

EXPLORATION OF WALMART RETAIL SALES DATA – A DESCRIPTIVE APPROACH

Chuks Chiazor

March 2022

www.chukschiazor.com

EXECUTIVE SUMMARY

Retail is one of the largest and most important industries, and as the middle class grows in size and purchasing power, this industry is likely to thrive. Because of its early applications of Big Data Analytics in its business processes, Walmart is widely regarded as the world's largest and most successful brick-and-mortar retailer. The goal of this study is to investigate Walmart's retail sales using Tableau software in a descriptive approach with six metrics - average sales, profit, and discount, distribution of average profit by state in the United States, percentage of profit per region, sales profits per month, average sales profit based on age, and Sub-category per average profit. The three theme solutions preceding the six key metrics were decision support, customer retention, and inventory management.

Table of Contents

EXPLORATION OF WALMART RETAIL SALES DATA – A DESCRIPTIVE APPROACH	1
EXECUTIVE SUMMARY	2
INTRODUCTION	4
METHODOLOGY	6
1. Data collection:	6
2. Data Cleaning:.....	6
3. Analysing the data using tools.....	6
4. Insight Analysis:	6
PROBLEM STATEMENT	7
DATA DESRCPTION.....	7
ANALYSING THE DATSETS.....	9
1. Average sales, profit and discount.....	9
2. Distribution of average profit by state.	10
3. Percentage of profit per region	11
4. Sales Profits per month.....	12
5. Average Sales Profit per Age.....	13
6. Product sub-category and total Profit	14
INSIGHT ANALYSIS	15
1. Decision Support	16
2. Customer Relationship Management	16
3. Inventory Management	17
CONCLUSION	18
REFERENCES	19
APPENDIXES	20
Appendix 1. Walmart retail data in Excel	20
Appendix 2. Product category and percentage of order quantity	20
Appendix 3. Assigning USA states on the map graph.....	21
Appendix 4. Customer Age bin parameter settings	21

Table of Figures

Figure 1. Methodology process	7
Figure 2.Importing dataset into tableau framework	8
Figure 3. Dashboard of average sales. profit and discount	9
Figure 4.Map chart of average profit per state in USA.....	10
Figure 5. Percentage profit per region.....	11
Figure 6. Total profit per month	12
Figure 7. Average sales profit per age.....	13
Figure 8. Sub-products and percentage profit.....	14
Figure 9. Case study visualization dashboard	15

INTRODUCTION

Retail is one of the largest and most important industries, and as the middle class grows in size and purchasing power, this industry is likely to thrive. (Keim, 2009; Wixom & Watson, 2001) Kroger, for example, uses data from its 40+ million cardholders' shopping habits to send personalised coupons to a large proportion of them on a regular basis. Ebay addressed their big data challenge by dividing their data analytics into three platforms, which allow eBay to conduct a variety of data analyses and offer new services to customers. (Roden et al., 2017). Furthermore, thanks to the massive amount of data that retailers can collect on their customers' (online and/or offline) purchases, data-driven personalization has become a realistic choice.

Walmart is widely recognised as the world's largest and most successful brick-and-mortar retailer. Walmart achieved global success by offering lower prices than competitors while providing services that others did not. They earned more than any other retailer in the world. Walmart has rapidly expanded to 27 countries in less than 60 years, with approximately 11,500 stores and over 2.2 million associates worldwide (Walmart, 2022) Each week, Walmart serves approximately 260 million customers, a truly astounding fact, especially when considering that Walmart records more customers than the world's population in a year. Walmart, as a retail industry leader, has been one of the first to recognise the value that data can bring to a company.

Walmart recognised the importance of allocating a significant portion of its resources to data management. Walmart Labs, led by the "Fast Big Data Team," was established in 2011. (Marr, 2016). A significant amount of time is spent at Walmart Labs on predictive analytics, which plays an important role in ensuring the efficiency of business operations.

Walmart is able to predict which days and times of day will be the busiest by analysing past experiences and conducting predictive analysis (Walmart lab, 2011). A precise forecast will allow Walmart to always have the optimal number of employees at the store checkout. Customers will finish their shopping faster as a result, and each Walmart location will have the optimal number of employees.

Another common, yet valuable, application of big data is getting to know customers and learning about their purchasing habits and preferences. (Walmart, 2022) As a result, Walmart Labs keeps track of these new discoveries because they can easily change and decide on new shopping patterns that customers will adopt. Walmart discovered new ways to use big data to improve every business segment, including supply chain and logistics. Walmart discovered this through the use of data and simulations.

Walmart discovered innovative ways to use big data to improve every business segment, including supply chain and logistics. Walmart discovered new methods to optimise transportation routes by using data and simulations (Walmart, 2022). Walmart was also able to save time and money by tracking the routes that truck transporters take. Truck drivers are also assigned a work schedule based on product availability and distribution. Furthermore, the steps taken by workers while loading, unloading, and transporting goods are recorded in order to determine where improvements can be made.

METHODOLOGY

We followed a methodology in this study. It consists of four steps:

Step 1

1. Data collection: The data for this paper was obtained from Kaggle (www.kaggle.com), an online community of data scientists. This study's information was extracted from a Walmart retail (USA) dataset. Section 4 provides a description of the dataset.

Step 2

2. Data Cleaning: The datasets used in this study have already been cleaned, so no additional data cleaning is required. All relevant columns that will be useful in this analysis are present in the data. (Please see Appendix 1)

Step 3

3. Analysing the data using tools: This is the third step in the methodology. Tableau is used to analyse the datasets here. Tableau is a data visualisation and business intelligence software that helps organisations make better decisions by providing good visualisation dashboards. Balaji et al., 2020.

Step 4

4. Insight Analysis: This is the final step in the proposed methodology. An insight to demonstrate the importance of big data analysis and to have a clearer and broader picture of the analysis is discussed here.

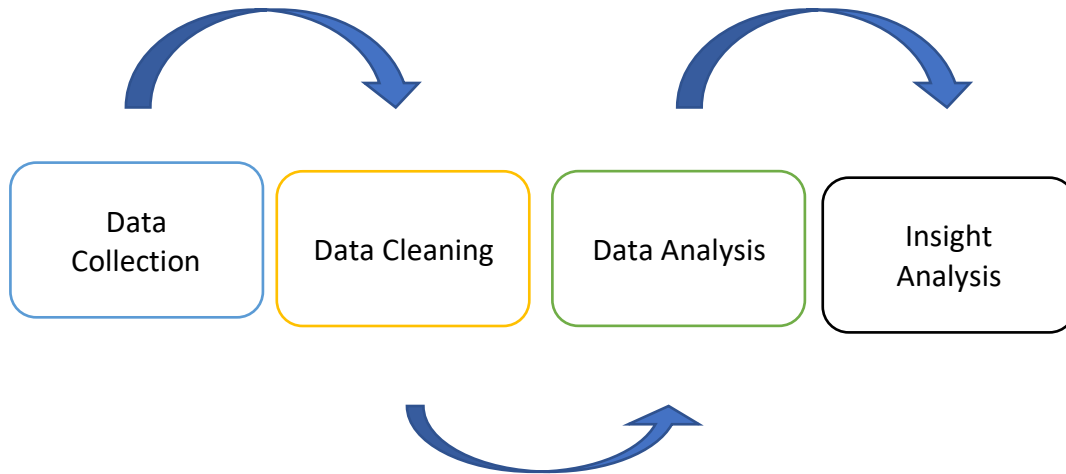


Figure 1. Methodology process

PROBLEM STATEMENT

The purpose of this report is to investigate Walmart's retail sales through analysis and visualisation using Tableau software. Furthermore, this study aims to achieve the following with the set Metrics - Average sales profit and discount, Distribution of average profit by state in the United States, Profit percentage per region, Sales Profits per month, Average Sales Profit based on Age, Sub-category, and Average Profit.

DATA DESCRIPTION

The retail data collected from Walmart retail is divided into three categories based on the number of orders: office supplies (55.05 percent), technology (24.29 percent), and furniture (20.64 percent) see appendix 2. The data spans 5 months, from 01/01/2012 to 27/05/2012, and consists of customer purchasing information. Customer name, Customer City, order date, product category, and so on are among the 27 fields and 8399 rows that comprise the data. (See Appendix 1).

Figure 2. below describes the process of importing the original data source which is present in excel format to tableau-platform, and because there are no multiple data, there was no need joining data.

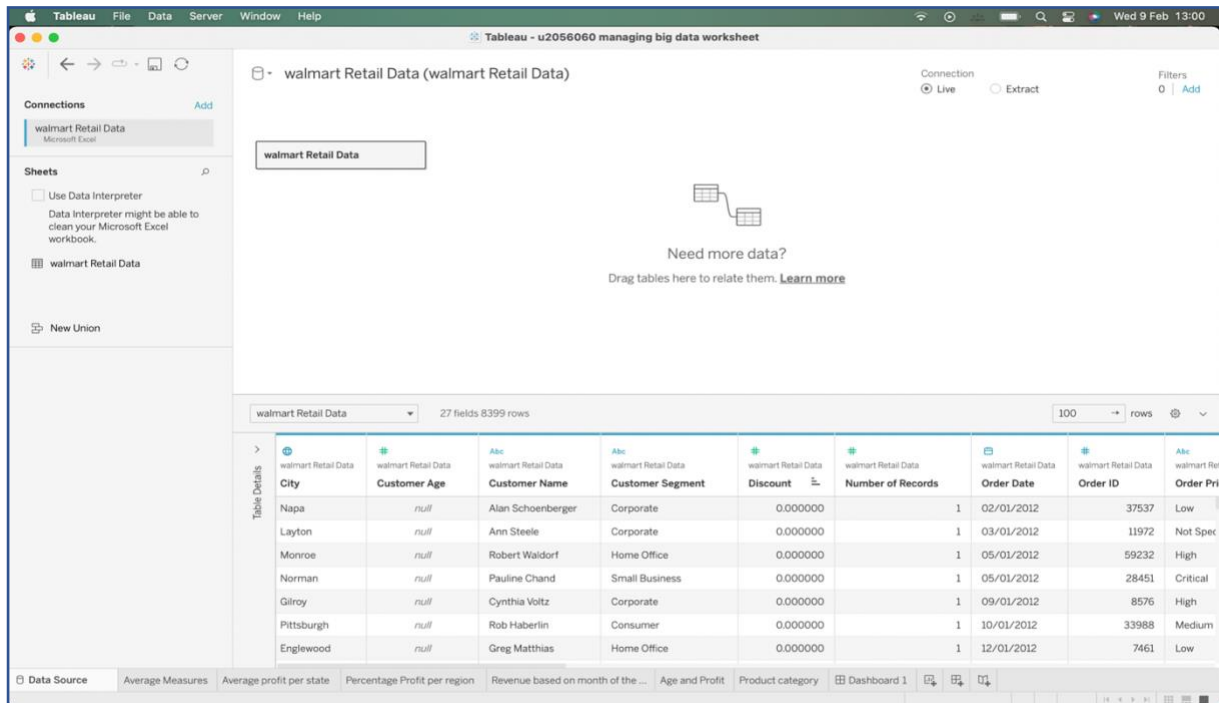


Figure 2.Importing dataset into tableau framework

The next step is to analyse the data after it has been successfully imported into the tableau platform. The data in the tableau platform is divided into two sections: numeric data goes into the measure section, and the rest of the data goes into the dimension section. In most cases, the tableau will divide the data for you. (Jena, 2019)

ANALYSING THE DATSETS

The proposed methodology concludes with this step. In this case, the datasets are analysed using tableau as the analysis and visualisation tool, based on the metrics provided.

1. Average sales, profit and discount

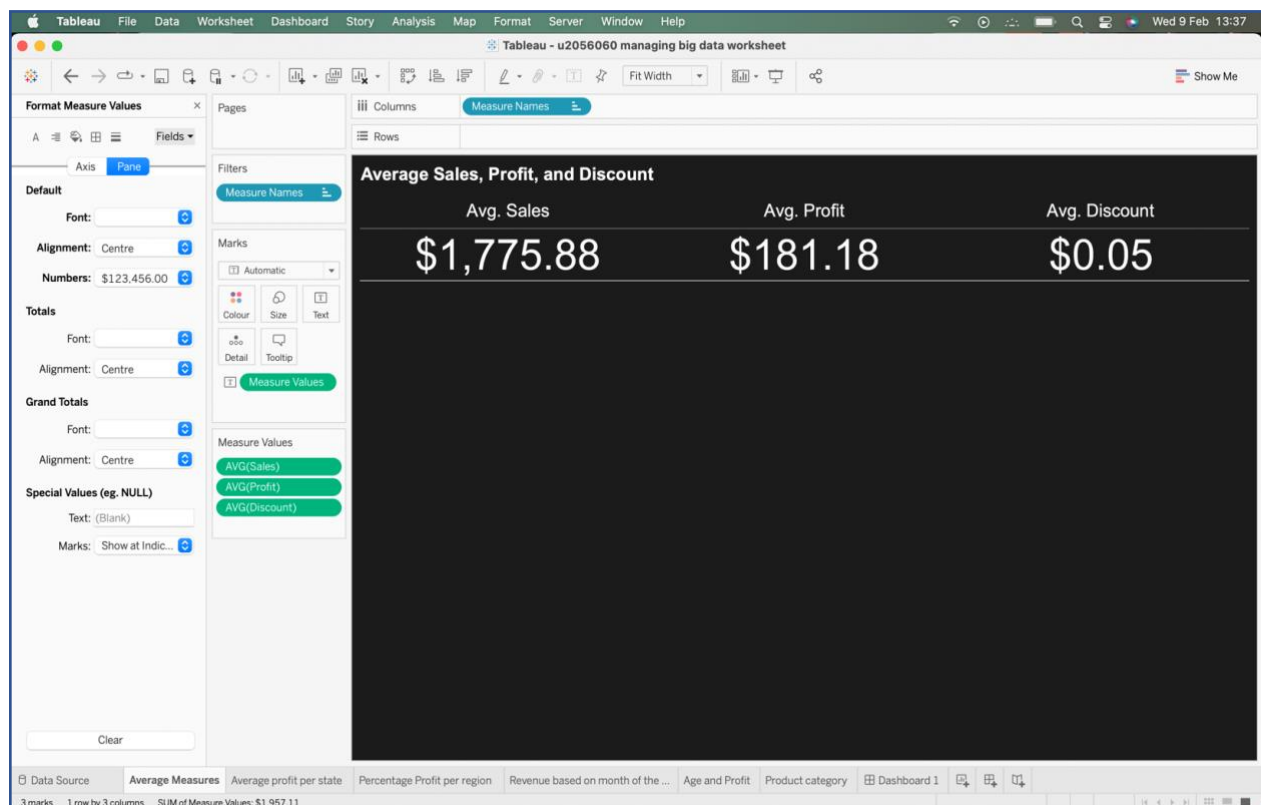


Figure 3. Dashboard of average sales. profit and discount

The graph above depicts the calculated total average sales, as well as the average profits and discount for the previous five months. As shown in Figure 3, the average sales from Walmart data were \$1,775.88, the average profit was \$181.18, and the average discount was \$0.05. This metric allows stakeholders to see at a glance how much average sales, profit, and discount the various stores achieved.

2. Distribution of average profit by state.

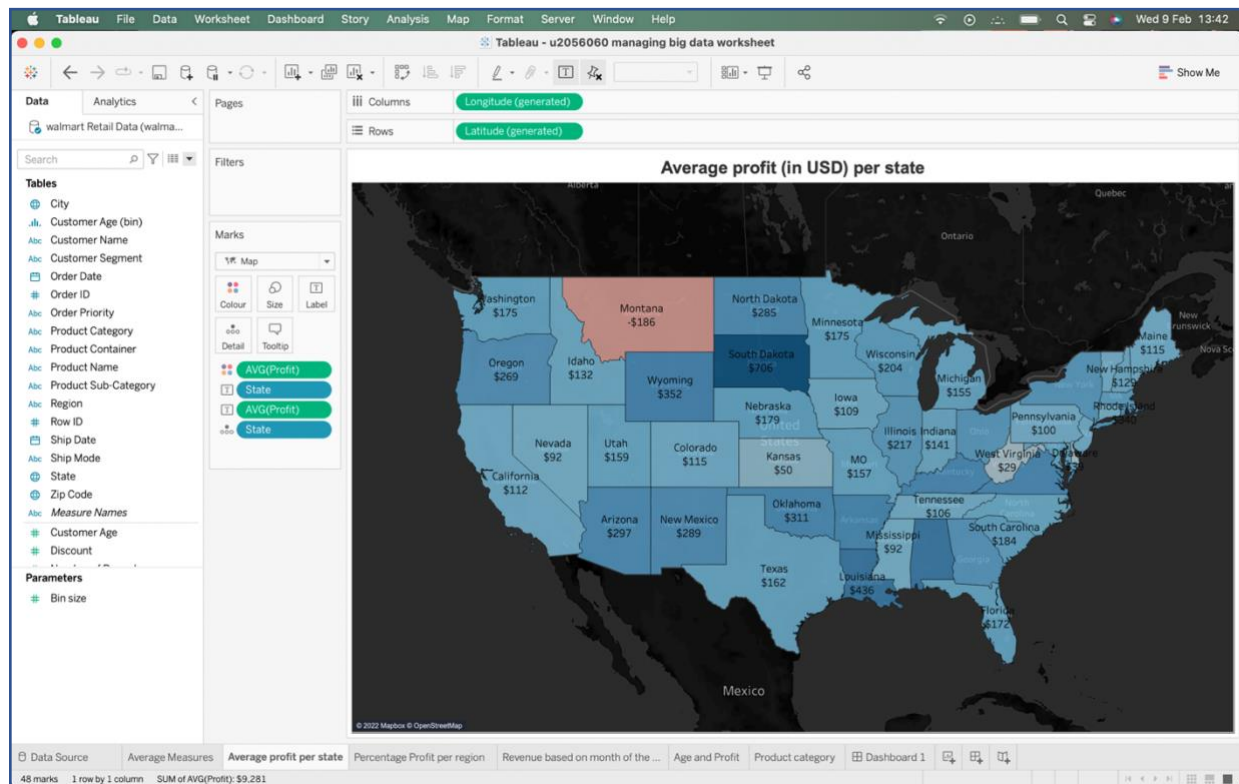


Figure 4. Map chart of average profit per state in USA

The graph above attempts to distinguish between states that generate the most profit and those that lose money. The map chart above depicts the average distribution of sales profits across states in the United States. It should be noted that the data set provided an individual profit for each of the 904 cities in the United States, and in order for us to successfully gain an understanding of sales profit distribution, the average sales profit per state was calculated. For more information on how this was accomplished, see (appendix 3). Figure 4 shows the different average sales profit contributions per state where the blue colour concentration is high. The higher the concentration of blue colour, the higher the profit, and the brighter the blue colour, the lower the profit, which results in an orange colour. According to the map graph, South Dakota had the highest percentage sales profit of \$706, followed by Louisiana, which had \$436, and Montana, which had the lowest profit of -\$186.

3. Percentage of profit per region

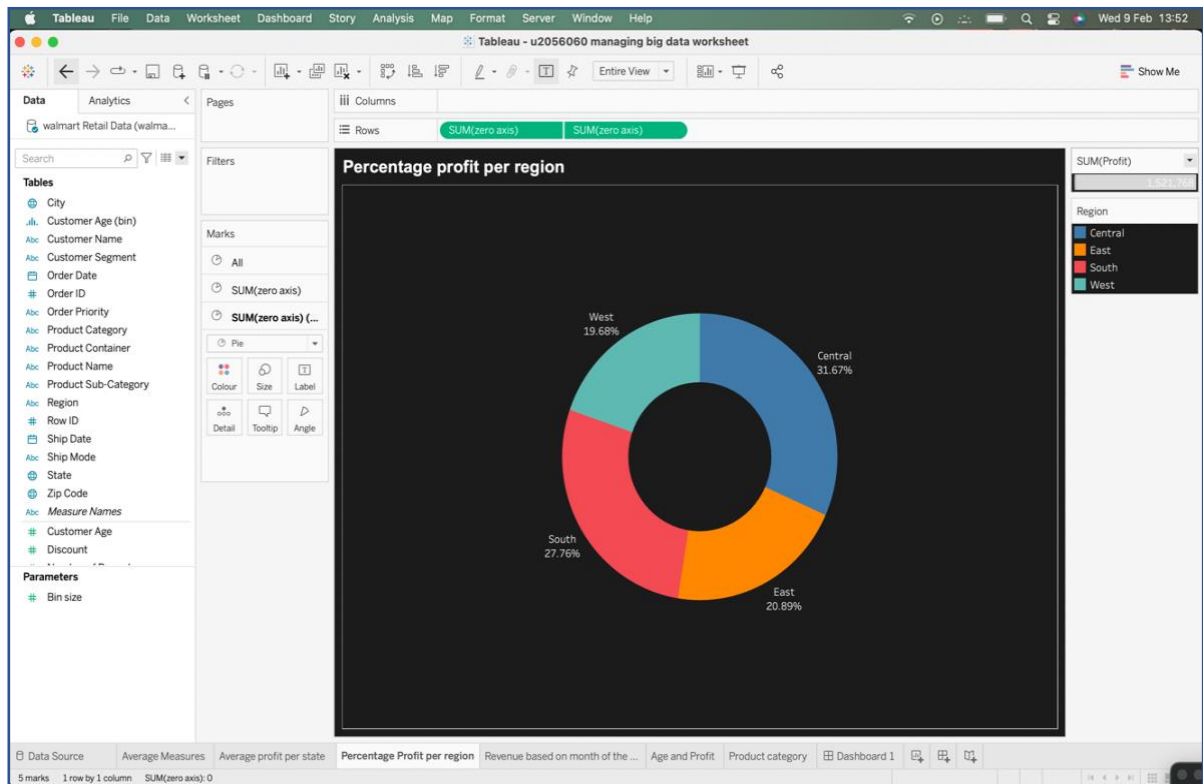


Figure 5. Percentage profit per region

After visualising the distribution of sales profits by state, it is important to understand the distributions among regions, so a doughnut chart is used to investigate the proportion percentage of sales profits by region. The west region, which includes Montana, has the lowest sales percentage profit of 19.68 percent, while the central region, which includes South Dakota, has the highest percentage of 31.67 percent. Figure 5 shows that the average sales distribution is similar to that shown in Figure 4.

4. Sales Profits per month

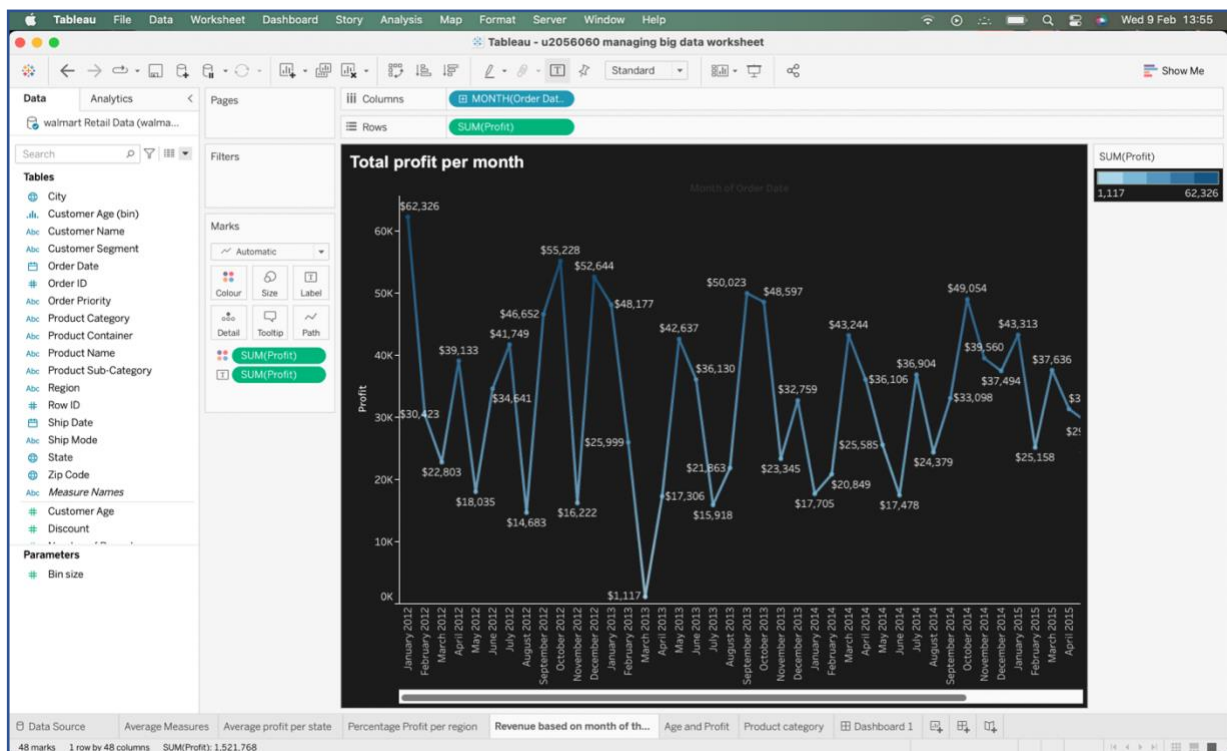


Figure 6. Total profit per month

In this case, an attempt is made to determine the monthly sales profit and to comprehend the purchasing habits of customers by obtaining data from the order dates column and totaling the sales profit, as shown in figure (6) above. We can see from the line graph above that the colour concentration becomes darker as the sales profit increases and lighter as the sales profit decreases. We can clearly see how the number of profits peaked at the beginning of January 2012, with a total profit of \$62, 326, and then fell precipitously, reaching its lowest point for the year in August, with a total profit of \$14, 683. The following year saw the lowest profit of \$1,117 in March 2013, but then randomly reached its peak between September and October. In 2014, a more consistent trend pattern is observed, with October having the highest peak of sales profit (\$49.054) and June having the lowest (\$17,705). We can see that the trend pattern is a random variation, which means that the line movements follow no pattern and are thus unpredictable.

5. Average Sales Profit per Age

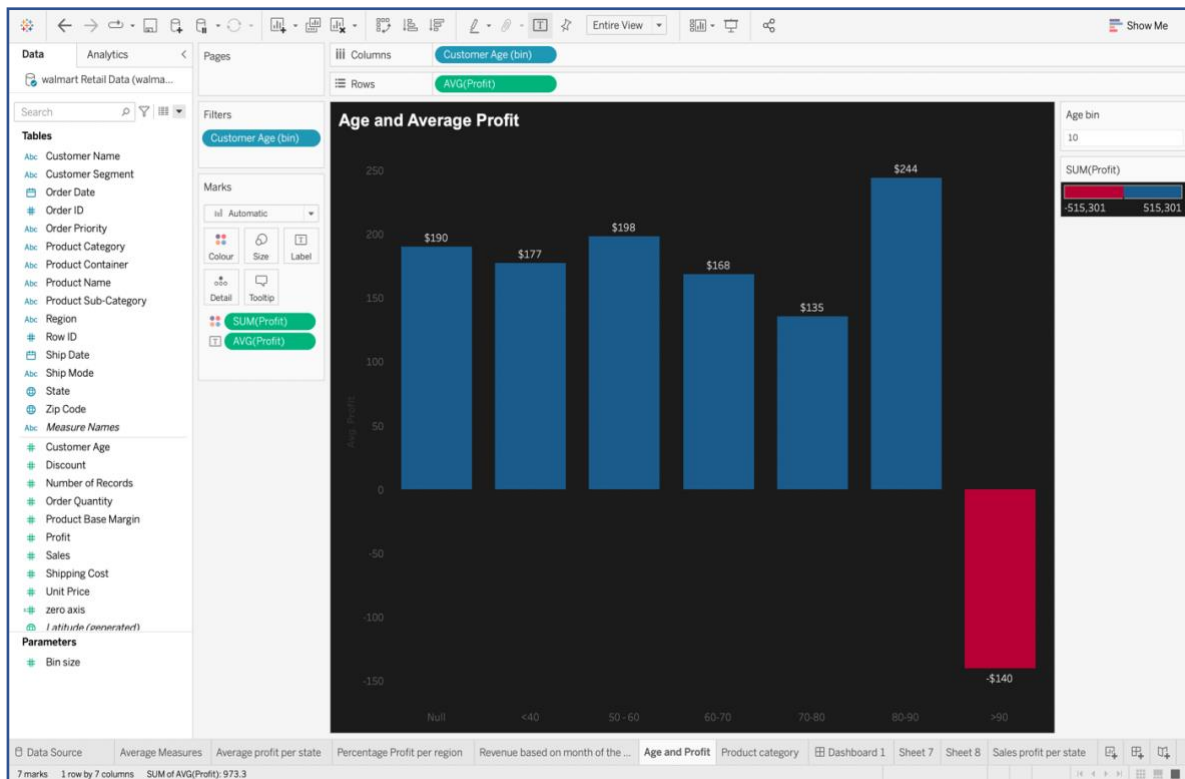


Figure 7. Average sales profit per age

This analysis attempts to determine which Walmart customer age group generates the most profit for the company, and it is represented visually by a bar chart.

The column for customer age and profit is in measure. A parameter called bin size was created, with the data type set to integer, the current value of 10, and the display format set to automatic. (See Appendix 4). The created parameter will then be used for the customer age bin. The bar chart colour was set to two stepped colours, Blue (positive) and Red (negative), with a centre point of 0. The positive average sales profit values are displayed in blue on the bar graph above, while the negative values are displayed in red. It can also be seen that the 80-90 years age range produced the highest average sales profit (\$244), followed by the 50-60 years age range (\$198), and the least with a negative average sales profit value of -\$140 was >90. Which leads to the conclusion that being over 90 is a loss for business, whereas being between 80 and 90 is beneficial.

6. Product sub-category and total Profit

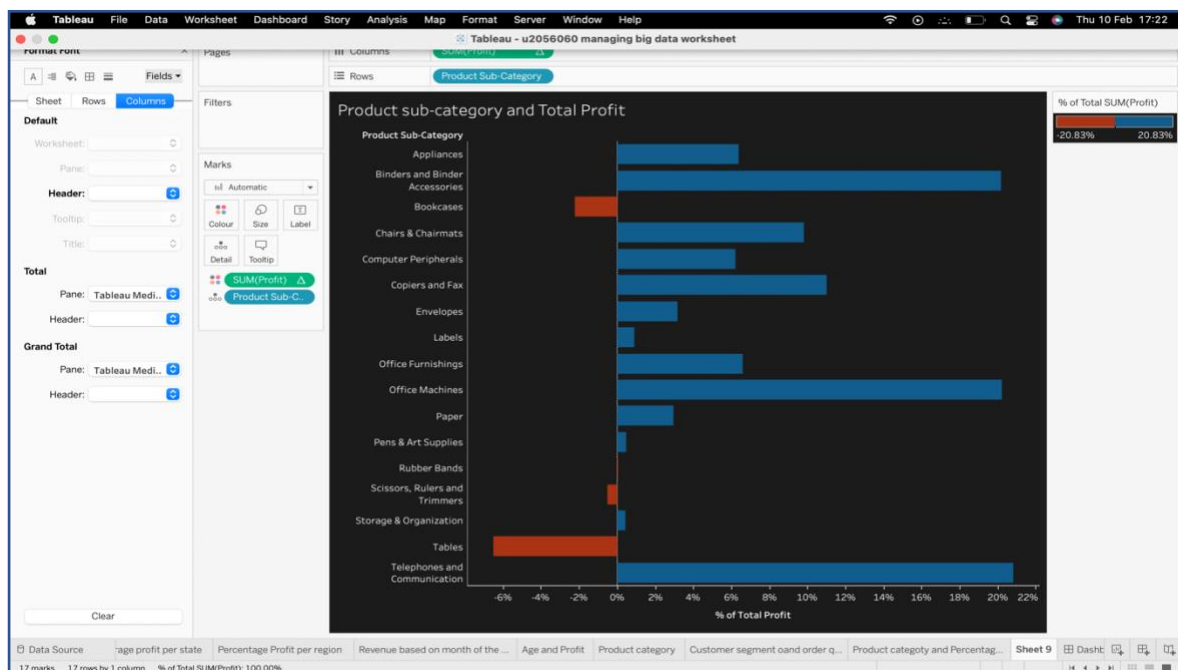


Figure 8. Sub-products and percentage profit

The side bar chart above depicts each sub-product category and its percentage profit. Based on this analysis, we can determine which products are profitable and which are losing money. The blue bar charts show positive profits, while the red bar charts show negative profits. According to the figure above, telephones and communication have the highest percentage profit of 20.83 percent, office machines come in second with 20.22 percent, and binder and binder accessories come in third with 20.20 percent, indicating that these products are beneficial to businesses because they maximise profits. Tables (-6.51 percent), bookcases (-2.21 percent), scissors, rulers, and trimmers (-0.51 percent) are all losing money and are highlighted in red. This does not, however, imply that these products are suitable for business use.

INSIGHT ANALYSIS

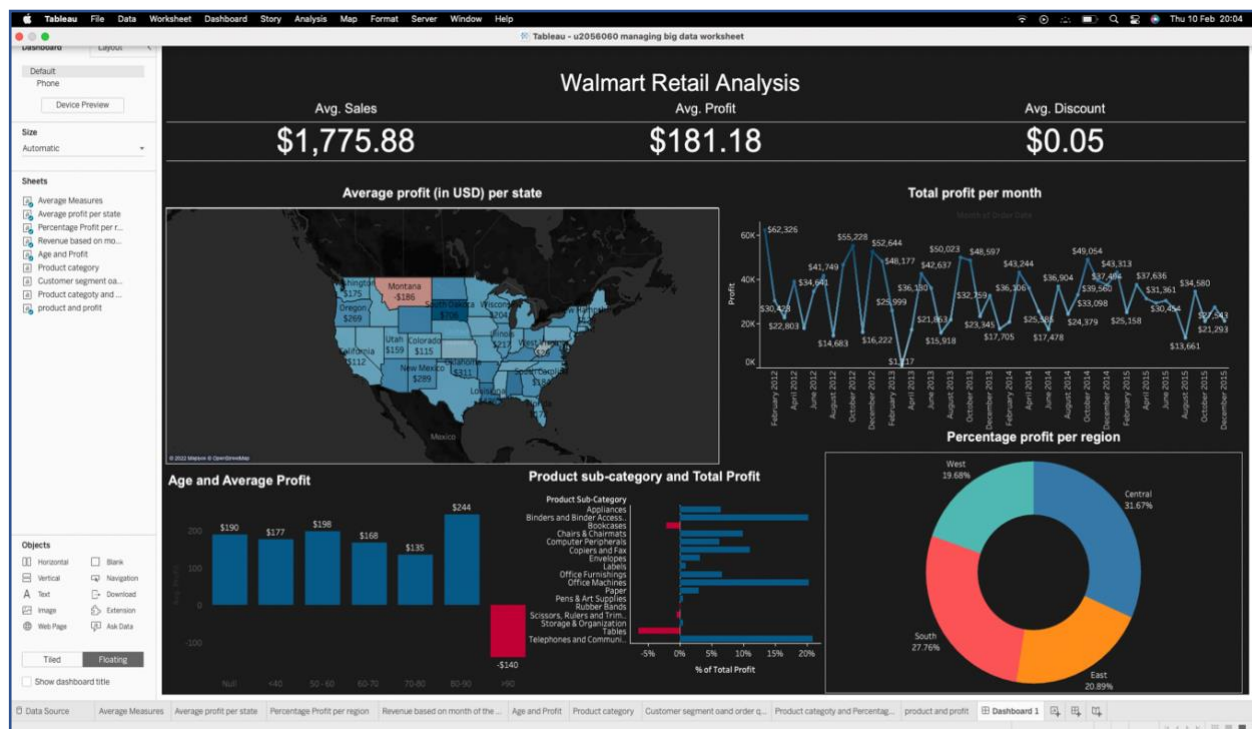


Figure 9. Case study visualization dashboard

A good business dashboard informs at a glance, and a great business dashboard combines high performance and ease of use to allow anyone to get data-driven answers to deeper questions. (Tableau, 2022) Figure 9 depicts a high-level overview of Walmart retail analysis.

Following an examination of the study's findings, it became clear that Big Data Analytics and visualisation have the potential to add significant value to Walmart and retail performance in general. According to the findings, the insight value provided by analysing and visualising Walmart sales data sets can be divided into three themes.

1. Decision Support

The analysis of average profit per state in the United States, as shown in Figure 4, acknowledges that the use of Big Data Analytics in Walmart can significantly improve decision making processes. According to Aversa et al., (2021), the adoption and development of Big Data decision support is heavily reliant on a data environment that promotes transparency and a clear corporate data strategy. This report, on the other hand, is seen as useful to stakeholders in decision making, which could include making the right investment based on location. In this case, investing in states such as South Dakota could yield a high profit for the organisation, and be wary of the risks of investing in low-profit states such as Montana.

2. Customer Relationship Management

This theme encapsulates the role of big data analytics and visualisation in improving customer relationships and experiences at Walmart. We can observe customer purchasing patterns on a monthly basis, as well as gain an understanding of total profit per month, using the visualisation analysis in Figure 6. Understanding customer purchasing patterns allows one to improve customer experience by knowing when to place personalised ads for seasonal products, for example. Subrahmanyam et al., (2020) identified the potential of BDA and AI in social media interactions in the retail sector in a variety of tenets such as sales and marketing, promotional activities, consumer interactions, customer service, and the impact of end user observations and reactions, as well as real time discussion.

Again, knowing the total sales profit per month can help Walmart managers with inventory management. By observing trends, these managers can better understand when to stock products to maximise sales. According to the study by Belarbi et al., (2016), big data is viewed as the most important tool for improving assortment in retail stores.

Furthermore, the sales profit per age analysis, which seeks to determine which Walmart customer age group generates the most profit for the organisation, is extremely useful in the

CRM department for customer experience and retention. Target marketing can be applied to certain age groups that may increase sales and profit while avoiding age ranges that produce losses for in this case, ages above 90 years; this can help save money.

3. Inventory Management

This theme captures the methods of inventory management in retail by utilising efficient BDA applications, such as Assortment in online retail stores. (Sultan et al., 2020) It is essential to note the calculations used to analyse the product category and the percentage number of orders. This was the solution as shown in Appendix 2: Office supplies are the most ordered product category (55.05 percent), followed by Technology (24.29 percent), and Furniture (20.64 percent).

However, a side bar chart was used to visualise a further analysis to determine which product sub-category yielded the highest profits and which ones lost money. Figure 8 shows that telephones and communication had the highest percentage profit of 20.83 percent, followed by office machines (20.22 percent) and binder and binder accessories (20.20 percent). Tables (-6.51 percent), Bookcases (-2.21 percent), scissors, rulers, and trimmers (-0.51 percent) all had losses. This means that, despite the fact that office supplies have the most orders, telephone and communications (Technology category) appear to have more profit. This is significant in the decision-making process of Walmart stakeholders regarding stock inventories. Store managers can now determine which products to stock more of in order to increase sales and profit and which products to stock less of in order to avoid further losses.

CONCLUSION

The goal of this study was to analyse and visualise Walmart retail sales using Tableau software in a descriptive approach with six metrics - average sales, profit, and discount, distribution of average profit by state in the United States, percentage of profit per region, sales profits per month, average sales profit based on age, and sub-category per average profit. The three theme solutions preceding the six key metrics were decision support, customer retention, and inventory management. Furthermore, it is suggested that the protocols used in this study be used in the analysis of other services, such as banking and financial institutions. Similarly, the use of other big data analysis and visualisation technologies in retail, such as R, Python, and artificial intelligence/machine learning, opens up exciting new possibilities for prediction.

REFERENCES

- Aversa, J., Hernandez, T., & Doherty, S. (2021). Incorporating big data within retail organizations: A case study approach. *Journal of Retailing and Consumer Services*. <https://doi.org/10.1016/j.jretconser.2021.102447>
- Balaji , N., Karthik , P. B. H., Bhaskar, B., & Barmavatu, P. (2008). Data Visualization in Splunk and Tableau: A Case Study Demonstration. *Journal of Physics: Conference Series* , <https://doi.org/10.1088/1742-6596/1767/1/012008>
- Balaji , N., Karthik , P. B. H., Bhaskar, B., & Barmavatu, P. (2008). Data Visualization in Splunk and Tableau: A Case Study Demonstration. *Journal of Physics: Conference Series* , <https://doi.org/10.1088/1742-6596/1767/1/012008>
- Belarbi, H., Tajmouati, A., Bennis, H., & Mohammed, E. H. T. (2016). Predictive Analysis of Big Data in Retail Industry. *1st International Conference on Computing Wireless and Communication Systems (ICCWCS-2016)*, December, 560–562. https://www.researchgate.net/publication/311900279_Predictive_Analysis_of_Big_Data_in_Retail_Industry
- Jena, B. (2018). An Approach for Forecast Prediction in Data Analytics Field by Tableau Software. *I.J. Information Engineering and Electronic Business*, 1, 19-26 . <https://doi.org/10.5815/ijieeb.2019.01.03>
- Keim, D. A. (2009). Visualization techniques for mining large databases: a comparison. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 923–937.
- Marr, B. (2016). Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results. *John Wiley and Sons Ltd, West Sussex, United Kingdom*
- Muller, P. (2015). Annual Report on European SMEs 2014/2015 (SMEs start hiring again). ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/performance-review/files/countries-sheets/2012/portugal_en.pdf
- Roden, S., Nucciarelli, A., Li, F., & Graham , G. (2017). Big data and the transformation of operations models: a framework and a new research agenda. *Production Planning & Control*, 28(11-12), 929-944. <https://doi.org/10.1080/09537287.2017.1336792>

- Subrahmanyam, D. V. R., V, R. P. M., Sahrudh, D. V, Kids, S., Bank, B. A., Doddaballapur, N., Floor, F., Kids, S., & Bank, B. A. (2020). *Developments in Retail Industry Through Social Media Integration With*. 7(09), 2813–2818.
- Sultan, M. F., Jabeen, M., & Mannan, M. A. (2020). Sentiment Analysis through Big Data in online Retail Industry: A Conceptual Quantitative Study on linkage of Big-Data and Assortment Proactive of Online Retailers. *KIET Journal of Computing and Information Sciences*. <https://doi.org/10.51153/kjcis.v3i2.47>
- Tableau. (2022). Business Dashboards. Tableau. <https://www.tableau.com/solutions/business-dashboards>
- Walmart Labs. (2011). *Goodbye, Kosmix. Hello, WalmartLabs*. Walmart Labs Blog. <http://walmartlabs.blogspot.com/2011/05/goodbye-kosmix-hello-walmartlabs.html>
- Walmart. (2021). *About*. Walmart. corporate.walmart.com/our-story
- Wissel, K. M., & DeLuca, L. (2018). Telling the Story of a Collection with Visualizations: A Case Study. *Collection Management*, 43(4), 264-275. <https://doi.org/10.1080/01462679.2018.1524319>
- Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, 25(1), 17–41.