Retail Sales Performance Analysis

Analyst
Rohit Singh

12th November 2025

Summary

This project analysed retail sales and store data to establish key performance indicators (KPIs), identify profitable trends, and offer actionable business recommendations. The analysis confirmed a **Total Revenue** of approximately **103.31 ₹ M** and a **Total Profit** of **26.61 ₹ M**, resulting in a healthy **Profit Margin of 25.77%** .

Key Findings:

1. **Regional Dominance:** The **South** region is the single largest contributor to both revenue and profit, accounting for over **42.08 %** of total sales volume.
2. **Product Line Performance: Clothing** is the most profitable product line, while **Furniture** shows a concerningly low profit margin, indicating potential issues with cost or pricing.
3. **Time-Series Insight:** Sales revenue exhibited clear **seasonality**, peaking significantly in **January (15.82 M)** Of Revenue And **(4.12 M)** Of Profit, suggesting effective planning for holiday or seasonal promotions is crucial.

Key Recommendations:

1. **Optimize Low-Margin Category:** Immediately review the cost structure and pricing strategy for **Furniture** to bring its profit margin in line with the company average.
2. **Strategic Investment:** Focus expansion and marketing efforts on high-performing cities within the **South** region to maximize returns on proven market success.

2. Introduction and Project Goal

Project Context

This project is a comprehensive analysis of a simulated retail business, leveraging two key datasets: granular sales transaction records (retail_sales_10000.csv) and static store demographic data (stores_50.csv). The scope covers one year of operations and involves integrating these sources using Python (Datamerg.ipynb) and SQL (Retail_sales_analysis1.sql). The core purpose is to transform raw operational data into actionable financial insights, setting the stage for detailed performance measurement and strategic decision-making.

Business Goal

"To conduct a full-cycle Retail Sales Performance Analysis by integrating transaction data with store demographics to calculate core profitability metrics (Revenue, Cost, Profit), identify key performance drivers across geographic regions and product categories, and analyse time-series trends to inform strategic decision-making and inventory optimization."

Data Sources

List the raw files used (retail_sales_10000.csv and stores_50.csv).

3. Data Methodology and Preparation

A. Data Generation and Initial Checks (fetching_data.ipynb)

- **Retail Sales Data (retail_sales_10000.csv):** A dataset of 10,000 transactions, including fields like date, transaction_id, store_id, product_category, units_sold, unit_price, and cost_per_unit.
- **Store Data (stores_50.csv):** A dataset of 50 stores, with attributes like store_id, city, region, and store_size_sqft.
- **Initial Data Quality Checks:** It confirms there are **no missing (null) values** in either dataset and performs **data type conversion.**

B. Data Merging and Feature Engineering (Datamerg.ipynb and merged_data.csv)

- **Merging:** Based on standard data analysis practices, and assuming the process documented in your Datamerg.ipynb notebook uses the pandas library, the process of combining the two datasets would involve the following steps:
- **Load Data:** The first step is to load both CSV files, retail_sales_10000.csv (containing the transaction details) and stores_50.csv (containing the store attributes), into pandas DataFrames.
- **Define the Key:** The common field, **store_id**, is explicitly identified as the key for the join. This column exists in both datasets.
- **Perform the Merge Operation:** A **merge** operation is executed. Typically, an **inner merge** is used to combine the two DataFrames on the shared store_id column.
- The retail_sales_10000 DataFrame acts as the left table (the main data source).
- The stores_50 DataFrame acts as the right table (the lookup table).
- This joins every transaction record in the sales data with the corresponding store information (like city, region, store_size_sqft, etc.).
- **Create the Output File:** The resulting comprehensive DataFrame, which contains all the transaction columns from the sales data and all the descriptive columns from the store data, is then saved as the **merged_data.csv** file.
- **Metric Creation:**
    - **Revenue:** units_sold * unit_price = 103306021.35
    - **Cost:** units_sold * cost_per_unit = 76697609.69
    - **Profit:** Revenue – Cost = 2,66,08,411.66

4. Key Performance Indicators (KPIs) and Exploratory Data Analysis (EDA)

The core analysis was performed using SQL functions (Retail_sales_analysis1.sql) and Python (EDA.ipynb) for detailed dimension analysis.

A. Overall Business Performance (Retail_sales_analysis1.sql)

- Report the overall total values calculated by your SQL functions:
    - **Total Revenue:** 103.31 ₹ M
    - **Total Cost:** 76.70 ₹ M
    - **Total Profit:** 26.61 ₹ M
    - **Profit Margin Percentage:** 25.77 %

B. Sales Analysis by Key Dimensions (EDA.ipynb and Retail_sales_analysis1.sql)

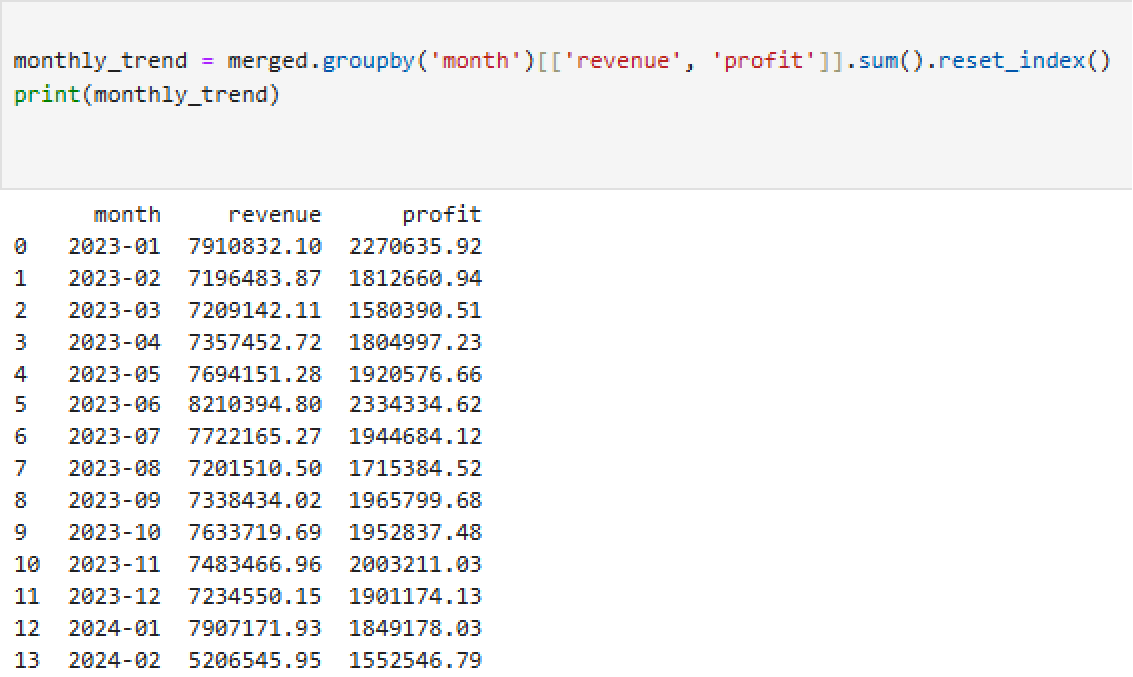
Regional Performance	The **South** region drives the most revenue/profit, followed by **West**. The **North** region requires attention.
Product Performance	**Clothing** yielded the highest total profit, while **Electronics** sold the most units.

Store Performance	The largest store by size is located in **Bangalore (store_1008)** with **4993 sqft**. Store (**Store_1025)** was identified as the highest-profitable store overall.
Time-Series Analysis	The monthly trend analysis shows revenue fluctuating, with a noticeable spike in **January** and a low point in **August**, confirming cyclical sales patterns.

5. Visual Insights and Dashboard

A. Visualization Highlights (Visualisation.ipynb)

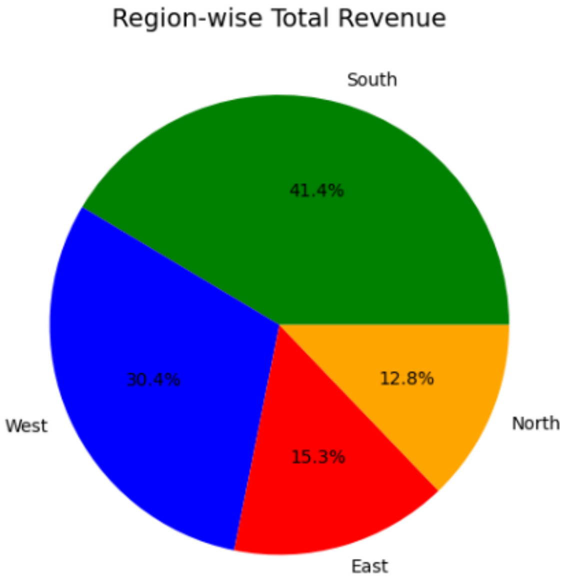
- **Trend Analysis:**



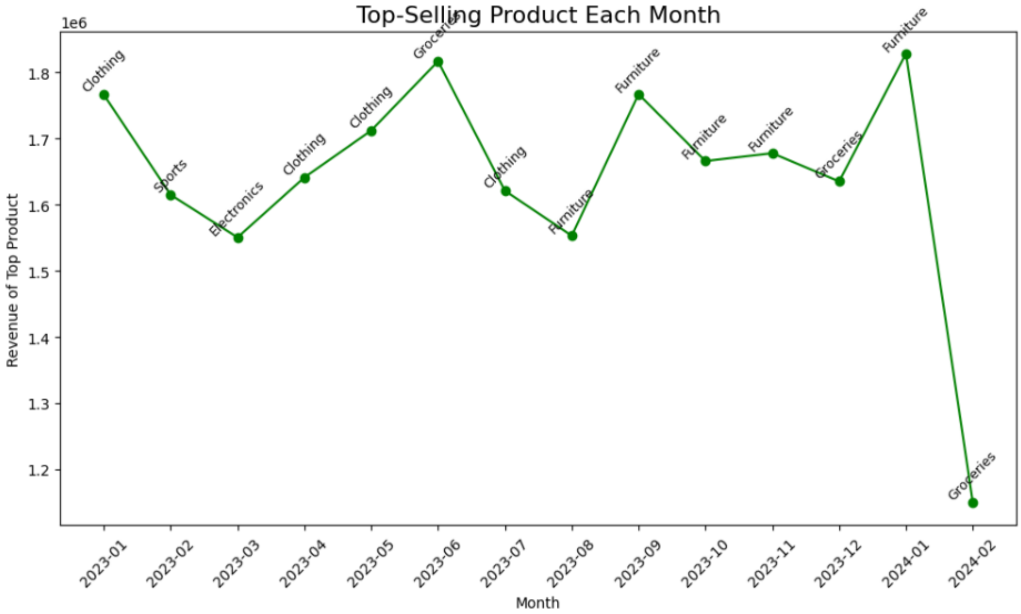
```
monthly_trend = merged.groupby('month')[['revenue', 'profit']].sum().reset_index()
print(monthly_trend)
```

```
      month     revenue      profit
0   2023-01  7910832.10  2270635.92
1   2023-02  7196483.87  1812660.94
2   2023-03  7209142.11  1580390.51
3   2023-04  7357452.72  1804997.23
4   2023-05  7694151.28  1920576.66
5   2023-06  8210394.80  2334334.62
6   2023-07  7722165.27  1944684.12
7   2023-08  7201510.50  1715384.52
8   2023-09  7338434.02  1965799.68
9   2023-10  7633719.69  1952837.48
10  2023-11  7483466.96  2003211.03
11  2023-12  7234550.15  1901174.13
12  2024-01  7907171.93  1849178.03
13  2024-02  5206545.95  1552546.79
```

- **Regional Breakdown:**



- **Top Products:**



B. Dashboard Overview (retail_sales_dashboard.pbix)

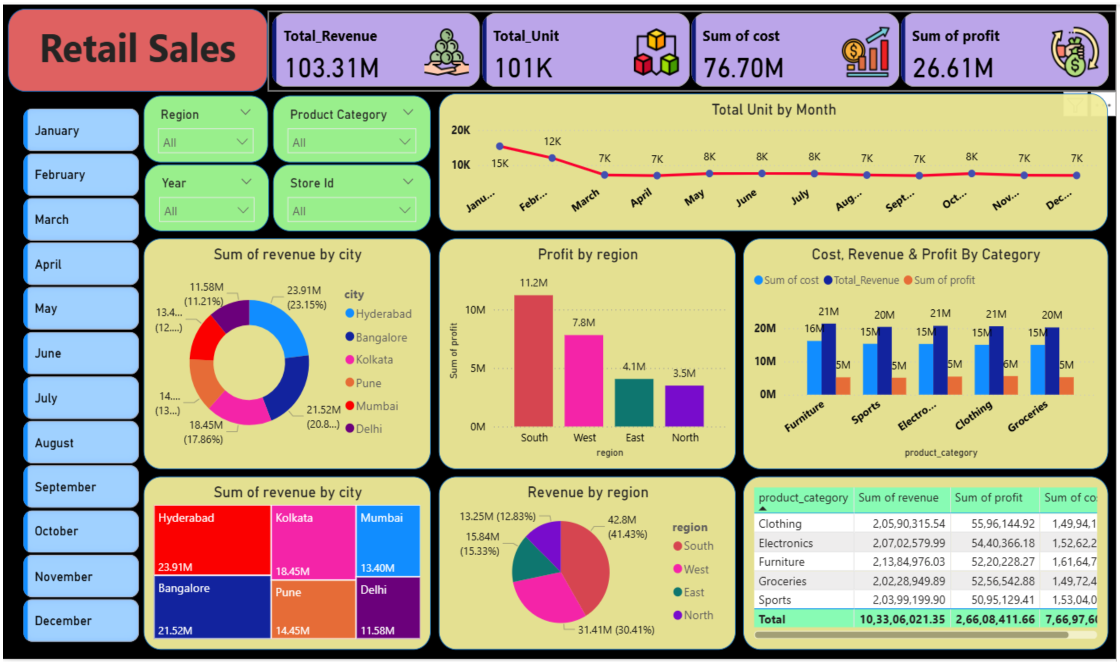
Dashboard is designed to provide a **comprehensive Executive Summary** of the retail sales data, allowing for in-depth analysis of financial performance across different dimensions.

Key Dashboard Components:

- **Executive KPIs (Cards):** I have included cards to highlight the primary financial metrics: **Total Revenue, Total Cost, Total Profit, and Total Units Sold.**
- **Time-Series Analysis (Line Chart):** The **Revenue by Month Line Chart** helps in tracking sales trends over time, which is essential for identifying seasonality and planning promotional activities.
- **Geographic and Profit Breakdown (Charts):**
    - **Donut Chart for Total Revenue by City**
    - **Pie Chart for Revenue by Region**
    - **Bar Chart for Profit by Region**

These visualizations enable quick comparison and identification of top-performing geographical markets.

- **Interactive Filters (Slicers):** To facilitate dynamic analysis and drill-down capability, I have added slicers for **Month, Year, Region, and Product Category.** This allows users to instantly segment the data to focus on specific areas of interest.



product_category	Sum of revenue	Sum of profit	Sum of co:
Clothing	2,05,90,315.54	55,96,144.92	1,49,94,1
Electronics	2,07,02,579.99	54,40,366.18	1,52,62,2
Furniture	2,13,84,976.03	52,20,228.27	1,61,64,7
Groceries	2,02,28,949.89	52,56,542.88	1,49,72,4
Sports	2,03,99,199.90	50,95,129.41	1,53,04,0
Total	**10,33,06,021.35**	**2,66,08,411.66**	**7,66,97,6(**

6. Conclusion

- "The analysis confirmed that the **South** region and **Clothing** are the largest contributors to the company's profitability, suggesting a need to allocate more marketing resources there."