

Retail Profitability Analysis

Dataset Name: retail_sales.csv

Author: Vaishnavi H

Date: October 2025

Tools Used: MySQL | Python (Pandas, Seaborn, Matplotlib) | Tableau

Objective: Analyze retail sales data to understand sales, profit, and discount patterns, and identify factors driving profitability across categories and regions.

Table of Contents

1. Executive Summary
2. Project Objectives
3. Tools and Technologies
4. Data Description
5. Data Cleaning (SQL Phase)
6. Exploratory Data Analysis (Python Phase)
7. Dashboard Development (Tableau Phase)
8. Insights and Recommendations
9. Conclusion

Executive Summary

This project aims to evaluate the profitability of a retail business using data analytics techniques.

The analysis integrates SQL for data preparation, Python for exploratory data analysis (EDA), and Tableau for dashboard visualization.

The key focus areas were:

- Understanding **sales and profit performance** by category, region, and customer segment.
- Analyzing **discount impact** on profitability.
- Identifying **seasonal sales trends** and high-performing areas.

The findings help in **strategic decision-making** for improving margins and operational efficiency.

Project Objectives

- Clean and prepare retail sales data using SQL.
 - Explore and visualize sales and profit trends using Python.
 - Build an interactive Tableau dashboard for real-time insights.
 - Identify key factors influencing profitability.
-

Tools and Technologies

Tool	Purpose
MySQL	Data cleaning, validation, and summary generation
Python (Pandas, Matplotlib, Seaborn)	Exploratory data analysis and visualization
Tableau	Dashboard creation and storytelling
Excel/CSV	Data exchange between tools

Data Description

Column	Description
Row_ID	Unique identifier for each record
Order_ID	Unique order identifier
Order_Date	Date when the order was placed
Ship_Date	Date when the order was shipped
Ship_Mode	Mode of shipment (Standard, Second Class, etc.)
Customer_ID	Unique ID of the customer
Segment	Customer segment (Consumer, Corporate, Home Office)
Country, City, State, Region	Geographical details
Product_ID	Unique product identifier
Category, Sub_Category	Product classifications
Product_Name	Product description
Sales	Total sales amount
Quantity	Units sold
Discount	Discount applied
Profit	Profit earned from the sale

Total Records: ~10,000 rows

Time Period: 2014–2017

Data Cleaning (SQL Phase)

Steps Performed

- Removed duplicate rows and invalid entries.
- Converted date and numeric formats.

- Standardized category, region, and segment names.
- Exported cleaned data with headers.

Example Query:

```
DELETE t1 FROM retail_sales t1
JOIN retail_sales t2
ON t1.Order_ID = t2.Order_ID
WHERE t1.Row_ID > t2.Row_ID;
```

Export with Headers:

```
(SELECT
'Row_ID','Order_ID','Order_Date','Ship_Date','Ship_Mode','Customer_ID',
'Customer_Name','Segment','Country','City','State','Postal_Code','Region',
'Product_ID','Category','Sub_Category','Product_Name','Sales','Quantity',
'Discount','Profit')

UNION ALL

(SELECT * FROM retail_sales)

INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/retail_cleaned.csv'
FIELDS TERMINATED BY ',' ENCLOSED BY "" LINES TERMINATED BY '\n';
```

Output: retail_cleaned.csv ready for Python and Tableau.

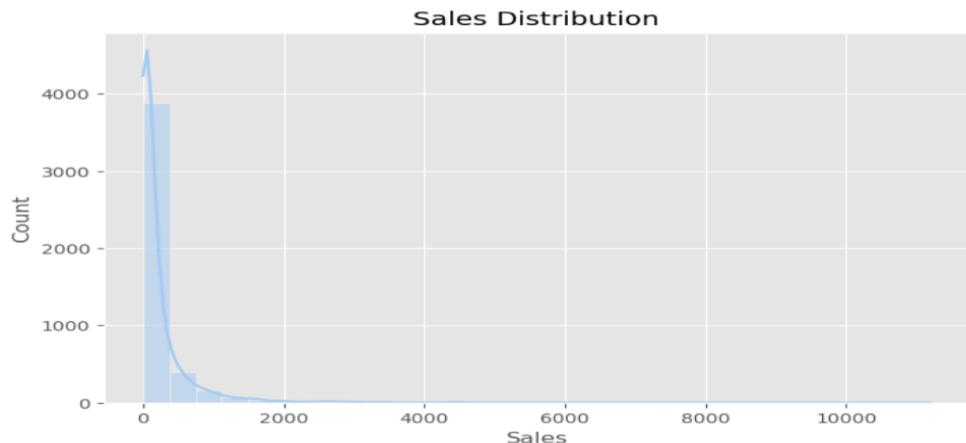
Exploratory Data Analysis (Python Phase)

Key Steps

- Imported retail_cleaned.csv into a Jupyter Notebook.
- Performed data validation and type conversions.
- Conducted univariate and bivariate analyses.
- Analyzed monthly trends and correlations.

Sample Visuals

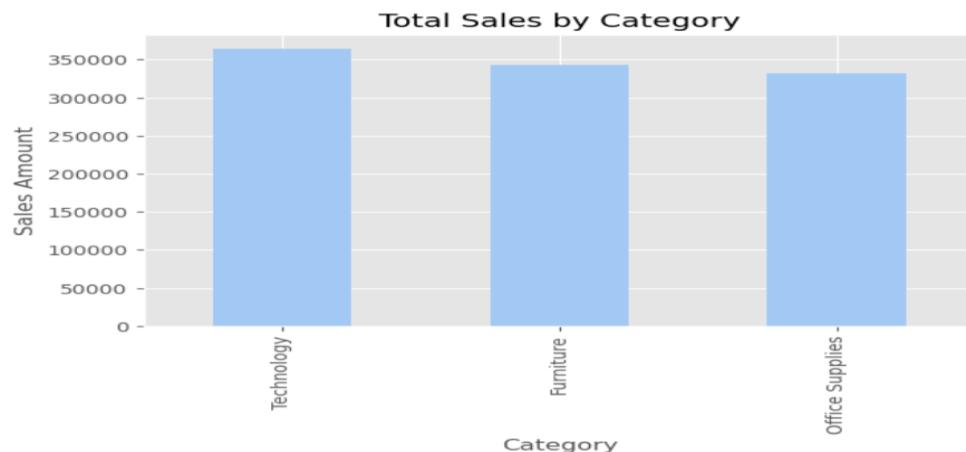
- **Sales Distribution:** Right-skewed, indicating a few high-value transactions.



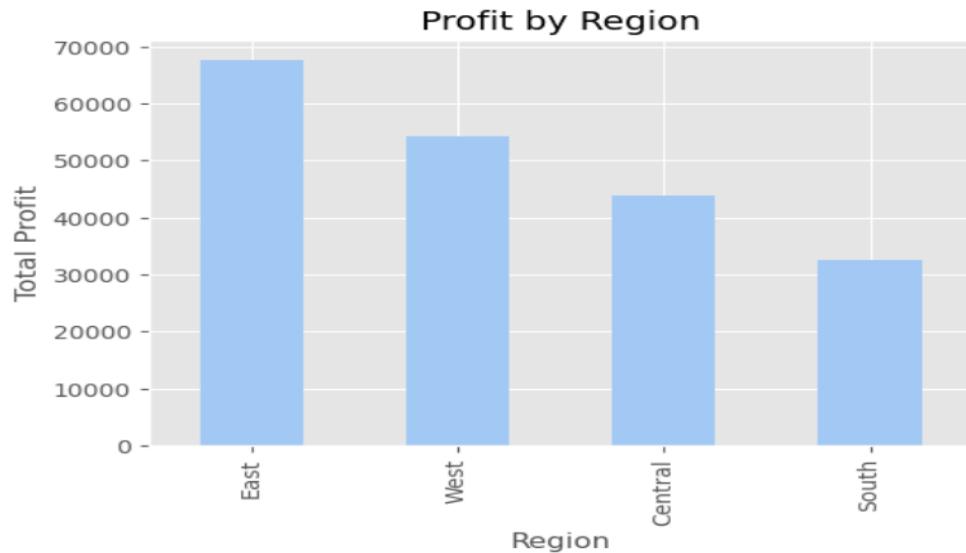
- **Profit Distribution:** Similar pattern with outliers in high-margin items.



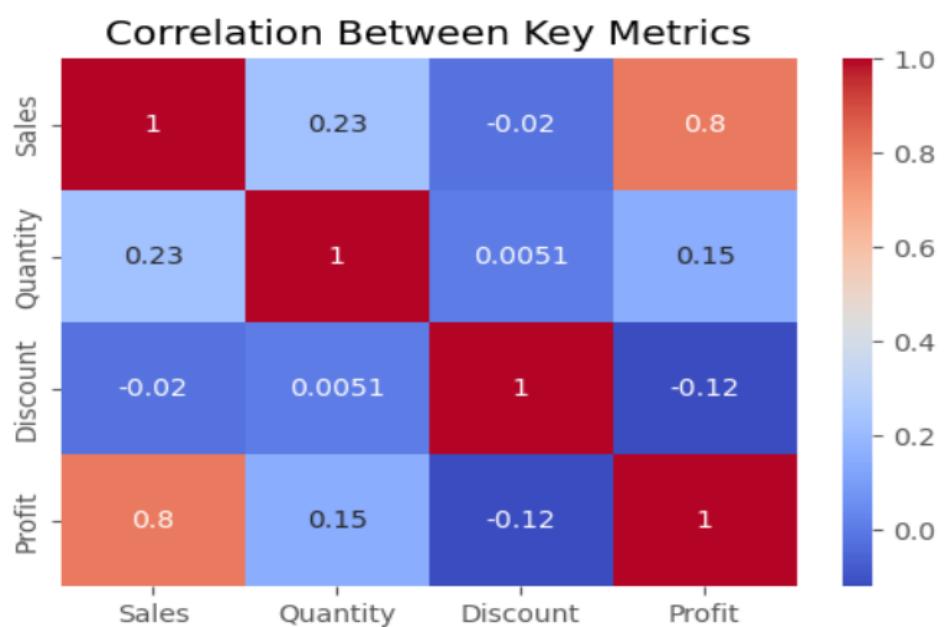
- **Category Analysis:** Technology category dominates both sales and profit.



- **Regional Analysis:** Eastern region leads in profitability.



- **Correlation Heatmap:** Negative correlation between discount and profit (-0.22).



Dashboard Development (Tableau Phase)

Data Sources

- category_summary.csv
- monthly_summary.csv
- region_summary.csv

Dashboard Components

Chart	Description
Monthly Trend	Line chart showing monthly sales and profit trends.
Category Breakdown	Bar chart visualizing category-wise sales and profit.
Regional Profitability	Bar chart showing profit distribution by region.
Filters	Interactivity by segment, category, and region.

Layout

| Monthly Sales & Profit Trend |

| Category-wise Sales | Region-wise Profit |

Output File: Retail_Profitability_Dashboard.twbx

Key Insights

Area	Observation	Impact
Category	Technology drives 36% of total profit	Focus promotions on high-margin SKUs
Region	West region most profitable	Replicate successful strategies to other regions
Discount	Negative impact on profit	Review and optimize discounting policy

Area	Observation	Impact
Seasonality	Sales peak in November–December	Plan inventory and marketing campaigns accordingly
Segments	Corporate and Home Office are steady profit sources	Enhance loyalty programs for these segments

Recommendations

1. **Optimize discount policies** to maintain healthy profit margins.
2. **Focus on Technology category** and expand product offerings.
3. **Analyze low-performing categories** (like Furniture) for pricing or cost issues.
4. **Prepare for Q4 sales surge** with proactive stock and marketing strategies.
5. **Replicate West region strategies** in other markets.

Conclusion

This project demonstrates a **complete data analytics workflow**:

- SQL for **data cleaning and preparation**
- Python for **exploration and insight discovery**
- Tableau for **visual storytelling and decision-making**

It highlights how integrated data analysis can transform raw transactional data into **strategic business insights** for profit growth and operational excellence.

Deliverable: Retail_Profitability_Analysis/reports/Final_Report.pdf