**DATA VISUALIZATION and ANALYSIS PROJECT**

**MID SUMMATIVE ASSIGNMENT**

**Name - Rohit Jaiswal**

**Cohort - Chandrayan**

**Dataset Link:**

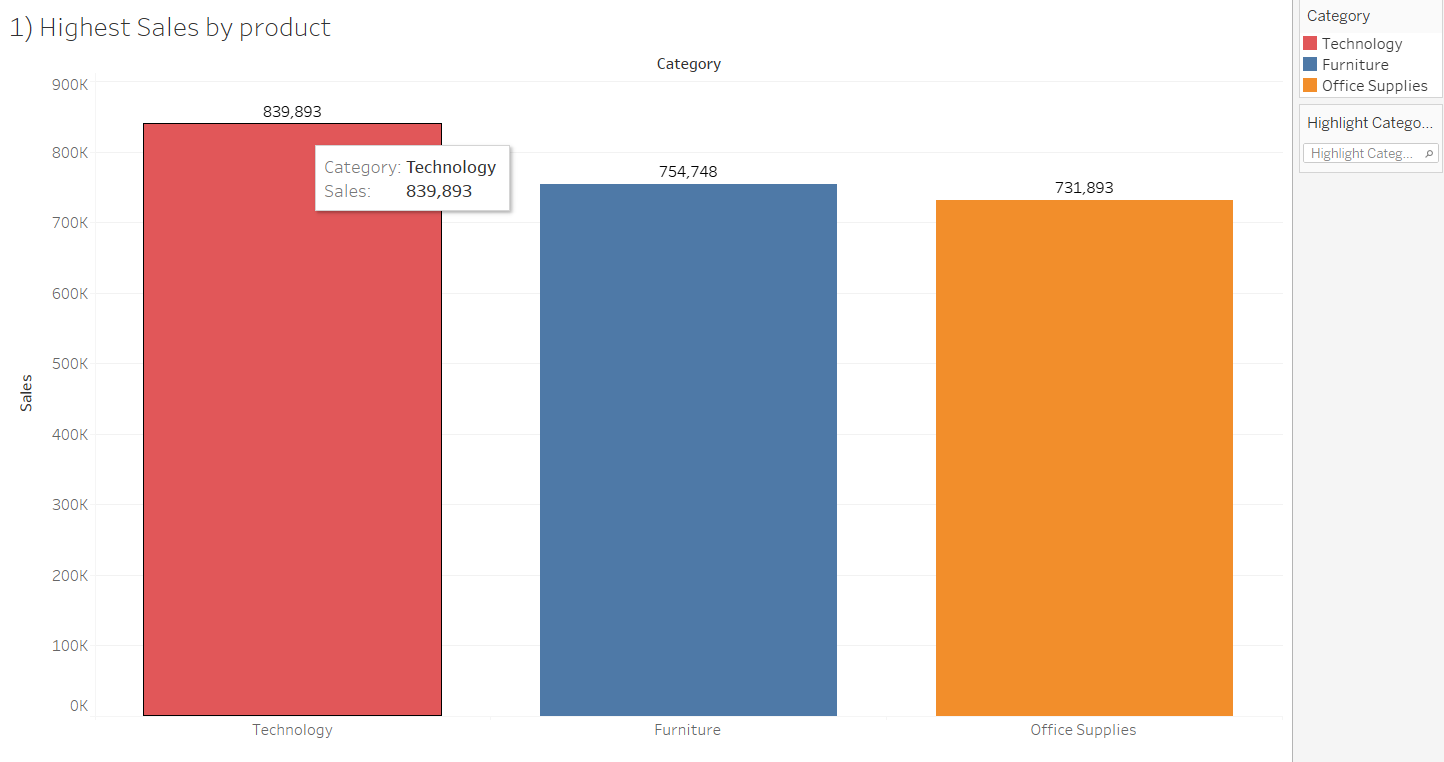
<https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

**About the Project:** In this project, I am working with a dataset of Superstore, aiming to answer 30 scenario-based questions through data visualization and analysis using Tableau. In this project I have selected the best chart for visualizing each question and explain each question based on the chart selection and its visualization.

**1.Which product categories have the highest total sales in the "Superstore" dataset?**

**Explanation**: I chose a Horizontal Bar Chart because it allows for a clear comparison of total sales across different product categories. The length of each bar makes it easy to identify the categories with the highest total sales.

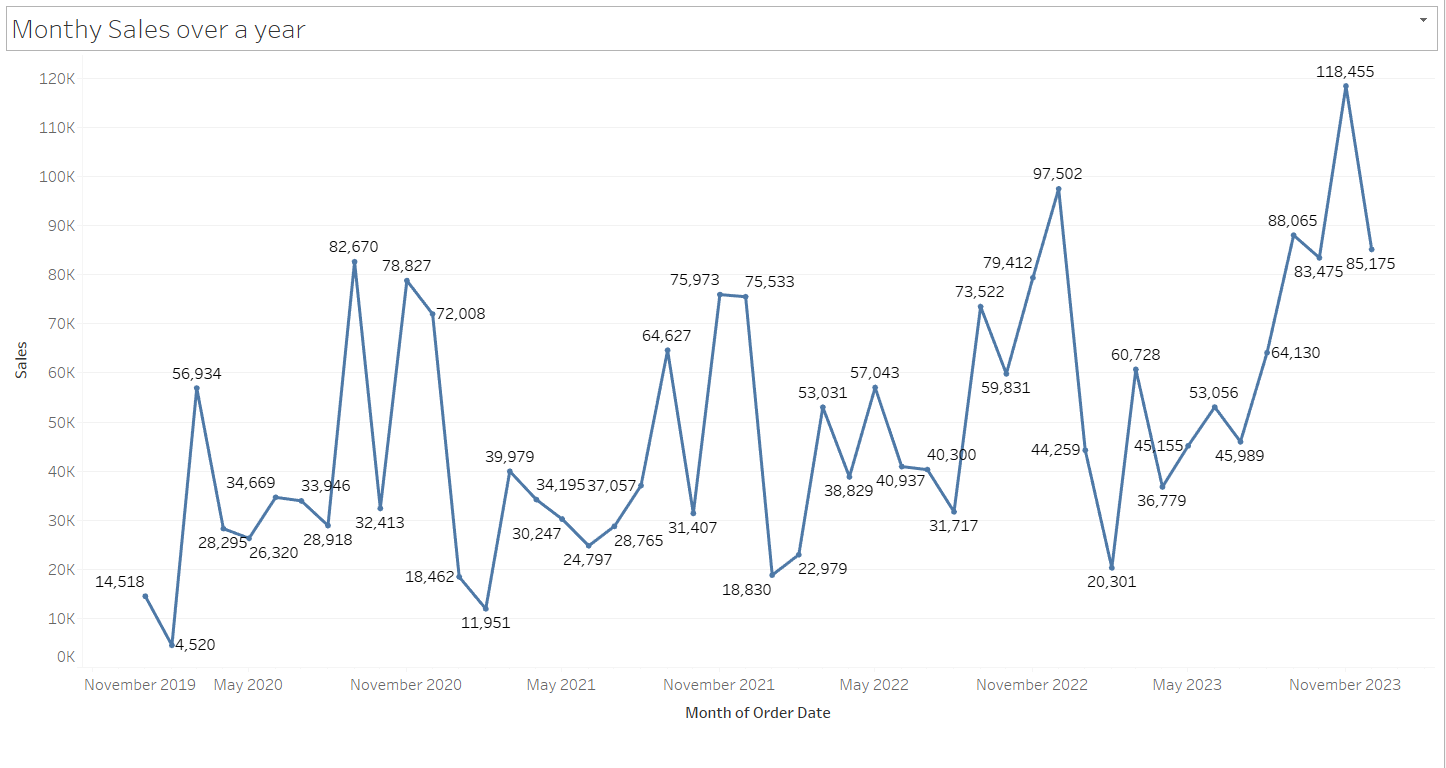
**Chart Type: Horizontal Bar Chart**

**Description:** With the help of this visualisation we are able to get the product category with highest sales is **Technology** with value of 839,893. The value is the sum of sales of **all years** in the dataset.

**2.How do the monthly sales amounts change over the course of a year?**

**Explanation:** I choose a Line Chart. It is an ideal for illustrating trends over time. By plotting monthly sales amounts on the x-axis and y-axis, the line chart allows for a clear view of how sales fluctuate throughout the year.

**Chart Type: Line Chart**

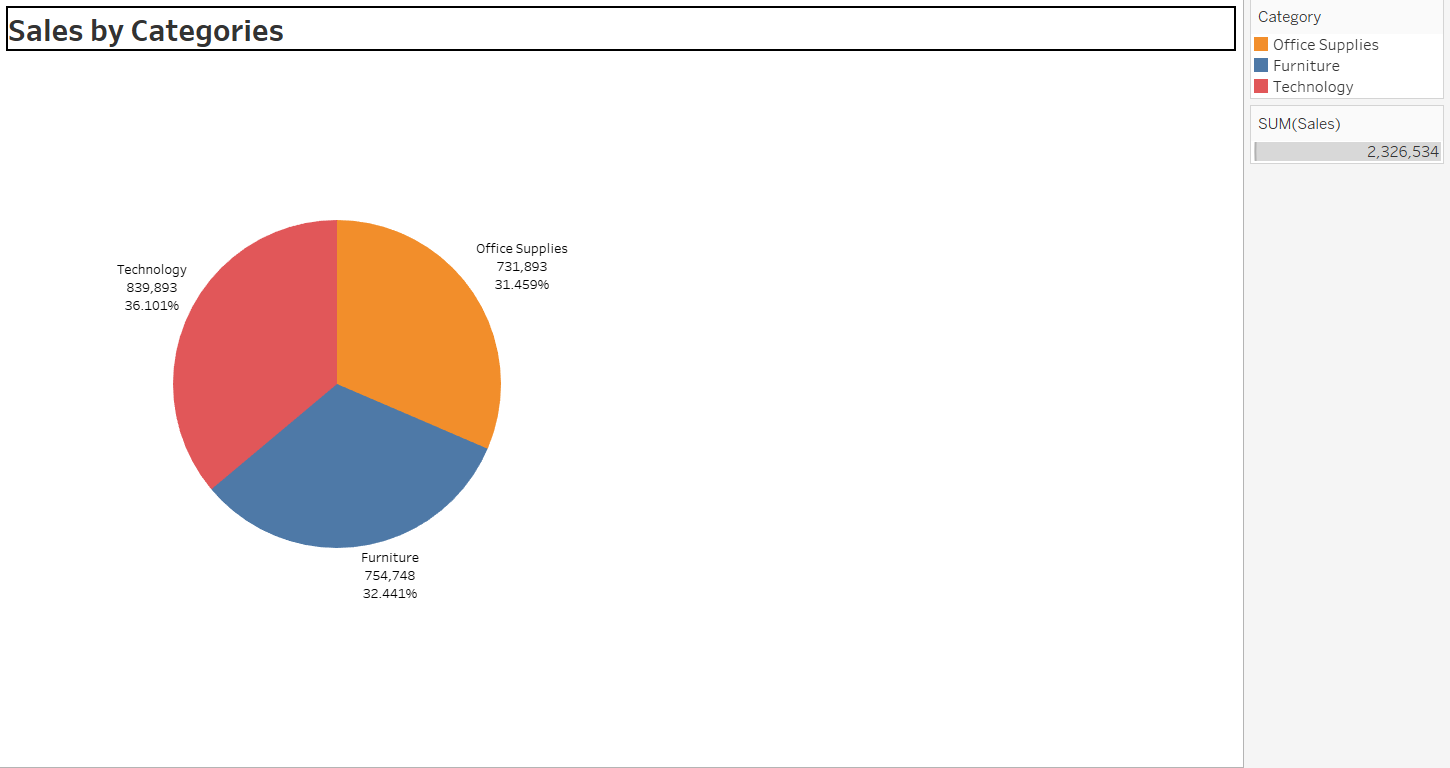
**Description:** With the help of the **line chart** we are able to understand the changes of sales throughout a year for multiple years. The insight we got is that, generally the sales are high in the month of November for every year.

**3.How is the total sales amount distributed among different product categories?**

**Explanation:** I choose **Pie Chart** because it provides a clear and concise representation of the percentage contribution of each product category to the total sales amount. Each colour represents a category, making it easy to grasp the distribution at a glance.

**Description:** Pie Chart visualisation has provided a clear and quick distribution of sales percentage wise for the following categories with the help of drill down we would also be able to get the sub categories sales percentage.

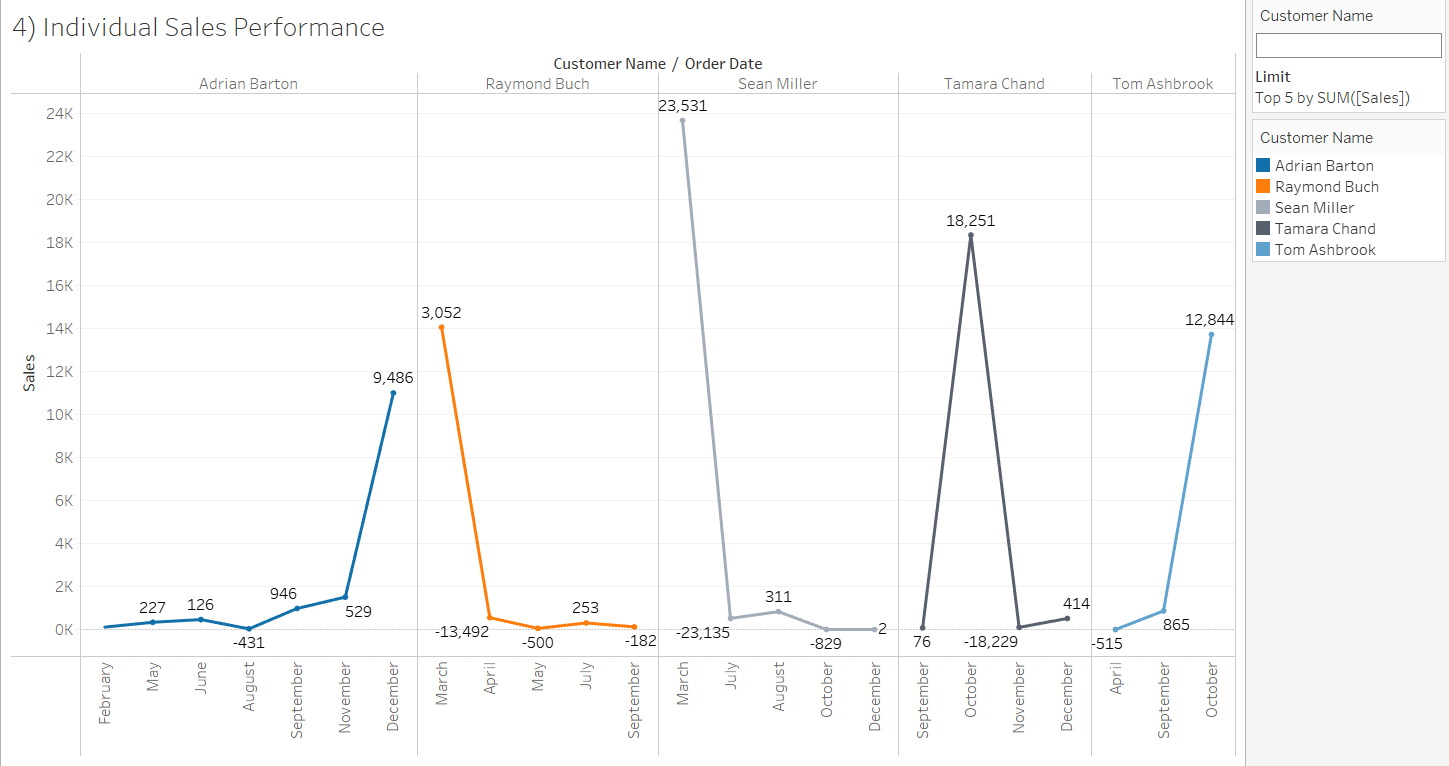
**Chart Type: Pie Chart**

****

**4.Can we analyze the sales performance of individual customers over time?**

**Explanation:** I choose Line Chart to effectively visualizes the sales performance of individual customers over time, providing insights into trends and identifying top-performing customers.

**Chart Type: Line Chart**

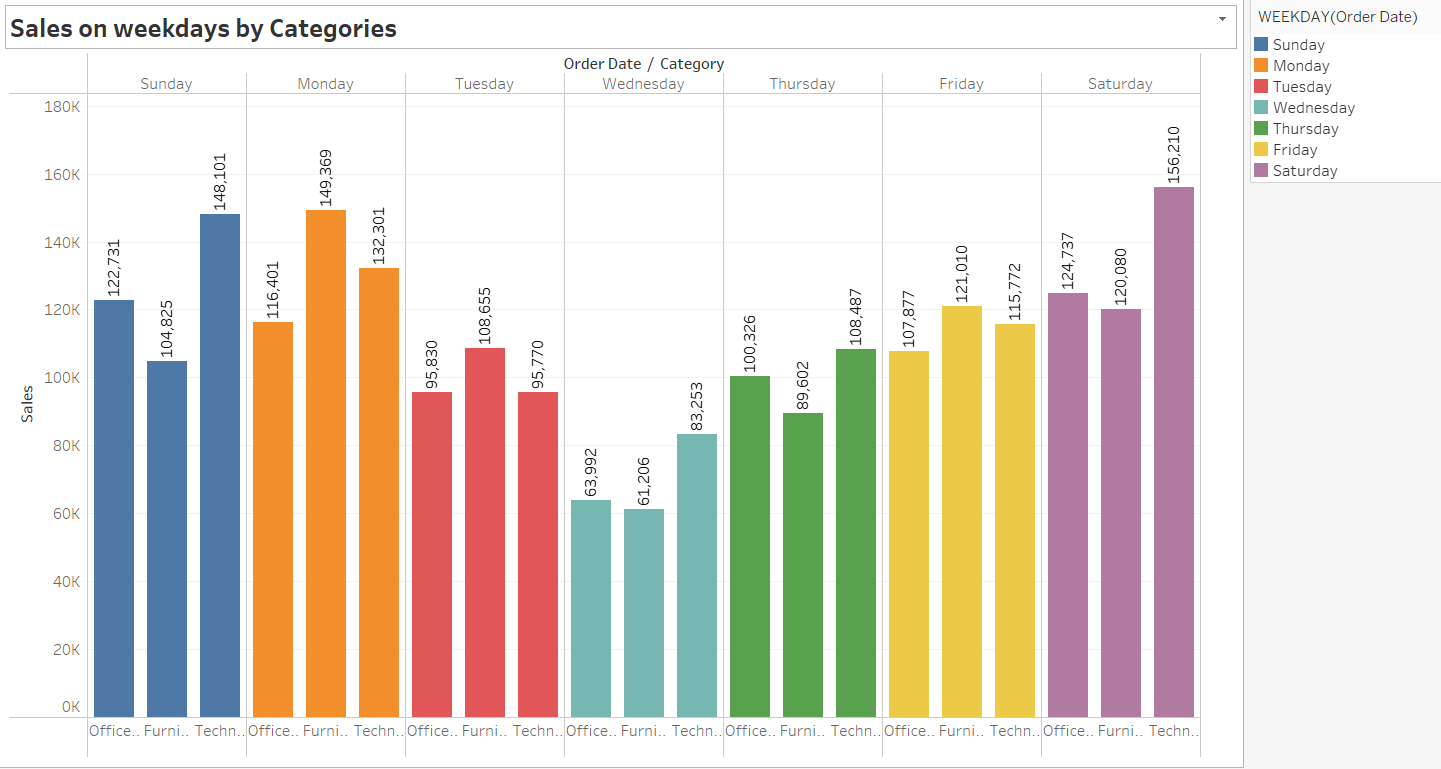
****

**Description:** The top 5 customers with highest sales are being considered for the analysis. We can observe the change in sales for the considered customers throughout the year, giving us insight into their buying pattern. It can be done for other buyers as well.

**5.How do sales vary based on different days of the week and product categories?**

**Explanation**: I Choose Side-by-Side Bar Chart is a suitable choice as it provides a clear comparison of sales across different days of the week for each product category. The side-by-side arrangement facilitates easy visual comparison and identification of trends.

**Chart Type: Side by side Bar Chart**

****

**Description:**

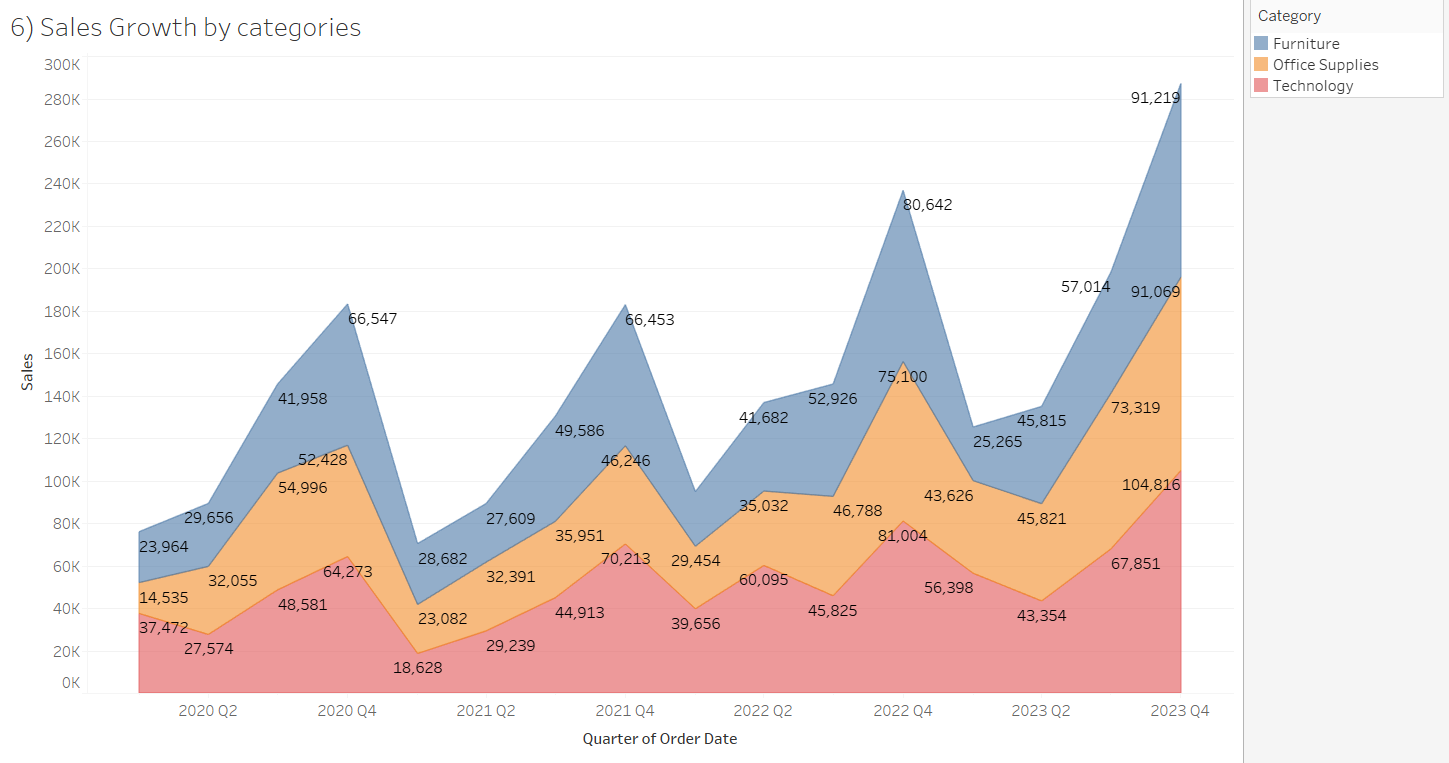
After visualizing we understood that the sales on different days of the week and product categories, we observe the highest value are in technology category on the day of Saturday with the value 1,56,210 and the lowest value are in Furniture category on the day of Wednesday with the value 61,206.

It can be observed that the sales by categories fluctuate every day but the lowest sales happen on the day of wednesday or in the mid of the week, while in the start and end of the week the sales are comparatively more, this trend could be useful for employee usage and leave planning.

**6. Can we visualize the sales growth of different product categories over time?**

**Explanation:** An **Area Chart** is selected to depict the sales growth of various product categories over time because it allows for a clear representation of trends, patterns and comparison. Each colour on the chart will represent a different product category, enabling easy comparison and identification of growth or decline in sales over different periods.

**Chart Type: Area Chart**

****

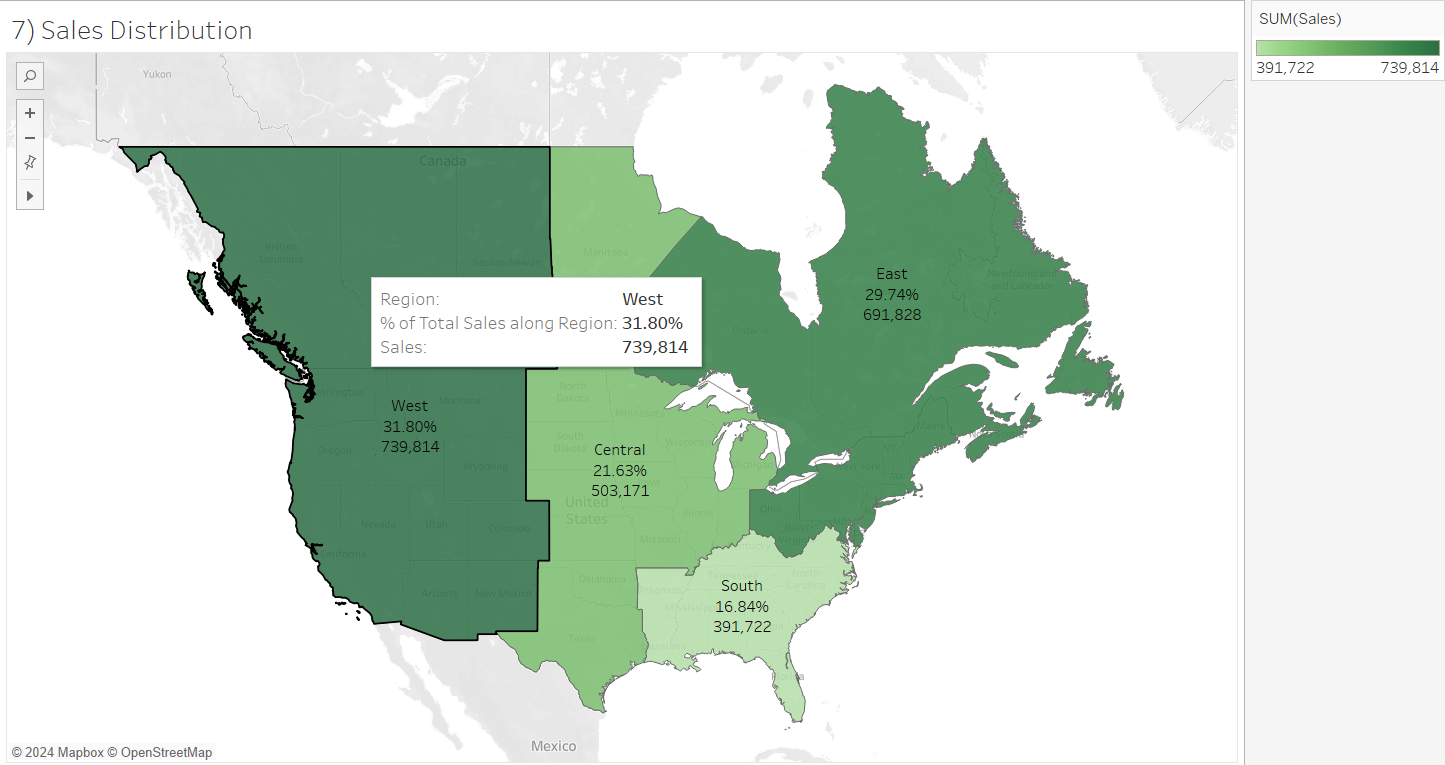
**Description:** After visualizing we analyze that the sales growth of technology is high and increasing almost every quarter. The office supplies and furniture are also having sales growth but compared to technology their sales growth is slow.

Furniture was having highest sales in the Q4 2020 but because of high sales growth of technology and office supplies the sales figure of these two categories surpassed the furniture, technology being at the top and the Office supplies at the second place this gives the customers behaviour towards the product category.

**7.How does the sales distribution vary across different regions in the "Superstore" dataset?**

**Explanation:** Opting for a **Map Chart** allows for a geographic representation of sales distribution across different regions. It provides a visual overview of sales magnitude in each region, enabling quick identification of high-performing areas. This chart type is particularly effective for showcasing geographical disparities in sales.

**Chart Type: Map Chart**

****

**Description:**

The sales distribution varies across different regions. In the superstore dataset we have 4 different regions and each region contains different sales profit.

The west region has high sales with the value of 7,39,814.

The east region also had good sales with the value 6,91,828.

The central region has medium sales 5,03,171 and the south region has the lowest sales with the value of 3,91,722.

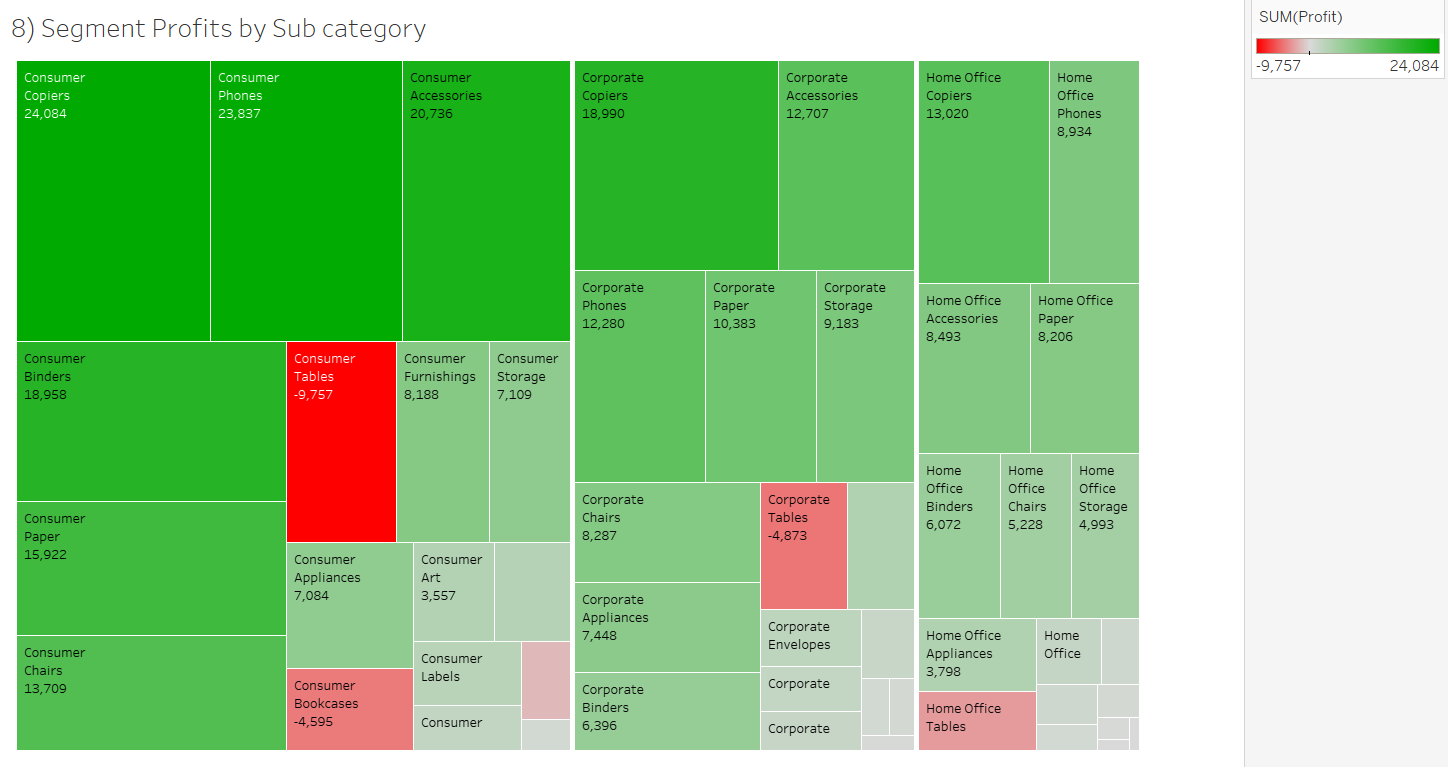
**8.Can we visualize the composition of profits across various subcategories within different customer segments?**

**Explanation:** A Tree Map is selected to visually represent the composition of profits across various subcategories within different customer segments. This chart type allows for an intuitive display of hierarchical data, with each rectangle representing a subcategory within a customer segment. The size of each rectangle corresponds to the profit, enabling a clear understanding of the contribution of subcategories to overall profits.

**Description:**After visualizing we understood that the Consumer segment is giving the highest profits to the company followed by corporate and then Home Supplies.

The Copier Sub Category in the consumer segment has the highest profits with the value of 24,084. While the tables category is performing poorly and is having loss in every customer segment, consumer bookcases also being the sub category under consumer category giving loss to the company.

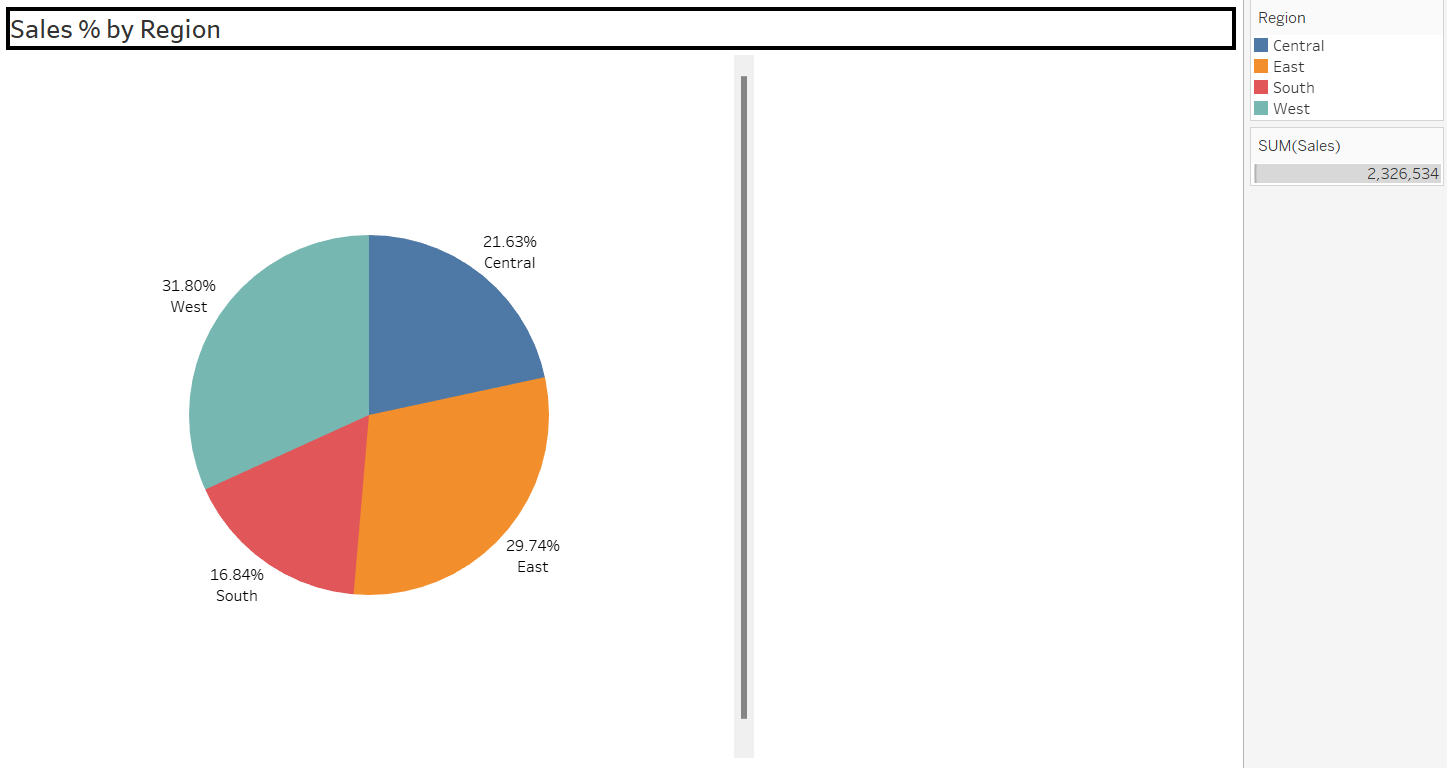
**Chart Type: Tree Map Chart**



**9. What is the percentage contribution of each region to the overall sales?**

**Explanation:** Opting for a **Pie Chart** allows for a representation of the percentage contribution of each region to overall sales. This chart type provides a visual overview of the sales distribution, making it easy to compare the contributions of different regions. Each region will be colour-coded or shaded based on its percentage contribution, offering a clear and intuitive representation.

**Chart Type: Pie Chart**

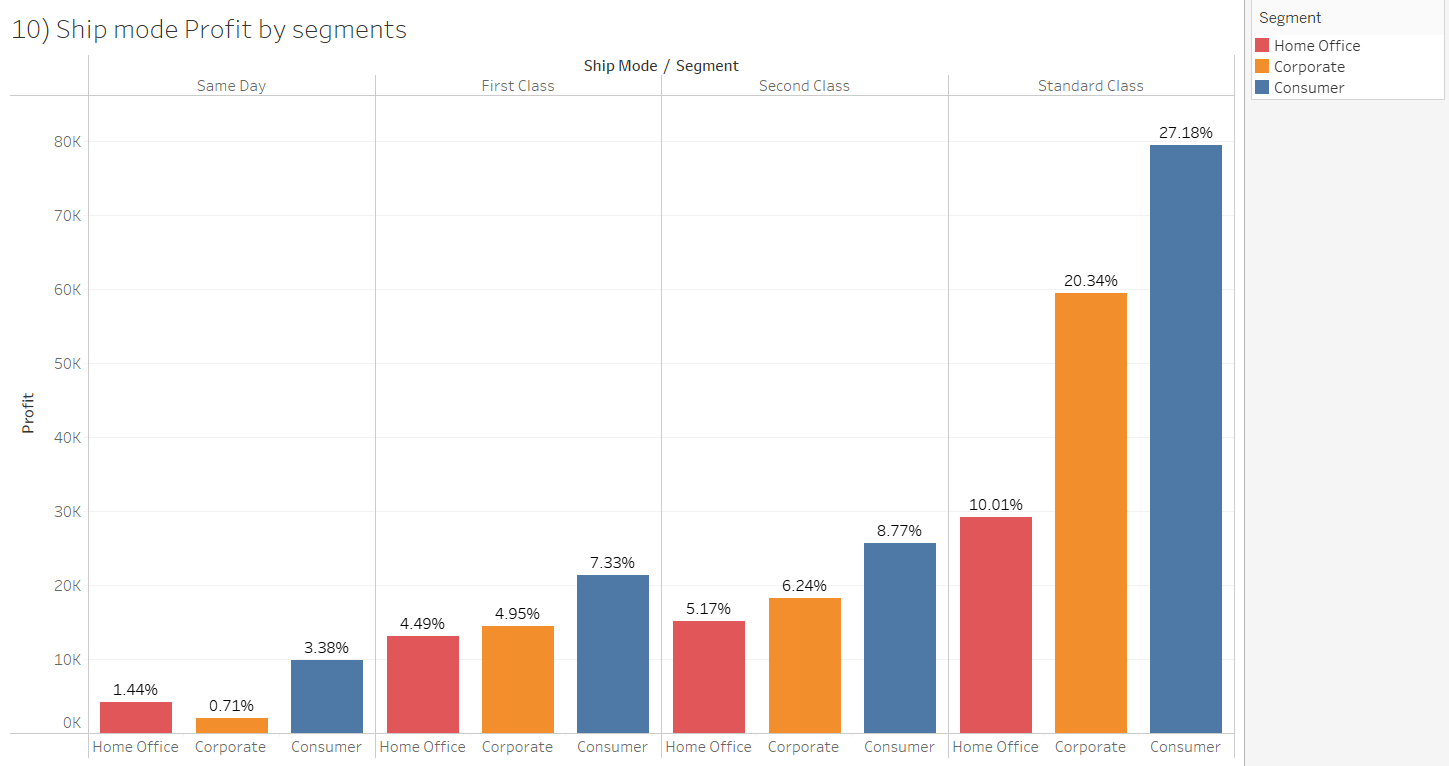
****

**Description:** After visualizing we understood that the superstore dataset has 4 different regions and each region contains different sales percentage. West region has the highest sales with the percentage of 31.80% and the south region has the lowest sales with the percentage of 16.84%.

**10.Can we visualize the profit margins associated with different shipping modes and customer segments?**

**Explanation:** A Side-By-Side Bar Chart is selected to effectively visualize the profit margins associated with different shipping modes and customer segments. This chart type allows for a clear comparison of profit margins across shipping modes, with separate bars for each customer segment. It provides a straightforward representation, enabling easy identification of variations in profit margins.

**Chart Type: Side by Side Bar Chart**

****

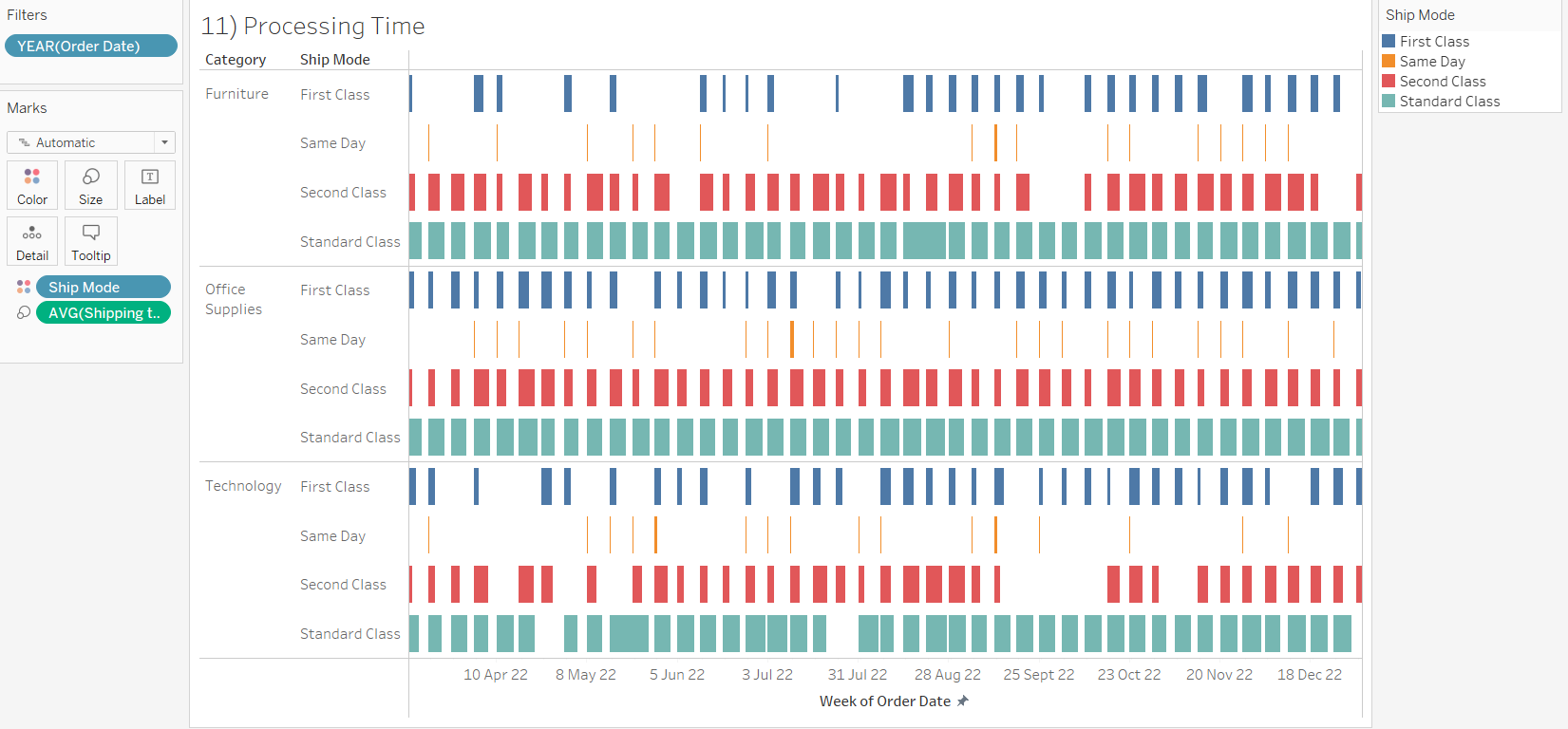
**Description:** After visualizing the chart, we analyze to understood that the consumer segment has getting high profits in standard class shipping mode with the value of 79,457(27.18%) and the lowest profit margin we are getting from corporate segment at same day with the value of 2087(0.71%).

Each shipping mode contains a different percentage of profit margins in our customer segments. Standard class shipping gains more profits compared to other shipping modes.

**11.How long does it take to process orders for different product categories?**

**Explanation:** A Gantt Chart has been selected to show the duration of the time being required to deliver the orders to the client, it properly shows the number of days taken to complete the process of delivery for different shipping modes. With drill down/filtering we can select and analyse the delivery duration to process the orders for different categories.  
 A Gantt Chart in Tableau displays tasks as horizontal bars along a timeline, with the length of the bar representing the duration of the task.

**Chart Type: Gantt Chart**

****

**Description:** After analyzing the chart we were able to observe that the delivery mode duration is same for all categories, although there are times when second class delivery mode delivers orders faster then first class delivery mode.

**12.How do discounts affect overall profit?**

**Explanation:** A Scatter Plot has been chosen for the visualization to show the effect of discount on profit. The Scatter plot would be the most suitable chart to show the relationship between the discount with the profit. Scatter Plots are a type of chart that display the relationship between two measures.

It gives quick understanding of the situation and relation of the measures within a glance identifying profits for different discount rates and showcasing the negative relation of profits with discount, as discount increases the profit decreases.

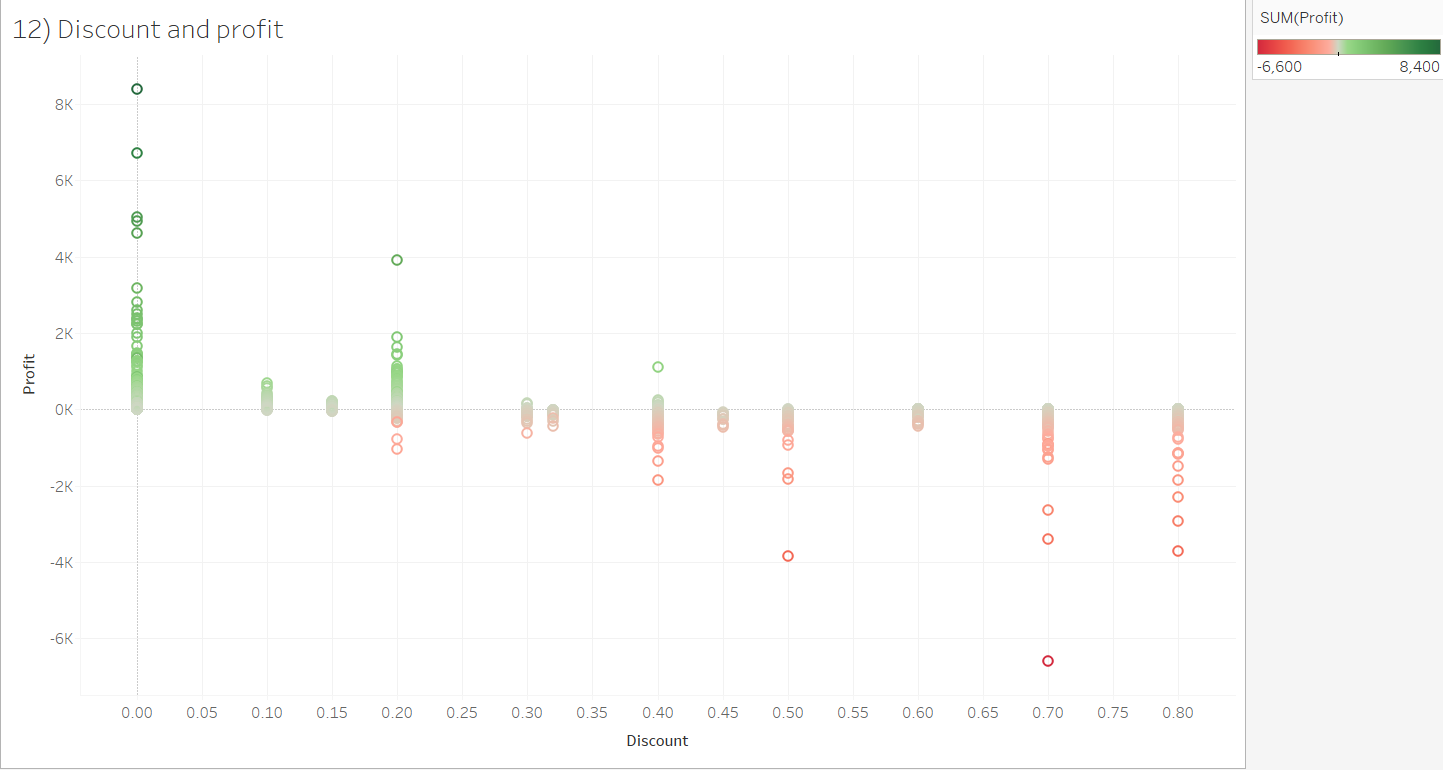
**Description:**

Profits at every discount level can be observed with the help of scatter plot, with the help of the visualization we can observe that with increasing discount the profit keeps on decreasing for many orders.

The highest loss is 6600 of loss at 70% discount, while the highest profit of 8,400

has been observed at 0% discount.

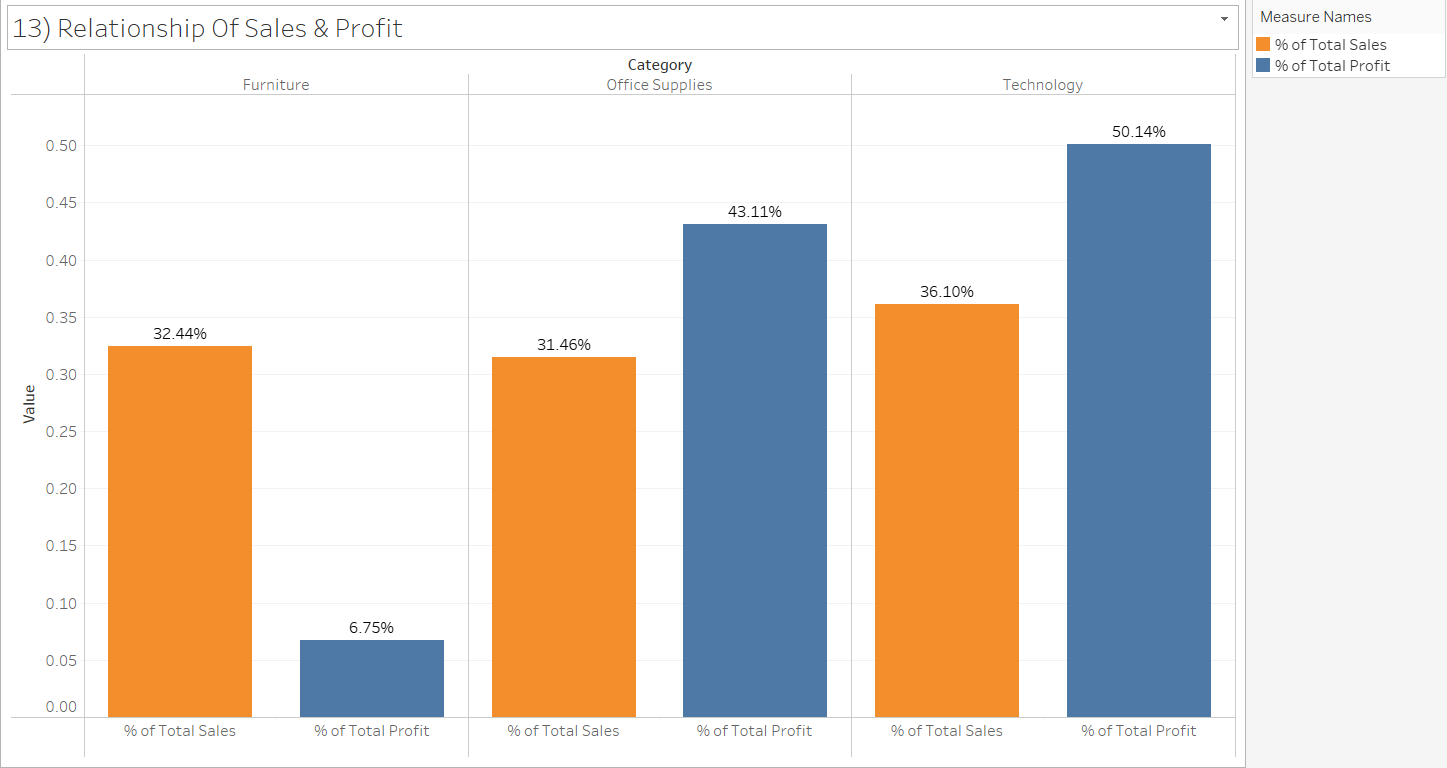
**Chart Type: Scatter Plot**

****

**13.Can we visualize the relationship between product sales and profitability for different product categories?**

**Explanation:** Opting for a **Side-By-Side Bar** Chart to compare product sales and profitability for different categories simultaneously. A **Side-By-Side Bar** Chart allows for a direct visual comparison between product sales and profitability across various categories.

This choice facilitates a quick assessment of the relative performance of each category in terms of both sales and profitability, aiding in identifying patterns or disparities.

**Chart Type: Side-By-Side Bar Chart**

**Description:** Visualization of the product sales and profitability of categories relationships. There are 3 types of categories: furniture, office supplies, and technology.

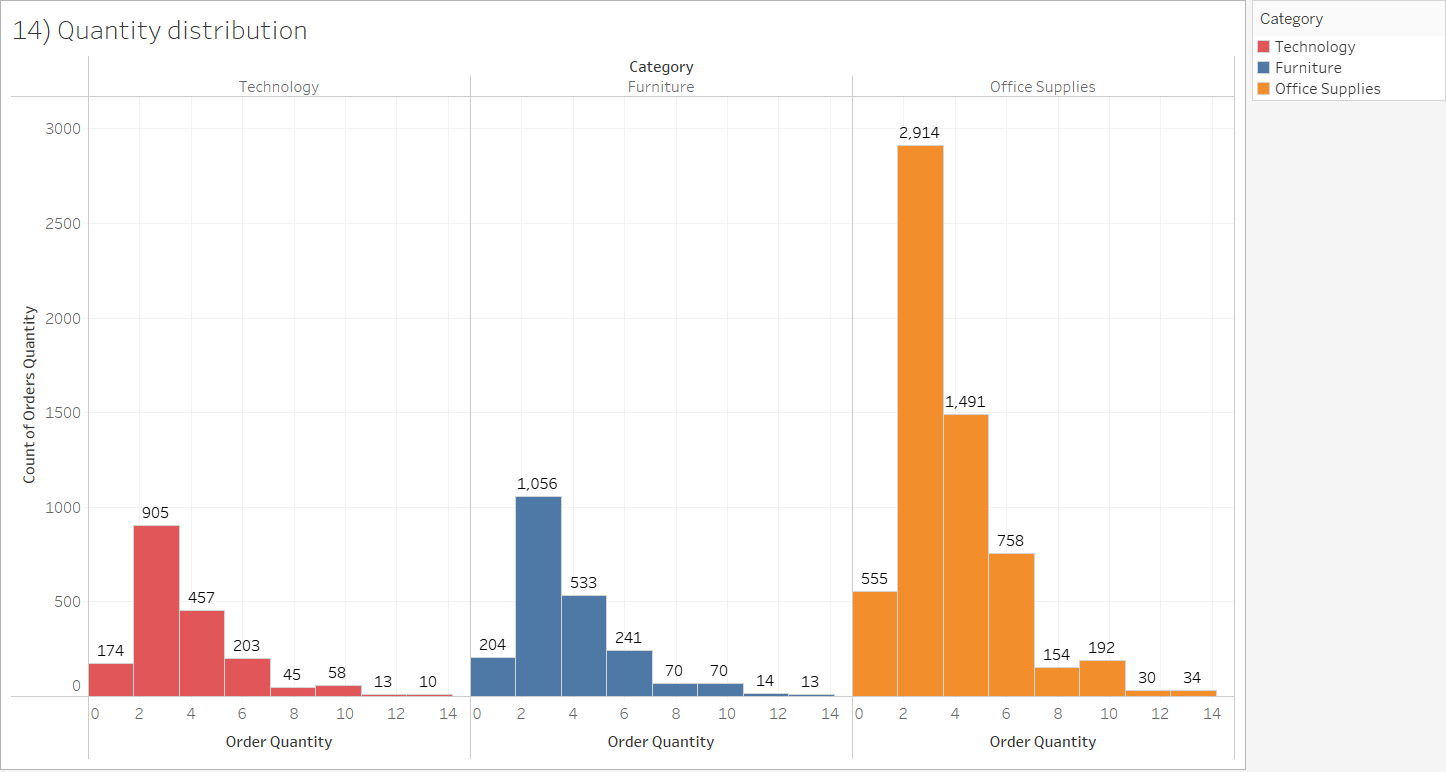
The furniture sales are with the value of 32.44% and its profits are 6.75% similarly office supplies sales are 31.46% and its profits are 43.11% and finally technology sales are 36.10% and its profits are 50.14%. Comparing 3 categories, the relationship of sales and profits technology has gained more profits compared to office supplies and furniture.

**14.What is the distribution of order quantities for products in the dataset?**

**Explanation: Histogram** is chosen to illustrate the distribution of order quantities. A **histogram** provides a visual representation of the frequency or count of order quantities within different ranges, offering insights into the overall pattern of order quantities.

A **Histogram** is a suitable choice to show the spread and concentration of order quantities. It enables a quick understanding of the distribution pattern, helping identify common order quantity ranges and potential outliers within the dataset.

**Chart Type: Histogram**

****

**Description:** We analyze the distribution of order quantities for products by viewing this visualization. Same pattern has been observed for all categories.

The technology category distribution of order count is high on order quantity 2-4 range with the value of 905 and the lowest order count is 10.

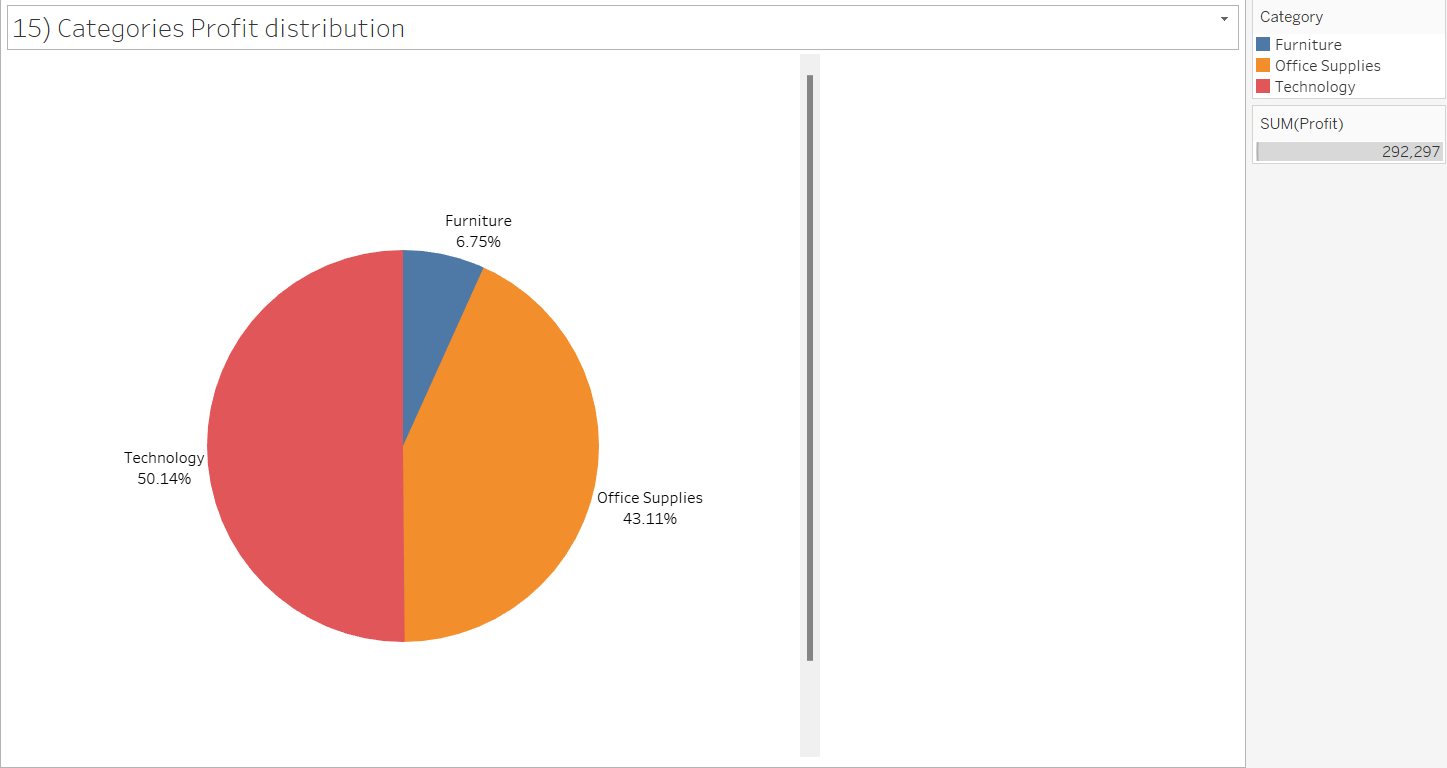
The furniture category distribution of order count is high on order quantity 2-4 range with the value of 1056 and the lowest order count is 13.

The office supplies category distribution of order count is high on order quantity 2-4 range with the value of 2,914 and the lowest order count is 30.

**15.How do the profit distributions vary across different product categories?**

**Explanation:** A Pie Chart provides a quick and intuitive overview of how profits are distributed among different product categories. Each slice represents a product category's share of the total profits, allowing for easy identification of which categories contribute more significantly to the overall profitability.

**Chart Type: Pie Chart**



**Description:** The technology has a distribution of profit with the value of 50.14% and the office supplies has contained distribution of profit with the value of 43.11% and the furniture has contained distribution of profit with the value of 6.75%. The technology contains the highest profits contribution.

**16.Can we compare the shipping time distributions for different shipping modes?**

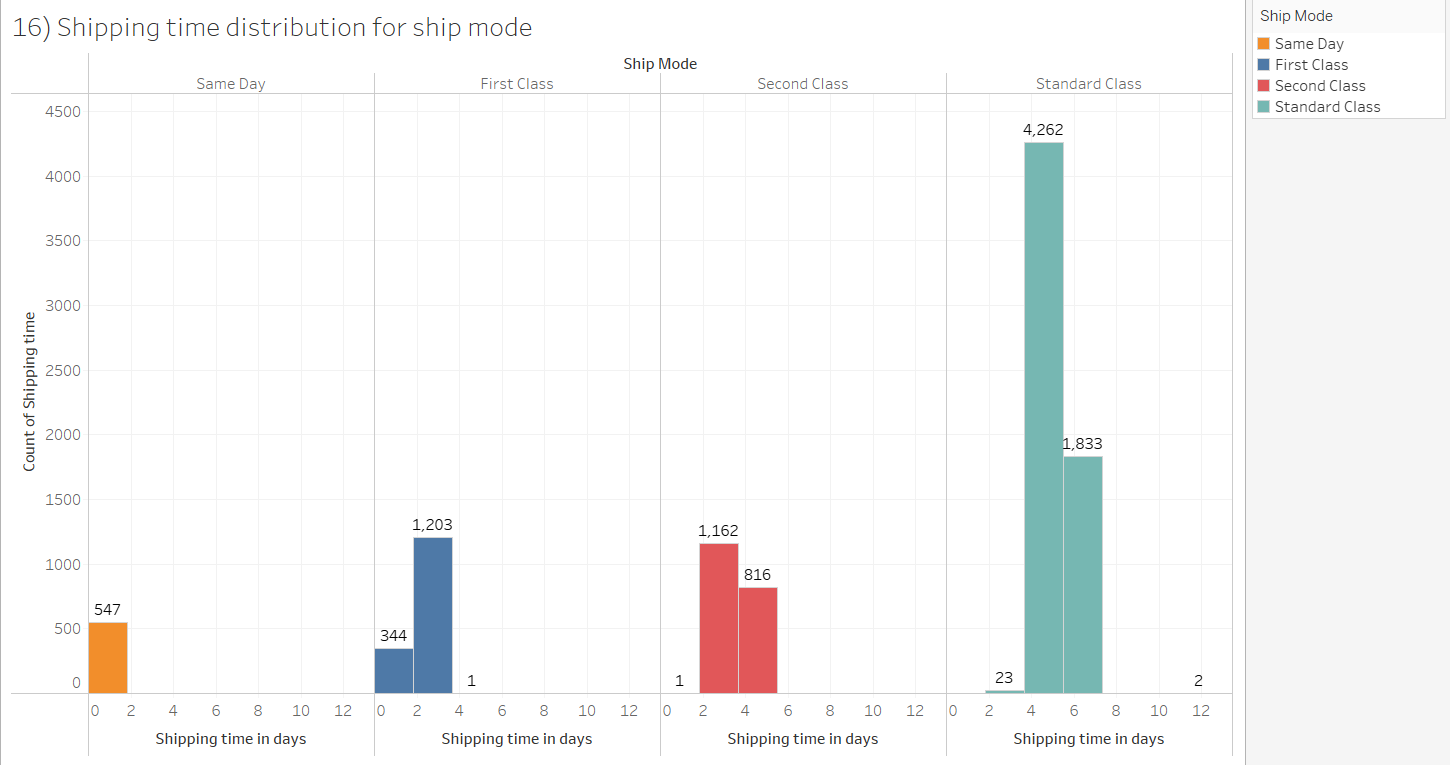
**Explanation:** I have used histograms to get the distribution of different shipping modes, with the help of this visualization it is easy to compare the shipping time distributions as it provides side by side comparison of each shipping mode.

**Description:** The shipping time distributions for different shipping modes provide valuable insights.

Standard class had maximum orders while 4262 orders required 4-6 days, 1833 orders were delivered in 6-8 days range.

Second class and first class both shipping modes provide most of the delivery of orders in the range of 2-4 days.

**Chart Type: Histogram**



**17.What is the monthly trend in the number of orders shipped?**

**Explanation:**

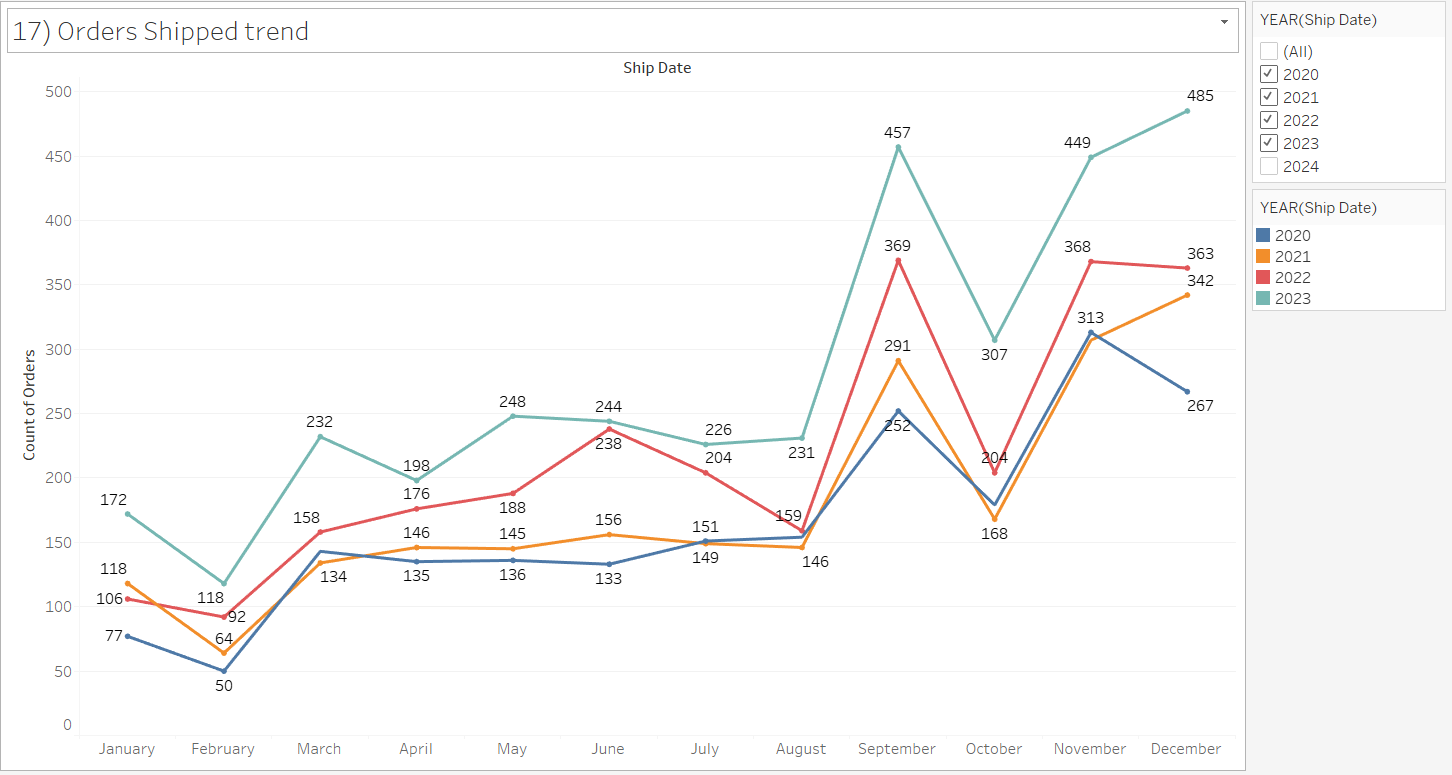
I have chosen a Line Chart to depict the monthly trend in the number of orders shipped. A line chart is effective for showcasing trends over time, providing a clear visualization of how the volume of orders changes on a monthly basis.

A Line Chart will present a visual narrative of the fluctuations in the number of orders shipped each month. It allows for a quick understanding of any patterns, seasonality, or trends in the order shipment process over the course of the year.

**Description:**

In all categories the February month has lowest orders compared to remaining months and highest orders are in the month of September & then in the month of November, there is a substantial decrease in the number of orders in the month of October.

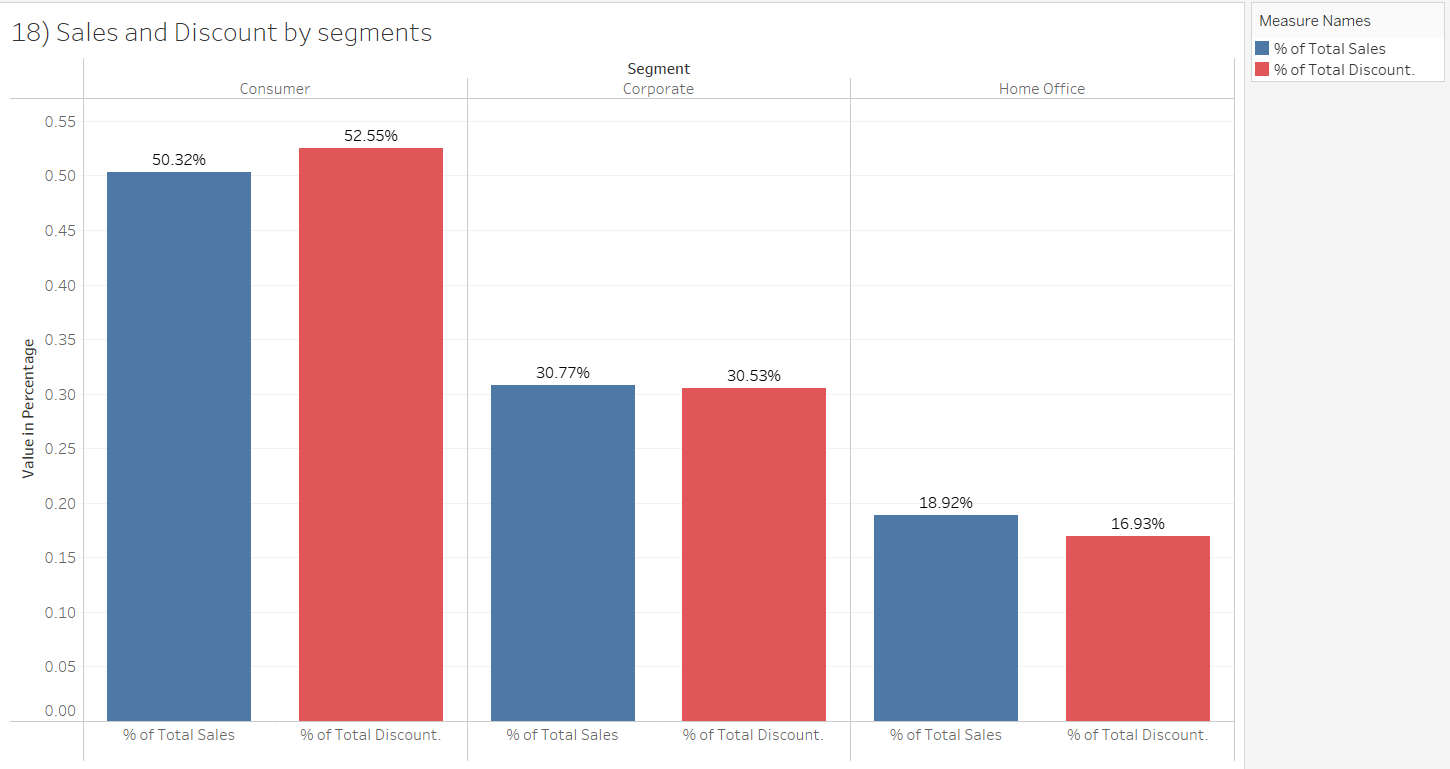
**Chart Type: Line Chart**



**18.How do different customer segments perform in terms of sales and discount rates?**

**Explanation:** I choose **Side by Side Bar Chart** is effective for visually comparing multiple variables within each category, making it suitable for this scenario. It allows for a side-by-side comparison of sales and discount rates for each customer segment. This visualization method provides a clear and detailed view of the performance of different segments in terms of both sales and discount rates, facilitating easy interpretation and analysis.

**Chart Type: Side by Side Bar Chart**

****

**Description:** In the performance of sales and discount in the chart.

The Consumer segment discount value is 52.55% and the consumer sales value is 50.32%.

The Corporate segment discount value is 30.53% and the Corporate sales value is 30.77%.

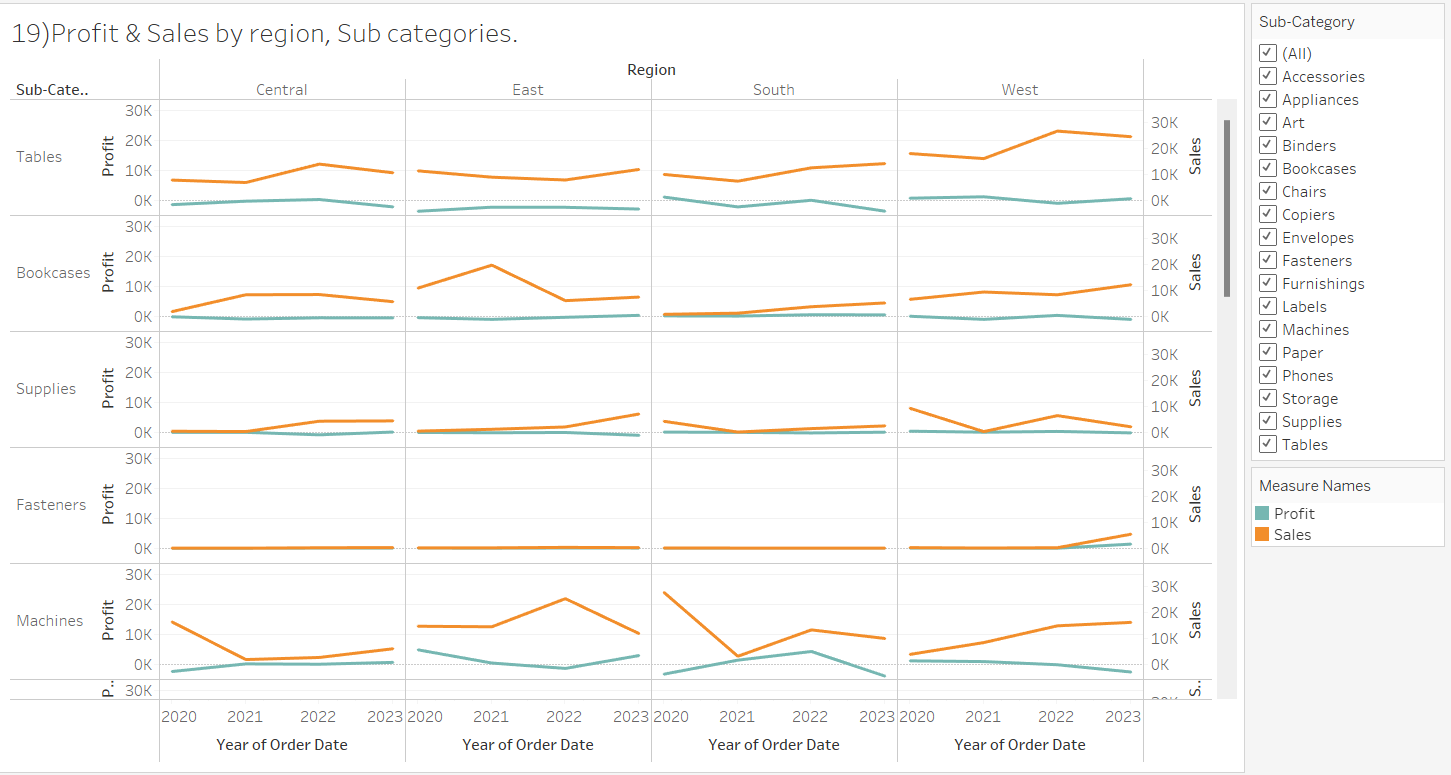
The Home office segment discount value is 16.93% and Home office value is 18.92%.

Compared to all the segments, sales and discount values, consumers do have a high discount.

**19.What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?**

**Explanation:** A Dual Lines Chart would be the most suitable chart to properly visualize the trends of sales and profit over the period of time further it can be visualized for different product sub categories and regions, filters facilitate the choice of selecting the required sub category, for analysis of sales and profit trends for different regions.

**Chart Type: Dual Lines Chart**

****

**Description:**

The visualization is sorted for sub categories in an ascending order on the base of profit, the pattern for sub categories having the lowest profits and the respective sales can be observed.

**Tables** being among the worst, continuous trend of negative profit can be observed for this sub category, no relation of sales and profit can be identified for almost all regions.

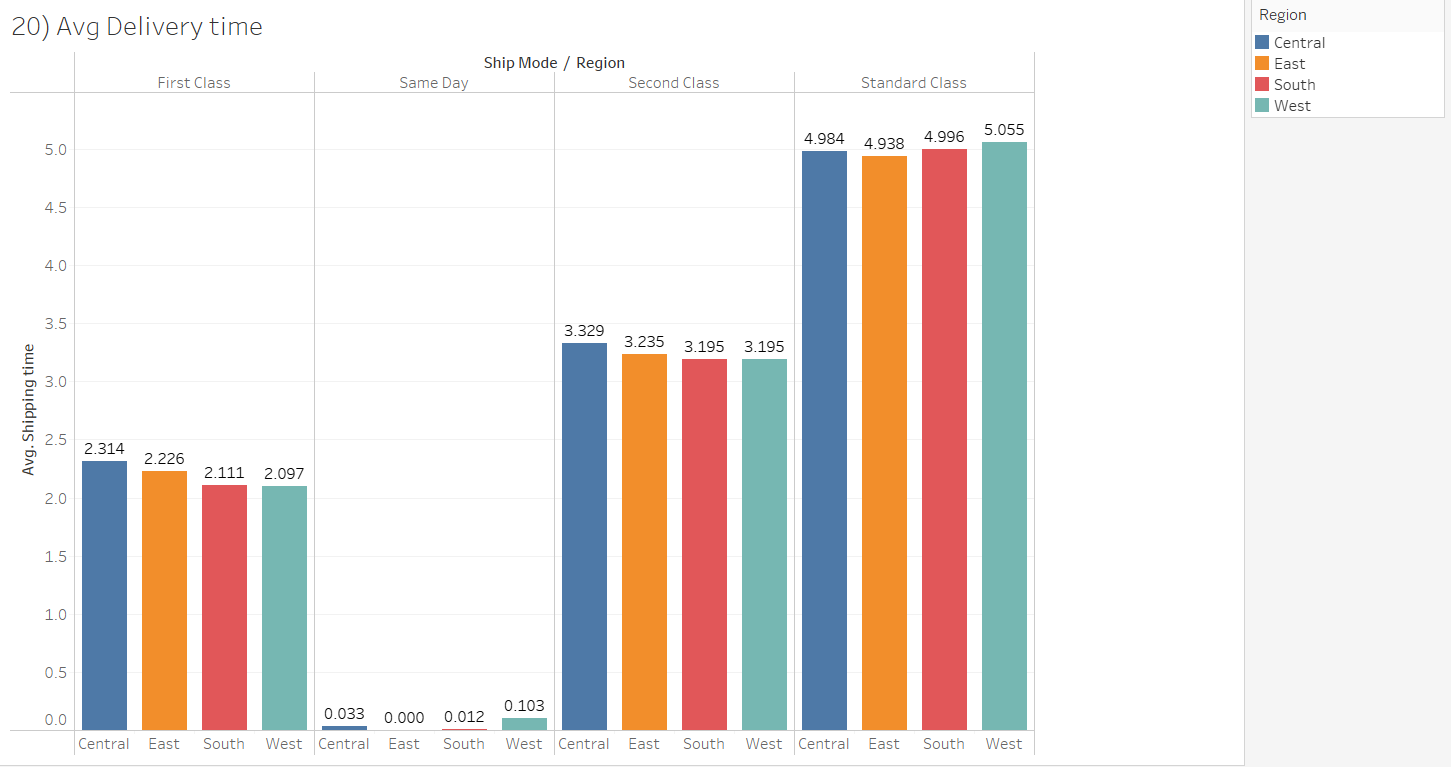
**BookCases & Supplies** have almost remained constantly in the unprofitable range. The fluctuation can be seen with respect to sales through the years but no change in profits can be observed.

**Fasteners** is the sub category which have very low profits and sales but are in sync with each other, for the region of west in the year of 2024 increase of sales and a slight increase in profit can be observed etc.

**20.What is the average delivery duration for different regions and ship modes?**

**Explanation: Side by Side Bar Chart** provides a side by side comparison of different regions & Shipping modes simultaneously providing a quick and clear analysis on the time duration taken by the delivery process.

**Chart Type: Side by Side Bar Chart**

****

**Description:**

The average delivery duration analysis for the shipping modes,

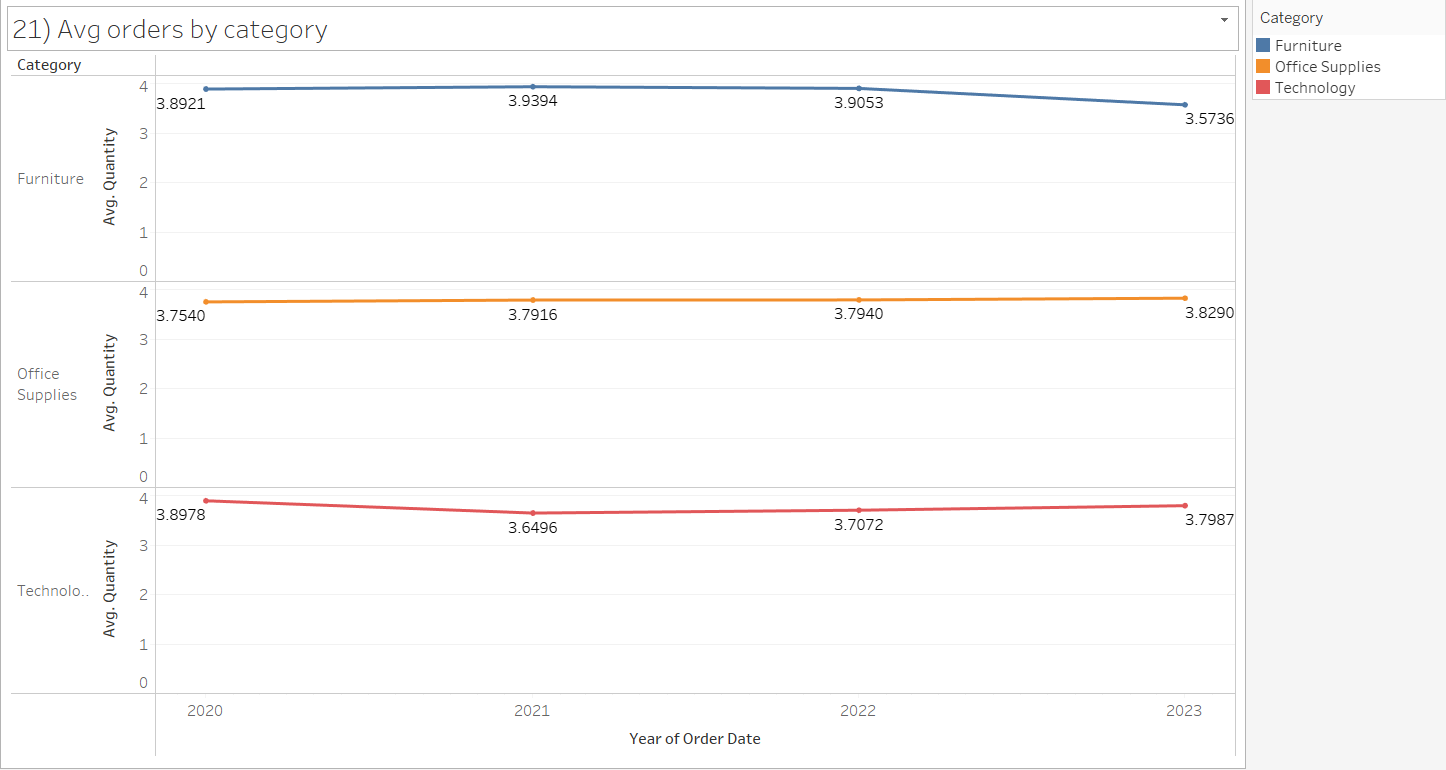
First Class - Central region is having the longest avg delivery duration, followed by east,south and then West being the quickest one to get the delivery.

Same Day Delivery- All regions are under the mark of 1 day for same day delivery, the west region for some reason is having the highest delivery time, same can be analysed for the other regions.

**21.How has the average order quantity changed over the years for various product categories?**

**Explanation:** I Choose a Line Chart to illustrate the changes in average order quantity over the years for various product categories. A line chart effectively communicates trends over time, providing a clear visualization of how average order quantities change annually. A Line Chart will visually represent the trends in average order quantity for different product categories over the years. It allows for easy identification of patterns, fluctuations, or any significant changes in customer ordering behaviour across various categories.

**Chart Type: Line Chart**

****

**Description:** The average order quantity changed over the years. In 2023 the furniture average order quantity with the value of 3.5736 decreased compared to previous years.

In 2023 the office supplies average order quantity with the value of 38290 slightly increased compared to previous years.

In 2023 the technology average order quantity with the value of 37987 had slightly decreased compared to previous years.

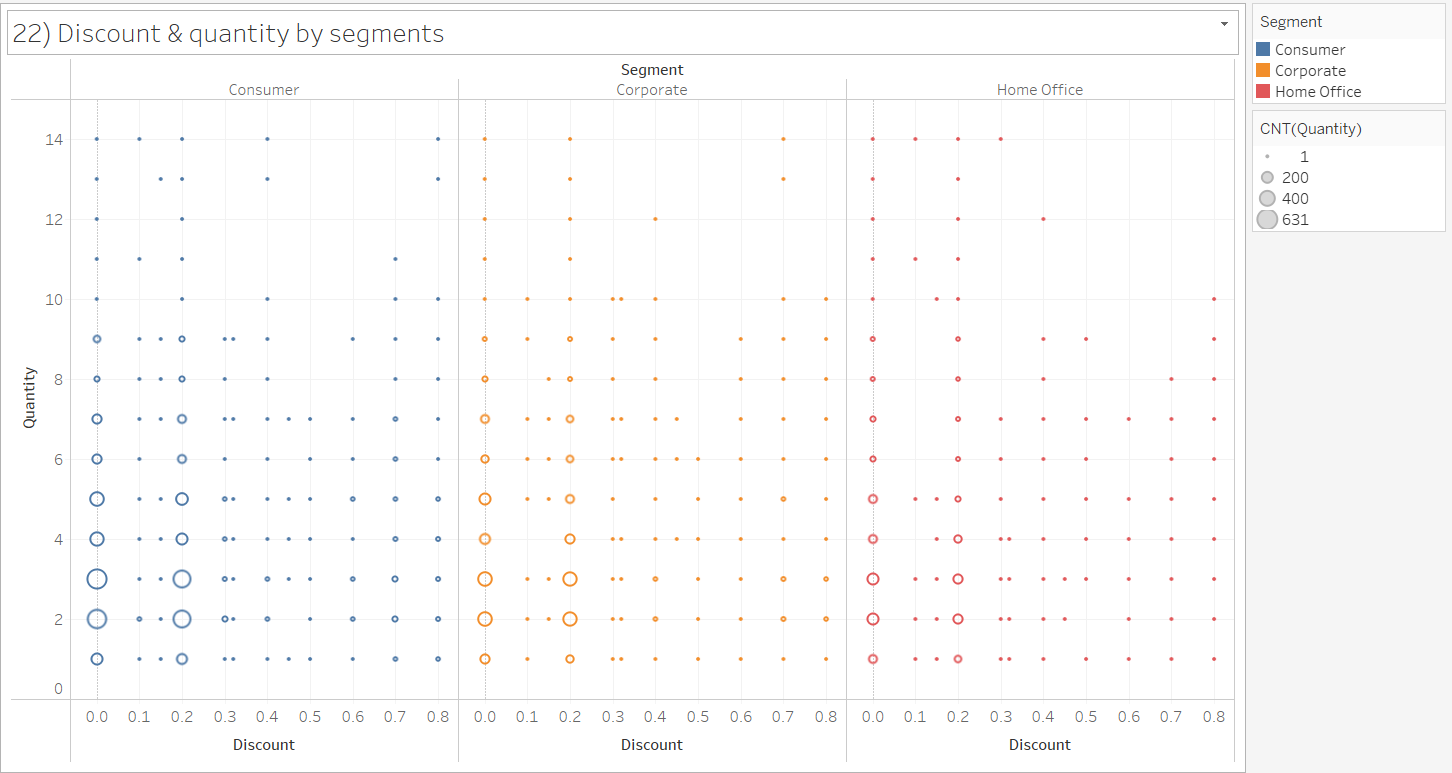
**22.Can we visualize the correlation between discount rates and order quantities for different customer segments?**

**Explanation: Scatter Plot** would be the most suitable chart to show the correlation between discount rates and order quantities for different customer segments. It illustrates the following patterns of order quantities related to different discount rates with respect to the customer segments, with the help of bubbles scattered around the visualisation and the size of the bubble describing the count of the order quantities it appropriately shows the relation between the multiple variables in a concise and clear format.

**Description:** All the segments have the same trend with respect to the discount rates at 0% and 20% and the high amount of order quantities associated with these discount rates only.

There are comparatively more orders in consumer segments with high discount rates as compared to other segments with a bigger order quantity bubble.

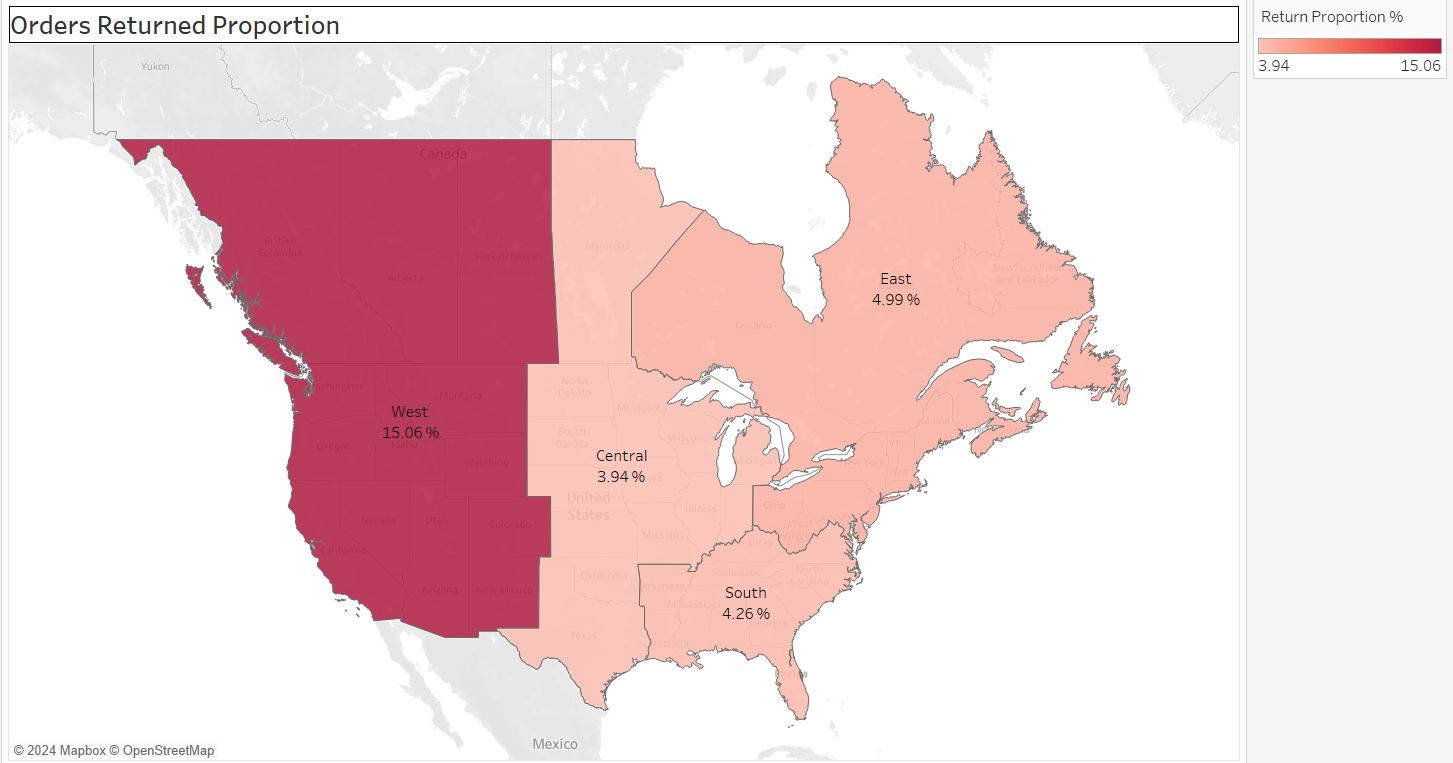
**Chart Type: Scatter Plot**



**23.What is the proportion of orders returned in each region within the Superstore dataset?**

**Explanation:** I have chosen the Map Chart to visualize the proportion of orders returned for each region as it provides quick and easy comprehension of the number of orders returned in percentage terms. A Map chart provides a spatial representation of return locations, enabling easy identification of regions with higher return rates.

**Chart Type: Map Chart**

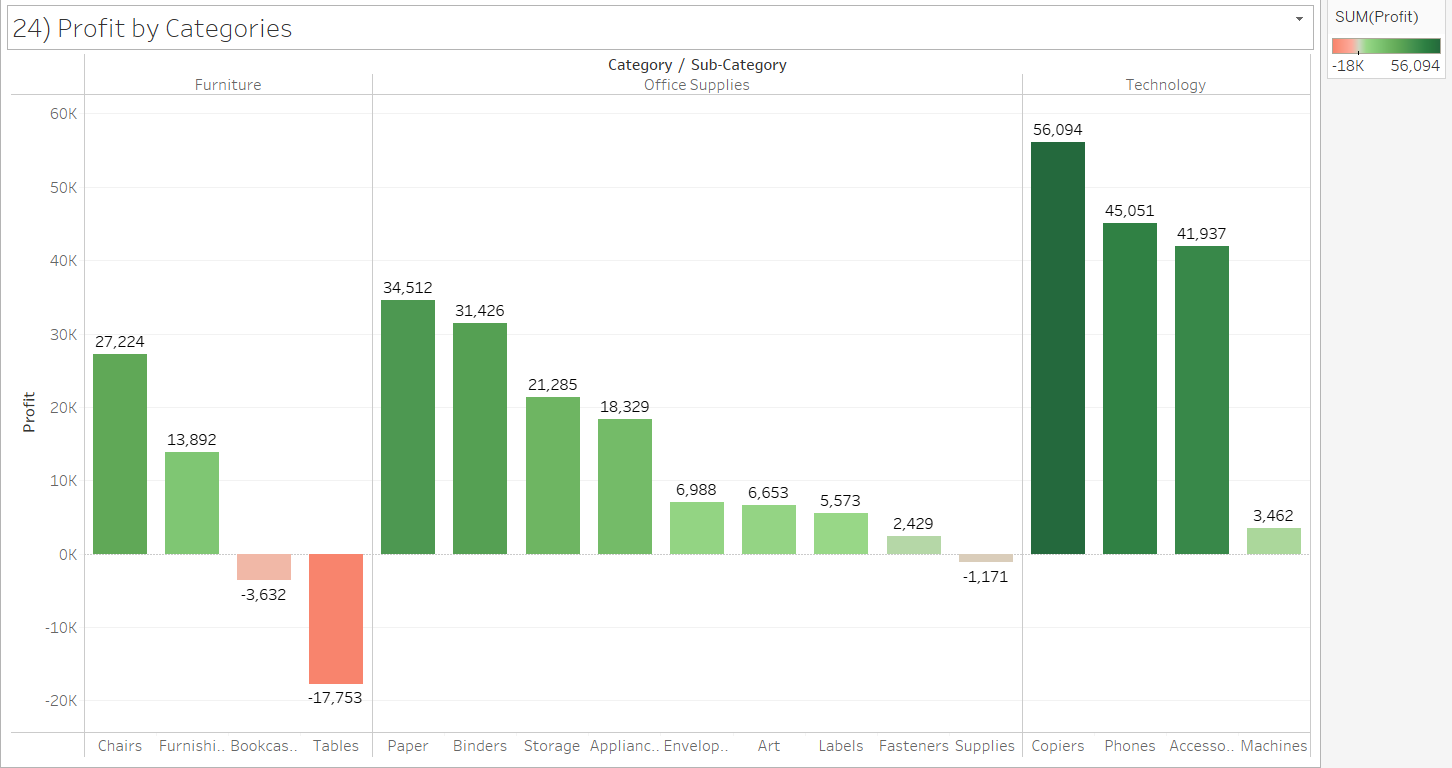
****

**Description:**The west region has getting more returns with the value of 61.25% and the south region getting lowest returns with the value of 8.62%. It will affect the overall profitability of the west region because it has the highest returns compared to other Regions.

**24.Can you compare the profit of different products for different subcategories?**

**Explanation:** I have used Bar Chart to compare profits for different product categories and subcategories, it provides easy and detailed comparison of profits and being supported with the colour toning to show the segregation of the profit making categories to the loss making categories.

**Chart Type: Bar Chart**

****

**Description:**

It can be clearly identified that the tables with the 17753 loss is the worst performing sub category followed by Bookcases with 3632 loss. 2 categories in the furniture section are the top 2 poorly performing categories.

The profits of the office supplies are distributed in many sub categories, paper being the most profitable while the supplies being the only sub category bearing loss to the company.

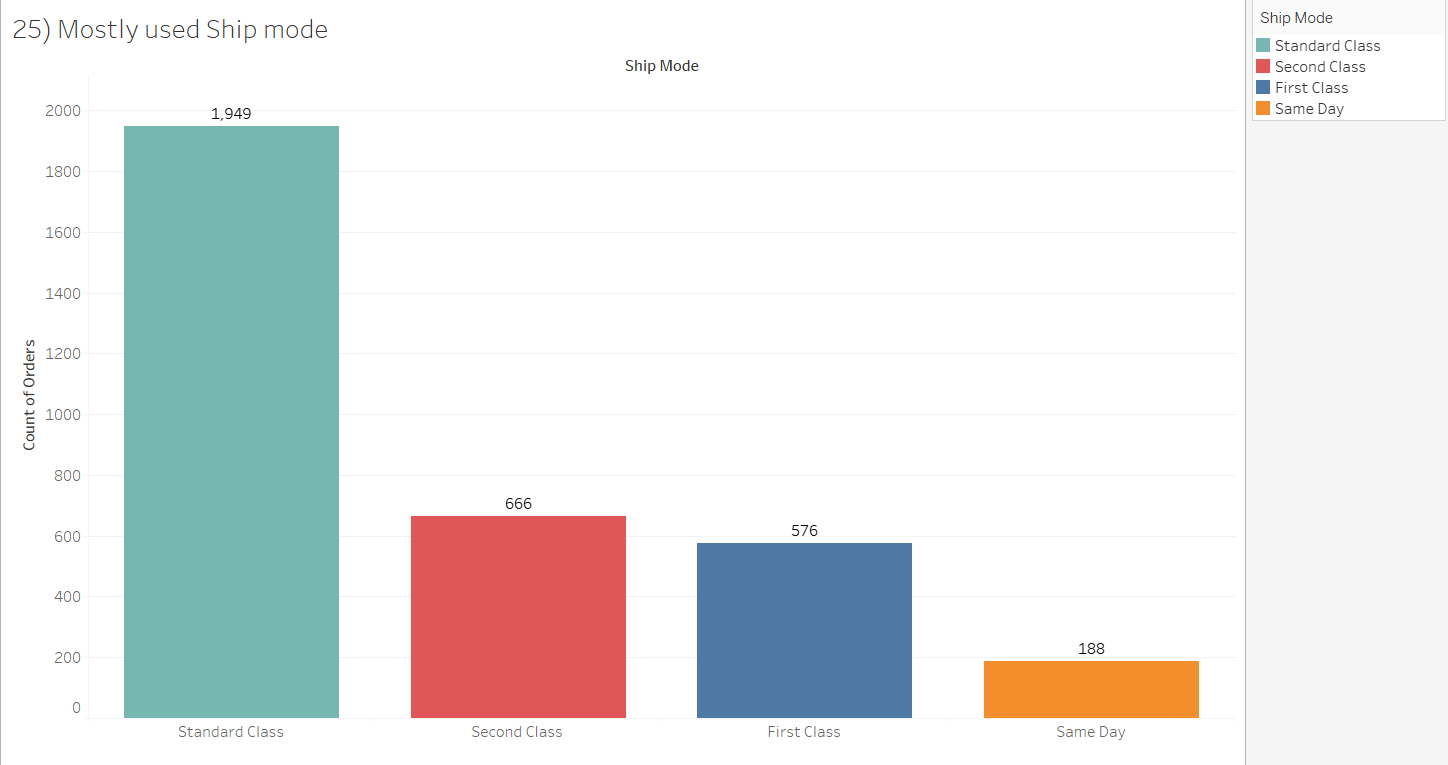
Technology is the best performing category with the concentrated profits, copiers are giving highest profit output, while the machines section is profitable but it is not generating enough profits as compared with other subcategories in the technology category.

**25.Which shipping mode is the most commonly used in the Sample Superstore dataset?**

**Explanation:** I choose Bar Chart to effectively display the frequency or count of different categories, making it suitable for comparing the usage of different shipping modes in the Sample Superstore dataset.

A Bar Chart allows for a straightforward comparison of the frequency or count of each shipping mode. Each bar represents a shipping mode, and the length of the bar depicts its usage frequency. This visualization method facilitates easy identification of the most commonly used shipping mode in the Sample Superstore dataset.

**Chart Type: Bar Chart**

****

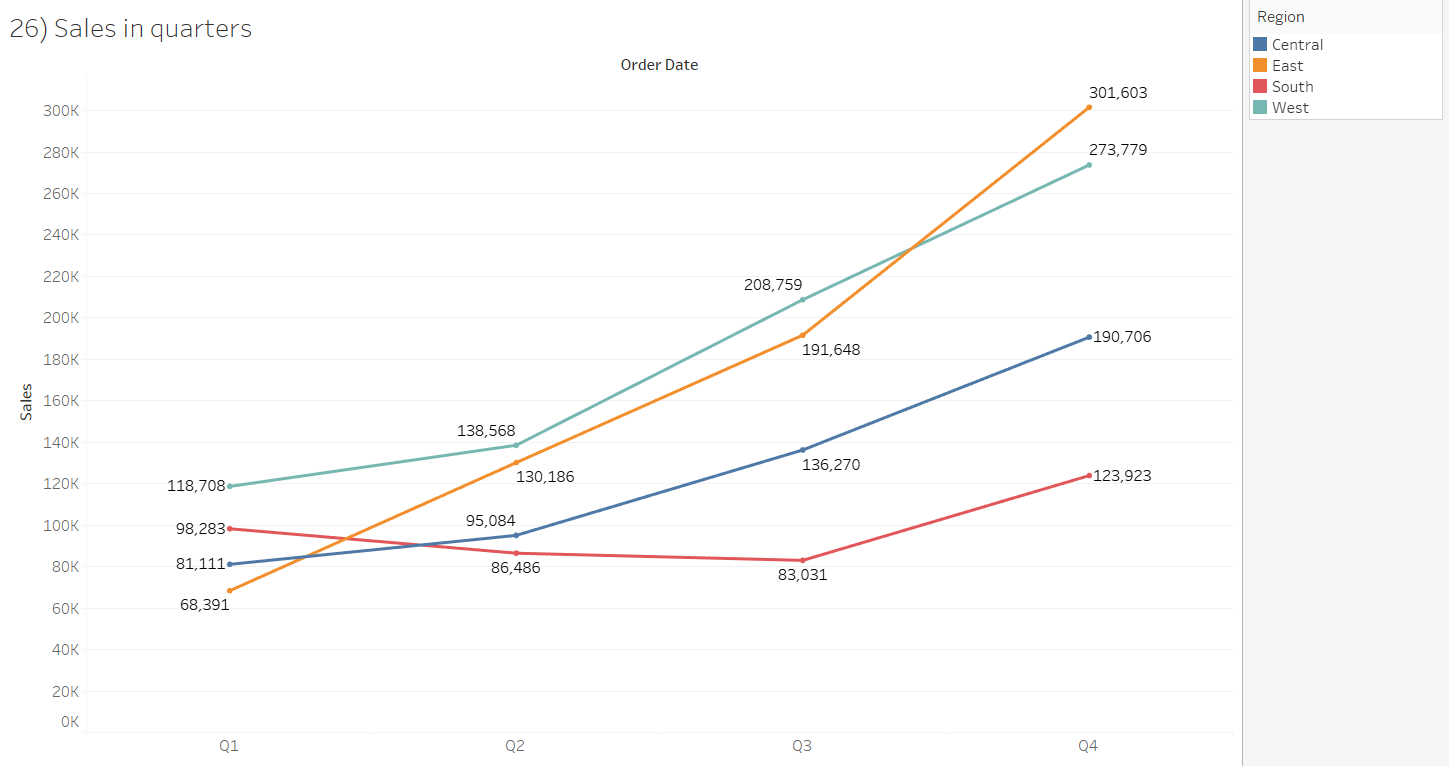
**Description:** After visualizing by viewing the chart we understood that the most commonly used shipping mode is standard class with 1st rank.

Standard class is the most commonly used shipping mode because it has more orders shipped compared to other shipping modes. Standard class has the highest orders followed by the second class, then first class and same day has the lowest shipping orders.

**26.How does the sales performance of different regions evolve throughout the quarters of a year?**

**Explanation:** I choose a Line Chart to illustrate the sales performance of different regions throughout the quarters of a year. A line chart effectively communicates trends over time, providing a clear visualization of how sales evolve across different regions each quarter. A Line Chart will visually represent the fluctuations in sales for different regions over the quarters, allowing for easy identification of any patterns, seasonality, or trends in sales performance throughout the year.

**Chart Type: Line Chart**

****

**Description:** The sales performance of different regions evolves throughout the quarters of a year. The sales performance of a different region quarter 4 was highest compared to all regions. Data represented is based on total sales throughout all years.

The east region in Q1 18,953 it evolves as highest in Q4 with the value of 99,368.

The west region in Q1 52,164 it evolves as highest in Q4 with the value of 85,482.

The south region in Q1 13,642 it evolves as highest in Q4 with the value of 56,064.

The central region in Q1 40,530 it evolves as highest in Q4 with the value of 46,191.

All regions sales performance increases throughout the quarter year compared to Q1.

**27.What is the distribution of order priorities across different product categories?**

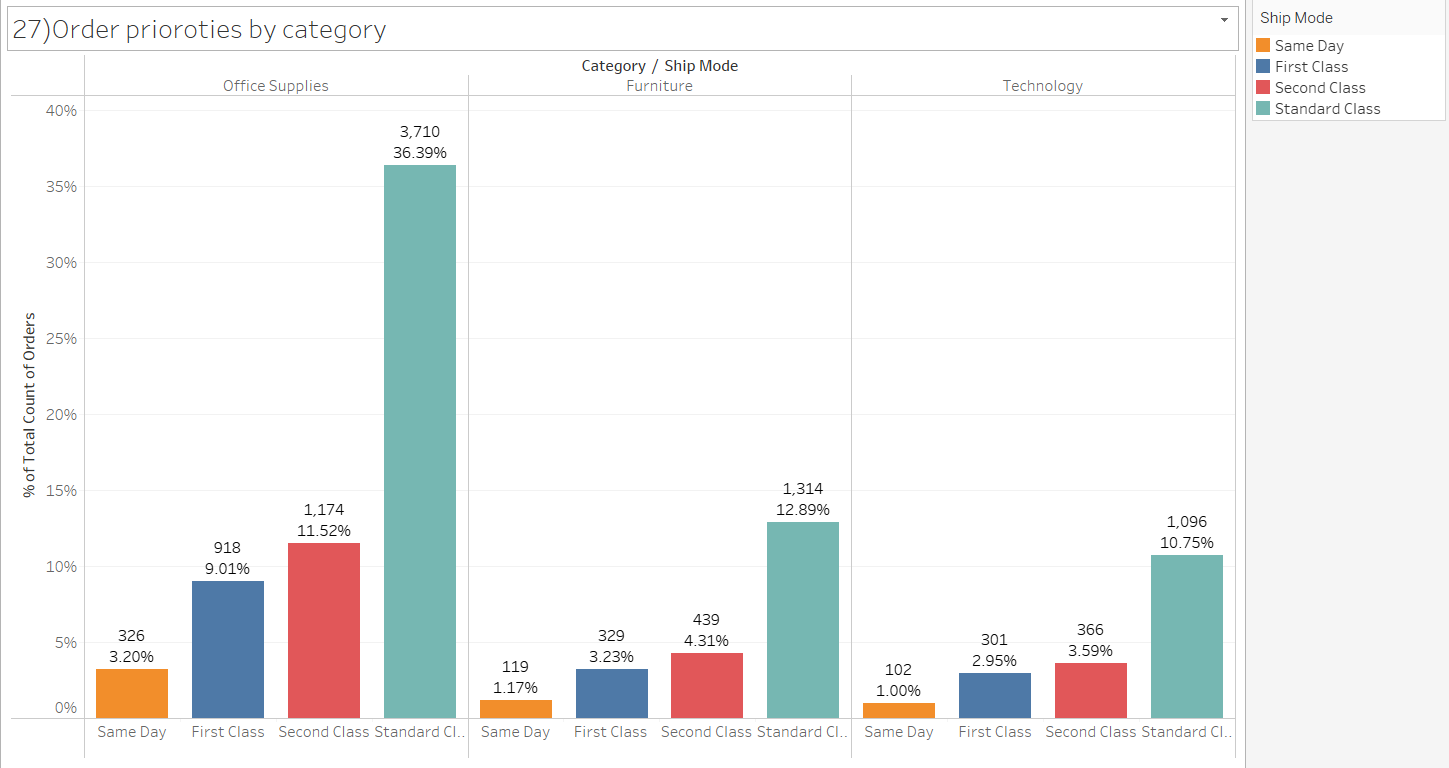
**Explanation:** I choose a Side by Side Bar Chart to visualize the distribution of order priorities across different product categories. A Side by Side bar chart effectively illustrates the total value (total orders) and its components (order priorities), making it suitable for comparing the distribution of priorities across categories.

**Description:**

Office Supplies has the most orders that need to be shipped on priority approx. 3.2% that is 326 orders followed by furniture 1.17% 119 orders and then technology with 1% 102 orders.

While most of the people prefer the standard class more as compared to other shipping modes.

**Chart Type: Side by Side Bar Chart**

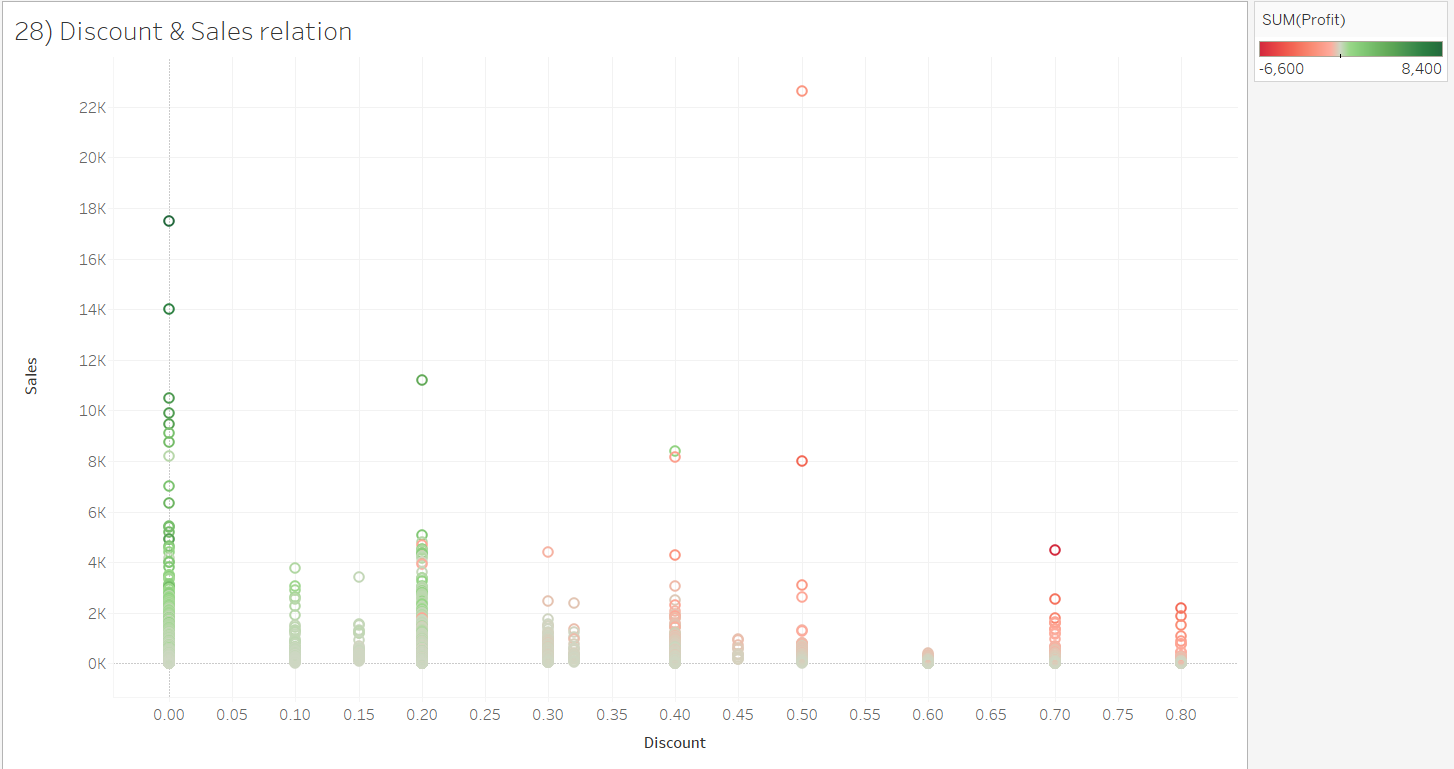
****

**28.What is the relationship between discounts and sales?**

**Explanation: Scatter Plot** is the most appropriate chart for showing the relationship between the discount and sales, showcasing whether there is a positive, negative, or no correlation between discounts and sales.

Scatter Plots are useful for identifying patterns and trends as with increasing discounts the sales are decreasing , and some outliers have been detected for the same.

**Chart Type: Scatter Plot**

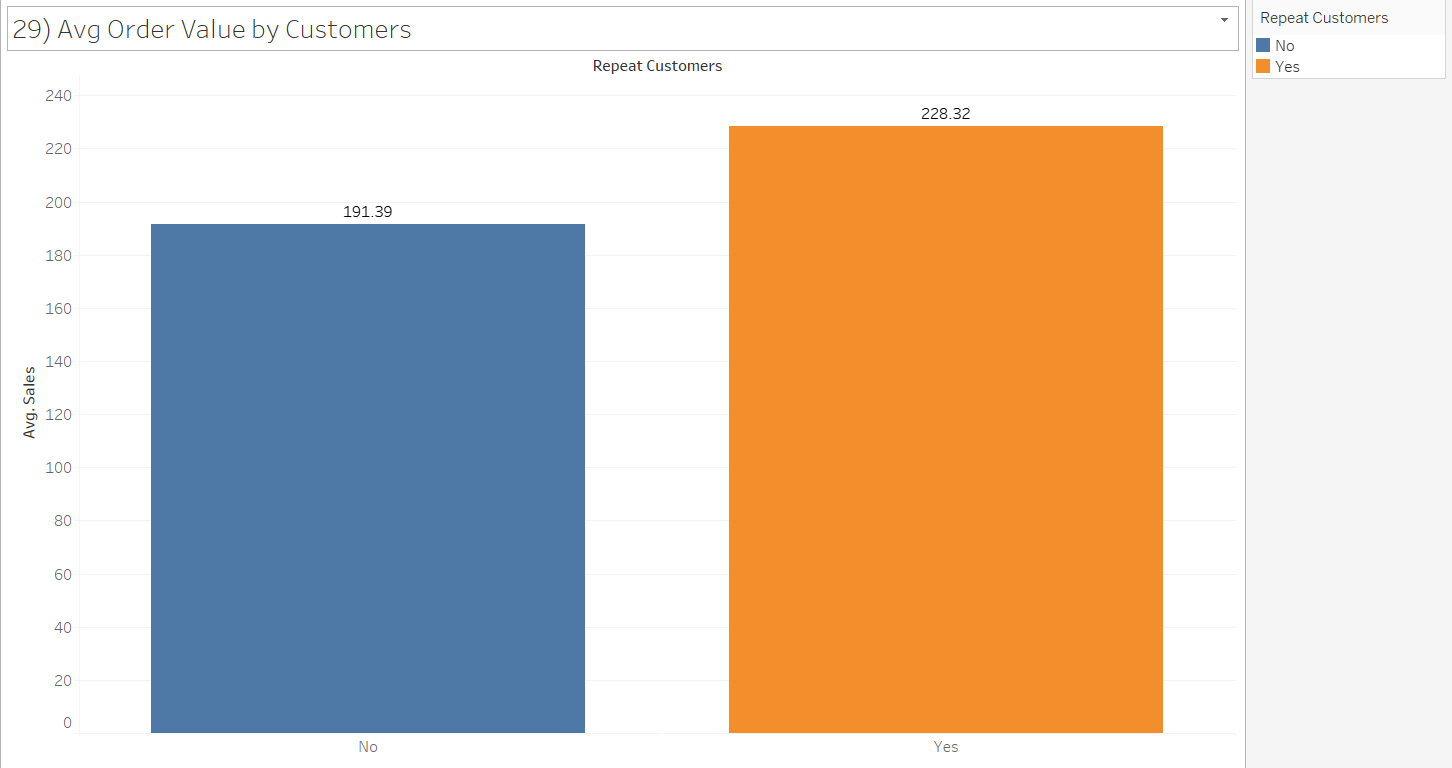
****

**Description:** The Relationship between discount and sales are when discount percentage decreases the sales profit will be increased when the discount is increased then the sales profit will be decreased automatically. Certain Outliers have been detected but overall if there is a high discount on sales the overall profit will be affected.

**29.How does the average order value differ between repeat customers and new customers?**

**Explanation:** A Bar Chart has been used to show the difference between the avg. order value of New and repeat customers. Bar Chart is the best option to visualize and compare the data in different categories.

**Chart Type: Bar Chart**

****

**Description:**

The average order value difference between the new and existing customers is 28 but the count of orders being made by these customers has a huge difference, as new customers made 27 orders while the existing customers have made 10,167 orders.

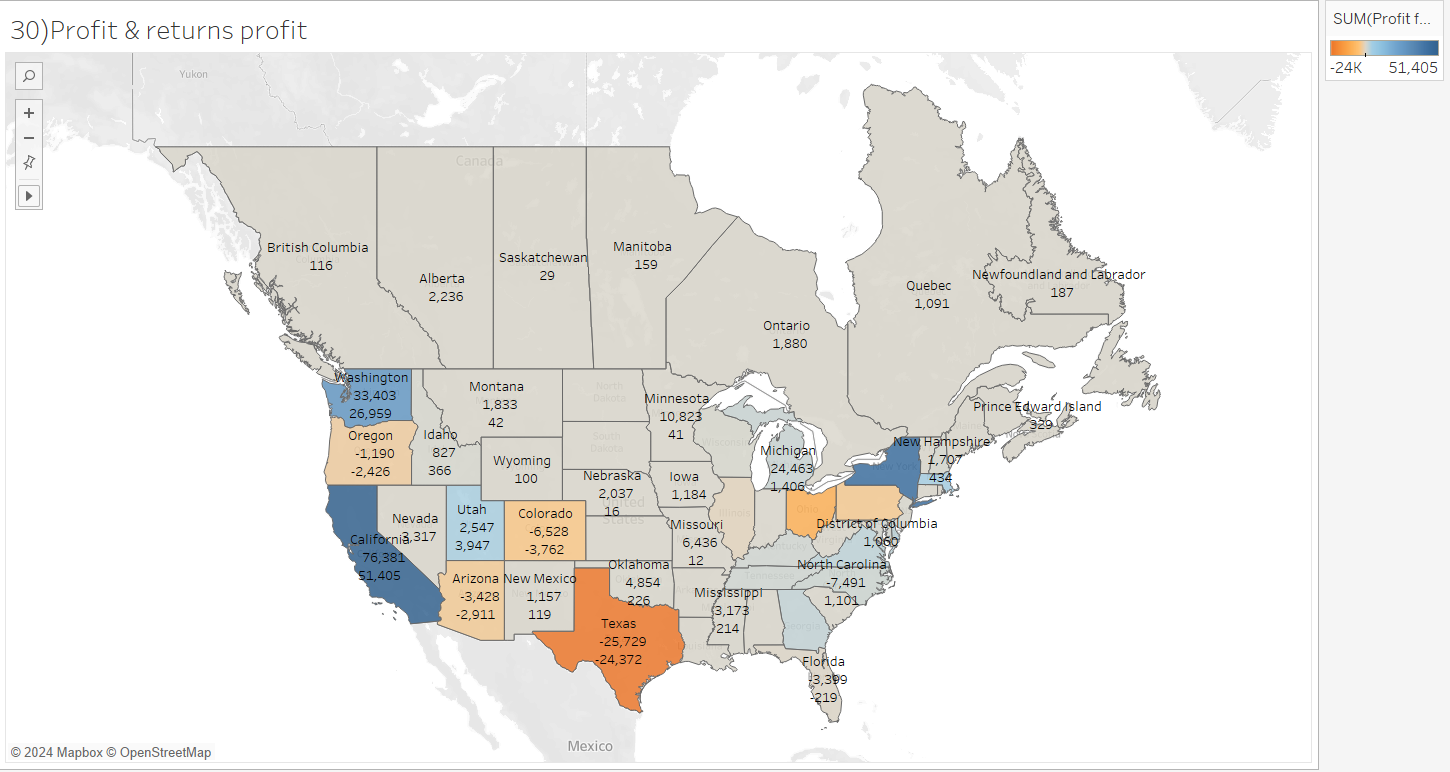
The magnitude of this difference suggests that building customer loyalty and encouraging repeat business can significantly impact the overall revenue generated. It implies that repeated customers, over time, demonstrate a willingness to invest more in each transaction, showcasing a stronger commitment and potentially deeper satisfaction with the products or services offered. This insight is valuable for businesses aiming to tailor marketing strategies and incentives to nurture and retain their existing customer base

**30.What is the geographical distribution of returns and its impact on overall profitability?**

**Explanation:** I choose a Map chart to visualize the geographical distribution of returns. A map chart is effective for showcasing spatial patterns, allowing for an understanding of where returns are concentrated. Additionally, you may complement this with a stacked bar chart to illustrate the impact of returns on overall profitability.

A Map chart provides a spatial representation of return locations, enabling easy identification of regions with higher return rates. A stacked bar chart can accompany this to showcase the impact of returns on overall profitability. Together, these visualizations offer insights into the geographical distribution of returns and their influence on profitability.

**Chart Type: Map Chart**

****

**Description:**

The geographical distribution of returns in overall profits are the United States California state with the profit of 76,381 and the impact of returns in overall profits 51,405. The United States Texas the impact of returns in overall profits -24372. The returns will impact the overall profitability because the returns will have shipping mode costs on sales throughout the different regions and country in different categories and also the returns of products will impact differently on profits it will impact on profitability across geographical distribution.