**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 charts 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.

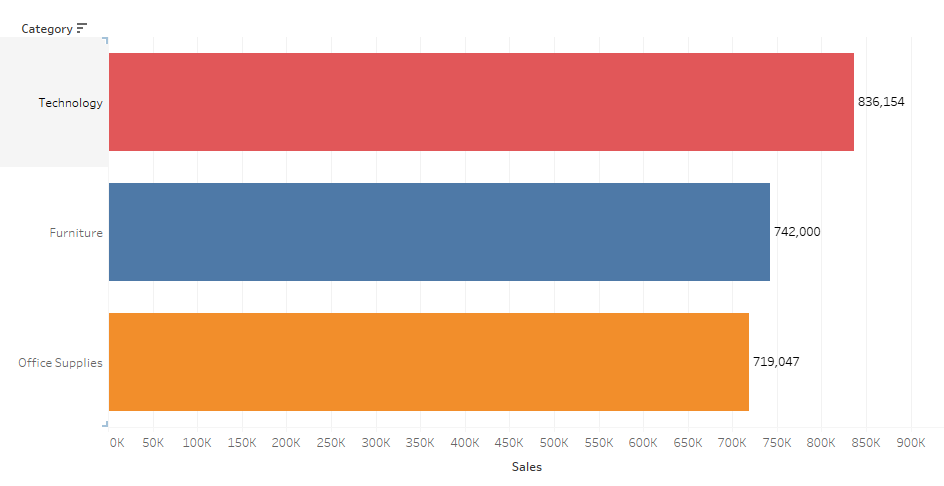
By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

**Good luck!**

**Tip :** Please zoom in for a better view.

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

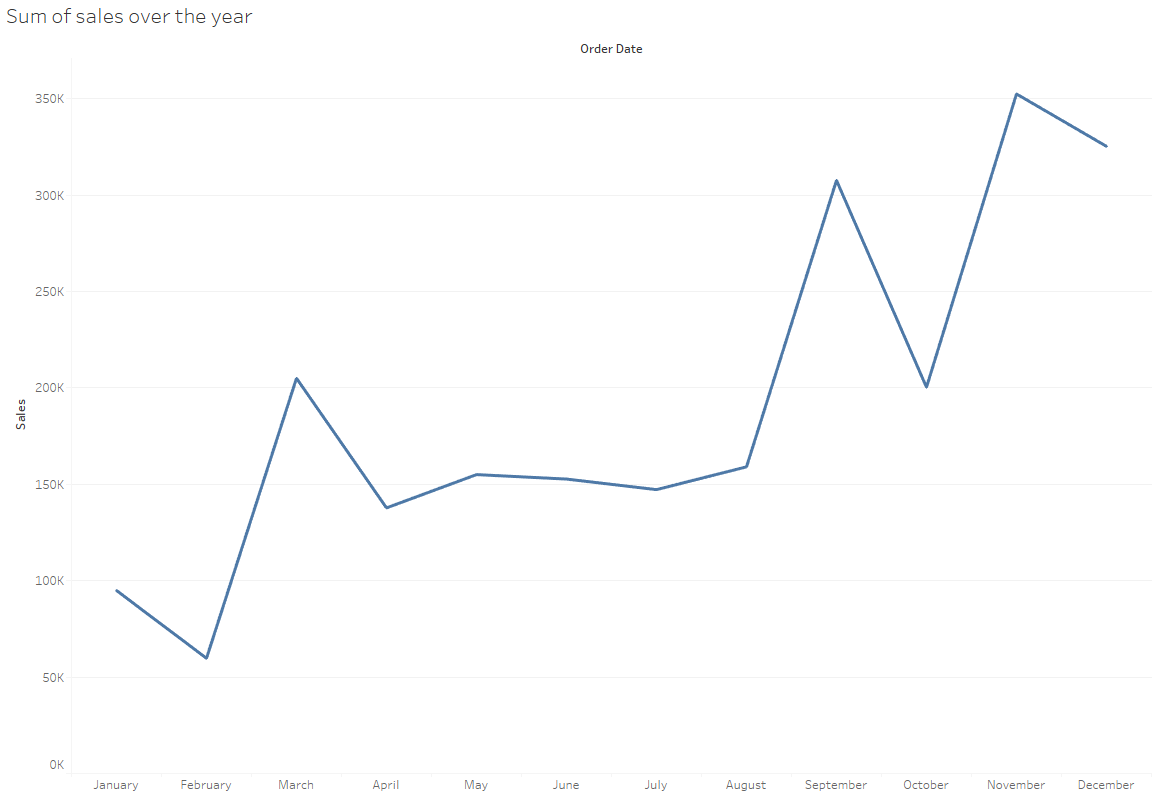
**Ans.** The product categories that have the highest total sales are Technology (836,154), Furniture (742,000), Office Supplies(719,047).



**Justification :** Bar charts are used to show comparison in categorical data, here being ‘category’ and the size of the bars indicate the total sales, hence I have used this chart.

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

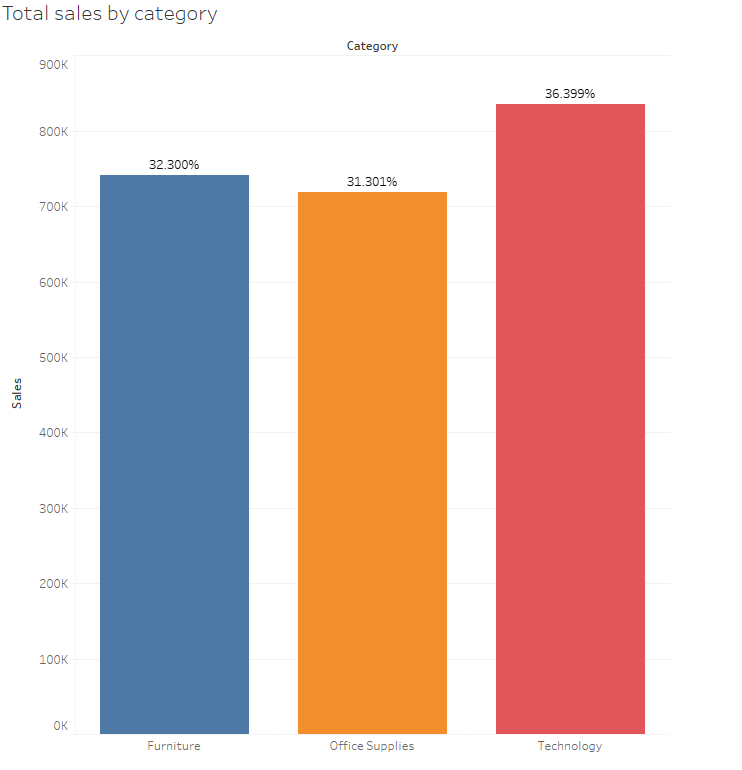
**Ans.** Below picture shows monthly sales amounts over the course of the year. Most sales are in November and least are in February.



**Justification :** Line charts are good to see trends over time, hence line chart was used.

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

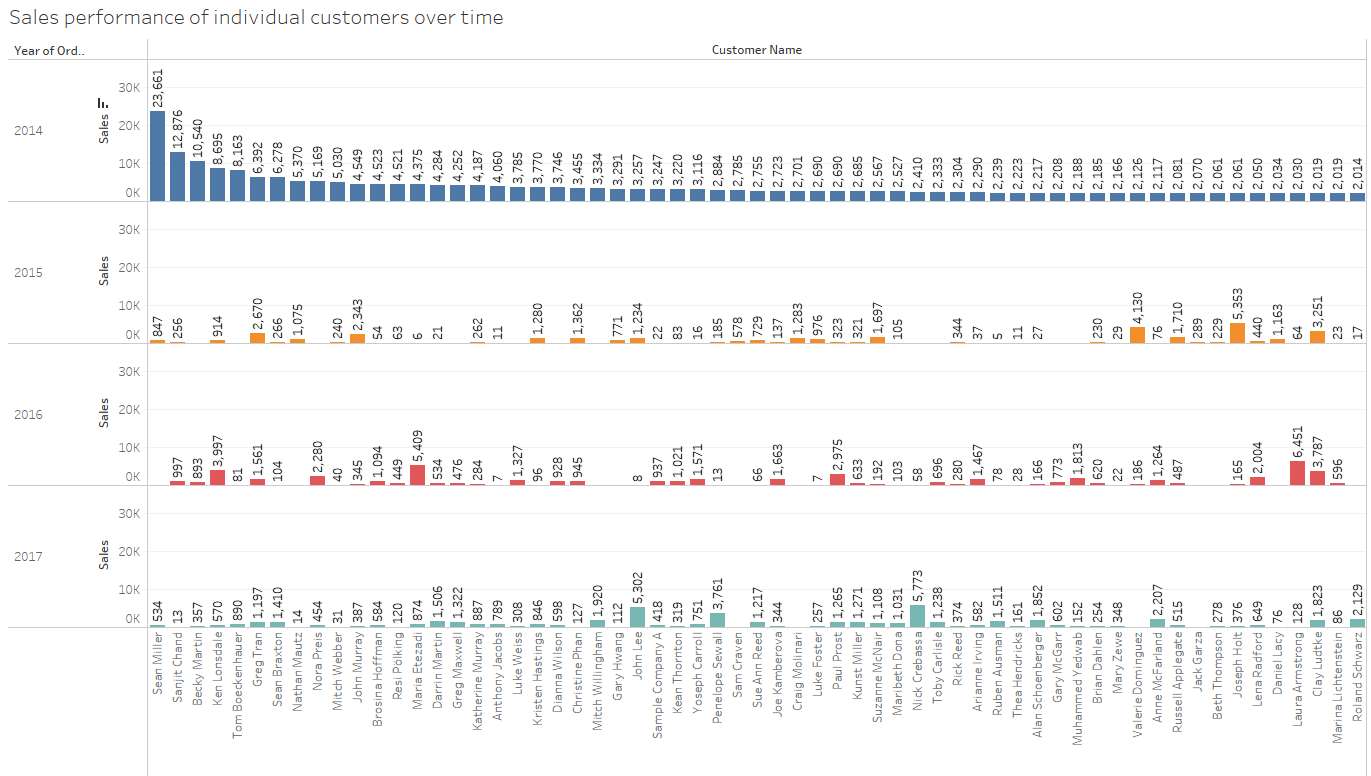
**Ans.** Below chart shows total sales amount distributed by product category. Technology has the highest sales percentage of 36.399% and Office Supplies have the lowest sales of 31.301%.



**Justification :** Bar chart with added table calculation shows the share of sales by category, hence bar chart is used.

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

**Ans.** Below chart shows sales performance of individual customers over time. Sean Miller had the highest sales performance in 2014(23,661).

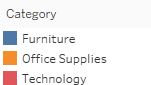
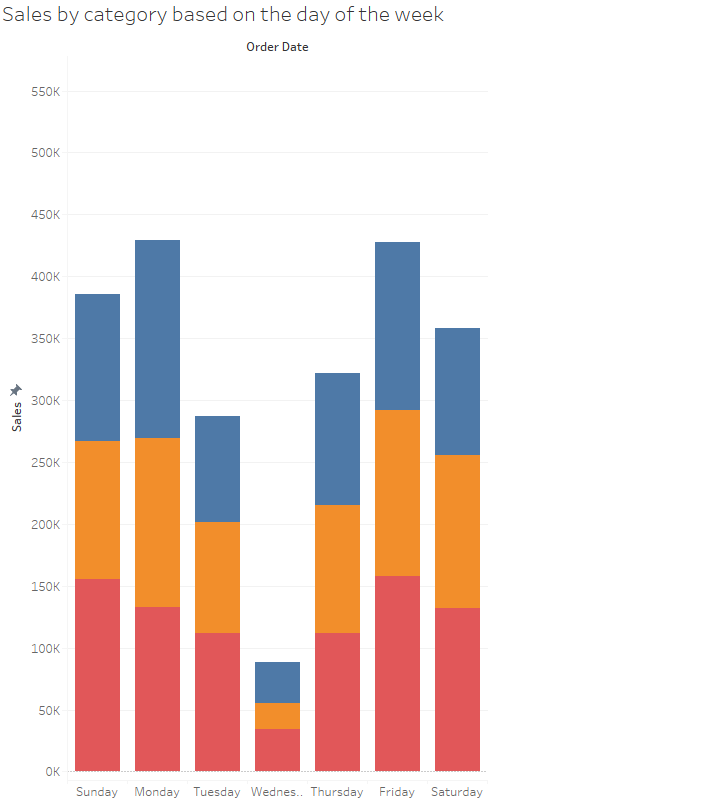


**Justification :** Bar chart divided by year into columns shows the sales data of

different customers over time.

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

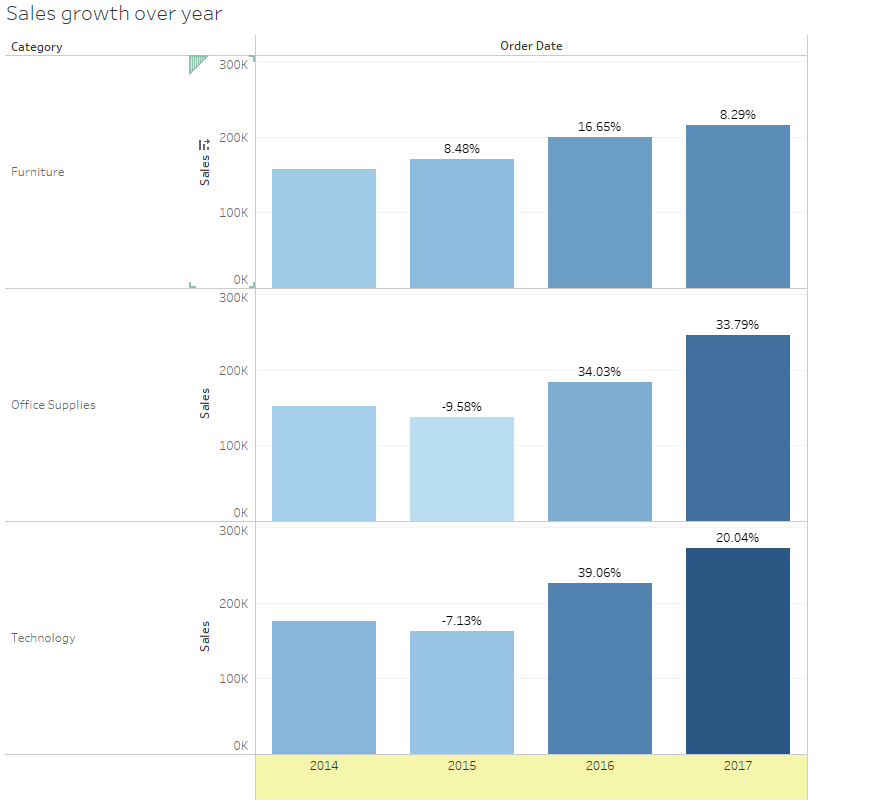
**Ans.** Sales based on different days of the week and product categories are shown in the chart below. Wednesday has the lowest cumulative sales and Monday and Friday have the highest.



**Justification :** Column chart with different colours based on the product category shows the required visualisation accurately, hence used.

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

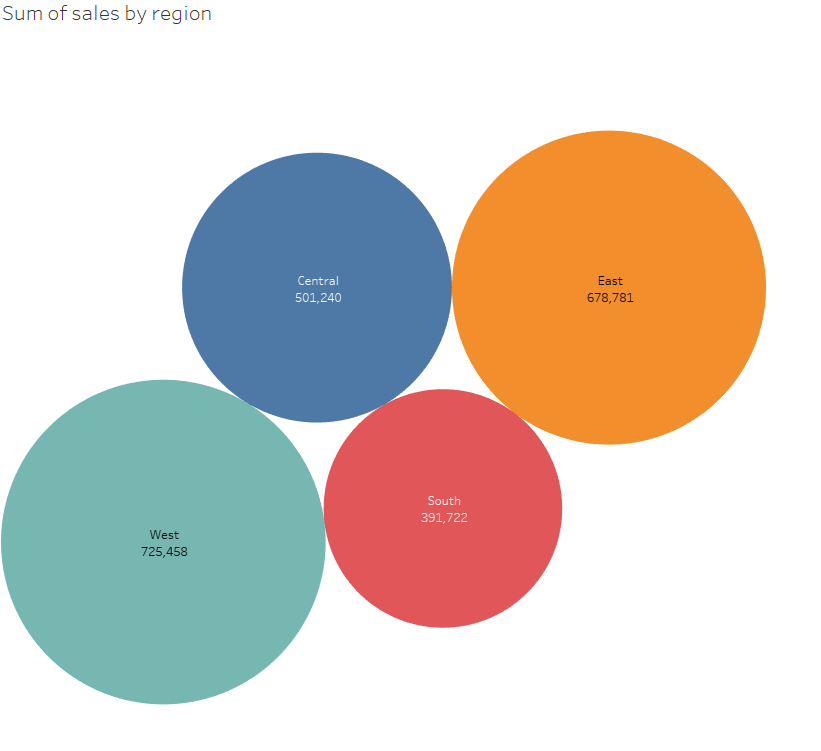
**Ans.** Below chart shows the growth of sales in percentage of different product categories over time. Sales of technology related product increased by 20% from 2016 to 2017.



**Justification :** Bar chart with marks for sales with table calculation for percentage difference shows the growth in sales, positive or negative over time accurately, hence used.

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

**Ans.** Below chart shows the sales distribution across different regions.West region has the highest sales of 725,458 and the South region has the least sales of 391,722.

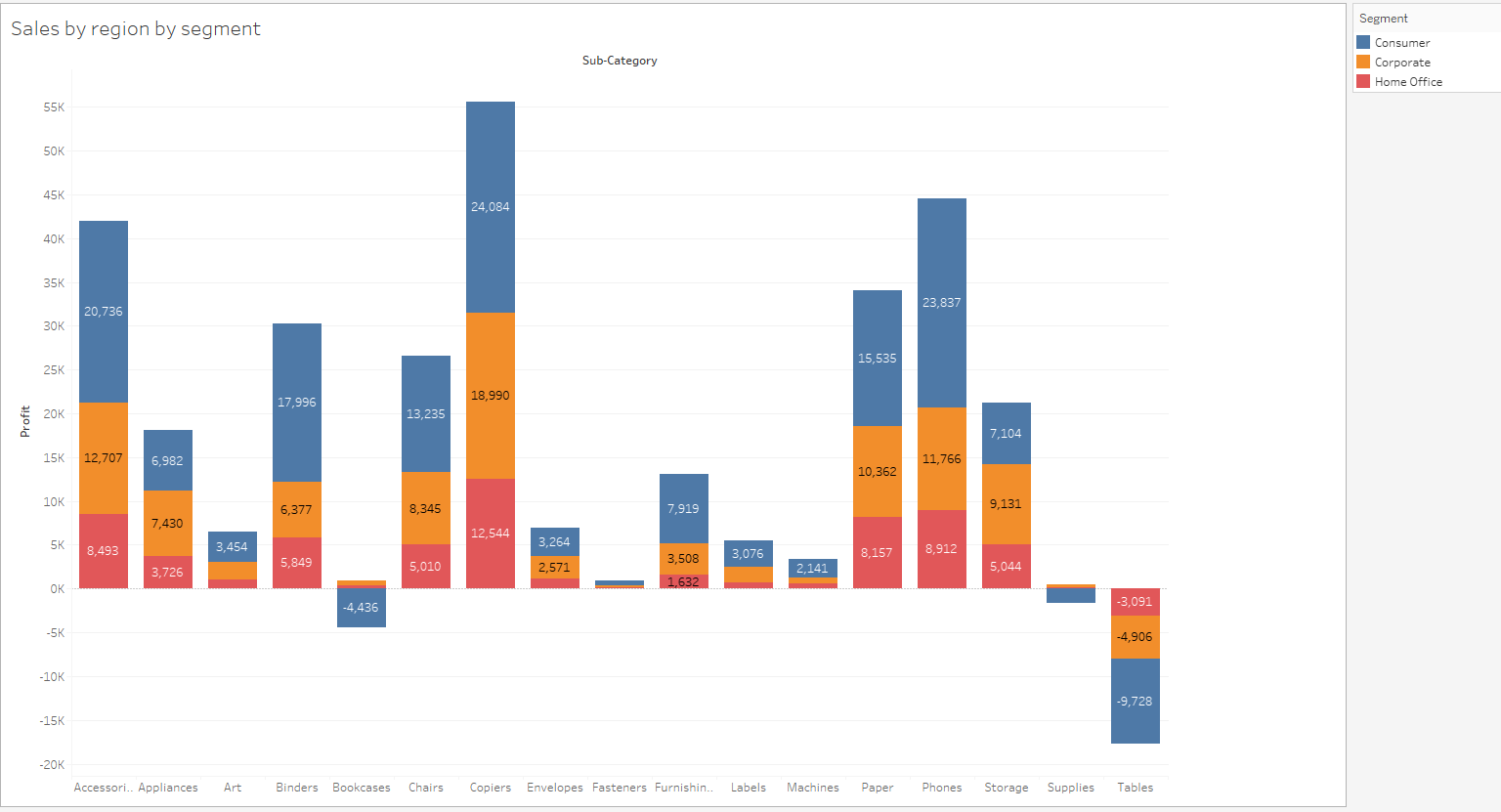


**Justification :** Bubble chart shows the sales distribution with the size of the bubbles. It is used to compare measures based on dimensions, hence used.

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

**Ans.** Below chart shows the profits across various subcategories within different customer segments. Copiers have the highest profit in the Consumer segment(24,084).

Tables have the lowest profit in the Consumer segment(-9728).

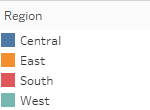
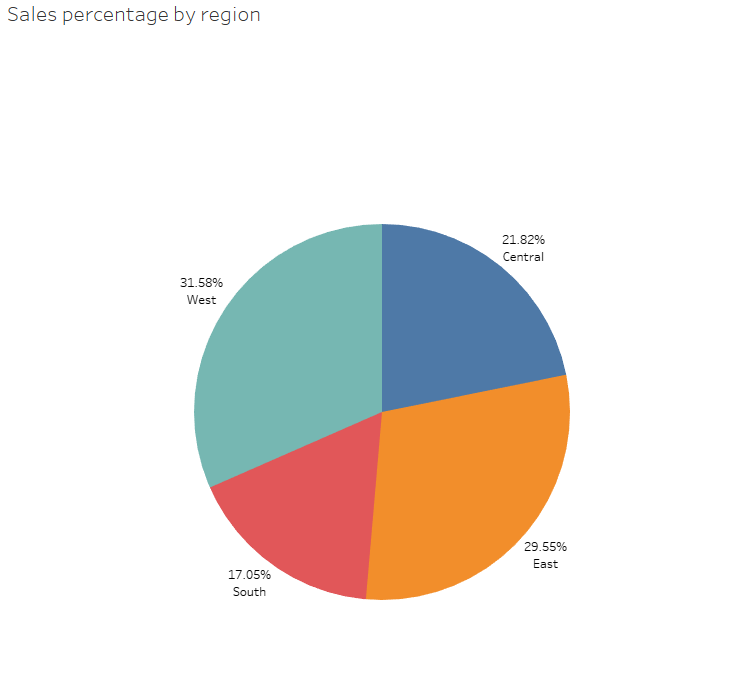


**Justification :** Stacked bubble charts are very useful in visualising profit among various segments for various sub categories.

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

**Ans.** West region has the highest sales contribution of 31.58% and South region has

the lowest sales contribution of 17.05%.



**Justification :** Pie charts are very useful for showing percentage contribution of a

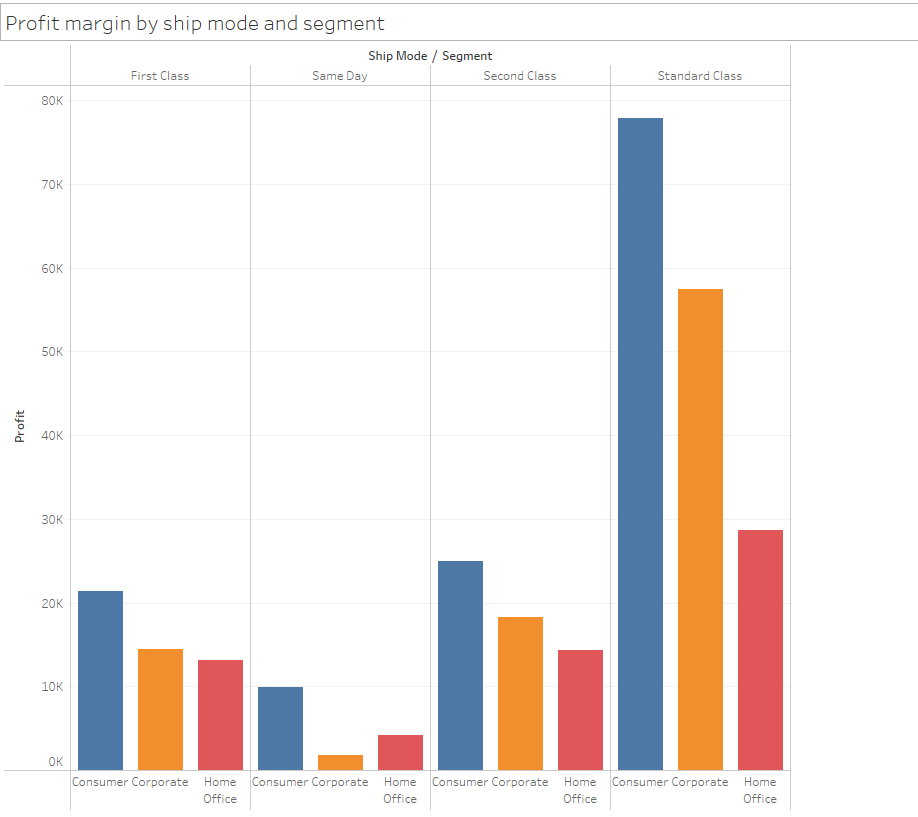
measure based on different categories, hence used.

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

**Ans.** Items for Consumer segment with Standard ship mode have the highest profit

margin, but items for Corporate segment with same day delivery have the lowest

profit margin.



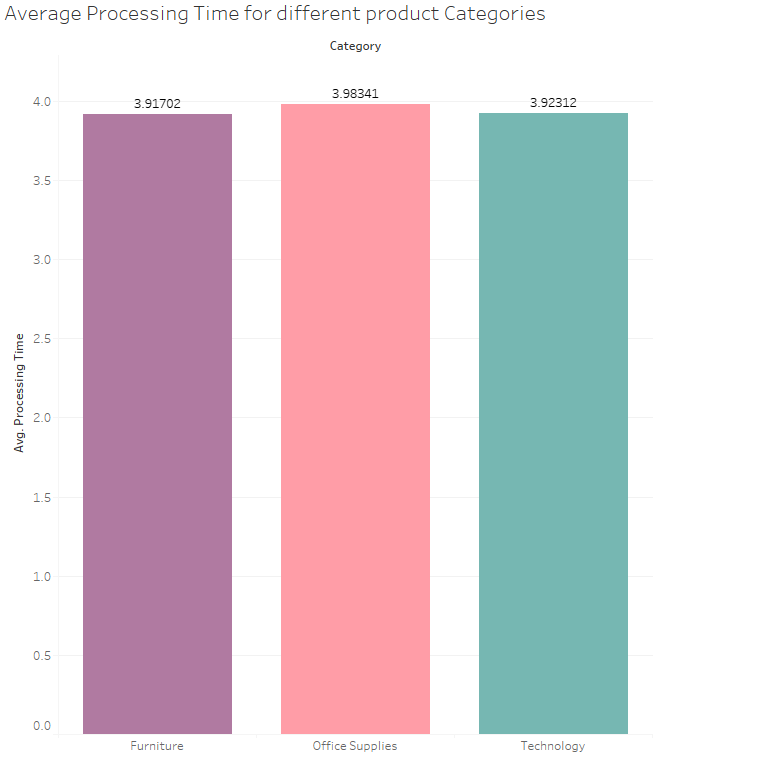
**Justification :** Column chart with different colour showing different segment and

with columns based on ship mode are very useful for visualising the above

requirement, hence used.

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

**Ans.** Below chart shows the average processing time for different product categories. Furniture takes the least processing time and Office Supplies take the maximum processing time.



` **Justification :** By creating a calculated field called Processing time which is the difference between Ship Date and Order Date, and creating a column graph of

Average processing time by Product Categories we can visualise the required

visualisation, hence column graph is used.

1. How do discounts affect overall profit?

**Ans.** According to the trend shown in the graph, profit is indirectly proportional to

discount, i.e., as discount increases profit decreases and as discount decreases profit

increases.



**Justification :** Scatter plot is good for seeing relationships between two discrete

continuous variables, hence scatter plot is used.

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

**Ans.** As we can see from the graph below, profitability increases with increases in

sales, but in some cases profit also increases for lower numbers of sales. Thus there

are some data points which don’t follow the normal trend.



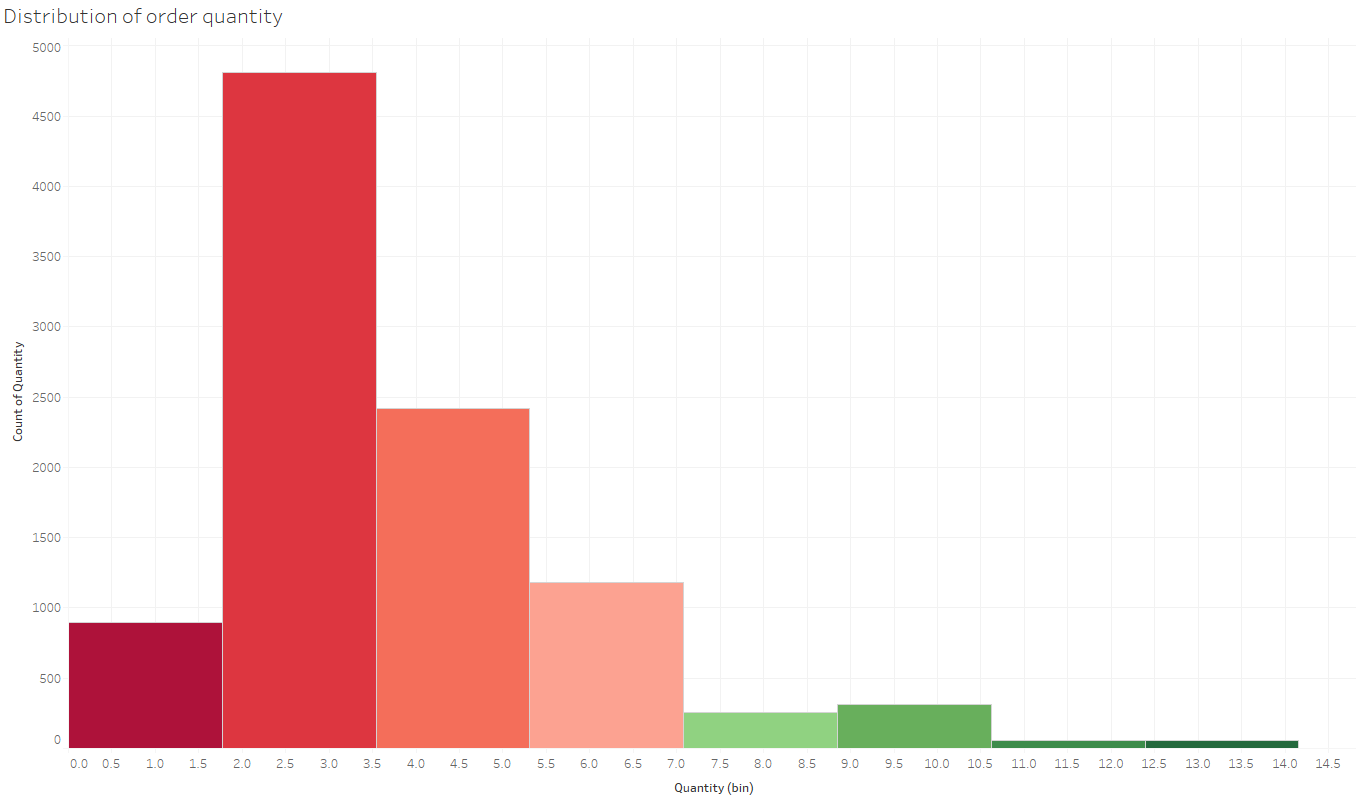
**Justification :** Line charts are very good for seeing trends of continuous variables,

hence it is used.

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

**Ans.** Below graph shows the distribution of order quantity in a histogram chart.

Most order quantities fall within 2 - 4 with very few above 10.

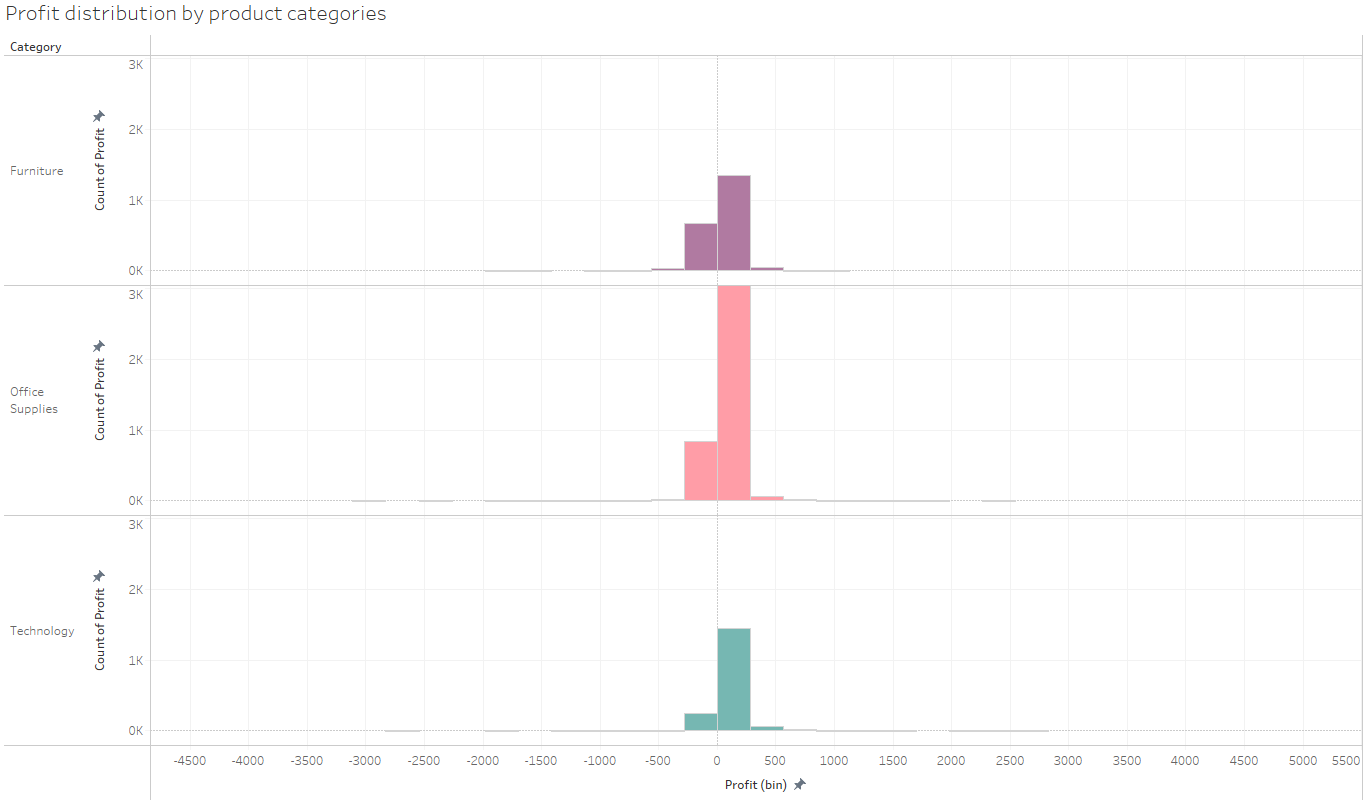


**Justification :** Histograms are very useful to show the distribution of a variable by

dividing it into bins. Hence, histogram is used.

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

**Ans.** Most profit falls within the profit bin of 0, and most occurs for Office Supplies.



**Justification :** Histograms for each category shows the profit distribution for each

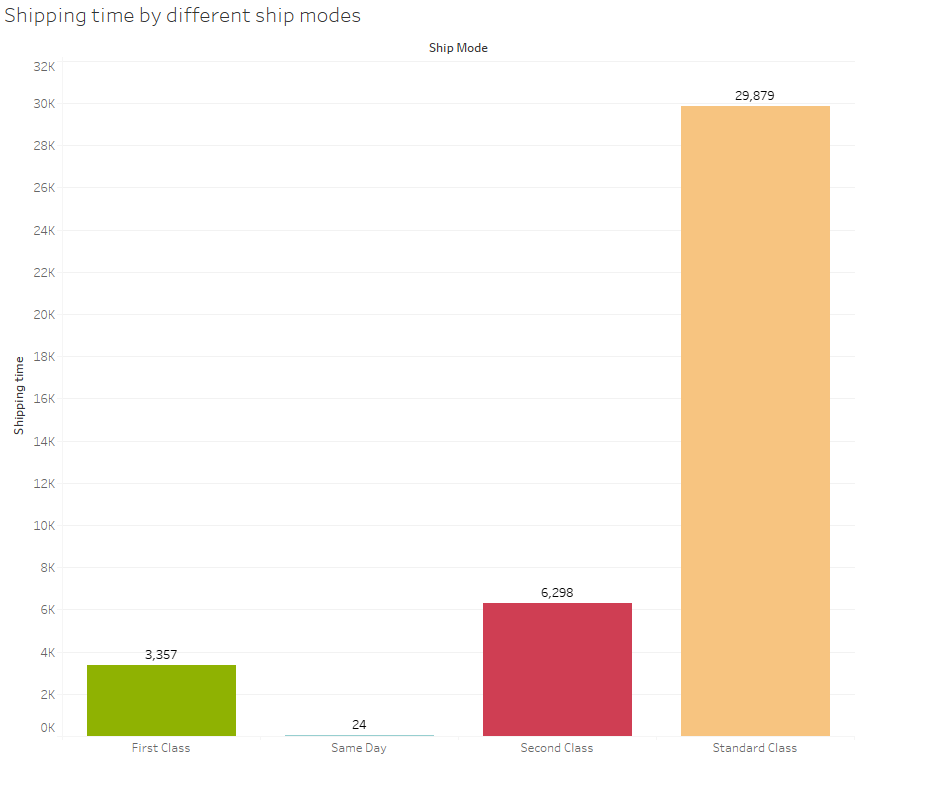
category. Hence, histogram is used.

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

**Ans.** Cumulative shipping time of standard ship mode is highest(29,879). Cumulative

Shipping time of same day shipping mode is lowest(24) as we can see from the chart

below.



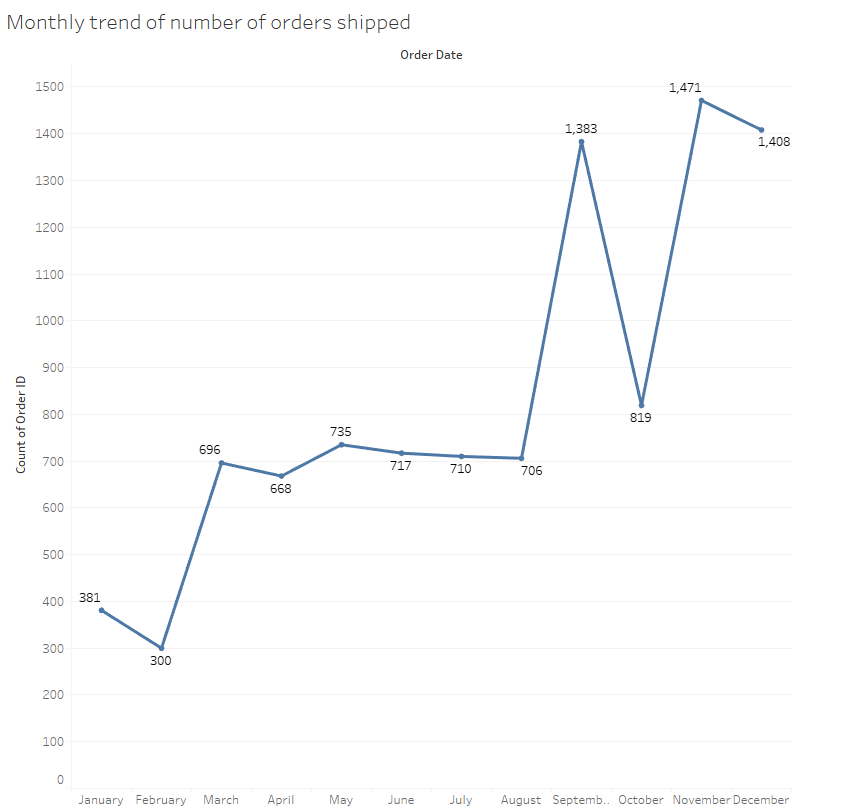
**Justification :** Column/Bar chart with created field for shipping time can help to

visualise the requirement effectively, hence used.

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

**Ans.** According to the chart below, the trend is, generally the number of orders

increase with progression of the year, with a dip in number of orders in September.

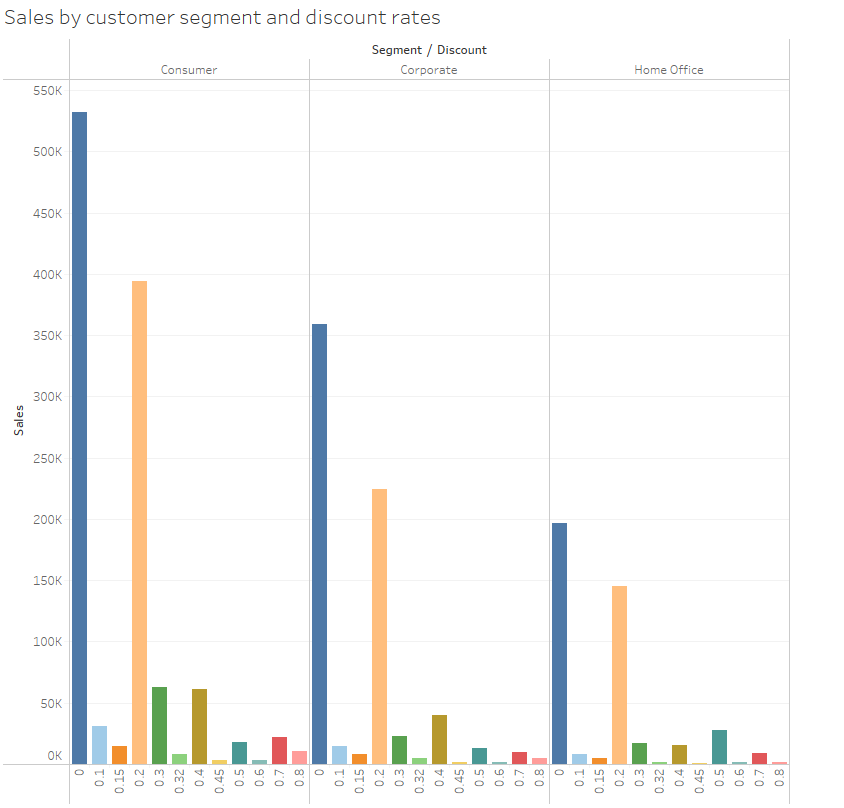


**Justification :** Line charts are very useful to see trends over time, hence used.

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

**Ans.** Below chart shows Sales by different customer segments and discount rates,

Most sales occur with a discount rate of 0 for all segments.



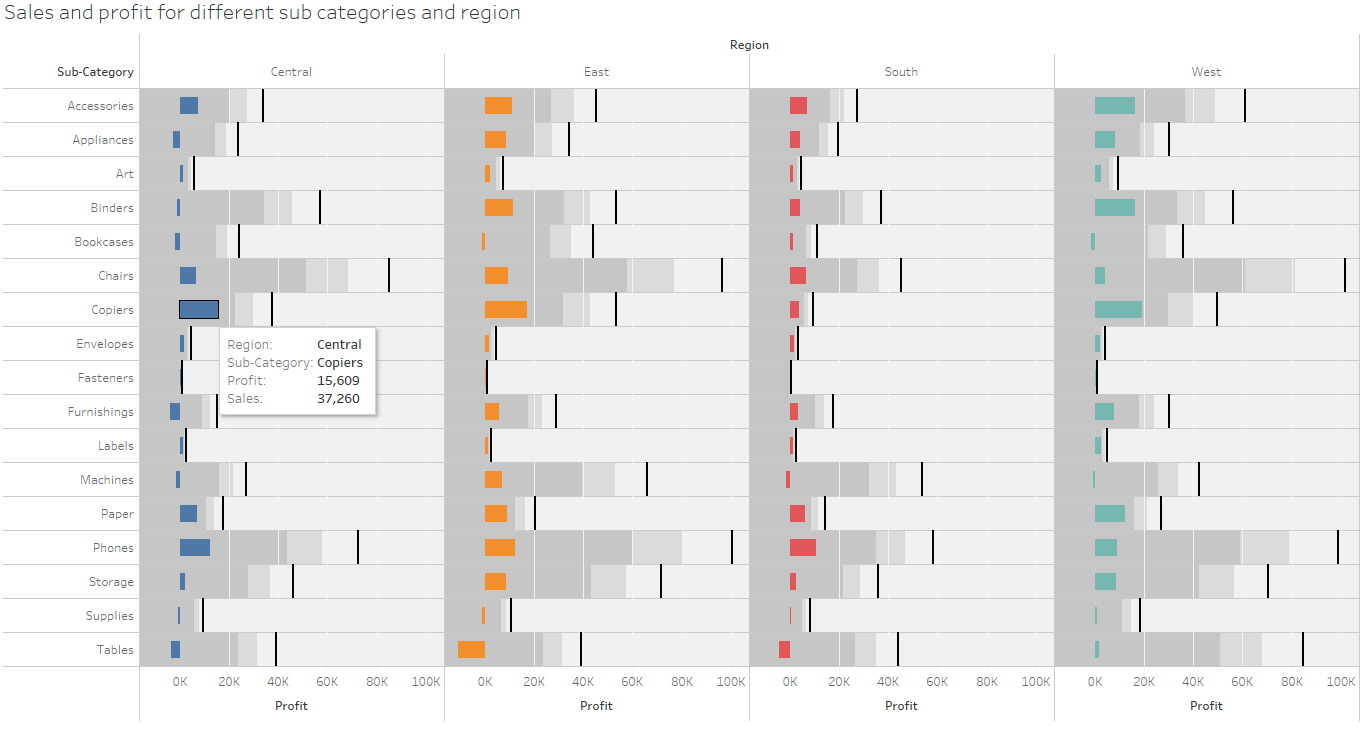
**Justification :** Column charts are very useful for showing comparison between

different categories, hence used.

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

**Ans.** Below chart shows the sales and profit for different sub-categories and regions.

You can see the value of the sales and profit by hovering over the bars.



**Justification :** Bar charts with separated column for regions show the required

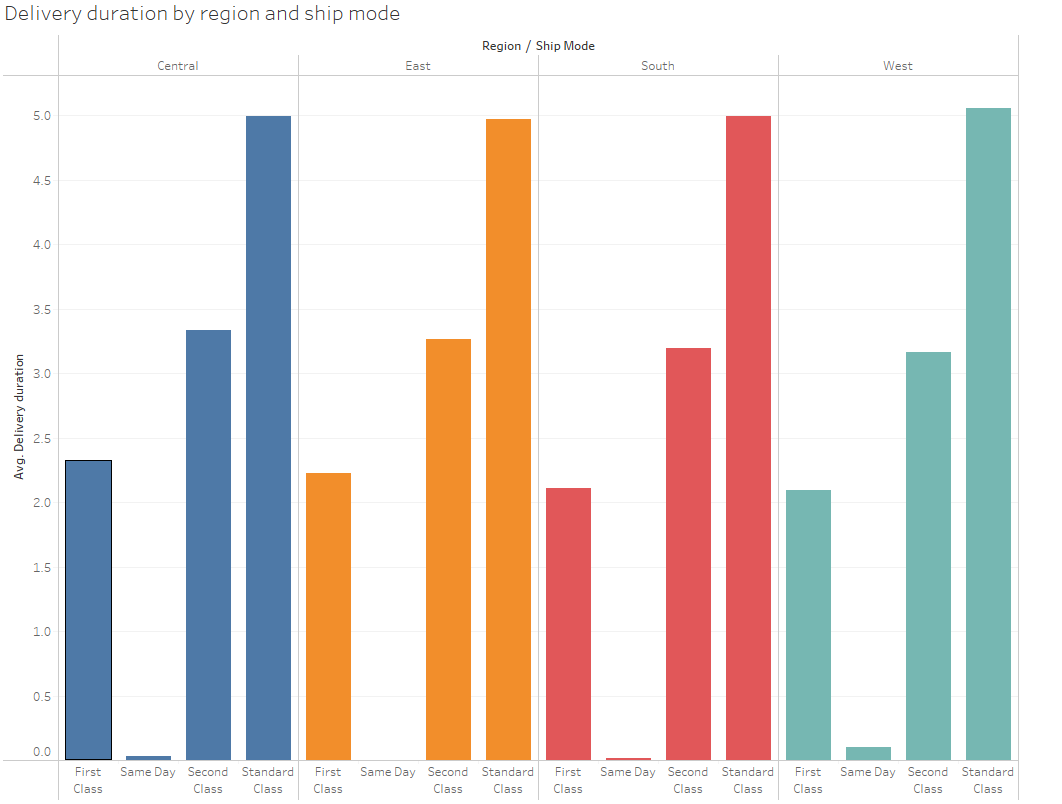
Visualisation effectively, hence used.

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

**Ans.** The chart below shows average delivery duration for different regions and

ship modes. Average delivery duration of standard ship mode is highest among all

regions and average delivery duration of same day ship mode is lowest for all regions.



**Justification :** Bar/Column charts are very useful for comparing values among

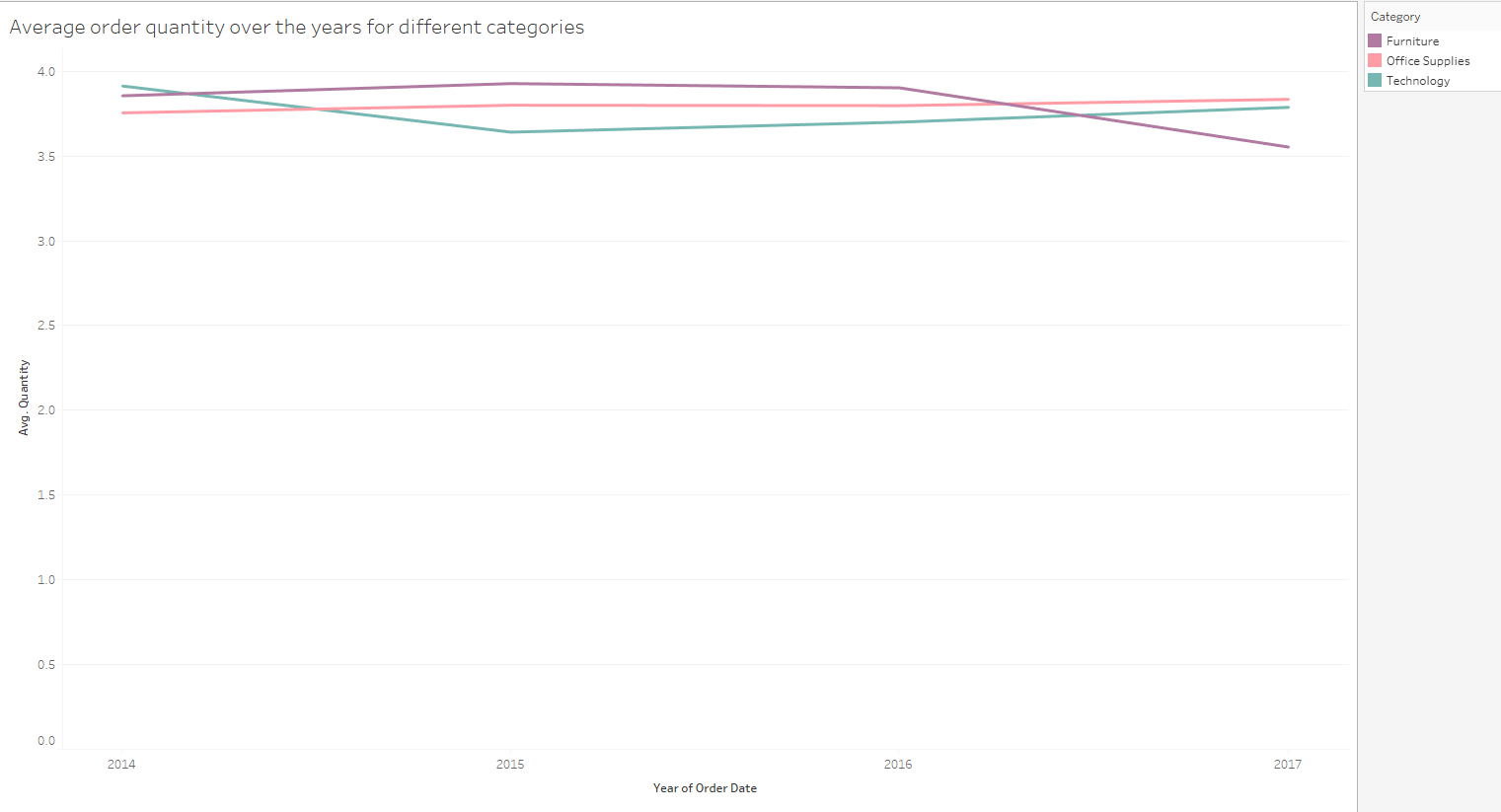
different categories, hence used.

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

**Ans.** Average order quantity has been between 3.5 to 4 for all product categories,

over the years. Average order quantity for furniture category has dipped below 3.5

in the past two years.

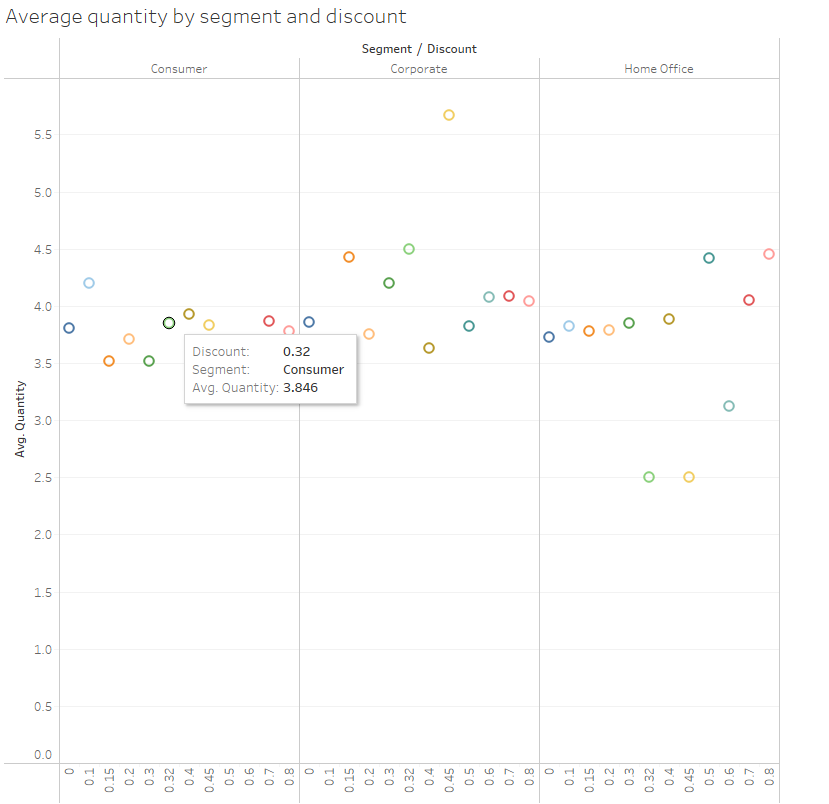


**Justification :** Line charts are very useful for seeing trends over time, hence used.

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

**Ans.** Below chart shows the correlation between average quantity ordered

discount rates for different customer segments.

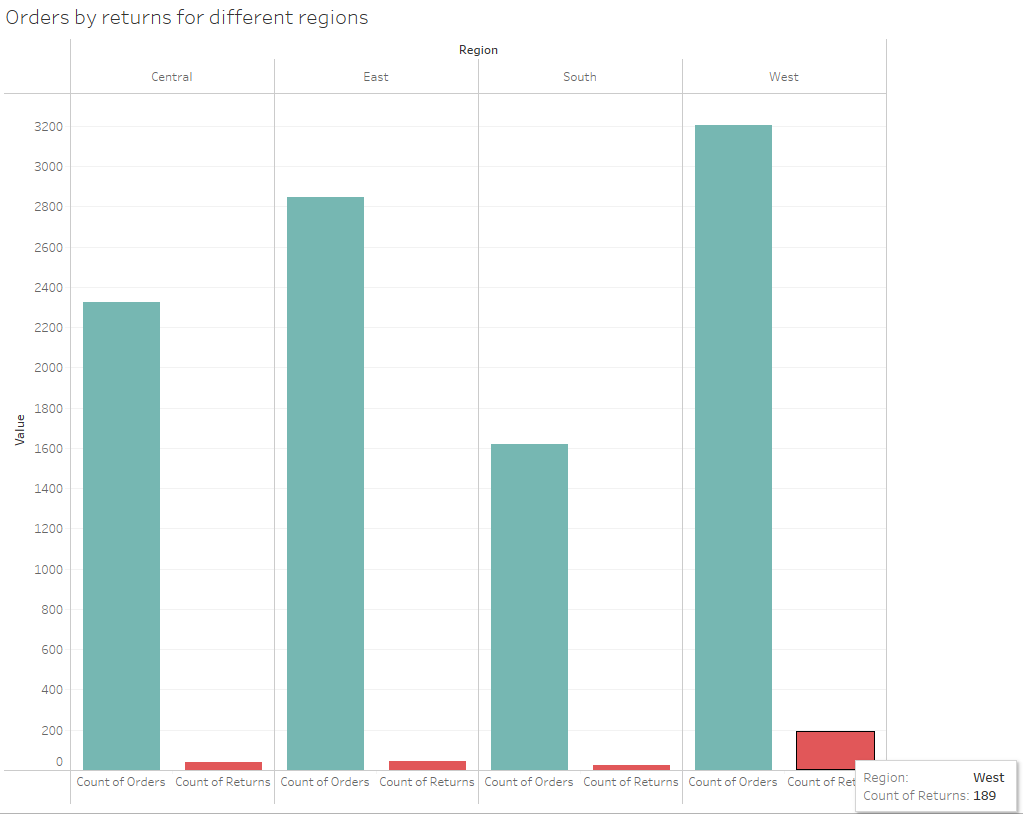


**Justification :** Scatter plots are very good for understanding correlation, hence used.

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

**Ans.** Below chart shows the proportion of orders and returns in each region.

Most returns take place in the West region.

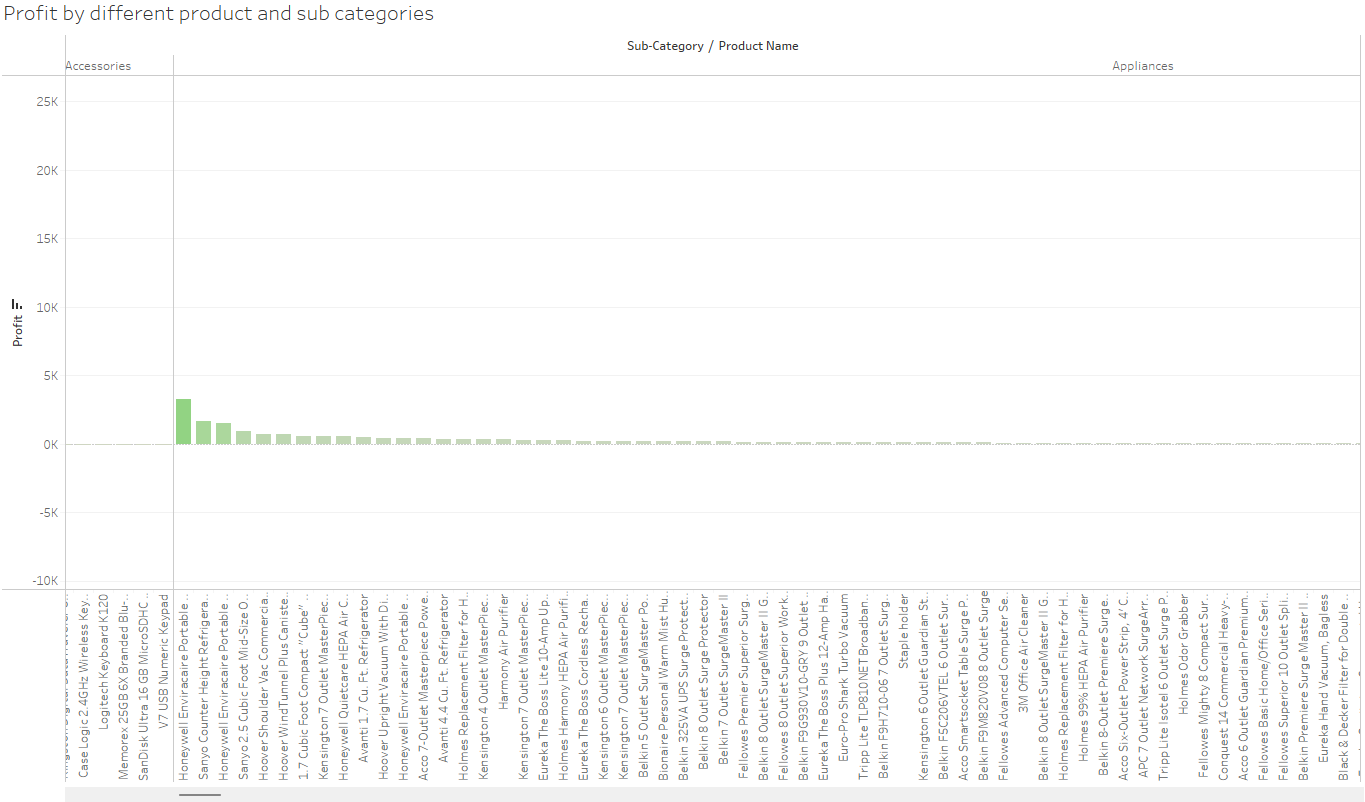


**Justification :** Side by side Bar/Column chart are very useful for comparing quantities based on different categories, hence used.

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

**Ans.** Below chart shows the profit of different products for different subcategories.

Honeywell Portable Air Cleaner has the most profit in Appliances subcategory.

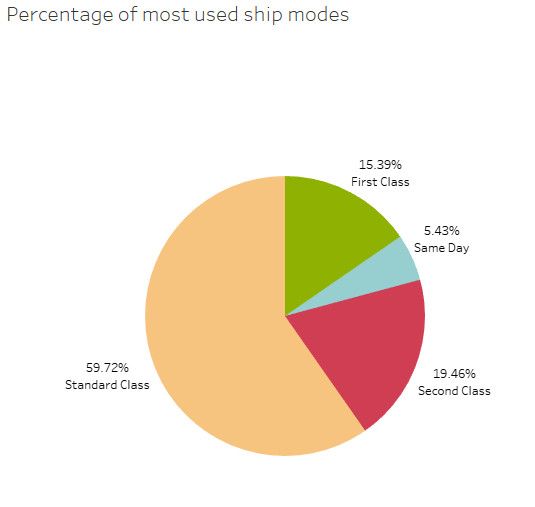


**Justification :** Bar/Column charts are very useful for comparing quantities with different categories, hence used.

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

**Ans.** Standard Class ship mode is most commonly used with a share of 59.72%.

Same Day ship mode is least commonly used with a share of 5.43%.



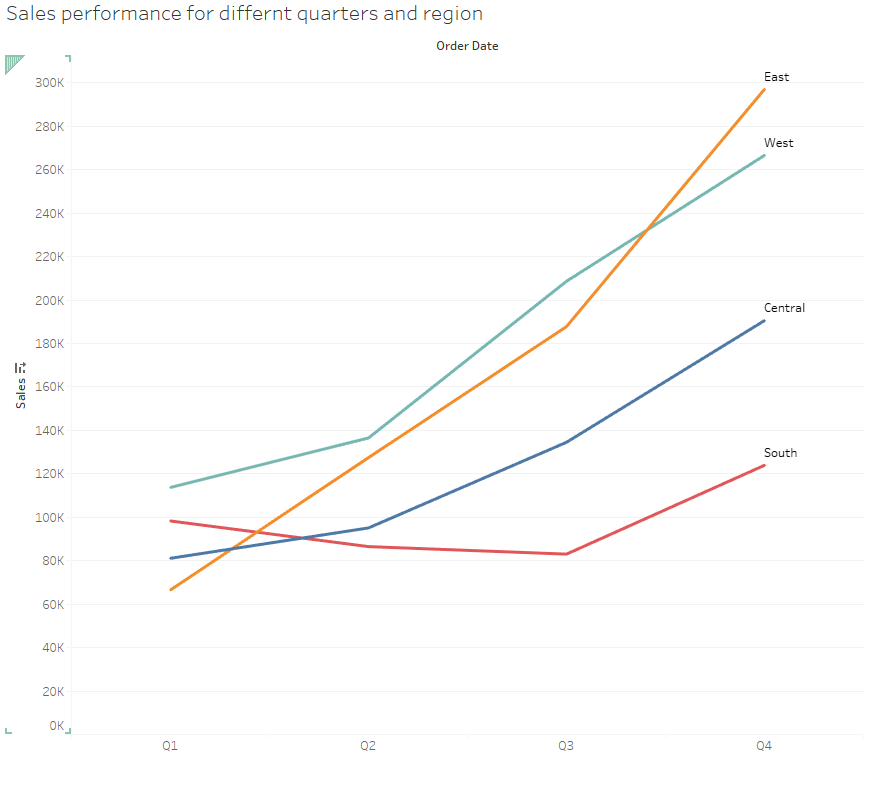
**Justification :** Pie charts are very useful in seeing percentage shares, hence used.

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

**Ans.** Below chart shows the sales performance for different regions for different

quarters of the year. East region has the most sales performance while South region

has the least.

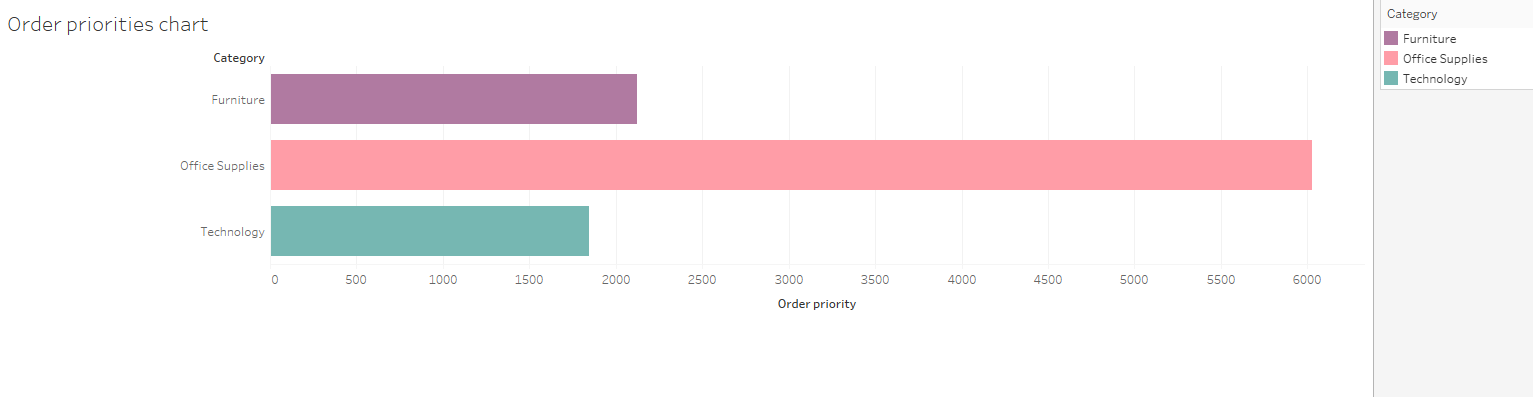


**Justification :** Line charts are very useful for seeing trends over time, hence used.

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

**Ans.** The chart below shows the distribution of order priorities. Office Supplies

category has the highest priority, while Technology has the lowest priority.



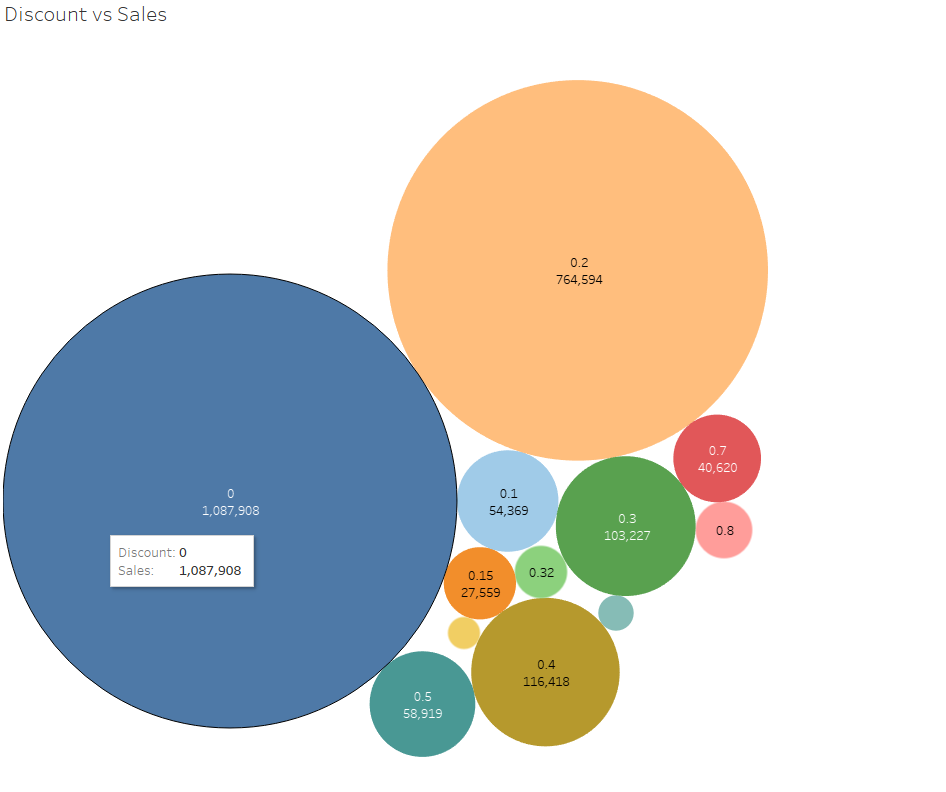
**Justification :** Bar/Column chart are very useful for comparing values within

different categories, hence used.

1. What is the relationship between discounts and sales?

**Ans.** Most sales occur with lower discount rates 0 - 0.2 while for higher discount rates

sales value is less.



**Justification :** Bubble charts are very useful for quantity comparison based on size,

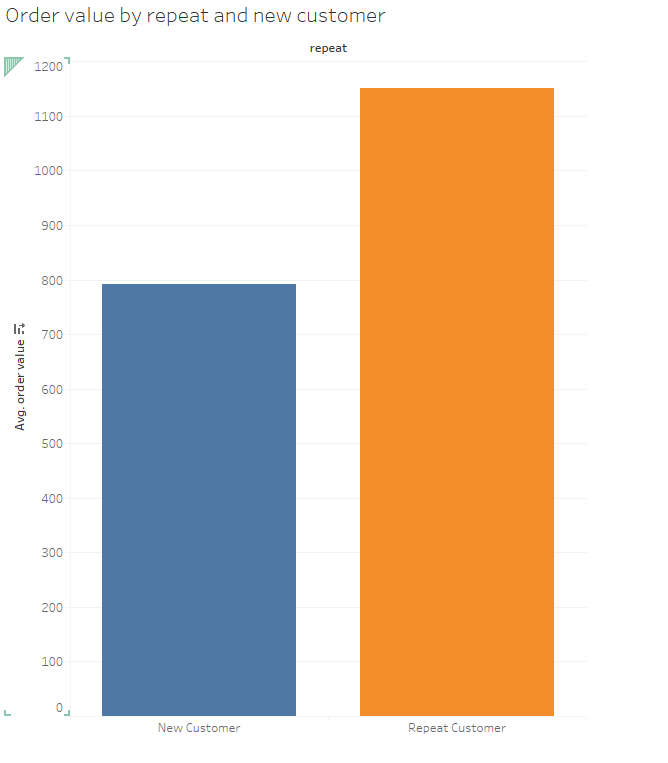
hence used.

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

**Ans.** Repeat customers have greater average order value than new customers.

Average order value for repeat customers are 1150.5 while average order value for

new customers are 792.0.



**Justification :** Bar/Column charts are very useful for comparing value across

different categories, hence used.

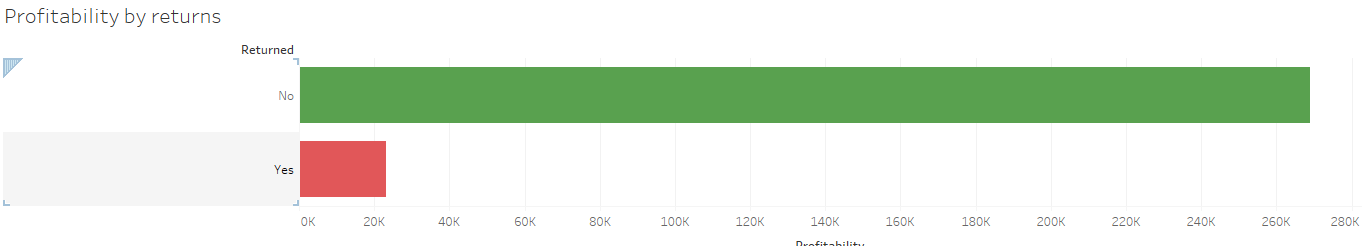
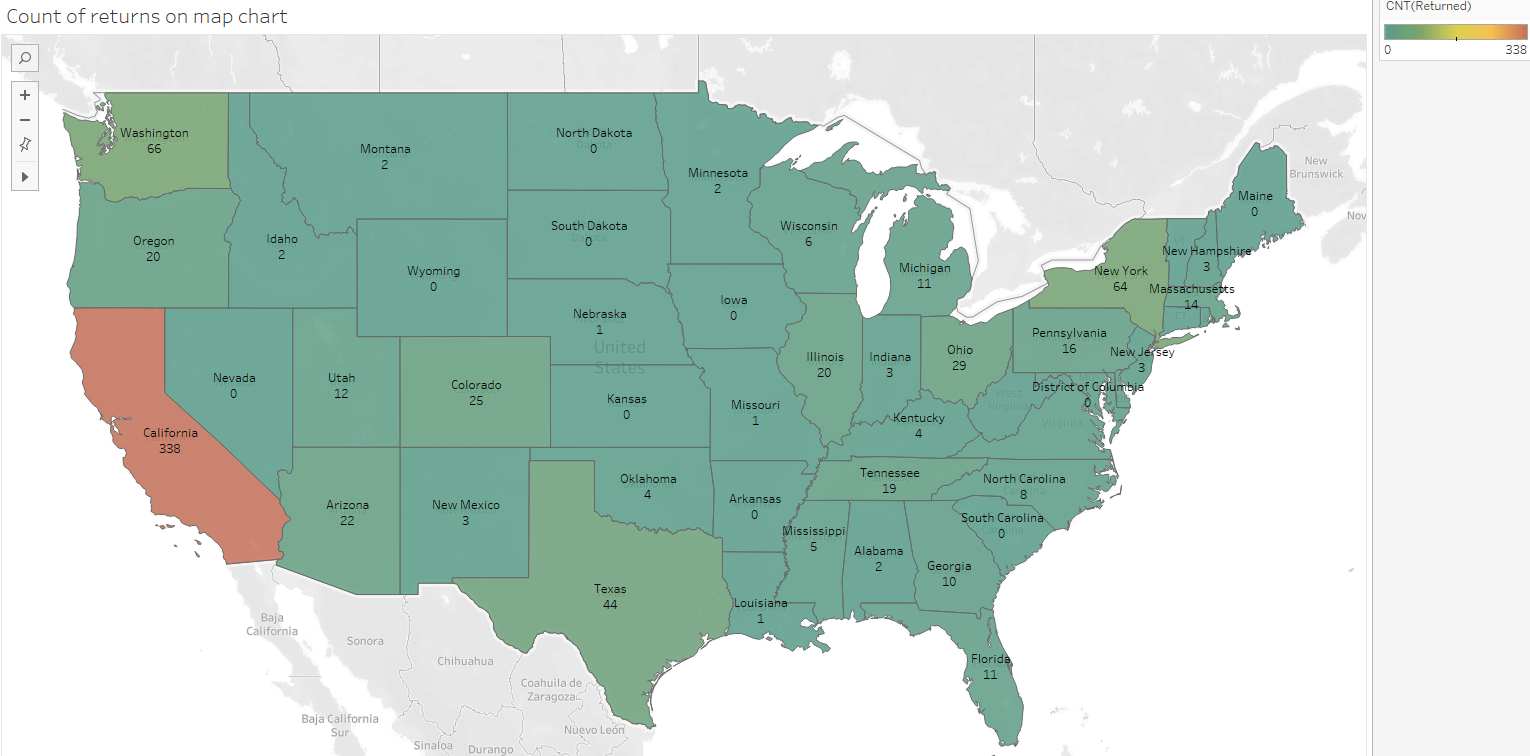
1. What is the geographical distribution of returns and its impact on overall profitability?

**Ans.** Below chart shows the geographical distribution of returns and the second chart

shows the impact of returns on profitability. California has the highest number of

returns (338). Maine, Nevada, Arkansas, Wyoming and Iowa has 0 returns, hence

boosting profitability.



**Justification :** Map charts are very useful for showing geographic distribution, hence

used.