

Data Visualization and Analysis Project

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

Dataset Link:

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

Tableau Public Link:

https://public.tableau.com/app/profile/mehak.saini/viz/Samplesuperstore_16985552193180/Sheet54?publish=yes

Questions:

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

Answer- The product category 'Technology' has the highest sales, followed by 'Furniture'.

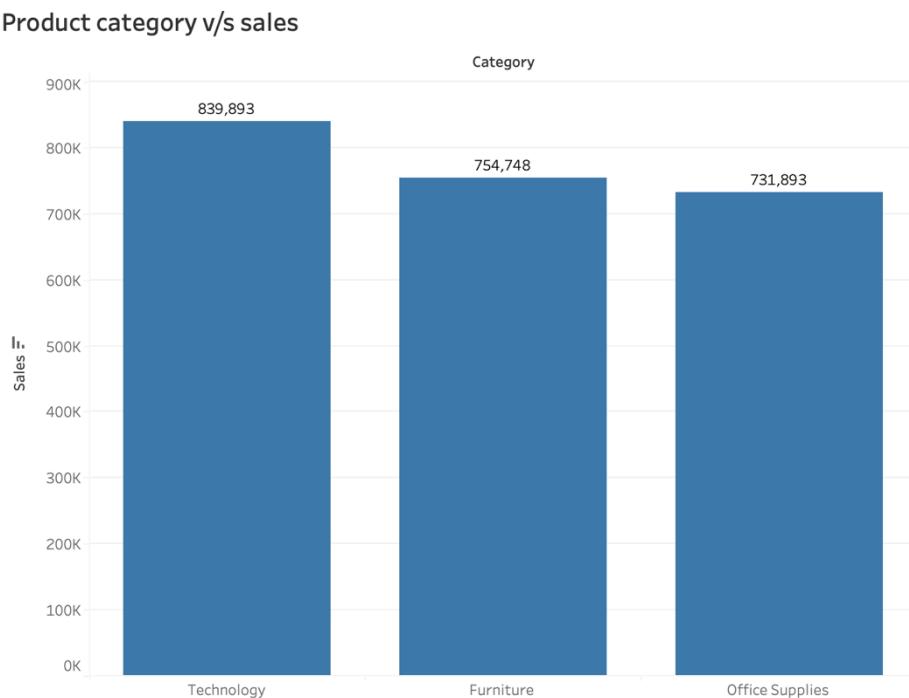


Fig. 1: Vertical bar chart displaying category wise sales

Chart used- Bar chart.

Reasons for choosing Bar chart-

- Bar charts are excellent for comparing the total sales of different product categories. Each bar represents a category, allowing easy visualization to compare the lengths of the bars to identify which categories have the highest total sales.

- Bar charts rank the categories based on their total sales, with the tallest bar representing the highest total sales. This makes it easy to identify not only which categories have the highest total sales but also their relative positions compared to other categories.
2. How do the monthly sales amounts change over the course of a year?
- Answer-** Sales are maximum in the months of November and December for all years. Also, sales are lowest in the starting of the year and gradually increase towards the end with some ups and downs.

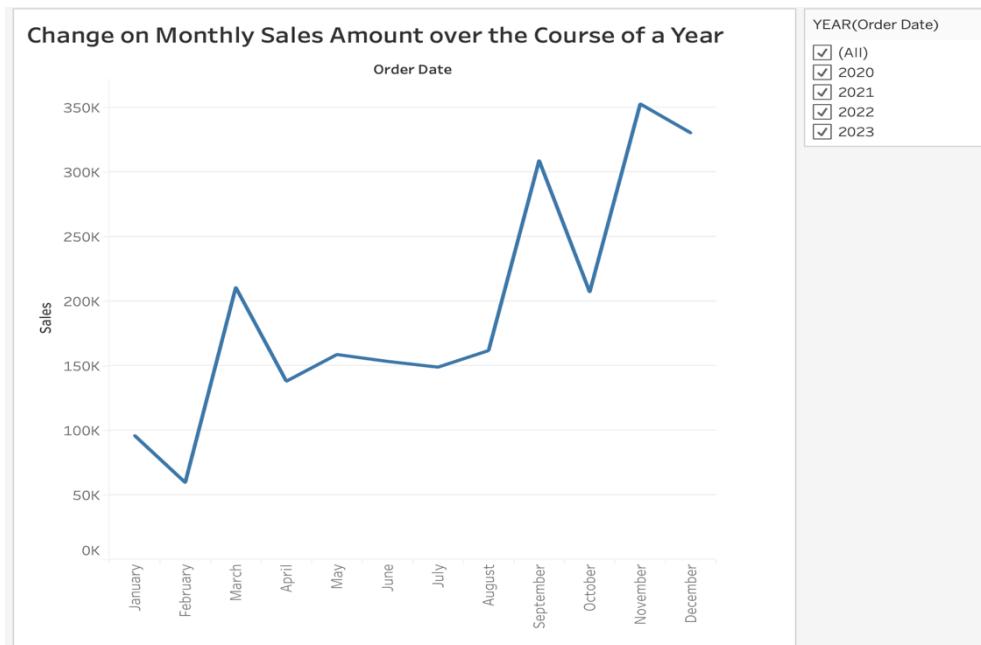


Fig. 2: Line chart displaying the monthly sales for all years.

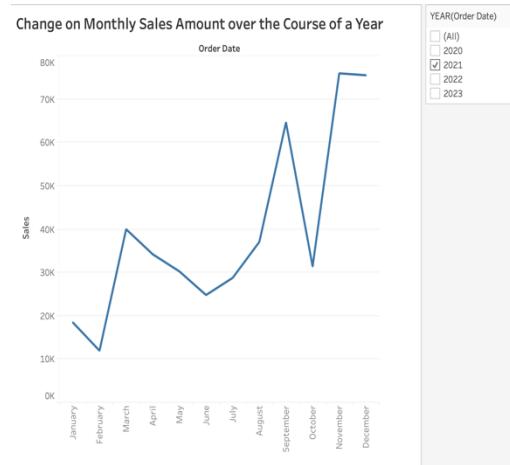
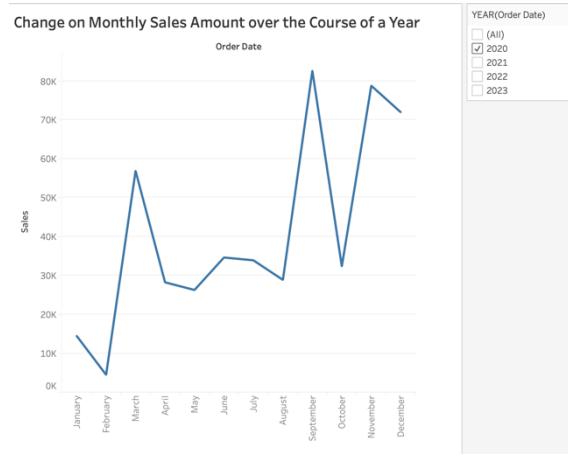


Fig. 3: Line chart displaying the monthly sales for the year 2020.

Fig. 4: Line chart displaying the monthly Sales for the year 2021.

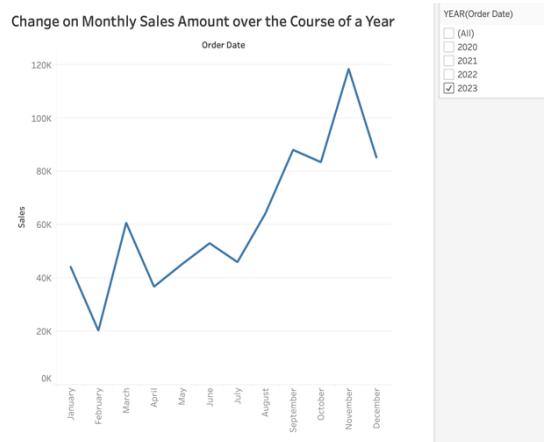
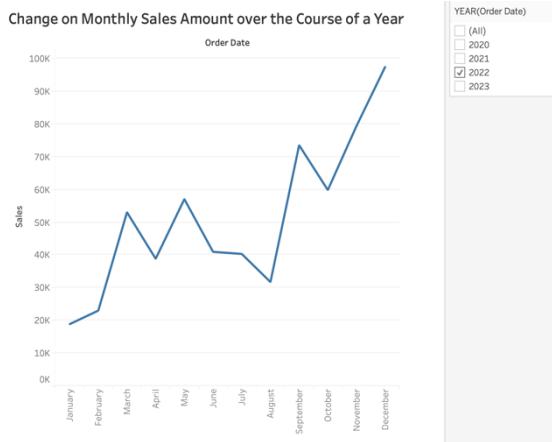


Fig. 5: Line chart displaying the monthly sales for the year 2022.

Fig. 6: Line chart displaying the monthly sales for the year 2023.

Chart used- Line chart.

Reasons for choosing Line chart-

- Line chart is used here as it is excellent for visualizing trends over time. By plotting monthly sales amounts on the y-axis and time (months) on the x-axis, one can easily observe the overall trend of sales amounts over the course of a year.
- Line charts effectively display the temporal relationship between sales amounts and time. One can identify patterns such as seasonality, trends, or fluctuations in sales amounts throughout the year.
- Line charts allow easy spotting of changes or deviations in sales amounts over time. Any spikes, dips, or shifts in the sales trend can be quickly identified, helping to pinpoint significant events or factors affecting sales.

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

Answer- The sales amount is maximum for Technology (36.101% of total), followed by Furniture (32.441% of total). It is least for Office Supplies (31.459% of total).

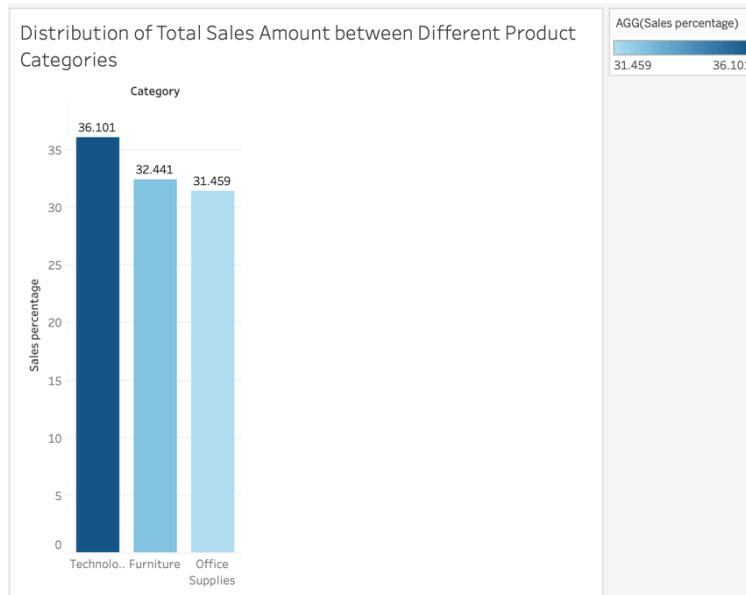


Fig. 7: Bar chart depicting the percent sales of different product categories.

Chart used- Bar chart.

Reasons for choosing Bar chart-

- Bar chart is chosen here to display what percent of the total sales amount is distributed among the different categories of products since bar charts are ideal for comparing values between different categories, allowing quick identification of categories having the highest and lowest total sales amounts.
- The length of each bar directly corresponds to the total sales amount of the respective product category, making it easy to interpret and compare the sales distribution across categories briefly.

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

Answer- Yes, we can view and analyse the sales performance of individual customers over time. We can use the 'Filter by' Customer ID feature. The following circle view chart shows the sum of sales for all customers for different years.

Chart 4

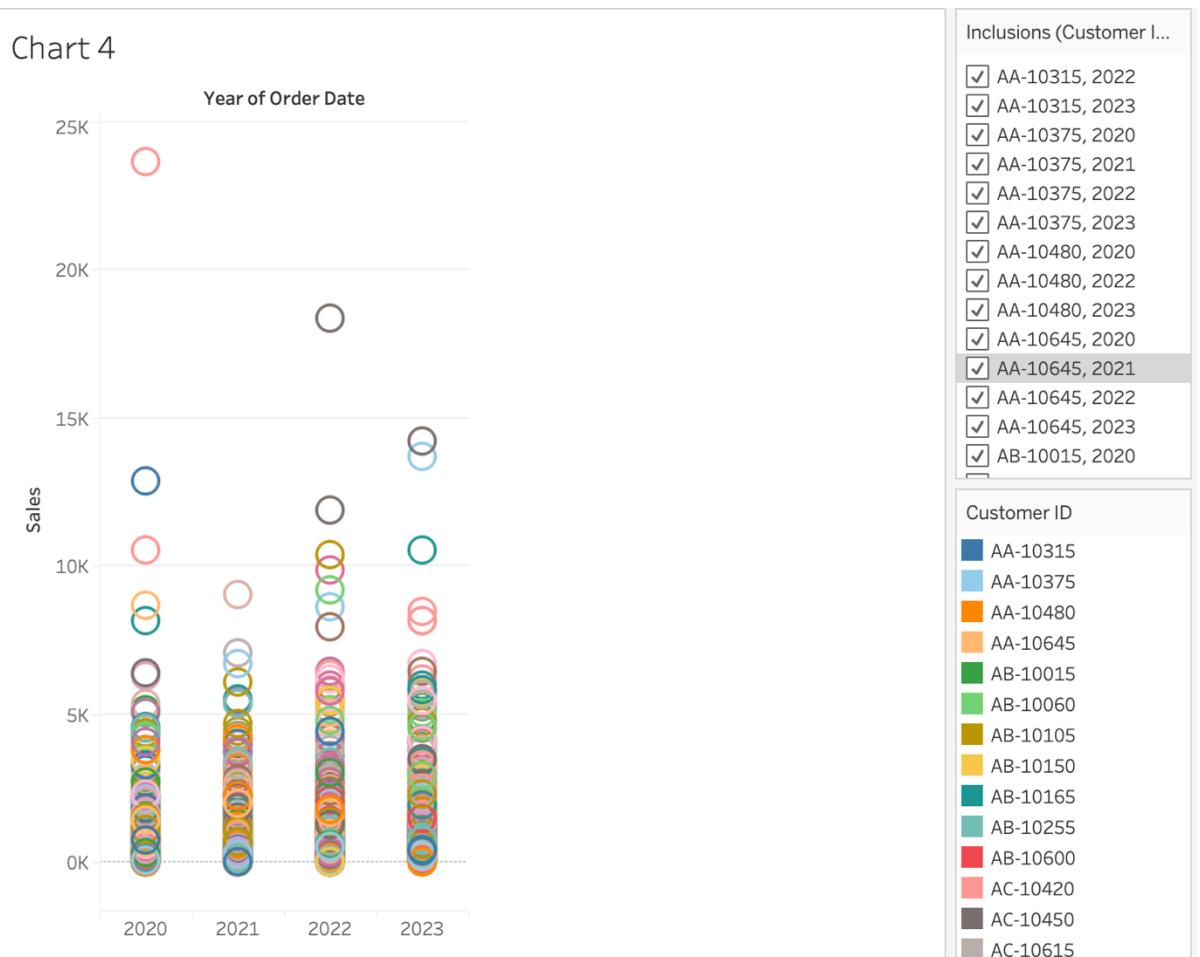


Fig. 8: Circle views chart depicting sales of different customers over different years.

We can also filter individual users. For e.g., if we select any customer ID (let AA-10315), the data will be displayed for that only, as shown in the following graph. As we can see from the following plot, sales were maximum for the customer in 2022, followed by very low value of sales in 2020 and 2023. This customer didn't purchase anything in 2021.

Chart 4

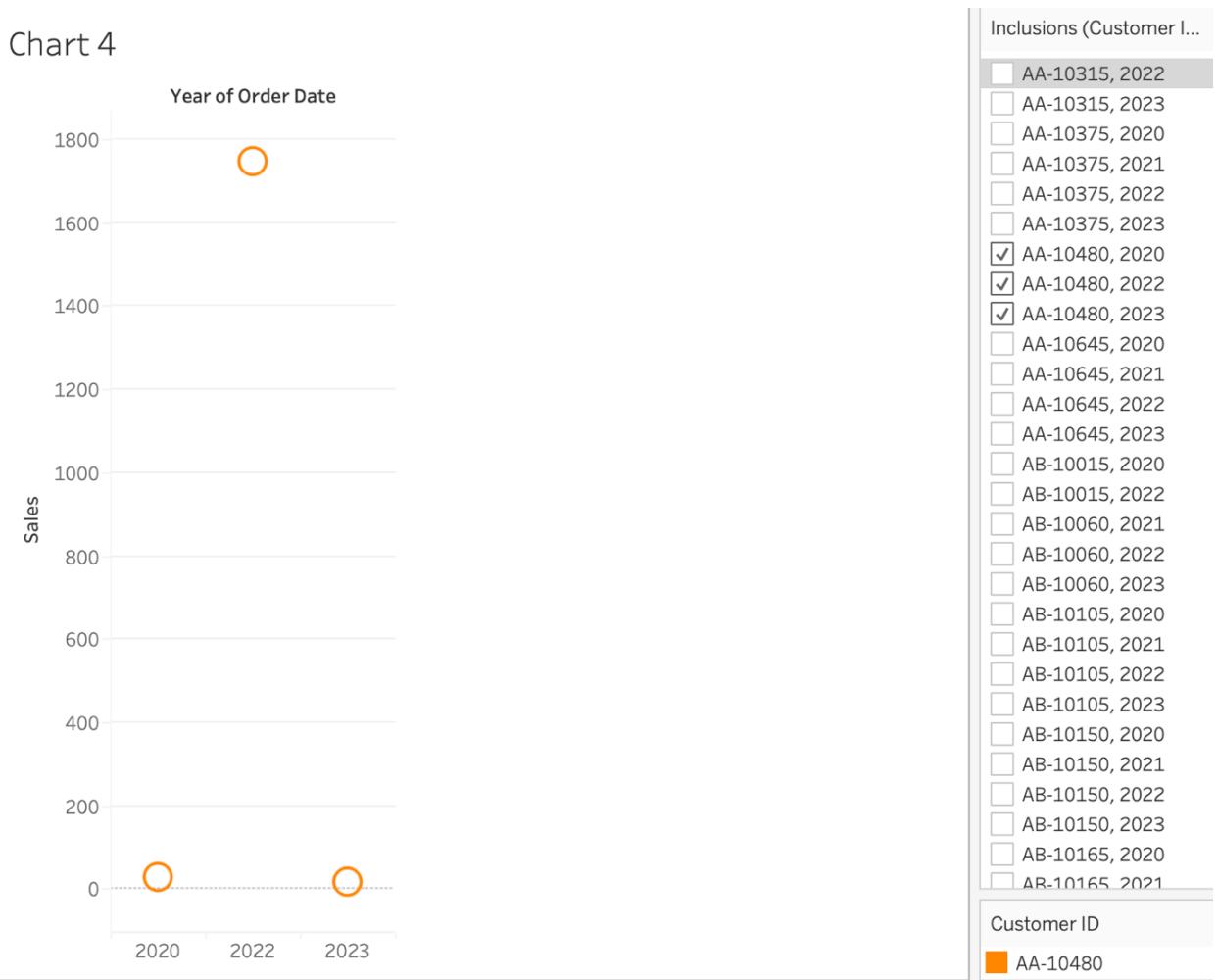


Fig. 9: Chart depicting sales variation for a particular customer (AA-10315) over the years.

Chart used- Circle views.

Reasons for choosing Circle views –

- Circle views, also known as scatter plots or bubble charts, allow for the visualization of individual data points, in this case, individual customers. Each circle represents a customer, and its position on the plot can correspond to metrics such as sales amount or frequency of purchases over time.
 - This type of chart can help in easy analysis and identification of patterns that might be related to sales.
5. How do sales vary based on different days of the week and product categories?

Answer- For technology, sales were highest at the beginning of the month i.e., on the 2nd day and remains same afterwards. In the end, it gradually decreased (with some fluctuations) as the days increased. For furniture, sales were lowest on 23rd day of the month. For office supplies, sales were maximum during the middle of the month i.e., on the 17th day.

Sales of Different Products Categories on Different Days

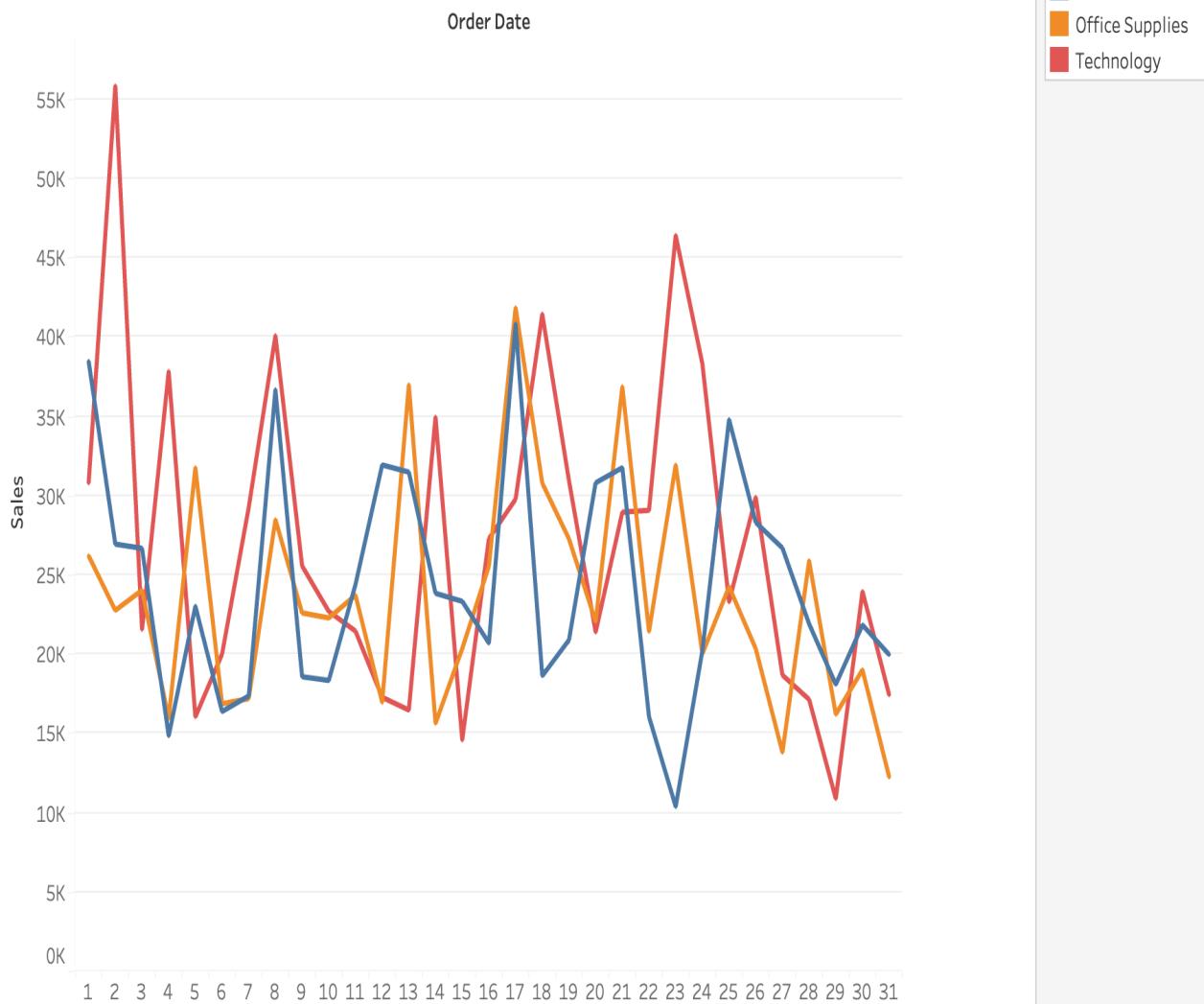


Fig. 10: Line chart displaying sales of different product categories on different days.

Chart used- Line chart.

Reasons for choosing Line chart –

- By plotting sales data over each day of the week, we can easily observe patterns and fluctuations in sales.
 - The lines connect data points for each day of the week, providing a smooth visualization of sales trends and making it easy for viewers to interpret the data.
 - Each line on the chart can represent a different product category, enabling us to compare sales trends and identify which categories perform better on specific days.
6. Can we visualise the sales growth of different product categories over time?

Answer- Yes, by using line charts, we can visualize the sales growth of different product categories over time. From the following graph, it can be seen that sales of all 3 product categories are increasing. Technology is leading the sales at 272,419 in 2023.

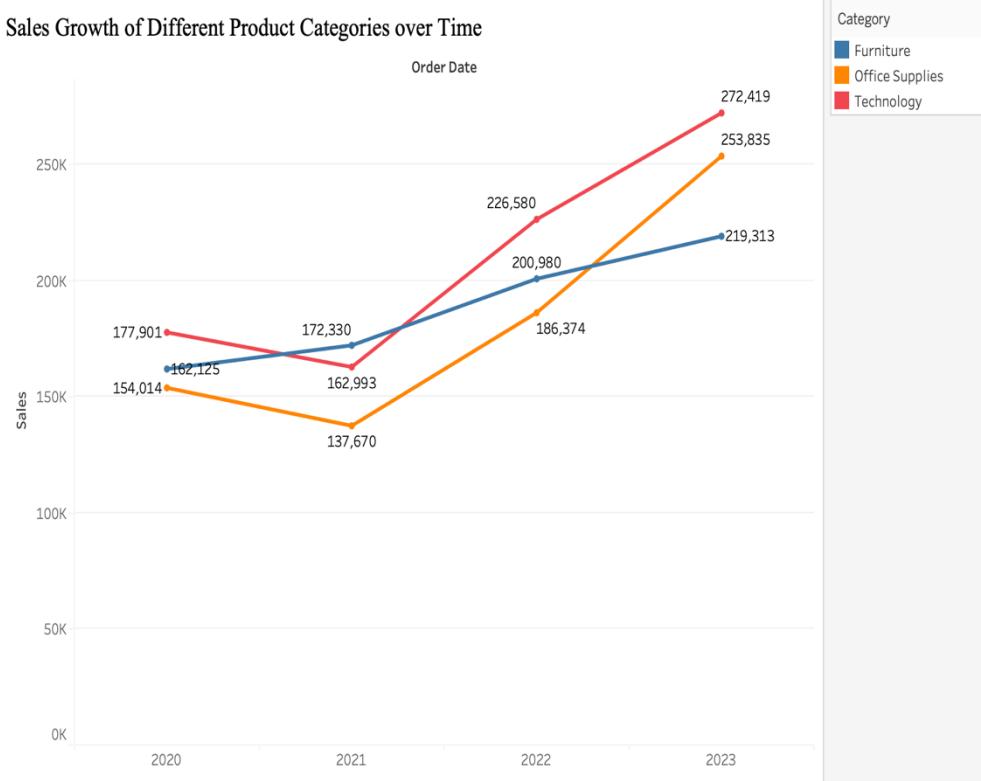


Fig. 11: Line chart depicting sales growth of different product categories over time.
 Chart used- Line chart.
 Reason for choosing Line chart –

- Line chart is chosen here since we want to see trends over a time and these charts are best for this type of visualization.

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

Answer- The following bar plot depicts the distribution of sales for different regions. It is maximum in the western area, followed by eastern area. It is least in the southern region.

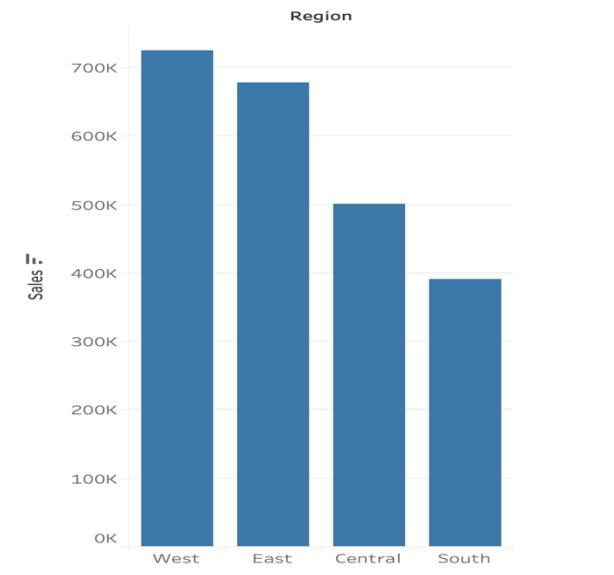


Fig. 12: Bar plot depicting region wise sales.

Reason for choosing bar chart-

- Bar chart used here is efficiently describing the sales distributions over different regions without any complex visualization.

Also, for analysing further region wise sales distributions, symbol maps can be used. From the following symbol maps, it can be inferred that the sales in US (2,297,201) are far more than those in Canada (29,333). Also, in the US, the state California has the maximum sales (457,688). In Canada, the state Saskatchewan has the lowest sales (132).

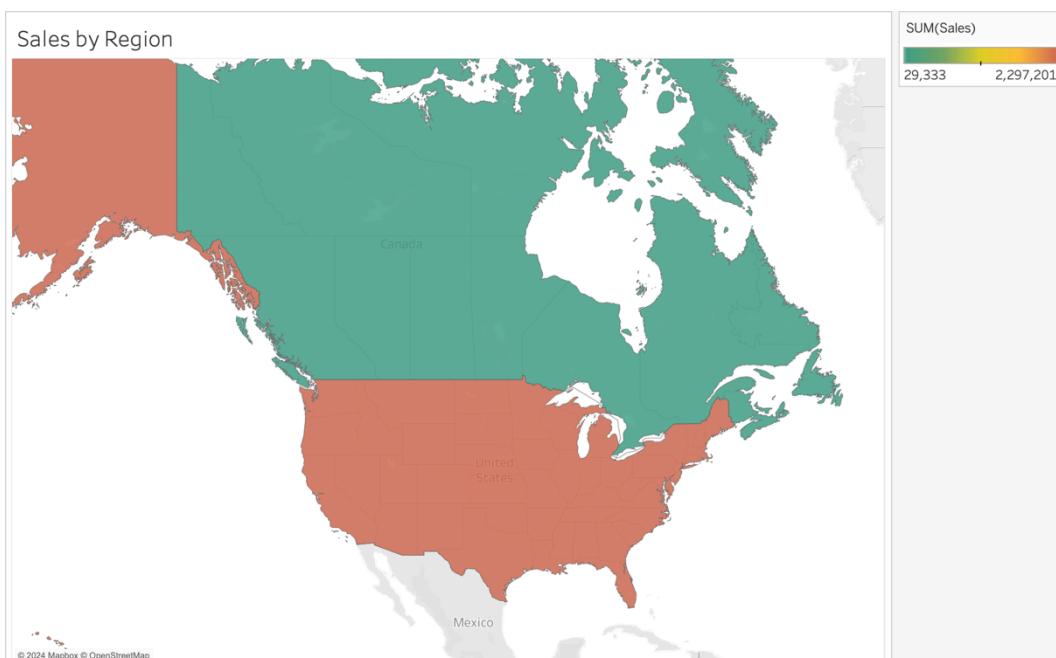


Fig. 13: Symbol plot depicting sales by country i.e., US and Canada)

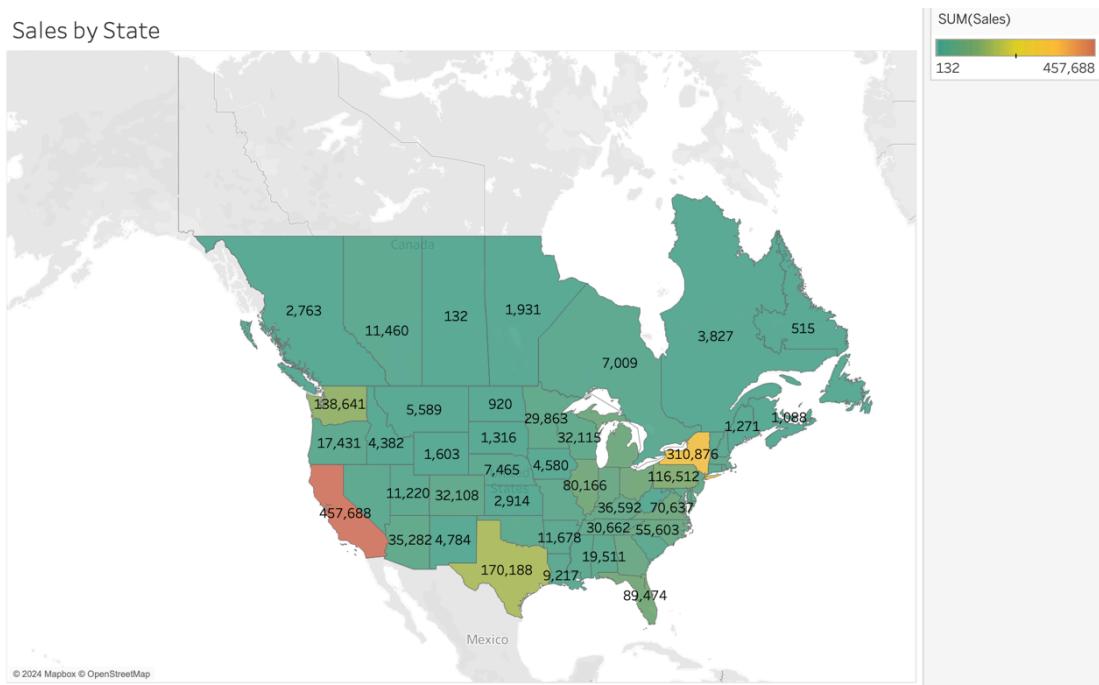


Fig. 14: Symbol map depicting sales distribution of different states.

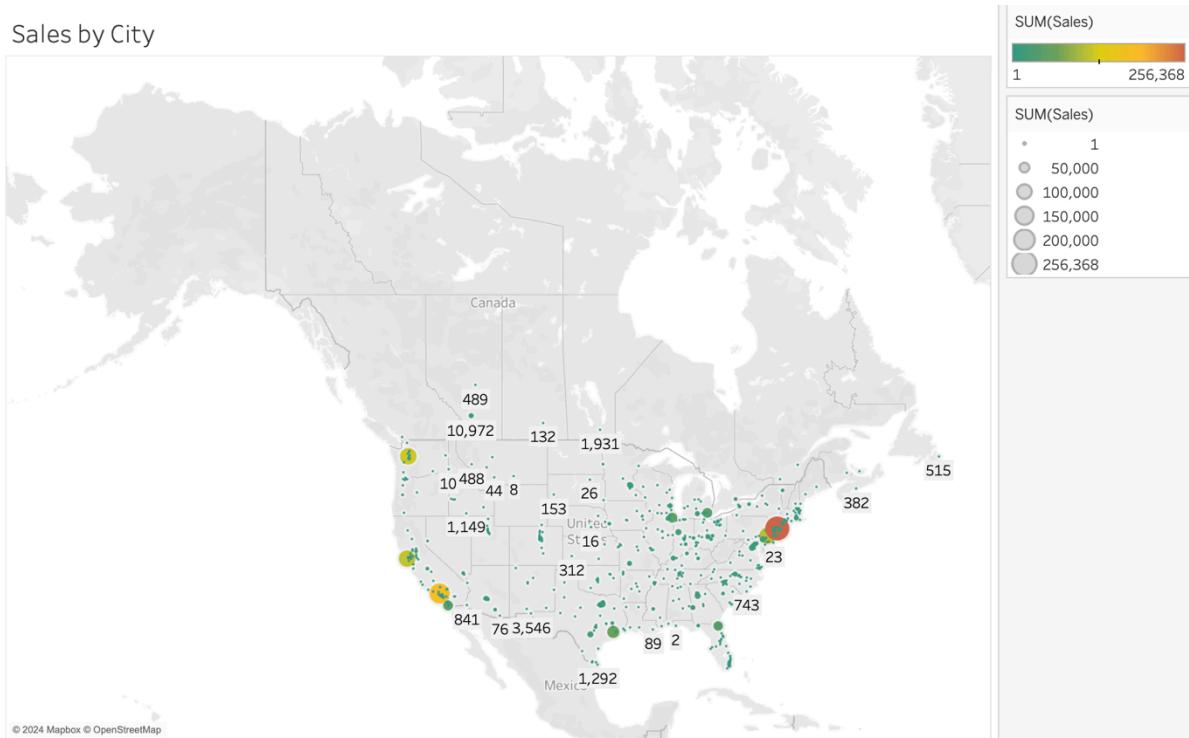


Fig. 15: Symbol map displaying sales distribution of different cities.

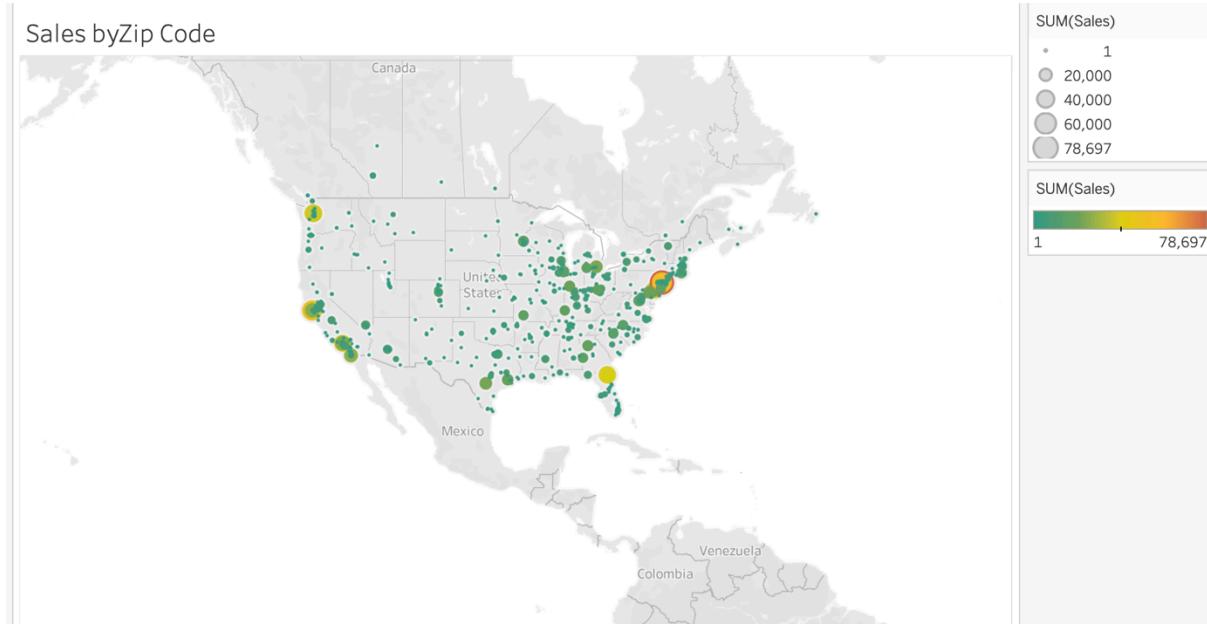


Fig. 16: Symbol map displaying sales distribution of different Zip codes.

Chart used- Symbol map.

Reasons for choosing Symbol map -

- Symbol maps provide a geospatial representation of sales distribution across different regions. Each region is represented by a symbol or marker on the map, allowing viewers to visually assess the distribution of sales across geographical areas.
- Symbol maps facilitate comparison of sales distribution across different regions.
- 8. Can we visualise the composition of profits across various subcategories within different customer segments?

Answer- In the following Stacked bar chart, the profit percentage for different categories of products.

Sheet 26

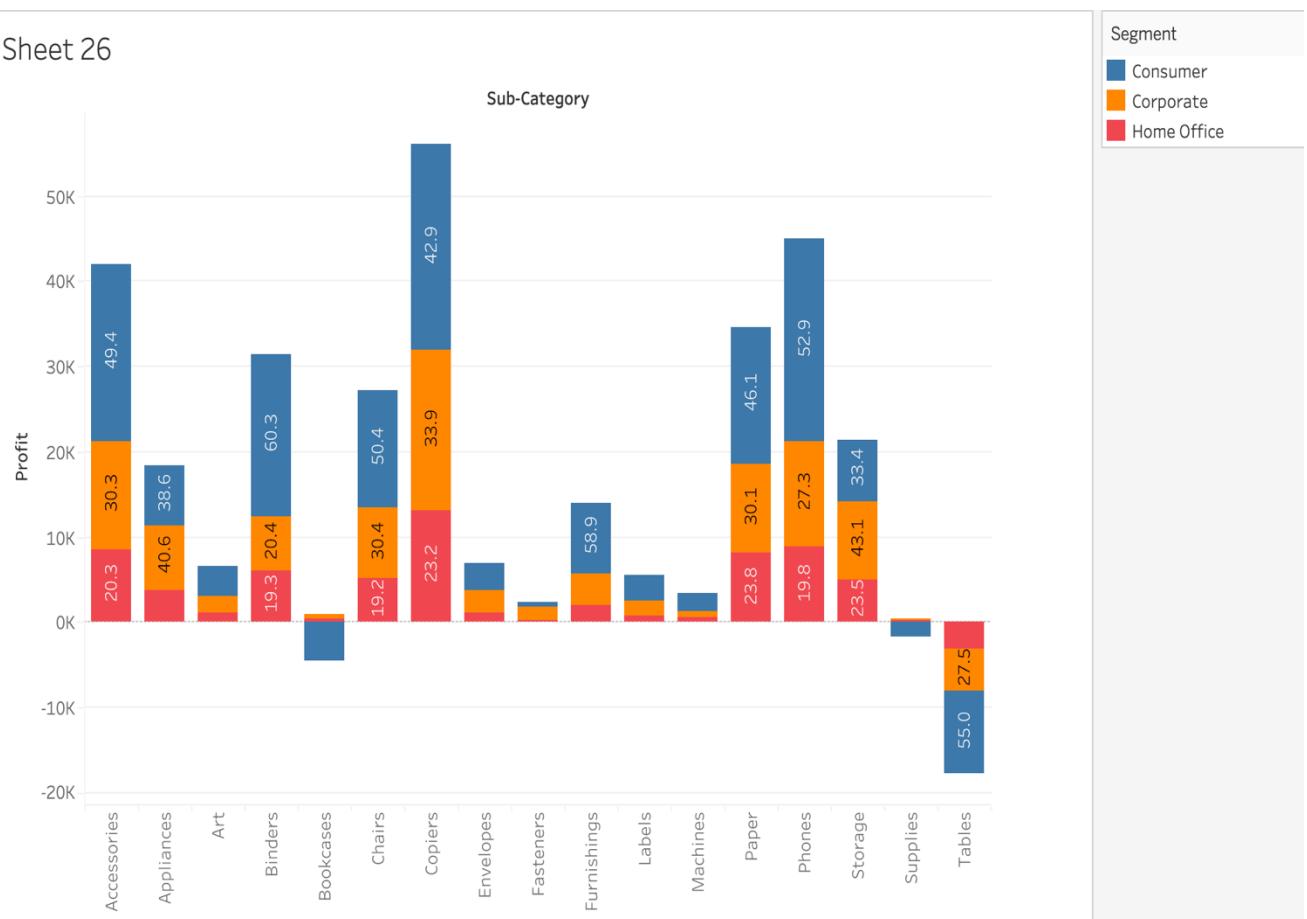


Fig. 17: Profit percentage of different segments for different product subcategories.

Chart used- Stacked bar chart.

Reasons for choosing Stacked bar chart -

- Stacked bar charts effectively display the composition of profits within different customer segments and subcategories. Each bar represents a customer segment, and the segments within each bar represent different subcategories of profits. This allows us to compare the contribution of each subcategory to the total profits within each customer segment.
- Stacked bar charts allow for segmented analysis, making it easy to analyze profits across different customer segments. Each bar can be segmented into subcategories, providing insights into how profits are distributed within each segment.
- Stacked bar charts show the relative proportions of profits within each customer segment and subcategory. We can easily see the proportion of profits contributed by each subcategory compared to others, as well as the total profits for each customer segment.

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

Answer- The sales percentage of western region is most (roughly 32%) and the sales for the southern region is the least (roughly 17%).

Sales Percentage by Region

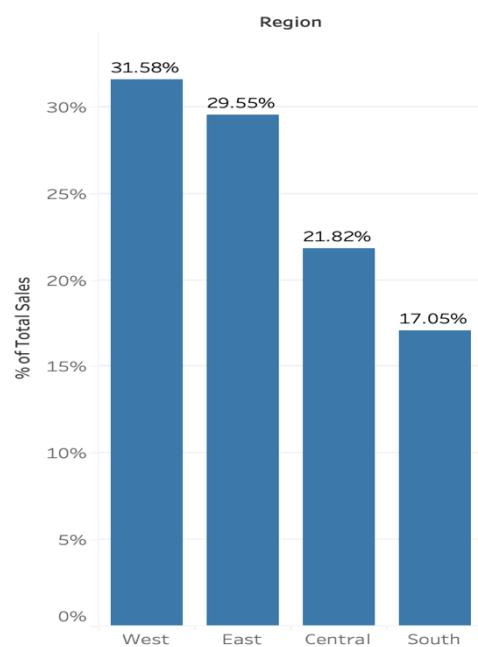


Fig. 18: Sales percentage distribution for different regions.

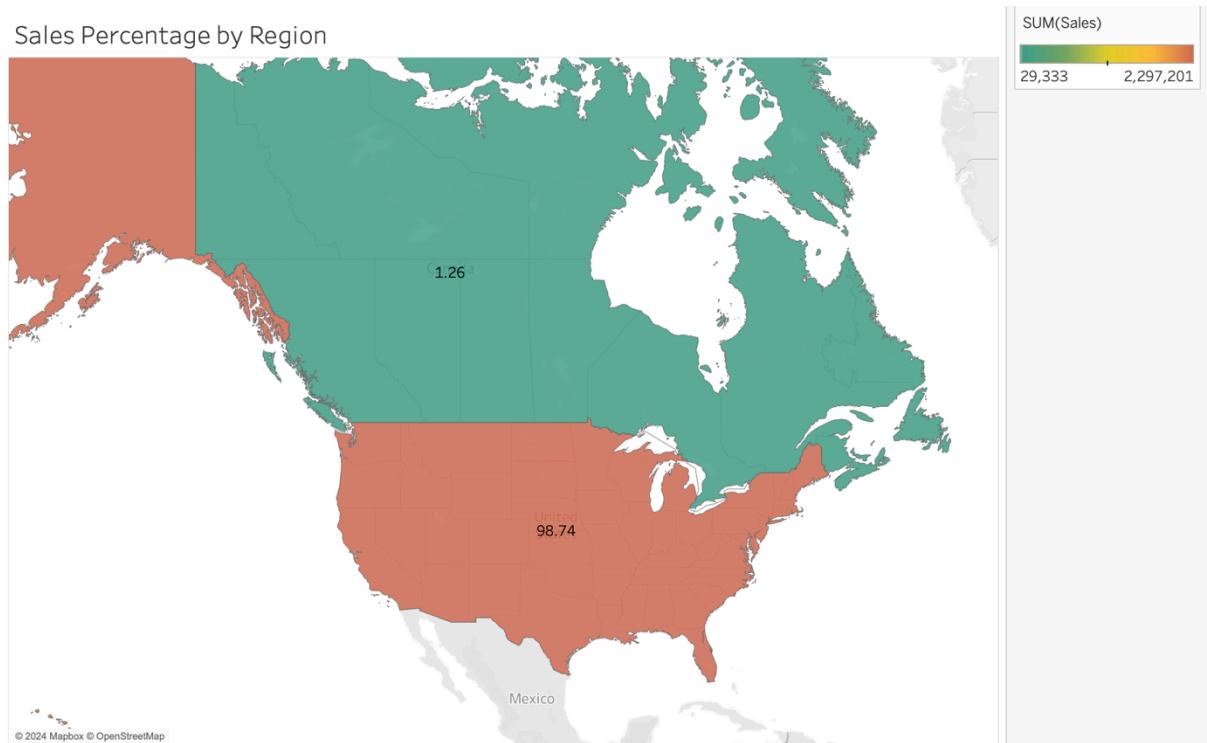


Fig. 19: Sales percentage by Region (Country)



Fig. 20: Sales percentage by Zip code.

Charts used- Bar chart and Symbol map.

Reason for choosing these charts –

- While bar chart help in easy-to-understand visuals for region wise percentage sales distribution, symbol map help in visualization of sales distribution geographically.

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

Answer- Yes, the following side by side bar graph is used for the same. Standard class shipping has the highest profits no matter what the customer segment.

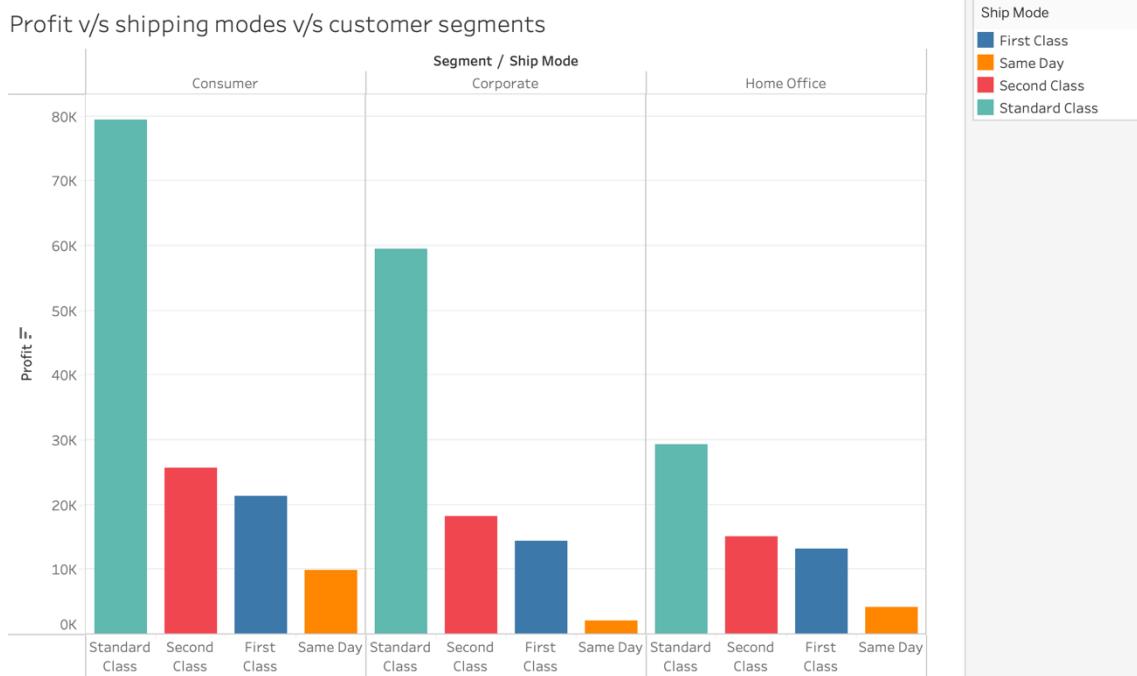


Fig. 21: Graph displaying profit margins associated with different shipping modes and customer segments.

Chart used- Side-by-side bars.

Reasons for choosing Side-by-side bars –

- Side-by-side bars allow for easy comparison of profit margins between different shipping modes and customer segments. By placing bars side by side, viewers can visually assess the differences in profit margins across the various categories.
- Side-by-side bars provide a clear and concise visualization of profit margins. Each pair of bars represents a specific combination of shipping mode and customer segment, making it easy for viewers to interpret the data and identify patterns or trends.

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

Answer- The average number of days taken to process the orders of Copiers (in the Technology category) is the minimum while a greater number of days are required for Office Supply products.

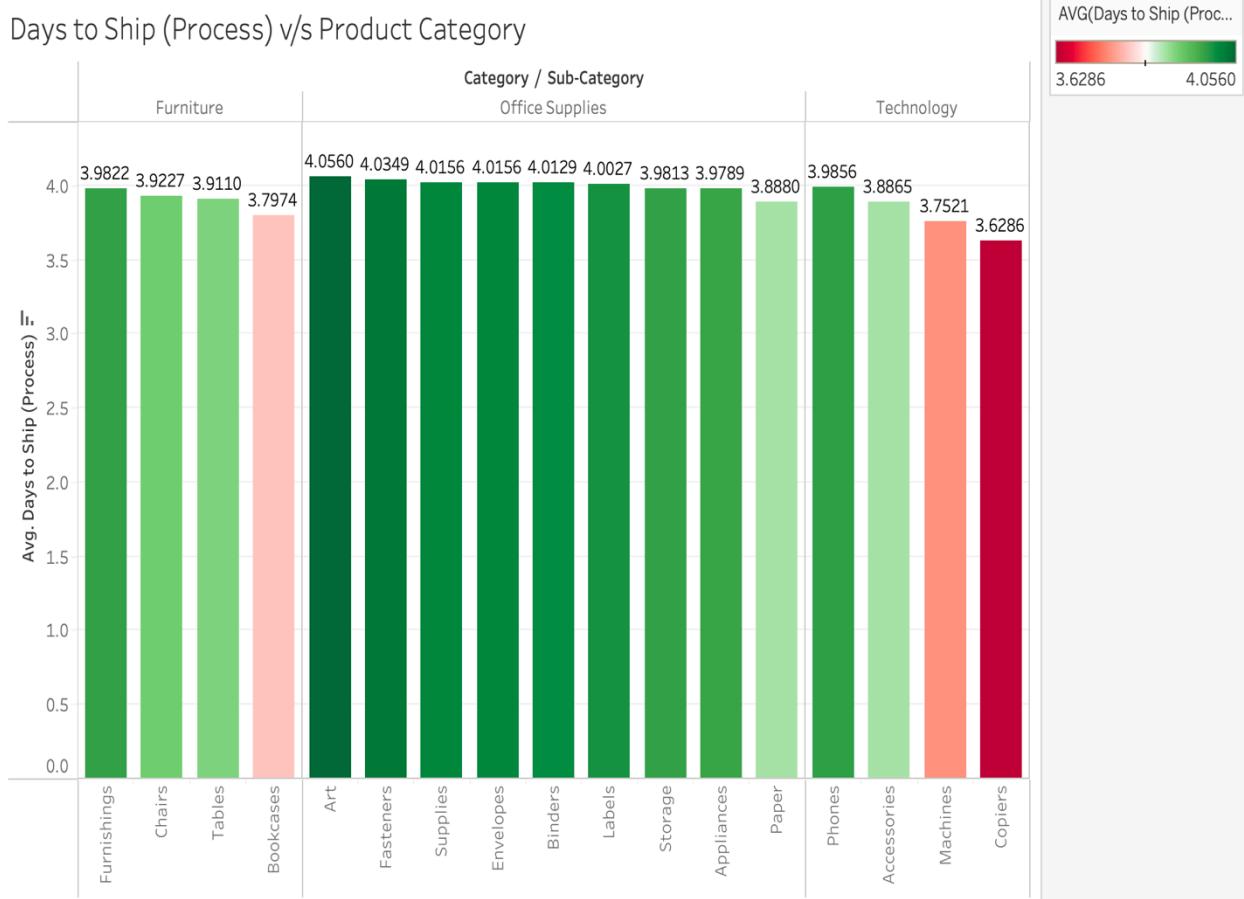


Fig. 22: Days to ship (process) v/s product categories and subcategories.

Chart used- Side-by-side bars.

Reason for choosing Side-by-side bars –

- Side-by-side bar charts are used here for easy comparison of average number of days required for processing of orders for different categories and sub categories of products.

- The red colour bars depicts lesser average days required for order processing while dark green bars denotes more number of days for order processing.

12. How do discounts affect overall profit?

Answer-When discount is zero, profit is highest as customers have to pay the full price. Hence, the profit decreases as discounts increase. However, there is a rise in profit when the discount is roughly around 0.2 of the price.

Effect of Discount on Profit

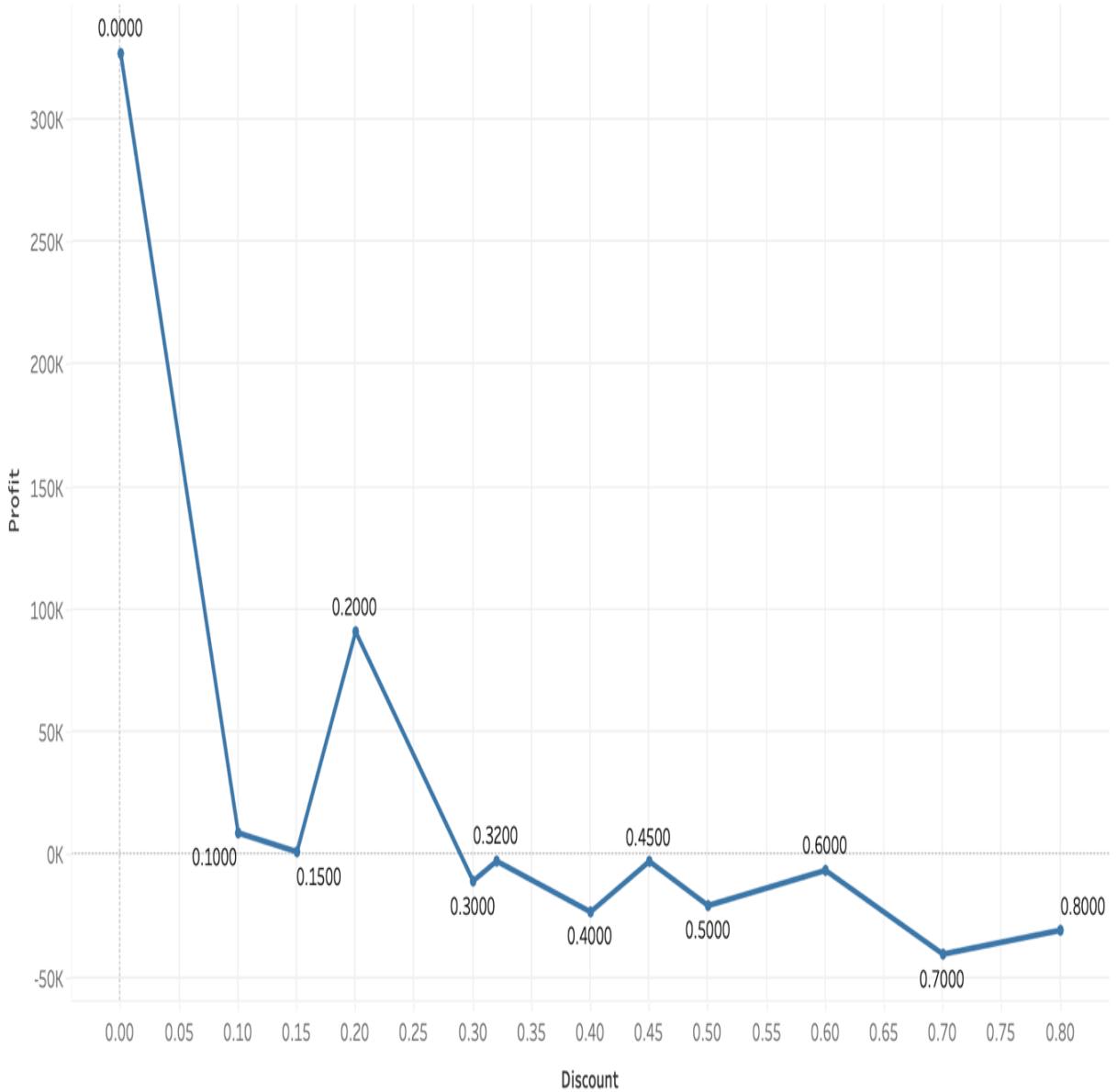


Fig. 23: Line chart depicting effect of Profit on different values of discount.

Chart used- Line chart.

Reason for choosing Line chart -

- Line chart is used here for viewing the effect of discount values on profit.
- It shows the relationship between discount and profit in a trend line format which is quite easier to interpret.

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

Answer- Yes, the following stacked bar chart describes the same.

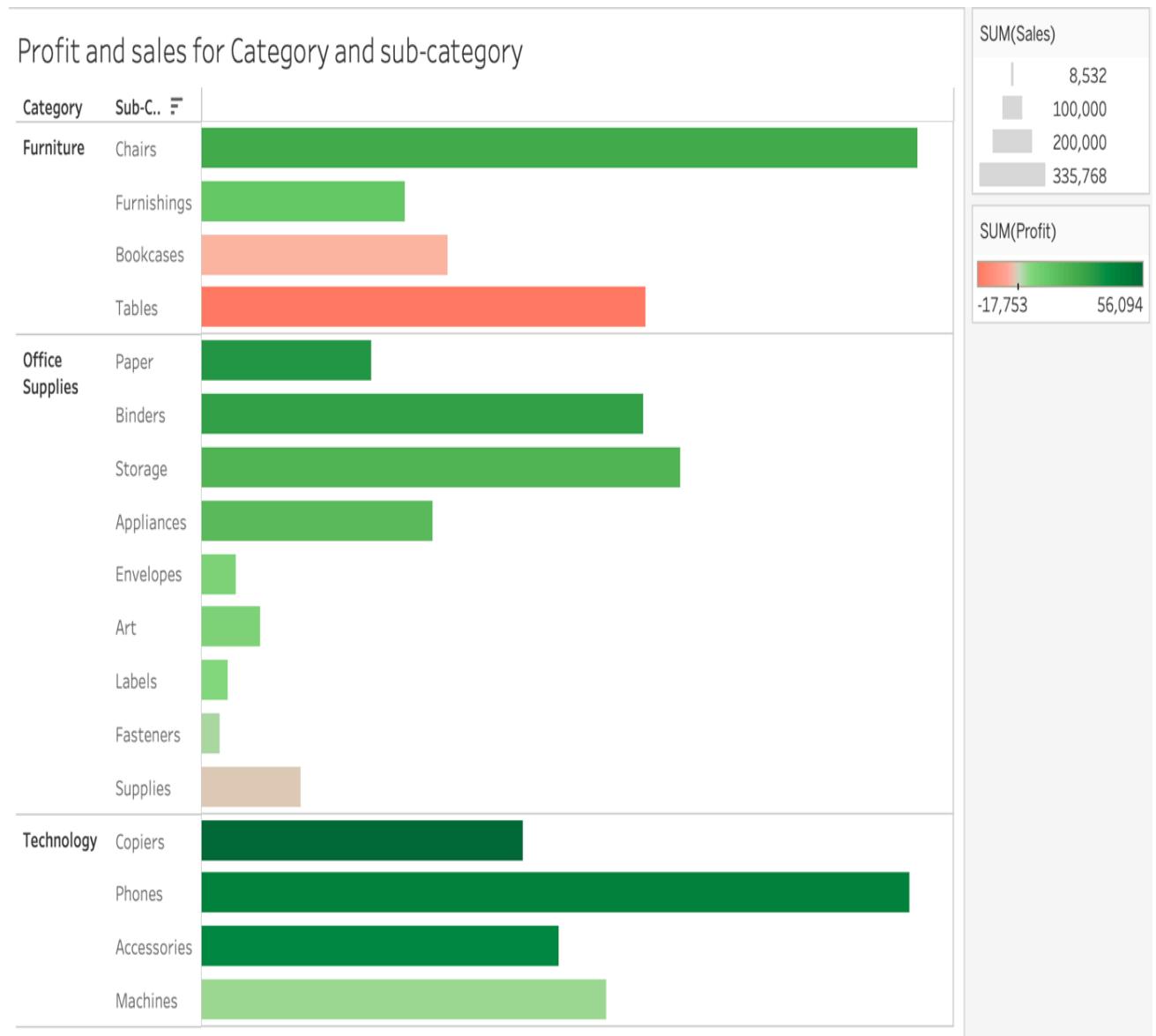


Fig. 24: Chart displaying relation between sales (shown by size of bar), profit (shown by colour of bar) and product categories.

Chart used- Bar chart.

Reason for choosing Bar chart -

- The size of the horizontal bars for different categories and subcategories of products efficiently describes the sum of sales.

- The colour of the horizontal bars describes the sum of profits in an easy-to-understand manner.
- We can compare the sales and profit distributions for different products easily just by looking at the chart.

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

Answer- From the following graph, it can be inferred that in Furniture category, the Furnishings are highest in quantity, followed by Chairs and the lowest are Bookcases. Similarly, for Office Supplies category, the maximum number of products are Binders and the least quantity of products sold are Supplies. Also, for Technology category, Phones are the most sold product and the least sold are Copiers.

Distribution of order quantities for different product categories.

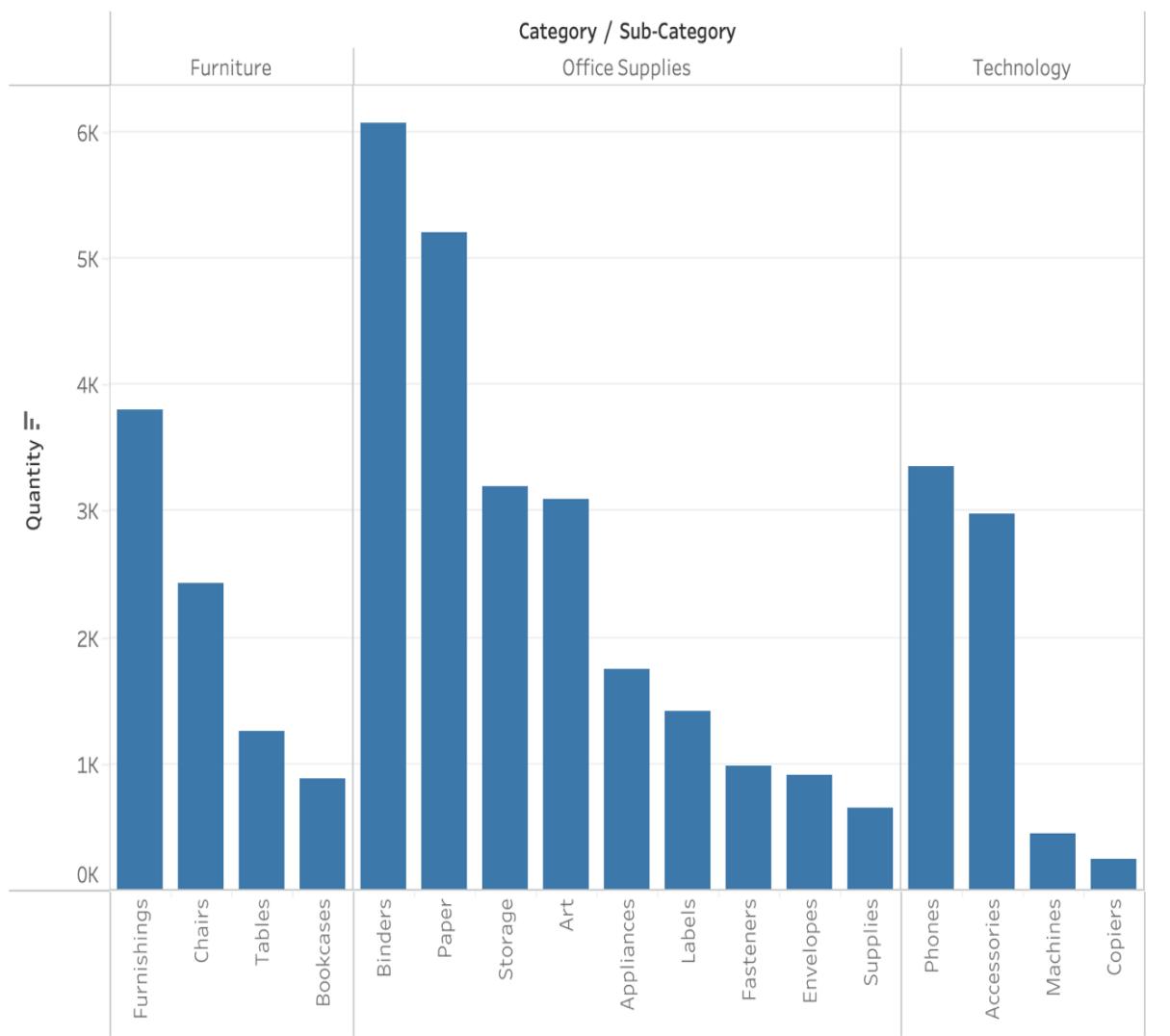


Fig. 25: Graph showing distribution of order quantities for products in the dataset.

Chart used- Side-by-side bar.

Reason for choosing Side-by-side bar –

- Vertical side-by-side bars conveniently display the total quantities of different products sold along with their sub categories.

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

Answer- Profits of products ‘Copiers’ and ‘Phones’ are highest in the Consumer market segment. Other popular high profit categories in this segment are ‘Accessories’, ‘Binders’, and ‘Paper’. ‘Copiers’ are also producing high profit in ‘Corporate’ segment, followed by ‘Home Office’ segment of customers. Tables are being sold in highest loss in all three segments followed by ‘Corporate’ and ‘Home Office’ segments.

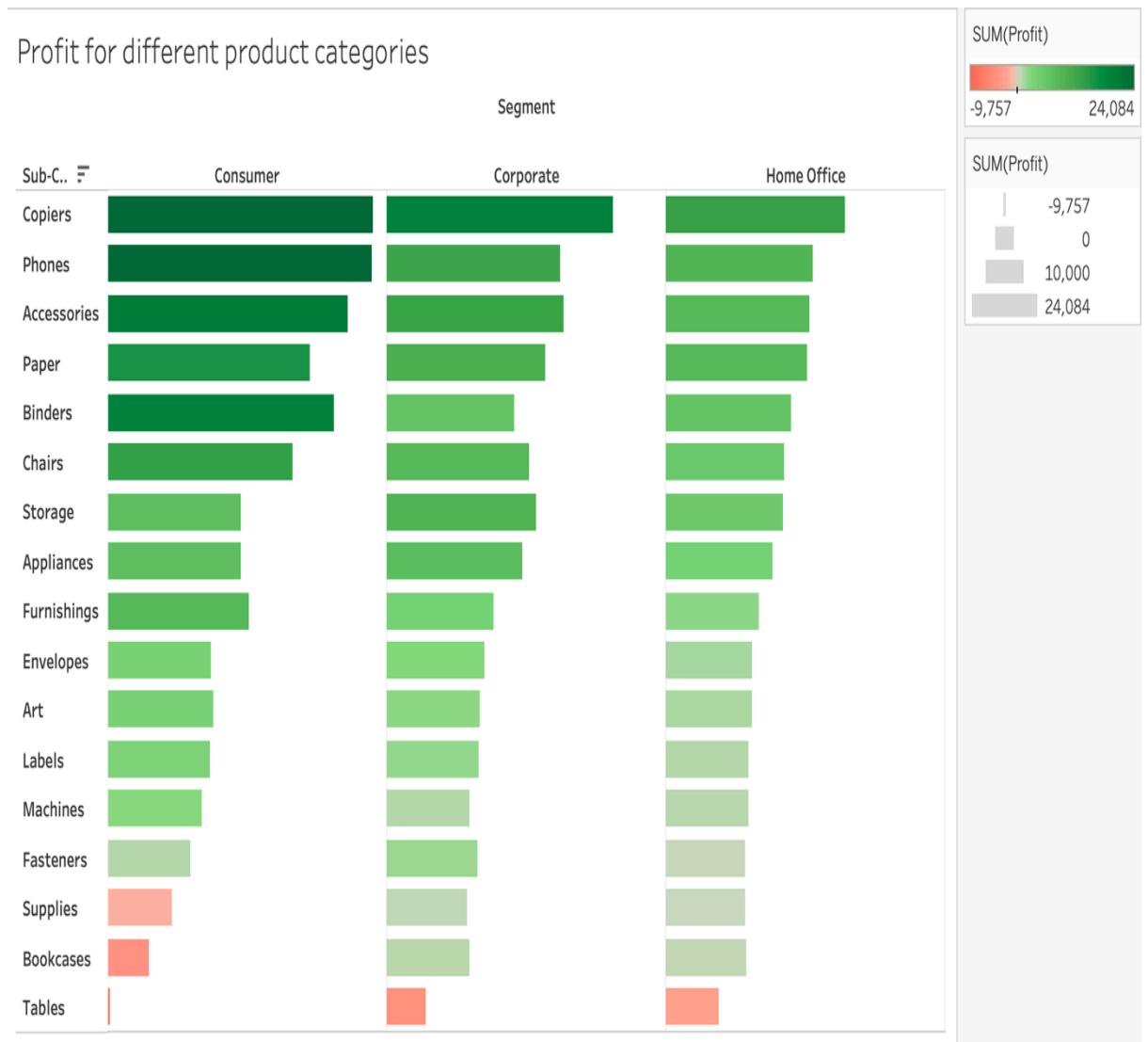


Fig. 26: Profit for different product categories

Chart used- Side-by-side bar.

Reason for choosing Side-by-side bar –

- Side-by-side bar chart is chosen for easy analysis and comparison of profits and losses incurred on different products and their subcategories.
- The greener and longer bars denotes high profit while the shorter and redder bars denote less profit (i.e., losses).

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

Answer- Yes, the following graph compares the shipping time (in days) distribution of different shipping modes of orders.

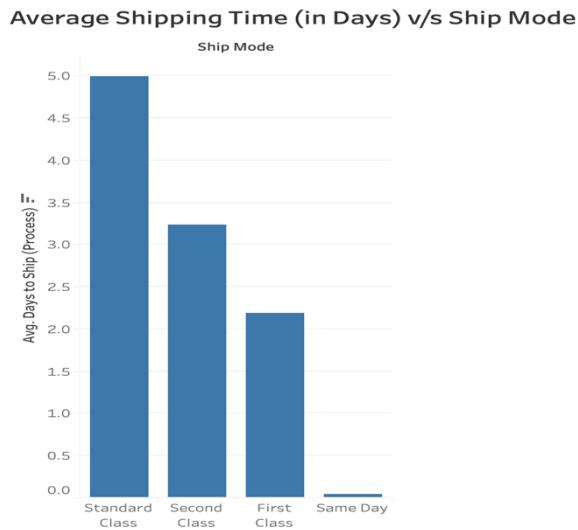


Fig. 27: Average shipping time (in days) v/s ship mode.

Chart used- Bar chart.

Reason for choosing bar chart-

- For simple analysis and comparison of shipping days of different shipping modes without drawing any complex charts.

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

Answer- The monthly trend shows higher number of orders during December (maybe due to Christmas) and lowest number of orders in the first few months.

No. of Orders Shipped v/s Months.

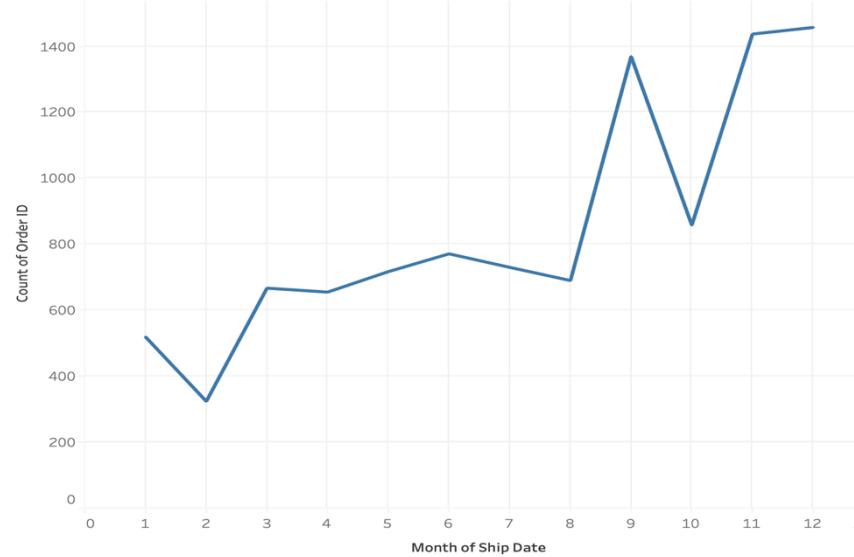


Fig. 28: number of orders sold in different months.

Chart used- Line chart.

Reason for choosing Line chart-

- Line chart is used here since we are dealing with trends over time.

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

Answer- The ‘Consumer’ customer segment produces the most sales and profit, followed by ‘Corporate’. The least amount of sales and profit are in the ‘Home Office’ segment.

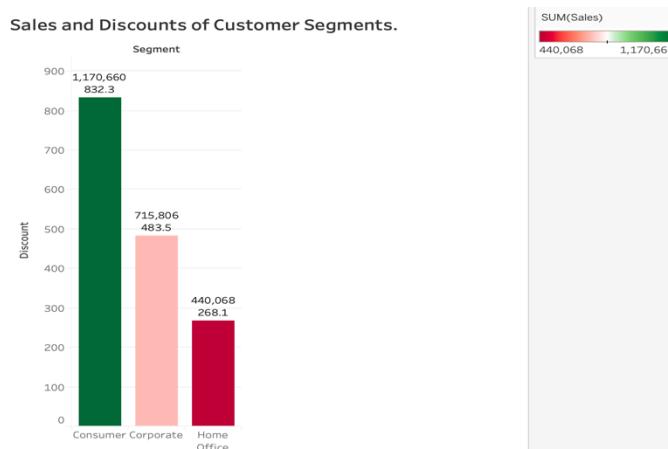


Fig. 29: Sales and Discount of different customer segments

Chart used- Bar chart.

Reason for choosing bar chart-

- For simple analysis and comparison of shipping days of sales and discounts of different customer segments.

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



Fig. 30: Heatmap showing sales (shown by label and size of square) and profit (shown by colour of square) trends across different product subcategories and regions.

Answer- The squares that are big in size and are green are the best for business. Also, for some products, the profit is less even if the sales are high.

Chart used- Heatmap with labels.

Reasons for choosing Heatmap with labels-

- Heatmaps provide a comprehensive overview of sales and profit trends across multiple dimensions (product subcategories and regions) simultaneously. Each cell in the heatmap represents the intersection of a product subcategory and a region, allowing viewers to quickly identify patterns and trends in sales and profits.
- Heatmaps use color-coding to represent quantitative values, such as sales and profits. By assigning different colours to different ranges of values, viewers can easily interpret the data and identify areas of high or low performance.
- By observing the colour intensities and textual labels, viewers can quickly identify which subcategories and regions are performing well or underperforming.
- Heatmaps offer a compact visualization of complex data, allowing viewers to analyse a large amount of information in a relatively small space. This makes it easy to identify trends and patterns without overwhelming the viewer with too much detail.

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

Answer- The following graph displays the the average delivery duration for different regions and ship modes.

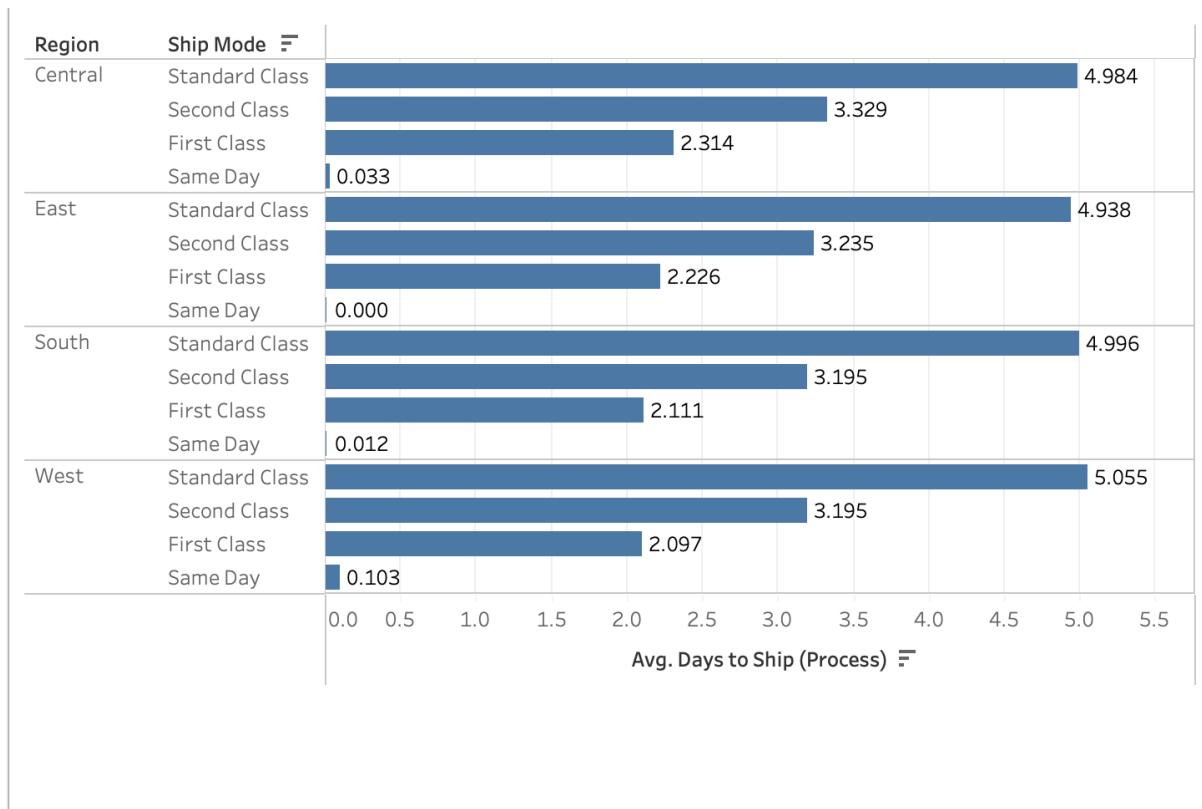


Fig. 31: Graph depicting the average days to ship i.e., deliver the order in different regions by different ship modes.

Chart used- Horizontal bar chart.

Reason for choosing Horizontal bar chart-

- It is used here for easy comparison of average ship days for different shipping modes in different regions.

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

Answer- Over the years, the average order quantity for 'Furniture' has fallen down while for 'Office Supplies' remains constant. Also, for 'Technology', there was a dip in average order quantity in 2021 but it has gradually increased since then.

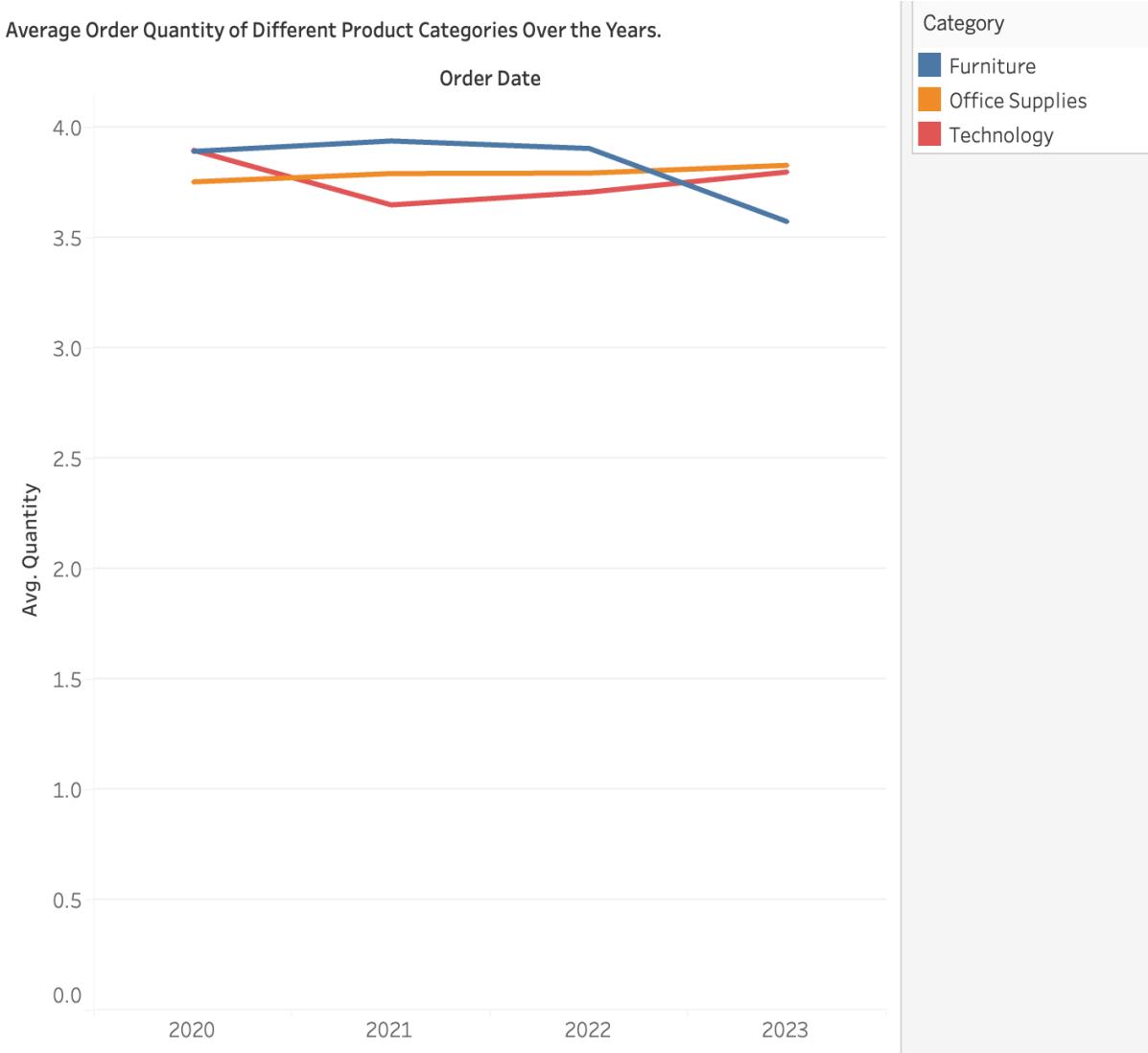


Fig. 32: Line chart showing average order quantity changed over the years for various product categories.

Chart used- Line chart.

Reason for choosing line chart-

- Line chart is chosen here since we want to analyse the order quantity trends over time.

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

Answer- Yes, the following scatter plot helps in understanding the relation between discount and order quantities.



Fig. 33: Scatter plot showing discount rates and order quantities for different customer segments.

Chart used- Scatter plot.

Reason for choosing scatter plot-

- Scatter plots are the most popular choice when trying to understand the correlation between two quantities.

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

Answer- The proportion of returned orders is the highest in the West, as shown in the following graph:

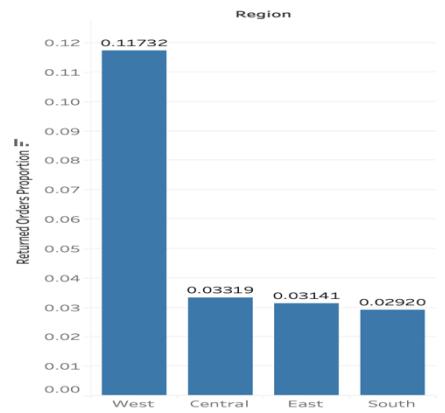


Fig. 34: Bar graph depicting the proportion of orders returned in each region within the Superstore dataset.

Chart used- Bar graph.

Reason for choosing bar graph-

- It is used here for quick and easy analysis and visualization.

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

Answer- Yes, from the following graph, it can be inferred that Copiers produce the most profit, followed by Phones and Accessories.

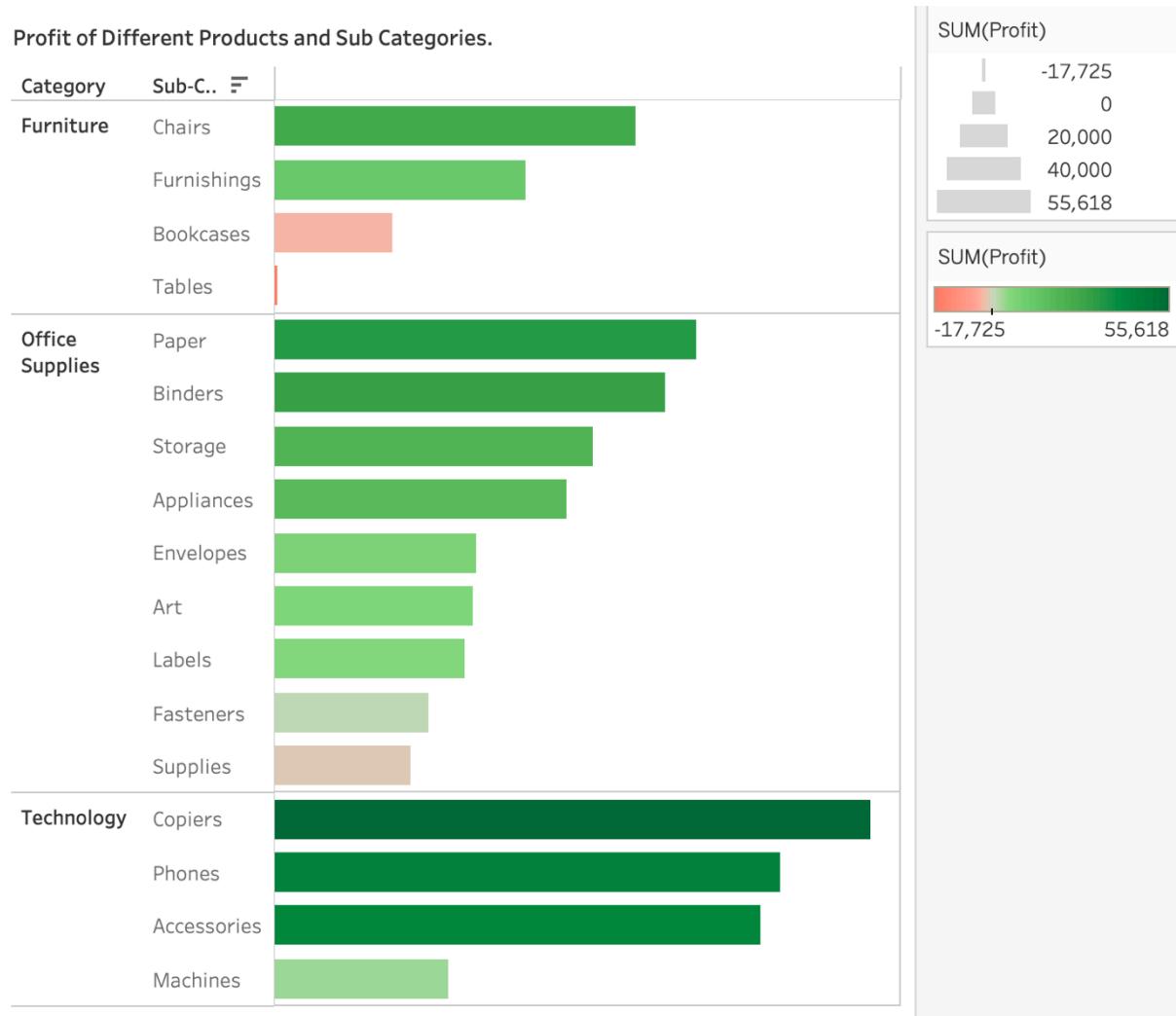


Fig. 35: Graph showing profits of different categories and sub categories of products.

Chart used- Horizontal side-by-side bar graph.

Reasons for choosing horizontal side-by-side bar graph-

- Horizontal side-by-side bar graphs allow for a clear and direct comparison of profits between different products within each subcategory. Each bar represents a product, and the length of the bars directly corresponds to the profit, making it easy for viewers to compare profits visually.
- Horizontal side-by-side bar graphs naturally facilitate ranking, as viewers can easily identify which products within each subcategory have the highest or lowest profits based on the length of the bars.

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

Answer- From the following graph, it can be concluded that the Standard Class is the most frequently used shipping mode.

Count of Orders for Different Shipping Modes

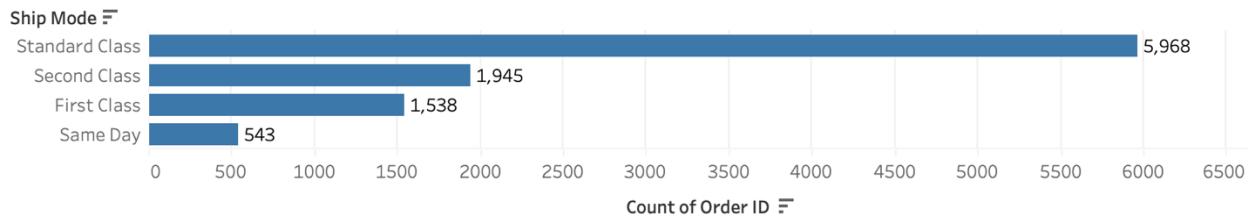


Fig. 35: Horizontal bar graph

Chart used- Horizontal bar graph.

Reason for choosing horizontal bar graph-

- To compare the number of orders via different shipping modes easily.

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

Answer- Sales performance is seen to increase over the different quarters of different years.

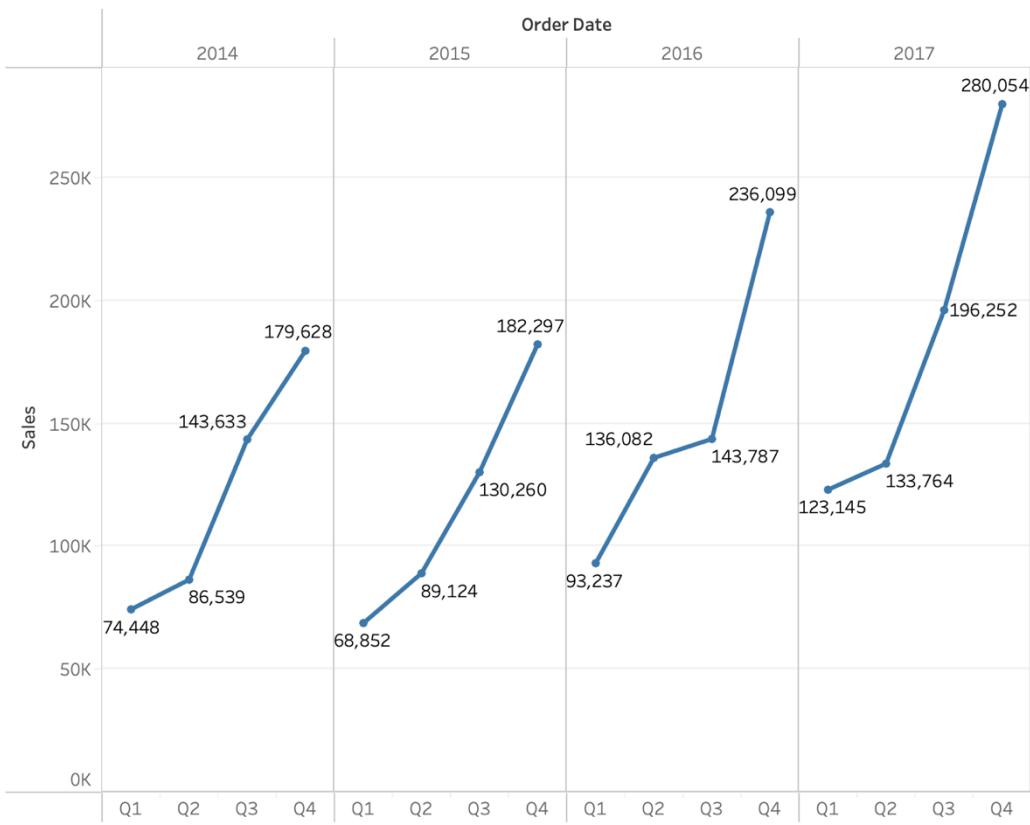


Fig. 36: Line charts depicting sum of sales in different quarters of different years.

Chart used- Line chart.

Reason for choosing Line chart-

- This side-by-side line chart is used here to display the comparison between sales in different quarters in different years.

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

Answer- The distribution of order priorities across different product categories is shown in the following graph. Here, orders with High priority refer to those that opt for ‘Same Day’ delivery in shipping. The Medium priority is for ‘First class’ and ‘Second class’ shipping modes while Low priority is for ‘Standard’ shipping mode since it takes the maximum order processing time.

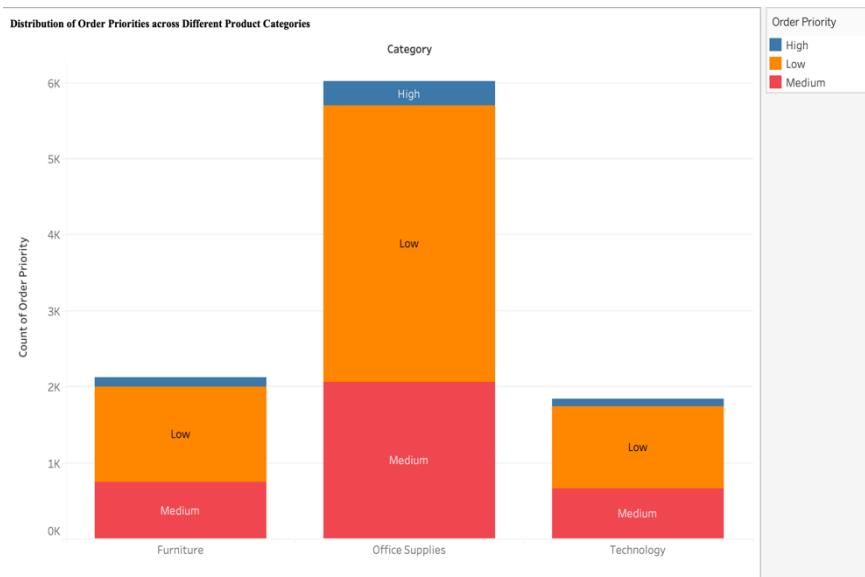


Fig. 37: Distribution of order priorities across different product categories.

Chart used- Stacked bar chart.

Reason for choosing Stacked bar chart-

- Here, stacked bar charts denote the proportion of orders with High, Medium and Low priority for different product categories.

28. What is the relationship between discounts and sales?

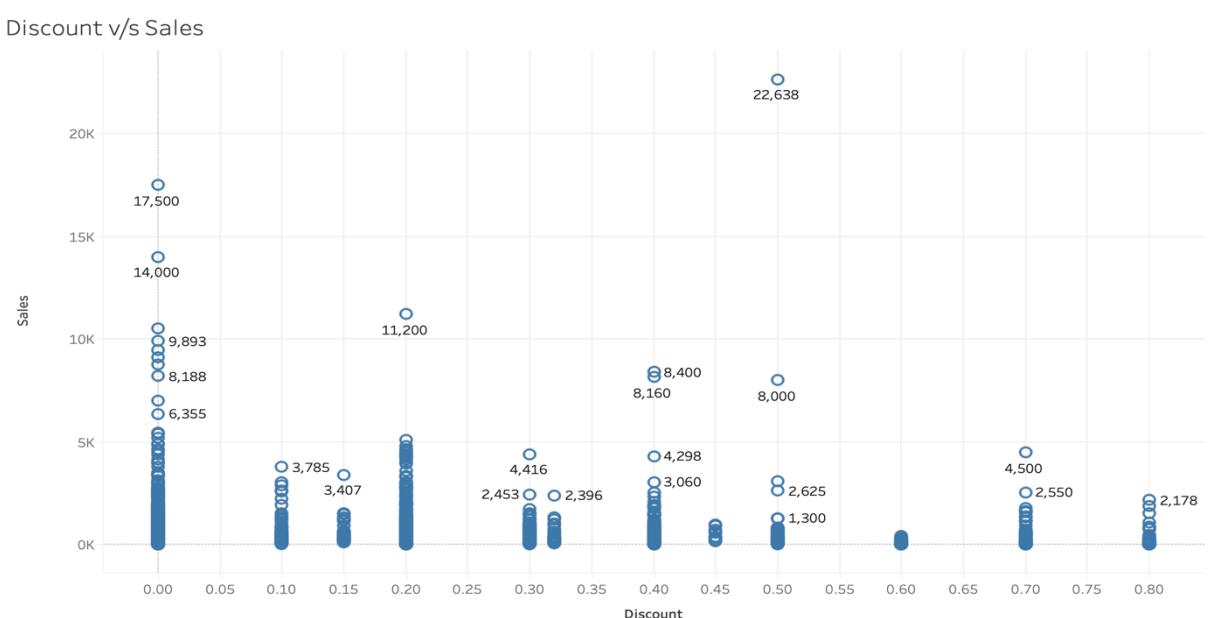


Fig. 38: Discount v/s Sales plot

Answer- For low discounts also, sales are high as customers are still willing to buy.

Chart used- Scatter plot.

Reason for choosing Scatter plot-

- Scatter plot is chosen here since it is the best plot for showing relationship between two variables in this manner.

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

Answer- From the following graph, it can be concluded that new customers are very less with low order value.

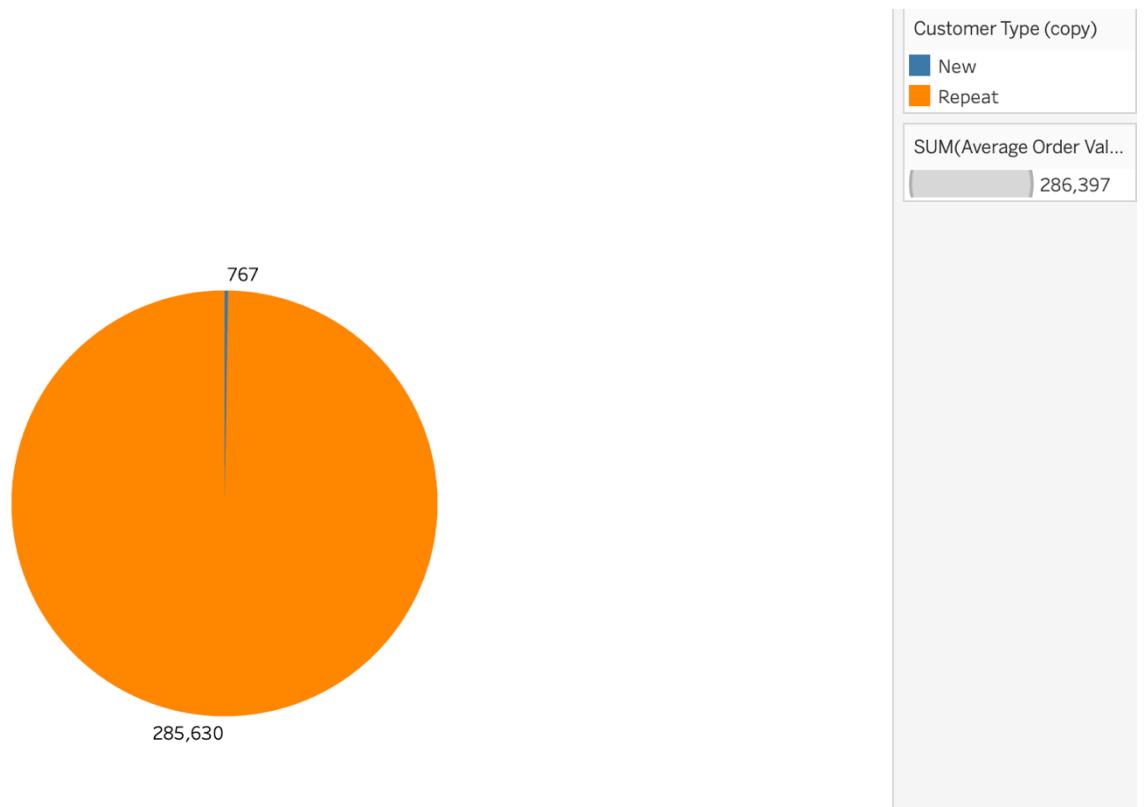


Fig. 39: Average order values for new and repeating customers

Chart used- Pie chart.

Reason for choosing Pie chart-

- Pie charts excel at showing proportions or percentages of a whole. In this case, if the focus is on the proportion of average order values contributed by repeat and new customers to the total, a pie chart can make it easy to see these proportions at a glance.
- If one category (e.g., repeat customers) significantly dominates the average order value compared to the other category (e.g., new customers), a pie chart can effectively highlight this dominance by emphasizing the size difference between the corresponding pie slices.
- Pie charts are simple and concise, presenting information in a compact format that is easy to understand. They can quickly convey the relative importance of repeat and new customers in terms of average order values without overwhelming the viewer with excessive detail.

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

Answer- The number of orders returned is highest in the western region but the profit is also highest here. Similarly, the number of returns in lowest in the southern region but so is the profit.

Effect of Returns in Different Regions on Profitability

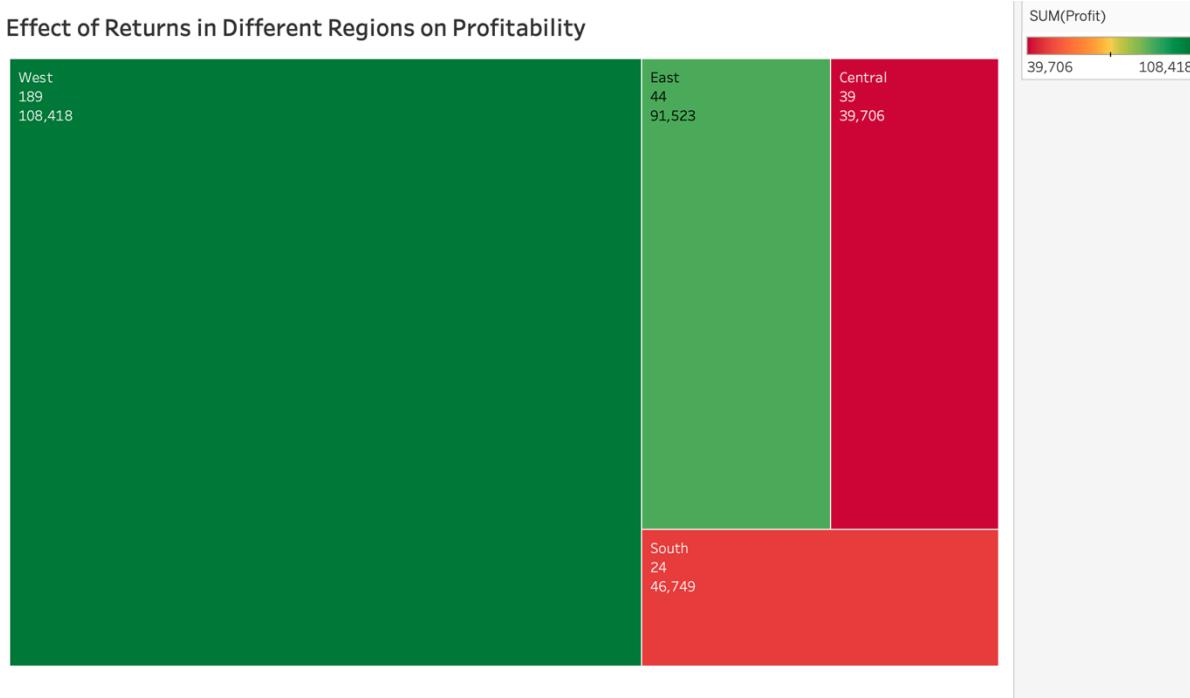


Fig. 40: Graph depicting the effect of returns in different regions on profitability.

Chart used- Tree map

Reasons for choosing Tree map chart-

- The size and colour of each rectangle in the tree map represents the profitability associated with each geographic region.
- Tree maps use space efficiently by arranging rectangles of varying sizes within a bounded space. This allows for the visualization of a large amount of geographical data in a compact and organized manner, making it easier for viewers to grasp the overall distribution of returns across different regions.