**DATA VISUALIZATION and ANALYSIS PROJECT**

**MID SUMMATIVE ASSIGNMENT**

**Name – Ravi Sharma**

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

**About the Project:** In this project, I’m analyzing a Superstore dataset to address 30 scenario-based questions. I will use Tableau for data visualization, selecting the most effective chart for each question and providing detailed explanations based on the chosen visualizations

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

**Explanation**: I opted for a Horizontal Bar Chart to facilitate a straightforward comparison of total sales among various product categories. The length of the bars provides a visual representation that makes it simple to spot which categories have the highest sales totals.

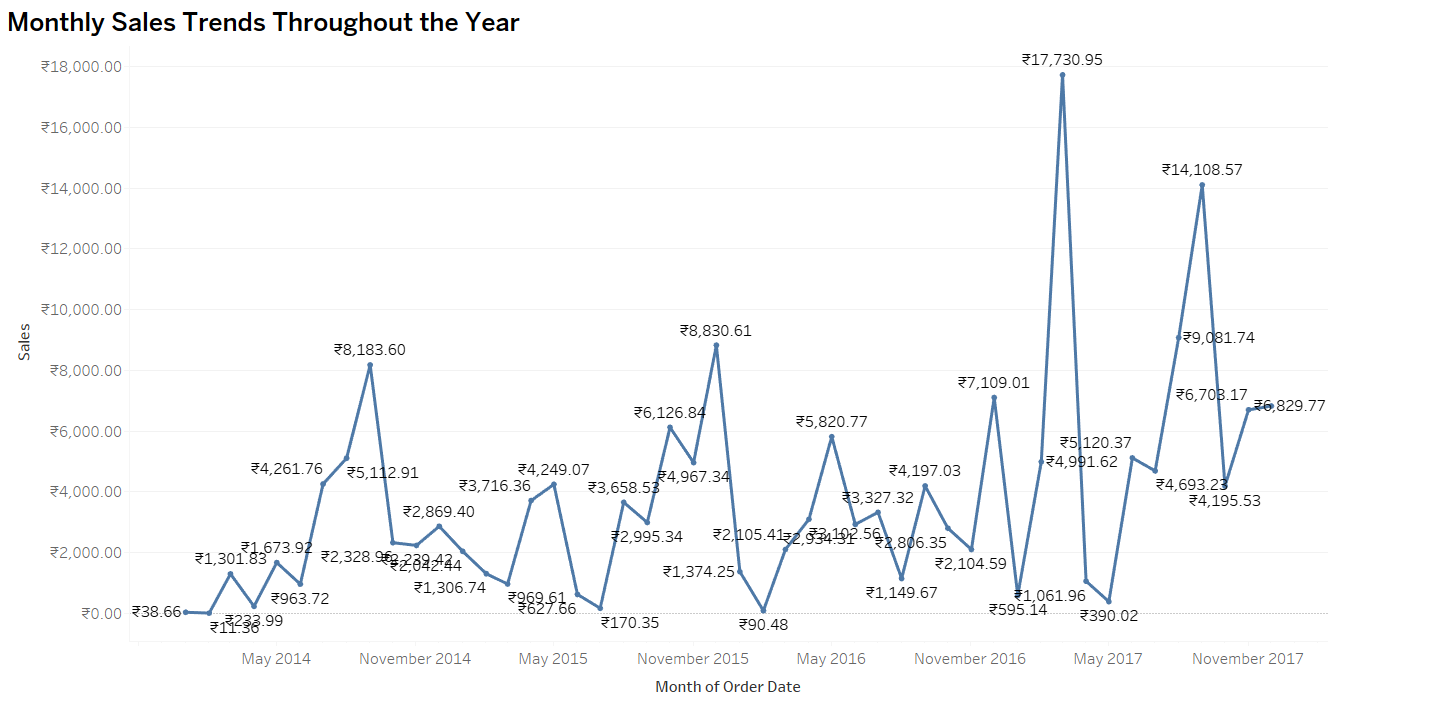
**Chart Type: Horizontal Bar Chart**

**Description:** This visualization helps us identify that the product category with the highest total sales is Technology, with a cumulative sales value of 72,708. This value represents the sum of sales across all years in the dataset.

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

**Explanation:** I opted for a Line Chart as it adeptly showcases trends over time. By plotting monthly sales amounts along the x-axis and y-axis, the line chart clearly visualizes how sales vary throughout the year.

**Chart Type: Line Chart**

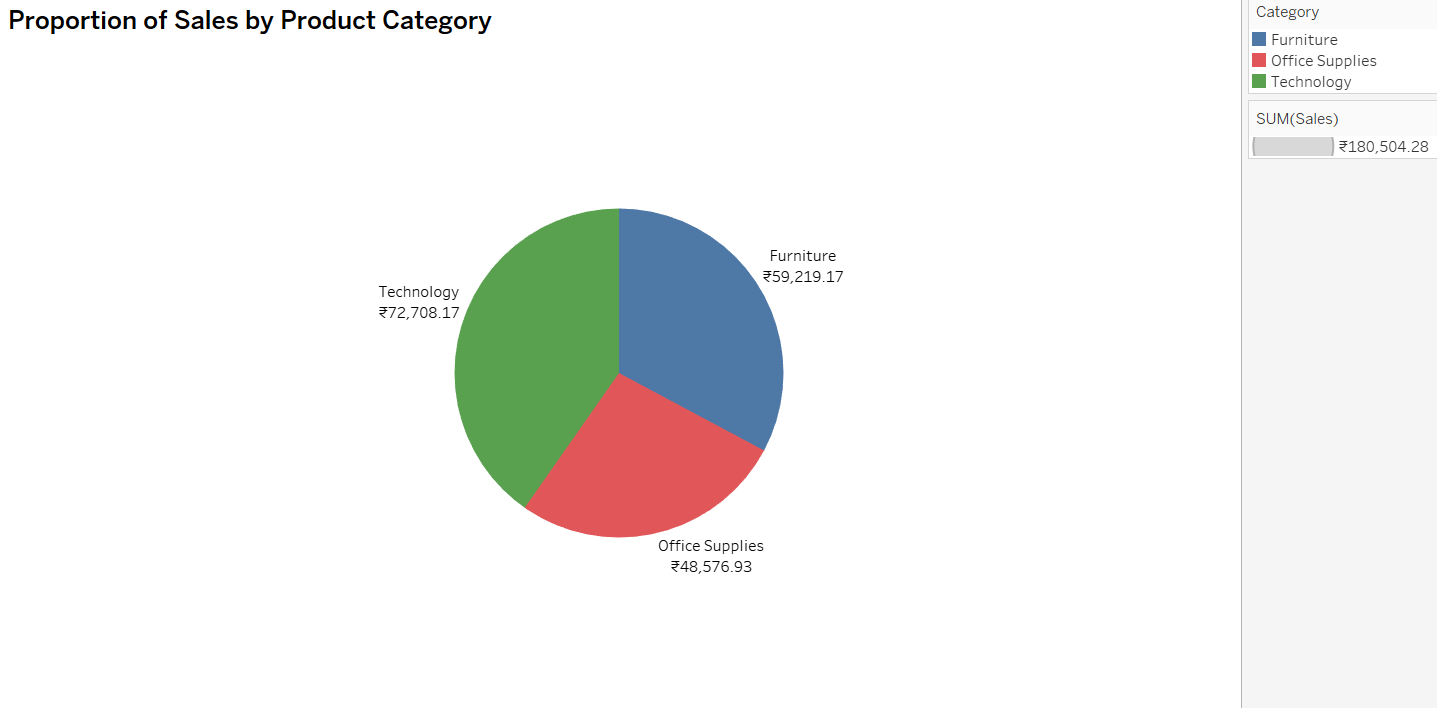
**Description:** Using the line chart, we can observe the sales trends across multiple years. The key insight is that sales consistently peak in November each year.

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

**Explanation:** I selected a Pie Chart as it offers a straightforward and succinct way to depict the percentage contribution of each product category to the total sales. The distinct colors for each category make it effortless to understand the distribution at a glance.

**Description:** The total sales amount is distributed across various product categories, with each category contributing a different percentage to the overall sales. By examining this distribution, we can see which categories are the highest performers and understand their relative impact on the total sales. The diverse colors used in the visualization represent each product category, providing a quick and clear picture of their sales contributions.

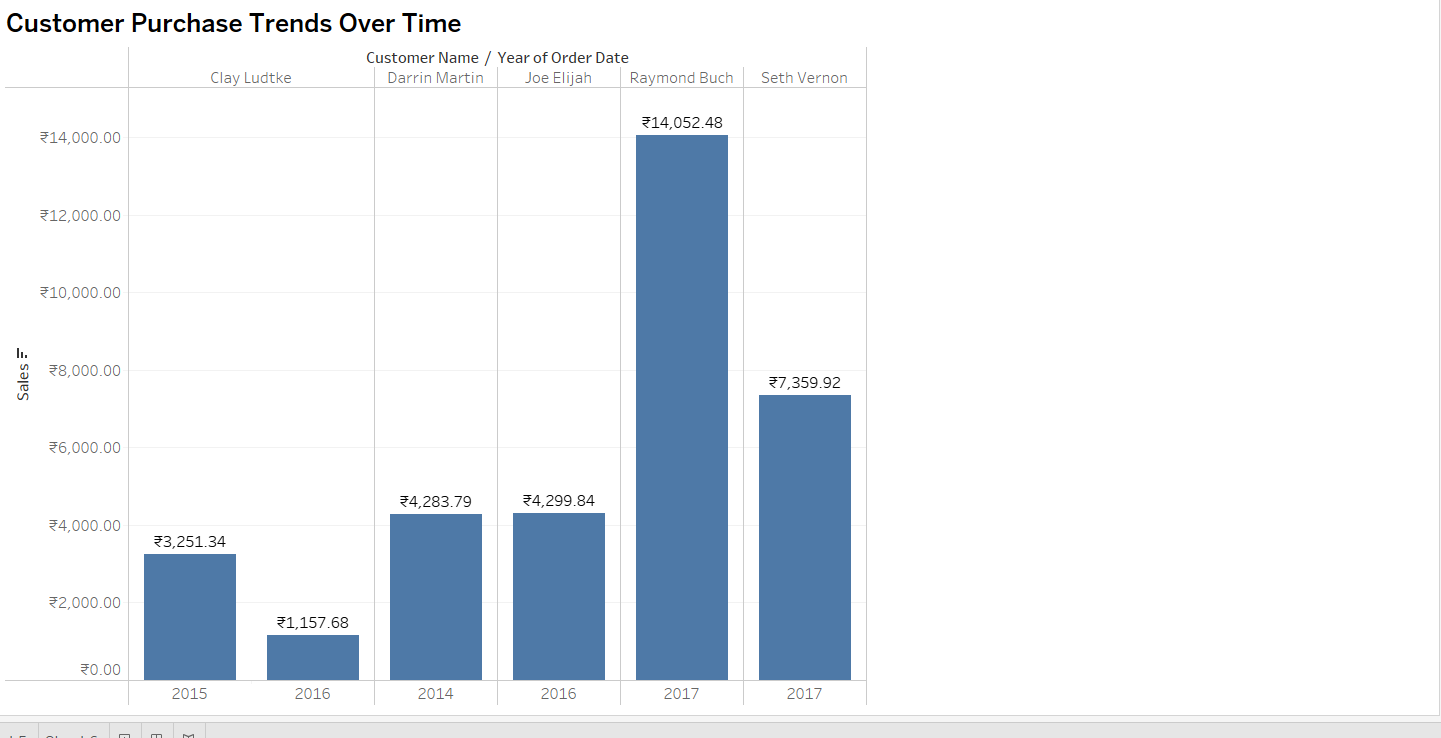
**Chart Type: Pie Chart**

****

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

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

**Chart Type: Bar Chart**

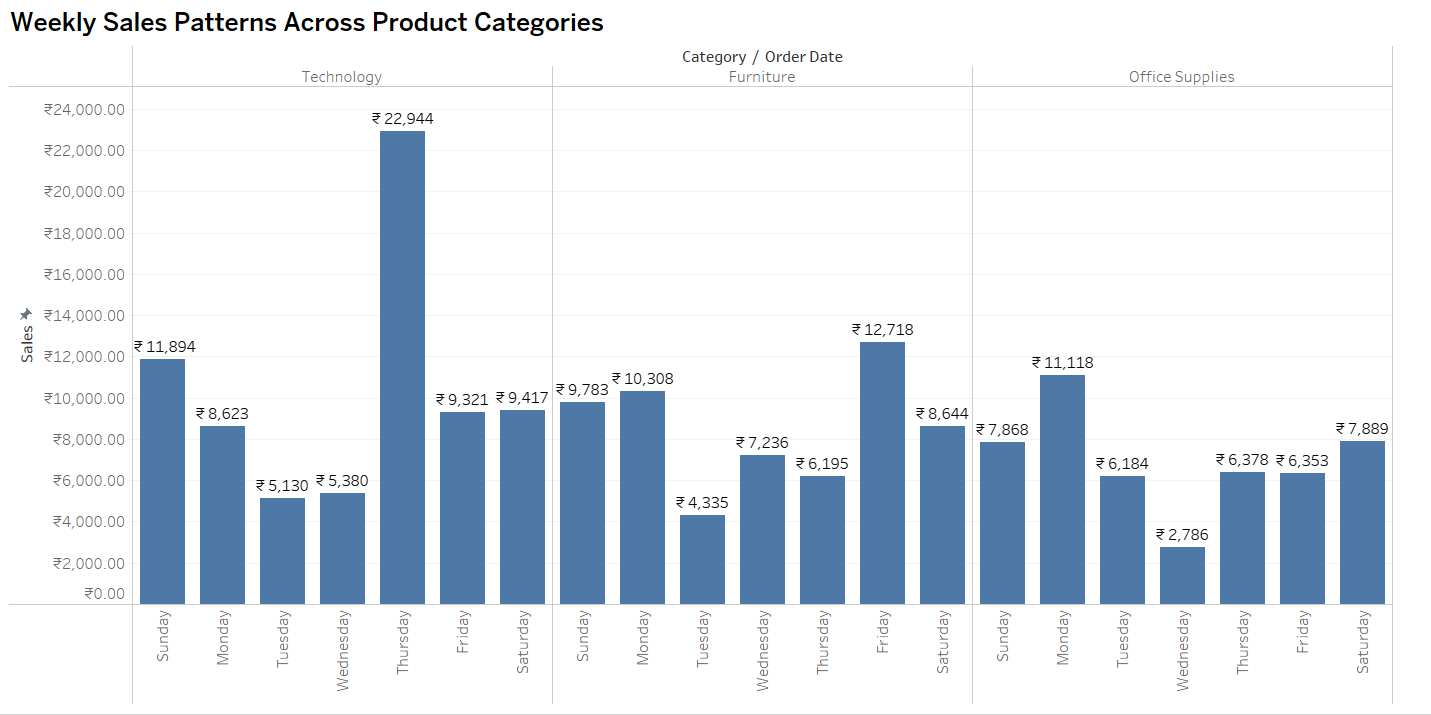
****

**Description:** We're focusing on the top 5 customers with the highest sales for this analysis. By examining their sales changes over the year, we gain valuable insights into their purchasing patterns. This approach can also be applied to analyze other customers.

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

**Explanation**: I opted for a Side-by-Side Bar Chart because it offers a clear comparison of sales across various days of the week for each product category. This arrangement allows for easy visual comparison and helps in identifying trends effectively.

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

****

**Description:**

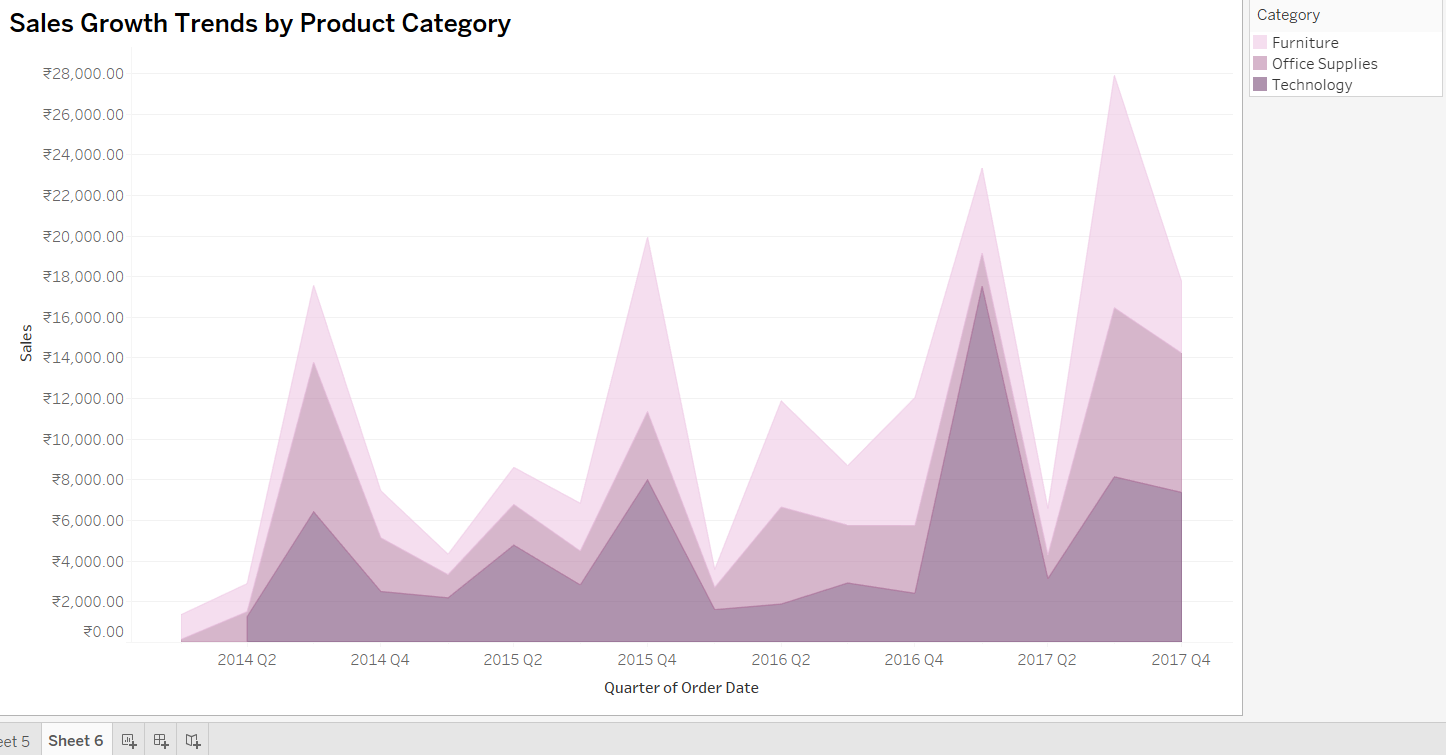
Upon visualizing the sales data, we noticed that the sales for different product categories vary throughout the week. The highest sales were recorded in the Technology category on Thursday, amounting to 22,944. Conversely, the lowest sales were observed in the Office Supplies category on Wednesday, with a value of 2,786.

This analysis indicates that while daily sales fluctuate, the lowest sales typically occur mid-week, particularly on Wednesdays. Conversely, higher sales are often seen at the start and end of the week. This pattern could be valuable for planning employee schedules and leave to align with peak and off-peak sales periods.

**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:** Upon visualizing the sales data, it’s clear that the technology category has shown significant and consistent growth, with sales increasing nearly every quarter. While office supplies and furniture also exhibit growth, their pace lags behind technology.

Interestingly, furniture experienced the highest sales in Q4 2020. However, the rapid growth in sales of technology and office supplies soon outpaced furniture, with technology taking the lead and office supplies following in second place. This shift highlights customer preferences and buying behavior towards these product categories.

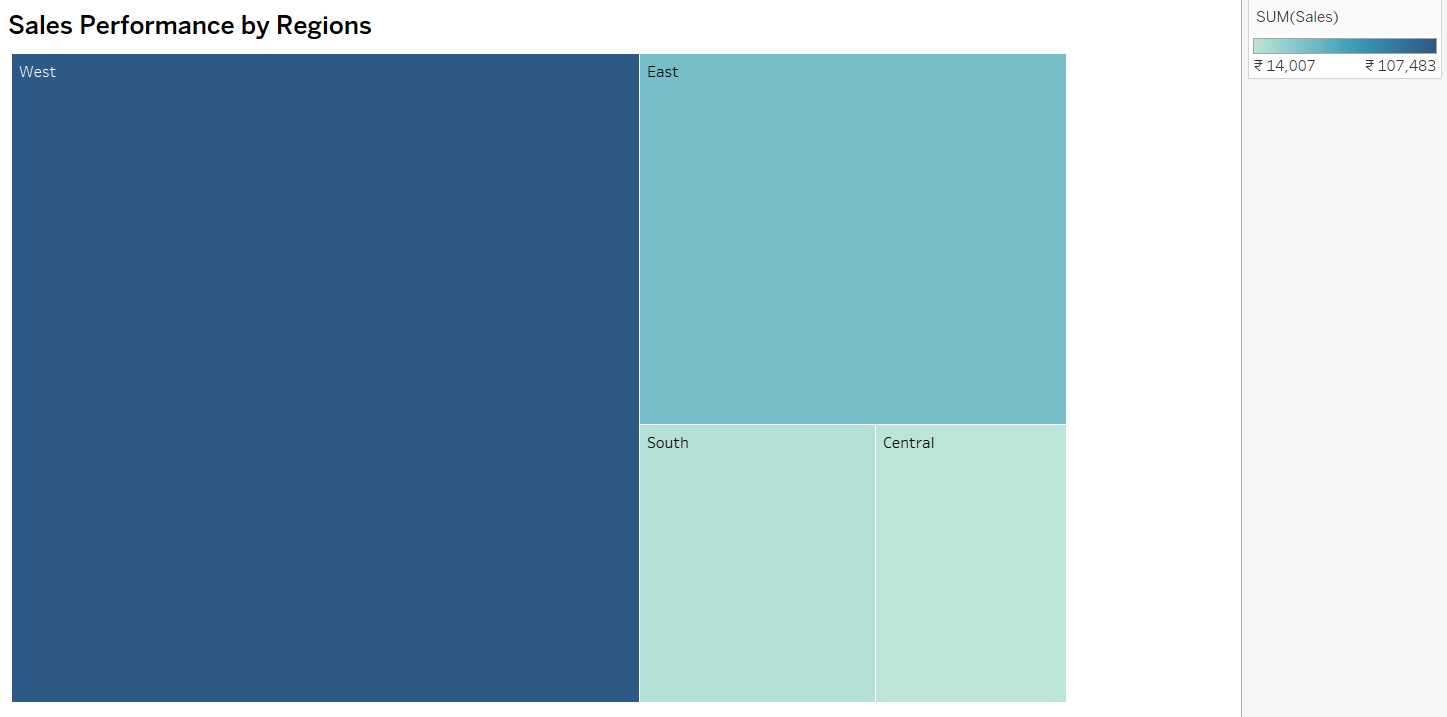
**7.How does the sales distribution vary across different regions in the "Superstore" dataset?**

**Explanation:** This Treemap graph offers an insightful hierarchical view of sales distribution across different regions, including West, East, South, and Central. It effectively communicates how sales revenue is apportioned among these regions, allowing viewers to quickly grasp regional sales trends and patterns.

**Chart Type: Treemap Chart**

**Description:**

By examining the Treemap, decision-makers can gain valuable insights into regional sales dynamics. This includes identifying potential growth opportunities, recognizing top-performing regions, and understanding the distribution of sales resources effectively

****

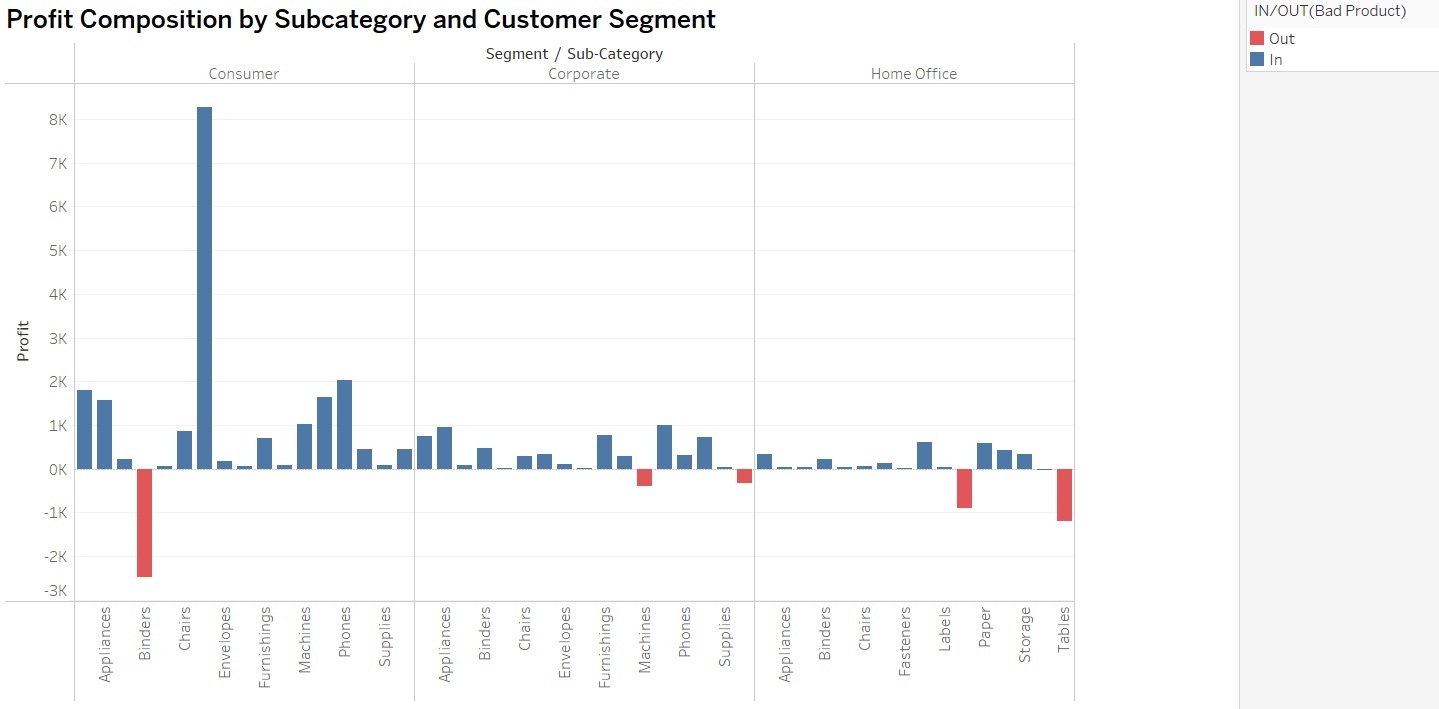
In summary, this Treemap provides a comprehensive visual representation of sales distribution by region, offering viewers a hierarchical perspective of sales data. It serves as a valuable tool for decision-makers seeking to understand regional sales trends and make strategic choices to enhance overall sales performance.

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

**Explanation:** The Stacked Bar Chart provides a clear and efficient visual summary of profit distribution across various subcategories within distinct customer segments. This visualization enables quick comprehension of profit composition patterns and trends, aiding decision-makers in identifying the most profitable subcategories within each customer segment.

**Chart Type: Stack Bar Chart**

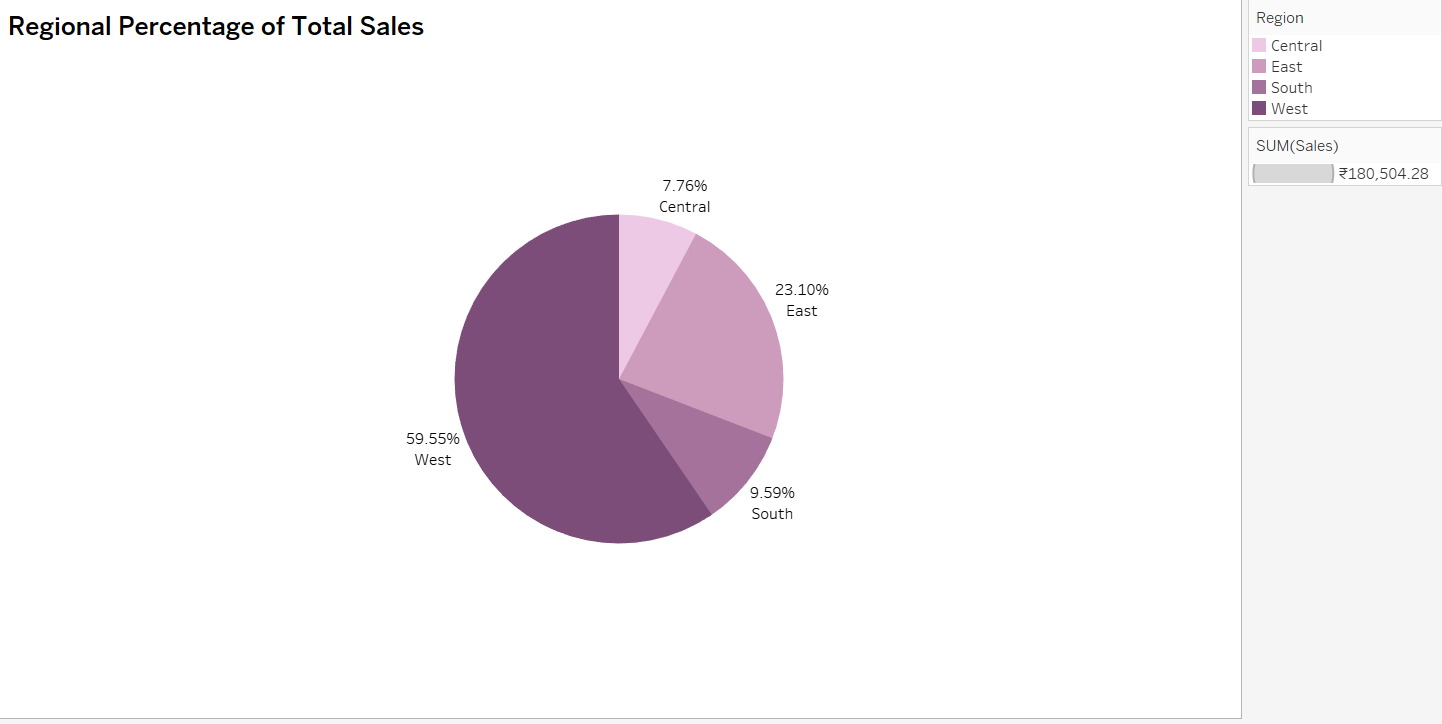
**Description:** Through our visualization, we've discerned that the Consumer segment yields the highest profits for the company, followed by the corporate segment, and then Home Supplies. Losses are indicated in red, representing products that perform poorly.



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

**Explanation:** Choosing a Pie Chart enables the visualization of each region's percentage contribution to the overall sales. This chart type offers a clear overview of sales distribution, facilitating easy comparison between regions. The different regions will be color-coded or shaded according to their percentage contributions, providing an intuitive and straightforward representation.

**Chart Type: Pie Chart**

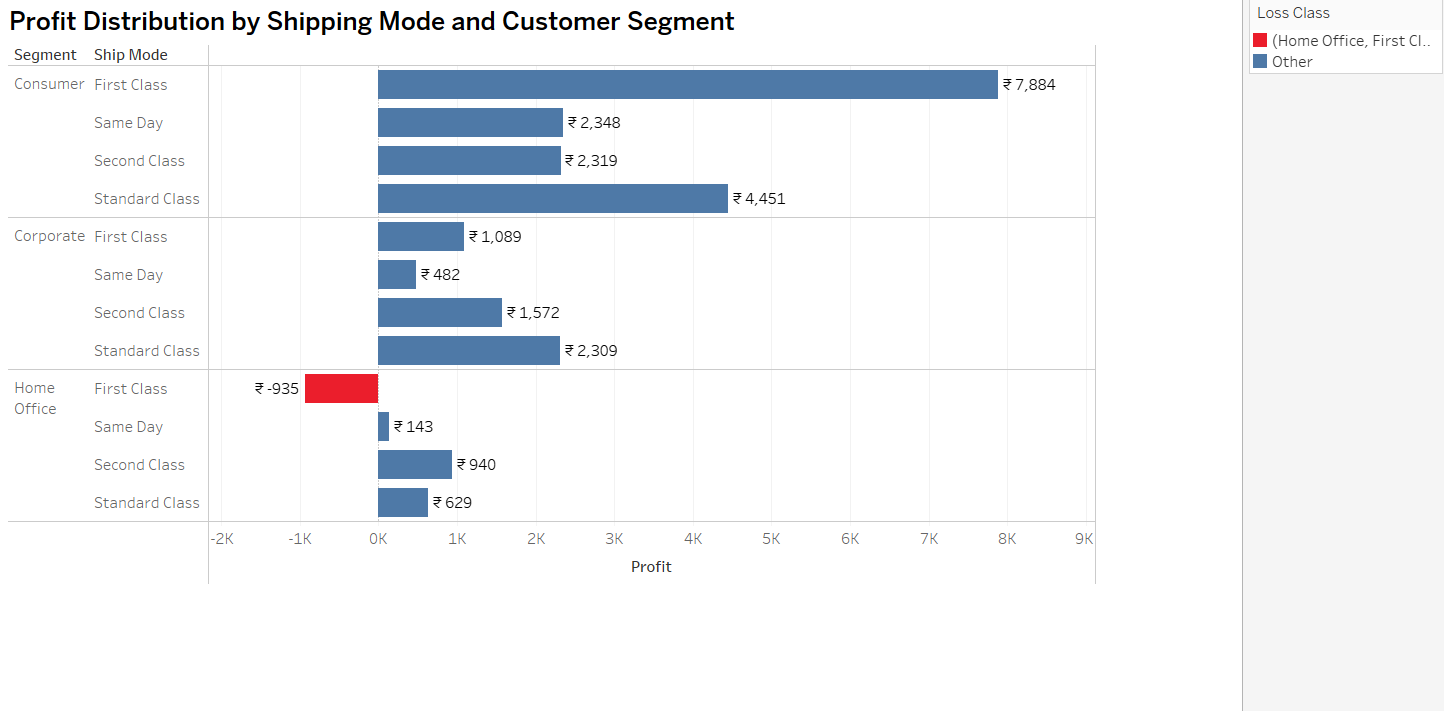
****

**Description:** Upon visualizing the Superstore dataset, we observed that it comprises four distinct regions, each exhibiting varying sales percentages. The West region leads with the highest sales, accounting for 59.55% of the total, while the Central region has the lowest sales, contributing only 7.76%.

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

**Explanation:** A Stacked Horizontal Bar Chart is chosen to visualize the profit margins across various shipping modes and customer segments. This chart type clearly compares profit margins, with individual bars representing each shipping mode and distinct segments within the bars for different customer segments. It offers an intuitive representation, facilitating easy identification of profit margin variations.

**Chart Type: Stacked Horizontal Bar Chart**

****

**Description:** Upon analyzing the chart, we found that the Consumer segment achieves the highest profits in the First Class shipping mode, with a value of ₹7,884. On the other hand, the lowest profit margin in the same shipping mode is observed in the Home Office segment, amounting to ₹143. Additionally, the First Class shipping mode has a loss of ₹935, highlighted in red. Each shipping mode presents varied profit margins across different customer segments. Interestingly, Standard Class shipping stands out by generating higher profits compared to other shipping modes.

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

**Explanation:** A Bar Chart has been chosen to illustrate the time required to deliver orders to clients, effectively showing the number of days needed for the delivery process across different shipping modes. By utilizing drill-down and filtering options, we can further analyze delivery durations for various product categories. In Tableau, a Bar Chart displays tasks as horizontal bars along a timeline, with the length of each bar representing the duration of the task.

**Chart Type: Bar Chart**

****

**Description:** After analyzing the chart, we observed that the delivery duration for different shipping modes is almost the same across all product categories. This uniformity indicates consistent processing times, regardless of the category.

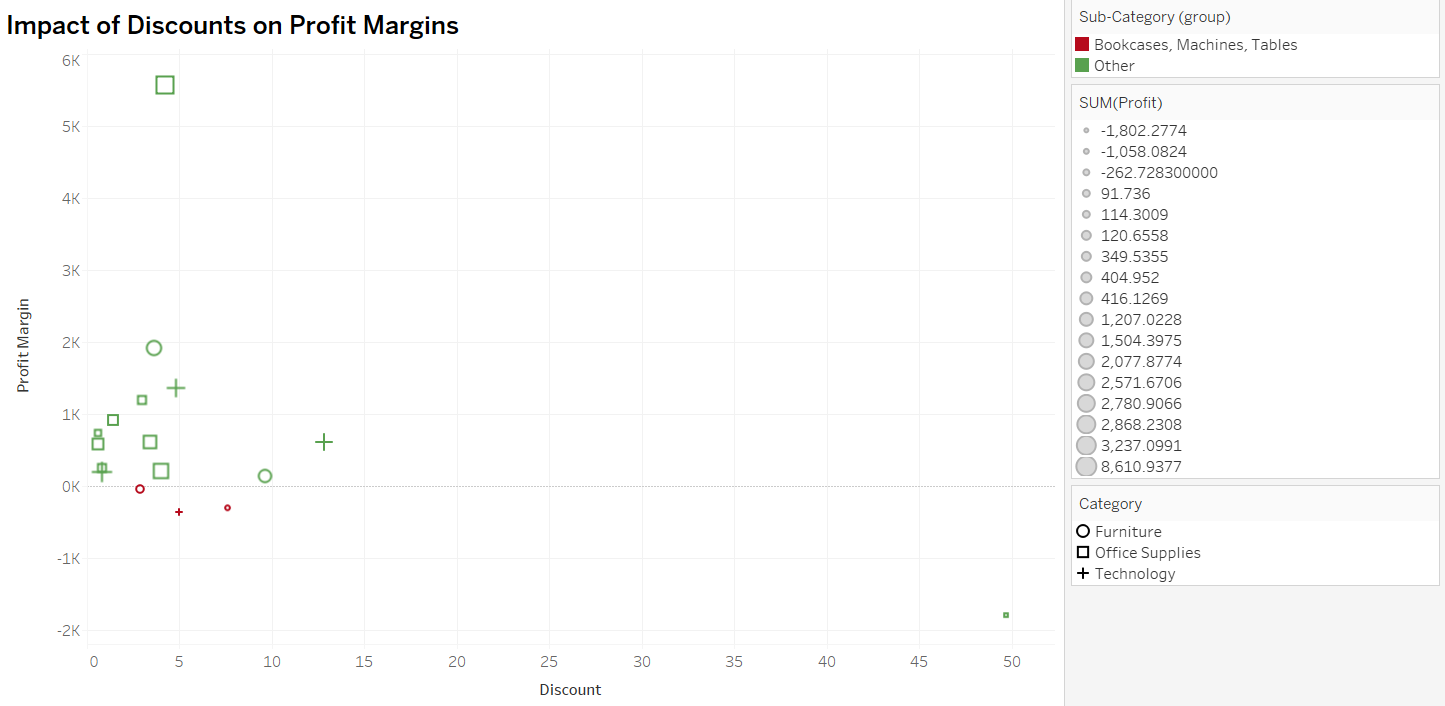
**12.How do discounts affect overall profit?**

**Explanation:** A Scatter Plot has been selected to visualize the effect of discounts on profit. This type of chart is ideal for illustrating the relationship between discounts and profit, as it displays the connection between two measures. The Scatter Plot provides a quick and clear understanding of the relationship between these measures at a glance, highlighting how profits change with different discount rates. It effectively showcases the negative correlation where profits decrease as discounts increase.

**Description:**

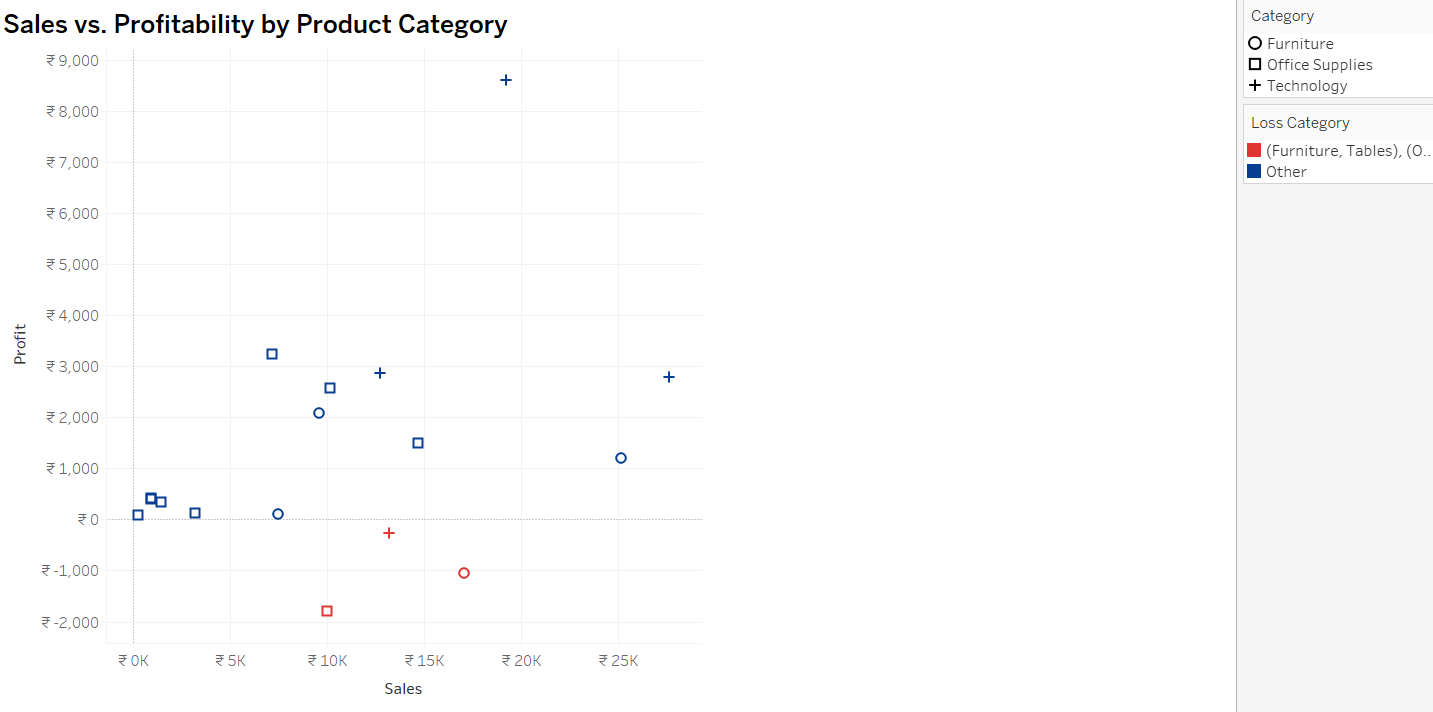
Using a scatter plot, we can observe the profits at every discount level. The visualization reveals that as the discount increases, the profit decreases for many orders. The lowest profit margin is ₹ -364 at a 5% discount and red colour show loss margin, while the highest profit margin, amounting to ₹5578 is observed at a 4% discount.

**Chart Type: Scatter Plot**

****

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

**Explanation:** A Scatter Chart has been chosen to compare product sales and profitability across different categories simultaneously. This chart type facilitates direct visual comparisons, making it easy to assess the relative performance of each category in terms of both sales and profitability. It helps identify patterns or disparities quickly.

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

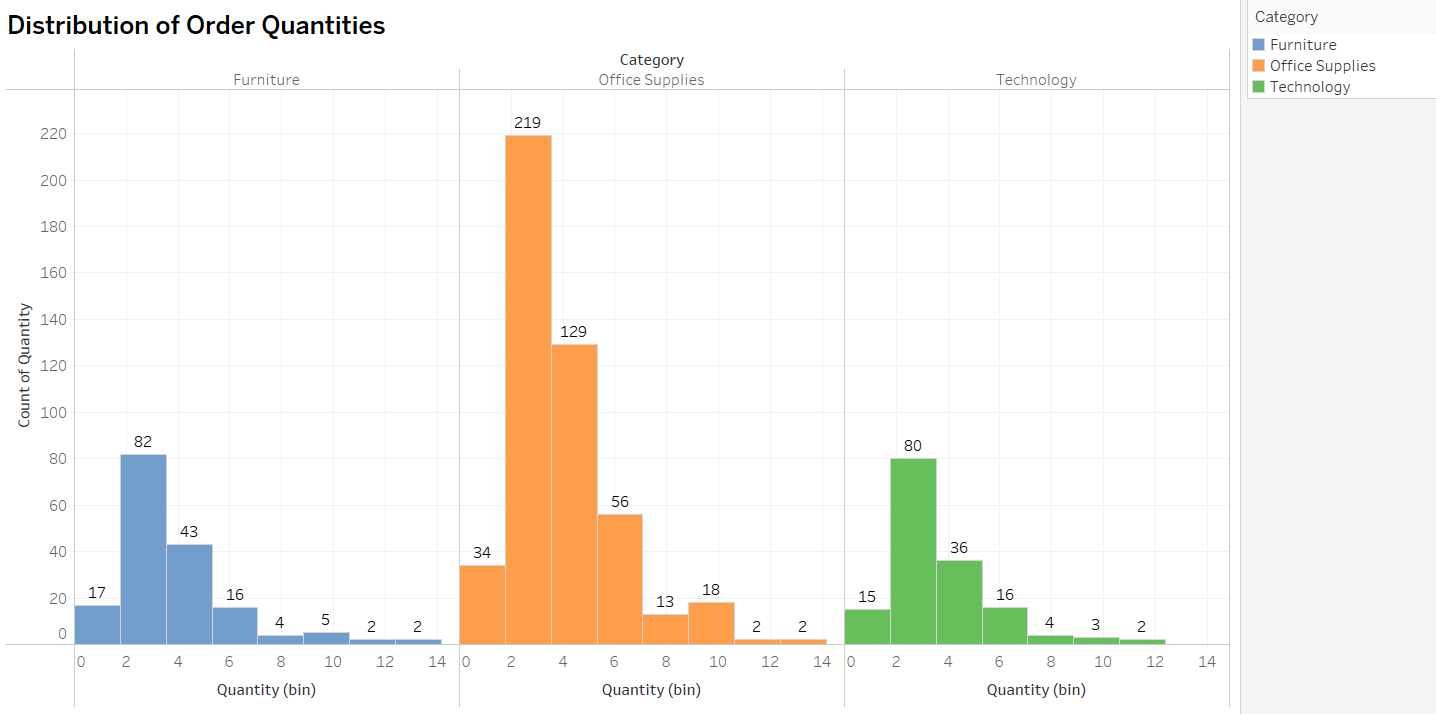
**Description:** The visualization of product sales and profitability reveals relationships among three categories: furniture, office supplies, and technology. Among these, the technology category has achieved higher profits compared to office supplies and furniture.

**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:** Upon analyzing the distribution of order quantities for products, we observed a consistent pattern across all categories.

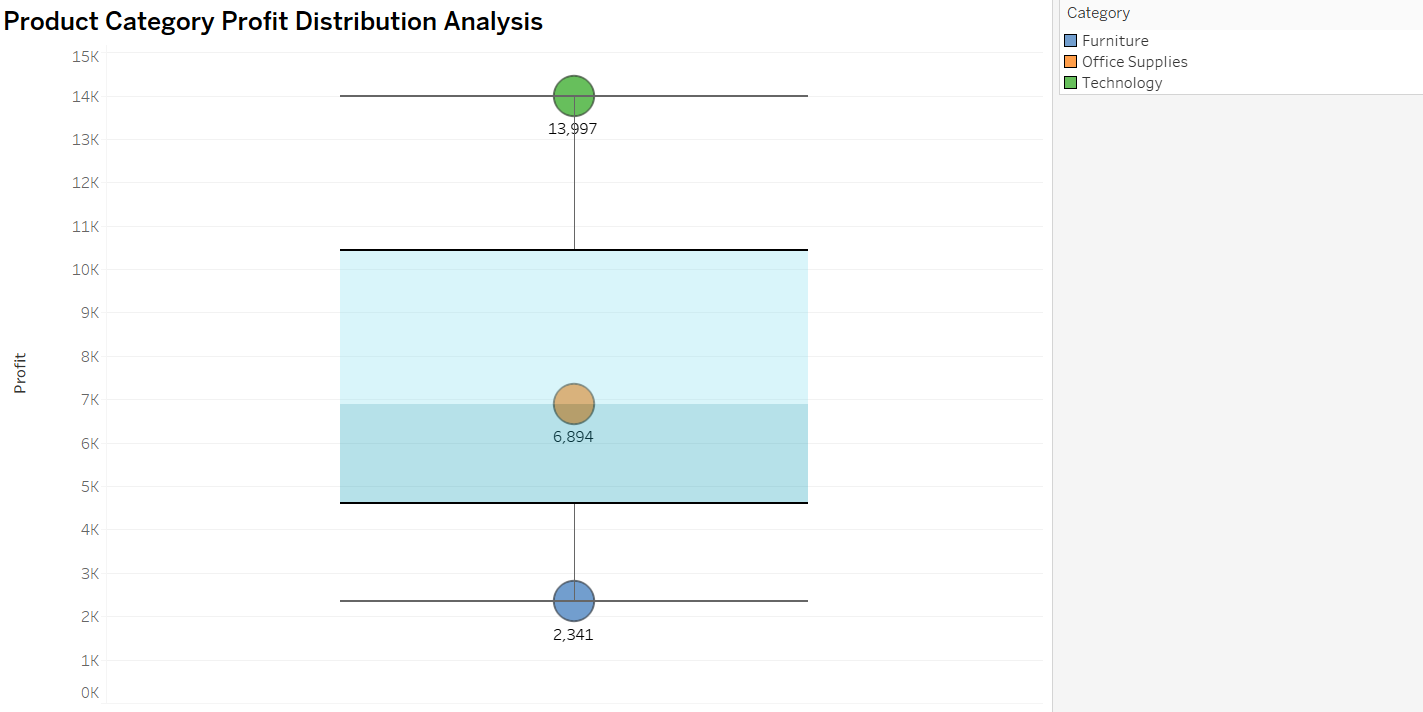
* **Technology Category**: The highest order count falls within the 2-4 range, with a value of 80, while the lowest order count is 2.
* **Furniture Category**: Similarly, the highest order count is in the 2-4 range, with a value of 82, and the lowest order count is 2.
* **Office Supplies Category**: The highest order count is also in the 2-4 range, with a value of 219, and the lowest order count is 2.

This visualization highlights the common trend of order quantities being concentrated in the 2-4 range across all product categories.

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

**Explanation:** To efficiently create an insightful Box Plot in Tableau that illustrates profit distributions across different product categories, utilize Tableau's built-in features to generate box plots for each category. These plots will display key statistics such as median, quartiles, and potential outliers. This visual representation provides valuable insights into the variation and spread of profits within different categories, aiding in the identification of high-performing categories and those that may require further analysis or strategic adjustments.

**Chart Type: Box Plot**



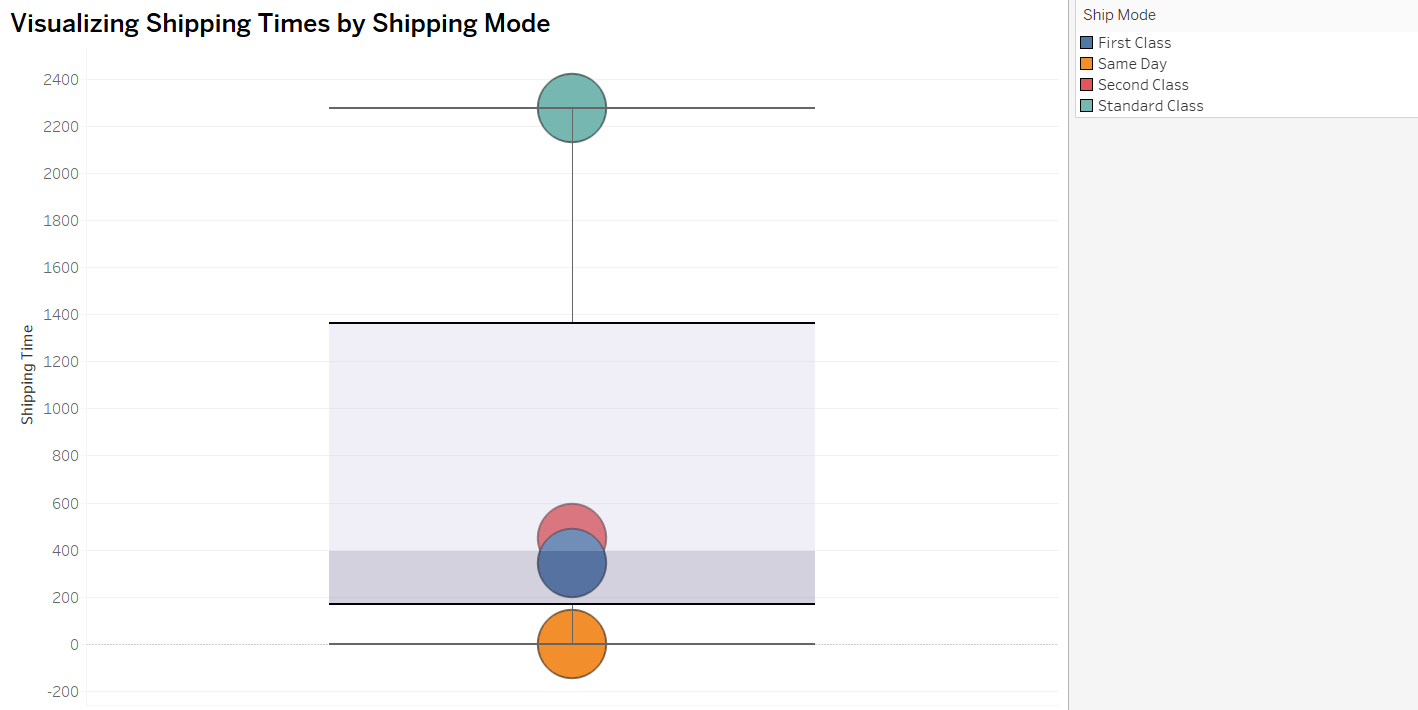
**Description:** This Box Plot serves as a powerful visual tool for illustrating profit distributions across various product categories. Each box within the plot represents a distinct product category, showcasing key statistics such as median profit, quartiles, and potential outliers. It provides valuable insights into the variation and spread of profits within different categories, aiding in the identification of high-performing categories and those that may require further analysis or strategic adjustments.

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

**Explanation:** I have used a Box Plot to visualize the distribution of shipping times across different shipping modes. Each box represents a distinct shipping mode and offers key statistical insights such as median shipping time, quartiles, and potential outliers. This approach helps in identifying variations in shipping times among different modes, facilitating better decision-making and optimization of shipping processes.

**Description:** By comparing the boxes and whiskers for different shipping modes, you can quickly identify which modes have shorter or longer median shipping times. The spread of the boxes shows the variability in shipping times for each mode. Outliers can highlight any unusual shipping times that may need further investigation.

**Chart Type: Box plot**

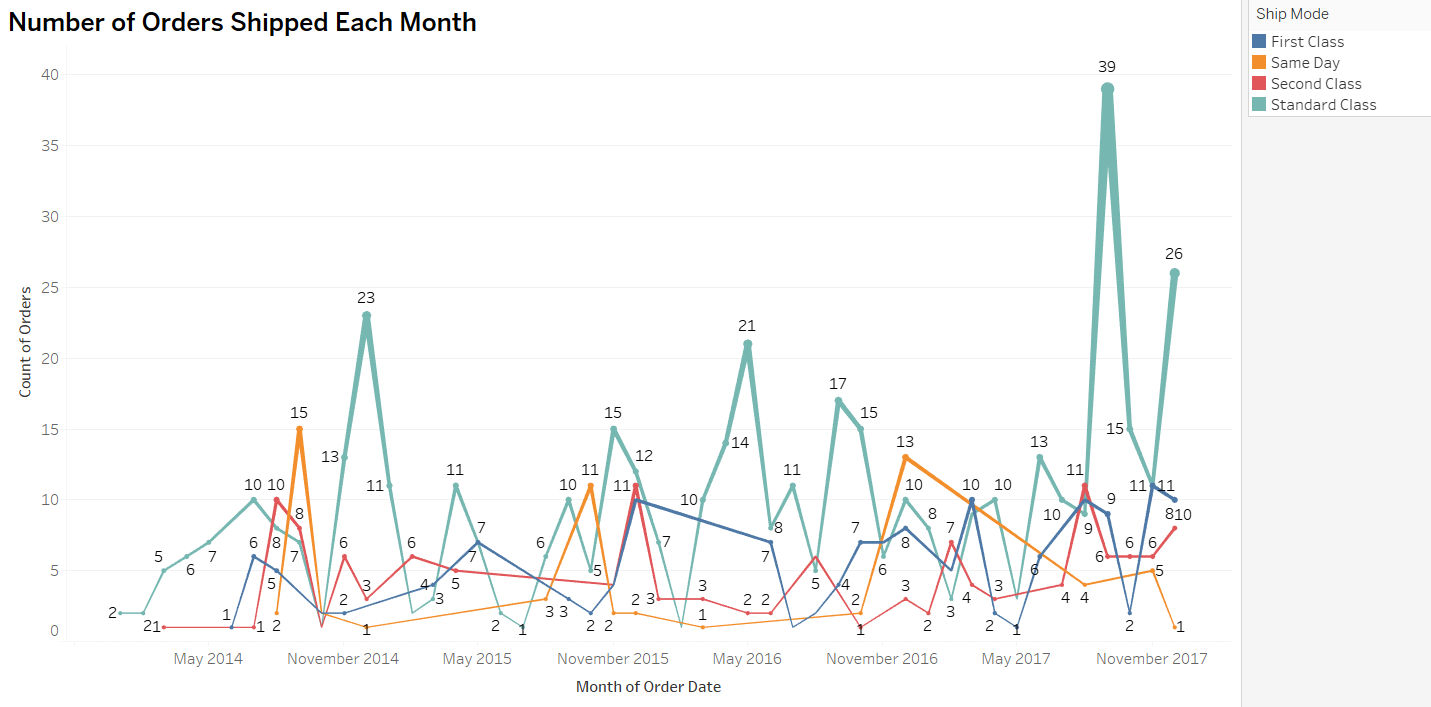


**17.What is the monthly trend in the number of orders shipped?**

**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. This visual narrative allows for a quick understanding of any patterns, seasonality, or trends in the order shipment process over the course of the year.

**Chart Type: Line Chart**



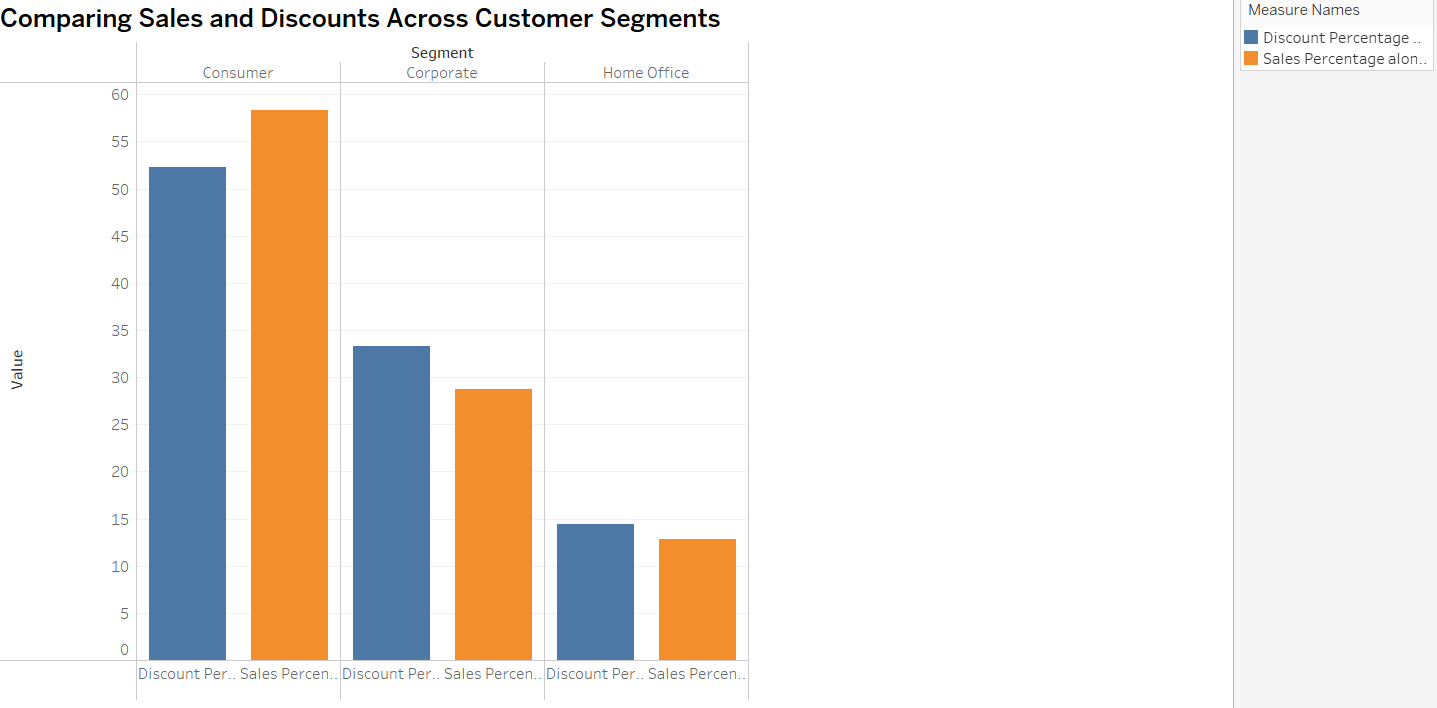
**Description:**

In all categories, February has the lowest number of orders compared to other months. The highest number of orders are observed in September, followed by November. However, there is a substantial decrease in the number of orders in October.

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

**Explanation:** I chose a Side-by-Side Bar Chart because it's effective for visually comparing multiple variables within each Segment, 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 52.27% and the consumer sales value is 58.38%.

The Corporate segment discount value is 33.35% and the Corporate sales value is 28.76%.

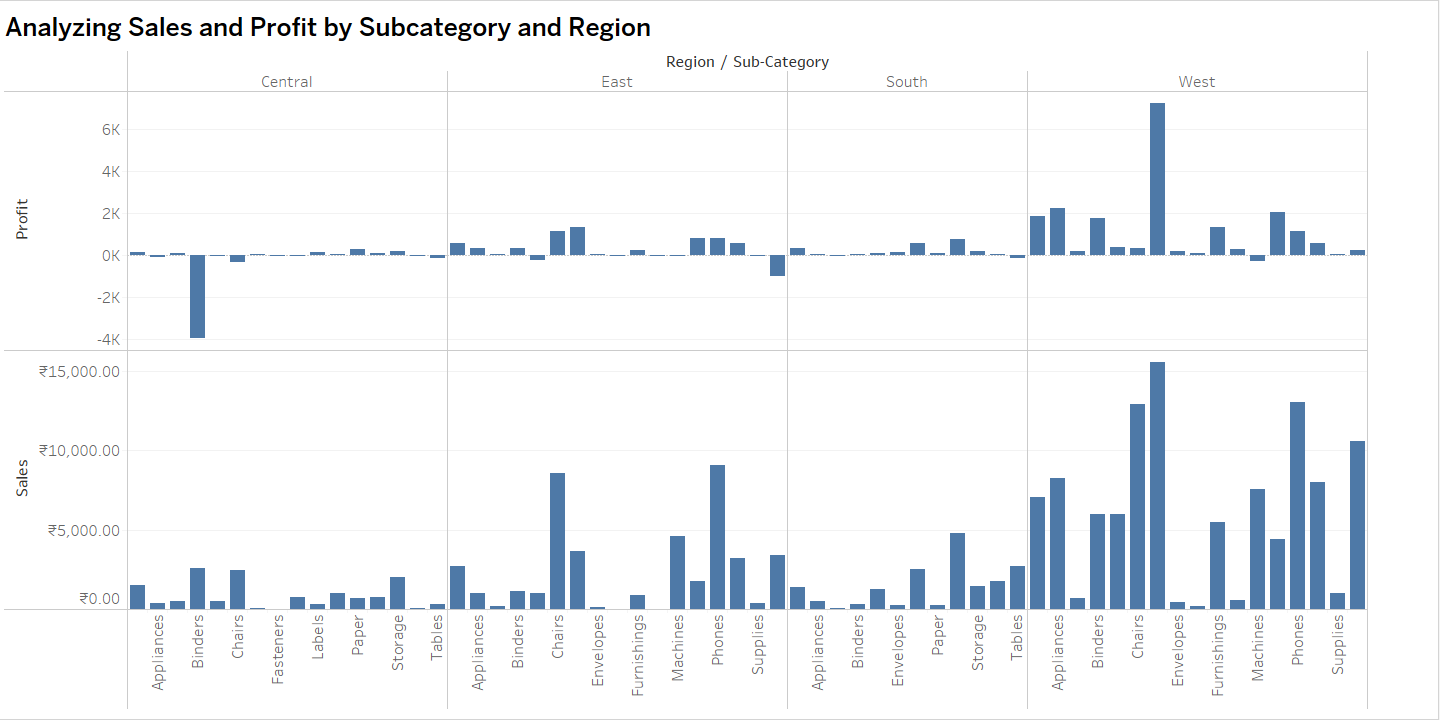
The Home office segment discount value is 14.38% and Home office value is 12.86%.

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 Bar Chart is ideal for visualizing sales and profit across different product subcategories and regions. Filters facilitate selecting the required subcategory, enabling detailed analysis of sales and profit trends for various regions..

**Chart Type: Bar Chart**

****

**Description:** Analyzing sales and profit trends across various product subcategories and regions provides valuable insights into the performance of different segments. This analysis helps in identifying high-performing categories and regions, as well as areas that may require strategic adjustments.

**Sales Trends:**

* **Product Subcategories**: Sales trends can vary significantly across different product subcategories. By examining these trends, you can identify that Copiers subcategories generate the highest sales. High-sales subcategories often indicate popular products that meet customer demand.
* **Regions**: Sales trends across West regions reveal the geographical areas with the highest sales. This helps in understanding regional preferences and tailoring marketing strategies to maximize sales in specific regions.

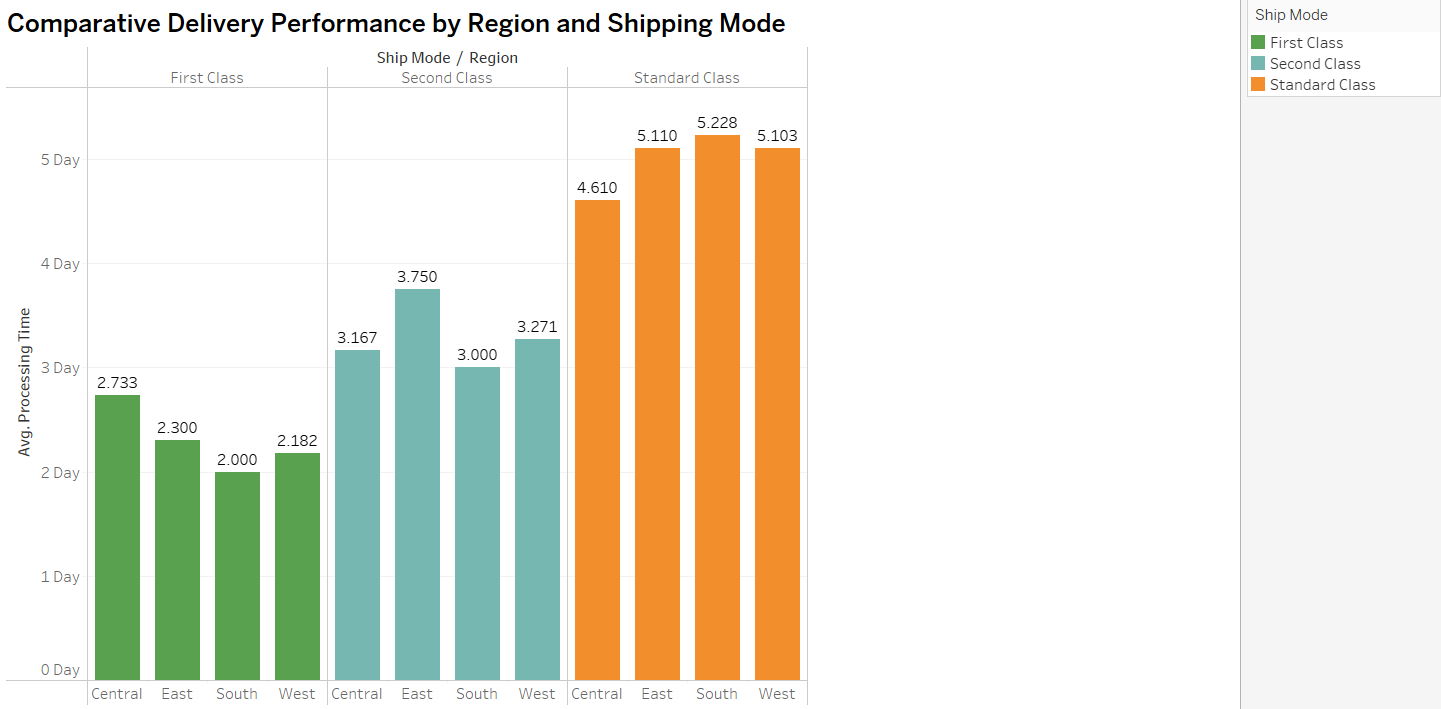
**Profit Trends:**

* **Product Subcategories**: Profit trends across subcategories highlight which products are the most and least profitable. High-profit subcategories (Copiers) contribute significantly to the overall profitability of the business. Conversely, low-profit subcategories (Binders) may require cost optimization or pricing strategies to improve profitability.
* **Regions**: Profit trends across regions indicate the geographical areas that contribute most to the overall profitability. Analyzing regional profit trends helps in identifying lucrative markets and areas that may need focused efforts to enhance profitability.

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

**Explanation: Side by Side Bar Chart** provides a clear and effective way to compare different regions and shipping modes simultaneously. This visualization offers a quick and detailed analysis of the time duration taken by the delivery process, enabling easy identification of trends and performance across various segments.

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

****

**Description:**

**Average Delivery Duratin Analysis for Shipping Modes**:

* **First Class**: The Central region has the longest average delivery duration, followed by the East and South regions. The West region is the quickest in terms of delivery time.
* **Same Day Delivery**: All regions fall under the same day delivery category, so I removed this class from the analysis.

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

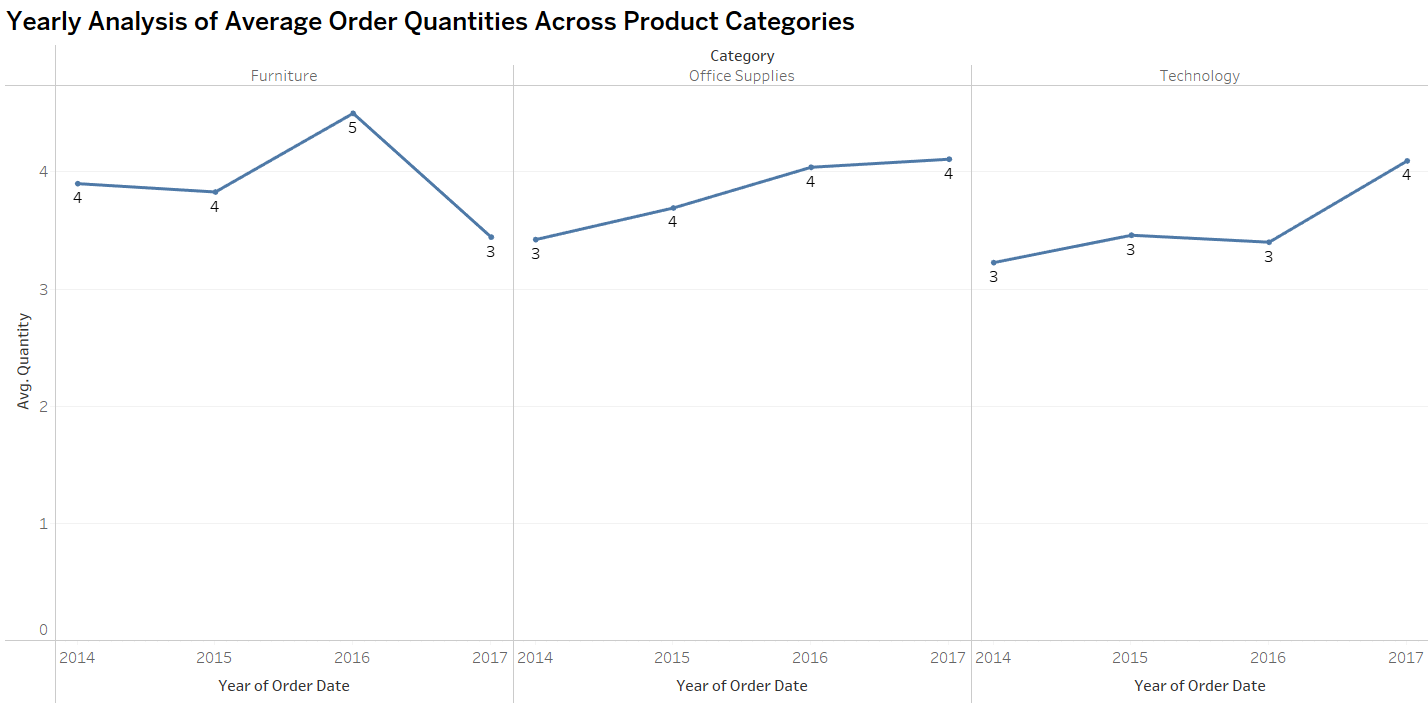
**Explanation:** I chose a Line Chart to illustrate the changes in average order quantity over the years for various product categories. A line chart effectively communicates trends over time, providing a clear visualization of how average order quantities change annually. This approach allows for easy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

**Description:** The average order quantity has changed over the years:

* **Furniture**: In 2017, the average order quantity decreased to 3 compared to previous years.
* **Office Supplies**: In 2017, the average order quantity remained constant at 4, similar to previous years.
* **Technology**: In 2017, the average order quantity slightly increased to 4 compared to previous years.

.

**Chart Type: Line Chart**

****

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

**Explanation: A bubble chart** is a fantastic way to visualize the correlation between discount rates and order quantities for different customer segments, as it adds a third dimension to your analysis

**Description:**

**Positive Correlation**: Larger bubbles at higher discount rates suggest that customers are more likely to order larger quantities when discounts are higher.

**Negative Correlation**: Smaller bubbles at higher discount rates might indicate that higher discounts do not necessarily lead to larger orders for some segments.

**No Correlation**: Uniform bubble sizes regardless of discount rates suggest that other factors may be influencing order quantities.

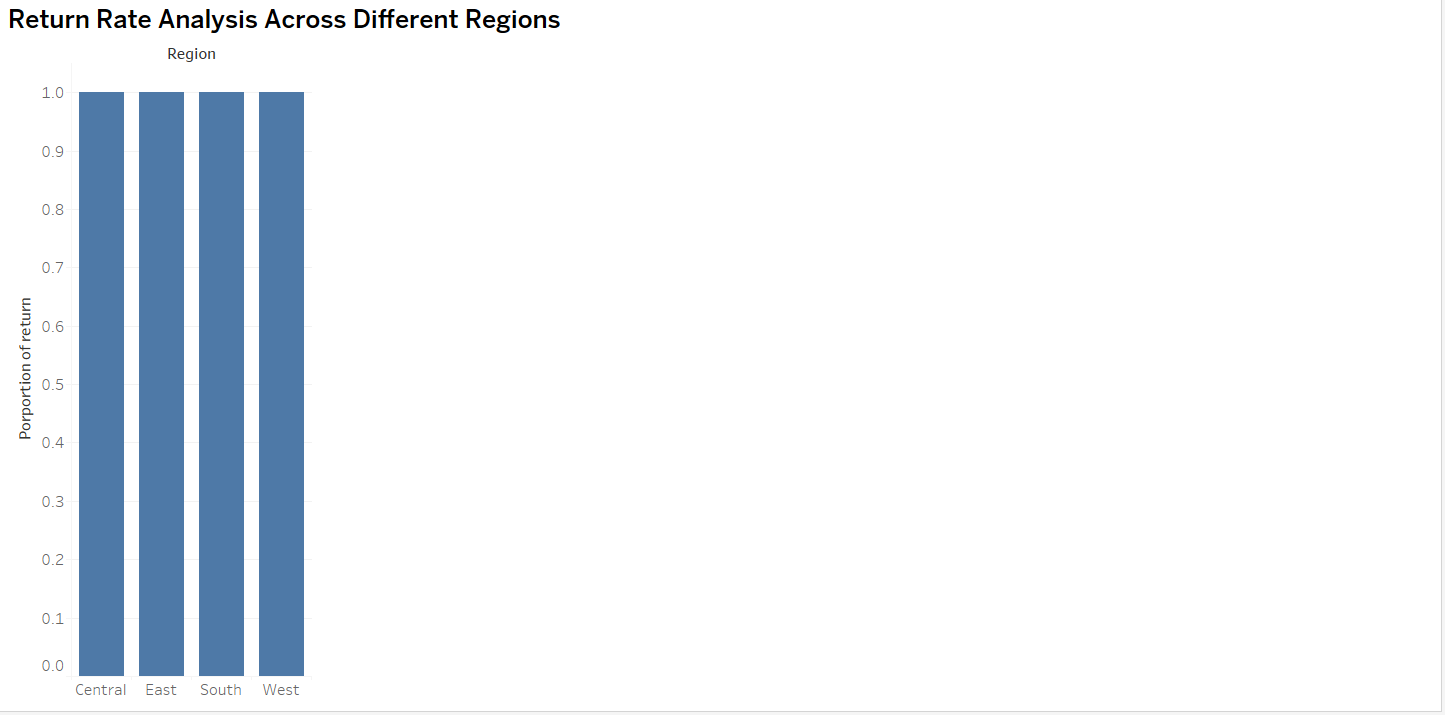
**Chart Type: Bubble Chart**



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

**Explanation:** I have chosen the Bar Chart to visualize the proportion of orders returned for each region, as it offers quick and easy comprehension of return rates. This chart effectively highlights the number of orders returned, allowing for straightforward comparison across different regions.

**Chart Type: Bar Chart**

****

**Description:** The return rates are consistent within each region:

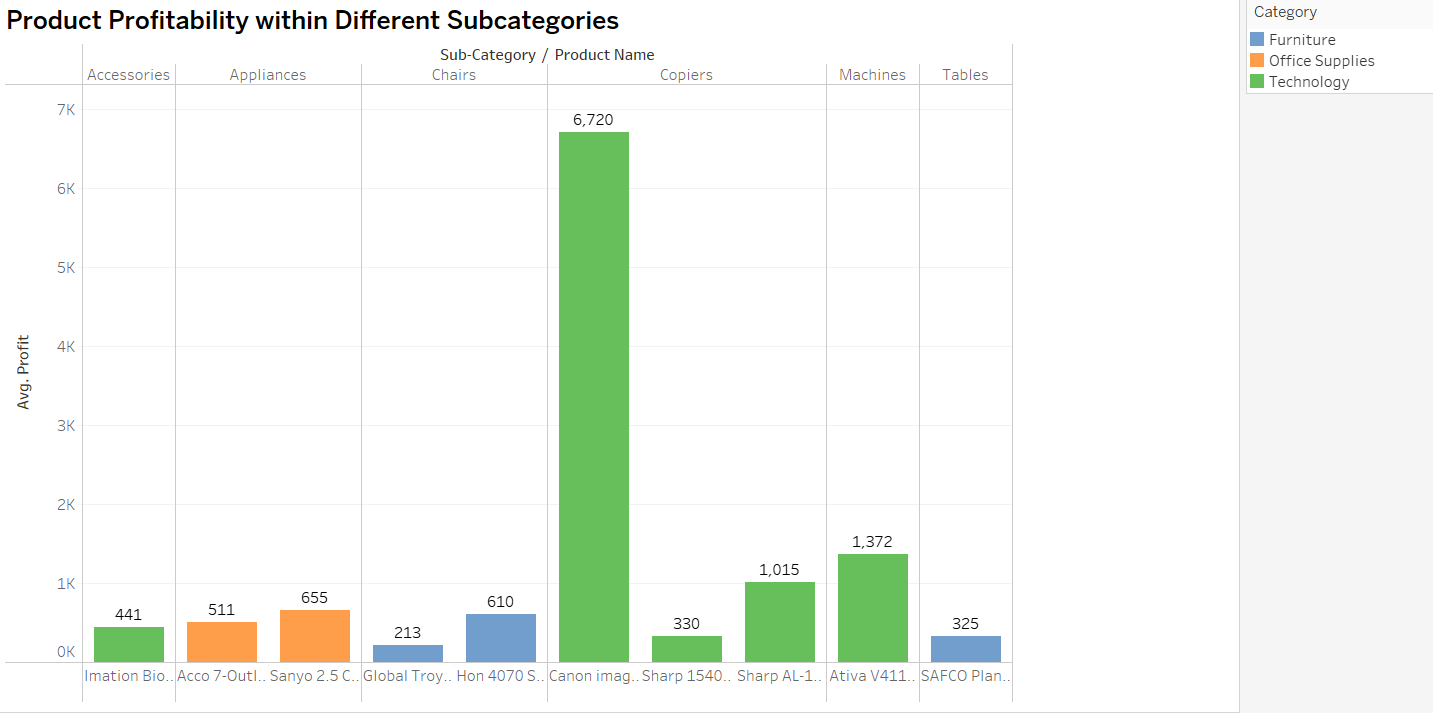
* The West and South regions both have the same return rates.
* The East region also has a consistent return rate.
* Similarly, the Central region has a uniform return rate.

.

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

**Explanation:** I have used a Bar Chart to compare profits for different products and subcategories. This approach provides an easy and detailed comparison of profits, supported by color toning to clearly differentiate the profit-making products. The bar chart effectively highlights the profitability of each product within its respective subcategory and category, allowing for straightforward analysis and insights.

**Chart Type: Bar Chart**

****

**Description:**

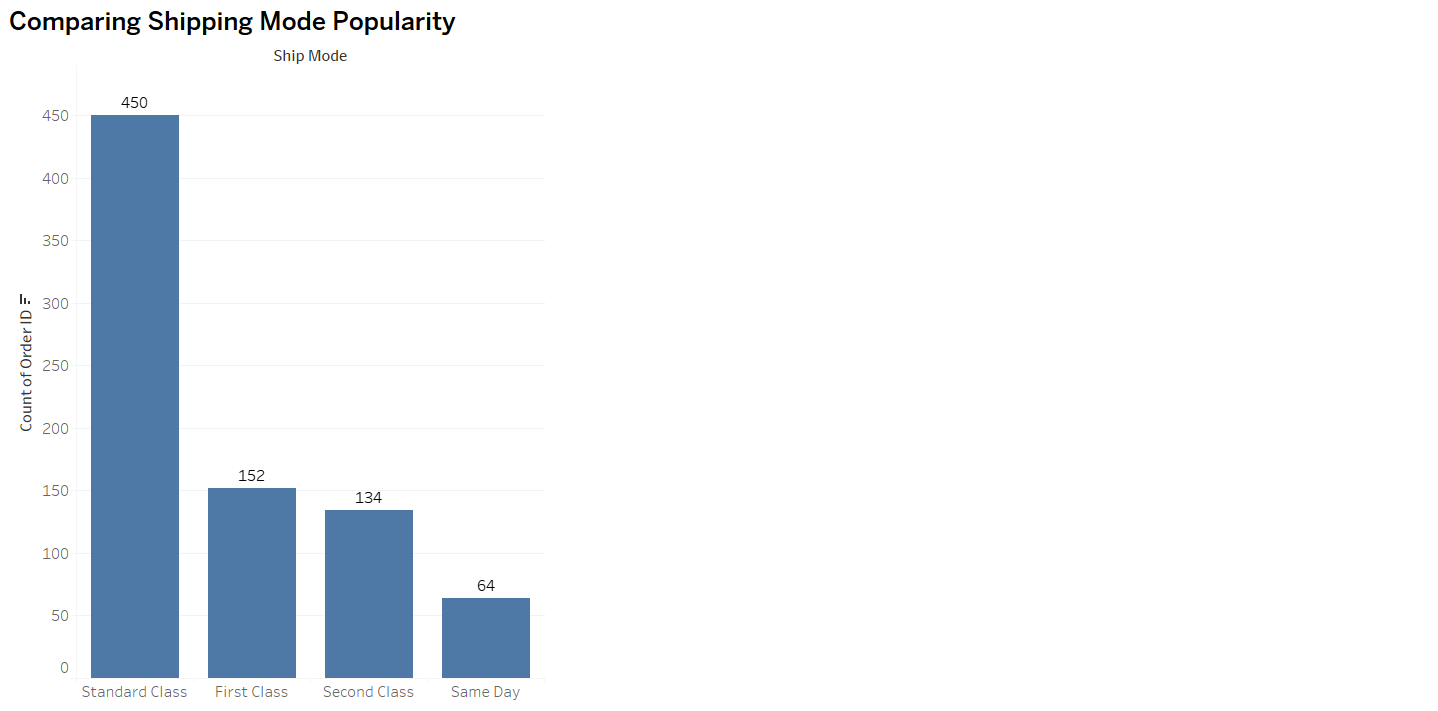
* **High Profit Products**: In Copiers subcategory that Canon Product generate the highest profit, signifying strong market demand or successful pricing strategies.
* **Low Profit Products**: In Chairs subcategory that Global Troy Product generate the lowest profit, which may need cost optimization or marketing efforts to boost their performance.

By creating these visualizations, you can gain valuable insights into the profitability of different products across various subcategories, helping you optimize your product offerings and strategic decision-making.

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

**Explanation:** I have chosen a 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 used shipping mode in the Sample Superstore dataset.

**Chart Type: Bar Chart**

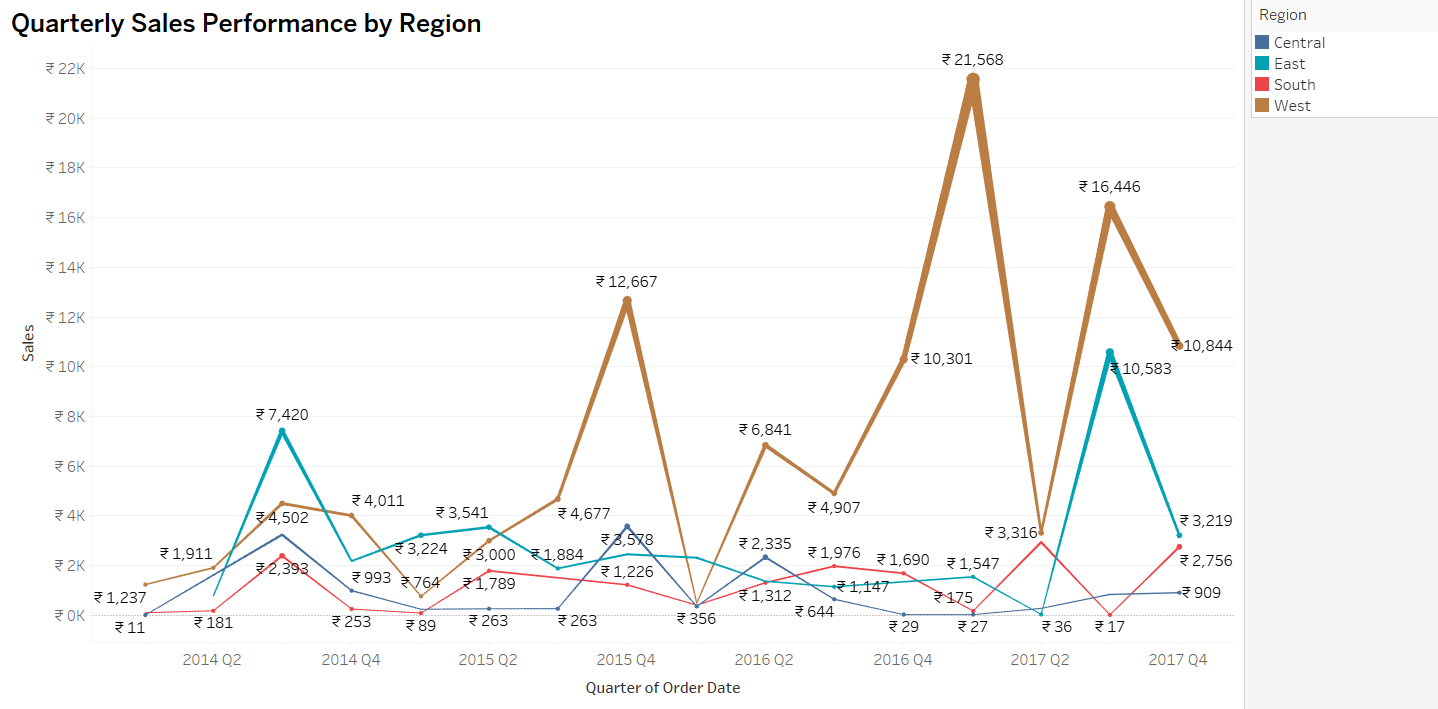
****

**Description:** After analyzing the chart, we found that **Standard Class** is the most used shipping mode, ranking first. This mode has the highest number of orders shipped, making it the preferred choice compared to other shipping modes.

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

**Explanation:** I have chosen 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. This chart visually represents the fluctuations in sales for different regions over the quarters, allowing for easy identification of 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 west region in Q1 1,237 it evolves as highest in Q2 with the value of 21568.

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

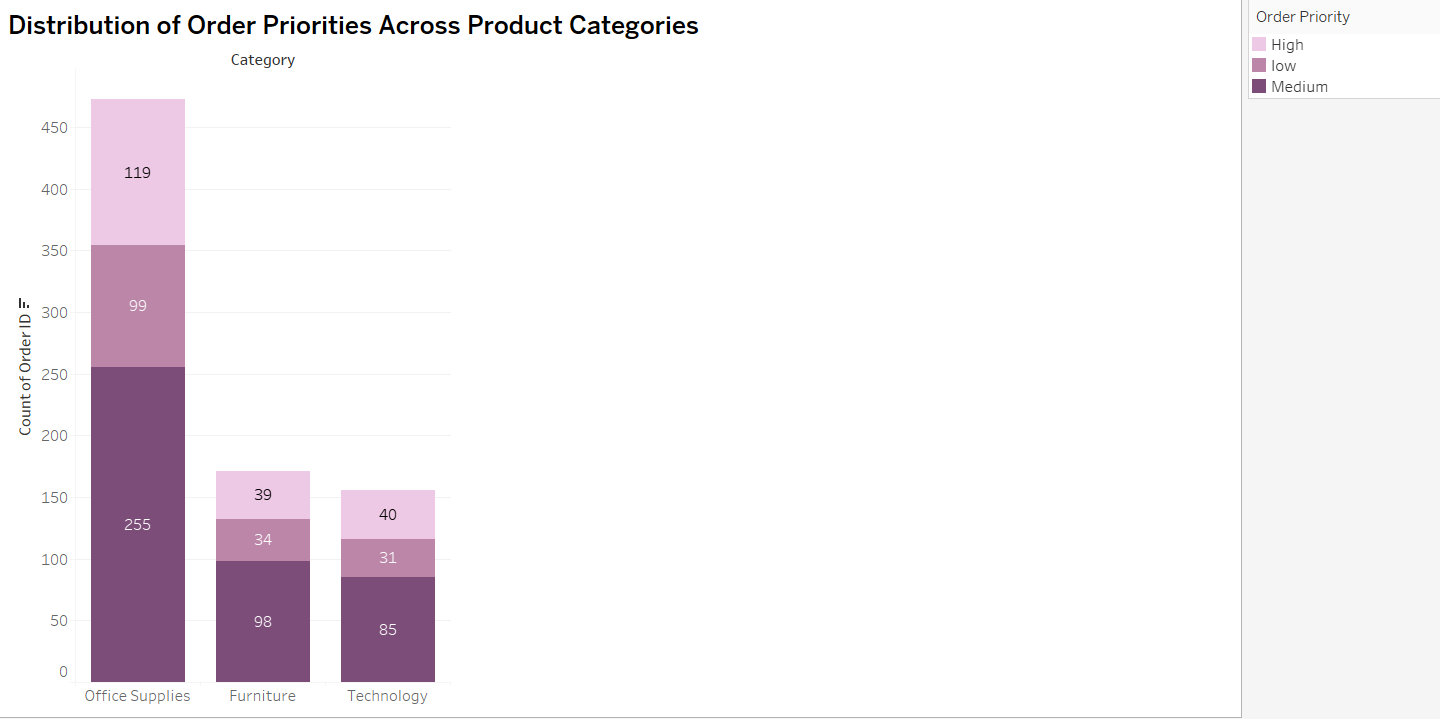
**27.What is the distribution of order priorities across different product categories?**

**Explanation:** I have chosen a Stacked Bar Chart to visualize the distribution of order priorities across different product categories. A stacked 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. This visualization provides a clear comparison of how order priorities are allocated within each product category, highlighting the differences and similarities across categories.

**Description:**

* **Furniture Category**: In this category, medium priority orders are the highest, followed by high priority orders.
* **Office Supplies Category**: Medium priority orders dominate in this category.
* **Technology Category**: Medium priority orders are the highest, followed by high priority and low priority orders.

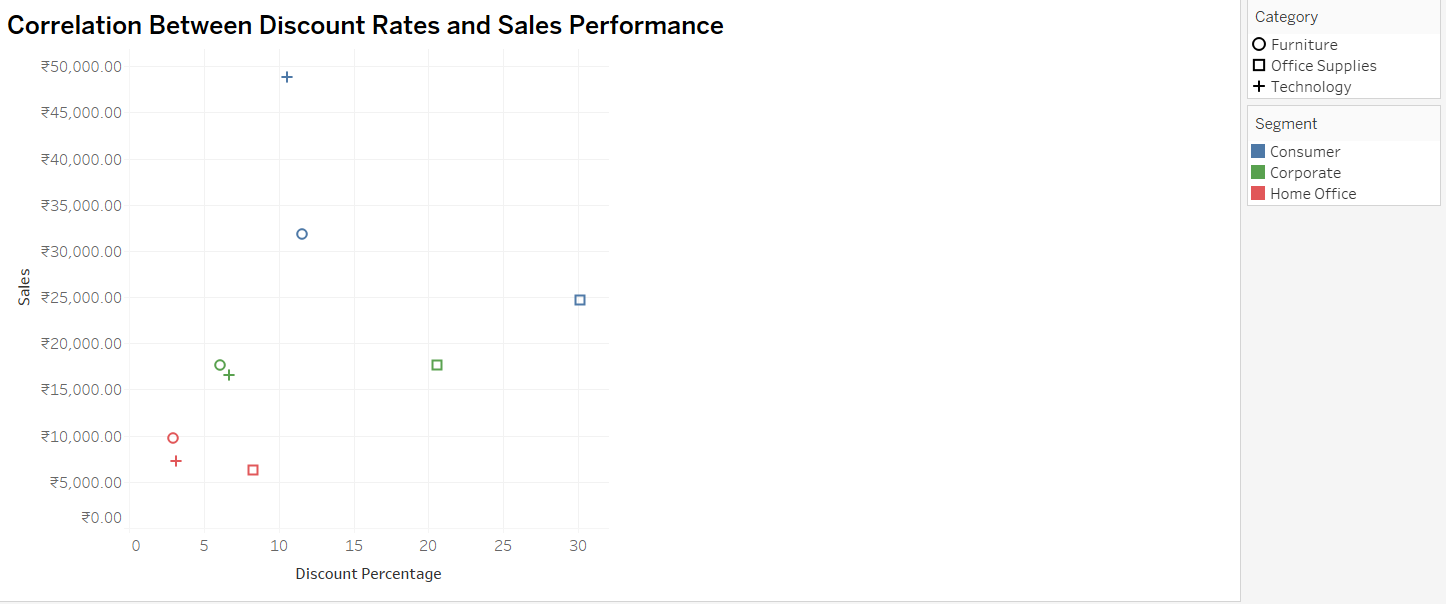
**Chart Type: Stacked Bar Chart**

****

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

**Explanation: Scatter Plot** is the most appropriate chart for showing the relationship between discounts and sales. It effectively showcases whether there is a positive, negative, or no correlation between discounts and sales. Scatter plots are particularly useful for identifying patterns and trends. For instance, if increasing discounts lead to decreasing sales, a scatter plot will clearly depict this relationship. Additionally, scatter plots help in detecting outliers, which can provide further insights into the data.

**Chart Type: Scatter Plot**

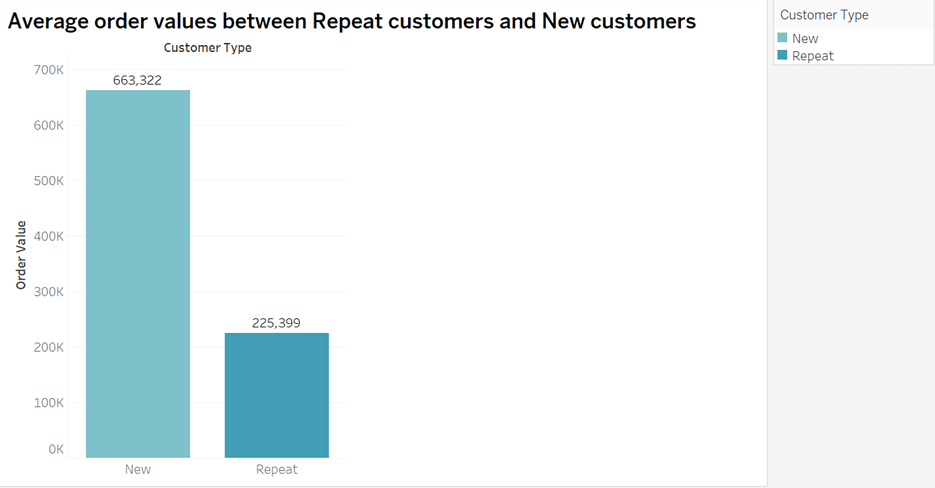
****

**Description:** The relationship between discounts and sales can be summarized as follows: When the discount percentage decreases, the sales profit tends to increase. Conversely, when the discount percentage increases, the sales profit typically decreases. Although certain outliers have been detected, the overall trend shows that higher discounts negatively impact overall profit.

**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:**

This bar chart serves as a valuable visual tool for comparing and contrasting the average order values between two distinct customer types: "Repeat" and "New." The chart is structured with two bars, each representing one of these customer categories. The horizontal axis categorizes customers into "Repeat" and "New," allowing for a straightforward differentiation. On the vertical axis, the bars rise to depict the average order values associated with each customer type.

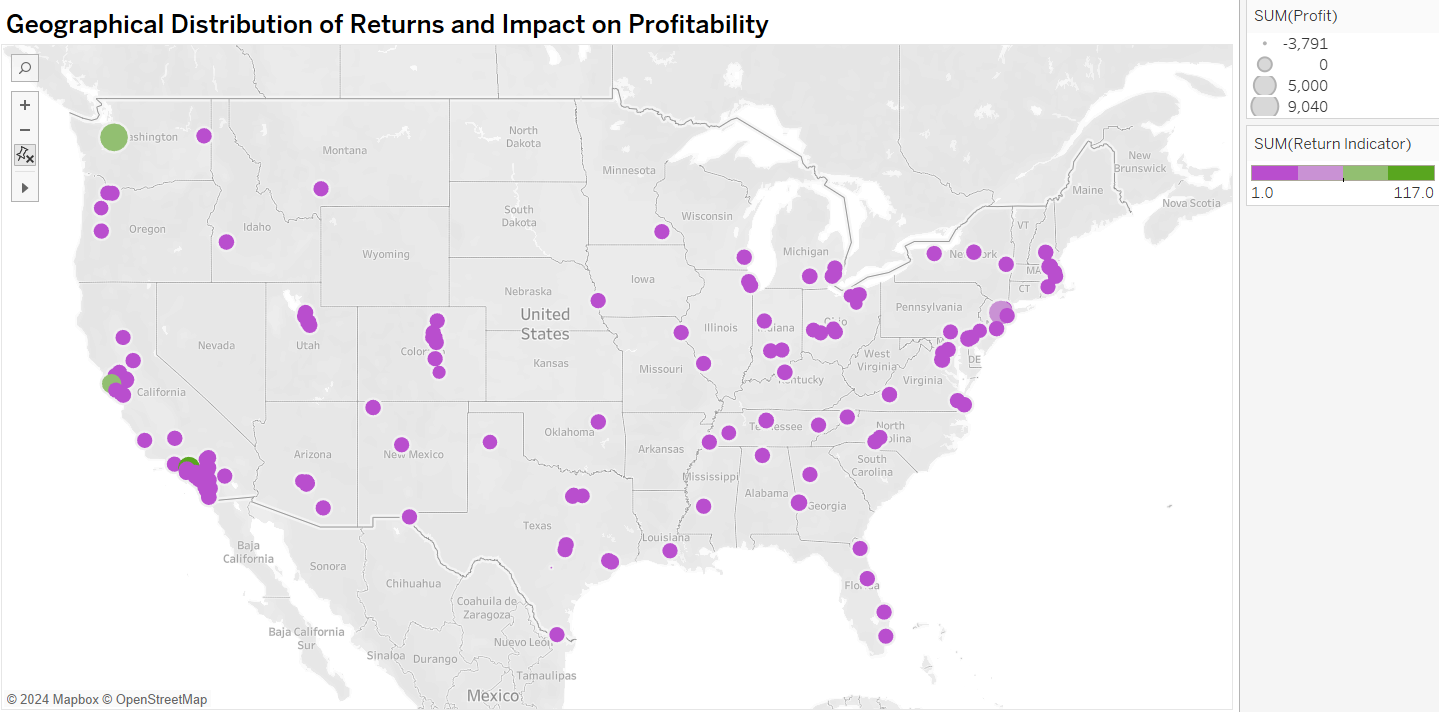
The bar chart provides an immediate and clear comparison, highlighting any disparities or similarities in the average order values between repeat and new customers.

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

**Explanation:** I have chosen 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 affects overall profitability in various regions of the United States. For instance, California shows a significant profit impact due to returns, while Texas experiences a negative impact on profitability. The cost of handling returns, including shipping and processing, affects sales across different regions and product categories. This distribution highlights how returns influence profitability differently across various geographical areas.