NGEE ANN POLYTECHNIC

School of InfoComm Technology

Data Visualisation & Story Telling

Specialist Diploma in Data Analytics

October 2023 Semester

ASSIGNMENT 1
(Individual Assignment)

Submission Deadline:

10th December 2023 (Sunday), 2359 hrs

Student Name	**:**	Yeo Ming Wei, Alger
Student Number	**:**	2489843W

Penalty for late submission:

10% of the marks will be deducted every calendar day after the deadline.

NO submission will be accepted after 17th December 2023 (Sunday), 2359 hrs

Table of Contents

1. Project Objectives

1.1 Introduction

E-commerce is becoming increasingly prevalent in today's world as we move towards the digital era. This trend has given rise to a surge in online retailers who are seeking innovative ways to boost their profitability while competing in a highly competitive market. Among the key questions arising are: **How do online retailers select their sales location? How do they optimise their sales price for different countries?** Our project involves analysing sales transaction dataset specialising in electronic goods. This dataset presents as opportunities to explore and find answers to these questions which are valuable for the electronic retailer management.

Customer Analysis

- Which customers generate the most sales revenue and profit?
- Are most of the customers from Consumer or Corporate?
- Which customers received a higher number of discounts?

Product Analysis

- What are the best-selling category and sub-category products?
- Which category and sub-category products have the highest profit?
- Which category and sub-category products have the most quantity sold?
- Are there any products with consistently high or low sales and profitability?

Finance and Price Analysis

- Does the amount of sales correlate with profit?
- What is the distribution of sales?
- What is the distribution of profit?
- What does the relationship between the distribution of sales and profit mean?
- Is there a notable increase in profit during specific months?
- Does product discount rate impact sales volume, and is there an optimal discount rate that maximizes profit?
- Does the quantity of discounts given affect profit?

Shipping Analysis

- How does the choice of ship mode affect sales volume, sales and profit?
- Are there countries where one shipping mode is preferred over others?

Geographical Analysis

- Which countries contributed to the most sales?
- Which countries contributed to the most profit and which countries are experiencing a loss?

2. Data Preparation

The data is in good condition with no significant data cleaning needed. It is also both complete and free from major issues.

Steps were taken to ensure that the Retailer Dataset V1.xlsx dataset was cleaned. FIlter function in Excel was employed to identify and exclude missing data. The sort function is then utilised to pinpoint and eliminate inaccurately inputted data. The presence of duplicated data was checked using the condition function. There appears to be some duplication in the Order ID which are meant to be unique numbers generated when an order is placed. Due to this duplication, Order ID will not be utilised in the dataset.

While conducting data cleaning, an error was identified in the spelling of the sub-category "Camera Accessoires", “TV Accessoires” and “HiFi Accessoires” within the dataset. As these errors were consistent, they were amended in the visualisation as it does not affect the outcome of the data apart from labeling.

During the visualisation process, I came across a variable (Discount) that displayed only duplicates in the graph. I carefully examined the dataset to identify any subtle distinctions that might require editing but strangely, there were no apparent errors. Interestingly, when I used a pivot table in Excel, it was able to reveal the duplicates as seen in the image below. To address this issue, I manually overwrite the values in the Discount column using the filter and fill handle feature. After confirming that the duplicates no longer existed in the data table, the data source in Tableau was replaced with the updated and edited Excel file.

3	Discount
4	0
5	0.01
6	0.05
7	0.1
8	0.15
9	0.15
10	0.19
11	0.2
12	0.22
13	0.25
14	0.27
15	0.3
16	0.35
17	0.37
18	0.4
19	0.45
20	0.45
21	0.47
22	0.5
23	0.55
24	0.57
25	0.6
26	0.67
27	0.7
28	0.8
29	Grand Total

Prior to data exploration and analysis, further research was done to better understand the context of the data and metadata. The metadata is provided in the accompanying Retailer Dataset Dictionary.xlsx file which provides descriptions of the data elements. Understanding the data definitions helps me to understand the content and relevance of each variable. I then proceed to validate the metadata to ensure that it accurately represents the data as inaccurate or incomplete metadata can lead to misunderstandings about the contents of the data.

3. Exploratory Data Analysis

The purpose of this analysis is to gain initial insights into the dataset and identify trends, patterns or anomalies, which will aid in further analysis and making informed decisions.

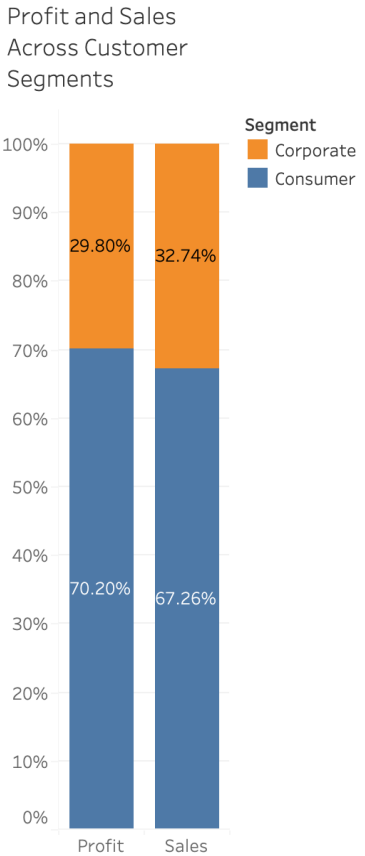
3.1 Data Description

Dataset Name: Retailer Dataset V1
Data Size: 20281 rows, 25 columns
Data Format: Excel

3.2 Customer Analysis

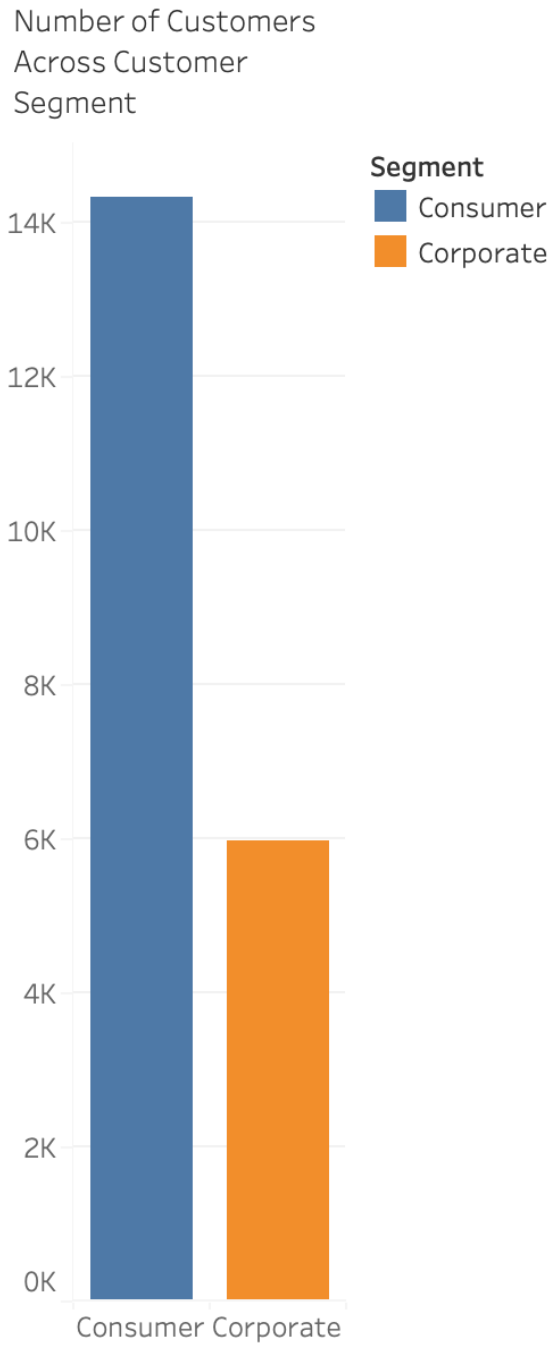
3.2.1 Which customers generate the most sales revenue and profit?

A stacked bar chart was created to compare the percentage of sales and profit between Consumer and Corporate. It is evident that Consumer has significantly contributed to more sales and profit as compared to Corporate.



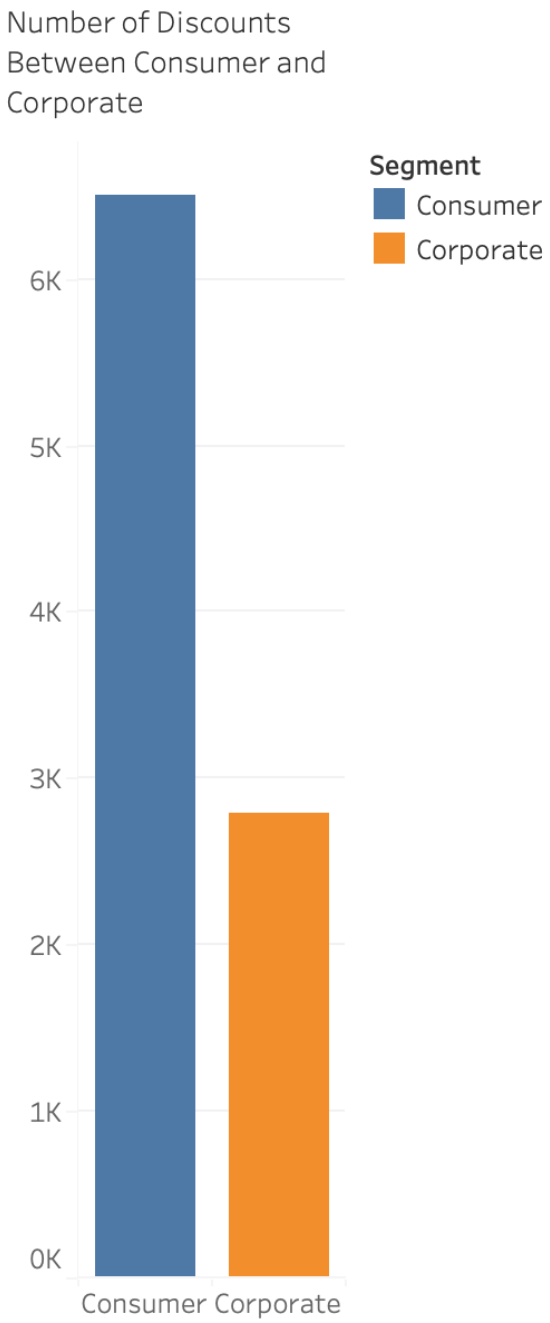
3.2.2 Are most of the customers from Consumer or Corporate?

A bar chart is created to further explore whether the number of customers in each segment relates to the sale and profit on the previous chart. The data indicates a higher number of Consumer than Corporate, suggesting a potential connection with the identified trend.



3.2.3 Which customers received a higher number of discounts?

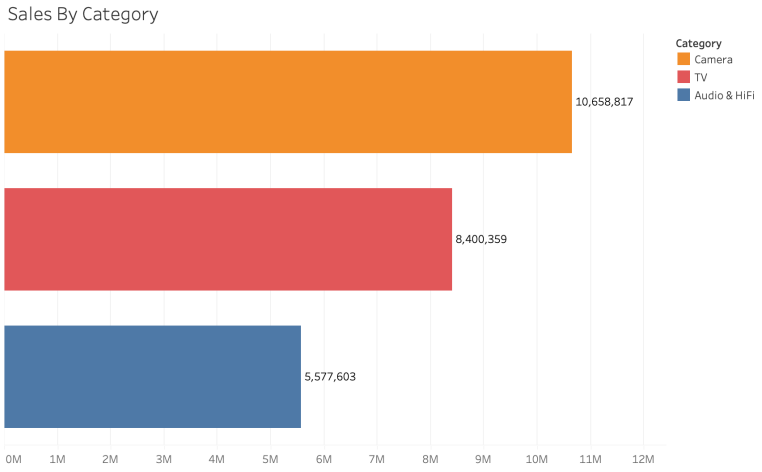
A bar chart was generated to determine whether Consumer or Corporate received a greater number of discounts. A new calculated field was generated to exclude individuals who received a 0% discount from the count. This adjustment applies to the other visualisations that involve the counting of discounts. The visual representation below indicates that Consumer received a higher number of discounts compared to Corporate.



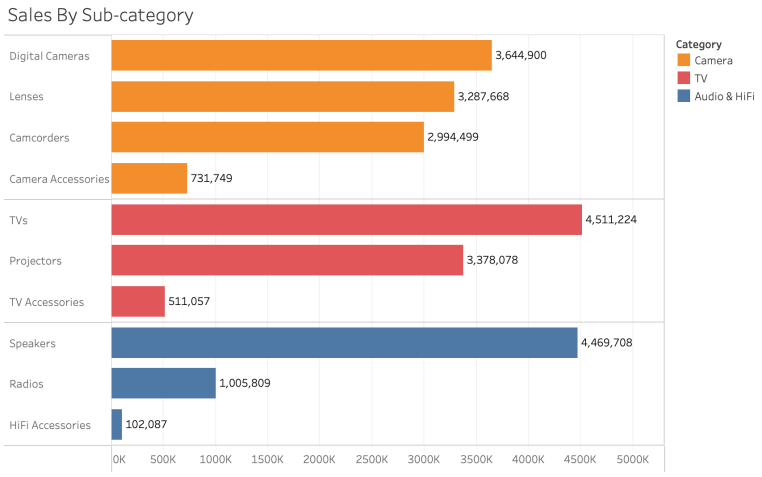
3.3 Product Analysis

3.3.1 What are the best-selling category and sub-category products?

Bar charts were created to give an overview of the sales by product categories. Camera was shown to be the best-selling product category followed by TV and Audio & Hifi. In order to further analyse the sub-categories, another bar chart was created for better visualisation.

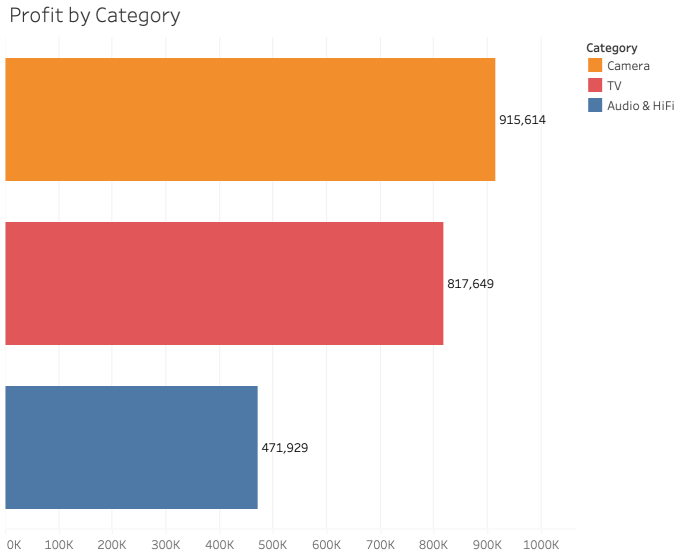


When the chart was broken down into their sub-categories, the data reveals that within the categories of TV and Audio & Hi-Fi, the best-selling sub-category products have outperformed those in the Camera category even though it was mentioned previously that Camera has the highest sales. It is possible that the Camera category achieves the highest sales as it has more sub-category products than the other categories. All accessories in each category have notably the lowest sales which can be attributed to the lower selling price.

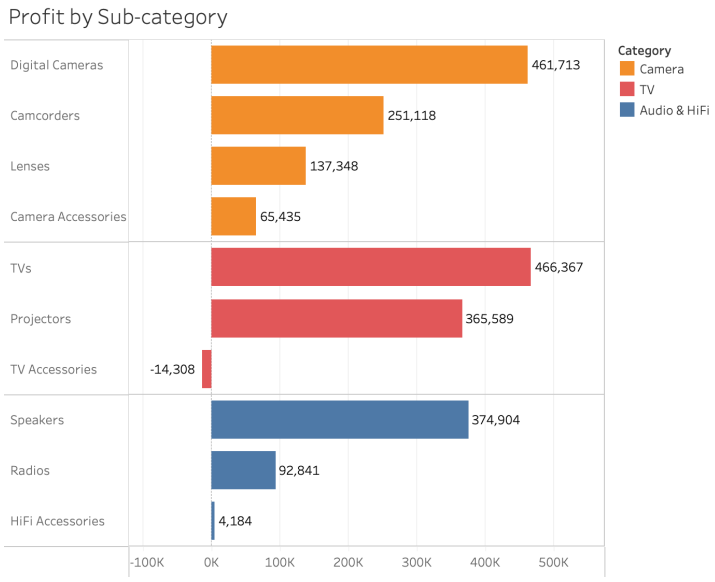


3.3.2 Which category and sub-category products have the highest profit?

Bar charts were created to give an overview of the profits across the product categories. Camera was shown to generate the highest profit followed by TV and Audio & Hifi. In order to further analyse the sub-categories, another bar was created for better visualisation.

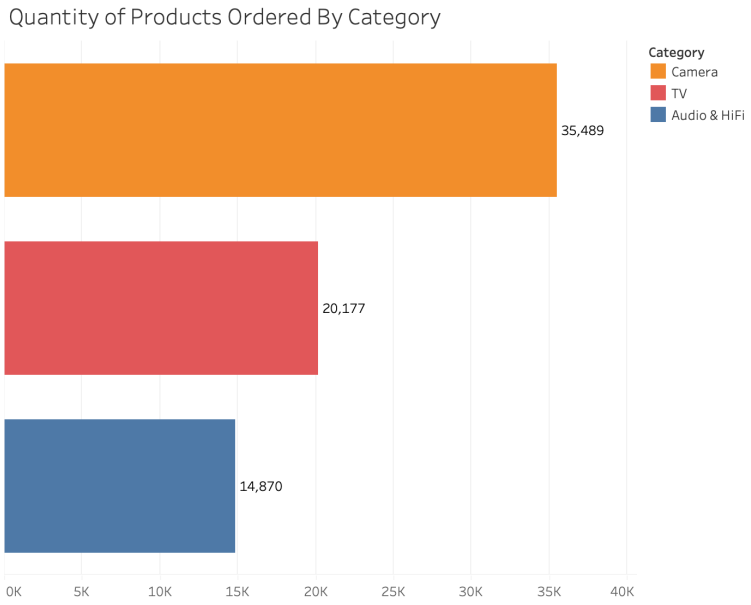


When the chart was broken down into their sub-categories, the most profitable sub-category products in each category are identical to their corresponding best-selling items in terms of sales. All accessories once again mirror their sales performance and exhibit the least profitability in each category. Notably, Hifi Accessories has almost no profit and TV accessories stand out with a recorded loss.

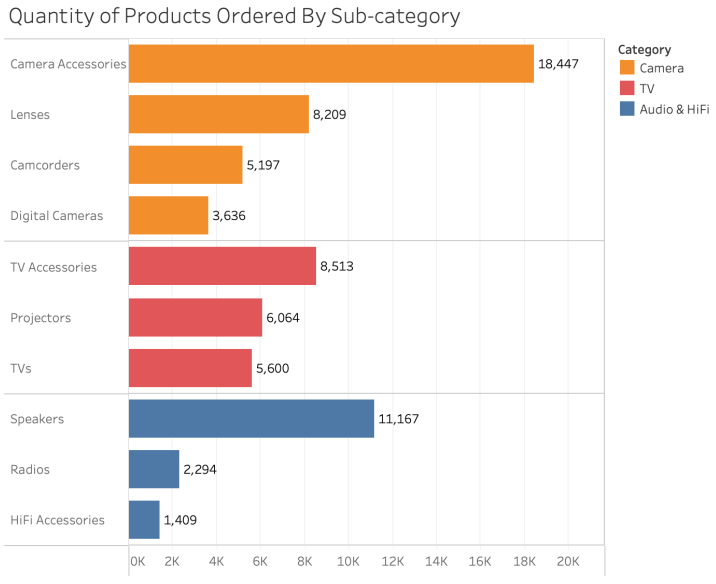


3.3.3 Which category and sub-category products have the most quantity sold?

Bar charts were created to give an overview of the quantity of products ordered across the product categories. Camera has the most quantity ordered followed by TV and Audio & Hifi. In order to further analyse the sub-categories, another bar chart was created for better visualisation.

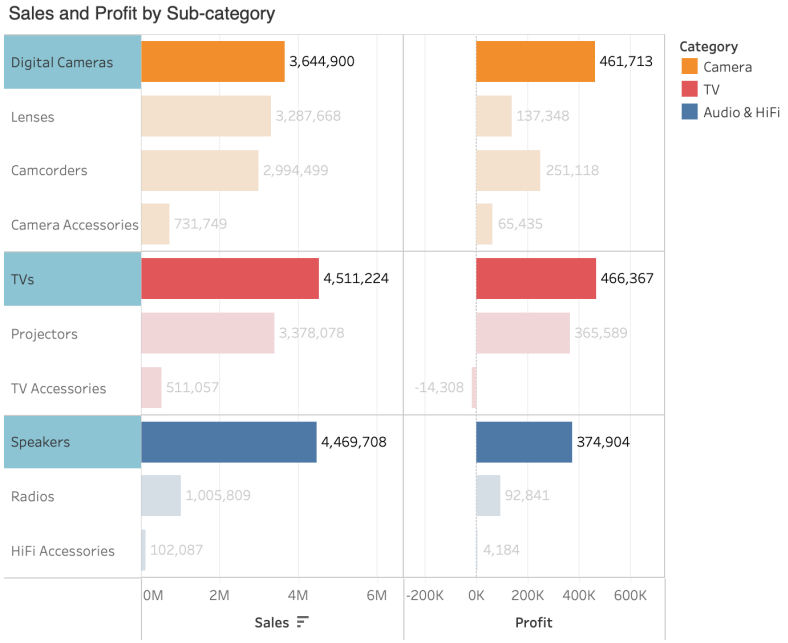


All accessories have the highest quantity ordered in their product category with the exception of Hifi Accessories. In this case, Speakers take the lead in terms of the highest quantity ordered. As Speakers contributed to the highest sales and profit in its product category, this might suggest that the sales and profit were driven by volume.

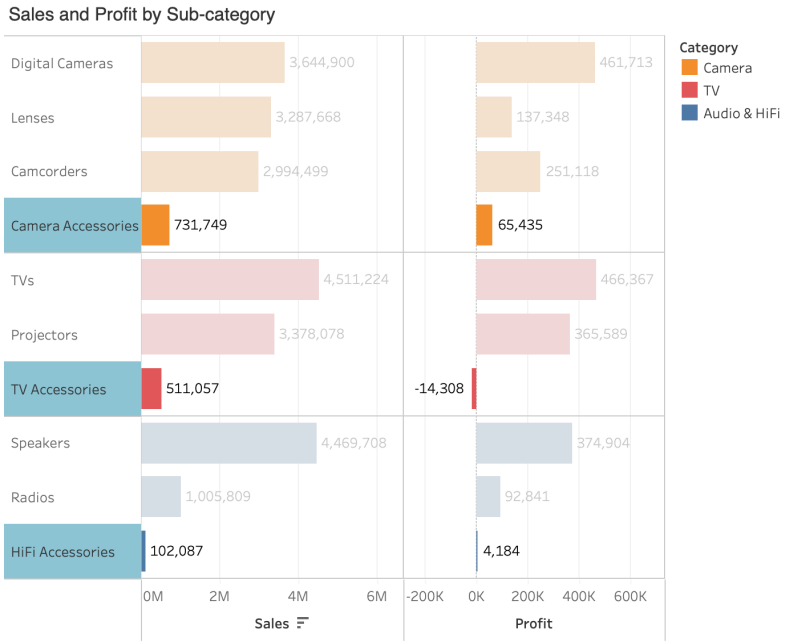


3.3.4 Are there any products with consistently high or low sales and profitability?

Bar charts that were created earlier have been combined for easier comparison. Below we can see that the selected product sub-categories namely, Digital Cameras, TVs and Speakers have consistently high sales and profitability.



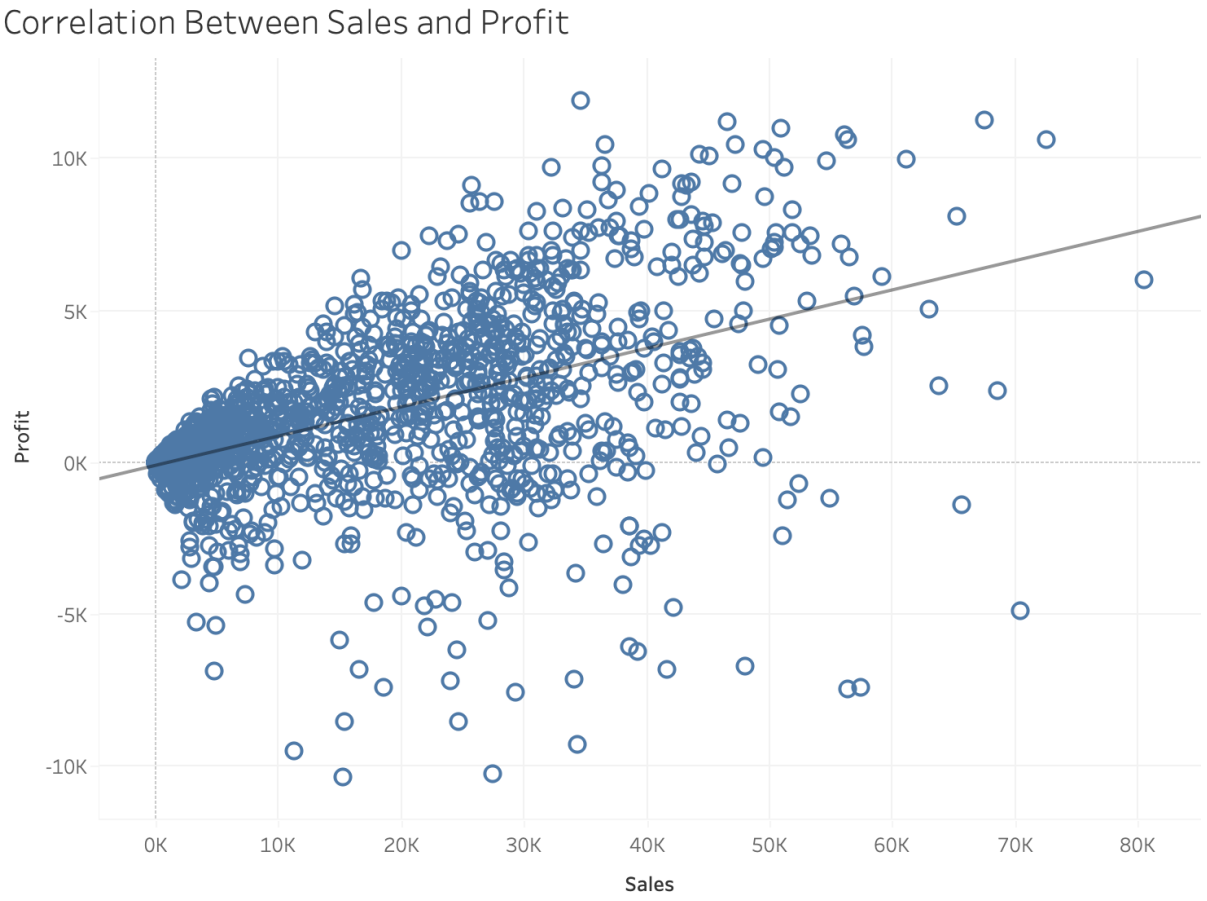
Conversely, the selected product sub-categories namely, Camera Accessories, TV Accessories and Hifi Accessories have consistently low sales and profitability.



3.4 Financial and Price Analysis

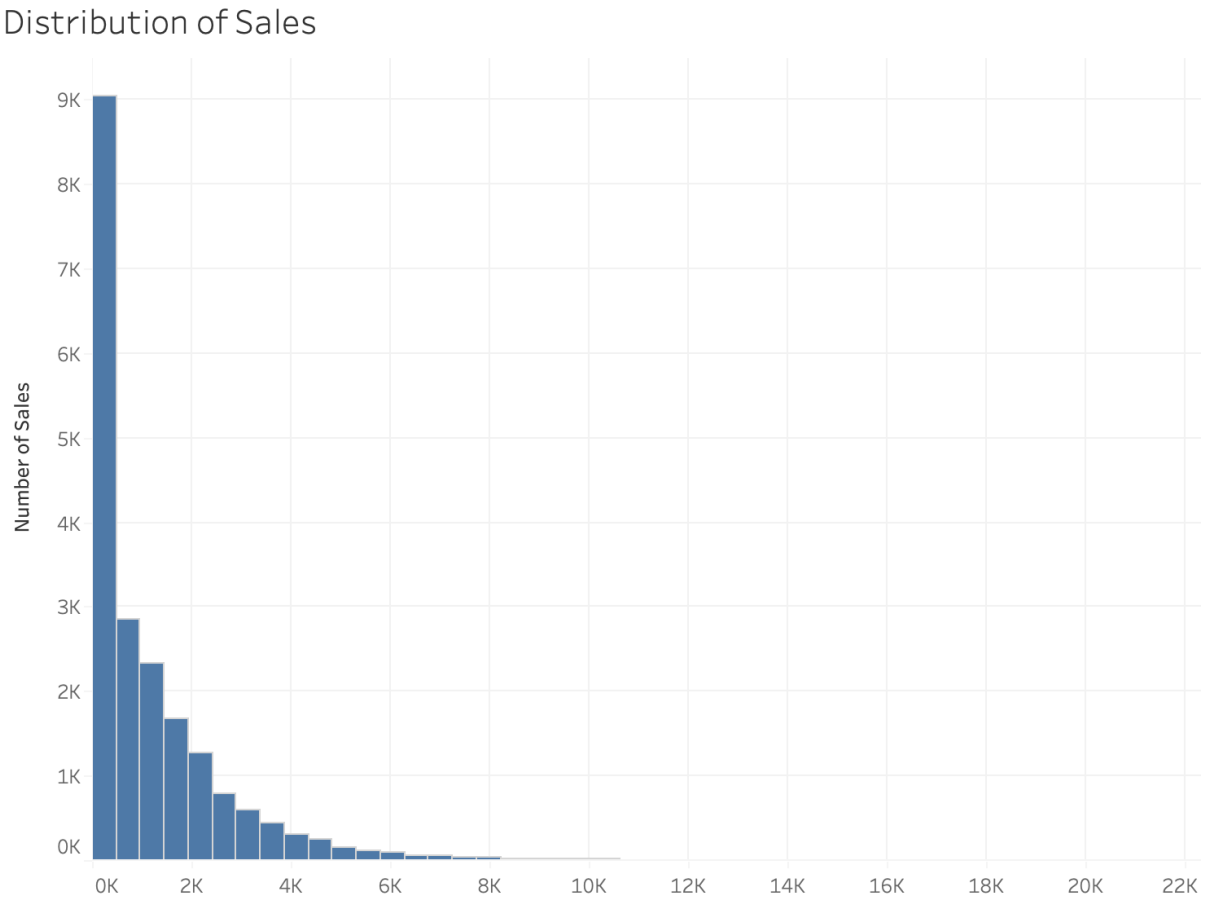
3.4.1 Does the amount of sales correlate with profit?

A scatter plot was created to analyse the trend between sales and profit. The graph below suggests a positive correlation between sales and profit, indicating that an increase in sales might result in higher profitability. Further investigation is done by looking at the distribution of sales and profit.



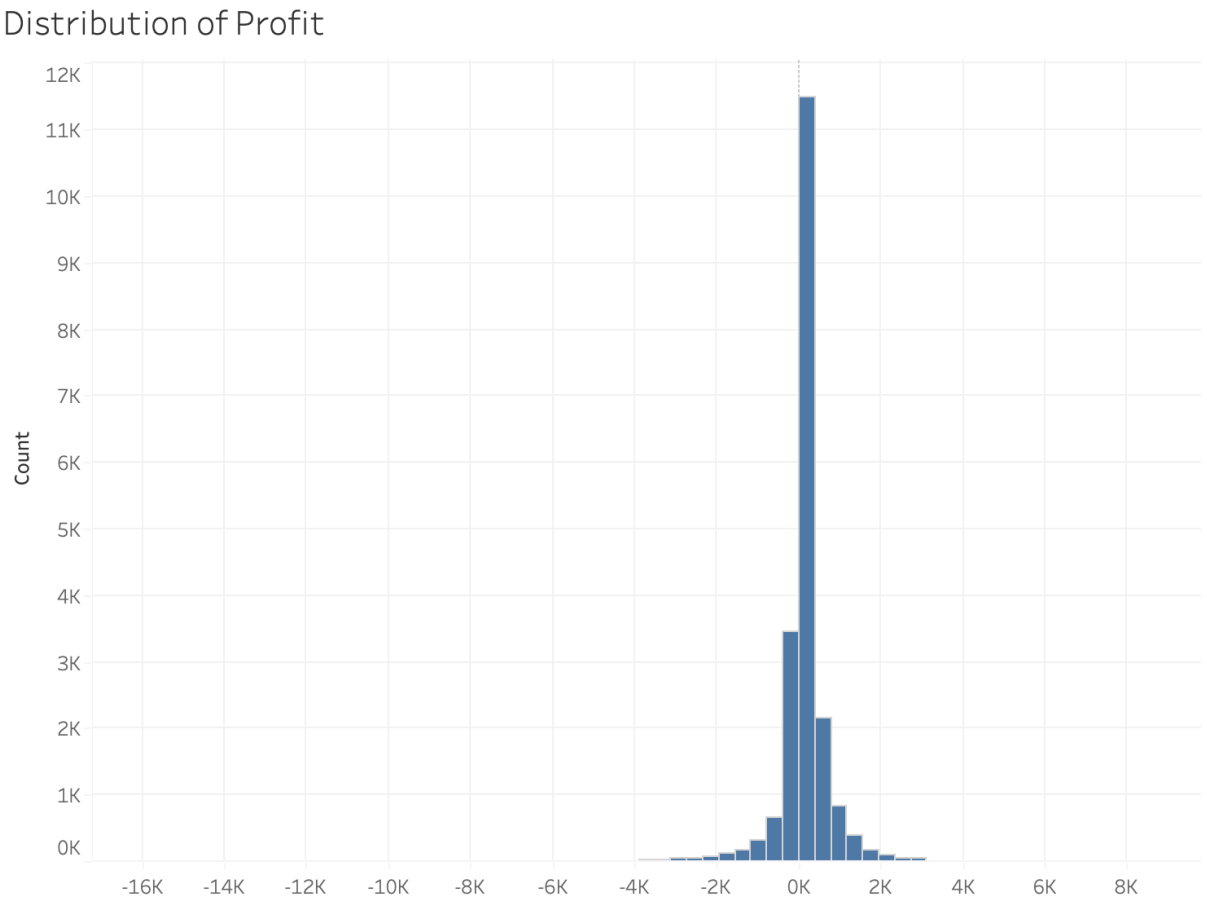
3.4.2 What is the distribution of sales?

Below is a histogram showing a positively skewed distribution of sales, where most of the sales transactions are on the lower end. The extended range on the x-axis is due to a few outliers which contributed to higher transactions which cannot be visualised on the graph due to the small value. They are not removed as they are not related to incorrect data entry and could provide valuable insights. This approach acknowledges that the sales represented by the outliers may be legitimate and warrant further exploration.



3.4.3 What is the distribution of profit?

Below is a histogram showing a normal distribution of profit which signifies that the majority of profit values cluster around the center indicating a balanced spread of profits. The extended range on the x-axis is due to a few outliers which contributed to higher profit and loss and they cannot be visualised due to its small value. They are not removed as they are not related to incorrect data entry and could be further investigated.

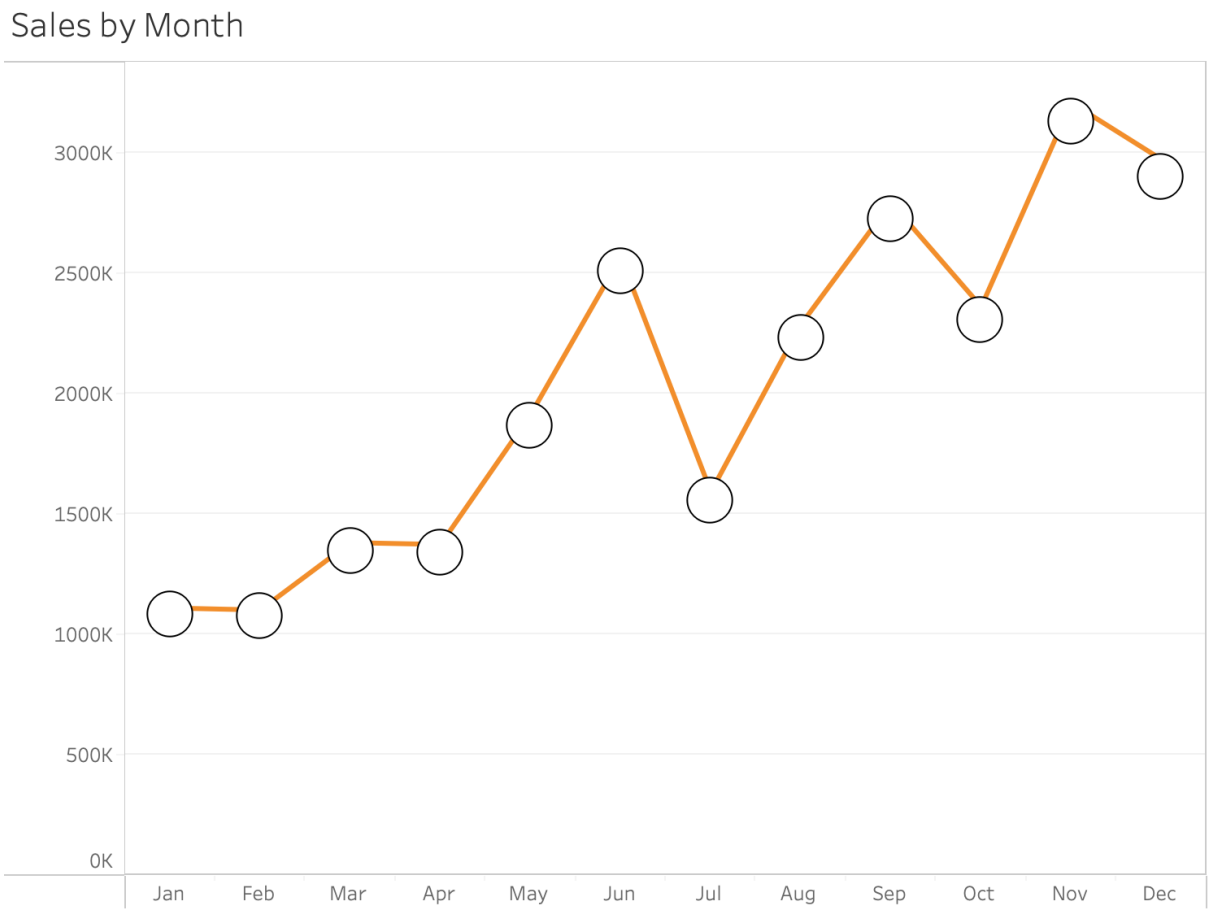


3.4.4 What does the relationship between sales distribution and profit signify?

A positively skewed distribution of sales and normal distribution of profit could mean that there may be a disconnection or asymmetry between the two variables. The normal distribution of profit could indicate stability and predictability in profit has a weak relation with variability in sales.

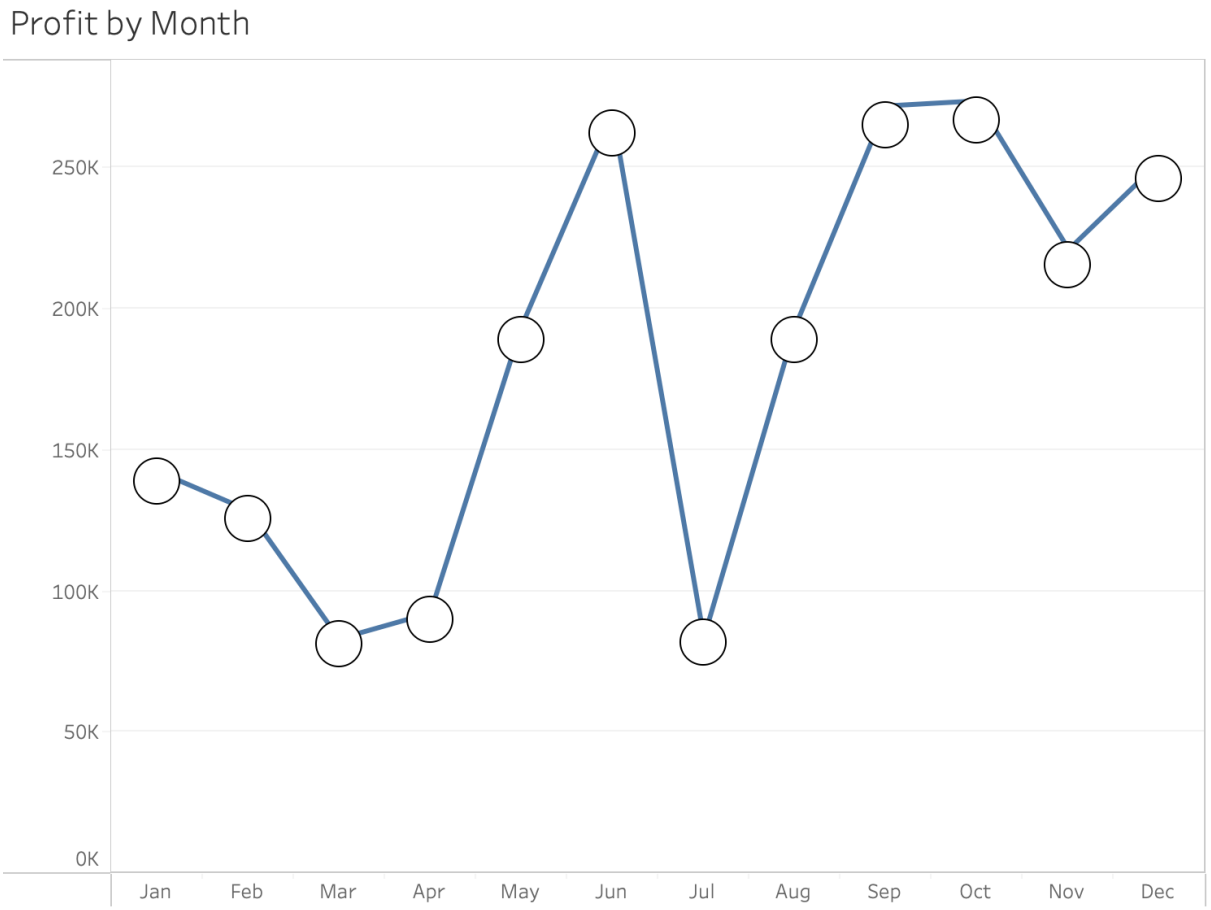
3.4.5 Is there a notable increase in sales during specific months?

A line chart has been generated to examine the monthly sales trend. Sales exhibit the lowest figures at the beginning of the year, but there is an upward trend as the year progresses. Three peaks were observed, particularly in June, September, and November. The most significant decline in sales occurred in the month of July.



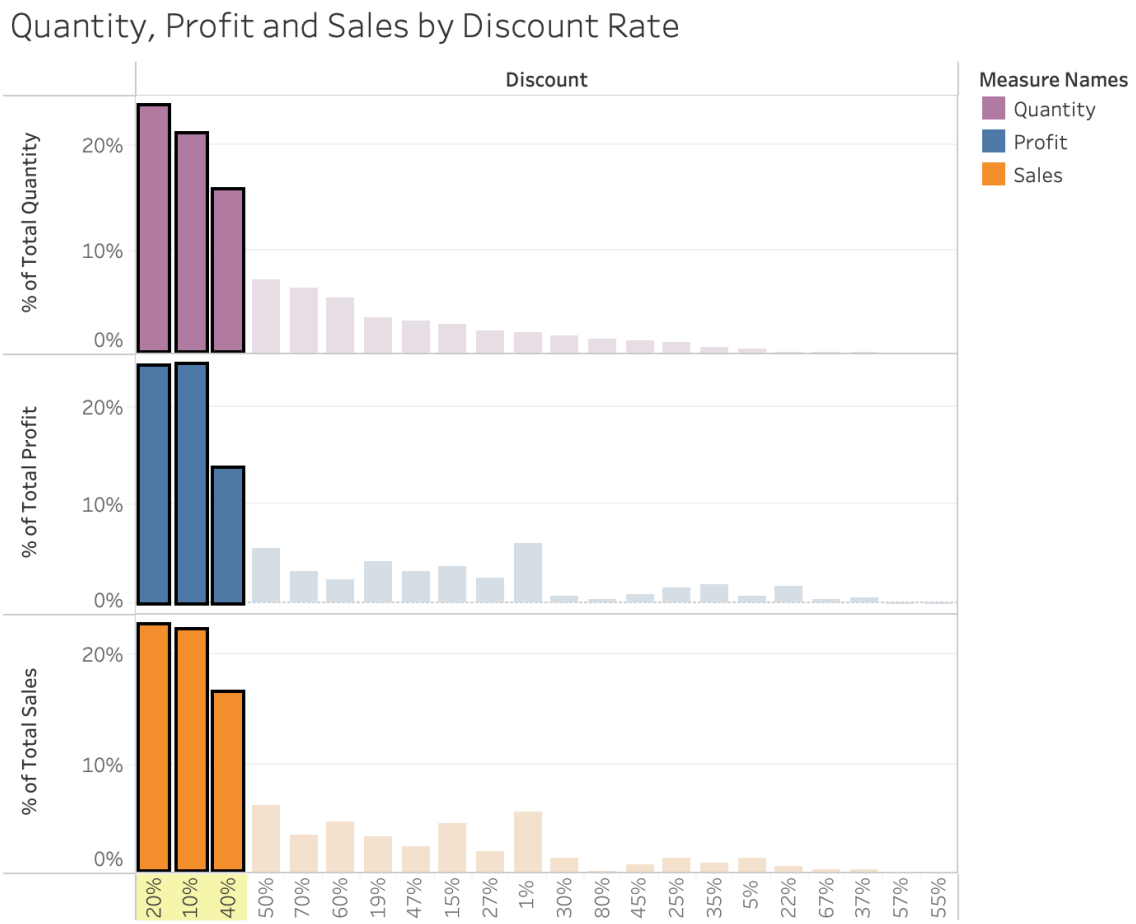
3.4.6 Is there a notable increase in profit during specific months?

A line chart has been generated to examine the monthly profit trend. The highest profits were found in June, September and October. There is a steep decline in profit in July and interestingly, a slight decline in November despite having the highest sales throughout the year as mentioned earlier.



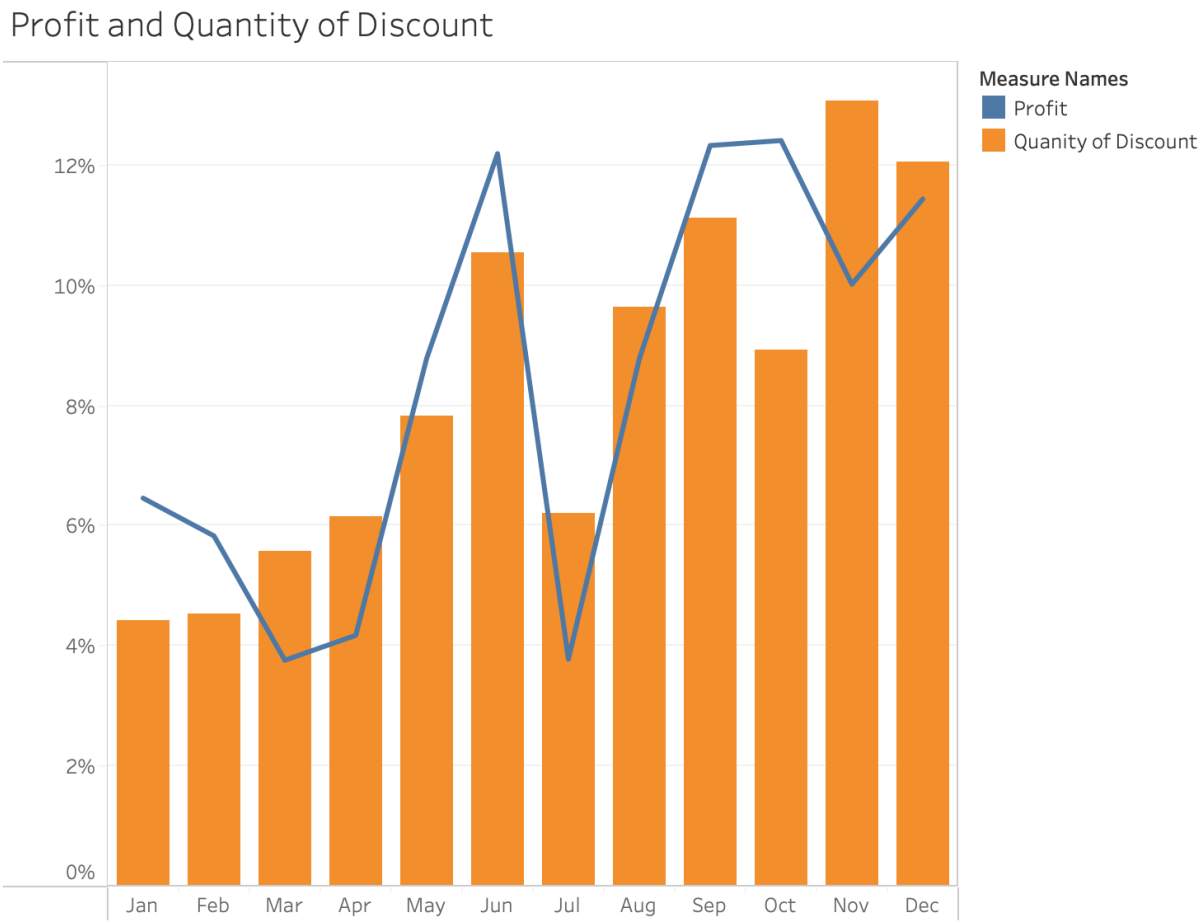
3.4.7 Does product discount rate impact sales volume, and is there an optimal discount rate that maximizes profit?

Multiple bar charts were created to better compare discounted rates with sales volume, sales revenue and quantity of orders. The bar chart below reveals that high discount rate does not impact sales volume. However, a discounted rate of 10%, 20% and 40% seem to have produced the highest profit among the other discounted rates.



3.4.8 Does the quantity of discounts given affect profit?

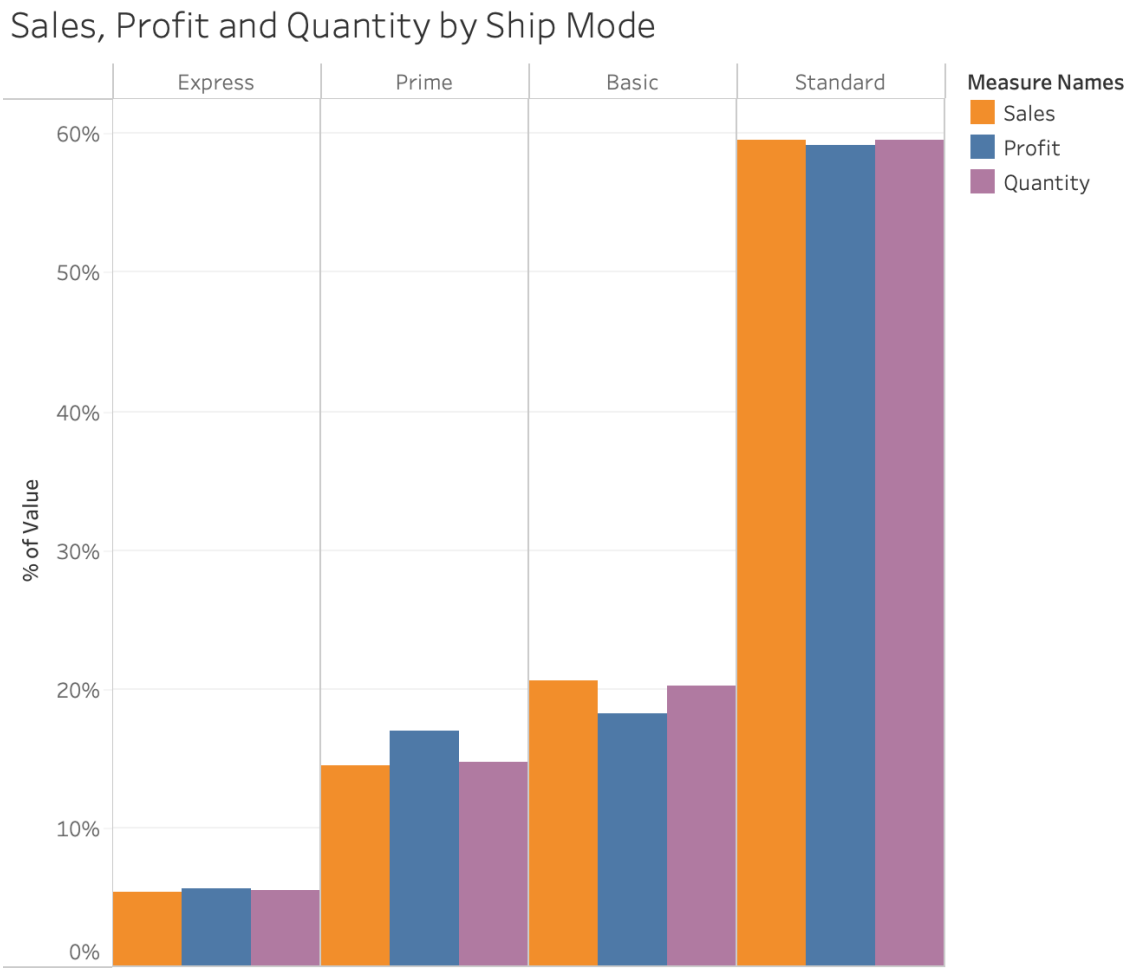
The quantity of discounts in each month of the year was explored with a dual axis bar chart. From the visualisation below, it is evident that profit follows closely with the quantity of discount.



3.5 Shipping Analysis

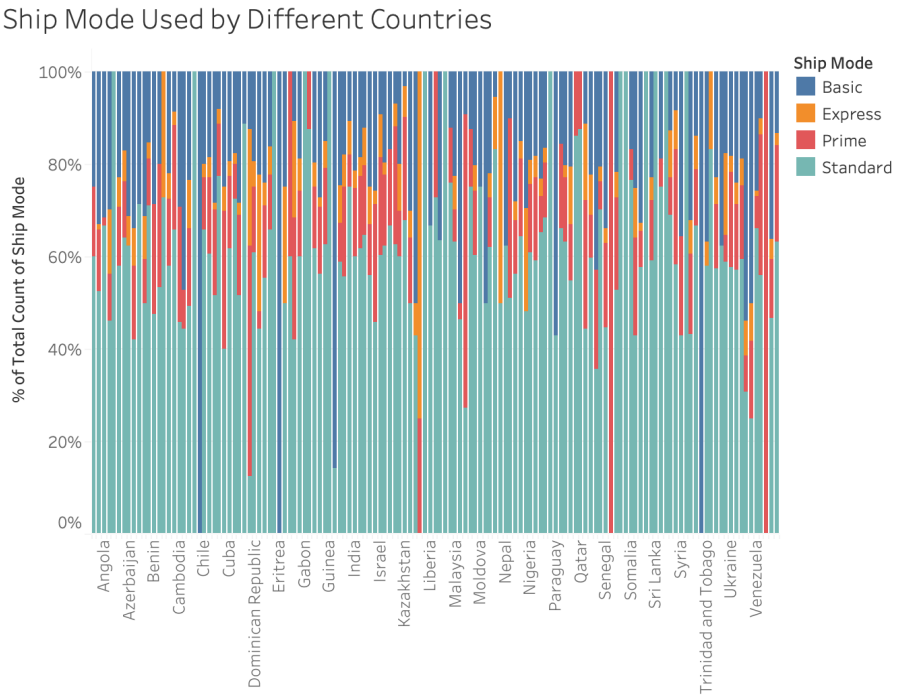
3.5.1 How does the choice of ship mode affect sales volume, sales and profit?

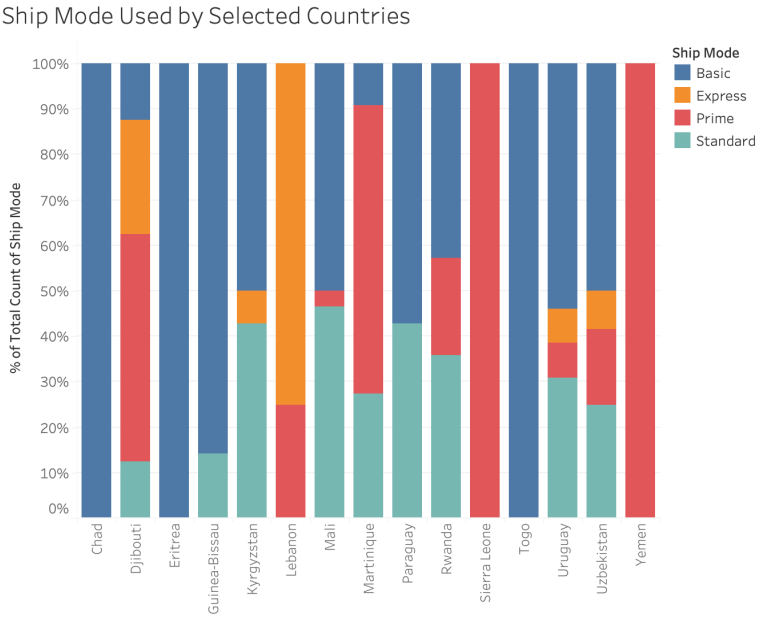
A grouped bar chart is created and it is evident that standard shipping has generated the highest percentages in terms of sales, profit and quantity. The standard shipping mode is typically more cost-effective and this factor could be contributing to its highest percentage across all the variables.



3.5.2 Are there countries where one shipping mode is preferred over others?

In the stacked bar chart, it is evident that the majority of countries have a preference for Standard shipping. However, there are a few countries that prefer a specific shipping mode other than Standard. Due to the complexity of the chart, I have narrowed down the focus to countries displaying a distinct preference for a particular shipping mode over Standard with a separate stacked bar chart.

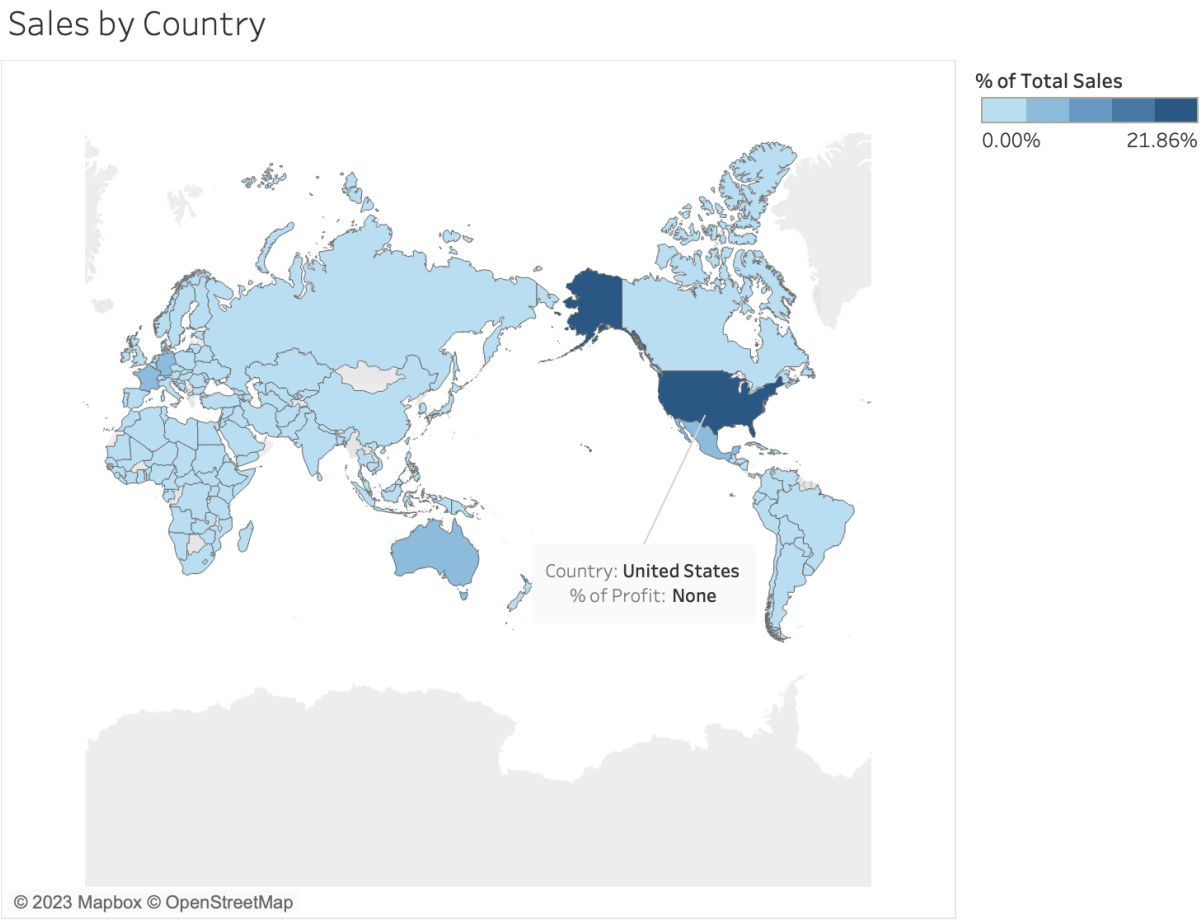




3.6 Geographical Analysis

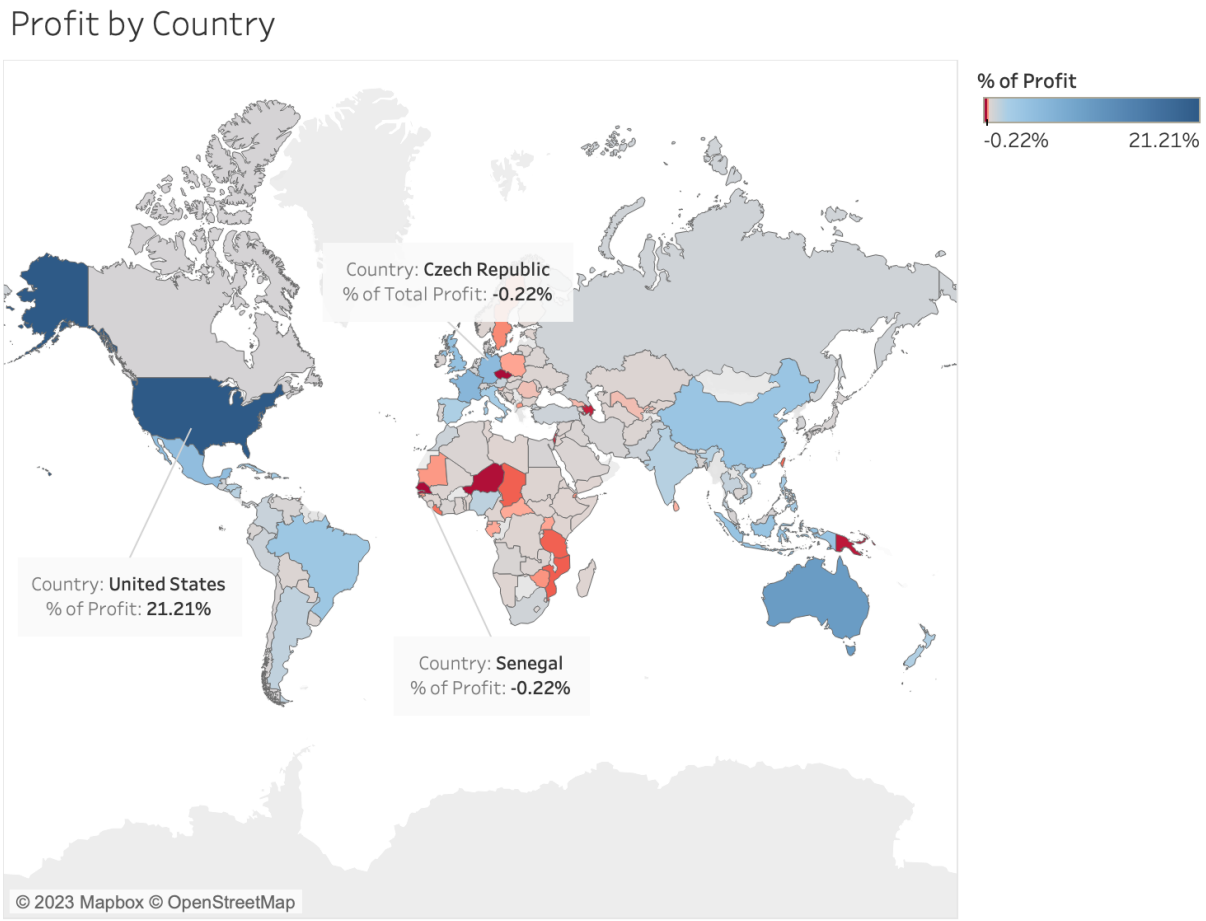
3.6.1 Which countries contributed to the most sales?

A choropleth map was created to better visualise countries the contribution of sales from different countries. The United States has the highest percentage of sales as noted by the darker shade.



3.6.2 Which countries contributed to the most profit and which countries are experiencing a loss?

A choropleth map was created to better visualise countries the contribution of profits from different countries. The United States has made the highest percentage contribution to profit while countries highlighted in red indicate those experiencing losses. Czech Republic and Senegal share the lowest contribution to profit.

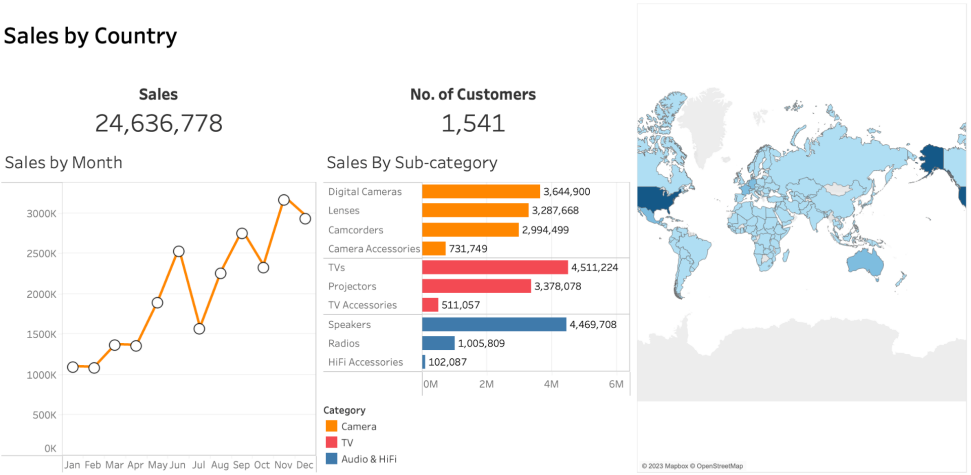


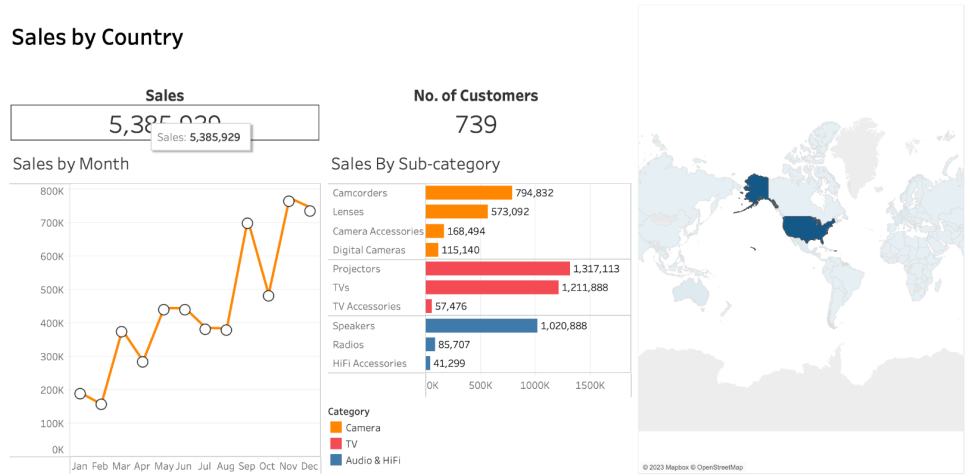
4. Dashboard

As mentioned earlier in the report, below are the key questions which might interest management.

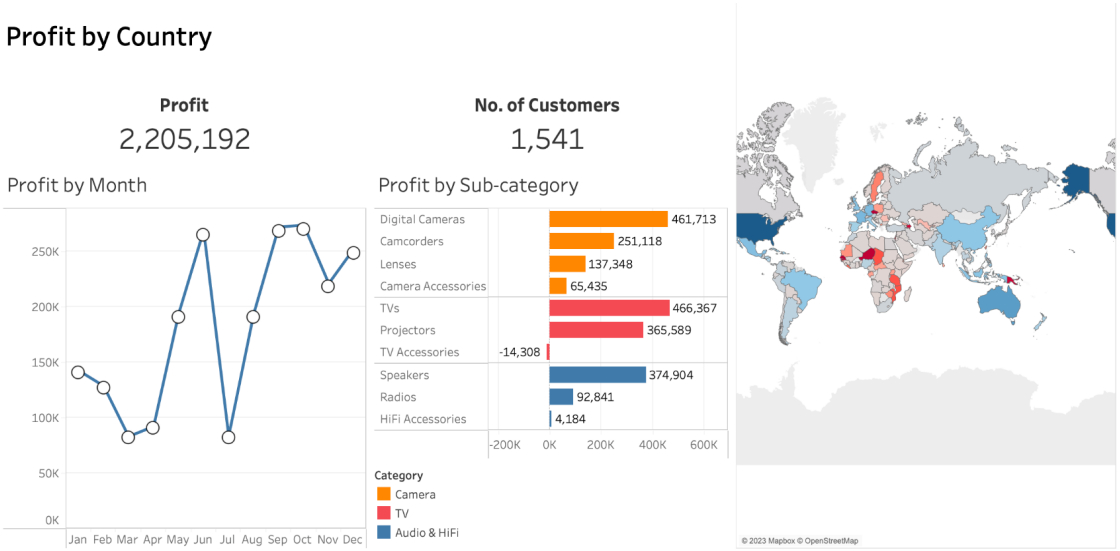
4.1 How do online retailers select their sales location?

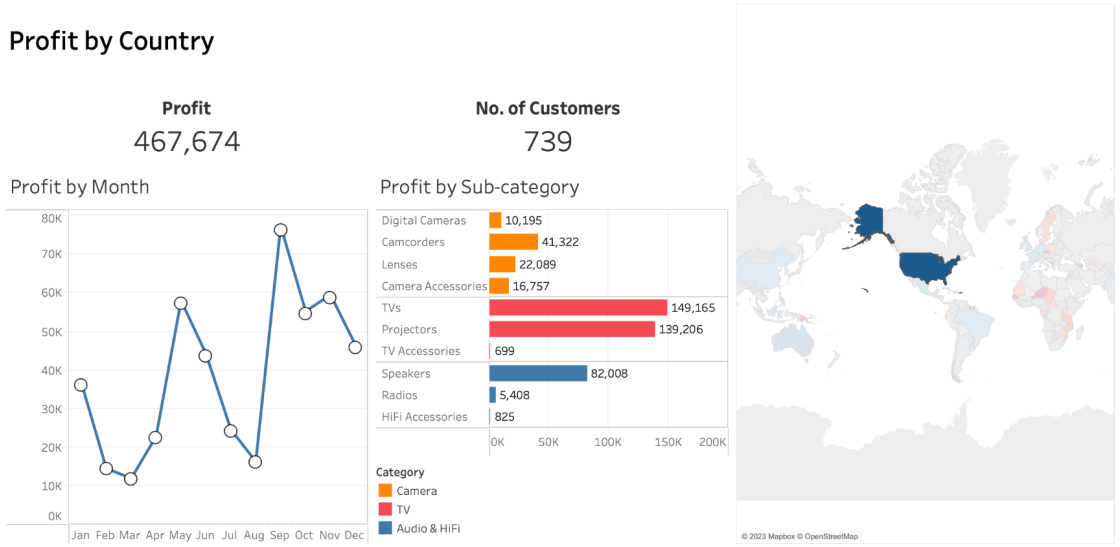
In order to answer the first question, the dashboard has been designed to showcase sales data in a comprehensive manner. It includes visualisation of sales by month, sales by sub-category products and a choropleth map representing sales distribution. We can explore countries of interest by clicking on the map to trigger simultaneous visualisation of sales by month and sales by sub-category. Additionally, the dashboard provides information about the sales amounts and the number of customers. This multifaceted approach to analysing data from different countries can be a valuable tool for making informed decisions regarding the selection of sales locations.





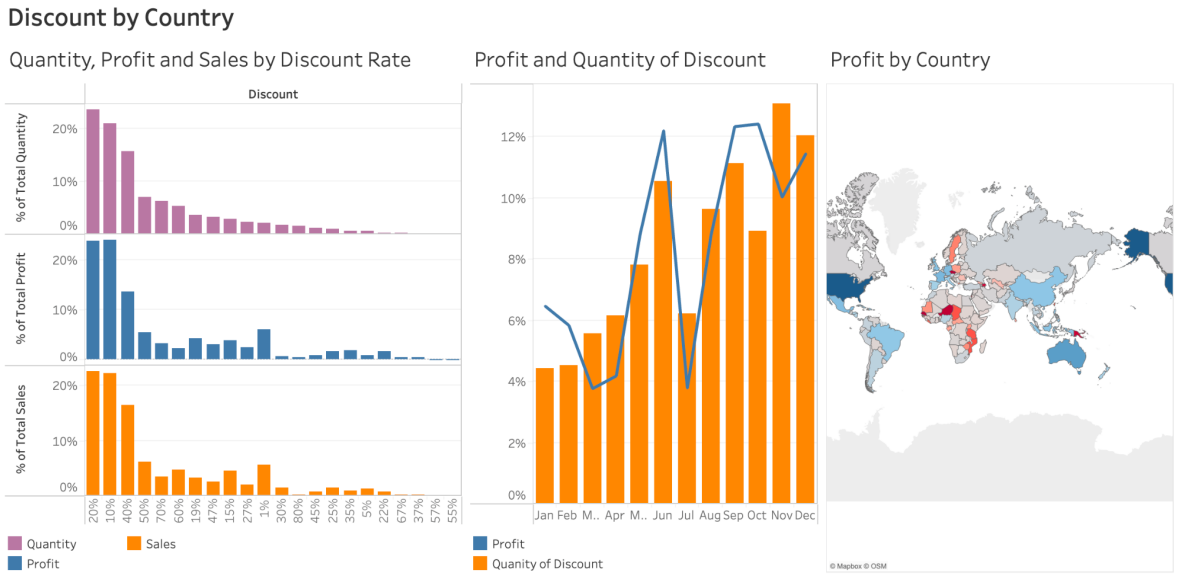
Similarly, a dashboard focused on profit has been developed to assist in the decision-making process for selecting sales locations. Its layout is the same as the Sales by Country dashboard, serving a comparable purpose in aiding the analysis and decision-making related to the profitability of different locations.

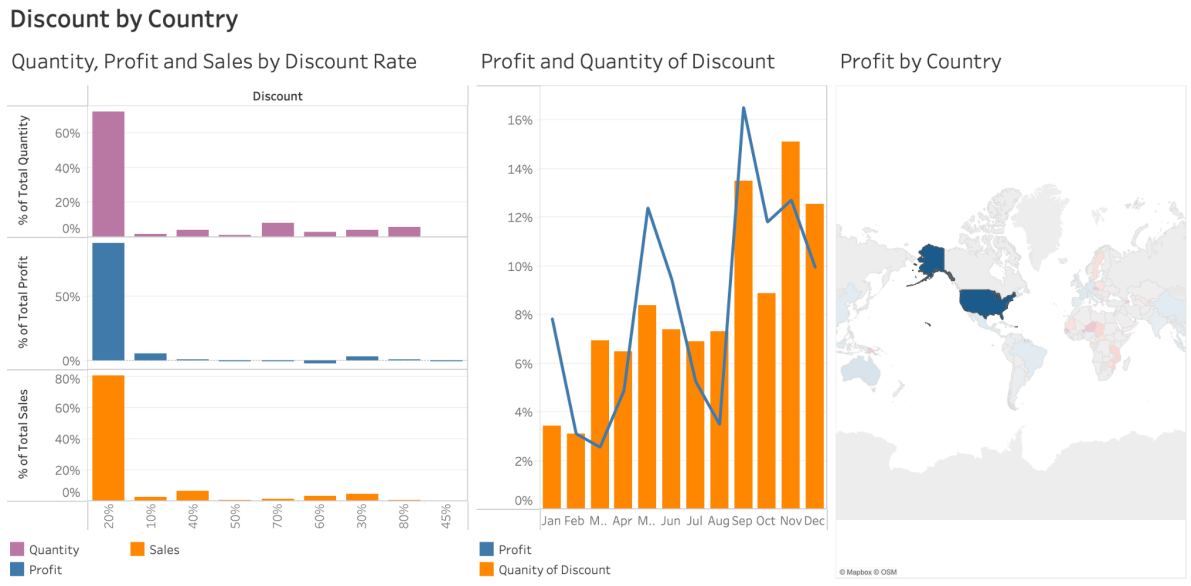




4.2 How do online retailers optimise their sales price for different countries?

A dashboard with the emphasis of discount was included for visualisation. With the interactive choropleth map, we can explore countries of interest to determine optimal discount rates. The quantity of discount was also included as it was found to be closely related to profits. By analysing insights from the data, the dashboard facilitates the prediction of optimal discount rates for each country, thereby optimising their sales price.





5. Presentation

https://youtu.be/y3vSJEdla1M