

**Github Link -**

[https://github.com/mishra2022/almabetter/blob/main/Rajesh\\_MidAssesment- Module 4.pdf](https://github.com/mishra2022/almabetter/blob/main/Rajesh_MidAssesment-Module4.pdf)

**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 visualisation and analysis. Your objective is to select the best chart for each question, explain your choice. This project will showcase your proficiency in data visualisation, critical thinking, and effective communication.

**Skills Required:**

- Proficiency in data visualisation concepts and techniques.
- Familiarity with Tableau or a similar data visualisation 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 visualisation 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 visualise 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 visualisation and analysis project. Remember to maintain consistency in your responses, explanations, and visualisation styles. This project will not only demonstrate your skills but also your ability to effectively communicate complex information through visualisations. 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 visualisation 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, organised, 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 any 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 visualisation 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?
2. How do the monthly sales amounts change over the course of a year?

3. How is the total sales amount distributed among different product categories?
4. Can we analyze the sales performance of individual customers over time?
5. How do sales vary based on different days of the week and product categories?
6. Can we visualise the sales growth of different product categories over time?
7. How does the sales distribution vary across different regions in the "Superstore" dataset?
8. Can we visualise the composition of profits across various subcategories within different customer segments?
9. What is the percentage contribution of each region to the overall sales?
10. Can we visualise the profit margins associated with different shipping modes and customer segments?
11. How long does it take to process orders for different product categories?
12. How do discounts affect overall profit?

13. Can we visualise the relationship between product sales and profitability for different product categories?
14. What is the distribution of order quantities for products in the dataset?
15. How do the profit distributions vary across different product categories?
16. Can we compare the shipping time distributions for different shipping modes?
17. What is the monthly trend in the number of orders shipped?
18. How do different customer segments perform in terms of sales and discount rates?
19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?
20. What is the average delivery duration for different regions and ship modes?
21. How has the average order quantity changed over the years for various product categories?

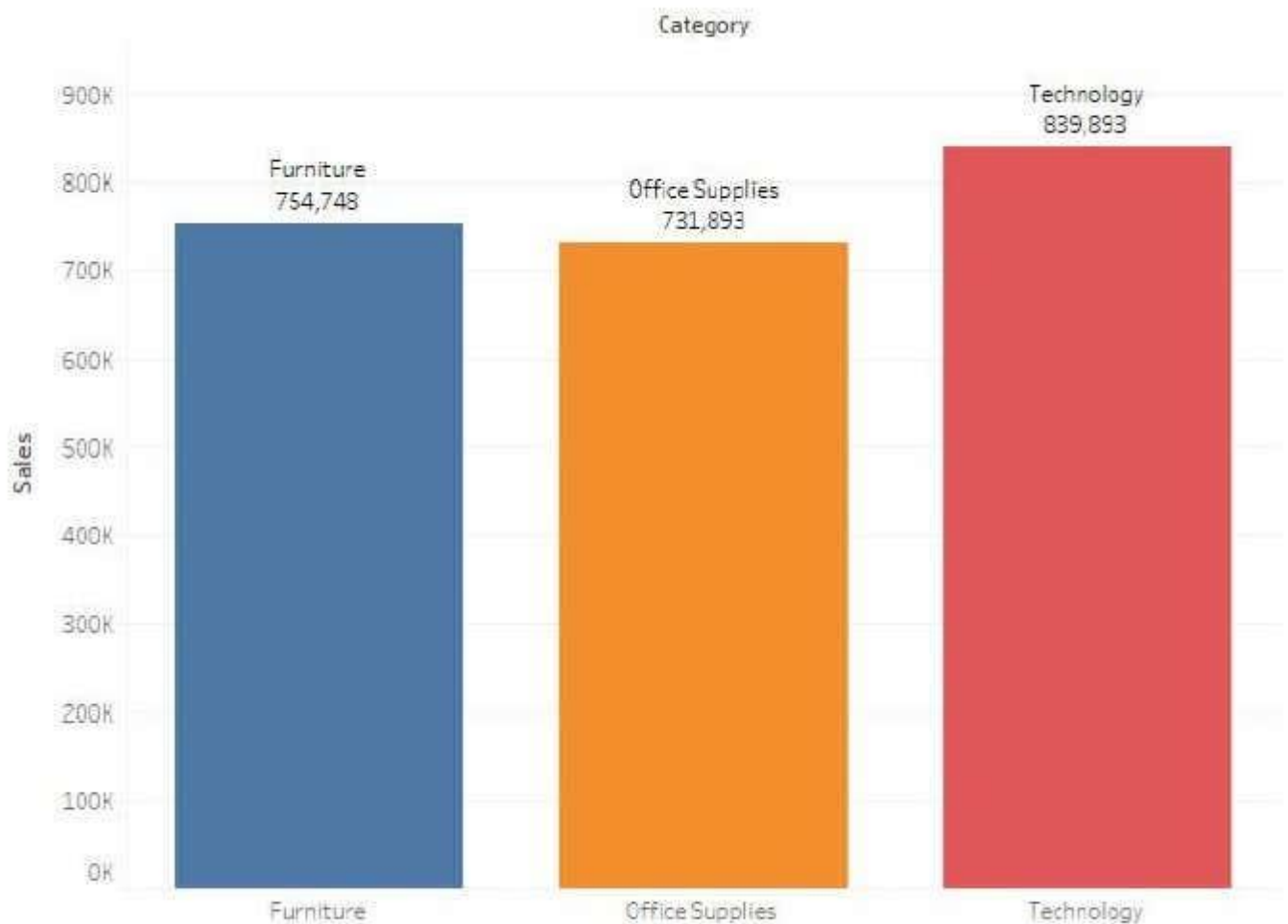
22. Can we visualise the correlation between discount rates and order quantities for different customer segments?
23. What is the proportion of orders returned in each region within the Superstore dataset?
24. Can you compare the profit of different products for different subcategories?
25. Which shipping mode is the most commonly used in the Sample Superstore dataset?
26. How does the sales performance of different regions evolve throughout the quarters of a year?
27. What is the distribution of order priorities across different product categories?
28. What is the relationship between discounts and sales?
29. How does the average order value differ between repeat customers and new customers?
30. What is the geographical distribution of returns and its impact on overall profitability?

**Questions:**

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

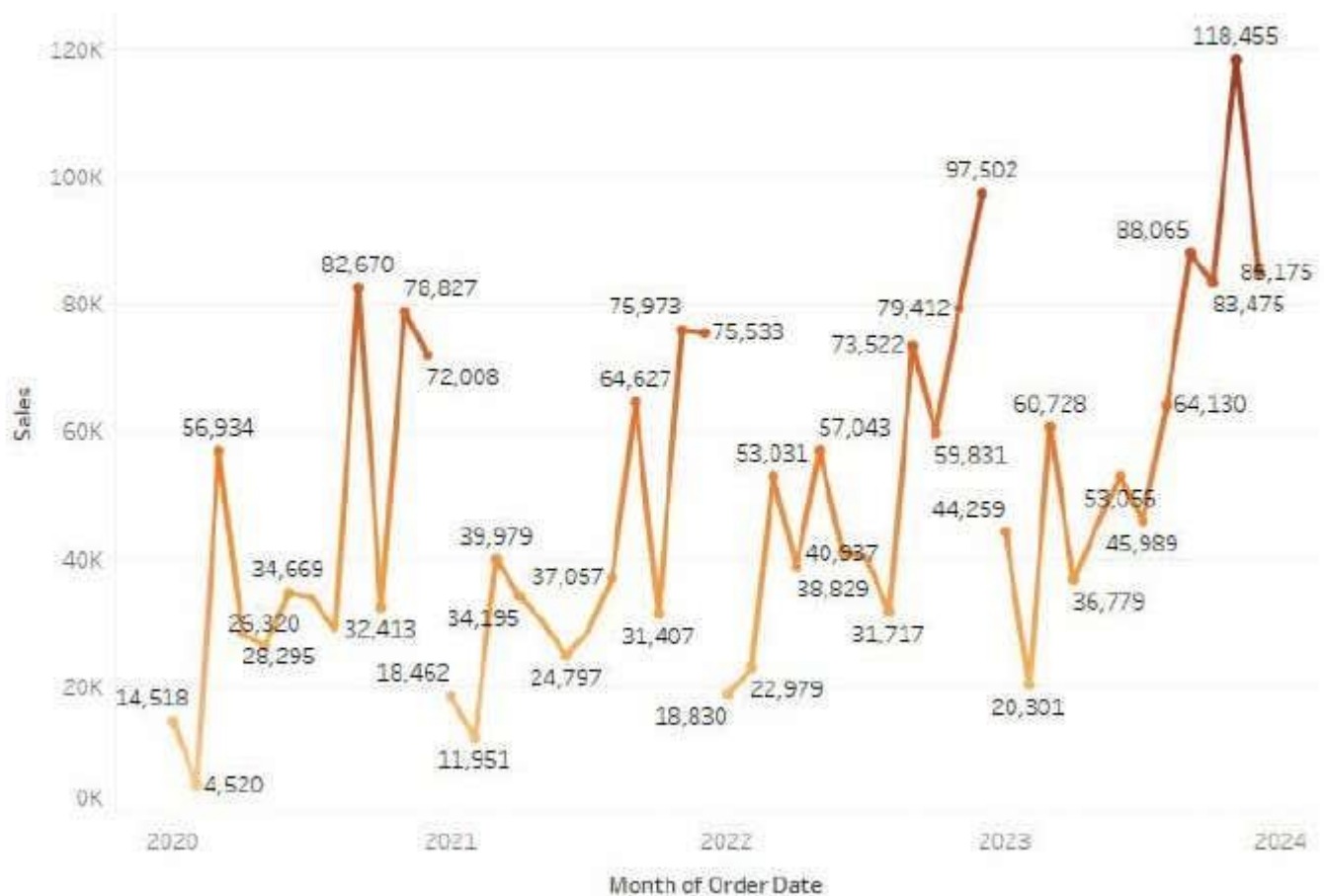
- I'm choosing a bar plot for easy understanding and to identify the best solutions for customers. The highest-selling technology category has total sales of 839,893, while the lowest-selling category is office supplies with total sales of 731,893.

Q1.



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

Ans2.



● I am choosing a line plot because it visually represents the trend over time, making it easy to understand. In 2020, February had the highest sales at 4,520, and the maximum sales occurred in September at 82,670. In 2021, February had sales of 11,951, and the highest sales were in November at 75,973. In 2022, January had sales of 18,830, and the maximum sales were in December at 97,502. Finally, in 2023, February had sales of 20,301, and the highest sales were in November at 118,455.

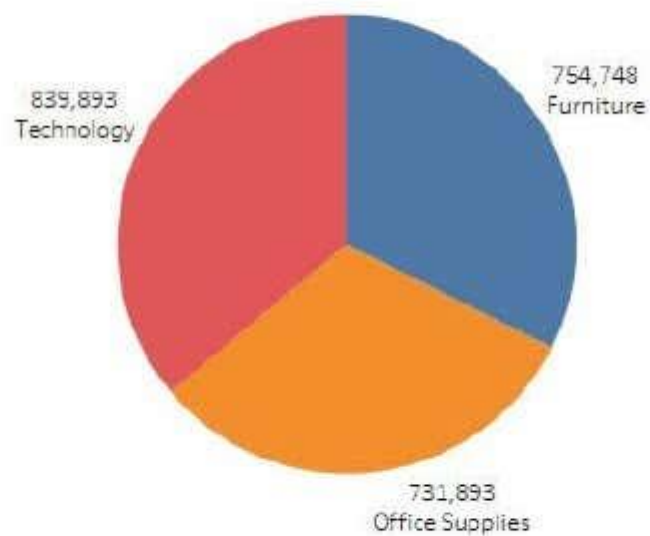
3 How is the total sales amount distributed among different product categories?

● I chose a pie plot to visually represent the distribution of the total sales amount among different product categories. In this plot, it's evident that the highest sales are attributed to the "Technology" category, while the "Office Supplies" category has the lowest sales. This



visualisation effectively conveys the proportion of sales each category contributes to the overall total, helping stakeholders quickly grasp how the sales amount is distributed among the product categories, aiding in decision-making and resource allocation for the business

Ans 3.



.4.How does the performance of employees vary within each department across different regions?

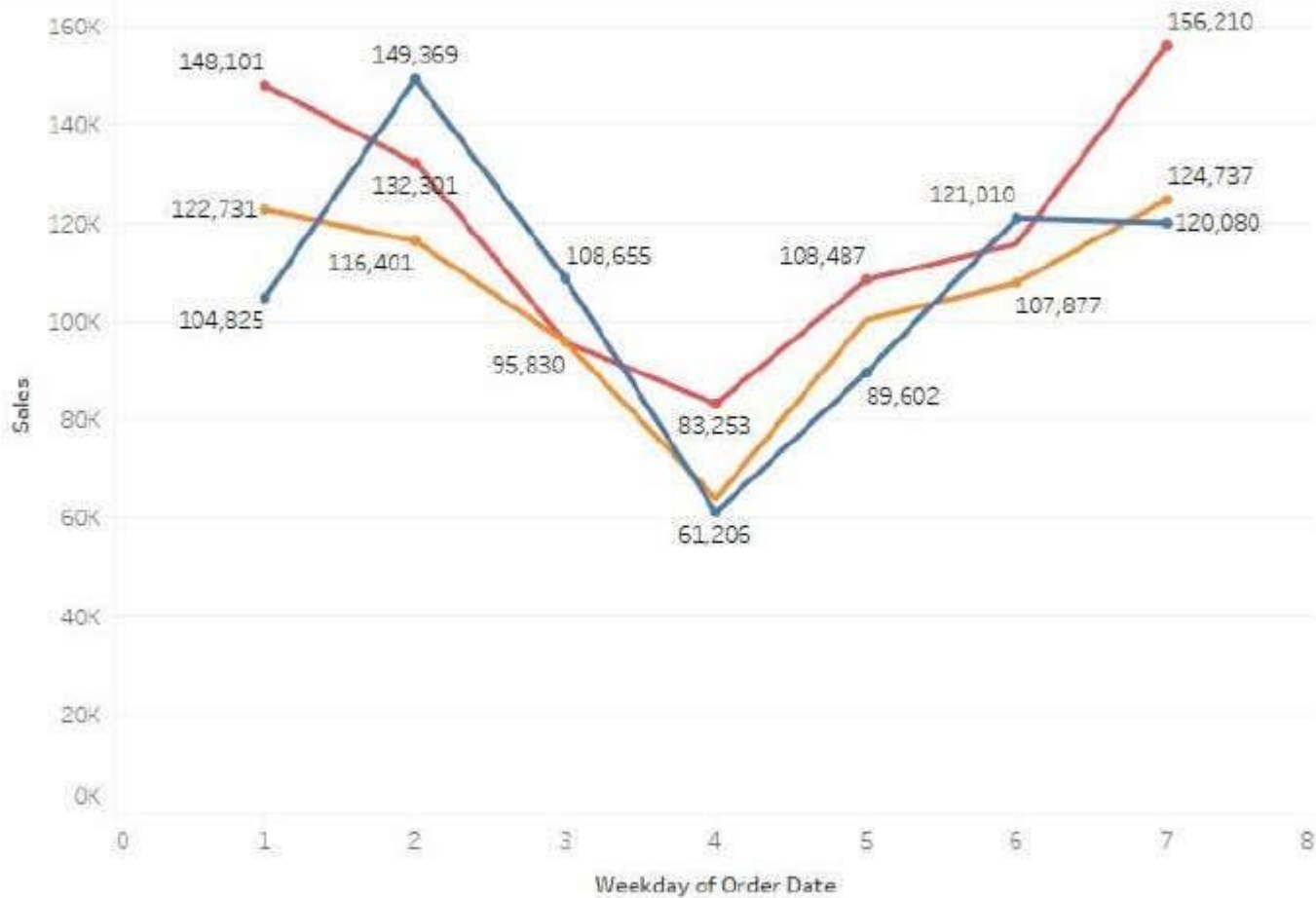
Ans4



- I chose a tree map to analyze employee performance within each department across various regions. The tree map illustrates that the Consumer department in the West region contributed the most, with 57,710 units, while the Home Supplies department in the South had the lowest contribution, with only 4,621 units. This visualization helps identify performance disparities among departments and regions, aiding in strategic decision-making

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

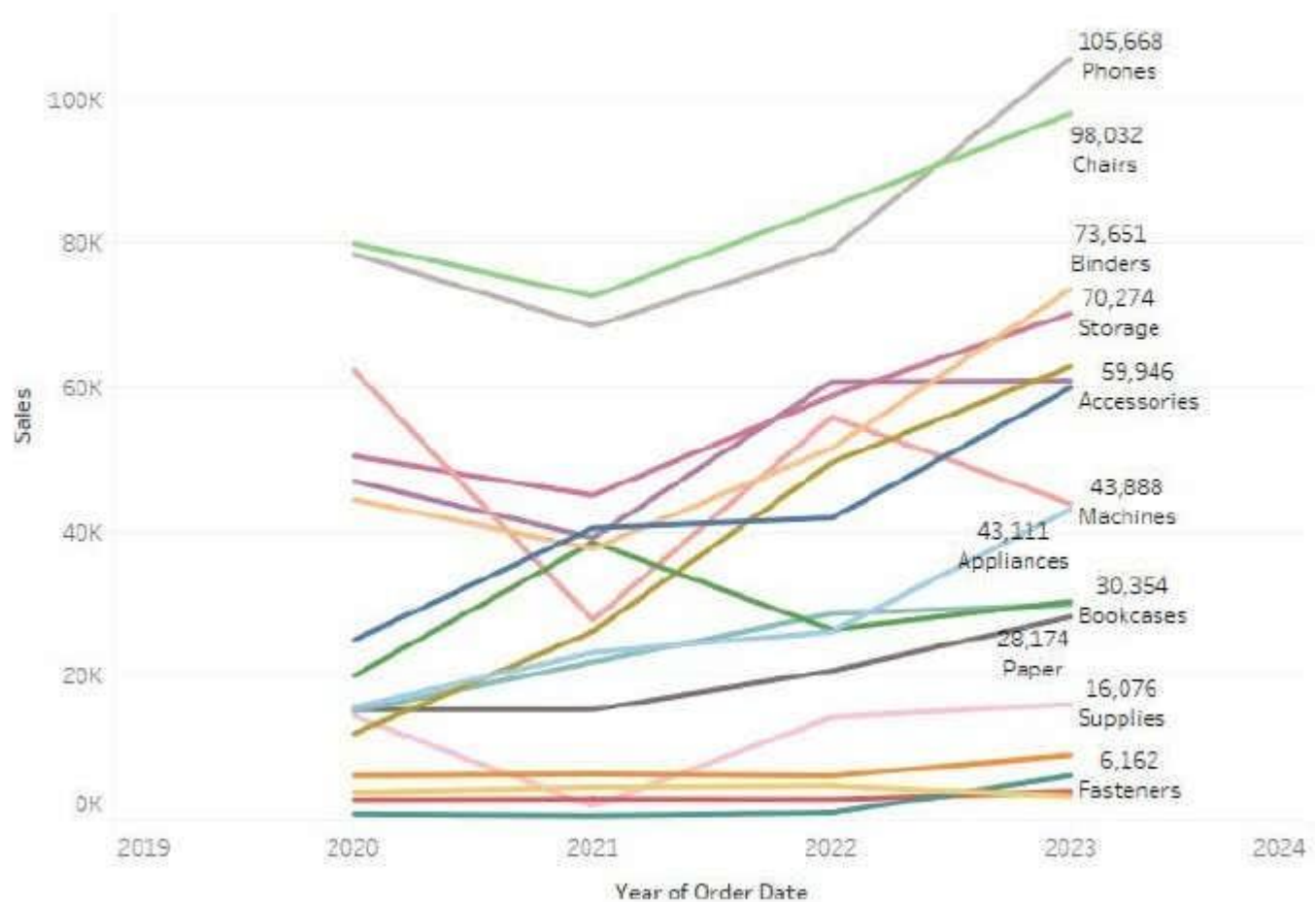
Ans5.



● Sales data analysis using a line chart reveals interesting patterns. Among product categories, Furniture sales are lowest on Wednesdays at 61,208, followed by Office Supplies at 63,992, and Technology at 83,252. Conversely, Furniture sales peak on Mondays at 149,369, Technology on Saturdays at 156,210, and Office Supplies on Saturdays as well, with 124,737. This visualization helps identify weekly sales trends by product category, aiding in strategic decision-making.

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

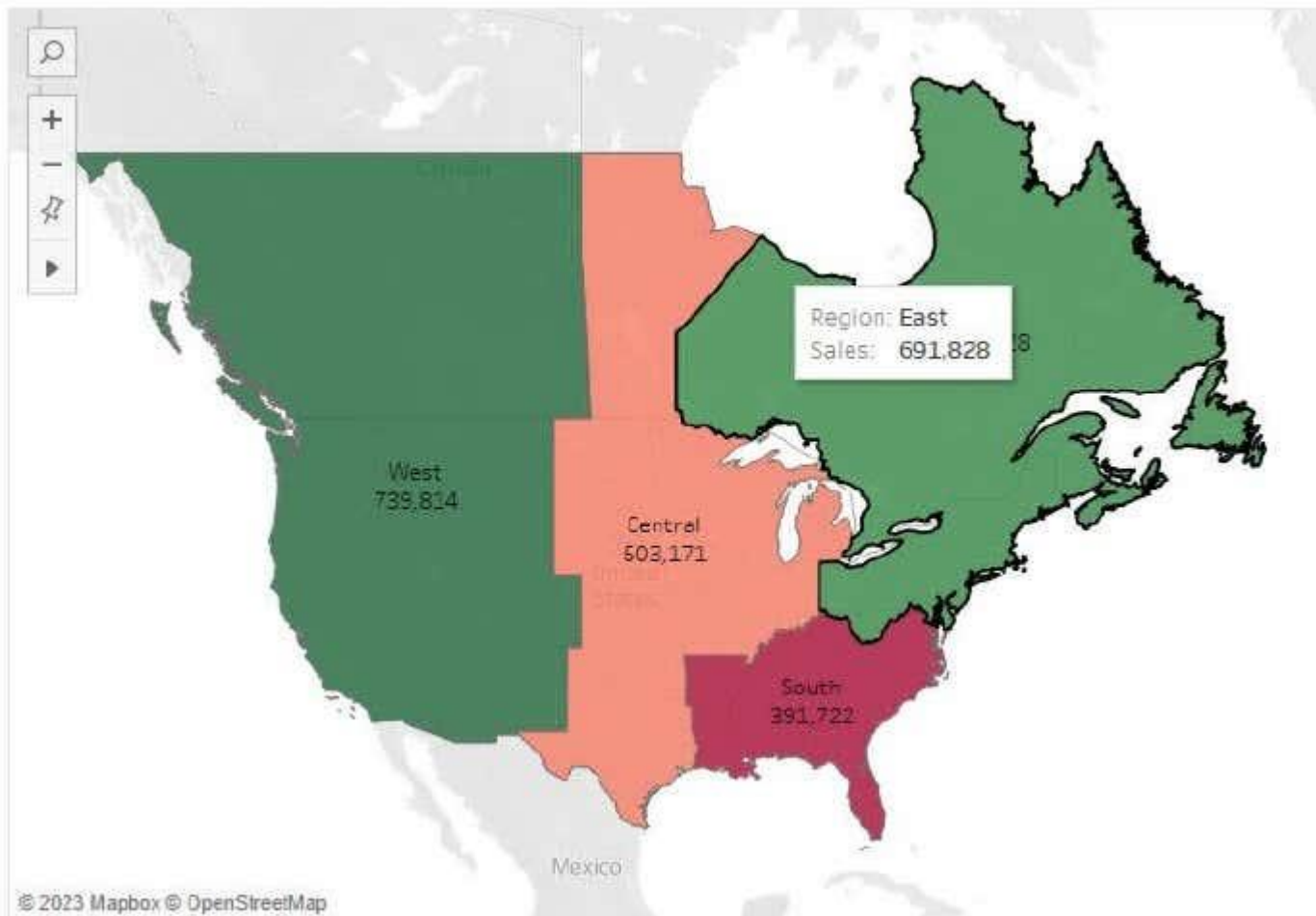
Ans6



●A line chart is a suitable choice for visualizing sales growth over time for various product categories. In this plot, I've segmented the data into subcategories. Notably, Fasteners show the lowest growth, while Phone Chair and Other Stuff exhibit the highest growth rates. This visual representation enables us to track the sales trajectory of different product categories, aiding in identifying trends and making informed business decisions

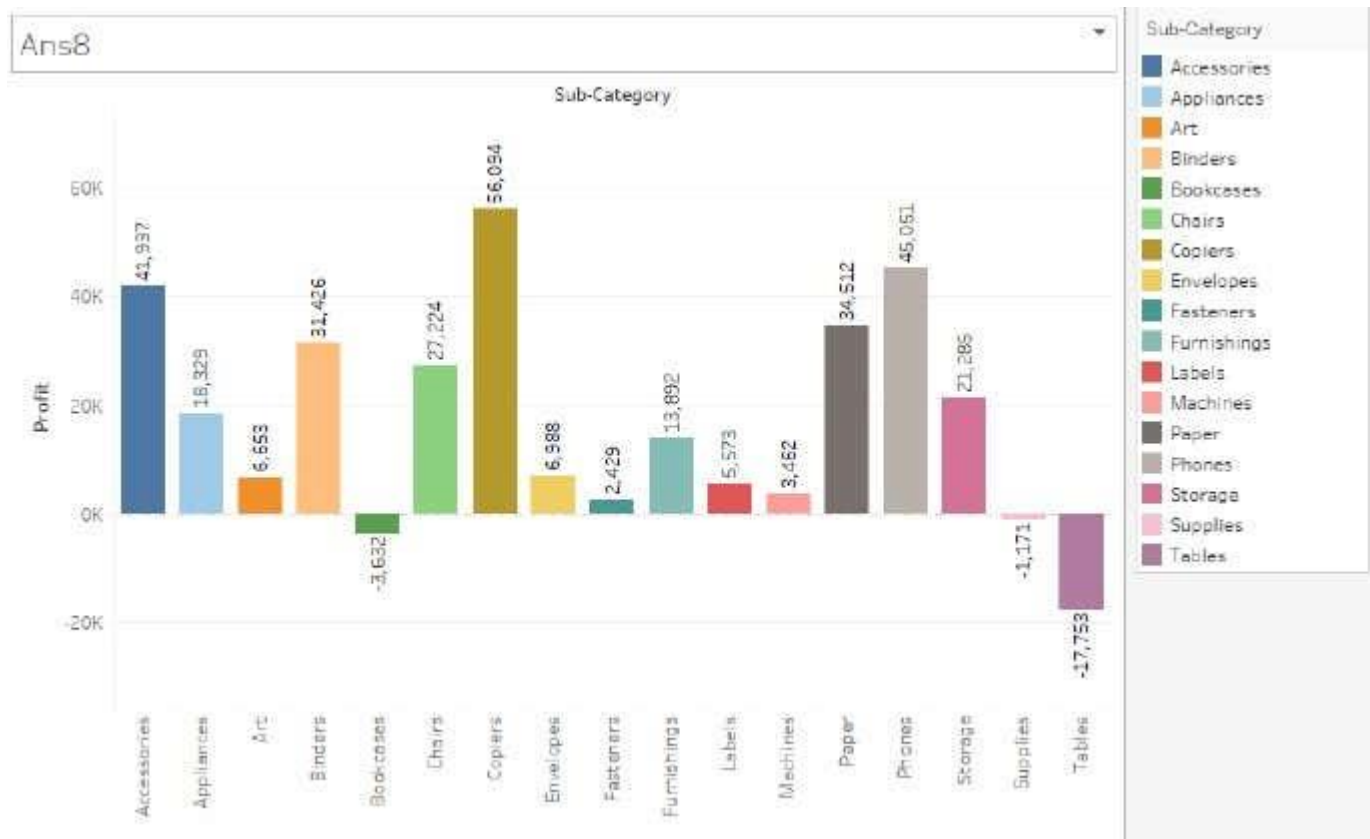
.7.How does the sales distribution vary across different regions in the "Superstore "dataset?

Ans7



The sales distribution across different regions in the "Superstore" dataset can be visualised using a map plot. In this representation, the West region shows the highest sales at \$739, 814, indicating significant growth. In contrast, the South region exhibits the lowest sales at \$391,722, signifying comparatively minimal growth in Superstore product sales across that area.

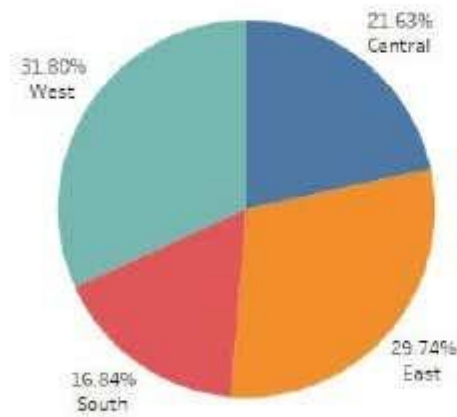
8. Can we visualise the composition of profits across various subcategories with indifferent customer segments?



I can visualize the composition of profits across various subcategories within different customer segments using a bar plot. In this plot, it's easy to discern which product subcategories yield the most profit in customer segmentation. For instance, in this superstore data, "Copiers" generate the highest profit of \$56,094, while "Tables" have a negative profit of -\$17,753, indicating a loss in this subcategory.

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

Ans9

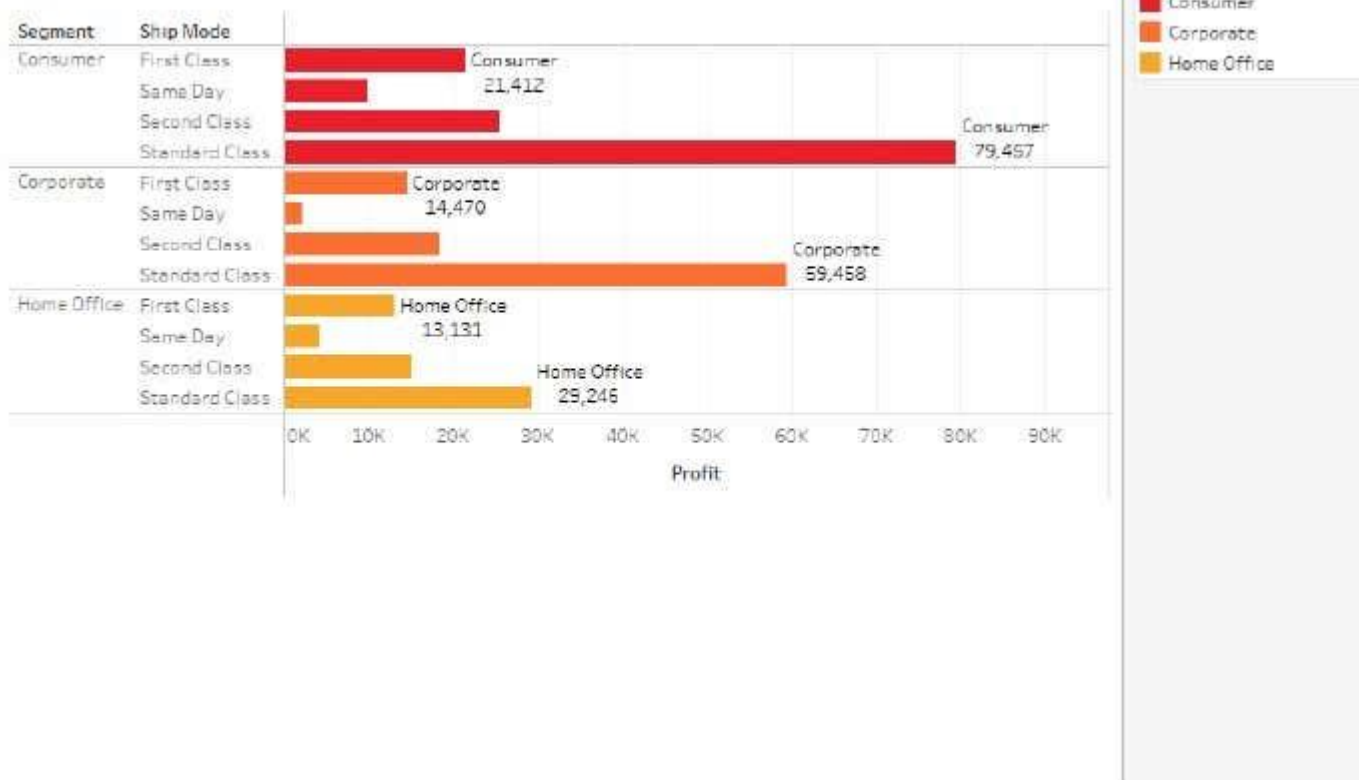


---

To illustrate the percentage contribution of each region to overall sales, a pie plot is selected. In this plot, the West region stands out as the highest contributor with 31.80% of total sales, while the South region has the lowest contribution at 16.84%. This visual representation effectively conveys the sales distribution across different regions.

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

Q10



I can visualize profit margins associated with different shipping modes and customer segments using a column chart. This chart reveals that the "Consumer" segment has the highest profit margins, particularly with "Standard Days" shipping mode. Conversely, "Same Day" shipping shows the lowest margins across all customer segments. Among shipping modes, "Corporate" customers prefer "Standard Class" the most, while "Same Day" is the least preferred, with "Home Appliance" products having similar preferences in the "Corporate" segment.

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



Ans11



I can use a bar plot to visualize the time it takes to process orders for different product categories. In the plot, it's evident that "Furniture" products have the longest processing time, with "Standard Class" taking the most time, while "Same Day" is the quickest. "Office Supplies" and "Technology" categories show similar processing times, mirroring the pattern observed in the "Furniture" category.

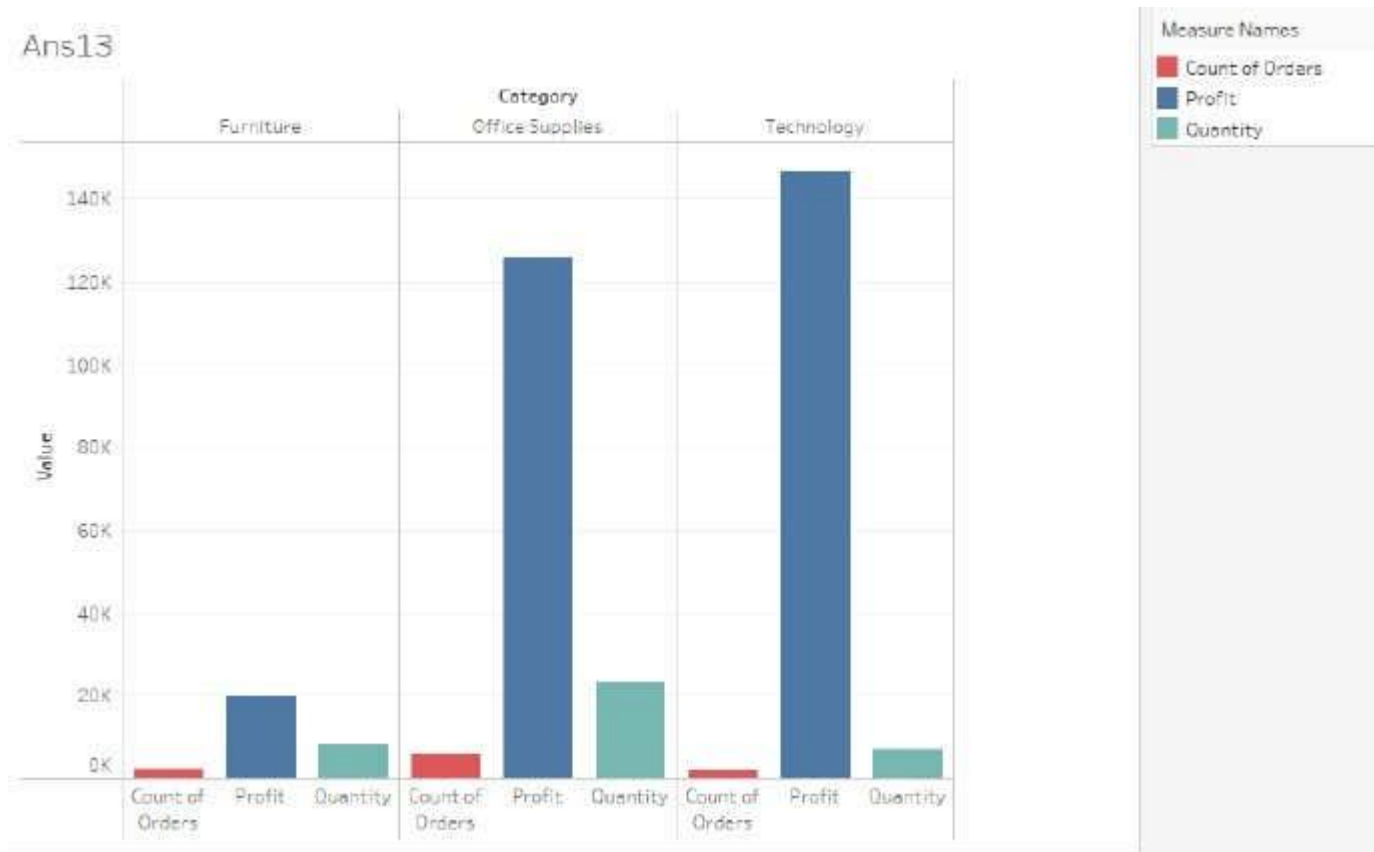
12. How does the performance of different salespeople compare in terms of sales targets, actual sales, and profitability?



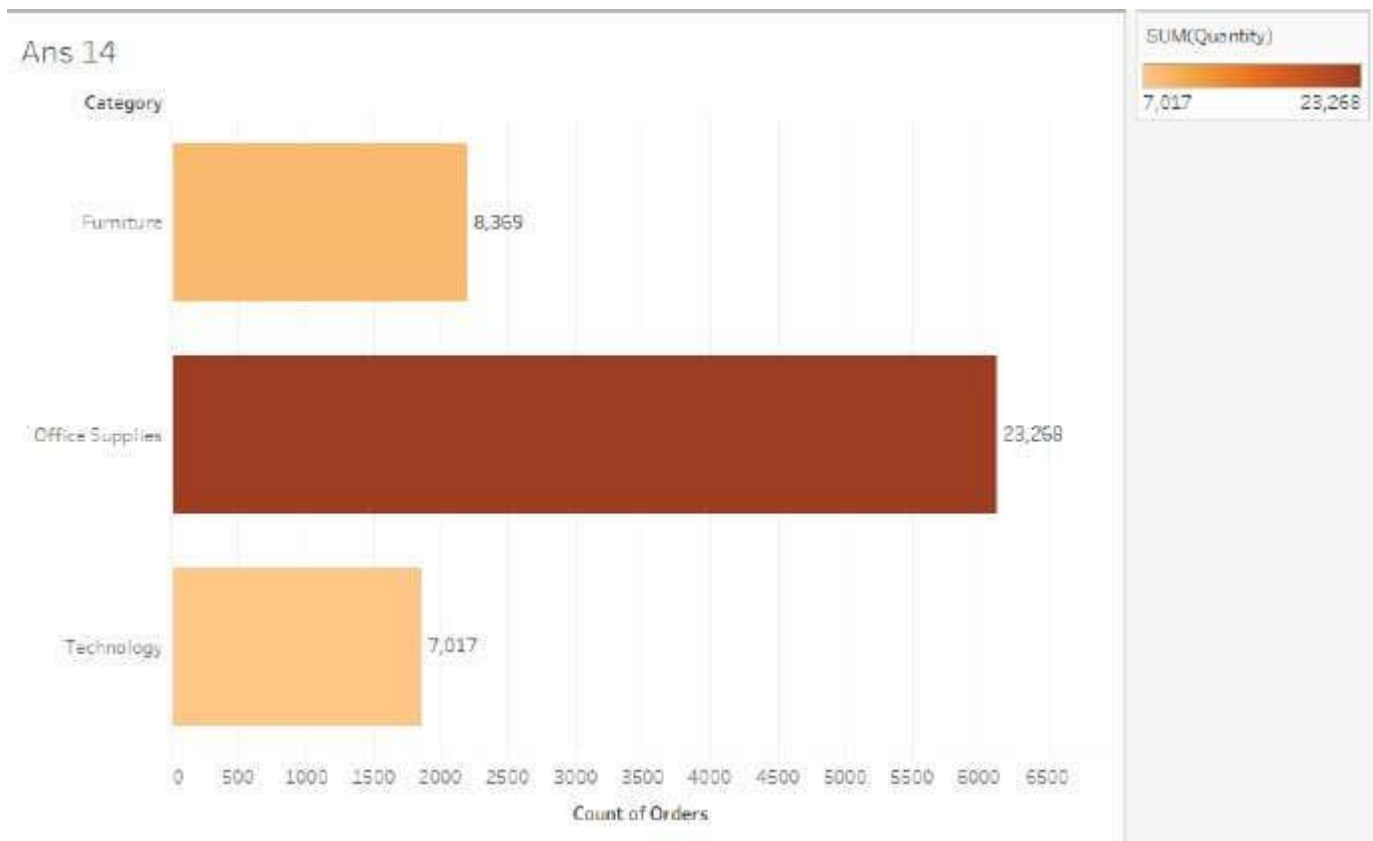
A symbol map of the sales data shows that the salesperson in California is performing the best, with the highest profitability and sales. The salesperson in Ohio is performing the worst, with the lowest profitability and sales. This suggests that there is a significant difference in performance between different salespeople.

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

Ans13

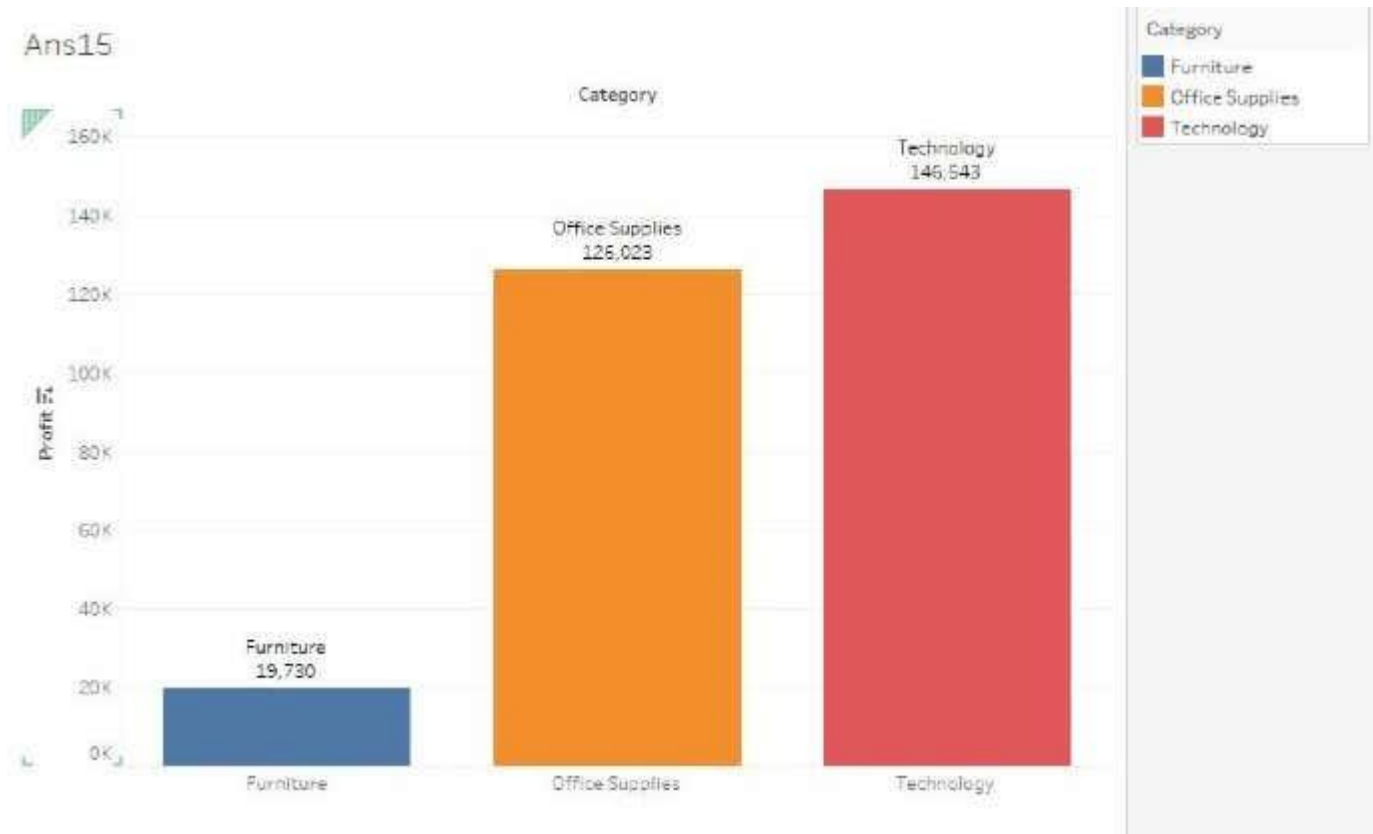


14. What is the distribution of order quantities for products in the dataset?



Certainly, a column chart is a suitable choice for visualizing the time it takes to process orders for different product categories. In this chart, we can observe the distribution of order quantities across categories. For instance, "Office Supplies" has the highest order count at 23,268, while "Technology" products have the lowest order count at 7,017, providing a clear comparison of order processing times across product categories.

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



A bar chart, is an effective choice for visualizing variations in profit distributions across different product categories. In this plot, it's evident that "Technology" products have the highest profit contribution at \$146,543, followed by "Office Supplies" at \$126,023. Conversely, "Furniture" has the lowest profit contribution, with only \$17,730 in profit distribution, showcasing the disparities in profitability among product categories.

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

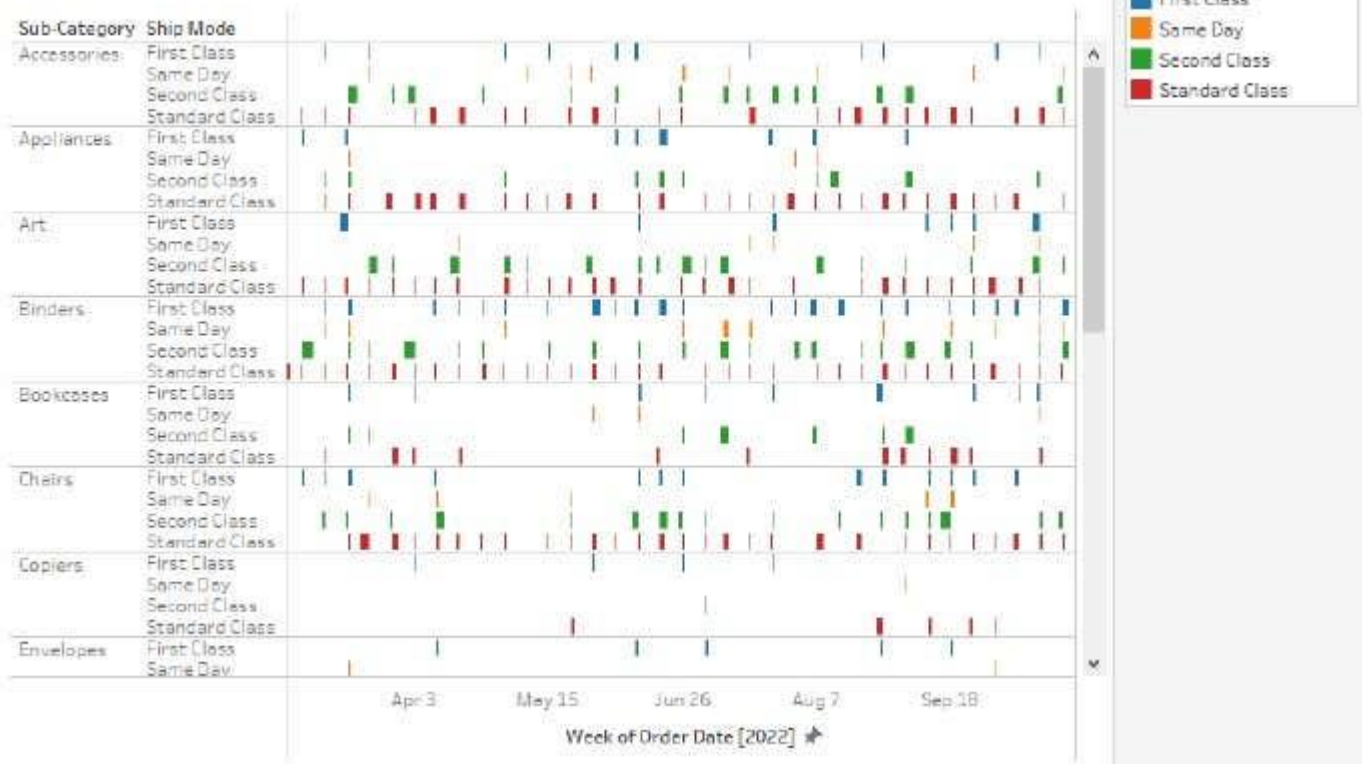
Ans16



A column chart is an effective choice to compare shipping time distributions for different shipping modes. It shows that "Standard Class" is the most preferred shipping mode, while "Same Day" is the least preferred. Additionally, it's notable that "Same Day" shipping is mainly associated with offline store orders, indicating longer processing times compared to other modes.

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

Ans17



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



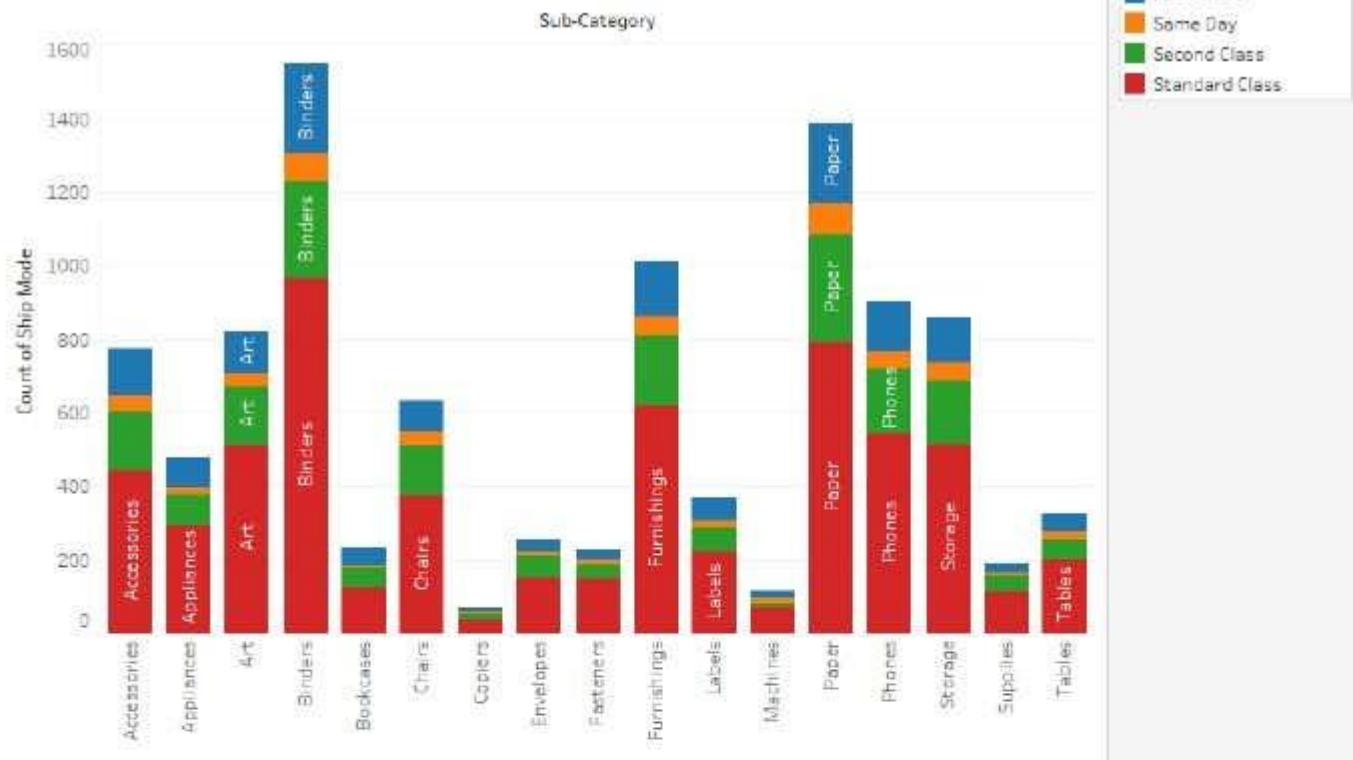
A bar plot is an ideal choice to visualize the performance of different customer segments in terms of sales and discount rates. In this plot, it's evident that the "Consumer" segment receives the highest discounts at 832.3, followed by the "Corporate" segment at 483.5. The "Home Office" segment receives the lowest discounts at 269.1, providing a clear comparison of sales and discount rates across customer segments.

19. How efficiently are different product subcategories being fulfilled in terms of order processing time and on-time delivery?

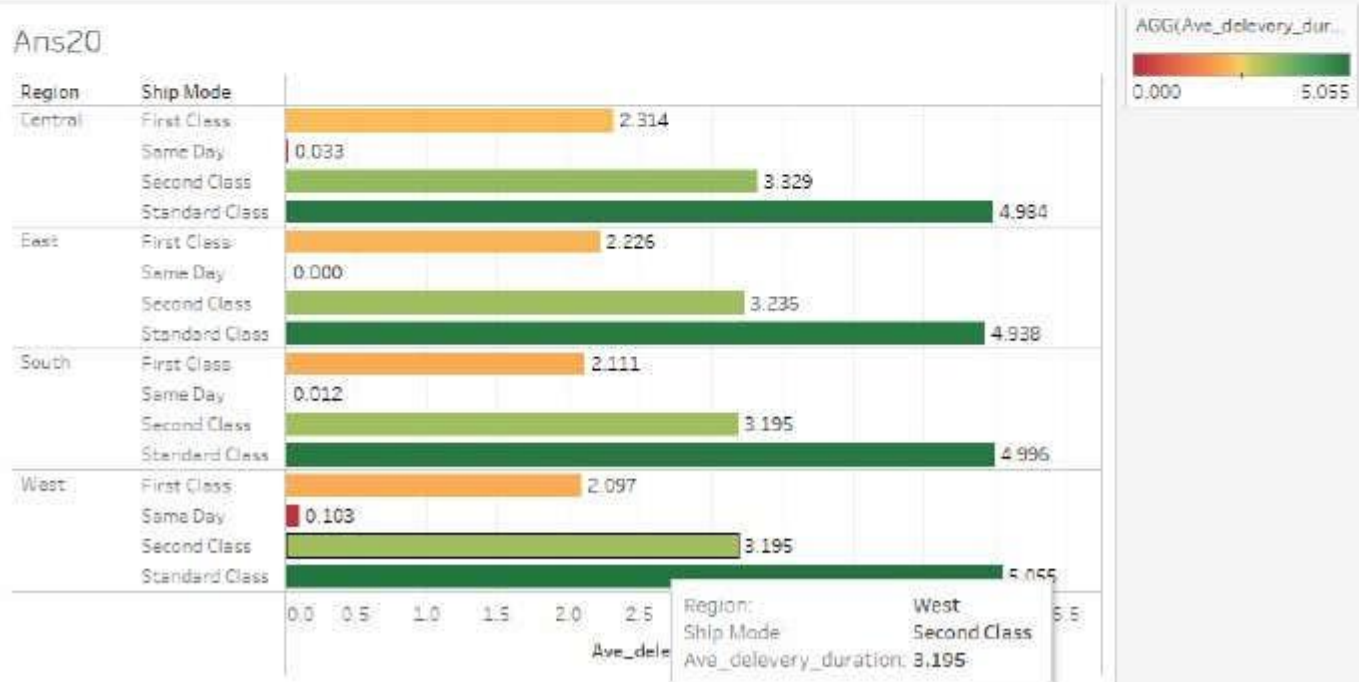
20. What is the average delivery duration for different regions and ship modes?



Ans19



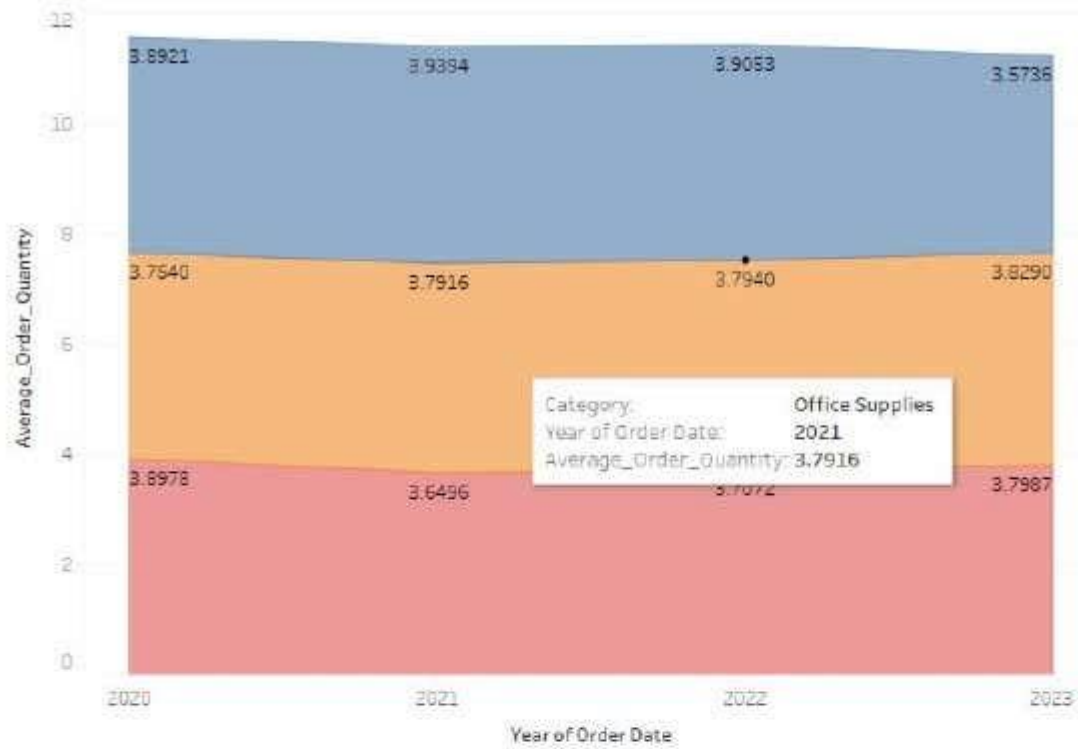
Ans20



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

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

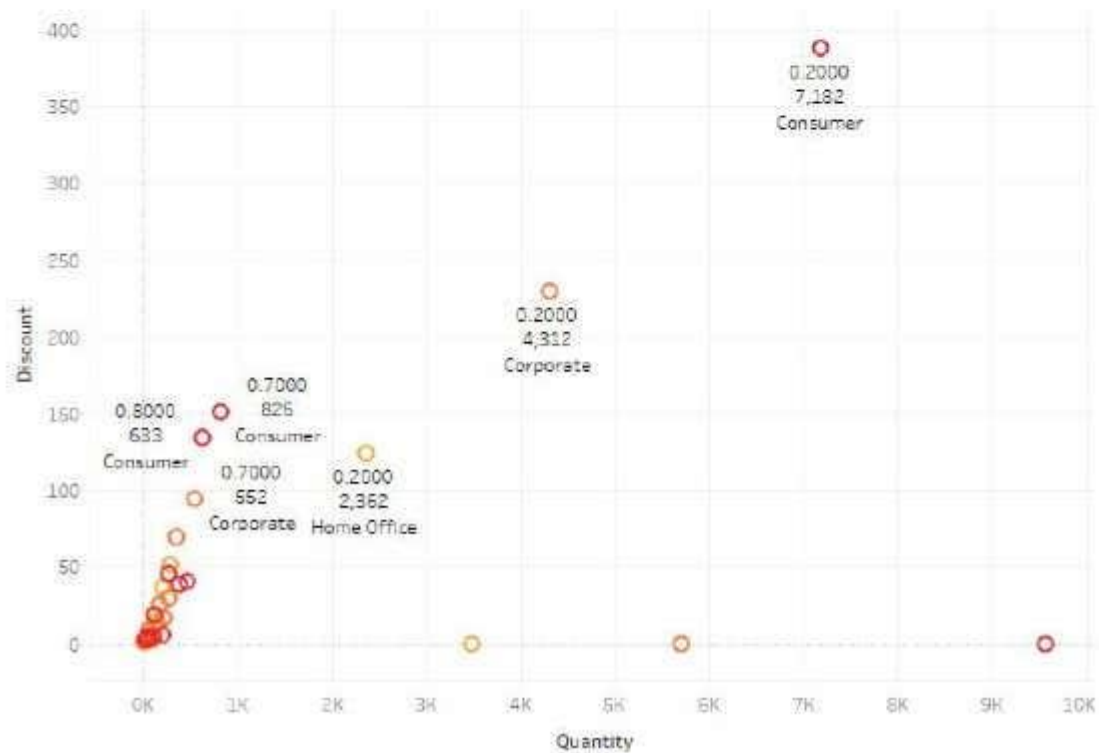
Ans 21



Category

- Furniture
- Office Supplies
- Technology

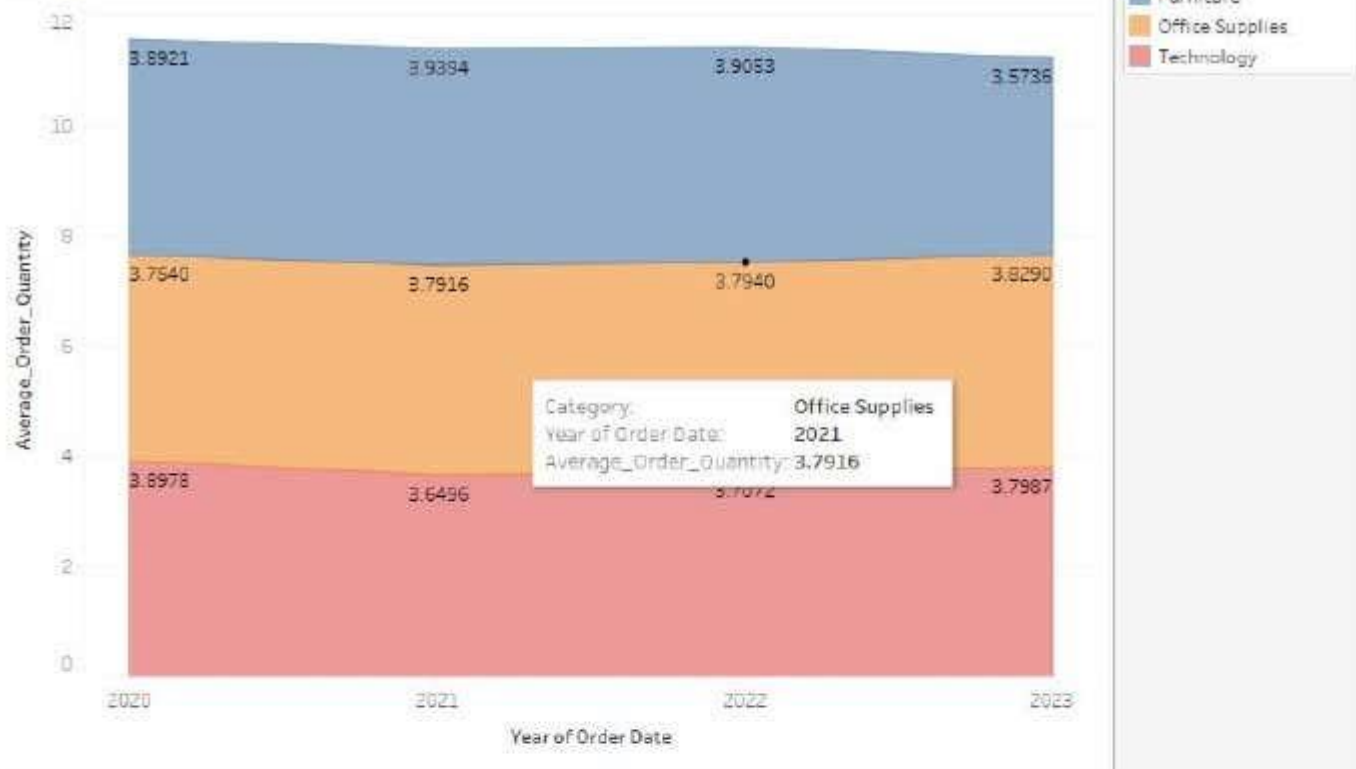
Ans 22



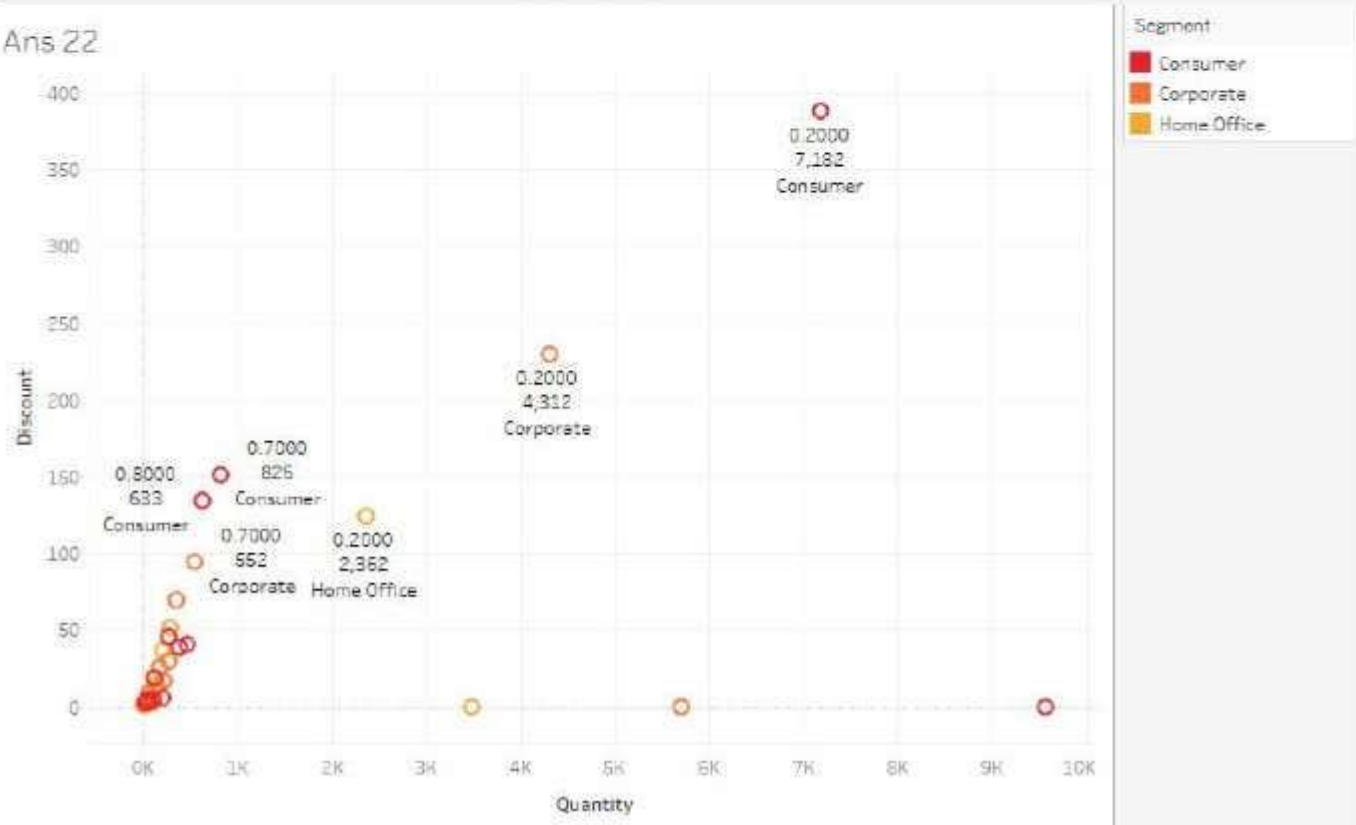
Segment

- Consumer
- Corporate
- Home Office

Ans 21



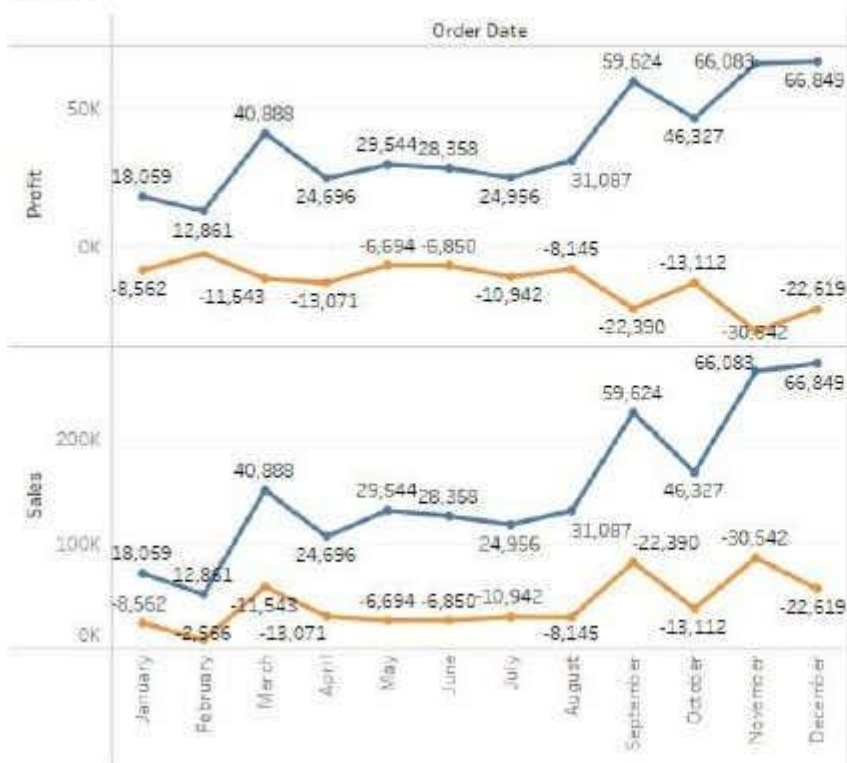
Ans 22



23. What is the trend of returns and refunds across different regions and product categories?

24. How do the sales of high-profit products compare with low-profit products over time?

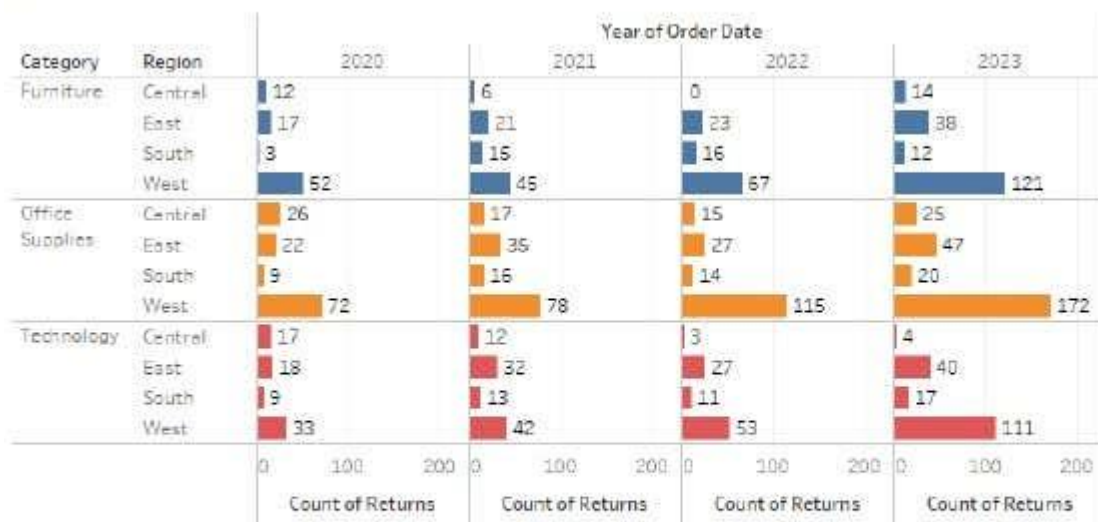
Ans24



profit\_ratio\_of

High-Profit  
Low-Profit

Ans23



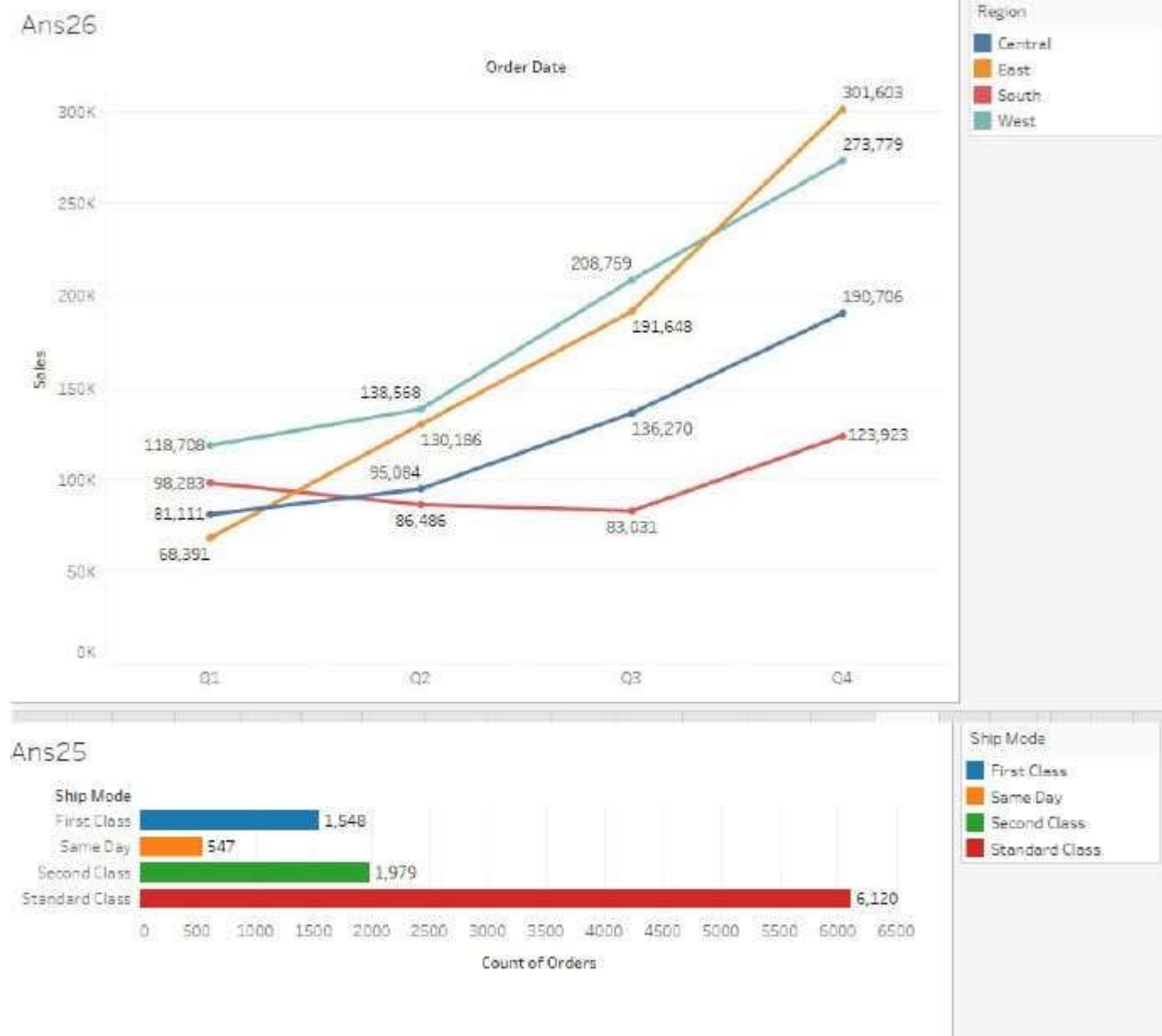
Category

Furniture  
Office Supplies  
Technology

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

A column plot is an appropriate choice to visualize the most commonly used shipping mode in the Sample Superstore dataset. It's evident that the "Standard Class" shipping mode is the most commonly preferred choice, with 6,120 instances, while "Same Day" shipping mode is the least common, with only 547 instances, indicating a clear preference for standard shipping among customers.

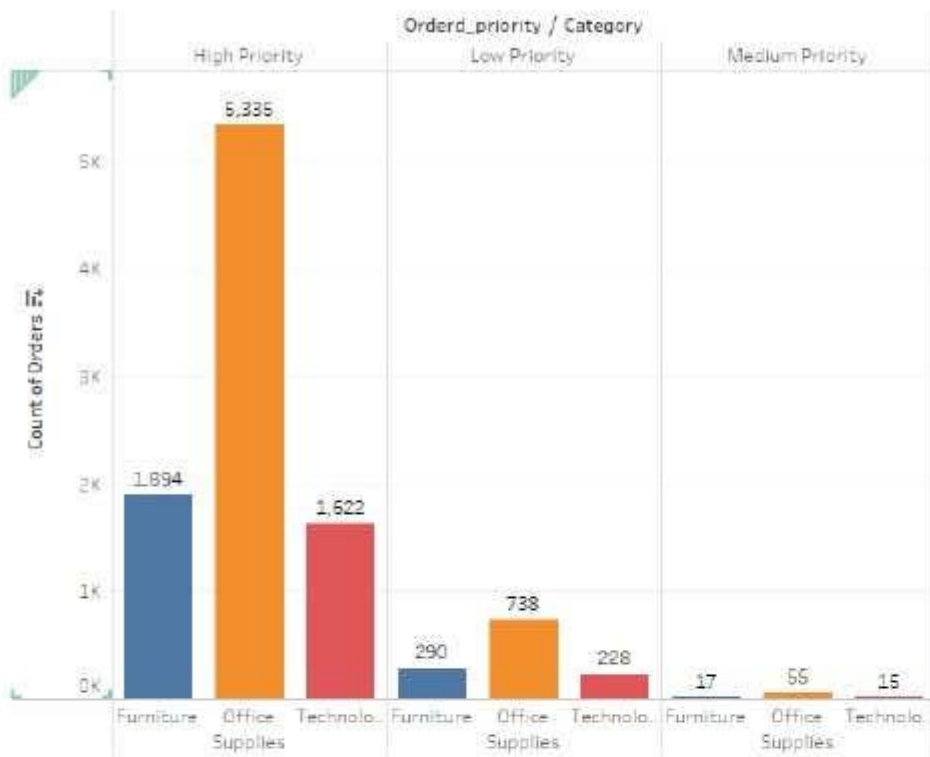
26. How does the sales performance of different regions evolve throughout the quarters of a year?



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

28. Can we visualise the sales conversion rates for different marketing channels and customer segments?

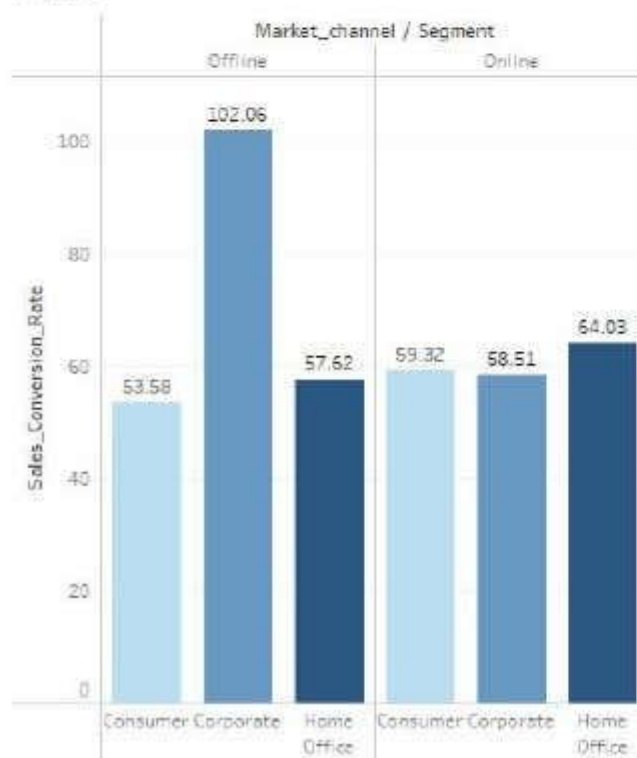
Ans27



Category

- Furniture
- Office Supplies
- Technology

Ans28

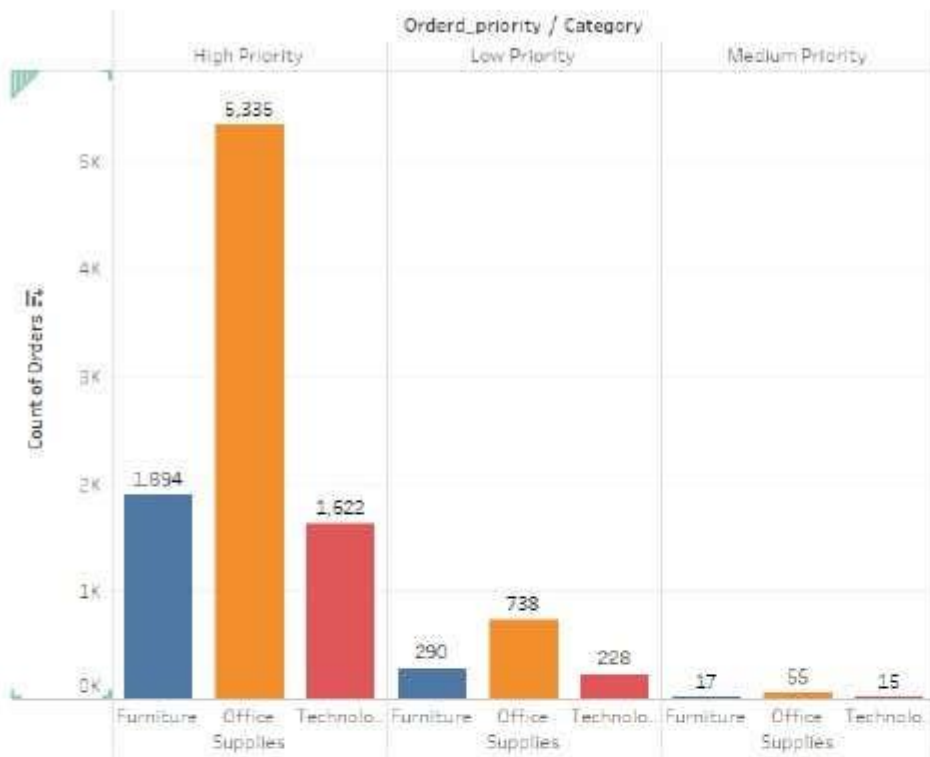


Segment

- Consumer
- Corporate
- Home Office



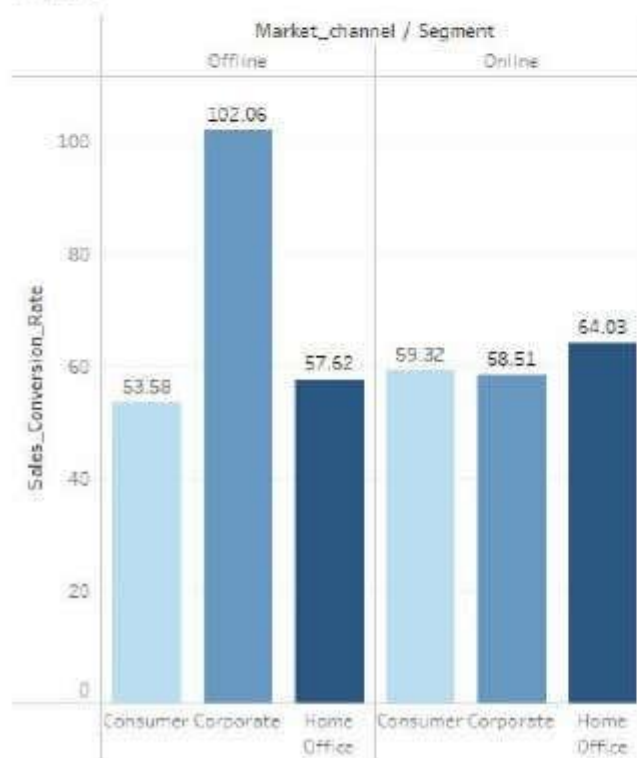
Ans27



Category

- Furniture
- Office Supplies
- Technology

Ans28



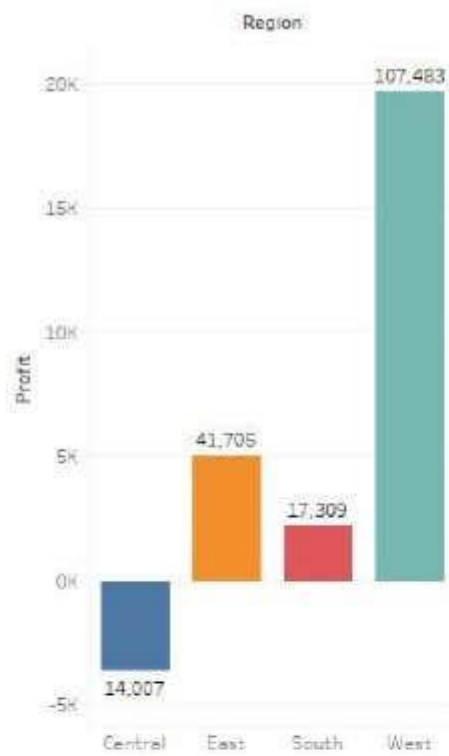
Segment

- Consumer
- Corporate
- Home Office

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

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

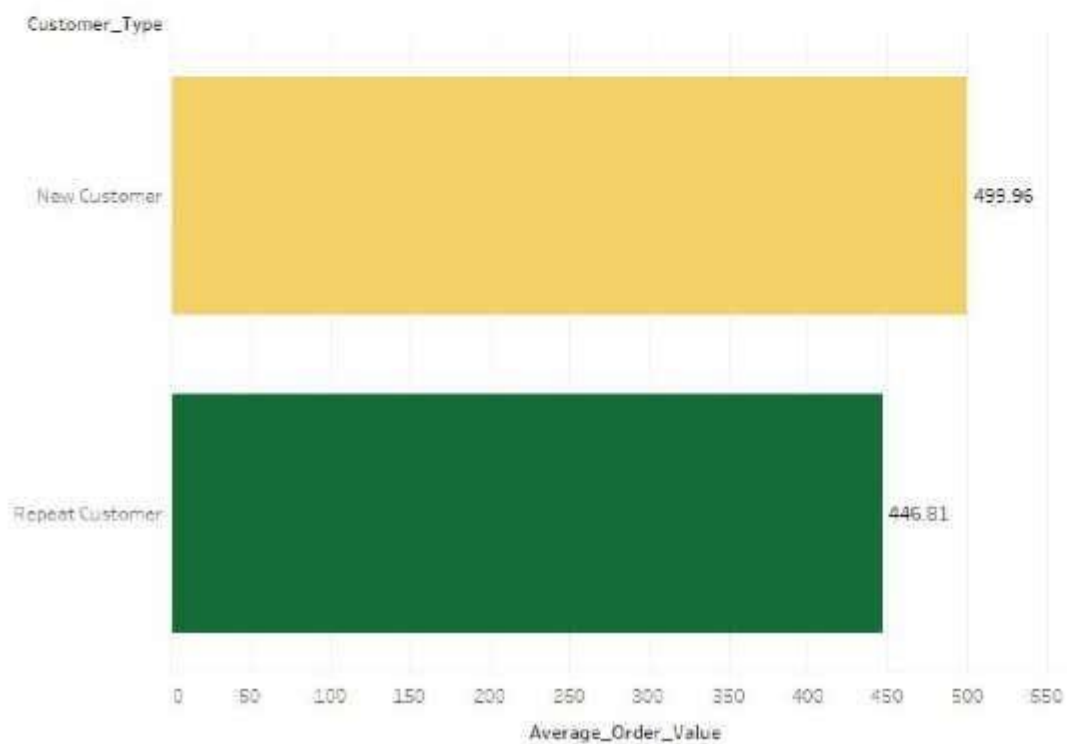
Ans30



Region

- Central
- East
- South
- West

Ans29



Customer\_Type

- New Customer
- Repeat Customer