Exploring Customer Behavior Through Database Analysis

SQL Case Study

A Real-Life Data Analyst Case Study


ORDER

INTRODUCTION

As a data analyst, one of the most intriguing aspects of my job is uncovering insights about customer behavior through database analysis. Recently, I had the opportunity to work on a Case Study that involved analyzing multiple tables in a database to gain a better understanding of customer behavior for a popular online retailer.

The database consisted of five tables: Sales, Product, Goldusers_Signup, Users, and User_name. Each table contained unique information about the customers and products sold on the website.

The Sales table provided details about each Sale, including the date of the sale, the product sold, the price, and the user who made the purchase. By analyzing this table, we were able to identify the top-selling products and the most active users on the website.

The product table contained information about each product sold on the website, including the product ID, Product Name and price of a Product. .

Using this table, we were able to identify the prices for each product and compare them to determine which products are more expensive and which are less expensive.

The goldusers_signup table contained data about customers who signed up for the website's premium membership program. We were able to use this table to identify the demographics of customers who were most likely to sign up for the program and which benefits of the program were the most appealing to customers.

The users table provided information about each customer who created an account on the website, including their name, Signup Date. By analyzing this table, we were able to identify trends in customer demographics.

The user_name table contains information about each user's UserID and User_Name. We utilized this table to link with other tables and use their User ID to join with other tables. Specifically, the table includes the

UserID and the corresponding Names of the users. The UserID can be used as a linking factor to connect with other tables in the database. The Names column provides the name of each user corresponding to their UserID.

By analyzing these tables in conjunction with each other, we were able to gain a comprehensive understanding of customer behavior on the website. We were able to identify the top-selling products, the most active users, and the demographics of customers who were most likely to sign up for the premium membership program. This information allowed us to make data-driven decisions about marketing, product development, and customer engagement.

Overall, this Case Study was a fascinating exploration of customer behavior through database analysis. By leveraging the power of data, we were able to uncover insights that would have been impossible to discover through intuition alone. Let Start with some Problem Statement.

Problem Statement

I wants to use the data to answer a few questions about the customers, especially about their visiting patterns, how much money they've spent and also which menu items are their favourite. Having this deeper connection with his customers will help him deliver a better and more personalised experience for his loyal customers.

Based on the insights gained from the data, the client plans to decide whether to expand the existing customer loyalty program. Additionally, they require assistance in generating some basic datasets so that their team can easily inspect the data without the need for using SQL.

This problem statement was solved using SQL, which was used to analyze and query the database to generate insights and datasets. SQL was utilized to extract the necessary data and generate reports to provide insights into the customer behavior, sales trends, and product performance, which helped the client make data-driven decisions

regarding the loyalty program. SQL was also used to create datasets that could be easily accessed and analyzed by the client's team, allowing them to understand the data without needing to have expertise in SQL.

I has shared with you 5 key datasets for this case study:

Sales

Product

Username

Users

Gold users_Signup

Entity Relationship Diagram

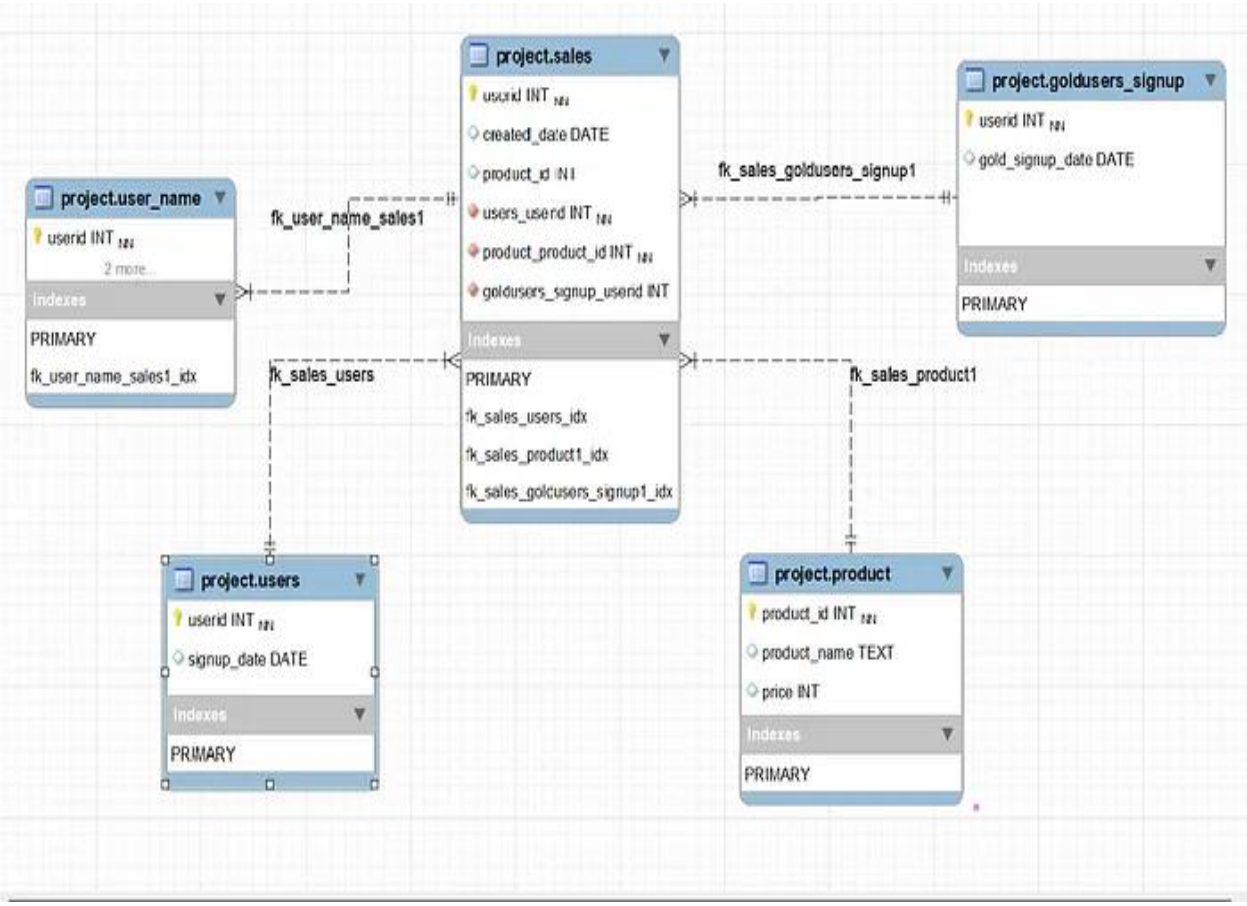


Table of Content

Data Summary

Table 1: Sales

The Sales table consists of 62 rows and 3 columns: user_id, created_date, and product_id. The product_id ranges from 1 to 11, and the user_id ranges from 1 to 10. This table provides information on which user_id has confirmed an order, which product they have purchased, and the date on which the order was placed.

CoachX LIVE

Sales

userid	created_date	product_id
1	19-04-2017	2
3	18-12-2019	1
2	20-07-2020	3
1	23-10-2019	2
1	19-03-2018	3
3	20-12-2016	2
1	09-11-2016	1
1	20-05-2016	3
2	24-09-2017	1
1	11-03-2017	2
1	11-03-2016	1
3	10-11-2016	1
3	07-12-2017	2
3	15-12-2016	2
2	08-11-2017	2
2	10-09-2018	3
4	01-05-2019	1
5	23-11-2018	3
6	30-06-2017	9
7	12-08-2018	8
8	19-03-2019	7
9	04-12-2017	6
10	22-09-2018	2
4	17-08-2020	1
5	12-05-2017	10
6	27-01-2014	11
7	02-04-2014	7
8	15-12-2020	8
9	08-09-2017	8

Table 2: Product

The second table provides more information about the products that are being sold in the first table. It includes 11 rows, each representing a unique product, with columns for the product ID, product name, and price. The product IDs range from 1 to 11 and correspond to the product IDs in the first table.

Each product has a unique name that describes the dish, such as Dal Makhani, Butter Chicken, or Fish Curry. The prices of the products range from 80 to 450, with Mango Lassi being the least expensive and Mutton Biryani being the most expensive.

This information can be used to analyze the sales data in the first table, to determine which products are the most popular and which products generate the most revenue. It can also be used to make decisions about pricing and product offerings in the future.

CoachX LIVE

Product

product_id	product_name	price
1	Dal Makani	160
2	Shahi Panner	170
3	Butter Chicken	340
4	Aloo Gobi	150
5	Chole Bhature	100
6	Fish Curry	380
7	Chicken Tikka	300
8	Mutton Biryani	450
9	Veg Pulao	200
10	Mango Lassi	80
11	Gulab Jamun	100

Table 3: User Name

The table Below shows a list of user IDs and corresponding names. There are 10 users in total, each with a unique user ID ranging from 1 to 10. The names of the users are listed in the column next to their respective user ID. This table could be used to track user information in a database or to provide a reference for user identification in a software system. The names provided in this table are fictional and are used for demonstration purposes only.

User_name

userid	Names
1	Anshul
2	Rohan
3	Shreya
4	Priya
5	Aryan
6	Sara
7	Sahil
8	Tanvi
9	Ritika
10	Gaurav

Table 4: Users

This table shows the signup dates for the users in the system. The dates range from 2013 to 2016, indicating that the system has been active for several years. It also shows that some users signed up on the same day, such as users 10 and 9, and users 7 and 8. This could be an indication that they signed up together, or that the system was launched on those dates. The signup dates could also be useful for analyzing user behavior, such as how long they have been active on the system and if there are any patterns in signup dates.

Users

userid	signup_date
1	02-09-2014
2	15-01-2015
3	11-04-2014
4	17-11-2015
10	02-01-2016
9	02-01-2016
7	02-04-2013
8	15-12-2013
5	08-09-2015
6	13-07-2014

Table 5: Golduser_Signup

The given table shows the gold_signup_date for a few users, where the users with their respective user IDs have upgraded to gold membership. Gold membership usually comes with additional perks and benefits like exclusive offers, faster shipping, and priority customer support, among others.

It can be observed that the gold_signup_date for the users ranges from 2017 to 2019, which indicates that they have been gold members for a while now. It’s interesting to note that not all users who signed up earlier have upgraded to gold membership, as seen in the case of user ID 2, who signed up in 2015 but is not a gold member. This could be due to various reasons like personal preference, lack of interest in the offered benefits, or inability to pay the additional fees for gold membership.

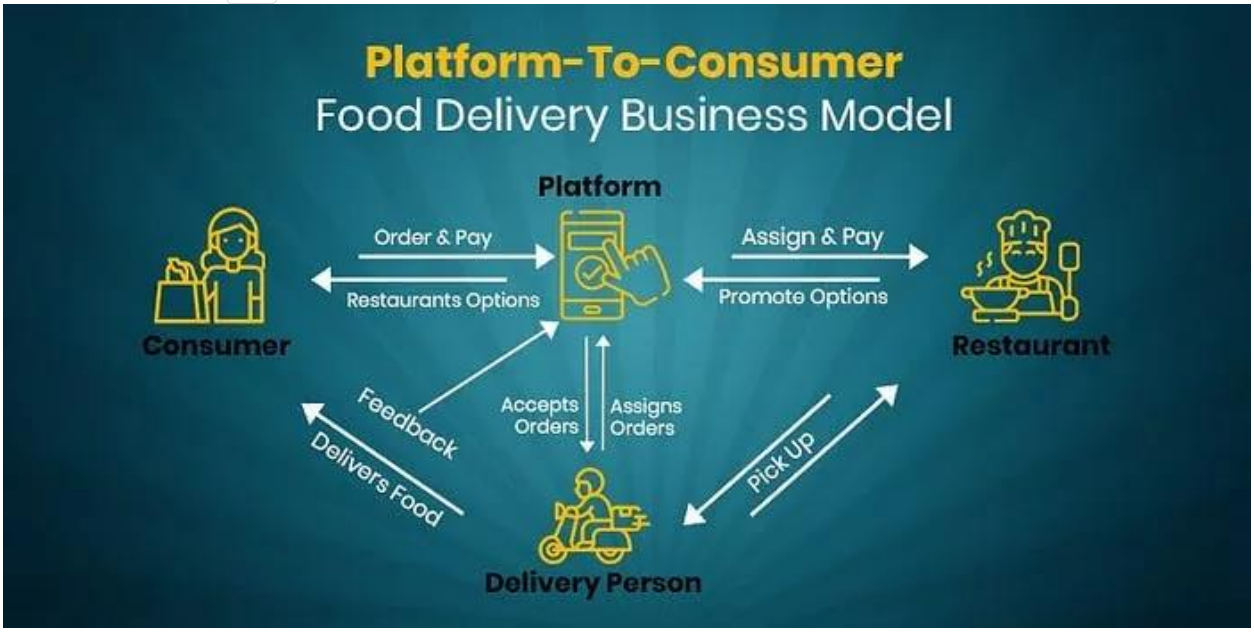
CoachX LIVE

Users

userid	signup_date
1	02-09-2014
2	15-01-2015
3	11-04-2014
4	17-11-2015
10	02-01-2016
9	02-01-2016
7	02-04-2013
8	15-12-2013
5	08-09-2015
6	13-07-2014

2. Summary

Based on the provided information, this is an analysis of an online food ordering platform. The analysis includes information about the sales made on the platform, the products available for purchase, and the users who have signed up and become gold members. The sales table contains 62 rows and includes columns for user ID, product ID, and order date, which can be used to track which users have ordered which products and when. There are 11 products available for purchase, each with a different price. The user table includes information about 10 users, including their names and signup dates. Finally, the gold member table includes information about five users who have signed up for a premium membership on the platform. Overall, this analysis provides a basic understanding of the platform's sales and user data, which can be used to identify trends and opportunities for growth.



3. Problems

Before looking at the solution, it is important to try and solve the problem yourself and explore different methods to solve this real-world problem. You can find the problem question on this view this page if you would like to attempt it yourself. For the dataset used in this project, you can access it on the same page. By attempting the problem yourself and exploring different solutions, you can gain a deeper understanding of the challenges faced by companies in the online delivery industry and the potential for data analysis to provide valuable insights.

Q1. What is the total sales revenue generated by each product?

To determine the total sales revenue generated by each product, we need to analyze the sales data in the database. By using SQL queries, we can aggregate the sales data by product and calculate the total revenue generated by each product.

Based on the information provided in the original table, we can calculate the total revenue generated by each product. Butter Chicken has generated the highest revenue of Rs. 2720, while Chole Bhature has generated the least revenue with only 3 sale

Q2. Which 3 product has the highest sales revenue?

To determine the 3 products with the highest sales revenue, we need to analyze the sales data in the database. By using SQL queries, we can aggregate the sales data by product and sort them by the revenue generated to identify the top-selling products.

The answer to this question cannot be determined based on the information provided in the original table. However, once we have access to the sales data, we can use SQL to calculate the total revenue generated by each product and sort them in descending order to identify the top 3 products with the highest sales revenue.

By identifying the top-selling products, the company can focus on promoting these products to generate more revenue and profits. This information can also be used to optimize the inventory and stock levels of these products to ensure that they are always available to customers.

Q3. How many users have signed up for the service and has taken the gold membership?

Knowing the number of users who have taken the gold membership can provide valuable insights into the success of the loyalty program and the overall business performance. A high number indicates an effective program leading to increased customer retention, repeat purchases, and positive referrals. A low number suggests a need for program improvement, which can be done by offering better rewards or simplifying the redemption process. This information can help businesses identify opportunities to improve the loyalty program and ultimately drive growth and profitability.

Q4. What is the revenue generated from gold users?

By analyzing the revenue generated from gold users, businesses can evaluate the impact of the loyalty program on their overall business performance. This information can help businesses make informed decisions about how to allocate resources to improve customer retention and increase revenue. Additionally, it can provide insights into which products or services are most popular among gold users and identify potential areas for expansion or improvement. Ultimately, understanding the revenue generated from gold users can help businesses maximize the effectiveness of their loyalty program and drive growth and profitability.

Q5. What is the total revenue generated from gold users?

If the revenue generated from gold users is significantly higher than that of non-gold users, then it suggests that the loyalty program is successful in driving sales and customer loyalty. On the other hand, if the revenue generated from gold users is not significantly different from that of non-gold users, then it may suggest that the rewards offered by the loyalty program are not attractive enough to incentivize customers to spend more.

Q6. Which users has been a gold user for the How much of time?

Identifying which users have been gold users for how long can provide valuable insights into the loyalty of customers and the effectiveness of the loyalty program. By analyzing this data, businesses can identify their first gold user and use this information to better understand their most loyal customers. Additionally, by knowing how long each gold user has been a member, businesses can identify trends in loyalty and use this information to optimize their loyalty program. For example, if the average time that customers remain gold users is low, it may indicate that the rewards being offered are not attractive enough to retain customers. On the other hand, if the average time that customers remain gold users is high, it may indicate that the loyalty program is successful in building strong customer relationships.

Q7. What is the most popular product among gold users?

Identifying the most popular product among gold users is crucial in determining which products to prioritize in our menu and marketing campaigns. Gold users are our premier customers, and their purchasing behavior can provide valuable insights into their preferences. By analyzing their order history, we can identify the product that is most frequently ordered by gold users, and use this information to develop strategies to increase sales and customer loyalty.

Q8. What is the total sales revenue generated in each year?

This is an important analysis to determine the pattern of sales over the years, identifying which years had the highest and lowest sales revenue. This information can help us understand which factors contribute to sales fluctuations and predict future sales trends.

Q9. How has the sales revenue changed over the years?

To analyze the change in sales revenue over the years, we can calculate the percentage increase or decrease in revenue each year. This will help us understand the trends in our business and identify which years have shown the most growth or decline in sales. By studying this data, we can make informed decisions about where to allocate resources and focus our efforts to maximize revenue in the future.

Q10. What is the average Gold-signup compare to just sign up for the users?

From an average perspective, we can understand what percentage of our total users have signed up for the gold membership compared to those who haven't, which can give us insights into the popularity and effectiveness of our loyalty program.

Q11. How many gold members users have order how many numbers of time ?

"We can analyze the difference in the number of orders made by gold members before and after signing up for the membership. This will help us understand whether there is a significant difference in ordering behavior and whether there are any weaknesses in our service that we need to address. Additionally, we can also identify how many times on average gold members have placed an order to assess their loyalty and engagement with our platform."

Q12. What is the total amount each customer spend on Online Food Delivery?

From this, we can understand the total amount that each customer spends on online food delivery, which can help us identify high-spending customers and tailor our services accordingly.

Q13. What is the frequency of customer visits to the online platform?

By analyzing the frequency of customer visits, we can gain insights into user engagement and loyalty towards the platform. We can use this information to identify areas for improvement, such as providing personalized recommendations, improving the user interface, or offering more discounts or loyalty rewards. Additionally, we can use this data to predict future customer behavior and make data-driven decisions.

Q14. What was the first order purchase by each customer ?

Knowing the first order purchase by each customer can help us understand their taste and preferences, which can help us personalize their experience and offer them relevant recommendations. This can lead to increased customer satisfaction and loyalty. Additionally, analyzing the data can help identify popular first orders and assist in developing marketing strategies to promote these items

Q15. What is the most purchase item on the menu and how many times was it purchased by all customers?

By identifying the most purchased item on the menu and the number of times it has been purchased by all customers, we can make informed decisions about which menu items to feature prominently on the platform. This can help increase sales and customer satisfaction by making it easier for customers to find and order their favorite items. Additionally, by analyzing the most popular menu items, we can identify trends in customer preferences and adjust our menu offerings accordingly to meet the changing demands of our customers.

Q16. Which item was most popular for each customer ?

Analyzing the most popular item for each customer can help us personalize their experience by recommending their favorite items or creating personalized offers. It can also help in predicting future demand for certain items and optimizing inventory management. Furthermore, it can provide insights into the taste preferences of our customers and help us develop targeted marketing strategies.

Q17. Which item was purchase first by the customer after they become a member ?

Analyzing the first order purchased by the customer after they become a gold member can give us insight into their taste preferences and what items they prioritize in their food choices. It can also give us an idea of what types of promotions or offers are most likely to attract gold members to make a purchase. Additionally, tracking the first purchase of gold members can help us identify any patterns or trends in their behavior that we can use to improve our services or offerings.

Q18.Which item was purchase before the customer become a member ?

By knowing the last product purchased before the customer became a member, we can analyze if there is any specific product that encouraged the customer to become a member. This can help us create targeted marketing strategies to attract more customers to become members.

Q19. What is the total orders and amount spent for each member before they become a member ?

It can also help in identifying potential loyal customers who are likely to become gold members in the future.

Q20. Rank all the transactions for each member whenever they are a XYZ gold member for every non gold member Transaction marks as na ?

Ranking all the transactions for each member during their gold membership can help us to identify the most popular items purchased by gold members, as well as their spending patterns during their gold membership period. This information can be used to optimize our menu, promotions, and discounts for gold members to increase their satisfaction and loyalty to our platform. Additionally, comparing the ranked transactions of gold members to non-gold members can help us identify any differences in spending patterns or preferences between the two groups, which can guide our marketing and sales strategies.

4.Conclusion

In conclusion, my Case Study on online delivery using SQL has successfully addressed some of the key queries and challenges faced by companies in the online delivery industry. By analyzing customer data, I have been able to provide insights into customer behavior, such as the frequency of orders, popular food items, and the effectiveness of the gold membership program. With this information, companies can make data-driven decisions to improve their services and increase customer satisfaction. My Case Study has also demonstrated the power and usefulness of SQL in handling large and complex datasets, allowing companies to extract meaningful insights from their data. Overall, my project has shown that leveraging data through SQL can be an effective tool for companies in the online delivery industry to stay competitive and provide better services to their customers.

Submission Guidelines

Format: PowerPoint or PDF
Length: 1-20 slides.
Sections: Introduction, Key Findings, Actionable, Methodologies , Approaches, Insights, Conclusions

Tools and Technologies :

SQL/ MS-SQL SERVER

Deadline:

Submit your report and presentation within 21 Days from the day you will start.