Mid Course Summative Assessment Data Visualization Tools

Babita Pruseth,
Cohort Moscow, AlmaBetter

Guidelines for Data Visualization and Analysis Project

About the Project:

In this project, you will be working with a dataset from the Superstore, aiming to answer 30 scenario-based questions through data visualization and analysis. Your objective is to select the best chart for each question, explain your choice. This project will showcase your proficiency in data visualization, critical thinking, and effective communication.

Skills Required:

- *Proficiency in data visualization concepts and techniques.*
- *Familiarity with Tableau or a similar data visualization tool.*
- *Strong analytical and problem-solving skills.*
- *Ability to choose appropriate charts based on data characteristics and question requirements.*
- *Clear and concise communication skills.*

Deliverables:

- *A Google document containing solutions to the scenario based questions including the screenshot of relevant chart picked for each scenario, presented in a concise and well-structured format. Make sure to provide explanations that highlight your problem-solving skills.*

Rubrics for Assessment:

Question Responses:

- *Accuracy and completeness of answers for all 30 questions.*
- *Clear and concise explanations that address the question's context.*

Chart Selection and Explanation:

- *Thoughtful rationale for choosing specific chart types.*
- *Justification based on data characteristics, context, and communication goals.*

Creative Enhancements:

- *Effective use of creative elements to enhance visualization quality.*
- *Enhancements that contribute to better understanding or engagement.*

Note:

- *Duplicate this document and proceed to write your solutions.*
- *For each scenario and question, provide a justification for the choice of chart type. Explain why it is the best option to visualize the data effectively.*
- *Attach screenshots of the charts you have created in Tableau for each scenario and*

question using the Superstore dataset. Label them clearly to match the corresponding questions in the Google Document.
● Submit the duplicated google doc file after completion.

Use these guidelines to structure your data visualization and analysis project. Remember to maintain consistency in your responses, explanations, and visualization styles. This project will not only demonstrate your skills but also your ability to effectively communicate complex information through visualizations. Good luck!

Problem Statement: Choose the Best chart for any 30 scenario based questions from Superstore Dataset.Imagine you are a data enthusiast aiming to excel in data visualization and analysis. In this task, you have been given any 30 scenario-based questions derived from the Superstore dataset, and your objective is to provide insightful answers using appropriate charts. For each question, you need to select a chart that best represents the data, explain why you chose that specific chart, and then proceed to build the chosen chart using Tableau.Your responses should be succinct, organized, and illustrative of your problem-solving capabilities.

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

Please keep in mind:

1. *Answer Completion: Ensure that you furnish answers for all 30 questions and build charts for them.*
2. *Encouraged Creativity: Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.*

By completing this task effectively, you'll not only demonstrate your proficiency in data visualization and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

Good luck!

Questions:

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

Aim:

To compare the total sales across product categories to determine which category has the highest sales in the "Superstore" dataset.

Chart Chosen:

A bar plot is selected for easy visualization of sales comparisons across categories. This format provides a clear, straightforward way to identify top-performing categories and make data-driven decisions.

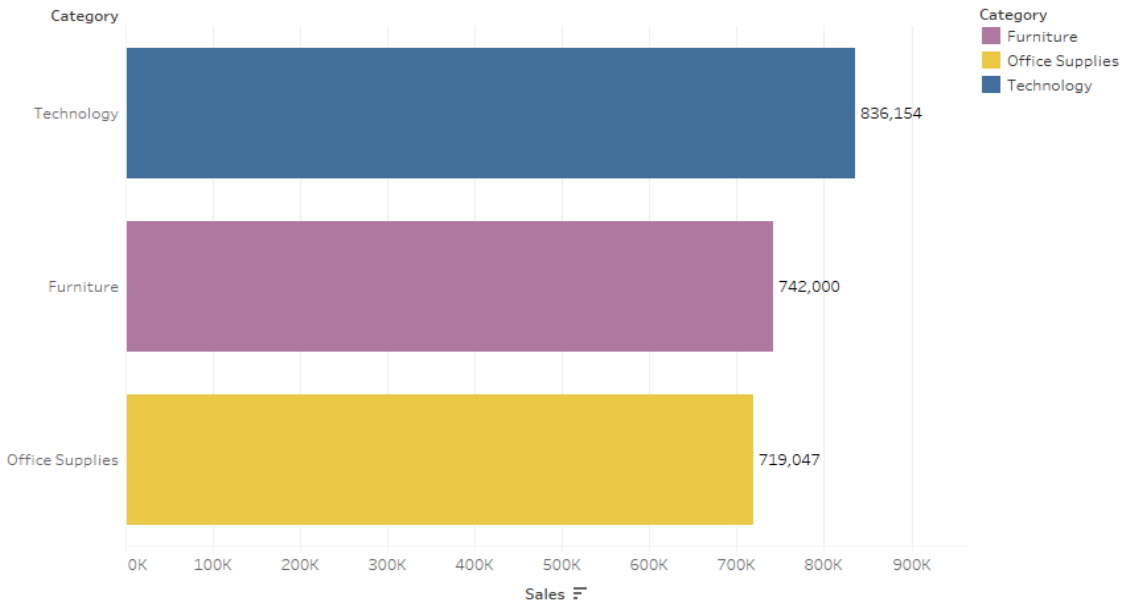
Outcome:

The bar plot illustrates that among the three product categories:

- Technology is the highest-selling category with total sales of 839,154.
- Furniture comes second with sales totaling 742,000.
- Office Supplies ranks third, generating 719,047 in total sales.

This analysis suggests that Technology should be a key focus for marketing and inventory strategies, followed by Furniture and Office Supplies.

Ans1



Sum of Sales for each Category. Color shows details about Category. The marks are labeled by sum of Sales. The view is filtered on Category, which keeps Furniture, Office Supplies and Technology.

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

Aim:

To analyze how sales change on a monthly basis over the course of each year, using the "Superstore" dataset.

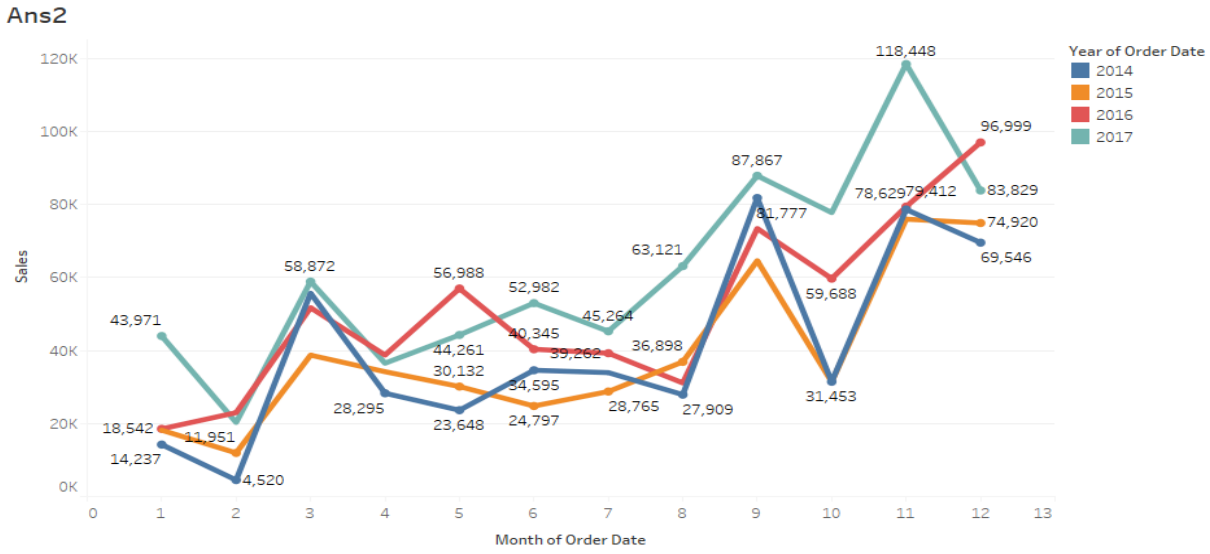
Chart Chosen:

A line plot is chosen to effectively show the trend of monthly sales across the years. The continuous nature of a line plot makes it easy to observe fluctuations and identify seasonal patterns or peak sales periods.

Outcome:

- 2014: Sales started at 14,237 in January, with the highest sales in September at 81,777.
- 2015: January sales were 18,174, and sales peaked in November at 75,973.
- 2016: Sales in January were 18,542, and the maximum occurred in December with sales reaching 96,999.
- 2017: January saw sales of 43,971, and the highest sales occurred in November with a total of 118,448.

Sales show a general increase towards the end of each year, with significant peaks in November and December, likely due to holiday shopping periods. This insight suggests that promotional campaigns and stock availability should be maximized towards the final months of each year.



The trend of sum of Sales for Order Date Month. Color shows details about Order Date Year. The marks are labeled by sum of Sales. The data is filtered on Order Date Month, which keeps 12 of 12 members.

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

Aim:

To understand how total sales are distributed across different product categories in the "Superstore" dataset.

Chart Chosen:

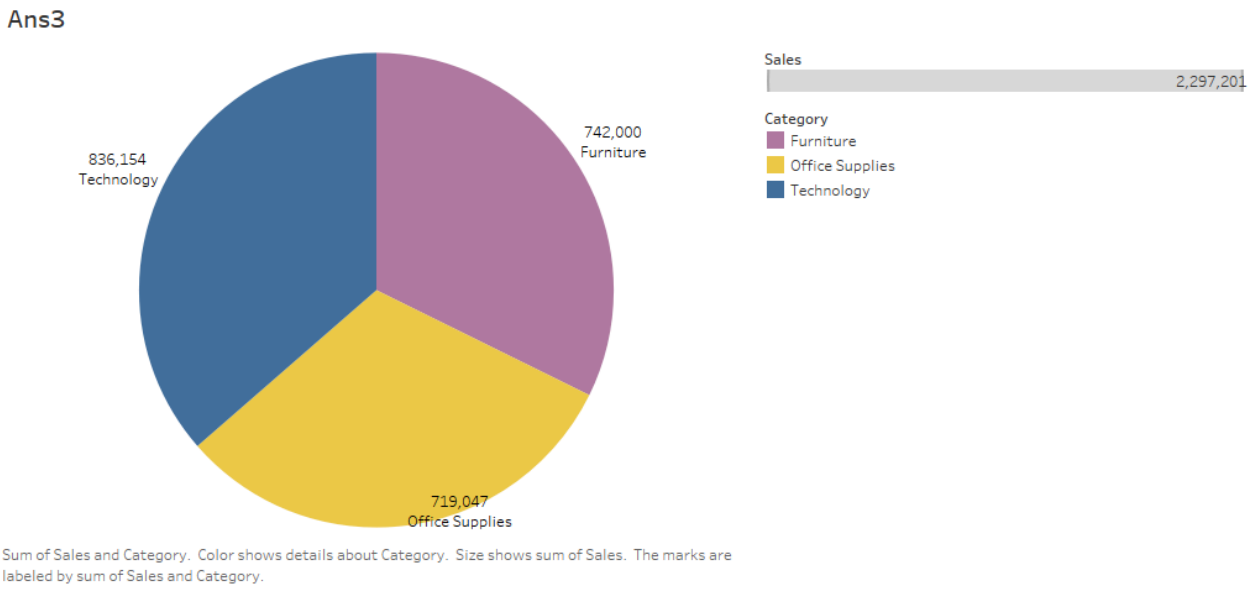
A pie chart is selected to visually represent the proportion of total sales contributed by each product category. This type of chart is ideal for showing how different categories compare in terms of their share of total sales.

Outcome:

The pie chart illustrates the distribution of sales among the product categories:

- Technology contributes the highest sales with a total of 836,154.
- Furniture follows with 742,000 in sales.
- Office Supplies has the lowest sales at 719,047.

This visualization clearly shows that Technology accounts for the largest portion of total sales, making it a key category for investment and resource allocation. Office Supplies, being the smallest contributor, may require targeted strategies to boost its sales. This distribution insight can help stakeholders make informed decisions regarding inventory management, marketing focus, and resource allocation.



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

Aim:

To analyze the sales performance of individual customers over time.

Chart Chosen:

A line chart with multiple lines, but only showing four customer IDs for simplicity due to the large dataset size.

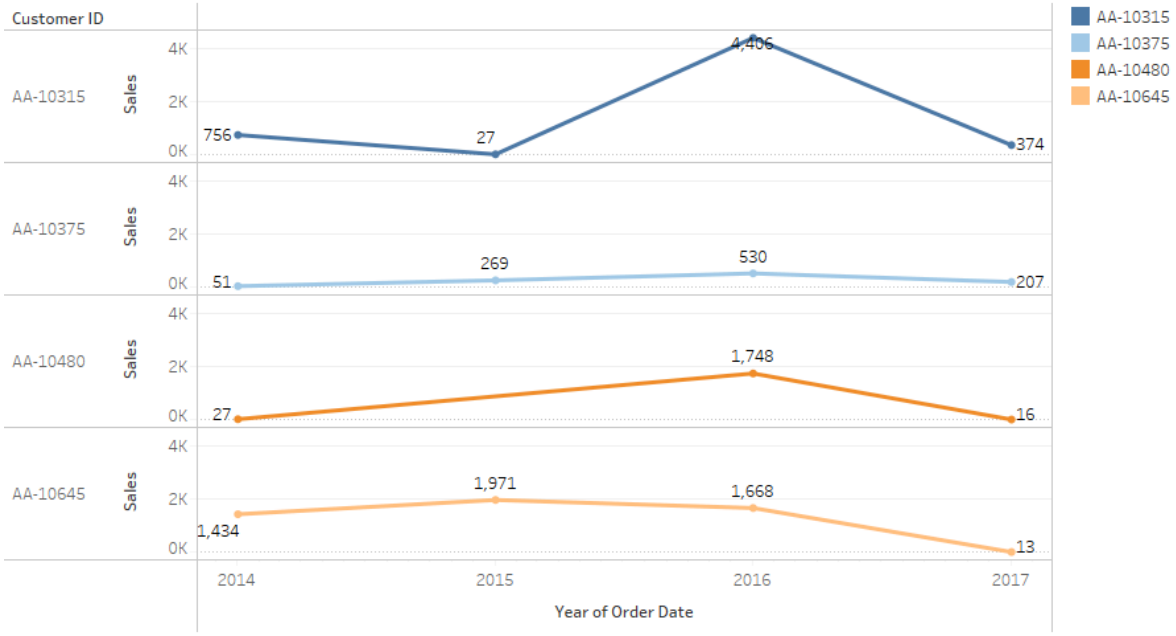
Outcome:

Although only four customer trends are shown in the chart for ease of display, the overall analysis provides insights into:

- Tracking sales trends for all customers over time.
- Identifying key customers with consistent or high sales performance.
- Detecting changes in purchasing behavior to adjust business strategies.

The method can be extended to the full dataset for a more comprehensive analysis of all customers.

Ans4



The trend of sum of Sales for Order Date Year broken down by Customer ID. Color shows details about Customer ID. The marks are labeled by sum of Sales. The view is filtered on Customer ID, which keeps AA-10315, AA-10375, AA-10480 and AA-10645.

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

Aim:

To analyze the sales performance of different product categories across days of the week.

Chart Chosen:

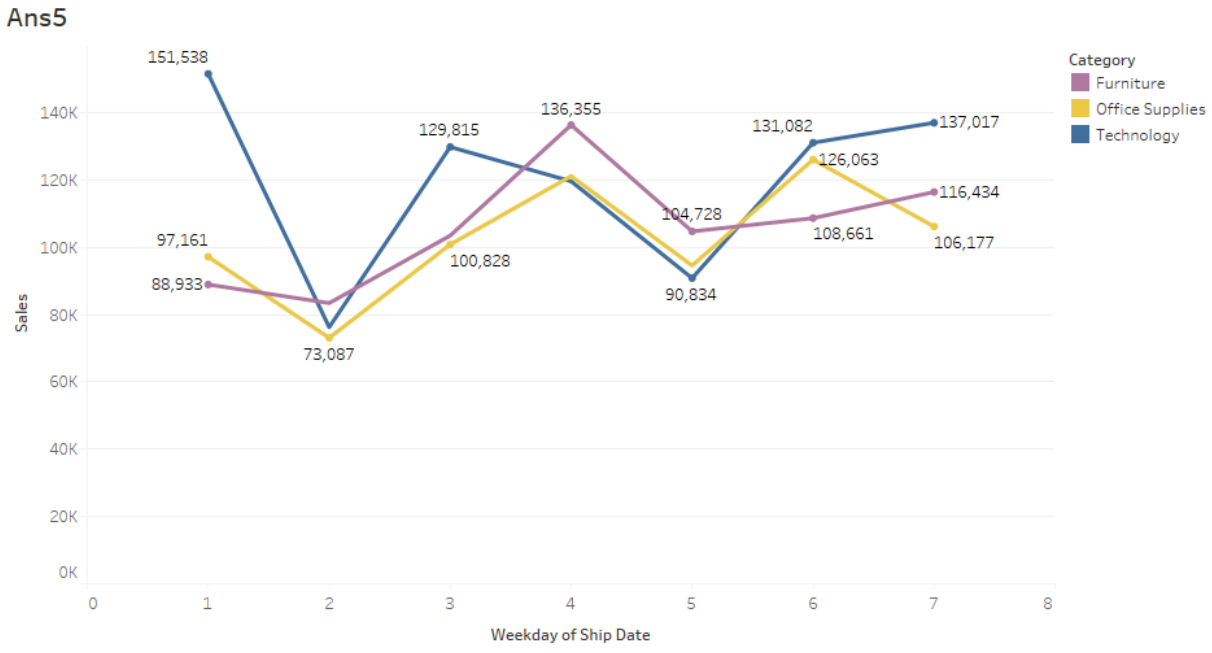
A line chart is chosen to track the sales trends of product categories across each day of the week.

Outcome:

The line chart reveals:

- Technology has the highest sales on Sunday, with a total of 151,538.
- Office Supplies sees peak sales on Friday, totaling 126,063.
- Furniture records its highest sales on Wednesday, at 136,355.

This insight helps in identifying the best-performing days for each category, which can guide marketing and inventory strategies.



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

Aim:

To visualize and analyze the sales growth of different product categories over time.

Chart Chosen:

A line chart is chosen to effectively represent the sales growth of various product categories over time. The data is segmented into subcategories for a more detailed analysis.

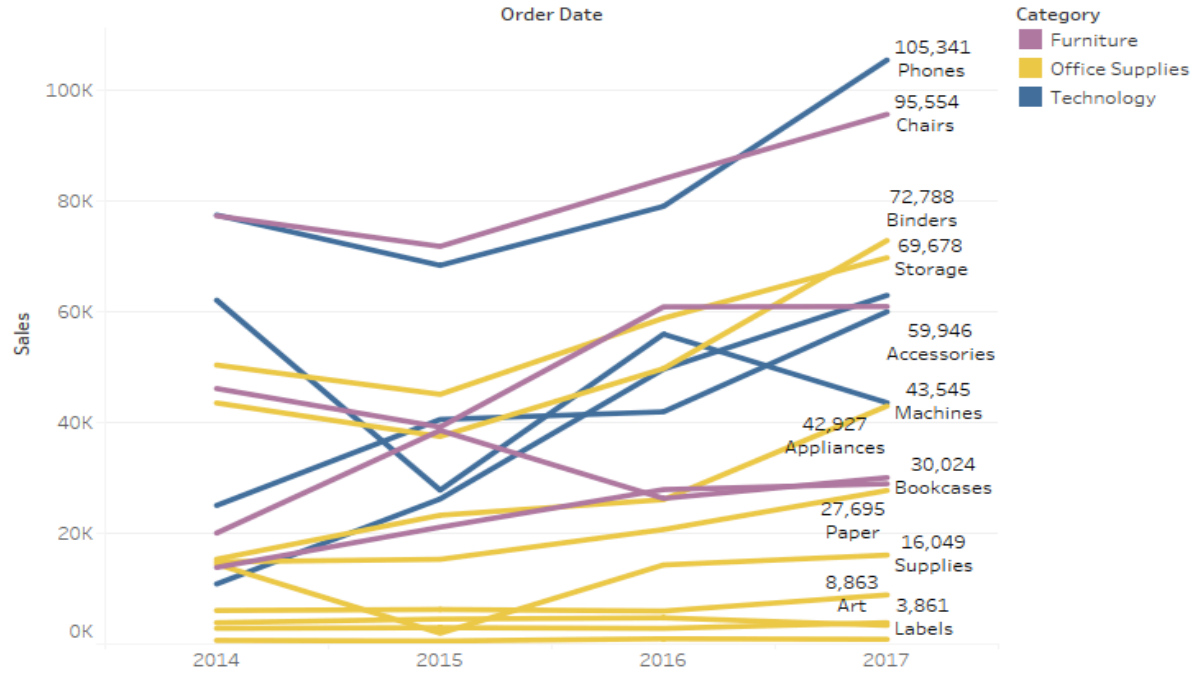
Outcome:

The line chart shows:

- Fasteners (Office Supplies) display the lowest growth.
- Phone (Technology), Chair (Furniture), and Other Stuff exhibit the highest growth rates over time.

This visualization helps track the trajectory of different subcategories, enabling businesses to identify growth trends, plan for inventory, and adjust marketing strategies based on product performance.

Ans6



The trend of sum of Sales for Order Date Year. Color shows details about Category. The marks are labeled by sum of Sales and Sub-Category. Details are shown for Sub-Category.

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

Aim:

To analyze how sales distribution varies across different regions in the "Superstore" dataset.

Chart Chosen:

A Map Chart is chosen to visually represent the sales distribution across regions. This type of chart provides a clear geographical context, making it easy to compare sales performance across different areas.

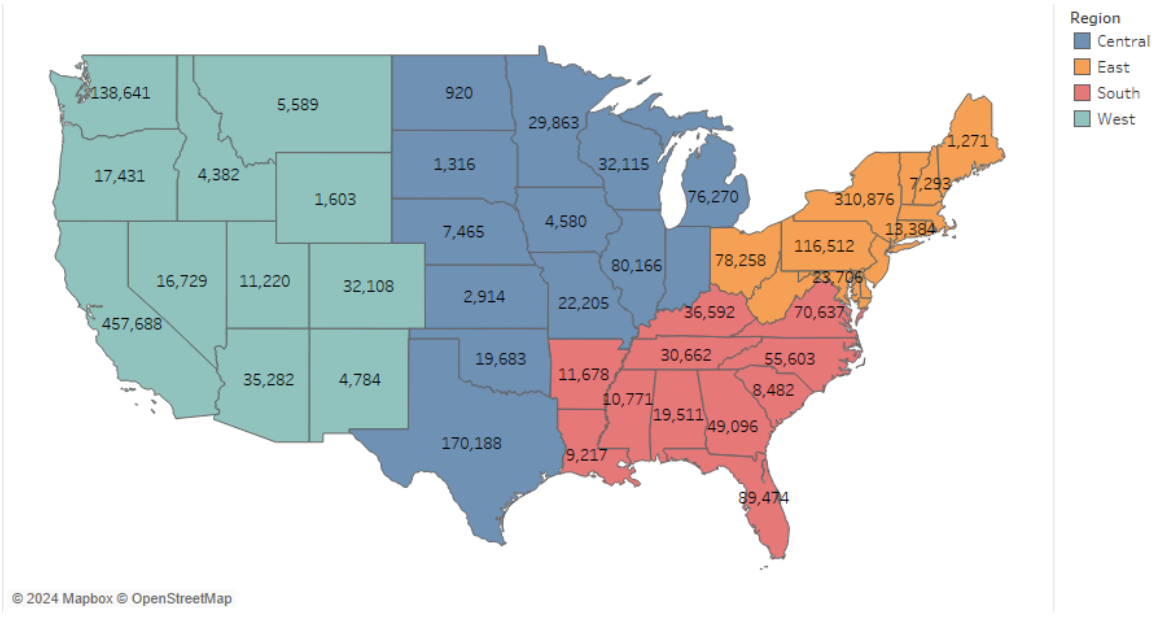
Outcome:

The map chart effectively:

- Highlights regional differences in sales.
- Allows for quick identification of high-performing and low-performing regions.
- Provides an intuitive visual representation of how sales are distributed geographically.

This helps in understanding regional sales trends, optimizing distribution strategies, and targeting marketing efforts based on geographical sales performance.

Ans7



Map based on Longitude (generated) and Latitude (generated). Color shows details about Region. The marks are labeled by sum of Sales. Details are shown for Country and State.

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

Aim:

To visualize the composition of profits across various subcategories within different customer segments.

Chart Chosen:

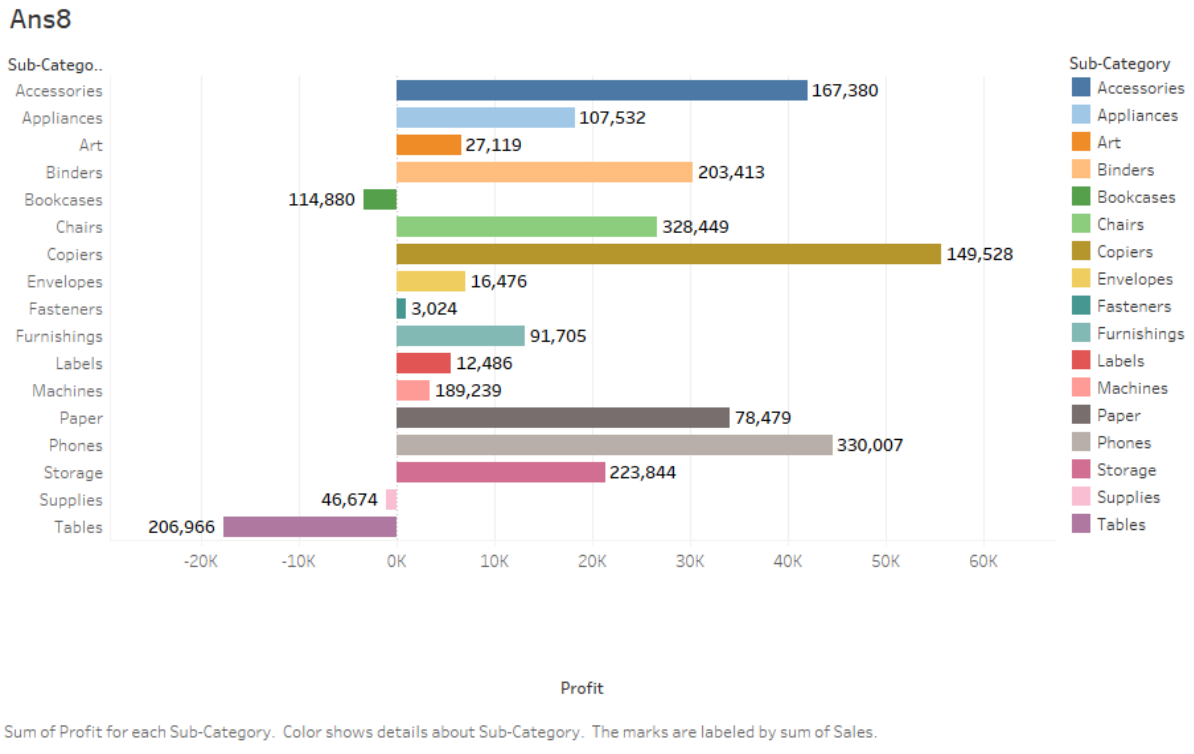
A bar plot is chosen to effectively represent and compare the profits generated by different subcategories within various customer segments. This type of chart allows for a clear understanding of which subcategories yield the most and least profit.

Outcome:

The bar plot reveals:

- Copiers generate the highest profit of $55,618 with total sales of $149,528.
- Bookcases and Suppliers show negative profits of -$3,473 and -$1,189, respectively, with total sales of $114,880 and $46,674.

This visualization helps identify which subcategories are most profitable and which ones are not performing well, guiding strategic decisions in inventory management, marketing, and customer targeting.



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

Aim:

To visualize the percentage contribution of each region to overall sales in the "Superstore" dataset.

Chart Chosen:

A pie chart is selected to represent the percentage contribution of each region to total sales.

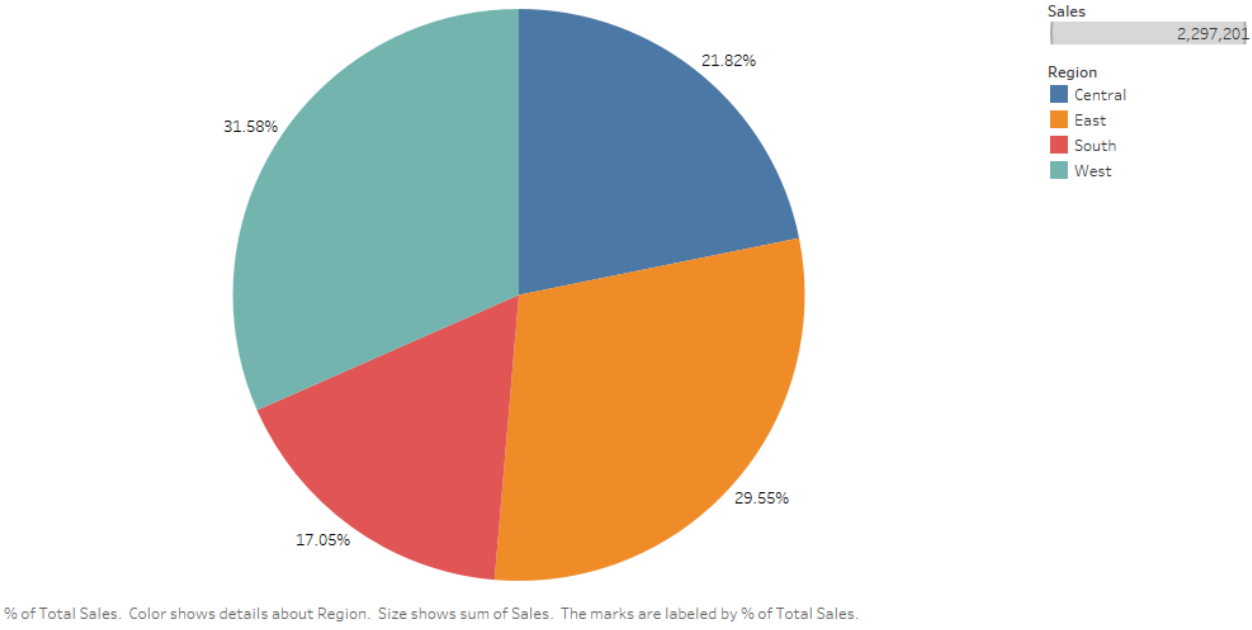
Outcome:

The pie chart shows:

- West region: 31.58% of total sales.
- East region: 29.55% of total sales.
- Central region: 21.82% of total sales.
- South region: 17.05% of total sales.

This visualization effectively communicates how each region contributes to overall sales, highlighting the distribution and helping in strategic planning and resource allocation.

Ans9



% of Total Sales. Color shows details about Region. Size shows sum of Sales. The marks are labeled by % of Total Sales.

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

Aim:

To visualize the profit margins associated with different shipping modes and customer segments in the "Superstore" dataset.

Chart Chosen:

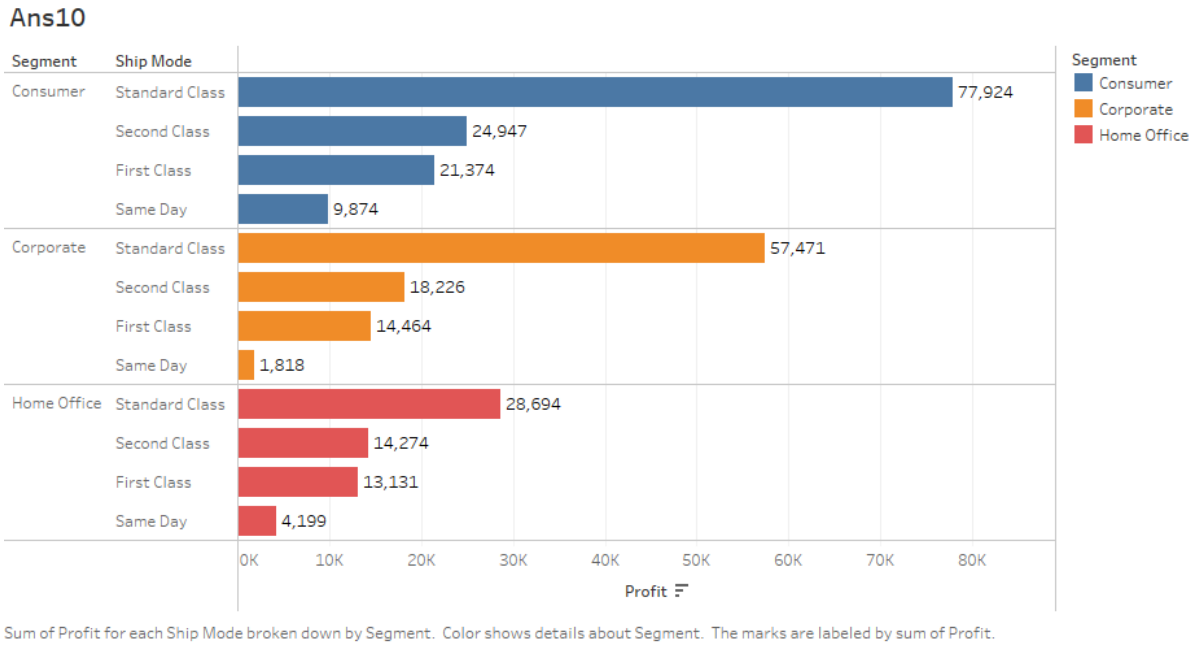
A column chart is selected to represent profit margins by shipping mode and customer segment. This chart is effective for comparing profit margins across categories.

Outcome:

The column chart reveals:

- Consumer segment has the highest profit margins, especially with the Standard Days shipping mode.
- SameDay shipping mode shows the lowest profit margins across all customer segments.
- Among shipping modes, Corporate customers prefer Standard Class the most, while Same Day is the least preferred.
- Home Appliance products in the Corporate segment show similar preferences to the general trend.

This visualization helps in understanding the profitability of different shipping options and customer preferences, aiding in optimizing shipping strategies and customer targeting.



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

Aim:

To visualize the average order processing times for different product categories.

Chart Chosen:

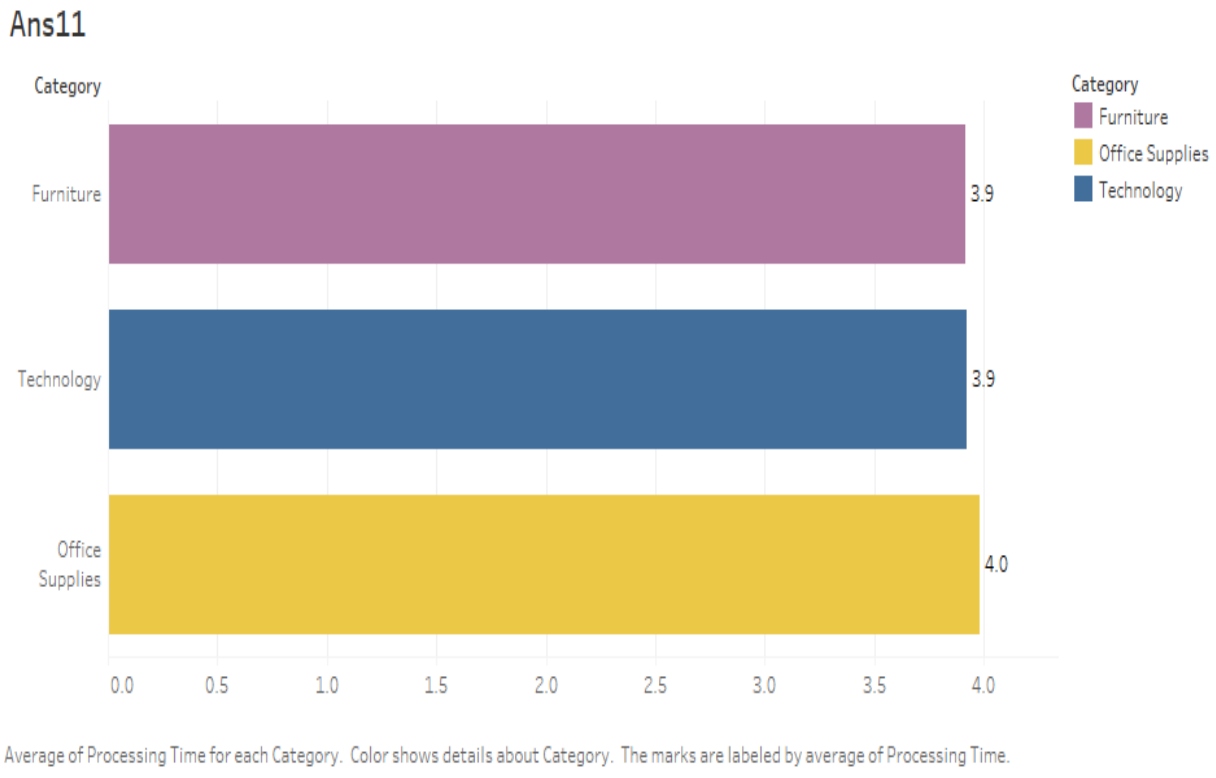
A bar chart is selected due to its effectiveness in comparing categorical data and clearly presenting differences in average processing times across categories.

Outcome:

The bar chart shows:

- Furniture and Technology categories each take 3.9 days on average for order processing.
- Office Supplies category takes slightly longer, averaging 4 days for order processing.

This visualization provides a clear comparison of processing times across product categories, helping to identify areas for potential improvement in order fulfillment efficiency.



12. How do discounts affect overall profit?

Aim:

To analyze how discounts affect overall profit.

Chart Chosen:

A line chart is ideal for visualizing trends in profit margins across different subcategories or time periods. It helps in understanding how discounting impacts profit margins over a continuous range.

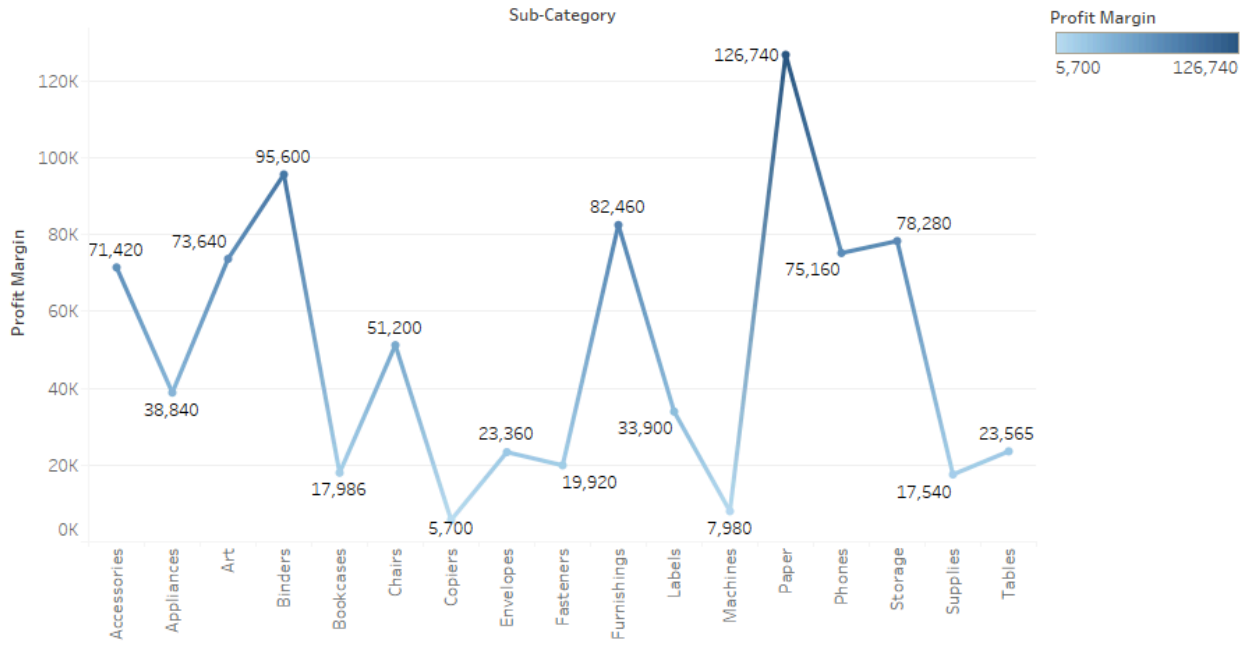
Outcome:

The line chart shows:

- Among subcategories, Paper has the highest profit after discount, with a total profit of $126,740.

This visualization helps in assessing how different discount levels influence profit margins and identifying which subcategories are most profitable after applying discounts.

Ans12



The trend of sum of Profit Margin for Sub-Category. Color shows sum of Profit Margin. The marks are labeled by sum of Profit Margin.

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

Aim:

To visualize the relationship between product sales and profitability for different product categories.

Chart Chosen:

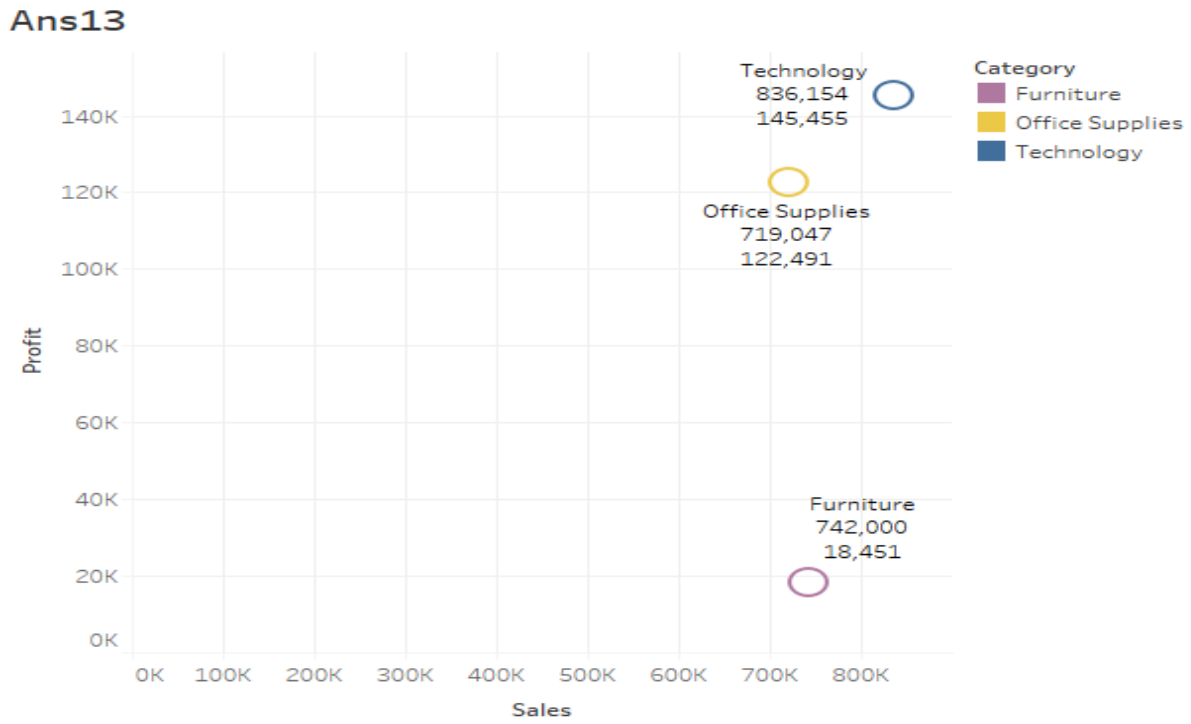
A scatter plot is used to represent the relationship between sales and profitability. By plotting averages, it helps normalize comparisons across categories and identify trends and correlations.

Outcome:

The scatter plot reveals:

- Technology: Sales of $836,154 and Profit of $145,455.
- Office Supplies: Sales of $719,047 and Profit of $122,491.
- Furniture: Sales of $742,000 and Profit of $18,451.

This visualization helps to understand how different product categories perform in terms of both sales and profitability, highlighting categories with higher profit margins and identifying areas for potential improvement.



Sum of Sales vs. sum of Profit. Color shows details about Category. The marks are labeled by Category, sum of Sales and sum of Profit.

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

Aim:

To visualize the distribution of order quantities for different products in the dataset.

Chart Chosen:

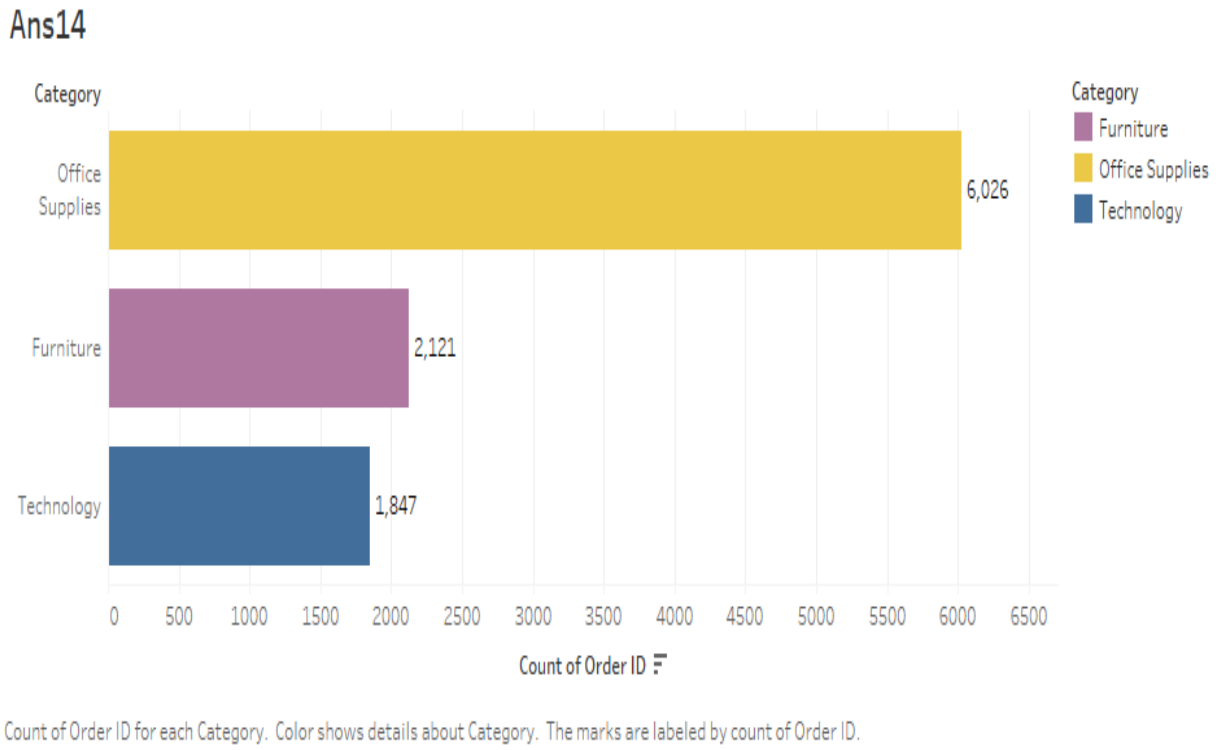
A bar chart is selected to effectively represent the distribution of order quantities across product categories. This chart clearly shows the comparison of order counts among categories.

Outcome:

The bar chart illustrates:

- Office Supplies has the highest order count at 6,026.
- Furniture has 2,121 orders.
- Technology has the lowest order count at 1,847.

This visualization helps to understand which product categories have the highest and lowest order volumes, aiding in inventory management and sales strategy development.



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

Aim:

To visualize the profit distributions across different product categories.

Chart Chosen:

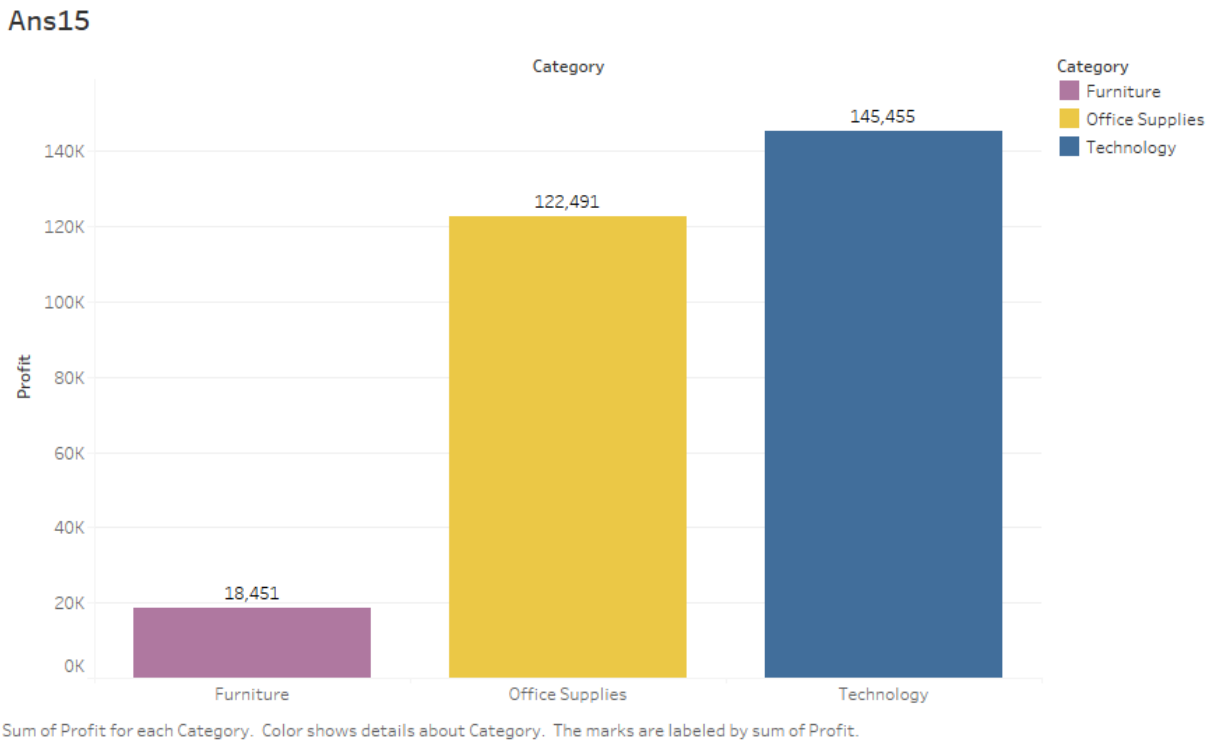
A bar chart is selected to effectively display variations in profit distributions among product categories. This chart clearly shows the differences in profit contributions across categories.

Outcome:

The bar chart reveals:

- Technology products have the highest profit contribution at $145,445.
- Office Supplies followed with a profit of $122,491.
- Furniture has the lowest profit contribution, with $18,451.

This visualization highlights the disparities in profitability among different product categories, aiding in identifying which categories are more profitable and guiding strategic business decisions.



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

Aim:

To compare the shipping time distributions for different shipping modes.

Chart Chosen:

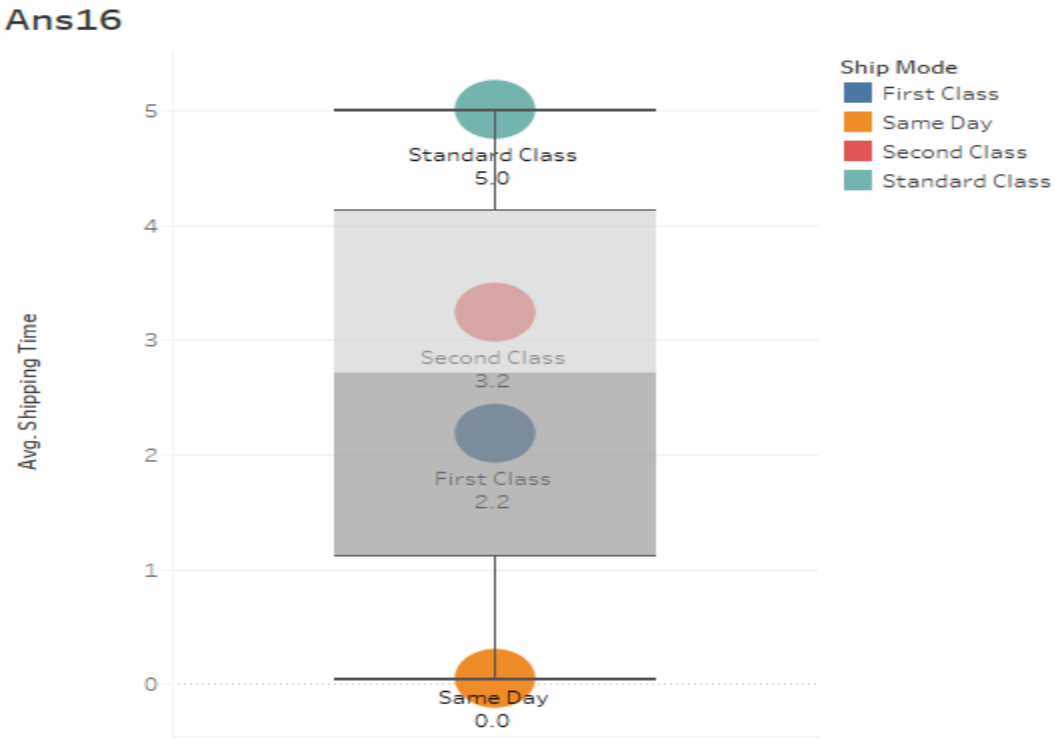
A **box plot** is chosen to compare shipping time distributions across different shipping modes. It provides a clear view of the variation in shipping times, including medians, variability, and outliers.

Outcome:

The box plot shows:

- Standard Class takes the longest time, averaging around 5 days for shipping.

This visualization highlights the differences in shipping times among various modes, helping to identify which shipping options are more efficient or slower, and informing decisions on shipping strategies and customer satisfaction.



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

Aim:

To visualize the monthly trend in the number of orders shipped.

Chart Chosen:

A line plot is selected for its ability to clearly display changes over time, highlight trends, and show fluctuations in order volume.

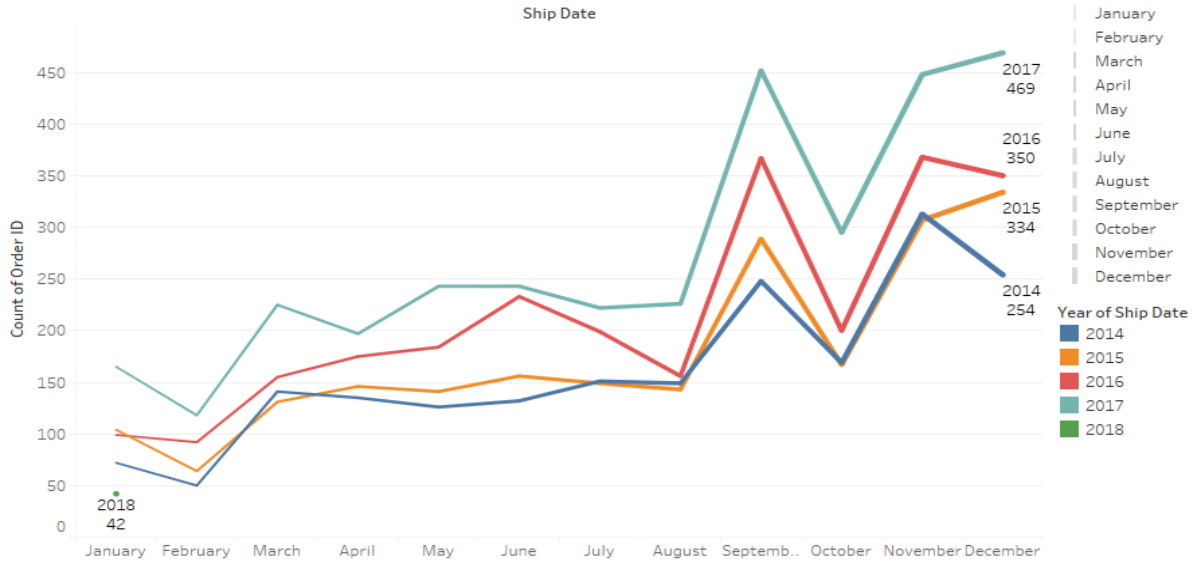
Outcome:

The line plot shows the monthly trend in the number of orders shipped for different years:

- 2017: Highest order volume of 469 in December.
- 2016: Highest order volume of 367 in September.
- 2015: Highest order volume of 334 in December.
- 2014: Highest order volume of 313 in November.
- 2018: Data available only for January, with 42 orders.

This visualization provides insights into seasonal patterns and helps in understanding order volume trends across different years.

Ans17



The trend of count of Order ID for Ship Date Month. Color shows details about Ship Date Year. Size shows details about Ship Date Month. The marks are labeled by Ship Date Year and count of Order ID.

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

Aim:

To compare the performance of different customer segments in terms of sales and discount rates.

Chart Chosen:

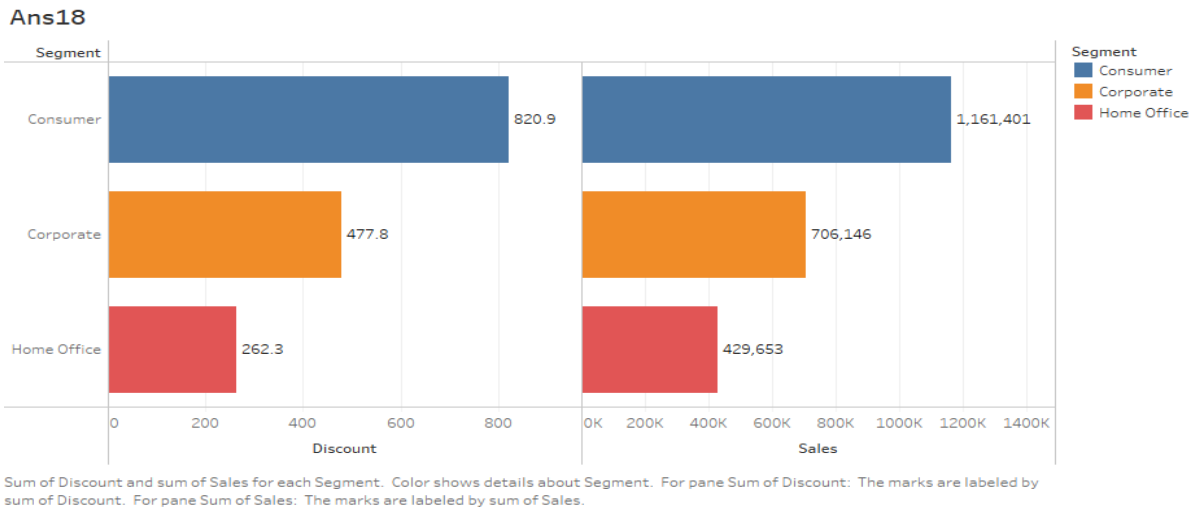
A bar plot is chosen for its effectiveness in comparing categorical data and clearly showing differences in sales and discount rates across customer segments.

Outcome:

The bar plot reveals:

- Consumer segment:
    - Highest Discounts: 820.9
    - Total Sales: $1,161,401
- Corporate segment:
    - Discounts: 477.8
    - Total Sales: $706,146
- Home Office segment:
    - Lowest Discounts: 262.3
    - Total Sales: $429,653

This visualization provides a clear comparison of sales and discount rates across customer segments, aiding in understanding which segments receive higher discounts and generate more sales, thus informing marketing and pricing strategies.



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

Aim:

To analyze sales and profit trends across different product subcategories and regions.

Chart Chosen:

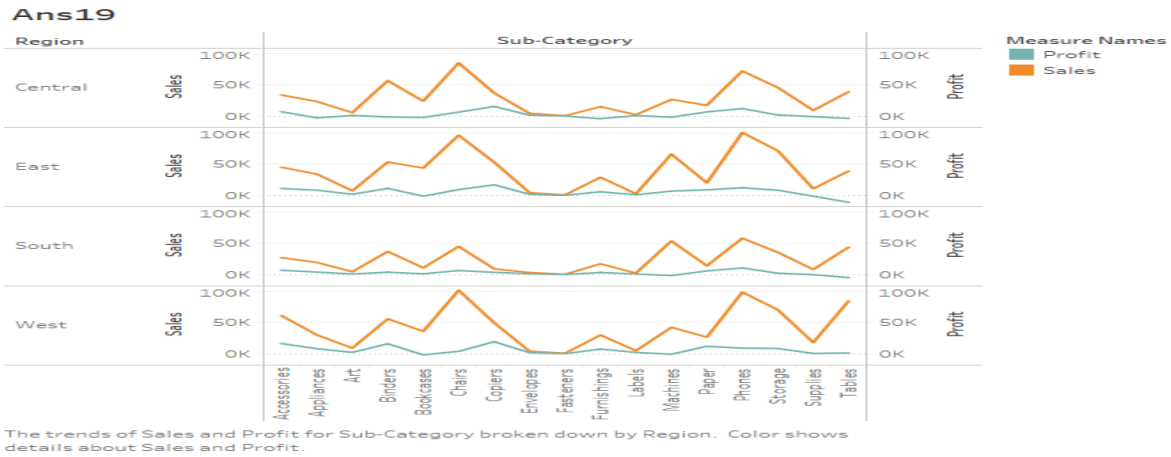
A dual-axis chart is selected to compare sales and profit trends on the same graph. This allows for an easy comparison of how trends in sales relate to trends in profit across different product subcategories and regions.

Outcome:

The dual-axis chart shows:

- Central Region:
  - Highest Sales: $85,231 (Chairs)
  - Highest Profit: $15,609 (Chairs)
- East Region:
  - Highest Sales: $100,615 (Phones)
  - Highest Profit: $17,023 (Copiers)
- South Region:
  - Highest Sales: $58,304 (Phones)
  - Highest Profit: $10,767 (Phones)
- West Region:
  - Highest Sales: $101,781 (Chairs)
  - Highest Profit: $19,327 (Chairs)

This chart helps to visualize how sales and profit trends are interrelated across various subcategories and regions, aiding in understanding regional performance and identifying key profitable areas.



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

Aim:

To compare the average delivery duration for different regions and shipping modes.

Chart Chosen:

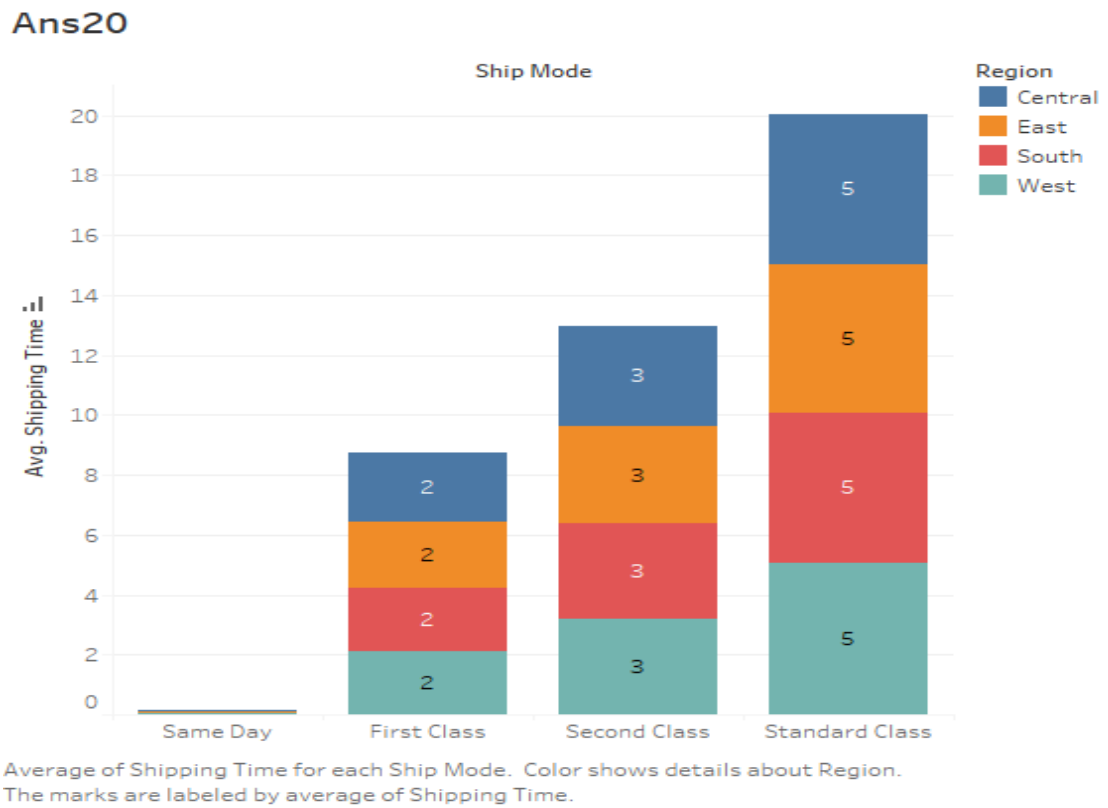
A stacked bar chart is chosen to effectively display the average delivery durations across regions, with a breakdown of each shipping mode's contribution.

Outcome:

The stacked bar chart reveals:

- Longest Shipping Time: Across all regions, the Standard Class shipping mode has the longest average delivery duration, averaging around 5 days.

This chart provides a clear comparison of average delivery times, helping to identify which shipping modes and regions have longer or shorter delivery durations.



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

Aim:

To analyze how the total order quantity has changed over the years for various product categories.

Chart Chosen:

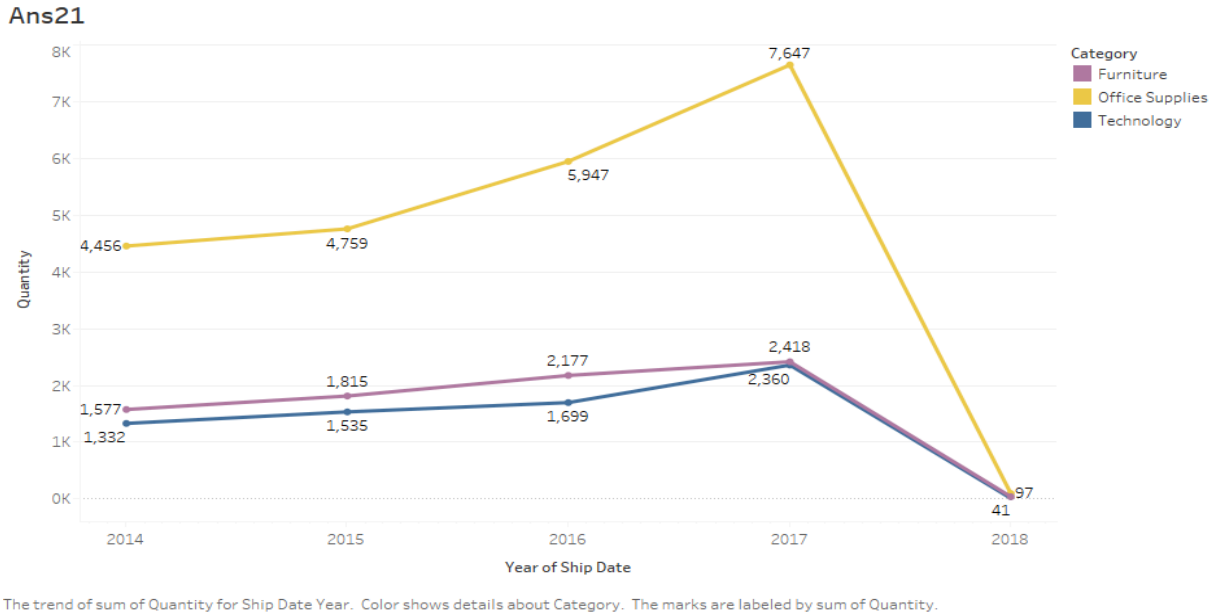
A line chart is selected as it effectively tracks changes in the **total** order quantities over time for each product category, allowing for an easy comparison across years.

Outcome:

The line chart shows the following:

- Office Supplies had the highest total order quantity in 2017, with 7,647 units.
- Furniture recorded its highest total order quantity in 2017, with 2,418 units.
- Technology also reached its peak in 2017, with 2,360 units.

This visualization highlights how order quantities have evolved, helping identify the growth or decline of demand for different product categories over the years.



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

Aim:

To analyze the correlation between discount rates and order quantities for different customer segments.

Chart Chosen:

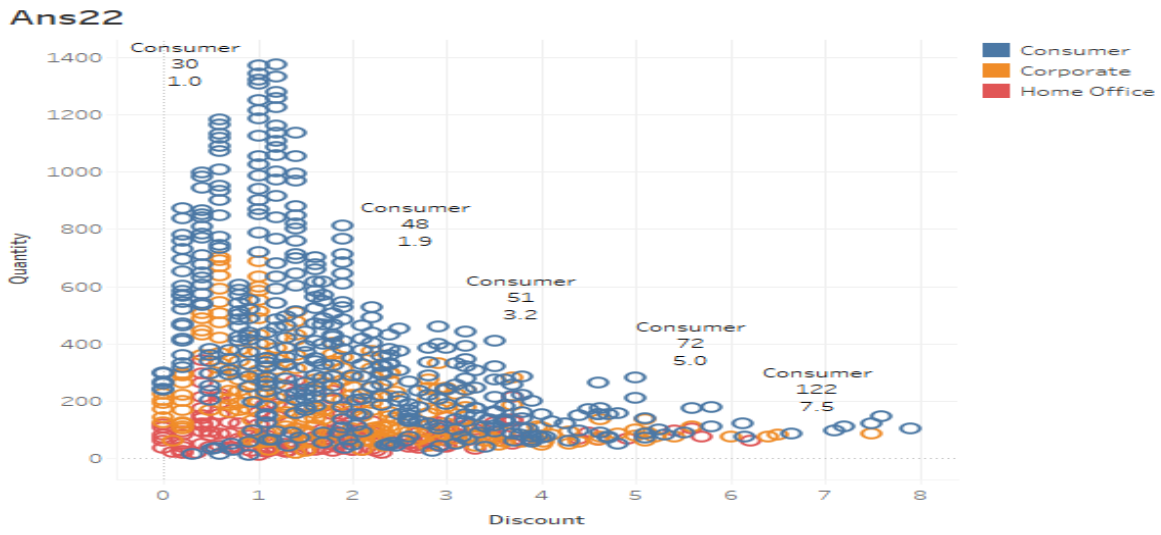
A scatter plot is selected because it visually represents the relationship between two variables: discount rates and order quantities. By plotting these variables for different customer segments, we can easily compare how discounts influence order quantities.

Outcome:

The scatter plot reveals patterns across different customer segments:

- Consumer segment shows a stronger correlation between higher discounts and larger order quantities, indicating a sensitivity to discounts.
- Corporate and Home Office segments exhibit less variation in order quantities with increasing discount rates, suggesting a different response to discounts.

This visualization provides valuable insights into customer behavior, helping businesses tailor their discount strategies for different customer segments.



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

Aim:

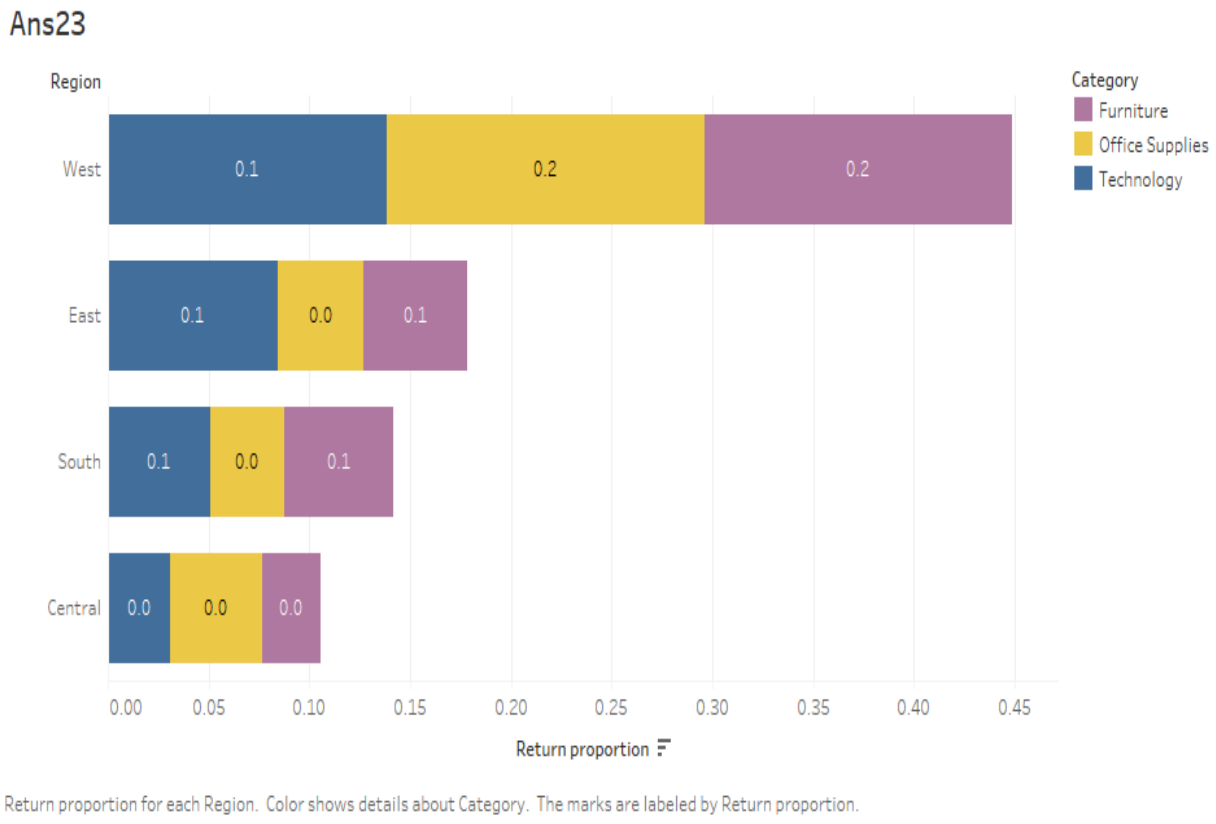
To visualize the proportion of orders returned in each region within the Superstore dataset.

Chart Chosen:

A bar chart is selected to highlight the proportion of returned orders across different regions. This allows for clear comparison of return rates and easy identification of regional differences.

Outcome:

The bar chart reveals that the West region has the highest proportion of returned orders across all categories. Other regions such as East, South, and Central have relatively lower return proportions. This insight could help in investigating potential issues or opportunities related to customer satisfaction, product quality, or delivery services in the West region.



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

Aim:

To compare the profit of different products within various subcategories in the Superstore dataset.

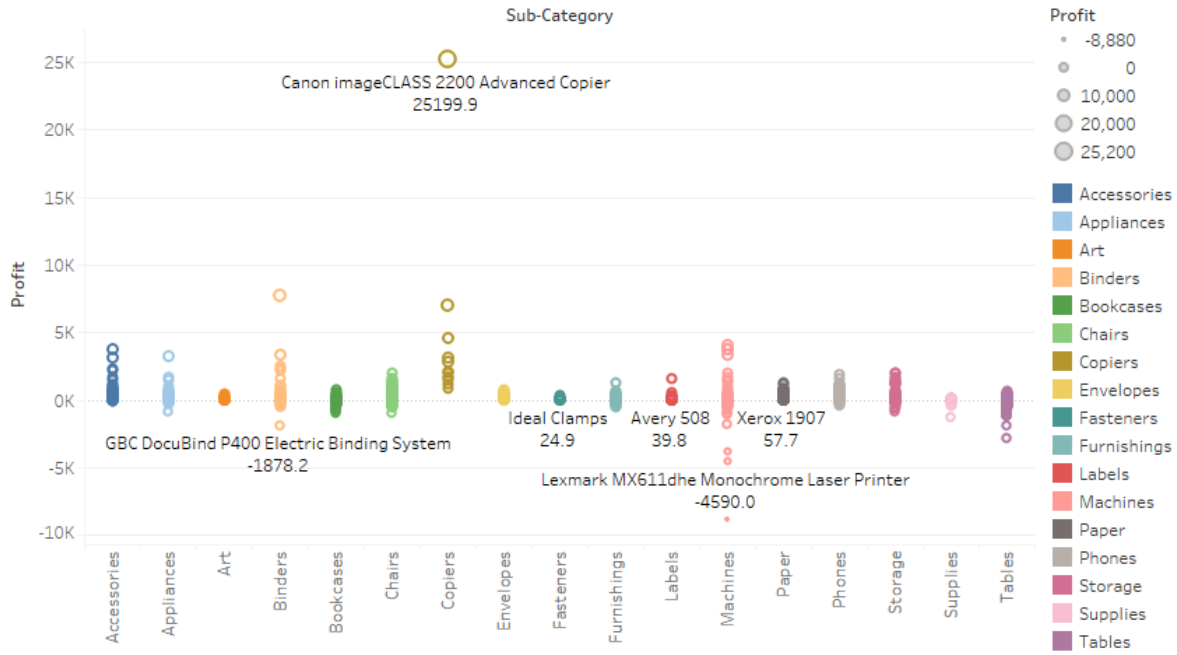
Chart Chosen:

A circle view chart (bubble chart) is selected to visually compare the profit of products within different subcategories. This chart effectively highlights the profit distribution by showing product size (based on profit) within subcategories.

Outcome:

The circle view reveals that the highest profit is observed under the Copier subcategory, with the product, Canon Image CLASS2200 Advanced Copier generating the most profit. This visualization allows for easy identification of the most profitable products within each subcategory, aiding in decision-making for product strategy and inventory management.

Ans24



Sum of Profit for each Sub-Category. Color shows details about Sub-Category. Size shows sum of Profit. The marks are labeled by Product Name and sum of Profit. Details are shown for Product Name.

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

Aim:

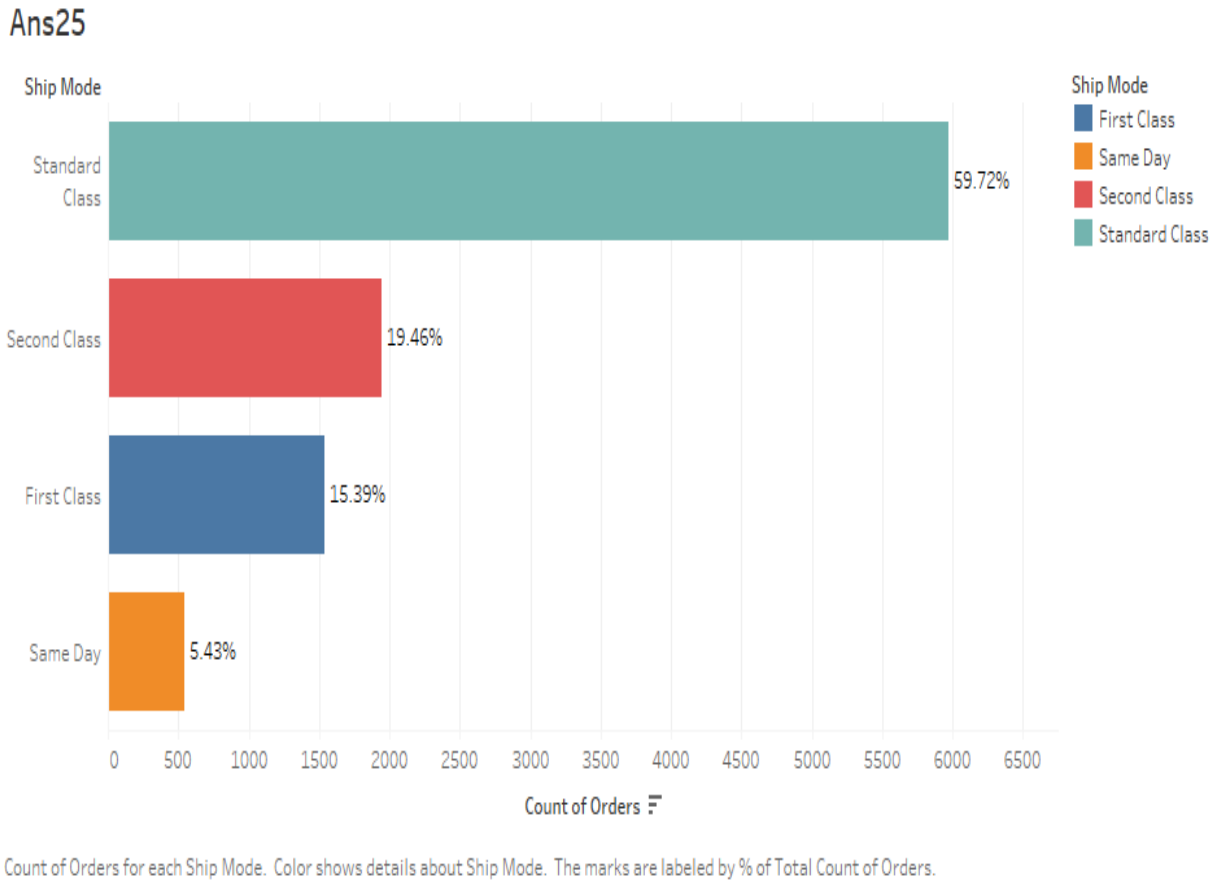
To determine the most commonly used shipping mode in the Sample Superstore dataset.

Chart Chosen:

A bar chart is chosen for this analysis because it clearly displays the frequency of orders for each shipping mode, making it easy to compare their usage.

Outcome:

The analysis shows that Standard Class is the most commonly used shipping mode, accounting for 59.72% of all orders. This indicates a strong preference for Standard Class in the dataset, with other shipping modes being less frequently utilized.



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

Aim:

To analyze how the sales performance of different regions evolves throughout the quarters of a year.

Chart Chosen:

A line chart is selected for this analysis as it effectively tracks and compares sales trends over time, providing insights into regional performance and seasonal variations.

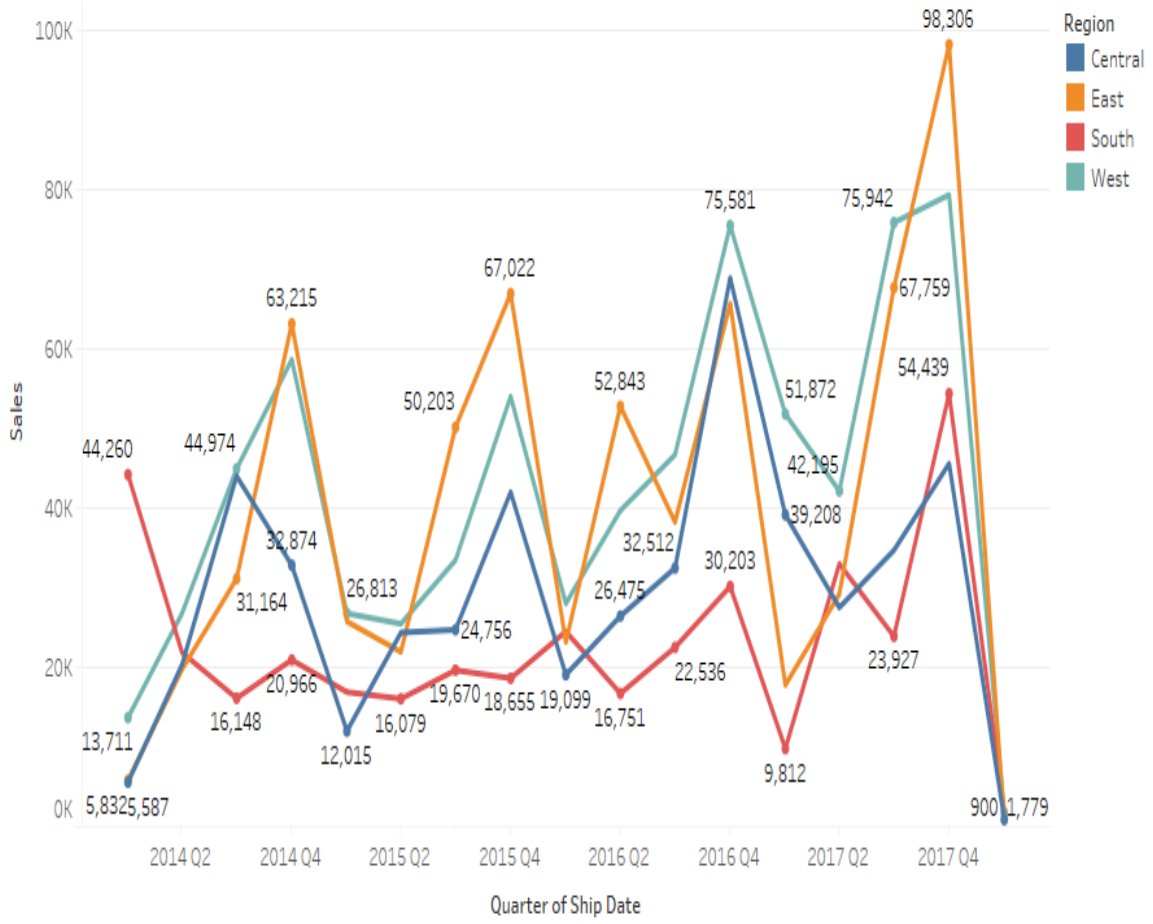
Outcome:

The line chart illustrates the following trends in sales performance by region across different quarters:

- 2014:
  - Q1: South leads with $44,260.
  - Q2: West leads with $26,806.
  - Q3: West leads with $44,974.
  - Q4: East leads with $63,215.
- 2015:
  - Q1: West leads with $26,813.
  - Q2: West continues to lead with $26,813.
  - Q3: East leads with $50,203.
  - Q4: East leads with $64,022.
- 2016:
  - Q1: West leads with $28,056.
  - Q2: East leads with $52,843.
  - Q3: West leads with $46,755.
  - Q4: West leads with $75,581.
- 2017:
  - Q1: West leads with $51,872.
  - Q2: West leads with $42,195.
  - Q3: West leads with $75,942.
  - Q4: East leads with $98,306.

This visualization helps in identifying the dominant regions each quarter and understanding seasonal sales patterns across different years.

Ans26



The trend of sum of Sales for Ship Date Quarter. Color shows details about Region. The marks are labeled by sum of Sales.

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

Aim:

To visualize the distribution of order priorities across different product categories.

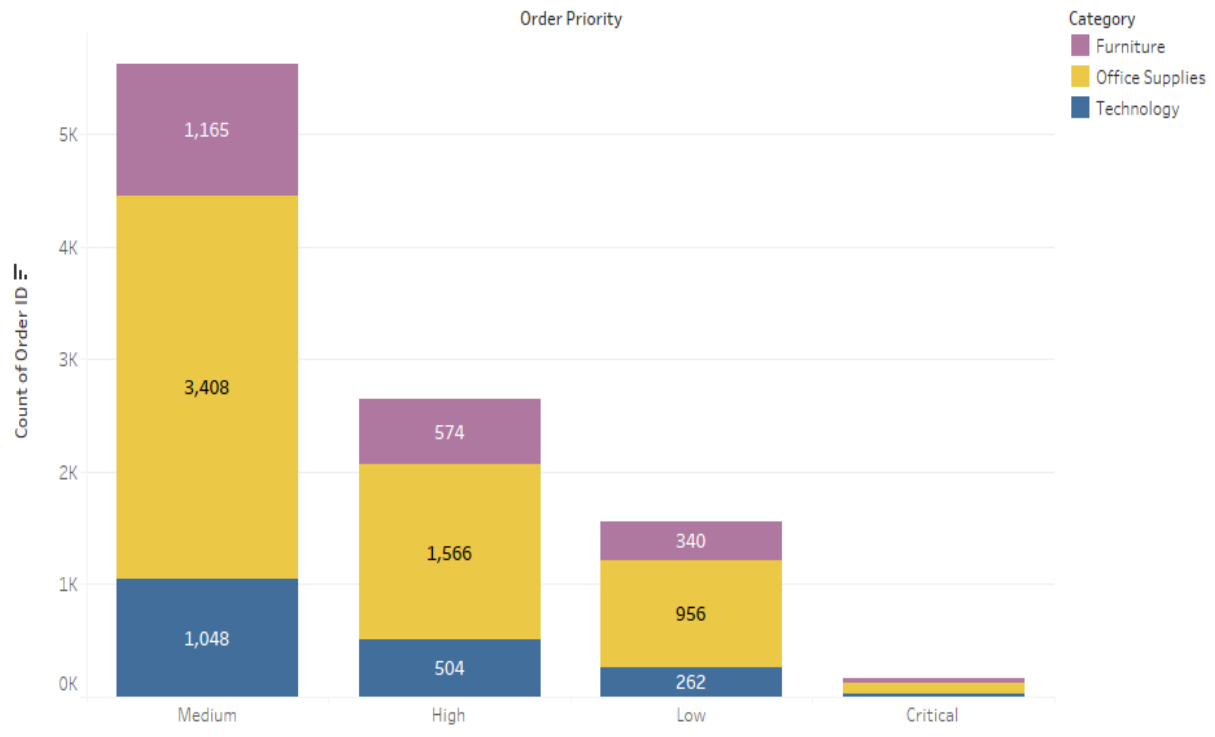
Chart Chosen:

A bar chart is selected as it effectively shows the distribution of order priorities across various product categories, making it easy to compare and understand how priorities are allocated.

Outcome:

The bar chart reveals that order priorities are uniformly distributed across all product categories, with the majority of orders categorized as "Medium" priority. This insight indicates that the threshold for determining priority is met consistently across categories, with fewer orders falling into "High," "Low," or "Critical" priority levels.

Ans27



Count of Order ID for each Order Priority. Color shows details about Category. The marks are labeled by count of Order ID.

28. What is the relationship between discounts and sales?

Aim:

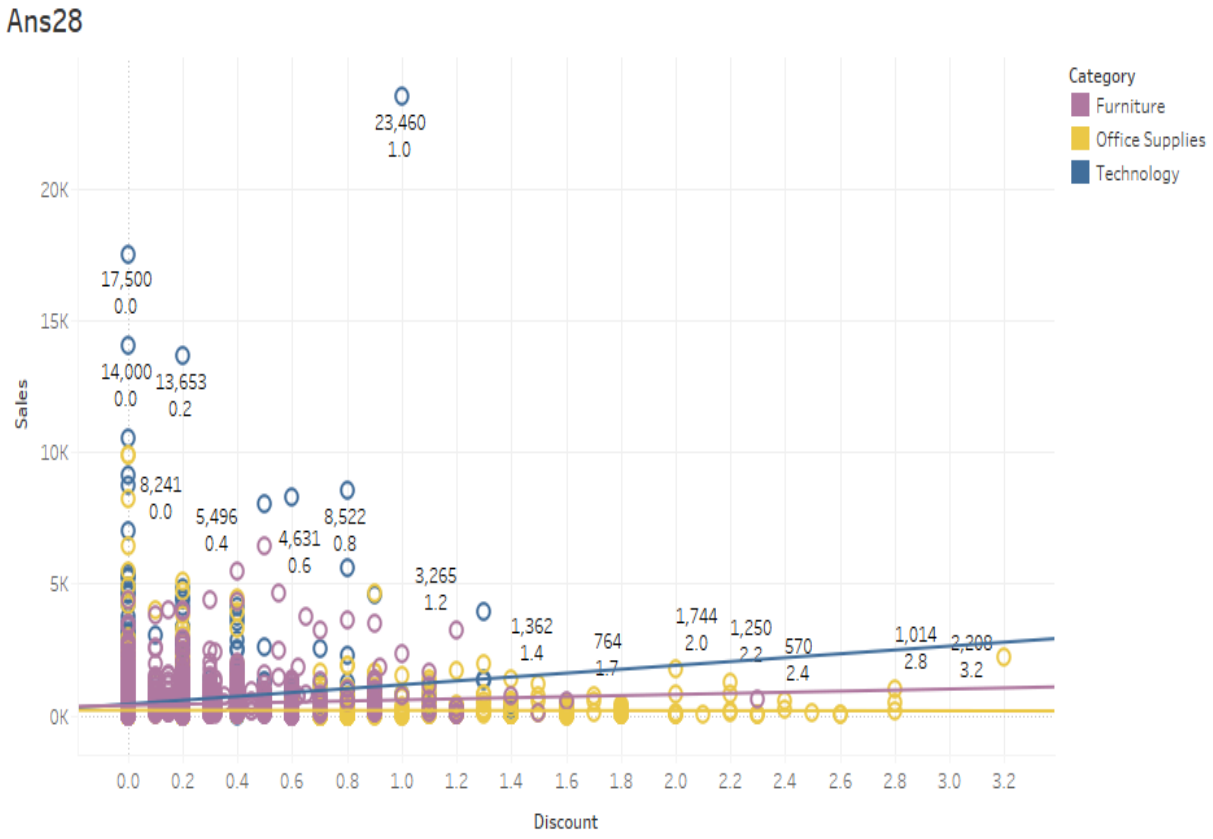
To analyze the relationship between discounts and sales.

Chart Chosen:

A scatter plot is selected for this analysis because it effectively shows the correlation between discount levels and sales figures. It helps to identify patterns, trends, or any significant relationships between the amount of discount offered and the corresponding sales outcomes.

Outcome:

This visualization provides valuable insights into the effectiveness of discount strategies and their impact on sales performance.



Sum of Discount vs. sum of Sales. Color shows details about Category. The marks are labeled by sum of Sales and sum of Discount. Details are shown for Order ID.

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

Aim:

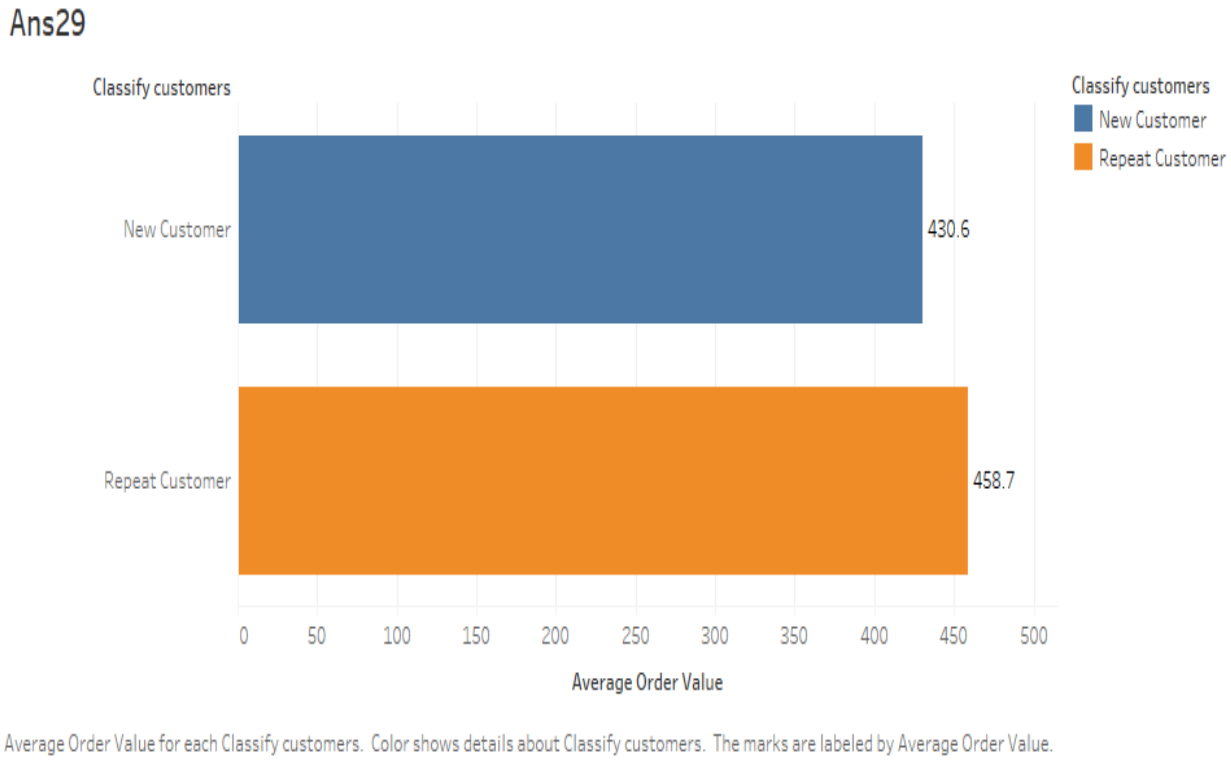
To compare the average order value between repeat and new customers.

Chart Chosen:

A bar chart is selected for this analysis because it effectively displays the average order value for each customer group, allowing for a straightforward comparison between repeat and new customers.

Outcome:

The bar chart reveals that repeat customers have a higher average order value of $458.7 compared to new customers, who have an average order value of $430.6. This insight indicates that repeat customers tend to place larger orders on average, which could be useful for tailoring marketing strategies and customer engagement efforts.



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

Aim:

To visualize the geographical distribution of returns and understand their impact on overall profitability.

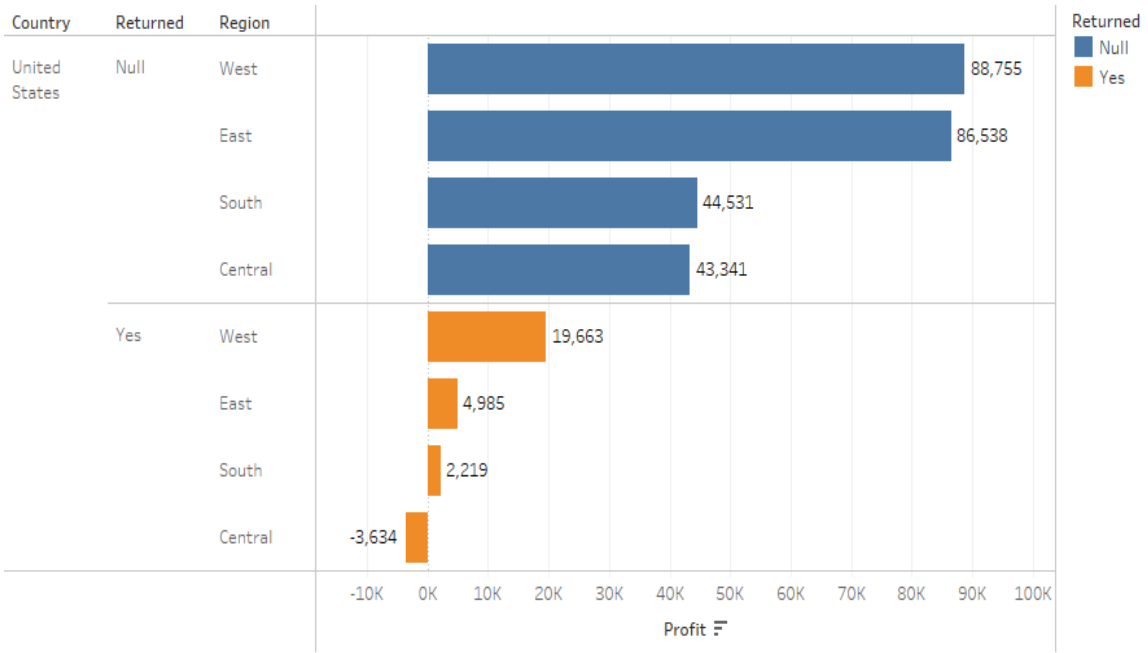
Chart Chosen:

A column chart is selected for this analysis as it effectively displays the distribution of returns across different regions and their effect on profitability.

Outcome:

The column chart demonstrates that products with returns generally result in lower profitability compared to those that are not returned, across all regions. This insight highlights the financial impact of returns on profitability and can guide strategies to reduce return rates and improve overall profit margins.

Ans30



Sum of Profit for each Region broken down by Country and Returned. Color shows details about Returned. The marks are labeled by sum of Profit.

Thank You