End-to-End Retail Performance and Behavioral Analytics (RetailNova Inc.)

Submitted By:
V.Viswesh

Course:
Diploma in Data Analytics with Generative AI

Tools Used:
SQL Server, Python, Power BI

Project Type:
Retail Business Intelligence Project

Submission Date:
January 2026

Project Summary:
This project focuses on analyzing retail sales data to derive actionable business insights related to sales performance, customer behaviour, product profitability, store efficiency, and return trends. The project demonstrates a complete analytics lifecycle, including data cleaning, transformation, analysis, and visualization using industry-standard tools.

INDEX (TABLE OF CONTENTS)

INTRODUCTION & BUSINESS CONTEXT

Introduction

Retail organizations generate large volumes of transactional and customer data on a daily basis. However, without proper analytics, this data cannot be effectively utilized for decision-making. Retail analytics enables businesses to identify sales trends, understand customer purchasing behavior, measure product performance, and optimize store operations.

Business Context

In a competitive retail environment, organizations must rely on data-driven insights rather than intuition. This project simulates a real-world retail analytics scenario where sales, customer, product, store, and return data are analyzed to support strategic business decisions.

BUSINESS PROBLEM & OBJECTIVES

The primary challenge addressed in this project is the lack of visibility into sales performance, profitability, customer value, and product returns. Without structured analysis, it becomes difficult to identify high-performing products, valuable customers, profitable stores, and areas requiring operational improvements.

Project Objectives

The key objectives of this project are:

- Analyze total revenue and quantity sold over time
- Identify top-selling and most profitable products
- Understand customer distribution and demographics
- Evaluate store performance and profitability
- Analyze return patterns and trends

DATASET DESCRIPTION

Dataset Overview

The project uses multiple datasets representing different business entities:

- **Sales Data:** Transaction details such as order date, quantity, unit price, discount, and total amount
- **Customer Data:** Customer demographics including age, gender, region, and signup date
- **Product Data:** Product attributes such as category, brand, cost price, and margin
- **Store Data:** Store details including location, type, and operating cost
- **Returns Data:** Returned orders with return date and reason

These datasets together provide a complete picture of retail operations.

Methodology:

1. Business understanding

2. Data collection and understanding

3. Data cleaning and transformation

4. SQL-based analysis

5. Python-based calculations

6. Power BI dashboard creation

7. Insight generation

DATA CLEANING & PREPARATION

Data Cleaning Process

Data cleaning was performed using SQL and Python. Unnecessary columns were removed, missing values were handled appropriately, and data types were standardized. Foreign key constraints were applied to maintain referential integrity across tables.
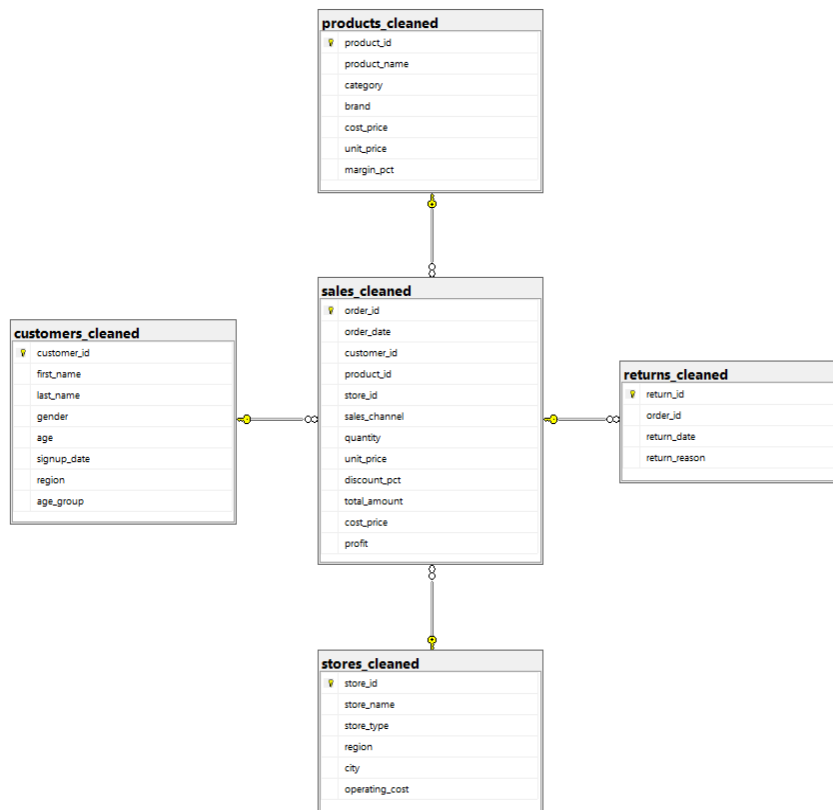
Profit Calculation

Profit was calculated at the transaction level using Python by subtracting total product cost from total sales amount. This profit value was stored in the sales table to ensure consistency across SQL analysis and Power BI dashboards.

SQL Analysis

Cleaned data was loaded into Microsoft SQL Server for structured analysis. SQL queries were written to answer all mandatory business questions, including revenue trends, top performing products, customer distribution, return rates, and store profitability. Common SQL techniques used included joins, aggregations, window functions, and common table expressions (CTEs) to ensure clarity and performance.DATA MODELING

products_cleaned
- product_id
- product_name
- category
- brand
- cost_price
- unit_price
- margin_pct

sales_cleaned
- order_id
- order_date
- customer_id
- product_id
- store_id
- sales_channel
- quantity
- unit_price
- discount_pct
- total_amount
- cost_price
- profit

customers_cleaned
- customer_id
- first_name
- last_name
- gender
- age
- signup_date
- region
- age_group

returns_cleaned
- return_id
- order_id
- return_date
- return_reason

stores_cleaned
- store_id
- store_name
- store_type
- region
- city
- operating_cost

Explanation

A star-schema-based relational data model was created with the sales table acting as the central fact table. Dimension tables such as customers, products, stores, and returns are connected using primary and foreign key relationships. This structure ensures efficient querying, accurate aggregations, and seamless dashboard filtering.

SALES OVERVIEW DASHBOARD
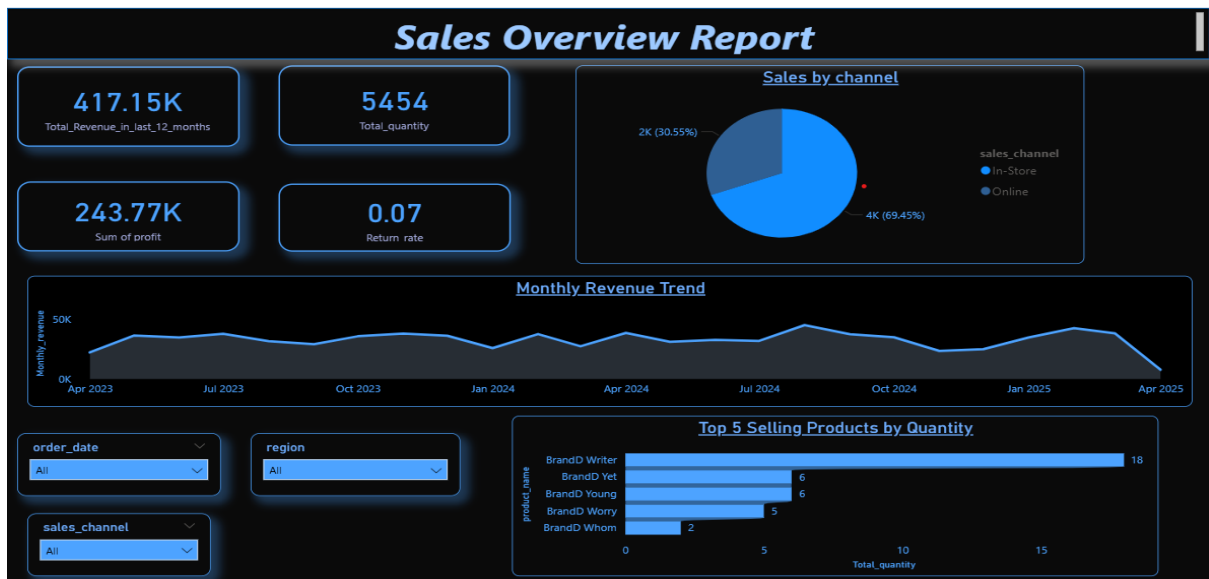
Explanation:-The Sales Overview dashboard provides a high-level summary of business performance. It includes total revenue for the last 12 months, total quantity sold, monthly revenue trends, top 5 selling products, and sales by channel (Online vs In-Store). This dashboard helps stakeholders quickly assess overall sales health.

CUSTOMER INSIGHTS DASHBOARD
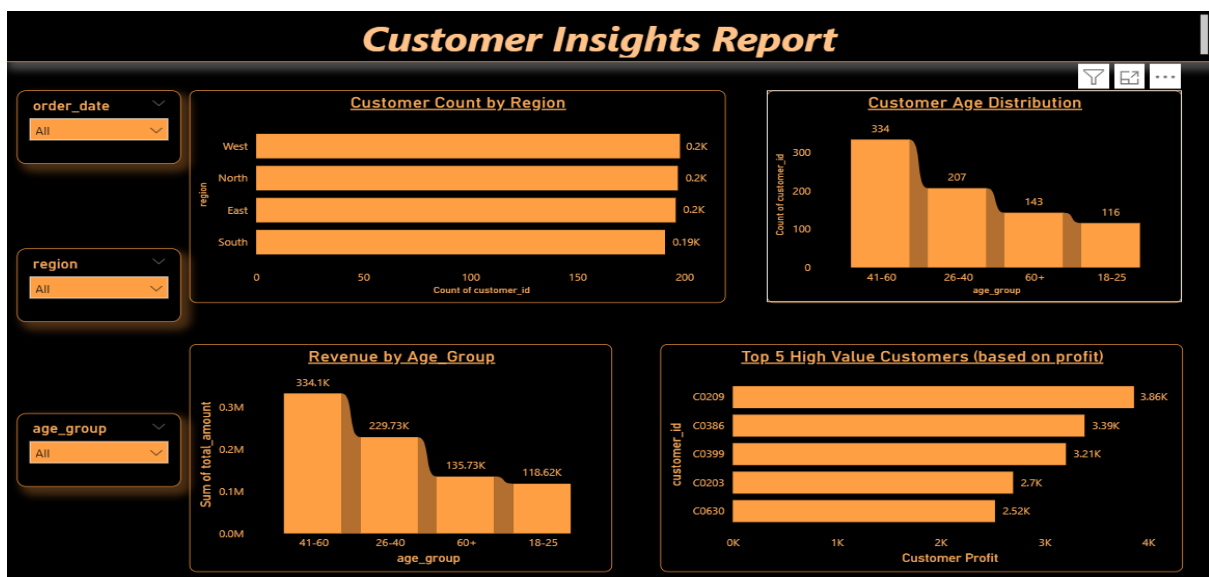
Explanation;-This dashboard focuses on customer behavior and demographics. It displays customer count by region, age distribution, revenue contribution by age group, and the top 5 high-value customers based on profit. These insights help businesses understand customer segments and target high-value customers effectively.
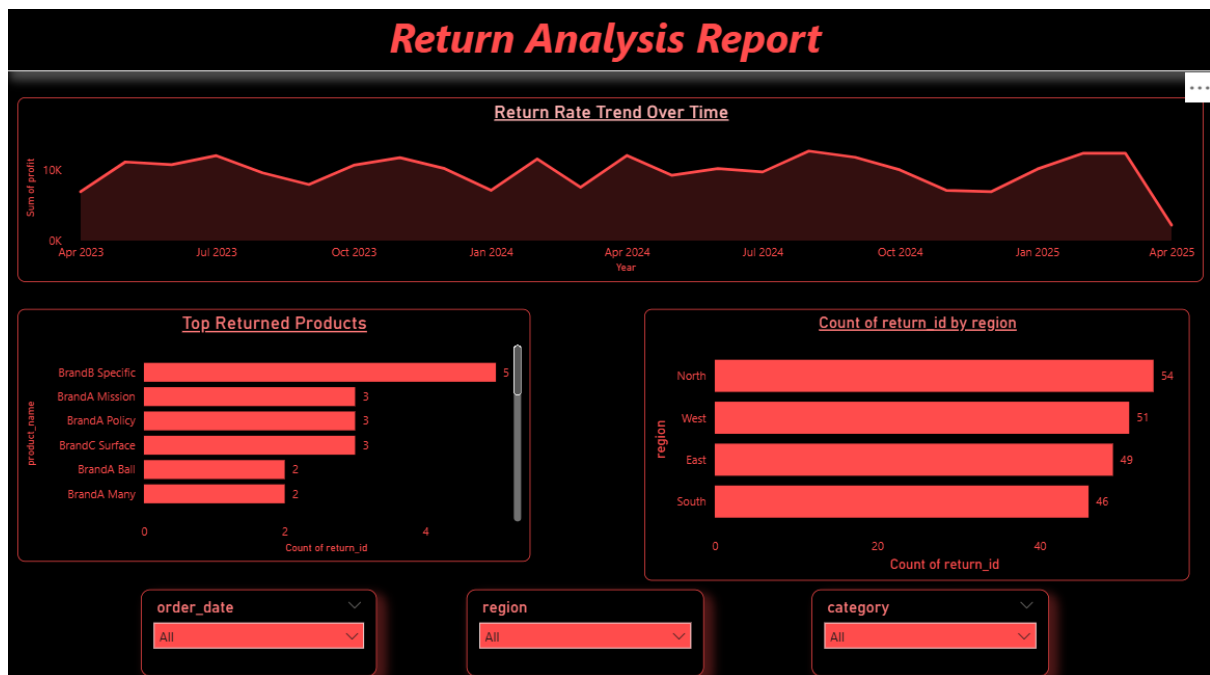
PRODUCT PERFORMANCE DASHBOARD

Explanation:-The Product Performance dashboard analyzes product-level metrics such as total profit by category, return rate by category, and top products by sales quantity. This dashboard helps identify profitable categories and products with high return rates that may require quality or supply chain improvements.

STORE & RETURN ANALYSIS DASHBOARDS

Return Analysis Report

Explanation

This dashboard evaluates store performance and return trends. It includes monthly profit by store, store profitability compared to operating cost, return rate trends over time, top returned products, and total returns by region. These insights help optimize store operations and reduce product returns.

Business Recommendations

Why add it:
Dashboards show *what happened*
Recommendations show *what business should do*

Example (you can reuse):

- Focus marketing on high-value customers identified
- Reduce return rates in high-return categories
- Optimize store operating costs
- Improve quality checks for frequently returned products

☞ Trainers LOVE this section

Assumptions & Limitations

Why add it:
Shows **analytical maturity**

What to add:

- Analysis based on available historical data
- Profit calculated at transaction level
- Online sales treated as virtual store
- No external data sources used

Future Scope

Why add it:
Shows forward thinking

What to add:

- Sales forecasting
- Customer churn prediction
- Automated dashboards
- Real-time data integration

CONCLUSION

Conclusion

This Retail Sales Analytics project demonstrates a complete end-to-end analytics workflow using SQL, Python, and Power BI. By integrating data cleaning, analysis, and visualization, the project delivers actionable insights that support data-driven decision-making. The dashboards provide a clear, interactive, and professional view of retail performance, making this project suitable for real-world business applications.