**MIS581 Capstone**

Krista O’Neill

Colorado State University

MIS581 M5 CT

Dr. Orenthio Goodwin

May 8, 2022

### Contents

[Abstract 3](#_Toc102412636)

[Introduction & Overview 4](#_Toc102412637)

[Research Proposal 4](#_Toc102412638)

[Objectives 5](#_Toc102412639)

[Research Questions & Hypotheses 6](#_Toc102412640)

[Data Analysis for a Department Store: A Literature Review 8](#_Toc102412641)

[Methods / Methodology 10](#_Toc102412642)

[Limitations 13](#_Toc102412643)

[Ethical Considerations 14](#_Toc102412644)

[Findings 15](#_Toc102412645)

[Conclusions 27](#_Toc102412646)

[Recommendations 28](#_Toc102412647)

[Final Conclusion 30](#_Toc102412648)

[References 31](#_Toc102412649)

# Abstract

In this research project, a data analyst explores a dataset recording sales from a superstore over the course of several years. The store sells many products to various clients in several categories. The analyst wishes to explore research questions investigating ways to improve profits, whether that be through customer segment, shipping time, and discounts. The analyst will seek, through descriptive statistic tests as well as predictive and statistical models, to understand if the type of customer purchasing has an effect on profits, explore the impact of longer or shorter shipping times on profits, if any, and whether or not discounting products increases revenue.

# Introduction & Overview

The invention of the superstore brought a new level of convenience to shoppers worldwide. Created as a solution to customers spending time travelling between many various stores, fighting traffic, or struggling to park, superstores have changed the frontier of commerce irrevocably and superstores quickly took over as a tour de force of retail (Golden et al., 2006). The integration of big data into the framework of retail brought about another revolution of efficiency, allowing for a wide-scope vision of retail operations through the lens of data analysis (Bijaksic et al., 2014). This project will analyze a dataset collected from a superstore, uploaded on Kaggle for public consumption, in order to explore business questions that could increase the profitability of the store. A project whose scope includes retail will allow for multiple levels of analysis with many real-life applications, considering the vast nature of commerce and e-commerce in today’s world. Its goal is ultimately to explore avenues of potential profit to the store, as well as identify areas that could use improvement. It will both explore contextual correlations based on the nature of the data, as well as look for previously unidentified patterns that could increase overall sales and profit.

# Research Proposal

This research project will investigate the dataset of a large superstore which specializes in selling furniture, office supplies, and similar home goods. The superstore’s goal, like any store’s goal, is to increase profit; this is a cornerstone of any store (Bijaksic et al., 2014). Another element that should be considered is an attention to customer satisfaction, which in the end is likely to lead to profit. Therefore, the primary research question should be: *Which, if any, variables’ relationships cause either an increase or decrease in profits?* All the hypotheses to be investigated will relate back to the superstore’s information and how its various factors and their correlations bring either growth or decline to revenue.

# Objectives

In order to gain a full understanding of the data, it will be wise to conduct a primary descriptive statistics overview. While it is important to view the quantile statistics of each column to examine mean, median, and standard deviation, it is also necessary to visualize each column contextually. Data visualizations often help reveal patterns that are not previously obvious. Tableau will be able to provide colorful and useful data summaries of the superstore data.

In addition to an overview, the data analyst will explore possible correlations between variables. While SAS Studio can perform a rudimentary correlation analysis, a further examination should be done by conducting various statistical tests. One important preliminary step that would be useful is to perform a two-sample t-test, to see if the segment of customer is significantly correlated to the order priority. Another might explore the profitability of the company over time, using the dates provided, to see if a linear regression would reveal a prediction for future sales.

Finally, a more in-depth analysis will be done by segmenting the markets and exploring their individual statistics, and what they can tell the data analyst about the disparities between markets (Africa, APAC, EU, Canada, etc.). This may also be extended to investigating not only the profitability of each market, but the tendency of each market to purchase specific subcategories of products, or the likelihood of each market to purchase more highly when items are discounted.

Overall, the goal of the analyst is to explore the different correlations that will help answer business questions that will achieve the superstore’s business goals. That is, to increase profitability and decrease losses, while expanding reasonably to new locations that will provide an increase in revenue. Additionally, it should not be neglected to address issues and decreasing sales in different markets, and exploring potential reasons for those decreases.

# Research Questions & Hypotheses

**Statement 1**

There is often a U-shaped balance when it comes to discounting products, meaning that if a product is too heavily or often discounted, customers may perceive it to be of low quality and avoid it, and if a product is discounted appropriately, customers will be encouraged to act on the sale (Zheng et al., 2021). Therefore, an important question to investigate is whether or not discounts impact the superstore, either positively or negatively.

*H0: Discounts do not have an effect on profits.*

*H1: Discounts have a statistically significant effect on profits.*

In order to test this, it will be necessary to transform the column *Discount*. Currently it has numerical values; RStudio can be used to quickly convert all values where Discount = 0 to a “0” value and Discount != 0 to a value of “1”. Then a two-sample t-test can show whether or not the application of a discount impacts profitability.

Should it happen that discount does have a significant impact, then a linear regression could show the change in profitability, as the percentage of discount increases. This can be done and validated through either RStudio or SAS Studio, and then visualized through Tableau.

**Statement 2**

In an era of increasing convenience for customers, it is important for companies to maximize the efficiency of their logistics, and provide high value to meet the high expectations of the customers (Ma, 2017). Delivery time has a chance to heavily impact customer satisfaction, depending on the sector of item they are purchasing.

*H0: Longer shipping time has no effect on profits.*

*H1: Longer shipping time has an effect on profits.*

This will involve using RStudio, by using the difftime() function. First it is important to verify that the *Order Date* and *Ship Date* variables are ‘date’ class; if not, they must be converted. Using RStudio, a new variable can be created which gives the length between order and ship dates as an integer. Using percentiles, one can analyze the 1st and 4th quartiles to see what the fastest and slowest shipping values are. These can also be visualized by country or market in Tableau to investigate which countries experience average longer or faster shipping values.

By separating the 1st and 4th percentiles into a Boolean variable using RStudio, a two-sided t-test can then determine whether or not longer shipping time impacts profits. To take it a level further if possible, it might be a good idea to choose companies with many records over time, and visualize in Tableau the sales over time, potentially in histogram format. By choosing companies with historically longer shipping times versus companies with shorter shipping times, an analyst can then inference whether or not shipping times decrease companies’ interest in purchasing from the superstore. It might be of value as well to investigate the Shipping Priority variables as well and compare the predictive models of orders with high priority vs. low priority.

**Statement 3**

*H0: Customer segment has no impact on average sales.*

*H1: Customer segment has an impact on average sales.*

As well as performing a two-sample t-test on the 1st and 4th percentile of customers through RStudio, an analyst should also perform descriptive statistical visualizations through Tableau to gain a quick overview over which segments might garner more profit overall. It might be of interest to investigate a decision tree model. This would be a model done in SAS Studio or RStudio which could determine, depending on which average spending quartile they fell into, which segment of customer they were likely to be, or vice versa.

# Data Analysis for a Department Store: A Literature Review

**Importance of Data Analysis for Retail: General**

The importance of data analysis in retail cannot be underestimated. Business intelligence is becoming a more technologically driven function than ever; BI provides tools for companies to make timely and precise decisions within the framework of their organization (Bijaksic et al., 2014). In terms of retail, data can allow a company to understand the scope and speed of their operations, as well as its profits or losses (ibid).

**Data Analysis for Customer Satisfaction**

Big data also allow a company a window into the perspective of the customer, and provides many opportunities to increase customer satisfaction. At the Sixth International Conference on Information Management and Technology, a research explored the consequences of data on customer satisfaction and loyalty. With an increase in data tracking capabilities, there are many opportunities for companies to capitalize on personalization. However, with the additional increase in online retail sources, companies are pressed to optimize their personalization options in a bid to win and hold customer attention (Ling, 2021). Wang et al. also explore the overwhelming amount of development in online retail formats and the resulting increase in data to be analyzed, using as an example a chart which shows the visible trend of online shopping over the weekdays versus weekends (2021). Zheng et al. bring attention the psychological element of retail data, describing the research done to analyze the impact of discount on customer purchases and trust in a store’s product (2021).

Data analysis can be vital for uncovering patterns and trends in data that might not otherwise have been clear before (Bijaksic, 2014). Bijaksic refers to data processing in retail as “sales process diagnosis” (2014), indicating that not only can it identify positive trends to encourage and maximize, but potential areas for improvement as well. This literature will go well with the exploration of how discounts may or may not statistically impact a company’s profit.

**Data Analysis in Globalization**

The company whose data is being analyzed is a company which operates on a global level, with multiple markets and sectors. Youssef et al. explore the importance of big data analytics in a global market. There are cultural differences in the ways that different countries place values on key things like security, management support, and decision-making culture (Youssef et al., 2022). When running a company that spans multiple cultures, it is important for the company to focus its central identity on a data-driven approach. This may mean analyzing data from specific sectors and comparing it in order to find the best approach inn each market, allowing for an agile and flexible approach, as the article suggests (ibid). This literature can be used to further explore the question regarding longer shipping times, introducing the potential need for strengthening branches of the company that are furthest from the center. This would lead to some crucial and costly business decisions that could increase the profit of the company.

**Literature Review Summary**

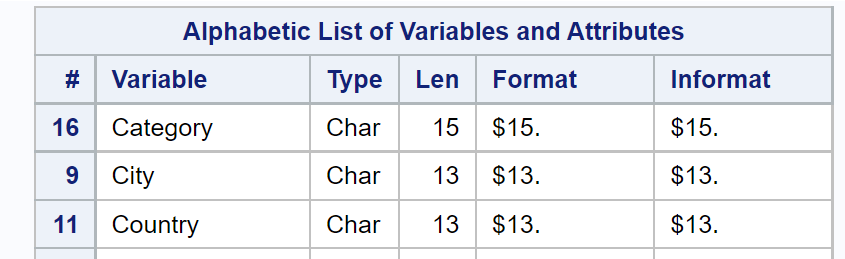
There is no small amount of data regarding the effect of big data and its usages on the world of retail. Some common themes which follow throughout each article is the wide variety of ways in which data excellence increases business value. This may include customer satisfaction, time forecasting, tracking for previously unseen correlations, and predicting avenues of future growth and expansion. The portfolio study will include these aspects as well as statistical models to support their conclusions.

# Methods / Methodology

The dataset in question is obtained from Kaggle and contains superstore sale information from January 1, 2011, through mid-2015. There are 51,290 observations, with 24 variables. The data has secondary keys to hypothetical other tables not included in the original dataset: CustomerID, ProductID, and OrderID (this is not a primary key, due to the fact that there are non-unique values.

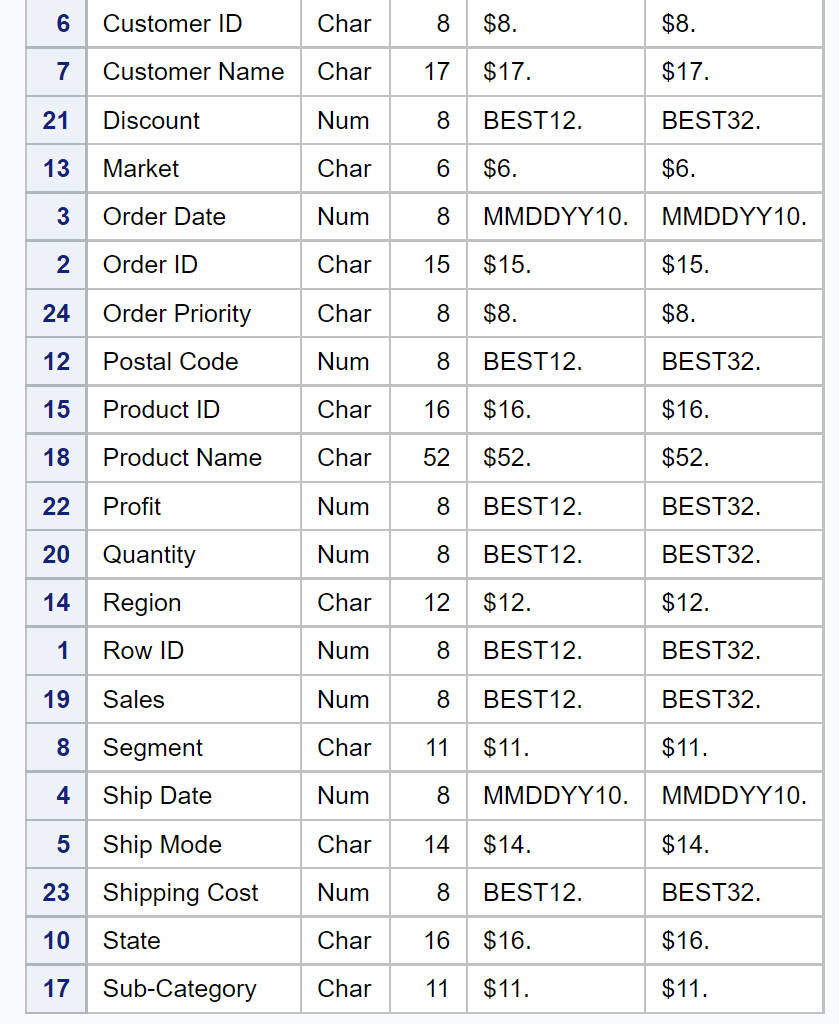
**Figure 1**

*A data dictionary for the dataset*



**Figure 2**

*Data dictionary, continued*



While qualitative data is certainly useful in terms of providing business intelligence and context, for the purposes of this research, the research question mostly focuses on profit. Therefore, using qualitative data to produce quantitative models that will predict and produce profit will be very important. The dataset provided, as mentioned later, contains both qualitative and quantitative data, and so mixed methods will be required, as well as converting qualitative variables to categorical ones.

Before beginning any analysis, it will be important to perform those conversions in an organized and effective way, as well as eliminating variables that have few to no observations, such as “Postal Code”. Additionally, it will be of use to identify and analyze outliers and their purpose for existence, as well as how to handle them.

The superstore described has collected records of sales to countries worldwide, and has divided these countries into different marketplaces. Categories of sale include office supplies, technology, and furniture. Subcategories provide more detailed information, while product descriptions contain a list of the exact products sold. Orders have different priorities, though whether that is based on customer ranking, time between order and shipping, or customer selection, it is not made clear.

In order to accurately analyze the data, it will be necessary to identify the categorical variables which can be easily analyzed, such as Shipping Priority, Market, or Customer Segment. These are qualitative data, but can be quantified. Data which indicate the quantity of customers of the Corporate segment, for example, becomes useful in a measurable way, while still retaining qualitative information about that segment. Meanwhile, other data, such as the item descriptions, which are very long character entries and often contain multiple items, are qualitative data that would be more difficult to parse, and would require a deeper level of text analysis than will be investigated in this particular project. However, the data concerning shipping and order dates can be easily converted into quantifiable values by using timediff() and finding the time between order and ship.

Finally, there is a large number of data which is objectively quantitative. This data, including Profits, Sales, Quantity, Discount, and Shipping Cost can be used in many different ways to correlate with the other qualitative data. Discount can be used to create regressions, while cross-referencing Quantity of sales with individual customers can give key information about repeat customers.

Tableau, since it is primarily a visualization tool, will be utilized for descriptive statistics and basic visualization that will allow the analyst to explore the hypotheses proposed in this paper, as well as draw other inferences. RStudio, as a statistical programming language, will be used to clean the data and provide basic statistical tests; this can be used in combination with SAS Studio which offers many of the same features with a slightly more user-friendly interface, and which also performs automated visualization sans coding in some cases. Some Python packages such as sklearn and numpy may be effective in creating and validating prediction models for data. All of these resources should be used in conjunction to provide a clear picture of the story the data is telling. It is likely that a superstore will have access to all these resources and would want to leverage them to the best effect.

# Limitations

While the grocery dataset contains secondary keys to other data tables, those tables are not available for analysis. This limits the exploration of additional contexts and insights. For example, customer demographics can only be explored on a limited basis; being able to join a products database with sales reports might give a clearer indication of popular selling trends. Additionally, while there is information available on order priority, the analyst has no context as to how the company operates and handles orders of different priorities.

Further information to be explored might include the sales team, store, or employees handling the particular sales. It should not be ignored that customer experience is often the turning point which keeps a customer, and if a certain employee or specific store has a higher success rate, the company would benefit from investigating and differentiating the conditions that establish this.

# Ethical Considerations

In any situation where customer data is involved, it is important to maintain confidentiality. While in this dataset there is not really a lot of confidential demographic information, it is important that the company understands that there could still be financial or social ramifications for a breach in said data (Lee et al., 2016). Anyone with that information could connect customer name or company name to their purchase history or even banking information (Maryville University, 2020). The organization must begin any data analysis with a strong and cohesive understanding of moral principles and organizational values when it comes to data privacy and sharing (Culnan et al., 2009). Fortunately, none of the analyses or models require the usage of customer name, so that column of observations will be omitted entirely through RStudio, thereby anonymizing the data.

Not only is it important to understand the importance of data privacy, it is important also to remember that human bias may easily come into play when analyzing data. For example, when measuring customer segments, one might potentially hold troublesome biases that affect the insights the analyst creates from the data (Kiviat, 2019). In this particular instance, the analyst is supported by the fact that there is no demographic information included to measure customer segmentation, and so conclusions can be drawn with less bias. Nor will any demographic or sensitive information appear in any visualizations.

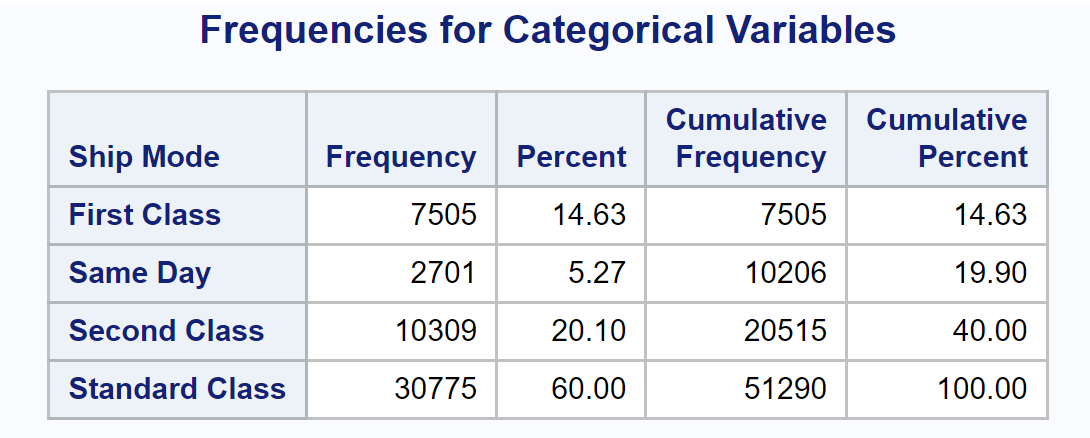
# Findings

**Descriptive Statistics / Overview**

In order to gain a quick look at the data, it is important to use RStudio’s and Tableau’s basic descriptive functions to see if there are any immediately noticeable trends or outliers in the data. Of most importance are the variables which will be analyzed in the previously proposed hypotheses.

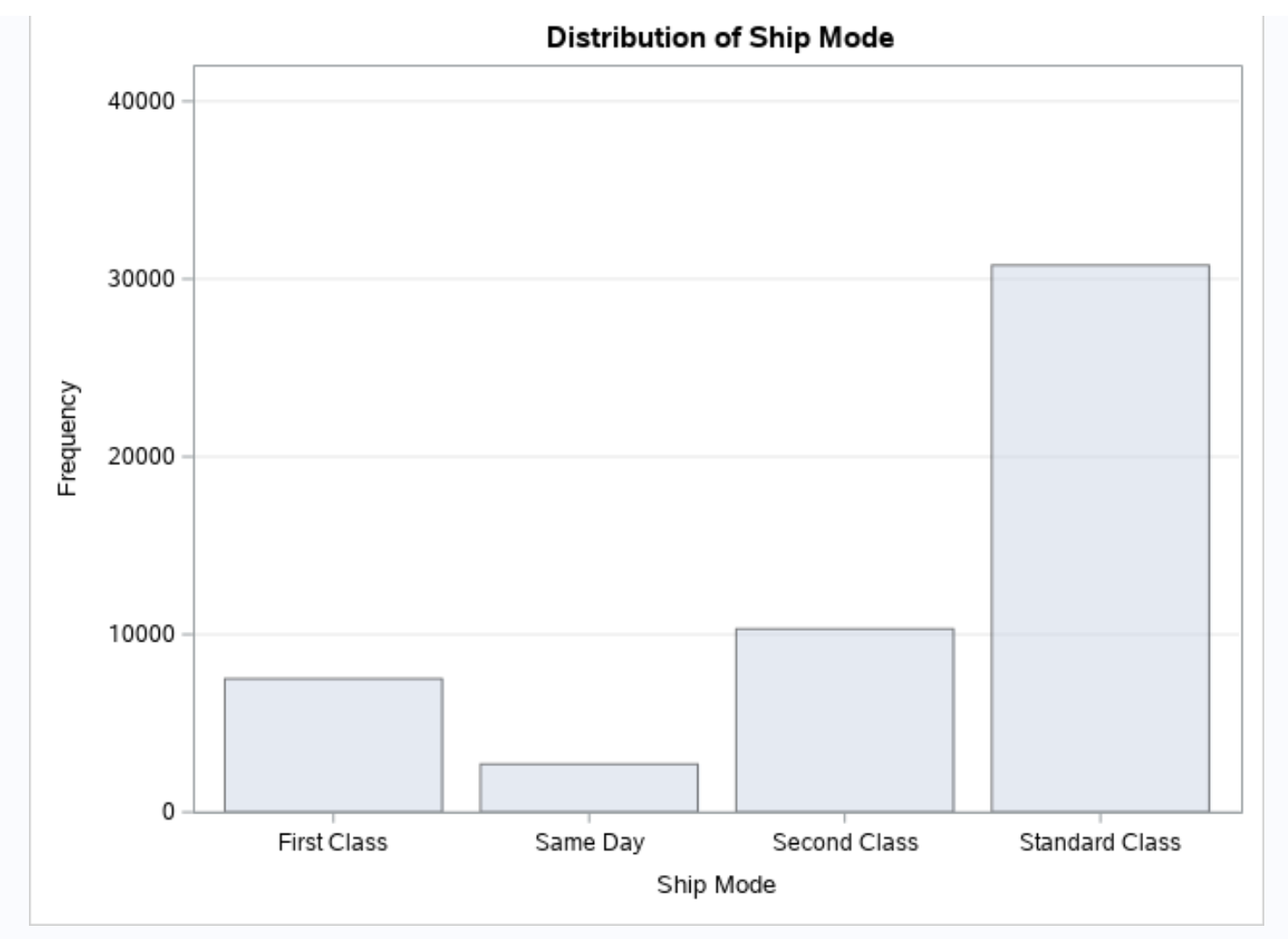
**Figure 3**

*Showing the frequencies for categorical shipping variables in SAS Studio*



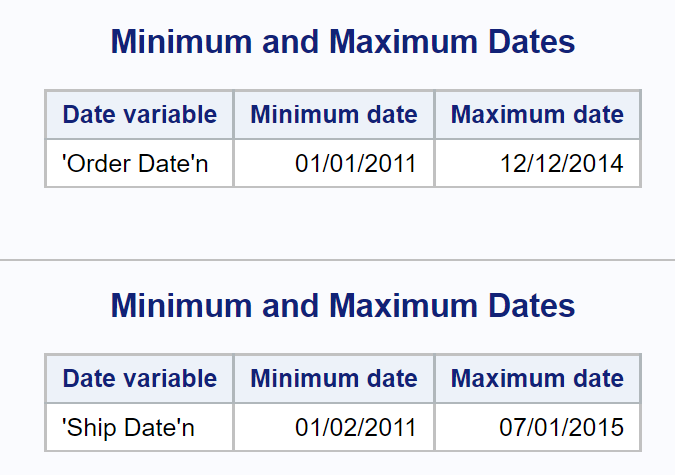
**Figure 4**

*Distribution of Shipping Mode*



**Figure 5**

*Showing “minimum” & “maximum” order and ship dates*



**Figure 6**

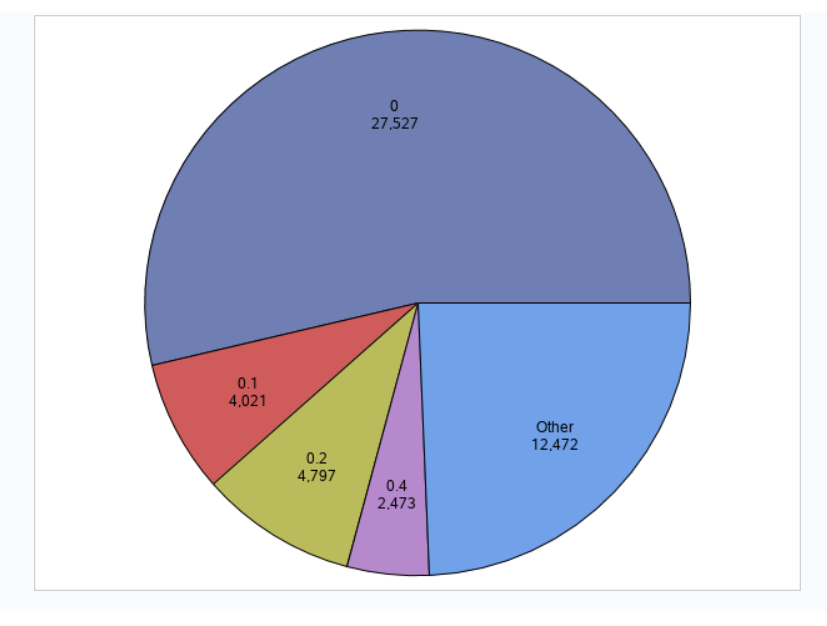
*A view of the Summary Statistics shows that some observations may be misclassified*



A cursory view of the Summary Statistics indicates that while Segment was meant to refer to Home Office, Consumer, or Corporate, a minimal amount of observations have been misclassified by their country instead. Further investigation of the original .csv file and through RStudio show that this appears to be an error on the part of SAS Studio, as none of the values are misclassified there.

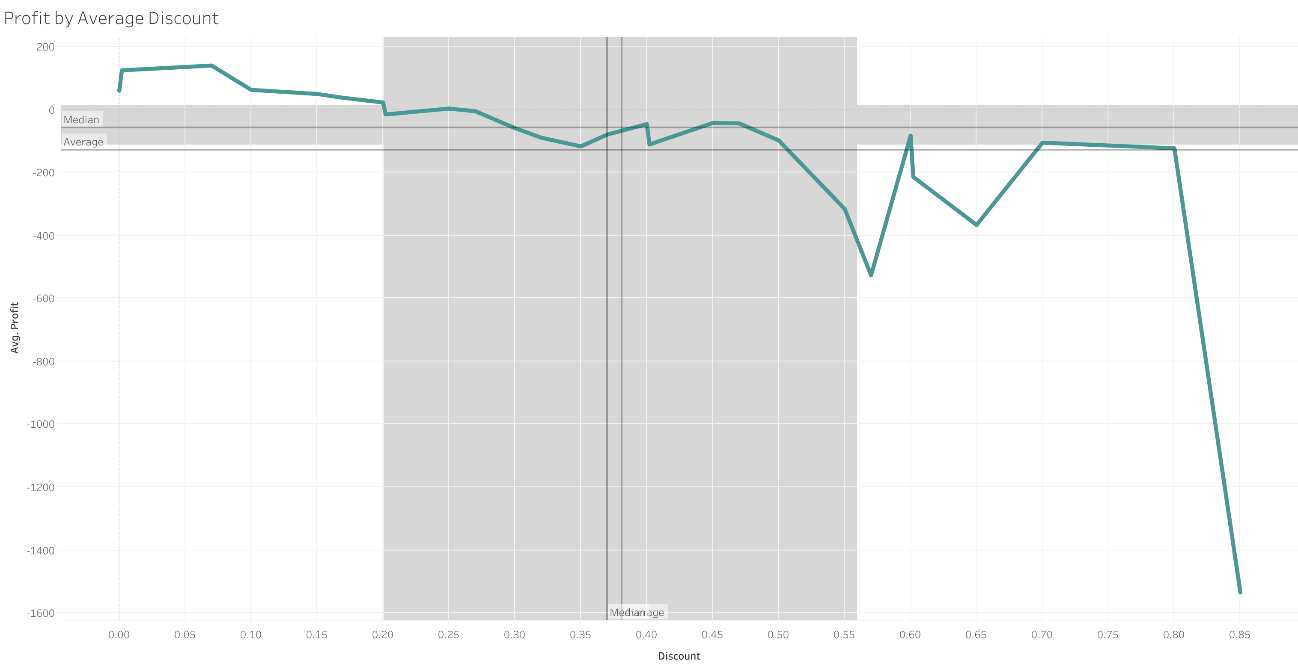
**Figure 7**

*A SAS Studio pie chart of the amounts of observations that were discounted. It appears that more than half of sales were not discounted, but the next highest discount was 10%.*



**Do Discounts Affect Sales?**

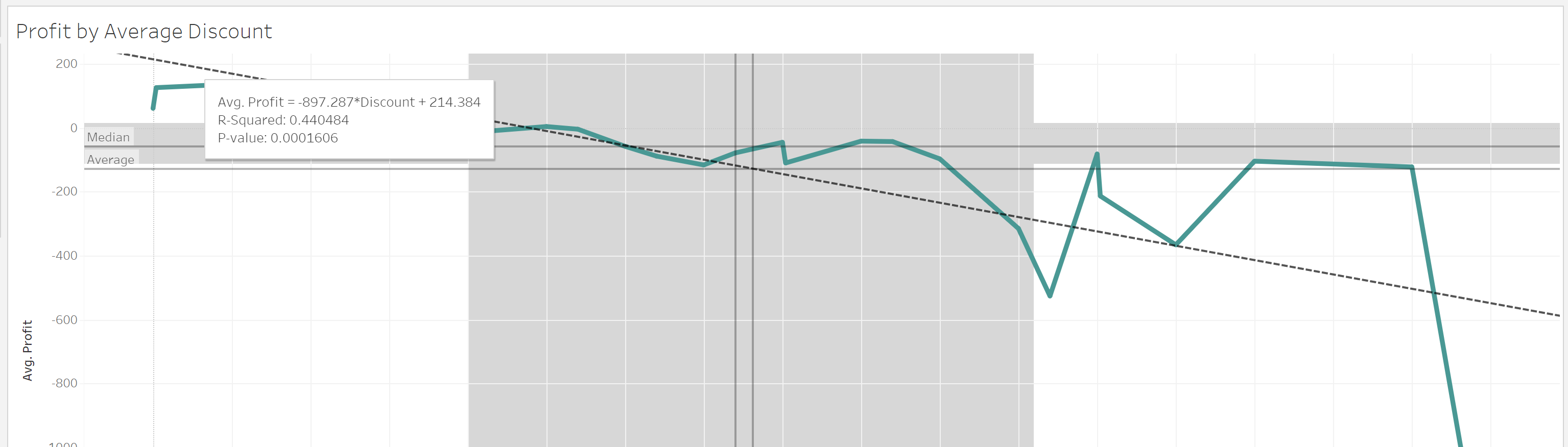
**Figure 8**

*A visualization of profit by average discount*

A quick Tableau visualization of the store’s average profit when measured by discount indicates that the only time that discounts make an average profit is around the 0.07 or 7% discount mark, where the profit measures an average of $141. Additionally, finding a trend line indicates that, with a high p-value, an increase in discount decreases profits significantly.

**Figure 9**

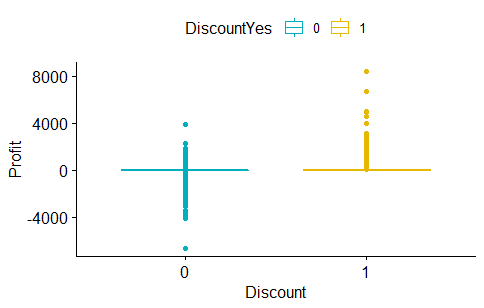
*Showing the trend line of the graph from Figure 8.*



In order to support the hypothesis that discounts impact sales, RStudio was used to separate data into two dataframes, where DiscountYes = 1 if a discount was applied. Profit was the other variable selected in this case. A box plot comparing these two dataframes indicates that there does seem to be a difference in profit.

**Figure 10**

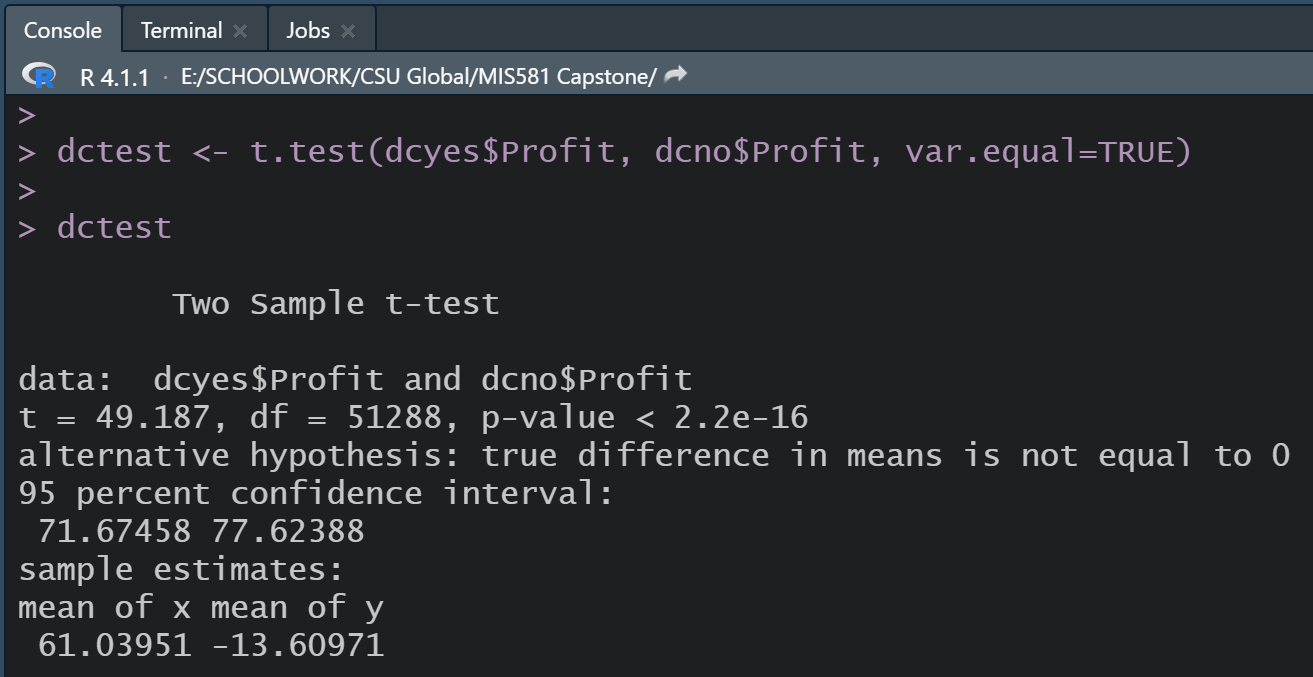
*Showing box plots of the profits based on whether or not a discount was applied.*



A two-sample t-test then was then performed to investigate whether or not sales with discounts had a significantly different mean profit than sales without. This t-test showed with a very high confidence level that the mean profit of sales done with a discount were roughly $61.04, while the mean profit of sales done without discount were

**Figure 11**

*Showing the results of the two-sample t-test*



In order to support this result showing that the differences are significant, the analyst performed a linear regression through SAS.

**Figure 12**

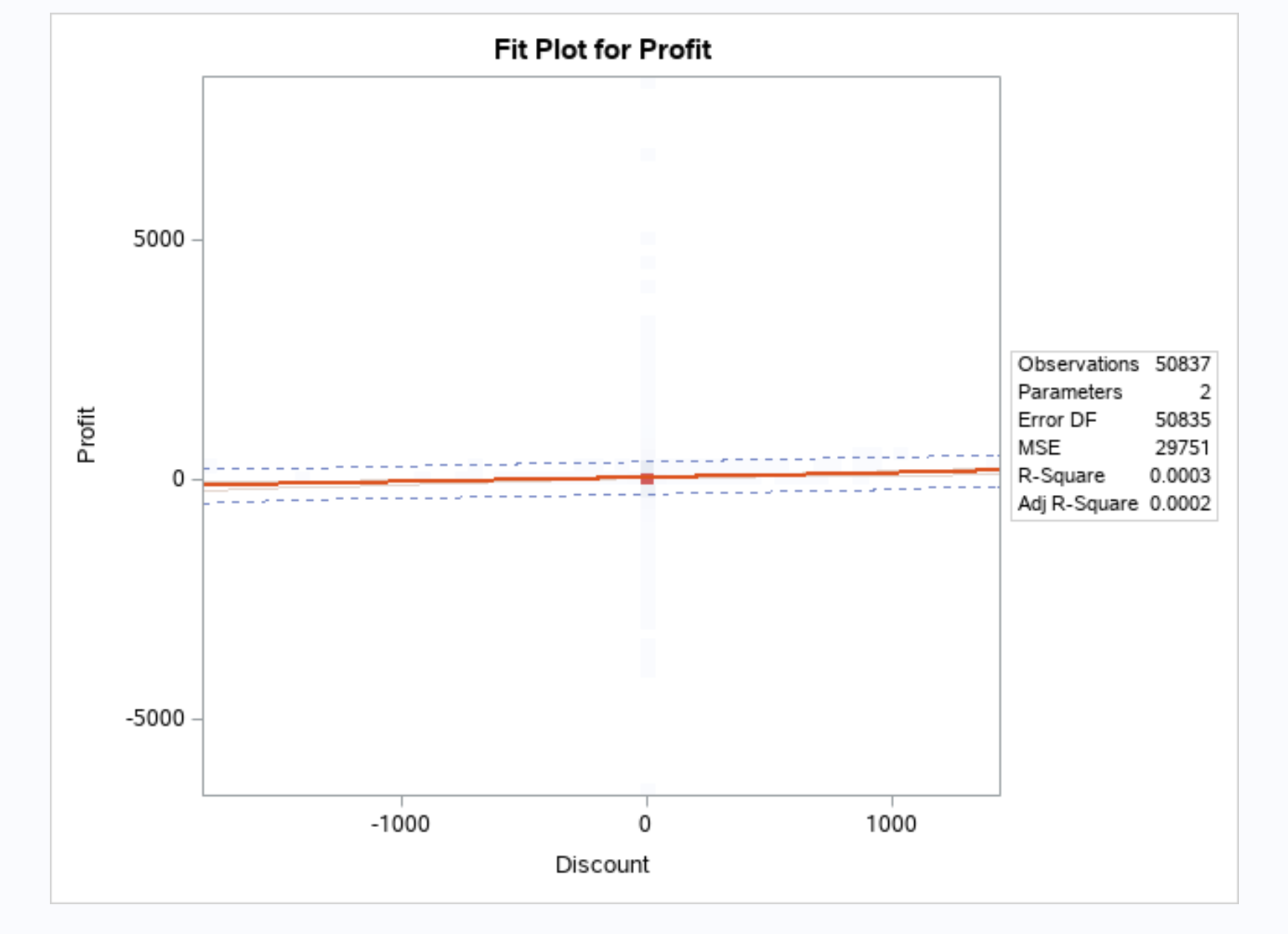
*Showing a linear regression completed in SAS Studio*



Using SAS Studio, the linear regression indicates that as discount increases from 0 to 1, profit actually increases with a slope of 0.09589, which is fairly low. However, it does increase with a p-value of far less than 0.05, indicating that it is within acceptable confidence limits and is statistically sound. This suggests that a linear regression is more useful to note than Tableau’s trend line, as the null hypothesis can be rejected: discounts do have a statistically significant impact on profit.

**Figure 13**

*The linear regression, continued*

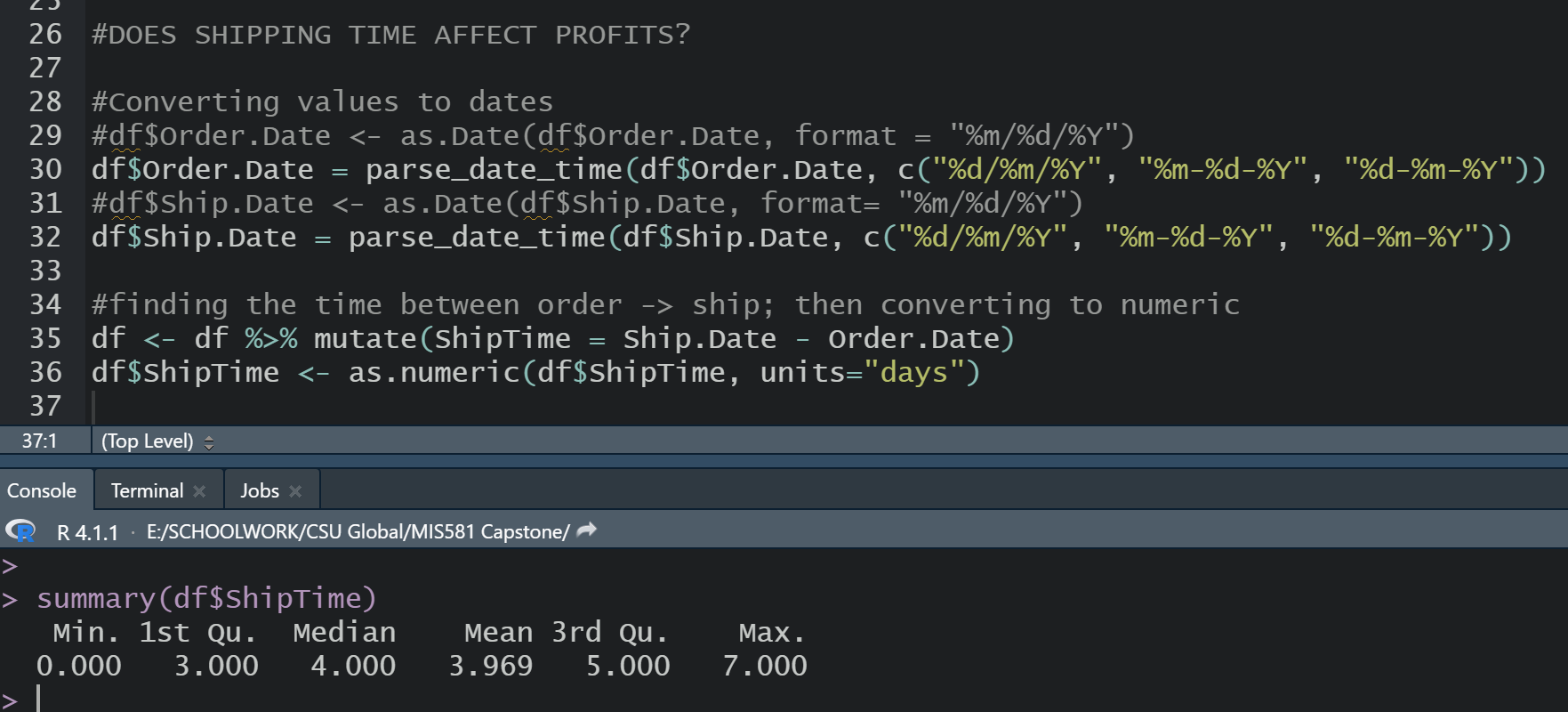


**Does Shipping Time Have An Effect on Profits?**

In order to investigate the importance of shipping times, it was necessary to find the time between ordering and shipping. RStudio difftime() and lubridate() were used to create a new column called ShipTime, and investigating the summary indicates that there is an average shipping time of 3.969 days, a minimum of 0, and a maximum of 7 days.

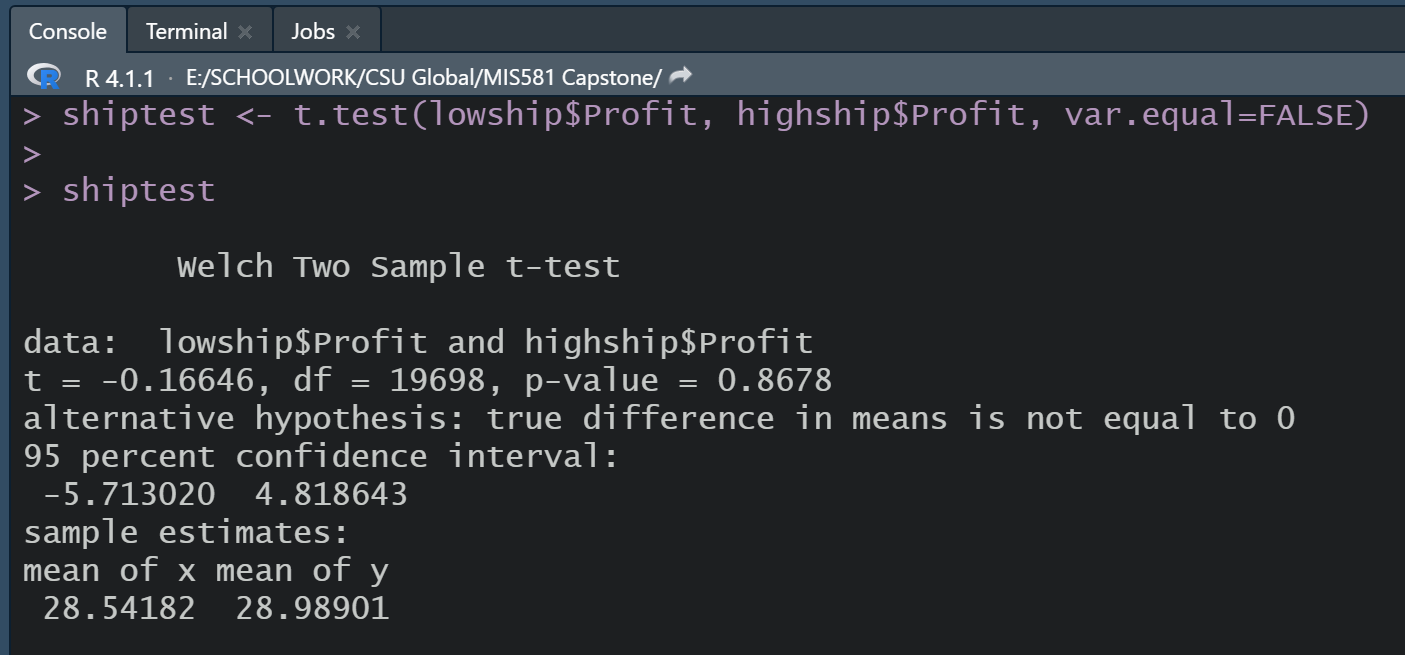
**Figure 14**

*Summary of new ShipTime variable.*



**Figure 15**

*A two-sample t-test of high vs. low shipping times*



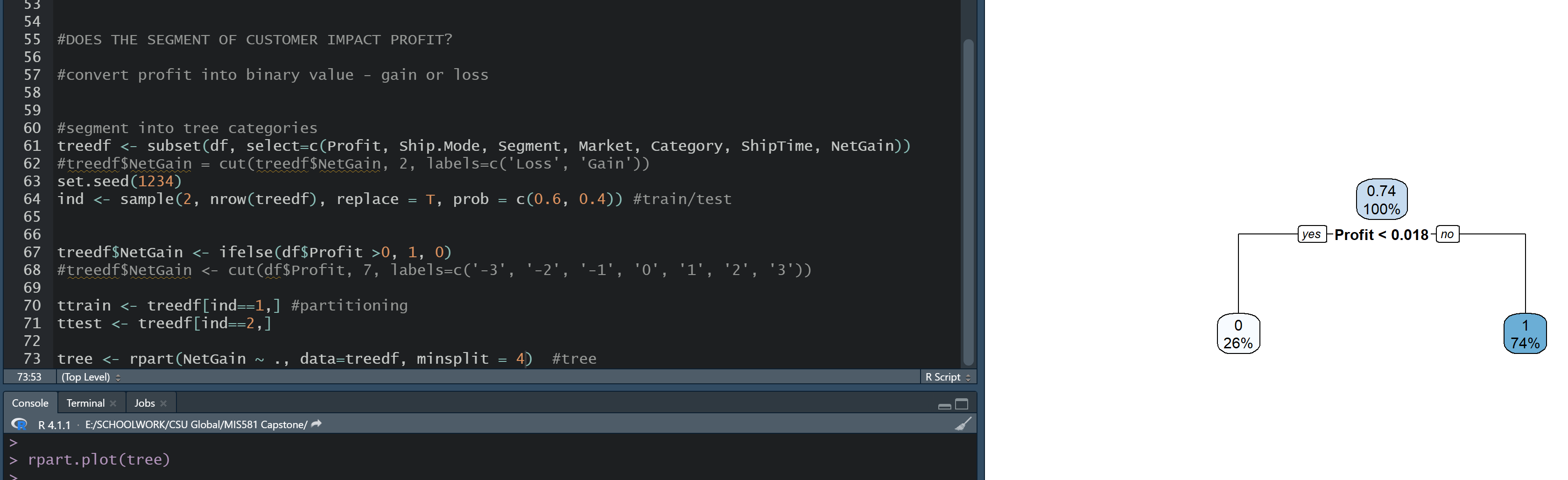
With a confidence level that is far above the acceptable p-value limit of 0.05, it becomes clear that with a two-sample t-test which compares observations with shipping times of 2 or less vs. shipping times of 6 or more days, there still is no statistically significant difference in means between profits. Therefore, the alternative hypothesis is discarded, and the analyst can effectively answer the question with the response, “No, shipping time does not have a statistically significant impact on profits.” In the future, in order to optimize performance, it might be wise to investigate the costs of shipping instead, to see if they can be reduced. This is not information provided in the original dataset, but would be valuable to a company that seemingly has already mastered effective shipping times.

**Does Customer Segment Have an Effect on Profits?**

In an attempt to understand the effect of customer segment on profits, an earlier proposal in the research project was to create a classification tree. Using RStudio, the analyst was able to create a factor-based 0/1 variable based on whether or not a positive profit was turned in each observation. However, the basic classification tree algorithm did not produce useful results. This suggests that either the algorithm was not correctly produced, or did not have enough useful classifiable information to go from. It would have been useful to include a node where the segment was shown to influence profits. Further investigation may reveal there is another solution to a simple problem. Instead, another tactic was chosen.

**Figure 16**

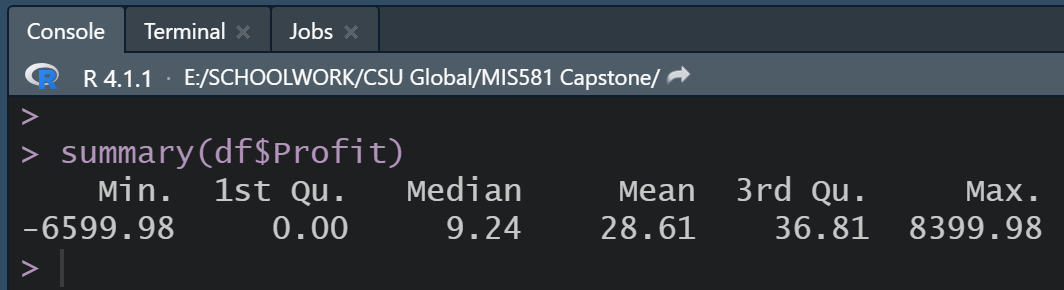
*Showing the 2-node classification tree that resulted.*



The analyst decided to identify the sales with the highest profits versus the lowest profits. In order to identify this, RStudio was used to perform a summary of the $Profit variable.

**Figure 17**

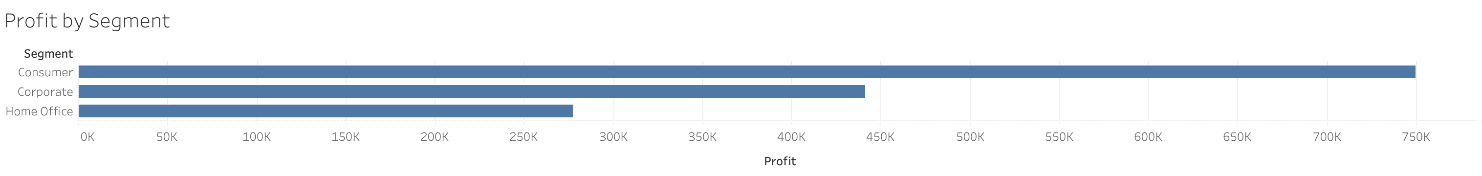
*The summary*



Profits above the 3rd quartile are greater than $36.81, while anything below the 1st quartile is a negative profit. By examining the data in Tableau by segment, it appears that the consumer segment far outsells the other segments as a whole. However, if measuring the average profit by segment, Home Office actually wins with a higher profit average than Consumer.

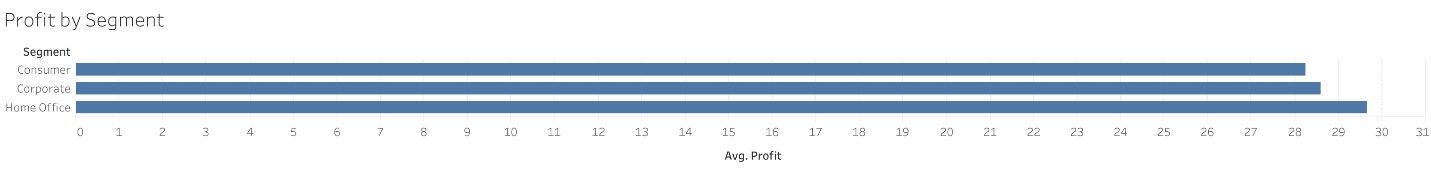
**Figure 18**

*Sum of profit by segment*



**Figure 19**

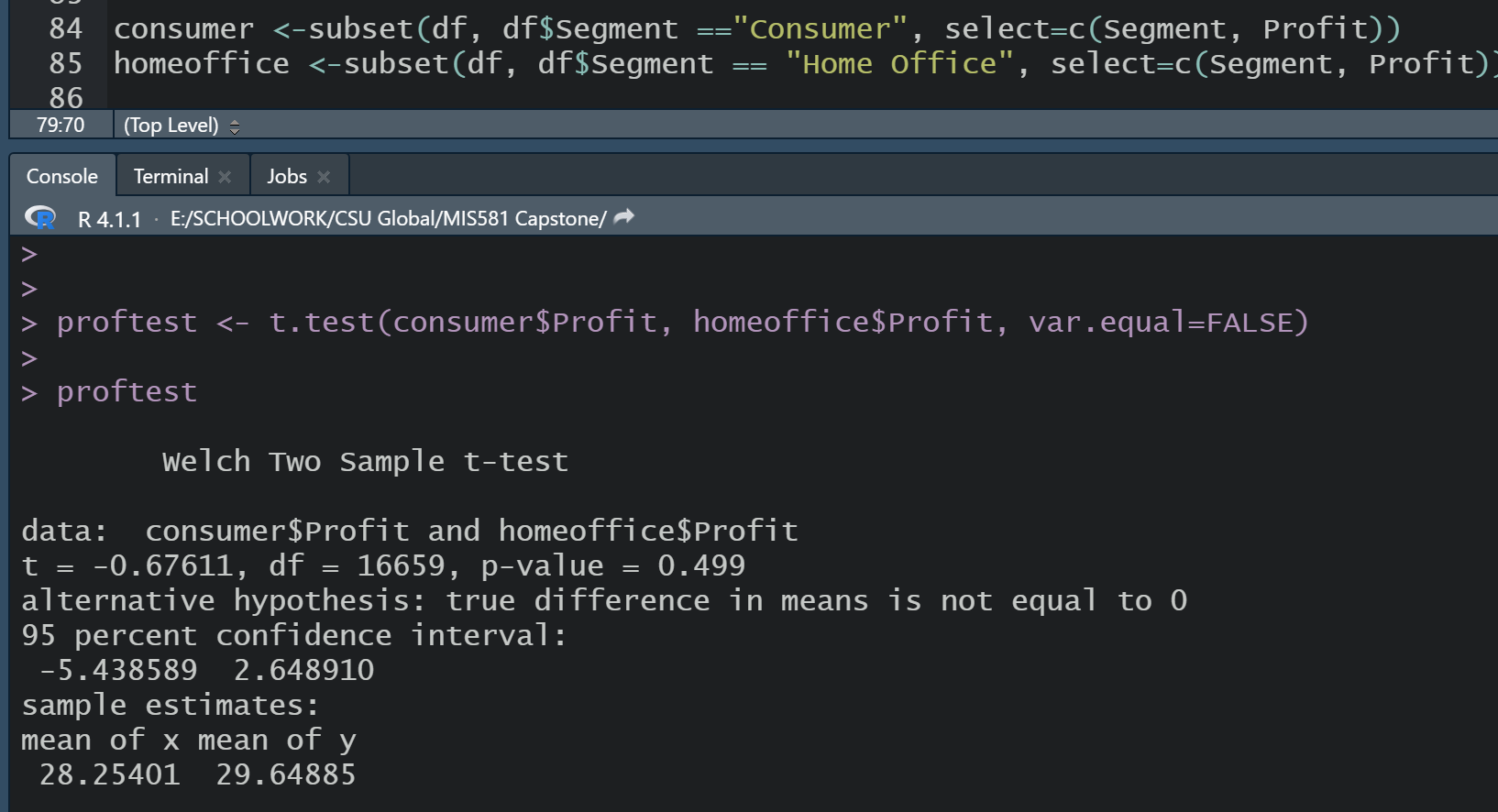
*Average of profit by segment*



To test this, RStudio was used to split data by Segment, comparing Consumer vs. Home Office. Then, an independent two-sample t-test was run to see if there was any significant difference in the means in profit for each segment. Results showed that with a p-value of 0.499, the difference in means is not statistically significant. The mean profit of Consumer sales versus Home Office sales is only 28.25401 to 29.64885. Therefore, the alternative hypothesis must be discarded and the null hypothesis remains. The variable Segment, at least between the overall highest profiting segment and average highest profiting segment, does not have a significant positive or negative impact on profits.

**Figure 20**

*The two-sample t-test results*



# Conclusions

**Do Discounts Affect Sales?**

It becomes evident that discounts do have a statistically significant effect on profits—those observations which were discounted experienced a positive profit, while sales done without discount averaged negative profit, or a loss. This makes senses, as customers often attach value to purchasing an item at a perceived lower value, even if that value is still higher than its production cost (Business Victoria, 2022). Additionally, price promotions have been shown to increase interest in not only individual sales, but also brand and category sales (Gauri et al., 2017), which is critical in superstore performances.

**Does Shipping Time Impact Profit?**

The analysis performed indicated that there was no correlation between higher shipping times and lower profits. This may indicate that the shipping methods and shippers utilized by the company are efficient and do not lose any profit, no matter the distance. This may indicate also that they are geographically located in appropriate places to provide the same level of service to all customers.

**Does Customer Segment Have an Effect on Profits?**

Preliminary analysis suggested that the customer segment did not have a particular impact on sales. While the segment Consumer netted the highest overall profit, Home Office customers actually spent more on average. Corporate sales fell in the middle in both cases, but when averaged, the profits of Corporate sales were not that much different than that of Consumer or Home Office. Potential conclusions to be drawn are that the products provided by the superstore are effective for all segments; there is no specialization of products in all segments. It also seems that while average consumers create the majority of the store’s profits, when a corporate or home office customer arrives, they will tend to spend more.

# Recommendations

In terms of addressing the impact of discount on sales, the company should take into consideration that despite the perceived negative impact of selling items at a “lesser” value, they will be gaining a considerable amount of consumer attention by providing goods at a lower perceived value. It has also been noted that when “bundling” products and services as packages or deals, customers may also continue to engage at a higher rate (Business Victoria, 2022); this is something the superstore could consider implementing. Once implemented, they could continue to collect data to see if this method is effective for improving profits even more. Performing a time series analysis along with the discount trends could also allow the company insight into when the best time to perform discounts is, allowing maximization of profits; a large amount of retail experiences some level of time-series modeling due to seasonal needs (Wang et al., 2021).

A further recommendation would be to investigate the impact of discounts on different segments of customers. There has been shown to be different reactions in different demographics of customers (Gauri et al., 2017); there are other human elements involved in peoples’ attitudes in purchasing discounted items, including geography, income, and environment such as urban or rural (Pitts et al., 2018). By collecting further information on the type of customer and their reactions to discounts, the analyst would be able to provide even more insightful models that could help the business target discounts accurately, in terms of category discounted and customer attracted, for the highest profit.

When it comes to shipping times, the conclusion was that shipping time did not impact profit significantly whatsoever. However, the analyst and company would benefit from continuing to measure specific customers in the same geographical locations and markets, and perform time series on their profits. This would allow the company to identify potential seasons of slow shipping that could be addressed; it would also allow them to see if customers who, on average, experienced lower shipping times, eventually stopped buying as much, resulting in a decrease in company profits. Continued analysis of the shipping price and time in comparison to the sales/profit of observations could also help the company determine whether or not their shipping price is appropriate in comparison to the amount of money being spent and the time provided (Yuliang & Zhang, 2012); the creation of an algorithm to maximize customer service would greatly benefit the company—the propagation of theories is not effective unless it also creates hypotheses and predicted outcomes to be tested and acted upon (Wang et al., 2021).

Finally, in the investigation, it appears that there was no significant difference in profit between customer segments. One possible avenue of exploration would be the investigation of different customer segment profits depending on market or region. It is possible that the company could improve profits by slightly specializing their individual physical locations; for example, if the Western market has a higher amount of Corporate-segmented customers, then the company could investigate cultural and environmental factors to improve their physical locations. Customer tastes can be highly volatile (Evans & Harrison, 2005), and by being aware of individual customer needs, the company will have a chance for a high competitive advantage. (Ma, 2017). Something to note is that there are other avenues to explore with relation to segment; the analyst should investigate how Customer Segment interacts and correlates with other variables, as well.

# Final Conclusion

The analyst has begun an investigation into a large superstore dataset and has drawn conclusion about three particular variables: shipping time, customer segment, and discount. While only one of those things—discount—was found to be significant in affecting profit, the analyst should not discount the ‘absence’ of a statistically significant correlation. There is far more to be explored in the data. As the future of retail continues to evolve, and superstores continue to become a larger and larger force in the market (Barnes Reports, 2022), a discerning data analyst should be on the lookout for new ways to view their datasets and glean insights about how to continue to gather and leverage data improve the company’s performance and reach their operational goals.

# References

Barnes Reports (2022). 2022 World industry & market outlook. *Worldwide Warehouse Clubs & Superstores Industry Report* 1-140. <https://csuglobal.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=155472134&site=eds-live>

Bijaksic, S., Markic, B., & Bevanda, A. (2014). Business intelligence and analysis of selling in retail. *Informatologia 47*(4), 222-231. <https://csuglobal.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=108799178&site=eds-live>

Business Victoria (2022). *Benefits of discounting.* <https://business.vic.gov.au/business-information/finance/pricing-for-profit/develop-discount-strategies>

Culnan, M., & Williams, C. (2009). How ethics can enhance organizational privacy: lessons from the Choicepoint and TJX data breaches. MIS Quarterly 33(4), p. 673-687. <https://csuglobal.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.10.2307.20650322&site=eds-live>

Evans, C., & Harrigan, J. (2005). Distance, time, and specialization: Lean retailing in general equilibrium. *The American Economic Review 95*(1), 292-313. <https://csuglobal.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.4132681&site=eds-live>

Federal Reserve Bank of St. Louis (2008). Big-box retail and its impact on local communities. *Economic Information Newsletter*. <https://files.stlouisfed.org/files/htdocs/pageone-economics/uploads/newsletter/2008/200801.pdf>

Gauri, D., Ratchford, B., Pancras, J., & Talukdar, D. (2017). An empirical analysis of the impact of promotional discounts on store performance. *Journal of Retailing 93*(3), 283-303. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0022435917300490>

Golden, S., Jeutang, N., Pattaik, R., Rosenbaum, D., & Thompson, E. (2006). Big box stores: Their impacts on the economy and tips for competing. *Bureau of Business Research, University of Nebraska Lincoln*. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1005&context=bbrpub>

Kiviat, B. (2019). The moral limits of predictive practices: The case of credit-based insurance scores. American Sociological Review 84(6), p. 1134-1158. <https://csuglobal.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=tsh&AN=139825030&site=eds-live>

Lee, W., and Chang, H. (2016). An ethical approach to data privacy protection. ISACA Journal 6, 1-9. <https://www.isaca.org/resources/isaca-journal/issues/2016/volume-6/an-ethical-approach-to-data-privacy-protection>

Ling, T. (2021). Analysis on the development strategies of traditional retail enterprises’ digitalization under the background of big data. *The Sixth International Conference on Information Management and Technology 121,* 1-4. <https://dl-acm-org.csuglobal.idm.oclc.org/doi/10.1145/3465631.3465791>

Ma, S. (2017). Fast or free shipping options in online and Omni-channel retail? The mediating role of uncertainty on satisfaction and purchase intentions. *International Journal of Logistics Management 28*(4), 1099-1122. <https://www.proquest.com/docview/1958171208?accountid=38569>

Maryville University (2020). The increasing focus on cyber ethics issues. Maryville University Bachelor’s in Cyber Security. <https://online.maryville.edu/blog/cyber-security-ethics/>

Massachusetts Institute of Technology (n.d.). Big-box economic impact studies. <http://web.mit.edu/course/4/4.293/!Phoenix/Research/Tenant%20Research/bigboxstudies.pdf>

Pitts, J., Bell, S., Wu, Q., & McGuirt, T., Sharpe, P., & Rafferty, A. (2018). Impact on dietary choices after discount supermarket opens in low-income community. *Journal of Nutrition Educatio and Behavior 50*(7), 729-735. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S1499404618301428>

Song, Y., Sindakis, S., Aggarwal, S., Chen, C., & Su, J. (2021). Managing big data in the retail industry of Singapore: Examining the impact on customer satisfaction and organizational performance. *European Management Journal 39*(3), 390-400. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0263237320300530>

Wang, R.., Jiayi, S., Xiurui, W., & Yang, L. (2021). The impact index and dynamic forecast of the retail: New business formats based on internet big data. *The 5th International Conference on E-Commerce, E-Business and E-Government* 73-78. <https://dl-acm-org.csuglobal.idm.oclc.org/doi/10.1145/3466029.3466031>

Wang, X., Hyun, J., Bendle, N., & Kopalle, P. (2021) The role of machine learning analytics and metrics in retailing research. *Journal of Retailing 97*(4), 658-675. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0022435920300932>

Yuliang, Y., & Zhang, J. (2012). Pricing for shipping services of online retailers: Analytical and empirical approaches. *Decision Support Systems 53*(2), 368-380. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0167923612000401?via%3Dihub>

Youssef, M., Eid, R., & Agag, G. (2022). Cross-national differences in big data analytics adoption in the retail industry. *Journal of Retailing and Consumer Services* 64. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0969698921003933>

Zheng, D., Chen, Y., ZHang, Z., & Che, H. (2021). Retail price discount depth and perceived quality uncertainty. *Journal of Retailing*. <https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0022435921000786>