

PROJECT 5

RETAIL SALES INSIGHTS WITH SQL + PYTHON

Submitted By: AKANSHA SINGH (2022408007)

Submitted To: Ms. MOUSHREETA DEBROY

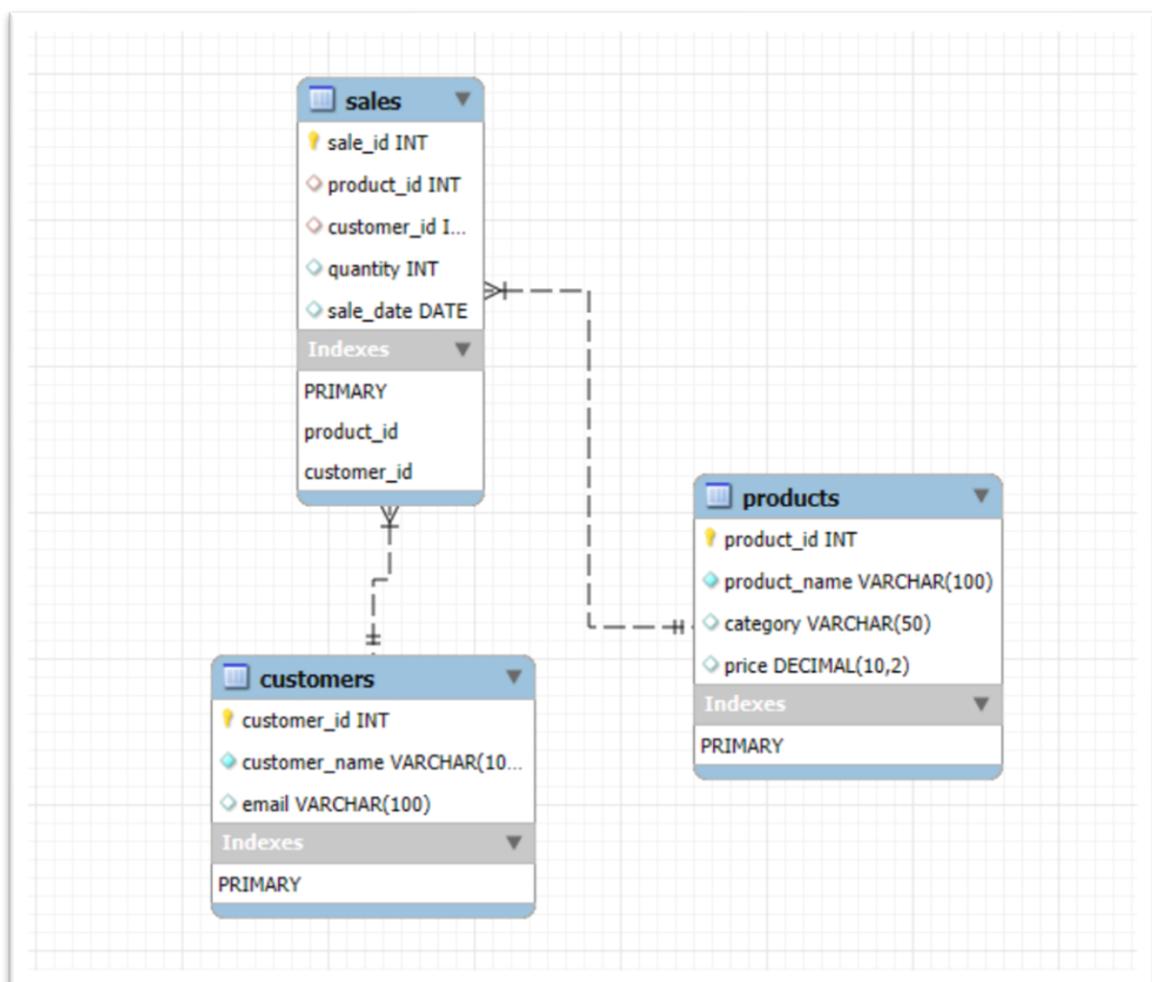
Date: 21st Oct 2025

1. PROJECT OBJECTIVE

The purpose of this project is to understand how databases and data science work together to extract useful business insights. This project involves designing a small Retail Sales Database, writing SQL queries to find insights, and then visualizing the results using Python.

2. DATABASE DESIGN AND SCHEMA

2.1 Schema Diagram (EER) The database was designed using a relational model with three tables: Products, Customers, and Sales. The Sales table is a "fact" table that links to the two "dimension" tables using foreign keys (product_id and customer_id), as shown in the EER diagram below.



2.2 SQL DEFINITIONS THE FOLLOWING SQL CODE WAS USED TO CREATE THE TABLES:

SQL:

```
CREATE TABLE Products ( product_id INT PRIMARY KEY  
AUTO_INCREMENT, product_name VARCHAR(100) NOT NULL, category  
VARCHAR(50), price DECIMAL(10, 2) );
```

```
CREATE TABLE Customers ( customer_id INT PRIMARY KEY  
AUTO_INCREMENT, customer_name VARCHAR(100) NOT NULL, email  
VARCHAR(100) );
```

```
CREATE TABLE Sales ( sale_id INT PRIMARY KEY AUTO_INCREMENT,  
product_id INT, customer_id INT, quantity INT, sale_date DATE, FOREIGN  
KEY (product_id) REFERENCES Products(product_id), FOREIGN KEY  
(customer_id) REFERENCES Customers(customer_id) );
```

3. ANALYSIS, VISUALIZATIONS, AND OBSERVATIONS

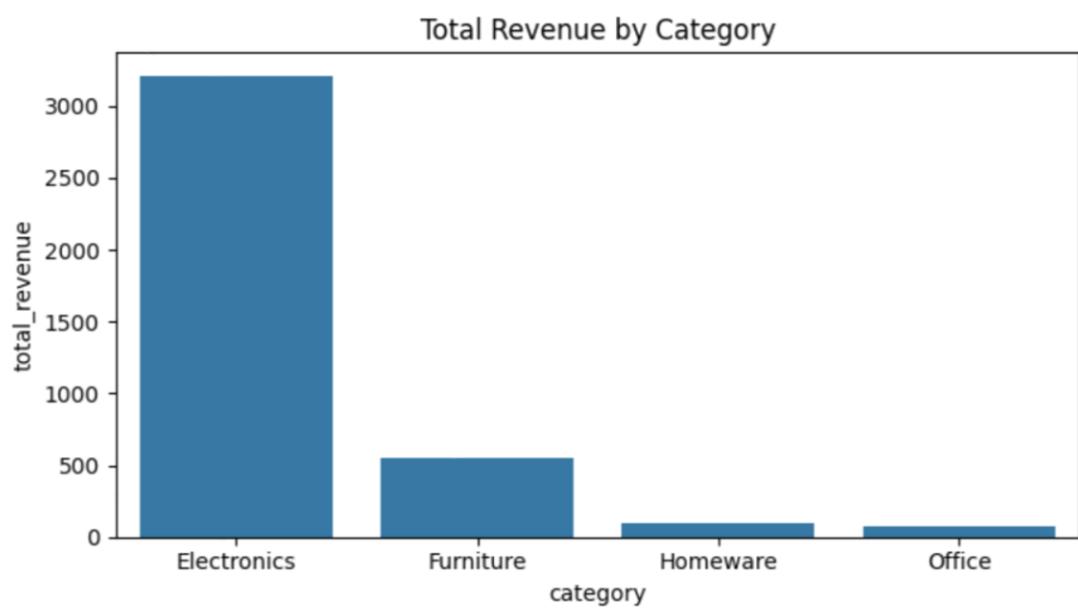
The SQL database was queried to extract key business insights, which were then plotted using Python.

3.1 INSIGHT 1: TOTAL SALES REVENUE BY CATEGORY

The first analysis query calculated the total revenue generated by each product category.

OBSERVATION:

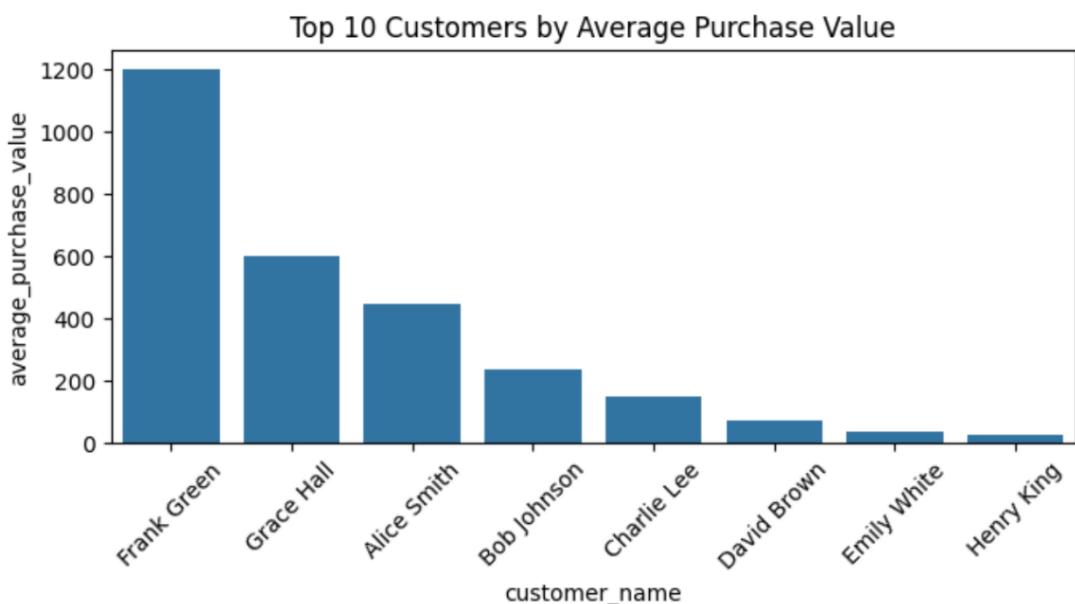
This bar chart shows that 'Electronics' is the dominant category, driving the vast majority of sales revenue. 'Furniture' is a distant second, while 'Homeware' and 'Office' are very low-performing.



RECOMMENDATION:

This insight suggests the business should focus its marketing budget and inventory on high-value electronics.

3.2 INSIGHT 2: AVERAGE PURCHASE VALUE BY CUSTOMER



OBSERVATION:

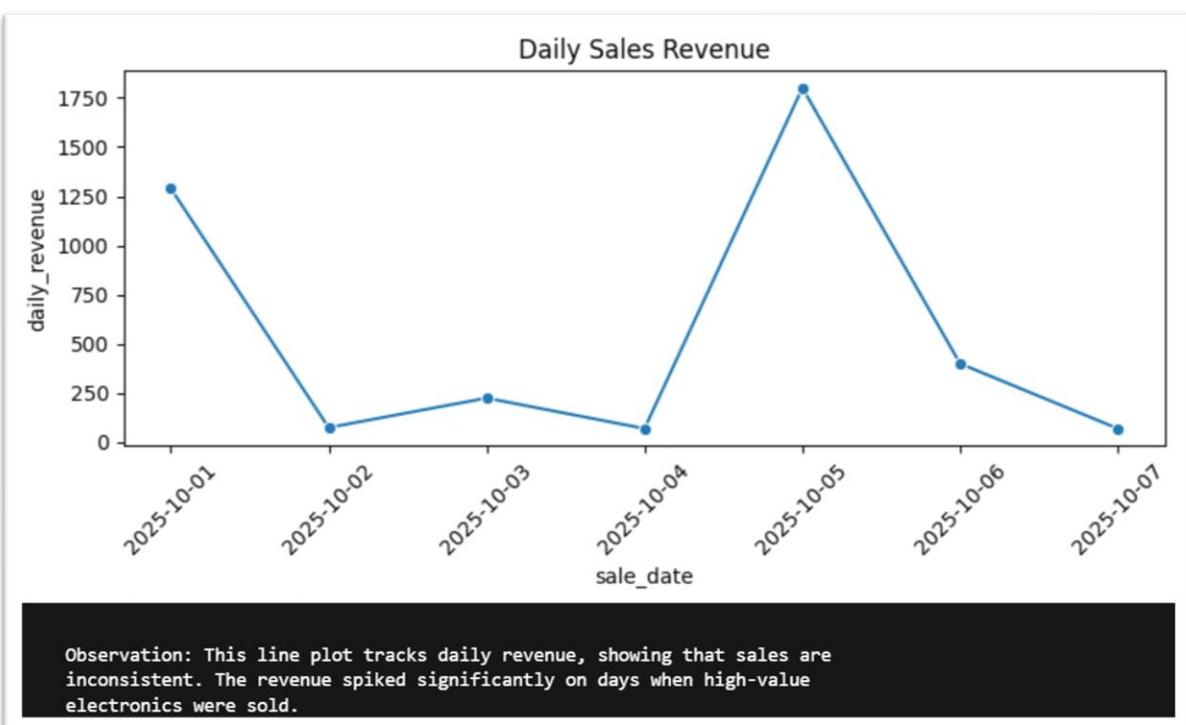
This chart identifies our high-value customers. Customers like 'Frank Green' and 'Alice Smith' have the highest average purchase value, as they buy expensive items like laptops.

RECOMMENDATION:

We should target these high-spending customers with a loyalty program or special offers to encourage repeat business.

3.3 Insight 3: Daily Sales Revenue Trend

This query tracked the total revenue generated by the store on each day.



OBSERVATION:

This line plot tracks daily revenue, showing that sales are inconsistent. The revenue spiked significantly on days when high-value electronics were sold.

RECOMMENDATION:

This shows our daily income is volatile and depends on big-ticket items. The business should explore strategies to increase smaller, more consistent sales to create a more stable daily revenue.

4. CONCLUSION AND RECOMMENDATIONS

This analysis successfully combined SQL and Python to extract actionable insights from the company's sales data. The findings show that the business is heavily reliant on the 'Electronics' category and a few high-value customers.

Based on the observations, the following business recommendations are proposed:

1. **Focus on High-Value Customers:** Target high-spending customers (like Frank Green and Alice Smith) with a loyalty program or special offers to encourage repeat business.
2. **Boost Low-Performing Categories:** Investigate why 'Homeware' and 'Office' sales are so low. Consider a "bundle" marketing campaign (e.g., "Buy a Laptop, get 10% off a Desk Lamp") to increase sales in these areas.
3. **Stabilize Daily Revenue:** Explore strategies to increase the frequency of smaller, more consistent sales to create a more stable daily income stream, rather than relying on large, infrequent purchases.