Tab 1

Retail Store Sales Analysis

Uncovering Insights from Transactional Data Using SQL

INTRODUCTION

In this project, I performed an in-depth exploratory data analysis for a fictional retail business using SQL. The objective was to gain practical knowledge and understanding of a real-world data analyst role: cleaning, exploring, and analyzing: customer, order, and product data to uncover patterns and direct actionable business insights.

Using raw transactional data from a mini-retail store, I wrote and executed SQL queries to answer key business questions such as:

- Who are the highest-spending customers?

- What are the most popular product categories in different regions?

- How do sales trend over time (daily, weekly, monthly)?

- Which countries generate the most revenue?

- What customer segments exist based on spending behavior?

This analysis required practical application of core SQL skills, including:

- Data cleaning and transformation

- Aggregate functions and grouping

- Common Table Expressions (CTEs)

- Window functions for advanced ranking

- Conditional logic with CASE

The goal was to build a hands-on, realistic SQL project that reflects how a data analyst would approach messy data, extract insights, and deliver value to business stakeholders

DATASET OVERVIEW

This dataset simulates transaction data from a fictional retail store that operates across multiple countries. It captures essential sales data including;

- Customer details
- Product names
- Unit prices
- Quantity sold
- Order dates
- Country of transaction.

   The dataset was manually constructed and imported using PostgreSQL, fully managed and analyzed using a phone-based SQL environment (DB Fiddle), showcasing the ability to work with industry-standard SQL tools even under hardware limitations.

   It reflects both clean and imperfect real-world data including missing product names, quantities — making it a strong candidate for data cleaning and exploration tasks. This simulates challenges data analysts face when working with live business datasets.

   Although it's structured as a single flat table, the columns represent multiple dimensions typically seen in normalized databases: customers, products, orders, and regional data.

   This dataset is ideal for developing and demonstrating SQL fluency, including the use of subqueries, window functions, CASE expressions, aggregation, and date/time functions.

   Time Frame: Orders span various dates across the year 2023, allowing for analysis of time-based trends, customer patterns, and revenue performance.

BUSINESS QUESTIONS SOLVED

This project addressed several key business questions that reflect common real-world challenges in the retail industry. Each question was answered using SQL queries written and executed in PostgreSQL via a phone-based SQL platform (DB Fiddle).

1.   What are the total sales and number of orders overall?

→ Helps stakeholders understand the business's overall performance.

```
SELECT

    SUM(quantity * price) total_revenue,

    COUNT(order_id)

FROM retail_orders;
```

2.   Which product categories are most popular?

→ Identifies best-selling items and guides inventory or marketing focus.

```
SELECT

        product_category,

        SUM(quantity * price) total_revenue

   FROM  retail_orders

   GROUP BY product_category

 ORDER BY  total_revenue DESC

   LIMIT 3;
```

3.   Which countries generate the highest revenue?

→ Assists in prioritizing regional expansion and resource allocation.

```
SELECT
```

```
    country,

    SUM(price * quantity) total_sales

FROM retail_orders

GROUP BY country

ORDER BY total_sales DESC

LIMIT 1;
```

4.   Who are the top spending customers?

→ Useful for loyalty programs, premium targeting, and retention strategies.

```
SELECT

    customer_name,

    SUM(quantity * price) total_spent

FROM retail_orders

GROUP BY customer_name

ORDER BY  total_spent DESC;
```


5.   How do sales trends change over time (weekly)?

→ Time-based insights for planning, seasonal promotions, or forecasting.

```
SELECT

    DATE_TRUNC('week', order_date) AS week_start,

    SUM(price * quantity) AS total_sales

FROM retail_orders

GROUP BY week_start
```

ORDER BY week_start;

6.   What is the average order value per customer ?

→ Key metric for measuring customer value and optimizing pricing.

SELECT

    customer_name,

    AVG(price * quantity) AS avg_order_value

FROM retail_orders

GROUP BY customer_name

ORDER BY avg_order_value DESC;

7.   Who are the 'High', 'Medium', and 'Low' spenders? (Using CASE statements)

→ Customer segmentation for tailored marketing strategies.

SELECT

    customer_name,

    SUM(price * quantity) total_spend,

 CASE

    WHEN SUM(price * quantity) < 3000 THEN 'low spenders'

  WHEN SUM(price * quantity) BETWEEN 3000 AND 6000 THEN 'medium spenders'

  WHEN SUM(price * quantity) > 6000 THEN 'high spenders'

  ELSE 'unknown'

  END AS spending_category

FROM retail_orders

GROUP BY customer_name

ORDER BY total_spend DESC;

8.   Which customers spent more than the average customer spend? (Using subqueries)

→ Helps detect VIPs or high-value customers.

WITH spend_total AS (

  SELECT

    customer_name,

    SUM(price * quantity) AS total_spend

  FROM retail_orders

  WHERE price IS NOT NULL AND quantity IS NOT NULL

  GROUP BY customer_name

)

SELECT *

FROM spend_total

WHERE total_spend > (

  SELECT AVG(total_spend)

  FROM spend_total

);

9.   Rank customers by total spend within each country. (Using Window Functions)

→ Supports performance comparison and localized strategies.

SELECT

customer_name,

country,

SUM(price * quantity) AS total_spent,

RANK() OVER (PARTITION BY country ORDER BY SUM(price * quantity) DESC) AS country_spend_rank

FROM retail_orders

WHERE price IS NOT NULL AND quantity IS NOT NULL

GROUP BY customer_name, country

ORDER BY country, country_spend_rank;

10.   Identify countries with revenue above global average. (Using CTEs and aggregation)

→ Measures country-level performance against benchmarks.

WITH country_revenue AS (

  SELECT

    country,

    SUM(price * quantity) AS total_revenue

  FROM retail_orders

  GROUP BY country

),

global_avg AS (

  SELECT

    AVG(total_revenue) AS avg_revenue

  FROM country_revenue

)

SELECT

   cr.country,

   cr.total_revenue

FROM

   country_revenue cr,

   global_avg ga

WHERE cr.total_revenue > ga.avg_revenue

ORDER BY cr.total_revenue DESC;

11.   What are the busiest days in terms of total sales?

→ Operational insight for staffing, stock management, and peak load handling.

SELECT

   DATE_TRUNC('day', order_date) AS day,

   SUM(price * quantity) AS total_sales

FROM retail_orders

WHERE price IS NOT NULL AND quantity IS NOT NULL

GROUP BY day

ORDER BY total_sales DESC

LIMIT 5;

KEY INSIGHTS AND RECOMMENDATIONS

* High spenders contributed disproportionately to total revenue. These customers could be targeted for loyalty campaigns or premium offerings.

* Home decor, books and clothing products dominated both quantity sold and revenue earned. Inventory and marketing efforts should be prioritized for these segments.

* Kenya had the highest customer engagement and spending, followed by Ghana and Nigeria. Regional campaigns and pricing strategies can be adjusted accordingly.

* Peak activity in terms of sales was observed during weekdays and late in the month. Promotional campaigns should align with these windows.

* The customer base is skewed toward a few countries like Kenya.
 Expansion into underrepresented countries could drive further growth.

* A significant portion of orders came from repeat customers. This supports retention strategies such as personalized emails or loyalty rewards.

* Using window functions, top customers were ranked per country, offering insights for local influencer partnerships or VIP treatment.

* A CTE revealed only a few countries exceeded average revenue, indicating where the store performs strongest and where improvements can be made.

Top Customers: Michael Smith and John Doe were among the highest spenders.

Category Trends: Home Decor and clothing performed best overall.

Revenue Leaders: Kenya and Ghana topped the revenue charts.

Time-Based Trends: Weekdays and month ends saw higher order volumes and revenue spikes.

Customer Segmentation: Most customers fall into the Medium spender and Low spender range.

CONCLUSION

This project marks a significant milestone in my data journey. Using a fictional retail dataset, I performed a complete SQL-based sales analysis—from exploring raw transactional records to uncovering key business insights. All of this was done entirely on a mobile device, with PostgreSQL as the database engine.

Throughout the project, I answered real-world business questions by leveraging:

- CTEs and subqueries to simplify complex aggregations

- Window functions to rank and segment data

- Case statements to create dynamic customer spend categories

- Date functions to analyze trends over time

This project taught me more than just SQL syntax. It strengthened my ability to think like a data analyst: formulating questions, testing hypotheses, and communicating insights from data. It also proved that even with limited tools and infrastructure, it's possible to do meaningful analysis — all I needed was my phone, persistence, and guidance.

TOOLS USED

- SQL Dialect: PostgreSQL

- Platform: DB Fiddle (phone-based)

- Device: Iphone 11