**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.

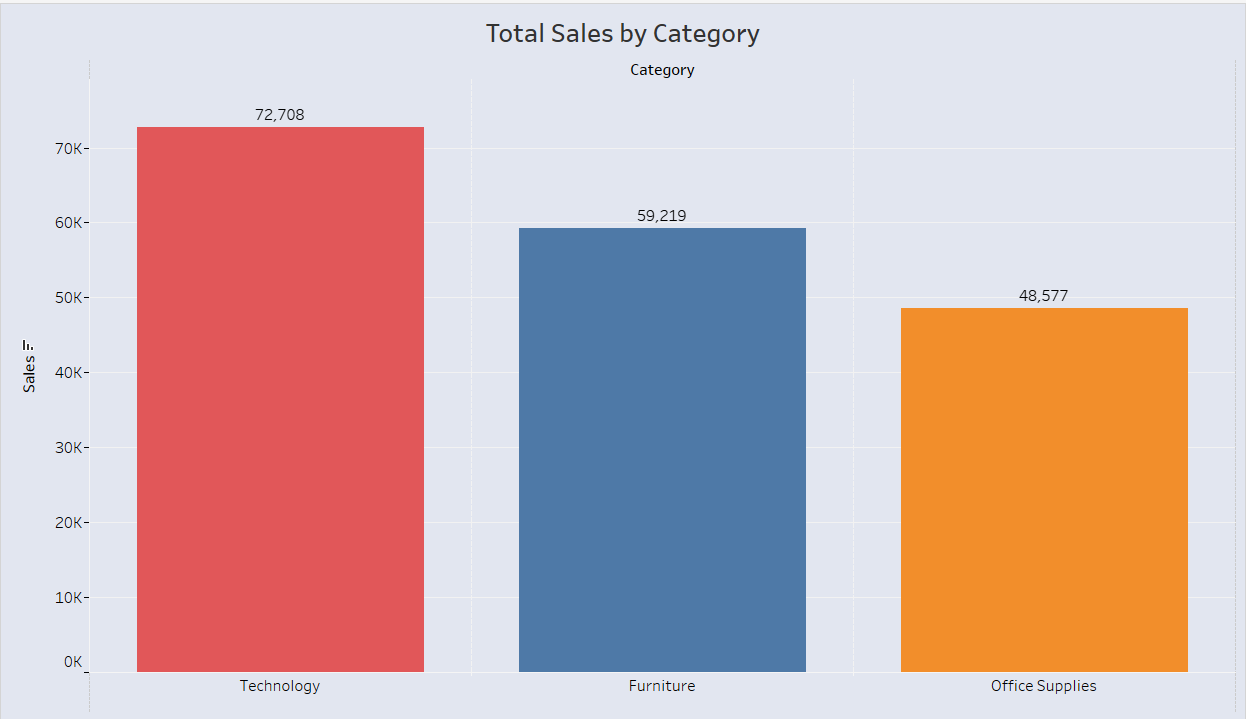
**Problem Statement: Choose the Best chart for any 30 scenario-based questions from Superstore Dataset.**

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.

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

**Questions:**

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

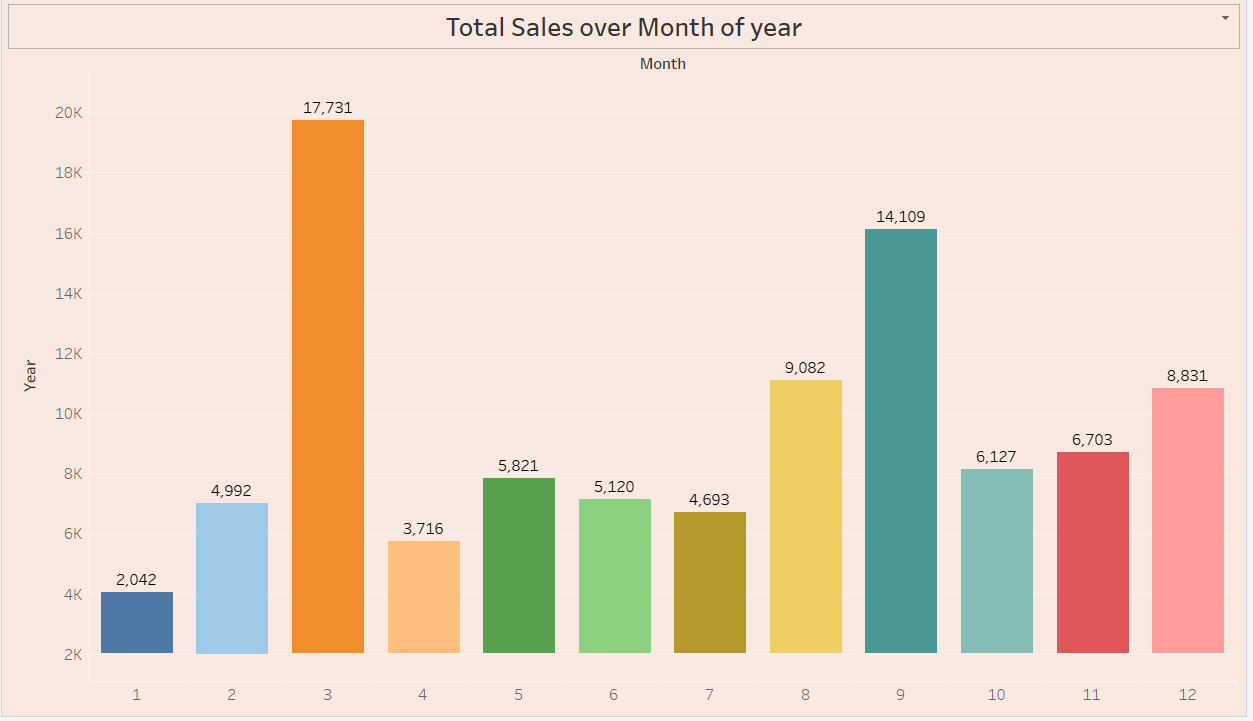
Ans. 

The bar chart titled "Total Sales by Category" shows the total sales for three product categories: Technology, Furniture, and Office Supplies. The y-axis represents the total sales in dollars, and the x-axis represents the product categories.

The bar corresponding to **Technology** is the tallest, indicating that it has the highest total sales. The bar is coloured **red** and has a value of **72,708**. The bar corresponding to **Furniture** is slightly shorter than the Technology bar, indicating that it has the second-highest total sales. The bar is coloured **blue** and has a value of **59,219**. The bar corresponding to **Office Supplies** is the shortest of the three, indicating that it has the lowest total sales. The bar is coloured **orange** and has a value of **48,577**.

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

Ans.



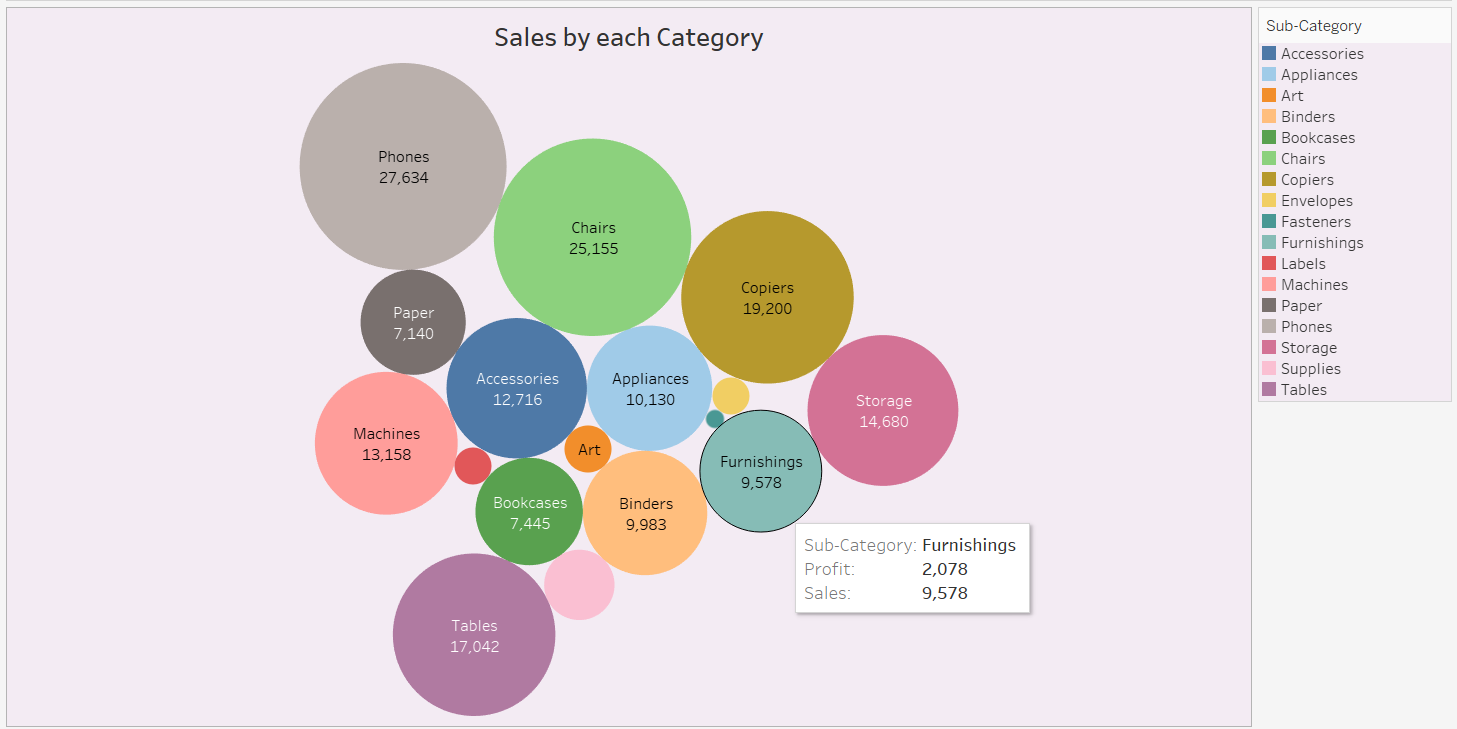
Here are some observations about how the monthly sales amounts change over the course of the year:

* **Overall Trend:** The sales generally increase from January to March, then decline from April to June, and then increase again from July to December.
* **Peak Sales:** The highest sales occur in March, followed by October and September.
* **Lowest Sales:** The lowest sales occur in January, followed by February and April.
* **Seasonal Patterns:** There seems to be a seasonal pattern, with higher sales in the spring and fall months compared to the summer and winter months.
* **Month-to-Month Fluctuations:** There are fluctuations in sales from month to month, even within the general trends. For example, sales increase significantly from February to March, but then decline slightly from March to April.

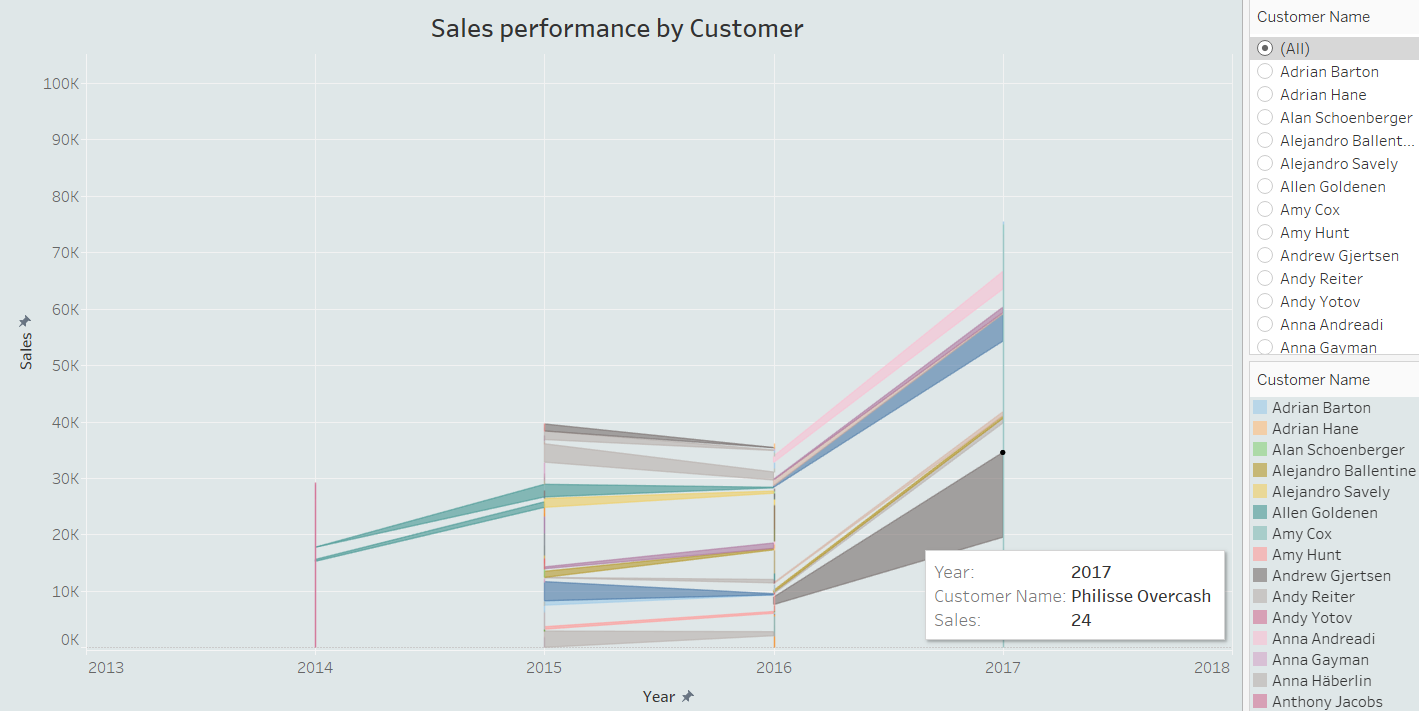
1. How is the total sales amount distributed among different product categories?

Ans. Total sales amount which is distributed among different product categories:

* **Phones** have the highest total sales, followed by **Chairs** and **Copiers**.
* **Paper** has the lowest total sales.
* The total sales of **Accessories** and **Storage** are relatively high, while the total sales of **Appliances** and **Machines** are relatively low.
* Within each category, there are variations in total sales among sub-categories. For example, within the **Furnishings** category, **Tables** have the highest total sales, followed by **Storage** and **Chairs**.



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

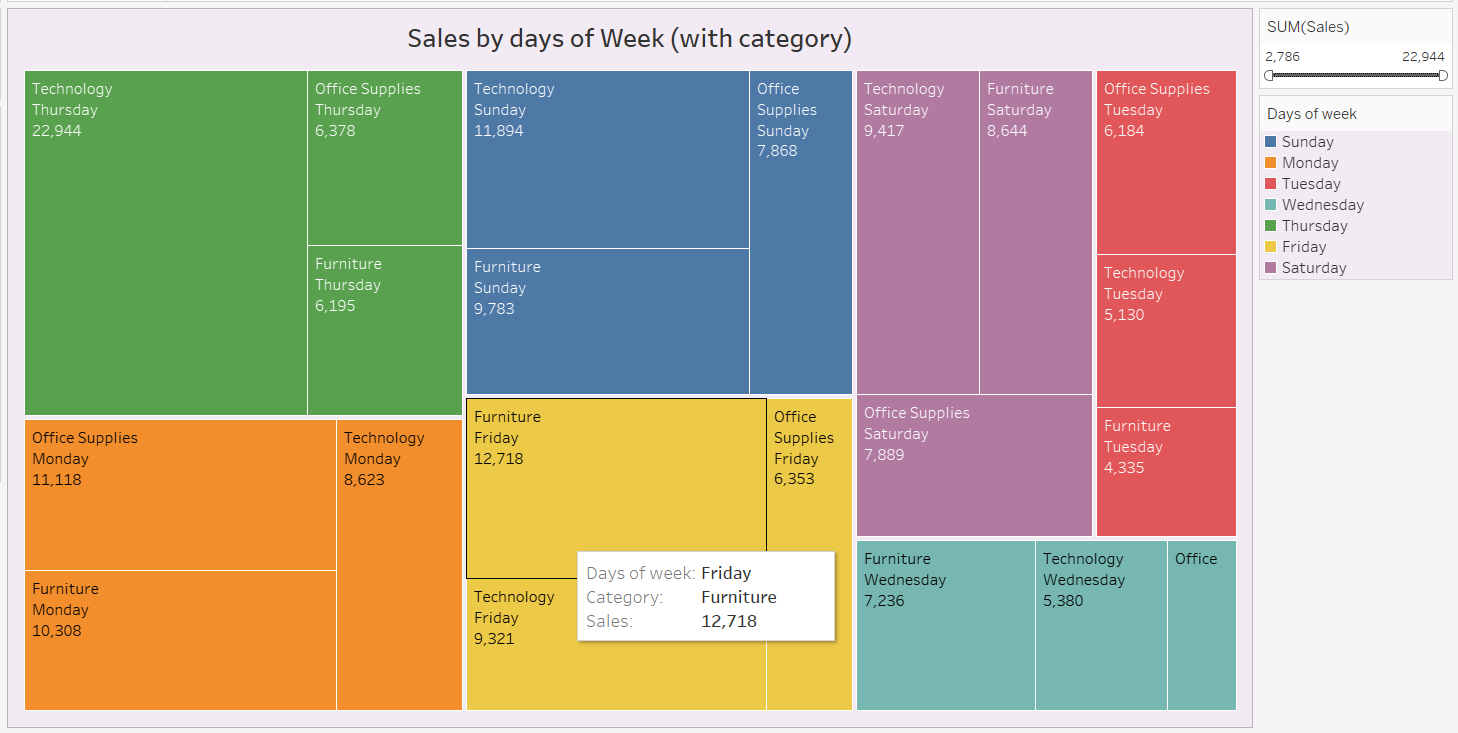
Ans.

**Yes, we can analyze the sales performance of individual customers over time using the provided image.**

The image shows a line chart titled "Sales performance by Customer." The chart displays the sales for each customer over the years 2013 to 2018. Each line represents a different customer, and the colour of each line corresponds to the customer's name.

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

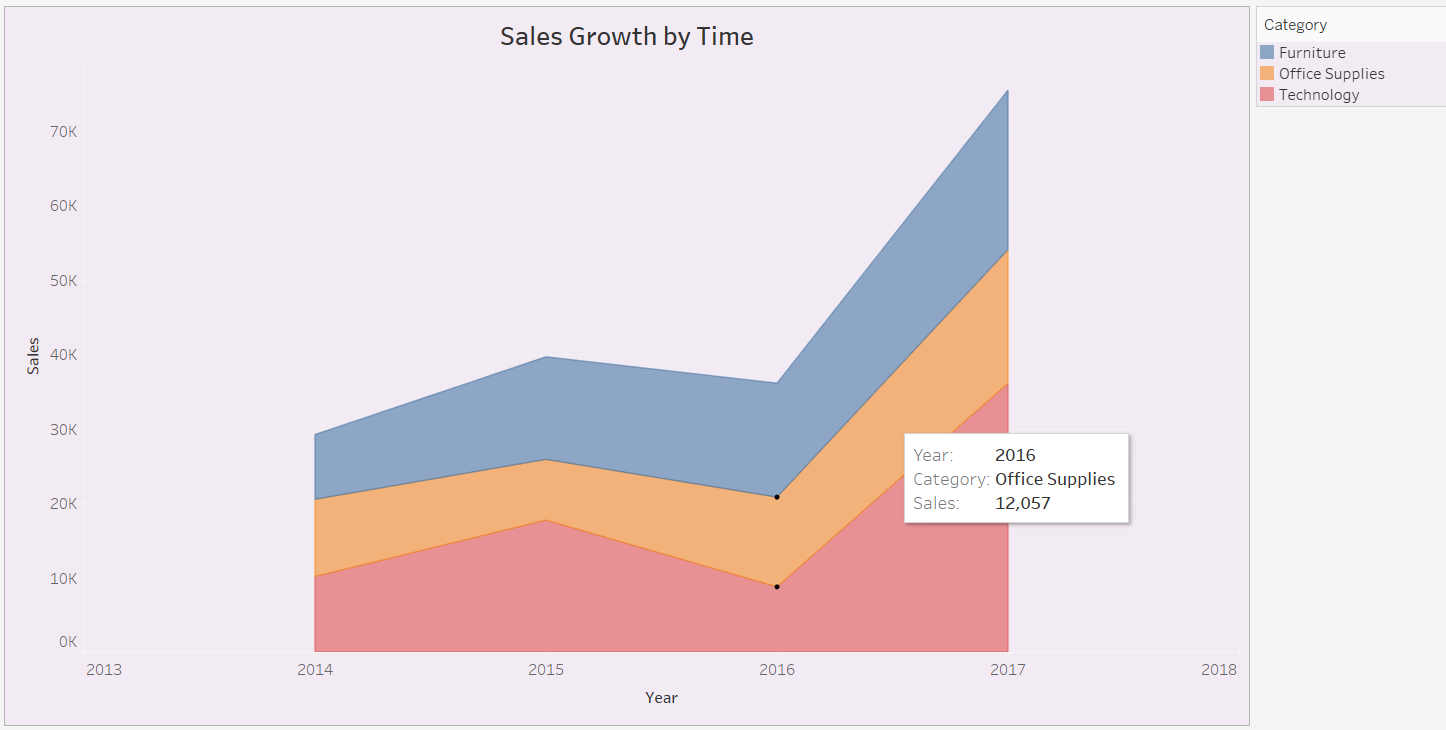
Ans.



The treemap shows that **Technology** sells best on **Thursday**, **Office Supplies** on **Monday** and **Saturday**, and **Furniture** on **Friday**. **Thursday** is the busiest day overall, while **Monday** is the slowest. Sales vary by category on each day.

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

Ans.



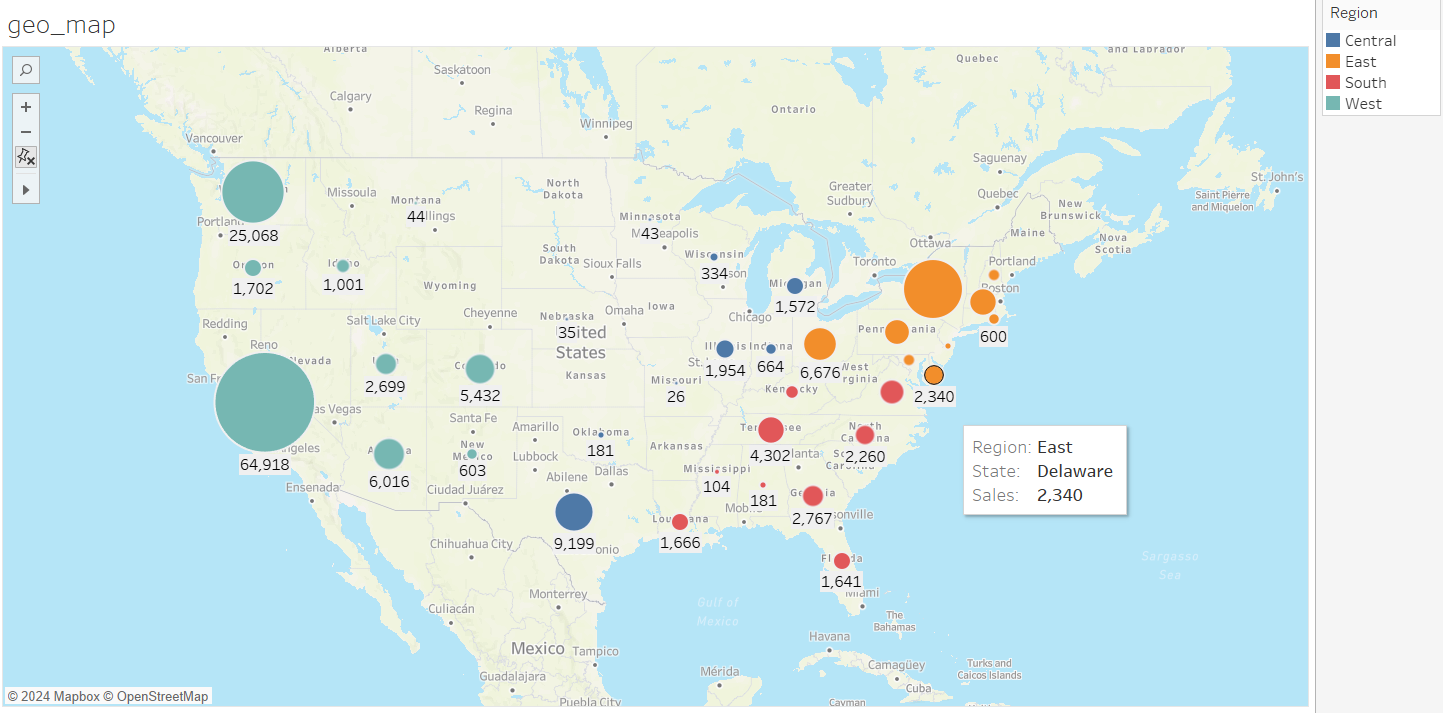
**Yes, we can visualise the sales growth of different product categories over time using the provided chart.**

**Key observations:**

* **Furniture** has consistently been the top-selling category throughout the period.
* **Office Supplies** experienced the most rapid growth, especially in 2016-2017.
* **Technology** has shown steady growth but has not outpaced Office Supplies.

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

Ans.



The image shows a choropleth map titled "geo\_map". The map displays the total sales for different regions in the "Superstore" dataset. The color and size of each bubble represent the total sales of that region.

* **West Region:** The West region has the highest total sales, with several large bubbles indicating high sales in various cities within the region.
* **Central Region:** The Central region also has significant sales, with several large bubbles scattered throughout the region.
* **East Region:** The East region has lower total sales compared to the West and Central regions, with fewer and smaller bubbles.
* **South Region:** The South region has the lowest total sales, with only a few small bubbles indicating minimal sales activity.

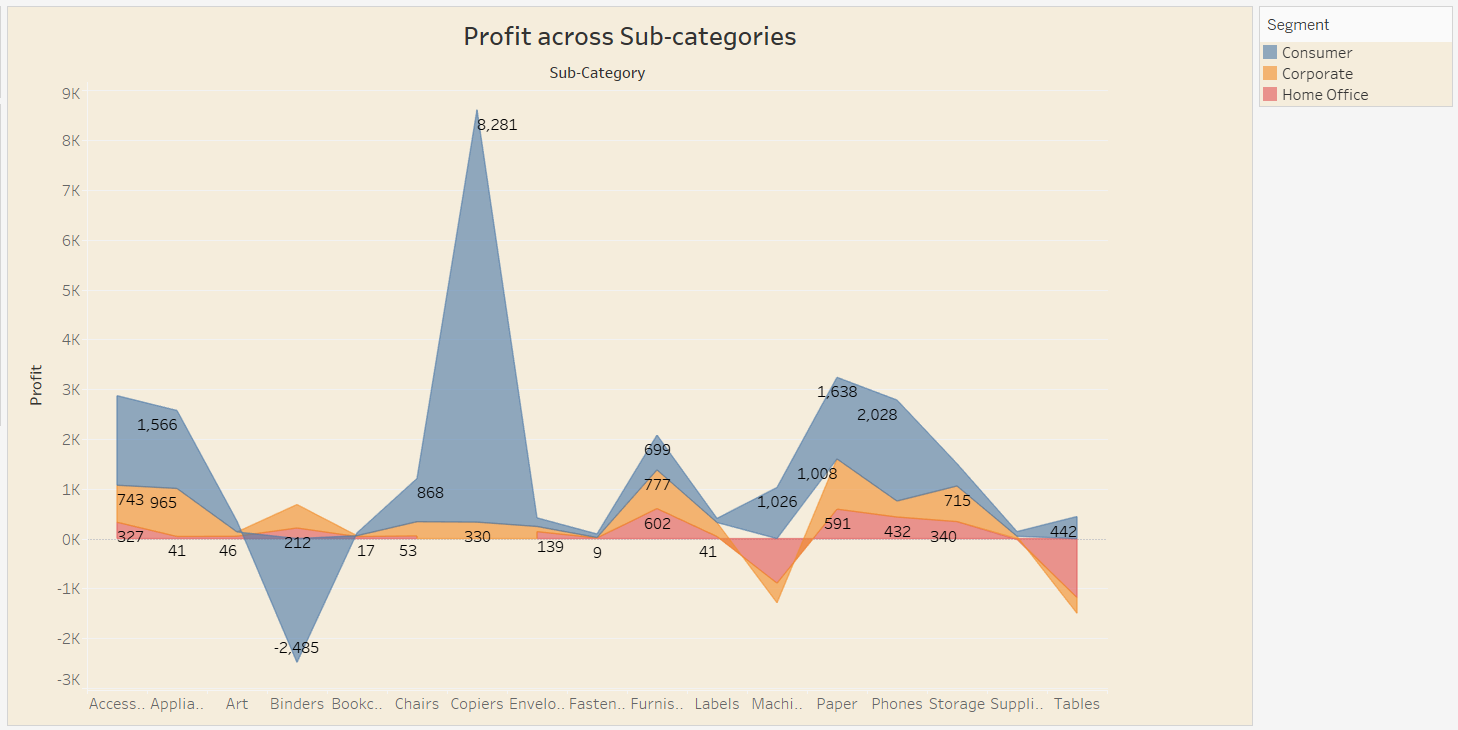
1. Can we visualise the composition of profits across various subcategories within different customer segments?

Ans.

**Yes, we can visualize the composition of profits across various subcategories within different customer segments using the provided chart.**

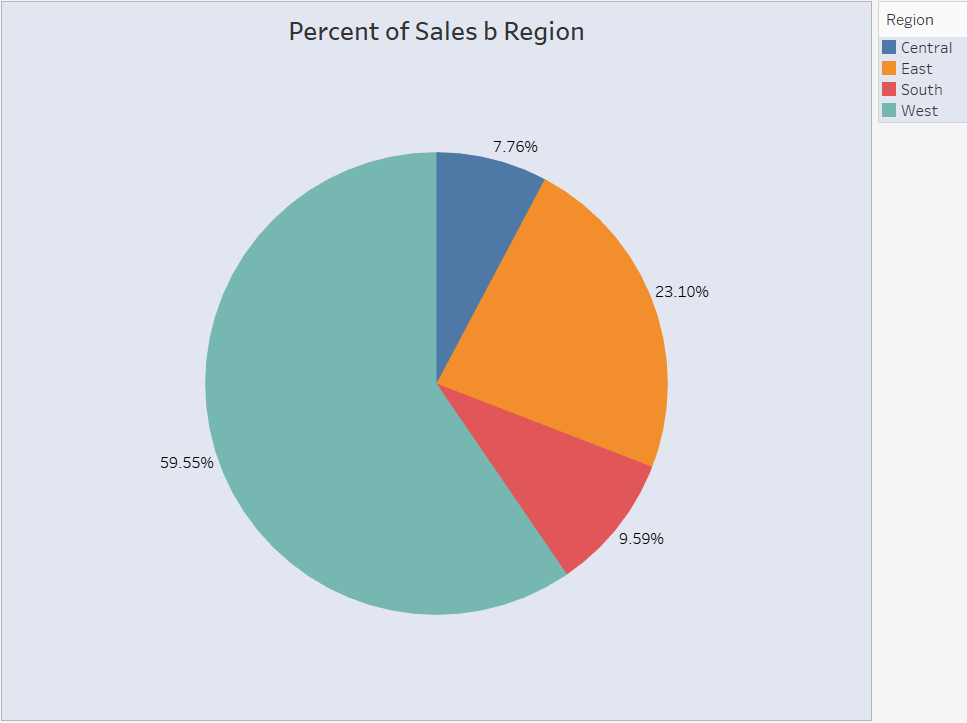
**Key observations:**

* **Consumer Segment:** While there are some profitable subcategories like Copiers and Phones, the overall profit for this segment is relatively low.
* **Corporate Segment:** This segment contributes significantly to the company's profits, with several subcategories like Phones, Storage, and Tables generating substantial revenue.
* **Home Office Segment:** This segment also shows promise, with subcategories like Copiers and Tables contributing positively to the bottom line.



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

Ans.



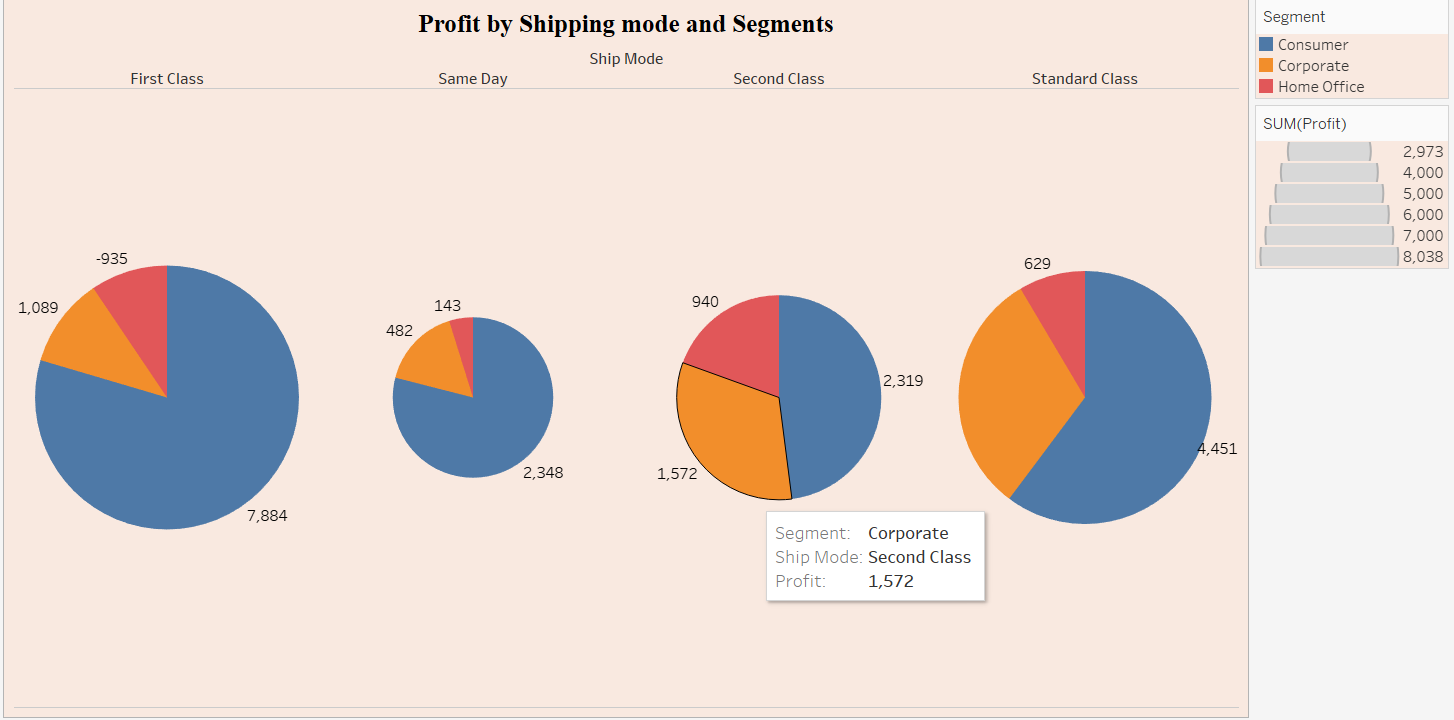
The pie chart titled "Percent of Sales by Region" shows the percentage contribution of each region to the overall sales.

Here is the breakdown of the percentage contribution of each region:

* **West:** 59.55%
* **East:** 23.10%
* **South:** 9.59%
* **Central:** 7.76%

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

Ans.



**Yes, we can visualize the profit margins associated with different shipping modes and customer segments using the provided pie charts.**

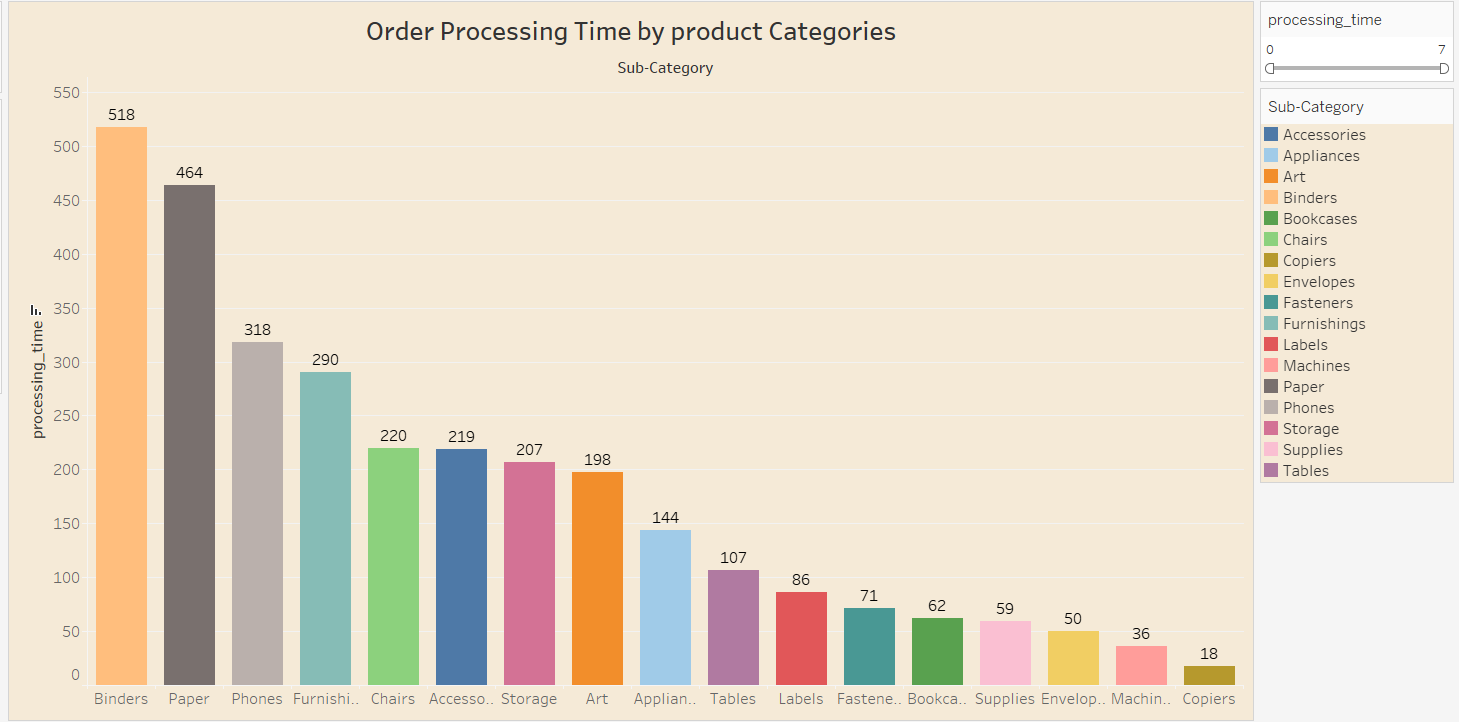
**Key observations:**

* **First Class:** While it's the most expensive shipping mode, it doesn't always yield the highest profits. For example, in the "Consumer" segment, "Standard Class" is more profitable.
* **Same Day:** This shipping mode generally generates lower profits compared to others, especially for the "Home Office" segment.
* **Second Class:** This mode is consistently profitable across all segments, making it a reliable choice.
* **Standard Class:** While it's the most affordable option, it can still be profitable, especially for the "Corporate" segment.

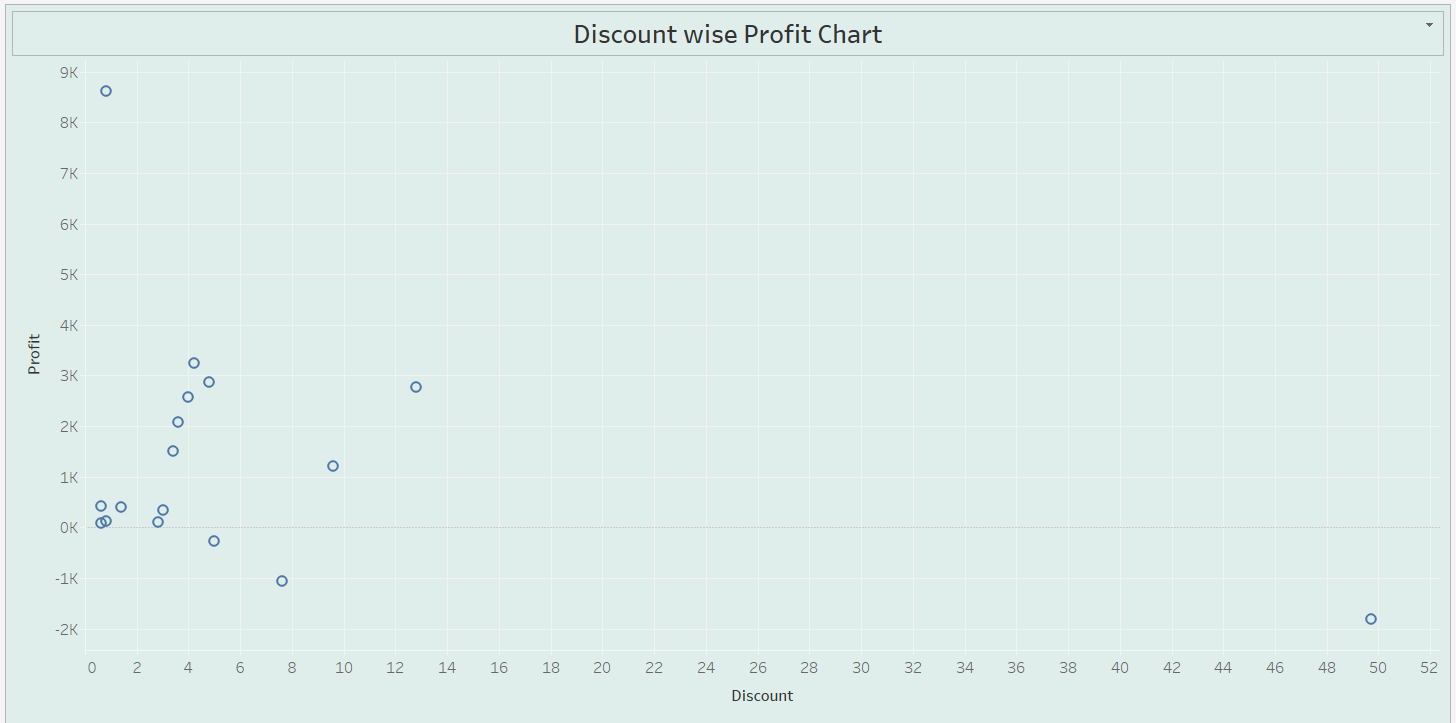
1. How long does it take to process orders for different product categories?

Ans. The bar chart titled "Order Processing Time by Product Categories" shows the processing time for different product categories. The x-axis represents the product categories, and the y-axis represents the processing time in days.

* **Binders** have the longest processing time, followed by **Paper** and **Phones**.
* **Copiers** have the shortest processing time, followed by **Machines** and **Envelopes**.
* There is a wide range of processing times among different product categories. For example, the processing time for **Binders** is more than 5 times the processing time for **Copiers**.
* Some product categories have similar processing times. For example, **Furnishings**, **Chairs**, and **Accessories** all have processing times around 200 days.



1. How do discounts affect overall profit?

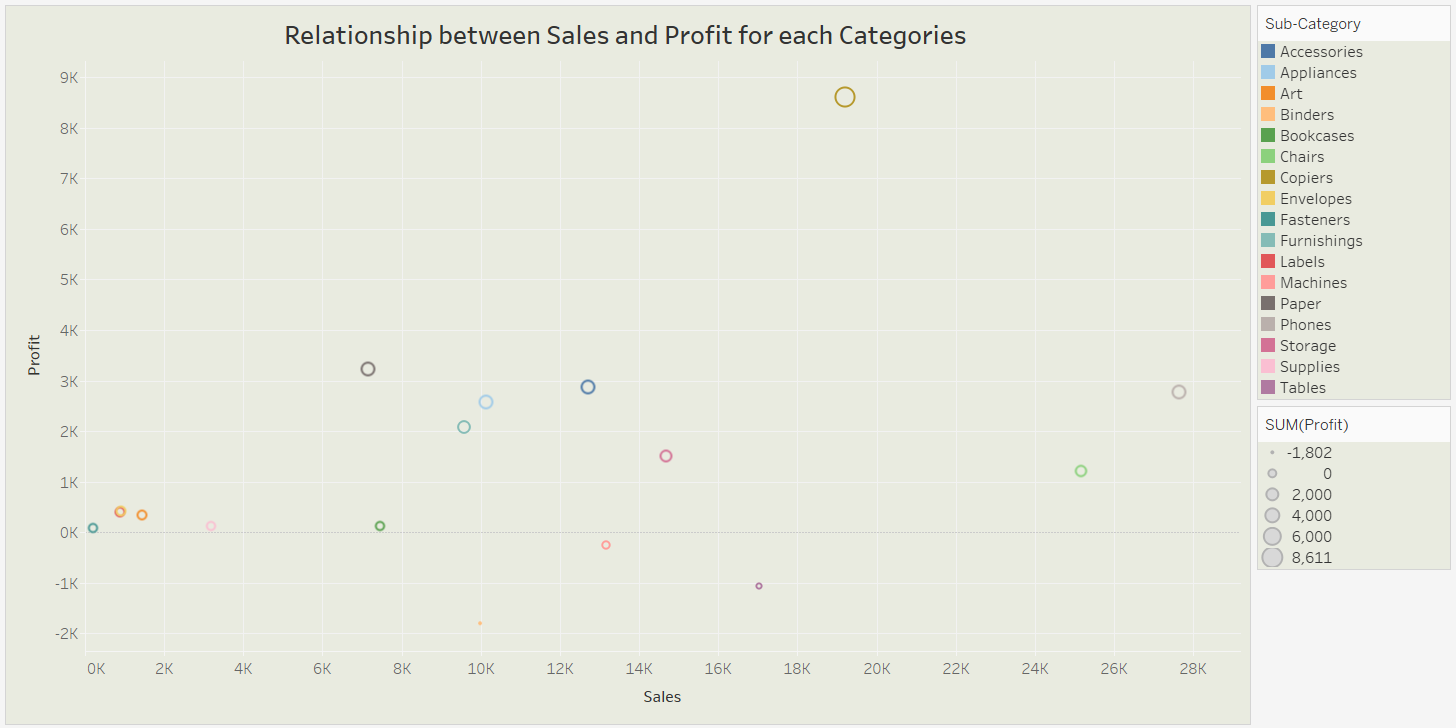
Ans. 

The scatter plot titled "Discount wise Profit Chart" shows the relationship between discounts and profits. The x-axis represents the discount, and the y-axis represents the profit.

* **Negative Correlation:** There is a negative correlation between discounts and profits. This means that as discounts increase, profits tend to decrease.
* **Outlier:** There is one outlier in the data, which is a point with a high discount and a relatively high profit. This point might be skewing the overall relationship between discounts and profits.
* **Clustering:** The data points tend to cluster in the lower left corner of the chart, indicating that most products with discounts have relatively low profits.
* **Individual Variation:** While there is a general trend of decreasing profits with increasing discounts, there is also individual variation. Some products with high discounts may still have relatively high profits, while others with low discounts may have relatively low profits.

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

Ans.

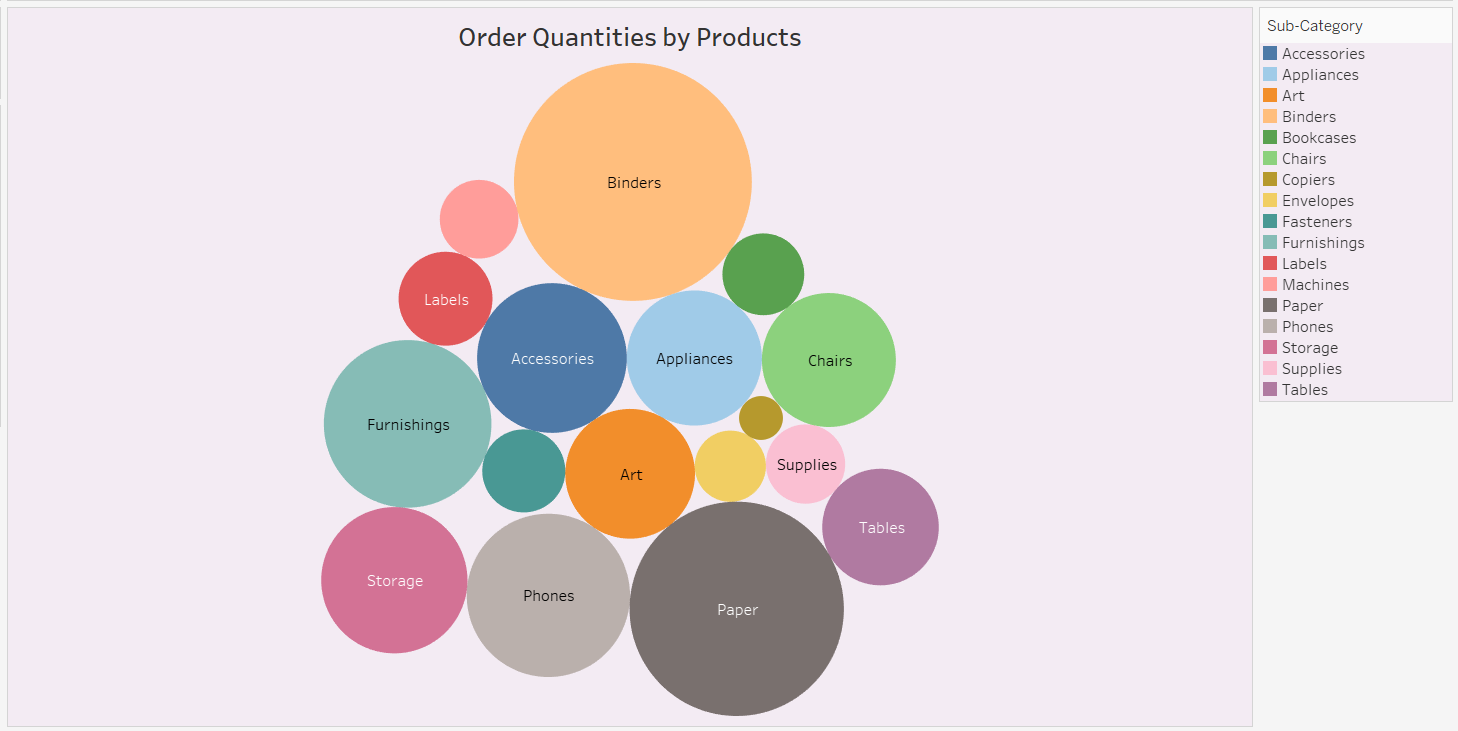


**Yes, we can visualize the relationship between product sales and profitability for different product categories.**

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

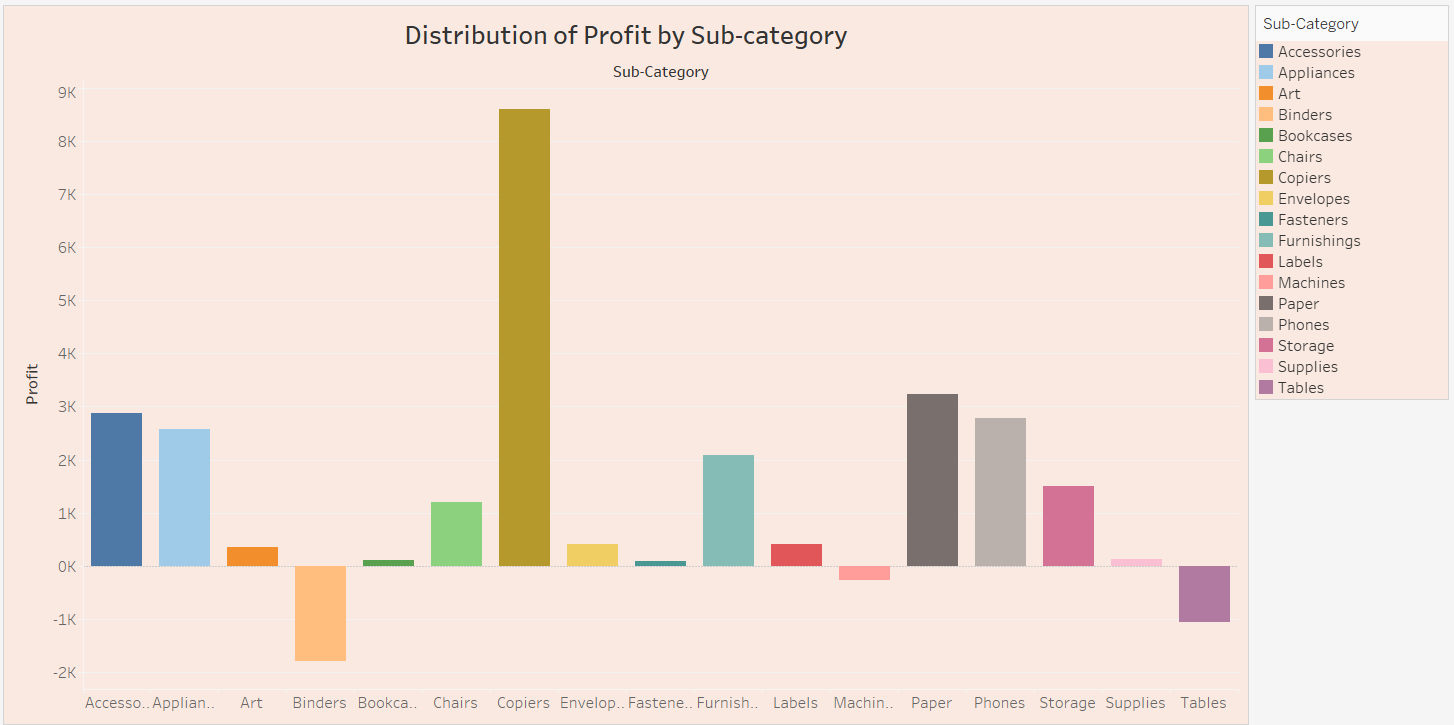
Ans. The bubble chart titled "Order Quantities by Products" shows the distribution of order quantities for different products in the dataset. The size of each bubble represents the order quantity of that product.

* **Binders** have the highest order quantity, followed by **Phones** and **Accessories**.
* **Paper** has the lowest order quantity.
* There is a wide range of order quantities among different products. For example, the order quantity for **Binders** is more than 10 times the order quantity for **Paper**.
* Some products have similar order quantities. For example, **Furnishings**, **Chairs**, and **Appliances** all have order quantities around 1000.



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

Ans.



The bar chart titled "Distribution of Profit by Sub-category" shows the profit distribution for different sub-categories within each product category. The x-axis represents the sub-categories, and the y-axis represents the profit.

* **Office Supplies:** Within the Office Supplies category, **Copiers** have the highest profit, followed by **Paper** and **Storage**. **Binders** and **Envelopes** have negative profits.
* **Technology:** Within the Technology category, **Phones** have the highest profit, followed by **Copiers** and **Accessories**. **Machines** have a negative profit.
* **Furniture:** Within the Furniture category, **Chairs** have the highest profit, followed by **Tables** and **Furnishings**. **Bookcases** have a negative profit.

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

Ans.



**Yes, we can compare the shipping time distributions for different shipping modes.**

The bar chart titled "Shipping Time by Shipping Modes" shows the distribution of shipping times for different shipping modes. The x-axis represents the shipping modes, and the y-axis represents the distinct count of processing time.

* **Standard Class** has the highest number of distinct processing times, followed by **Second Class** and **First Class**. **Same Day** has the lowest number of distinct processing times.
* **First Class** and **Second Class** have similar shipping time distributions, with a few distinct processing times.
* **Standard Class** has a wider range of shipping times, with a higher number of distinct processing times.
* **Same Day** has a narrower range of shipping times, with a lower number of distinct processing times.

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

Ans.

The line chart titled "Monthly trend by Order Shipped" shows the monthly trend in the number of orders shipped. The x-axis represents the months, and the y-axis represents the number of orders shipped.

* **Seasonality:** There appears to be a seasonal pattern in the number of orders shipped. The number of orders tends to be higher in the spring and fall months compared to the summer and winter months.
* **Growth:** The overall trend is upward, indicating that the number of orders shipped has been increasing over time.
* **Fluctuations:** There are fluctuations in the number of orders shipped from month to month, even within the general trends. For example, there is a significant increase in orders shipped in December, followed by a decline in January.



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

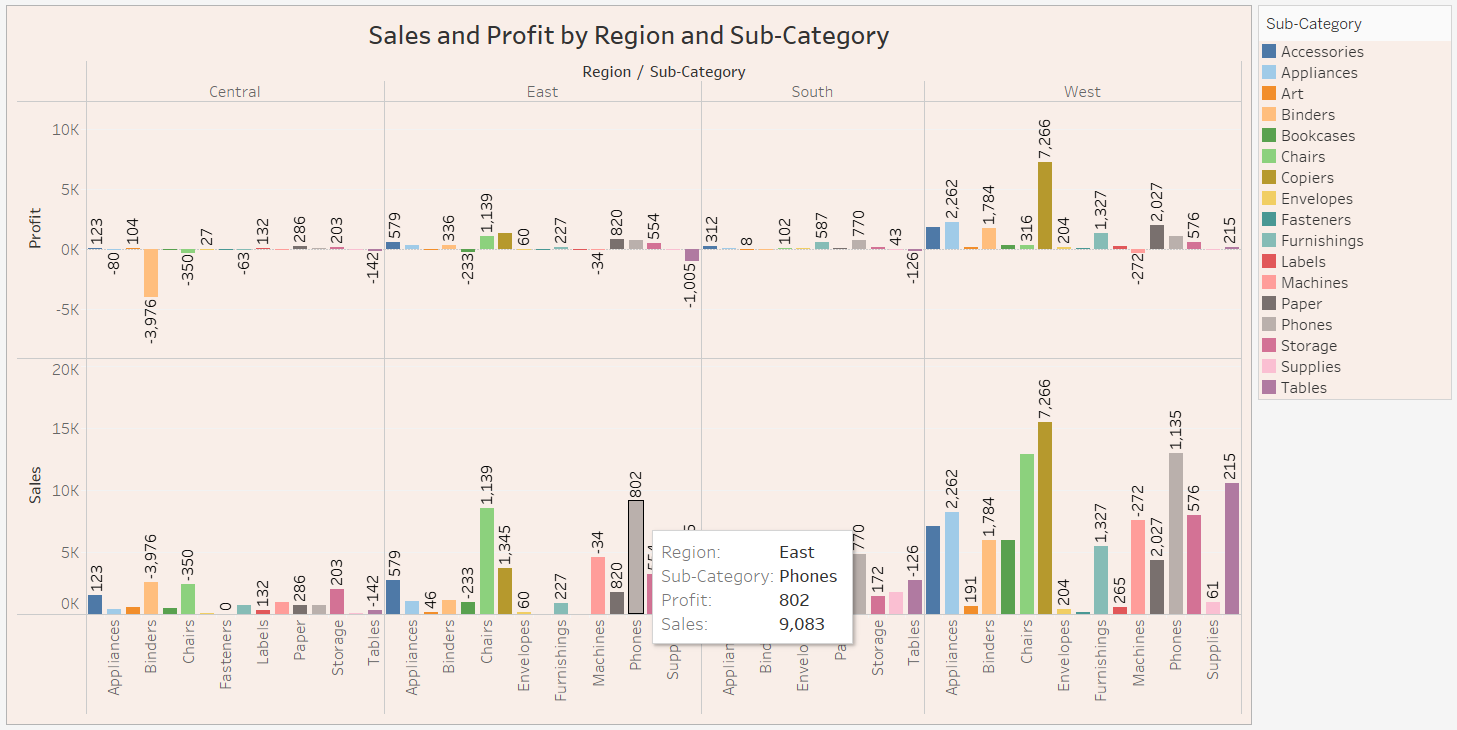
Ans.



The scatter plot titled "Sales and Discount by Customer segment" shows the relationship between sales and discount rates for different customer segments. The x-axis represents the sales, and the y-axis represents the discount. Each colour represents a different customer segment.

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

Ans.



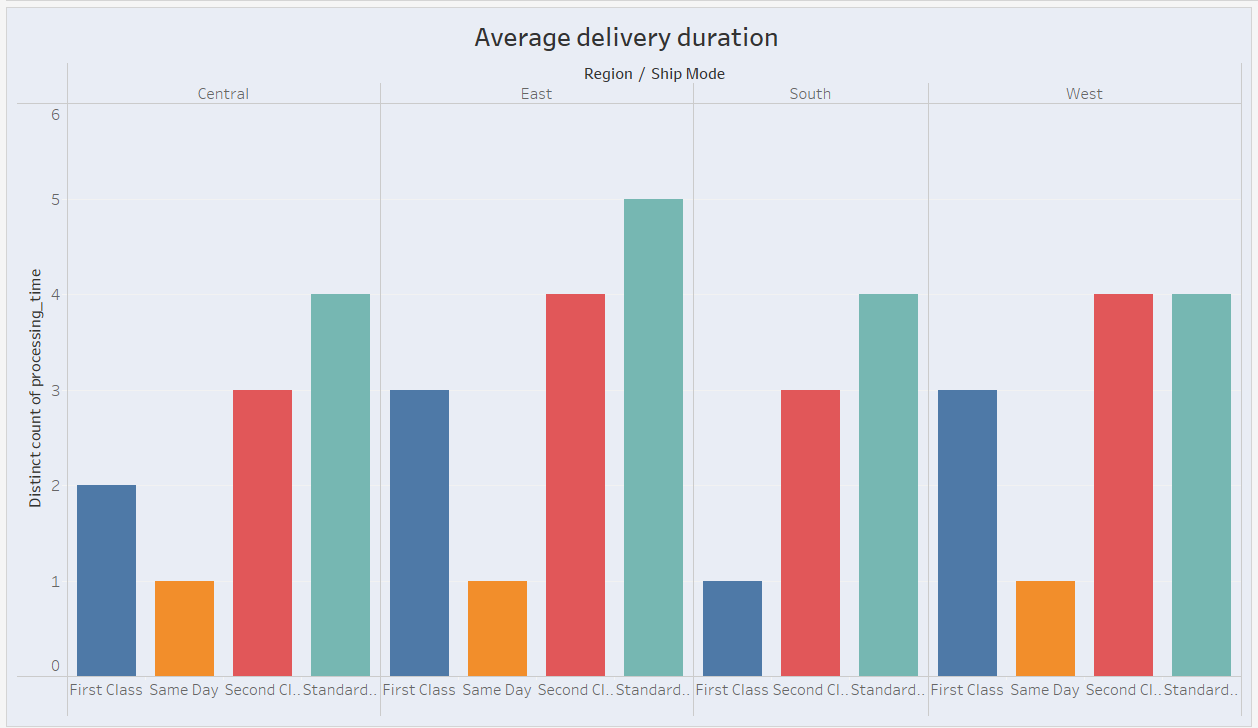
The bar plot titled "Sales and Profit by Region and Sub-Category" shows the sales and profit trends for different product subcategories and regions in the Superstore dataset. The x-axis represents the subcategories, the y-axis represents the regions, the colour represents the sales, and the size of the bubble represents the profit.

Here are some observations about the sales and profit trends across different product subcategories and regions:

* **West Region:** The West region has the highest sales and profits overall, with several subcategories contributing significantly to the total sales and profits.
* **East Region:** The East region has moderate sales and profits, with a few subcategories contributing significantly to the total sales and profits.
* **Central Region:** The Central region has low sales and profits overall, with most subcategories contributing minimally to the total sales and profits.
* **South Region:** The South region has the lowest sales and profits overall, with most subcategories contributing minimally to the total sales and profits.
* **Office Supplies:** Within the Office Supplies category, **Copiers** and **Paper** have the highest sales and profits in the West region, while **Binders** and **Envelopes** have negative profits in the South region.
* **Technology:** Within the Technology category, **Phones** have the highest sales and profits in all regions.
* **Furniture:** Within the Furniture category, **Chairs** and **Tables** have the highest sales and profits in the West region, while **Bookcases** have negative profits in the South region.

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

Ans.



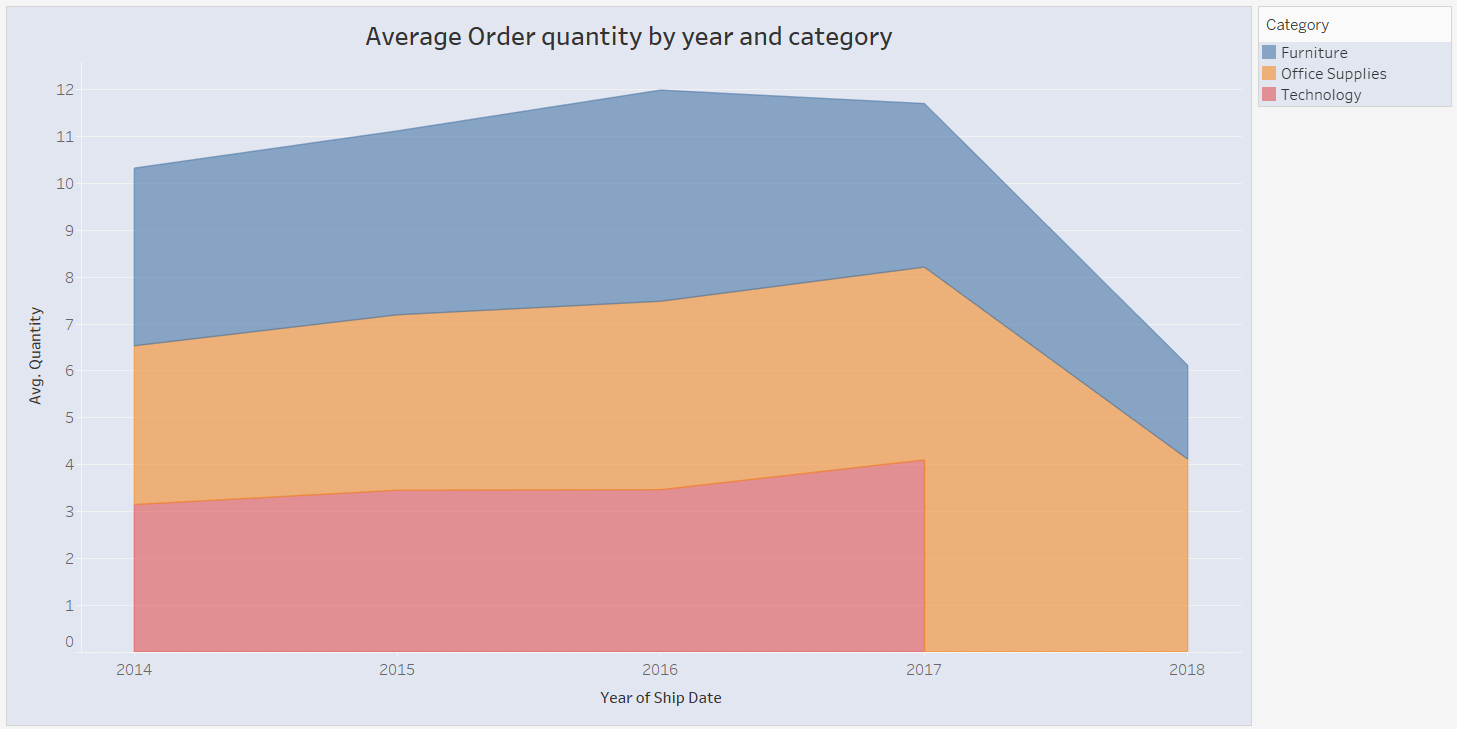
The bar chart titled "Average delivery duration" shows the average delivery duration for different regions and ship modes. The x-axis represents the regions and ship modes, and the y-axis represents the average delivery duration.

* **Central Region:** The average delivery duration for the Central region is the lowest for all ship modes.
* **West Region:** The average delivery duration for the West region is the highest for all ship modes.
* **East Region:** The average delivery duration for the East region is moderate for all ship modes.
* **South Region:** The average delivery duration for the South region is also moderate for all ship modes.
* **Standard Class:** The average delivery duration for Standard Class is the highest for all regions.
* **First Class:** The average delivery duration for First Class is the lowest for all regions.
* **Second Class:** The average delivery duration for Second Class is moderate for all regions.
* **Same Day:** The average delivery duration for Same Day is the lowest for all regions.

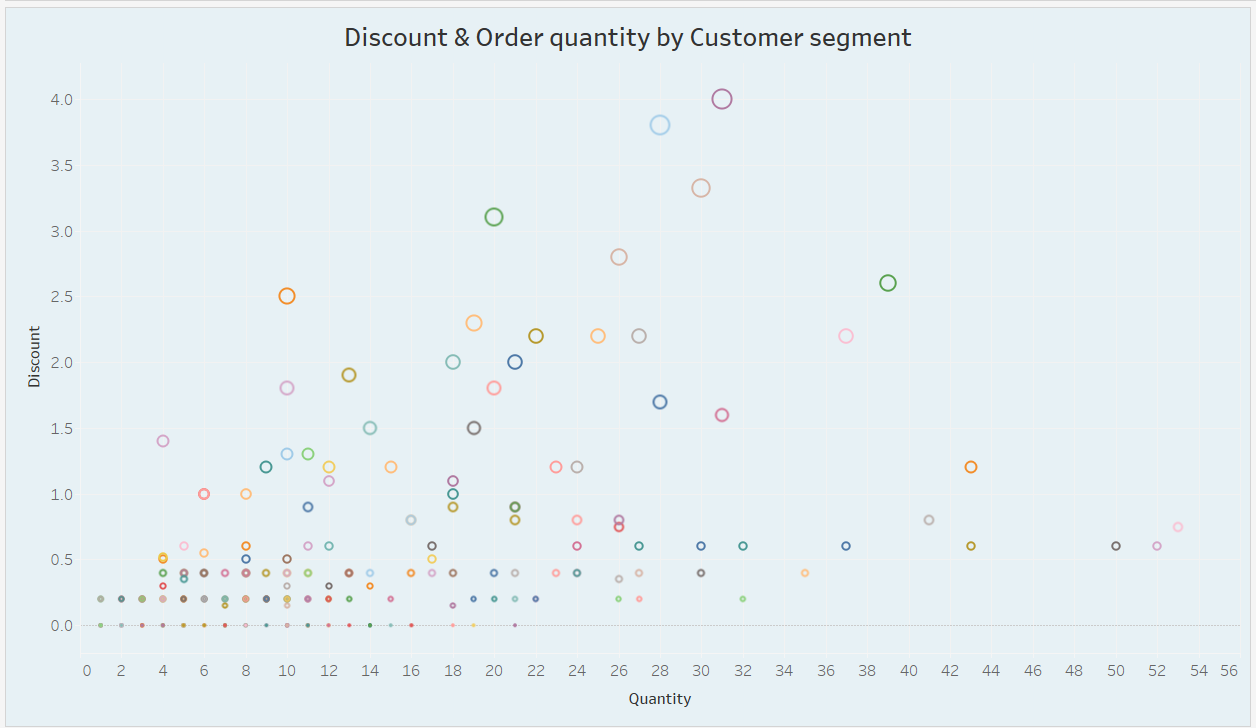
1. How has the average order quantity changed over the years for various product categories?

Ans. The line chart titled "Average Order quantity by year and category" shows the average order quantity for various product categories over the years. The x-axis represents the year of ship date, and the y-axis represents the average order quantity.

* **Furniture:** The average order quantity for furniture increased from 2014 to 2017, then declined in 2018.
* **Office Supplies:** The average order quantity for office supplies remained relatively stable from 2014 to 2016, then increased in 2017 before declining in 2018.
* **Technology:** The average order quantity for technology increased from 2014 to 2015, then declined from 2015 to 2017 before increasing again in 2018.



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

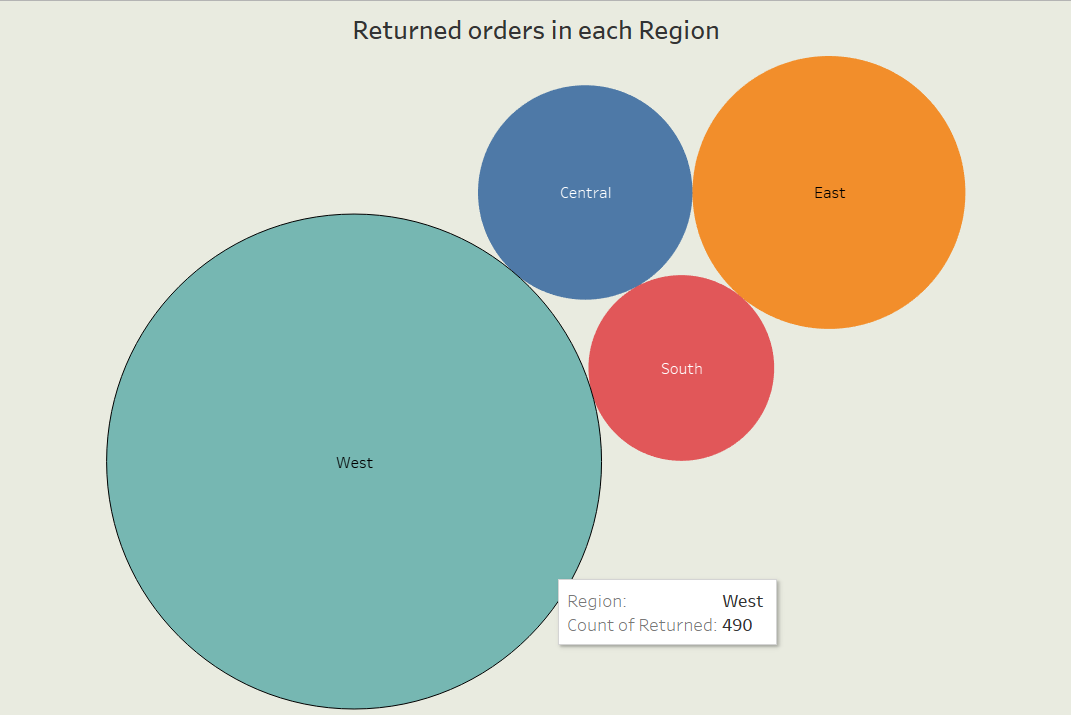
Ans.

**Yes, we can visualize the correlation between discount rates and order quantities for different customer segment.**

* **Overall Trend:** There seems to be a general trend where higher discounts are associated with larger order quantities. However, this relationship is not perfectly linear.
* **Customer Segment Variation:** The scatter plot shows that different customer segments exhibit varying patterns. For instance, some segments might have a stronger correlation between discounts and order quantities than others.
* **Outliers:** There are a few data points that deviate from the general trend, indicating exceptional cases where discounts might not have the expected impact on order quantities.

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

Ans.



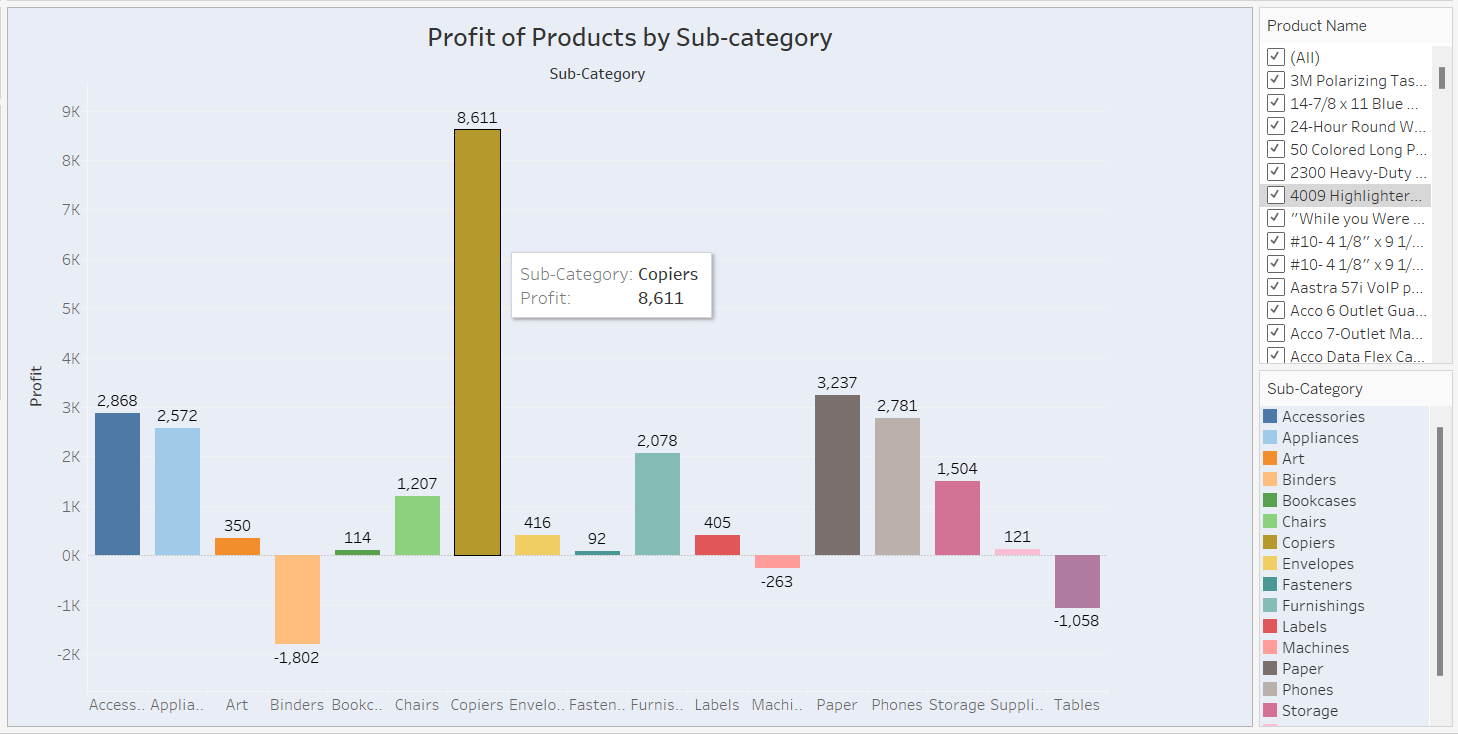
The bubble chart titled "Returned orders in each Region" shows the proportion of orders returned in each region within the Superstore dataset. The size of each bubble represents the number of orders returned in that region.

* **West:** 490
* **East:** 179
* **Central:** 167
* **South:** 102

The chart shows that the West region has the highest number of returned orders, followed by the East, Central, and South regions.

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

Ans.

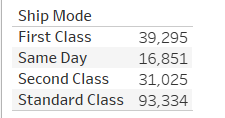


**Yes, we can compare the profit of different products for different subcategories.**

* **Copiers:** Within the "Copiers" subcategory, "3M Polarising Task Lamp" stands out with the highest profit.
* **Accessories:** "Acco Data Flex Cable" is the most profitable product in this subcategory.
* **Art:** "14-7/8 x 11 Blue Portfolio" is the top-selling product with the highest profit.
* **Binders:** "24-Hour Round Wirebound Notebook" is the most profitable binder.

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

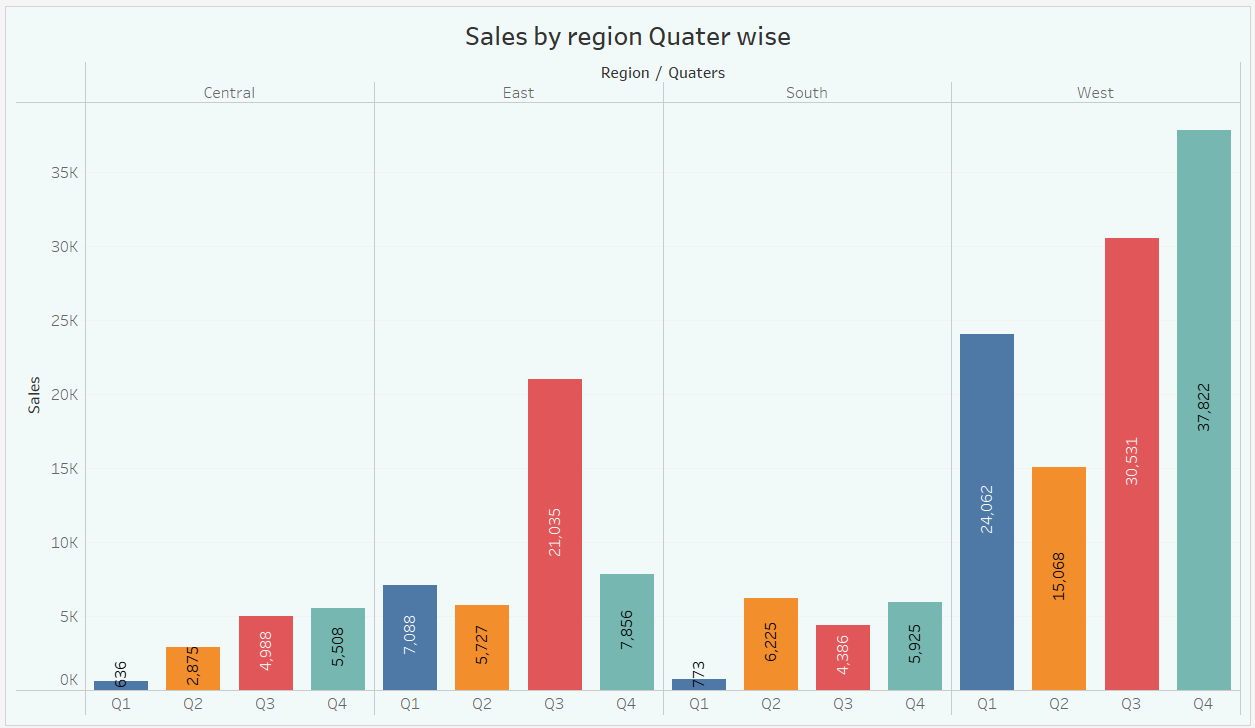
Ans.



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

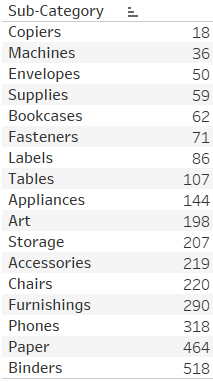
And.The bar chart titled "Sales by region Quarter wise" shows the sales performance of different regions throughout the quarters of a year. The x-axis represents the quarters, the y-axis represents the sales, and each colour represents a different region.

* **West Region:** The West region consistently has the highest sales throughout all quarters, with a significant increase in sales in the fourth quarter.
* **East Region:** The East region has moderate sales throughout the year, with a slight increase in sales in the third quarter.
* **South Region:** The South region has the lowest sales throughout the year, with a slight increase in sales in the second quarter.
* **Central Region:** The Central region has moderate sales throughout the year, with a slight increase in sales in the first quarter.

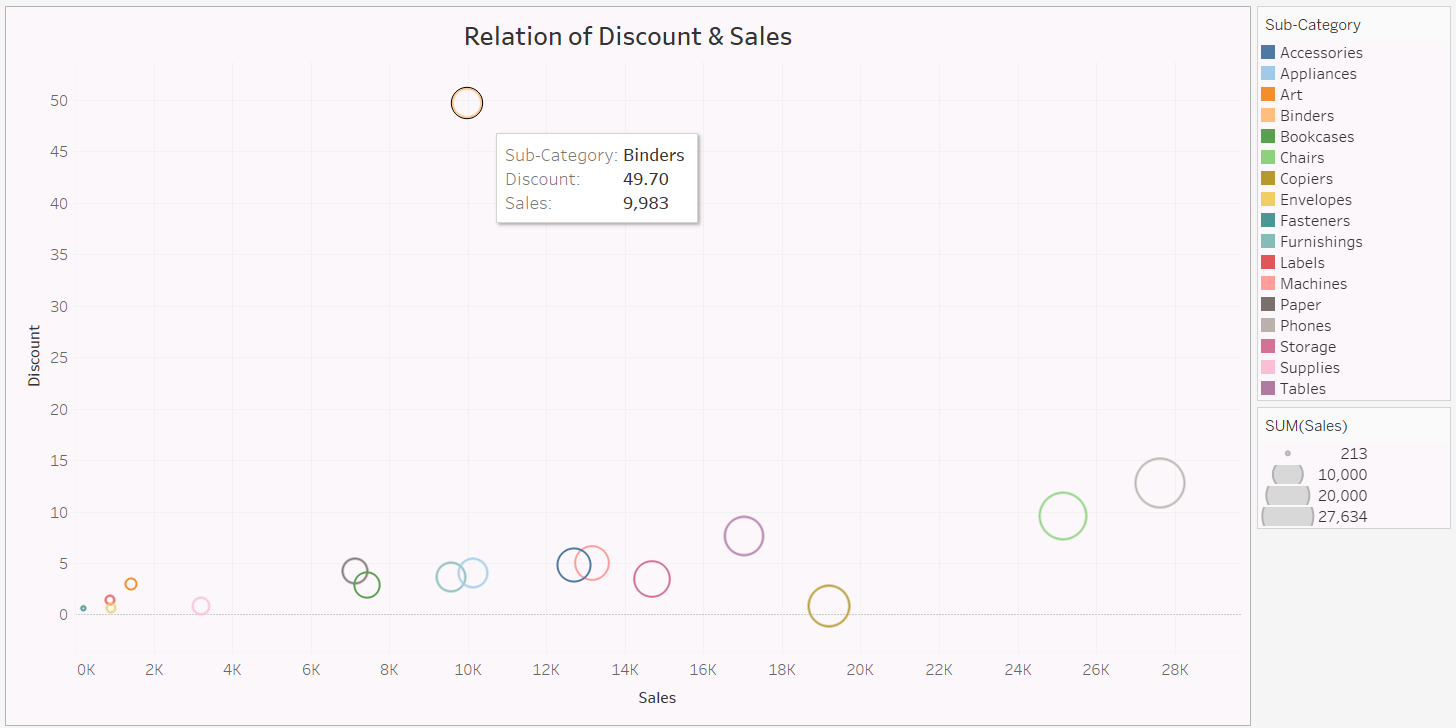


1. What is the distribution of order priorities across different product categories?

Ans.



1. What is the relationship between discounts and sales?

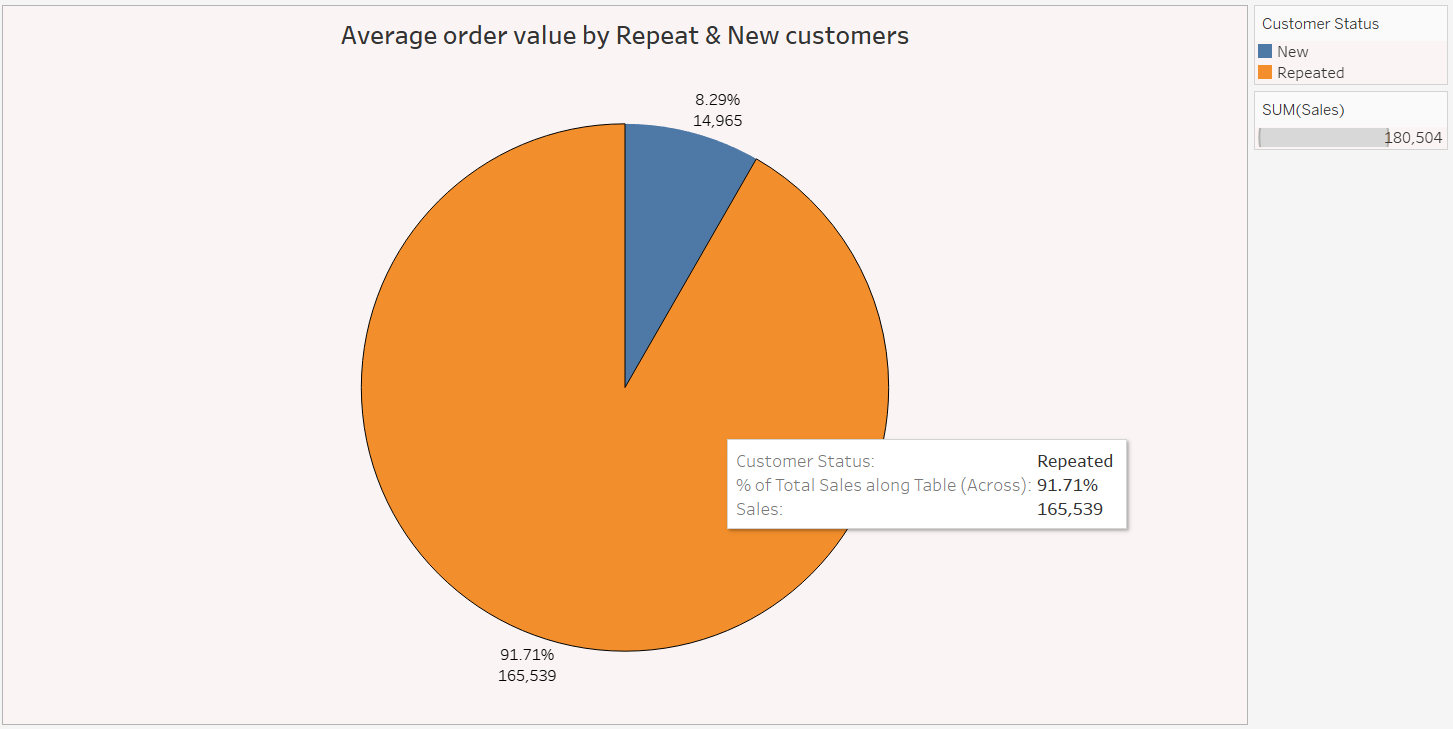
Ans. 

The scatter plot titled "Relation of Discount & Sales" shows the relationship between discount and sales for different product categories. The x-axis represents the sales, and the y-axis represents the discount. Each colour represents a different product category.

* **Higher Discounts, Lower Priority:** Products with higher discounts tend to have lower sales, which might suggest that they are not considered as important or urgent by customers.
* **Lower Discounts, Higher Priority:** Products with lower discounts tend to have higher sales, which might suggest that they are considered as more important or urgent by customers.

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

Ans.

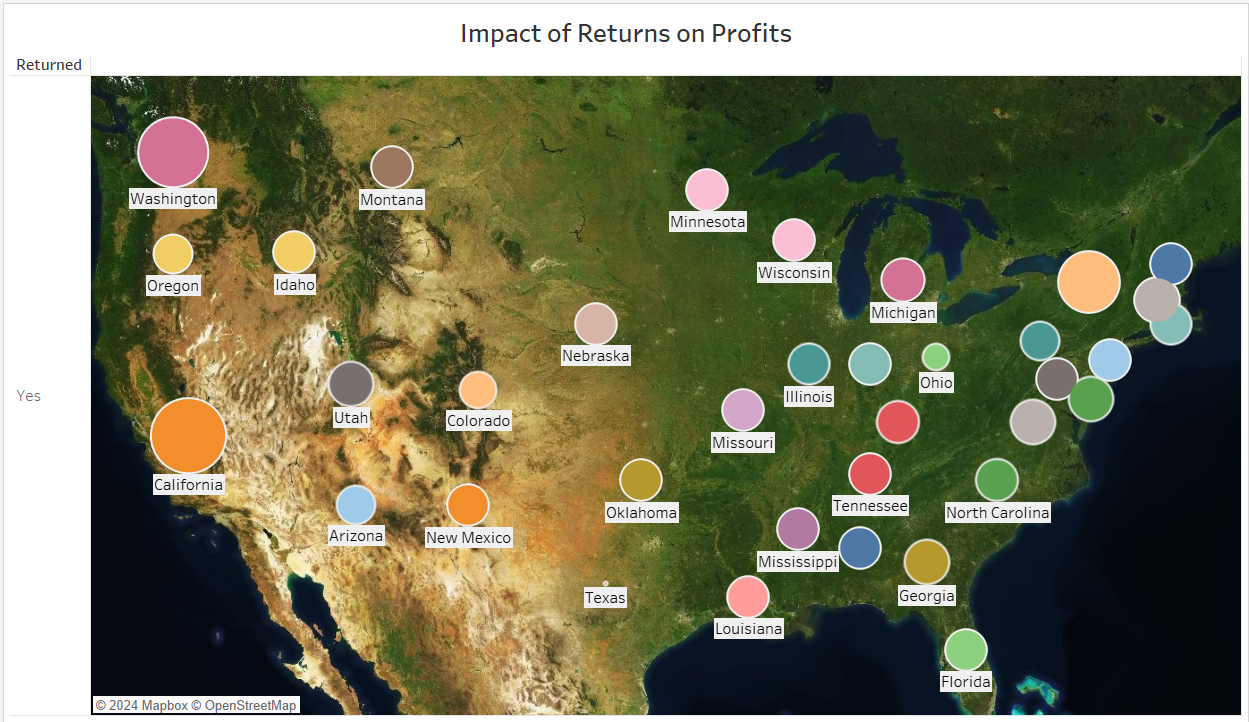


The pie chart titled "Average order value by Repeat & New customers" shows the percentage contribution of repeat and new customers to the overall sales. The size of each slice represents the percentage contribution of that customer type.

* **Repeat Customers:** 14.965
* **New Customers:** 8.295

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

Ans.



The map titled "Impact of Returns on Profits" shows the geographical distribution of returns and its impact on overall profitability. The size of each bubble represents the number of returns in that region, and the color represents the impact of returns on profitability.

* **West Region:** The West region has the highest number of returns, but the impact of returns on profitability is moderate. This could be due to a high volume of sales in this region, which offsets the impact of returns.
* **East Region:** The East region has a moderate number of returns, and the impact of returns on profitability is also moderate.
* **Central Region:** The Central region has a low number of returns, and the impact of returns on profitability is low.
* **South Region:** The South region has a moderate number of returns, but the impact of returns on profitability is high. This could be due to a lower volume of sales in this region, which amplifies the impact of returns.