

# **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:**

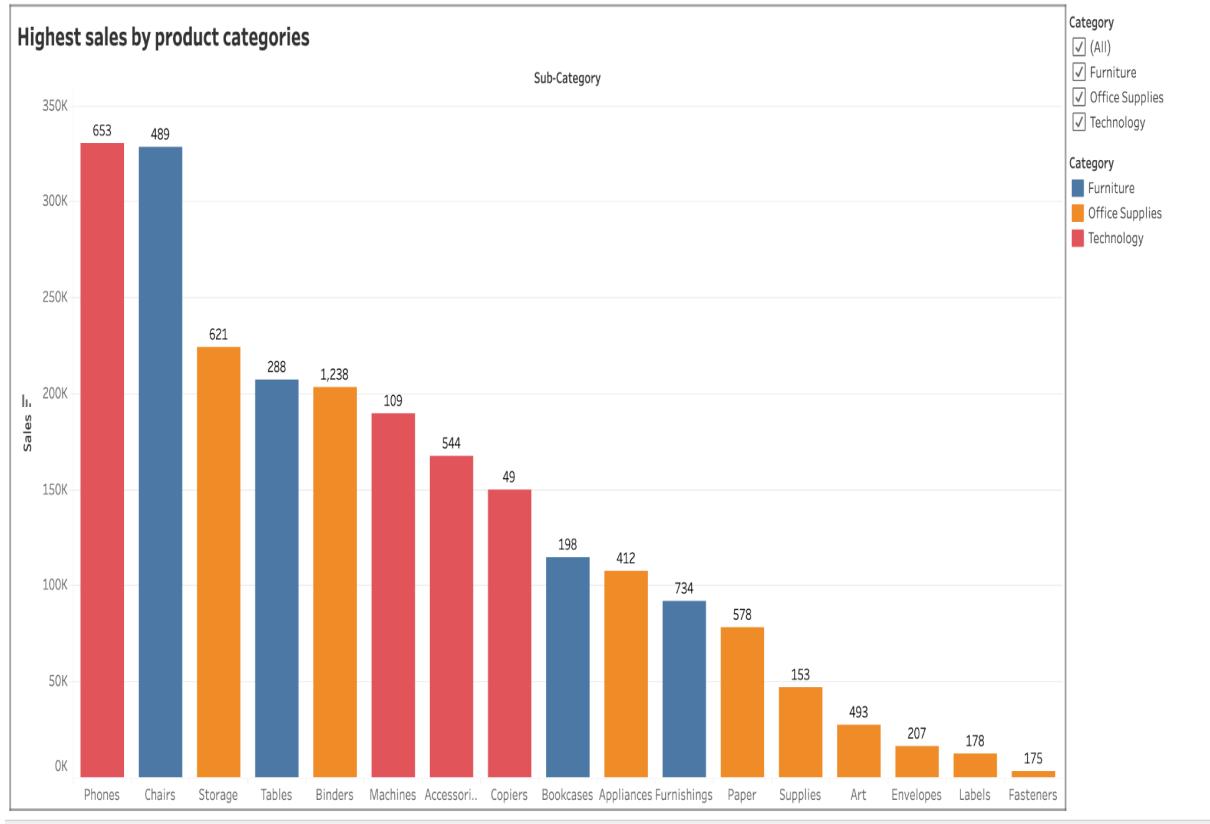
1. **Answer Completion:** Ensure that you furnish answers for all any 30 questions and build charts for them.
2. **Encouraged Creativity:** Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

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?

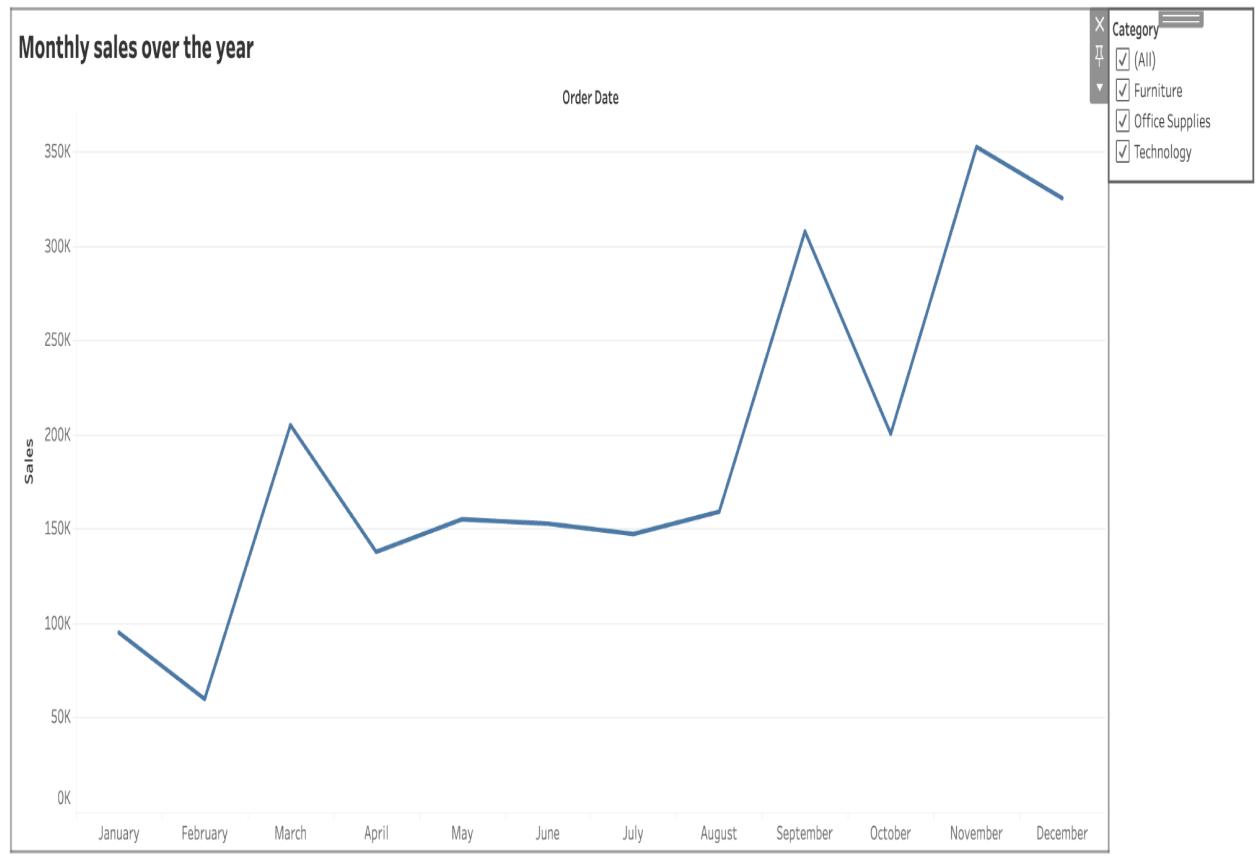


A bar chart would be effective choice for showing comparison between total sales and different product categories because it allows clear and straightforward visualisations of the data.

## Justification

1. **Easy comparison** with product categories. Length of each bar showing total sales for each product categories, it is simple to viewer for identify which product have low sales and high sales.
2. **Categorical data** product categories are categorical data. Which represent By the bars of chart.
3. **Visual clarity** the simplicity of bar shows easily high and low of sales and that audience can grasp easily.

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

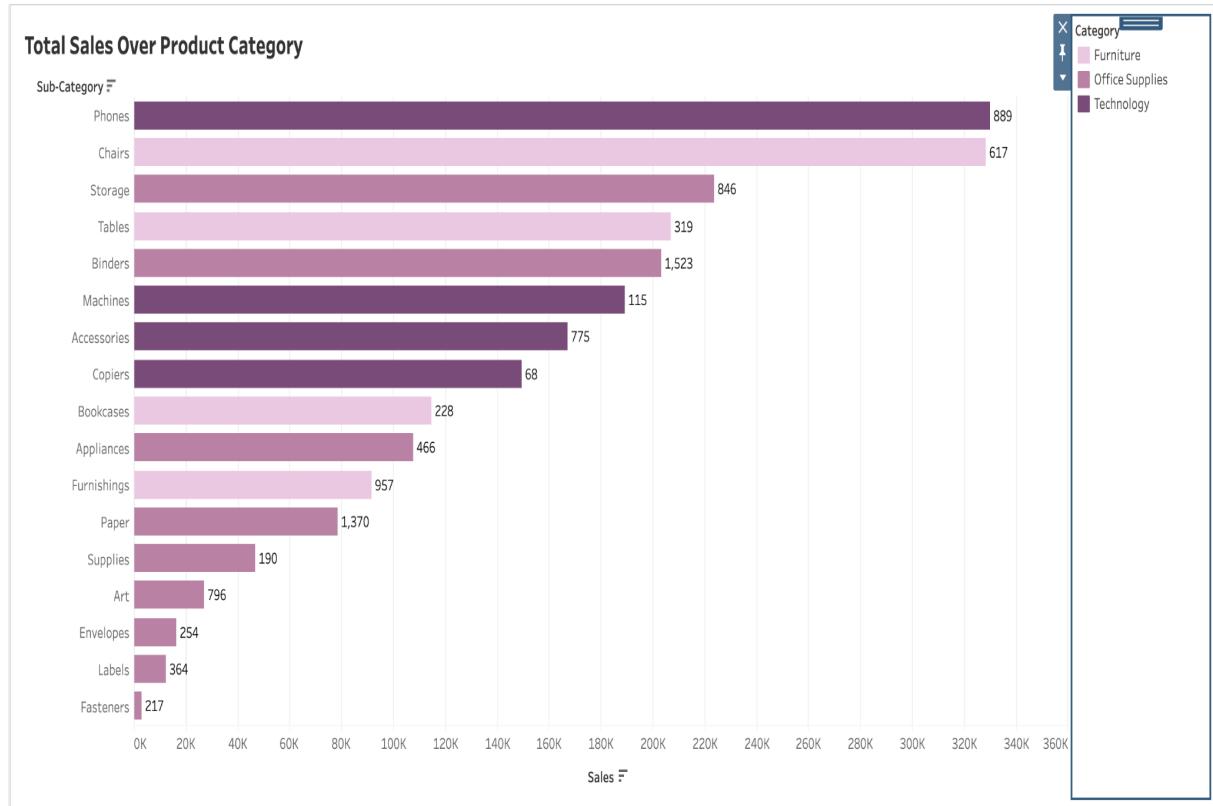


A Line chart is most effective way to show trend over time with monthly  
Because it shows high and low peak of the data according monthly or yearly.

### Justification

- Data time series** line charts particularly use for showing trend over time, a line chart can effectively represent the sequential of data.
- Clear trend analysis** a continuous line charts can show the trend between data and patterns formation. You can see high and low between data and situation when this low and high forming.
- Easy to understand** line chart are easy to understand by the audience. The simplicity of chart allows stakeholders or anyone to understand the data trend and sales over the year.

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

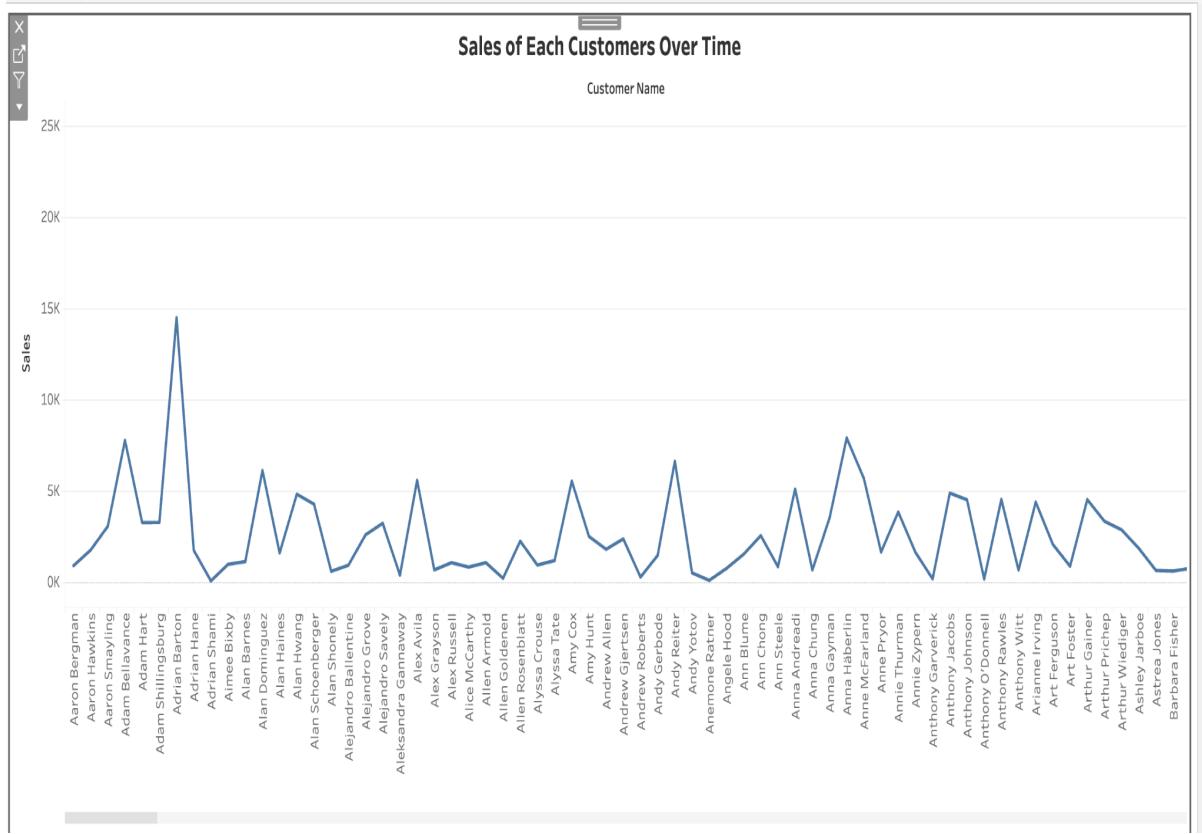


A bar charts is effective for comparison between sales and different category Through we can easily recognise the sales high and low with product information.

## Justification

- Clarity in comparison** bar chart are excellent for comparing product categories representing individuals product categories and sales amount.
- Categorical data representation** Bar charts are well suited for representing and comparing values across different categories.
- Space efficiency** A bar chart are space-efficient and can accommodate a large number of categories without become cluttered.

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

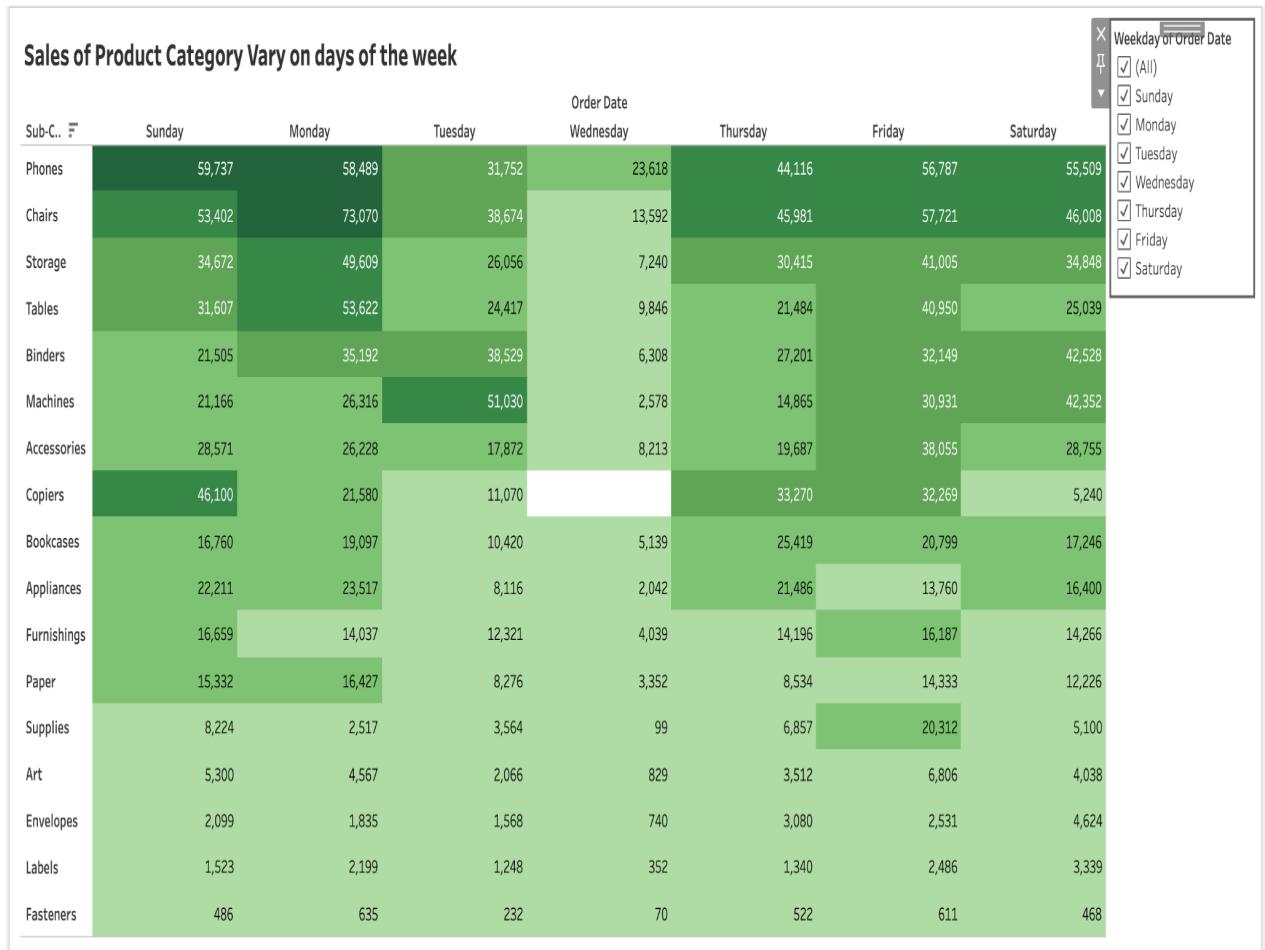


Using line charts to analyse the sales performance of individual customers over time is sensible choice because line chart show the best customer who give us business continuously.

## Justification

- Trends** Line charts are particularly effective for visualising trends over time. You can understand easily patterns, fluctuations, trends in their purchasing behaviour.
- Sequential data Connection** A line chart help to establish a clear connection between different time points.
- Story Telling communications** Line chart are effective tools for storytelling in data analysis, line chart helps to communicate the narrative of how individuals customers have contributed to overall sales growth or decline.

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

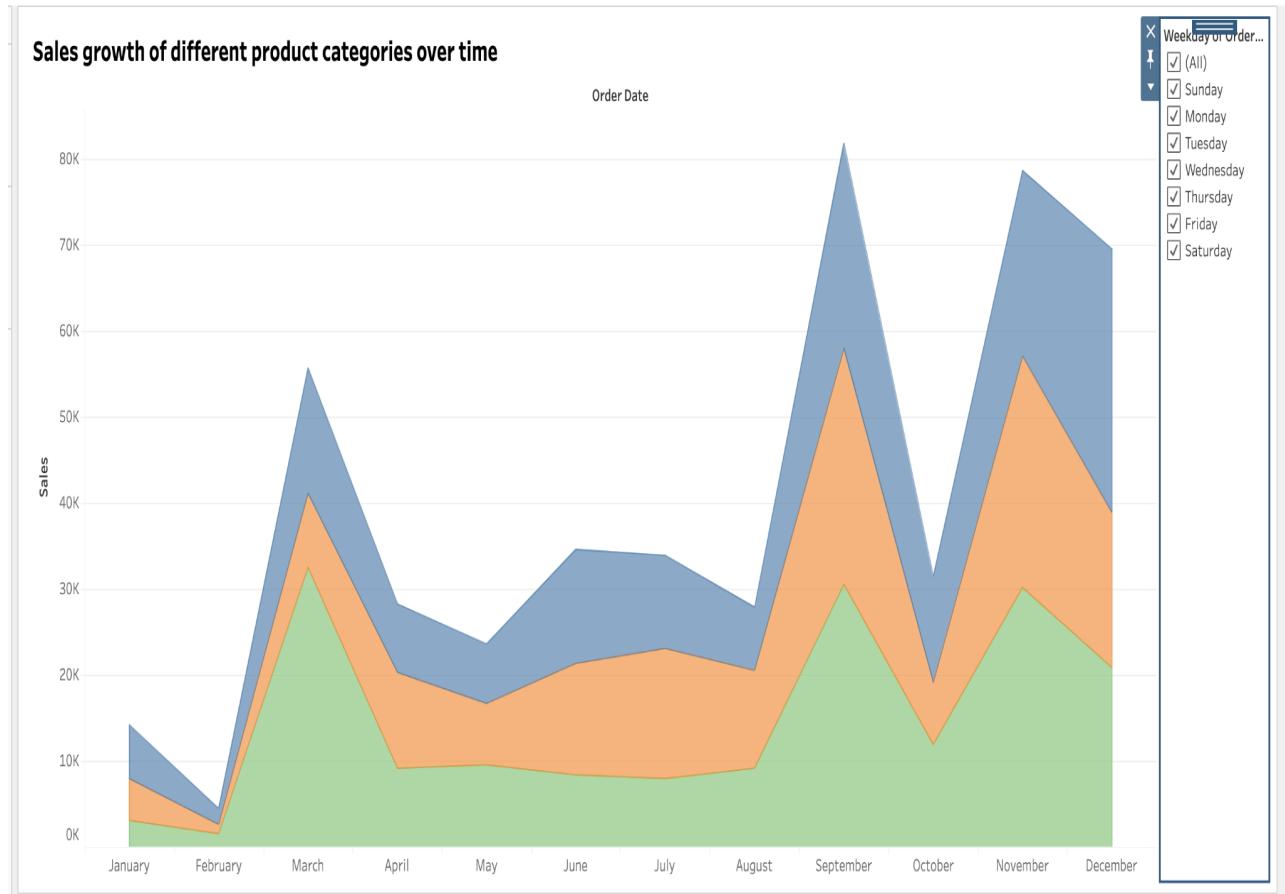


A heatmap is suitable visualisations choice for anyalzing how sales vary days of the week and product categories. It provides an efficient and intuitive way to interpret complex data relationships.

### Justification

- Visulaizinf patterns** Heatmaps are effective for visualisations patterns and variations in dataset. The color gradient make it easy identity the trends, outliers,concentrations of sales.
- Comparative analysis** The heatmaps allows you to compare sales across different days of the week and the product category simultaneously.
- Simplicity and Readability** Heatmaps are visually simple yet powerful. They condense a lot of information into a compact format, making easy to viewers to grasp the overall trends.

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

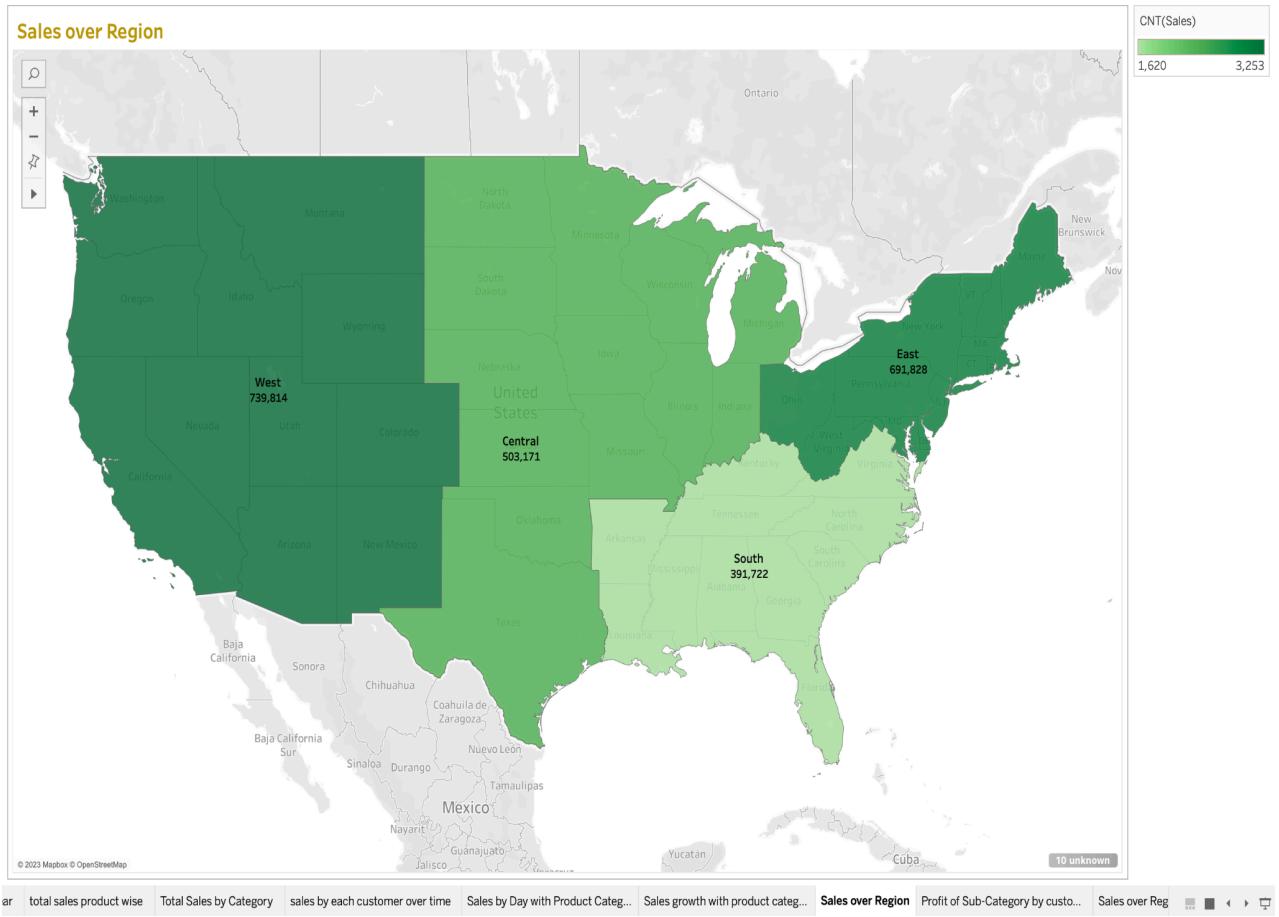


A line chart is a suitable way for visualizing the sales growth of different product categories over time. It effectively communicates temporal trends, allows for comparative analysis, and maintains clarity in presenting the sales growth narrative.

## Justification

- Temporal Trends** Area charts are the suitable way for showing temporal trends into dataset. They can show sales growth or decline for each product over the time.
- Focus on change over time** Area chart are specially designed for the showing changing trends.
- Smooth Continuity** the line charts represent the smooth and continuous representation of sales growth. This continuity helps to tracking how sales evolve for each category.

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

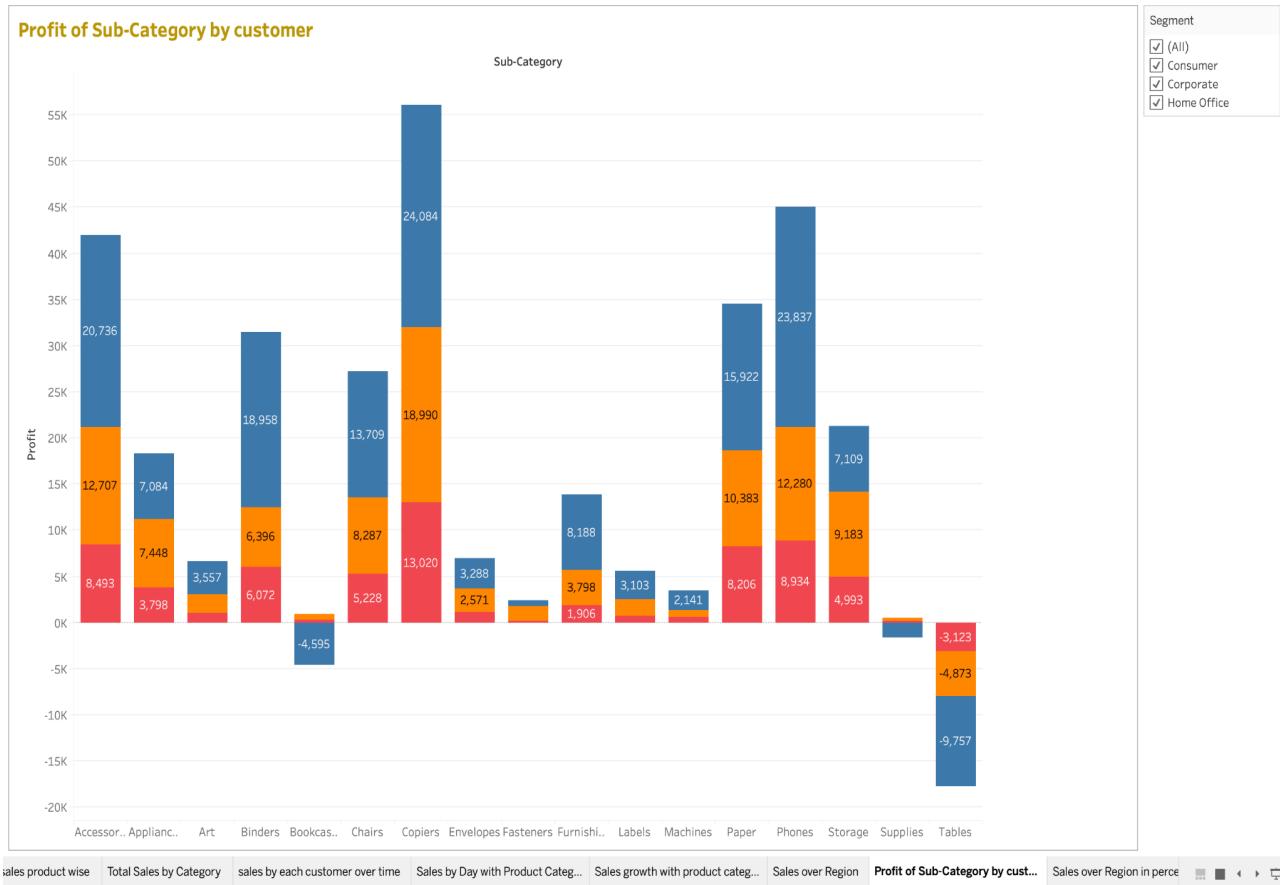


A map chart is indeed a suitable choice for visualizing the sales distribution across different regions reasons are Geographical Representation, Quick Identification of High and Low Performing Regions,Spatial Relationships etc.

## Justification

- 1. Geospatial context** A map charts provides the geospatial context, allowing you to see sales distribution across different region. This is especially useful for analyzing sales patterns in a global, national or regional context.
- 2. Identifying opportunities and Issues** By visualizing sales distribution on map, you can easily identify regions that contribute to sales or areas that may require attention due to lower sales.
- 3. Intuitive Representation** Maps are Intuitive and widely understood. Viewers can easily interpret and understand the sales distribution based on geographical locations.

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

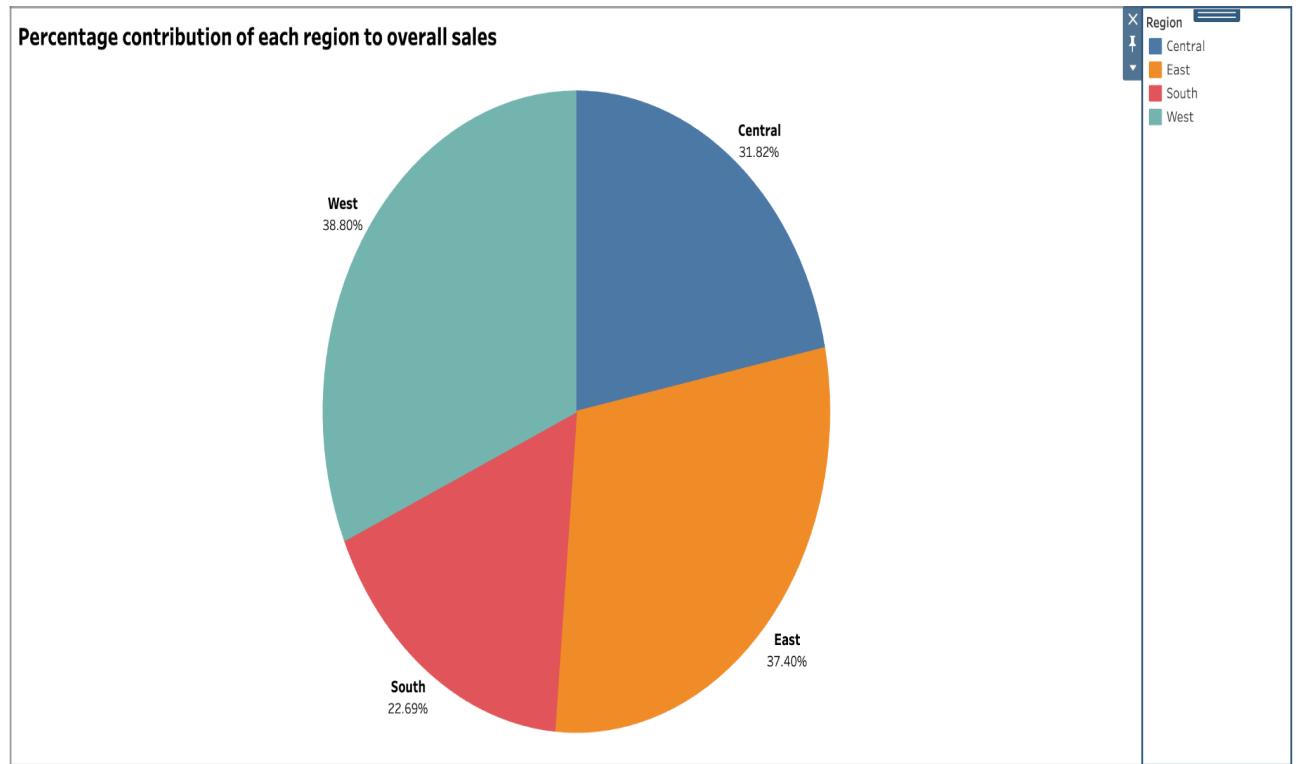


A stacked bar chart is a suitable choice for visualizing the composition of profits across various subcategories within different customer segments. It effectively communicates the distribution of profits and allows for easy comparison and analysis.

## Justification

- Composition Visualization** Stacked bar charts are effective for visualizing the composition of values across different product categories within each segment. Each bar represents a customer segment, and the segments are divided into subcategories.
- Summation Profits** The total height of the stacked bar represents the total profits for a particular customer segment and sub categories of products.
- Color Encoding for Details** Using color to differentiate the subcategories within the stacked bars adds details and makes the visualization appealing.

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

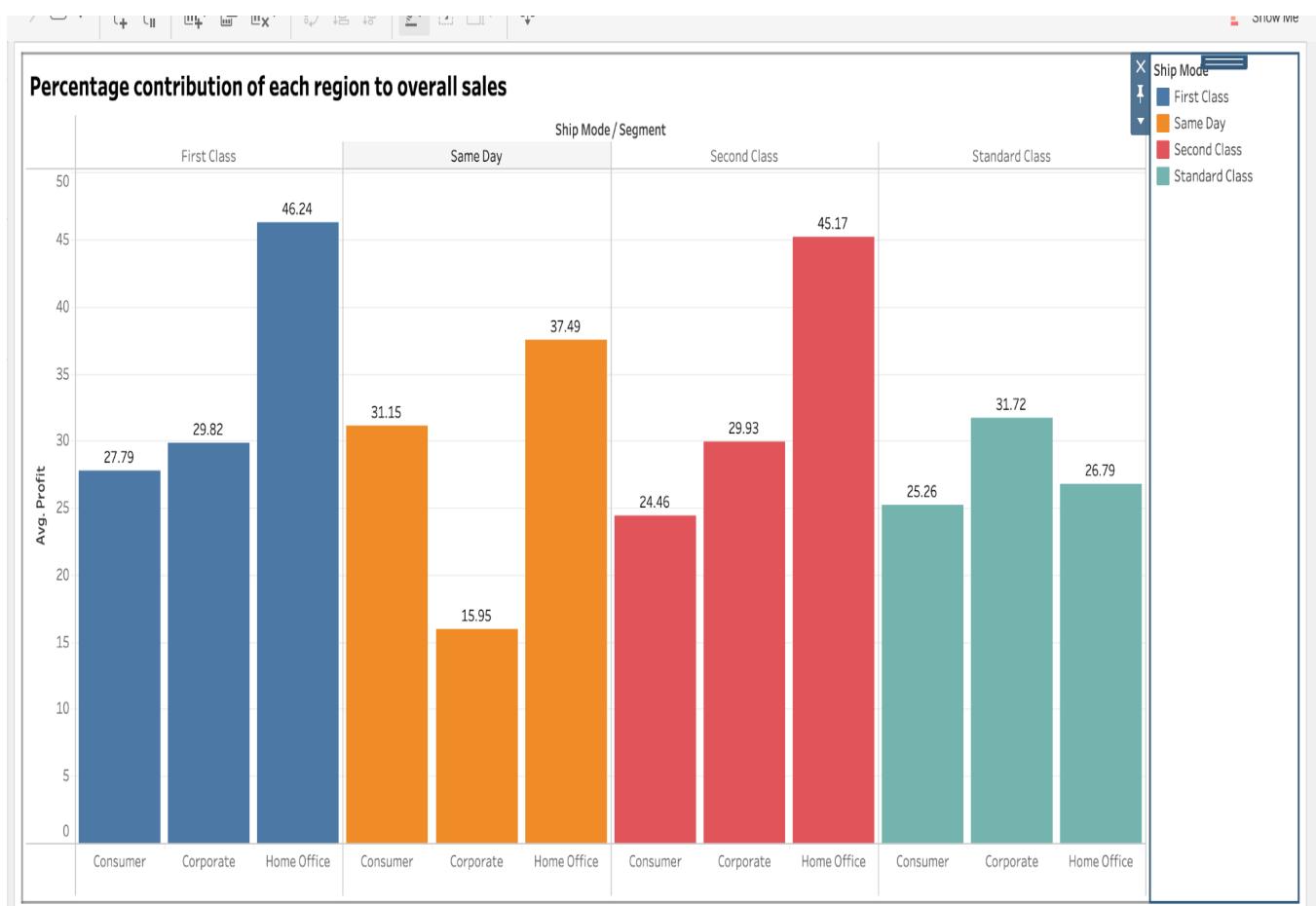


A pie chart is a suitable choice for visualizing the percentage contribution of each region to the overall sales reasons are Percentage Representation, Summarizing Totality, Clear Comparison etc.

### Justification

- 1. Percentage Representation** A pie chart is designed specially to represent the parts of whole as percentages. Each slice of the pie responds to region, and size of slice represent the percentage contribution to the overall sales.
- 2. Clear Overview** A pie charts provides a clear and concise overview of how sales are distributed across different regions. Viewers can easily see the proportion of the sales for each region.
- 3. Single data series** Since you are analyzing the percentage contribution of regions, which is single data series, a pie chart is a straightforward and effective way of conveying this information.

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

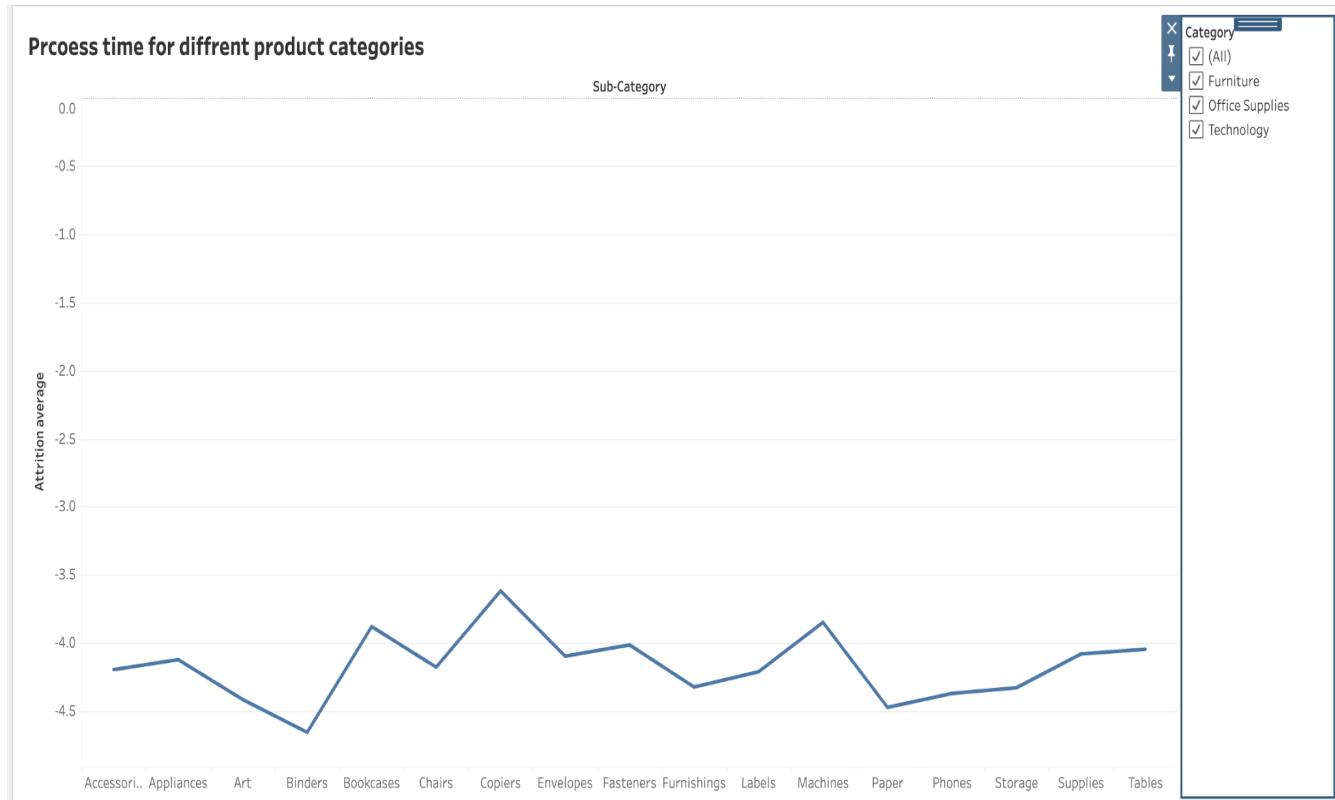


visualizing profit margins associated with different shipping modes and customer segments can be effectively done using a color-coded chart as reasons are Easy Differentiation.

## Justification

- Comparison Across Categories:** Grouped bar charts are effective for comparing values (profit margins) across different categories (shipping modes) within each subgroup (customer segment). Each group of bars represents a customer segment, and the bars within each group represent different shipping modes.
- Visualizing Subgroups:** This type of chart is well-suited for visualizing subgroups (customer segments) within larger categories. The grouping allows viewers to quickly compare profit margins across shipping modes for each customer segment.
- Color Encoding for Detail:** Using color to differentiate shipping modes within the grouped bars adds detail and makes it visually appealing.

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

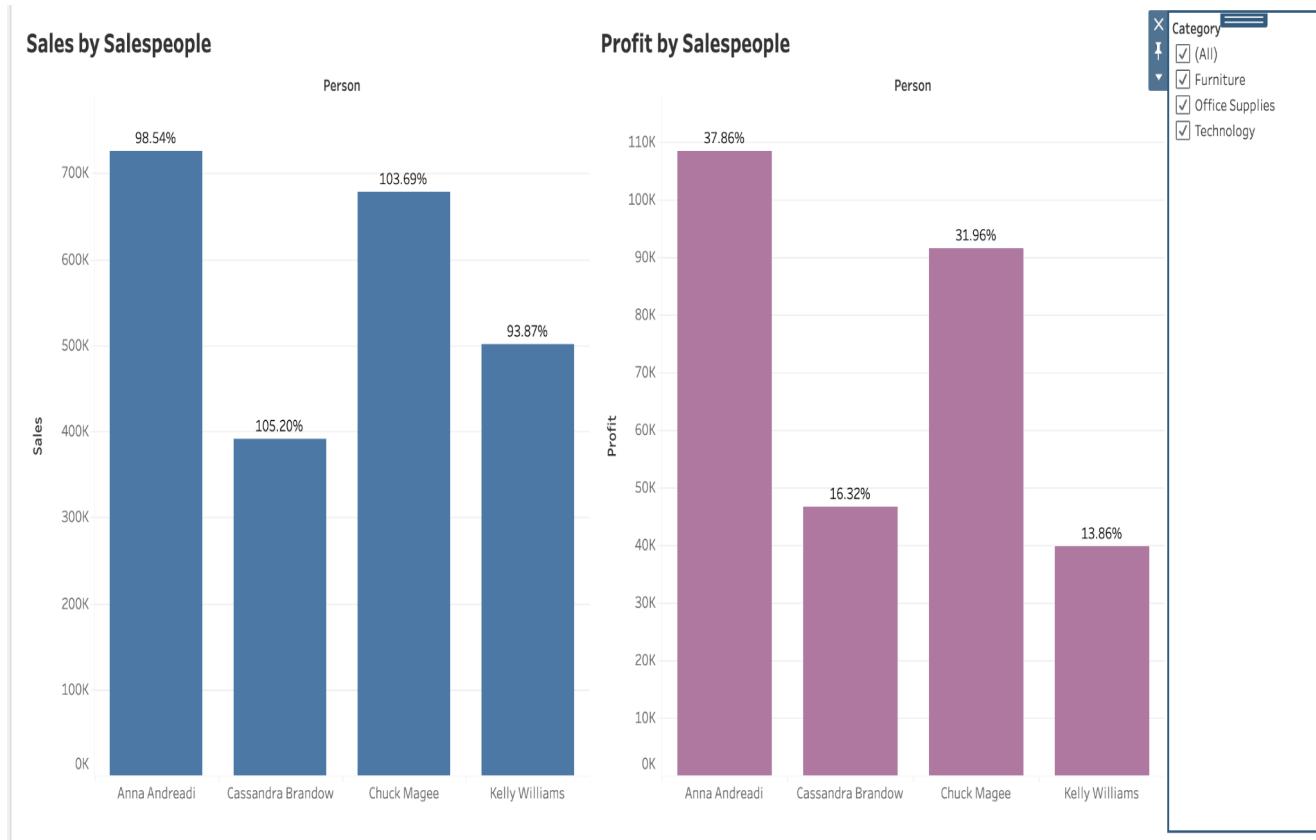


Creating a line chart to visualize the processing times for different product categories can be an effective way to convey trends and variations over time.

## Justification

- Time-Series Data** Line charts are particularly well-suited for visualizing time-series data, where the data points are ordered by time. In this case, you're interested in understanding how the processing times for different product categories change over time.
- Highlighting Variations** Variations in processing times within each product category can be effectively highlighted using a line chart. For example, you may observe spikes or dips in processing times
- Facilitating Communication** Line charts are easy to interpret and are commonly used in business settings for presenting trends and performance metrics. This makes them suitable for communicating insights to a diverse audience, including stakeholders who may not be familiar with complex data visualizations.

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

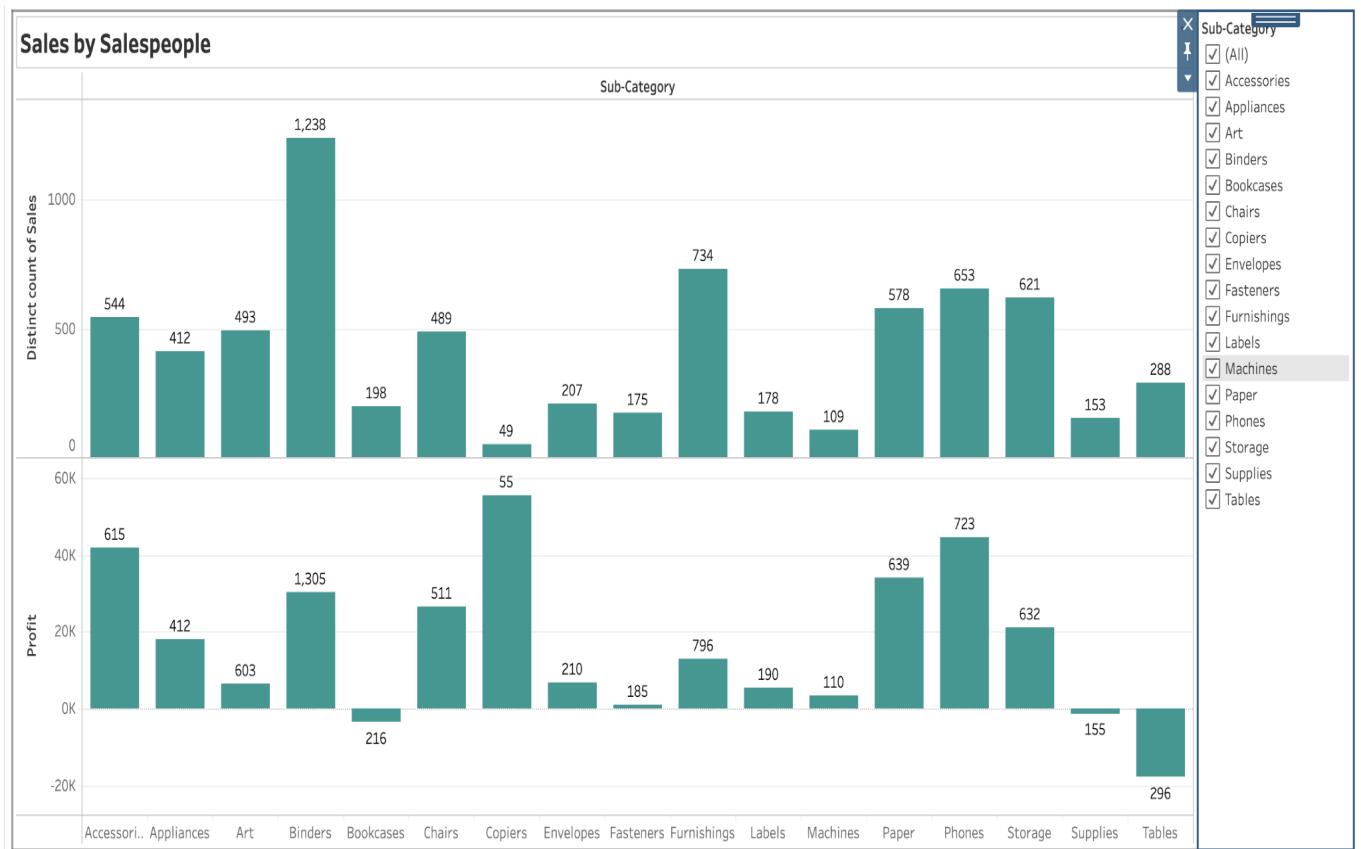


Using a bar graph to compare the performance of different salespeople in terms of actual sales and profitability is a sound choice. The visual nature of the bar graph simplifies complex data, making it easily understandable for a broad audience.

## Justification

- 1. Visual Comparison** Bar charts and similar visualizations are effective for comparing values across categories, making them suitable for comparing actual sales and profitability for different salespeople.
- 2. Customization for Clarity** Customizing the chart improves clarity and makes it easier for stakeholders to interpret the information. Clearly labeled axes and a well-formatted chart contribute to effective communication.
- 3. Grouping and Sorting** Grouping and sorting help in identifying top performers and areas for improvement quickly. This enhances the usability of the visualization for decision-makers.

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

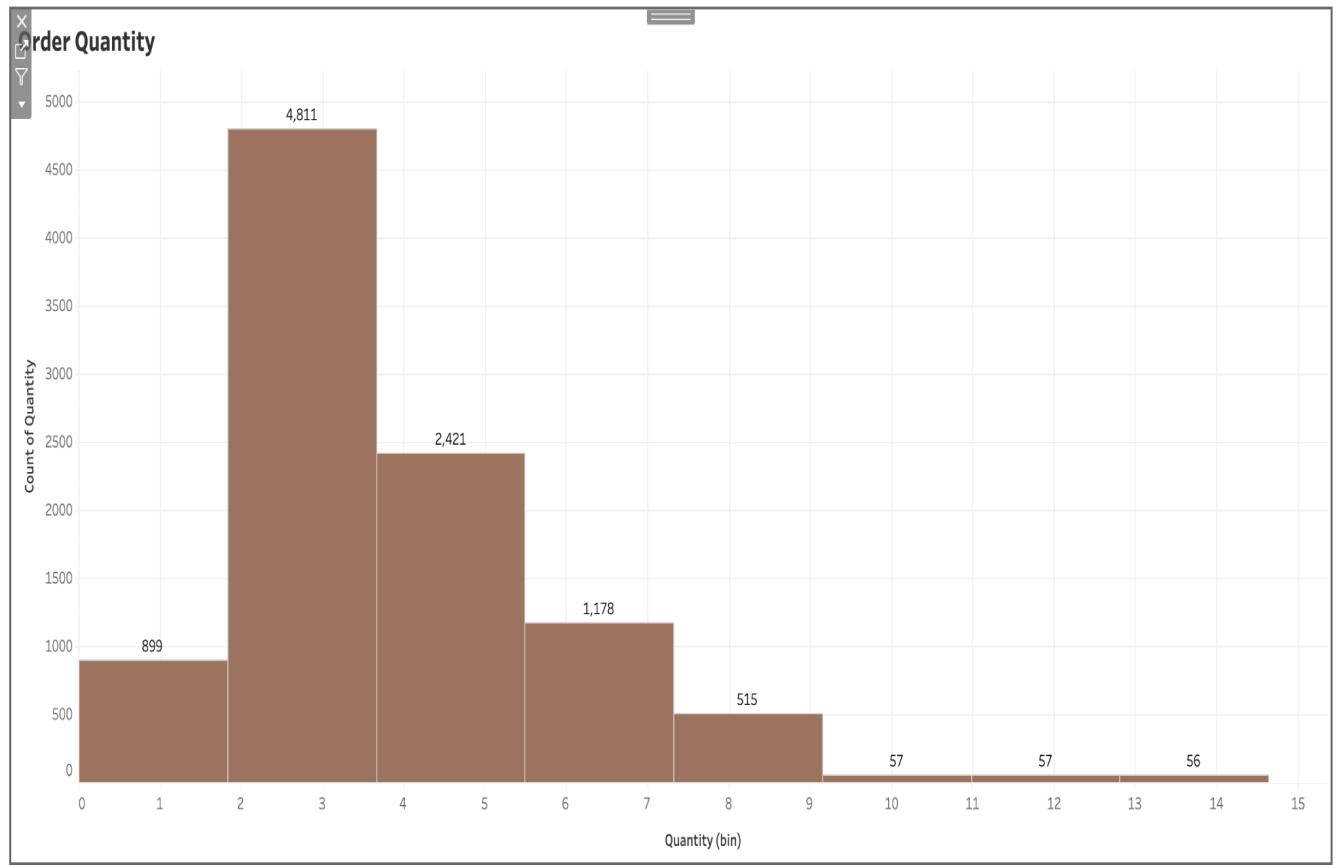


Using a bar chart for Visualizing the relationship between product sales and profitability for different product categories is a common and effective practice in business analysis.

## Justification

- Clear Comparison** Bar charts make it easy to compare values across different categories. each product category represented by a separate bar.
- Categorical Data Representation** Bar charts are particularly suitable for representing categorical data, where each category is distinct, each product category is a distinct entity, making a bar chart an appropriate choice.
- Ease of Interpretation** Bar charts are intuitive and widely understood. Viewers can quickly grasp the information presented without the need for complex explanations.

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

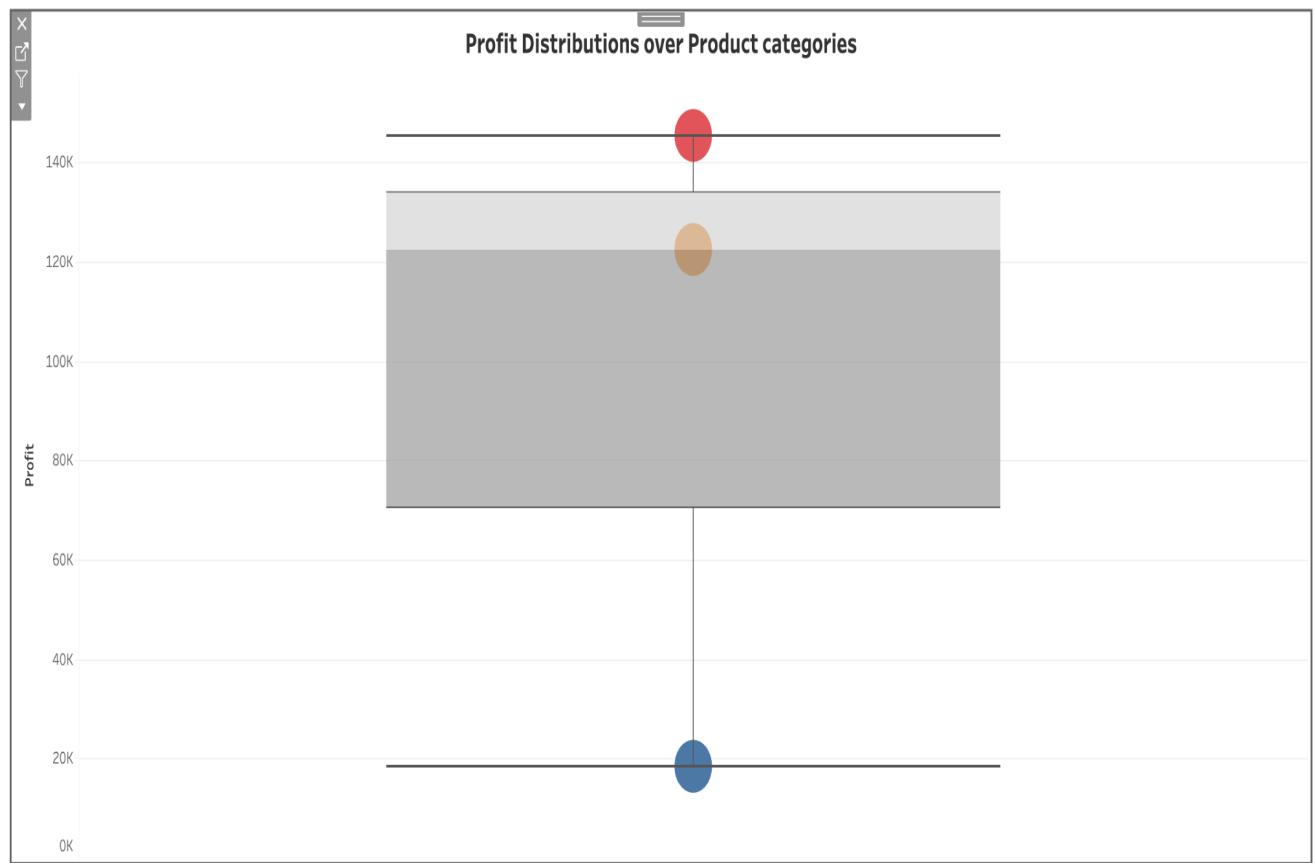


A histogram is an excellent choice for visualizing the distribution of order quantities for products in a dataset.

## Justification

- Frequency Distribution** Histograms are specifically designed to display the frequency distribution of a dataset. different order quantity ranges (bins) on the x-axis and the frequency (or count) of products falling into each bin on the y-axis.
- Visualization of Patterns** Histograms allow you to identify patterns and trends in the distribution of order quantities.
- Central Tendency and Spread** the shape of the histogram gives sense of the central tendency (mean or median) and the spread (variability) of the order quantities.

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

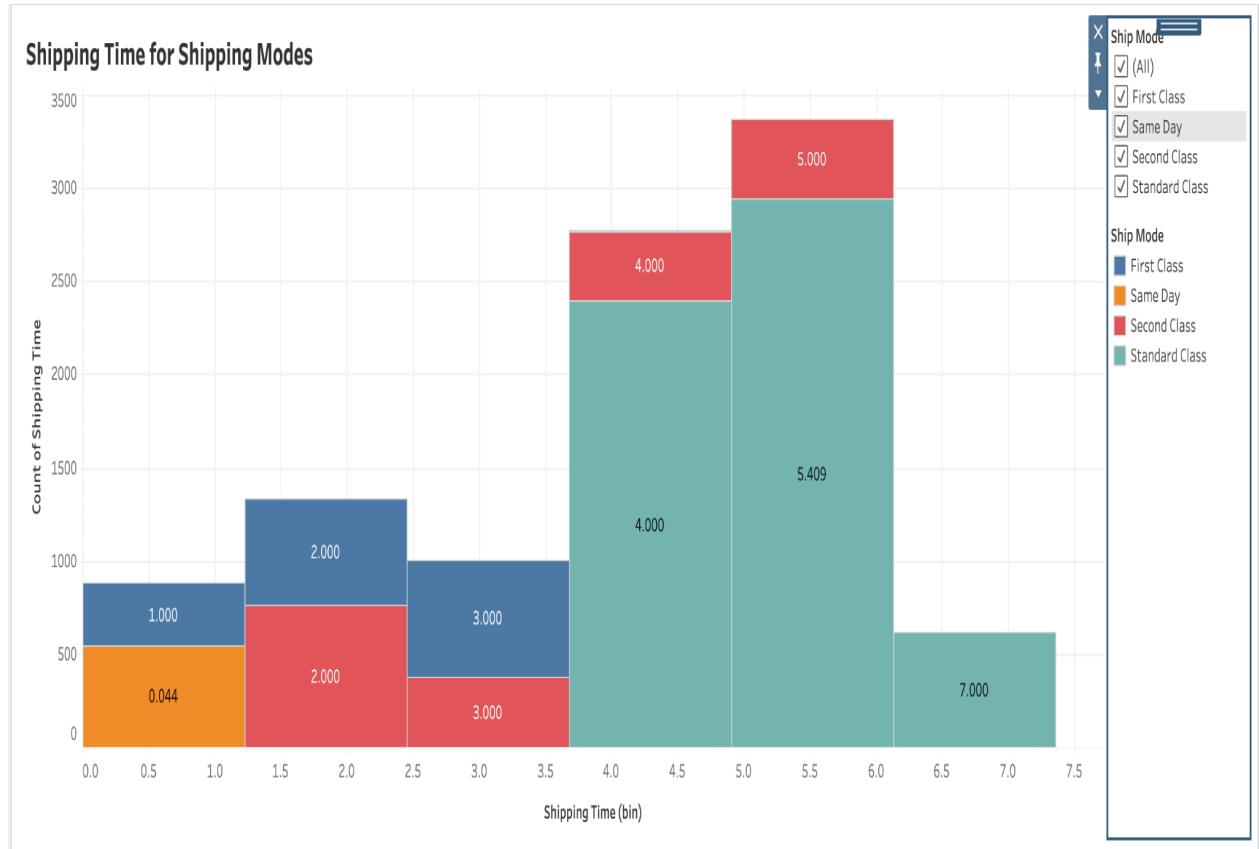


Analyzing the box plot of profit distributions across different product categories reveals valuable insights into the variation in profitability. Each box plot represents a distinct product category.

### Justification

- Comparison of Central Tendency and Spread** Box plots provide a clear visual representation of the central tendency (median) and spread (interquartile range) of profit distributions.
- Identification of Outliers** Box plots include information about outliers, helping to identify any product categories with unusually high or low profits compared to the majority.
- Efficient Visualization of Multiple Categories** Box plots efficiently display distribution characteristics of each product category side by side. This allows for a quick and effective comparison, making it easy to identify variations and similarities among different categories.

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

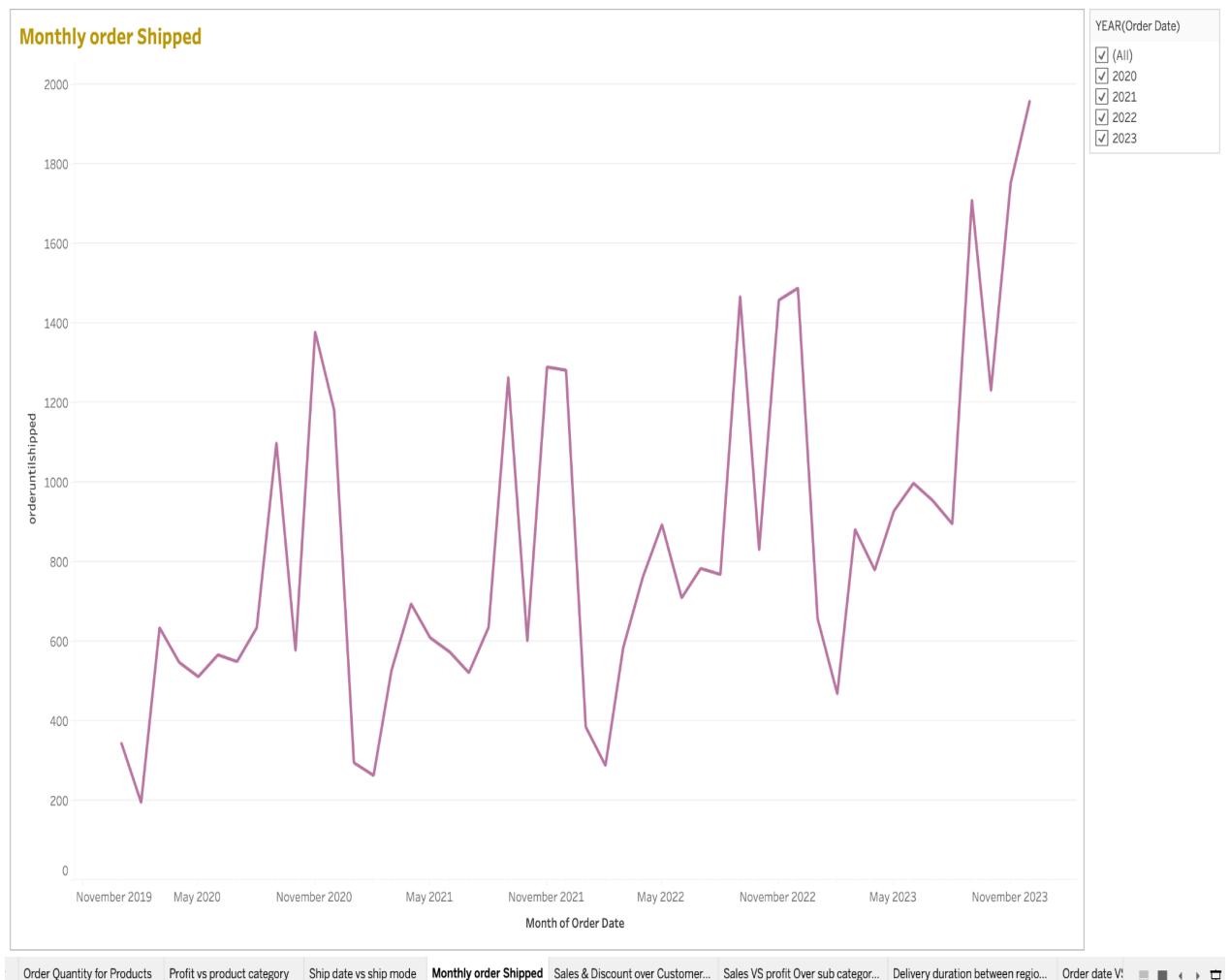


When comparing shipping time distributions for different shipping modes, histograms are a useful visual tool to illustrate the frequency or probability of different shipping times.

## Justification

- Visualizing Distribution Shape** Histograms provide a clear representation of the shape of the distribution for each shipping mode. They allow you to observe patterns such as symmetry, skewness, or multimodality, which can offer insights into the reliability and consistency of each shipping method.
- Detecting Outliers** Outliers, or unusually long shipping times, can be quickly identified in histograms. This is essential for understanding the potential for delays or exceptional cases in each shipping mode.
- Facilitating Decision-Making** Decision-makers can easily interpret histograms to compare shipping modes and make informed decisions based on the observed patterns and trends in the data.

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

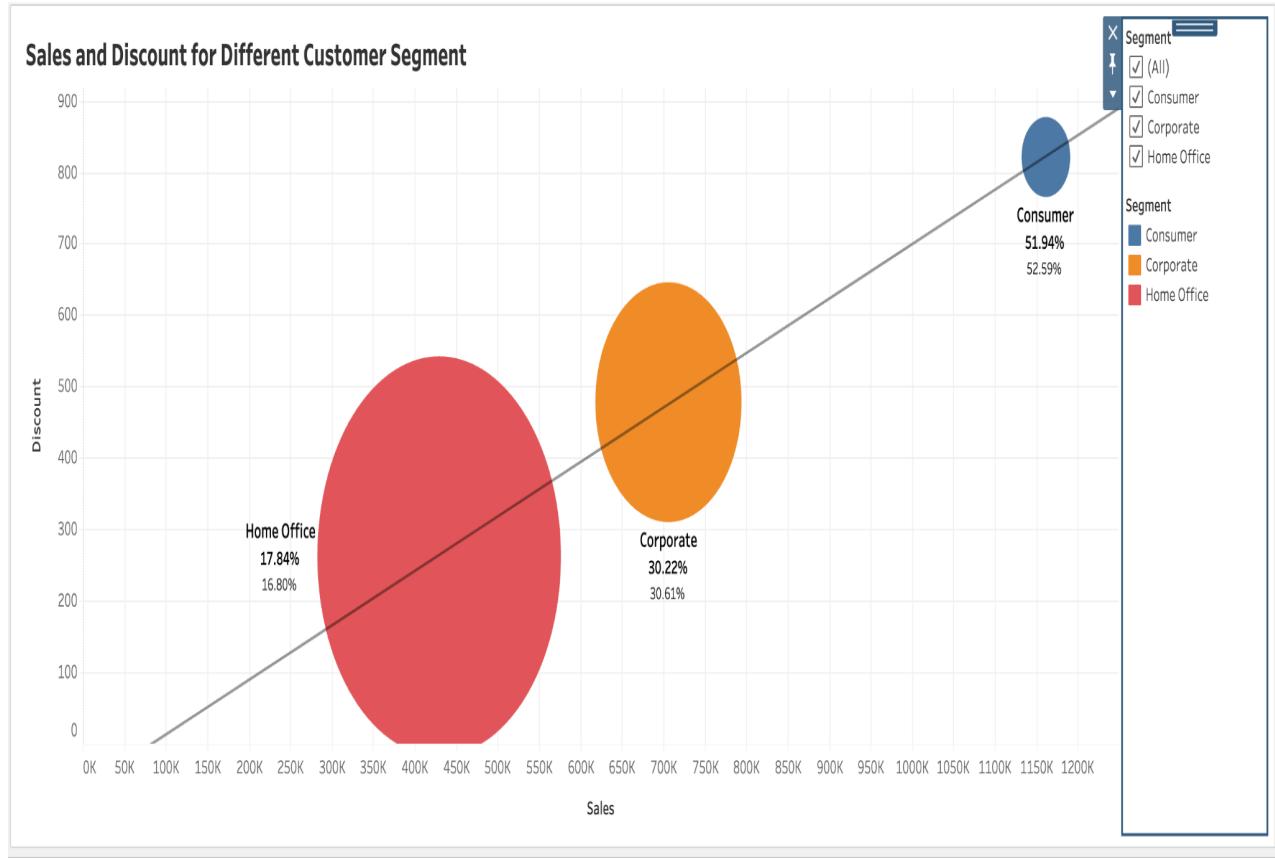


A line chart is an excellent choice for visualizing the monthly trend in the number of orders shipped reasons are Continuous Data Representation, Identification of Patterns, Visualizing Changes Over Time.

## Justifications

- 1. Time Series Representation** Line charts are well-suited for representing time series data, where the x-axis typically represents time.
- 2. Clarity in Trends** Line charts provide a clear visualization of trends over time. The continuous line connecting data points helps in identifying patterns, fluctuations, and trends in the monthly shipment of orders.
- 3. Facilitates Forecasting** If the dataset spans a considerable time period, a line chart can assist in identifying seasonal patterns or long-term trends. This information can be valuable for forecasting future order shipment trends.

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

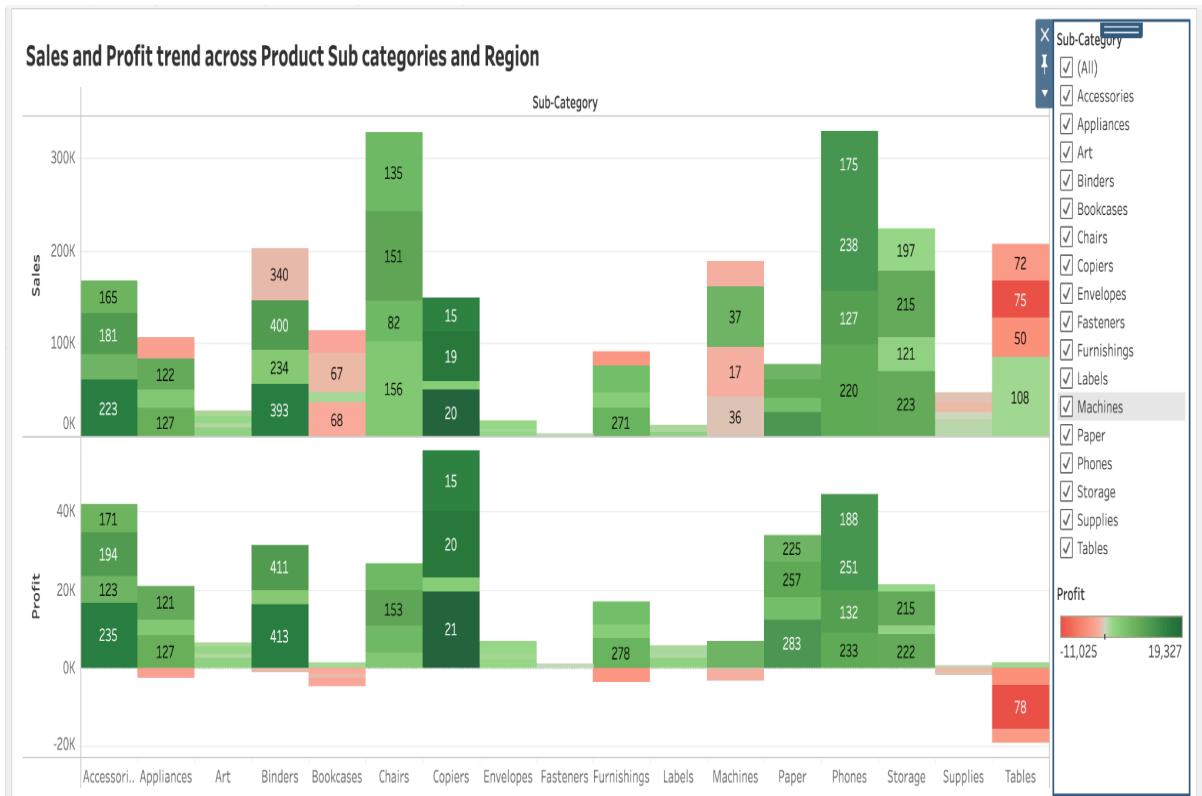


A scatter plot to analyze the performance of different customer segments in terms of sales and discount rates can provide valuable insights. A scatter plot is a versatile and powerful tool for exploring the relationship between sales and discount rates across different customer segments.

### Justification

- Individual Data Point Representation** Scatter plots display individual data points, allowing you to represent each customer segment as a distinct point on the plot.
- Visualizing Relationships** Scatter plots are excellent for visualizing relationships between two continuous variables, such as sales and discount rates. You can observe whether there's a discernible pattern, trend, or correlation between these variables for each customer segment.
- Segment Comparison** Scatter plots allow for a direct comparison between different customer segments. Each segment can be represented by a different color or marker shape, making it easy to distinguish their performance 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?

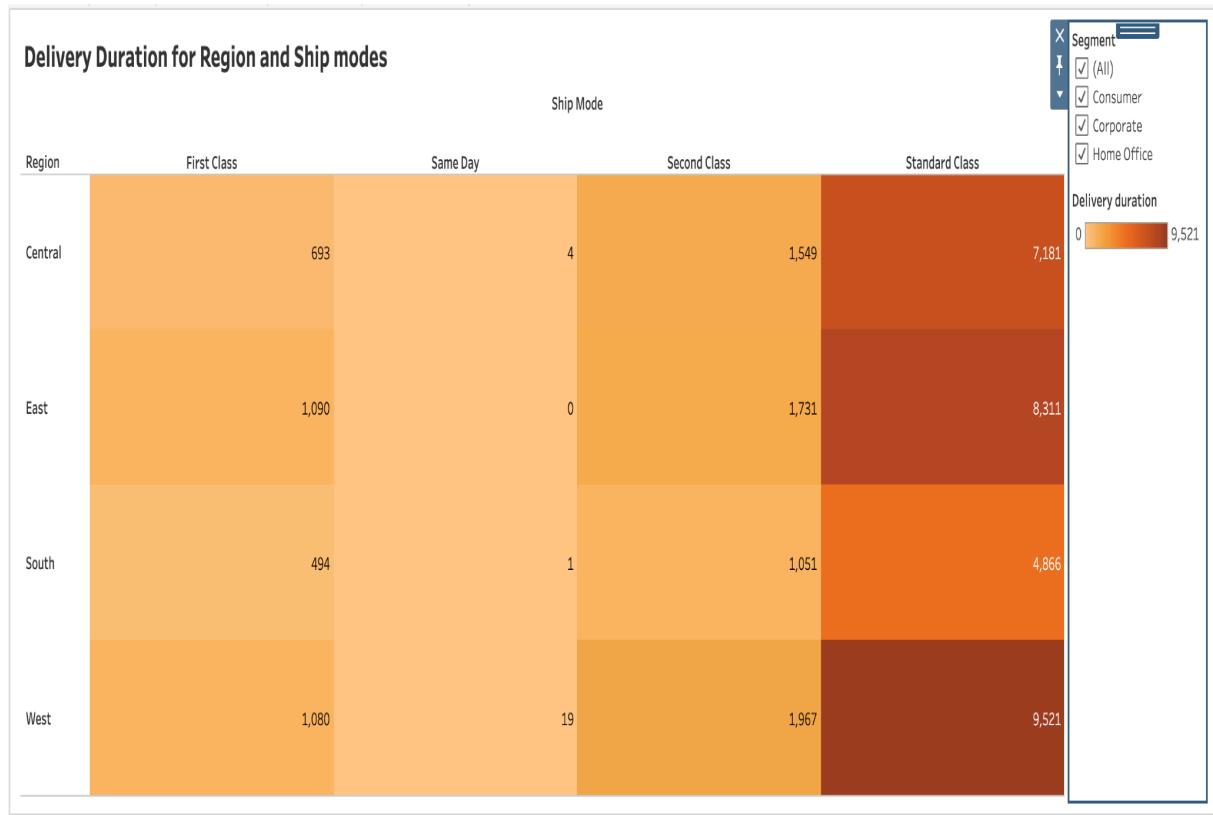


In bar chart different colors for different product subcategories and regions, and consider adding additional annotations or labels to highlight specific trends or noteworthy observations.

## Justifications

- Customization for Readability** Bar charts offer customization options, such as color-coding, grouping, and labeling, which can enhance the readability of the chart. This is important when dealing with multiple categories and regions.
- Categorical Data Representation** Bar charts are well-suited for representing categorical data, such as product subcategories and regions. Each bar represents a discrete category, and the visual separation between bars enhances the clarity of the presentation.
- Facilitating Trend Identification** Over time or across different data points, changes in sales and profit trends can be easily identified in a bar chart. This is particularly useful for spotting trends and patterns across various product subcategories and regions.

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

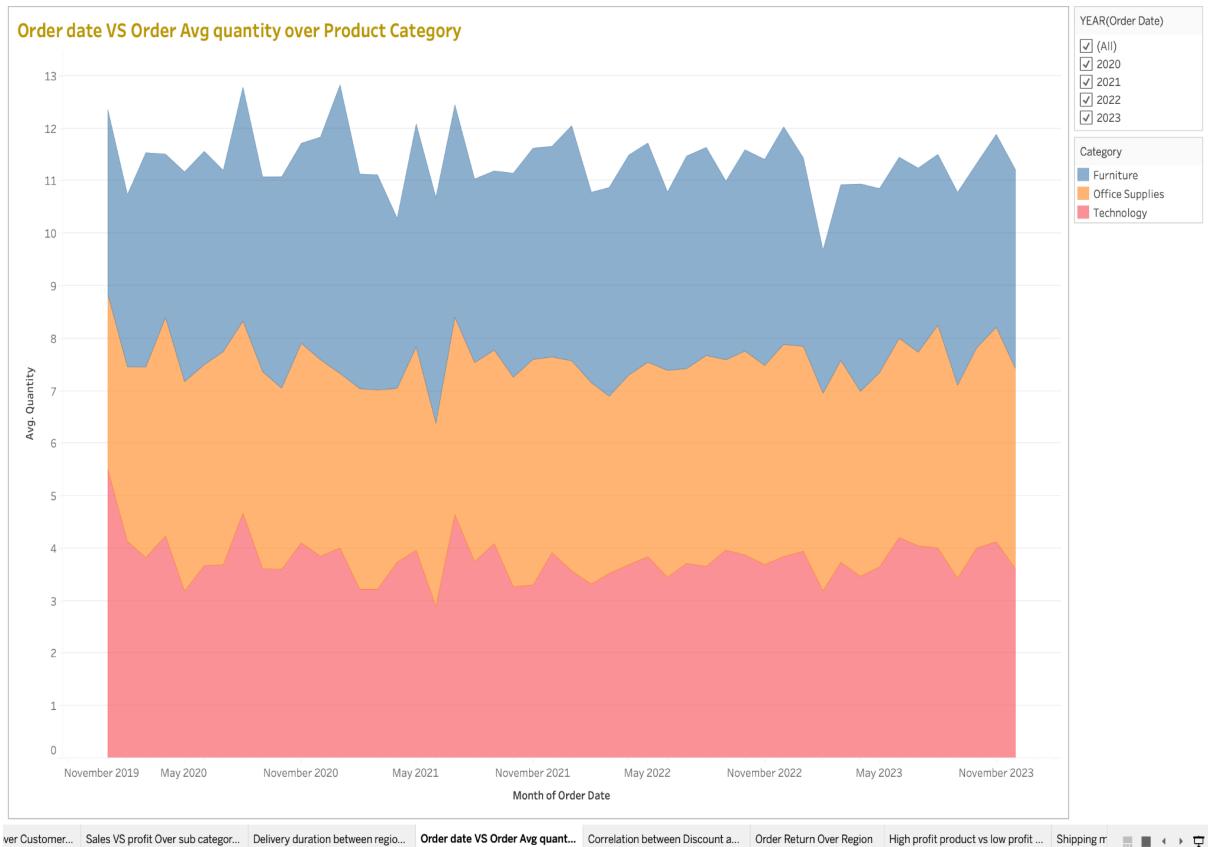


A heat map to visualize the average delivery duration for different regions and ship modes is a suitable choice. The heat map should effectively communicate the average delivery durations, allowing stakeholders to make informed decisions based on the observed patterns.

## Justification

- Interactive Exploration** Many visualization tools, including Tableau, allow for interactive exploration of heat maps. Users can hover over cells for detailed information, filter data dynamically, and zoom in on specific regions or ship modes, enhancing the depth of analysis.
- Quantitative Information Display** The color gradients in a heat map can represent quantitative information effectively. This makes it easy to convey not only the existence of variations but also the magnitude of differences in average delivery durations.
- Efficient Comparison** The grid layout of a heat map enables efficient visual comparison between regions and ship modes. Users can quickly identify areas of interest or concern, such as longer delivery durations in specific regions or with certain ship modes.

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

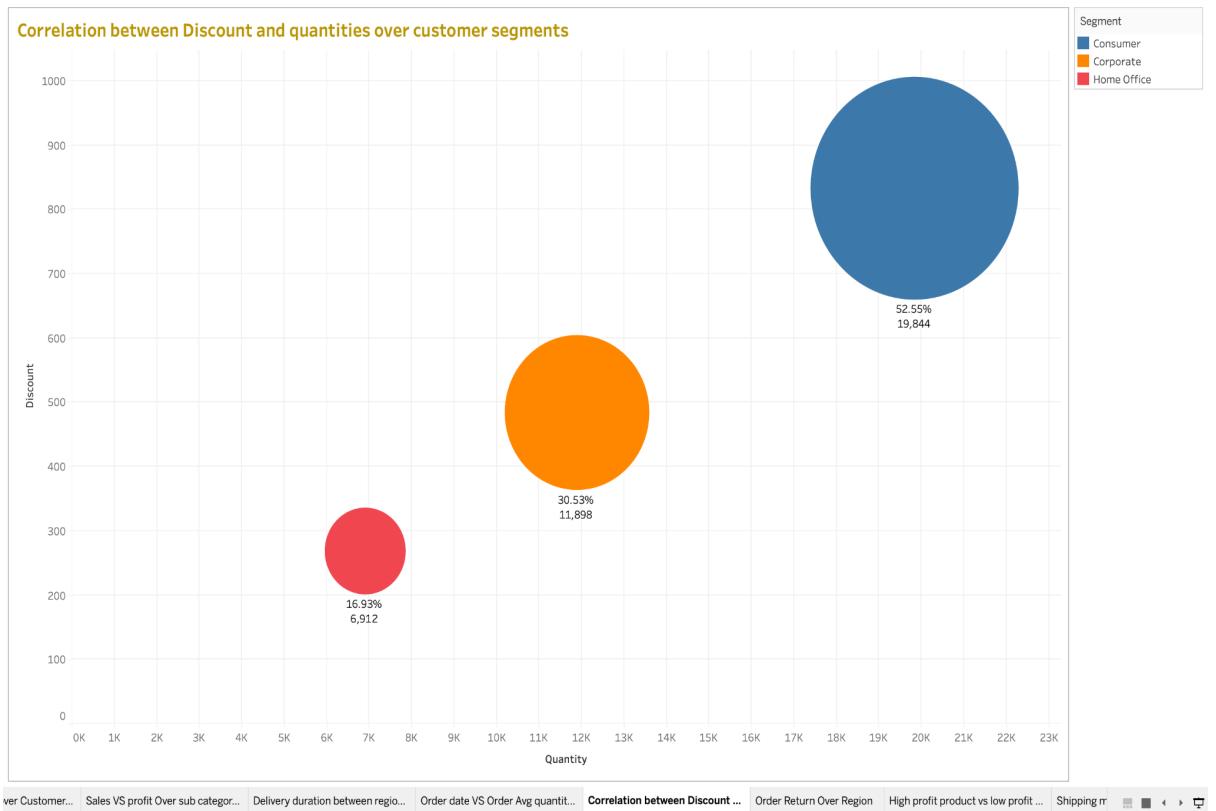


the average order quantity has changed over the years for various product categories, an area chart can be an effective visual representation. An area chart is suitable for illustrating the cumulative values of different categories over a continuous axis, such as time.

## Justification

- Visual Comparison** An area chart allows for a clear visual comparison of the trends in average order quantity for different product categories over time.
- Cumulative Effect** The filled areas in the chart create a cumulative effect, making it visually intuitive to see the total impact of changes in average order quantity for each product category.
- Multiple Categories** An area chart is effective when visualizing multiple categories simultaneously, as each category is represented by a distinct colored area, aiding in the interpretation of the overall trend.

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

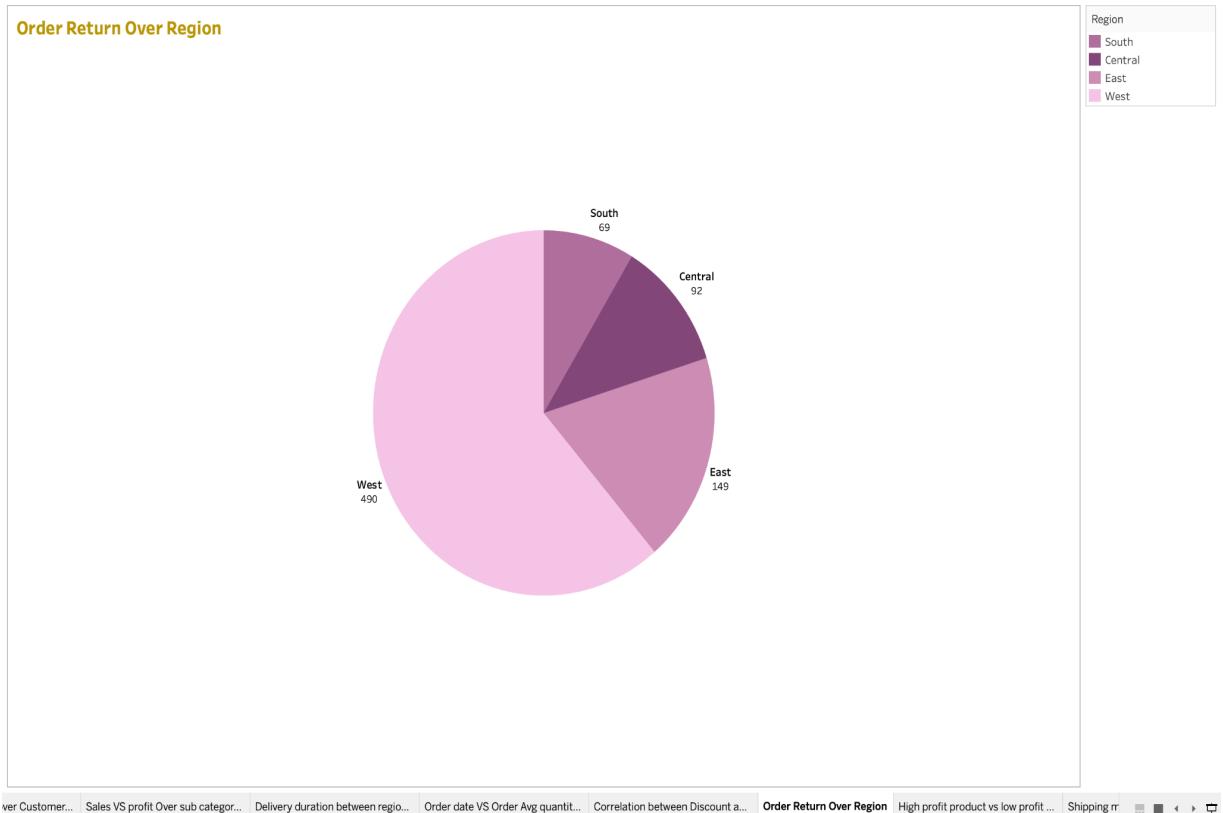


Scatter plot is a useful tool for visualizing the correlation between discount rates and order quantities for different customer segments. Reasons include Correlation Assessment, Identification of Outliers, etc.

## Justification

- 1. Individual Data Points** Scatter plots display individual data points on a two-dimensional plane, where each point represents a combination of discount rate and order quantity for a specific customer segment. This allows for a detailed examination of the relationship between the two variables.
- 2. Segmented Analysis** Since you want to analyze the correlation for different customer segments, scatter plots allow you to create separate plots for each segment. Each plot can have its own set of data points, making it clear how discount rates impact order quantities within each segment.
- 3. Correlation Strength** The scatter plot allows you to visually assess the strength and direction of the correlation. A strong positive correlation would show a trend where higher discount rates are associated with higher order quantities, while a strong negative correlation would indicate the opposite.

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

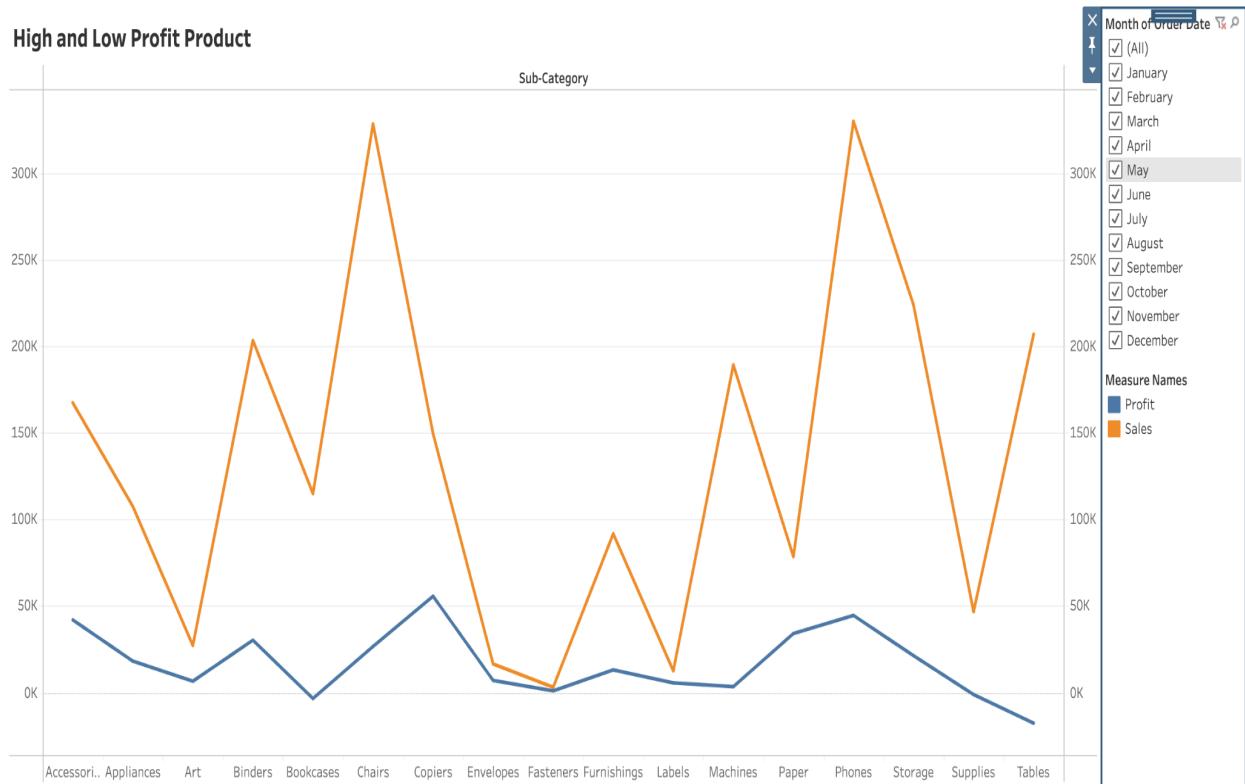


A pie chart is a suitable choice for visualizing the proportion of orders returned in each region. Pie charts communicate the distribution of returned orders across different regions in a visually compelling and easy-to-understand format.

## Justification

- 1. Clear Proportional Representation** A pie chart effectively communicates the proportion of each region's contribution to the total orders returned. Each region is represented as a slice of the pie, and the size of each slice corresponds to the proportion of returned orders.
- 2. Summarization of Data** The primary purpose of the visualization is to provide a summarized view of the proportion of returned orders in each region. A pie chart serves this purpose by presenting a clear and concise overview.
- 3. Easy Comparison** Pie charts make it easy to compare the contribution of different regions at a glance. Stakeholders can quickly identify which regions have higher or lower percentages of returned orders.

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

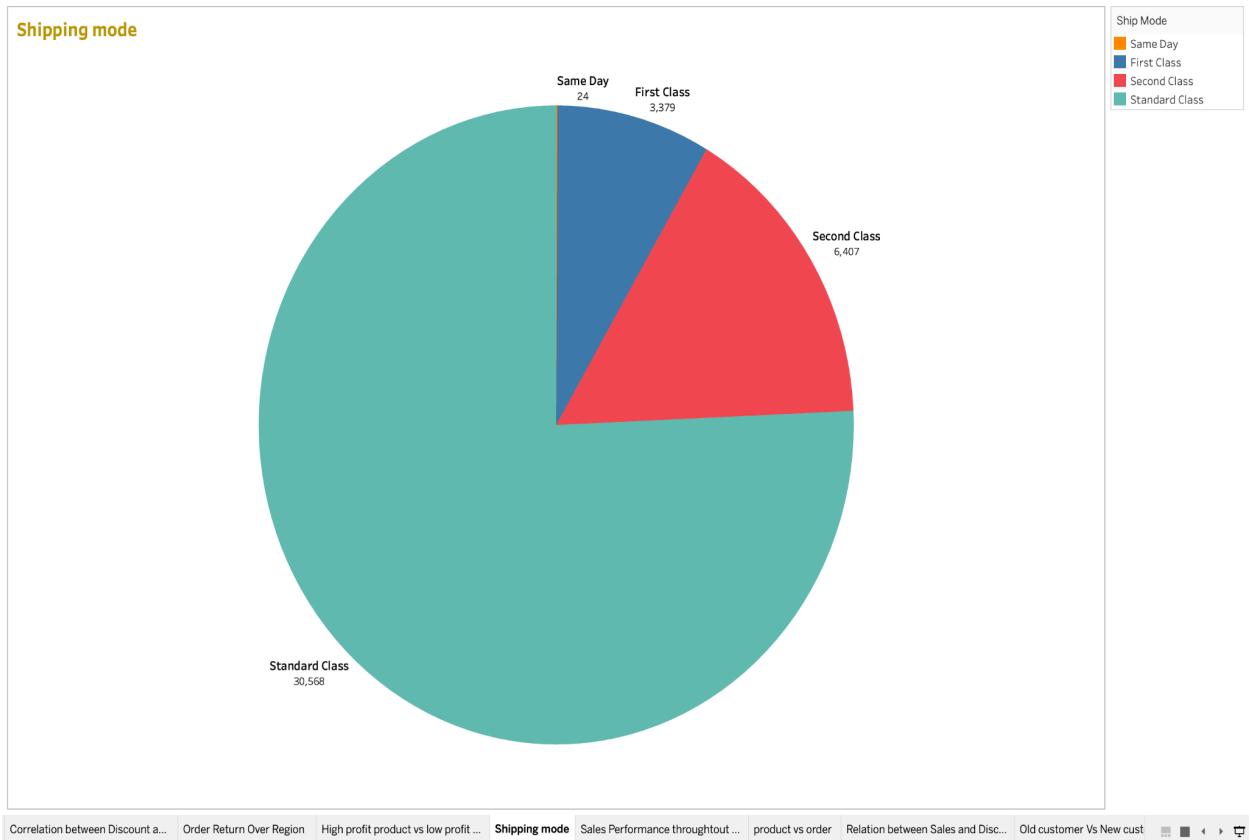


A dual-axis line chart in Tableau, you can effectively communicate the sales performance of high-profit and low-profit products, providing valuable insights for decision-making.

## Justification

- Comparison of Trends:** The dual-axis line chart enables a clear comparison of the sales trends for high-profit and low-profit products over time.
- Efficient Use of Space:** Combining the two axes on the same chart allows for an efficient use of space, making it easier for viewers to analyze the data.
- Visual Clarity:** Differentiating high-profit and low-profit products with distinct colors enhances visual clarity, helping viewers quickly understand the information.
- Temporal Analysis:** The use of a timestamp on the x-axis facilitates a temporal analysis, allowing users to identify any seasonality or patterns in sales over time.

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

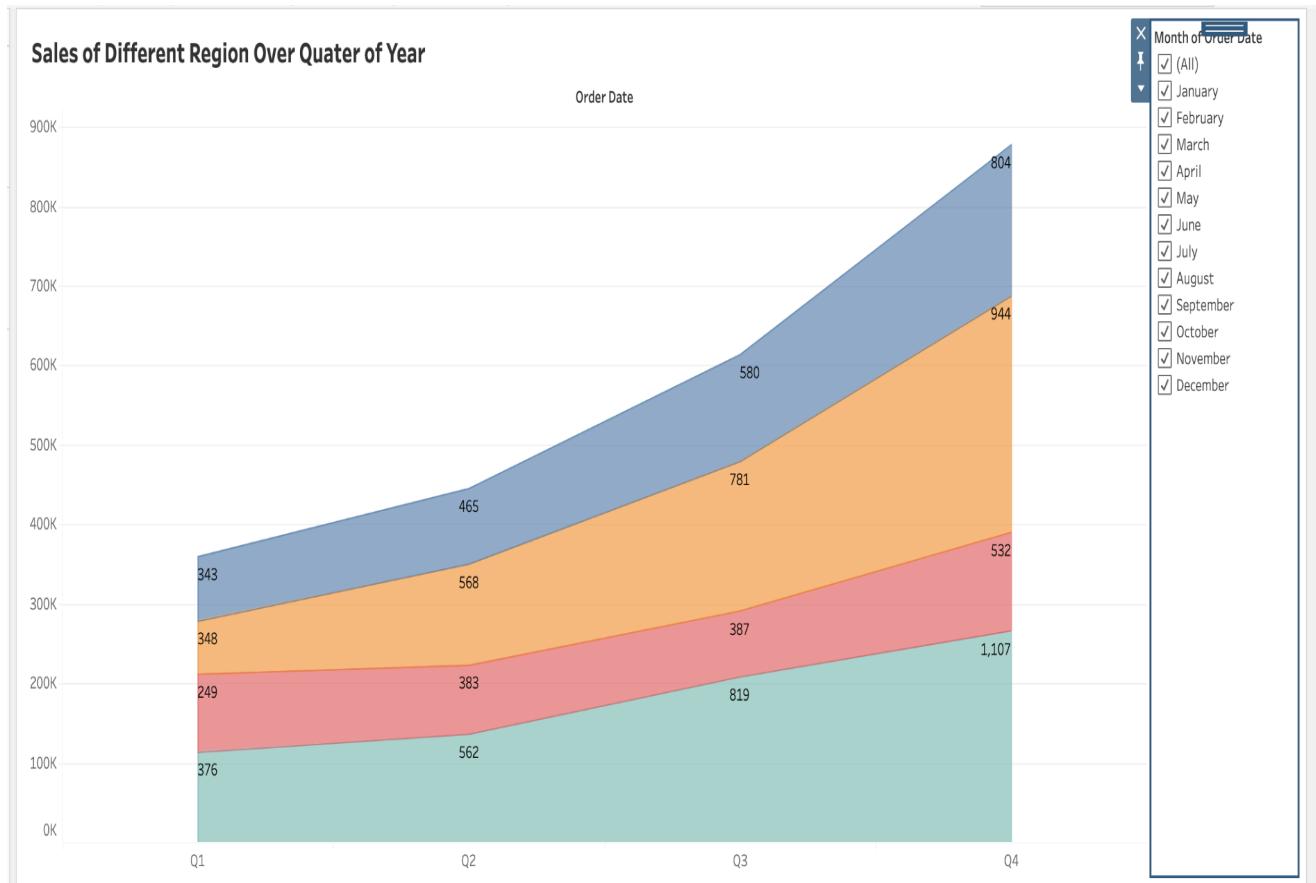


A pie chart is a suitable choice for visualizing the distribution of shipping modes in the Sample Superstore dataset because it allows for a clear representation of the proportion of each shipping mode, allowing stakeholders to quickly grasp the most commonly used shipping mode.

## Justification

- 1. Proportional Representation** A pie chart provides an intuitive and visually appealing representation of the proportion of each shipping mode. Each slice of the pie corresponds to a specific shipping mode, and the size of the slice represents its share of the total.
- 2. Easy Comparison** Pie charts make it easy to compare the usage of different shipping modes. Stakeholders can quickly identify which shipping mode is the most commonly used by observing the size of the corresponding slice.
- 3. Simplicity and Clarity** Pie charts are simple and easy to understand, making them accessible to a wide audience, including those who may not be familiar with data analysis. The clear and concise representation helps in communicating information effectively.

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

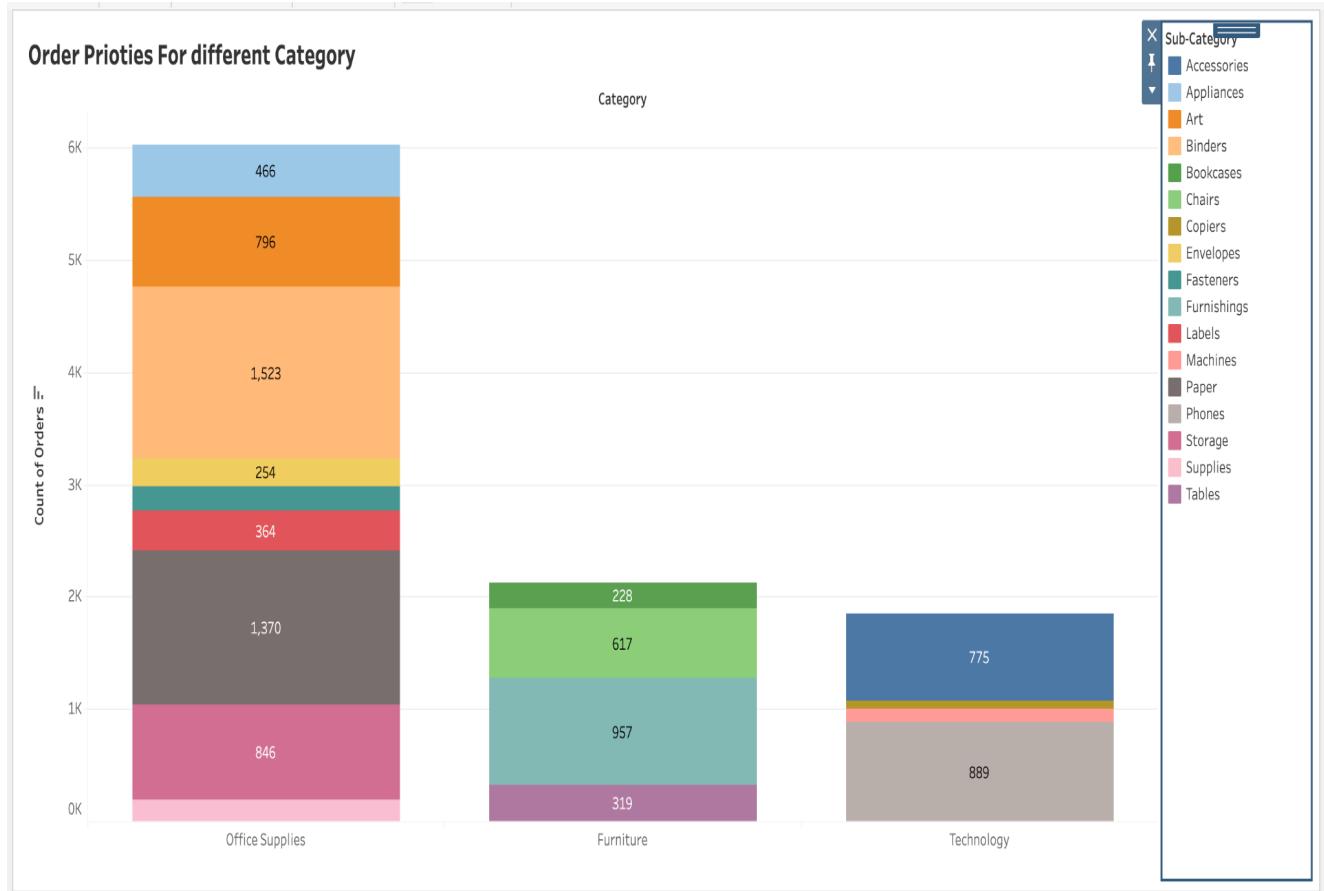


An area chart compelling visualization that effectively communicates the sales performance of different regions throughout the quarters of a year.

## Justification

- Temporal Trends** An area charts effectively communicate temporal trends, allowing viewers to observe how sales evolve over quarters.
- Region Comparison** These chart types enable a clear comparison of sales performance across different regions within each quarter.
- Visual Appeal** An area charts are visually appealing and easy to interpret, making them suitable for communicating complex information in a straightforward manner.
- Summarized View** The visualization provides a summarized view of the sales distribution across regions, making it easy to identify patterns and trends over time.

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

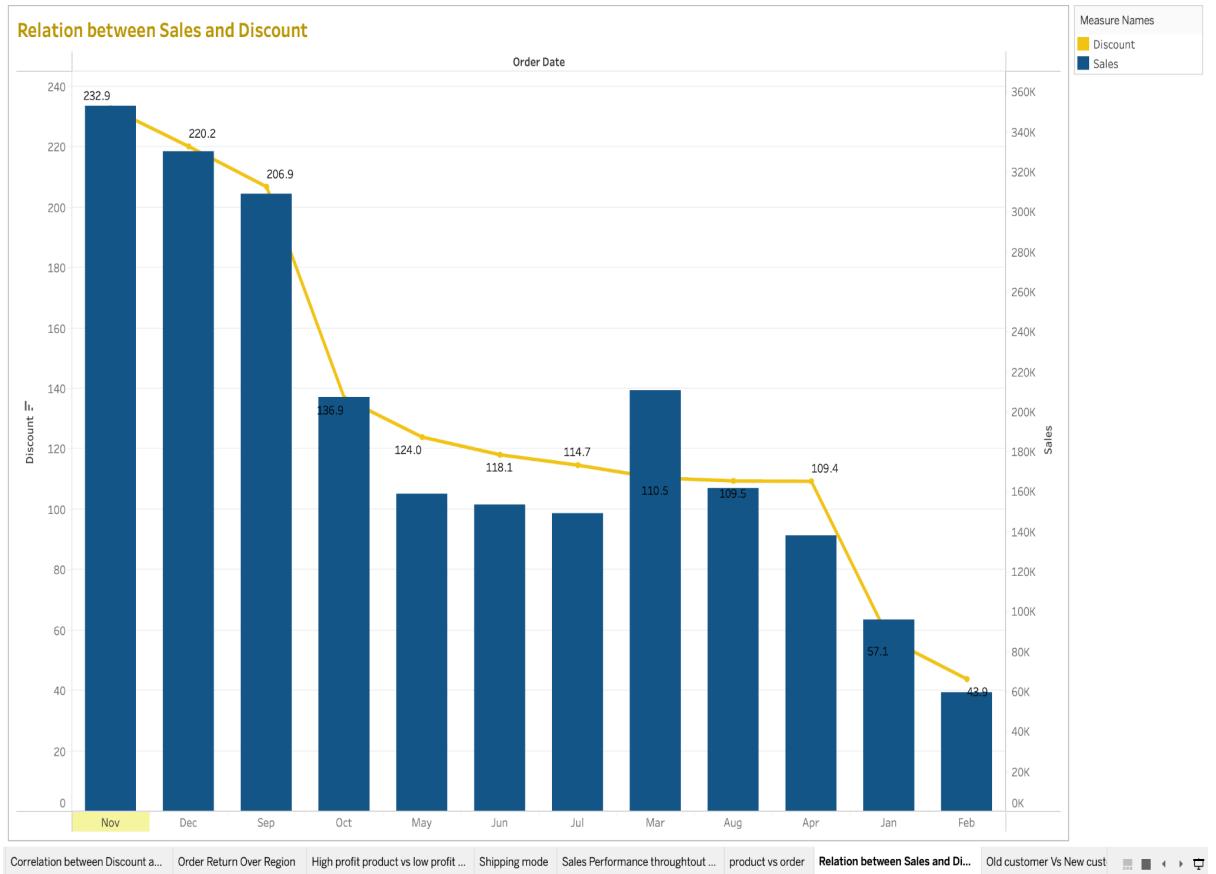


A stacked bar chart would be a suitable choice. This type of chart allows you to visualize the composition of order priorities within each product category. effectively visualize and communicate the distribution of order priorities across different product categories.

### Justification

- Composition Analysis** Stacked bar charts are effective for analyzing the composition of different categories within a whole. In this case, it helps to understand how order priorities contribute to the total orders within each product category.
- Visual Clarity** The use of stacked bars makes it easy to compare the distribution of order priorities across different product categories.
- Easy Interpretation** Stacked bar charts are intuitive and easy to interpret. Each bar represents a product category, and the segments within each bar represent the order priorities.
- Comparison Across Categories** This chart type facilitates the comparison of order priorities across various product categories in a single view.

28. What is the relationship between discounts and sales?

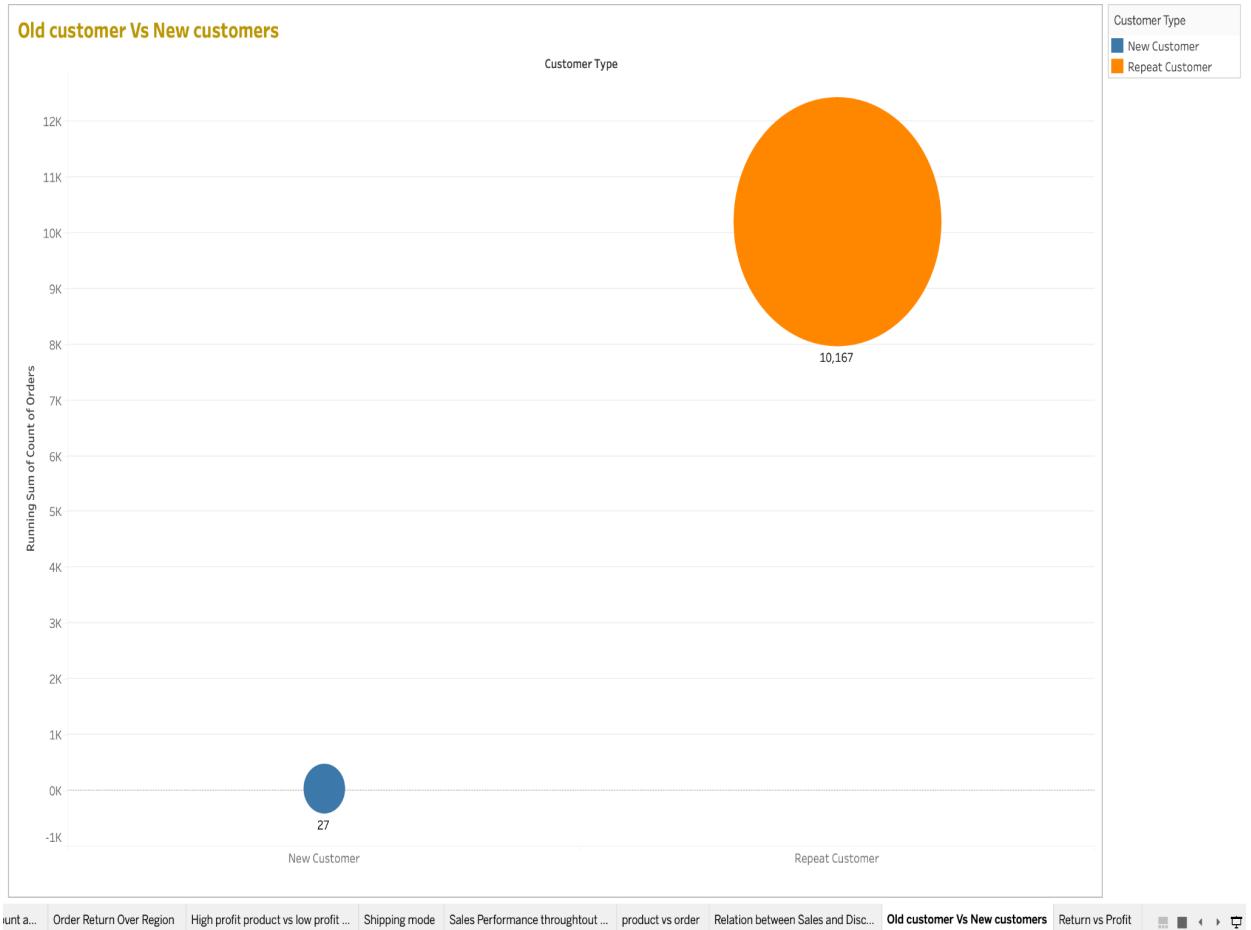


A bar chart with a trend line can be a good choice for several reasons are Trend Analysis, Quantitative Comparison and Communication of Correlation.

## Justification

- Visual Representation** A bar chart is effective for visually representing the sales values corresponding to different discount levels. Each bar provides a clear comparison of sales amounts.
- Trend Line for Relationship Analysis** Adding a trend line helps in analyzing the overall relationship between discounts and sales. It provides a visual representation of whether increased discounts lead to increased or decreased sales.
- Comparison Across Discount Levels** A bar chart allows for a direct comparison of sales values across different discount levels, helping identify patterns and outliers.
- Simplicity and Clarity** Bar charts are easy to interpret, and the addition of a trend line enhances the clarity of the relationship between discounts and sales.

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

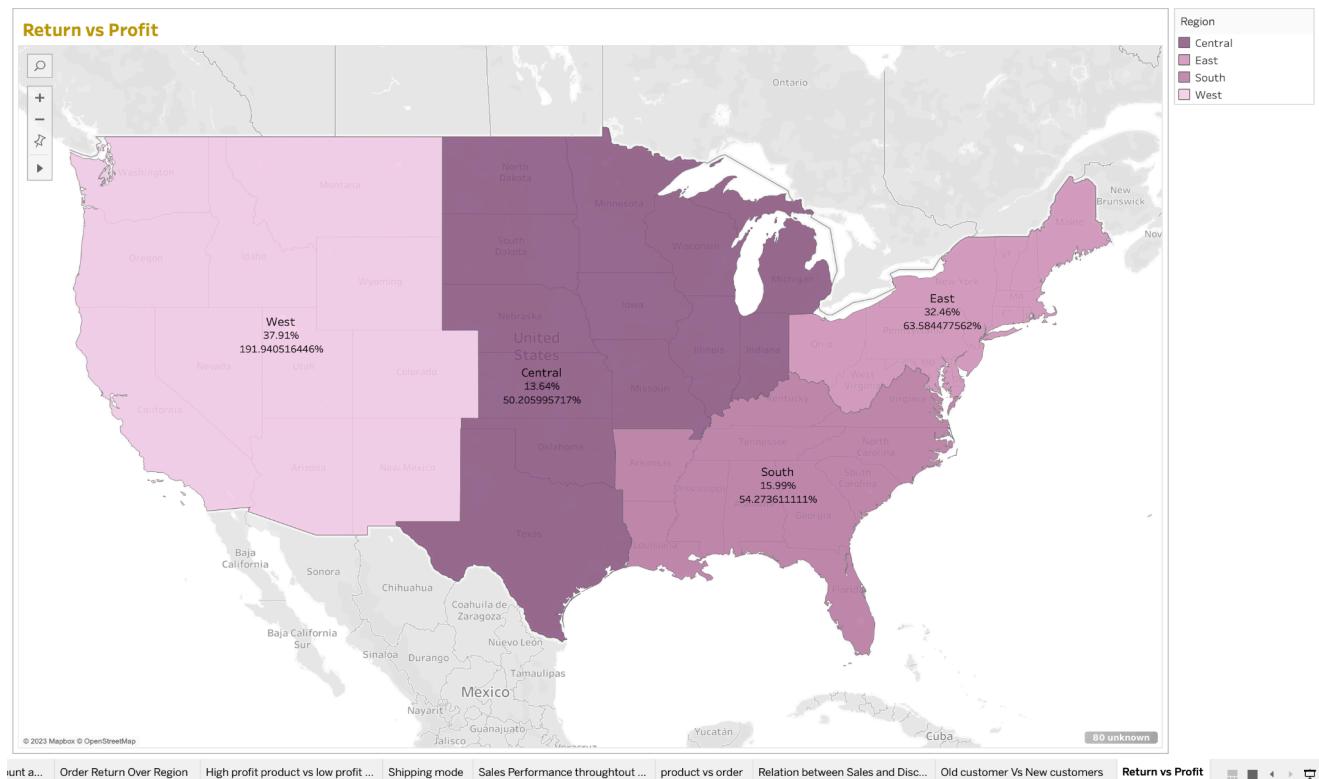


A scatter plot is an effective choice for visualizing the difference in average order value between repeat customers and new customers. This type of chart allows you to display individual data points for each customer, making it easy to observe patterns, trends, and variations in average order values.

## Justification

- Individual Data Points** Scatter plots display individual data points, making it easy to observe the specific average order values for each customer and their corresponding customer type.
- Visual Comparison** The scatter plot allows for a visual comparison of average order values between repeat and new customers. Clusters or patterns can be quickly identified.
- Identification of Outliers** Scatter plots are effective in identifying outliers or individual data points that deviate significantly from the overall trend, helping to identify exceptional cases.

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



A map is an excellent choice for visualizing the geographical distribution of returns and understanding its impact on overall profitability. Effectively communicate the geographical distribution of returns and visualize its impact on overall profitability. This information can inform strategic decisions, such as optimizing supply chain processes or implementing targeted marketing strategies in regions with higher return rates.

## Justification

- Geographical Context** A map provides a geographical context, allowing stakeholders to see the distribution of returns across different regions. This helps in understanding if there are specific geographic areas with higher return rates.
- Spatial Patterns** Maps can reveal spatial patterns and correlations, helping identify if returns are concentrated in certain locations. This information can be crucial for supply chain optimization and targeted business strategies.
- Visual Impact** Geographical data is often more impactful when presented visually on a map. Stakeholders can quickly grasp the regional differences in return rates, making it easier to identify trends and outliers.
- Decision Support** A map facilitates decision-making by presenting returns data in a context that is easy to interpret. This is especially important when considering the impact of returns on overall profitability.

**5.**