Data Visualization and Analysis Project

**Project Submitted By:- Nishith Sharma**

**About the Project:**

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

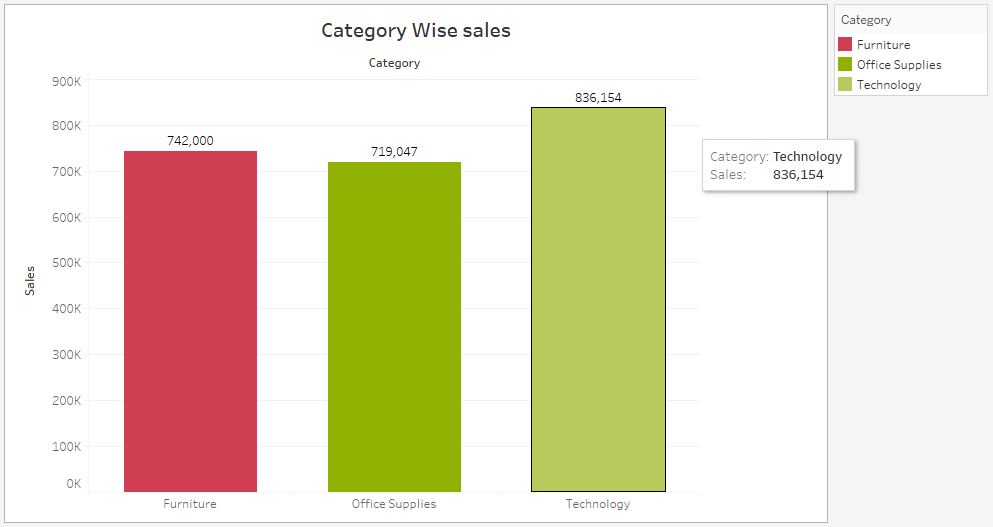
**Problem Statement: Choose the Best chart for any 30 scenario-based questions from Superstore Dataset.**

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.

**Dataset Link:**

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

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

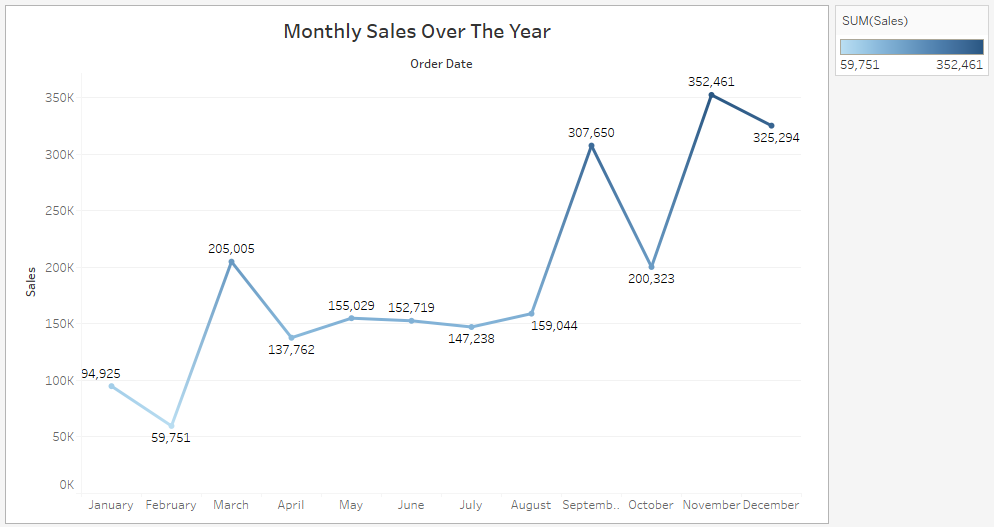


**Que. Why did you choose this chart type?**

**Ans:-** To show which category have the highest sales in terms of sales, Bar Chart is the best option as I have 1 measure i.e. sales and 1 dimension i.e. category.

Using this bar chart we can clearly see that the highest bar is of “Technology” and hence it is the category with the highest sales.

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

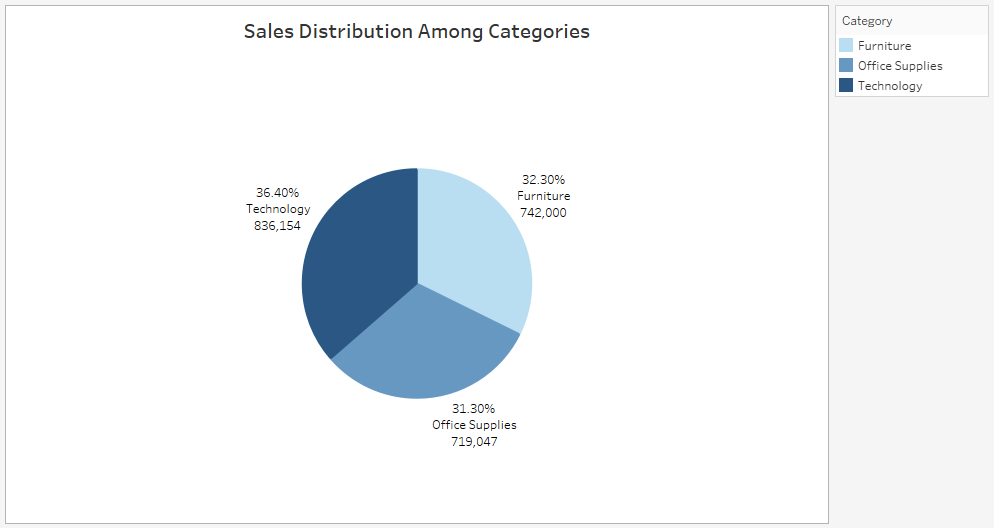


**Que. Why did you choose this chart type?**

**Ans:-** Line chart is the best option to visualize data which includes time series data or when we have to analyze on the basis of trends and pattern. Similarly in this question we have to do the analysis of sales over the course of year.

When we look at this line chart we can say that sales is increasing over the year but in the months of February, April, October and December sales has declined in comparison to the previous month.

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

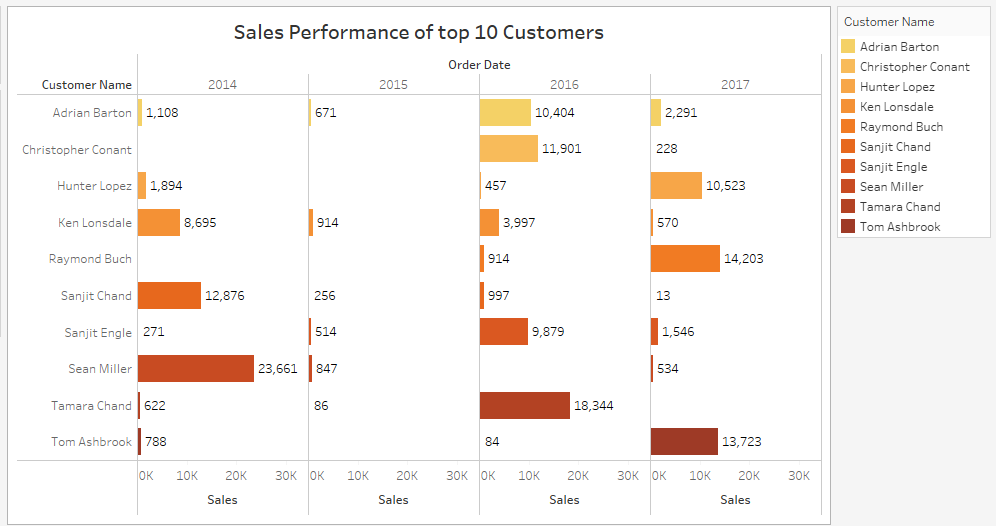


**Que. Why did you choose this chart type?**

**Ans:-** As I have to compare the distribution of sales amount among different categories, pie chart is on top to show the proportion of whole. In this visualization also I have 2 measures and 1 dimension which is ideal condition to plot pie chart.

This pie chart shows that the sales distribution is somewhat similar in all the 3 categories but “Technology” leads the group with 36.40% and “Office Supplies” at last with 31.3%.

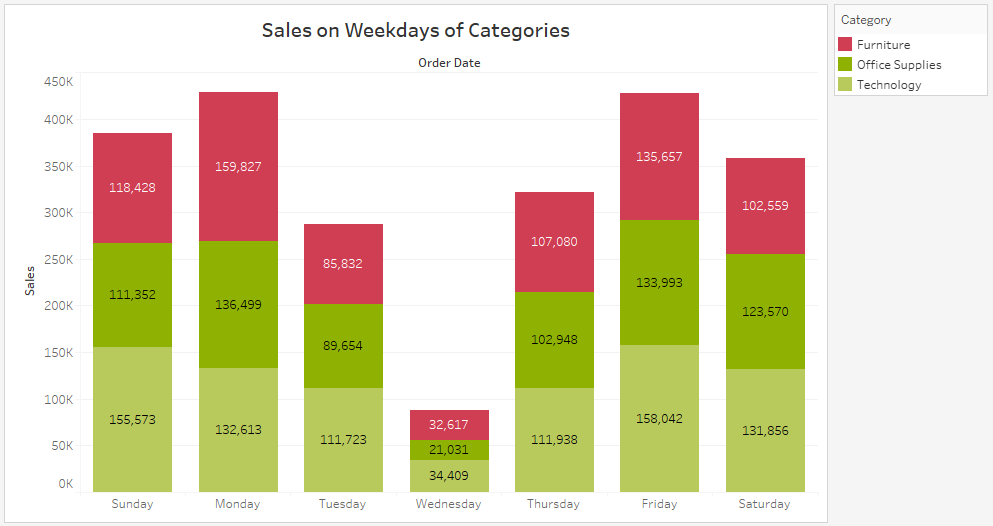
1. **Can we analyze the sales performance of individual customers over time?**



**Que. Why did you choose this chart type?**

**Ans:-** This chart shows the sales trend over the years for individual customers. But showing trend for all the customers is not possible in this chart as it can make the chart messy and less informative so I have taken top 10 customers and selected Horizontal Bar chart type to show the data in effective manner.

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

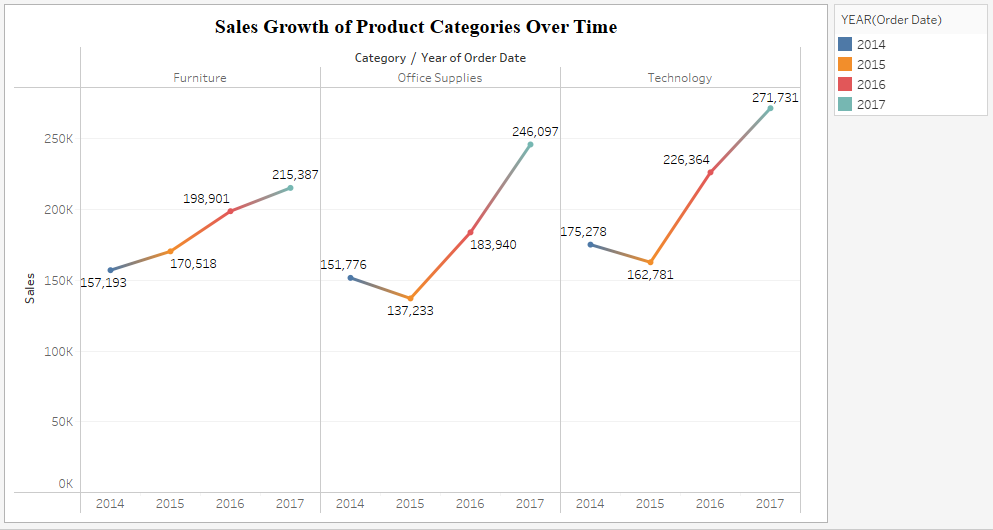


**Que. Why did you choose this chart type?**

**Ans:-** A stacked bar chart is used to display to the sale across different days of the week, segmented by product categories. Stacked bar charts are effective in illustrating the composition of a whole (total sales for each day) broken down into segments (sales for different product categories) while allowing direct comparison between the segments (product categories) on each day.

Monday and Friday are the two days of the week which have highest sales throughout the categories. While Wednesday remains the least profitable day of the week.

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

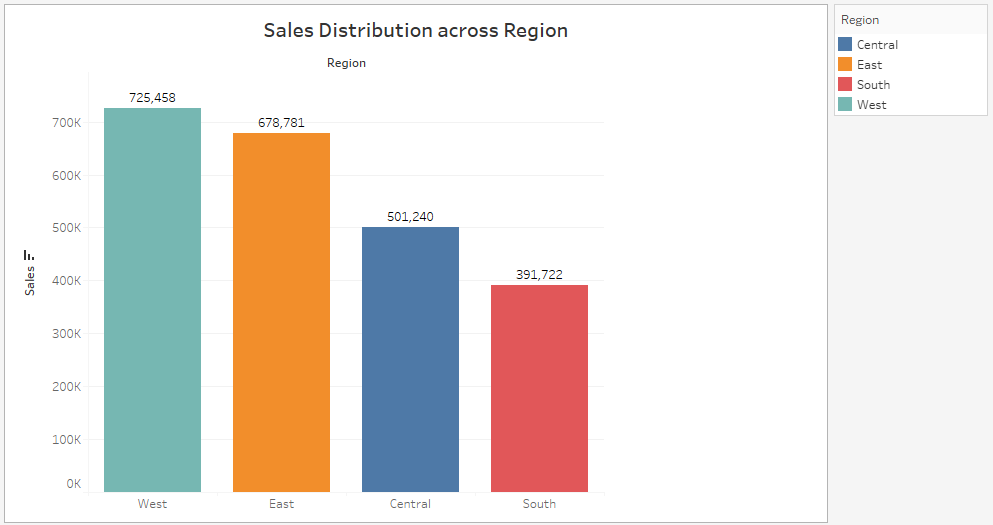


**Que. Why did you choose this chart type?**

**Ans:-** The sales growth of different categories over the time can be greatly visualized by multiple line chart which shows differential trajectory for all the categories.

This multiple line chart is showing the overall increase in the sales for all the categories barring “Office Supplies” and “Technology” where sales for the year 2015 decreased in comparison to 2014 and then sales kept increasing for all.

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

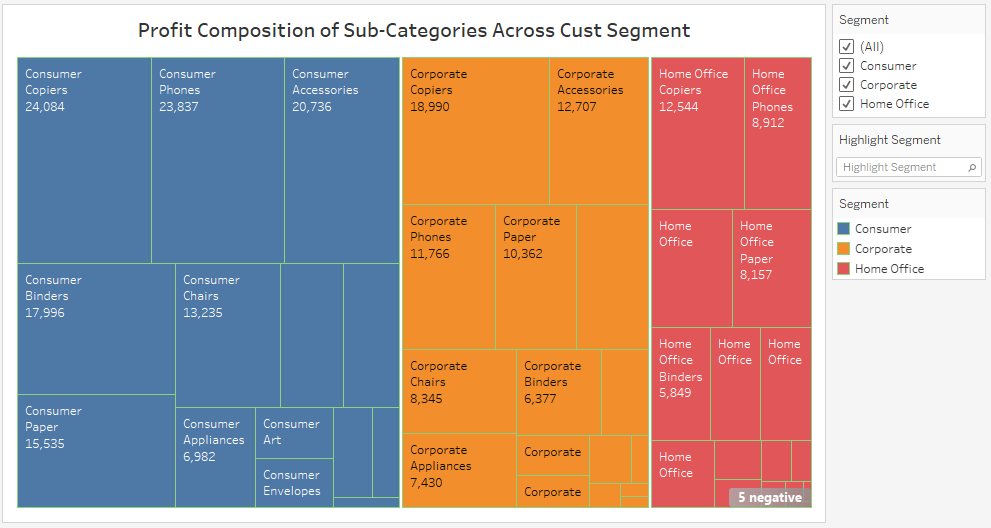


**Que. Why did you choose this chart type?**

**Ans:-** Here, I have used bar chart to display the sales distribution across different regions i.e. East, West, Central and South. Bar chart is best chart type to display the comparison between categorical data.

In this question Bar chart is showcasing that “west” Region has the highest sales with total sales amounting to $725,458 followed by “East” and “central”. “South” region has the lowest sales.

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

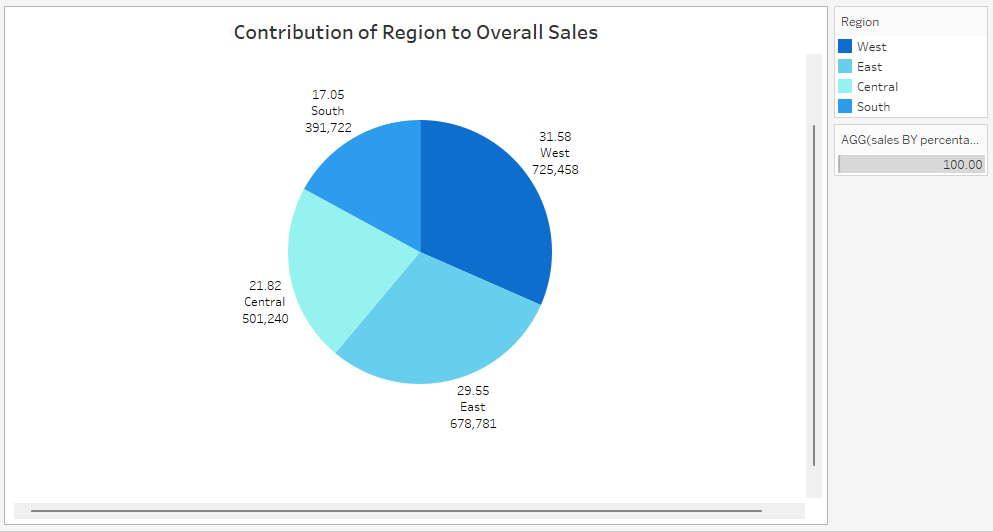


**Que. Why did you choose this chart type?**

**Ans:-** I have used Tree map Chart here to show composition of profit for various subcategories.

This tree map shows that Copiers, phones and Accessories are the top 3 sub-categories for all the 3 segments in varying order. While 5 sub-categories shows loss.

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

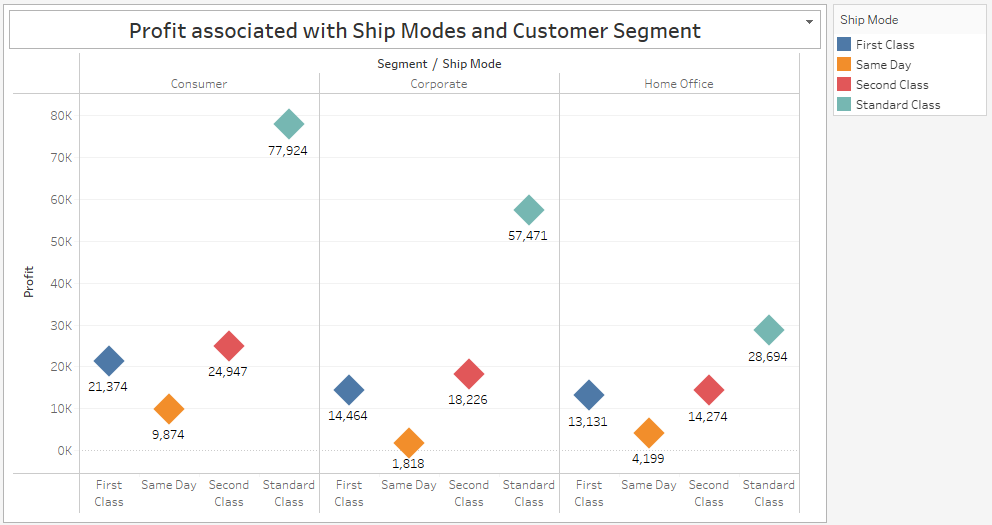


**Que. Why did you choose this chart type?**

**Ans:-**  Pie chart is the best chart type to show the percentage contribution of categorical data. To show the region wise share of sales I have used pie chart.

“West” region has the highest contribution percentage i.e. 31.6% share in overall sales. Which is followed by “East” and “Central” regions with 29.6% and 21.8% respectively. “South” region having the least share in overall sales in USA.

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

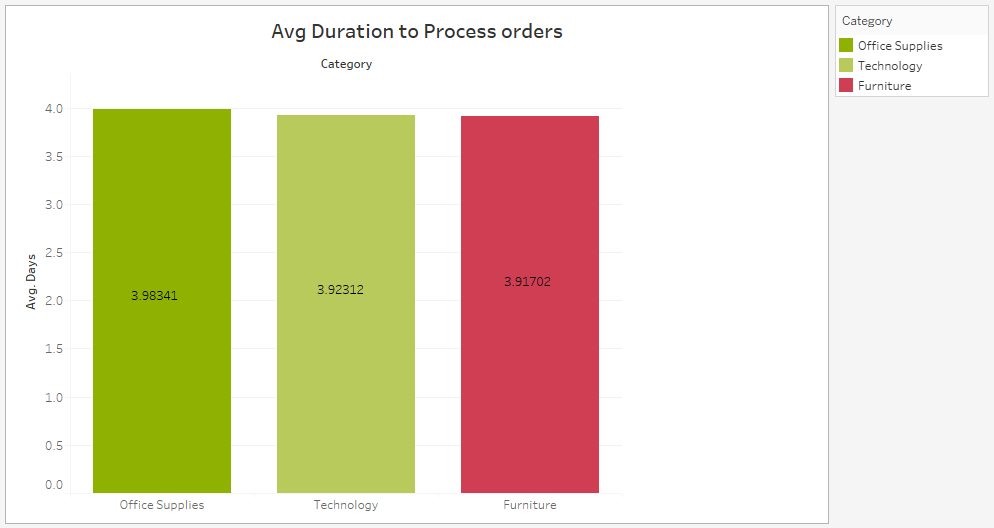


**Que. Why did you choose this chart type?**

**Ans:-** This is a grouped bar chart which I have modified using shapes instead of conventional Bars. The shapes are showing the highest point of bars which indicating the profit amount for all shipping modes across different consumer segments.

Standard Class is contributing most to earn profit across different consumer segments, while same day mode has the least contribution.

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



**Que. Why did you choose this chart type?**

**Ans:-** Bar chart is used here to showthe comparison and average duration takes to process orders for different categories.

This chart displaying that the “Office Supplies” has the highest avg duration of 3.98 days, while “furniture” has the best avg amongst all the 3 categories of 3.91 days.

1. **How do discounts affect overall profit?**

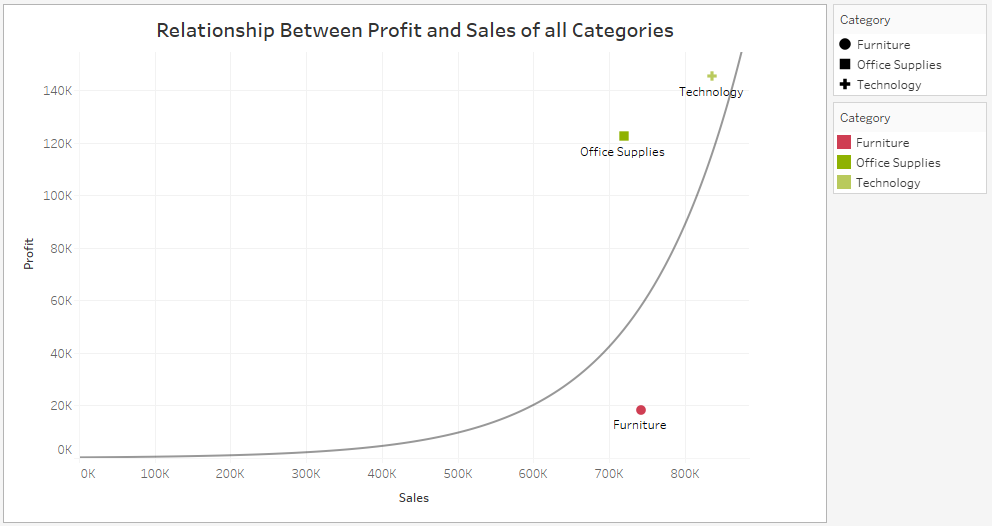


**Que. Why did you choose this chart type?**

**Ans:-** Bar chart is used to analyze the relation between Discounts and Profit. I have categorized discount given in multiple bins which gives wider scope to look at relation.

As we can see, with the increase in discount percentage the overall profit is declining and with highest percentage of discount the profit turning into losses.

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

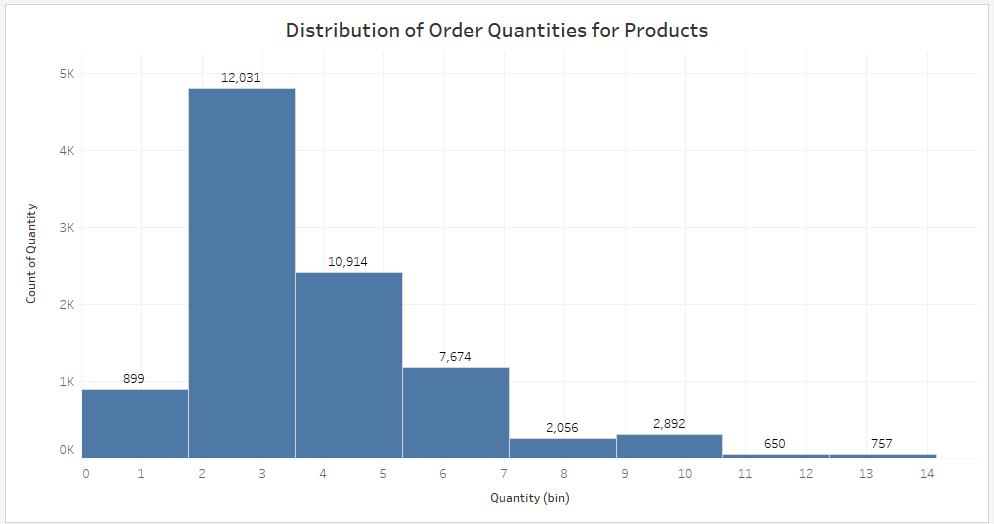


**Que. Why did you choose this chart type?**

**Ans:-** The relationship between product sales and profitability for different product categories can be effectively visualized through the scatter plot.

The scatter plot visually represents the relationship between product sales and profitability. Each point on the plot corresponds to a specific product category, with the x-axis representing sales and the y-axis representing profitability. This format allows for a clear examination of how sales and profitability are related to different product categories. Here it is clearly shown how furniture is doing sales approximately similar to other categories still not making a profit near to other categories.

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

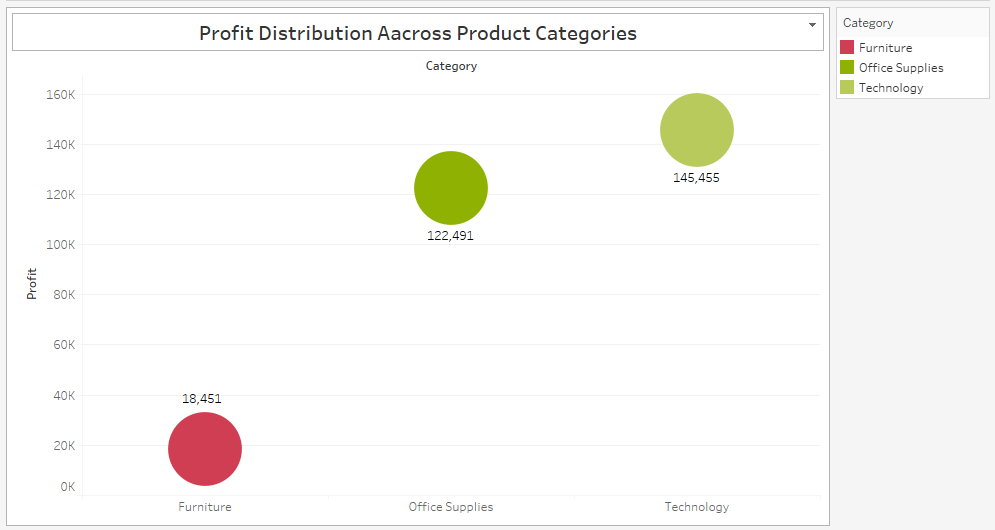


**Que. Why did you choose this chart type?**

**Ans:-** The distribution of order quantities for products in the dataset is effectively visualized through the histogram.

The histogram visually represents the distribution of order quantities, displaying the frequency of different quantity ranges. Each bar on the histogram corresponds to a specific range of order quantities, allowing for a clear understanding of the distribution pattern. This format provides insights into the most common order quantities and their variations within the dataset.

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

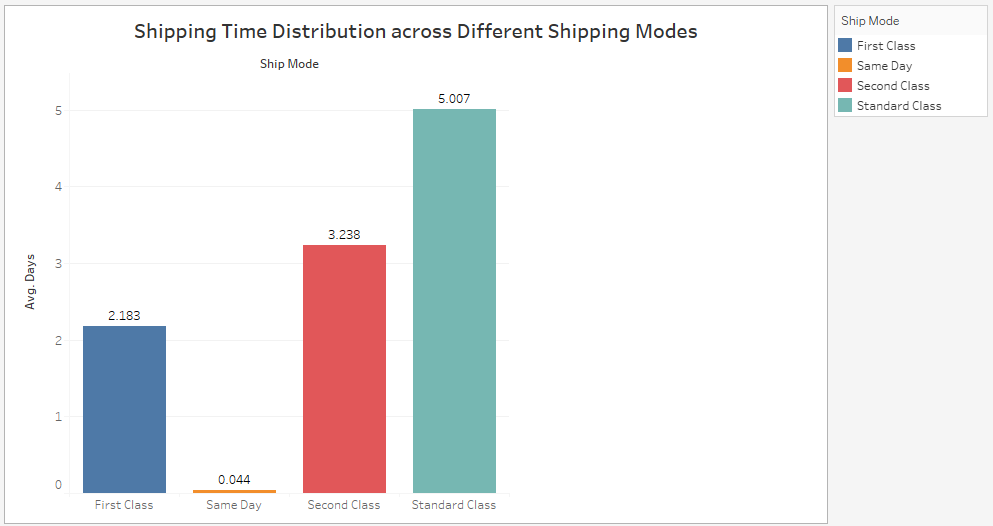


**Que. Why did you choose this chart type?**

**Ans:-** The profit distributions across different product categories are effectively visualized through the Bar chart modified by using shapes in place of vertical bars.

This format provides a clear understanding of how profits are distributed, allowing for comparisons between different product categories. Here “Technology” is making the highest profit.

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

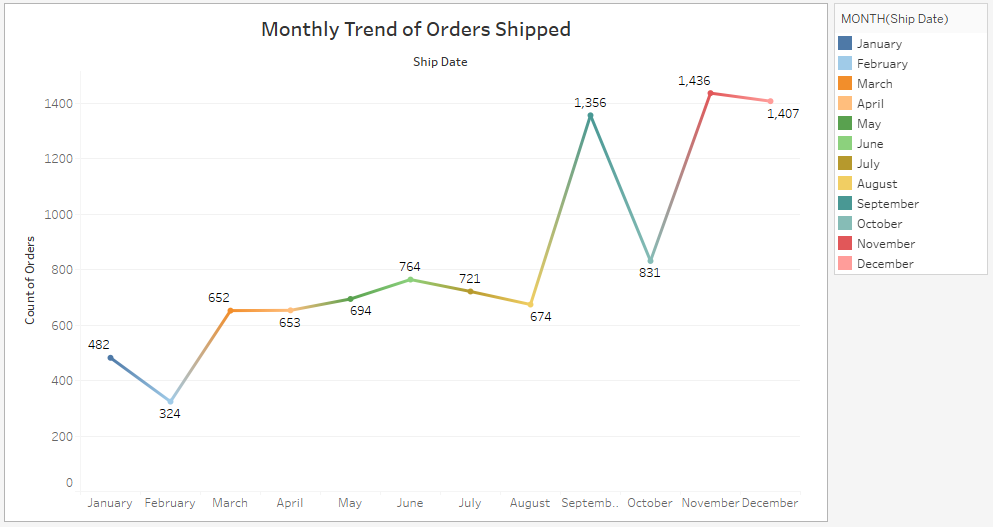


**Que. Why did you choose this chart type?**

**Ans:-** Bar chart is used to compare shipping time in days with different shipping modes.

This Bar chart shows that Standard Class takes almost 5 days to ship the orders, Second class takes 3 days. While Same day mode as promised ship orders on the very same day.

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

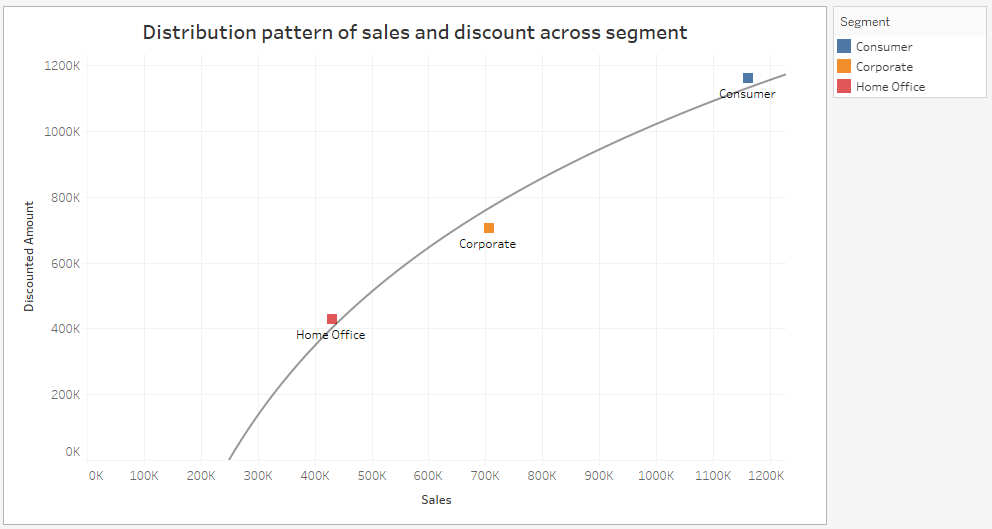


**Que. Why did you choose this chart type?**

**Ans:-** The monthly trend in the number of orders shipped is effectively visualized through the line chart.

The line chart illustrates the monthly trend in the number of orders shipped over the observed time period. Each point on the line represents a specific month, allowing for a clear understanding of the fluctuation and patterns in order shipments. This format provides insights into the seasonality or trends associated with the number of orders shipped each month.

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

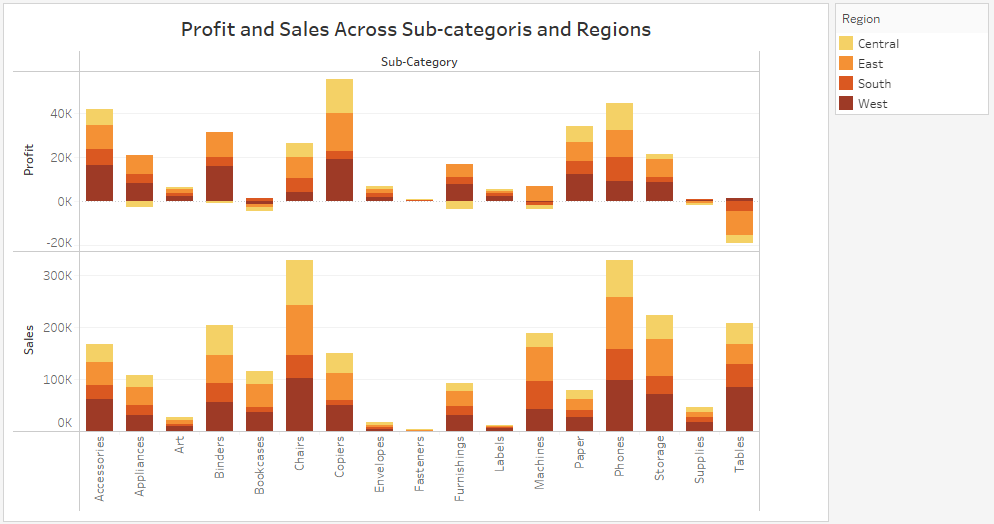


**Que. Why did you choose this chart type?**

**Ans:-** The performance of different customer segments in terms of sales and discount rates can be effectively visualized through the scatter plot.

The Scatter plot visually represents the performance metrics of different customer segments, showcasing both sales and discount rates on the same chart. Each group of point corresponds to a specific customer segment, with one axis representing sales and the other representing discount rates. This format allows for a simultaneous comparison of both metrics, providing a comprehensive view of how different customer segments perform in terms of both sales and discount rates.

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

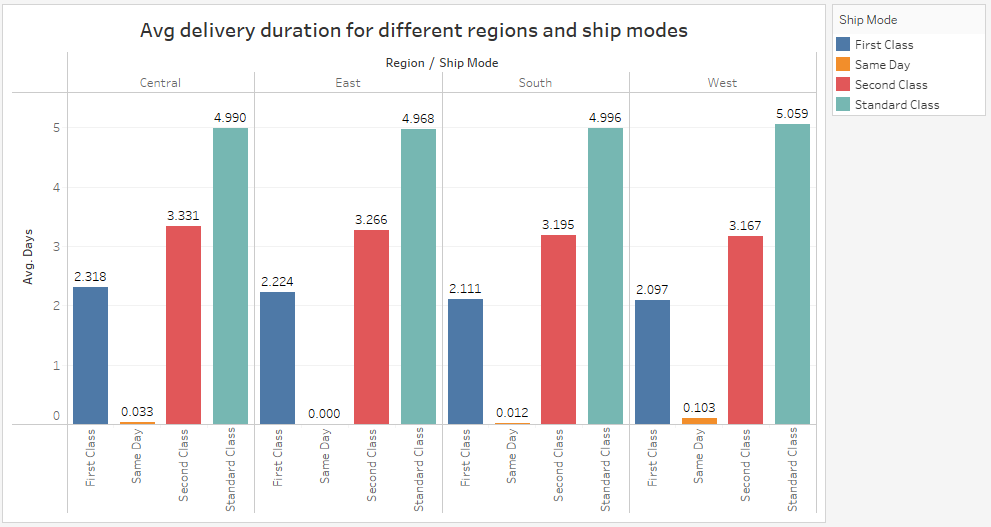


**Que. Why did you choose this chart type?**

**Ans:-** The sales and profit trends across different product subcategories and regions in the Superstore dataset can be effectively visualized through the stacked bar chart.

The stacked bar chart visually represents the trends in both sales and profits for different product subcategories within various regions. Each bar corresponds to a specific product subcategory, and the chart includes dual axes for sales and profits. This format allows for a comprehensive view of how sales and profits evolve across different product subcategories and regions.

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

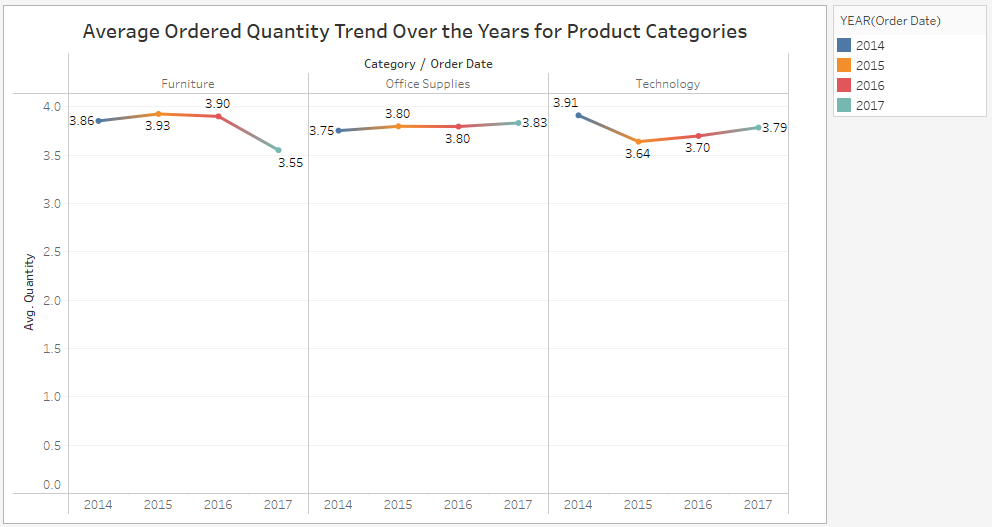


**Que. Why did you choose this chart type?**

**Ans:-** The average delivery duration for different regions and ship modes is effectively visualized through the grouped bar chart. In which ship mode and region in compared to delivery duration.

The bars visually represent the average delivery duration for different regions and ship modes. Each group of bars corresponds to a specific region, and bars within each group represent different ship modes. This format allows for a clear and concise comparison of average delivery durations across various regions and ship modes.

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

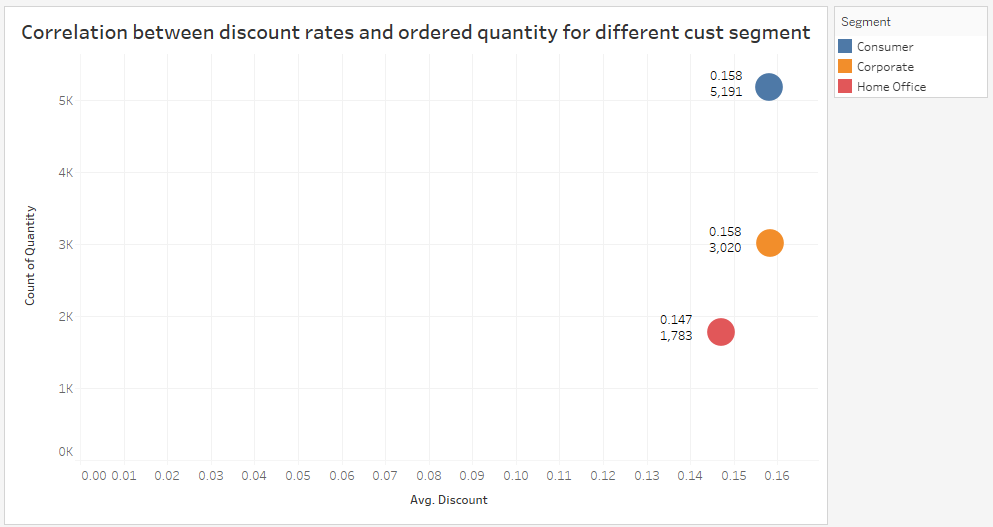


**Que. Why did you choose this chart type?**

**Ans:-** Line chart Is the best option to analyze the trend and pattern how avg order quantity changed over the years for various categories. This is like grouped line chart where trend of avg order quantity of all the 3 categories is compared to each other.

Each line corresponds to a specific product category, allowing for a clear understanding of how the average order quantity has changed over the years. This format provides insights into the fluctuations and patterns associated with the average order quantity within each product category.

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

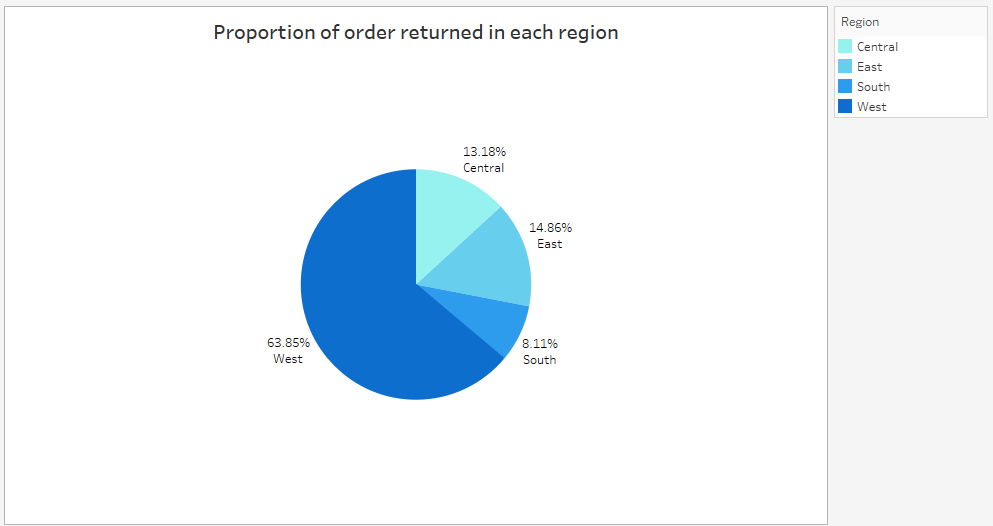


**Que. Why did you choose this chart type?**

**Ans:-** The correlation between discount rates and order quantities for different customer segments can be effectively visualized through the scatter plot.

The scatter plot visually represents the correlation between discount rates and order quantities. Each point on the plot corresponds to a specific customer segment, with the x-axis representing discount rates and the y-axis representing order quantities. This format allows for a clear examination of how discount rates and order quantities are related for different customer segments.

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

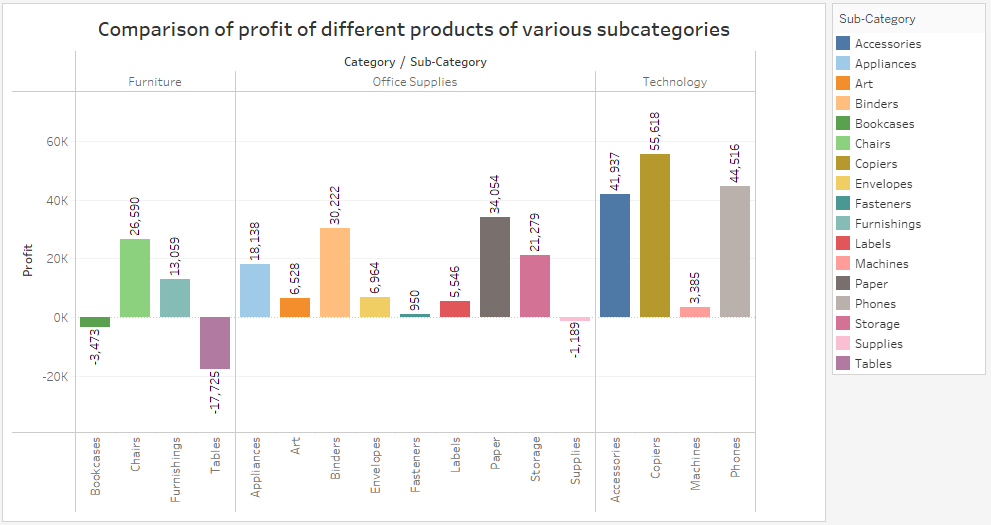


**Que. Why did you choose this chart type?**

**Ans:-** The proportion of orders returned in each region within the Superstore dataset can be effectively visualized through the Pie chart.

The Pie chart visually represents the proportion of orders returned in each region. Each segment of the bar corresponds to a specific region, and the share of each segment indicates the proportion of returned orders. This format allows for a clear and concise comparison of the proportion of returned orders across various regions. “West” region has the highest return share overall and “south” has the least.

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

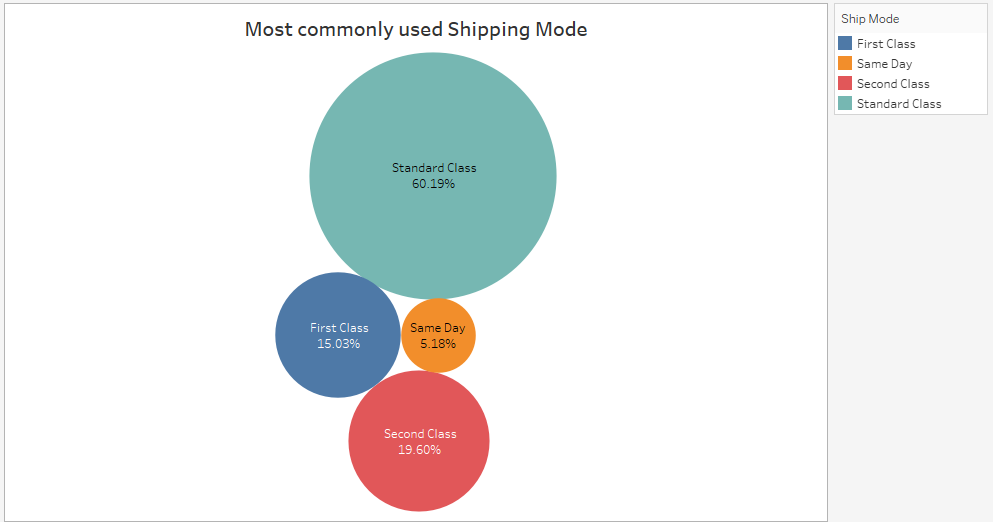


**Que. Why did you choose this chart type?**

**Ans:-** To compare the profit of different products for various categories and subcategories, we can effectively utilize the bar chart.

The bar chart visually represents the profit of different products within each subcategory. Each bars corresponds to a subcategory within category. The height of each bar indicates the profit generated by a specific subcategory. This format allows for a straightforward comparison of profit distribution among products in different subcategories. We can see there are some losses bearing subcategories also.

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

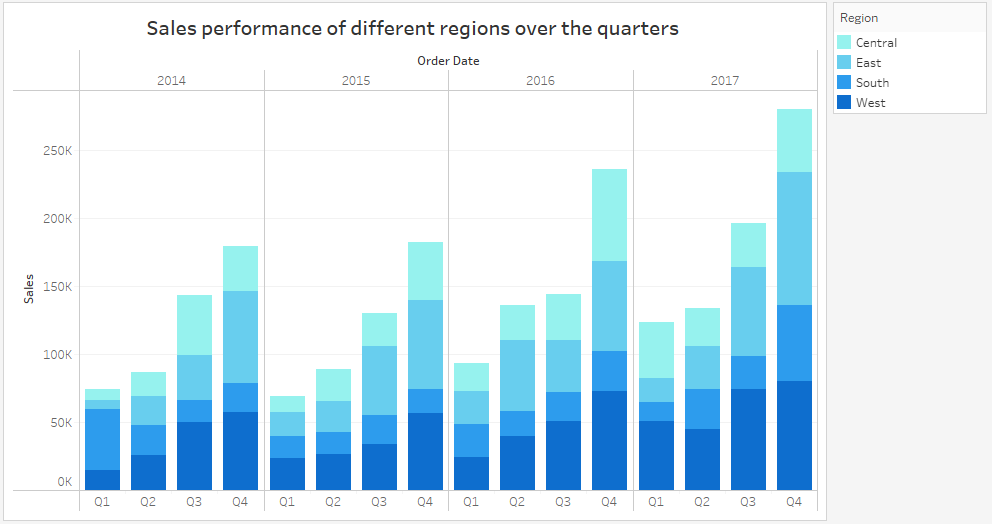


**Que. Why did you choose this chart type?**

**Ans:-** To visualize which shipping mode is being used widely, Packed Bubble Chart is used. This type of chart makes job easy to analyze the problem on the basis of size of bubbles. Bigger the size higher the share in composition.

Standard class has the biggest size bubble which means this is the most widely used Shipping mode.

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



**Que. Why did you choose this chart type?**

**Ans:-** The evolution of the sales performance of different regions throughout the quarters of a year can be effectively visualized through the Grouped Bar chart.

The Bar chart visually represents the trends in sales for different regions over the four quarters of a year. Each Segment corresponds to a specific region, allowing for a clear understanding of how the sales performance evolves throughout each quarter. This format provides insights into seasonality, trends, and variations in sales across different regions.

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

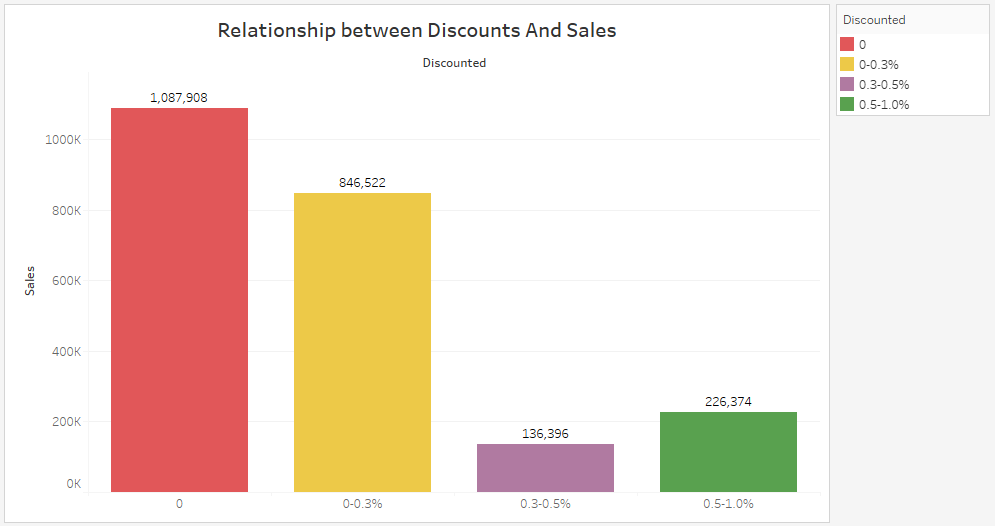


**Que. Why did you choose this chart type?**

**Ans:-** The distribution of order priorities across different product categories can be effectively visualized through the group bar chart.

The group bar chart visually represents the distribution of order priorities within each product category. Each segment of the bar corresponds to a specific order priority, and the height of each segment indicates the proportion or count of orders with that priority. This format allows for a clear and concise comparison of order priorities across various product categories.

1. **What is the relationship between discounts and sales?**

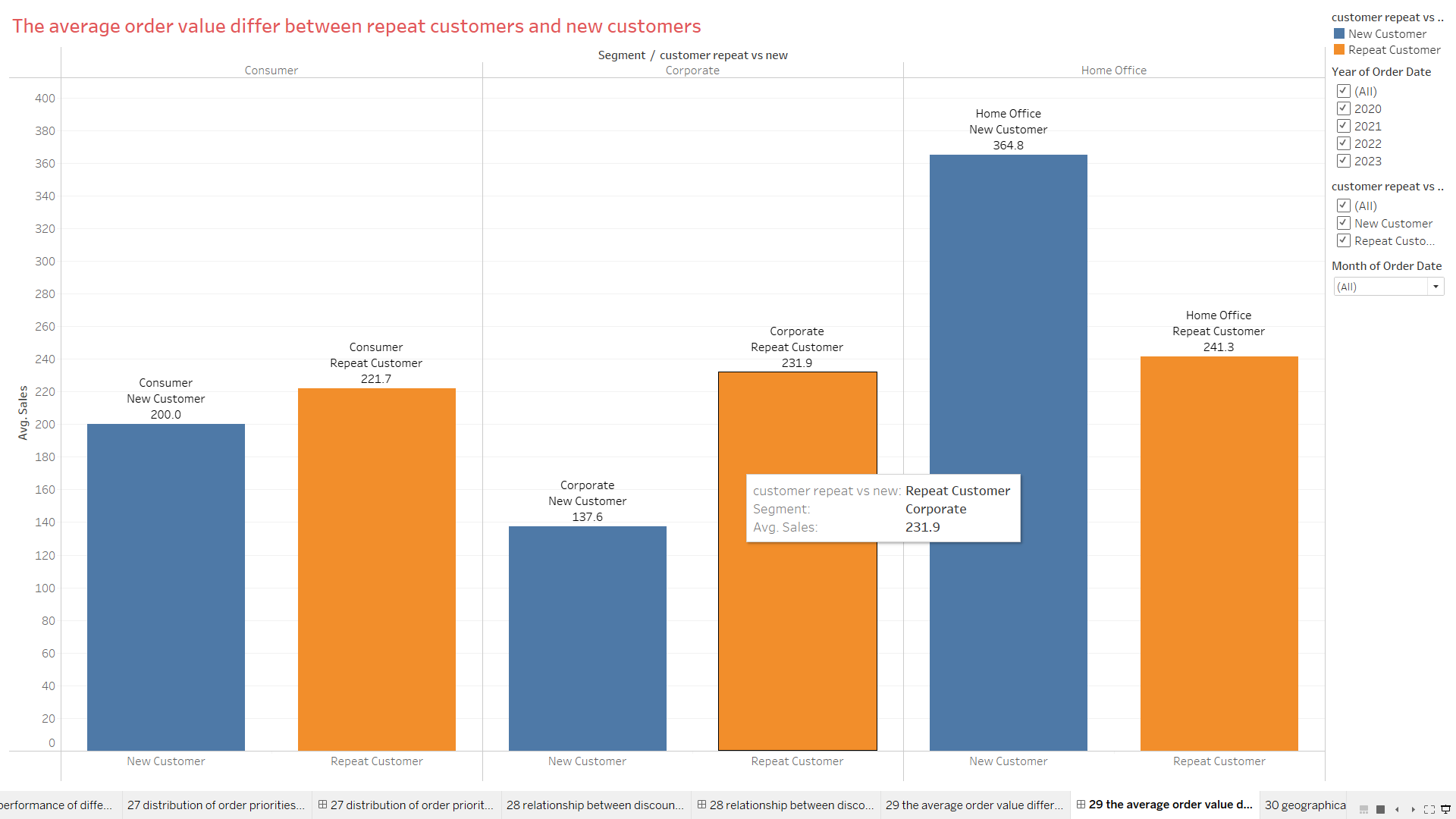


**Que. Why did you choose this chart type?**

**Ans:-** Bar chart is used to analyze the relation between Discounts and Sales. I have categorized discount given in multiple bins which gives wider scope to look at relation.

As we can see, with the increase in discount percentage the overall sales is declining.

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

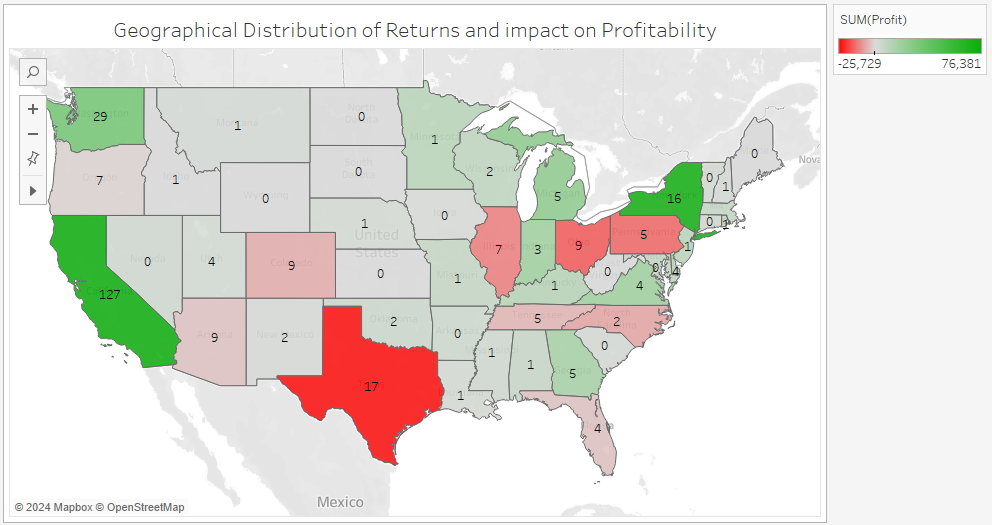


**Que. Why did you choose this chart type?**

**Ans:-** The difference in the average order value between repeat customers and new customers can be effectively visualized through the grouped bar chart.

The grouped bar chart visually represents the average order value for both repeat customers and new customers. Each group of bars corresponds to a specific customer segment, with one of bars representing repeat customers and the other representing new customers. This format allows for a simultaneous comparison of the average order value for both customer segments, providing insights into their purchasing behavior.

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



**Que. Why did you choose this chart type?**

**Ans:-** The geographical distribution of returns and its impact on overall profitability can be effectively visualized through the Tableau map chart and a separate bar chart.

Tableau Map Chart: - The map chart visually represents the geographical distribution of returns. Each region is color-coded based on the proportion of profit where red represents loss and green shows profit with varying concentrations and also shows the count of returned items. This allows for a clear understanding of where returns are concentrated geographically.

***THANK YOU***