BUSINESS INTELLIGENCE ON RETAIL DATA – A DECISION MAKING TOOL THROUGH DASHBOARD

A. Kannammal
Department of Decision and Computing Sciences,
Coimbatore Institute of Technology,
Coimbatore 641 014, India,
kannammal@cit.edu.in,
kannaphd@gmail.com

R. Gokulavasan
MSc (Integrated) Decision and Computing Sciences,
Coimbatore Institute of Technology,
Coimbatore 641 014, India,

M. Sarathram
MSc (Integrated) Decision and Computing Sciences,
Coimbatore Institute of Technology,
Coimbatore 641 014, India,

Abstract— **Transactions made in superstore data are very helpful for the development and can also be used to forecast future sales and to study the past about business challenges. The use of Business Intelligence tools can help analyze large amounts of data including quality analysis and company analysis. This project was designed using BI approach to analyze the data. The research focuses on reporting the data regarding the sales and profit acquired. This project also uses OLAP operations to describe data visualization with help of Bivariate Analysis such that it provides better benefits and competitive advantage. Business Intelligence is expected by company leaders to be able to understand the data that will have been processed in understanding visual forms and can easily absorb the information needed to make decisions for the company.**

Keywords: *Business Intelligence, Retail, Bivariate Analysis, OLAP, BI Tools, Dashboard.*

1. Introduction:

SuperStore is a retail business location in the United States. It comprises companies that operate by having large size spaces which store and supply large amounts of goods. It consists of an extensive store that sells a product line of grocery items and merchandise products such as Furniture, Office Supplies and Technology products to their customers who are mass Consumer, Corporate and Home Offices. Data contains sales, profit, geographical information of the store. These forms of data containing transaction results carried out over year to year, are very useful for the company development, and can be used for future forecasting. To support this, Superstore needs a dashboard so that it helps in making decisions and managing data so that they can add valuable advantages to support the existing business processes. Also, this system analyzes the data to identify weak areas and opportunities to boost business growth. The analysis also gives insights for the sales and profit of various products.

2. Problem Definition

With growing demands and cut-throat competition in the market, it would be better to understand which products, regions, categories and customer segments they should target or avoid. Every store be it online or offline needs evaluation and analysis to predict daily sales and know what goods customers want at a particular time and what the trend would be every day, month and year.

A superstore functioning across the City of United States tends to understand some certain features of his business and obtain vital information from its data, that would be helpful to plan or focus on what is ongoing sales, with a full understanding of sales trends and forecasting for the future all things being equal. This kind of analysis will help in proper stocking of needed goods, in order to reduce understocking and overstocking, which helps in supply chain management to process beneficiaries; also focus extremely on the selling of products by taking into consideration demand and supply.

3. Objectives

- The purpose of the project is to analyze sales and profit and to perform various superstore statistics, thus providing the perspective on the way business is trending. The data have continuously been collected covering over 48 states as from January 2014 -2018. This data will be helpful in reports and performance analysis. The reports will be based on Categories, Sub-Categories, Segments and Regions.

- This focus on analysis of [1], Sales Trend, Most and Least selling products by number of sales, Most preferred shopping mode, Profitable categories and sub-categories, yearly sales for various states, etc.

- With this analysis, the store can identify various aspects of the purchasing pattern and take respective measures. Also try to reduce the Expense by Supply Chain Management wise and increase the sales and profit towards geographical wise; also forecasted the Sales for further upcoming few years.

4. Literature Review

In existing systems [4] the basic analysis is done about the data and exploratory data analysis is done based on various approaches. Performing further analysis using the difference between order date and shipment date. Also, some customer level analyses are made such as first and last purchase of each customer, calculating retention of each customer, number of orders by each customer, average revenue from each customer which are based on company and management purposes. Some include basic EDA operations in python including the factors of

State, category, year, etc. with sales as target variable. Can be also known as univariate analysis on the data [3].

The proposed system consists of Exploratory Data Analysis upon two factors such as sales and profit and forecasting the sales for a further few time periods. This Bivariate Analysis is done based on building effective Dashboards using Kibana (ELK stack) where live data can be taken into account for evaluation & processing the data for better decision-making process. Some OLAP operations such as slicing and dicing are applied using bar charts and other plots. Also considering sales attributes since data contains time series factors in it, ARIMA (Autoregressive Integrated Moving Average) is used to forecast the generation of sales for a few upcoming periods of time. With this analysis, the store can identify various aspects of the purchasing pattern and take respective measures.

The outcome of the analysis is to give the store management a comprehensive, but easy to understand analysis using some key visualization tools for easy understanding and also forecast future sales. Also try to reduce the Expense by Supply Chain Management wise and increase the sales and profit towards geographical wise; also forecasted the Sales for further upcoming few years. In addition, the developed system provides dashboards in Kibana for Management side & Customer side which helps in in-direct marketing of products & helps in boosting up sales, which are discussed clearly in the below sections.

The dataset was obtained from kaggle.com [2] which provides information about the sales and profit from US supermarket from 2014 to 2017. Dimensions consisting of 10000 * 21.

5. Preprocessing:

Descriptive Statistics of the data has been done. NULL values and any other missing value are checked. Splitting of the order date to their individual entities such as day, month, year has been made. Ship Date has been dropped, since there was no relevance for the visualization. Final data consists of dimensions on basis of 10000 * 23 instances. Python & Excel are used for performing Exploratory Data Analysis on the Data.

6. Exploratory Data Analysis

Sweetviz: A package in python used for web-based visualization in which it gives the relation between the Dependent Variable (target variable) to each & every Independent Variables, including some of the descriptive statistics along with some corresponding plots such as graphs and charts. An Interactive EDA is built with help of this technique as shown in fig.1 below.
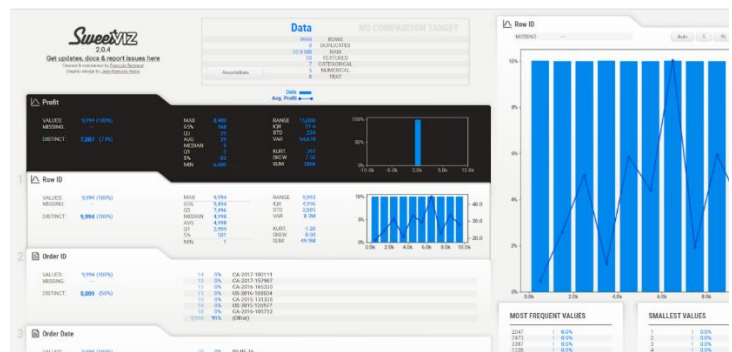

Fig. 1. Represents the EDA done using Sweetviz.

Also, Box Plot was used for visualizing and analyzing the distribution of continuous variables; Histograms are used for visualizing and analyzing the distribution of continuous variables.
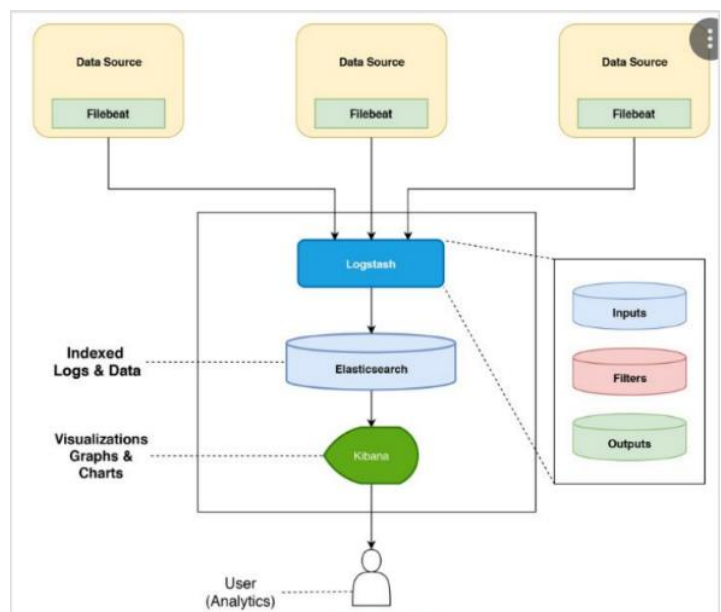
7. Process Flow:

Fig: 2. ELK stack Architecture

Figure: 2. Represents the architecture of the model of ELK stack with neat visual representation of how the architecture works, the layer composed in each stage of the process [10].
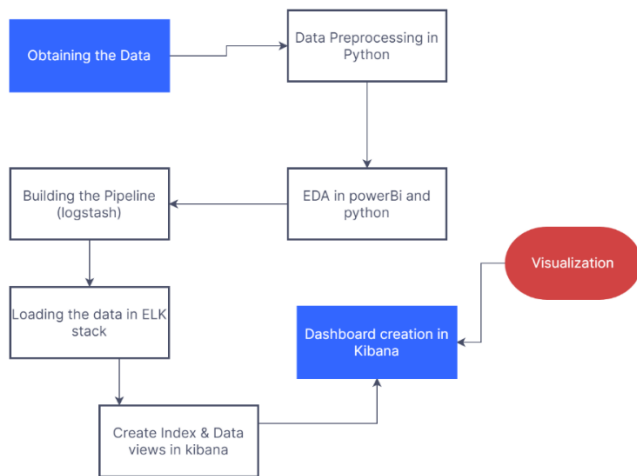
Fig: 3. Process Flow of proposed system

Figure: 3. Represents the flow of the proposed system where the process is made with help of above-mentioned tools.

8. Report and Inferences:

This proposed system consists of two dashboards, one for the company view & other for customers' view.

An image of the dashboard for the Company along with some constraints(controls) to Year, Month, State, City, Region which are linked to the whole dashboard, is shown below figure 4.
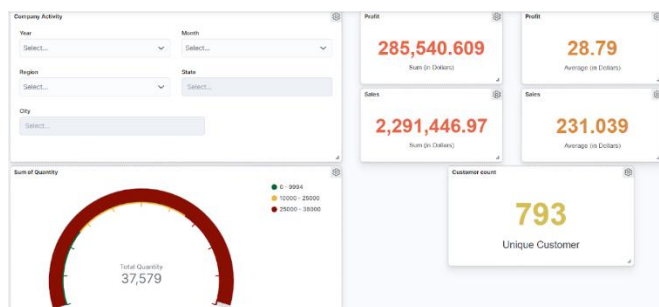

Fig: 4. Dashboard for the Company side

Figure 4. The Metric plot on the right is used to show the Overall Sum and Average of Sales, Profit according to given constraints. The Gauge plot on left shows the Total Quantity sold in each period respectively.
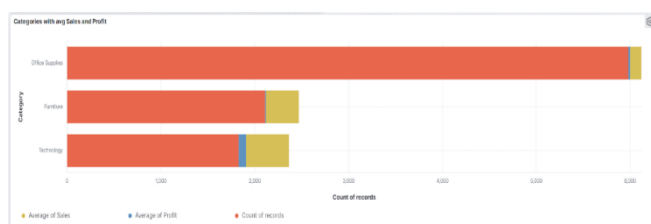

Fig: 5. Category with Average Sales and Profit

Figure 5 shows category versus average of both sales and profit, where Average of sales for office supplies is less than other categories so the store can reduce the stock of office supplies to prevent overstocking and increase the stocking of furniture to gain better sales.
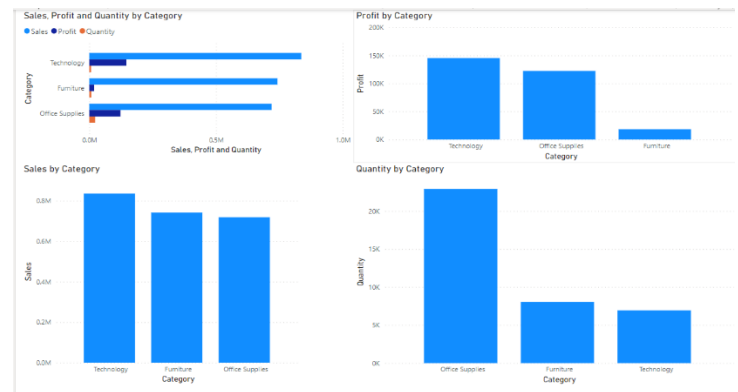

Fig: 6. Category vs Sum of sales

Figure 6 shows that all 3 categories make similar amounts of sales. Although Furniture makes similar sales as Technology, it is the least profitable also quantity sold are at a minimum. Office Supplies sell the most, in terms of quantity as it is a relatively cheap product.
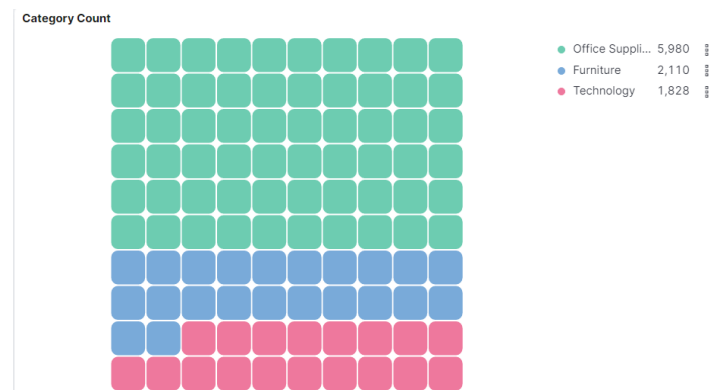

Fig: 7. Count of Category

Figure 7 is a Waffle Plot used for viewing the count of categories sold under given constraints by the user using controls above on Dashboard.

Fig: 8. Sub-category vs sales and profit

Figure 8 plots the sub-categories with sales and profit. Phones and Chairs are Top 2 best-selling sub-category. The most profits are produced by Copiers>Phones>Accessories> etc. For some products sales increase but profit decreases & vice versa, so the company has to focus on marketing strategy for better improvement in sales & profit. On the other end, the spectrum, machines, fasteners, supplies, tables make close to 0 margin to losses. These products may be dropped from the product catalog (or) bargain to a lower price from the supplier.
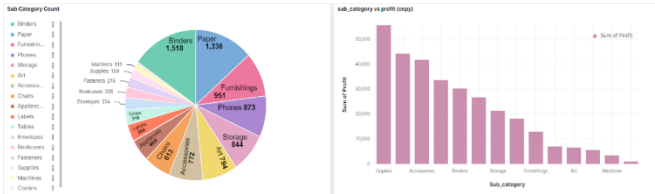

Fig: 9. Subcategories count vs profit

Figure 9 represents the count of subcategories with respect to their profit. Eg: Blinder's count is high but produces less profit compared to Copiers, where the stock count of copiers is Low; so, they can increase the count of copiers according to the suitable decision making. Art produces Less profit but count is high so the store can reduce the stock to prevent overstocking and help in supply chain management.
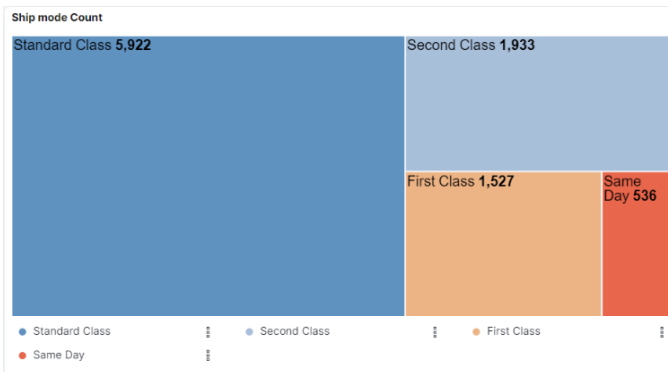

Fig: 10. Count of Shipping Mode

Figure 10 depicts the count of Ship mode taken by customers of the store, showing that most customers choose Standard class because it is Cost Effective than other modes.
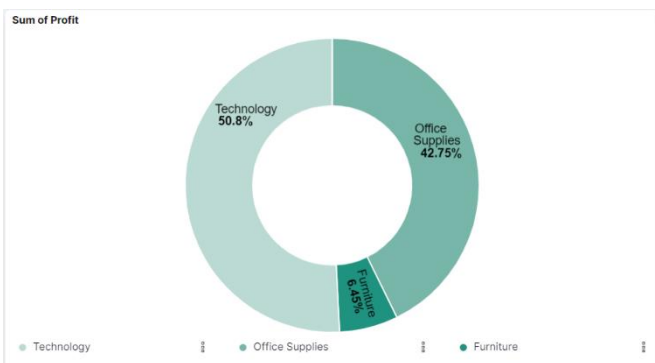

Fig: 11. Donut Chart for Sum of Profits.

Figure 11 represents the percentage of Sum of profits according to the given input constraints. Overall report shows, Furniture with less percentage so the store can improve or implement marketing campaigns in respective areas.
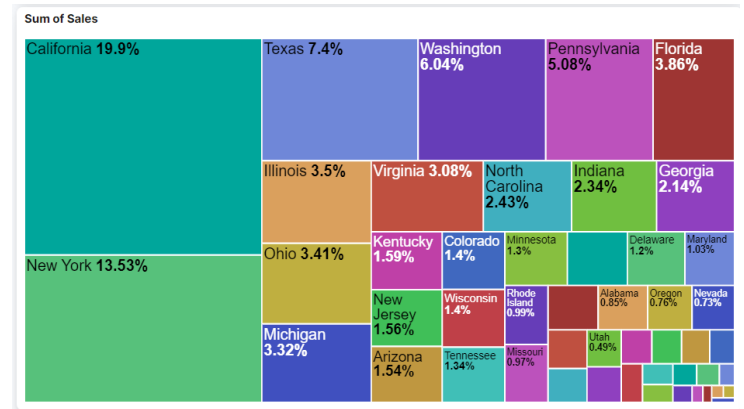

Fig: 12. State vs Percentage of Sales

Figure 12 depicts the percentage of Sales in Each state which are displayed according to the given constraints. Some states such as Virginia, Maine of the westside and Dakota of northside produce less sales due to reasons of population, product pricing, no. of stores in respective locations is less.
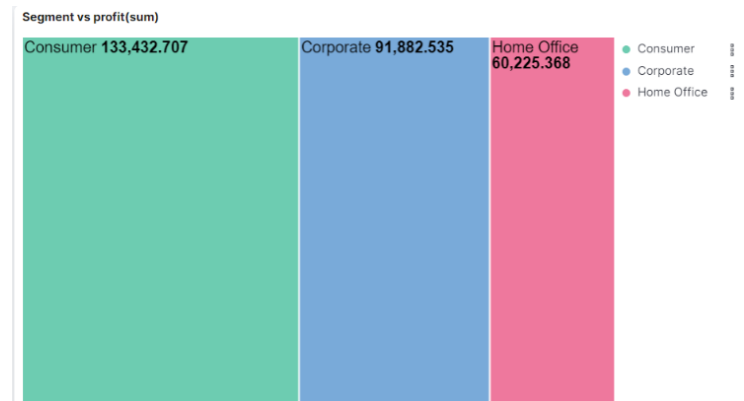

Fig: 13. Segment vs Profits

Figure 13 shows the sum of profits produced by 3 different segments. Most profitable is Consumer>Corporate>Home office. So, marketing strategy has to place more focus on retaining Consumer and Corporate Segment customers. Also, can provide special offers for home office to improve profits by sales and provide or improve marketing campaigns for corporate segments.
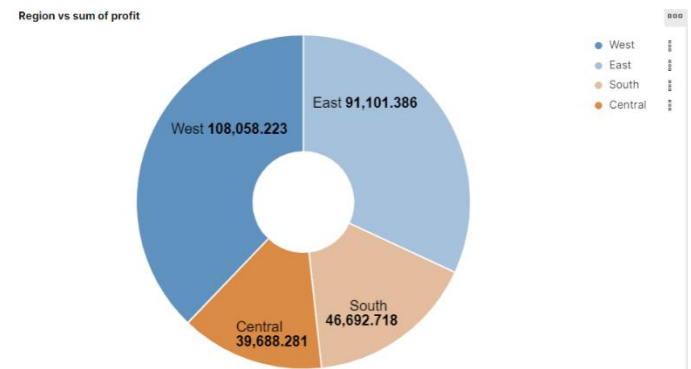
Fig: 14. Region wise Profit

Figure 14 shows the profit acquired region wise which can be viewed based on constraints. Profitable regions from the overall time period are East > West > South > North.
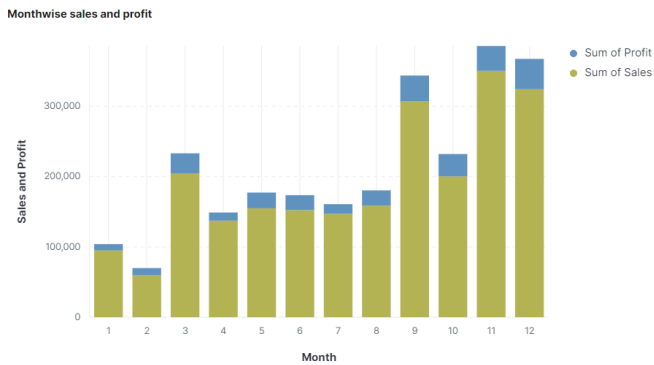

Fig: 15. Month Wise Sales & Profit

Figure 15 shows Month wise Sales and profit according to given year and region constraints. We can observe the sales and profit on the overall time period is in March and from September to December which can be due to seasonal purchase & festivals reasons by family members.
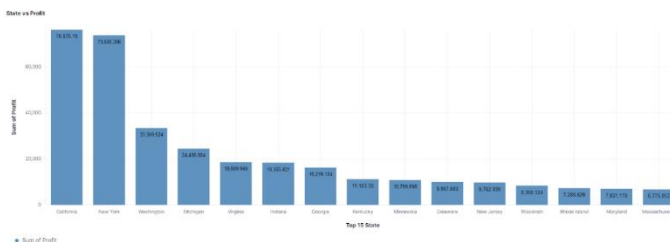

Fig: 16. State versus Profit

Figure 16 represents the plot of profit generated to the top 15 states in respective year and region accordingly using a vertical bar plot in descending order.
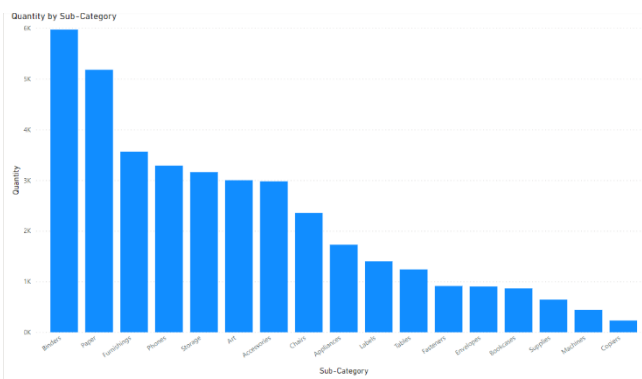

Fig: 17. Subcategory vs Quantity

Figure 17 shows the bar plot of subcategory versus quantity in descending order. One can deduce that, store should ensure inventory is always well-stocked for the top selling sub-category. Eg: Binders, Paper, Furnishings.
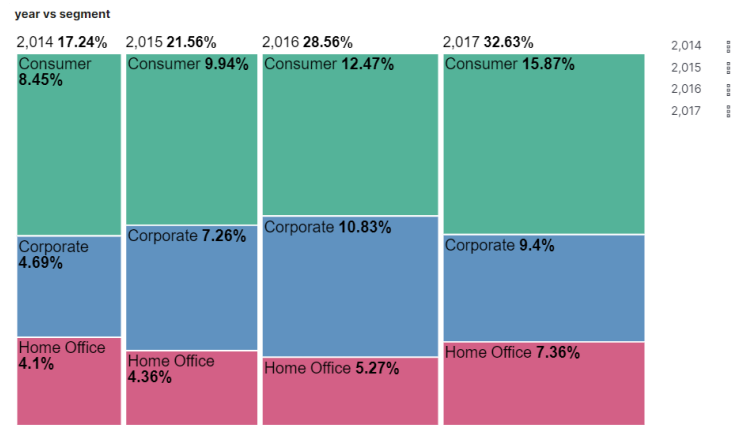

Fig: 18. Year vs Segment (in %)

Figure 18 represents the distribution of profit of the segment versus years of overall time period using a mosaic chart. In 2017, Consumer & Home office produced more profit percentage than compared to Corporate. So, the store sorts the problem by proper analyses such as checking and maintaining proper stock of products, proper maintenance, and providing better discounts for some goods at seasonal time.
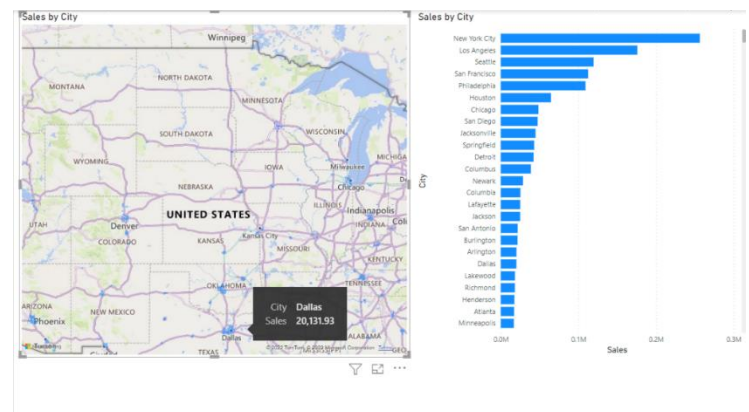
Fig: 19. City with Highest Sales

Figure 19 depicts there is a huge disparity (great difference) between the cities with highest & lowest sales. So, Marketing strategy has to target respective cities according to decisions made by the company.

An image of the dashboard for the Company along with some constraints(controls) to Year, Month, Category, Subcategory, Region which are linked to the whole dashboard, is shown below in figure 20. (Note: this is done with some logical representations).
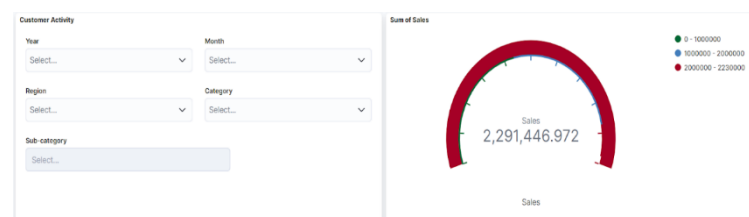
Fig: 20. The Gauge plot on right shows the Total Sales sold in each period respectively.

Fig: 21. Month versus Sum and average of Discount
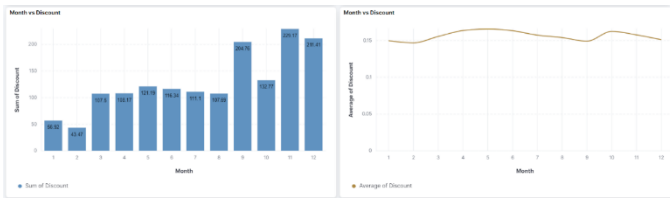
From Figure 21, customers can view the month wise total profit and discount obtained for each category or subcategory respectively which in turn can boost him/her to make a note on respective categories.

Fig: 22. Subcategory versus Discount

From Figure 22, customers can view the maximum of discounts provided to top 15 subcategories which can motivate them to view about the categories and investigate the product, which indirectly helps in marketing them.
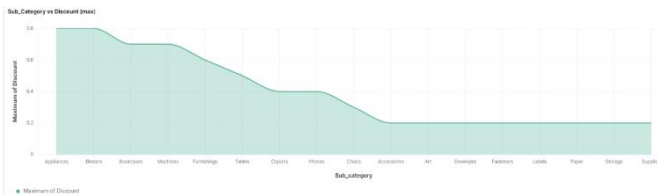

Fig: 23. Forecast for a year

Figure 23 provides a view of observed sales versus predicted one which provides an information that in march the observed sales is less than forecasted which may be due to various reasons such as economy crisis and from July to upgrowing months the revenue is generated more or less equally to the observed and forecasting.
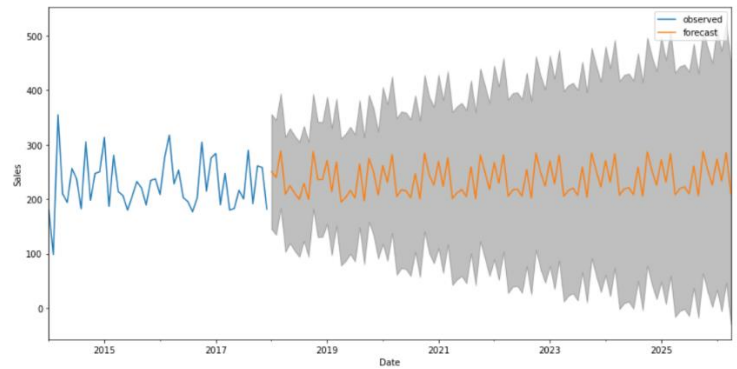

Fig: 24. Forecast for further years

Figure 24 the plots denote the sales under some specific range in form of upper limit and lower limit, where lower limit represents the worst-case scenario and upper limit represents the best-case scenario over a period of time of 5years. It can be also observed that the forecasted sales tend to remain same throughout and this may vary in future due to enormous situations or conditions which are unpredictable, but lies under the provided range of values or plotted area.

9. Conclusion

The analysis and forecast ensure that the problem statement is followed with necessary questions answered with the analysis. For loss-making products like Supplies, Bookcases, Tables, consider to either drop these from the catalog or change suppliers and bargain for a cheaper price. Consumer and Corporate Segments make up more than 70% of the customer base. Target them, especially customers from the East and West region in the Top 10 cities with Highest Sales by introducing special promotions and bundles for mass Consumer and Home Offices and send promotional emails or flyers. Sales of the store have increased every year resulting in high profit by margin by the end of 2017. With this analysis, the store can identify various aspects of the purchasing pattern and take respective measures. Also try to reduce the Expense by Supply Chain Management wise and increase the sales and profit towards geographical wise; also forecasted the Sales for further upcoming few years.

References:

[1] https://www.kaggle.com/code/layusmen/superstore-sales-analysis-and-forecasting/notebook

[2] https://www.kaggle.com/datasets/vivek468/superstore-dataset-final?select=Sample+-+Superstore.csv

[3] https://www.kaggle.com/code/karan842/superstore-sales-eda

[4] https://www.kaggle.com/code/apoorvmehta1993/superset-store-sales-analysis

[5] Chen,Y.L., Chen, J.M., Tung, C.M., 2006. A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Decision Support Systems,

[6] Griffith, D. A., 2005. An examination of the influences of store layout in online retailing. Journal of Business Research, Volume 58, Issue 10, 1391-1396

[7] Tripathi, G., Dave, K., 2013. Store format choice and relationship quality in apparel retail: A study of young and early-middle aged

shoppers in New Delhi region. Journal of Retailing and Consumer Services. V

[8] Yu, H., Tullio-Pow, H., Akhtar, A., May 2015. Retail design and the visually impaired: A needs assessment. Journal of Retailing and Consumer Services.

[9] https://subscription.packtpub.com/book/big-data-and-business-intelligence/9781788831031/1/ch01lvl1sec10/what-is-elk-stack.

[10] https://www.oreilly.com/library/view/elasticsearch-7-quick/9781789803327/ce73a675-ffc9-4e8c-8c2b-5e94356d50ce.xhtml

[11] https://www.softwareadvice.com/resources/retail-data-analysis-to-boost-sales/

[12] https://www.sciencedirect.com/science/article/pii/S016781161930062X

[13] https://www.menervasoftware.com/retail-analytics-solutions/