Superstore Sales Performance Analysis

An End-to-End Data Analytics, Data Science & Visualization Project

KAVYA TALWAR
BSc. MSCs.
St. Joseph's Degree and P.G. College, Hyderabad

Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn)
Google Colab · Power BI

Superstore Sales Performance Analysis

Index

---

1. Title

Superstore Sales Performance Analysis
A comprehensive project integrating data analytics, data science, and business intelligence visualization.

---

2. Objective

The goal of this project is to analyze historical sales data from a retail superstore to uncover actionable insights related to product performance, regional profitability, customer behavior, and discounting strategies. The findings are visualized through an interactive dashboard to support data-driven decision-making.

---

3. Tools Used

Python (executed via Google Colab)

- **NumPy** – Efficient numerical operations and data structure support.
- **Pandas** – For data loading, cleaning, transformation, and exploration.
- **Matplotlib & Seaborn** – Static visualizations and correlation heatmaps.
- **Scikit-learn** – Optional modeling (linear regression) for profit prediction.

Power BI

Used for building a dynamic and interactive dashboard featuring:

- KPI Cards
- Bar, Donut, Line, Scatter, and Matrix charts
- Slicers for interactivity
- Drill-down and filtering capabilities

---

4. Dataset Description

- **Name**: Superstore Dataset
- **Source**: Public dataset available on Kaggle
- **Link**: Kaggle - Superstore Dataset
- **Records**: 9,994
- **Features**: 21 columns

Key Columns Used:

- Order Date
- Sales
- Profit
- Quantity
- Discount

- Segment
- Category and Sub-Category
- Region, State
- Product Name

This dataset represents sales transactions from a fictional U.S. retail store chain across four years (2014–2017), offering a wide variety of consumer products.

---

5. Methodology

Step 1: Data Cleaning & Exploration (Python)

- **File Import**: Handled Unicode errors by specifying encoding (`ISO-8859-1`)

- **Cleaning**:

    - Removed duplicates and unused columns
    - Converted date columns into proper datetime formats
    - Checked and addressed missing values or invalid data types

- **Exploratory Data Analysis**:

    - Aggregated sales and profit across region, category, and sub-category
    - Identified trends, spikes, and outliers
    - Created a correlation matrix to examine the relationship between discount and profit

Step 2: Data Science Component (Optional)

- **Linear Regression**: Modeled how varying discounts influence profit margins

- **Pattern Identification**:

    - Used `groupby()` to detect segment-level profitability
    - Ranked sub-categories by performance

- **Visualization**:

    - Scatter plot of discount vs. profit
    - Heatmaps to visualize feature relationships

Step 3: Data Visualization in Power BI

Constructed a professional dashboard with:

KPI Cards:

- Total Sales: $2.30M
- Total Profit: $286.40K
- Total Orders: 5,009
- Average Discount: 0.16

Visual Charts:

- **Sales & Profit by Region** (Bar Chart with Metric Switch)
- **Profit by Category & Sub-Category** (Matrix Visual)
- **Monthly Sales Trend** (Line Chart)
- **Sales by Segment** (Donut Chart)
- **Impact of Discount on Profit** (Scatter Plot)
- **Top 10 Products by Sales** (Horizontal Bar Chart)

Slicers for Interactivity:

- Year
- Region
- Category

Styling:

- A clean layout with distinct sections
- Thematic color-coded KPI cards
- Footer with author credits:
  Prepared by Kavya Talwar | *Superstore Dataset* | *July 2025*

---

6. Key Insights

- **Technology** is the most profitable category, performing well across all regions.
- **Furniture** underperforms consistently, with losses likely linked to high discounting.
- **Discounts** show a strong negative correlation with profit — excessive discounts reduce margins significantly.
- **The South region**, while achieving strong sales figures, reports lower profitability.
- **Top 10 products** contribute a significant portion of overall sales, suggesting room for targeted inventory or promotional focus.
- The **Consumer segment** dominates in both order volume and sales, indicating a key target audience.

---

7. Deliverables

- **Python Notebook (.ipynb)**
  Includes data import, cleaning, EDA, and optional modeling with

plots.

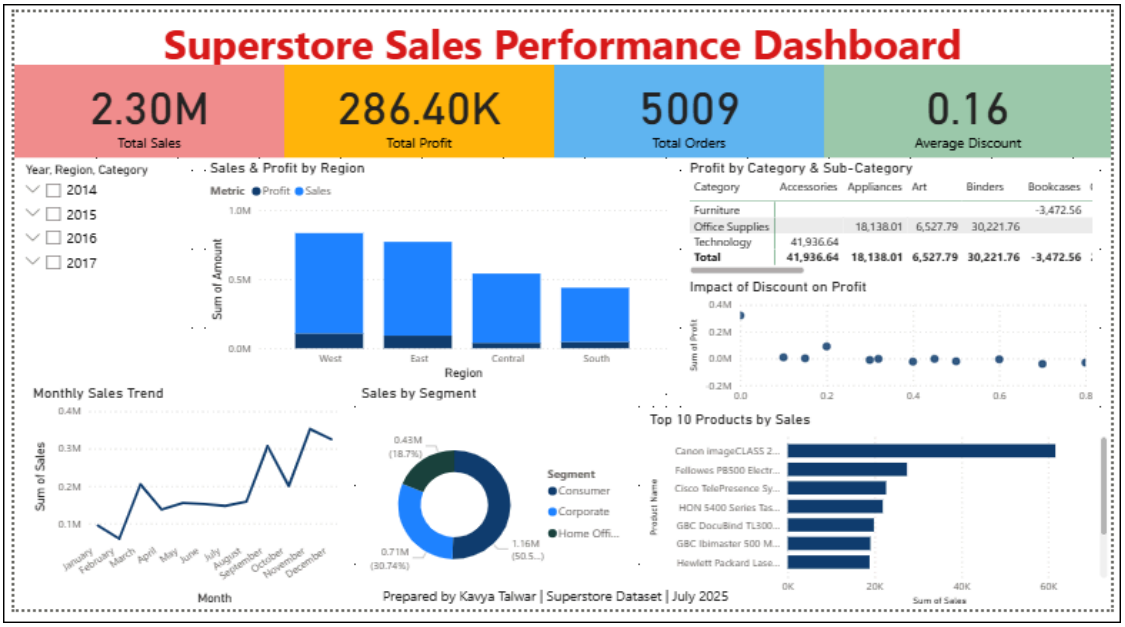
- **Power BI Dashboard (.pbix)**
  Fully interactive and presentation-ready report with filters, KPIs, and visual insights.

- **Summary Report (PDF)**
  This comprehensive document consolidating all findings, methodology, and visuals.

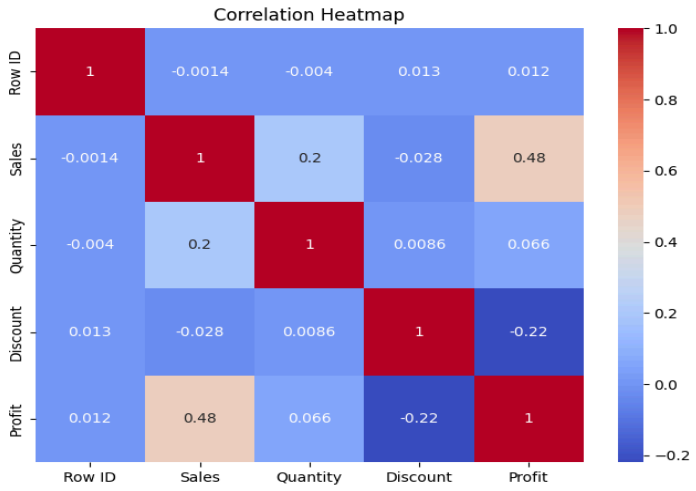
---

8. Screenshots

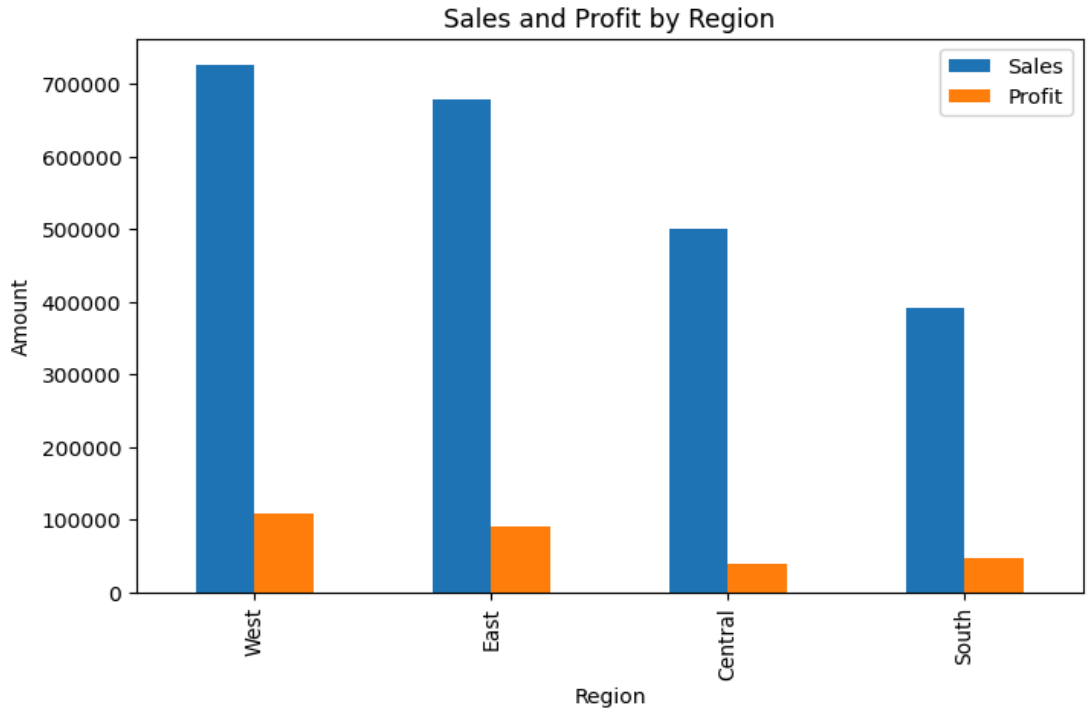
- ✅ Final Power BI Dashboard



- ✅ Python correlation heatmap / scatter plot



- ✅ Key EDA outputs



---

9. Conclusion

This project demonstrates how data can be leveraged across analytics, science, and visualization workflows to extract actionable insights. From

Python-based exploration and modeling to an interactive Power BI dashboard, the Superstore data has been analyzed in depth.

The findings highlight how pricing strategies, regional performance, and product segmentation affect business profitability. With clear visuals and real-time interactivity, the dashboard serves as a valuable decision-making tool for retail business managers.

By combining technical skills with business intuition, this project showcases a practical approach to transforming raw data into strategic insight.
