Analyzing Profitability and Reducing Losses on SuperStore sales

Final Report

Team #14 | Nama'a Azaizeh, Anas Ali, Mohammed Mahameed, Ahmad Abueid, Lojain Alhteabat, Mohannad Zananiri

Introduction

Overview of the Problem

The SuperStore is a retail business that sells a wide range of products to customers. In order to make informed decisions and improve overall performance, it is essential to conduct data analysis. This project aims to analyze the available datasets to gain insights into various aspects of the business and identify potential areas for improvement.

Specific Problem to be solved

The specific problem to be addressed in this project is to analyze the factors that impact profit and sales within the superstore. By identifying these factors, the store can make data-driven decisions to improve its bottom line and increase sales. The project will explore the relationship between variables like ship mode, segment, lead time, city, and returns, and their effect on profit and sales. The insights gained will help the store optimize these factors to drive profitability and sales growth.

Importance of the Problem:

This problem is of significant importance for the Super Store for several reasons. By understanding the factors that influence profit and sales, the store can make informed decisions to enhance its financial performance. Optimizing variables such as ship mode, segment, lead time, city, and returns can lead to cost savings, improved operational efficiency, and increased revenue. This, in turn, allows the Super Store to remain competitive, attract investors, and ensure long-term sustainability in the retail market.

Business Impact

Analyzing profitability and reducing losses in SuperStore sales can significantly impact the store's success. Insights into factors influencing profit and sales enable targeted strategies, such

as improving low-performing categories and targeting specific segments for marketing. Effective resource allocation and inventory planning result from analyzing the quantity ordered. Leveraging these insights leads to increased profitability, reduced losses, improved efficiency, and enhanced competitiveness in the retail market.

Data

The dataset utilized for this project consists of two primary datasets: the Order dataset and the Returns dataset. The Order dataset contains a total of 22,150 data points, combining three files to obtain data from the years 2014 to 2022. This comprehensive dataset enables a more extensive analysis of the SuperStore's sales and profitability. The dataset includes 21 columns that provide valuable information about each order.

Strengths: The dataset has a substantial number of data points over an eight-year period, allowing comprehensive analysis of sales and profitability trends. It includes a wide range of fields that capture essential information for exploring factors impacting profitability.

Weaknesses: The dataset lacks customer satisfaction data or ratings, making it challenging to understand return reasons. Additionally, it may not include external factors like customer demographics or market conditions, limiting the analysis of profitability and loss reduction

Data Analysis & Computation

A dataset containing attributes of small retail businesses called Super Store from the year 2014 until 2022.

Desired Datasets

Datasets Considered for Analysis:
The project will primarily utilize two datasets: the Order dataset and the Returns dataset.

a. Order Dataset:

The Order dataset contains 21 columns and 22,156 rows, including the most important columns:

Ship Mode: The mode of shipment for the order (e.g., same day , first class , second class)
Segment: Customer segment (e.g., consumer, corporate, home office)
Region: Geographical region of the order location
Category: Product category (e.g., office supplies, furniture, technology)
Sub-Category: Product sub-category (e.g., chairs, tables, phones)
Sales: Revenue generated from the order

Quantity: Quantity of products ordered
Discount: Discount applied to the order
Profit: Profit earned from the order

Dataset Details: The superstore dataset contains 22,150+ Rows and 21 Columns.

Size: 2.74 MB (2,814 KB).

Access: Downloaded from Kaggle, Tableau, and Data.world.

Sources: we used three datasets of the same store but the difference is the date range.
the datasets we used are:

- Superstore (2014 - 2017) from Kaggle: Sample - Superstore | Kaggle
- Superstore (2017 - 2020) from Tableau: Sample - Superstore | Tableau
- Superstore (2020 - 2022) from Data.world: Superstore Sales - dataset by ehughes | data.world

And we merged these three datasets using Python with making sure that we don't have any duplicate values.

Data Profiling

Data Table Schema:

The Table below shows the fields name, types, and descriptions. Moreover, we investigate if the columns had null values and if available what percentage of the null fields for each column. Also, it had a column to show based on our discussion what column we will use on our project.
Generally, the dataset was compiled by researchers and has previously been utilized for research purposes. Hence, the data has already been cleansed, formatted correctly, and does not contain any duplicates. However, there are some null values present, but they are not found in the main columns
.

Field	Type	Description	Contain Nulls?	Aim to use?
order_id	STRING	This is a unique identifier of the Order record.	No	No
order_date	DATE	Shows the date when the order was placed by the customer.	No	Yes
ship_date	DATE	Shows date when the order was shipped from the	No	Yes

		superstore to the customer.		
ship_mode	STRING	Shows the mode of the ship. It can be any one of these: (First Class, Second Class, Standard Class, Same Day).	No	Yes
customer_id	STRING	This is a unique identifier of the customer who made the order.	No	Yes
customer_name	STRING	Shows the name of the customer who made the order.	No	No
segment	STRING	Shows the segment to which the customer belongs. It can be any one of these: (Consumer, Corporate, Home Office).	No	Yes
country/region	STRING	Shows the name of the country or region where the order was delivered.	No	Yes
city	STRING	Shows the name of the city where the order was delivered.	No	Yes
state	STRING	Shows the name of the state where the order was delivered.	No	Maybe
postal_code	STRING	Shows the postal code associated with the city or region where the order was delivered.	Yes(1%)	No
product_id	STRING	This is a unique identifier of the product of the order.	No	Yes
category	STRING	The broad category to which the product belongs.	No	Yes
sub_category	STRING	A more specific sub-category within the broader category.	No	Yes
product_name	STRING	Shows the name or description of the specific product that was ordered.	No	Yes
sales	FLOAT	Shows the total sales amount for the specific product in the order.	No	Yes
quantity	INTEGER	Shows the quantity of the product ordered by the customer.	No	Yes

discount	FLOAT	Shows the discount (in percentage) applied to the product price for the specific order.	No	Yes
profit	FLOAT	Shows the profit earned by the superstore from the sale of the product in the specific order.	No	Yes

Data Wrangling

The data wrangling that was made on the Orders sheet:

- The data is overall clean and does not have any outliers.
- Using the filter tool, we found out that the (postal code) column contains cells that are empty.
- The (postal code) column will not be important in our analysis and visualization.
- Removing the empty cells will not matter because in our analysis we will ignore the (postal code) column.
- Other than the empty cells in the ignored (postal code) column, the data is clean.

Mathematical checks

The Table below shows mathematical checks containing calculations from the year 2014 until 2022.

Operation Name	Quantity	Sales	Discount	Profit
count	20188	20188	20188	20188

mean	3.790717	229.033843	0.155790	28.665239
std	2.226675	621.546813	0.206345	233.349665
min	1.0	0.44	0.0	-6599.978
max	14.0	22638.48	0.8	8399.976
25%	2.0	17.246	0.0	1.74
50%	3.0	54.224	0.2	8.674
75%	5.0	209.70	0.2	29.341

Exploratory Data Analysis

The following figures and findings are derived from conducting an exploratory analysis on a subset that delves into the orders data of the Superstore, serving as a testbed for our research inquiry. Our goal is to document pertinent observations related to our subject and develop a model with high precision.

Discount

Figure 1.1 This graph shows the movement of discounts within the years. It was clear to us that there was a sharp drop in sales in the year 2018, and this resulted in a decrease in the movement of discounts. The reasons for the decline are likely to be political reasons in the economy at that time.

(Figure 1.1)

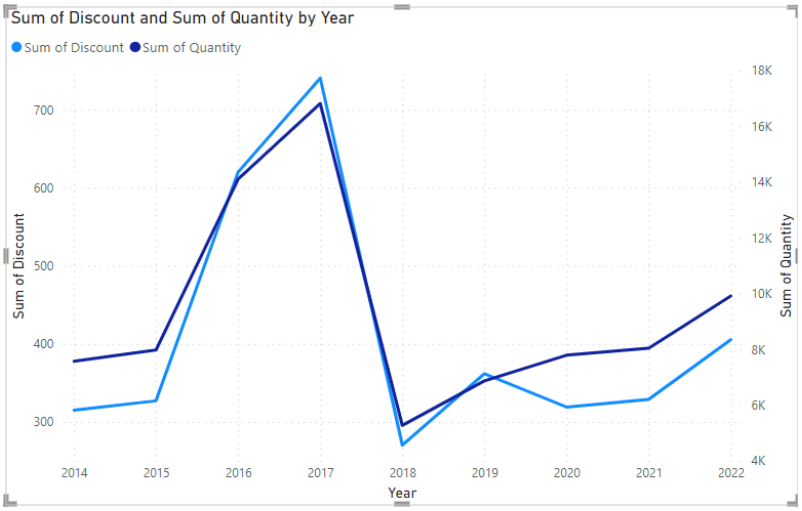
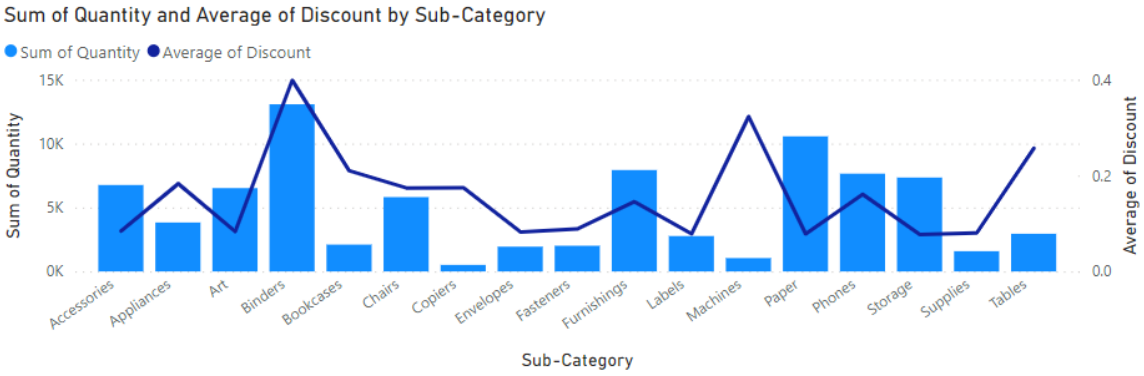
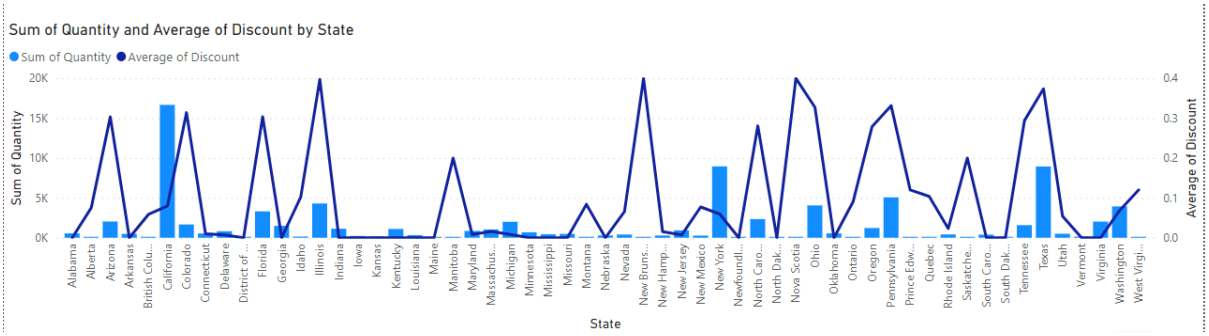


Figure 1.2 /1.3 These graphs aim to know the determining the percentage of discounts according to the quantities required in the states or according to categories
It turns out that there is no clear relationship and a fixed trend. We cannot say that the greater the number of quantities in a particular category or state, this means that it will be discounted but we can know which state or category has a big or small discount .



(Figure 1.2)

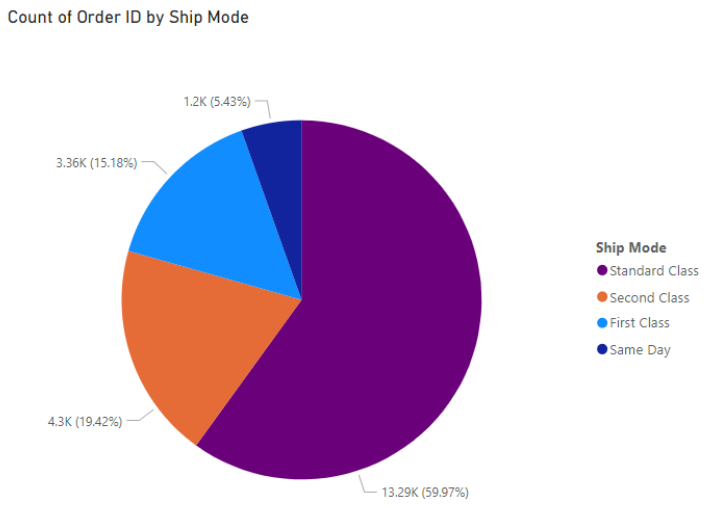


(Figure 1.3)

For segments there are no specific trend, all segments have the same discount rate

Ship mode / Lead Time

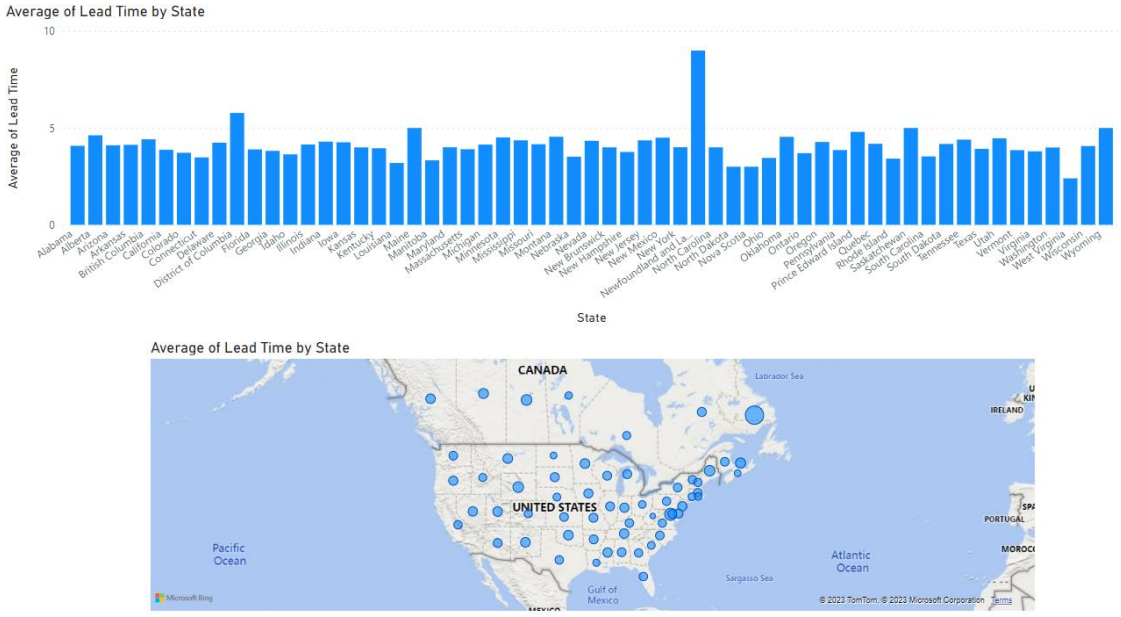
Figure 1.4 The graph shows the distribution of the orders flow by ship mode
We note that most applications are standard class.
Same Day: shipping on same day
First Class: Shipping within 1 Day
Second Class: Shipping within 3 Days
Standard Class: Shipping with 6 days



(Figure 1.4)

Figure 1.5The graph shows the distribution of the lead time by state. We note that most states have the same lead time except Newfoundland and Labrador, and this matter was logical when referring to the data. It was discovered that there are only two orders with quantities of 13 for this state.

The lead time here is a generated column by the formula (Ship date- Order date) To have a numerical column for ship mode



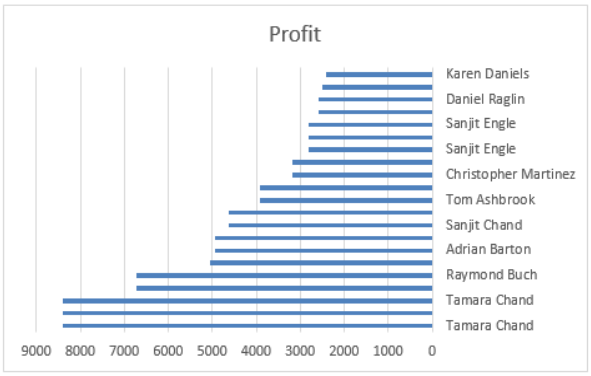
(Figure 1.5)

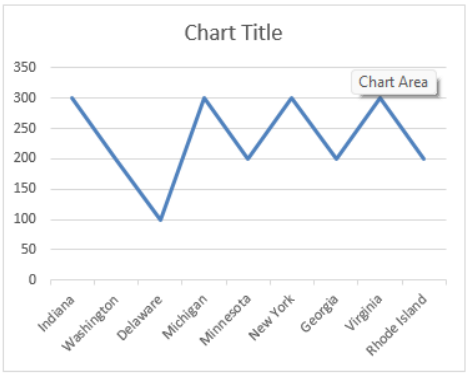
As for the impact of segments, Dates and categories on lead time
There is no turbulent relationship. All relationships are convergent and divided into all years, categories and segments at an average rate

Customer

Figure 1.6 / Figure 1.7 these graphs show the distribution of the Profit by Customer then shows the distribution of the customer by state. By ordering data from largest to

smallest we note the names of top 20 Customers who benefitted the store





(Figure 1.6) **(Figure 1.7)**

Correlation coefficient between numeric columns

Figure 1.8 The graph shows the correlation coefficient between numeric columns in a dataset which measures the strength and direction of the linear relationship between two variables. We note we have :

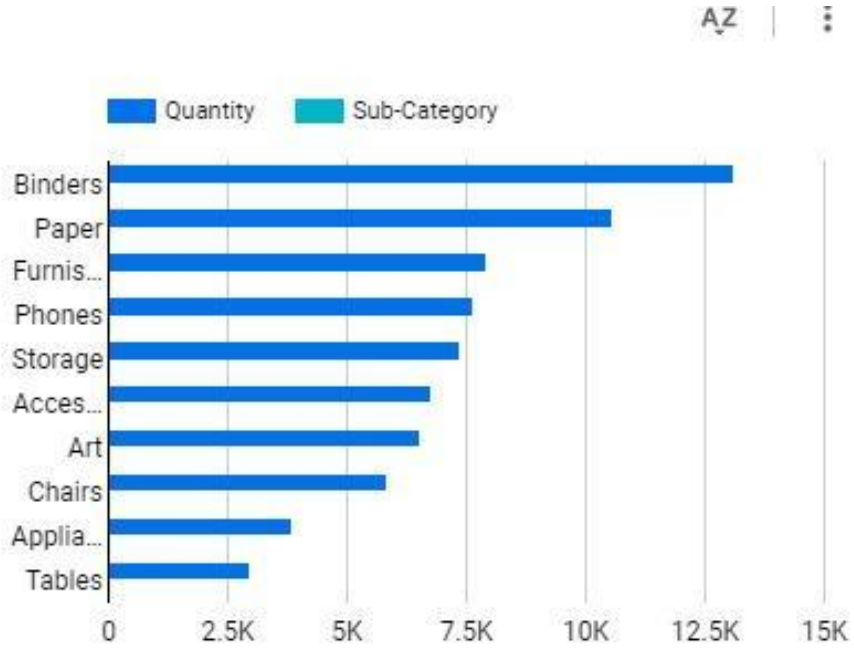
- A positive correlation between (Sales with profits) ,(sales with profit_percentages) ,(sales with Quantity) ,(Quantity with profit) , (Quantity with discount)and (profit with profit_percentage) suggests that as one variable increases, the other variable tends to increase as well, but these relationship is not as strong as it would be with a correlation closer to +1. which means that there is a positive relationship between the variables, but it might be relatively weak.
- A negative correlation which indicates that as one variable increases, the other variable tends to decrease, and vice versa.
  And this relationship occurred between (Sales with discount) ,(discount with profit_percentages) , (Quantity with discount)and (quantity with profit_percentage)
- We note the correlation between (Discount with profit_percentage) is -0.68 it indicates a strong negative correlation between these two variables being analysed.

(Figure 1.8)

Quantity

Figure (2.1) : The graph shows the quantity of products categorized by their respective category, sub-category, and individual product names. The data in this graph is based on the latest available information. the report graph an overview of the quantity of products across various categories, sub-categories, and individual items. It can be used

to identify the distribution of products and analyse their demand



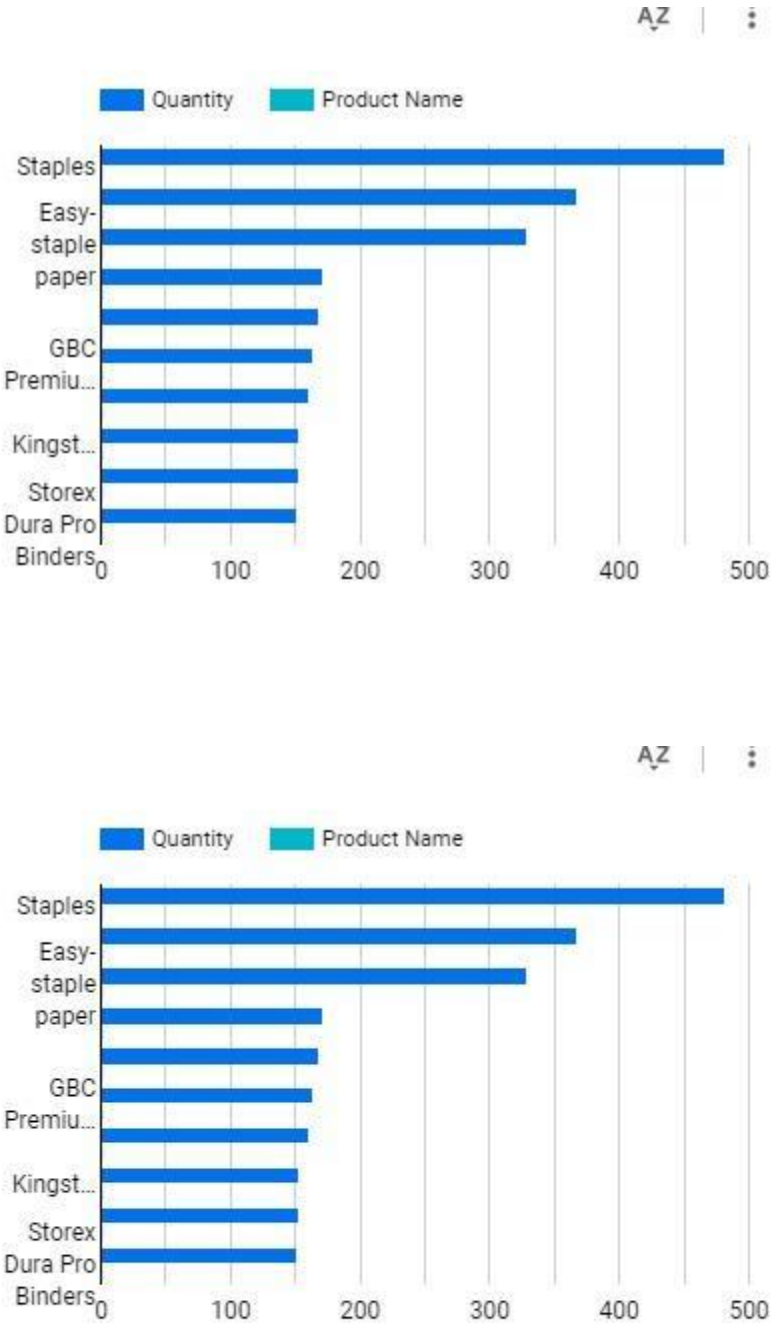


Figure 2.2 This graph provides an analysis of the quantity of items available in various geographic locations. The data is categorized by country, region, state, and city, offering insights into the distribution of products in different locations. By examining the quantity of items in different geographic locations, businesses can better understand

the demand patterns and tailor their supply strategies accordingly. This report can aid in identifying potential areas for growth and optimizing inventory distribution.

Figure 2.3 This graph presents an analysis of the quantity of items available across different segments. The data is categorised by segments, allowing for insights into the distribution of products based on specific market segments. By examining the quantity of items in different segments, businesses can gain valuable insights into the demand patterns for various products. This graph can help optimise inventory management and marketing strategies, targeting specific segments more effectively.

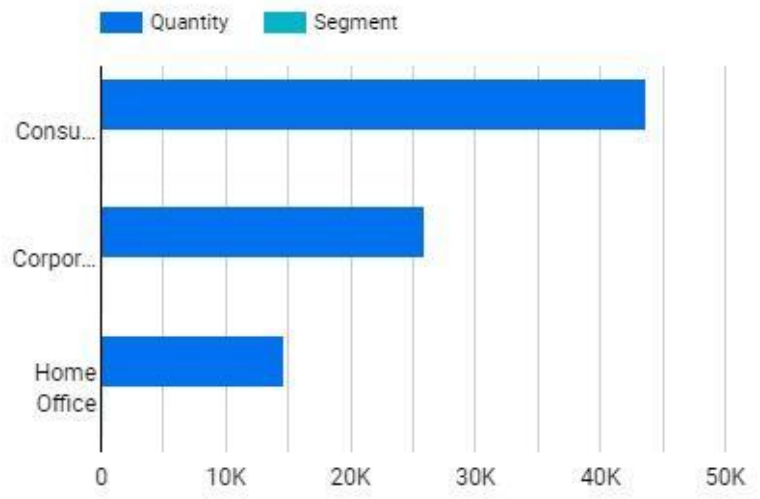


Figure 2.4 This graph provides an analysis of the quantity of items available over different dates. The data is categorised by dates, enabling insights into the distribution of products over time.
By examining the quantity of items over time, businesses can observe demand patterns and potential trends. This graph can assist in understanding seasonal variations, identifying peak demand periods, and optimising inventory management strategies accordingly.

Profit

Figure 3.1 This graph presents an analysis of the profit generated by various products, categorised by their respective category and sub-category. The data offers insights into the profitability of different product segments.

By examining the profit generated by products across different categories and sub-categories, businesses can identify high-performing segments and potential areas for improvement. This graph can aid in decision-making related to pricing, inventory management, and resource allocation to maximise overall profitability.

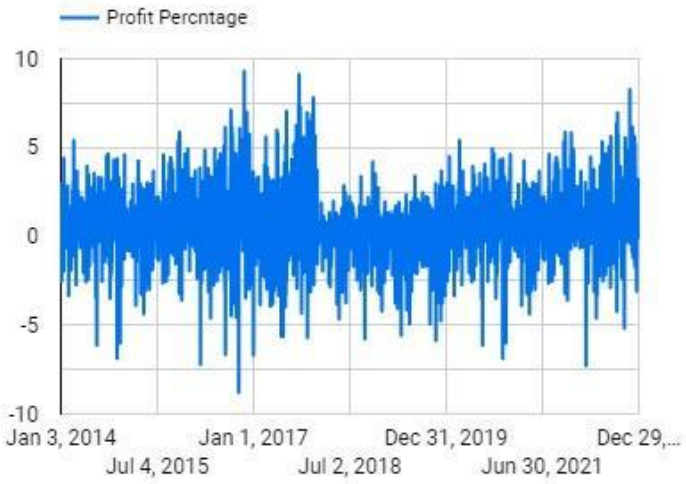


Figure 3.2 This graph provides an analysis of the profit generated in various geographic locations. The data is categorized by Country, Region, State, and City, allowing for insights into the profitability of different locations.By examining the profit in different geographic locations, businesses can gain valuable insights into the performance of their products and services in various regions. This graph can help optimise business strategies, focusing on profitable areas and addressing challenges in less profitable regions.

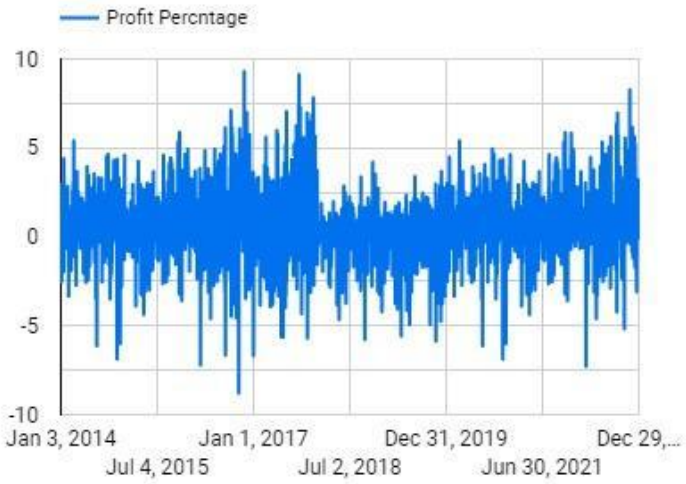
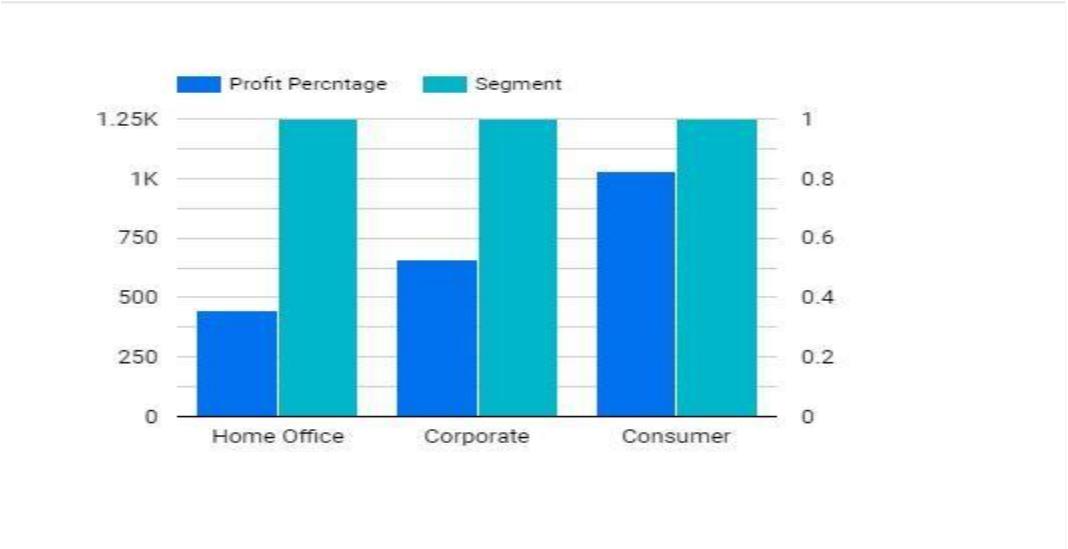


Figure 3.3 This graph provides an analysis of the profit generated in various segments. The data is categorized by segments, enabling insights into the profitability of different market segments.

By examining the profit in different segments, businesses can gain valuable insights into the performance of their products and services in specific market segments. This graph can assist in understanding which segments are the most profitable and help in making informed decisions on resource allocation and marketing strategies.



Profit Percentage

(**Figure 5.1)** This figure depicts the percentage of profit earned in each product category. It offers valuable insights into which categories yield higher or lower profit margins, aiding in identifying the most profitable product lines.

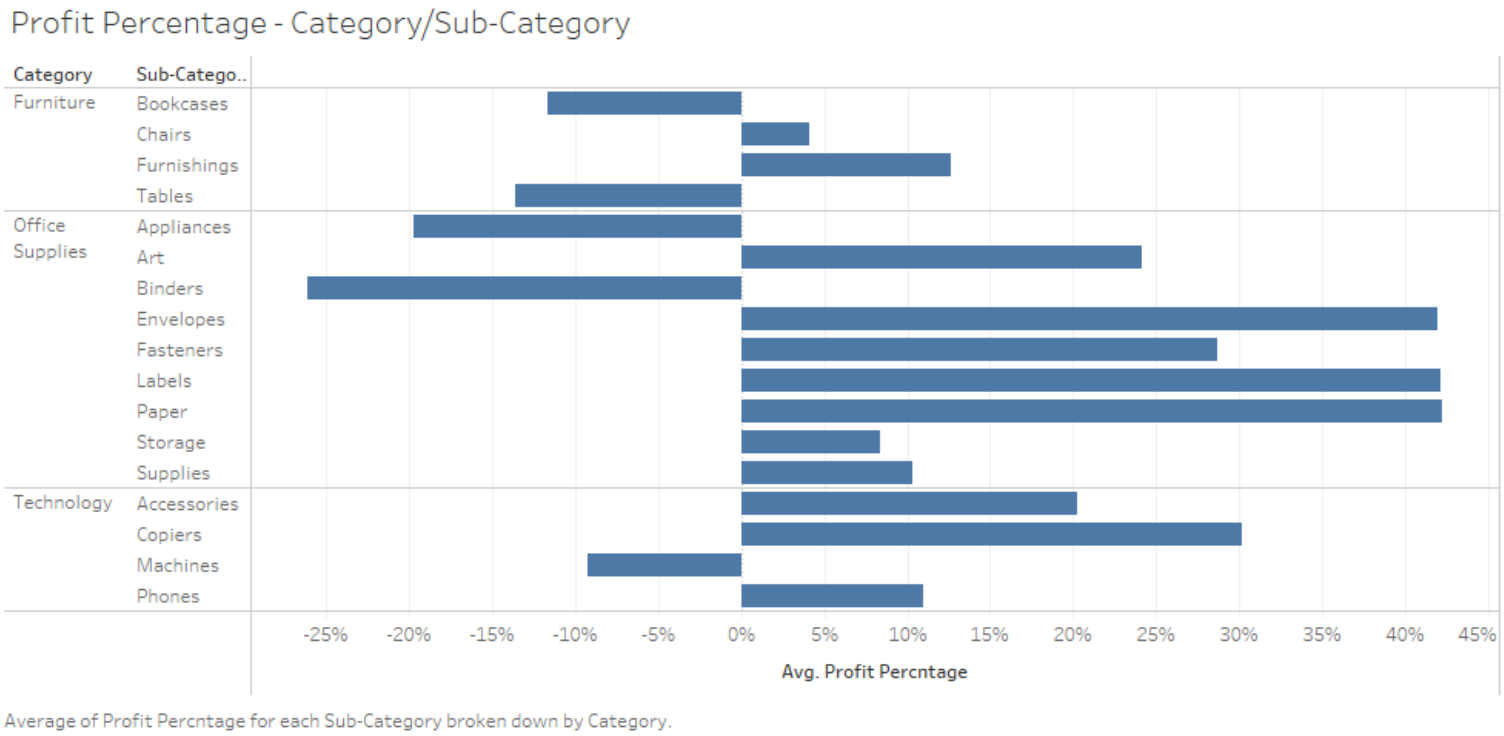


Figure 5.1: Profit Percentage - Category/Sub-Category

(Figure 5.2) This figure shows the variation in profit percentages across the United States and Canada. It helps identify geographically the strong or weak profit regions, which can inform expansion or optimization decisions.

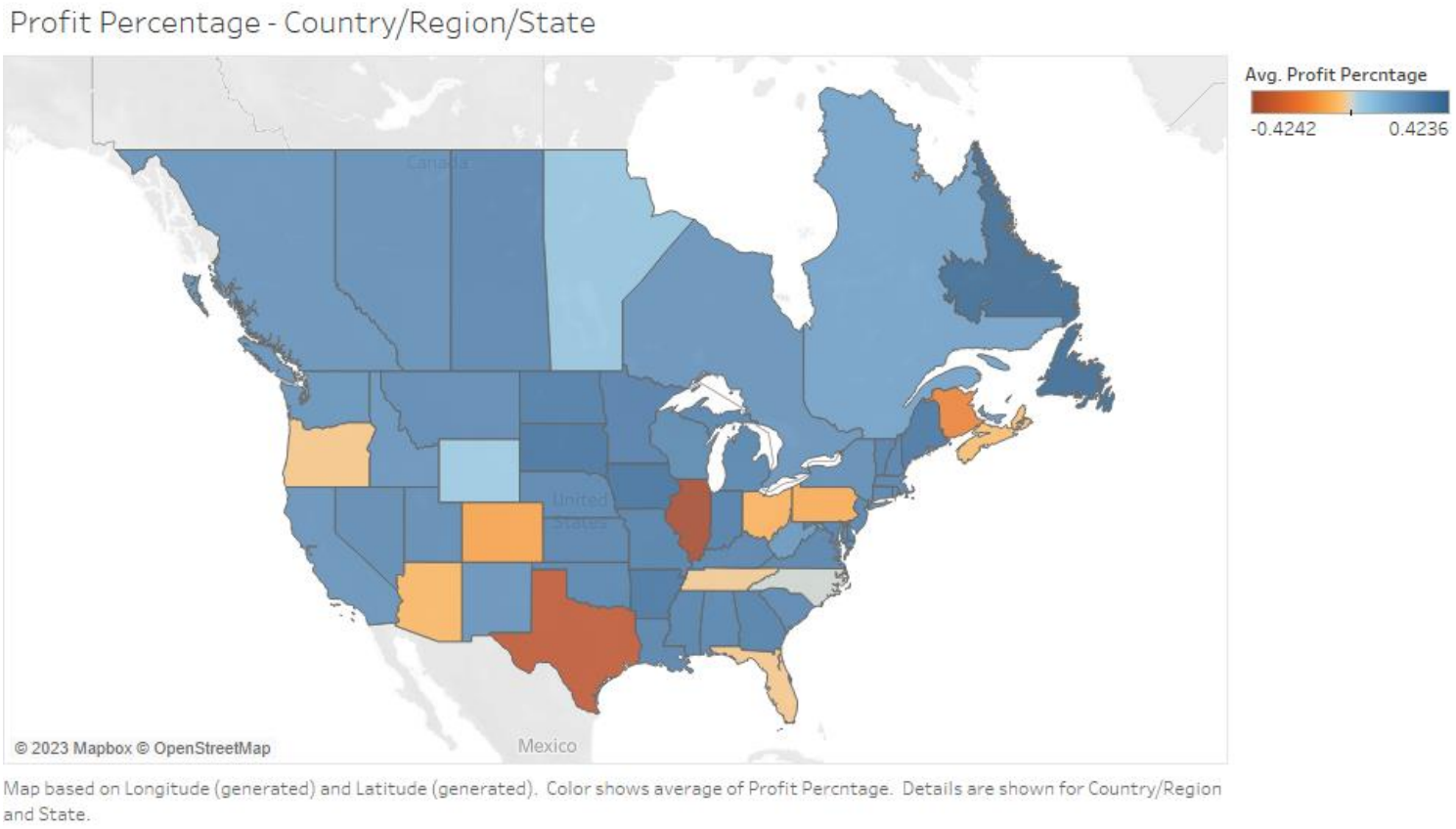


Figure 5.2: Profit Percentage - Country/Region and State

(Figure 5.3) This figure showcases the profit percentage attributed to different customer segments. By comparing segments, we can identify which customer groups contribute to the store's profits and sales strategies.

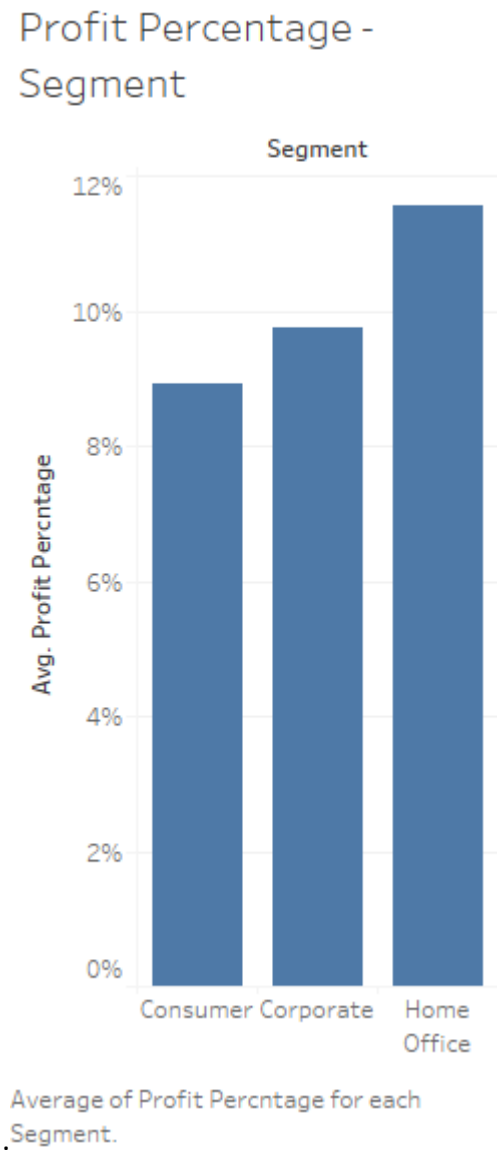


Figure 5.3: Profit Percentage - Segment

(Figure 5.4) This figure presents the trend of profit percentage over time, showing how the profitability of the superstore has changed across different dates.

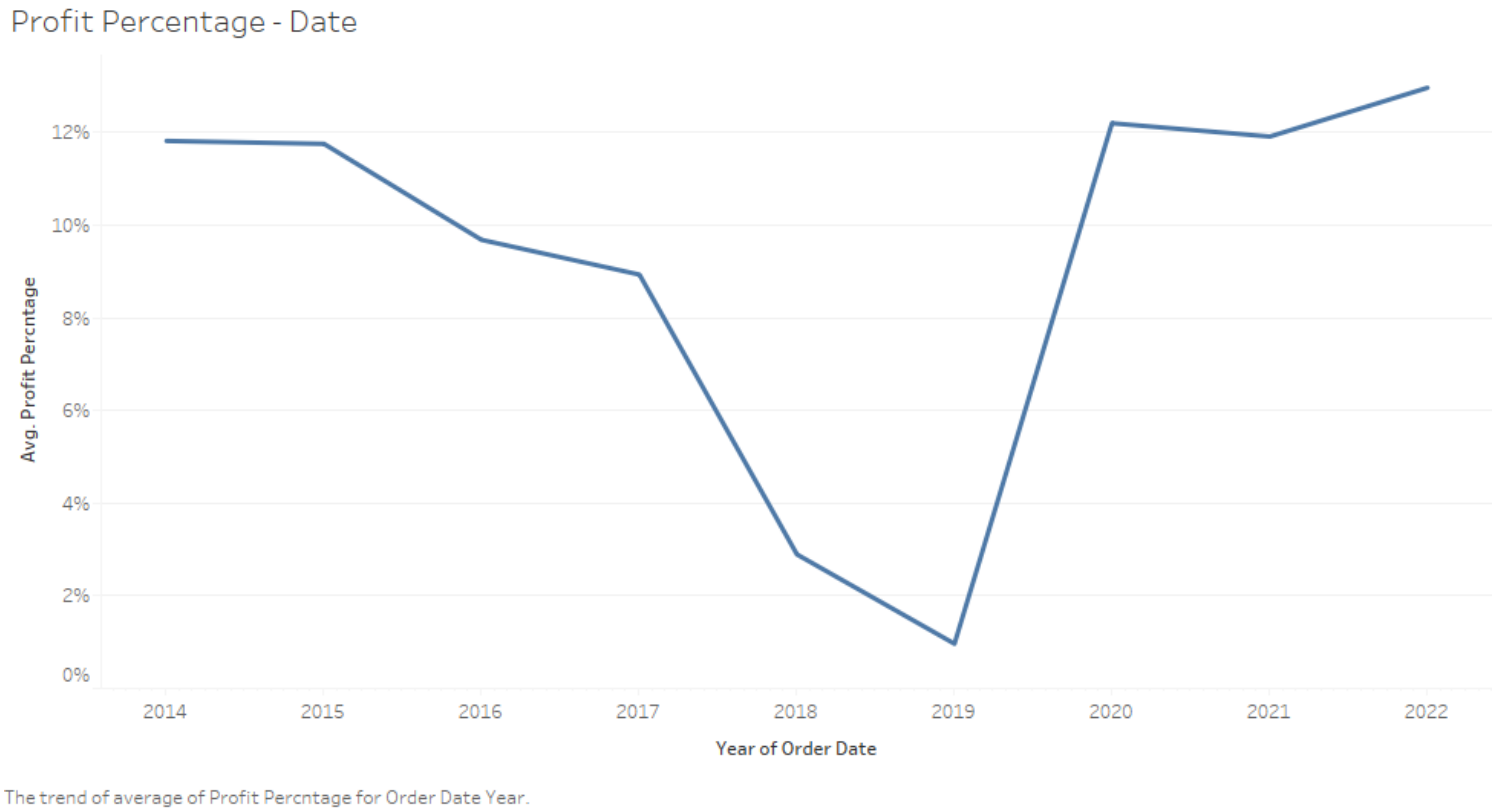


Figure 5.4: Profit Percentage - Date

Sales

(Figure 6.1) This bar chart displays the total sales for each product category and sub-category. It provides a clear comparison of sales performance across different categories and sub-category, which helps in understanding the most and least popular products.

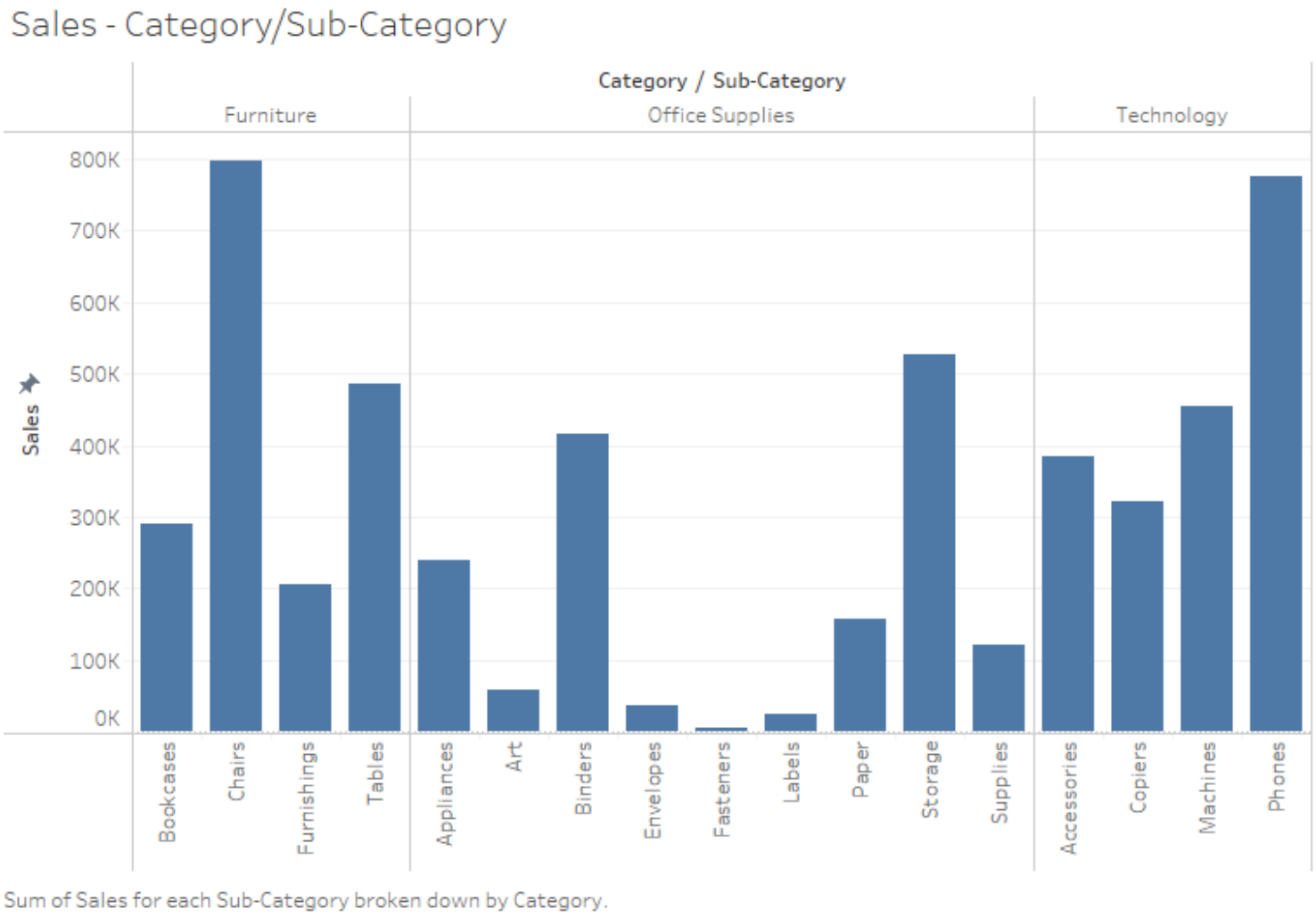
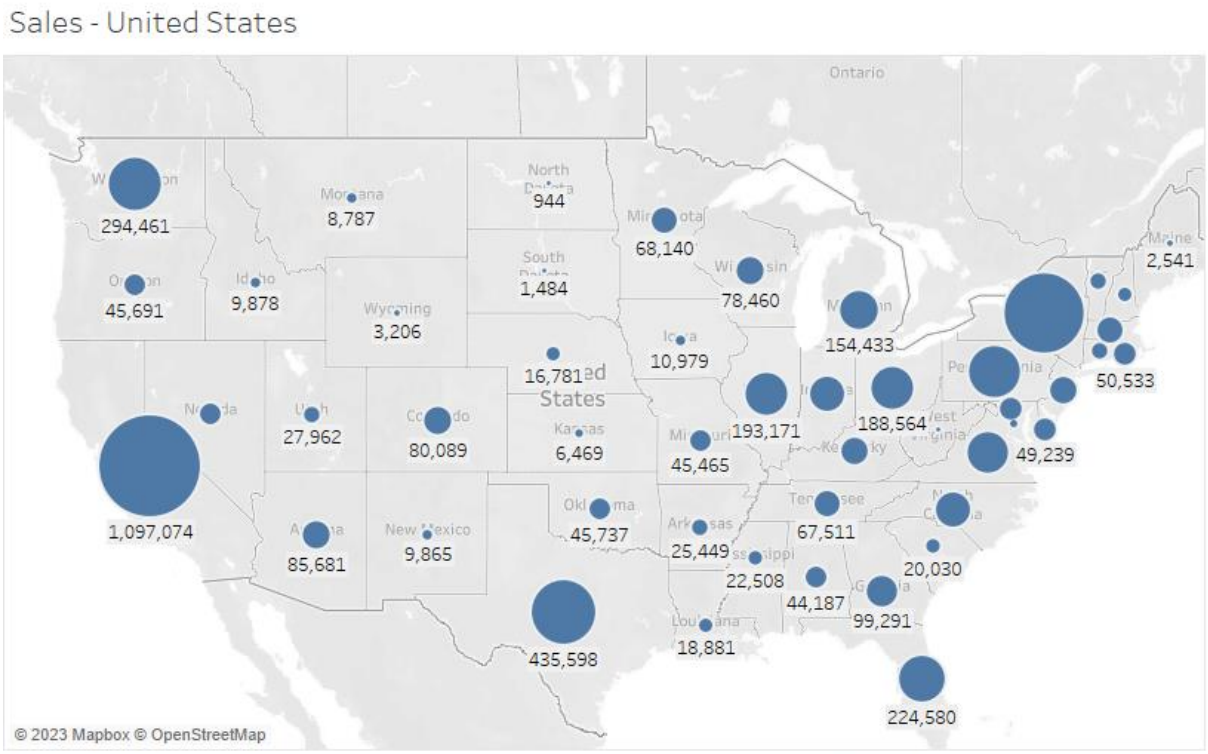


Figure 6.1: Sales - Category/Sub-Category

(Figure 6.2) This geological plot visually represents sales data for the United States and Canada, allowing you to see the distribution of sales across different regions.



Map based on Longitude (generated) and Latitude (generated). Size shows sum of Sales. Details are shown for Country/Region and State. The view is filtered on Country/Region, which keeps United States.

Figure 6.2.1: Sales - Country/State (United States)

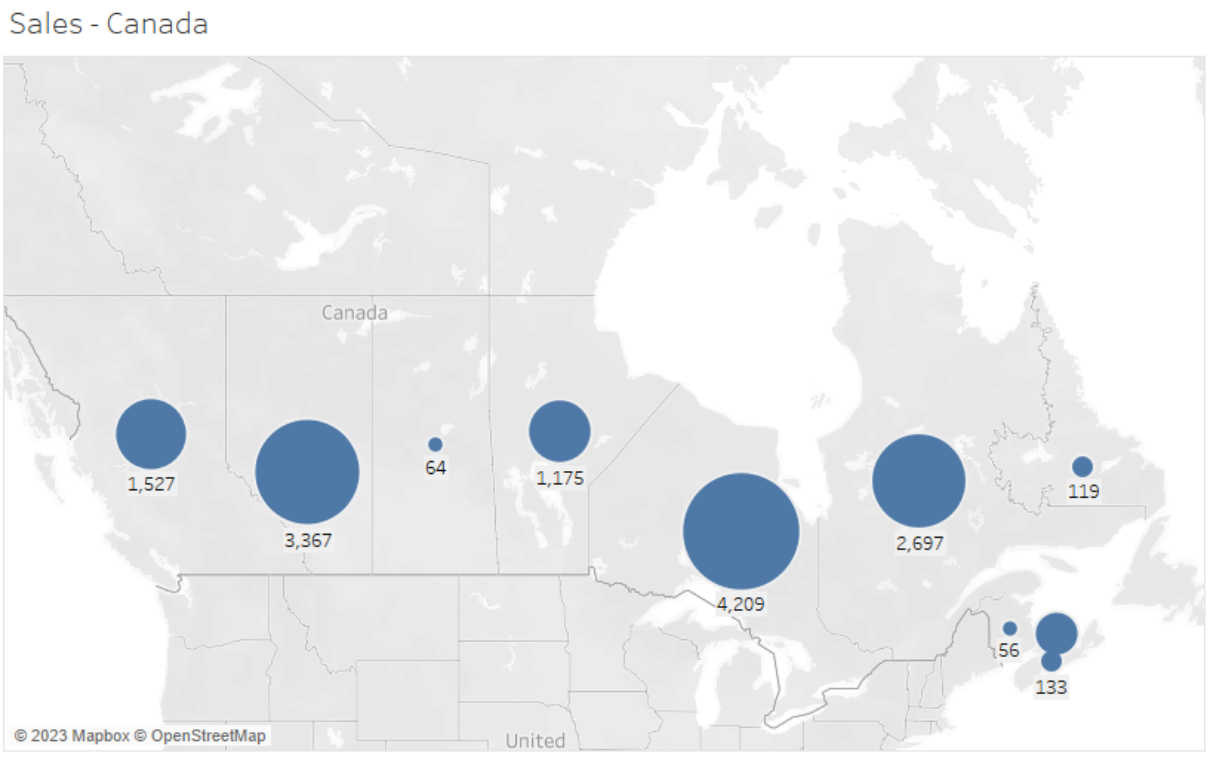


Figure 6.2.2: Sales - Country/State (Canada)

(Figure 6.3) This pie chart displays the proportion of sales attributed to each customer segment. It provides a clear understanding of the contribution of different customer groups to the overall sales revenue.

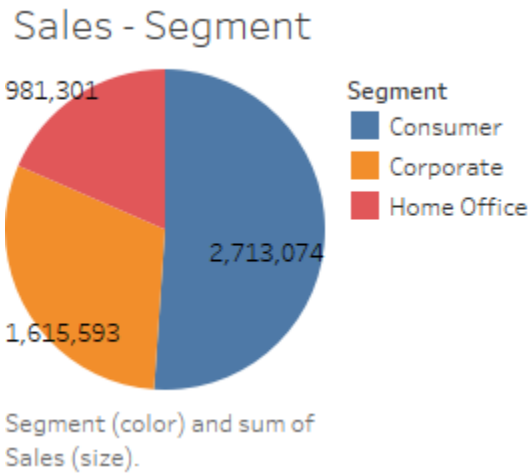


Figure 6.3: Sales - Segment

(Figure 6.4) This time series plot showcases the trend in sales over time throughout the years. It helps to identify the overall growth or decline in sales.

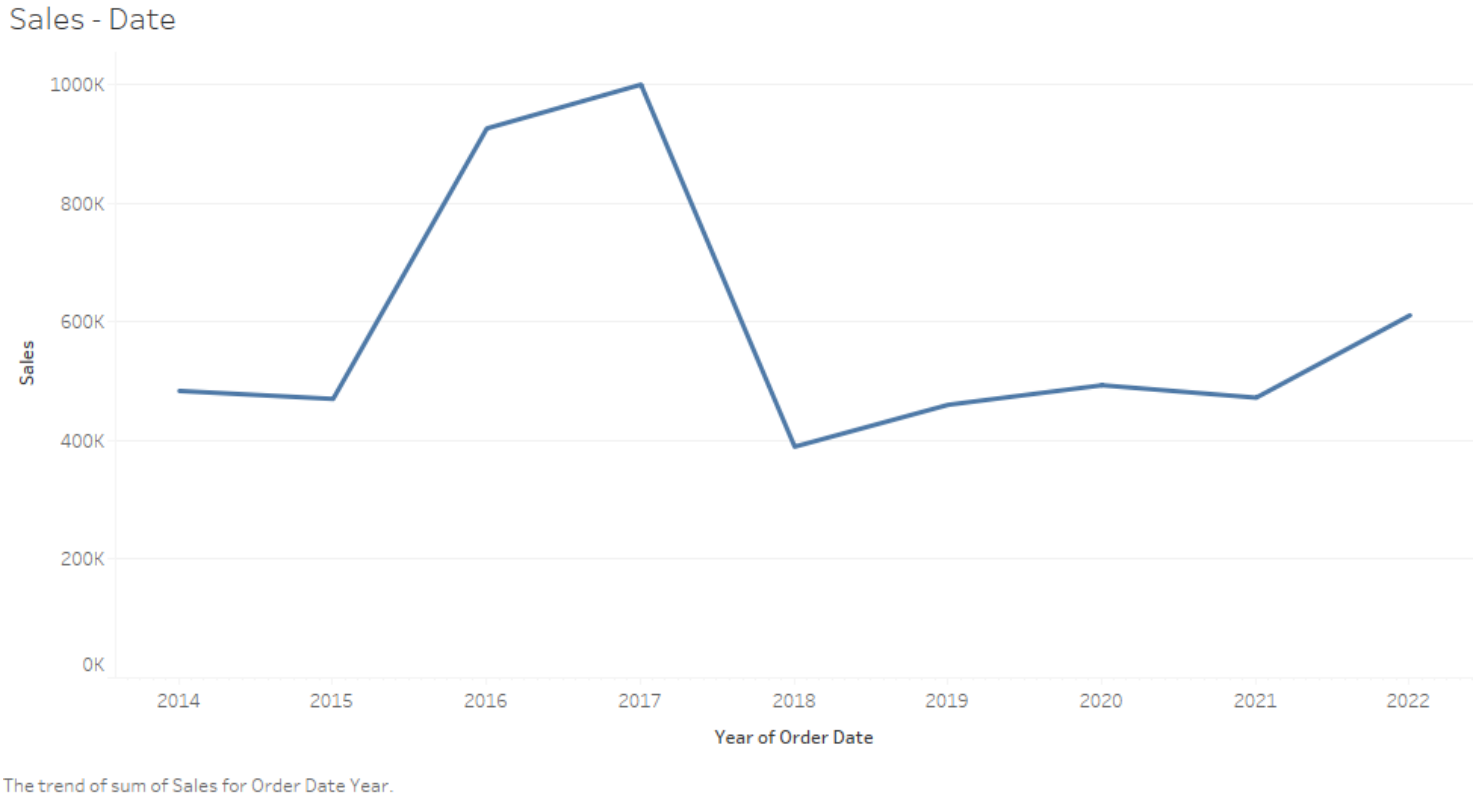


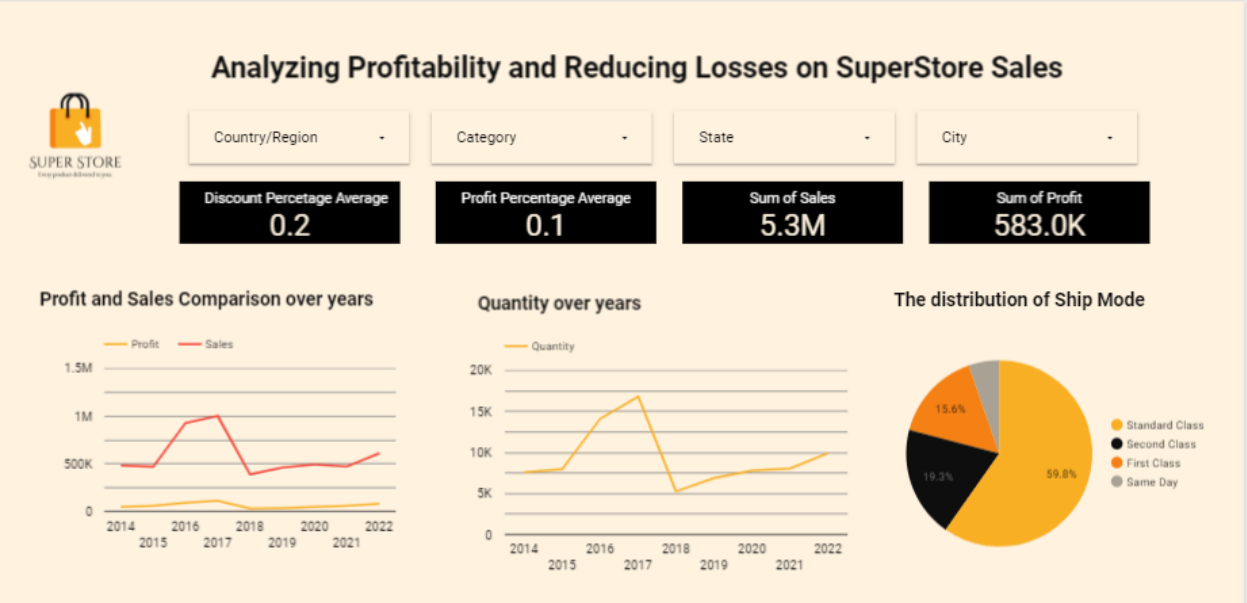
Figure 6.4: Sales - Date

Dashboard

Introduction

- The dashboard link here.
- We created 4 pages to solve our problem, below are the details of it.
- These dashboards are interactive; the only ways to do filtration are choosing options from drop-downs or choosing countries from maps.

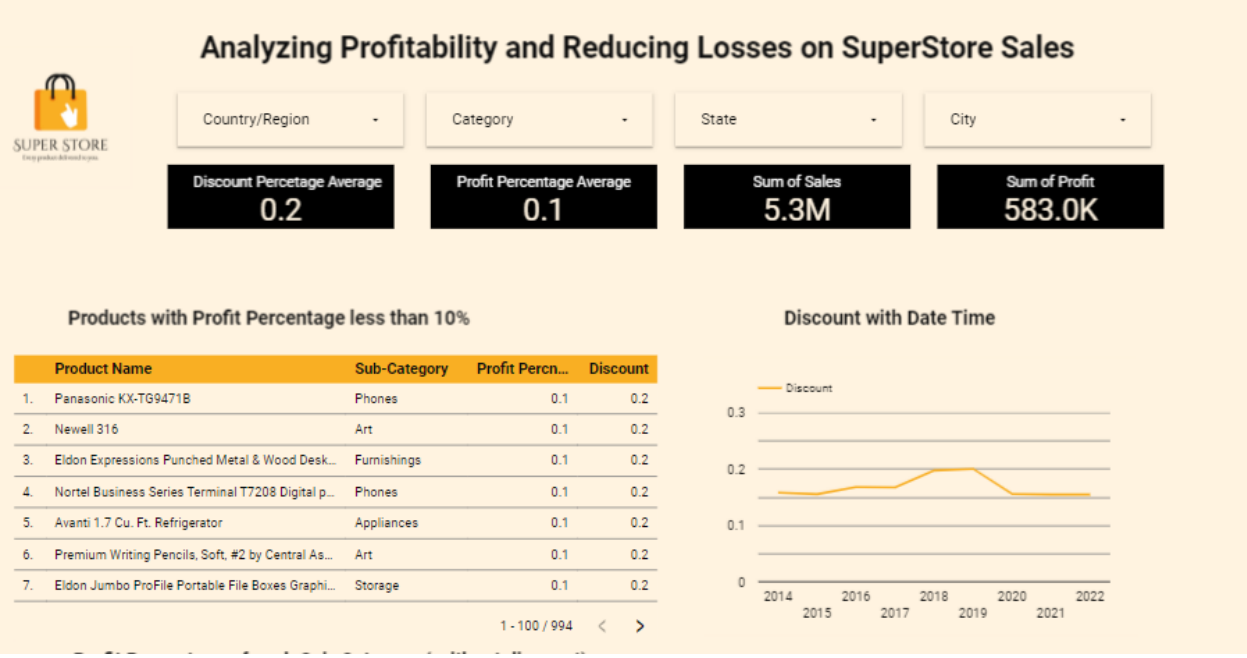
Overview Page



Sales Map Page

Profit Percentage With Discount Page

Profit Percentage Without Discount Page



	Product Name	Sub-Category	Profit Percn...	Discount
1.	Panasonic KX-TG9471B	Phones	0.1	0.2
2.	Newell 316	Art	0.1	0.2
3.	Eldon Expressions Punched Metal & Wood Desk...	Furnishings	0.1	0.2
4.	Nortel Business Series Terminal T7208 Digital p...	Phones	0.1	0.2
5.	Avanti 1.7 Cu. Ft. Refrigerator	Appliances	0.1	0.2
6.	Premium Writing Pencils, Soft, #2 by Central As...	Art	0.1	0.2
7.	Eldon Jumbo ProFile Portable File Boxes Graphi...	Storage	0.1	0.2

Conclusion

In conclusion, our comprehensive data analysis of SuperStore's sales and profitability revealed crucial insights that significantly impact the business.

By examining the relationships between various columns, we identified the key influencing factors on profit. Notably, a strong negative correlation of approximately -0.86 between discounts and profit percentage was discovered.

Consequently, we focused on subcategories with negative profit percentage and applied a reduction in special discounts. As a result, the average profit percentage increased from 0.09 to 0.26, representing a substantial improvement of 0.16.

These findings emphasize the importance of data-driven decision-making to optimize operations, reduce losses, and enhance competitiveness in the competitive retail market.

Future Work

In our analysis, we diligently addressed sub categories with a profit percentage less than 0, leading to a substantial improvement in profitability by removing discounts. However, it is crucial to recognize that in the realm of commerce, profit margins are typically gauged differently. A low margin is considered around 5%, a healthy margin at 10%, and a high margin at 20%. Despite our efforts, certain categories still exhibit profit rates below 10%. In light of this, I strongly recommend directing attention towards these specific categories and strategizing to reformulate the sale prices. By doing so, we can ensure a healthy profit margin of 10% or higher, thereby bolstering the overall financial performance and sustainability of SuperStore in the competitive retail market. This future endeavor will enable the store to make data-driven pricing decisions, optimize operations, and maintain a robust financial outlook