Retail Sales Performance Analysis

Data Cleaning, Analysis & Power BI Dashboard Report

By: Habib Ullah Khan
Tools Used: Excel, Python (Pandas, Matplotlib, Seaborn), Power BI
Dataset: Kaggle Sample Superstore

1. Introduction

In today's competitive retail industry, data-driven insights have become crucial for sustaining profitability and identifying hidden inefficiencies. The dataset used for this project was derived from Kaggle's ***Sample Superstore*** data a comprehensive record of retail transactions containing order-level details such as sales, profit, discount, shipping mode, and customer information.

The main objective of this analysis was to uncover performance issues that affect overall profitability, particularly those related to discounts, regional trends, and category-level sales. By using Python for data cleaning and exploration, and Power BI for interactive visual storytelling, the project transforms raw data into actionable business insights.

2. Problem Overview

At first glance, the company's sales numbers appeared healthy orders were increasing, and customer segments were well-distributed. However, deeper investigation revealed that high sales did not always translate into high profits.

Several underlying issues became clear:

- Some product categories such as Furniture were generating strong revenue but suffering from low or even negative profit margins.
- The discount strategy lacked clear boundaries heavy discounts were being applied in regions or products that did not require them, causing major profit erosion.
- Certain regions, especially the Central region, consistently underperformed despite similar customer bases and product availability.Shipping costs and lead times also varied dramatically, hinting at operational inefficiencies.

These challenges reflected a common business problem: growth without profitability. The goal of this analysis was to identify what was driving those inefficiencies and how strategic adjustments could turn sales into sustainable profit.

3. Data Cleaning and Preparation (Python)

Before building the Power BI dashboard, the dataset underwent extensive preparation and cleaning in Python. Using Pandas, I first removed irrelevant columns and handled missing or duplicate entries to ensure data accuracy.

Next, I converted date fields such as *Order Date* and *Ship Date* into a proper datetime format and derived new columns like Year, Month, and Shipping Duration to allow time-based trend analysis. Numerical outliers in *Sales* and *Profit* were reviewed and capped to prevent skewed visualizations.

Exploratory Data Analysis (EDA) was then performed using Matplotlib and Seaborn to gain early insights. For example, scatterplots revealed that higher discounts consistently reduced profit margins, while bar charts showed Technology as the strongest category contributor. These early visuals guided how the Power BI dashboard would later be structured.
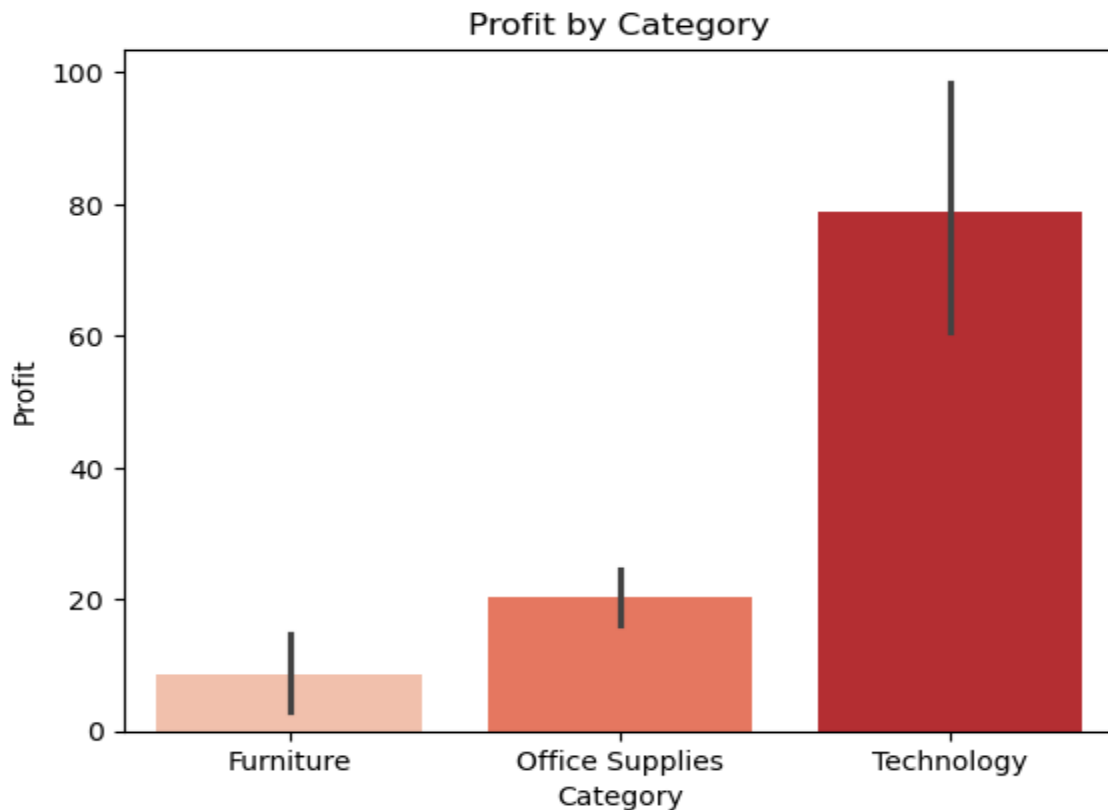
Fig 1: profit by category using python library Matplotlib for bar chart
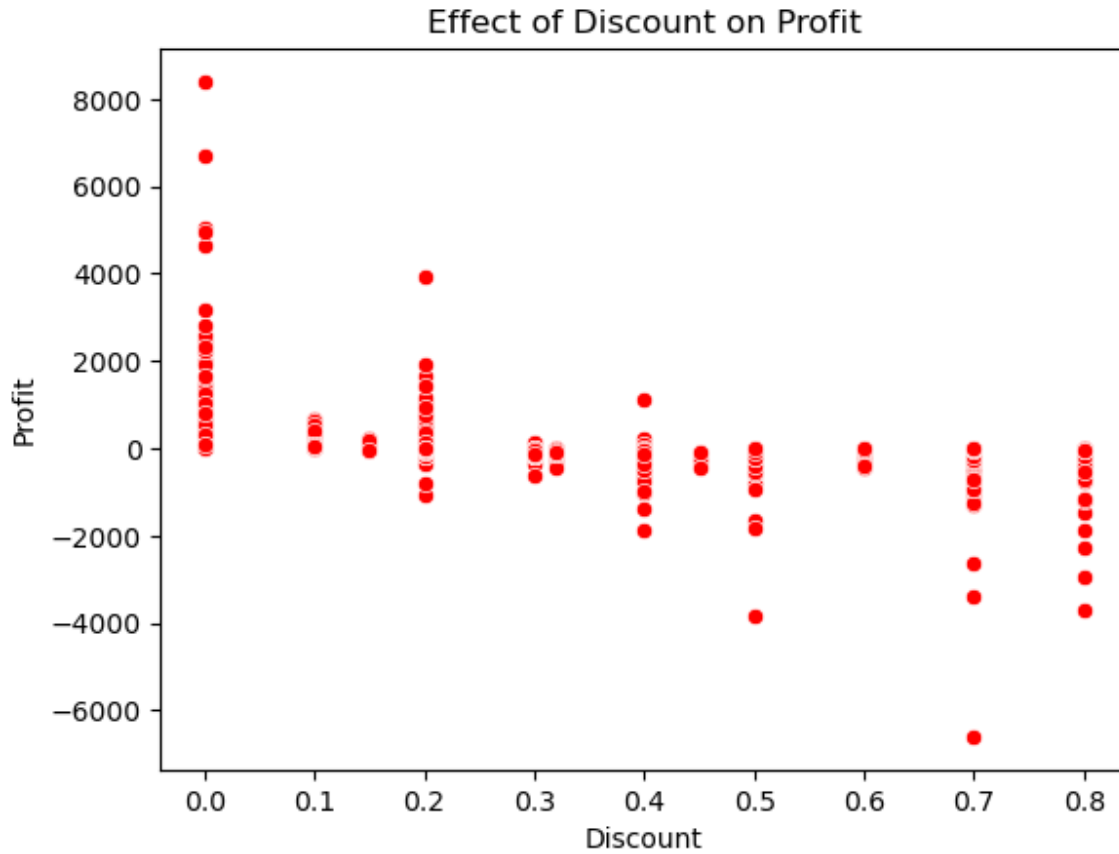
Fig:2 Effect of Discount on profit using

By the end of the cleaning phase, the dataset was standardized, formatted, and ready for Power BI ingestion. This ensured that all calculations and visuals were built on trustworthy data.

4. Power BI Dashboard and Analysis

Once the clean dataset was imported into Power BI, I developed an interactive dashboard that could present all major business insights in a single visual story. The dashboard design followed a hierarchical analysis flow moving from overall KPIs to regional, category, and product-level insights.

The main page displayed key performance indicators (KPIs) such as:

- Total Sales

- Total Profit

- Average Discount

- Total Orders

These KPIs were supported by visual elements like maps, bar charts, and line graphs, enabling quick comparison between regions, categories, and time periods.
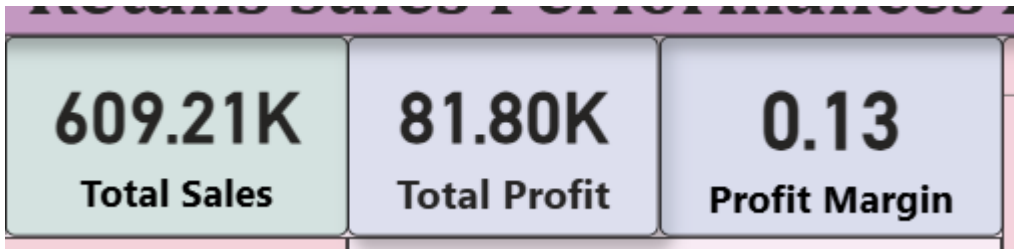

Fig3: : KPI section total sales, profit, and discount visual in Power BI

Through this dashboard, several crucial findings emerged:

- The West region contributed the highest sales and profits, indicating mature operations and efficient pricing.
- The Central region underperformed in both sales and profit, largely due to excessive discounting and inconsistent shipping timelines.
- Technology products drove the highest profitability, while Furniture lagged behind despite similar sales volume.
- Month-over-month analysis revealed a sharp rise in sales during November and December likely linked to seasonal promotions and holiday demand.
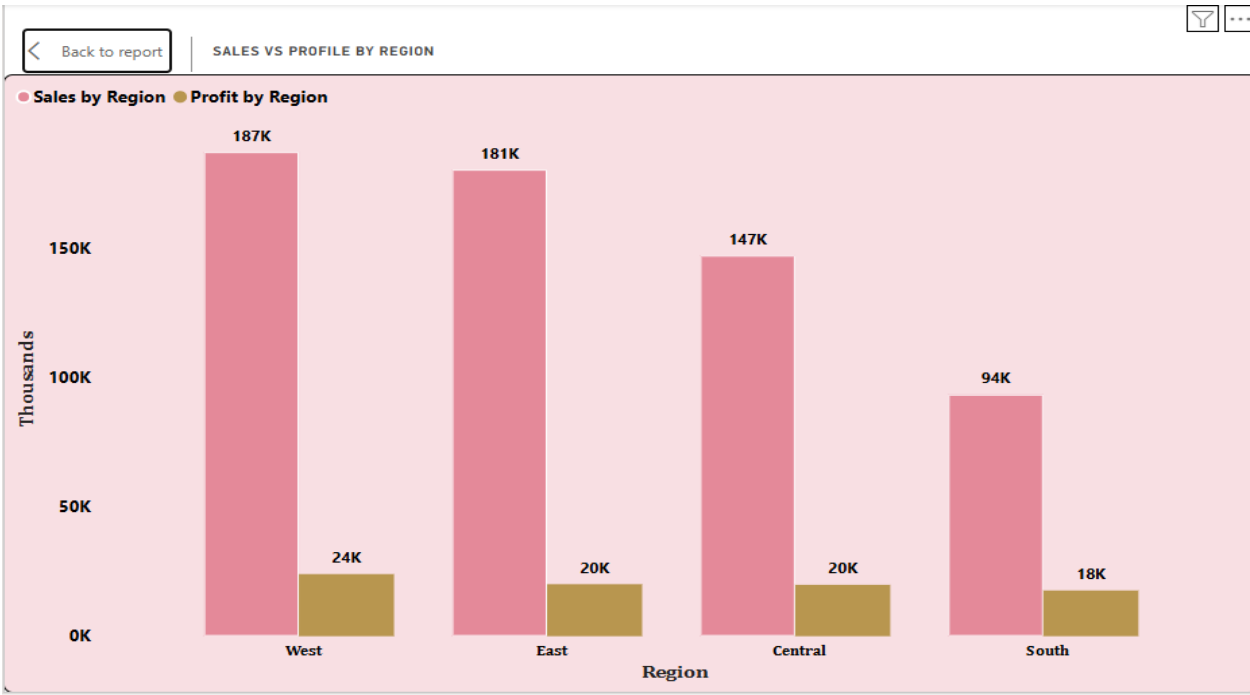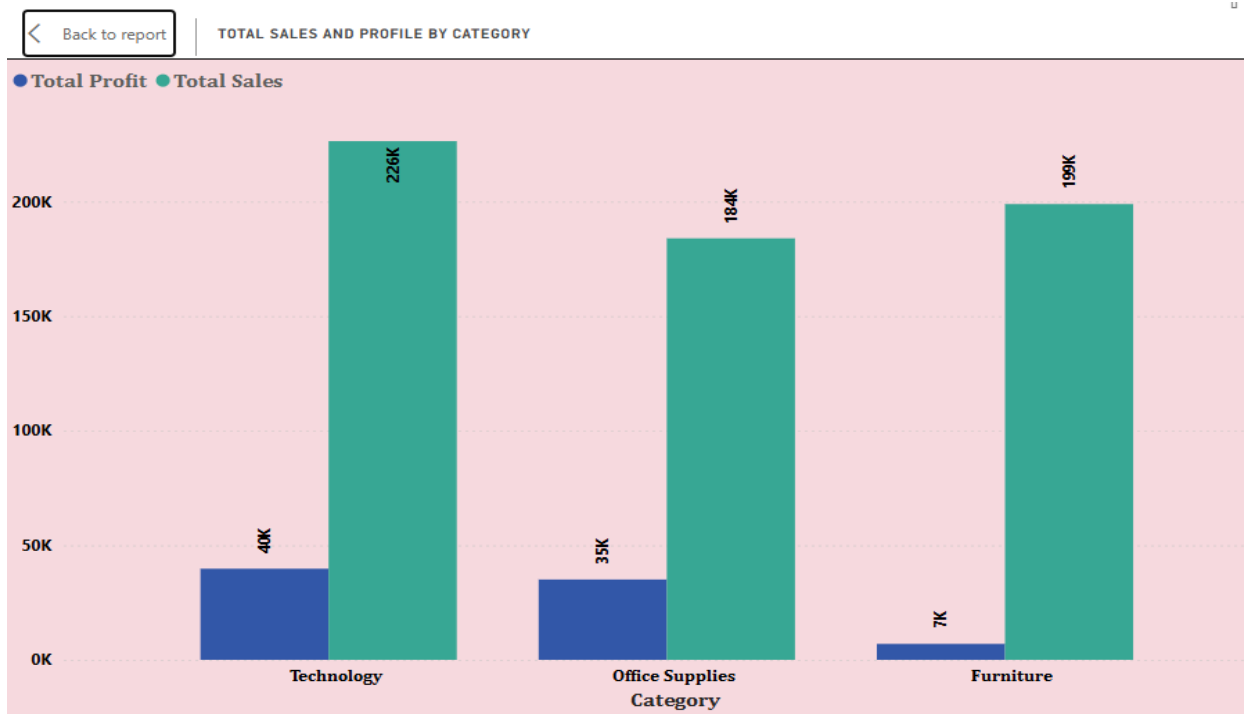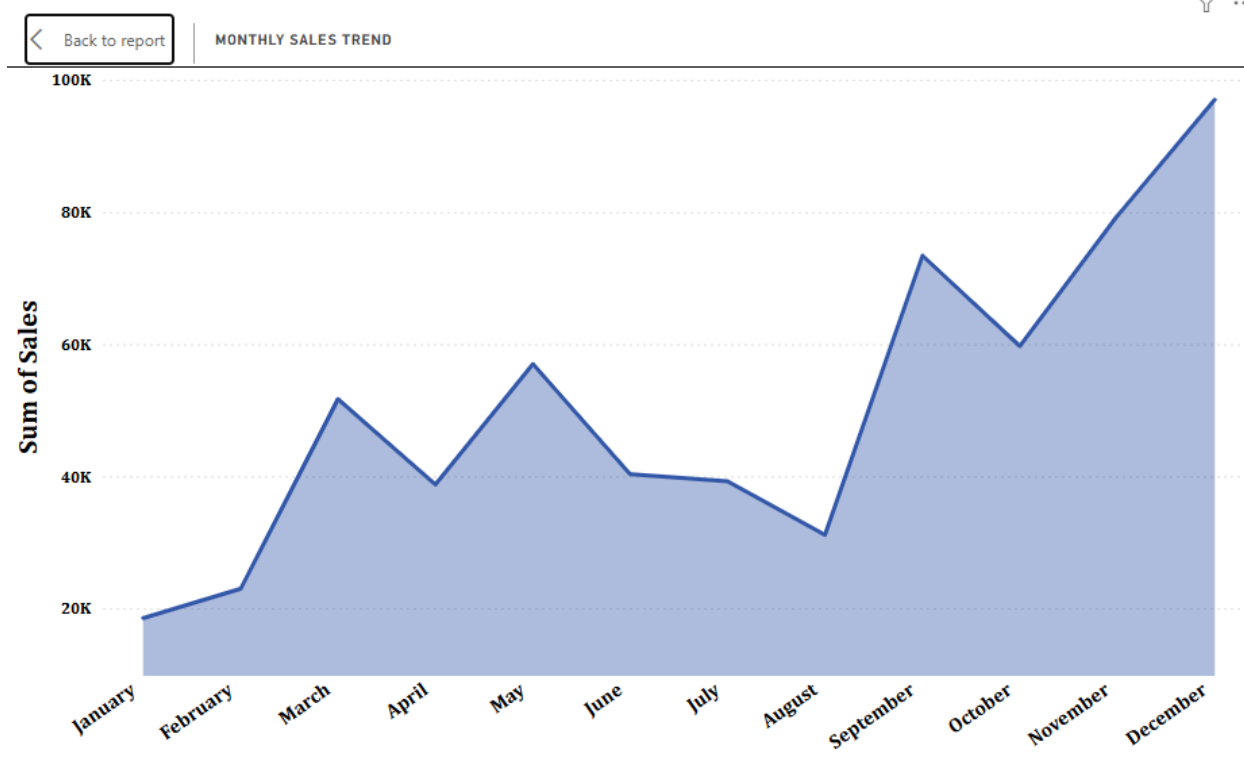

Fig 4: Bar chart of Sales vs Profit by Regional Power BI

● Total Profit ● Total Sales

Fig 5: Total Sales and Profit by Category Clustered Column Chart Power BI

Fig 6: Line Chart of Monthly Sales Trend Power BI

By making the dashboard interactive with slicers for *Category*, *Region*, and *Year*, decision-makers can dynamically explore these insights and identify exactly where performance gaps lie.

5. Insights and Business Recommendations

After analyzing both the Python results and Power BI visualizations, several actionable insights emerged that could directly enhance business strategy:

1. **Optimize Discount Strategy:**
   Discounts beyond 30% tend to cause losses, especially in Furniture. Implementing a structured discount policy based on profit margin and inventory level would preserve profitability.

2. **Focus on High-Value Regions and Products:**
   Investment should be prioritized in high-performing regions (like the West) and product lines (Technology), while re-evaluating logistics and pricing models in low-performing ones.

3. **Improve Shipping Efficiency:**
   Variability in delivery time between regions suggests room for supply chain optimization. Shortening delivery lead time can improve customer satisfaction and repeat sales.

4. **Seasonal Planning:**
   Since sales peak in November–December, preparing inventory and targeted marketing before these months could yield even higher returns.
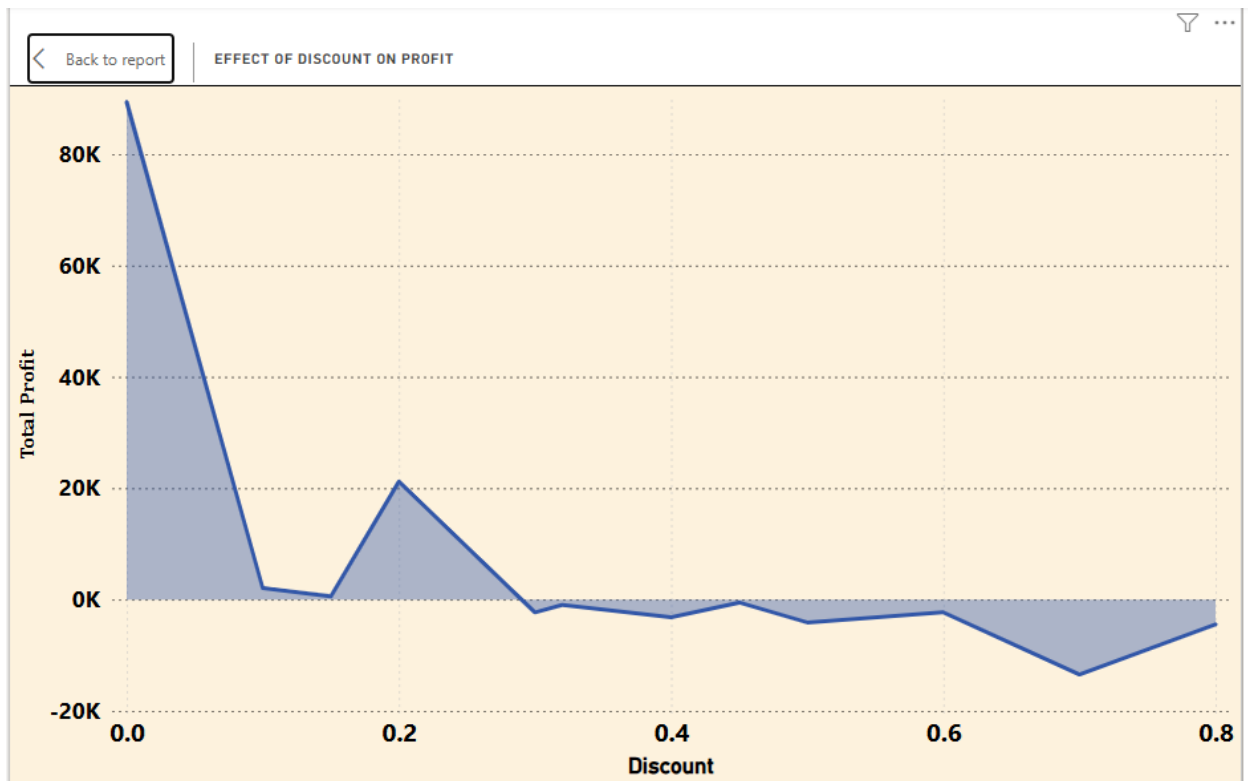

Fig 7: Effect of Discount on Profit Power BI

6. Conclusion

This project demonstrates how a combination of Python data cleaning and exploratory analysis with Power BI visualization can transform raw retail data into clear, actionable business insights. The analysis not only identified where the company was losing profit but also provided a roadmap for recovery and growth.

By systematically identifying discount inefficiencies, regional inconsistencies, and category performance differences, this analysis helps management make evidence-based decisions. The Power BI dashboard now serves as a live monitoring tool — allowing the team to track sales performance, identify emerging trends, and respond quickly to market changes.

Ultimately, this project illustrates the true value of data analytics: turning information into intelligence, and intelligence into impact.
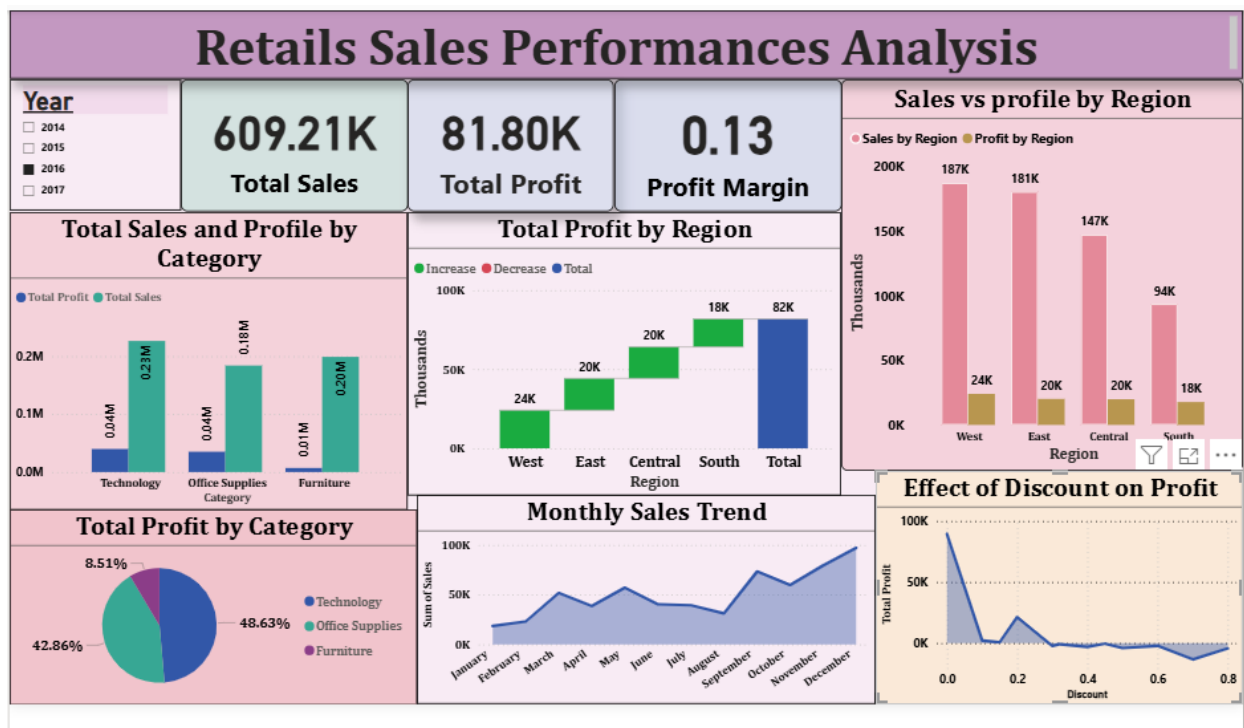
Fig 9: Final Power BI dashboard summary view