Project Report: Ola Ride Analytics

1. Objective:

The project aims to analyze ride data, focusing on successful bookings, ride statistics, customer behavior, and the reasons behind ride cancellations and incomplete rides. The goal is to provide key insights and create views that answer various business questions for decision-making.

---

2. Database Structure:

The project utilizes a database called `ola_db` with the primary table `bookings`, which stores data related to ride bookings. The table includes various fields such as:

- `booking_status`, `ride_distance`, `driver_ratings`, `customer_ratings`
- `vehicle_type`, `payment_method`, `v_tat` (vehicle turnaround time), `c_tat` (customer turnaround time)
- `canceled_rides_by_customer`, `canceled_rides_by_driver`, `incomplete_rides`, etc.

---

3. SQL Queries:

A. Business Questions and Corresponding Views:

1. **Retrieve all successful bookings:**
2. CREATE VIEW successfull_bookings AS
3. SELECT *
4. FROM bookings
5. WHERE booking_status ILIKE '%Success%';
6. **Find the average ride distance for each vehicle type:**
7. CREATE VIEW vehicle_avg_distance AS
8. SELECT vehicle_type, ROUND(AVG(ride_distance), 2) as average_ride_distance
9. FROM bookings
10. GROUP BY 1;
11. **Get the total number of canceled rides by customers:**
12. CREATE VIEW cancelled_rides_by_cust AS
13. SELECT COUNT(*)
14. FROM bookings
15. WHERE booking_status ILIKE '%Canceled by customer%';
16. **List the top 5 customers who booked the highest number of rides:**
17. CREATE VIEW top_5_cust AS
18. SELECT customer_id, COUNT(8) AS Total_rides
19. FROM bookings
20. GROUP BY 1
21. ORDER BY 2 DESC
22. LIMIT 5;
23. **Get the number of rides canceled by drivers due to personal and car-related issues:**

```
24.  CREATE VIEW rdes_canceled_by_driver_P_C_issues AS
25.  SELECT COUNT(*) AS total_rides
26.  FROM bookings
27.  WHERE canceled_rides_by_driver = 'Personal & Car related issue';
```

28. Find the maximum and minimum driver ratings for Prime Sedan bookings:

```
29.  CREATE VIEW max_min_rating_for_prime_sedan AS
30.  SELECT vehicle_type, MAX(driver_ratings) AS max_rating,
     MIN(driver_ratings) AS min_rating
31.  FROM bookings
32.  WHERE vehicle_type = 'Prime Sedan'
33.  GROUP BY 1;
```

34. Retrieve all rides where payment was made using UPI:

```
35.  CREATE VIEW upi_payment AS
36.  SELECT *
37.  FROM bookings
38.  WHERE payment_method = 'UPI';
```

39. Find the average customer rating per vehicle type:

```
40.  CREATE VIEW avg_cust_rating_by_vehicle AS
41.  SELECT vehicle_type, ROUND(AVG(customer_rating), 2) AS
     average_cust_rating
42.  FROM bookings
43.  GROUP BY 1;
```

44. Calculate the total booking value of rides completed successfully:

```
45.  CREATE VIEW successful_rides_booking_value AS
46.  SELECT SUM(booking_value) AS total_booking_value
47.  FROM bookings
48.  WHERE booking_status = 'Success';
```

49. List all incomplete rides along with the reason:

```
CREATE VIEW incomplete_rides_reason AS
SELECT booking_id, incomplete_rides_reason
FROM bookings
WHERE incomplete_rides = 'Yes';
```

B. Additional Advanced Queries:

1. Find the average ride distance for completed rides, grouped by booking status:

```
2.  CREATE VIEW avg_distance_by_status AS
3.  SELECT booking_status, ROUND(AVG(ride_distance), 2) AS
    avg_ride_distance
4.  FROM bookings
5.  GROUP BY booking_status;
```

6. Retrieve the customer who has made the highest number of bookings for each vehicle type:

```
7.  CREATE VIEW top_customer_by_vehicle_type AS
8.  WITH ranked_customers AS (
9.      SELECT vehicle_type, customer_id, COUNT(*) AS total_bookings,
10.            RANK() OVER (PARTITION BY vehicle_type ORDER BY COUNT(*)
    DESC) AS rank
11.     FROM bookings
12.     GROUP BY vehicle_type, customer_id
13. )
14. SELECT vehicle_type, customer_id, total_bookings
15. FROM ranked_customers
16. WHERE rank = 1;
```

17. Calculate the average ride duration and the average booking value for each day of the week:

```
18.  CREATE VIEW avg_ride_duration_and_booking_value_by_dow AS
19.  SELECT to_char(TO_TIMESTAMP(date, 'DD-MM-YYYY HH24:M1'), 'Day'),
20.         ROUND(AVG(v_tat + c_tat), 2) AS average_duration,
21.         ROUND(AVG(booking_value), 2) AS average_booking_value
22.  FROM bookings
23.  GROUP BY 1
24.  ORDER BY 3 DESC;
```

25. Identify the most common reasons for incomplete rides, sorted by frequency:

```
26.  CREATE VIEW common_reasons_for_incomplete_rides_ AS
27.  SELECT CASE
28.          WHEN incomplete_rides_reason IS NULL OR
     incomplete_rides_reason = 'Other Issue' THEN 'Other Issue'
29.          ELSE incomplete_rides_reason
30.         END AS incomplete_rides_reason,
31.         COUNT(*) AS total_bookings
32.  FROM bookings
33.  GROUP BY incomplete_rides_reason
34.  ORDER BY total_bookings DESC;
```

35. Retrieve the total number of rides canceled by drivers for different reasons and calculate their percentage share out of total canceled rides:

```
36.  CREATE VIEW driver_cancel_reasons_percentages AS
37.  SELECT canceled_rides_by_driver,
38.         COUNT(*) AS total_rides_canceled_by_driver,
39.         CONCAT(ROUND(COUNT(*)::NUMERIC /
40.             (SELECT COUNT(*) FROM bookings
41.              WHERE canceled_rides_by_customer IS NOT NULL OR
     canceled_rides_by_driver IS NOT NULL) * 100, 2), '%')
42.         AS percent_of_total
43.  FROM bookings
44.  WHERE canceled_rides_by_driver IS NOT NULL
45.  GROUP BY canceled_rides_by_driver;
```

C. One-Line Queries for Direct Answers:

1. **Retrieve all successful bookings:**
```
2. SELECT * FROM bookings WHERE booking_status ILIKE '%Success%';
```
3. **Find the average ride distance for each vehicle type:**
```
4. SELECT vehicle_type, ROUND(AVG(ride_distance), 2) AS
   average_ride_distance FROM bookings GROUP BY vehicle_type;
```
5. **Get the total number of canceled rides by customers:**
```
6. SELECT COUNT(*) FROM bookings WHERE booking_status ILIKE '%Canceled
   by customer%';
```
7. **List the top 5 customers who booked the highest number of rides:**
```
8. SELECT customer_id, COUNT(*) AS Total_rides FROM bookings GROUP BY
   customer_id ORDER BY Total_rides DESC LIMIT 5;
```
9. **Get the number of rides canceled by drivers due to personal and car-related issues:**
```
10.  SELECT COUNT(*) FROM bookings WHERE canceled_rides_by_driver =
   'Personal & Car related issue';
```
11. **Find the maximum and minimum driver ratings for Prime Sedan bookings:**
```
12.  SELECT MAX(driver_ratings) AS max_rating, MIN(driver_ratings) AS
   min_rating FROM bookings WHERE vehicle_type = 'Prime Sedan';
```
13. **Retrieve all rides where payment was made using UPI:**
```
14.  SELECT * FROM bookings WHERE payment_method = 'UPI';
```
15. **Find the average customer rating per vehicle type:**
```
16.  SELECT vehicle_type, ROUND(AVG(customer_rating), 2) AS
   average_cust_rating FROM bookings GROUP BY vehicle_type;
```

17. **Calculate the total booking value of rides completed successfully:**
18. `SELECT SUM(booking_value) AS total_booking_value FROM bookings WHERE booking_status = 'Success';`
19. **List all incomplete rides along with the reason:**

```
SELECT booking_id, incomplete_rides_reason FROM bookings WHERE
incomplete_rides = 'Yes';
```

Insights:

1. **Booking Success vs. Cancellations:**
   - A significant portion of the rides are successfully completed, but cancellations (both customer and driver-related) still represent a large percentage of the total rides. Understanding the reasons behind cancellations can help improve customer retention and reduce lost business.
2. **Top Customers:**
   - The top customers who book the most rides are crucial for identifying high-value segments. Focusing on these customers for loyalty rewards or offering personalized services could drive further growth.
3. **Payment Methods:**
   - The high usage of UPI as a payment method suggests that UPI is a preferred choice for users, and Ola can consider optimizing the user experience for UPI payments or offering exclusive discounts for UPI transactions.
4. **Vehicle Type Performance:**
   - Some vehicle types, like the "Prime Sedan," might have different ride distances and ratings compared to others. This highlights the need to analyze each vehicle type separately for service improvements.
5. **Incomplete Rides:**
   - A high number of incomplete rides with unclear or undefined reasons ("null" or "Other Issue") suggests an area that needs more detailed tracking. Clear categorization of these issues would help address the underlying problems more efficiently.
6. **Ride Duration and Booking Value Trends:**
   - By calculating the average ride duration and booking value for each day of the week, Ola can optimize scheduling and pricing. For instance, rides with longer durations and higher booking values may occur on specific days or times, which can inform peak hour strategies.
7. **Driver Cancellations:**
   - Cancellations by drivers due to "Personal & Car related issues" suggest a need for driver support programs, or perhaps adjustments to the criteria for when drivers can cancel rides.

Recommendations:

1. **Focus on Driver Support:**
   - With a significant number of cancellations related to drivers' personal and car-related issues, offering incentives, maintenance support, or even training can reduce cancellations and improve service reliability.

2. **Enhance Customer Retention:**
   - o Identify and target top customers for loyalty programs. Personalized rewards or benefits based on booking frequency could boost customer retention.
3. **Improve Ride Completion:**
   - o Investigate the reasons behind "incomplete rides" further, especially focusing on why rides are being categorized as 'Other Issue.' This can provide valuable insights into operational improvements.
4. **Optimize Vehicle Types:**
   - o Evaluate performance and customer satisfaction for each vehicle type. If certain vehicles have lower ratings, they may need a targeted improvement plan, such as driver training or vehicle maintenance.
5. **Promote UPI Payments:**
   - o Since UPI is a popular payment method, Ola could offer incentives or discounts specifically for UPI transactions to drive more usage of the payment gateway.
6. **Data-Driven Scheduling:**
   - o Use data on average ride durations and booking values by day of the week to optimize peak time pricing and scheduling. For example, adjusting pricing based on demand during specific times can maximize revenue.
7. **Enhance Data Categorization:**
   - o Improving the categorization of incomplete rides (e.g., removing "null" or "Other Issue") will help generate more actionable insights for operational improvements.

Conclusion:

This project analyzed ride booking data and generated meaningful business insights regarding ride statuses, cancellations, customer preferences, and payment trends. The recommendations derived from these insights can guide Ola's efforts to improve customer satisfaction, driver performance, and operational efficiency.