**Sales Analysis**

**Uncovering Profitable Strategies for a US Superstore**

**Data Management and Visualization**

**BY**

**OPEYEMI OMENAI**

# ABSTRACT

The purpose of this research was to provide useful information to the upper management of a U.S. superstore in their quest to maximize sales and profits. The study used exploratory data analysis (EDA) to look at the store's performance from a number of angles and determine what factors had the greatest impact on revenue. The key areas of inquiry were sales and profits by sector, sales and profits by region, and overall profitability. The most and least profitable product categories were identified to begin the analysis of the store's income streams. Using EDA methods, the most profitable product categories and those that could be rethought or improved upon were identified. Following this, a segment analysis was performed to identify the most important client characteristics, such as age, income, and occupation. This information was vital in identifying the profile of the megastore's most valuable clientele. In addition, it exposed fresh avenues for business development and income enhancement that could benefit from some well-placed advertising. In addition, a thorough geographical analysis was conducted to learn how location affects business outcomes. Another important observation that was made was the high-potential areas for growth in both top and bottom lines by analyzing historical sales data, consumer tastes, and market conditions. Based on the findings, the megastore allocated its marketing resources where they would have the greatest impact. The grocery store may benefit from the EDA's suggestions for what to stock, how to market it, and where to devote scarce resources. The results of these kind of studies are frequently used by superstores to make adjustments or even completely eliminate lines based on the items, customers, and regions that have the greatest impact on earnings. The study's eventual goal was to give the megastore concrete suggestions for improving its bottom line and standing in the US marketplace.

**Table of Contents**

[ABSTRACT 2](#_Toc139464465)

[Introduction 4](#_Toc139464466)

[Future works and Business Insights 5](#_Toc139464467)

[Literature Review 7](#_Toc139464468)

[Dataset 9](#_Toc139464469)

[Data exploration 10](#_Toc139464470)

[Data Preparation 12](#_Toc139464471)

[Data Manipulation 13](#_Toc139464472)

[Analysis 14](#_Toc139464473)

[Exploratory data analysis 14](#_Toc139464474)

[Data Visualization 15](#_Toc139464475)

[Discussions and Recommendations 26](#_Toc139464476)

[Ethical Considerations 27](#_Toc139464477)

[Conclusion 27](#_Toc139464478)

[References 29](#_Toc139464479)

# Introduction

Businesses in today's retail environment need to be adaptable and innovative to survive. The Superstore Corporation is a prominent player in the market, therefore it is well aware of the importance of spotting profitable sales trends. Insights gained through the use of Exploratory Data Analysis (EDA) and visualizations are applied to sales data from a US-based superstore chain in this study.

The Superstore Group carries a large variety of products across many different categories to satisfy the needs of their customers. The store sells everything from electronics and home appliances to furnishings, office supplies, and even toiletries and cosmetics, giving customers a one-stop shop for all their shopping needs. The Superstore Company is dedicated to providing its customers with a convenient and inexpensive one-stop shop for a wide range of high-quality products. To better serve its customers, Superstore Corporation employs rigorous analytic methods. By examining sales data, businesses can gain insight into their clientele, shipping patterns, top-selling items, geographic reach, category efficiency, and more. This information will help the Superstore Corporation boost its sales, services, and overall customer happiness.

This study will help The Superstore Company improve their methods for increasing profits and expanding their business. Understanding consumer purchasing patterns allows for better stock management, more efficient marketing, and the discovery of new avenues for expansion. These strategic decisions will help the Superstore Company in the long run by allowing them to better serve their consumers and compete successfully in the retail market.

A thorough analysis of Superstore Company's sales data will be presented and discussed below. The analysis of consumer profiles, shipping patterns, best-sellers, geographical reach, category success, discount efficacy, and other critical data is central to our mission to assist the Superstore Company in growing and becoming more profitable.

## Future works and Business Insights

Several crucial business insights can be gleaned from answering the research questions presented here, which revolve around analyzing sales at a US-based megastore.

* Ship Mode Preferences: The superstore can improve its logistics and supply chain by studying the shipping preferences of its many client segments (home office, corporate, and consumer). Knowing what its customers want helps businesses improve their services and make shipping easier for everyone involved.
* Top Customers and Purchase Patterns: The superstore may learn more about its most loyal and valuable consumers by analyzing their buying habits and the frequency with which they shop. With this information, businesses may boost client retention and acquisition through focused marketing, loyalty programs, and customized offers.
* Popular Sub-Categories for Each Segment: Analyzing Revenue Generation Within Sub-Categories Revenue generation within sub-categories for various consumer segments provides insights into the most desirable product categories. Insights like these allow for more precise inventory management, targeted advertising, and specialized product offerings.
* Profitability and Discount Analysis: Analysis of sales, profits, and discount percentages for a number of product categories allows for an accurate assessment of how discounts affect total profitability. This research is useful for determining which types of goods can benefit most from price reductions in order to increase sales without compromising profits. It lets the megastore fine-tune its markdown policies across product lines.
* Potential Markets and Regional Concentrations: The superstore can maximize its marketing and sales efforts by concentrating on the regions where the concentration of clients is greatest. The organization can find new growth prospects and exploit untapped markets by analyzing sales performance across areas.

With this information, the megastore can better manage stock, narrow in on its preferred clientele, fine-tune its pricing, and distribute its limited resources more wisely. The corporation may use this data to improve efficiency, boost profits, and fortify its standing in the cutthroat retail industry.

# Literature Review

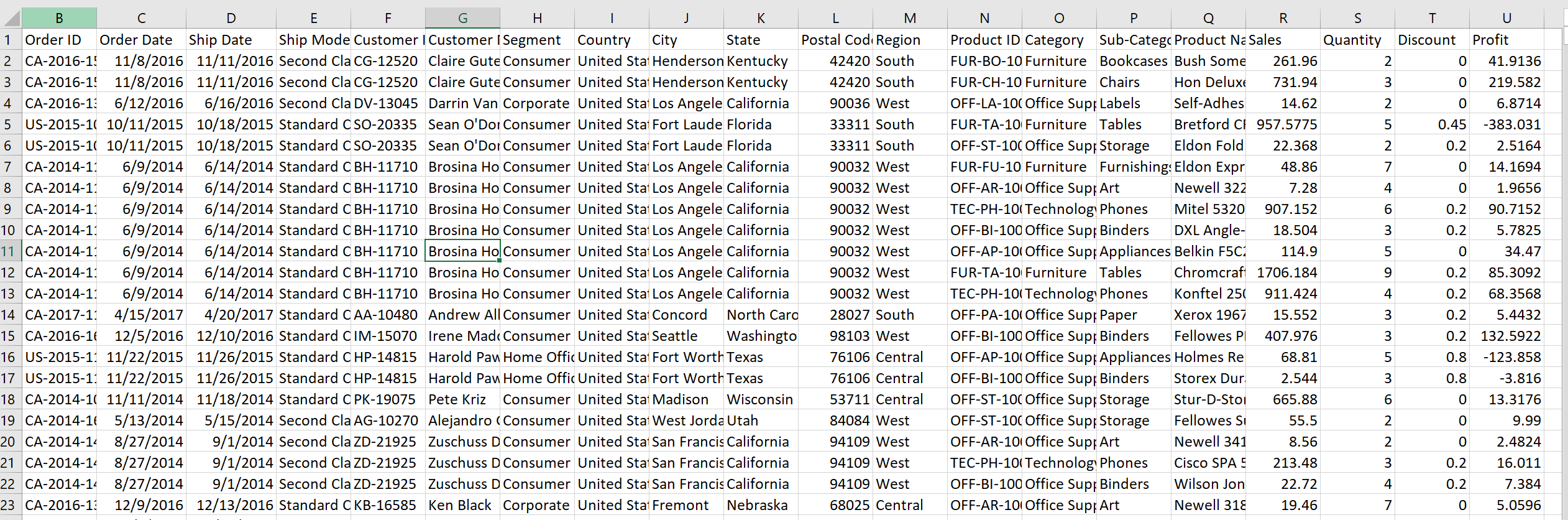
For any retail business to succeed, knowing what factors affect sales and profitability is essential Dinesh, et al. (2021). Research into sales analysis and profitability evaluation has been conducted to identify the most crucial factors in a company's success and the most effective means of achieving it. A variety of products and management of product categories have received a lot of focus over the years. Gupta, et al. (2020) found that well-managed category assortments boost customer satisfaction and sales, whereas Human, et al. (2020) highlight the significance of good category management in maximizing revenue and profits. By applying data analytics to predict product demand, fine-tune product assortments, and fine-tune inventory levels, retailers can potentially maximize profitability (Saleem, et al., 2021).

Segmentation and targeting strategies also have significant effects on sales and profitability. Customer lifetime value analysis is emphasized by Grossmann, et at. (2019) for its usefulness in identifying lucrative client segments and allocating resources accordingly. By categorizing clients according to their demographics, psychographics, or buying habits, retailers may better meet the needs and preferences of their varied clientele. A company's ability to raise or reduce sales and profits is directly tied to its pricing strategy. By implementing dynamic pricing strategies like personalized pricing and value-based pricing, retailers can increase both sales and profits Pradhan, et al. (2021). When managed properly, discounts and promotions can increase sales and customer loyalty Memarpour, et al. (2021), but when not, they can ‘eat’ into profits. Supply chain management and logistical operations are crucial to generating sales and profitability. Inventory management, distribution network architecture, and order fulfillment are highlighted as critical to maximizing sales performance by AboElHamd, et al. (2020). Companies can streamline operations and reduce lead times which can lead to customer satisfaction, cost reduction, and higher profitability.

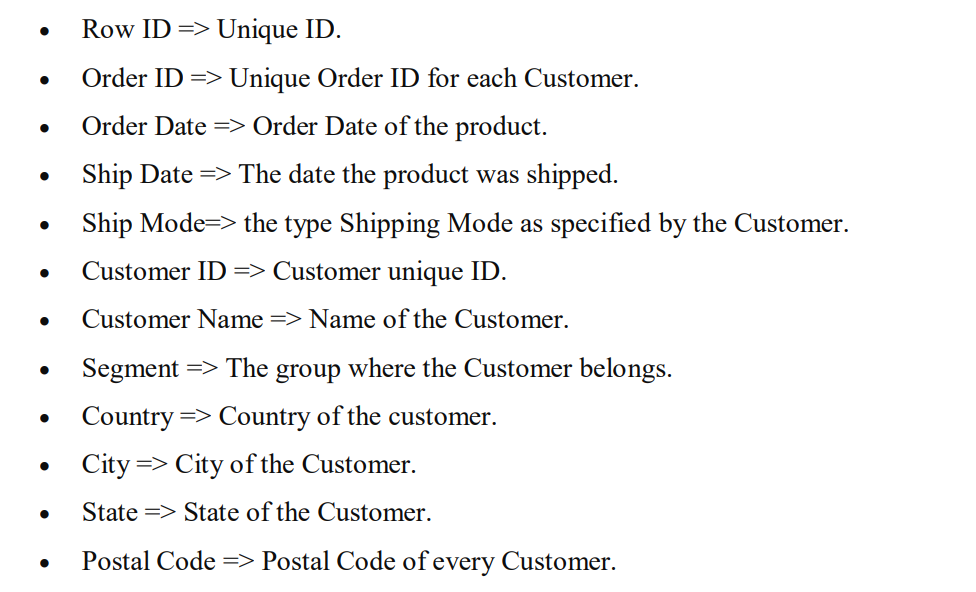
Marketing and customer relationship management (CRM) are crucial to a company's success. Customer retention and acquisition are key to maximizing profits in the long run, as noted by Rafiki, et al. (2019). Using customer information for targeted advertising and tailoring your approach to each customer's needs is a surefire way to increase loyalty, repeat business, and bottom line results. In order to get insight into consumer actions and organizational consequences, Exploratory Data Analysis (EDA) is widely employed. To better understand customer preferences, product efficacy, and market fluctuations, businesses can use EDA technologies including data visualization, correlation analysis, and clustering Westland, et al. (2020). Data mining and machine learning techniques are applied to large datasets to reveal hidden relationships and tendencies.

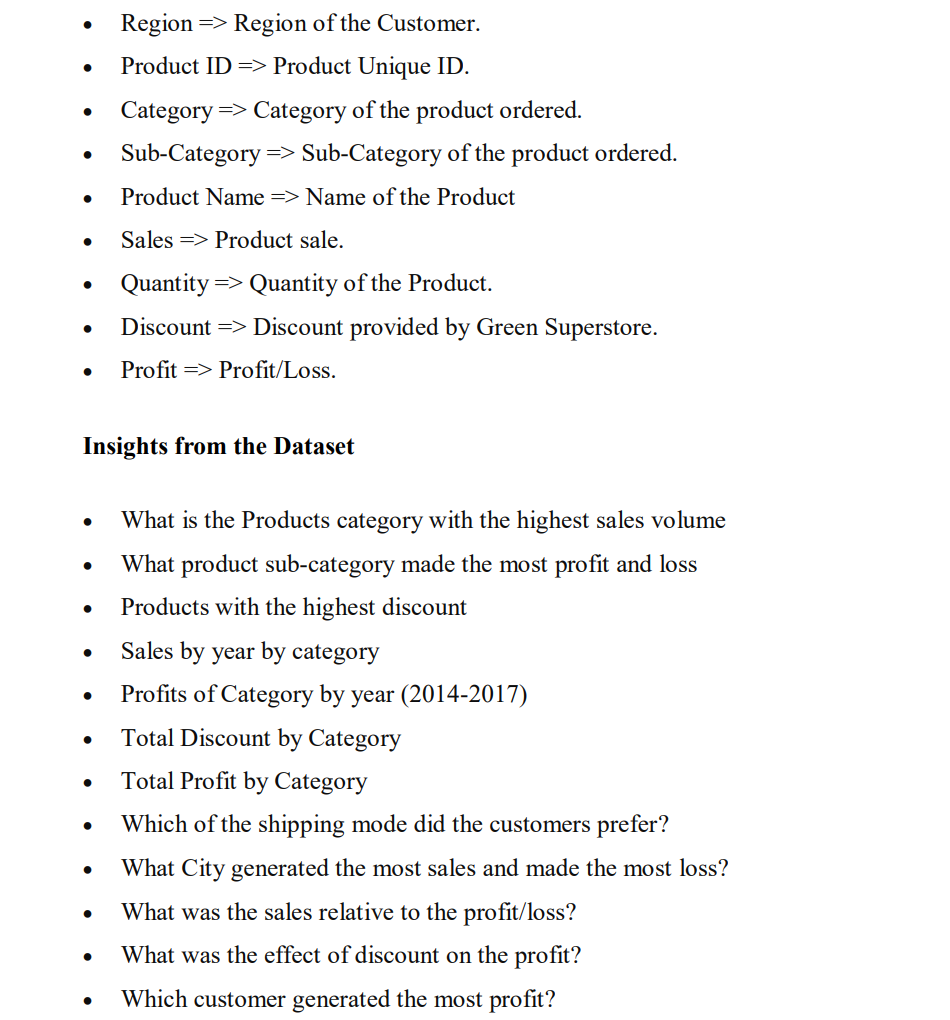
## Dataset

Green Supermarket's revenue and loss statements are available for review from 2014 to 2017. The Green Superstore is a huge retailer that employs a lot of people that need to know a lot about their products, sales, profits, regions, and types of customers. The American Green Superstore may be found in all four of the country's major regions, with stores in 49 of the 50 states. They have an extensive selection of office supplies, furniture, and technological tools suitable for both professional and domestic settings.



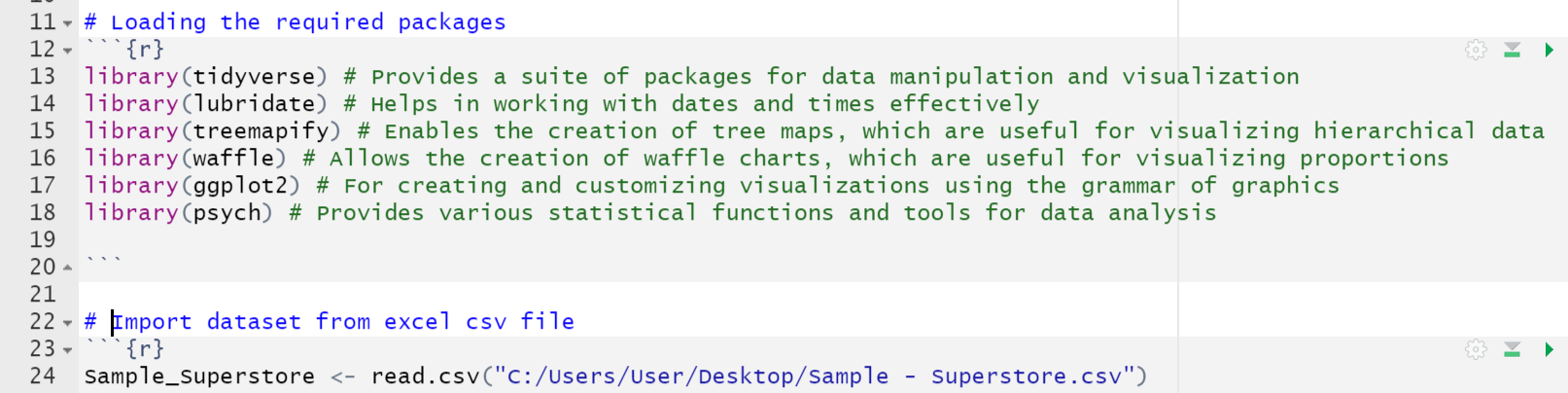
The dataset contains a total of 21 variables with 9994 observations. The description of each variable and the dataset insights are provided as follows:





# Data exploration

The exploration and analysis of the dataset commenced by first installing and loading all the R packages that are required for the analysis.



The packages provide access to a comprehensive set of tools for manipulating, presenting, and statistically analyzing data. The read.csv function in R, was then used to import the dataset. The imported dataset made it possible for appropriate analyses.

The head function is used to explore the first six rows of the dataset. This provided an overview of the data and an appreciation for its organization.

***Table 1: Dataset***

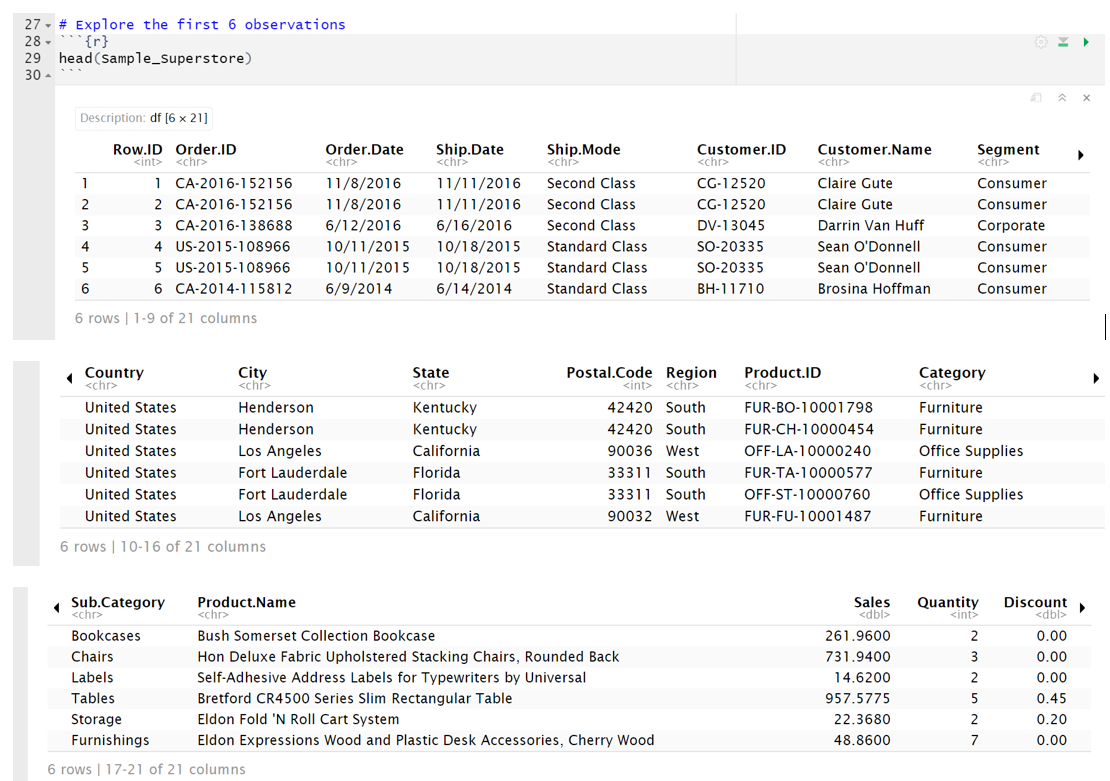
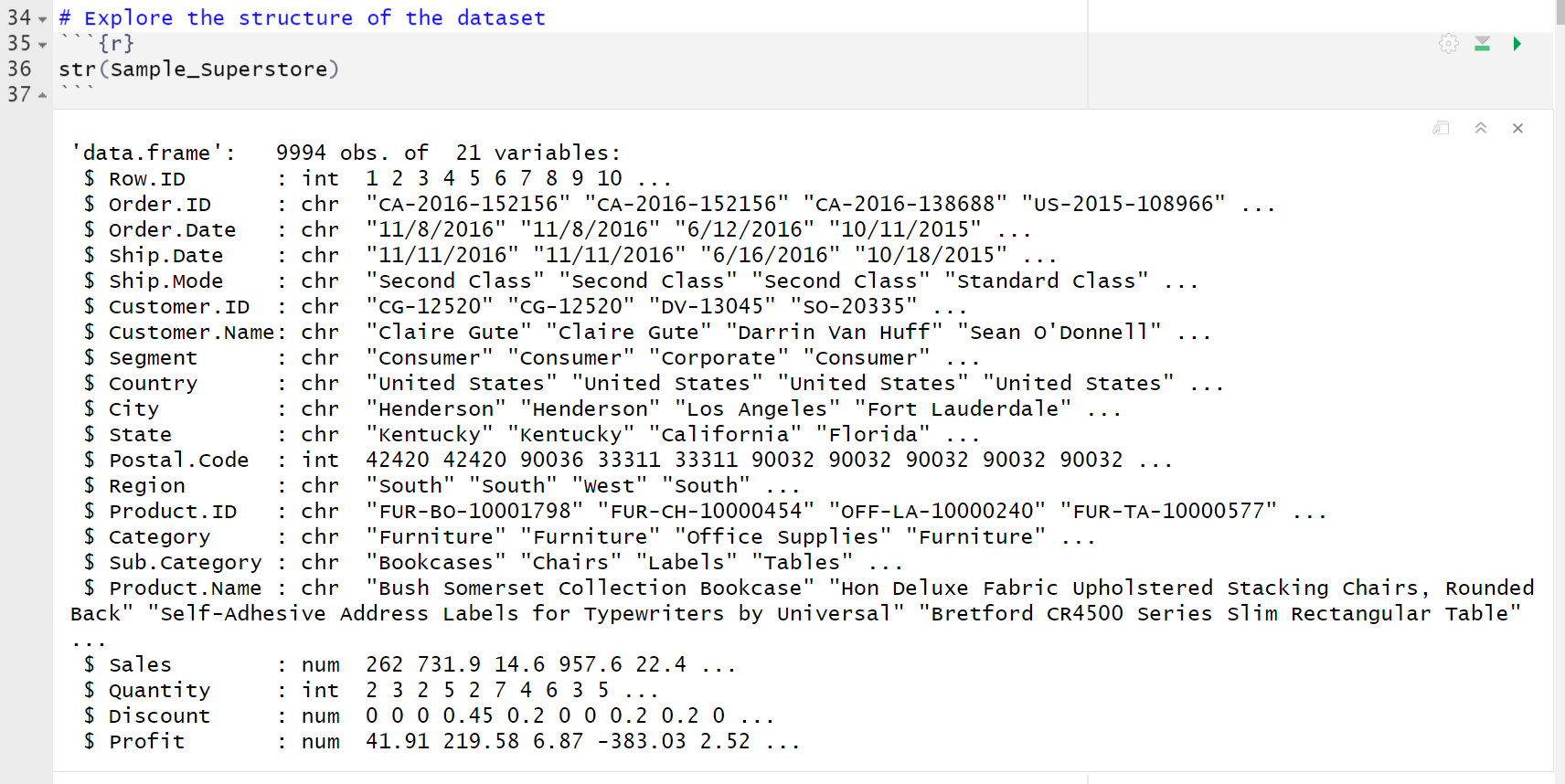


Table 1 shows the all the variables and the first 6 observations in each dataset.

To investigate the structure of the dataset, the string function was used. The names, data types, and initial values of the dataset's variables were all made available by this function. 

The str() was used to explore the data types that are present in the dataset. Based on the above output the data contains integer type variables, character variables and numbers. The output explains more about the variables in the dataset and their properties, including whether they are numeric or categorical, by inspecting the dataset's structure.

## Data Preparation

The data preparation focused on identifying and removing any missing information from the dataset. The total number of missing values in each column was determined using the colSums function in conjunction with the is.na function. This is a vital procedure for checking the integrity of the data and deciding how to deal with any gaps in the information.



The output as can be seen from the screenshot above; the dataset did not contain any missing value as well as any duplicated value.

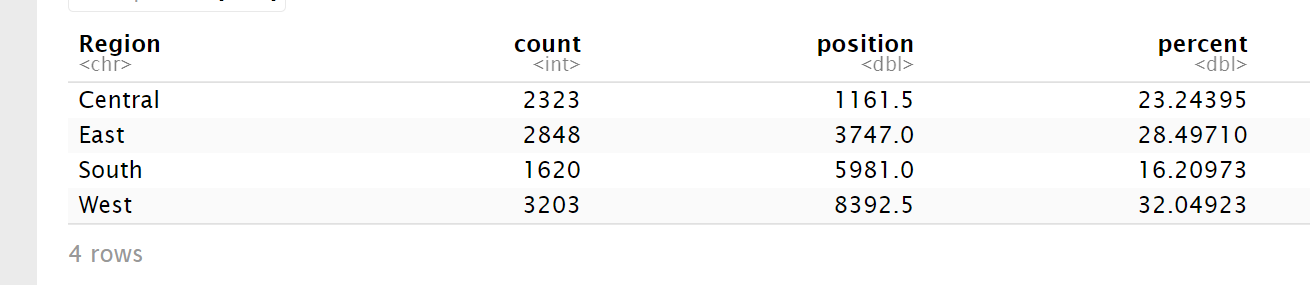
## Data Manipulation

Data manipulation involved creating new variables and renaming different variables for further analysis. Functions such as slice(), arrange(), summarize(), group\_by() and mutate functions were used.



The Mutate function was also used to create new variables



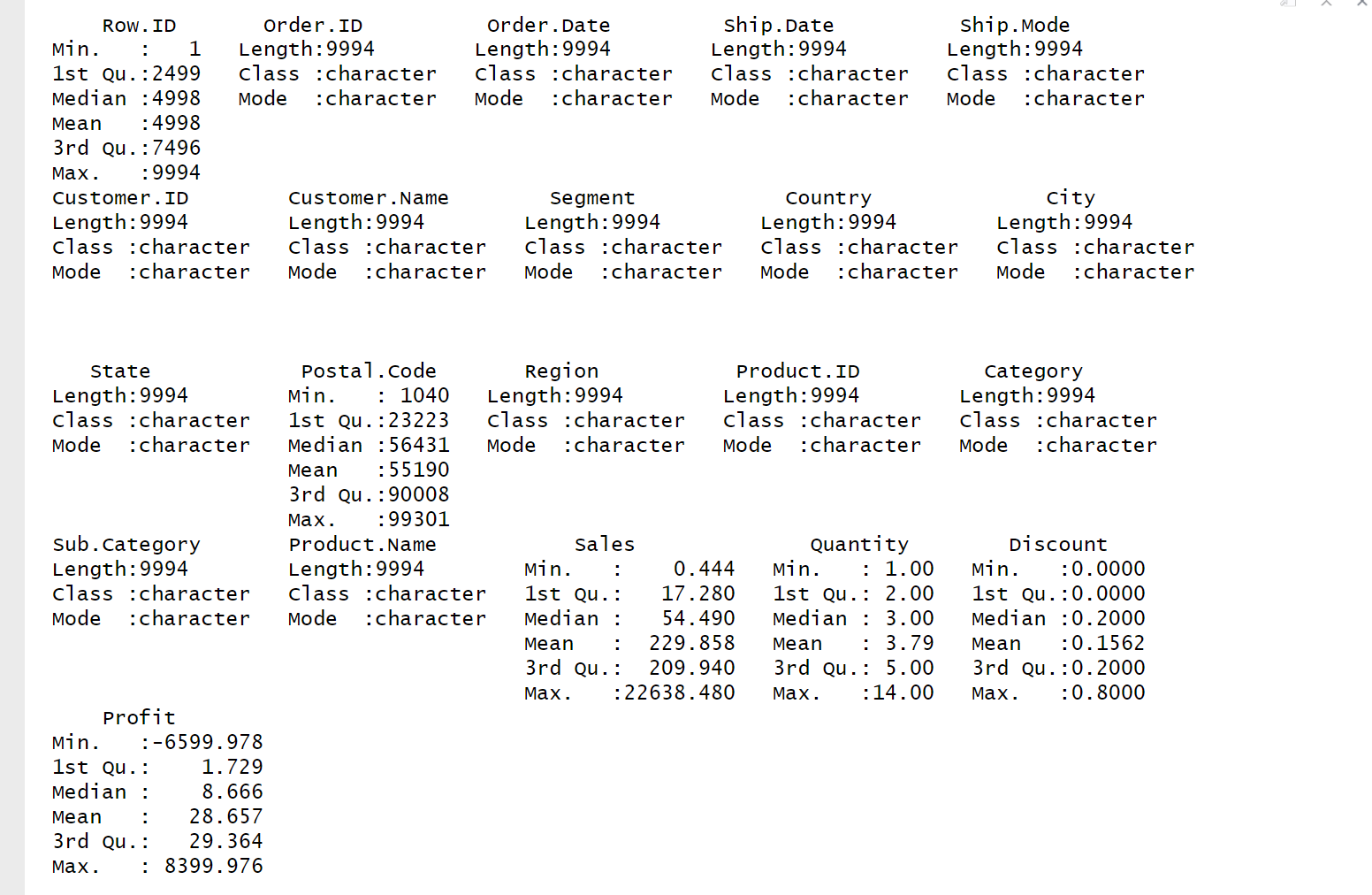


# Analysis

## Exploratory data analysis

This analysis portion includes some descriptive statistics that summarize the key factors in the data. The mean, the median, and the maximum and the minimum value of the data were all explored in this analysis as shown table 2.

***Table 2: Statistical summaries***



The total number of observations in all the variables was 9994. The mean or the average of the sales, quantity, discount and profit was 229.858, 3.79, 0.1562 and 28.657 respectively. The median values and the maximum and the minimum values are also displayed. The mean and the median values are important for

# Data Visualization

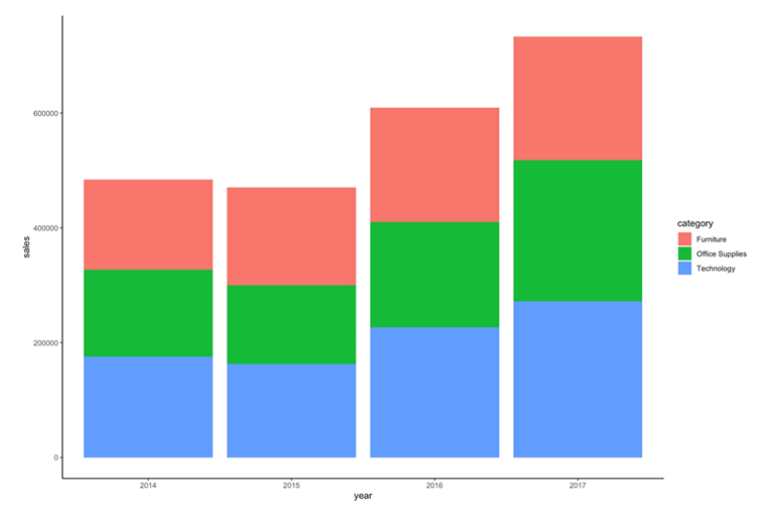
After the exploration of the descriptive summaries of the dataset, data visualization was conducted to explore more findings on the business insights. These helped to explore the performance and the other attributes that were present in the dataset.

**Analysis of sales value of categories by year**

This gives a clear understanding of how the products in the category have performed from

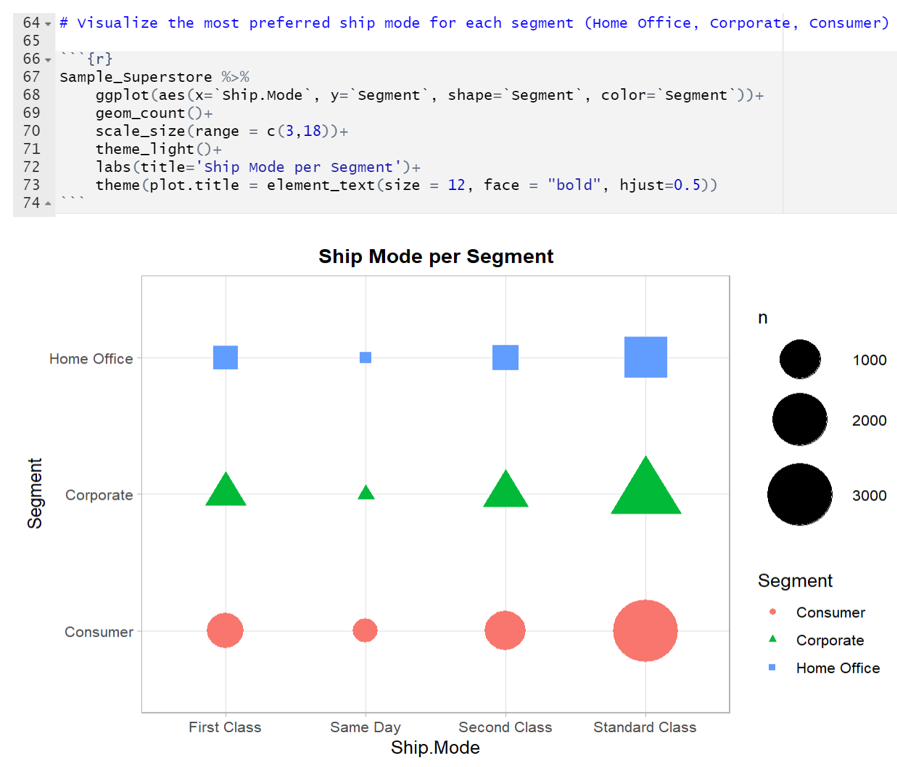
2014-2017. There was a slight decrease in 2015 from 2014 and then the sales value increased

from 2016-2017.



**Ship mode for each segment**

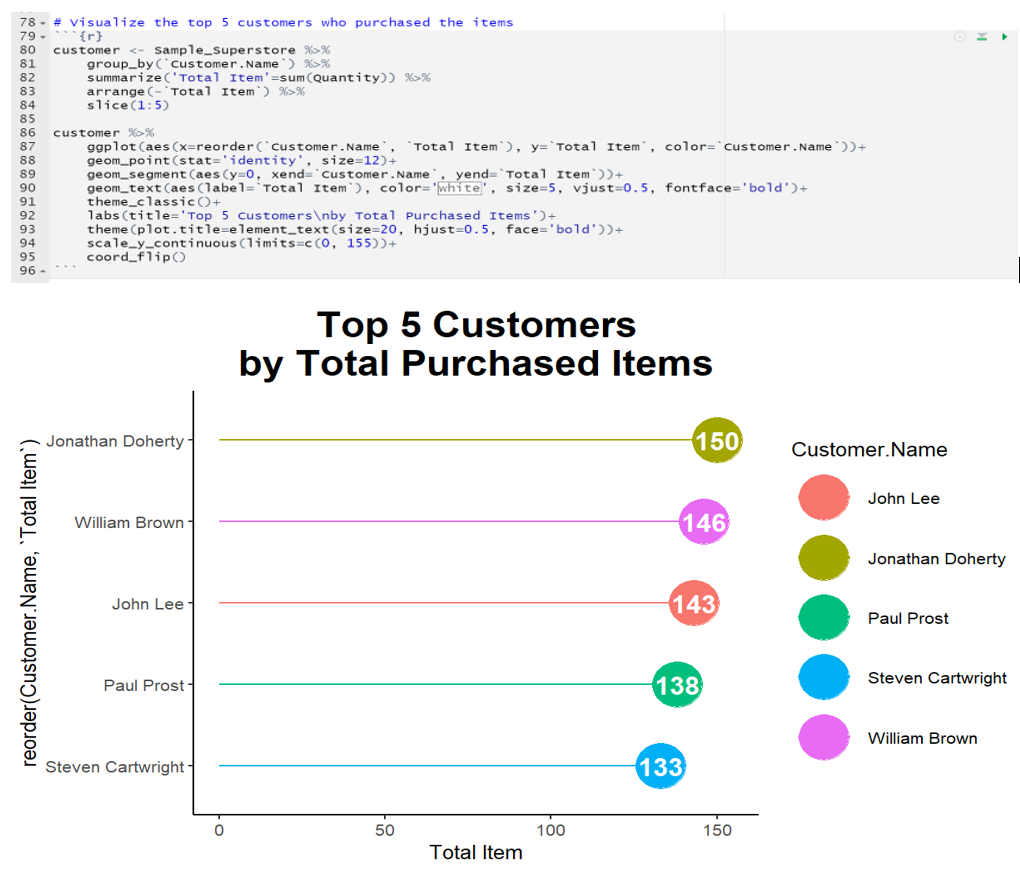
The data visualization commenced on wanting to investigate the type of ship mode for each segment.



The graph illustrates that the basic ship configuration is favored by every market. This suggests that customers tend to choose the most common and trustworthy shipping option. Additionally, all components of the ship have the same minimum ship mode option on the same day, demonstrating consistency in the shipping method.

**Top Customer**

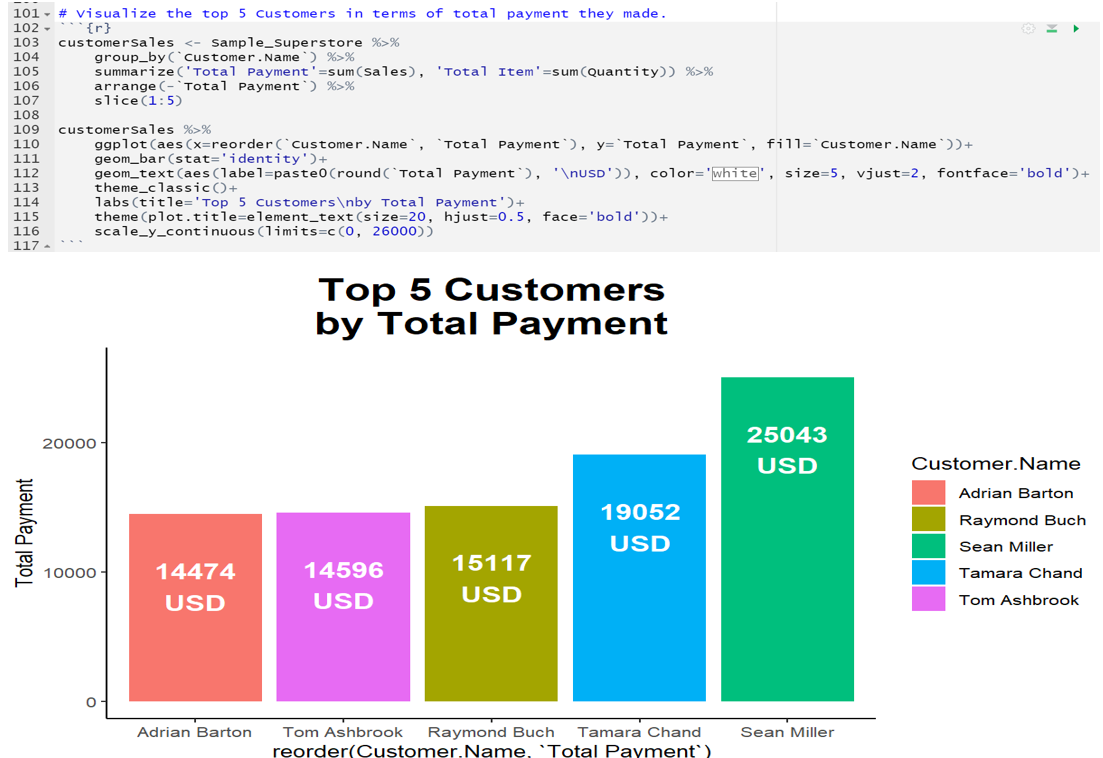
The top purchasing customers were also analyzed as follows.



Jonathan Doherty stands out as the customer with the highest number of purchases at the store, with a count of 150. This implies that Jonathan is a loyal and frequent customer who contributes significantly to the store's sales.

**Highest Paying Customer**

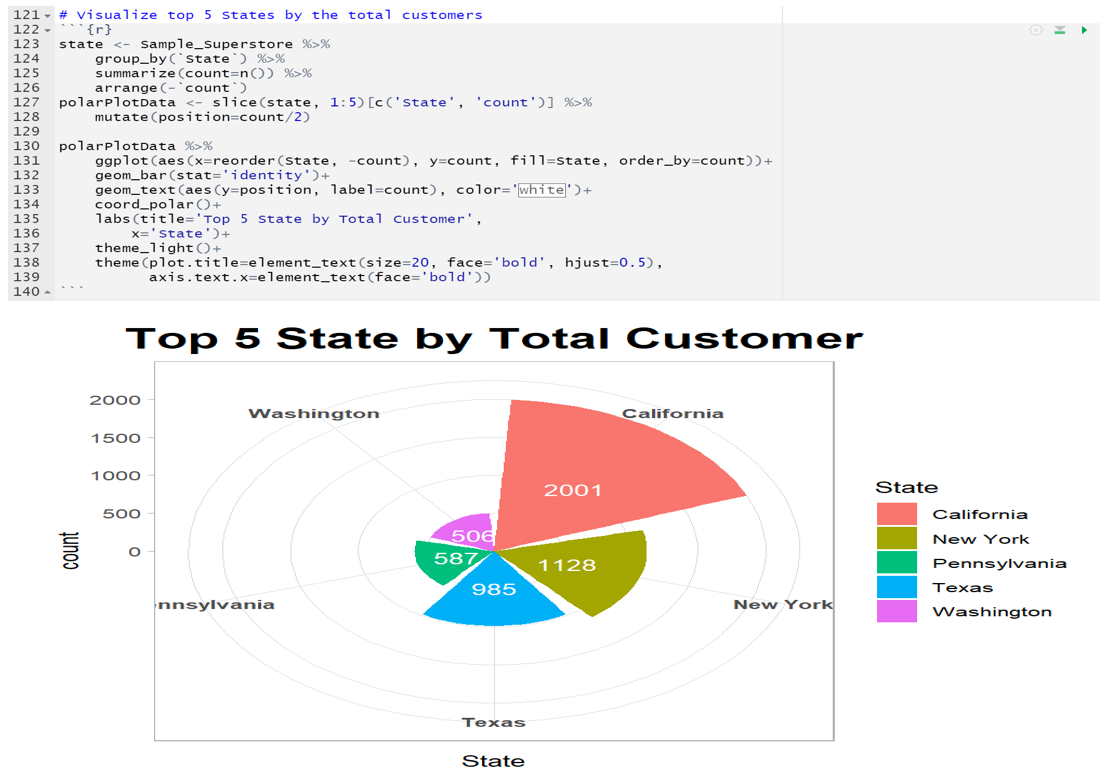
Among the customers, there were customers who were paying higher than others. This was analyzed using a bar graph.



According to the analysis, Sean Miller holds the top position among the five customers who paid the most, spending over $25,000. This amount is nearly double the payment made by the customers in fifth place. Sean's substantial spending demonstrates his high-value purchases and significant contribution to the store's revenue.

**Consumer Base by Location**

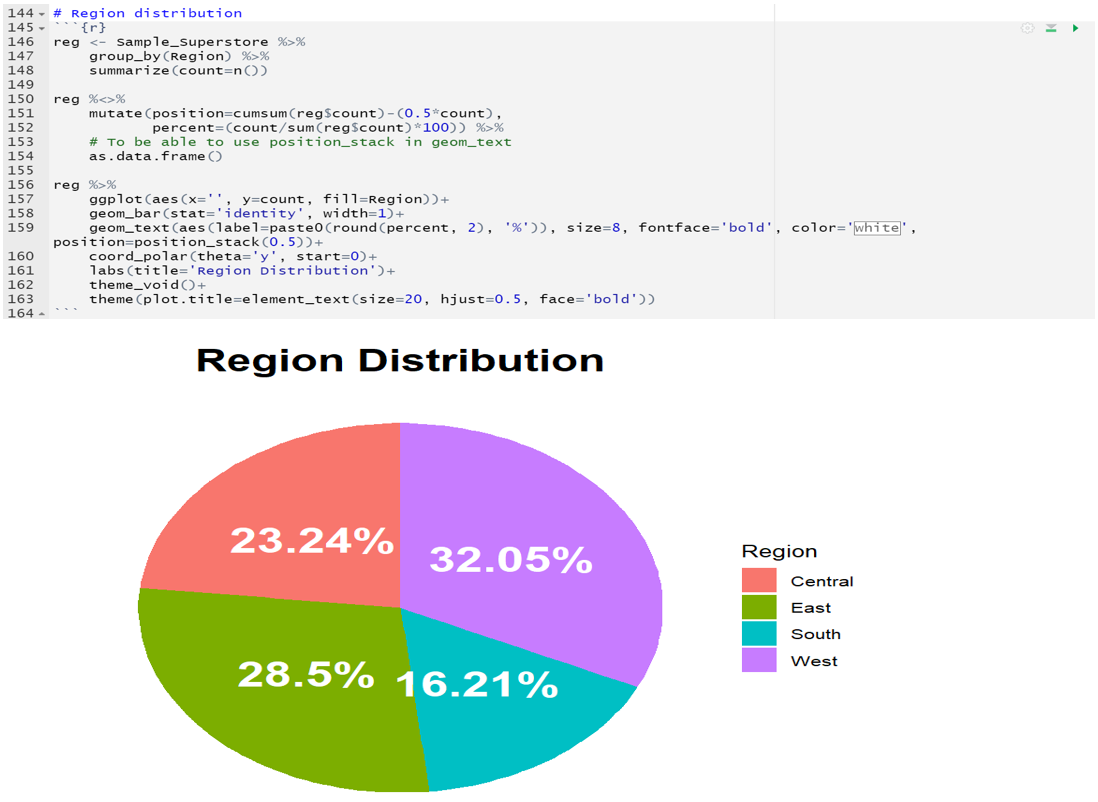
The consumer’s locations were also analyzed to see which state had the highest number of customers for the product.



California has the largest consumer base in 2001, being more than four times larger than Washington, DC. This indicates that California holds a significant market share and plays a crucial role in driving sales and revenue for the store. The graph highlights the dominance of the Western United States in terms of sales.

**Regional sales distribution**

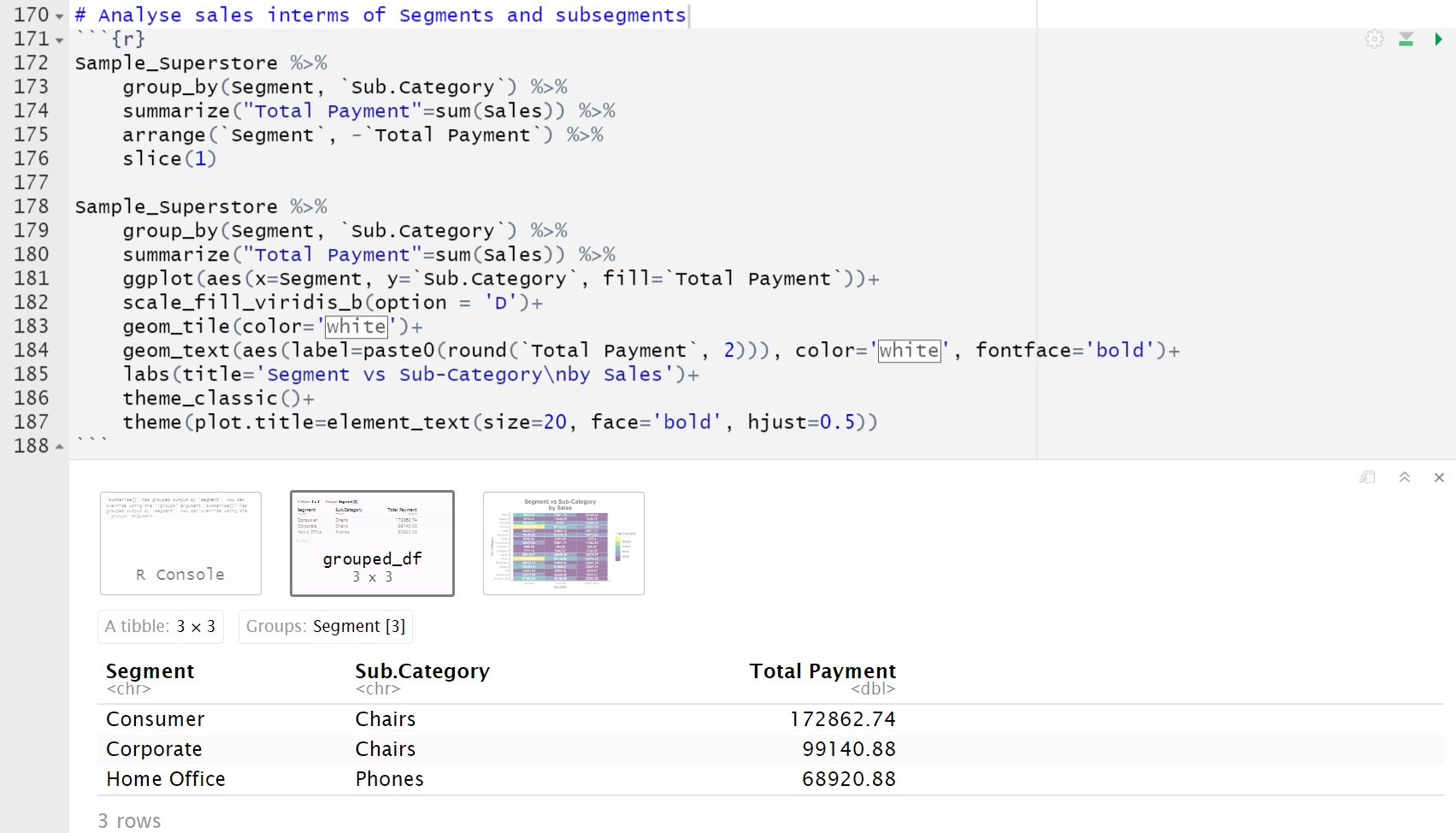
The sales distribution was investigated using a pie chart.

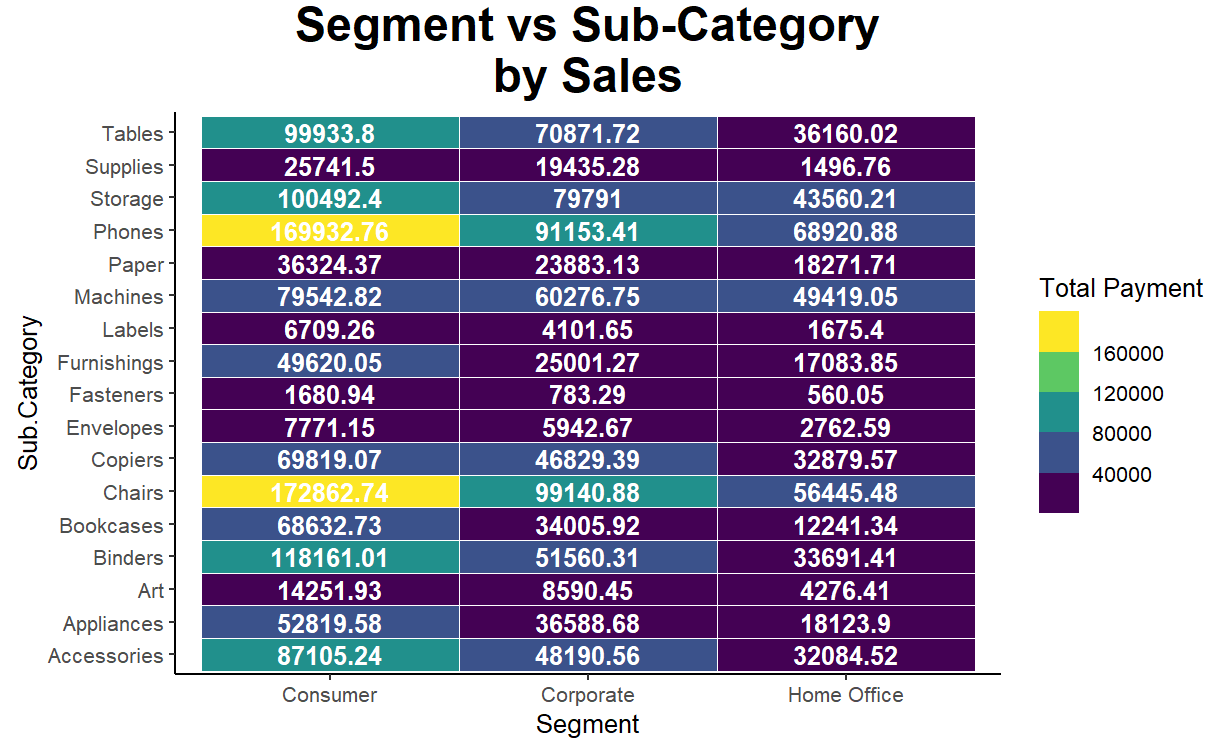


Based on the results from the pie chat above, it is clear that majority of sales come from the Western United States.

**Popular Categories**

The categories of the products such as tables, supplies, storage, phones, papers, machines etc. were all visualized using a table and the colored graph below.

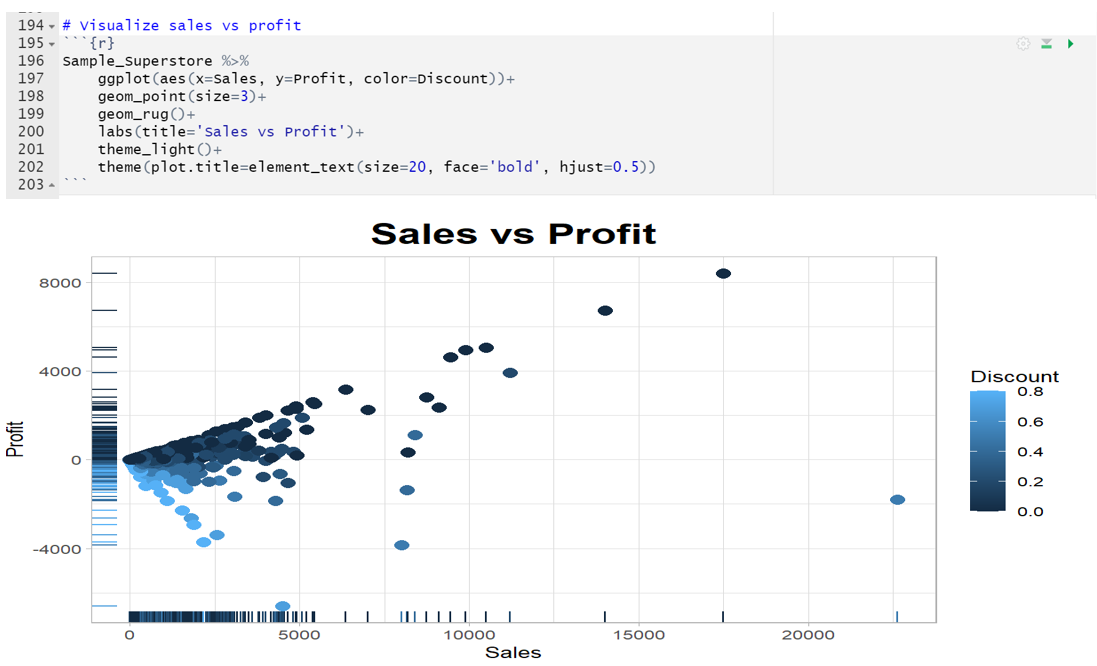




The graph and tables reveals that chairs and telephones are the dominant categories in their respective product segments. This suggests a higher demand for these items compared to others. Additionally, the retail industry incurs the highest costs, indicating that expenses related to running the store, such as inventory and operations, are substantial.

**Sales and Profits**

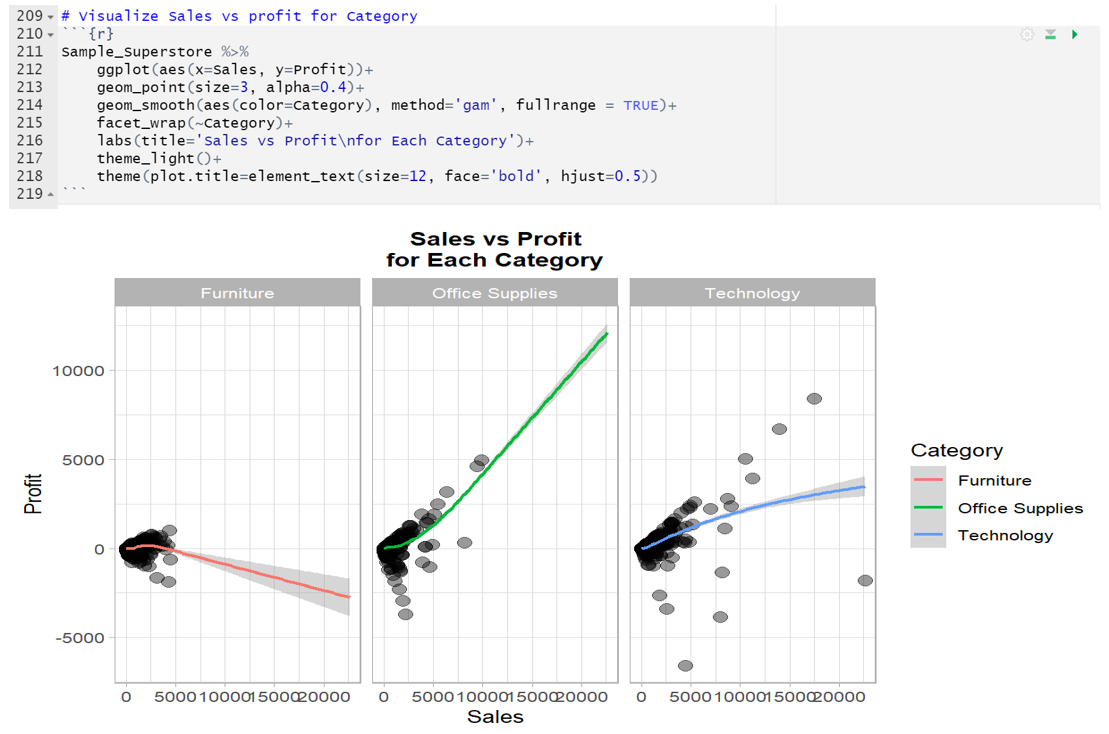
The sales profits were visualized using a scatter plot to check the profit margins that were realized in selling the products.



The scatter plot above indicates a positive relationship between sales and profits. When sales are high, profits tend to follow suit, suggesting that increased revenue contributes to higher profitability. However, according to the graph, it is worth noting that the discount offered by the store significantly impacts earnings. The more the discount, the lower the store's profitability, as the reduced prices affect profit margins.

**Demand for Technological and Supply Goods**

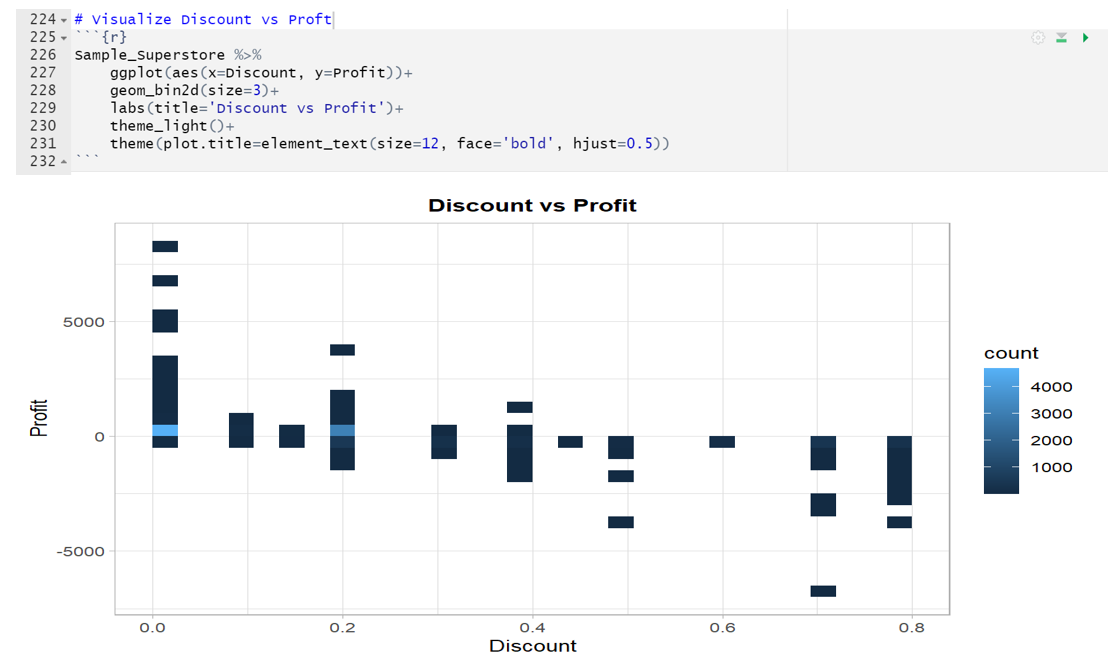
The trend line in the graph suggests an increasing demand for technological and supply goods for the workplace.



The growth may be driven by targeted advertisements aimed at specific age ranges. However, with rising demand, the likelihood of cheaper furnishings also increases, potentially due to market competition and pricing strategies.

**Impact of Store Discount and Profits**

The store discount and profits were also visualized to view the impact they had on the sales.



The graph highlights that the store's discount has a noticeable effect on sales. Sales tend to be higher even without a price cut, indicating that customers make numerous purchases regardless of discounts. However, it is evident that the store's profitability is impacted by the discount offered. The higher the discount, the greater the reduction in store earnings.

**Binder Sales and Discount**

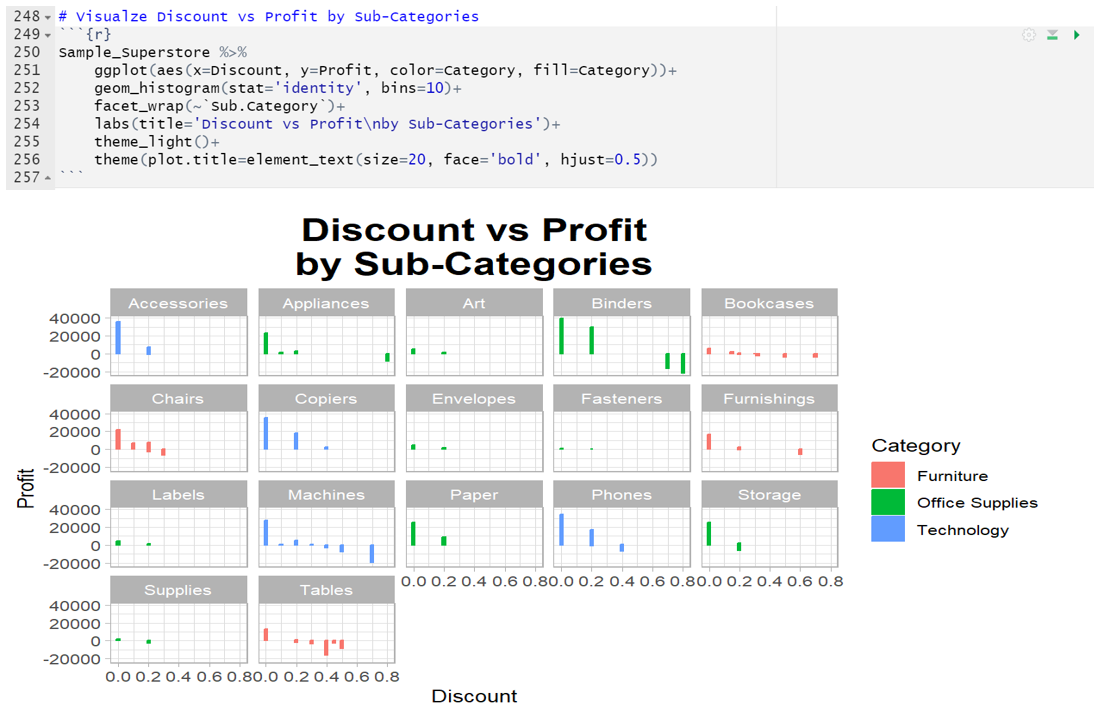
The relationship between the sales and the discount was also investigated using a scatter plot.



Based on the scatter plot, it is evident that, without a discount, binders emerge as the most profitable item for the store. However, when an 80% off coupon is applied to binders, the situation changes dramatically. The store incurs a loss of over $20,000 due to the significant discount offered on binders. This highlights the impact of discounts on the store's financial performance.

**Discount vs. profit Relationship by each Sub-category**

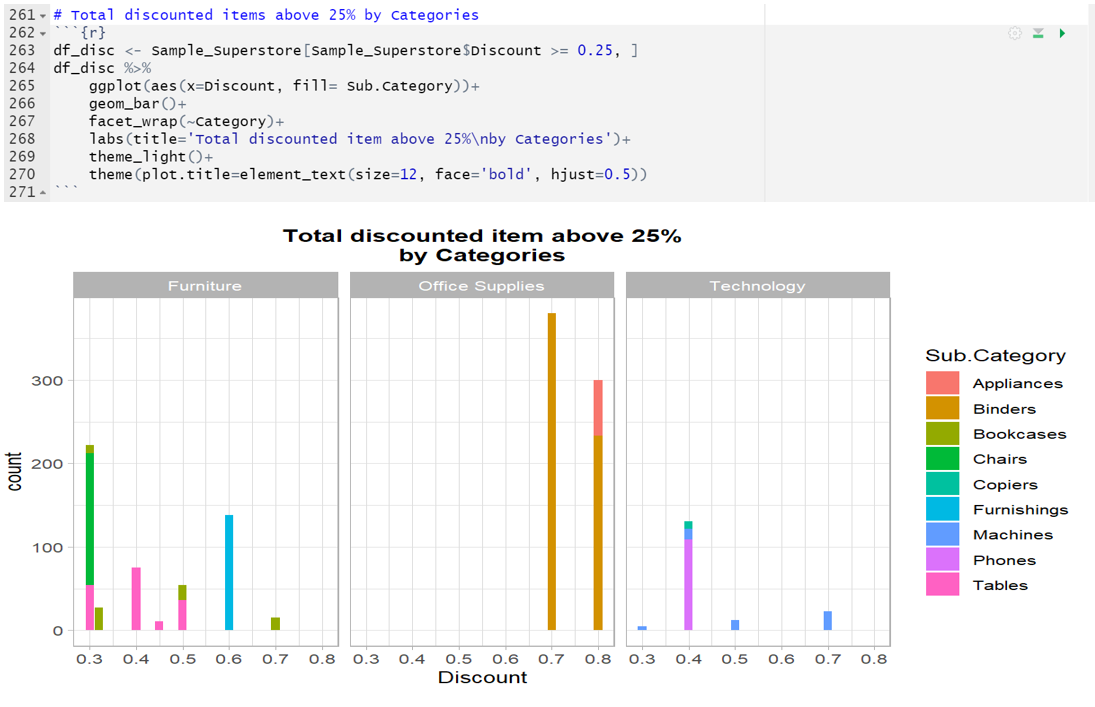
The discount and the profits relationship were also examined in each subcategory.



According to the graph, office supplies were more dominant than the rest of the categories.

**Office Supplies Discounts**

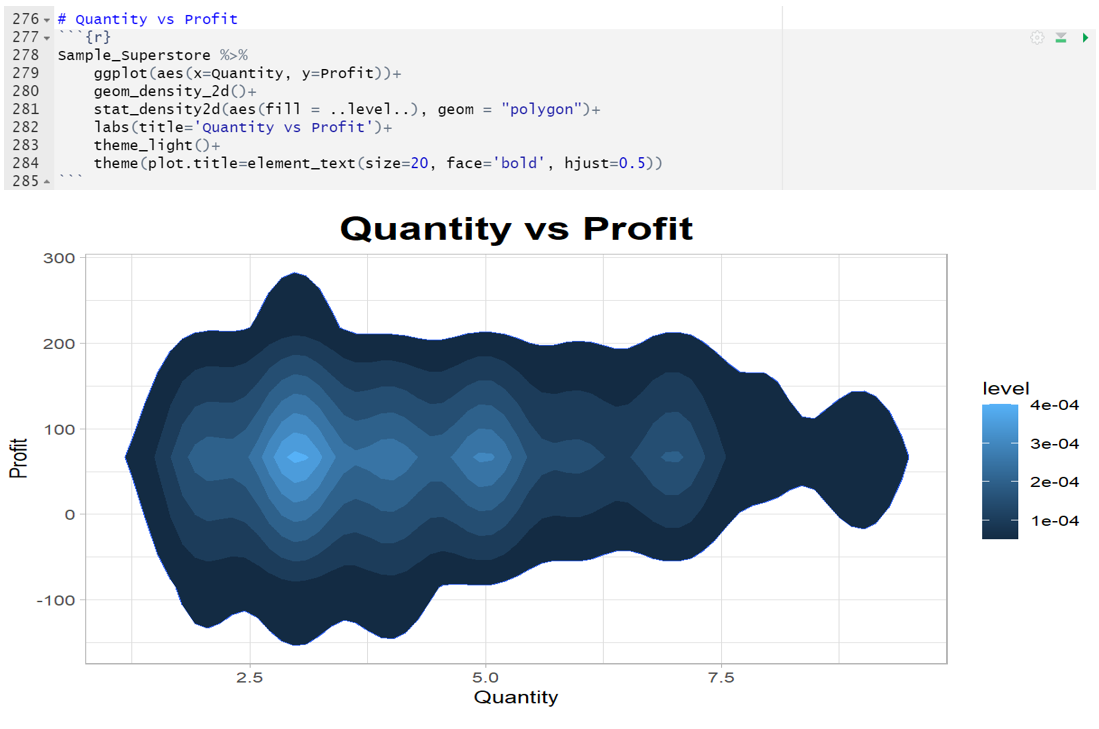
The graph below indicates that a significant number of products in the office supplies section are frequently on sale at steep discounts.



Over 350 products are discounted by 70%, and another 300 products are discounted by 80%. This suggests that the store often offers promotions and sales on office supplies, potentially to attract customers and increase sales in this category.

**Quantity vs. Profit Relationship**

The more merchandise is sold, the lower the store's profit will be. A scatter plot was plotted to show the relationship between quantity and sales.



The graph indicates that as the quantity of merchandise sold increases, the store's profit tends to decrease. This observation implies that while selling more items may lead to higher revenue, it may not always translate to higher profitability, as factors like cost and discounts play a significant role in determining profit margins.

## Discussions and Recommendations

In order to aid in strategic decision-making, this study aimed to evaluate sales data from a US-based superstore. Customers' demographics, shipping preferences, best-selling items, regional reach, category success, discount efficacy, and other factors were all taken into account. The study used exploratory data analysis (EDA) and data visualization to unearth key insights that can direct sales strategies and boost profits.

It was discovered that the most popular and dependable shipping option, the basic ship configuration, was the favored technique across all consumer categories. This knowledge will aid the megastore in streamlining its shipping operations, which in turn will increase consumer happiness. The study showed the significance of retaining and expanding business with the most valued clients, who were determined based on their purchase frequency and total spend. To keep them as customers and increase revenue, it is possible to employ personalized marketing and CRM tactics. Furthermore, California emerged as a prominent market driver, highlighting the importance of concentrated efforts in this area.

Chairs and telephones, according to the analysis of popular categories, have the largest shares of their respective markets, indicating great demand. Sales and profits can be maximized by using this data to inform inventory management and advertising decisions. The study also highlighted the influence of discounts on earnings and the connection between sales and profits. Finding the sweet spot between sales volume and profitability requires careful discount strategy. More specifically, the research examined the demand for technology and supply items, finding evidence of an upward tendency that may have been driven by focused advertising and competitive price techniques. The examination of the discount and profit relationship for binders showed how crucial discounts are to financial performance, highlighting the importance of prudent discount management for sustaining profits.

These results shed light on the preferences of various demographics, product effectiveness, geographical tendencies, and the effect of price cuts on revenue and profit. With this information in hand, the megastore will be able to fine-tune its approach to sales, optimize its product assortment, better its shipping operations, and personalize its marketing efforts to boost customer satisfaction, sales, and profits.

## Ethical Considerations

The use of secondary data in this study meant that there were no primary ethical concerns that needed to be addressed. The US-based superstore dataset, from which the information was culled, is publicly available and can be used in any number of ways for research and analysis. The data was used appropriately and responsibly, in accordance with the data provider's terms and conditions. Without considering privacy problems or potential harm to people, this study just concerned itself with drawing conclusions and making suggestions from the dataset.

## Conclusion

In conclusion, this study utilized exploratory data analysis and visualizations to analyze sales data from a US-based superstore. The findings highlighted key factors influencing sales and profitability, including customer demographics, shipping preferences, top-selling items, geographic reach, category success, and discount efficacy. The preferred shipping method was found to be the basic ship configuration for all customer segments, and top customers were identified based on purchase frequency and spending. California emerged as the state with the largest consumer base, and the Western United States dominated regional sales. Chairs and telephones were identified as popular categories, and the positive relationship between sales and profits was observed, along with the impact of discounts on profitability. These findings provide valuable insights for the superstore to optimize sales strategies, improve customer satisfaction, and drive profitability.

# References

AboElHamd, E., Shamma, H. M., & Saleh, M. (2020). Dynamic programming models for maximizing customer lifetime value: an overview. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 1* (pp. 419-445). Springer International Publishing.

Dinesh, S., & MuniRaju, Y. (2021). Scalability of e-commerce in the COVID-19 era. *International Journal of Research-GRANTHAALAYAH*, *9*(1), 123-128.

Grossmann, M., Brock, C., Hubert, M., & Reimer, T. (2019). The Relevance of Positive Word-of-Mouth Effects on the Customer Lifetime Value-A Replication and Extension in the Context of Start-ups. *SMR-Journal of Service Management Research*, *3*(3), 148-160.

Gupta, A., Mishra, V., & Tandon, A. (2020). Assessment of Shopping Mall Customers’ Experience through Criteria of Attractiveness in Tier-II and Tier-III Cities of India: An Exploratory Study.

Human, G., Ferreira, C. C., Robertson, J., & Whiterspoon, M. (2020). Customer experience, satisfaction and loyalty in business-to-business markets: Does the chain hold across customer segments?. *Acta Commercii*, *20*(1), 1-13.

Memarpour, M., Hassannayebi, E., Fattahi Miab, N., & Farjad, A. (2021). Dynamic allocation of promotional budgets based on maximizing customer equity. *Operational Research*, *21*, 2365-2389.

Pradhan, S., Patel, G., & Priya, P. (2021). Measuring customer lifetime value: application of analytic hierarchy process in determining relative weights of ‘lrfm’. *International Journal of the Analytic Hierarchy Process*, *13*(3).

Saleem, H., Uddin, M. K. S., Habib-ur-Rehman, S., Saleem, S., & Aslam, A. M. (2019). Strategic data driven approach to improve conversion rates and sales performance of e-commerce websites. *International Journal of Scientific & Engineering Research*, *10*(4), 588-593.

Rafiki, A., Hidayat, S. E., & Al Abdul Razzaq, D. (2019). CRM and organizational performance: A survey on telecommunication companies in Kuwait. International Journal of Organizational Analysis, 27(1), 187-205.

Westland, J. C. (2020). Audit analytics: Data science for the accounting profession. Springer Nature.