The purpose of this project is to analyze NCR ride booking data to uncover patterns in customer behavior, ride demand, cancellations, revenue, and operational efficiency. The dataset contains detailed information on bookings, ride outcomes, vehicle types, ratings, distances, and payment behaviors. Through SQL analysis, the goal is to identify key insights that can help ride-hailing platforms optimize driver allocation, reduce cancellations, understand peak usage times, and improve overall customer experience.

Q1. What is the distribution of ride booking statuses in the dataset?

```
SELECT booking_status, COUNT(*)
FROM ncr_ride_bookings
GROUP BY booking_status;
```

booking_status	COUNT(*)
No Driver Found	3450
Incomplete	2926
Completed	30145
Cancelled by Driver	8625
Cancelled by Customer	3418

Completed rides dominate the dataset, followed by driver and customer cancellations. "No Driver Found" and "Incomplete" rides form a smaller share. This shows overall platform reliability but also hints at operational inefficiencies in cancellations.

Q2. What are the top reasons customers cancel their rides?

```
SELECT cancel_reason_customer, COUNT(*)
FROM ncr_ride_bookings
WHERE cancel_reason_customer IS NOT NULL AND cancel_reason_customer <> "
GROUP BY cancel_reason_customer;
```

cancel_reason_customer	COUNT(*)
Driver is not moving towards pickup location	737
Driver asked to cancel	738
AC is not working	356
Change of plans	785
Wrong Address	802

The top customer cancellation reasons are mainly due to driver behavior (not moving toward pickup, asking to cancel) and changes in customer plans. This suggests improvement needed in driver compliance and tracking accuracy

Q3. What are the major reasons drivers cancel the rides?

```
SELECT cancel_reason_driver, COUNT(*)
FROM ncr_ride_bookings
WHERE cancel_reason_driver IS NOT NULL AND cancel_reason_driver <> "
```

GROUP BY cancel_reason_driver;

cancel_reason_driver	COUNT(*)
Personal & Car related issues	2222
Customer related issue	2158
More than permitted people in there	2099
The customer was coughing/sick	2146

Drivers mostly cancel rides due to personal or vehicle issues, customer-related problems, and overload of passengers.

Q4. What are the most common reasons for incomplete rides?

SELECT incomplete_ride_reason, COUNT(*)
FROM ncr_ride_bookings
WHERE incomplete_ride_reason IS NOT NULL AND incomplete_ride_reason <> "
GROUP BY incomplete_ride_reason;

incomplete_ride_reason	COUNT(*)
Vehicle Breakdown	981
Other Issue	940
Customer Demand	1005

Incomplete rides are mostly due to customer demand changes, followed by breakdowns and miscellaneous issues.

Q5. Which vehicle type has the highest demand?

SELECT vehicle_type, COUNT(*)
FROM ncr_ride_bookings
GROUP BY vehicle_type;

vehicle_type	COUNT(*)
eBike	3452
Go Sedan	8796
Auto	12183
Premier Sedan	5863
Bike	7294
Go Mini	9603
Uber XL	1373

Auto, Go Mini, and Go Sedan have the highest demand, indicating customer preference for affordable and mid-range options. Uber XL has the lowest demand.

Q6. What is the distribution of payment methods used by customers?

SELECT payment_method, COUNT(*)
FROM ncr_ride_bookings

```
GROUP BY payment_method;
```

payment_method	COUNT(*)
NULL	15493
UPI	15015
Debit Card	2684
Cash	8261
Uber Wallet	3871
Credit Card	3240

UPI is the most popular completed payment method, but many entries have NULL, suggesting incomplete rides or missing payment data. Cash remains widely used, followed by cards and Uber Wallet.

Q7. Which pickup locations have the highest number of rides? (Top 10)

```
SELECT pickup_location, COUNT(*) AS rides
FROM ncr_ride_bookings
GROUP BY pickup_location
ORDER BY rides DESC
LIMIT 10;
```

pickup_location	rides
Shivaji Park	320
Mehrauli	319
Ambience Mall	315
Jasola	313
Saket	313
Tughlakabad	310
Dwarka Sector 21	309
Pragati Maidan	308
Nehru Place	306
Rithala	305

High-traffic pickup locations include Shivaji Park, Mehrauli, and Ambience Mall. These hotspots indicate busy residential and commercial zones with consistently high demand.

Q8. Which drop locations have the highest number of rides? (Top 10)

```
SELECT drop_location, COUNT(*) AS rides
FROM ncr_ride_bookings
GROUP BY drop_location
ORDER BY rides DESC
LIMIT 10;
```

drop_location	rides
Basai Dhankot	328

Kalkaji	320
Noida Extension	317
Moolchand	307
Gurgaon Sector 56	306
Nawada	306
Sikanderpur	306
Connaught Place	305
Greater Noida	304
Dilshad Garden	303

Basai Dhankot, Kalkaji, and Noida Extension are the most common drop locations. Drop hotspots align with major residential/business areas, showing regular inter-city/zone movement patterns.

Q9. What are the peak hours for ride bookings?

SELECT HOUR(ride_time) AS hour, COUNT(*) AS total_bookings
FROM ncr_ride_bookings
GROUP BY HOUR(ride_time)
ORDER BY total_bookings DESC;

hour	total_bookings
18	3982
19	3612
17	3556
20	3232
16	3124
10	3070
15	2700
11	2692
9	2671
21	2652
14	2240
8	2237
12	2165
13	1807
7	1746
22	1696
6	1300
5	928
23	883
0	468
3	463
1	456
2	447
4	437

Peak booking hours are between 5 PM and 8 PM, aligning with evening commute traffic. Late-night and early-morning rides are significantly fewer

Q10. What is the average booking value of rides?

SELECT AVG(booking_value) FROM ncr_ride_bookings;

AVG(booking_value)
504.844305

The average booking value is approximately ₹505, indicating moderate pricing and a mix of short and medium-distance rides.

Q11. What is the average distance traveled per ride?

SELECT AVG(ride_distance) FROM ncr_ride_bookings;

AVG(ride_distance)
24.707713

The average ride covers about 24.7 km, suggesting many inter-area or long-route rides within the NCR region.

Q12. What are the average driver and customer ratings?

SELECT
    AVG(driver_rating) AS avg_driver_rating,
    AVG(customer_rating) AS avg_customer_rating
FROM ncr_ride_bookings;

avg_driver_rating	avg_customer_rating
4.23174	4.40454

Customer rating averages (4.40) are slightly higher than driver ratings (4.23), showing generally positive experiences on both sides with minor service gaps.

Q13. What is the cancellation rate for customers and drivers?

SELECT
    (SUM(booking_status = 'Cancelled by Customer') / COUNT(*)) * 100 AS customer_cancel_rate,
    (SUM(booking_status = 'Cancelled by Driver') / COUNT(*)) * 100 AS driver_cancel_rate,
    (SUM(booking_status LIKE 'Cancelled%') / COUNT(*)) * 100 AS total_cancel_rate
FROM ncr_ride_bookings;

customer_cancel_rate	driver_cancel_rate	total_cancel_rate
7.0381	17.7601	24.7982

Driver cancellations (17.76%) are significantly higher than customer cancellations (7.03%). Total cancellation rate is nearly 25%, indicating operational challenges that need attention.

Q14. What percentage of rides are incomplete?

SELECT
  (SUM(booking_status = 'Incomplete') / COUNT(*)) * 100 AS incomplete_rate
FROM ncr_ride_bookings;

incomplete_rate
6.025

About 6% of rides are incomplete, showing a relatively low but notable rate of mid-ride failures due to issues like breakdowns or customer decisions.

Q15. Which vehicle type generates the highest total revenue?

SELECT vehicle_type, SUM(booking_value) AS total_revenue
FROM ncr_ride_bookings
GROUP BY vehicle_type;

vehicle_type	total_revenue
eBike	1174086
Go Sedan	3033533
Auto	4165584
Premier Sedan	2020102
Bike	2506101
Go Mini	3337659
Uber XL	458641

Auto, Go Mini, and Go Sedan generate the highest revenue. Uber XL contributes the least due to lower ride volume.

Q16. What is the average ride distance for each vehicle type?

SELECT vehicle_type, AVG(ride_distance) AS avg_distance
FROM ncr_ride_bookings
GROUP BY vehicle_type;

vehicle_type	avg_distance
eBike	24.832713
Go Sedan	24.679813
Auto	24.734861
Premier Sedan	24.598689
Bike	24.733216
Go Mini	24.774262
Uber XL	24.197135

Average distance is similar across vehicle types (around 24–25 km), showing consistent usage patterns regardless of vehicle category.

Q17. Who are the most loyal customers based on the number of rides taken? (Top 10)

```
SELECT customer_id, COUNT(*) AS total_rides
FROM ncr_ride_bookings
GROUP BY customer_id
ORDER BY total_rides DESC
LIMIT 10;
```

customer_id	total_rides
CID6468528""	3
CID7828101""	3
CID4757284""	2
CID3470774""	2
CID8137475""	2
CID9950083""	2
CID2721563""	2
CID7709961""	2
CID7267162""	2
CID6609697""	2

Most loyal customers have taken 2–3 rides each, indicating low repeat-ride frequency in the dataset, possibly due to large user base and single-usage patterns.

Q18. Which customers generate the highest total revenue? (Top 10)

```
SELECT customer_id, SUM(booking_value) AS total_spent
FROM ncr_ride_bookings
GROUP BY customer_id
ORDER BY total_spent DESC
LIMIT 10;
```

customer_id	total_spent
CID7828101""	4722
CID2706299""	4277
CID2978596""	4220
CID5789715""	4133
CID9539119""	4109
CID2674107""	4093
CID1753183""	4088
CID4542328""	4032
CID8725190""	3985
CID7198247""	3984

Top-paying customers spent between ₹3,800–₹4,700. Their higher expenditure suggests frequent long-distance rides or premium vehicle usage.

The SQL analysis of NCR ride bookings reveals clear trends in customer usage, operational challenges, and business performance. Completed rides form the majority of activity, with peak demand occurring between 5 PM and 8 PM. Autos, Go Mini, and Go Sedan emerge as the most preferred vehicle types, while UPI and Cash dominate payment methods. Cancellation patterns show that both customers and drivers contribute significantly, primarily due to navigation issues, customer behavior, or vehicle-related concerns. Revenue analysis highlights Go Mini, Auto, and Go Sedan as the top contributors. Overall, these insights can support better fleet management, improve customer satisfaction, reduce cancellations, and guide strategic decisions for ride-hailing operations in NCR.