Retail Sales SQL Case Study

Introduction

Hi, my name is Gia. In this case study, I analyze a retail sales dataset using SQL to evaluate sales performance, customer purchasing behavior, and revenue trends. The objective of this analysis is to uncover insights that support data-driven business decisions while demonstrating practical SQL analysis skills.

Business Objective

The objective of this analysis is to evaluate retail sales performance by addressing the following questions:

- How much revenue is generated overall?
- Which product categories perform best?
- How do transaction volume and average order value differ by category?
- How does revenue change over time?

Data Source

The dataset used in this analysis comes from the **Retail Sales Dataset** available on Kaggle. Each row represents a single transaction and includes customer, product, pricing, and date information.

Data Description

Key fields used:

- Transaction ID
- Date
- Customer ID
- Product Category
- Quantity
- Price per Unit
- Total Amount

Data Credibility (ROCCC-style)

Reliable: Publicly available dataset used for analysis practice
Original: Provided by the dataset author on Kaggle
Comprehensive: Includes transaction-level sales data
Current: Covers sales through early 2024
Cited: Dataset source is clearly documented

PROCESS / CLEAN

Before analysis, the dataset was reviewed to ensure basic data quality and consistency:

- Confirmed total transaction count
- Verified numeric fields (price, quantity, total amount)
- Ensured dates were usable for time-based analysis

ANALYZE

Overall Sales Performance:
An initial SQL query was used to calculate total transactions, unique customers, and overall revenue.

Key finding:
The dataset contains 1,000 transactions from 1,000 unique customers, generating 456,000 in total revenue.

```
SELECT
    COUNT(*) AS total_transactions,
    COUNT(DISTINCT "Customer ID") AS total_customers,
    SUM("Total Amount") AS total_revenue
FROM sales;
```

123 total_transactions	123 total_customers	123 total_revenue
1,000	1,000	456,000

Revenue by Product Category:
Sales were grouped by product category to compare performance.

Findings:

- Electronics generated the highest total revenue.
- Clothing had the highest number of transactions.

- Beauty generated fewer transactions but higher-value purchases.

```
SELECT
  "Product Category",
  COUNT(*) AS transactions,
  SUM("Total Amount") AS revenue
FROM sales
GROUP BY "Product Category"
ORDER BY revenue DESC;
```

A-Z Product Category	123 transactions	123 revenue
Electronics	342	156,905
Clothing	351	155,580
Beauty	307	143,515

Average Order Value (AOV):
Average order value was calculated by category to understand pricing behavior.

Findings:

- Beauty had the highest average order value.
- Clothing had the lowest average order value, despite the highest transaction volume.
- Electronics balanced both volume and revenue.

```
SELECT
  "Product Category",
  COUNT(*) AS transactions,
  SUM("Total Amount") AS revenue,
  ROUND(AVG("Total Amount"), 2) AS avg_order_value
FROM sales
GROUP BY "Product Category"
ORDER BY avg_order_value DESC;
```

A-Z Product Category	123 transactions	123 revenue	123 avg_order_value
Beauty	307	143,515	467.48
Electronics	342	156,905	458.79
Clothing	351	155,580	443.25

Monthly Revenue Trends:
Revenue and transaction counts were analyzed by month.

Findings:

- Sales fluctuate month to month.
- May 2023 was the strongest month in both revenue and transactions.
- January 2024 appears to be a partial month and was excluded from trend interpretation.

```
SELECT
  strftime('%Y-%m', "Date") AS month,
  COUNT(*) AS transactions,
  SUM("Total Amount") AS revenue
FROM sales
GROUP BY month
ORDER BY month;
```

A-Z month	123 transactions	123 revenue
2023-01	76	35,450
2023-02	85	44,060
2023-03	73	28,990
2023-04	86	33,870
2023-05	105	53,150
2023-06	77	36,715
2023-07	72	35,465
2023-08	94	36,960
2023-09	65	23,620
2023-10	96	46,580
2023-11	78	34,920
2023-12	91	44,690
2024-01	2	1,530

CONCLUSIONS

This analysis shows that overall revenue performance depends on both pricing and transaction volume. While Electronics leads in total revenue, Beauty benefits from higher-priced purchases, and Clothing relies on high sales volume. Monthly trends suggest possible seasonality, which could be explored further with additional data.

Tools Used

SQLite
DBeaver
GitHub (project documentation)