Final Visualization Report: Superstore Dataset Analysis

Student Name: Fawad Anjum
Student ID: 23078277
GitHub Repo Link: https://github.com/fa24abb/final-Report-Visualization.git

Introduction

The analysis of the current paper focuses on a retail dataset that covers four years of data on sales, profit, customer segmentation and product categories. The purpose is to identify patterns in sales and profit over time, product lines, and geography. The application of data visualization and time series analysis is used to underpin the findings of the study.

Dataset Overview

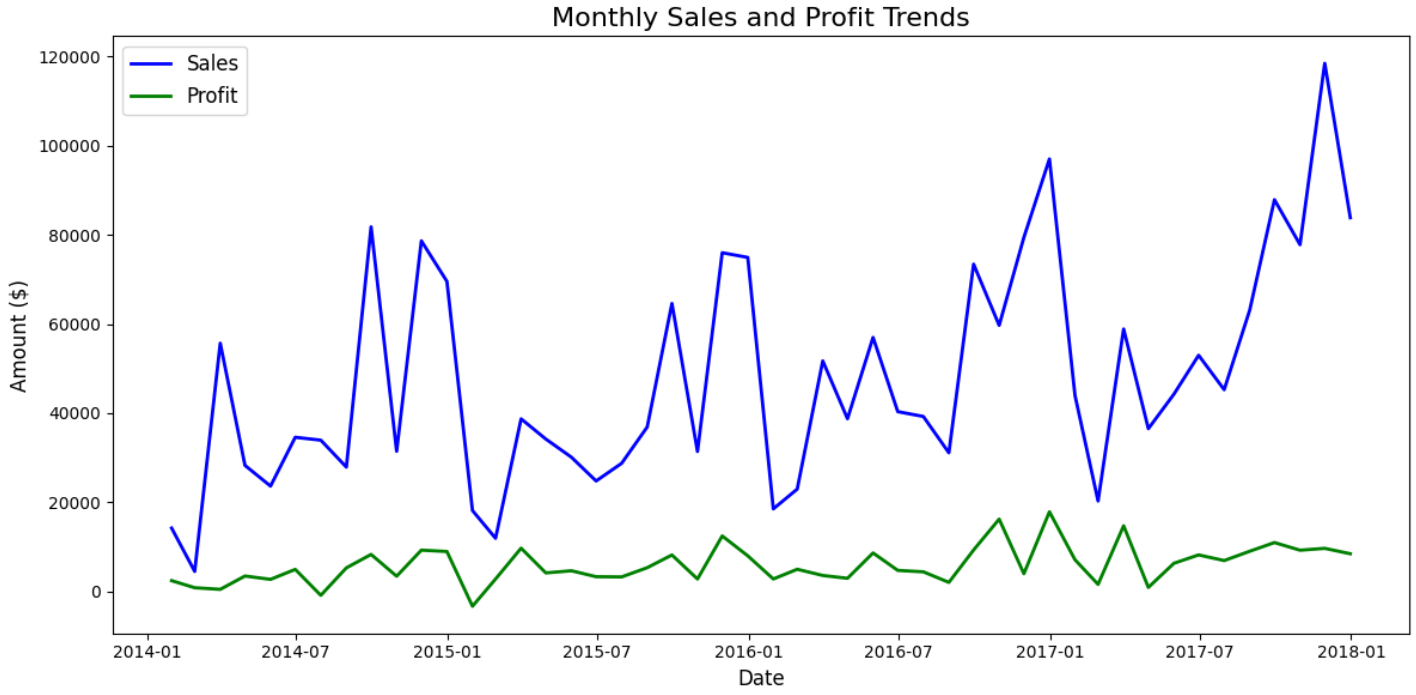
This dataset has 21 fields and multiple records of orders. The categorical variables are: shipping mode, product category, and region, while the quantitative variables include sales, profit and quantity sold. For the analysis, data is aggregated at the monthly level to determine the temporal trends. The breakdowns by sub-categories and scattergrams are used to make additional analyses.

Results and Discussion

Monthly Sales and Profit Trends

A time series analysis was conducted to evaluate the overall performance of sales and profit. Over 48 months, sales exhibited an upward trend, peaking during holiday seasons such as November and December. Profit, while following a similar seasonal pattern, displayed sharper fluctuations due to varying discount policies. For instance, December consistently saw high sales but reduced profits, likely due to promotional campaigns.

- Average Monthly Sales: **$21,765.93**.
- Average Monthly Profit: **$2,746.71**
- Months with maximum sales and profits were December 2017 and October 2016, respectively.

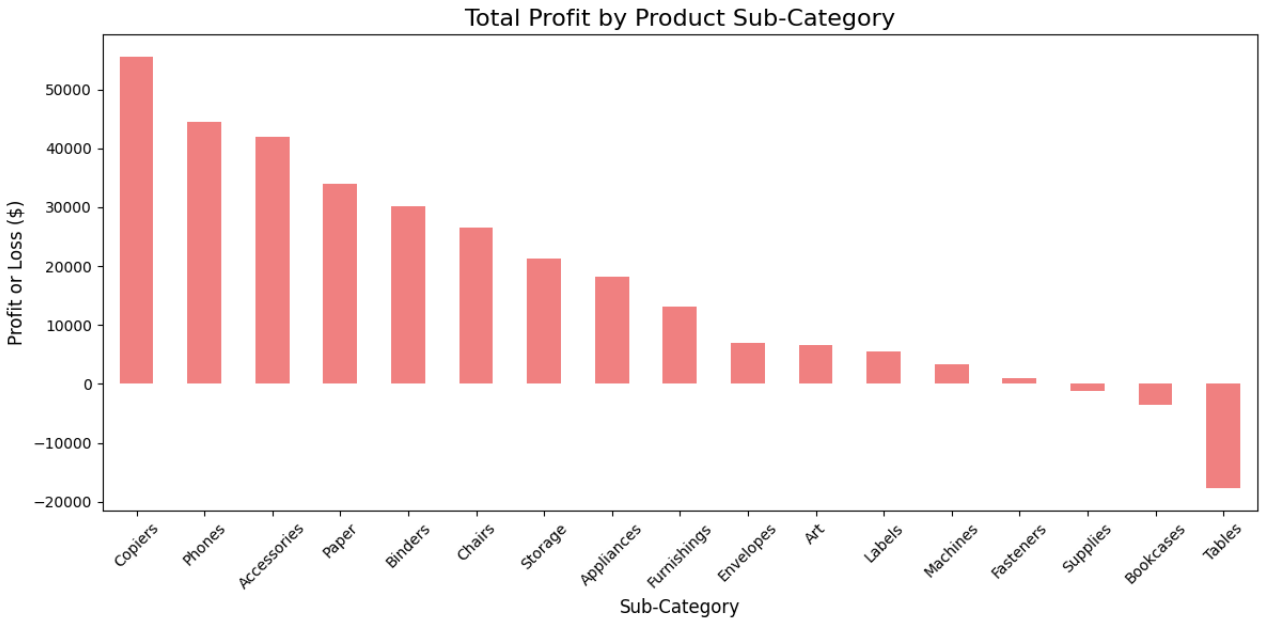


Profit by Product Sub-Category

The bar chart of sub-categories profitability shows that "Copiers" was the most profitable sub-category contributing over $ 55,000 whereas "Tables" was the most unprofitable sub-category (~$ 17,000). From this it indicates that product management and pricing strategies should target the loss-making products for better returns.

Notable insights:

- The most profitable sub-categories are Technology related ones, for example Copiers, Accessories.
- Furniture sub-categories (e.g., Tables, Bookcases) have a mixed bag of results with a lot of losses.



Sales vs. Quantity by Region

A scatter plot was used to compare the sales and quantity for the product across regions. Thus, the West region revealed higher sales for the same number of products than others which might mean higher avg order values. On the other hand, the South region contained more low-sale, low-quantity points, indicating poor performance or different consumers' pattern.

Regional Highlights:

- West: The highest number of sales with moderate number of products.
- South: Least number of products sold, implying that there is room for enhancing the approaches being used.

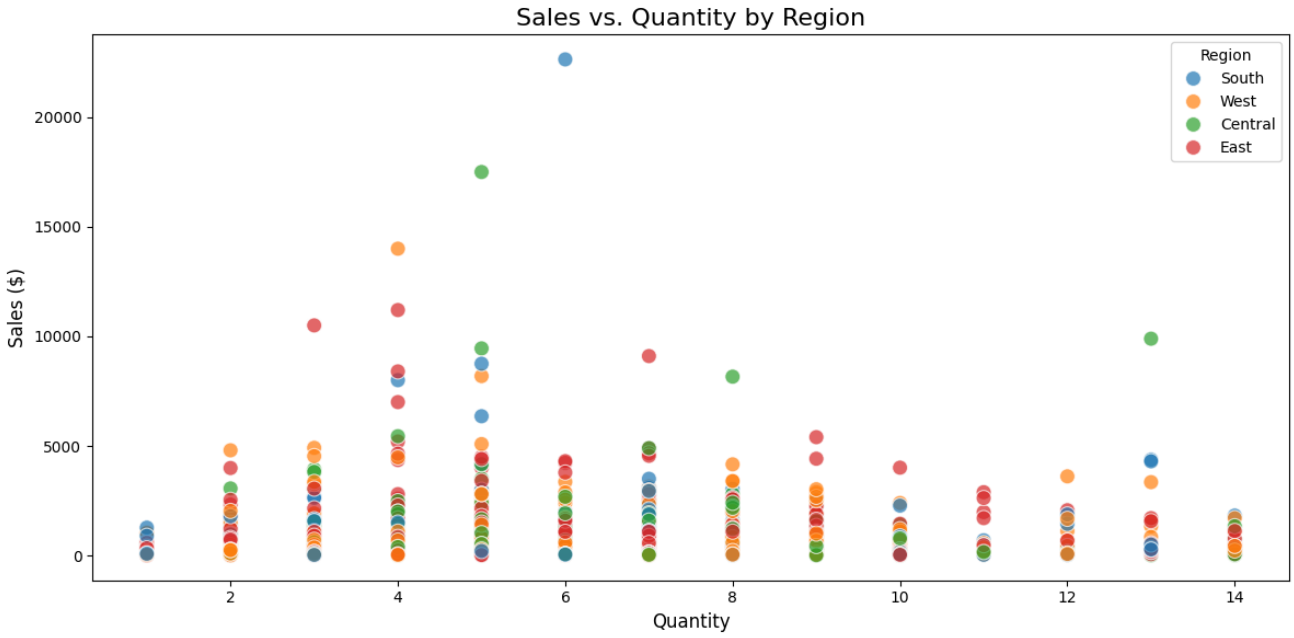


Table Summary

The following table outlines sales and profit ratios by geographic region.

Region	Total Sales ($)	Total Profit ($)	Average Discount (%)
Central	501239.8908	39706.3625	0.240353
East	678781.2400	91522.7800	0.145365
South	391721.9050	46749.4303	0.147253
West	725457.8245	108418.4489	0.109335

This shows that the West region has the highest contribution towards profits while exposing the potential for growth in the South region.

Conclusion

From this dataset, it is easy to see important aspects of sales, profitability, and per region performance. Whereas, technological products reflect high profitability, furniture items need cost and discount rejigging. Seasonal trends clearly show that it is effective to run holiday promotion campaigns. These findings can be useful for decision making to improve companies' profitability and to address their customers.