Retail Analytics through Visualization: Exploring the Superstore Dataset

[1]Mohammad Afnaan Anwar, [2]Fouzan Taha Hussain, [3]Prof. P. Lavanya

[1,2]Student, Department of Artificial Intelligence and Data Science, Methodist College of Engineering and Technology, Hyderabad, India

[3]Professor, Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Hyderabad, India

Abstract - **This project explores sales performance and operational trends within a retail business using the Superstore dataset, which includes detailed records of customer orders, sales, profits, and shipping information. Through comprehensive data visualization techniques, we analyze key business metrics across multiple dimensions such as region, product category, sub-category, customer segment, and shipping mode. The primary objective is to identify patterns and insights that can drive better business decisions, such as recognizing high-performing products, understanding regional sales differences, and evaluating profit margins. Tools like Tableau were employed to create interactive dashboards and static visualizations. The analysis reveals actionable insights such as the imbalance between high sales and low profitability in certain categories, and the impact of shipping choices on customer satisfaction. This project demonstrates how effective data visualization can simplify complex datasets and support strategic planning in retail environments.**

Keywords: Data visualization, Market distribution, Charts and graphs.

I. INTRODUCTION

In today's competitive retail landscape, making informed decisions based on data is crucial for business growth. The Superstore dataset provides a rich collection of sales records from a fictional retail company, including information on orders, products, customers, and shipping across various U.S. regions. This project aims to explore and analyze the dataset using data visualization techniques to uncover meaningful insights into sales trends, profitability, and operational performance. By leveraging tools such as Python and Tableau, we transform raw data into clear visual representations, helping identify top-performing segments, unprofitable areas, and strategic opportunities for improvement.

1.1 Literature Survey

The key words of this study are data visualization, RFM model and customer relationship management. Taking Wal-Mart as a research case, analyze the number of goods and stores in Wal-Mart supermarkets in four years (2011-2014) through data and visualize the results. Using visualize tool to analyze sales and commodities and present the results in a chart. It is particularly important to retain customers and improve customer retention rates in an environment of increasingly fierce competition among supermarket retailers. First, the supermarket retail data is visualized to make the data set more intuitive and easy to understand. Then the supermarket can choose the RFM (Regency, Frequency, Monetary) analysis method to distinguish customer value, so as to provide marketing services for different customers. The application of RFM first conducts basic evaluation through data visualization methods, and has a general grasp of retail data, then conducts RFM modeling to obtain the customer's RFM score, finally differentiating customer value.[1]

This paper explores the importance of business intelligence with a focus on leveraging sales data visualization to gather strategic insights. Beginning with an examination of the importance of data analysis and visualization of sales data, the study emphasizes its crucial role in identifying trends, optimizing performance, and making strategic decisions. Furthermore, the paper offers valuable insights into effective visualizations using the US regional sales dataset which encompasses sales, transactions and customer data which serves as a rich resource for analyzing sales patterns, product popularity, and channel performance. Practical ideas for visualizing this dataset are presented through this study, catering to a broad audience seeking actionable intelligence from their sales data. Therefore, this paper synthesizes key findings, addresses limitations, and emphasizes the complex nature of sales data and how visualization can empower businesses to harness the full potential of their sales data, fostering data-driven decision-making in an ever-evolving marketplace.[2]

Business analytics refers to the skills, technologies, practices for continuous and iterative performance of any business. This will not only help to gain insight but also help to plan the activities to drive a business. Based on

understanding the business process, the analytics is focuses on developing new insights to understand the business better based on data available and use of statistical methods. In contrast, Business Intelligence traditionally focuses on using a consistent set of metrics to both measure past performance and guide business planning, which is also based on data and statistical methods. BI includes a diversity of implements, process, methodologies and application that permits the organizations to gather data from internal and external sources.[3]

The amount of unprocessed data available every day is growing. This massive amount of data needs to be effectively assessed to give results that are extremely useful. In the present day, it is crucial for inventory management and demand forecasting to collect sales data for commodities or things, together with all their numerous dependent or independent parts. In a Big Mart Company, the use of sales forecasting is to estimate numerous goods that are readily available and supplied at multiple retailers in different towns. As the number of products and outlets increased drastically, it became increasingly difficult to forecast them manually.[4]

These data are impossible to store or process by a single machine or in a traditional way. So, the need to use distrusted storage and processing systems was an emergency, such as the Apache Hadoop system, which provides a fault-tolerant, dependable, horizontally scalable, and effective service. It is based on the Hadoop distributed file system (HDFS) and MapReduce.[5]

The data gathered is stored with the help of an excel sheet which is a fast and reliable source to maintain sales records. BI analysts can use a wide variety of business intelligence BI tools such as Tableau which gives them more flexibility in representing the results. Here sales records are vast information to be stored and it needs to be formatted in some order. Tableau, a visual analysis tool, helps to measure sales targets from the whole dataset easily and gives the results in the form of desktop view very fast. These visual views can be used further to achieve the best sales targets in the future.[6]

Business development in the world is increasing with the use of information technology which has an impact on local, national and international business activities, as well as generating big data. Big data that is generated needs to be processed, so that it can be used to analyse sales performance. Olist Store data processing is carried out by implementing the ETL process in the data warehouse using Pentaho and visualizing Business Intelligence on the smart dashboard using Tableau. The data warehouse design method used is a nine-step methodology.[7]

In the literature, all aspects of data visualization do not guarantee anomaly detection and practical forecasting recommendations. To fill this gap in the literature, this research aims at explaining data visualization characteristics (visibility, mass notification, information sharing, emergency management, and accessibility) combined with regression methods that allow organizations to detect anomalies and make forecasting. This research revisits and provides new insights into regression analysis because of the advent of the characteristics of data visualization. A database containing 9994 records was used to test the regression model to detect anomalies using data visualization. The regression model for data visualization focused on the relationships between sales and profits.[8]

The purpose of this study is to design a data analytics model for the home improvement store in order to solve these issues. The selected dataset that includes the sales information of the home improvement store will be evaluated and pre-processed before implementing it for analytics. The data analytics model is needed to be implemented in a home improvement store in order to mine the data especially in the current dataset. This analytic model allows the company to do descriptive and predictive analytics on the data.[9]

This paper discusses the integration of Tableau, a business intelligence tool for data visualization, into a first-year information literacy course at Hokuriku University as a practical approach to enhancing data science education. The university launched the "Data Science and AI Education Program" in 2022, which combines traditional information literacy topics with hands-on learning using Tableau, engaging students with real-world datasets from campus stores and cafeterias.[10]

1.2 Material & Method

The dataset titled "Superstore Dataset" provides a structured overview of a fictional retail operation, containing vital business metrics across categories such as sales, profit, shipping, and regional performance. This dataset includes both categorical dimensions—such as Region, Category, Sub-Category, and Segment—as well as quantitative measures like Sales, Profit, Quantity, and Discount. These fields were leveraged to create a comprehensive Tableau dashboard, designed to analyze retail performance from multiple business angles. The analysis was conducted using **Tableau Desktop**, where the dataset was imported directly and fields were assigned appropriate data types. The dashboard was developed to include four core visual components:

Sales Trend Over Time: A **line chart** displaying monthly or yearly sales trends, helping identify seasonal fluctuations and sales growth over time.

1. Total Sales and Profit Overview:

- **KPI cards** and **bar graphs** were used to present total sales and total profit, offering a quick snapshot of business health and financial performance.

2. Sales by Region (USA Map):

- A **filled map of the United States** illustrating region-wise sales contributions, enabling identification of high- and low-performing regions.

3. Sales by Product Category (Furniture, Technology, Office Supplies):

- A **bar chart or stacked chart** comparing sales and profit across the three main product categories, helping assess which segments contribute most to overall revenue.

Each visualization is equipped with **interactive filters** (such as year or region), allowing users to drill down into specific views for more granular insights. Tooltips and color gradients were applied to enhance data readability, while calculated fields were created to derive insights such as **profit ratio** and **category-wise contribution percentage**.

The use of Tableau enabled seamless integration of multiple visual components into a single interactive dashboard. The **Sales Trend** view revealed temporal spikes, particularly during year-end periods, indicating possible promotional or holiday effects. The **Region-wise map** showed that while the Western U.S. dominated sales, certain central states contributed less profit despite comparable sales volumes. Analysis of the **Category breakdown** exposed that although Technology drives the highest sales, Furniture sometimes results in lower profitability due to heavy discounting or shipping costs.

Through this structured and visual exploration, the dashboard transforms raw data into business-ready intelligence. It enables decision-makers to optimize marketing efforts, streamline logistics, and focus on product categories that offer the best return on investment.
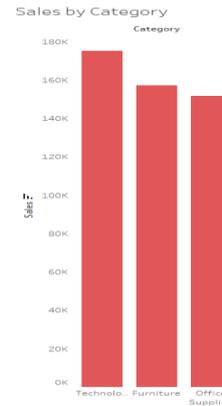
II. DATA VISUALIZATION

Figure 1: Sales by Category

This bar chart displays the total sales across three primary product categories: Technology, Furniture, and Office Supplies. Among these, Technology leads with the highest sales, exceeding $180,000, followed by Furniture and then Office Supplies, both of which have similar sales figures in the range of $150,000 to $160,000.
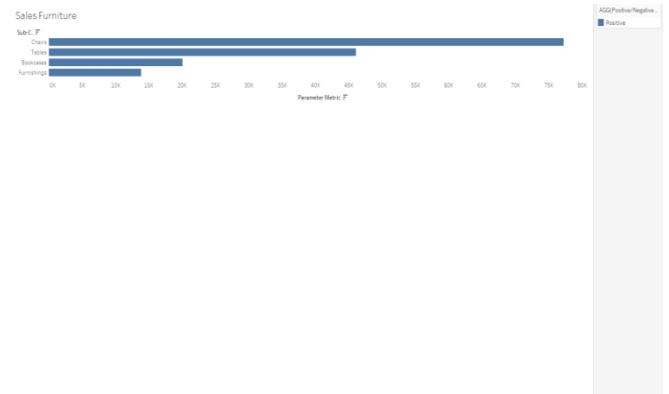

Figure 2: Sales Furniture

This horizontal bar chart displays sales performance within the Furniture category, segmented by its sub-categories: Chairs, Tables, Bookcases, and Furnishings. Among these, Chairs generate the highest sales, nearing $80,000, followed by Tables with approximately $45,000. Bookcases and Furnishings show notably lower figures, indicating less customer demand or lower transaction volume. The uniform color tone represents positive performance across all sub-categories. This visualization highlights the dominance of Chairs and Tables within Furniture sales, suggesting that these items are key revenue drivers and should be prioritized in marketing and inventory strategies.
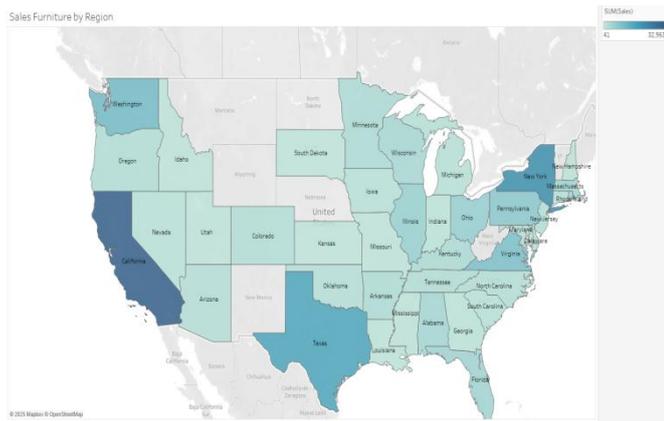
Figure 3: Sales furniture by region

This filled map visualization illustrates furniture sales across various U.S. states, with darker shades representing higher sales figures. Notably, California, New York, and Texas emerge as the top-performing states in furniture sales, each shaded in deeper blue tones. In contrast, states like South Dakota, Idaho, and New Mexico show significantly lighter shades, indicating lower sales volume. This geographic distribution highlights the concentration of furniture demand in highly populated or economically active states. The visual insight helps identify strong markets for expansion and reveals underperforming regions where targeted marketing or logistical improvements could be beneficial.
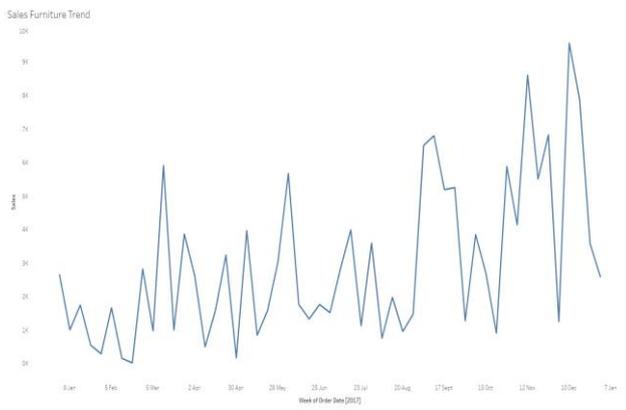

Figure 4: Sales Furniture Trend

This line chart represents the weekly sales trend for furniture throughout the year 2017. The sales pattern is highly fluctuating, with noticeable peaks and troughs. There is a significant rise in sales during the later months, especially in November and December, which may be attributed to seasonal shopping periods like Black Friday or holiday promotions. Earlier months show relatively lower and inconsistent sales, with occasional spikes in March and May. This visualization effectively highlights the seasonal nature of furniture sales, providing valuable insight for inventory planning, marketing campaigns, and demand forecasting.
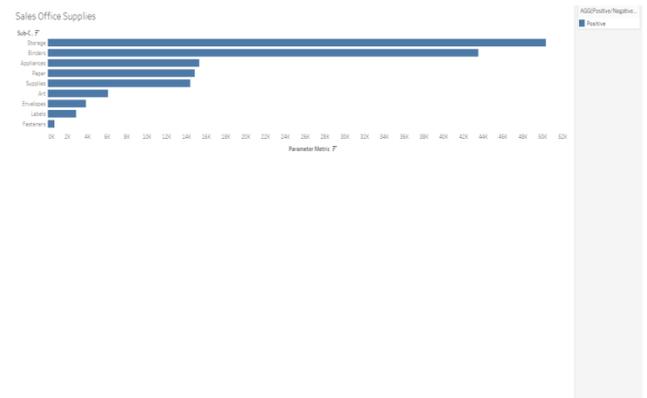

Figure 5: Sales Office Supplies

This horizontal bar chart visualizes sales across sub-categories within Office Supplies, revealing that Storage and Binders are the top-performing segments, each exceeding $40,000 in sales. These are followed by Appliances, Paper, and Supplies, which contribute moderately to overall revenue. Sub-categories like Art, Envelopes, Labels, and Fasteners show significantly lower sales figures, indicating limited customer demand or lower frequency of purchases. The positive color coding reflects that all sub-categories are performing profitably. This chart provides a clear view of sales concentration within Office Supplies, helping businesses prioritize high-demand items for promotion while re-evaluating the role of underperforming segments.
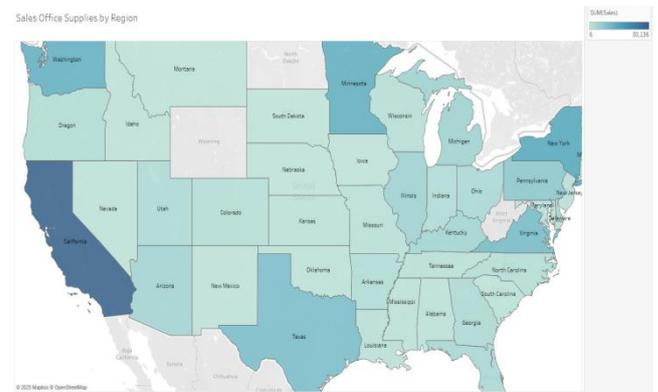

Figure 6: Sales Office Supplies by Region

This choropleth map, titled "Sales Office Supplies by Region," illustrates the total sales of office supplies across U.S. states, using a color gradient from light blue (6) to dark blue (30.136) to represent sales volume. California stands out with the highest sales, marked in the darkest shade, followed by states like New York, Texas, Washington, and Pennsylvania in darker hues, while states such as Wyoming, North Dakota, and South Dakota show the lowest sales in the lightest shade. Sourced from Mapbox and OpenStreetMap, the map is dated 2025 and provides a clear visual of regional sales disparities.
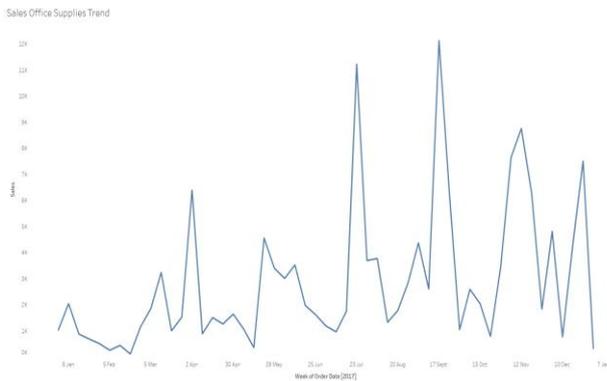
Figure 7: Sales Office Supplies Trend

This line graph, titled "Sales Office Supplies Trend," tracks weekly sales of office supplies in 2017, with the x-axis showing dates from January 8 to January 7 of the following year, and the y-axis representing sales in thousands (0K to 12K). The sales trend is highly volatile, starting at around 1K in early January, peaking sharply at 11K around late June, and experiencing multiple fluctuations throughout the year, with notable dips below 2K in April and December, and other peaks reaching 8K in March and September, indicating inconsistent sales performance over the period.
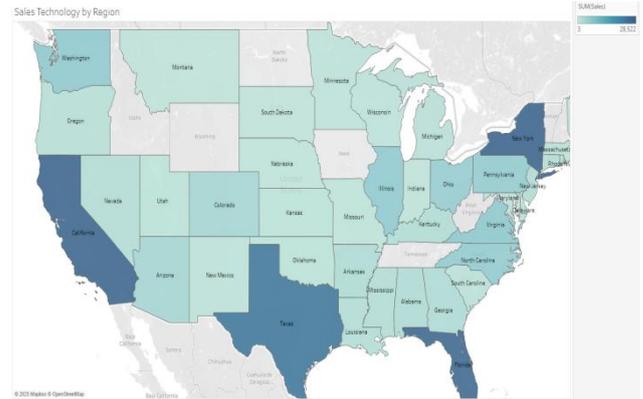

Figure 8: Sales Technology

This horizontal bar chart, titled "Sales Technology," displays sales data for three sub-categories of technology products—Phones, Machines, and Copiers—measured in thousands (0K to 80K) along the x-axis, with the y-axis listing the sub-categories. Phones lead with the highest sales, reaching around 75K, followed by Machines at approximately 50K, and Copiers with the lowest sales at about 15K, all represented in a consistent blue shade labeled as "Positive" in the legend, indicating a focus on positive sales metrics for these technology products.

Figure 9: Sales Technology by Region

This choropleth map, titled "Sales Technology by Region," visualizes technology sales across U.S. states, with a color gradient from light blue (3) to dark blue (28.526) indicating sales volume. California, New York, Texas, and Florida show the highest sales in the darkest shade, while states like Wyoming, North Dakota, and West Virginia have the lowest sales in the lightest shade, with other states like Washington and Pennsylvania falling in between, reflecting varying levels of technology sales across regions, as sourced from Mapbox and OpenStreetMap, dated 2025.
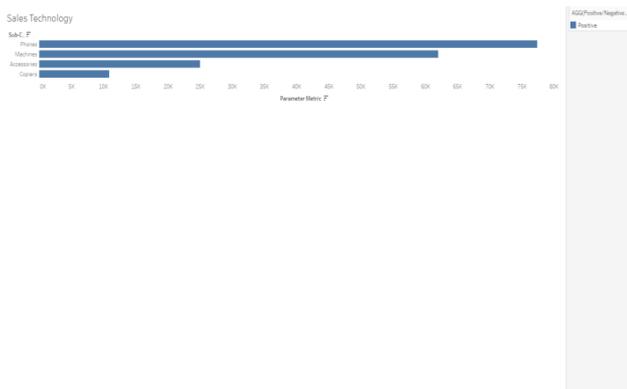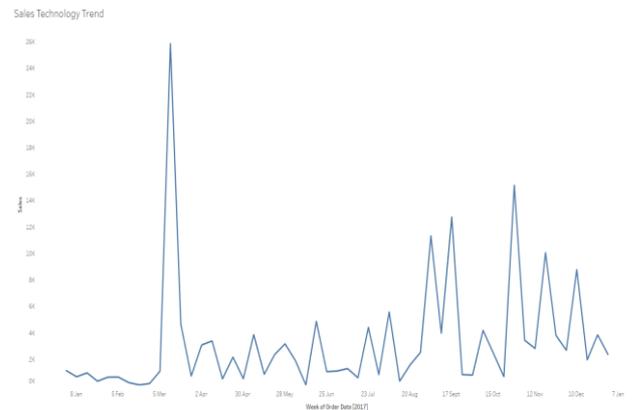

Figure 10: Sales Technology Trend

This line graph, titled "Sales Technology Trend," illustrates weekly technology sales throughout 2017, with the x-axis showing dates from January 8 to January 7 of the following year, and the y-axis representing sales in thousands (0K to 26K). The sales trend is highly erratic, starting below 2K in January, spiking dramatically to 25K around early March, then dropping sharply and fluctuating between 2K and 4K through April to July, with another peak at 14K in late August, followed by smaller spikes around 10K in September and October, before declining to around 2K by year-end, reflecting significant volatility in technology sales over the year.

Figure 11: Sales by category

The "Superstore Dashboard" for 2020 displays sales data across three categories—Furniture, Office Supplies, and Technology—totaling $215,387, $246,097, and $271,781 respectively, with area charts showing sales trends peaking around June and December. Furniture sales are led by Chairs at $95,554, followed by Tables at $60,894, while Office Supplies are dominated by Binders at $72,788 and Storage at $42,927, and Technology sales are highest for Phones at $105,341 and Copiers at $62,889. Geographically, three choropleth maps highlight California as the top sales region for all categories, with $40,674 in Furniture, $55,413 in Office Supplies, and $50,301 in Technology, while Nebraska has the lowest Furniture sales at $16, and South Carolina has the lowest Technology sales at $79, showing significant regional sales disparities.

Figure 13: Profit By Category

The "Superstore Dashboard" for 2020, set to the Profit metric, shows total profits for Furniture, Office Supplies, and Technology at $3,018, $39,737, and $50,684 respectively, with area charts indicating profit trends peaking around June and December, though Furniture dips into negative profits mid-year. Furniture profits are led by Chairs at $7,644, while Tables and Bookcases show losses at ($141) and ($584); Office Supplies are driven by Paper at $12,041 and Appliances at $7,403, with Fasteners at a loss of ($955); and Technology profits are highest for Copiers at $25,082, followed by Accessories at $15,672, with Machines at a loss

of ($2,869). Geographically, California leads Furniture profits at $16,641 but shows a loss of ($6,831) in Office Supplies, while New York leads Technology profits at $13,599, and North Carolina has the lowest Technology profit at ($3,834), highlighting varied regional performance.

III. RESULTS AND DISCUSSIONS

The data visualization of the Superstore dataset revealed several key patterns and insights across product categories, regions, and time. By leveraging Tableau's interactive capabilities, we were able to analyze performance in terms of sales, trends, and category-specific contributions.

The **Sales by Category** analysis indicated that **Technology** is the highest revenue-generating segment, followed by **Furniture** and **Office Supplies**. This suggests a strong customer preference for higher-value technological products. However, further breakdown into sub-categories revealed internal variations. Within **Furniture**, **Chairs** and **Tables** were the dominant contributors, while **Furnishings** lagged behind. Similarly, in **Office Supplies**, **Storage** and **Binders** drove most of the sales, whereas categories like **Labels** and **Fasteners** recorded minimal activity. These findings highlight the importance of drilling down beyond top-level categories to make informed stocking and marketing decisions.

The **geographical analysis**, presented through a filled map of the U.S., identified **California, New York, and Texas** as the top-performing states in terms of furniture sales. This regional concentration suggests that population density and economic activity significantly influence buying behavior. Conversely, states such as **South Dakota and New Mexico** showed low sales volumes, signaling either low market penetration or less demand, offering opportunities for localized campaigns.
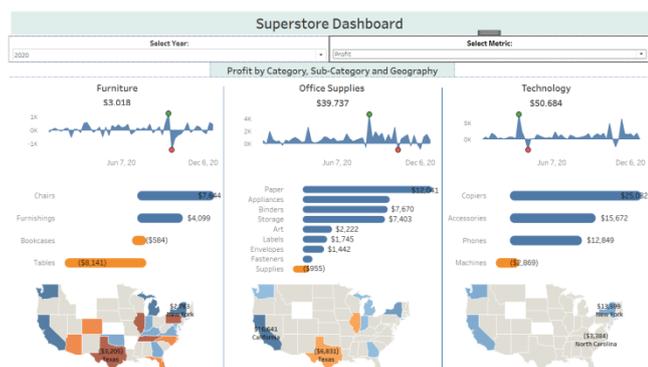
The **sales trend line chart** for the Furniture category uncovered clear **seasonal fluctuations**, with noticeable spikes during **November and December**, likely linked to holiday and end-of-year sales events. These trends can inform future forecasting models and help in optimizing inventory and promotional planning during peak sales periods.

Collectively, these visualizations revealed that while some categories perform consistently well, others show signs of inefficiency or stagnation. Additionally, regional and temporal factors significantly affect sales performance. The dashboard's interactive nature allows stakeholders to filter data by time, region, or category, offering tailored insights for different departments such as sales, marketing, and logistics.

In conclusion, the Tableau-based visual analytics approach provided not only clarity but also actionability. It enabled deeper exploration of performance drivers and revealed underutilized opportunities in product and regional strategies. These findings support data-driven decision-making that can enhance profitability, customer targeting, and resource allocation in a competitive retail environment.

IV. CONCLUSION

This study effectively utilized Tableau to visualize and analyze key performance indicators from the Superstore dataset, providing actionable insights into sales trends, product category performance, and regional distribution. Through interactive dashboards and dynamic charts, we uncovered that **Technology leads in overall sales**, while **Furniture and Office Supplies** contribute moderately, each with internal variations across sub-categories.

Geospatial analysis revealed that **states like California, New York, and Texas** are major revenue drivers, whereas several central and smaller states represent underexplored markets. **Seasonal trends**, particularly spikes in sales during November and December, emphasize the importance of timing in promotional strategies. Sub-category breakdowns, such as the dominance of **Chairs and Tables in Furniture** or **Storage and Binders in Office Supplies**, provided granular insights that can directly inform inventory management and marketing focus.

Overall, the integration of Tableau's visualization capabilities with structured retail data turned raw figures into clear, strategic insights. This approach enhances decision-making across various business functions and underscores the critical role of data visualization in modern retail analytics. Future work can extend this analysis to include predictive modeling and customer segmentation, further strengthening the value of data-driven strategy in retail operations.

REFERENCES

[1] Wang, Shuming, and Phisanu Chiawkhun. "Using data visualization for supermarket retail analysis." *Data Science and Engineering (DSE) Record* 3.1 (2022): 93-111.

[2] Upadhye, Akshata. "ENHANCING BUSINESS STRATEGY WITH SALES DATA VISUALIZATION." *Journal ID* 2811: 6201.

[3] Arora, Monika. "An Application of Dash Board: A Case of Superstore Sales." *Monika Arora, Nikhil Pal Singh, Pranav Chhabra (2016),"An Application of Dash board: A Case of Superstore sales", National Seminar on "TECHNO TRYST*. 2016.

[4] Arista, Artika, Theresiawati Theresiawati, and Henki Bayu Seta. "Big Mart Sales Data Visualization and Correlation." *JOIV: International Journal on Informatics Visualization* 8.2 (2024): 576-582.

[5] Mohamed, Rabab Hussien, and Kamel Hussien Rahouma. "Research for Big Data Storage and Analysis Based on Artificial Intelligence." *Journal of Advanced Engineering Trends* (2025).

[6] Patil, Manisha M. "A Case Study-Visual Analysis of Sales Records Using Tableau." *International Journal of Advanced Research in Science, Communication and Technology* (2021).

[7] Anggrainy, T. Difa, and A. Rusiana Sari. "Implementation of extract, transform, load on data warehouse and business intelligence using pentaho and tableau to analyse sales performance of offlist store." *Science* 7.2 (2022): 368-374.

[8] Lawson-Body, Assion, et al. "Anomaly identification through data visualization: regression analysis revisited." *Issues in Information Systems* 24.4 (2023).

[9] Yeng, Lim Jia, Mohd Norshahriel Abd Rani, and Nabilah Filzah Mohd Radzuan. "Data Analytics Model for Home Improvement Store." *Advances on Smart and Soft Computing: Proceedings of ICACIn 2021*. Springer Singapore, 2022.

[10] Tajiri, Shintaro, et al. "Integrating Tableau into a First-Year Information Literacy Course: A Practical Approach to Enhancing Data Science Education." *IIAI Letters on Institutional Research* 4 (2024).

****