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.

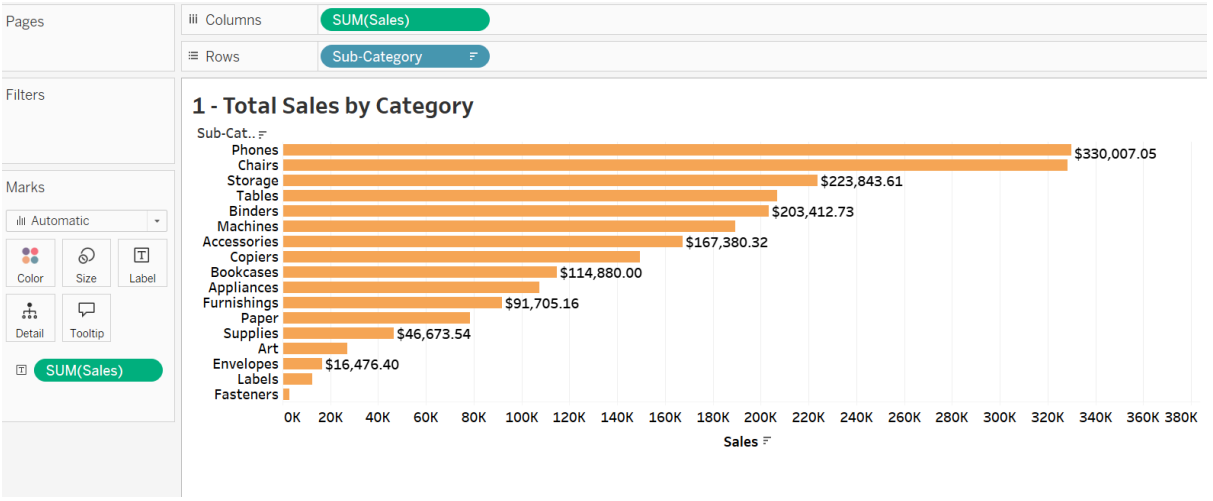
By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

Good luck!

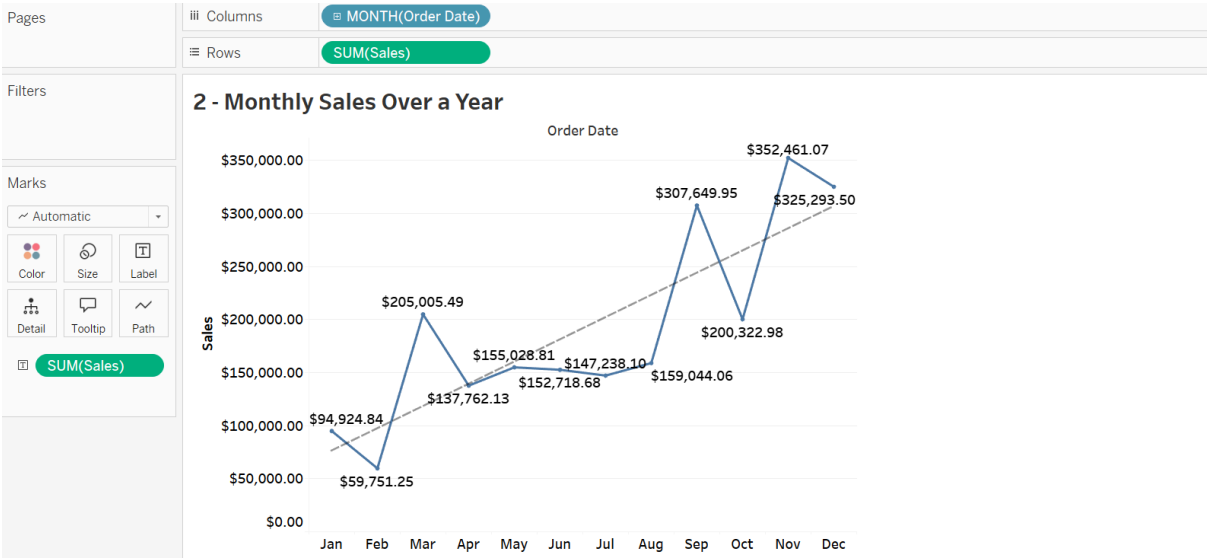
Questions:

1. Which product categories have the highest total sales in the "Superstore" dataset?
   A bar chart is one of the best visualizations for comparing total sales across different product categories because each category gets a separate bar, making it easy to compare which categories have higher or lower sales. The bar length directly

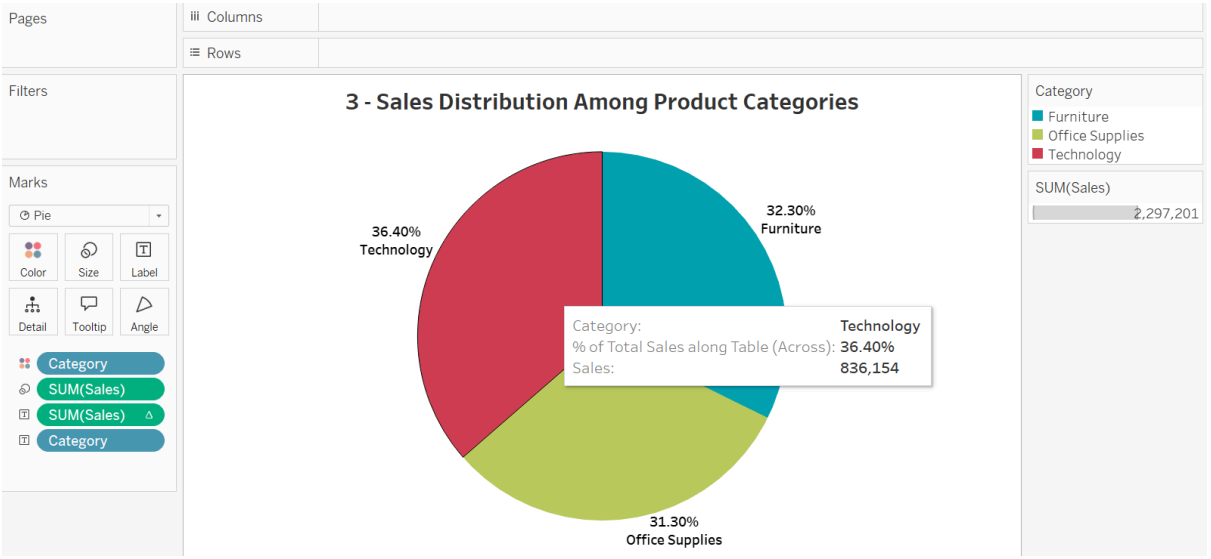
represents total sales. Since "Category" is a discrete field (Furniture, Technology, Office Supplies), a bar chart naturally organizes data into separate groups. If there's a big difference between the sales of different categories, a bar chart visually emphasizes this gap.



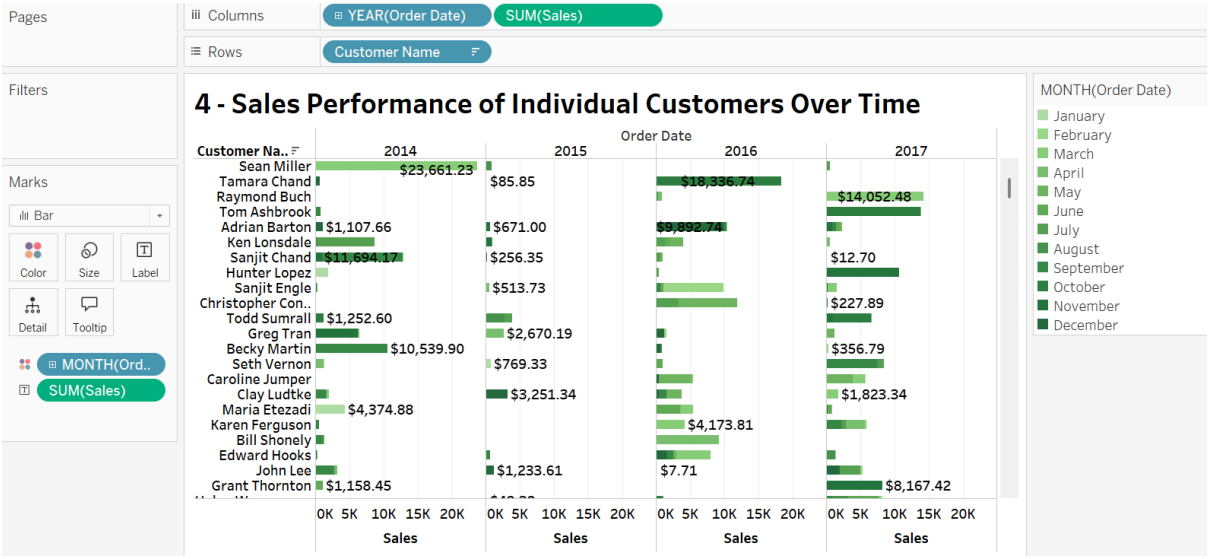
2. How do the monthly sales amounts change over the course of a year?
A line chart with a trend line is the best choice for visualizing monthly sales over a year because it effectively shows patterns, trends, and seasonality in sales. Sales data over time is continuous, making a line chart the most natural way to represent changes. Months are sequential, so connecting them with a line helps visualize the flow of sales over time. A line chart helps identify sales fluctuations whether sales increase, decrease, or remain stable over months. If there's seasonal variation (e.g., higher sales in December due to holidays), it becomes immediately visible. The Trend Line helps identify growth patterns, smooths out fluctuations to show the overall sales direction and helps determine if sales are steadily increasing, decreasing, or fluctuating.



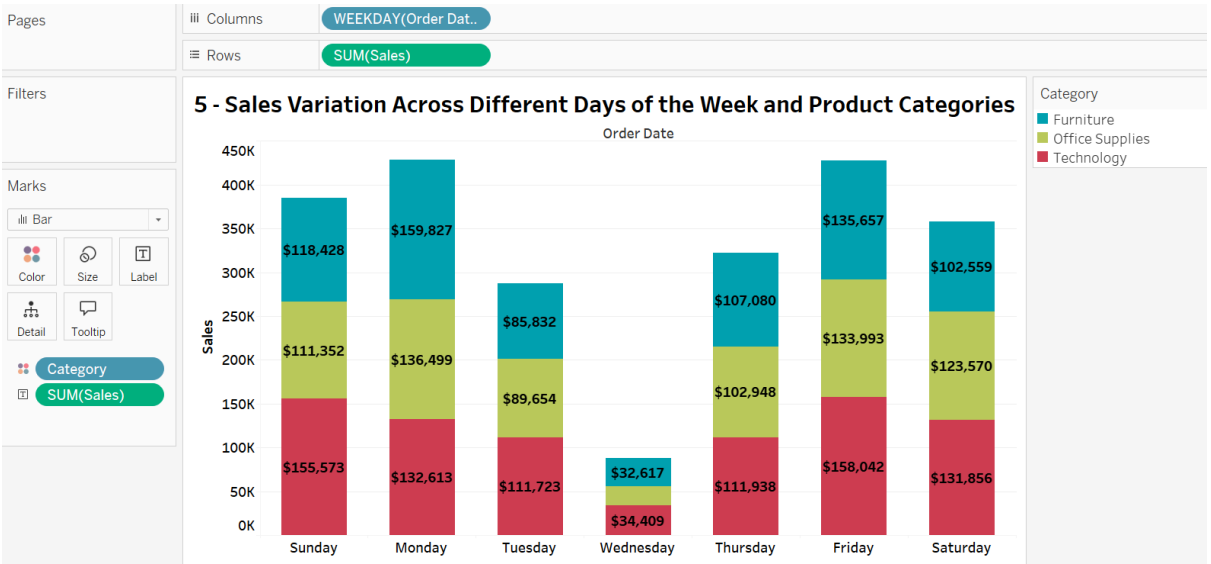
3. How is the total sales amount distributed among different product categories?
   A pie chart is useful for visualizing sales distribution among product categories because it effectively shows proportions and helps answer "Which category contributes the most to total sales?". A pie chart is best when comparing parts of a whole. It helps in understanding how much each category contributes to the total sales. If Furniture = 30%, Technology = 50%, and Office Supplies = 20%, the chart immediately shows which category dominates sales. If you have 3 to 5 categories, a pie chart makes comparisons simple. Larger slices indicate higher sales contribution, and smaller slices highlight lower sales.



4. Can we analyze the sales performance of individual customers over time?
   A stacked bar chart is a great choice for analyzing sales performance of individual customers over time because it provides a clear breakdown of contributions across different customers in a time series format. A stacked bar chart segments sales by customer while keeping the time trend visible. Each bar represents total sales for a specific time period (e.g., month, year). The stacked sections show individual customer contributions, making it easy to see who contributes the most. Helps identify top-performing customers, shows if some customers are growing while others are declining and reveals seasonal trends in purchases by different customers.

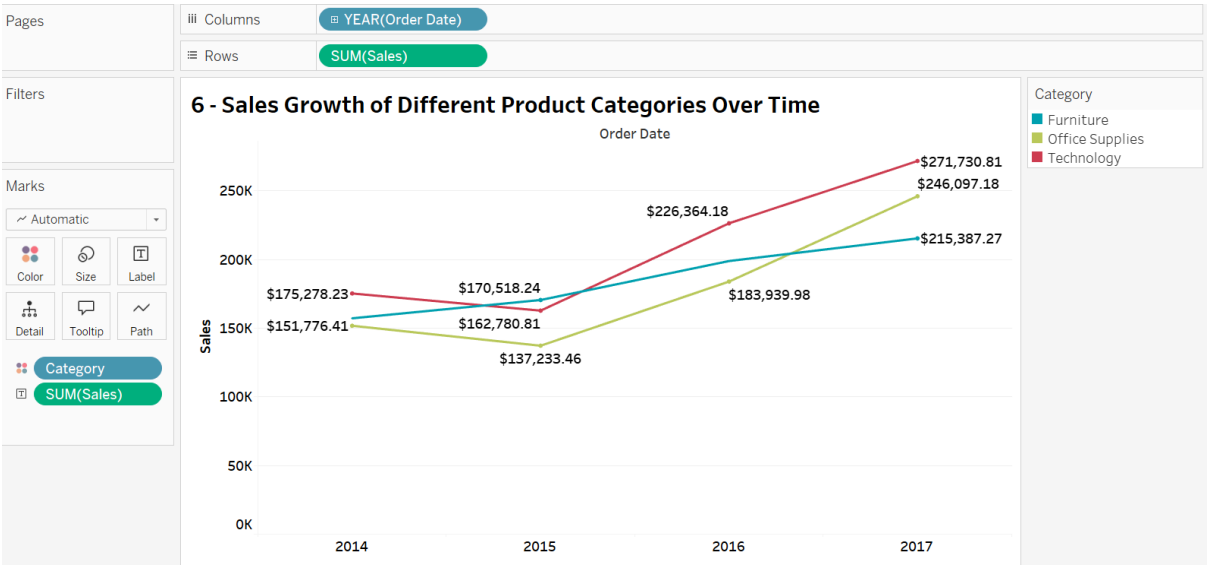


5. How do sales vary based on different days of the week and product categories?
   A bar chart is the best choice for analyzing sales variation across different days of the week and product categories because it makes comparisons clear, structured, and easy to interpret. Each product category can be displayed as a separate bar for each day of the week. We can easily compare which days have higher/lower sales for each category. Technology may have peak sales on Fridays, while Office Supplies sell more on Mondays. A bar chart makes this pattern immediately visible. The bar length represents sales, making it easy to compare between days. If one bar is significantly taller, it signals a higher sales volume for that day.

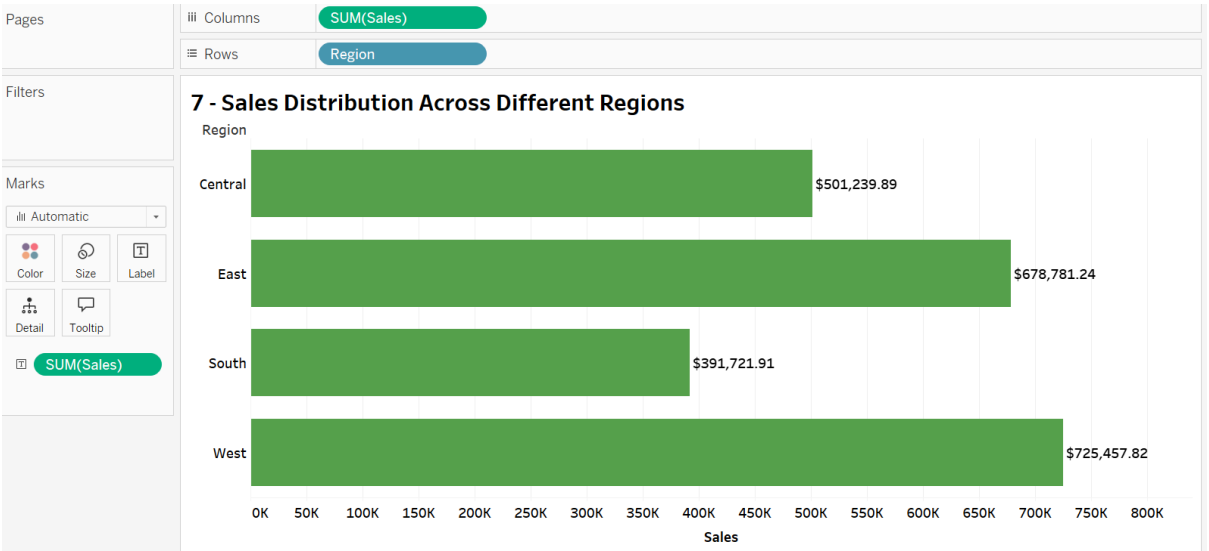


6. Can we visualise the sales growth of different product categories over time?
   A line chart is the best choice for analyzing sales growth of different product categories over time because it effectively shows trends, patterns, and changes in sales performance over a period. Sales data is continuous over time, and a line chart is ideal for tracking changes. A line smoothly connects monthly/weekly sales points, making trends easier to spot. Technology sales may show a steady upward trend, while

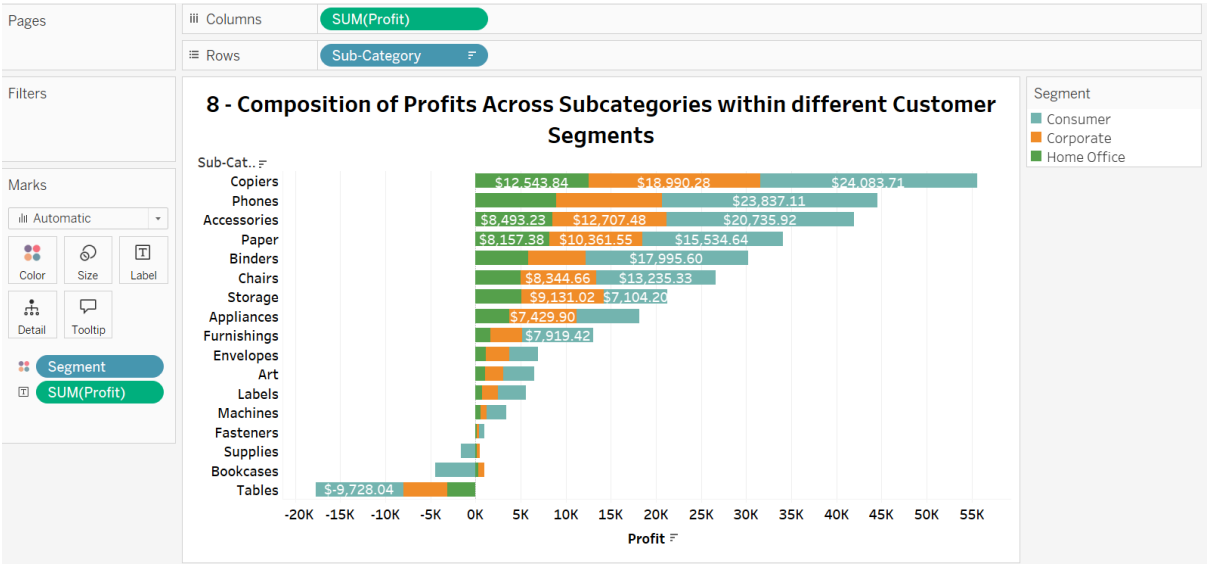
Furniture sales fluctuate and easily Identifies Growth Patterns & Seasonality. Office Supplies might have high sales in January. Furniture sales could peak in Q4 due to year-end corporate purchases. It allows Comparison Between Multiple Categories and each product category can have a separate line with a unique color.



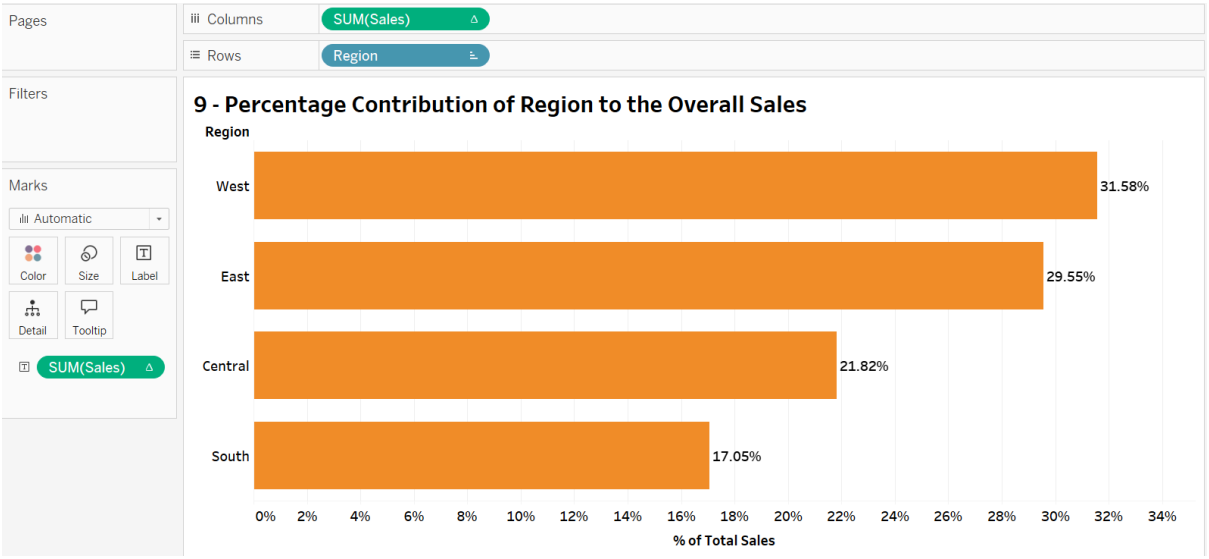
7. How does the sales distribution vary across different regions in the "Superstore" dataset?
   A bar chart is the best choice for analyzing sales distribution across different regions because it provides a clear, direct comparison of total sales in each region without unnecessary complexity. A bar chart makes it easy to compare total sales figures across different regions. Each bar represents a region, and its length corresponds to total sales. If West has the highest sales and South has the lowest, this difference is immediately visible in a bar chart. It is easy to Interpret and Read. If East and Central have similar sales, the bars will be almost equal in height, making comparison easy.



8. Can we visualise the composition of profits across various subcategories within different customer segments?
   A bar chart is the best choice for analyzing profit composition across subcategories within different customer segments because it provides a clear, structured comparison of profits for each subcategory while breaking it down by customer segments. Each subcategory gets a separate bar, making it easy to compare profits. Furniture might be highly profitable for Corporate customers but less profitable for Consumers. A bar chart makes this clear by visually separating each subcategory and segment.

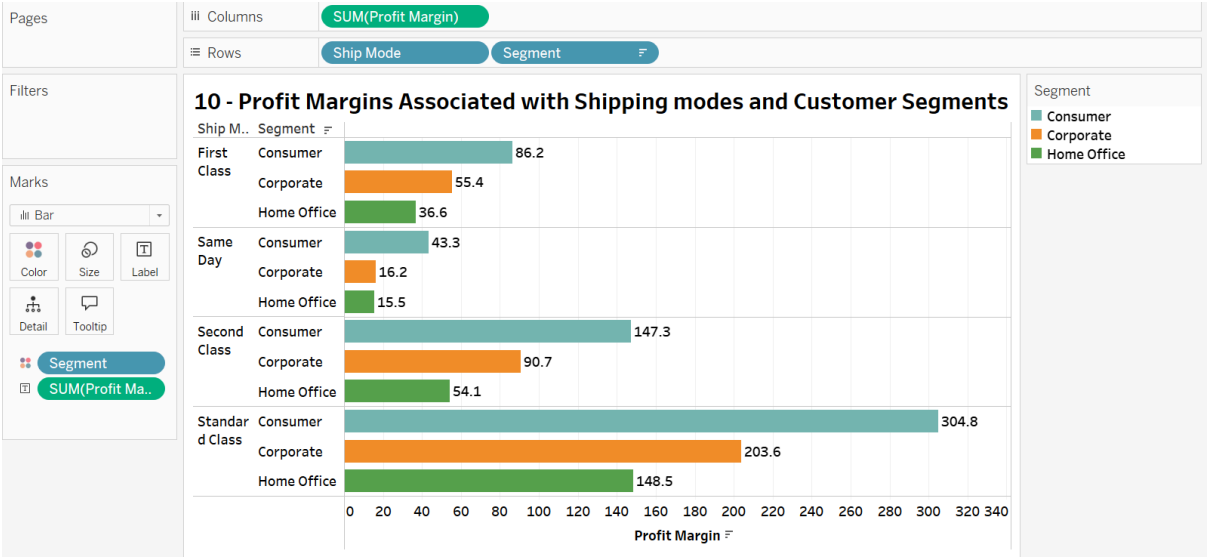


9. What is the percentage contribution of each region to the overall sales?
   A bar chart is an excellent choice for visualizing the percentage contribution of each region to total sales because it provides a clear, direct comparison of how different regions contribute to the overall revenue. Best for Comparing Categories Clearly. Each region gets a separate bar, making it easy to compare their contribution. The length of the bars represents the percentage of total sales each region contributes. If West contributes 40% of total sales, it will have the tallest bar, while South (10%) will have a shorter bar. Works Well for Discrete (Categorical) Data. Since regions are distinct categories, a bar chart is more appropriate than a line chart, which is meant for continuous data. Each bar stands alone, making it clear that regions are separate entities.



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

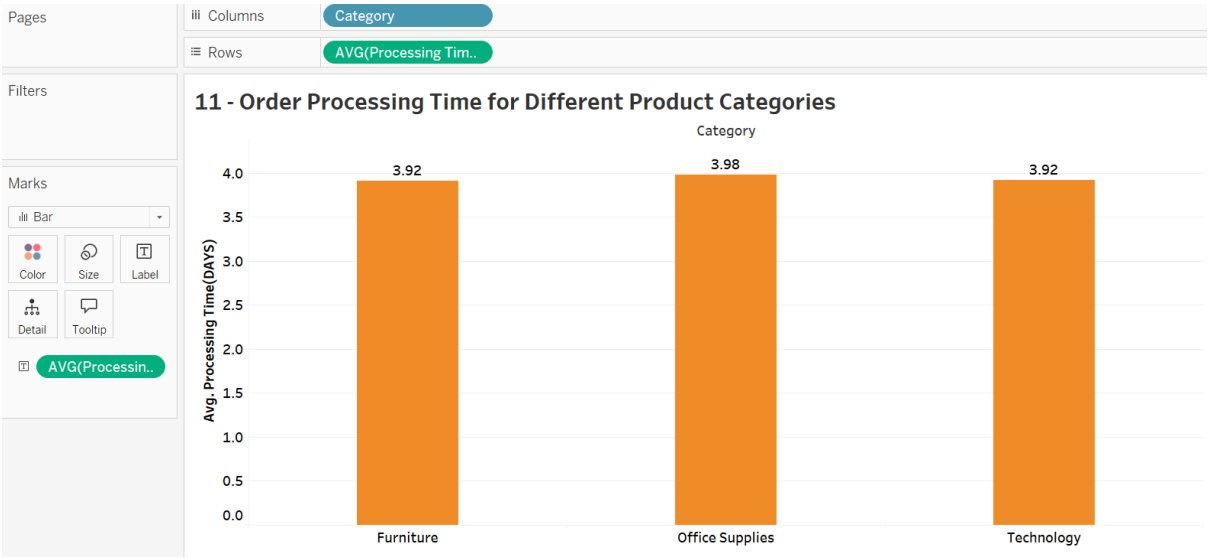
    A bar chart is the best choice for analyzing profit margins across different shipping modes and customer segments because it provides a clear, structured comparison of how profitability varies between these two categorical variables. Each shipping mode gets a separate bar, making it easy to compare profits. Customer segments (Consumer, Corporate, Home Office) can be shown as stacked or grouped bars. If Standard Class shipping is profitable for Corporate but unprofitable for Consumers, the bar chart will clearly show this difference. If First-Class shipping has high profit margins for Corporate but negative margins for Consumers, a stacked bar will show a large green section (Corporate) and a small red section (Consumer loss). If Home Office customers using Second-Class shipping have negative profit margins, it could indicate high shipping costs for low-value orders.



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

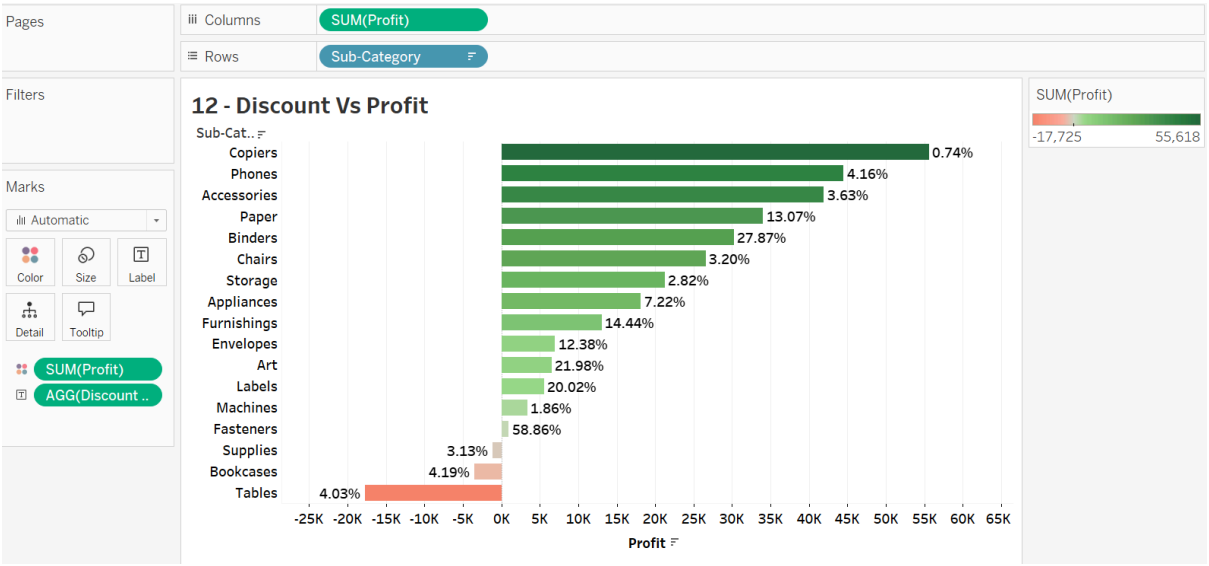
A bar chart is the best choice for visualizing order processing time across different product categories because it provides a clear, structured comparison of the average processing time for each category. If Furniture orders take 5 days on average, while Office Supplies take only 2 days, a bar chart will clearly highlight this difference. If some product categories have significantly higher processing times, a bar chart makes these outliers stand out.

Businesses can use this insight to optimize processing workflows and reduce delays. If Furniture orders take much longer than Technology orders, the company can investigate whether shipping logistics or inventory issues are causing the delay. Bars can be sorted in descending order to make it easy to see the slowest and fastest categories. Color coding can highlight categories with higher-than-average processing times.



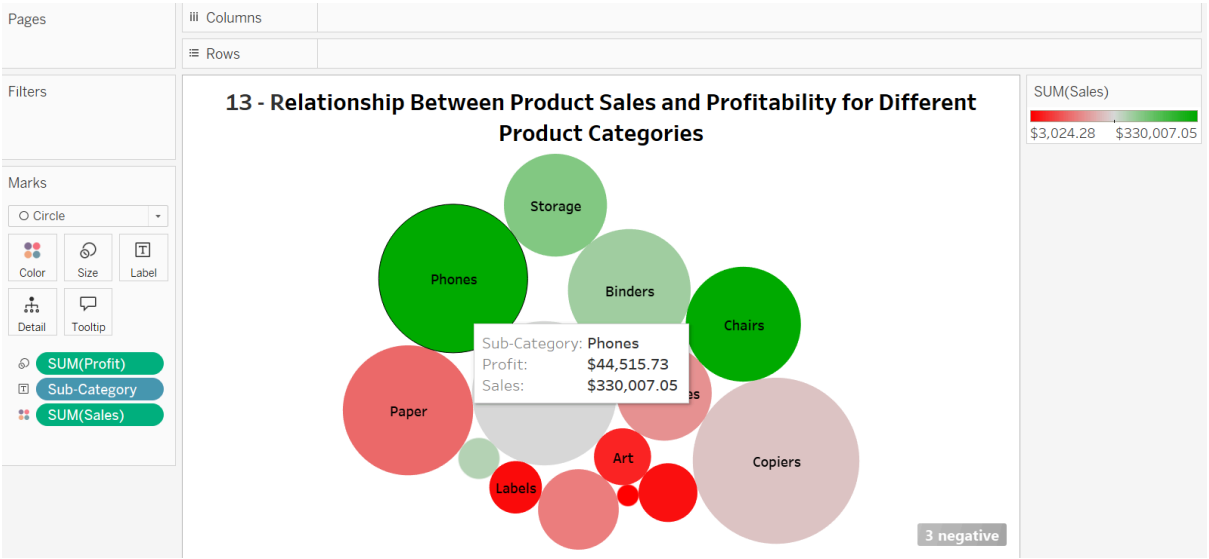
12. How do discounts affect overall profit?

A bar chart is the best choice for analyzing the relationship between discounts and profits because it allows for clear category-wise comparisons while showing how discounts impact profitability across different segments, products, or regions. A bar chart can display discounts and profits side by side for different product categories, regions, or customer segments. If furniture has high discounts but negative profits, it signals a pricing or cost issue. Since discounts and profits are analyzed per product, segment, or region, a bar chart makes it easy to compare across these groups. If a product subcategory with 50% discount shows negative profit, the business may need to adjust pricing or discontinue deep discounts. If Office Supplies with high discounts show declining profits, it might indicate over-discounting with low margins.

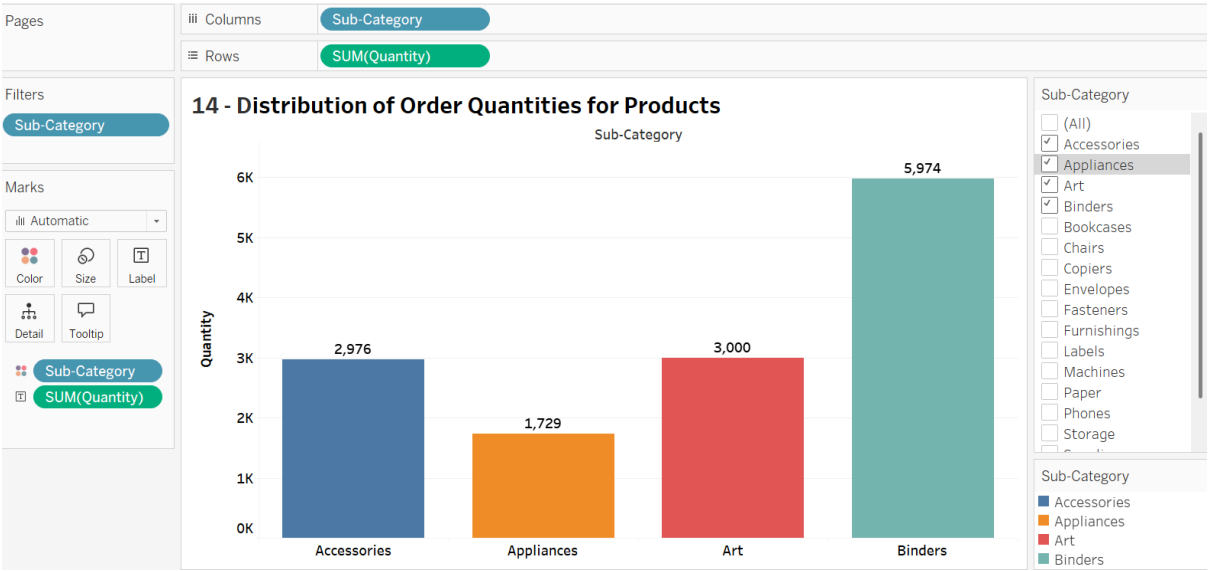


13. Can we visualise the relationship between product sales and profitability for different product categories?

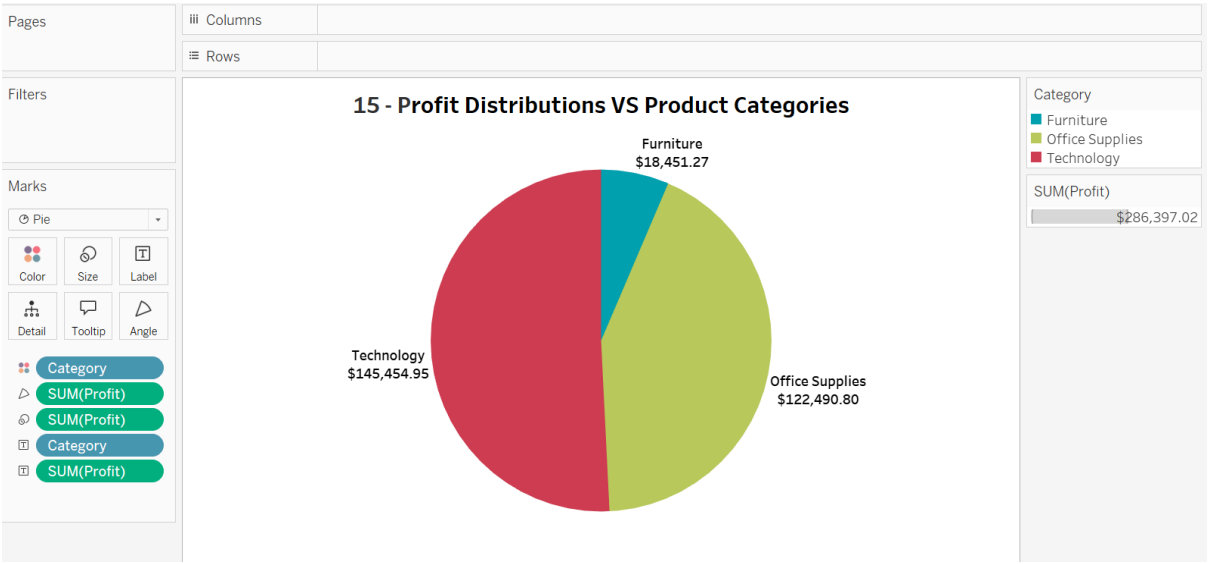
    A bubble chart is the best choice for visualizing the relationship between product sales and profitability across different product categories because it allows for three dimensions of data to be represented simultaneously in a visually intuitive way. If Furniture has high sales but low profit margins, the bubble will be far to the right but close to the bottom. If Technology has high sales and high profit, the bubble will be far right and high up. If a category has large sales volume but poor profitability, this will be immediately visible in a bubble chart. If Office Supplies has large bubbles (high quantity sold) but negative profit (positioned below the X-axis), it signals that discounts or costs are eating into margins. If Technology has moderate sales but very high profit, it will appear higher than other categories despite smaller bubble size. It highlights Outliers & Trends Effectively. Outliers are immediately visible. Categories with similar profitability but different sales volumes are easy to distinguish. If Furniture has a huge bubble (high sales) but is positioned near the X-axis (low profit), it signals low margins or high return rates.



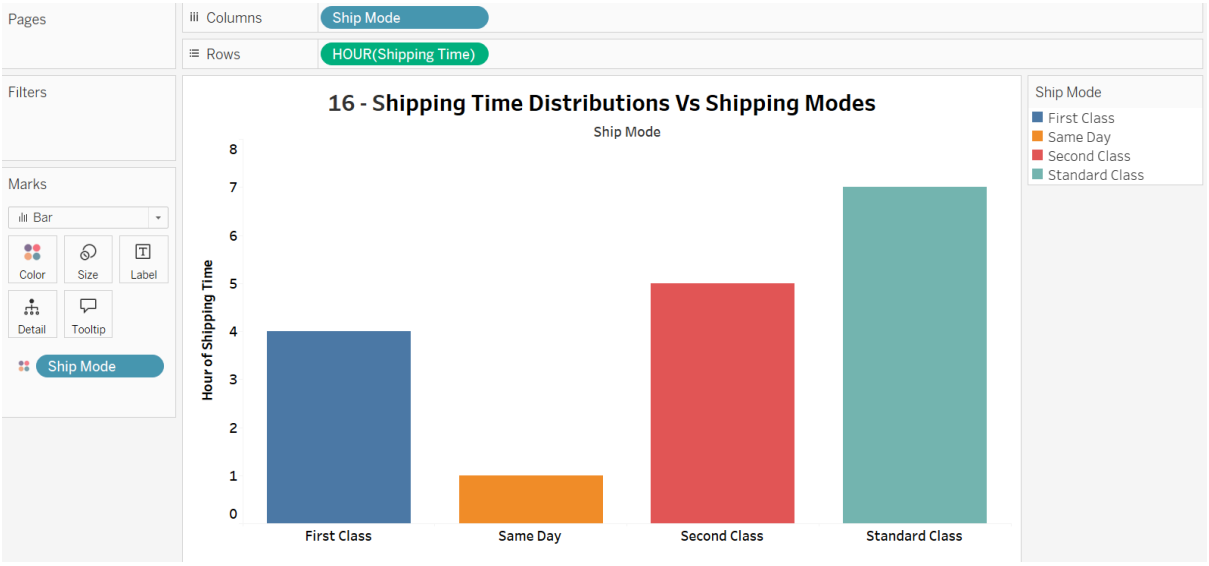
14. What is the distribution of order quantities for products in the dataset?
    A bar chart is the best choice for visualizing the distribution of order quantities for different products because it effectively compares distinct product categories and their order volumes in a clear and structured way. If Laptops have significantly higher order quantities than Chairs, a bar chart will clearly highlight this difference. Since products are distinct categories, a bar chart is the best way to compare their order volumes.



15. How do the profit distributions vary across different product categories?
    A pie chart is a good choice for visualizing profit distribution among different product categories because it effectively shows the proportional contribution of each category to the total profit in a single, easy-to-read visual. A pie chart is designed to show percentages, making it ideal for visualizing how each product category contributes to overall profit. If Technology accounts for 50% of total profit, it will take up half of the pie, clearly highlighting its importance.

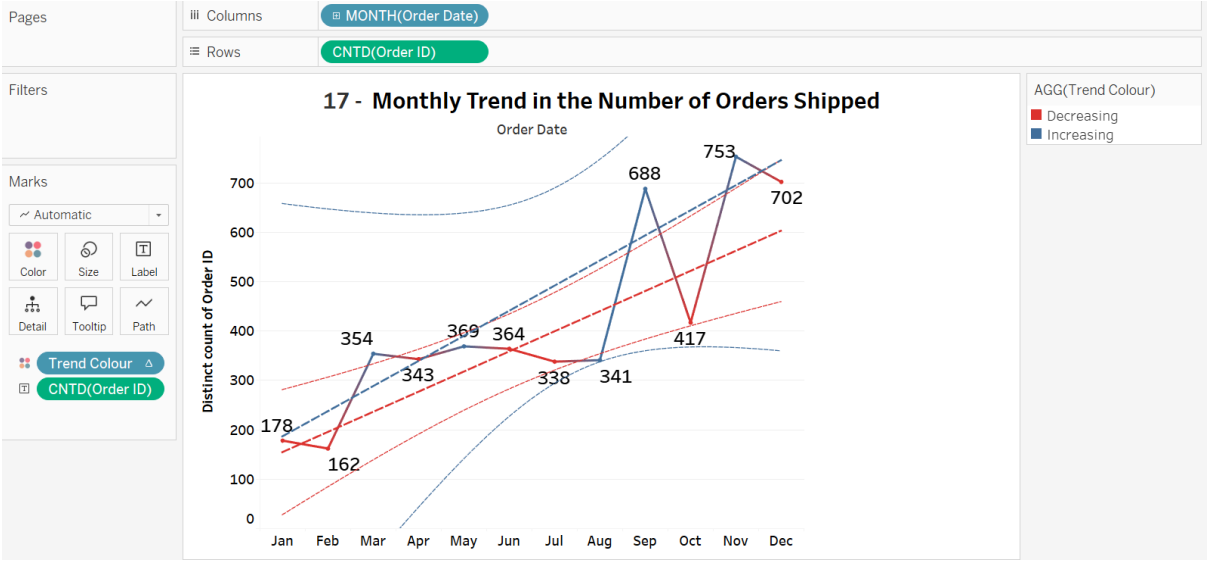


16. Can we compare the shipping time distributions for different shipping modes?
    A bar chart is the best choice for visualizing shipping time distribution across different shipping modes because it allows for clear comparisons between categories and makes it easy to see differences in delivery speed across different methods. A bar chart effectively compares shipping times for distinct categories. It makes it easy to identify which shipping mode has the shortest or longest delivery time. If Same-Day Shipping has an average of 1 day, while Standard Shipping takes 5 days, a bar chart clearly visualizes this difference. A bar chart allows us to see how shipping times vary within each mode by grouping values into bins or displaying averages.

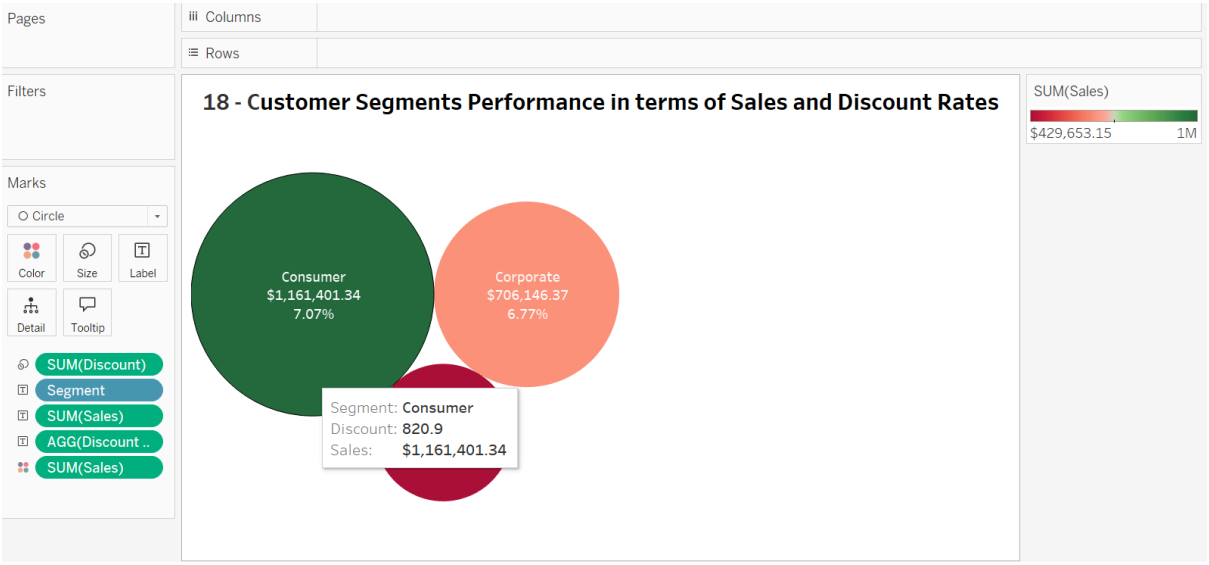


17. What is the monthly trend in the number of orders shipped?
    A line chart with a trend line is the best choice for visualizing the monthly trend in the number of orders shipped because it effectively shows patterns, seasonality, and overall trends over time. Line charts are ideal for continuous data over time, making

them perfect for monthly shipping trends. If orders increase significantly in November and December, a line chart will clearly show this pattern. A line chart highlights patterns in shipping volume, such as: Seasonal spikes (e.g., holiday shopping seasons), Periods of decline (e.g., slow months). A bar chart wouldn't work well here, as it would focus on individual months rather than overall trends.

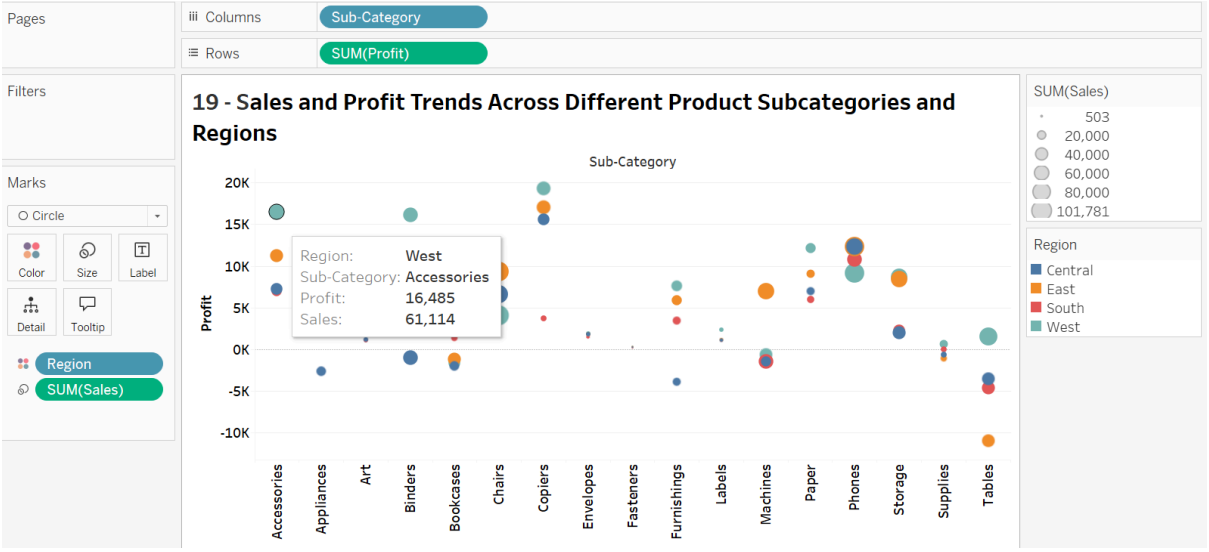


18. How do different customer segments perform in terms of sales and discount rates?
A bubble chart is an excellent choice for visualizing customer segment performance based on sales and discount rates because it allows for comparisons across three dimensions in a single chart. Best for Showing Three Variables Simultaneously. Unlike bar or line charts, a bubble chart visualizes three metrics at once. If the Consumer segment has high sales but also high discounts (large bubble), it might indicate lower profitability. Helps Identify Trade-offs Between Sales and Discounts. If a segment has high sales but large discounts, it might mean lower profits. If another segment has lower sales but smaller discounts, it might indicate higher profitability. If Corporate customers generate high sales but require large discounts, while Small Business customers generate steady sales with lower discounts, a bubble chart makes this comparison easy. Highlights Outliers and Performance Differences
A bubble chart makes it easy to spot which customer segment gets the most discounts or drives the highest sales.



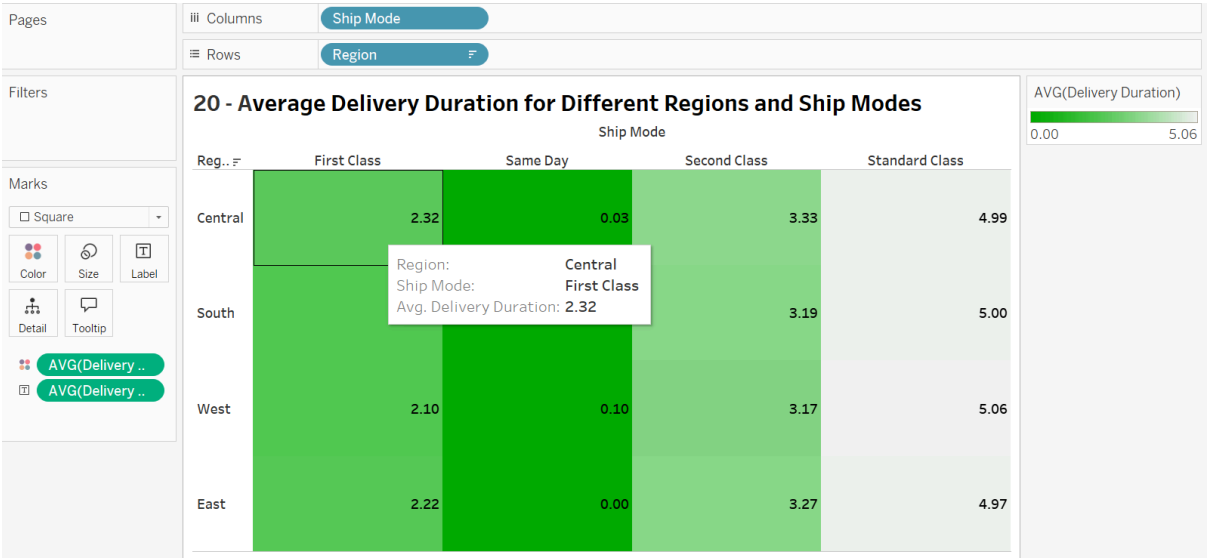
19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?

A scatter plot is the best choice for visualizing sales and profit trends across different product subcategories and regions because it effectively shows the relationship between two continuous variables while allowing category-based differentiation. If a subcategory (e.g., Technology) has high sales but low profit, a scatter plot will clearly highlight it as an outlier. Identifies High-Performing vs. Low-Performing Subcategories/Regions. A scatter plot helps distinguish:Profitable products/regions (High sales & high profit → top-right), High sales but low profit (or loss-making) products/regions (Bottom-right) and Low sales, low profit (underperformers) (Bottom-left). If Office Supplies has high sales but low profit in a specific region, the scatter plot will show that as a point in the bottom-right section. Detects Outliers and Profitability Issues. Some subcategories may have high sales but negative profits, indicating high discounts or high costs. A scatter plot makes it easy to identify which products or regions are not profitable despite strong sales.



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

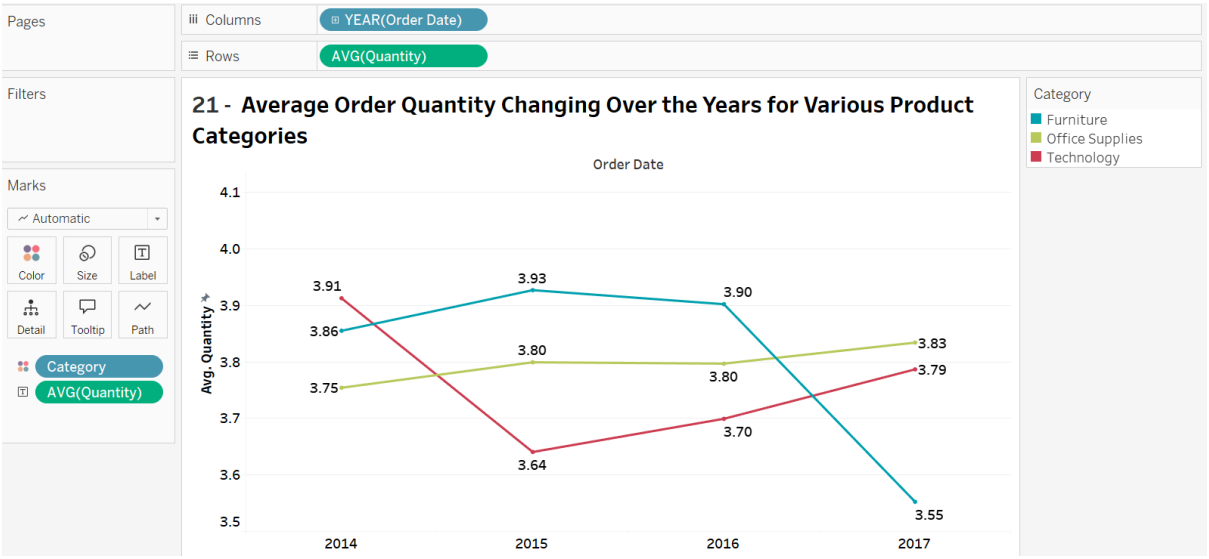
A treemap is the best choice for visualizing average delivery duration across different regions and shipping modes because it allows you to compare multiple categories in a structured, space-efficient way. A treemap visually organizes data into nested rectangles, making it ideal for showing how shipping modes perform within different regions. If Standard Class in the South region has the longest delivery time, its rectangle will be largest and/or darkest. Easily Highlights Largest and Smallest Delivery Time.



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

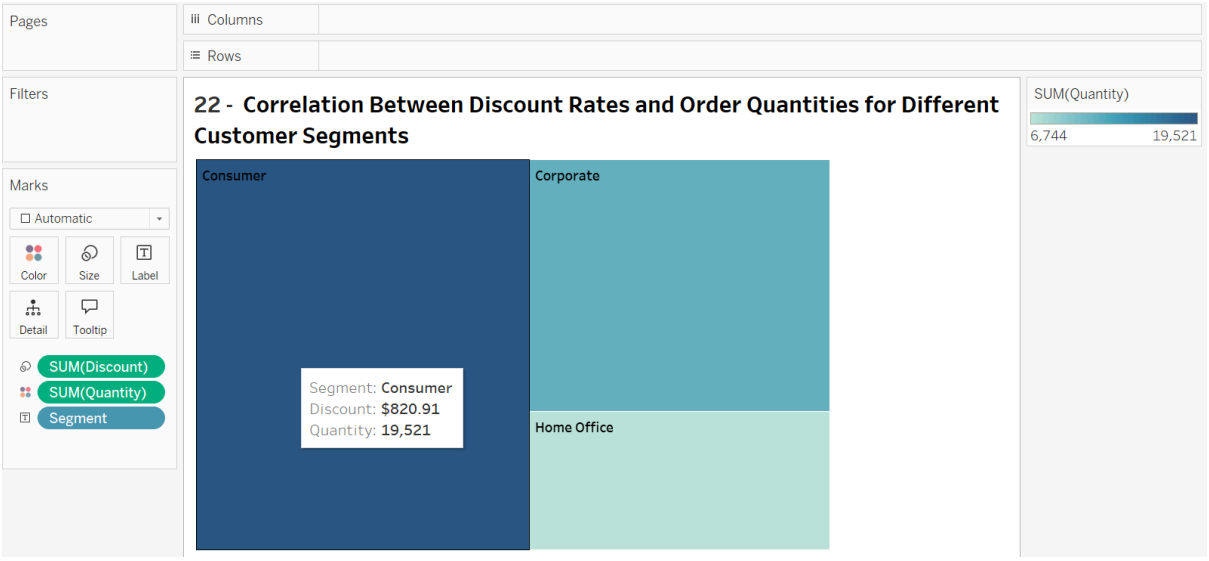
A line chart is the best choice for visualizing how average order quantity changes over the years for different product categories because it effectively shows trends over time and comparisons across multiple categories. A line chart is designed to show how values evolve over time.

It helps identify: Increasing trends → Which product categories are seeing growth, Declining trends → Which product categories are losing demand, Seasonal fluctuations → Do certain categories have peaks at specific times. Different product categories can be represented by different lines, making it easy to compare trends. If one category is growing while another is declining, this pattern is immediately clear.



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

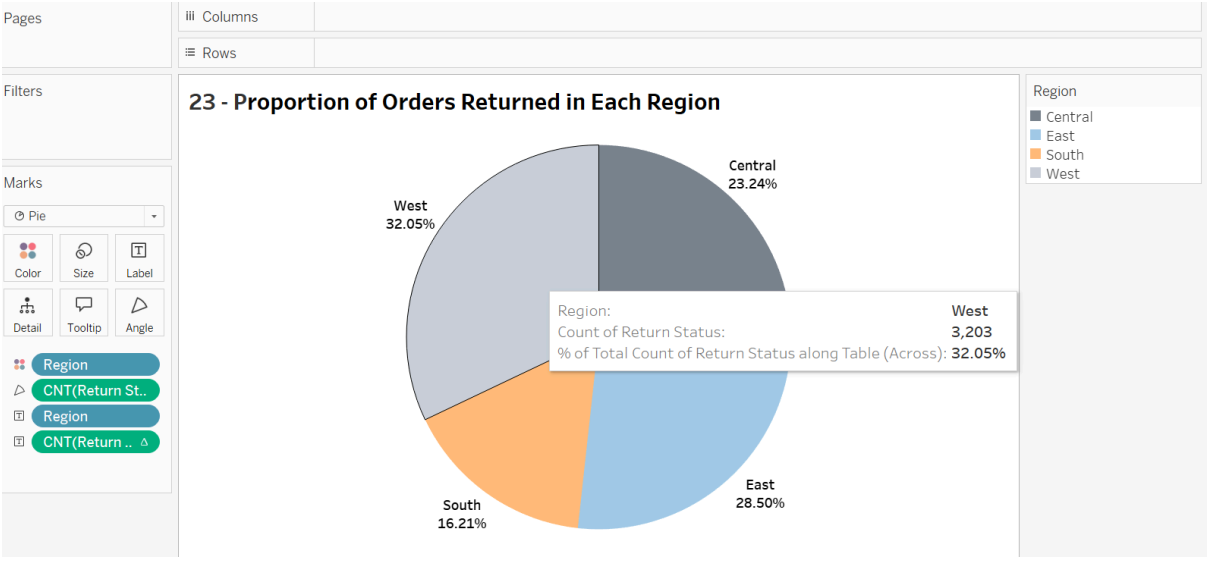
    A treemap is the best choice for visualizing the correlation between discount rates and order quantities for different customer segments because it allows for an efficient, hierarchical comparison while displaying the impact of discount rates across multiple customer groups. A treemap groups data by customer segments and further divides it by discount levels, making it easy to see how different segments respond to discounts. It answers Which customer segment has the highest order quantity at high discount rates, Do higher discounts lead to more orders in specific segments. Highlights the Impact of Discounts on Order Volume.



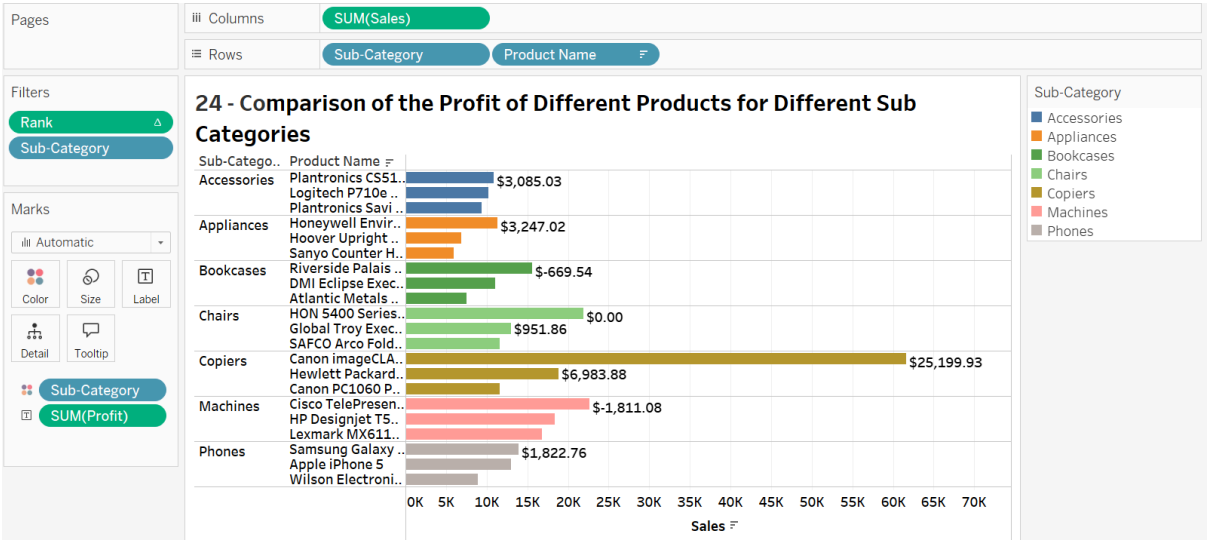
23. What is the proportion of orders returned in each region within the Superstore dataset?

    A pie chart is the best choice for visualizing the proportion of orders returned in each region because it effectively shows how different regions contribute to the total

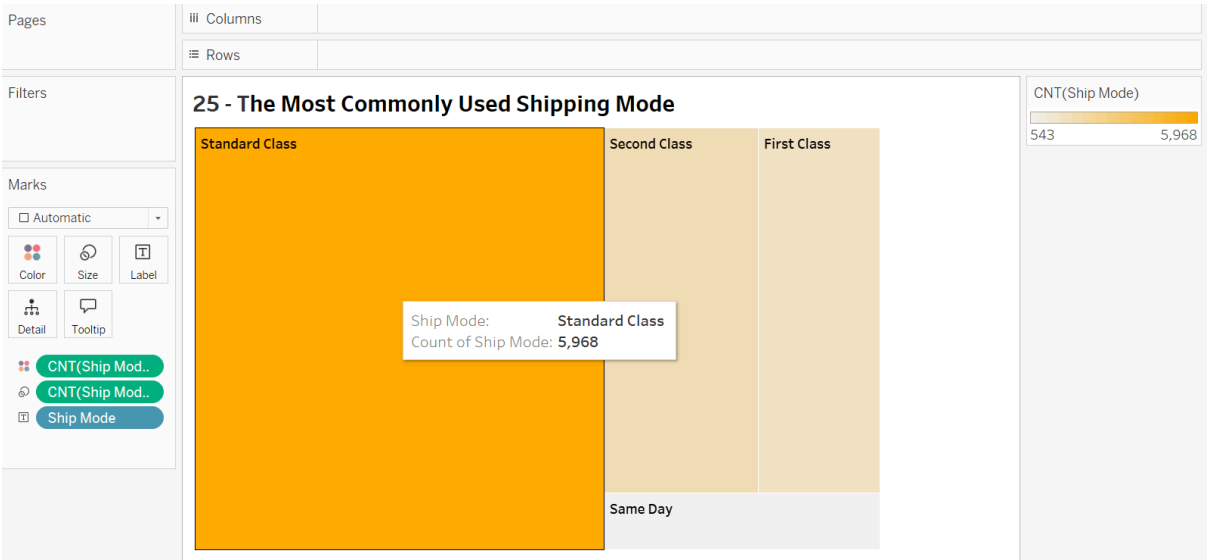
returns as percentages of the whole. A pie chart represents each region's share of total returns as a percentage.
It helps to identify Which region has the highest return rate and How do returns compare across regions. Clearly Highlights Regional Differences in Returns Color coding helps visually separate each region.Easy to Interpret at a Glance. A pie chart is simple and intuitive, making it great for quick insights.
Viewers can instantly identify which regions have the highest or lowest return percentages.



24. Can you compare the profit of different products for different subcategories?
A bar chart is the best choice for visualizing the comparison of profit across different products within sub-categories because it effectively displays individual product performance while maintaining clarity within each sub-category. A bar chart clearly shows profit differences between products within the same sub-category. If the "Office Chairs" sub-category has multiple products, the bar chart will show which specific chair models generate the most profit. Makes It Easy to Spot Trends & Outliers. Tall bars indicate high profits, while short bars (or negative bars) indicate losses. Outlier products that perform exceptionally well or poorly become immediately noticeable.

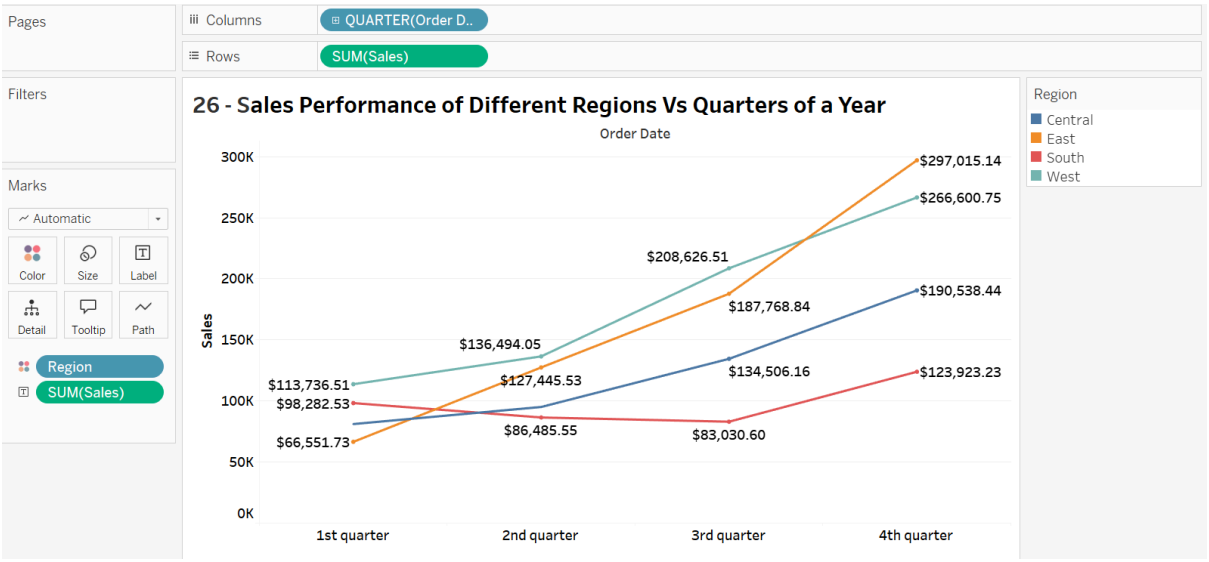


25. Which shipping mode is the most commonly used in the Sample Superstore dataset?
A treemap is the best choice for visualizing the most commonly used shipping modes because it allows for a hierarchical and proportional comparison, making it easy to see which shipping mode is used the most. A treemap represents each shipping mode as a rectangle, where the size of each rectangle corresponds to the frequency of usage. If "Standard Class" accounts for 60% of shipments, its rectangle will be much larger than those for other modes like "Same Day" or "First Class". A treemap avoids clutter, which is a common issue with pie charts when there are multiple categories. It provides a quick visual representation of usage without unnecessary complexity.



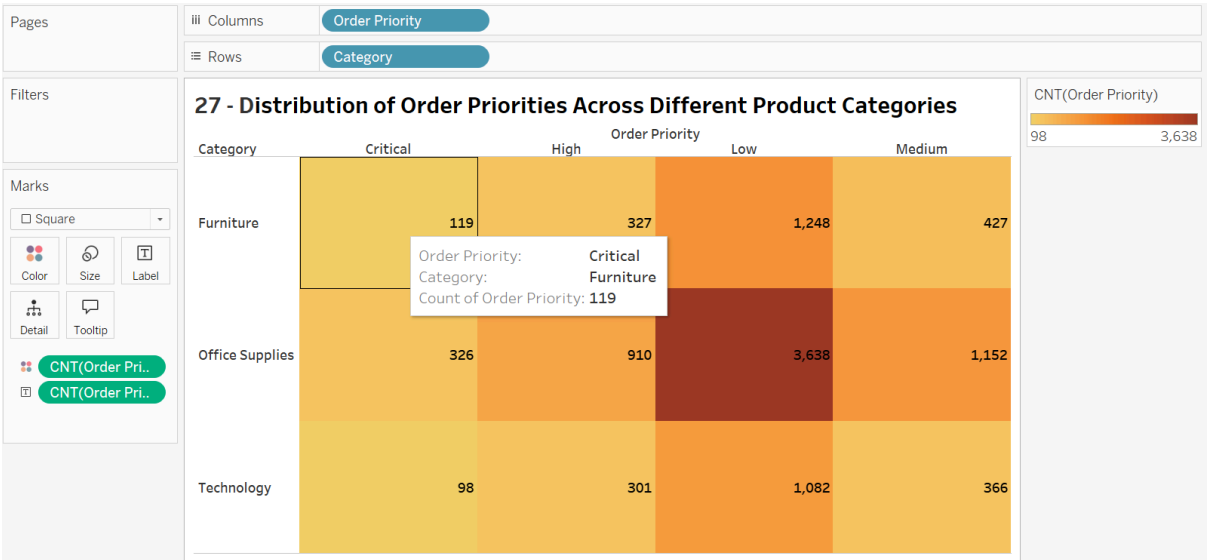
26. How does the sales performance of different regions evolve throughout the quarters of a year?
A line chart is the best choice for visualizing the sales performance of different regions across quarters because it effectively displays trends, comparisons, and seasonality over time. A line chart connects data points over time, making it easy to see sales growth, declines, or fluctuations across quarters. If sales in the West region

peak in Q4 and drop in Q1, the line will clearly show this trend, helping businesses make data-driven decisions. Easy Comparison Between Regions. A line chart reveals patterns that might indicate Peak seasons and Slow periods.



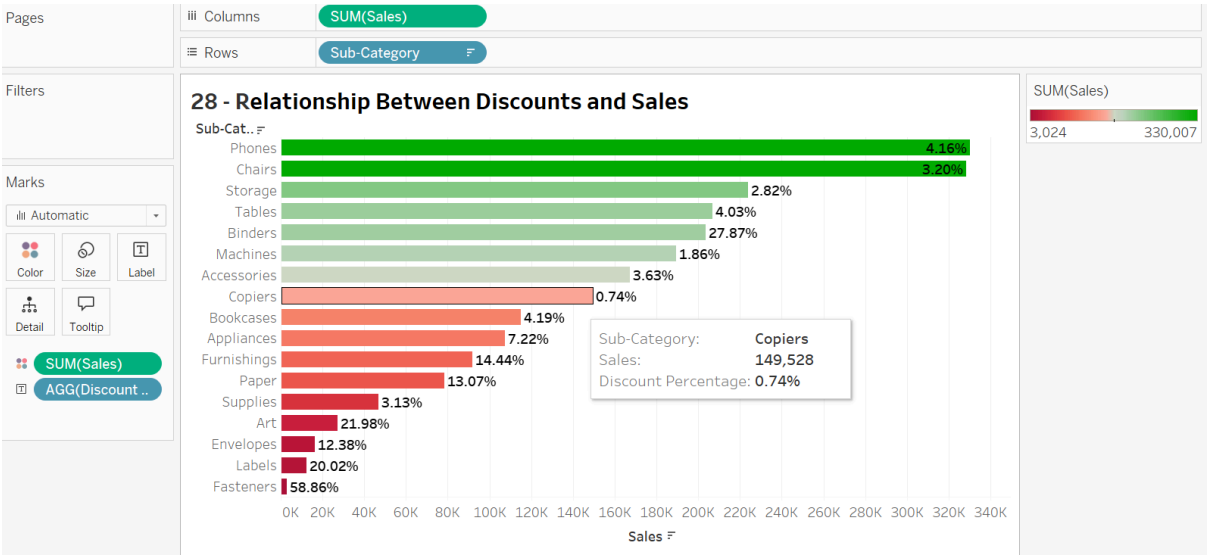
27. What is the distribution of order priorities across different product categories?

A treemap is the best choice for visualizing the distribution of order priorities across product categories because it effectively represents hierarchical and proportional data while allowing for easy comparisons. Efficiently Displays Hierarchical Data. Treemaps allow drilling down from categories to order priorities within them. Each product category is a large rectangle, and within it, smaller rectangles represent different order priorities. Avoids Clutter Compared to Other Charts. If multiple product categories have varying order priorities, a bar chart would look cluttered, while a treemap neatly organizes the information.



28. What is the relationship between discounts and sales?

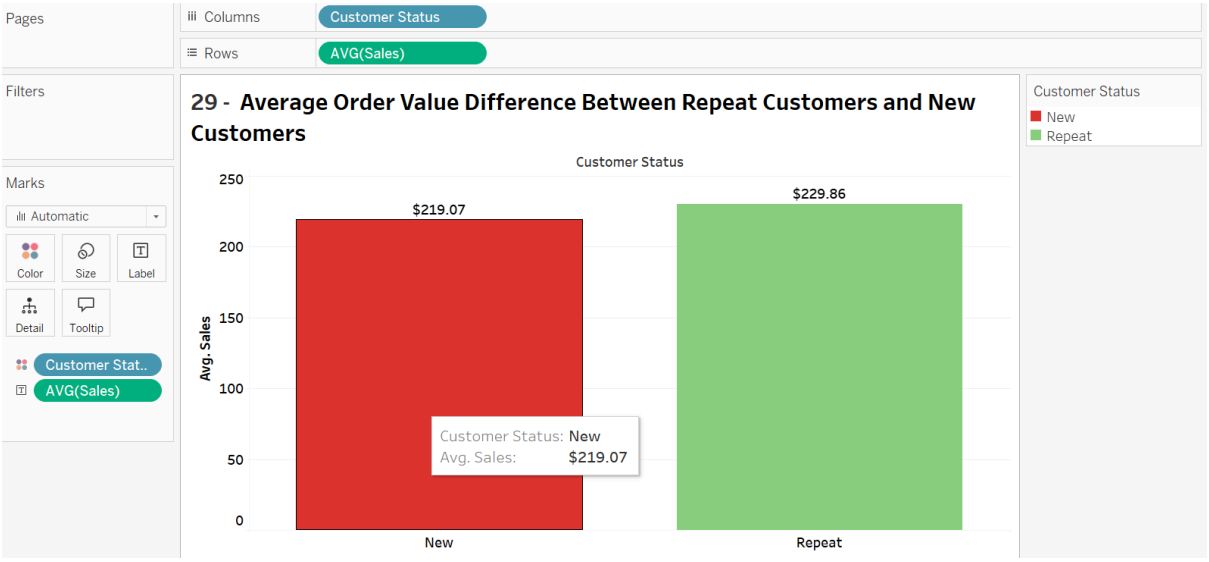
A bar chart is the best choice for visualizing the relationship between discounts and sales because it clearly shows how different discount levels impact total sales. Best for Comparing Sales Across Discount Levels. A bar chart allows direct comparison of sales at different discount percentages. If sales increase when discounts go from 0% to 20% but drop at 50%, a bar chart makes this pattern easy to see. Clearly Highlights Sales Impact at Different Discount Levels. Each bar represents sales at a specific discount level, making it easy to compare. The height of the bars shows how sales volume changes as discounts increase. Helps Identify the Optimal Discount Range. A bar chart reveals whether there is a "sweet spot" where moderate discounts drive the highest sales. If a specific discount consistently leads to high sales, businesses can adjust their pricing strategies accordingly.



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

A bar chart is the best choice for visualizing the difference in average order values (AOV) between repeat and new customers because it provides a clear and direct comparison. A bar chart clearly shows the gap in spending behavior between new and repeat customers. If repeat customers spend $200 per order while new customers spend $120, the bar for repeat customers will be taller than the one for new customers. Simple and Easy to Interpret. Only two categories (New Customers vs. Repeat Customers), so a bar chart keeps it clean and easy to read. Color coding can highlight higher vs. lower values.

Shows Business Impact of Customer Retention. If repeat customers spend significantly more per order, businesses can focus on loyalty programs or retargeting efforts.



30. What is the geographical distribution of returns and its impact on overall profitability?
A map visualization is the best choice for analyzing geographical distribution of returns and their impact on profitability because it provides a clear, location-based overview of where returns are happening and how they affect profits. Best for Location-Based Analysis. A map helps identify regions with the highest return rates and their profitability impact. It shows Which regions have the most product returns and How do high-return regions affect overall profits. If California has a high number of returns, the map will highlight it with a darker color, making it easy to spot patterns. A color-coded map (choropleth or symbol map) makes it easy to distinguish high and low return regions. Helps in Profitability Analysis By layering returns over profit data, businesses can see if high-return areas also have lower profitability.