PROJECT TITLE

Sales Data Dashboard

ORGANIZATION/ DEPARTMENT NAME & ADDRESS

Wayspire Ed-Tech PVT.LTD

SUBMITTED BY

Ms. Janhavi Sachin Khare

M.Sc. (Applied Statistics)

PRN:24060641047

SYMBIOSIS STATISTICAL INSTITUTE

॥वसुधैव कुटुम्बकम्॥

ACADEMIC YEAR 2024 - 26

Under the guidance of

Mr. Suryansh Rana

Designation: Team Manager

Email Id: contact@wayspire.in

Mobile: 9210061874

Contents

1. Executive Summary

This project aimed to build a dynamic and interactive sales dashboard using a combination of Python, SQL, and Tableau. The primary goal was to clean, analyze, and visualize sales data to derive meaningful business insights. The dataset used was obtained from Kaggle's Superstore Sales data. Using Python, the data was cleaned and analyzed; SQL was used to extract regional and time-based summaries, and Tableau was used to design an interactive dashboard that visualizes trends, KPIs, and regional performance. This project enhances business decision-making by providing clarity on sales drivers, profit trends, and customer behaviour.

2. Study Background

In the digital era, businesses collect vast amounts of transactional and operational data every day. However, data in its raw form often lacks immediate meaning. Extracting useful insights requires appropriate tools and techniques that can organize, clean, analyze, and visualize the data effectively.

This project focuses on a **retail sales dataset** obtained from Kaggle, simulating real-world business data for a Superstore. The dataset contains detailed records of customer orders including product categories, regions, discounts, shipping details, profits, and sales amounts.

Analysing such data not only helps to understand overall performance but also allows businesses to:

- Detect seasonal sales patterns
- Understand regional performance differences
- Evaluate the impact of discounting
- Identify the most and least profitable products

3. Literature Review

While specific academic literature is limited for dashboard projects, numerous resources were consulted from:

- NumPy and Pandas official documentation
- Plotly and Seaborn visualization tutorials
- Tableau's official training modules and blogs

4. Aims & Objectives

The project aims to design an end-to-end dashboard solution for sales data analysis, combining data science and visualization skills. The specific objectives include:

- Data Cleaning: Convert raw Kaggle dataset into a clean, analysis-ready format using Python. This includes handling date formats, generating new variables (e.g., monthly indicators), and removing or adjusting outliers.
- Exploratory Data Analysis: Use statistical summaries and visualizations to explore patterns in sales and profit.
- SQL Integration: Practice writing SQL queries for summarizing and filtering data across dimensions like time and region.
- Interactive Visualizations: Build advanced plots like line charts, bubble charts, and bar graphs using Seaborn and Plotly.
- Dashboard Design: Develop a Tableau dashboard with filters, KPIs, trendlines, and interactive visuals to help users explore data independently.
- Insight Generation: Translate visual and statistical findings into meaningful insights that can drive data-driven business decisions.

5. Methodology

The project followed a step-by-step analytical approach to extract insights from the sales dataset:

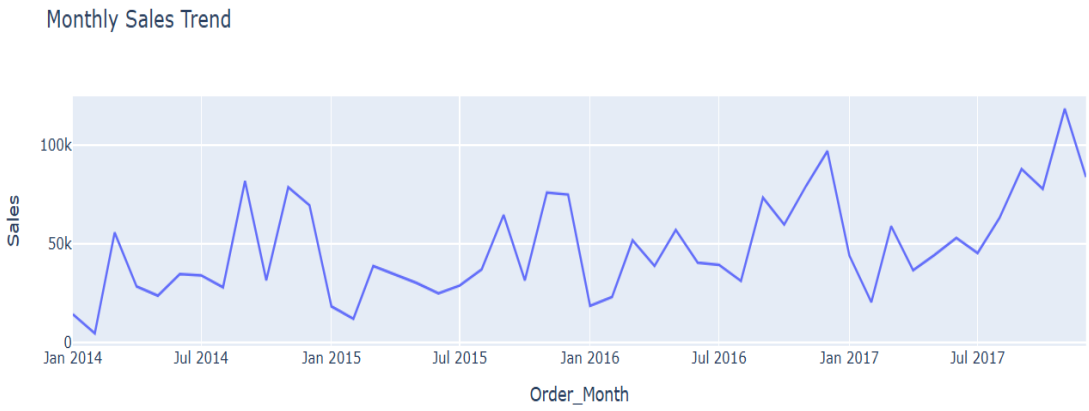
- The raw dataset was first cleaned and prepared for analysis. This included formatting date fields, creating new time-based variables, and checking for missing or inconsistent entries.
- A series of statistical summaries and visual explorations were conducted to understand sales patterns, profit distribution, and regional or category-level differences.
- Structured queries were used to extract key metrics like regional sales totals and monthly trends, helping in segmentation and performance comparison.
- Visualizations were created to represent trends over time, category-level profitability, and regional comparisons. These visuals supported quick understanding and helped identify areas needing attention.
- A final interactive dashboard was developed to bring all insights together in a dynamic and user-friendly format, enabling business users to filter and explore the data as needed.

6. Results

➢ Descriptive Statistical Analysis

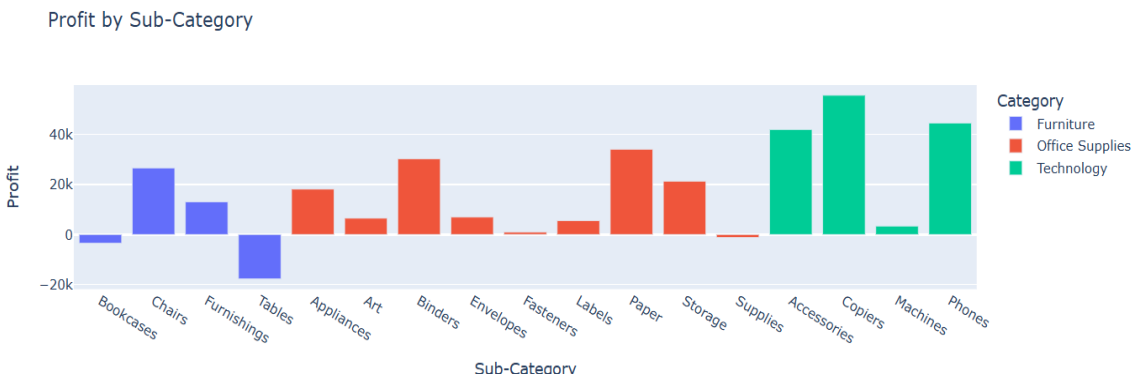
- The mean sales value of $229.86 indicates moderate spending per transaction, but the maximum value ($22,638.48) reveals a few very large purchases that skew the distribution.
- Profit values vary widely, with a mean of $28.66 and a standard deviation of $234.26, showing large inconsistency in profitability.
- A key insight is that some transactions incurred heavy losses (minimum profit = - $ 6,599.98), while others yielded high profit (up to $ 8,399.98).
- The interquartile ranges (IQR) for both sales and profit are relatively small compared to their respective maxima, again pointing to presence of outliers.
- This summary highlights the need for better pricing or discount strategies to avoid large negative profits.

➢ Monthly Sales Trend



The Monthly Sales Trend analysis highlights a cyclical sales pattern, with noticeable surges during the last quarter of each year. This suggests strong seasonality in consumer behaviour, potentially aligned with holidays or corporate procurement cycles. The trend shows positive growth from mid-2016 to 2017, with the highest peak recorded in Q4 2017, indicating improved business performance or marketing efforts during that period.

- Profit by Sub-Category



Profitability analysis across sub-categories reveals that Technology products, particularly Copiers and Phones, generate the highest profits. Office Supplies like Paper and Binders also contribute significantly to profitability. Conversely, certain Furniture items such as Tables and Bookcases exhibit negative or near-zero profit margins, suggesting inefficiencies or poor market performance in those product lines. This insight can guide inventory optimization and targeted sales strategies.

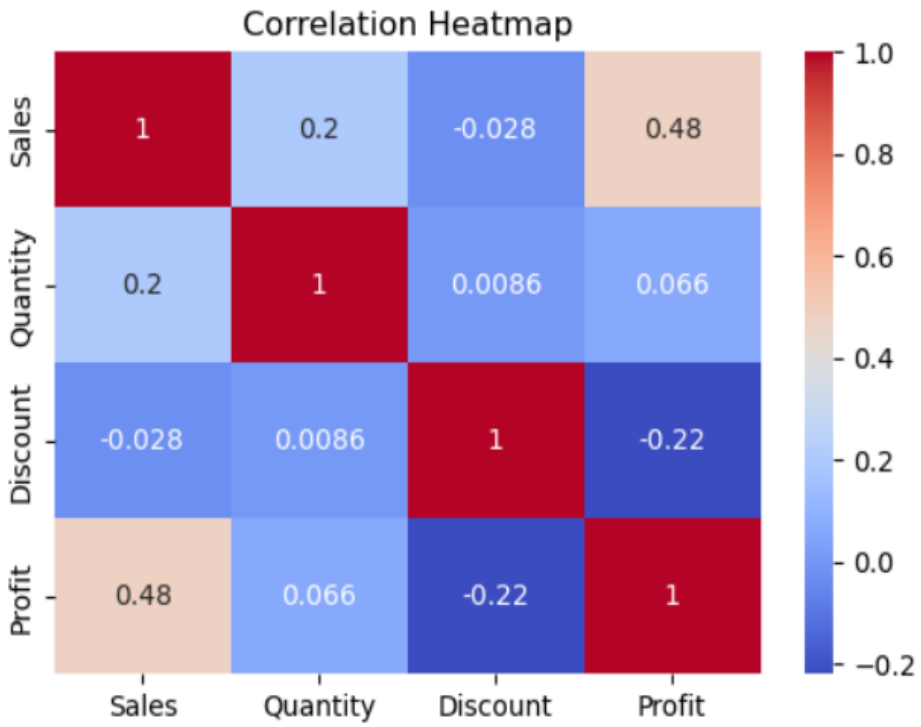
- Sales Vs Profit by State

A comparison of Sales vs. Profit by State reveals that while California and New York are both top performers in terms of revenue and profitability, states like Texas and Pennsylvania demonstrate a mismatch—high sales but low profits. This indicates potential issues in operational cost structures or discounting strategies in those regions. States with lower sales and profits represent untapped or underperforming markets that could be explored through targeted campaigns or localized offerings.

- Total Sales by Region

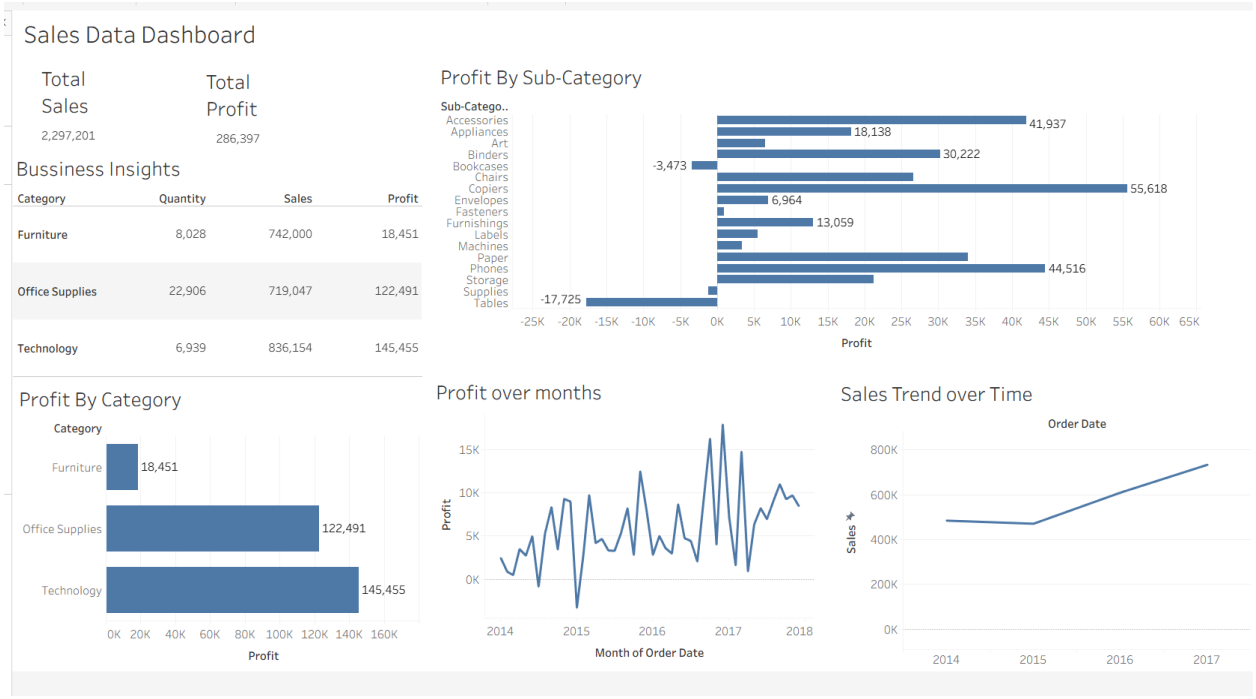
The sales dataset was analysed to determine performance across the four defined regions: West, East, Central, and South. The West region emerged as the top performer, with total sales amounting to over $7.25 lakhs, followed by the East with nearly $6.79 lakhs. These two regions dominate the overall revenue contribution. The Central region, although significantly behind the East and West, maintains a respectable sales figure. The South region, however, trails all others and may benefit from focused sales strategies or marketing interventions. These insights can be instrumental in regional planning, resource allocation, and identifying potential areas for expansion.

- Correlation Heat Map



The correlation heatmap shows relationships between numerical variables. Sales and Profit have a moderate positive correlation (0.48), meaning higher sales generally lead to more profit. However, Discount and Profit show a negative correlation (–0.22), indicating that high discounts reduce profit. Interestingly, Discount has almost no correlation with Sales, suggesting discounts do not significantly boost sales. Other variables like Quantity show weak correlations, meaning their impact on profit or sales is limited.

➢ Tableau Dashboard



- Total sales crossed $2.29 million, with total profit of $286,397, indicating overall profitability.
- Technology category leads in profit ($145,455), while Furniture lags behind with just $18,451.
- At the sub-category level, Copiers and Phones contribute the most to profit, while Tables show a significant loss (–$17,725).
- Monthly profit trend shows sharp fluctuations, with multiple spikes in late 2016 and 2017, suggesting seasonal buying behavior.
- Sales trend over time reveals consistent growth, especially after 2015, showing business scaling positively over years.

7. Discussion and Conclusion

The project successfully demonstrates how sales data, when analyzed properly, can reveal actionable business insights. The descriptive analysis shows wide variability in both sales and profit, with clear seasonal peaks during the year-end holiday period. The SQL queries reveal that the West and East regions are top performers, while the South remains underutilized.

The Tableau dashboard enables users to interact with filters such as region, product category, or time to explore sales trends in depth. It enhances the usability of the analysis by converting complex data into an accessible visual format. Stakeholders can quickly identify where sales are rising, which products generate the most profit, or where interventions are needed.

The correlation analysis brings attention to one key issue: while discounts are commonly used to drive sales, in this case, they appear to be negatively impacting **profit**. This suggests that discount strategies need to be reassessed. Meanwhile, high-profit products such as Copiers and Phones should receive more focus.

In conclusion, this project illustrates how a well-structured dashboard, built using data science and visualization tools, can enhance business intelligence and guide better decision-making. The approach used here is scalable and can be applied to real-time retail analytics in practical scenarios.

8. Recommendation

- Focus on high-profit categories like Copiers and Phones to improve overall profitability.
- Review discount strategies, as higher discounts are linked with lower profits. Apply them more selectively.
- Improve sales in the South region, which shows the lowest performance. This may need better outreach or logistics.
- Use the Tableau dashboard regularly to monitor trends, filter data, and make quick decisions.
- Plan for data refresh and automation to keep the dashboard updated and scalable for future use.

9. Acknowledgement

I would like to express my gratitude to my mentor and the institute for providing guidance and resources to complete this project.

10. References

- https://www.kaggle.com/datasets/vivek468/superstore-dataset-final
- https://pandas.pydata.org/
- https://plotly.com/
- https://www.tableau.com/learn/training

11. Annexure

- Python notebook: https://colab.research.google.com/drive/1ynh-F1VboxfMqyPCBYRLc-QiSpCVVGdy#scrollTo=87009377-80ca-4050-8b70-2e991ba271f0
- Dataset: https://drive.google.com/file/d/15lt-4YFjUvwqpf4-2fBrEBM0TOQX4HjS/view?usp=sharing