

Cab Booking System

ANALYTICS

Presenter: Nilesh Yadav



Project Introduction & Overview

📌 Introduction:

- The cab booking industry has experienced growth through digital platforms that connect passengers with available drivers. These platforms generate vast amounts of data related to customer bookings, trip routes, cancellations, driver performance, and revenue.
- This project show how structured database design and SQL-based data analysis can help uncover actionable insights from cab booking operations.



📈 Project Purpose:

To analyze cab booking data and support strategic decision-making across various domains - including customer behavior, driver performance, revenue trends, and operational efficiency.

🔍 What We Did:

- Designed a relational database schema for a cab booking system.
- Inserted sample data on customers, bookings, drivers, and trips.
- Ran SQL queries to extract trends and patterns.
- Interpreted key metrics to propose business improvements.

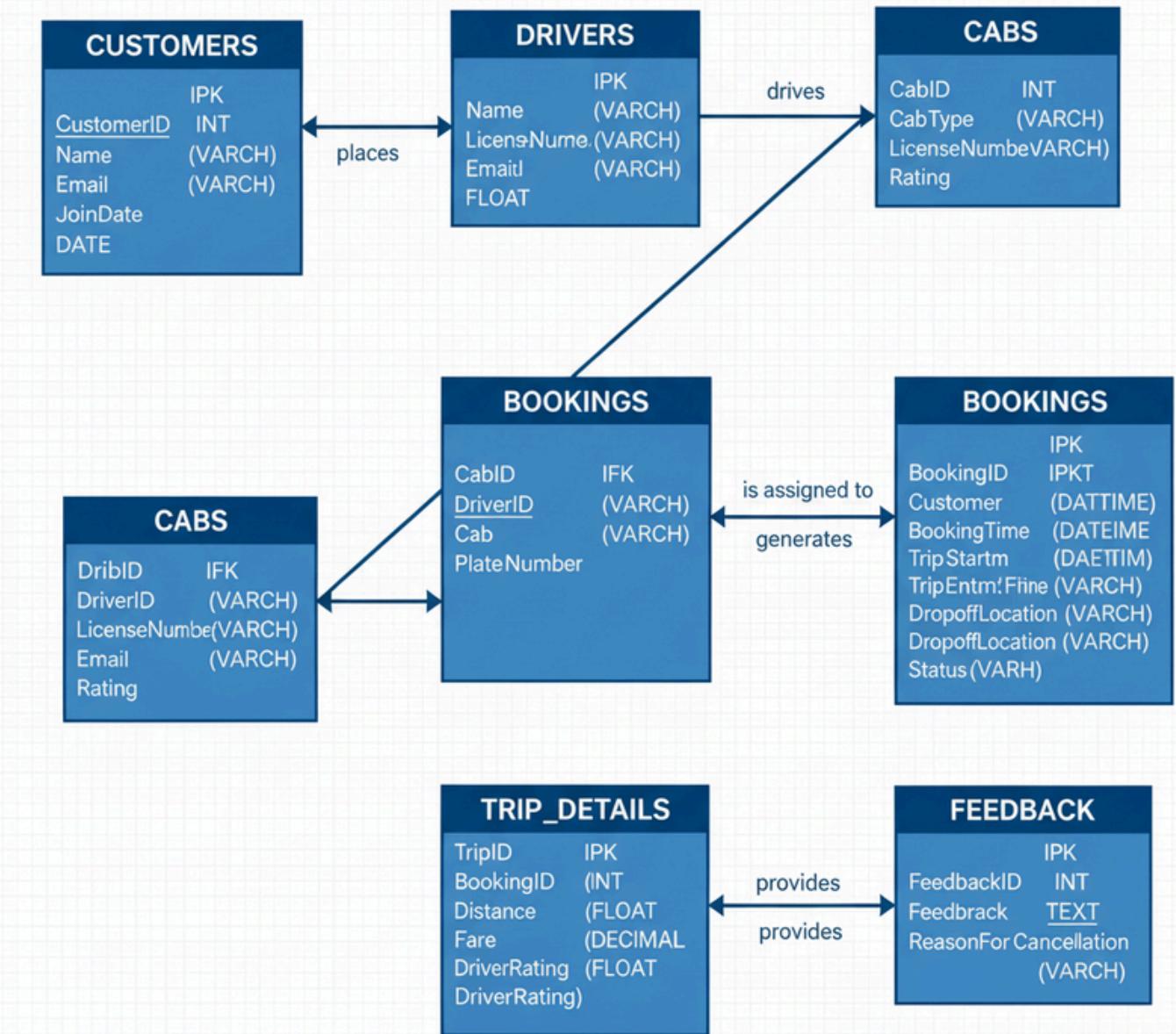
Entity-Relationship (ER) Diagram

The Entity-Relationship (ER) diagram represents the structure of the Cab Booking System database. It identifies six core entities: Customers, Drivers, Cabs, Bookings, Trip_Details, and Feedback. These entities are interrelated to capture all operational aspects of a cab booking platform.

- A Customer can place multiple Bookings.
- Each Booking is associated with a Cab and indirectly linked to a Driver who operates it.
- Every confirmed ride generates an entry in Trip_Details, recording the distance, fare, and customer rating.
- In the event of cancellations or feedback, entries are recorded in the Feedback table for performance tracking and service improvement.

This structured, relational schema allows for efficient data storage, consistency, and powerful analytics using SQL to drive smarter business decisions.

Cab Booking System ER Diagram



Key Analytical Questions Driving the Project

1. Identify customers who have completed the most bookings.
What insights can you draw about their behavior?

```
select name, count(*) as completed_booking  
from customers c  
join bookings b on c.customerid = b.customerid  
where status = 'completed'  
group by name  
order by completed_booking desc;
```

name	completed_booking
Rajan Pandey	2
Rakesh Yadav	1

- 📌 In this project, various SQL queries were designed to solve real-world business problems in the cab booking system. The goal was to extract meaningful data insights that can help the company improve customer satisfaction, monitor driver performance, increase revenue, and optimize operations. Below are some key questions that guided the data analysis along with their SQL implementations.

- 📌 Identifying the most active customers helps the business recognize loyal users and provide targeted benefits (e.g. promo codes, VIP status).

Customer and Booking Customer Cancellation Pattern Analysis

Find customers who have canceled more than 30% of their total bookings. What could be the reason for frequent cancellations?

🔍 Root Cause Analysis:

From the Feedback Table:

Jay Tiwari → Reason: “Driver Delay”

Harsh Singh → Reason: “Change of plans”

```
select c.customerid, name, count(case when status = 'cancelled' then 1 end) * 1.0 / count(bookingid) as cancelledbooking  
from customers c  
join bookings b on c.customerid = b.customerid  
group by c.customerid  
having cancelledbooking > 0.3;
```

	customerid	name	cancelledbooking
▶	2	Jay Tiwari	1.00000
	4	Harsh Singh	1.00000

- 📌 This analysis identifies customers who have canceled more than 30% of their total bookings. High cancellation rates can negatively affect resource planning, driver performance, and revenue. Understanding the reasons behind frequent cancellations—whether it's due to service issues or customer-side behavior—allows the company to take targeted action to improve reliability and retention.

Problem/Behavior	Recommended Action
Driver-related cancellations	Improve driver ETA accuracy. Use live tracking alerts.
Customer change of plans	Offer flexible cancellation or reschedule option.
Lack of feedback	Make feedback mandatory after cancellation.

Determine the busiest day of the week for bookings. How can the company optimize cab availability on peak days?

```
select dayname(bookingtime) as busiest_weekday, count(*) as total_booking  
from bookings  
group by busiest_weekday  
limit 1;
```

	busiest_weekday	total_booking
▶	Thursday	2

- Analyzing booking patterns by day of the week helps identify peak days when cab demand is at its highest. This analysis is useful for optimizing cab availability, driver scheduling, and pricing strategies. The SQL query below groups all bookings by weekday and reveals which day experiences the most demand.
- From our dataset, the busiest day for bookings is Thursday, indicating that mid-week operational planning should be prioritized to meet demand effectively.

Insight	Action Plan
Thursday = highest demand	Deploy more cabs and maximize driver shifts on Thursdays
Mid-week travel peak	Offer weekday loyalty rides or time-based discounts
Avoid underwhelming service	Send driver reminders for peak day shift confirmation

Low-Rated Driver Performance Analysis

- 📌 In this analysis, we attempted to identify underperforming drivers by checking for those who received an average trip rating below 3.0 in the last 3 months. The query returned no results, which indicates that all current active drivers have maintained a satisfactory service level according to customer feedback. This is a strong indicator of consistent driver quality and successful past interventions.

✓ Summary:

- No low ratings found ✓
- Drivers are performing well ✓
- Recommend proactive monitoring 🧠

Identify drivers who have received an average rating below 3.0 in the past three months. What strategies can be implemented to improve their performance?

```
select d.driverid, d.name, round(avg(t.driverrating), 2) as avg_rating, count(t.tripid) as total_trip  
from drivers d  
join cabs c on d.driverid = c.driverid  
join bookings b on c.cabid = b.cabid  
join tripdetails t on b.bookingid = t.bookingid  
where b.tripendtime >= date_sub(curdate(), interval 3 month) and b.status = 'completed'  
group by d.driverid, d.name  
having avg(t.driverrating) < 3.5  
order by avg_rating asc;
```

driverid	name	avg_rating	total_trip
----------	------	------------	------------

✓ Actionable Insights:

Recommendation	Description
⚠ Early Warning System	Set alerts when a driver's rating falls below 3.5
✓ Monthly Driver Scorecards	Regularly monitor performance trends
📝 Feedback Classification	Categorize reasons for low ratings (behavior, cleanliness, etc.)

Top 5 Long-Distance Drivers – Work Pattern Analysis

Find the top 5 drivers who have completed the longest trips in terms of distance. What does this say about their working patterns?

To understand driver work patterns and productivity, we analyzed which drivers have covered the longest total distance through completed trips. These drivers typically engage in long-distance bookings that are often higher-paying and indicate a strong work ethic, time availability, or strategic positioning in zones with high-value ride requests.

The query below identifies the top 5 such drivers, offering insights into their performance trends and how they can be used as a benchmark for others.

🔍 Interpretation / Insight:

- Alina Kapoor has completed the most kilometers – likely prefers or specializes in longer trips.
- Such drivers are often:
 - Positioned in high-demand zones (like airports or business districts)
 - Available for longer durations
 - Preferred by customers for high-rated, comfortable rides (especially SUVs)

```
select d.driverid,d.name,sum(t.distance) as totaldistance  
from drivers d  
join cabs c on d.driverid = c.driverid  
join bookings b on c.cabid = b.cabid  
join tripdetails t on b.bookingid = t.bookingid  
where b.status = 'completed'  
group by d.driverid,d.name  
order by totaldistance desc  
limit 5;
```

	driverid	name	totaldistance
▶	4	Alina Kapoor	15
	1	Raj Singh	12.5
	3	Anshu P	10

Observation	Recommendation
High distance drivers = high value	Set alerts when a driver's rating falls below 3.5
Concentration in long routes	Strategically allocate these drivers in long-trip zones
Preferred cab & driver profiles emerge	Use for VIP or corporate ride dispatch prioritization

6-Month Revenue Trend Analysis from Completed Bookings

Calculate the total revenue generated by completed bookings in the last 6 months. How has the revenue trend changed over time?

- 📌 This analysis calculates the total revenue earned from completed trips over the past 6 months. By grouping data month-wise, we can observe fluctuations in revenue and trip volumes over time. This helps the business evaluate growth trends, seasonal demand, and operational performance.

- Our SQL query aggregates two key metrics: number of completed trips and total fare, helping us understand both frequency and financial contribution per month.

- 🔍 Interpretation:
 - All revenue was earned in May, with zero trips in June to October (based on current dataset).
 - This indicates that either:
 - No trips were recorded in that period
 - Or sample data is limited
 - Still, the structure provides a working model for trend tracking with live production data.

```
select month(b.tripendtime) as month,
       count(t.tripid) as completed_trips,
       sum(t.fare) as total_revenue
  from bookings b
  join tripdetails t on b.bookingid = t.bookingid
 where b.status = 'completed'
   and b.tripendtime >= date_sub(curdate(),interval 6 month)
 group by month
 order by month asc;
```

	month	completed_trips	total_revenue
▶	5	3	750.00

Observation	Recommendation
High distance drivers = high value	Set alerts when a driver's rating falls below 3.5
Concentration in long routes	Strategically allocate these drivers in long-trip zones
Preferred cab & driver profiles emerge	Use for VIP or corporate ride dispatch prioritization

Top 3 Most Frequently Traveled Routes – Route Optimization Strategy

Identify the top 3 most frequently traveled routes based on PickupLocation and DropoffLocation. Should the company allocate more cabs to these routes?

- 📌 Analyzing frequent travel routes helps the company identify high-demand corridors. By tracking the number of completed rides and total revenue generated for each unique pickup and drop-off location pair, we can suggest more efficient cab allocation.
- This query reveals the most popular routes among customers and supports better planning for cab deployment, especially during high-demand hours or zones.
-

🔍 Interpretation:

- The Mall → University route generated the most revenue in a single trip.
- Downtown appears frequently as a pickup hub, suggesting it's a central location for demand flow.
- These patterns help pinpoint where cab supply needs to be proactively managed.

```
select b.pickuplocation,b.dropofflocation,count(*) as total_trips,  
sum(t.fare) as total_revenue  
from bookings b  
join tripdetails t on b.bookingid = t.bookingid  
where b.status = 'completed'  
group by b.pickuplocation,b.dropofflocation  
order by total_trips desc  
limit 3;
```

	pickuplocation	dropofflocation	total_trips	total_revenue
▶	Downtown	Airport	1	250.00
	Downtown	Hospital	1	200.00
	Mall	University	1	300.00

Insight	Recommendation
Repeated high-frequency routes	Allocate more cabs in those corridors during peak hours
High average fare zones	Focus SUV/Sedan variants in premium demand areas
Pickup hotspots	Place surge pricing + higher driver incentives

Average Waiting Time by Pickup Location – Reducing Delays through Optimization

- Analyze the average waiting time for different pickup locations. How can this be optimized to reduce delays?**

- 📌 One of the key factors affecting customer satisfaction is the waiting time between when a ride is booked and when it actually begins. This metric indicates how efficiently drivers are assigned and dispatched.

- This analysis calculates the average waiting time by pickup location, helping the company identify zones where delays are more frequent. The goal is to use this insight for better cab positioning, pre-scheduling, and traffic-aware routing in high-delay areas.



Interpretation:

- Mall and Downtown have longer average waiting times compared to other areas.
- Indicates possible delay in driver assignment or cab availability in these zones.

```
select b.pickuplocation, count(*) as total_trips, avg(timestampdiff(minute, b.bookingtime, b.tripstarttime)) as avg_waiting_minutes  
from bookings b  
where b.status = 'completed'  
group by b.pickuplocation  
order by avg_waiting_minutes desc;
```

	pickuplocation	total_trips	avg_waiting_minutes
▶	Mall	1	15.0000
	Downtown	2	12.5000

Identified Issue	Optimization Strategy
High wait time at Mall / Downtown	Pre-position drivers during peak periods for faster pickups
Delay in peak areas	Use predictive analytics to dispatch cabs ahead of demand
Shortage of drivers	Offer driver incentive bonus during delay-prone slots

Cancellation Analysis – Understanding Why Trips Are Canceled

Identify the most common reasons for trip cancellations from customer feedback.

What actions can be taken to reduce cancellations?

- Analyzing the reasons for trip cancellations helps uncover service-related issues and patterns in customer behavior. By aggregating feedback from canceled bookings, we can classify the most common causes of cancellations, such as driver delays or customer-side issues. Understanding these reasons allows the platform to develop targeted strategies to enhance service reliability and reduce booking abandonment.

🔍 Insight:

- 50% of cancellations were caused by driver delays - highlighting a critical operational issue.
- The rest were caused by customer-side reasons, such as last-minute plan changes.

This showcases the need to differentiate between driver-induced and customer-induced cancellations in future analytics and interventions.

```
select f.reasonforcancellation, count(*) as total_cancellations  
from feedback f  
join bookings b on f.bookingid = b.bookingid  
where b.status = 'cancelled'  
group by f.reasonforcancellation  
order by total_cancellations desc;
```

	reasonforcancellation	total_cancellations
▶	Driver Delay	1
	Customer Personal Reason	1

Issue Identified	Suggested Action
Driver Late Arrival	Improve driver ETA tracking & send automated delay alerts
Customer Plan Change	Offer flexible rescheduling or ride pause option
Cancellations Without Reason	Make feedback mandatory before allowing to cancel

Weekend vs Weekday Bookings – Demand-Based Dynamic Pricing

- Analyze whether weekend bookings differ significantly from weekday bookings. Should the company introduce dynamic pricing based on demand?
- Understanding how ride demand and revenue vary between weekdays and weekends enables the company to implement strategic pricing and resource allocation. This analysis groups trip data by day type – weekday vs weekend – and compares booking volume, total revenue, and average fare per trip.
- The goal is to evaluate whether weekends generate higher demand or higher-paying rides, which would justify implementing dynamic pricing and driver incentive structures during peak periods.

```
select case when dayofweek(b.bookingtime) in (1,7) then "weekend"
else "weekday"
end as day_type,
count(b.bookingid) as total_bookings,
sum(t.fare) as total_revenue,
avg(t.fare) as avg_fare_per_trip
from bookings b
join tripdetails t on b.bookingid = t.bookingid
where b.status = 'completed'
group by day_type;
```

day_type	total_bookings	total_revenue	avg_fare_per_trip
weekday	2	450.00	225.000000
weekend	1	300.00	300.000000

🔍 Interpretation:

- Fewer trips on weekends, but they generate higher average fare per trip.
- Weekdays have more consistent volume but slightly lower per-trip revenue.

Issue Identified	Suggested Action
Weekend trips earn more money per trip	Use slightly higher prices on weekends (dynamic pricing)
Fewer rides on weekends	Give offers or discounts to get more bookings
Weekdays have more regular bookings	Keep good number of drivers available on weekdays

Conclusion & Key Takeaways

- This project used SQL to analyze a cab booking system and uncovered valuable insights in customer behavior, driver performance, booking trends, and revenue generation. Based on these insights, we suggested practical actions to improve service quality, customer retention, and profit margins.

Project Summary:

- Identified loyal customers and frequent cancelers
- Found peak booking days and top travel routes
- Analyzed driver performance and distance trends
- Tracked revenue trends and weekend booking patterns
- Provided strategies for optimization and profitability



THANK YOU

For Your Attention
Any Questions?

