Exploring Superstore Insights: Trends, Profitability, and Predictions

Q: How do sales and discounts influence profitability, and how accurately can we predict profits based on key business metrics?

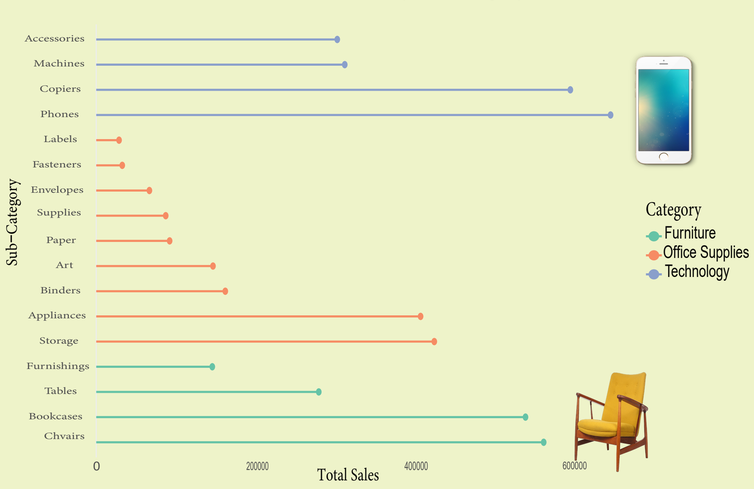
By: Rithvik Rangaraj

Introduction

This project delves into a superstore dataset to uncover sales and profit patterns across regions, segments, and categories. By analyzing seasonal trends, regional profitability, and the effects of discounts, we provide actionable insights. A predictive model further explores factors driving profitability, offering a data-driven approach to optimize performance.

Main Story and Motivation

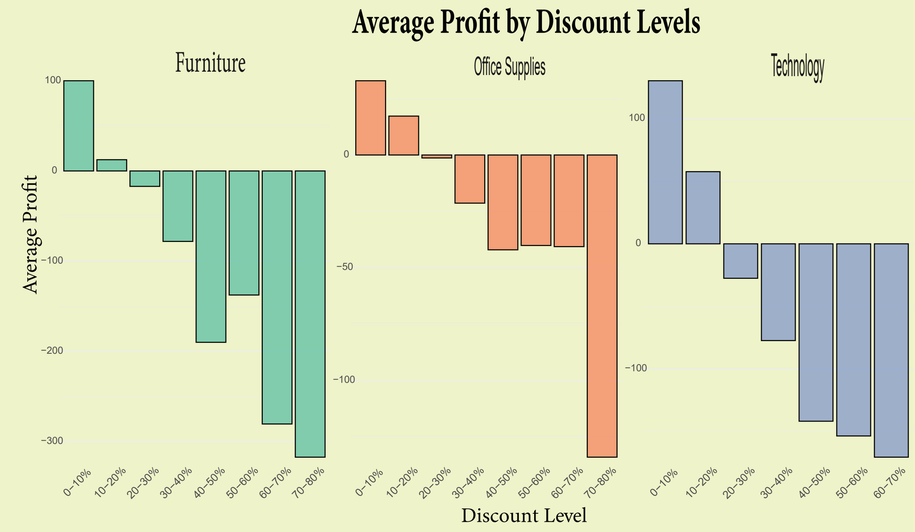
Seasonal trends highlight peak and valley months, while regional insights reveal performance differences across customer segments. The analysis also examines the effect of discounts on profits and uses predictive modeling to estimate profitability based on key business metrics. Retailers and business analysts are always looking for ways to optimize sales and maximize profits. This analysis provides actionable insights into key drivers of profitability, helping businesses make data-driven decisions to improve overall outcomes.



The sub-categories with the highest sales, such as Phones and Chairs, drive the majority of revenue, while smaller sub-categories like Fasteners contribute minimally, highlighting key areas for inventory focus.

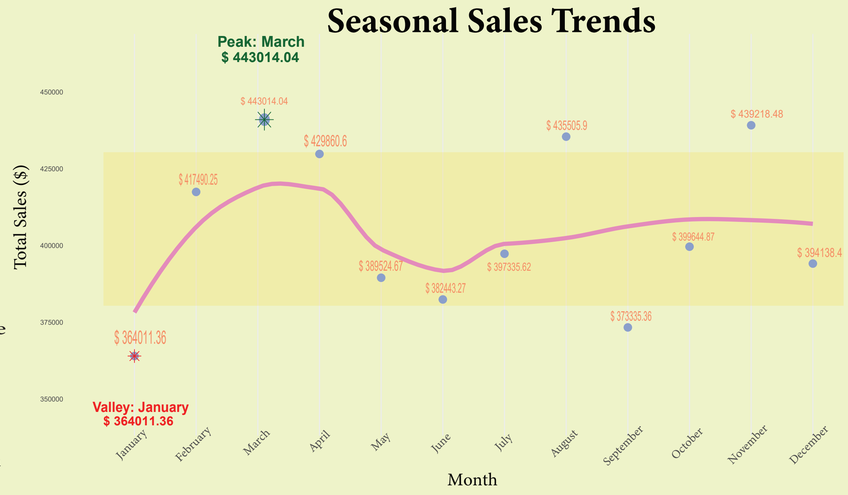
Sources

The dataset used is the **Superstore Dataset (2011–2015)**, obtained from Kaggle. Analyses and visualizations were performed in R using packages such as dplyr, ggplot2, RColorBrewer, caret, and randomForest. The final poster layout and design enhancements were completed in Adobe Illustrator.



Higher discount levels generally lead to lower profitability, especially in Furniture and Office Supplies categories.

Technology shows resilience, maintaining moderate profits at discount levels up to 30%, unlike other categories



Sales **peak** in **March**, reflecting **strong** spring activity, while **January** shows the **lowest** sales, likely due to post-holiday slowdowns.

Predictive Model: Overview

To uncover the key drivers of profitability, a Random Forest Regression model was developed, leveraging its robust machine learning capabilities. This model predicts profit using critical business metrics, including:

- **Sales**: Total revenue from transactions.
- **Discount**: Percentage reductions applied to products.
- **Quantity**: Number of items sold.
- **Shipping Cost**: Expenses incurred during delivery.

Model Performance

- Mean Absolute Error (MAE): **35.29**
  - On average, the model's predictions deviate by **$35.29** from the actual profit values.
- Mean Squared Error (MSE): **7324.26**
  - Represents the average squared difference between predicted and actual profits, emphasizing accuracy across all ranges.
- Root Mean Squared Error (RMSE): **85.58**
  - Measures the standard deviation of prediction errors, reflecting the model's reliability.
- R-squared (R^2): **0.70**
  - Indicates that **70% of the variance** in profit can be explained by the model.

Comparison of Actual vs Predicted Profits

Grouped by Sales Range