



Project Report: Retail Business Performance & Profitability Analysis

1. Introduction

The project “*Retail Business Performance & Profitability Analysis*” focuses on analyzing a retail company’s sales, profit, and inventory data to uncover valuable business insights.

Using Python for data analysis and Tableau for visualization, the project aims to identify top-performing products and categories, optimize inventory levels, and improve overall profitability.

The dataset contains 50,000 retail transactions covering one year (2021) with details like product, category, price, cost, region, quantity, and customer.

2. Abstract

The primary goal of this project is to transform raw retail transaction data into meaningful insights through data cleaning, visualization, and predictive analysis.

The workflow included preprocessing data in Python using Pandas, generating summary statistics, and visualizing the results in Tableau through an interactive dashboard.

A RandomForest regression model was developed to predict product-level profitability using factors such as stock days, sales quantity, and profit margin.

The Tableau dashboard highlights key business metrics such as total sales, profit, average margin, and stock performance — enabling data-driven decisions.

3. Tools Used

Tool	Purpose
Python (Jupyter Notebook)	Data cleaning, KPI calculation, model building
Libraries: Pandas, NumPy, Matplotlib, Scikit-learn	Data processing, visualization, and machine learning
Tableau	Dashboard creation and interactive visualization
CSV Dataset: retail_sales_clean.csv	50,000 records of sales data
MS Excel	Quick checks and verification of CSV outputs

4. Steps Involved in Building the Project

1. Data Collection & Preparation

- Generated a synthetic retail dataset (retail_sales_clean.csv) with 50,000 records.

- Verified data types, checked for missing values and duplicates, and created derived columns such as $Sales = Unit\ Price \times Quantity$.

2. Exploratory Data Analysis (Python)

- Calculated KPIs: Total Sales, Total Profit, Average Margin, Average Stock Days, and Unique Products/Customers.
- Grouped data by *Category*, *Product*, and *Date* to identify top-performing areas.
- Visualized profit trends, category performance, and correlations using Matplotlib.

3. Predictive Modeling

- Used a RandomForestRegressor to predict *Total Profit* based on *Avg Stock Days*, *Total Quantity*, and *Avg Margin*.
- Model achieved low error (MAE ≈ 65 , RMSE ≈ 89), with *Quantity Sold* being the top predictor of profit.

4. Tableau Dashboard Creation

- Connected aggregated CSV files (category_agg.csv, product_agg.csv, daily_summary.csv, kpi_summary.csv).
- Built four key sections:
 - **KPI Tiles** showing total sales, total profit, avg margin, and stock days.
 - **Profit by Category** bar chart.
 - **Daily Profit Trend** line chart with 7-day moving average.
 - **Inventory vs Profit** scatter plot and dynamic **Top N Products** parameter filter.
- Added interactive filters and tooltips for better exploration.

5. Conclusion

This project successfully demonstrates how data analytics and visualization can optimize retail operations.

The Python-based analysis provided accurate KPIs and predictive insights, while the Tableau dashboard made these insights interactive and easy to interpret.

Findings indicate that **Home** and **Electronics** categories contribute the most profit, and **quantity sold** is the strongest driver of product profitability.

Future work could extend this project to include **demand forecasting**, **stock optimization**, and **real-time data updates** for continuous business improvement.