

Sales Trend Analysis Report

The "Sample Supermarket Sales" dataset underwent a rigorous analysis aimed at uncovering critical insights to inform strategic business decisions. This analysis entailed data preprocessing, exploratory data analysis (EDA), statistical evaluation, and predictive modeling to address pivotal questions, including the identification of the most profitable regions and categories, the quantification of discount impacts on profitability, the elucidation of relationships among sales, profit, and discounts, and the predictive modeling of profitability outcomes.

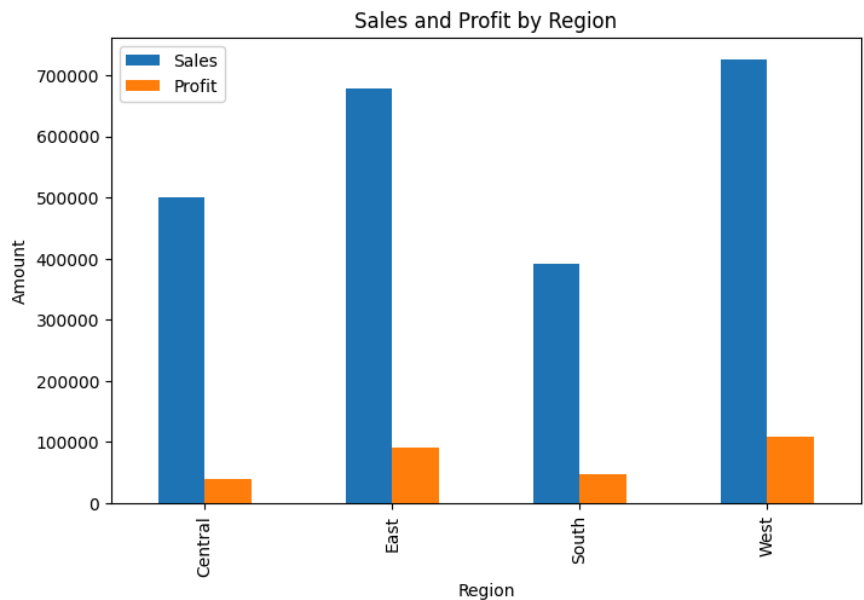
The dataset consisted of 9,994 entries distributed across 13 variables, including key numerical metrics such as Sales, Profit, and Discount, as well as categorical variables such as Region, Category, and Sub-Category. Data cleaning identified and removed 17 duplicate entries, ensuring the integrity of the dataset. No missing values were detected, and all variables were appropriately structured to facilitate analysis.

Key Metrics and Performance Highlights

The dataset revealed an aggregate sales value of \$2,296,195.59, coupled with a total profit of \$286,241.42, and an average discount rate of 15.63%. These overarching metrics serve as a baseline for subsequent in-depth analyses.

Regional Performance

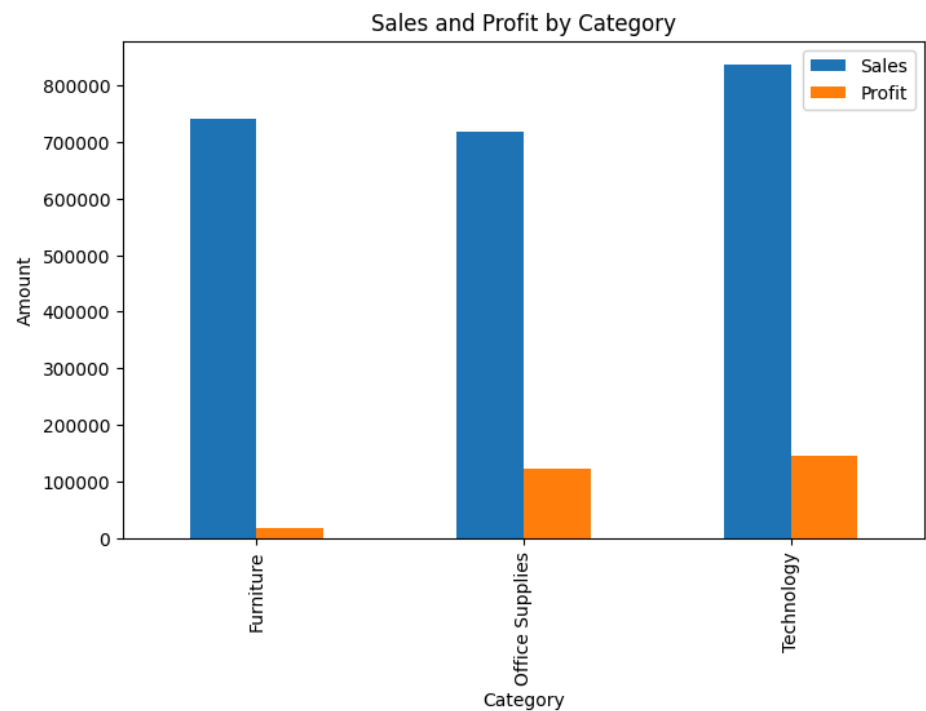
Substantial variation was observed in sales and profit metrics across regions. The West region emerged as the most lucrative, generating \$725,255.64 in sales and contributing \$108,329.81 in profit. In contrast, the



Central region recorded the lowest profitability (\$39,655.88) despite notable sales (\$500,782.85), suggesting inefficiencies or elevated operational costs in this region. A bar chart of "Sales and Profit by Region" visually underscores these discrepancies.

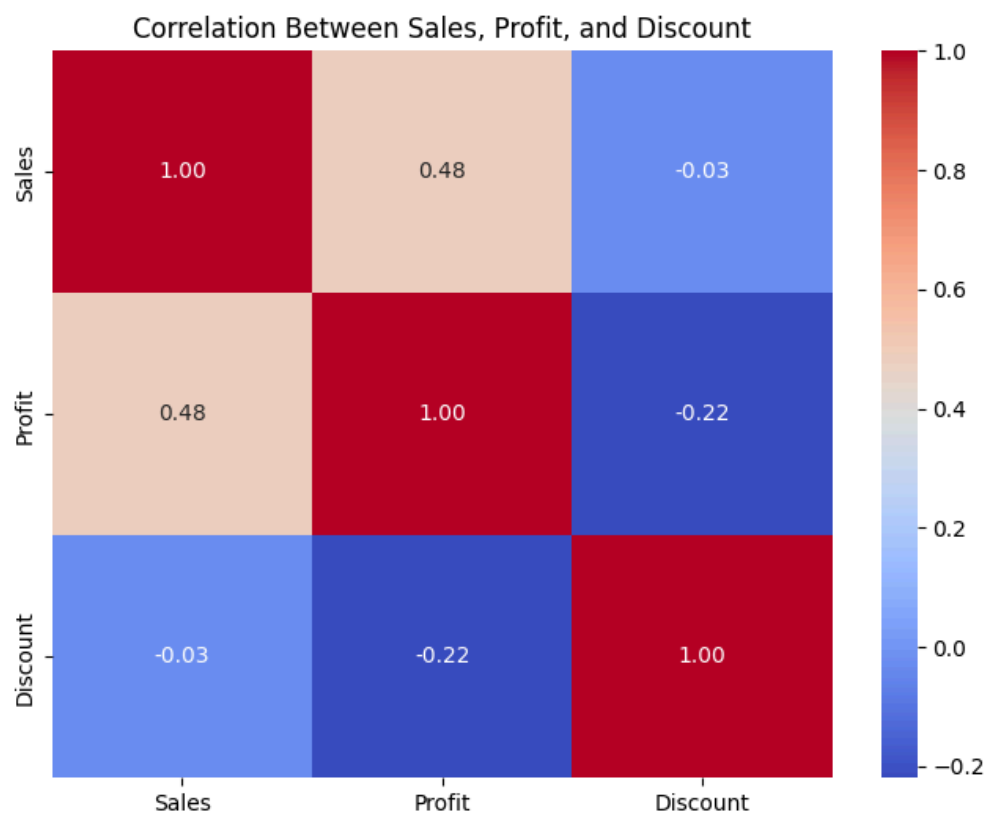
Category Performance

The Technology category demonstrated superior performance, achieving the highest sales (\$836,154.03) and profit (\$145,454.95). Conversely, the Furniture category exhibited suboptimal profitability (\$18,421.81) despite generating substantial sales (\$741,306.31), likely due to elevated costs or disproportionate discounting. Office Supplies represented a middle ground, contributing \$122,364.66 in profit from \$718,735.24 in sales. These trends are visually highlighted in the "Sales and Profit by Category" bar chart.



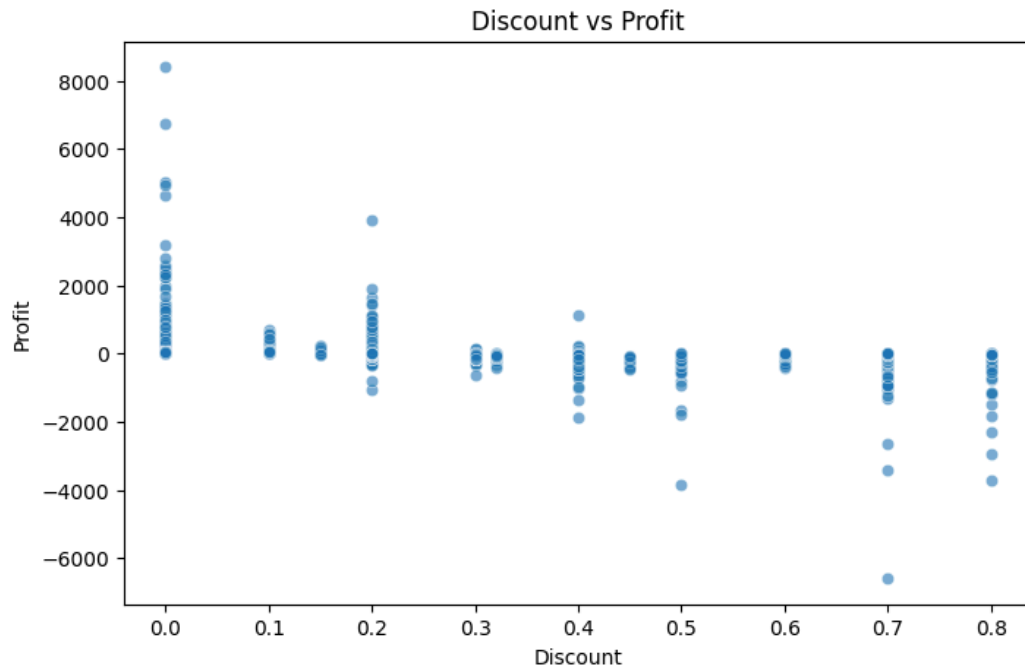
Correlation Analysis

Correlation analysis revealed notable relationships among numerical variables. A moderate positive correlation (0.48) was observed between Sales and Profit, suggesting that increased sales generally correlate with higher profits. However, a negative correlation (-0.22) between Discount and Profit indicated that higher discounts are detrimental to profitability. The weak correlation (-0.03) between Sales and Discount implies limited direct interaction between these variables. These relationships are visualized in the accompanying correlation heatmap.



Impact of Discounts

Further analysis of discount impacts revealed a clear negative trend: discounts exceeding 20% were frequently associated with losses. A scatter plot of Discount versus Profit underscores this trend, highlighting the need for disciplined discount management to avoid compromising profit margins.



Predictive Modeling Results

The dataset supported the development of predictive models to enhance business decision-making:

- **Linear Regression Model:** This model, aimed at forecasting sales, produced a **Mean Squared Error (MSE)** of **344,656.87**, demonstrating moderate predictive accuracy.
- **Random Forest Classifier:** Designed to predict profitability, this model achieved an accuracy of **93.34%**, effectively leveraging features such as Sales, Discount, and Quantity to classify orders as profitable or non-profitable.

Recommendations

This analysis suggests several actionable strategies to optimize performance:

1. **Prioritize the West Region:** Given its robust sales and profitability, the West region warrants increased resource allocation and marketing efforts.
2. **Refine Discount Strategies:** Discounts exceeding 20% should be minimized, as they significantly erode profitability.

3. **Enhance Furniture Margins:** Further investigation into production costs, pricing, and discount practices within the Furniture category is recommended to bolster profitability.
4. **Capitalize on Technology Products:** Investment in the Technology category is advised, given its consistently high profitability and sales volume.

Conclusion

The findings from this analysis offer a data-driven foundation for enhancing profitability and operational efficiency. By focusing on high-performing regions and categories while implementing disciplined discount management, the organization can achieve substantial financial growth. For a detailed account of the methods, visualizations, and predictive models, refer to the accompanying Jupyter Notebook and GitHub repository.

Appendix

All visualizations and analytical code are included in the Jupyter Notebook accompanying this report. For further details, please consult the GitHub repository.