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

A. Kannammal, M. Sarathram, R. Gokulavasan,

Dept. of Decision and Computing Sciences, Coimbatore Institute of Technology, Coimbatore.
Email: 1833014mdcs@cit.edu.in

Abstract.

Transactions made in superstore data are extremely useful for the turn of events and can likewise be utilized to forecast future sales and to study past business challenges. The utilization of Business Intelligence apparatuses can assist with breaking down a lot of information including quality investigation and company examination. This undertaking was planned to utilize BI way to deal with examining the information. The research focuses on reporting the data regarding the sales and profits 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 supposed by organization pioneers to have the option to comprehend the information that will have been handled in figuring out visual structures and can undoubtedly retain the data expected to pursue choices for the organization.

Keywords: Business Intelligence, Retail Analysis, Bivariate Analysis, OLAP, Kibana, Dashboard, Decision Tool.

1. INTRODUCTION

SuperStore is a retail business area in the United States. It includes organizations that work by having enormous size spaces which store and supply a lot of merchandise. To support this, Superstore needs a dashboard 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 recognize feeble regions and chances to help business development. The analysis also gives insights into the sales and profit of various products. The organization of the paper goes with the problem definition, briefing the existing problem in the society and the need for the system for effective decision making through visual tools, dashboard and continued by the objectives of the proposed system. The pre-processing and Exploratory Data Analysis depicts the process implemented and concepts used to derive insights from dynamic data; accompanied by the process flow which picturizes about the module and tools used for analysis and their overview. The results and discussion describe the dashboard approach and various charts used for real time insights on data. The future scope describes the possible future works that might be undertaken to benefit the society and mankind. The conclusion part describes the decision making in the management domain by dynamic dashboard.

2. LITERATURE REVIEW

In existing systems [1] 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 the order date and shipment date. Also, some customer-level analyses are made

such as the first and last purchase of each customer, calculating retention of each customer, the number of orders by each customer, and 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 the target variable. Can be also known as a univariate analysis of the data [1]. The proposed system consists of Exploratory Data Analysis upon two factors such as sales and profit and forecasting the sales for a further few 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 a better decision-making process. 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. With this analysis, the store can recognize different patterns of purchasing products and take measures respectively. 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.

3. OBJECTIVES

The purpose of the project is to analyse sales and profit and to perform various superstore statistics, thus providing a perspective on the way business is trending. With developing requests and merciless rivalry on the lookout, it would be smarter to comprehend which items, areas, classifications, and client fragments they ought to 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. The Store tends to understand certain features of its 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 the proper stocking of needed goods, 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 and increase the sales and profit geographically wise; also forecasted the Sales for further upcoming few years.

4. ANALYSIS

Descriptive Statistics of the data have been done. NULL values and any other missing values are checked. Splitting of the order date to their entities such as day, month, and the year has been made. Ship Date has been dropped since there was no relevance for the visualization. The final data consists of dimensions on basis of 10000 * 23 instances. Python & Excel are used for performing Exploratory Data Analysis on the Data. 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 Variable, 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 Plots and histograms were utilized for visualizing and analyzing the distribution of continuous variables
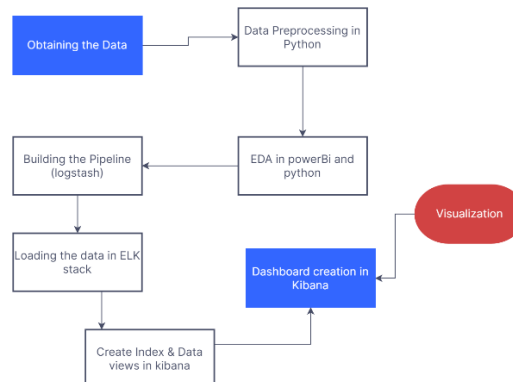
5.   MODULE AND PROCESS FLOW

Fig: 2. Process Flow of the proposed system

Figure 2. Represents the flow of the proposed system where the process is made with help of the above-mentioned tools. The data is extracted from the Kaggle website and pre-processing steps have been processed in python and some of the basic visualizations obtained from EDA operations(included in results) are done in PowerBi. For interactive visualizations and real-time processing systems, Kibana comes into play with help of ELK Stack architecture. With help of a pipeline(Logstash), the final pre-processed data has been pushed/loaded into the distributed search and analytics engine known as elastic search, where Kibana is used for visualization purposes by navigation on your data present in elastic search[7].·

6.   RESULTS AND DISCUSSION

This proposed system consists of two dashboards, one for the company view & other for the customers' view. An image of the dashboard along with some constraints(controls) to Year,

Month, State, City, and Region; Category, Sub-Category which are linked to the whole dashboard, is shown below in figure 3.
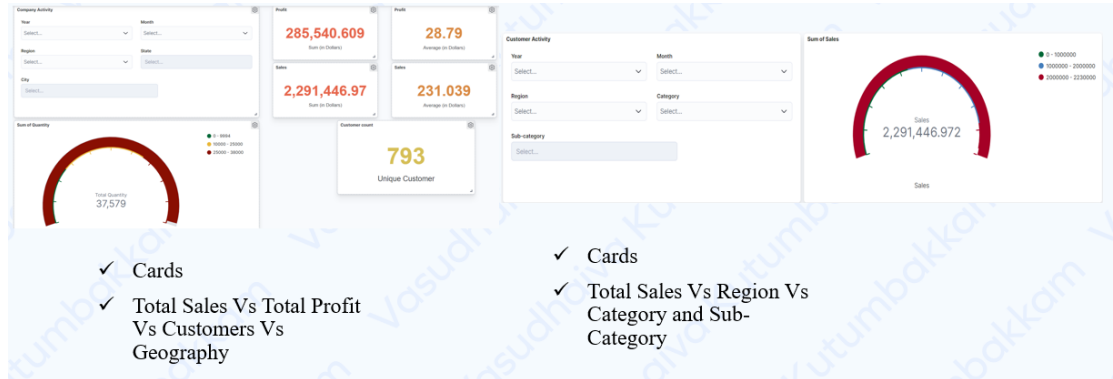

Fig: 3. Dashboard for Decision Making and Evaluation

Figure 3. The Metric plot is used to show the Overall Sum and Average of Sales, Profit according to given constraints in form of cards. The Gauge plot shows the Total Quantity sold in each period respectively

Fig: 4. Category vs Sum of sales

Figure 4 shows that all 3 categories make similar amounts of sales. Although Furniture makes comparable deals as Technology, it is the most un-beneficial likewise amount sold is at least. Office Supplies sell the most, as far as amount as it is a somewhat modest item.
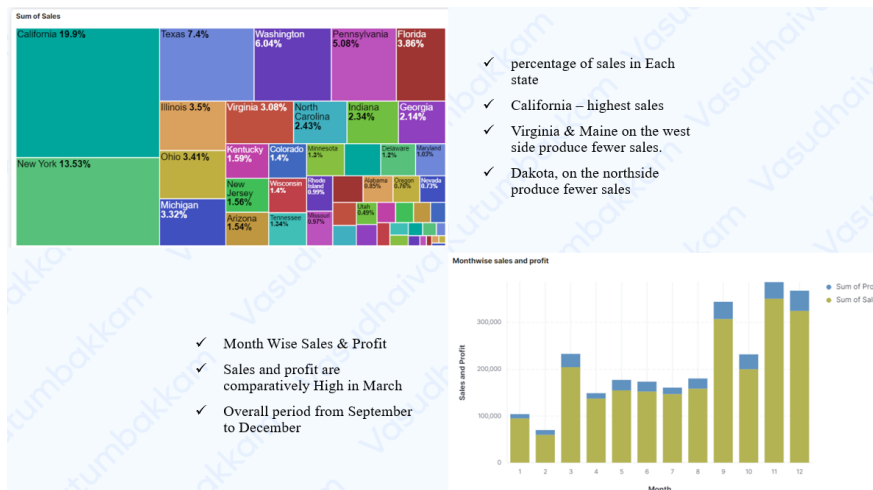
Fig: 5. State vs Percentage and Sum of sales

Figure 5 depicts the percentage of Sales in Each state which are displayed according to the given constraints. Some states such as Virginia, and Maine on the west side and Dakota on the northside produce fewer sales due to reasons of population, product pricing, no. of stores in respective locations is less. Using a vertical bar plot in descending order, the plot of profit generated by the top 15 states in respective years and regions can be generated accordingly; We can observe the sales and profit for the overall period in March and from September to December which can be due to seasonal purchases & festival reasons by family members.
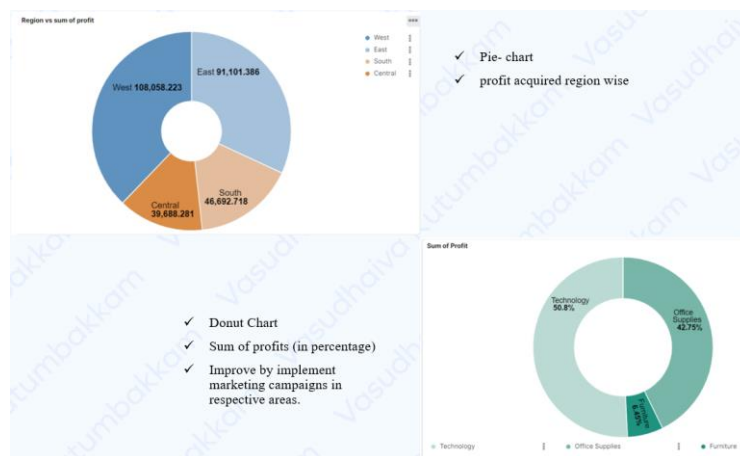

Fig: 6. Profit vs Region-wise and Category

Figure 6 shows the profit acquired region wise which can be viewed based on constraints. Profitable regions from the overall period are East > West > South > North. Overall report shows, Furniture with less percentage so the store can improve or implement marketing campaigns in respective areas.

Fig: 7. Profit vs Region-wise and Category

From Figure 7, 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 company in marketing them.
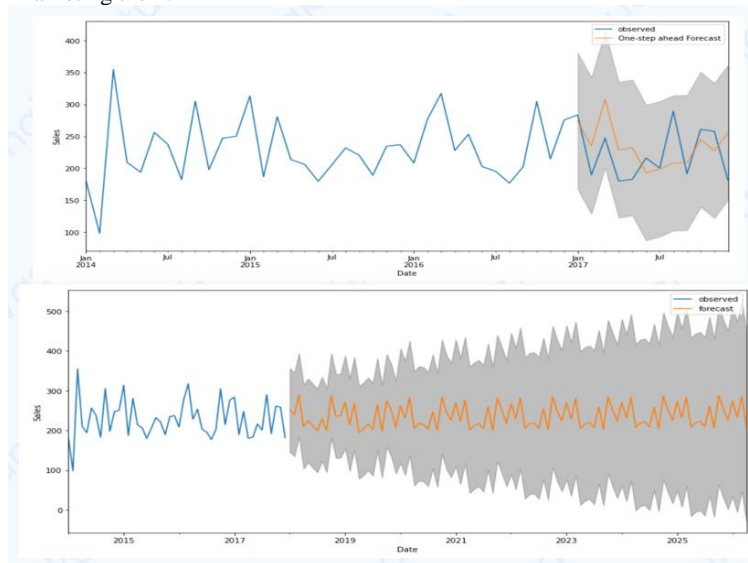

Fig: 8. Forecast for a Year and Further Years

Figure 8, the above plot provides a view of observed sales versus predicted one which provides a piece of information that in march the observed sales is less than forecasted which may be due to various reasons such as the economic crisis and from July to upgrowing months the revenue is generated more or less equally to the observed and forecasting; Below plot, denote the sales under some specific range in form of upper limit and lower limit, where the lower limit addresses the worst situation imaginable and the furthest cut-off addresses the most ideal situation throughout 5 years. It can be also observed that the forecasted sales tend to remain the same throughout and this may vary in the future due to enormous situations or conditions that are unpredictable, but lies under the provided range of values or plotted area.
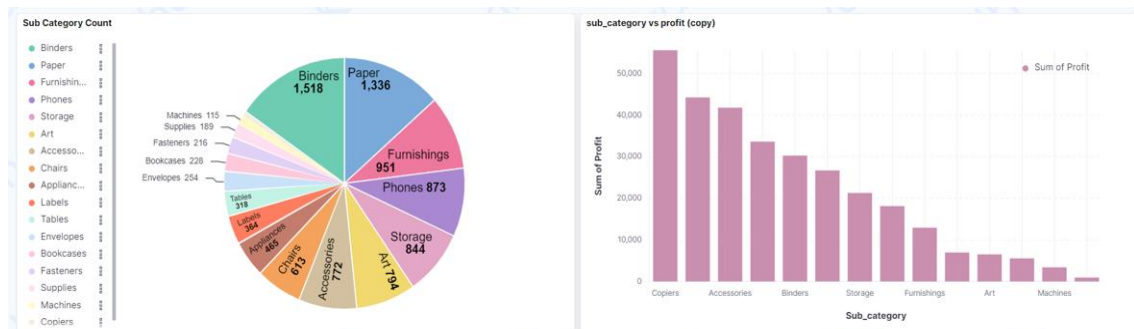
Fig: 9. Profit vs Region-wise and Category

Figure 9 represents the count of subcategories concerning 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 suitable decision making. Art produces Less profit but the count is high so the store can reduce the stock to prevent overstocking and help in supply chain management.

7. FUTURE WORKS

The implemented work is based on existing information (secondary data), but this will work in a better manner and more efficiently for sources like primary data. Having access to primary data will help to gain real-time information and updates so that predictions will be made more accurately so a person responsible for these activities can make decisions based on acquired results such as maintaining stock in accordance with supply chain management and deciding which areas to concentrate based on geological factors to improve the sales and revenue.

8. CONCLUSION

The analysis and forecast ensure that the problem statement is followed with necessary questions answered with the analysis. For misfortune-making items like Supplies, Bookcases, and Tables, consider either dropping these from the list or bargaining for a cheaper price from another supplier. Consumer and Corporate Segments make up more than 70% of the customer base. Target them, especially customers from the East and West regions in the Top 10 urban communities with Highest Sales by presenting extraordinary advancements and packs for mass Consumers and Home Offices and sending limited-time messages or flyers. Sales of the store have increased every year resulting in high profit by a margin by the end of 2017. With this analysis, the store can recognize different patterns of product purchases and take measures respectively. Also, try to reduce the Expense of Supply Chain Management wise and increase the sales and profit geographically wise; also forecasted the Sales for further upcoming few years.

9. REFERENCES

[1]   Layusmen, Sales Analysis and Forecasting, 2022. https://www.kaggle.com/code/layusmen/superstore-sales-analysis-and-forecasting/notebook.

8

[2]   Vivek Chowdhury, Dataset of Sales and Profit of Store, 2022. https://www.kaggle.com/datasets/vivek468superstore-dataset-final?select=Sample+-+Superstore.csv

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

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

[5]   Tripathi, G., Dave, K., "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, 2013.

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

[7]   ELK Stack Architecture, oreilly.com, 2022.

[8]   Agnes Teh Stubbs, softwareadvice.com, "How to use retail data analysis to Boost Sales", April 2018.

[9]   Marnik G. Dekimpe, "Retailing and retailing research in the age of big data analytics", International Journal of Research in Marketing, Volume 37, Issue 1, March 2020.

[10]   Mrs. Madhura K. Mane, Dr. R.D Kumbhar, "Dashboard in Human resource Information Systems with reference to Balanced Scorecard", International Journal of Information Systems, Volume 8, February 2018.

[11]   Stacey Barr, "7 Small Business Dashboard Design Dos and Don'ts ", 2010.

[12]   Tushar Mulik, Pravin Malusare, Omkar Takalkar, Monali Patil., "Business Intelligence: Dashboard", Department of Information Technology, Mumbai University, Mumbai, India,2014.

[13]   Vishnu Singh Khatuwal, Digvijay Puri, "Business Intelligence Tools for Dashboard Development", International Conference on Intelligent Engineering and Management (ICIEM), 2022.

[14]   L. Pappas and L. Whitman., "Riding the technology wave: Effective dashboard data visualization", Human Interface and the Management of Information, 2011.

Biographies

Dr. A. Kannammal is currently working as a Professor at Coimbatore Institute of Technology, Coimbatore, India. She has completed her Ph.D. in Computing Science from VIT University by 2007 and MCA from IGNOU by 2001. She is serving as research supervisor at Anna University, Chennai and Bharathiar University, Coimbatore. She has been guiding research scholars for PhD programme.

M. Sarathram is currently pursuing M.Sc. Decision and Computing Sciences (5year Integrated) as a Student at Coimbatore Institute of Technology, Coimbatore, India. He has completed full time internship for duration of six months from Huey Tech Private Limited, Noida, India by 2021.

R. Gokulavasan is currently pursuing M.Sc. Decision and Computing Sciences (5year Integrated) as a Student at Coimbatore Institute of Technology, Coimbatore, India. He has completed full time internship for duration of six months from Sycuris Tech Private Limited, Tamil Nadu, India by 2021.