**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:**

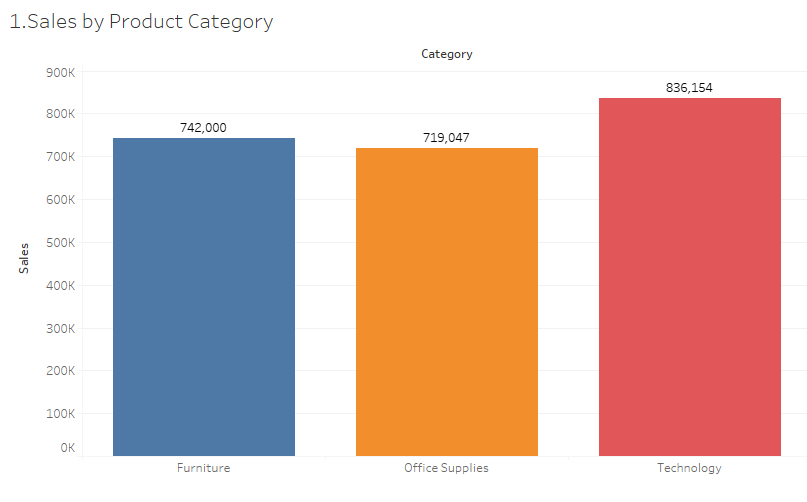
1. **Answer Completion**: Ensure that you furnish answers for all any 30 questions and build charts for them.
2. **Encouraged Creativity**: Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

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!**

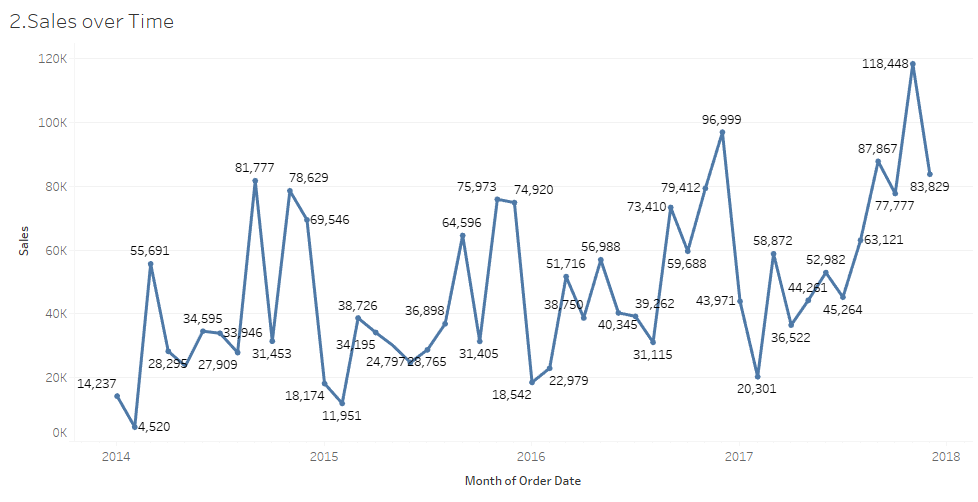
**Questions:**

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



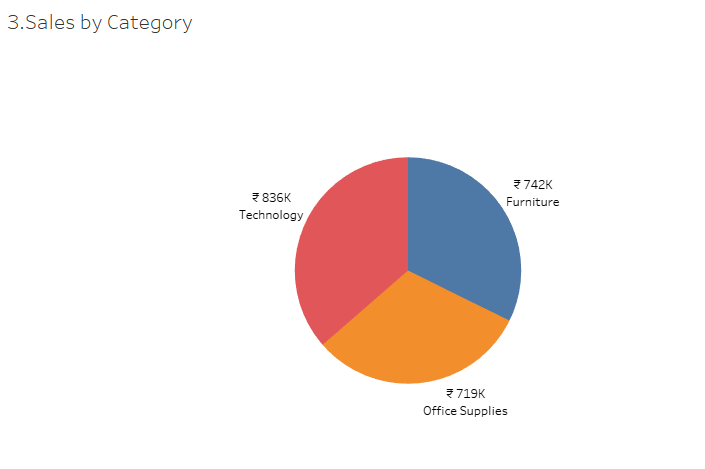
**Insights:** A bar chart is used to compare data across different categories. Using this, we found the Technology category has the highest sales with 8,36,154 followed by Furniture and Office Supplies.

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



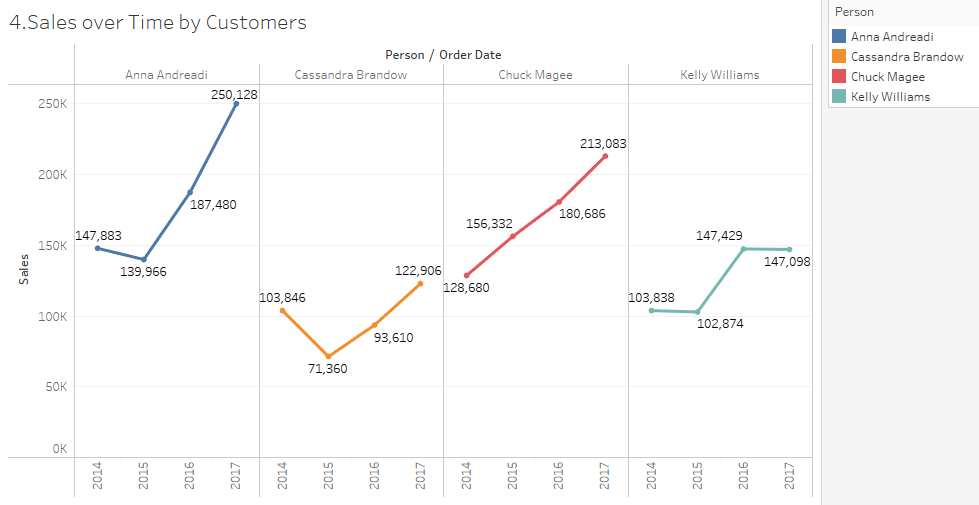
**Insights:** A line chart is used to show trends over time. It is best used when you have data that changes over a period of time. Using the trends, the year 2018 has the highest sales.

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



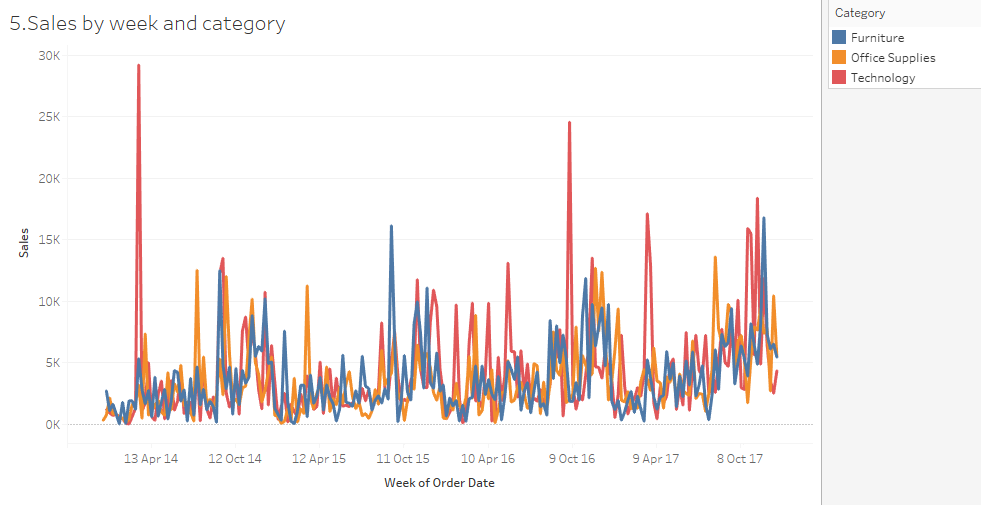
**Insights:** A pie-chart is used to show the proportion of different categories in the data. Here Technology has the highest distribution with 836k, then Furniture with 742k and Office Supplies with 719k.

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



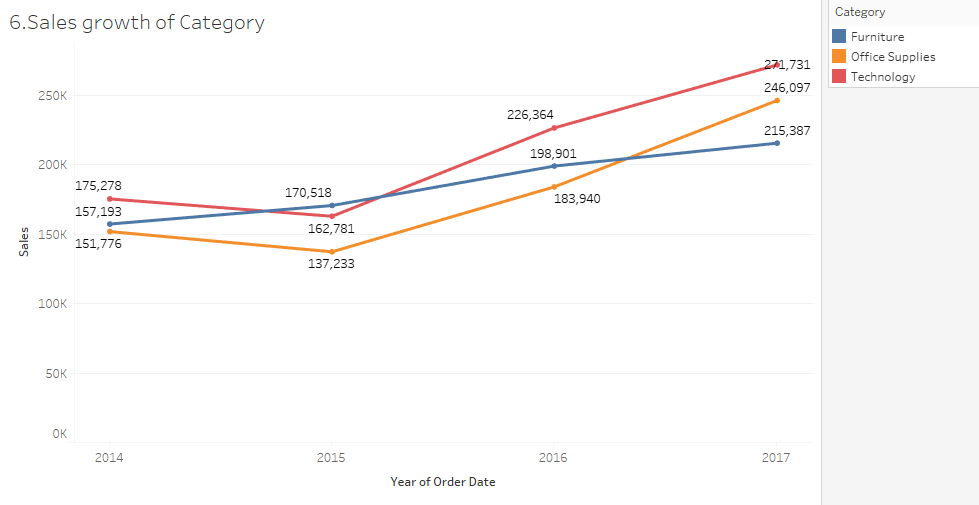
**Insights:** Bar chart is used to analyse data at different data points. Here, Anna Andreadi is the top customer with the highest sales of 2,50,128. Chuck Magea is the second highest with sales of 2,13,083.

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



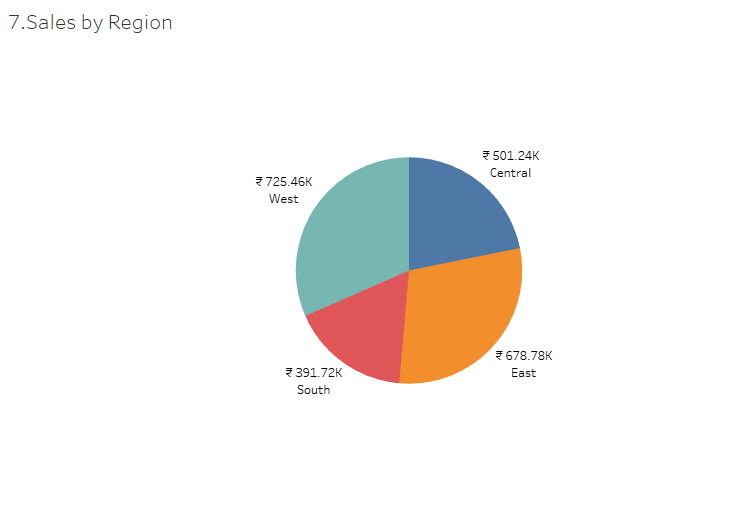
**Insights:** A line chart is used to show trends over time. Using the chart, we can know the trend of the weekly sales by different product categories.

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



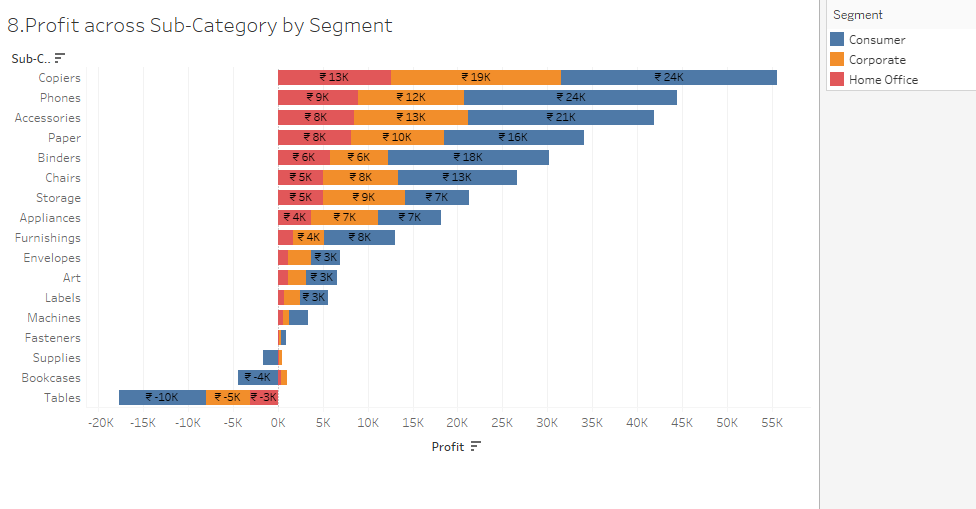
**Insights:** A line chart is used to show trends over time. Using the trend, we analyse the category Technology has the highest sales growth over the time period.

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



**Insights:** A pie-chart is used to show the proportion of different regions in the data. The West region has the highest sales distribution of 725.46k and the South region has the lowest sales distribution of 391.72k.

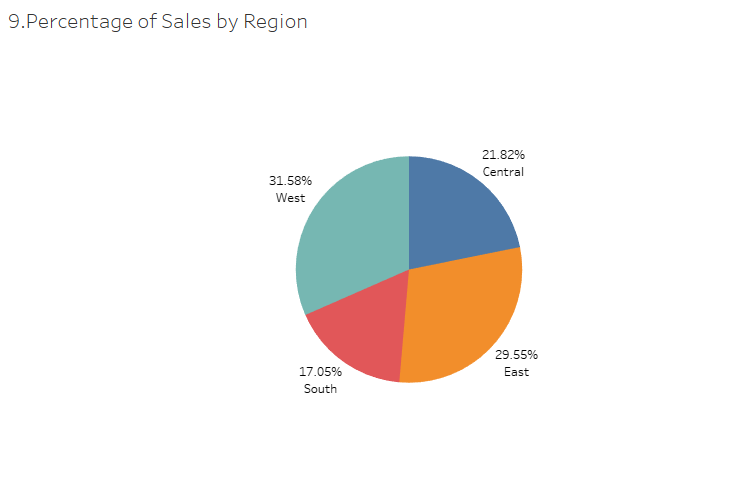
1. Can we visualise the composition of profits across various subcategories within different customer segments?



**Insights:** A bar chart is used to compare data across different sub-categories. The sub-category Copiers has the highest profit and Phones with the second highest profit.

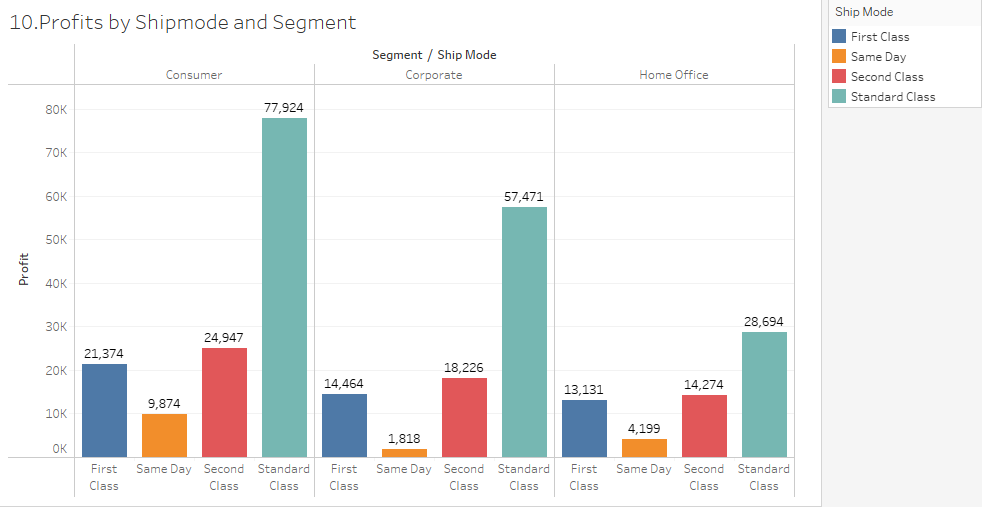
Tables and Bookcases are in the loss sub-category.

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



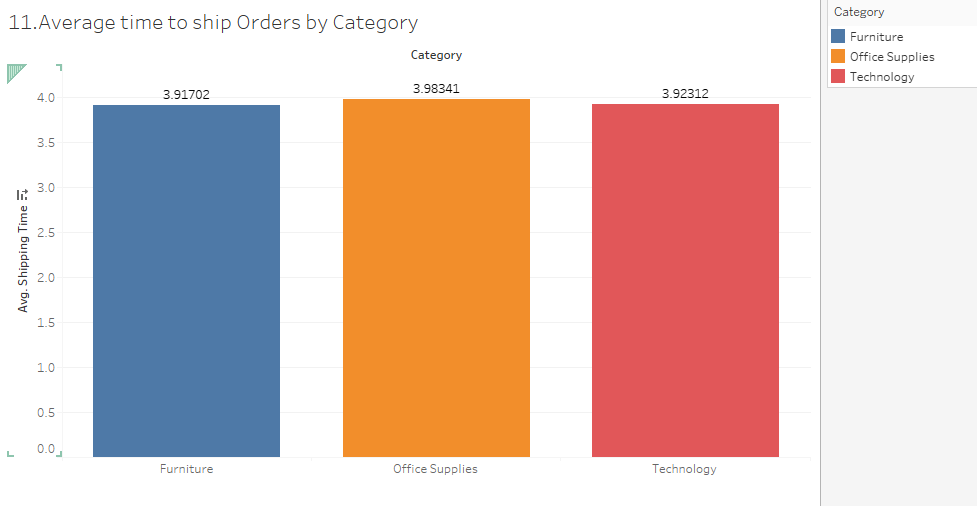
**Insights:** A pie-chart is used to show the proportion of different regions in the data. The West region has the highest sales percentage of 31.58% and the South region has the lowest percentage of 17.05%.

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



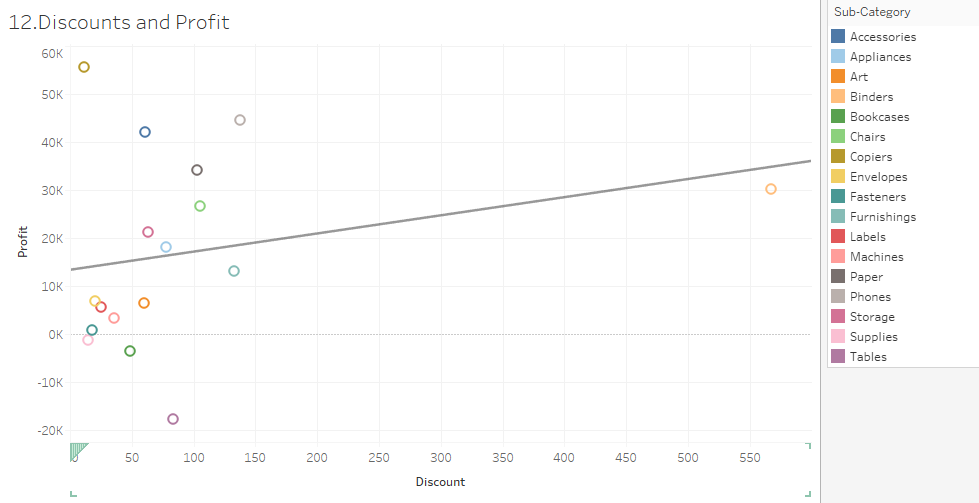
**Insights:** Bar chart is used to analyse data at different data points. Here, we analyse the profit margins of different customer segments at different ship modes. In all the customer segments, the ship mode Standard Class has the highest profit margin.

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



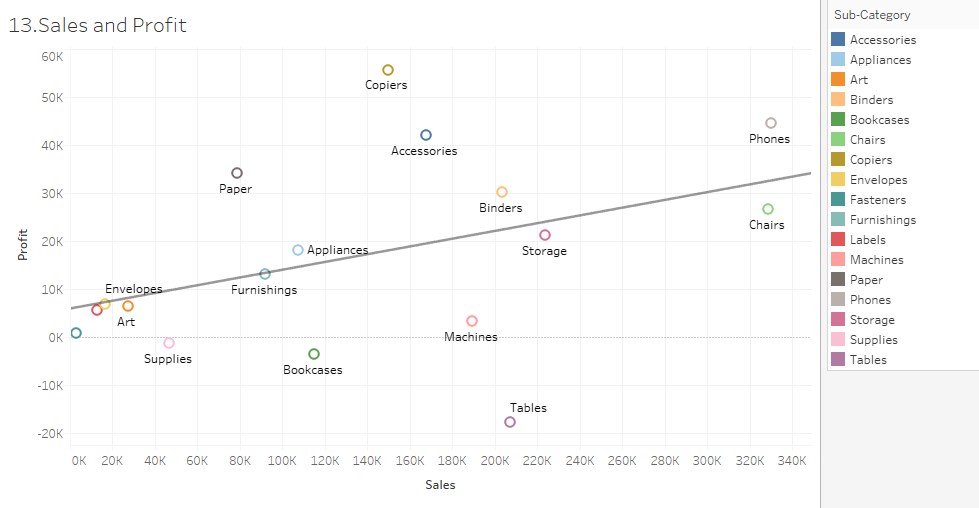
**Insights:** Bar chart is used to analyse data at different data points. The Product Category Office Supplies has an average time of 3.98341. Technology and Furniture also have an almost similar average time.

1. How do discounts affect overall profit?



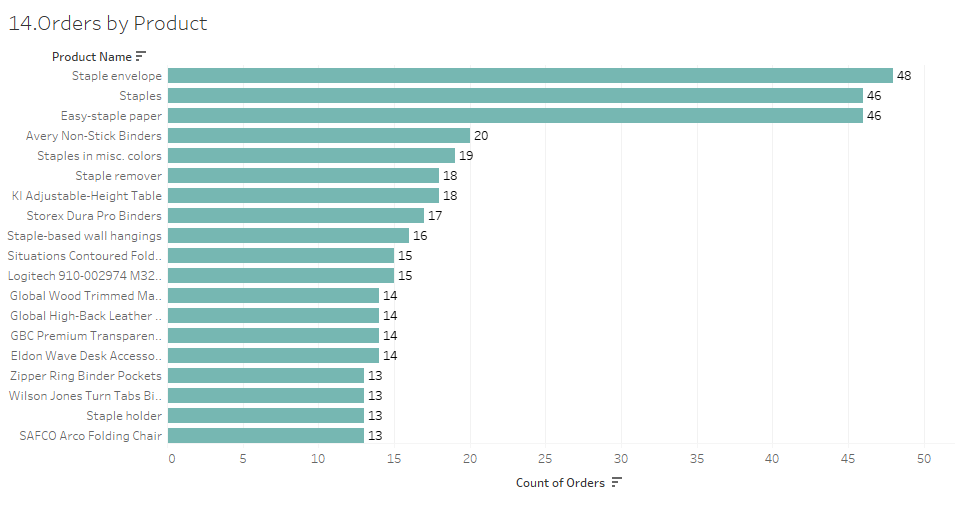
**Insights:** A Scatter plot is used to show the relationship between two variables. We can view the trend of discounts and profit within the different sub-categories.

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



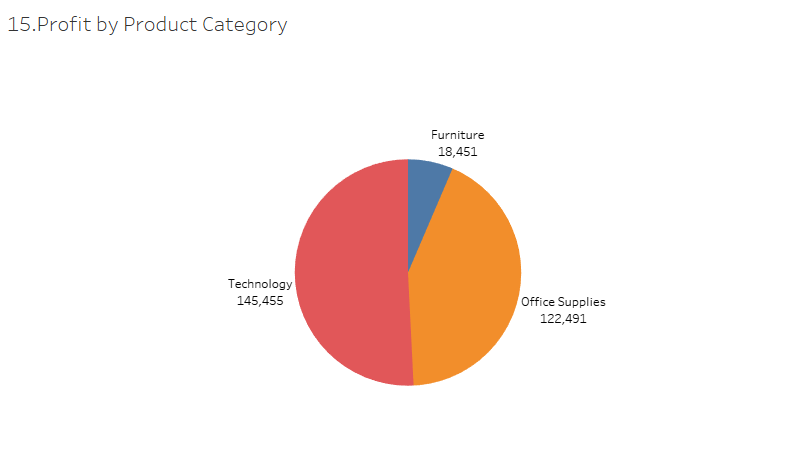
**Insights:** A Scatter plot is used to show the relationship between two variables. In the sub-categories, Chairs and Phones have the highest sales. In Spite of having medium sales, the Copiers have the highest profit margin. By seeing the trend line, we can say there is a positive relationship between sales and profit.

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



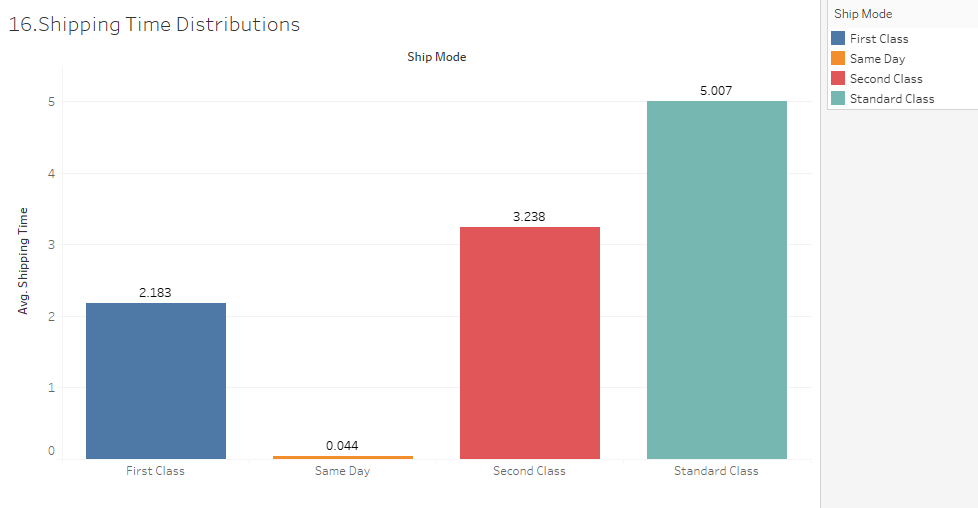
**Insights:** Bar chart is used to analyse data at different data points. Staple envelopes are the product with the highest number of orders. Staples and Easy-staple paper with the second highest number of orders.

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



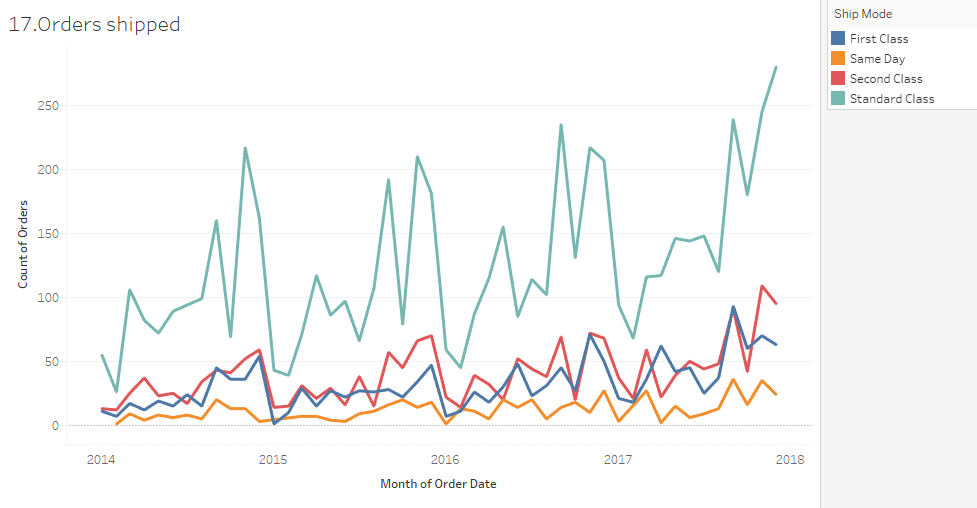
**Insights:** A pie-chart is used to show the distribution of different categories in the data. Technology category has the highest profit margin of 1,45,455 and Furniture category has the lowest profit margin of 18,451. It shows the distribution of profit among the three different product categories.

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



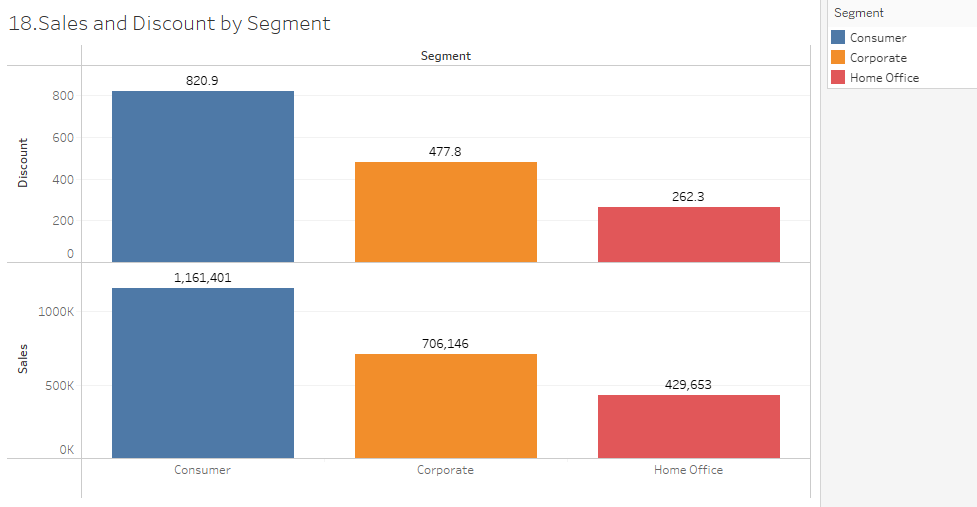
**Insights:** A bar chart is used to compare data across different sub-categories. The average shipping time distribution for Standard Class is 5.007, Second Class is 3.238, First Class is 2.183 and Same Day is 0.044.

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



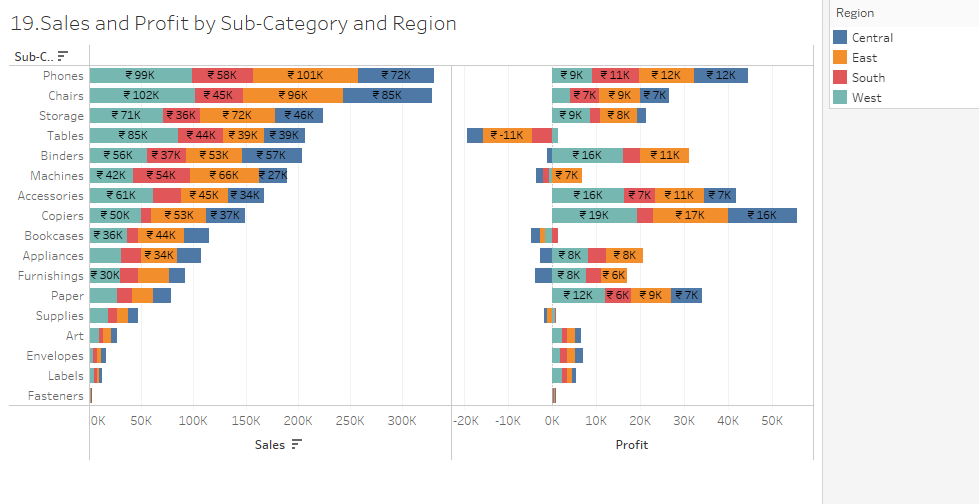
**Insights:** A line chart is used to show the trends over a period of a time. Standard class ship modes have the highest number of orders and same day ship modes have the lowest order trends.

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



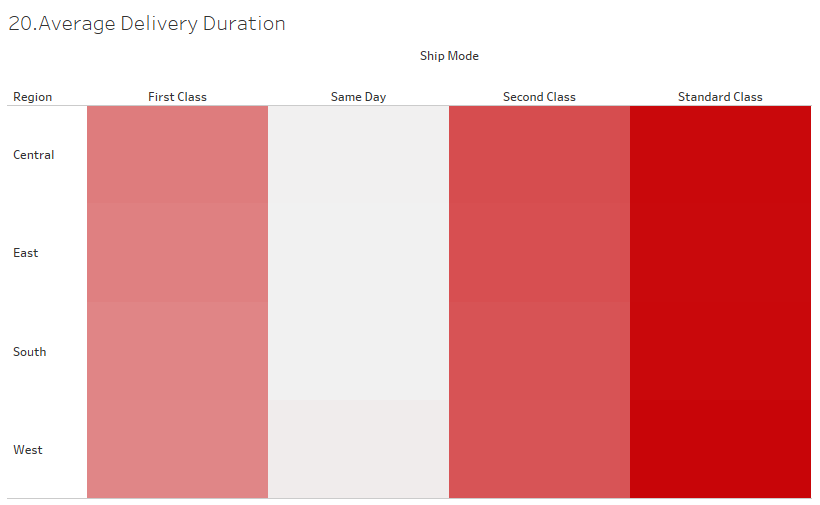
**Insights:** A dual bar chart is used to analyse the multiple levels of data at different data points. Consumers have the highest discount rates and the sales are also high. The Home Office has the lowest discount rates and the sales are low.

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



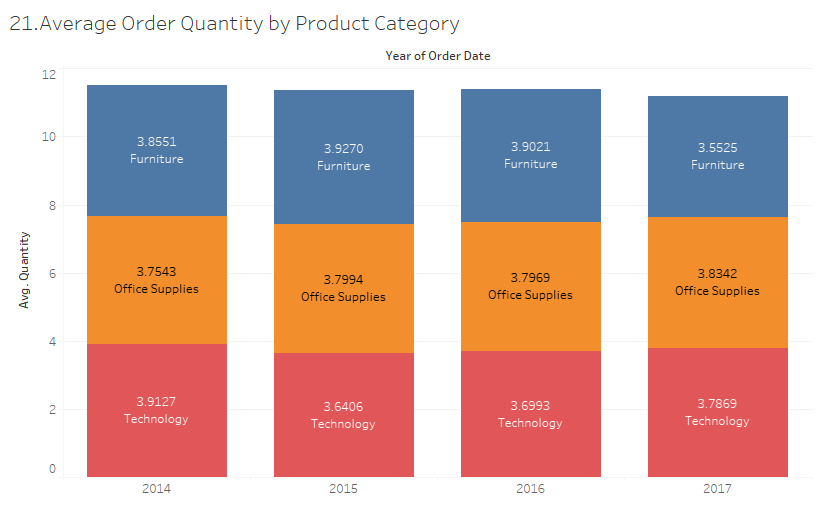
**Insights:** A dual bar chart is used to analyse the multiple levels of data at different data points. Phones and Chairs have the highest sales , and the profit wise Copiers are the highest.

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



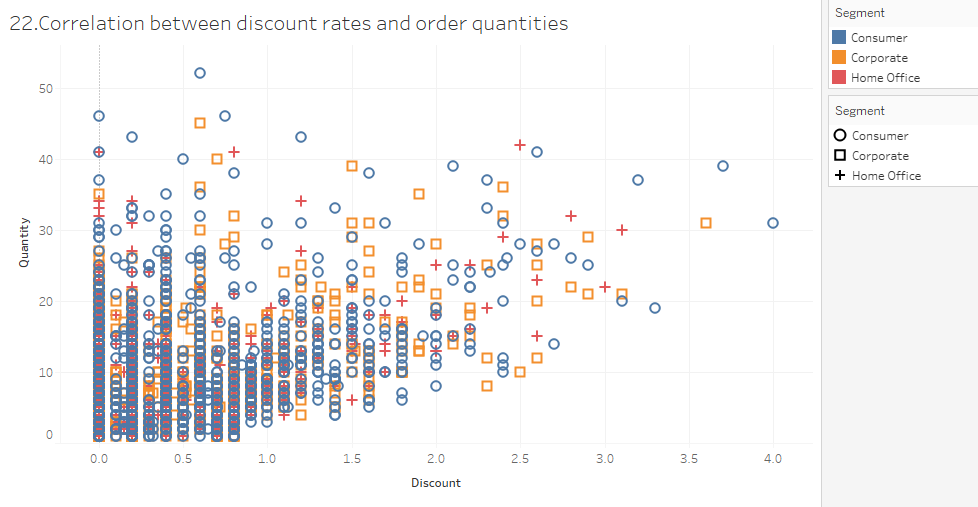
**Insights:** Heat Maps are a type of visualisation that use colour to represent data values in a table.The intensity of the colour represents the value of the data. The Standard Class has the highest intensity of the colour, i.e., the average delivery duration is high. The Same Day has low intensity and average delivery duration is low.

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



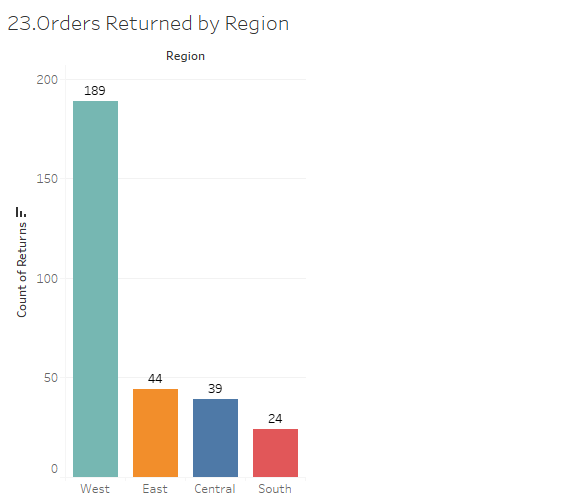
**Insights:** Bar chart is used to analyse data at different data points. We analyse the average order quantity for different product categories over the years. The Office Supplies has the highest average order of 3.8342 for the latest year 2017.

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



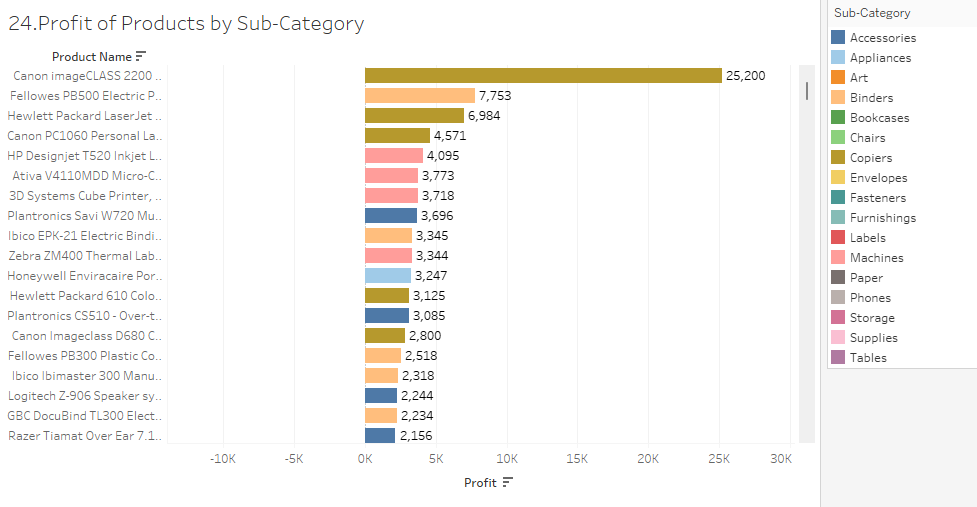
**Insights:** A Scatter plot is used to show the relationship between two variables. The relation between discount rates and order quantities for the different orders at the customer segment level.

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



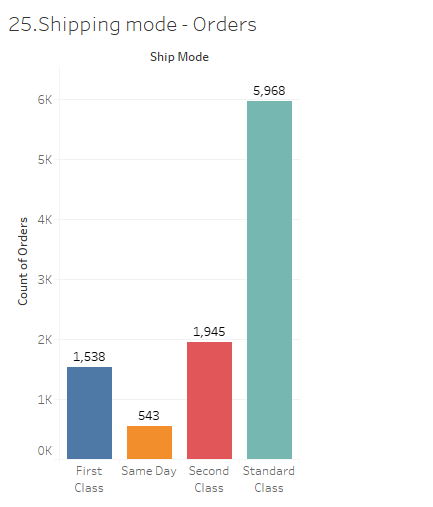
**Insights:** Bar chart is used to analyse data at different data points. The West region has the highest number of returns 189. Number of orders returned in the East is 44, Central is 39 and in the South 24.

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



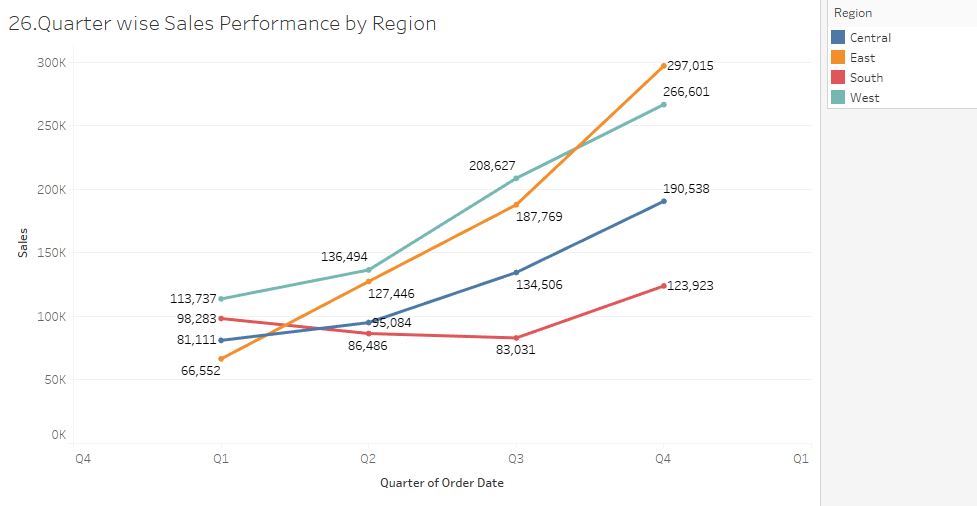
**Insights:** Bar chart is used to analyse data at different data points. Canon imageCLASS2200 is the product with the highest profit of 25,000 which belongs to the Copiers sub-category.

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



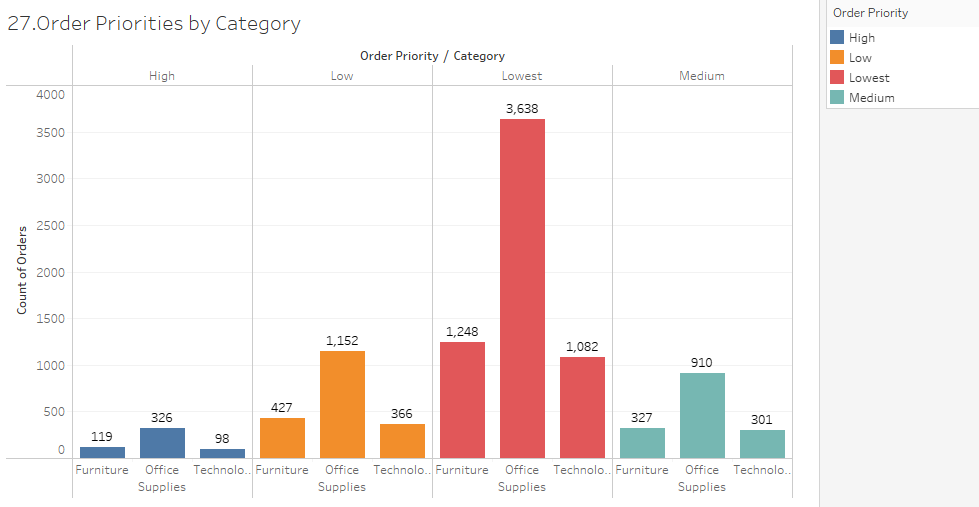
**Insights:** A bar chart is used to compare data across different categories. Standard Class shipping mode is the most common with the highest orders of 5968.

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



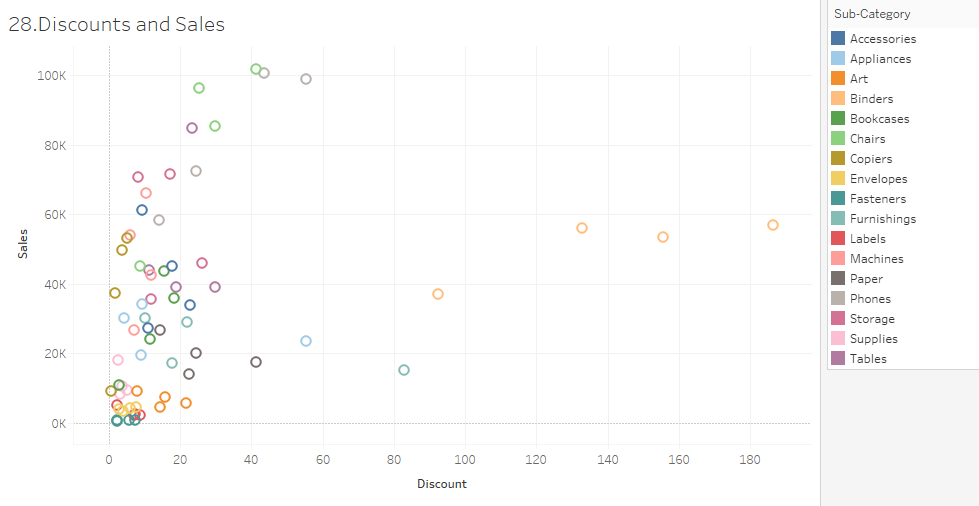
**Insights:** A line chart is used to show trends over time. It is best used when you have data that changes over a period of time. In the quarter Q4, the sales performance is high for all the regions when compared to the previous quarters. We can say that the sales are increasing with respect to the quarter.

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



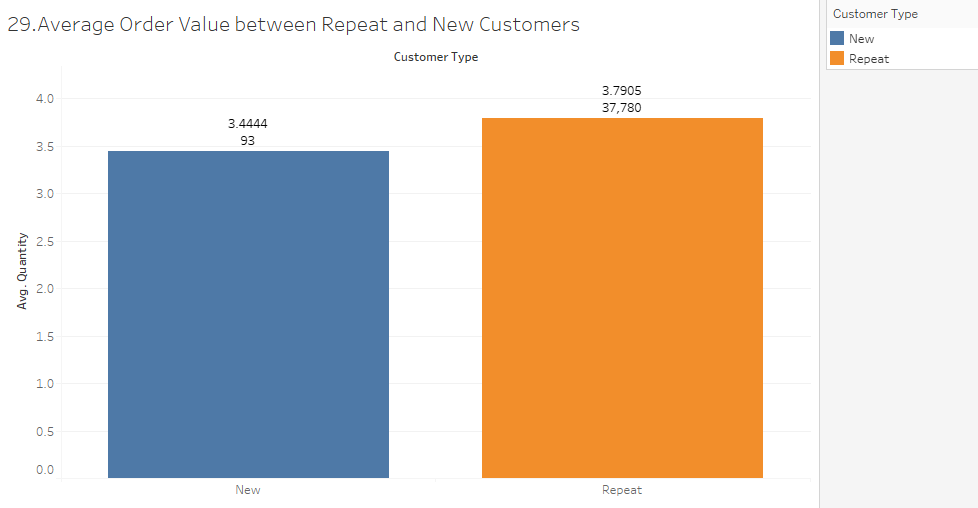
**Insights:** A bar chart is used to compare data across different categories. The high priority orders for Furniture is 119, office Supplies is 326 and Technology is 98. The Office supplies have the lowest priority orders of 3638.

1. What is the relationship between discounts and sales?



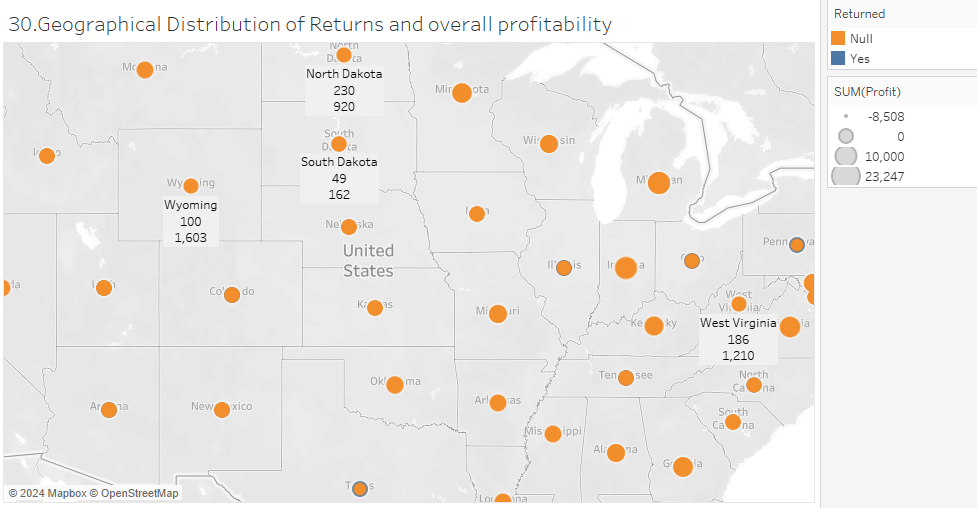
**Insights:** A Scatter plot is used to show the relationship between two variables. We analyse the relationship between sales and discounts for the different sub-categories.

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



**Insights:** A bar chart is used to compare data across different categories. The average order value for repeat customers is 3.7905 and new customers is 3.4444.

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



**Insights:** A map chart is used to show data on a map. It is best used when you want to show the distribution of data across different regions or locations. We can view the region wise distribution of returns and their profits.