Project: Superstore Sales Data Analysis using SQL

Let me explain this project in a simple way.

So basically, retail companies collect a lot of sales data every day, but most of the time this data is raw and messy. Just having data doesn't help the business unless we actually understand **what is selling, who is buying, and where problems are happening**.

That's why I worked on this **Superstore Sales Analysis project using SQL**.

This project focuses on analyzing a retail sales dataset using SQL to understand product performance, customer purchasing behavior, and operational challenges like delays and transaction issues. The goal was to clean the data, run analytical queries, and extract insights that support better business decisions in inventory planning, marketing, and operations.

What problem I was trying to solve

The main goal was to help the business answer questions like

Questions Solved in Superstore Sales Analysis Project (SQL)

1. What is the total number of transactions in the sales dataset?
2. How many unique products are present in the dataset?
3. What is the total quantity of products sold?
4. What is the average age of customers?
5. What is the customer count by gender?
6. How many orders were placed using each payment mode?
7. Which products are most frequently ordered?
8. What is the total quantity sold for each product?
9. How many transactions occur on each purchase date?
10. What are the top 5 most selling products by quantity?
11. Which products are most frequently cancelled?
12. What time of the day has the highest number of purchases?
13. What are the peak purchase hours during the day?
14. Who are the top 5 highest spending customers?
15. Which product categories generate the highest revenue?
16. What is the cancellation rate per product category?
17. What is the return rate per product category?
18. What is the most preferred payment mode among customers?
19. How does customer age group affect purchasing behavior?
20. Are certain genders buying more specific product categories

---

What data I worked with

I used a Superstore sales dataset that had:

- Order details
- Customer information
- Product categories
- Sales, quantity, and profit
- Shipping and delivery details

I first uploaded this data into a SQL database so I could analyze it properly.

What I did first – Data Cleaning

Real-world data is never perfect, so the first thing I did was **clean the data**.

I noticed:

- Some missing values
- Duplicate records
- Inconsistent category names

So using SQL, I:

- Removed duplicates
- Fixed missing values
- Standardized categories
- Created extra columns like delivery delay and profit-related fields

This step was very important because without clean data, any analysis can be misleading.

How I analyzed the data

Once the data was clean, I started asking business questions and answering them using SQL.

For example:

- I used `SUM` and `COUNT` to find total sales and number of orders
- I grouped data by category and product to see what sells the most
- I analyzed customers to find repeat buyers and high-value customers
- I checked shipping dates to see how many orders were delayed

I also used **window functions** to rank products and customers, which helped me easily find top performers.

Simple explanation of SQL logic (algorithms)

I didn't use machine learning here, but SQL itself works based on logic.

- **Aggregations** helped me summarize large data into meaningful numbers
- **Joins** helped me connect customer, order, and product data
- **CASE statements** helped me apply business rules like marking orders as delayed or on-time
- **Window functions** helped me rank and analyze data without losing details

All this logic helped me convert raw data into clear insights.

What insights I found

From the analysis, I found that:

- A few product categories were generating most of the revenue
- Nearly **30% of orders were getting delayed**, which is a big operational issue
- Some customers were buying very frequently and contributing more to sales
- Sales were higher in certain months, showing seasonal demand

---

How this helped the business

Because of this analysis:

- Inventory planning could be improved
- Delivery issues could be identified and fixed
- Marketing teams could focus on high-value customers
- Decision-making became more data-driven instead of guesswork

---

How I usually close this in an interview

"So overall, this project helped me understand how real business problems can be solved using SQL by cleaning data, analyzing patterns, and generating insights that actually matter to the business."