

Guidelines for Data Visualization and Analysis Project

About the Project:

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

Skills Required:

- Proficiency in data visualisation concepts and techniques.
- Familiarity with Tableau or a similar data visualisation tool.
- Strong analytical and problem-solving skills.
- Ability to choose appropriate charts based on data characteristics and question requirements.
- Clear and concise communication skills.

Deliverables:

- A Google document containing solutions to the scenario based questions including the screenshot of relevant chart picked for each scenario, presented in a concise and well-structured format. Make sure to provide explanations that highlight your problem-solving skills.

Rubrics for Assessment:

Question Responses:

- Accuracy and completeness of answers for all 30 questions.
- Clear and concise explanations that address the question's context.

Chart Selection and Explanation:

- Thoughtful rationale for choosing specific chart types.
- Justification based on data characteristics, context, and communication goals.

Creative Enhancements:

- Effective use of creative elements to enhance visualisation quality.
- Enhancements that contribute to better understanding or engagement.

Note:

- Duplicate this document and proceed to write your solutions.
- For each scenario and question, provide a justification for the choice of chart type. Explain why it is the best option to visualise the data effectively.
- Attach screenshots of the charts you have created in Tableau for each scenario and question using the Superstore dataset. Label them clearly to match the corresponding questions in the Google Document.
- Submit the duplicated google doc file after completion.

Use these guidelines to structure your data visualisation and analysis project. Remember to maintain consistency in your responses, explanations, and visualisation styles. This project will not only demonstrate your skills but also your ability to effectively communicate complex information through visualisations. Good luck!

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

Imagine you are a data enthusiast aiming to excel in data visualisation and analysis. In this task, you have been given any 30 scenario-based questions derived from the Superstore dataset, and your objective is to provide insightful answers using appropriate charts. For each question, you need to select a chart that best represents the data, explain why you chose that specific chart, and then proceed to build the chosen chart using Tableau.

Your responses should be succinct, organised, and illustrative of your problem-solving capabilities.

Dataset Link:

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

Please keep in mind:

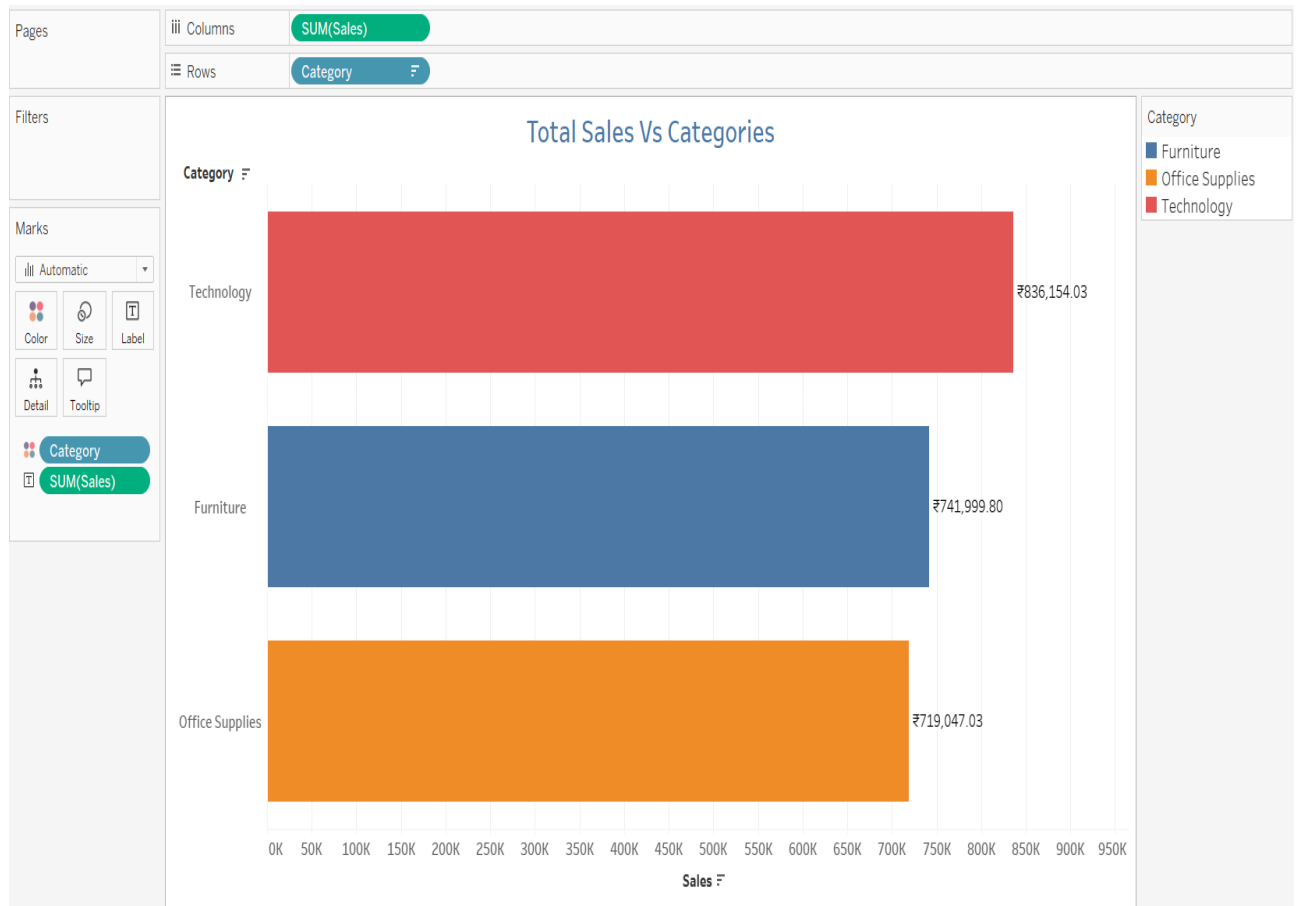
1. **Answer Completion:** Ensure that you furnish answers for all any 30 questions and build charts for them.
2. **Encouraged Creativity:** Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

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

Good luck!

Questions:

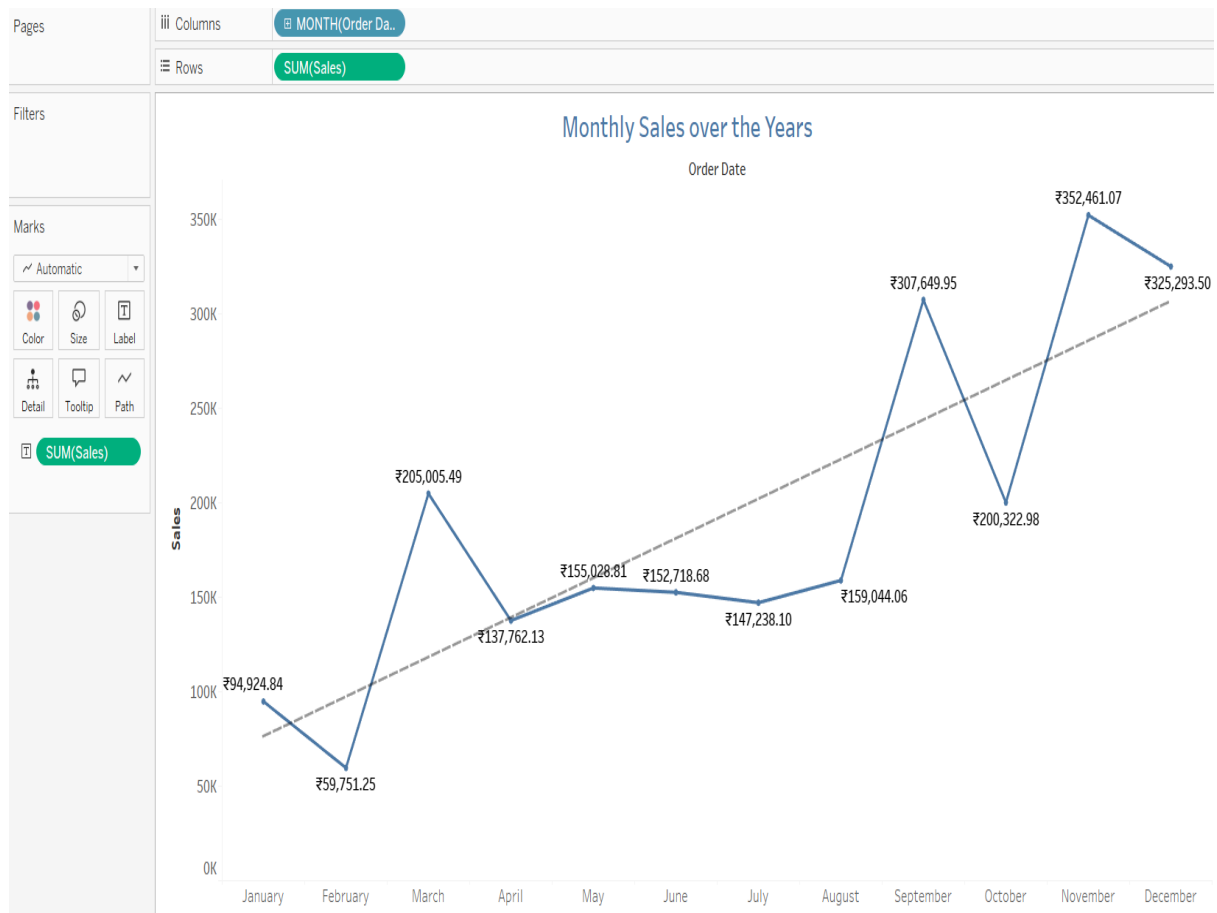
1. Which product categories have the highest total sales in the "Superstore" dataset?



Reason for Chart Selection:- A horizontal bar chart visually compares the size or value of different categories, making it easy to see which categories are larger or smaller. It also makes large amounts of data easier to read and interpret using bars of varying lengths. It effectively highlights differences between categories, making it clear where significant variations exist.

Observations:- We can observe above in the chart that Technology category products have the highest total sales and Office Supplies have the lowest sales.

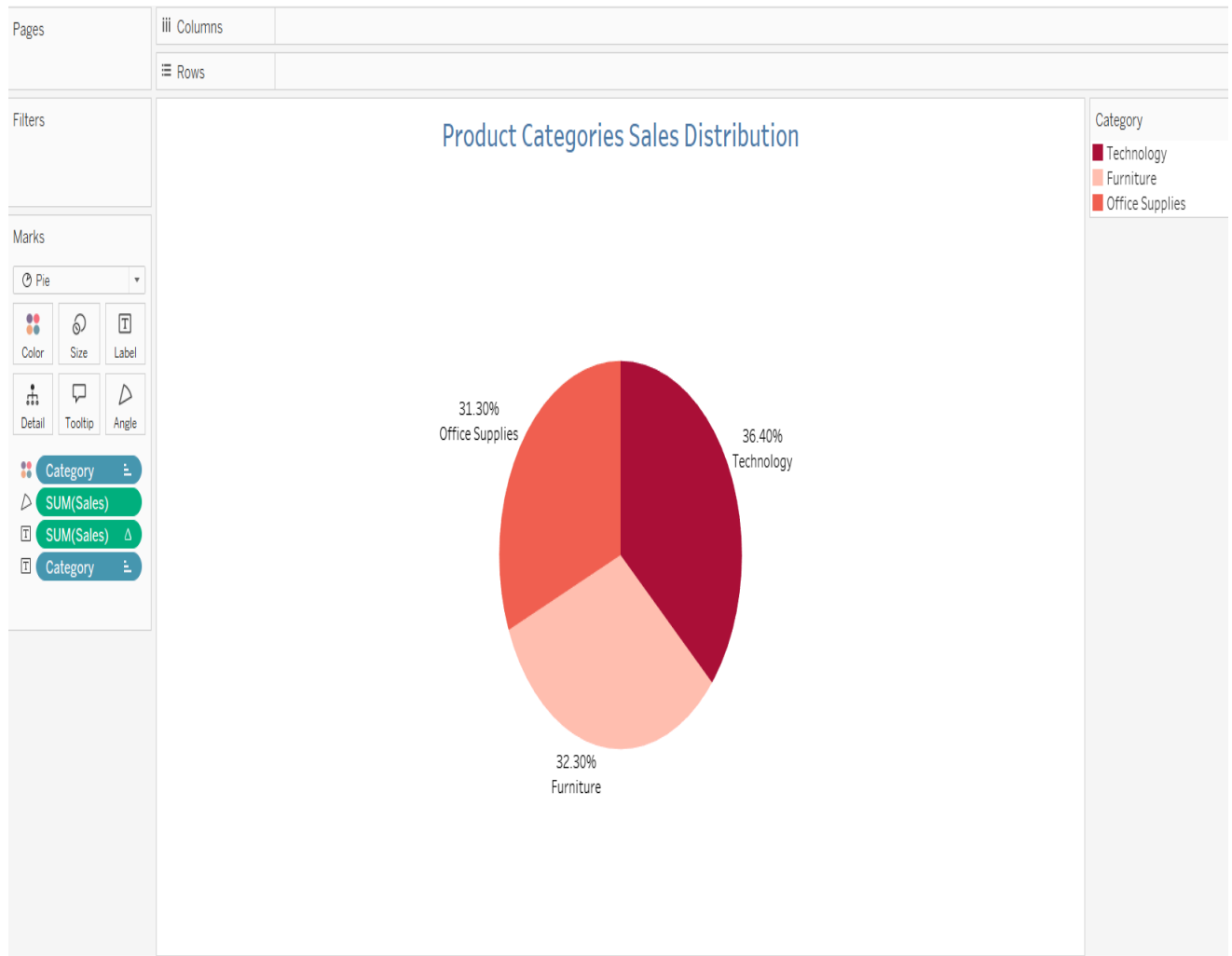
2. How do the monthly sales amounts change over the course of a year?



Reason for Chart Selection:- A line chart visually displays trends and changes over time, making it easy to track how values progress or fluctuate. It connects data points with a continuous line, highlighting patterns and trends clearly. This format is effective for showing relationships and trends across time periods, making it ideal for analyzing data that evolves over a sequence. The line chart simplifies the comparison of data points and reveals long-term trends and variations.

Observations:- We can observe above in the chart that over the years monthly sales have been on an increasing trend and the highest sales were done in November month and least in February month. Also, we can observe here that there was a dip in sales from April to August.

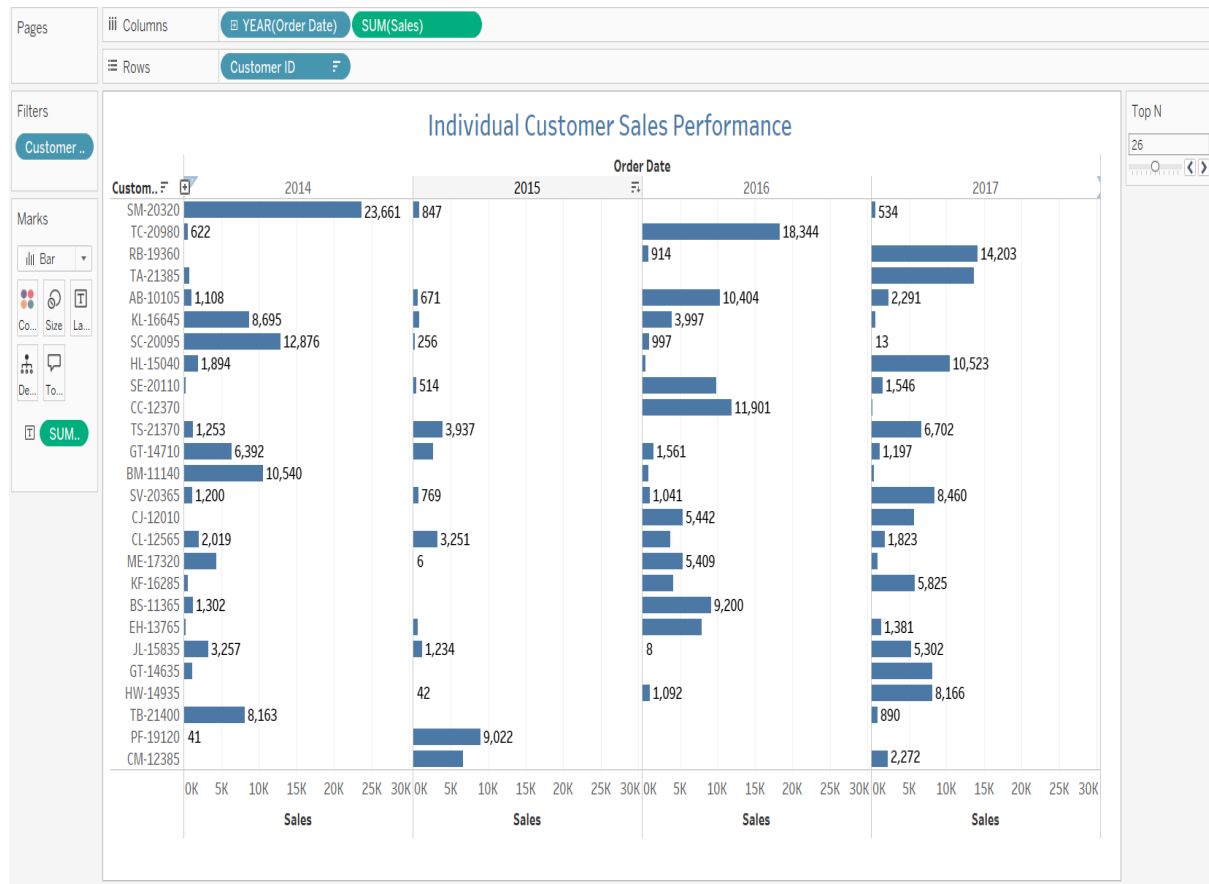
3. How is the total sales amount distributed among different product categories?



Reason for Chart Selection: A pie chart visually represents the proportions of different categories within a whole, making it easy to see the relative sizes of each segment. By displaying data as slices of a circle, it highlights the contribution of each category to the total, helping to quickly grasp the distribution and share of each part. This format is effective for illustrating percentages and part-to-whole relationships, making it ideal for comparing how individual components contribute to the overall dataset.

Observations:- We can observe in the above chart that among the categories contribution, Technology is the highest at 36.40% in terms of sales and Office Supplies second with 31.40% contribution.

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



Reason for Chart Selection:- A horizontal bar visually compares the size or value of different categories, making it easy to see which categories are larger or smaller. It also makes large amounts of data easier to read and interpret using bars of varying lengths.

It effectively highlights differences between categories, making it clear where significant variations exist.

Observations:- We can observe in the above chart the individual sales performance of the customer over the time and using that Top N parameter, we can select the top customers with sales performance dynamically.

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

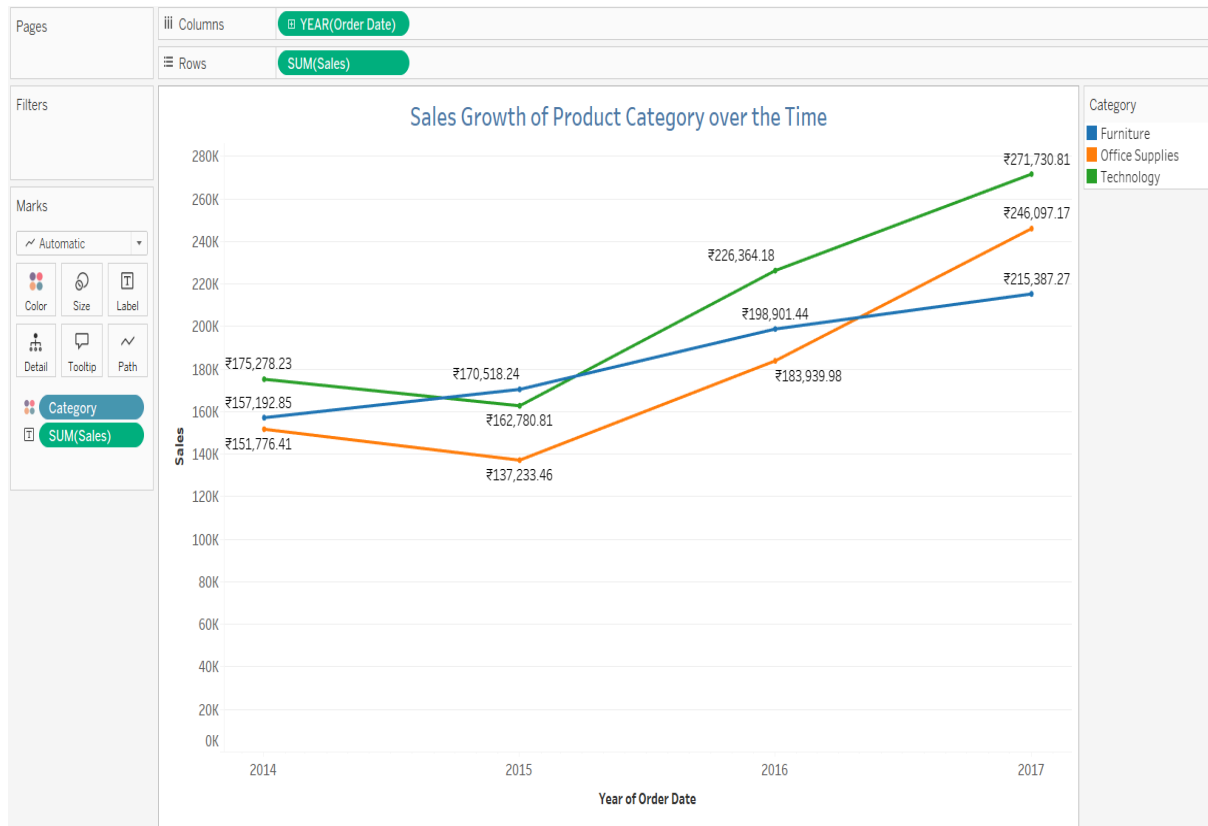


Reason for Chart Selection: A stacked bar chart visually compares the size or value of different categories while also breaking down each category into subcategories. This format makes it easy to see the total value of each main category as well as the distribution of subcategories within it. It effectively highlights the composition and relative contribution of each segment to the overall total, allowing for clear comparison and analysis of both individual and cumulative data.

Observations:- We can observe above that maximum sales are done on Mondays and Fridays, however, the Furniture category has the maximum sales on Mondays and the Technology category has the maximum sales on Fridays.

Also, we can see that on Wednesday the least number of sales are made among categories.

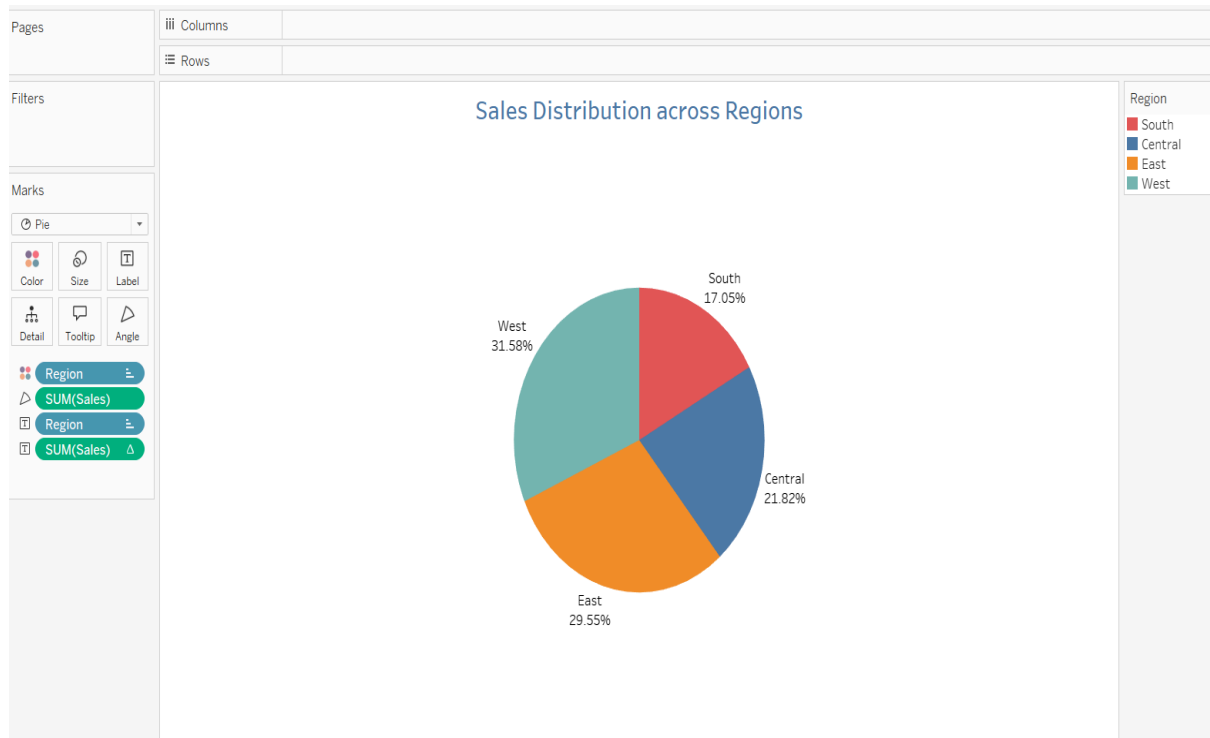
6. Can we visualise the sales growth of different product categories over time?



Reason for Chart Selection: A line chart with multiple lines differentiates categories by using distinct lines for each category, making it easy to track and compare the trends and changes over time for each category. This format connects data points with continuous lines, clearly showing patterns, trends, and variations for each category. It effectively highlights how different categories evolve over time, facilitating the comparison of their respective trajectories and long-term trends.

Observations:- We can observe in the chat above that there was a dip in sales for Office supplies and the Technology category, however good improvement trend can be observed post that dip. For Furniture, a gradual increase in sales can be observed.

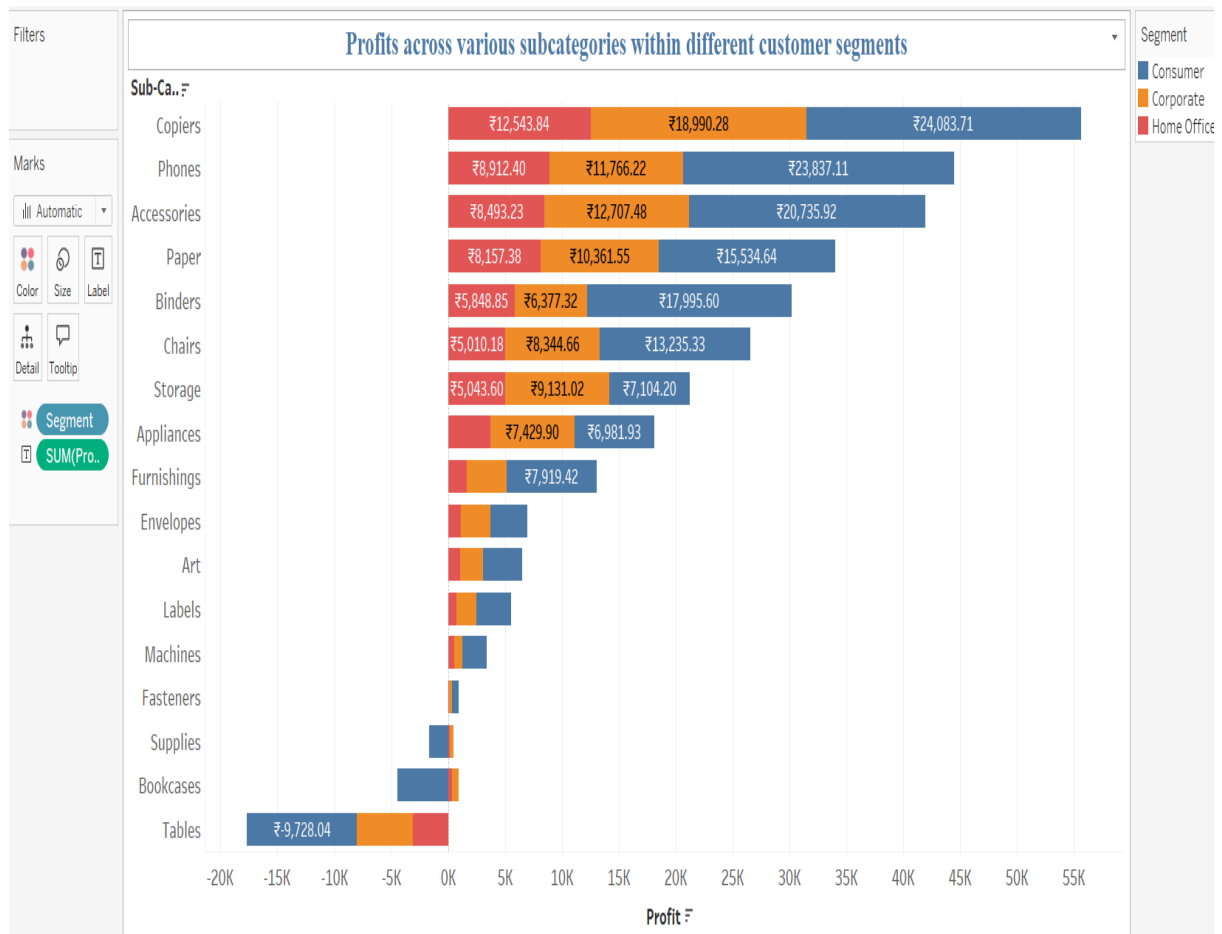
7. How does the sales distribution vary across different regions in the "Superstore" dataset?



Reason for Chart Selection: A pie chart visually represents the proportions of different categories within a whole, making it easy to see the relative sizes of each segment. By displaying data as slices of a circle, it highlights the contribution of each category to the total, helping to quickly grasp the distribution and share of each part. This format is effective for illustrating percentages and part-to-whole relationships, making it ideal for comparing how individual components contribute to the overall dataset.

Observations:- It can be observed that the West Region is leading in sales across different regions and the South Region is last in sales.

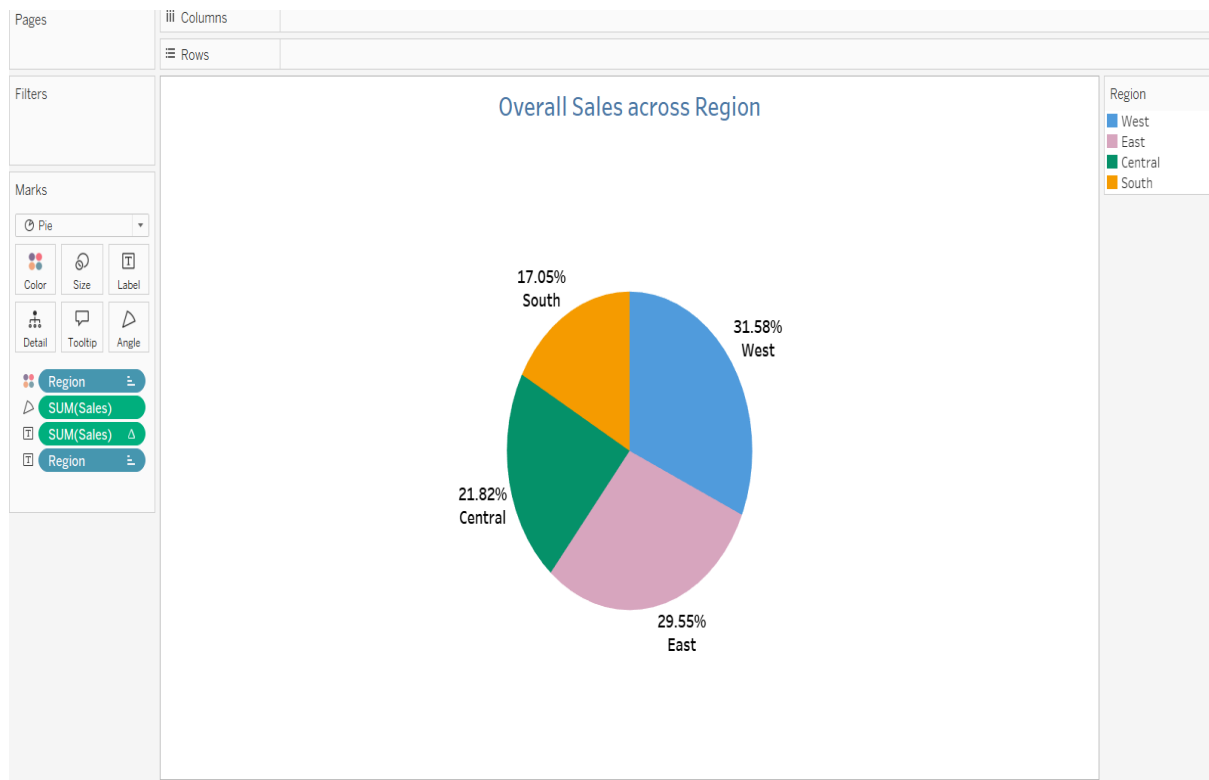
8. Can we visualise the composition of profits across various subcategories within different customer segments?



Reason for Chart Selection: A horizontal stacked bar chart visually compares the size or value of different categories while breaking down each category into subcategories. This format, with bars oriented horizontally, makes it easy to see the total value of each main category and the distribution of subcategories within it. It effectively highlights the composition and relative contribution of each segment to the overall total, facilitating clear comparison and analysis of both individual components and cumulative data. The horizontal orientation enhances readability, especially when category labels are long.

Observations:- We can observe in the above-stacked bar chart that, all three consumer segments are generating maximum profits in the copiers subcategory and the second is the Phones subcategory, however highest loss-generating subcategory is Tables across all Customer segments.

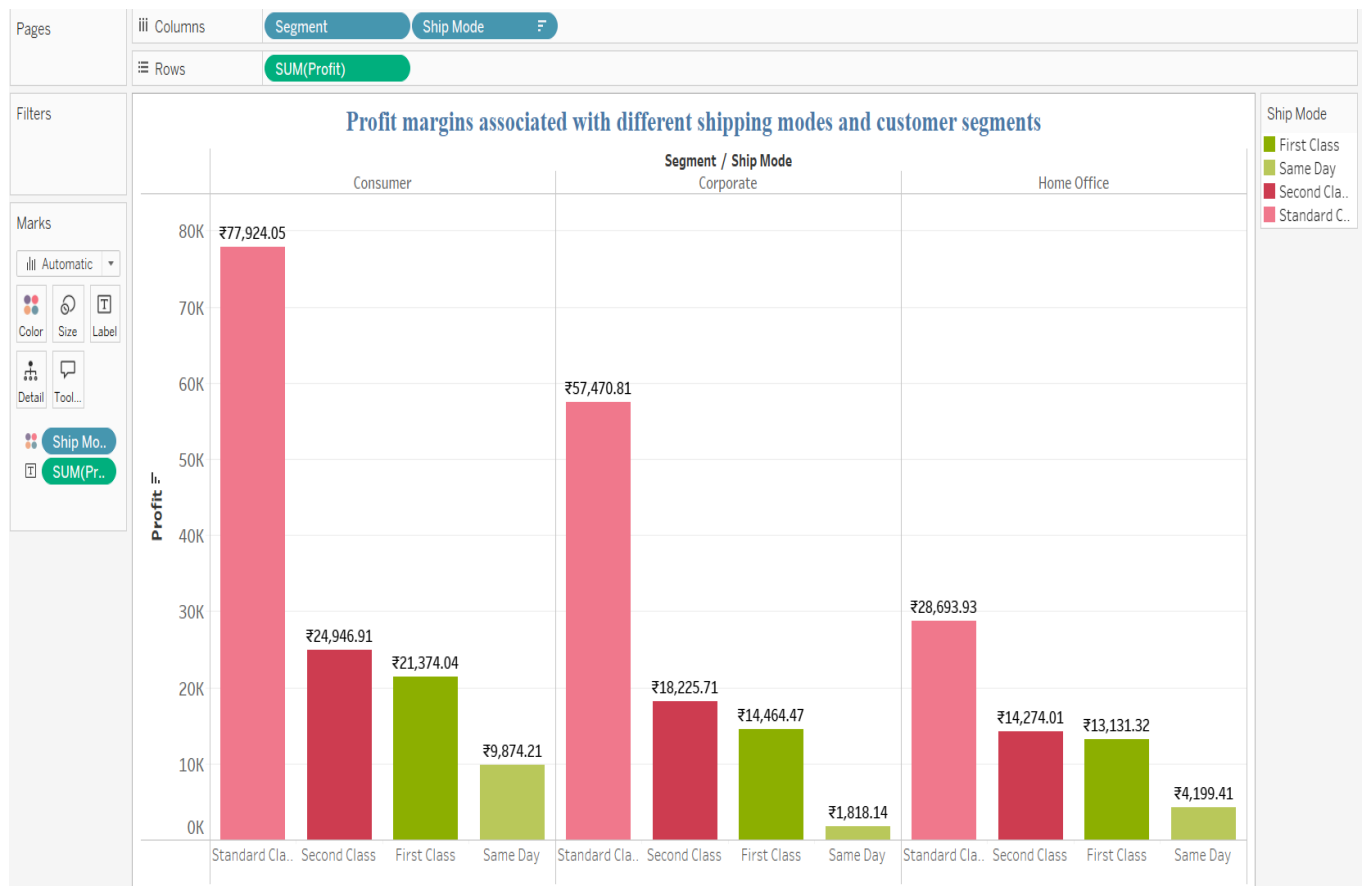
9. What is the percentage contribution of each region to the overall sales?



Reason for Chart Selection: A pie chart visually represents the proportions of different categories within a whole, making it easy to see the relative sizes of each segment. By displaying data as slices of a circle, it highlights the contribution of each category to the total, helping to quickly grasp the distribution and share of each part. This format is effective for illustrating percentages and part-to-whole relationships, making it ideal for comparing how individual components contribute to the overall dataset.

Observations:- We can observe in the above pie chart that maximum sales contributions are from the west and east regions, however south region is lagging in sales while comparing across all regions.

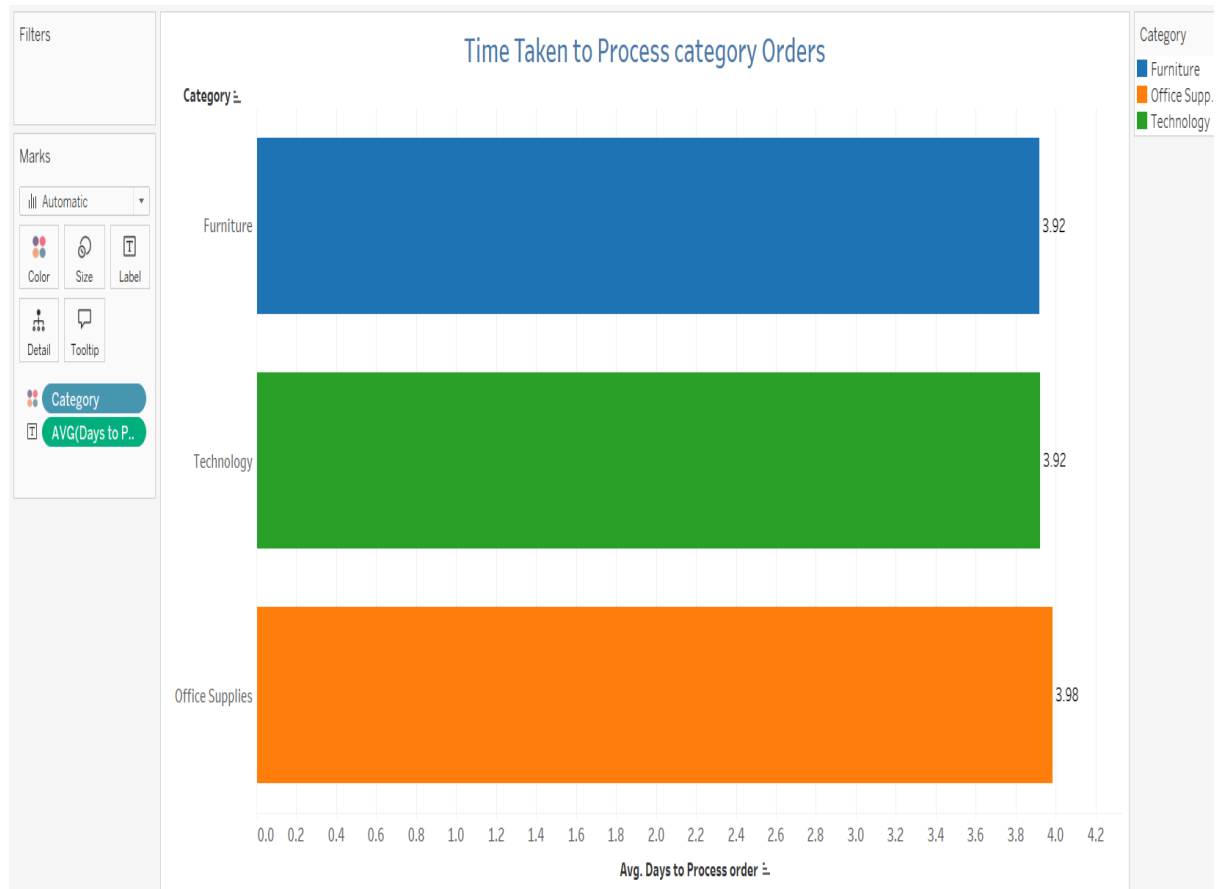
10. Can we visualise the profit margins associated with different shipping modes and customer segments?



Reason for Chart Selection: A side-by-side chart in Tableau visually compares the proportions of different categories within a whole by displaying multiple charts or bars adjacent to each other. This layout makes it easy to see and compare the relative sizes and contributions of each category across different metrics or time periods. By presenting data side-by-side, it highlights differences and trends, helping to quickly grasp the distribution and share of each part in relation to others. This format is effective for comparing multiple categories or datasets, revealing insights into part-to-whole relationships and relative performance.

Observations:- We can observe the above side-by-side chart and conclude that the Standard Class shipping mode is generating more profits across the customer segment and the consumer segment across all shipping modes is generating the highest revenue for the superstore.

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

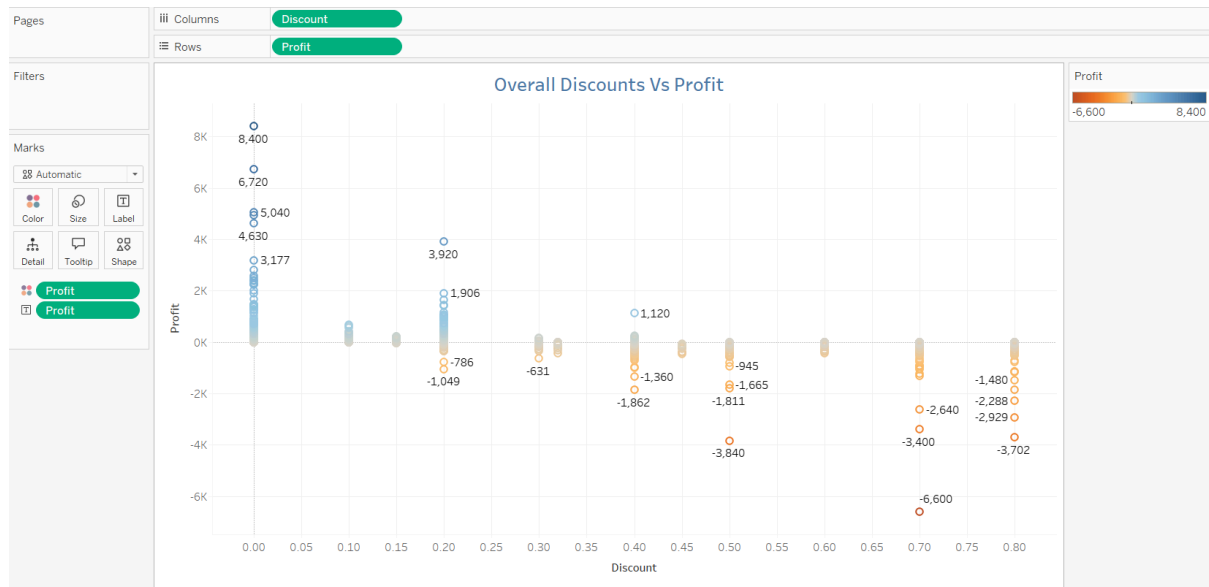


Reason for Chart Selection:- A horizontal bar visually compares the size or value of different categories, making it easy to see which categories are larger or smaller. It also makes large amounts of data easier to read and interpret using bars of varying lengths.

It effectively highlights differences between categories, making it clear where significant variations exist.

Observations:- We can observe here that the Technology and Furniture category Products are processed within the same number of days, however, Office Supplies category products have more days to process in comparison with the other two categories.

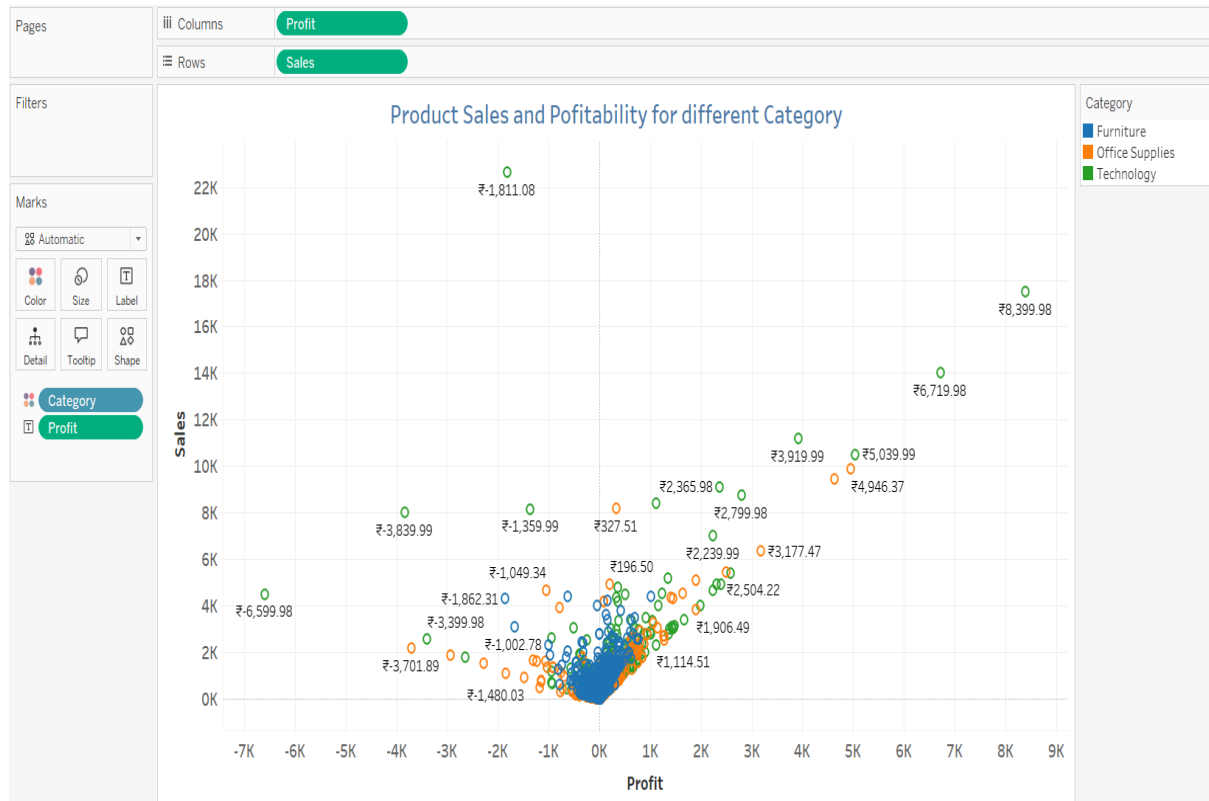
12. How do discounts affect overall profit?



Reason for Chart Selection:- A scatter plot visually represents the relationship between two numerical variables, making it easy to observe correlations, trends, and patterns. Each point on the chart corresponds to a pair of values, allowing for the identification of clusters, outliers, and the strength of relationships between the variables. This chart effectively highlights how changes in one variable relate to changes in another variable, providing clear insights into the dynamics of their interaction. Scatter plots are particularly useful for detecting linear or non-linear relationships and understanding the impact of one variable on another.

Observations:- We can observe a downward trend in the above chart, also while the discount is increasing we can observe that profit is decreasing and maximum profits are generated with Zeero discounts. Also, data is not highly correlated.

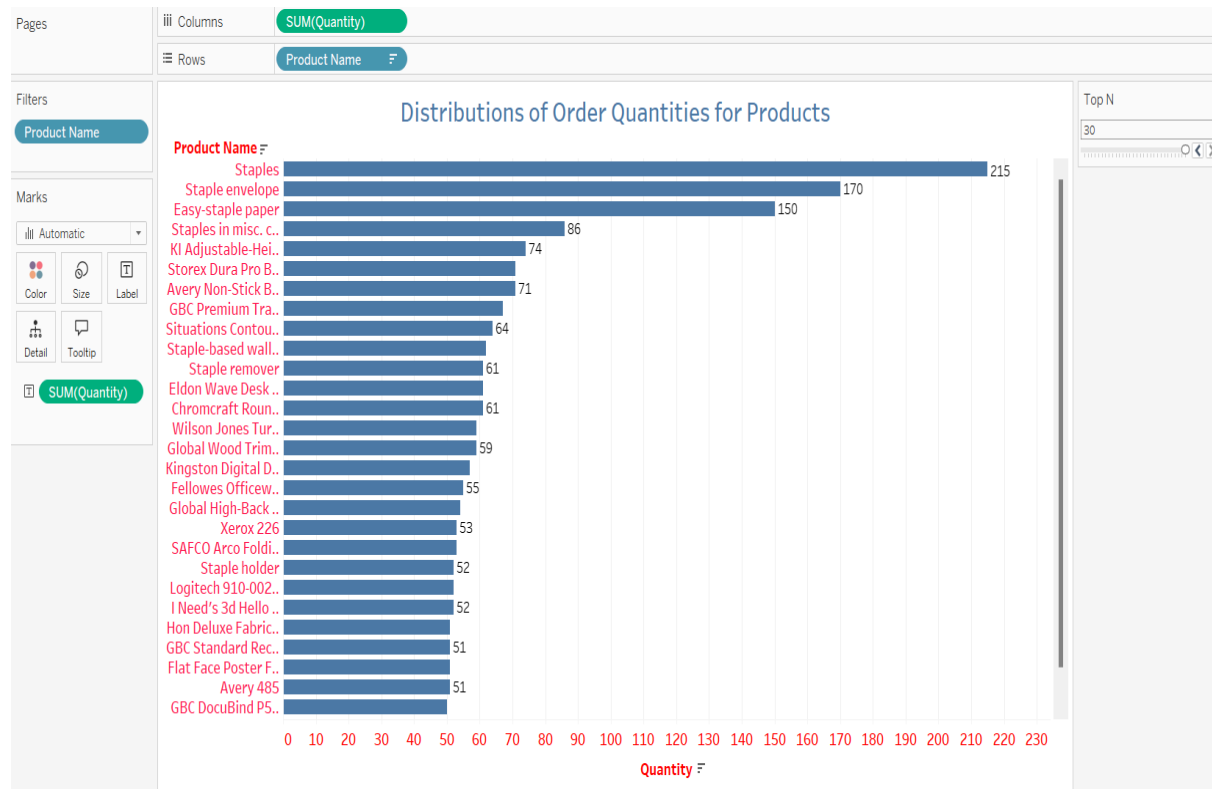
13. Can we visualise the relationship between product sales and profitability for different product categories?



Reason for Chart Selection:- A scatter plot visually represents the relationship between two numerical variables, making it easy to observe correlations, trends, and patterns. Each point on the chart corresponds to a pair of values, allowing for the identification of clusters, outliers, and the strength of the relationship between the variables. By plotting these points, a scatter plot effectively highlights how changes in one variable relate to changes in another variable, providing clear insights into their interaction. Scatter plots are particularly useful for detecting both linear and non-linear relationships, helping to understand how one variable might impact or be associated with another.

Observations: We can observe a positive correlation between sales and profit, also Technology product category leads in generating profits.

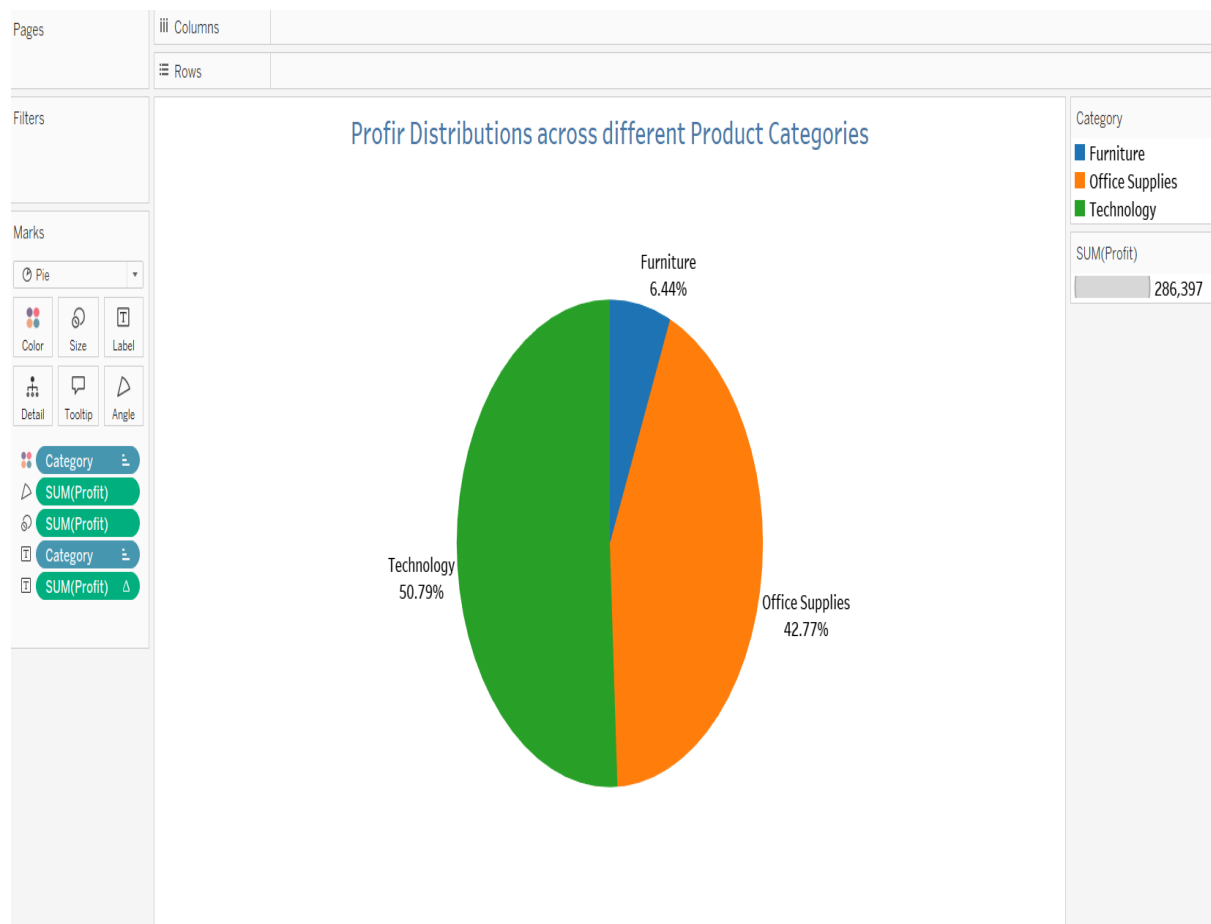
14. What is the distribution of order quantities for products in the dataset?



Reason for Chart Selection: A bar chart effectively visualizes the distribution of values across different categories, making it easy to compare the size or magnitude of each category. By displaying bars of varying lengths, it simplifies the interpretation of large amounts of data, clearly showing the frequency or quantity associated with each category. This chart highlights differences between categories, making it straightforward to identify patterns, trends, and significant variations within the distribution. Bar charts are particularly useful for comparing categorical data and understanding how values are distributed across different groups.

Observations: We can have clear look at the distribution of products in the above chart and using the Top N parameter we can dynamically check the top products with maximum quantities.

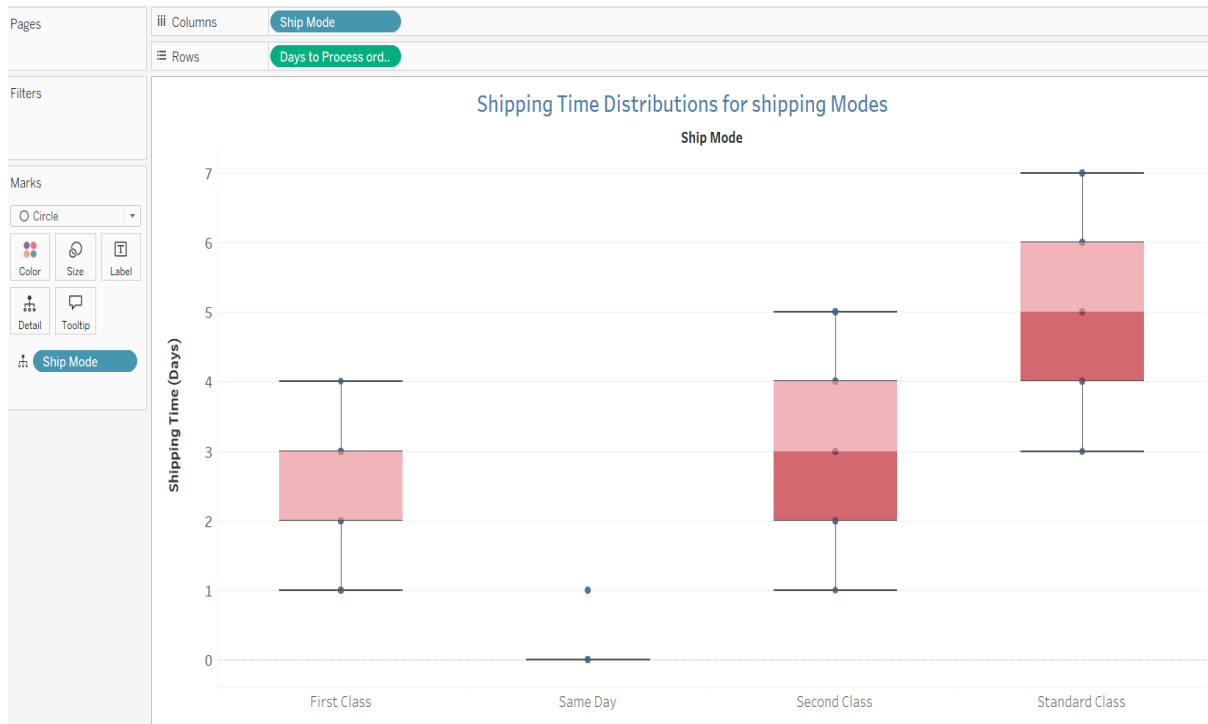
15. How do the profit distributions vary across different product categories?



Reason for Chart Selection: A pie chart visually represents the proportions of different categories within a whole, making it easy to see the relative sizes of each segment. By displaying data as slices of a circle, it highlights the contribution of each category to the total, helping to quickly grasp the distribution and share of each part. This format is effective for illustrating percentages and part-to-whole relationships, making it ideal for comparing how individual components contribute to the overall dataset.

Observations:- We can clearly observe in the above pie chart that the Technology category has the highest contribution to the overall profit with a 50.79% profit margin and the least in the list is the Furniture category.

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

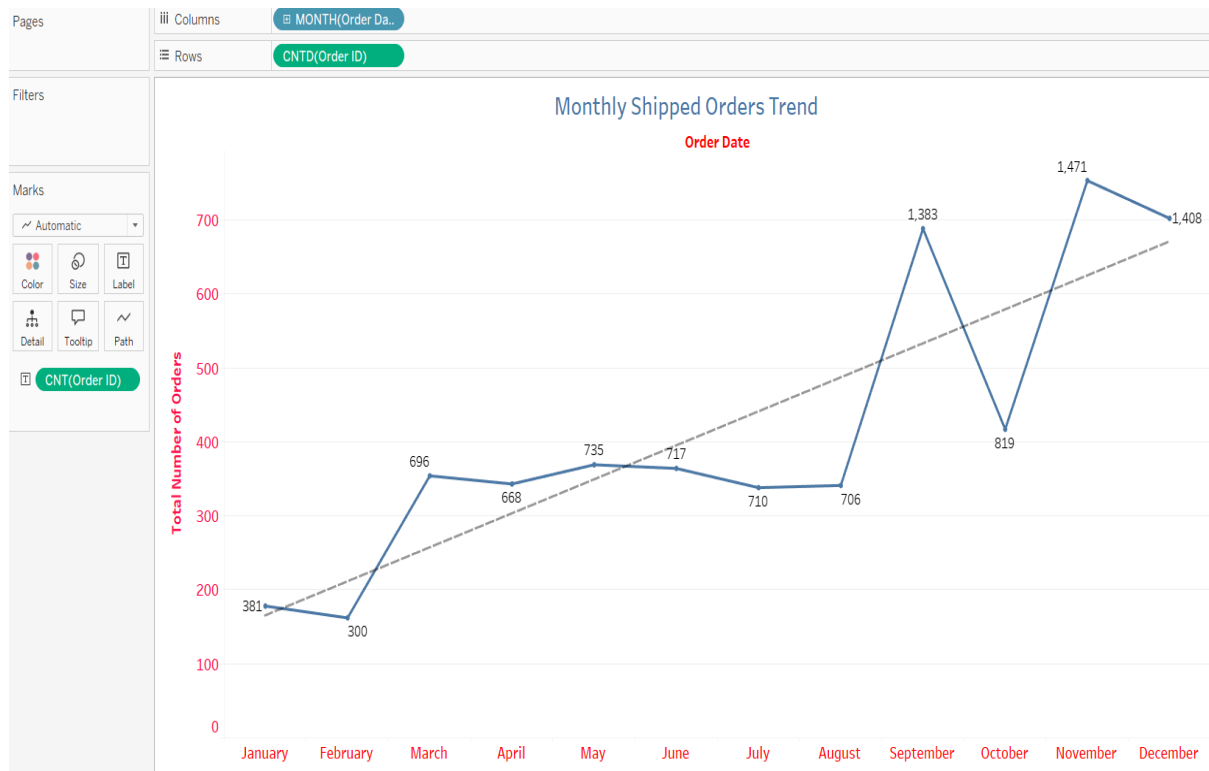


Reason for Chart Selection: A box plot visually represents the distribution of data across different categories, highlighting key statistical measures such as the median, quartiles, and potential outliers. By displaying data through a box with whiskers extending to the minimum and maximum values (excluding outliers), it effectively illustrates the range, center, and variability of the data. This format is particularly useful for comparing the spread and central tendencies across multiple categories or groups, making it ideal for identifying patterns, comparing distributions, and spotting anomalies within the dataset. Box plots are effective for understanding data distribution and variability at a glance.

Observations:- We can observe in the above boxplot that the median of Standard Class is highest compared rest of the shipping mode, also orders get shipped within 3-7 days.

We can observe an outlier in same-day ship mode as well, i.e. order getting shipped the very next day.

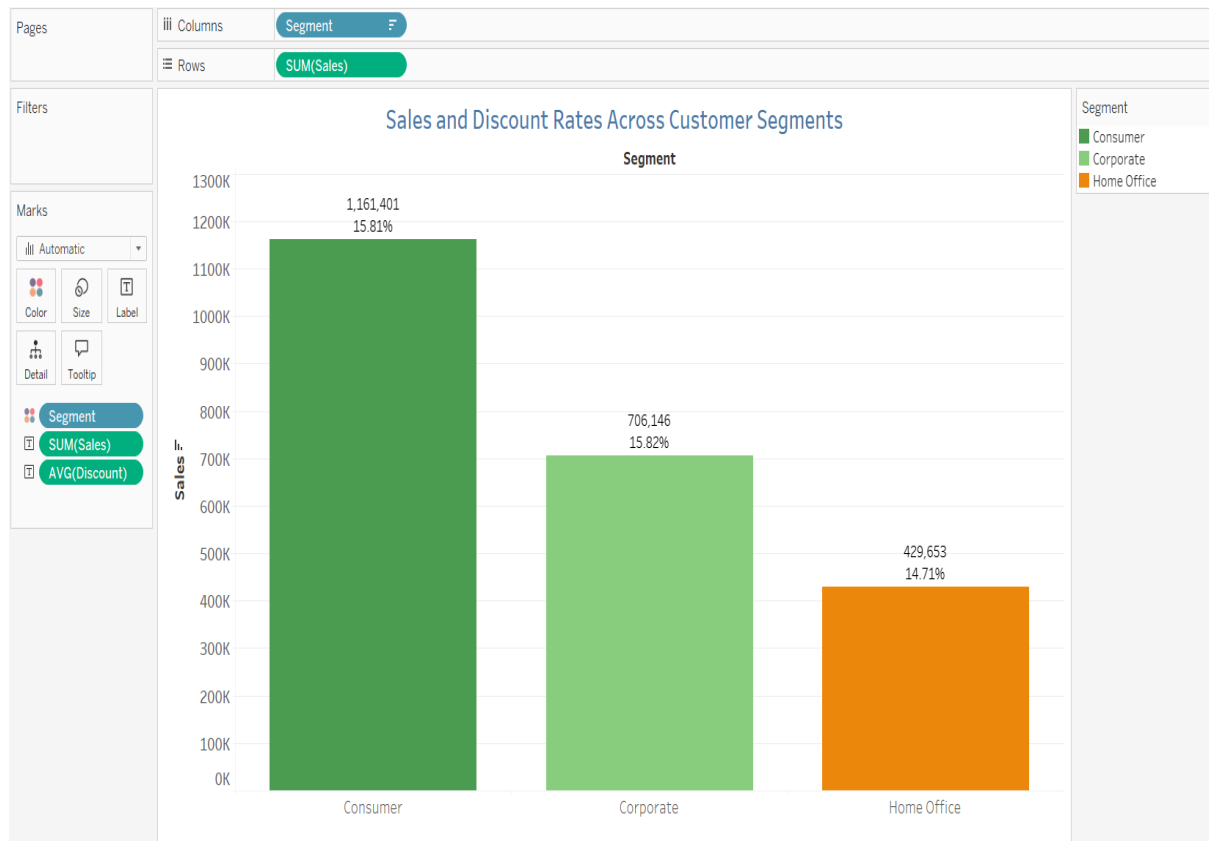
17. What is the monthly trend in the number of orders shipped?



Reason for Chart Selection:- A line chart visually displays trends and changes over time, making it easy to track how values progress or fluctuate. It connects data points with a continuous line, highlighting patterns and trends clearly. This format is effective for showing relationships and trends across time periods, making it ideal for analyzing data that evolves over a sequence. The line chart simplifies the comparison of data points and reveals long-term trends and variations.

Observations:- We can observe above an overall increasing trend of monthly shipped orders. We can observe a maximum order shipped in November and a minimum in the month of February. Also, a gradual decrease in the shipped order count can be observed between May and August.

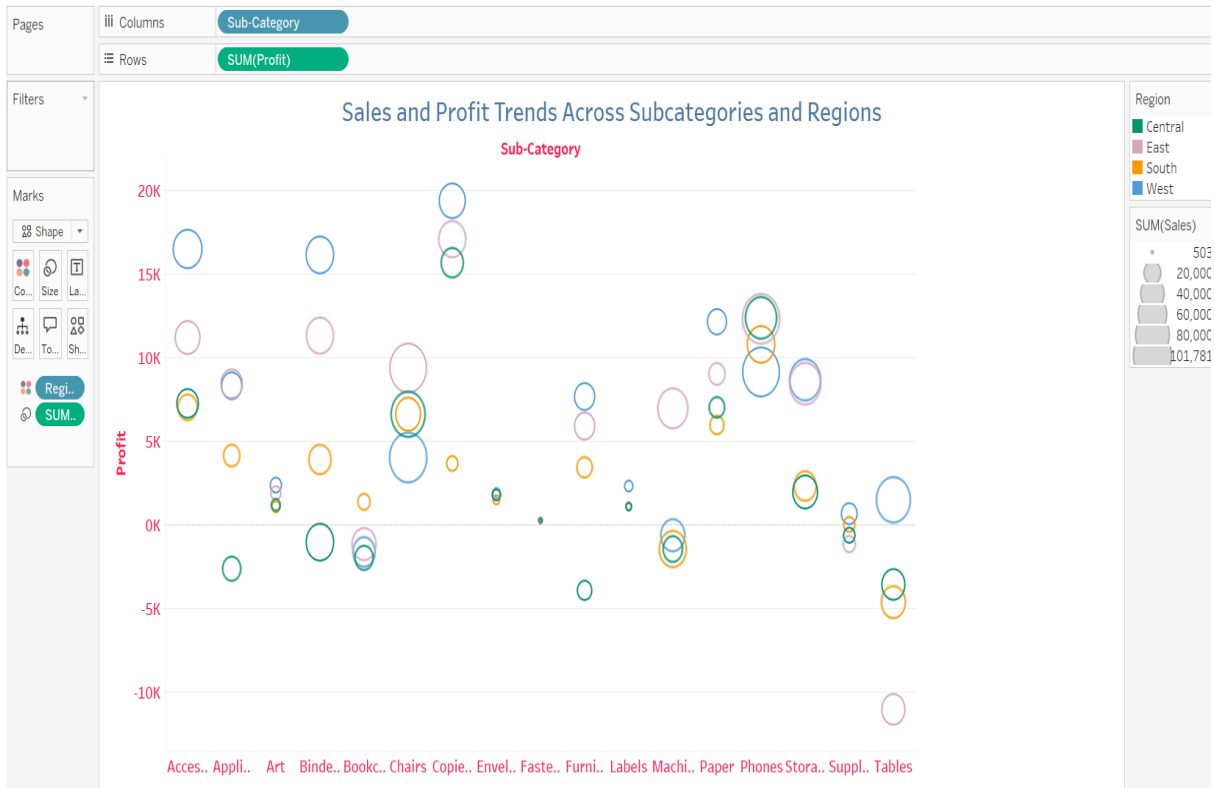
18. How do different customer segments perform in terms of sales and discount rates?



Reason for Chart Selection:- A column bar chart visually compares values across different categories, making it easy to see variations in size or magnitude. By displaying data as vertical bars, it highlights differences between categories clearly and allows for straightforward comparison of individual values. This format is effective for illustrating the relative size of each category, helping to quickly identify which categories are larger or smaller. Column bar charts are ideal for showing discrete data and are particularly useful for analyzing and comparing quantities across categorical variables.

Observations:- We can observe in the above chart consumer segment has maximum sales with around 15.81% discount and the corporate segment stands second with a maximum discount of 15.82% and second highest sales. Also, the Home office segment is providing good discounts but sales volume is least compared to the rest.

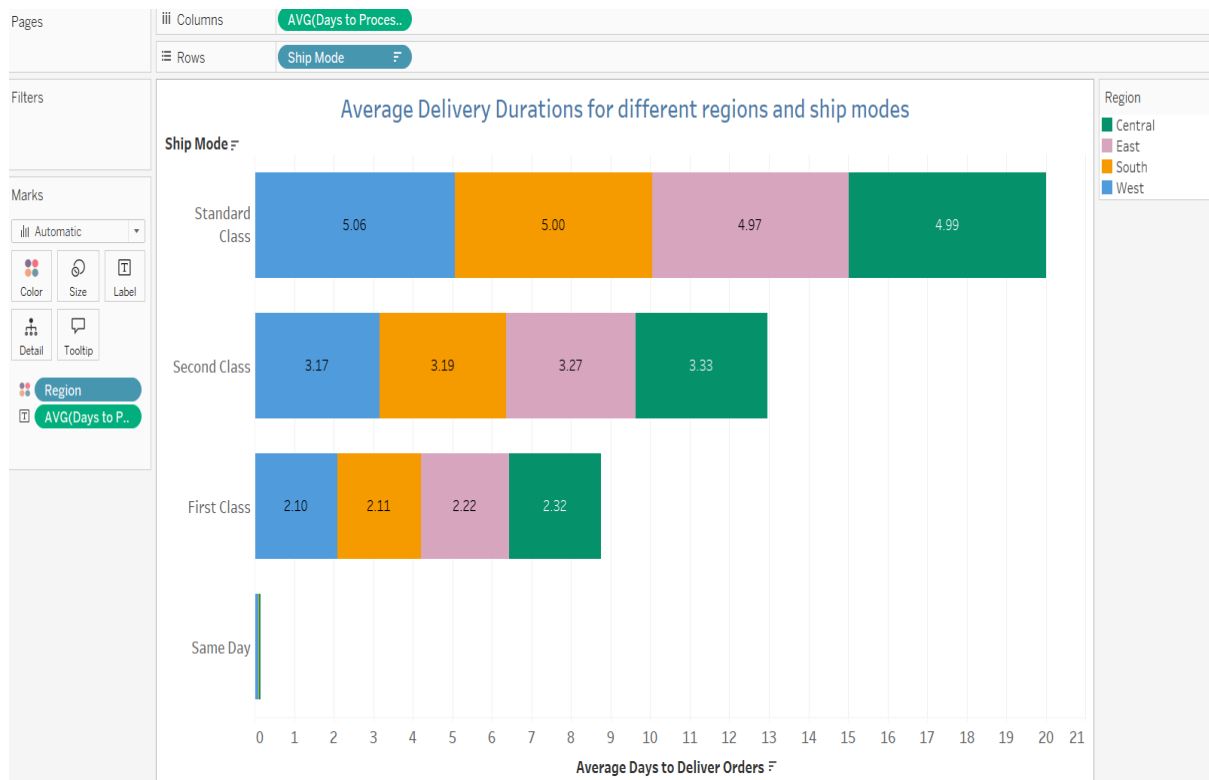
19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?



Reason for Chart Selection:- A scatter plot visually represents the relationship between two numerical variables, making it easy to observe correlations, trends, and patterns. Each point on the chart corresponds to a pair of values, allowing for the identification of clusters, outliers, and the strength of the relationship between the variables. By plotting these points, a scatter plot effectively highlights how changes in one variable relate to changes in another variable, providing clear insights into their interaction. Scatter plots are particularly useful for detecting both linear and non-linear relationships, helping to understand how one variable might impact or be associated with another.

Observations:- We can observe in the above chart that sub-category copiers is generating high profit and with a good count of sales in West East and central regions. Also, chairs and Phone categories have a good amount of sales with good profit across all regions.

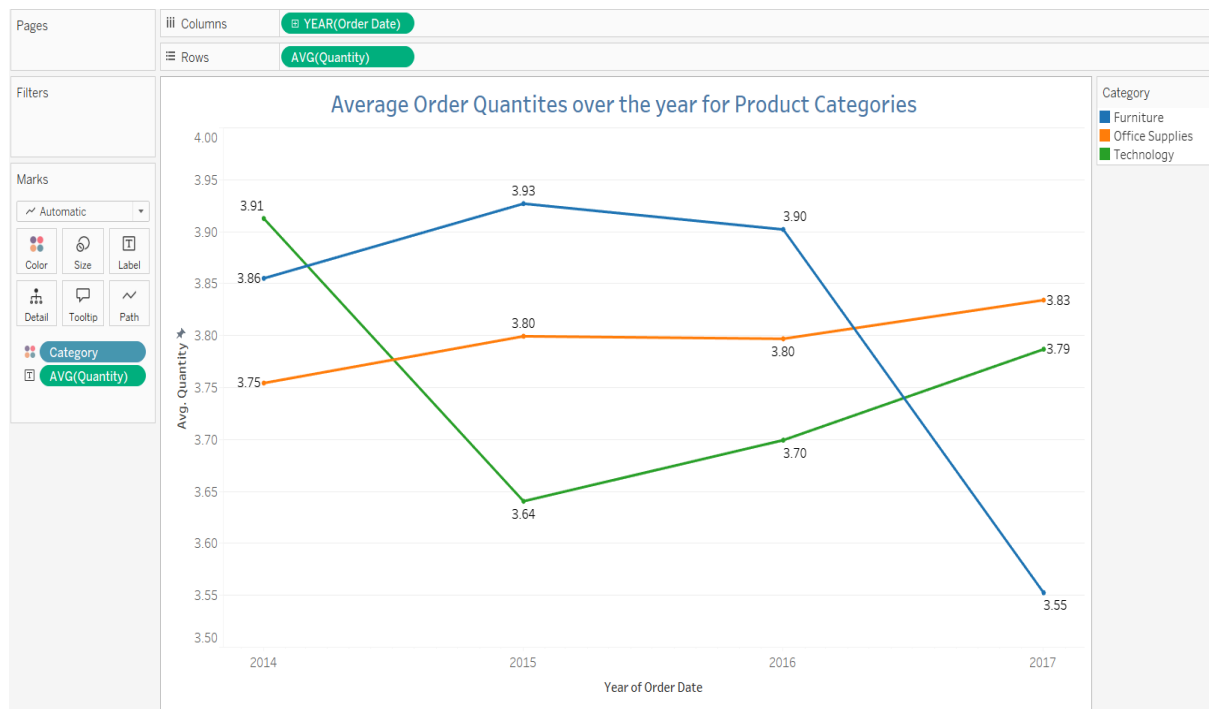
20. What is the average delivery duration for different regions and ship modes?



Reason for Chart Selection: A horizontal stacked bar chart visually compares the size or value of different categories while breaking down each category into subcategories. This format, with bars oriented horizontally, makes it easy to see the total value of each main category and the distribution of subcategories within it. It effectively highlights the composition and relative contribution of each segment to the overall total, facilitating clear comparison and analysis of both individual components and cumulative data. The horizontal orientation enhances readability, especially when category labels are long.

Observations:- We can observe in the above chart about average delivery durations for different regions and ship modes. In standard class ship mode around 5 days are taken to deliver orders in the west and south regions while in first class ship mode around 2 days are taken to deliver in the west and south regions.

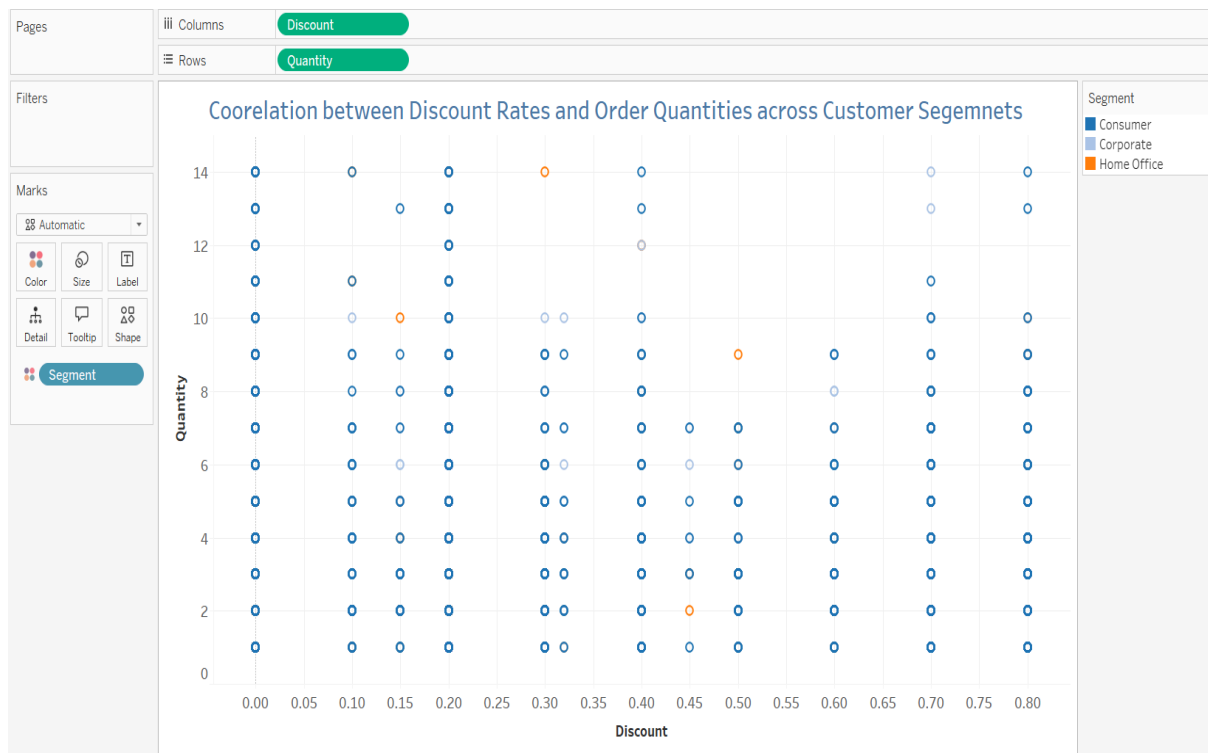
21. How has the average order quantity changed over the years for various product categories?



Reason for Chart Selection: A line chart with multiple lines differentiates categories by using distinct lines for each category, making it easy to track and compare the trends and changes over time for each category. This format connects data points with continuous lines, clearly showing patterns, trends, and variations for each category. It effectively highlights how different categories evolve over time, facilitating the comparison of their respective trajectories and long-term trends.

Observations:- We can observe in the above chart that the furniture category declined steeply from the year 2016-2017. Also, a declining trend can be observed for average quantities between 2014 and 2015 for the technology category and an increasing trend after that. In the case of office supplies we can observe that initially, it was an increasing trend between 2014-2015 and a slight decreasing trend can also be observed between 2015-2016 and again an increasing trend between 2016 -2017.

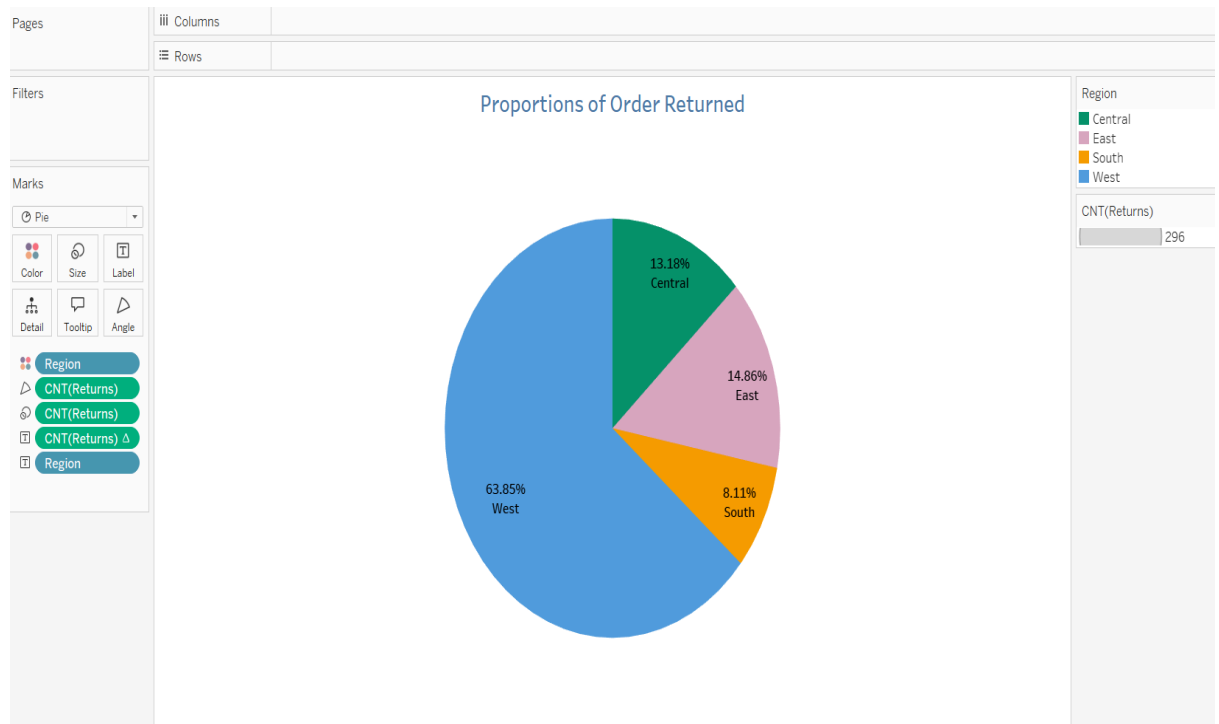
22. Can we visualise the correlation between discount rates and order quantities for different customer segments?



Reason for Chart Selection:- A scatter plot visually represents the relationship between two numerical variables, making it easy to observe correlations, trends, and patterns. Each point on the chart corresponds to a pair of values, allowing for the identification of clusters, outliers, and the strength of relationships between the variables. This chart effectively highlights how changes in one variable relate to changes in another variable, providing clear insights into the dynamics of their interaction. Scatter plots are particularly useful for detecting linear or non-linear relationships and understanding the impact of one variable on another.

Observations:- We can observe a slightly positive correlation in the above chart. Also, a moderate correlation can be observed in the corporate segment with more discounts and more quantities ordered.

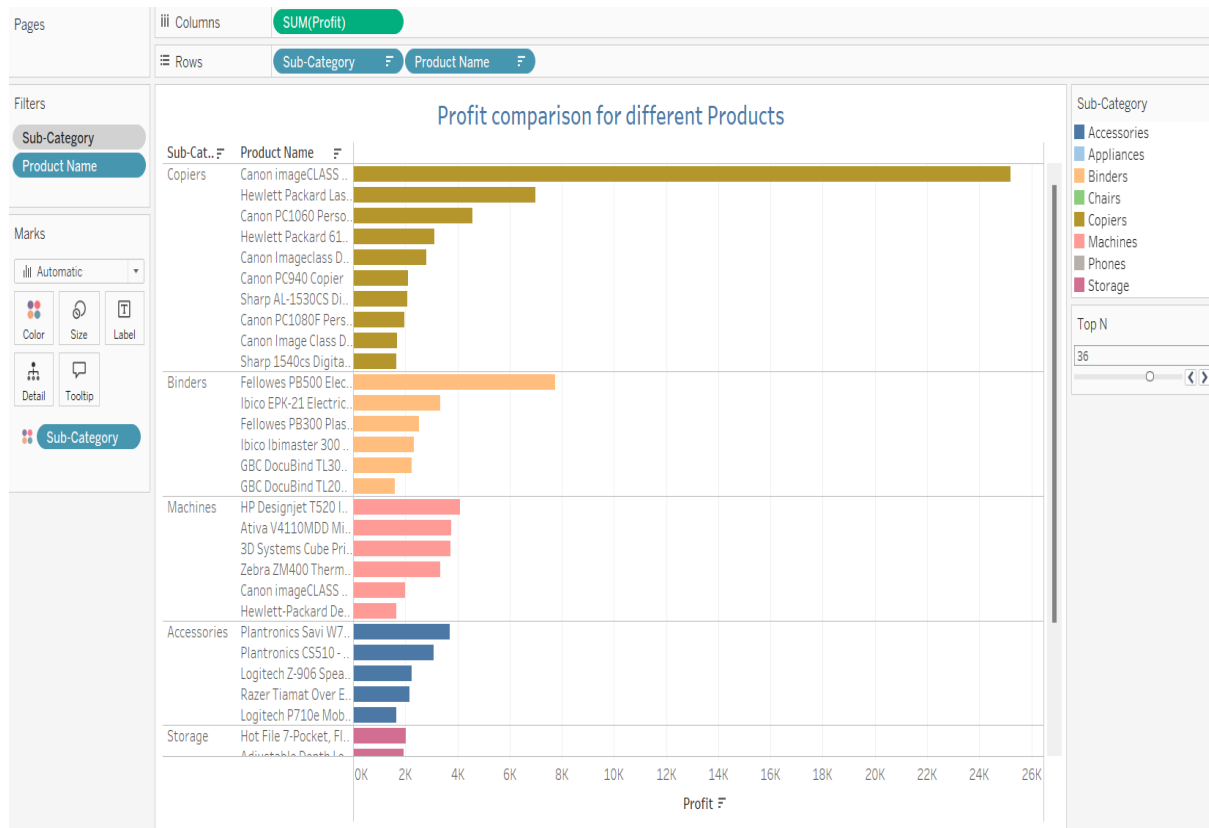
23. What is the proportion of orders returned in each region within the Superstore dataset?



Reason for Chart Selection: A pie chart visually represents the proportions of different categories within a whole, making it easy to see the relative sizes of each segment. By displaying data as slices of a circle, it highlights the contribution of each category to the total, helping to quickly grasp the distribution and share of each part. This format is effective for illustrating percentages and part-to-whole relationships, making it ideal for comparing how individual components contribute to the overall dataset.

Observations:- We can observe here that maximum number of orders were returned from the West Region with 63.85% and 14.86% from the East Region. Also, very less orders were returned from south region.

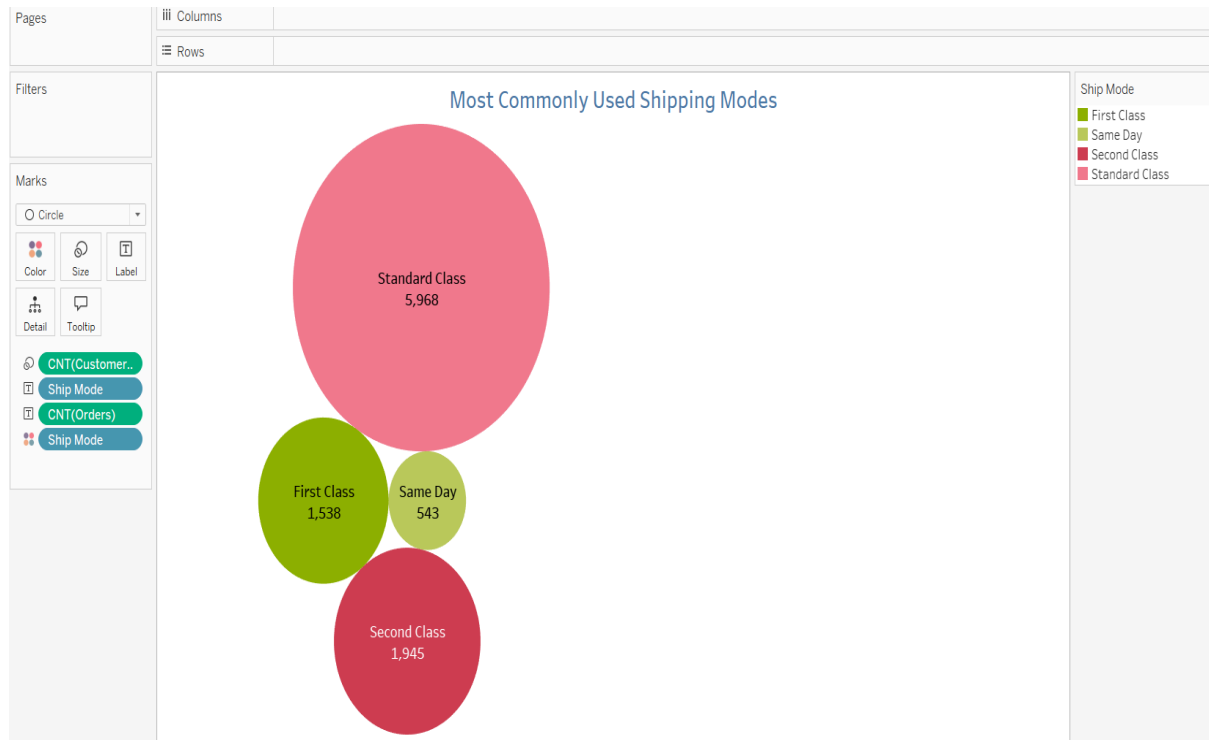
24. Can you compare the profit of different products for different subcategories?



Reason for Chart Selection:- A horizontal side-by-side bar chart in Tableau visually compares the proportions of different categories within a whole by displaying multiple horizontal bars adjacent to each other. This layout enhances the ability to see and compare the relative sizes and contributions of each category across various metrics or periods.

Observations:- We can observe in the above chart that the copiers sub-category generated the maximum profit. Binders and machines stand second and third. Also, we can have a look at the most profit-generating products under each sub-category.

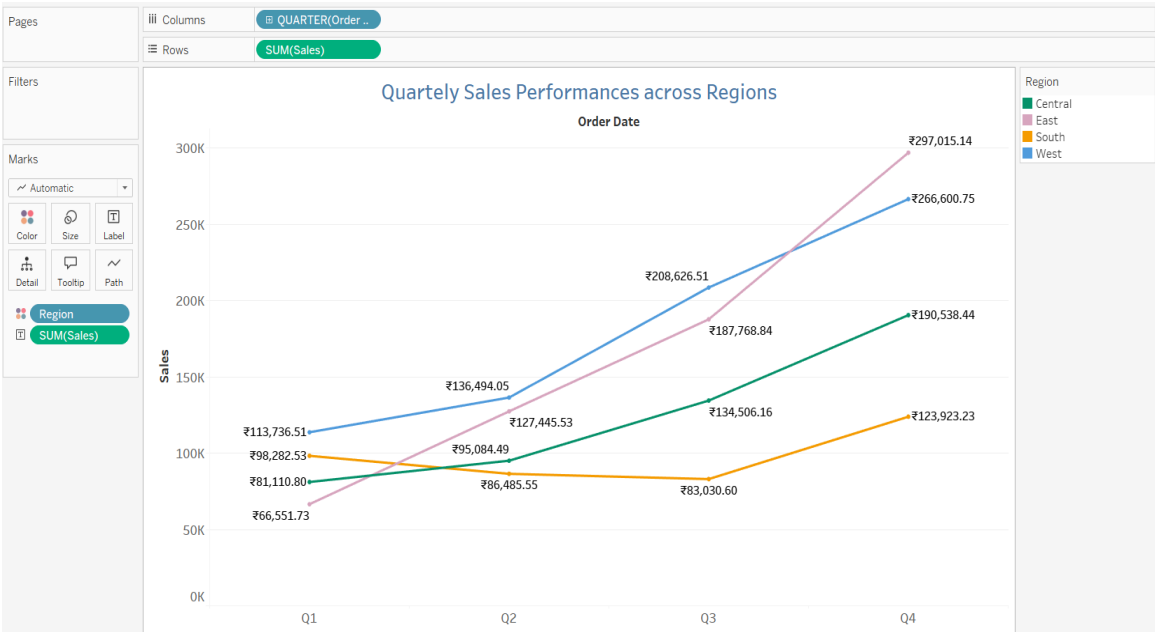
25. Which shipping mode is the most commonly used in the Sample Superstore dataset?



Reason for Chart Selection: A bubble chart effectively displays three dimensions of data simultaneously, making it ideal for visualizing relationships and patterns among multiple variables. By representing data points as bubbles with varying sizes and colors, it allows for an intuitive understanding of the magnitude and distribution of values across categories. This format is particularly useful for examining correlations, trends, and the relative importance of different variables, offering a comprehensive view of complex datasets.

Observations:- We can observe in the above chart that the Standard class is the most used shipping mode for orders and the least used mode is the Same day, which might be due to the same-day shipping mode having more charges.

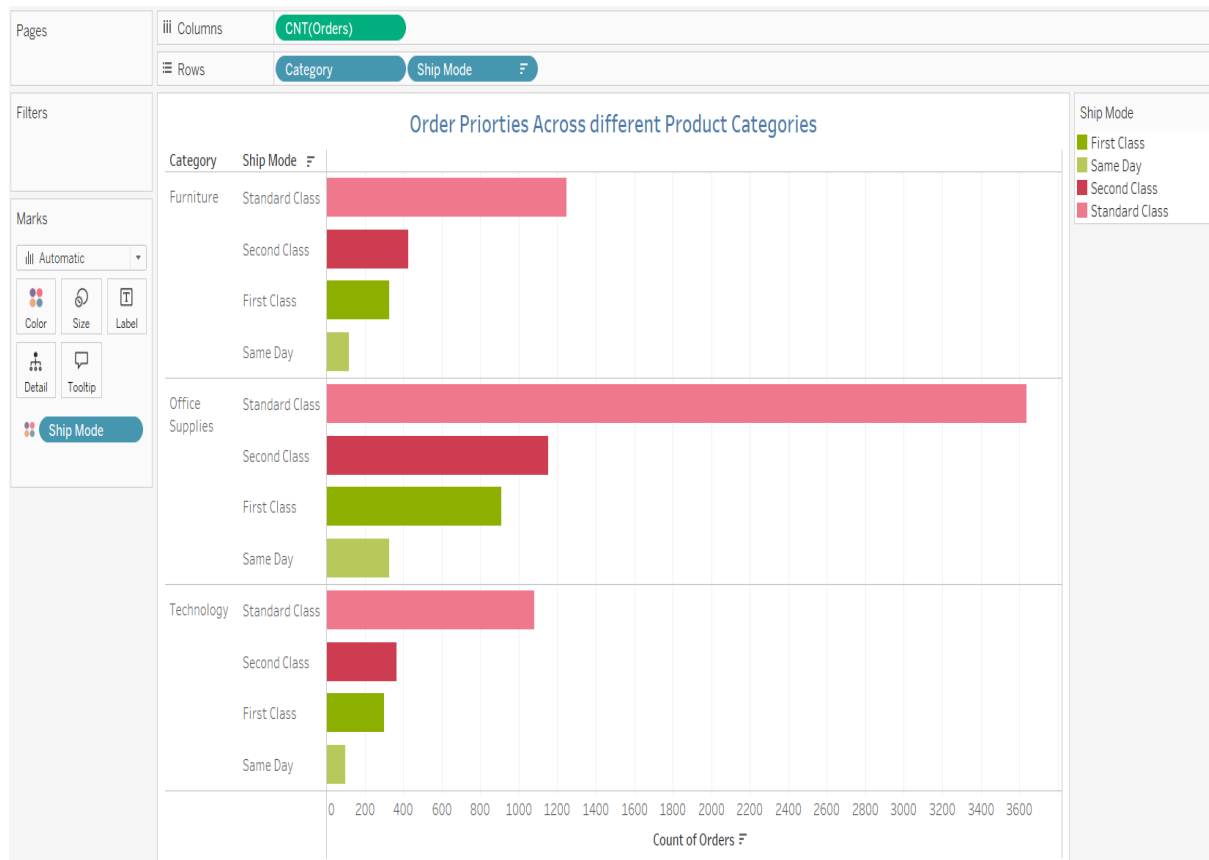
26. How does the sales performance of different regions evolve throughout the quarters of a year?



Reason for Chart Selection: A line chart with multiple lines differentiates categories by using distinct lines for each category, making it easy to track and compare the trends and changes over time for each category. This format connects data points with continuous lines, clearly showing patterns, trends, and variations for each category. It effectively highlights how different categories evolve over time, facilitating the comparison of their respective trajectories and long-term trends.

Observations: While looking at the quarterly sales, we can observe that the east region is showing a good improving trend in comparison to other regions with max sales in Q4. South region is lagging in sales in comparison to the rest of the region and was showing a decreasing trend in sales in first 3 quarters.

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



Reason for Chart Selection: A side-by-side horizontal bar chart allows for a straightforward comparison of categories by placing bars next to each other horizontally. This format is particularly effective for comparing the size and length of bars across different categories, making it easy to visualize and contrast their values side by side. It highlights differences and similarities between categories clearly, facilitating direct comparison and analysis of their relative sizes and distributions.

Observations:- We can observe in the above chart that maximum orders for the office supplies category have the highest number of orders across all shipping modes.

28. What is the relationship between discounts and sales?



Reason for Chart Selection:- A scatter plot visually represents the relationship between two numerical variables, making it easy to observe correlations, trends, and patterns. Each point on the chart corresponds to a pair of values, allowing for the identification of clusters, outliers, and the strength of relationships between the variables. This chart effectively highlights how changes in one variable relate to changes in another variable, providing clear insights into the dynamics of their interaction. Scatter plots are particularly useful for detecting linear or non-linear relationships and understanding the impact of one variable on another.

Observations:- We can observe a slightly negative correlation between sales and discounts stating high discounts are not needed for maximum sales. We can observe higher sales even with zero discounts.

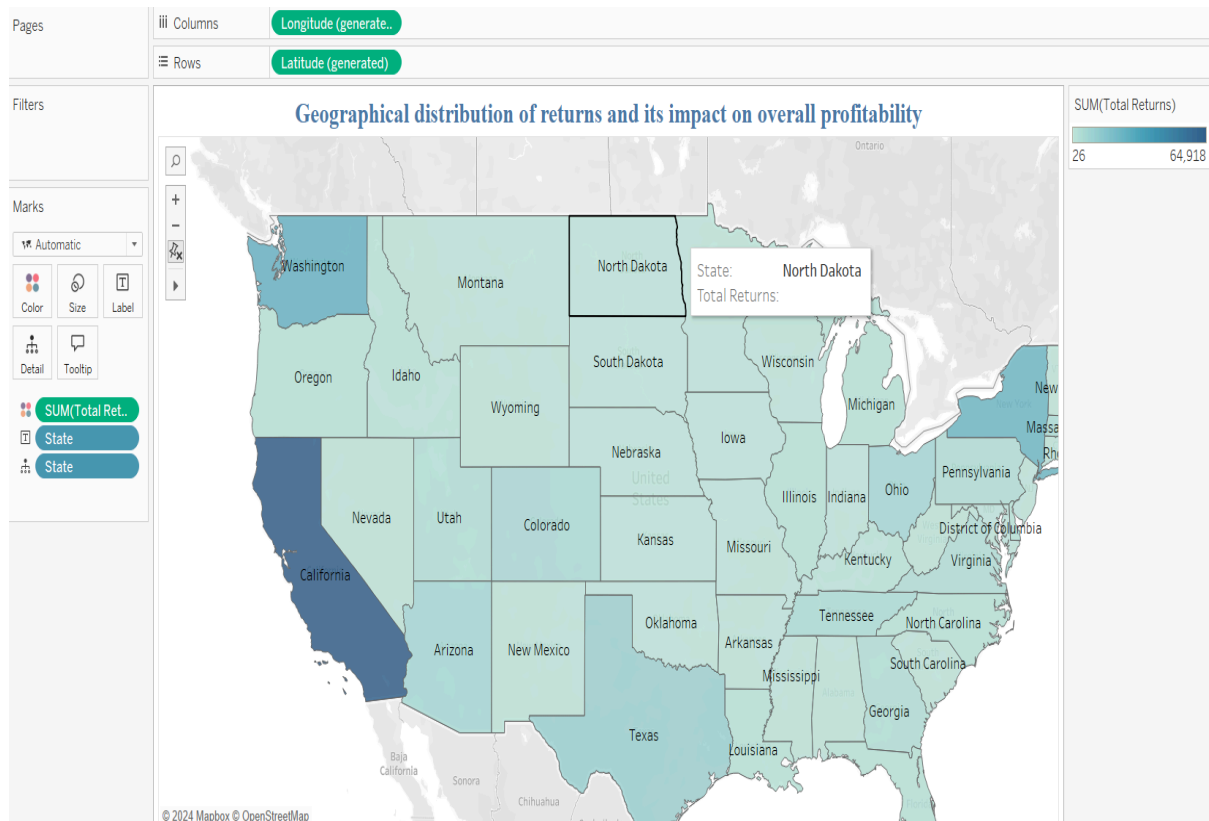
29. How does the average order value differ between repeat customers and new customers?



Reason for Chart Selection: A column bar chart effectively compares the size or value of different categories by using vertical bars to represent each category. This format makes it easy to see and compare the heights of the bars, allowing for a clear visual distinction between categories. It is particularly useful for displaying and interpreting data across different categories, highlighting variations in values and providing a straightforward way to assess the magnitude of differences. The vertical orientation of the bars makes it simple to understand and analyze the data distribution and trends.

Observations:- We can observe in the above chart that the average order value for new customers is higher in comparison with Repeat Customer.

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



Reason for Chart Selection: A geographical map effectively displays data across different regions by using visual markers and colour gradients to represent various values or metrics. This format allows for an intuitive understanding of how data is distributed geographically, highlighting regional patterns and variations. By visualizing data on a map, it is easy to see spatial relationships and trends, making it straightforward to analyze how different areas compare and to identify geographical clusters or disparities. This approach is particularly useful for presenting data with a spatial component and for gaining insights into regional distribution.

Observations:- We can observe in the above chart that maximum returns are from the western and northern sides of the country and the same is impacting overall profitability. Also, fewer returns are impacting overall profitability from the southern and eastern regions in comparison.