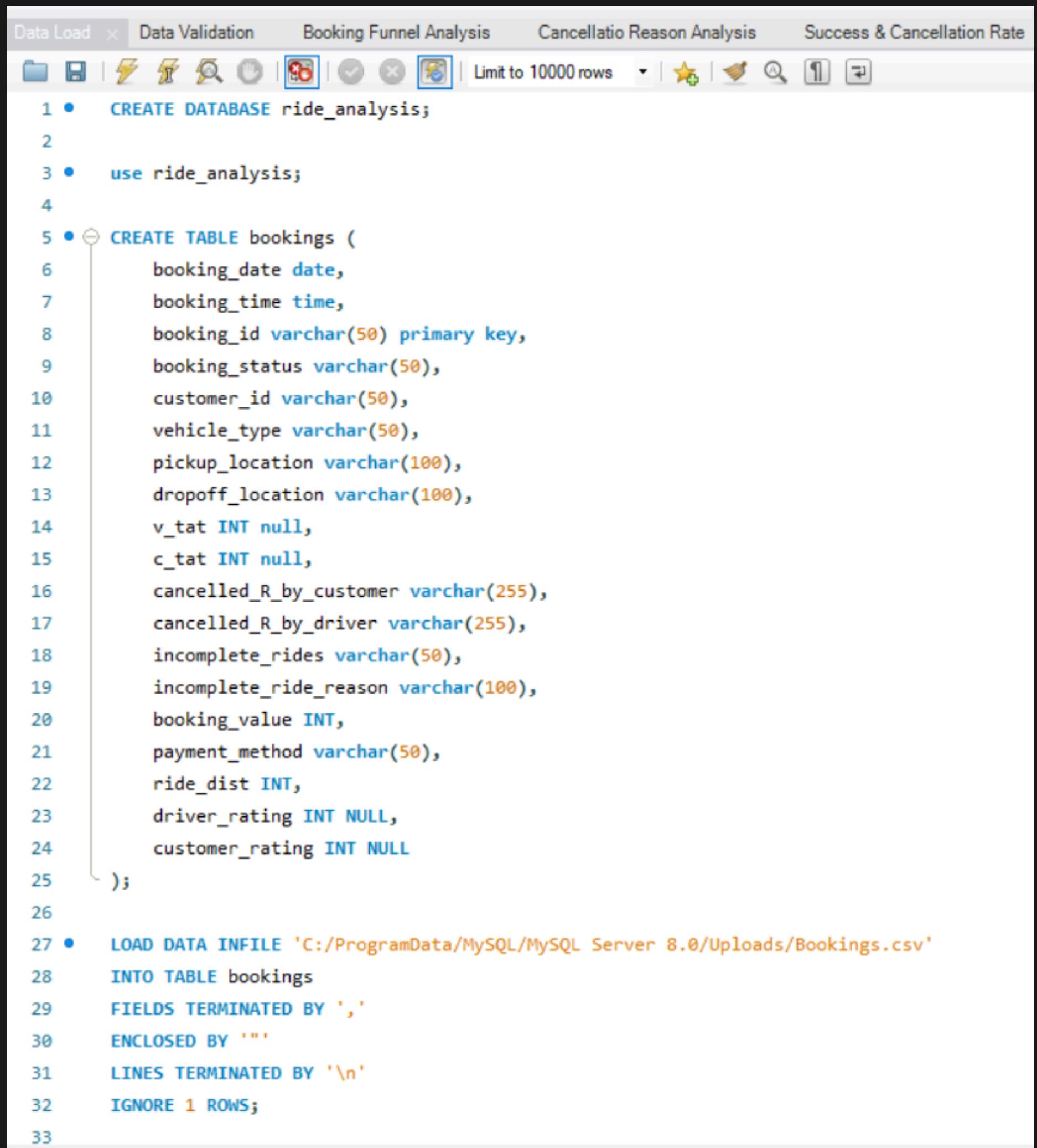
Dataset Details

- 103,000+ ride bookings
- Multiple vehicle types (Sedan, SUV, Auto, Bike, eBike, etc.)
- Time-based, location-based, and behavioral attributes
- Realistic cancellation and incomplete ride scenarios

Tools Used

- MySQL
- SQL aggregations & analytical queries

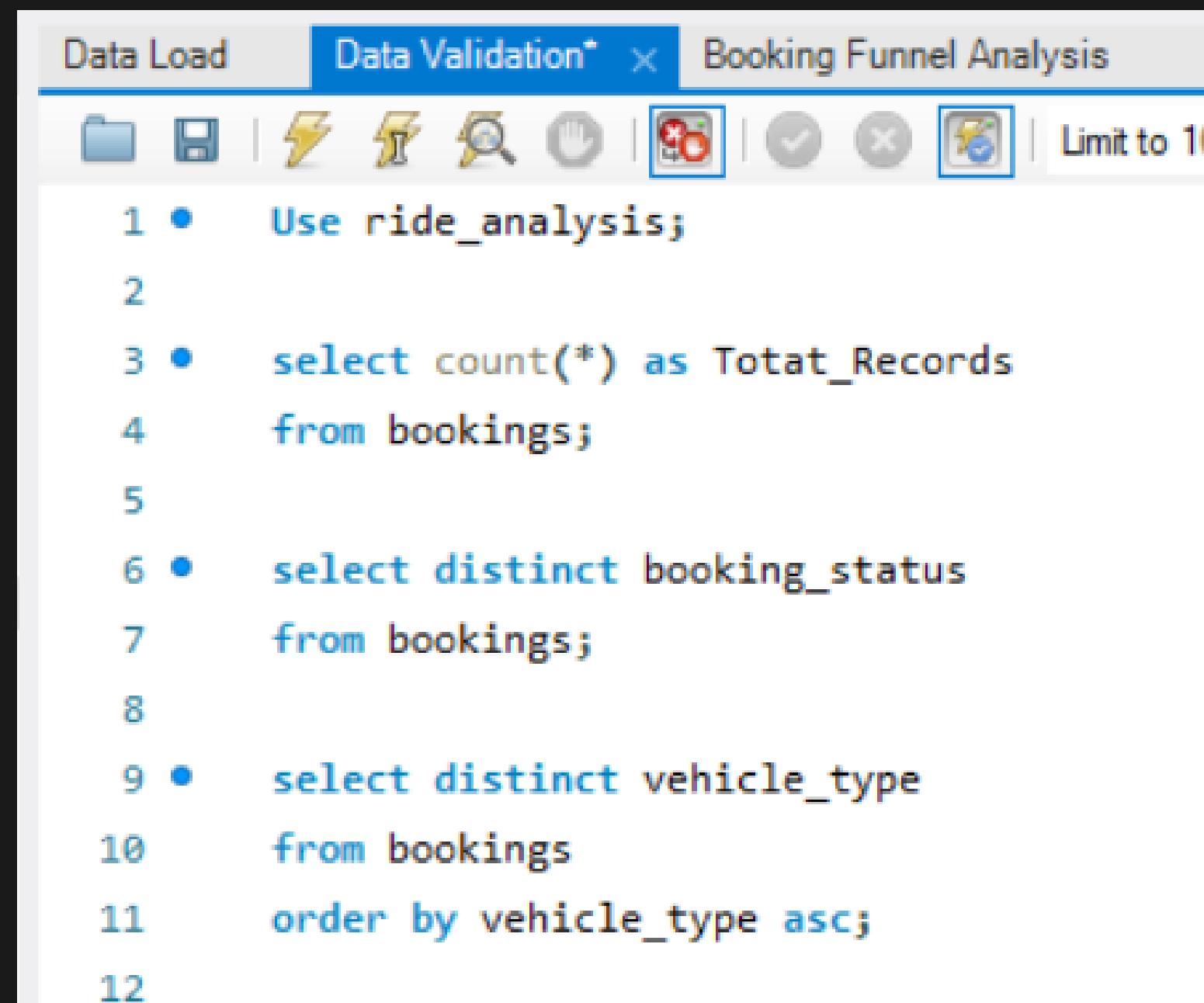
Data Loading & Preparation



The screenshot shows a MySQL Workbench interface with several tabs at the top: Data Load, Data Validation, Booking Funnel Analysis, Cancellation Reason Analysis, and Success & Cancellation Rate. The Data Load tab is active. Below the tabs is a toolbar with various icons. The main area contains the following SQL code:

```
1 • CREATE DATABASE ride_analysis;
2
3 • use ride_analysis;
4
5 • CREATE TABLE bookings (
6     booking_date date,
7     booking_time time,
8     booking_id varchar(50) primary key,
9     booking_status varchar(50),
10    customer_id varchar(50),
11    vehicle_type varchar(50),
12    pickup_location varchar(100),
13    dropoff_location varchar(100),
14    v_tat INT null,
15    c_tat INT null,
16    cancelled_R_by_customer varchar(255),
17    cancelled_R_by_driver varchar(255),
18    incomplete_rides varchar(50),
19    incomplete_ride_reason varchar(100),
20    booking_value INT,
21    payment_method varchar(50),
22    ride_dist INT,
23    driver_rating INT NULL,
24    customer_rating INT NULL
25 );
26
27 • LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Bookings.csv'
28 INTO TABLE bookings
29 FIELDS TERMINATED BY ','
30 ENCLOSED BY ""
31 LINES TERMINATED BY '\n'
32 IGNORE 1 ROWS;
33
```

Data Validation



The screenshot shows a software interface for data validation. The title bar includes tabs for "Data Load", "Data Validation*", and "Booking Funnel Analysis". Below the tabs is a toolbar with various icons. The main area displays a numbered list of SQL queries:

- 1 • Use ride_analysis;
- 2
- 3 • select count(*) as Total_Records
- 4 from bookings;
- 5
- 6 • select distinct booking_status
- 7 from bookings;
- 8
- 9 • select distinct vehicle_type
- 10 from bookings
- 11 order by vehicle_type asc;
- 12

Booking Funnel Analysis

- Total booking requests? → **103024**
- How many succeeded? → **63967**
- How many cancelled by customer? → **10499**
- How many cancelled by driver? → **18434**
- How many incomplete rides? → **3926**

```
SELECT
    COUNT(*) AS Total_Bookings,
    SUM(CASE
        WHEN booking_status = 'Success' THEN 1
        ELSE 0
    END) AS Successful_Bookings,
    SUM(CASE
        WHEN booking_status = 'Canceled by Driver' THEN 1
        ELSE 0
    END) AS Driver_Cancellation,
    SUM(CASE
        WHEN booking_status = 'Canceled by Customer' THEN 1
        ELSE 0
    END) AS Customer_Cancellation
FROM
    bookings;
```

Total_Bookings	Successful_Bookings	Driver_Cancellation	Customer_Cancellation
103024	63967	18434	10499

```
select
    SUM(CASE WHEN incomplete_rides = "Yes" THEN 1 ELSE 0 END) AS Incomplete_Rides
FROM bookings;
```

Incomplete_Rides
3926

Cancellation Rate Analysis

A) Overall Cancellation Rate:

Total cancellations / total bookings → **38%**

B) Customer vs Driver Cancellation Share:

→ Customer – **10%** & Driver – **10%**

C) Cancellation by Vehicle Type:

- Which vehicle has highest cancellation rate?
→ eBike with **28.4%** cancellation rate
- Is Bike or Auto more unreliable?
→ With cancellation rate **27%** in bike and **28.3%** in Auto,

Auto is more unreliable

Insight

Only **62%** of bookings are successful, with driver cancellations accounting for **18%**, Customer cancellations accounting for 10% and Miscellaneous **10%** of all failed rides

Prepared by:

Amey Ghotankar

```
-- Success & Cancellation Rate
select
    count(*) as Total_Bookings,
    round(SUM(case When booking_status = "Success" Then 1 Else 0 END)
        * 100 / count(*)) as Successful_Booking_Rate,
    round(SUM(case When booking_status != "Success" Then 1 Else 0 END)
        * 100 / count(*)) as Booking_Cancellation_Rate
from bookings;

-- Customer vs Driver Cancellation Share
select
    round(SUM(case When booking_status = "Canceled by Driver" Then 1 Else 0 END)
        * 100 / count(*)) as Driver_Cancellation_Rate,
    round(SUM(case When booking_status = "Canceled by Customer" Then 1 Else 0 END)
        * 100 / count(*)) as Customer_Cancellation_Rate
from bookings;

-- Vehicle wise cancellation rate
select
    vehicle_type,
    round(SUM(case When booking_status != "Success" AND booking_status <> "Driver Not Found" Then 1 Else 0 END)
        * 100 / count(*),1) as Cancellation_Rate
from bookings
group by vehicle_type
order by Cancellation_Rate desc;
```

Successful_Booking_Rate	Booking_Cancellation_Rate	Driver_Cancellation_Rate	Customer_Cancellation_Rate
62	38	18	10

Vehicle_Type	Cancellation_Rate
eBike	28.4
Prime SUV	28.3
Prime Plus	28.3
Auto	28.3
Mini	28.1
Bike	27.9
Prime Sedan	27.3

Cancellation Reason Analysis

A) Top 3 Customer Cancellation Reasons

- Driver is not moving towards pickup location and count is '3175'
- Driver asked to cancel and count is '2670'
- Change of plans and count is '2081'

B) Top Driver Cancellation Reasons

- Personal & Car related issue and count is '6542'
- Customer related issue and count is '5413'
- Customer was coughing/sick and count is '3654'

Insight

Over **30%** of customer cancellations occur due to long wait times and **35.5%** of driver cancellation occur due to personal and car related issues

Prepared by:

Amey Ghotankar

```
-- Top 3 Customer Cancellation Reasons  
• select  
distinct cancelled_R_by_customer as Customer_Cancellation_Reasons,  
count(cancelled_R_by_customer) as Count  
from bookings  
where cancelled_R_by_customer <> "null"  
group by cancelled_R_by_customer  
Order by Count desc  
limit 3;
```

```
-- Top 3 Driver Cancellation Reasons  
• select  
distinct cancelled_R_by_driver as Driver_Cancellation_Reasons,  
count(cancelled_R_by_driver) as Count  
from bookings  
where cancelled_R_by_driver <> "null"  
group by cancelled_R_by_driver  
Order by Count desc  
limit 3;
```

	Driver_Cancellation_Reasons	Count
▶	Personal & Car related issue	6542
	Customer related issue	5413
	Customer was coughing/sick	3654

	Customer_Cancellation_Reasons	Count
▶	Driver is not moving towards pickup location	3175
	Driver asked to cancel	2670
	Change of plans	2081

TAT Impact Analysis

A) Vehicle Turnaround Time (V-TAT)

- Avg V_TAT for successful rides vs cancelled rides
→ Average V_TAT for overall bookings is **106.1** whereas for successful bookings V_TAT is **170.9**.

B) Customer Turnaround Time (C-TAT)

- Does higher C_TAT increase incomplete rides?
→ **Yes**, Higher the C_TAT, Higher the No. of Incomplete Rides

```
-- Average of both Vehicle and Customer TAT

select
    round(avg(v_tat), 1) as Vehicle_TAT,
    round(avg(c_tat), 1) as Customer_TAT
from bookings;

-- Avg V_TAT for successful rides

select
    round(avg(v_tat), 1) as Vehicle_TAT
from bookings
where booking_status = "Success";

-- Avg C_TAT for incomplete rides

select
    incomplete_rides,
    round(avg(c_tat), 1) as Customer_TAT
from bookings
where c_tat <> "null"
group by incomplete_rides;
```

	Vehicle_TAT	Customer_TAT
▶	106.1	52.7

	Vehicle_TAT
▶	170.9

	Incomplete_Rides	Customer_TAT
▶	No	84.9
	Yes	84.7

Time Analysis I

A) Peak Booking Hours

- Afternoon 12PM has the highest no. of bookings - 4408

B) Cancellation by Time of Day

- Morning 10AM has the highest no. of cancellations - 1718

-- Peak Booking Hours

Select

```
hour(booking_time) as booking_hour,  
count(*) as Total_Bookings  
from bookings  
group by booking_hour  
order by Total_Bookings desc;
```

-- Peak hour cancellation

Select

```
hour(booking_time) as booking_hour,  
count(*) as Total_Bookings,  
sum(case when booking_status != "Success" then 1 else 0 end) as Cancellations  
from bookings  
group by booking_hour  
order by Cancellations desc;
```

	Booking_Hour	Total_Bookings
▶	12	4408
	15	4376
	8	4374
	17	4353
	9	4347
	21	4343
	10	4334
	1	4329
	0	4318
	2	4305

	Booking_Hour	Total_Bookings	Cancellations
▶	10	4334	1718
	12	4408	1709
	8	4374	1693
	7	4304	1676
	22	4283	1674
	9	4347	1667
	21	4343	1664
	17	4353	1662
	2	4305	1647
	20	4228	1641

Time Analysis II

C) Day-wise Trend (July)

- Tuesday has the highest no. of bookings – 16926 and Sunday Being the lowest – 13013

Insight

- Cancellation rates spike between **6 PM – 9 PM**, with least no of bookings.

```
-- day time trend
SELECT
    CASE
        WHEN HOUR(booking_time) BETWEEN 5 AND 11 THEN 'Morning'
        WHEN HOUR(booking_time) BETWEEN 12 AND 17 THEN 'Afternoon'
        WHEN HOUR(booking_time) BETWEEN 18 AND 21 THEN 'Evening'
        ELSE 'Late Night'
    END AS Time_of_Day,
    COUNT(*) AS Total_Bookings,
    SUM(
        CASE
            WHEN booking_status <> 'Success' THEN 1
            ELSE 0
        END
    ) AS Cancelled_Bookings,
    ROUND(
        SUM(
            CASE
                WHEN booking_status <> 'Success' THEN 1
                ELSE 0
            END
        ) * 100.0 / COUNT(*),
        2
    ) AS Cancellation_Rate
FROM bookings
GROUP BY Time_of_Day
ORDER BY Cancellation_Rate DESC;
```

	Time_of_Day	Total_Bookings	Cancelled_Bookings	Cancellation_Rate
▶	Evening	17018	6526	38.35
	Morning	30166	11480	38.06
	Afternoon	25915	9792	37.79
	Late Night	29925	11259	37.62

Location Analysis I

A) Pickup Locations with High Cancellations

- Vijayanagar, Total Bookings - **2113**, Cancellations - **849**
- Banashankari, Total Bookings - **2201**, Cancellations - **836**
- Tumkur Road, Total Bookings - **2105**, Cancellations - **836**

B) Drop Locations with High Incomplete Rides

- Marathahalli, Total Bookings - **2104**, Cancellations - **844**
- Sarjapur Road, Total Bookings - **2108**, Cancellations - **840**
- MG Road, Total Bookings - **2128**, Cancellations - **837**

```
-- Top 3 pickup locations with highest cancellation

select
    pickup_location as Pickup_Location,
    count(*) as Total_Bookings,
    sum(case when booking_status != "Success" then 1 else 0 end) as Cancellations
from bookings
group by Pickup_Location
order by Cancellations desc
limit 3;

-- Top 3 dropoff locations with highest cancellation

select
    dropoff_location as Dropoff_Location,
    count(*) as Total_Bookings,
    sum(case when booking_status != "Success" then 1 else 0 end) as Cancellations
from bookings
group by Dropoff_Location
order by Cancellations desc
limit 3;
```

	Pickup_Location	Total_Bookings	Cancellations
▶	Vijayanagar	2113	849
	Banashankari	2201	836
	Tumkur Road	2105	836

	Dropoff_Location	Total_Bookings	Cancellations
▶	Marathahalli	2104	844
	Sarjapur Road	2108	840
	MG Road	2128	837

Location Analysis II

Pickup Dropoff Successful Rides

- Majestic to Hennur → **49**
- Kammanahalli to Kadugodi → **45**
- Hulimavu to Banashankari → **44**
- Bellandur to Indiranagar → **43**
- Basavanagudi to Frazer Town → **43**

Insight

Majestic to Hennur is High in demand with **49** successful rides

-- High Demand Corridors

SELECT

pickup_location AS Pickup_Location,
dropoff_location AS Dropoff_Location,
COUNT(*) AS Successful_Rides

FROM

bookings

WHERE

booking_status = 'Success'

GROUP BY Pickup_Location , Dropoff_Location

ORDER BY Successful_Rides DESC

LIMIT 5;

	Pickup_Location	Dropoff_Location	Successful_Rides
▶	Majestic	Hennur	49
	Kammanahalli	Kadugodi	45
	Hulimavu	Banashankari	44
	Bellandur	Indiranagar	43
	Basavanagudi	Frazer Town	43

Incomplete Ride Analysis

% of incomplete rides

→ **3.81%** are incomplete out of total **103024** rides

Top reasons for incomplete rides

→ Customer demand is top reason with **1601** incomplete rides

Vehicle types most affected

→ Prime Sedan has the highest with **611** incomplete rides and

Mini is lowest **517**

Insight

41% of incomplete rides are due to Customer Demand

```
-- % of incomplete rides

select
    sum(case when incomplete_rides = "yes" then 1 else 0 end) as Incomplete_Rides,
    round(sum(case when incomplete_rides = "yes" then 1 else 0 end)
        * 100 / count(*),2) as 'Percentage'
from bookings;

-- reasons of incomplete rides

select
    incomplete_ride_reason as Incomplete_Ride_Reason,
    sum(case when incomplete_rides = "yes" then 1 else 0 end) as Incomplete_Rides
from bookings
where incomplete_ride_reason <> "null"
group by incomplete_ride_reason
order by Incomplete_Rides desc;

-- vehicles wise incomplete_ride count

select
    vehicle_type as Vehicle_Type,
    sum(case when incomplete_rides = "yes" then 1 else 0 end) as Incomplete_Rides
from bookings
where incomplete_ride_reason <> "null"
group by vehicle_type
order by Incomplete_Rides desc;
```

Incomplete_Rides	Percentage
3926	3.81

Incomplete_Ride_Reason	Incomplete_Rides
Customer Demand	1601
Vehide Breakdown	1591
Other Issue	734

Vehicle_Type	Incomplete_Rides
Prime Sedan	611
Bike	579
eBike	574
Auto	562
Prime SUV	547
Prime Plus	536
Mini	517

Customer Behavior Analysis

High-Risk Customers:

- Customers with >50% cancellation rate

There are total **9178** customers with cancellation rate greater than **50**

```
SELECT
    COUNT(*) AS customers_with_high_cancellation
FROM
    (SELECT
        customer_id,
        ROUND(SUM(CASE
            WHEN booking_status = 'Canceled by Customer' THEN 1
            ELSE 0
        END) * 100 / COUNT(*), 2) AS cancellations_rate
    FROM
        bookings
    GROUP BY customer_id
    HAVING cancellations_rate > 50) t;
```

Insight

8.9% of cancellations from total are cancelled by customers and **15.7%** by drivers

Customers with High Cancellation
9178

Revenue Analysis

A) Total Revenue

→ \$ 56,534,614

B) Top 3 Vehicles with highest revenue

→ Prime Sedan - \$ 8,298,612

→ eBike - \$ 8,181,523

→ Auto - \$ 8,092,210

C) Revenue by Payment Method

→ Cash - \$ 19,260,316

→ UPI \$ 14,168,436

```
-- Revenue Analysis

select
    CONCAT('$ ', FORMAT(sum(booking_value),0)) AS Total_Revenue
from bookings;

select
    vehicle_type as Vehicle_Type,
    CONCAT('$ ', FORMAT(sum(booking_value),0)) AS Total_Revenue
from bookings
group by vehicle_type
order by Total_Revenue desc
limit 3;

select
    payment_method as Payment_Method,
    round(sum(booking_value),0) as Total_Revenue
from bookings
where payment_method <> "null"
group by payment_method
order by Total_Revenue desc;
```

Total_Revenue
\$ 56,534,614

Vehide_Type	Total_Revenue
Prime Sedan	\$ 8,298,612
eBike	\$ 8,181,523
Auto	\$ 8,092,210

Payment_Method	Total_Revenue
Cash	19260316
UPI	14168436
Credit Card	1312793
Debit Card	338922



Thank You!

Let's Connect

Email

apghotankar@gmail.com

Social

www.linkedin.com/in/ameyghotankar

Phone

+91-9403900084