

# 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 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.

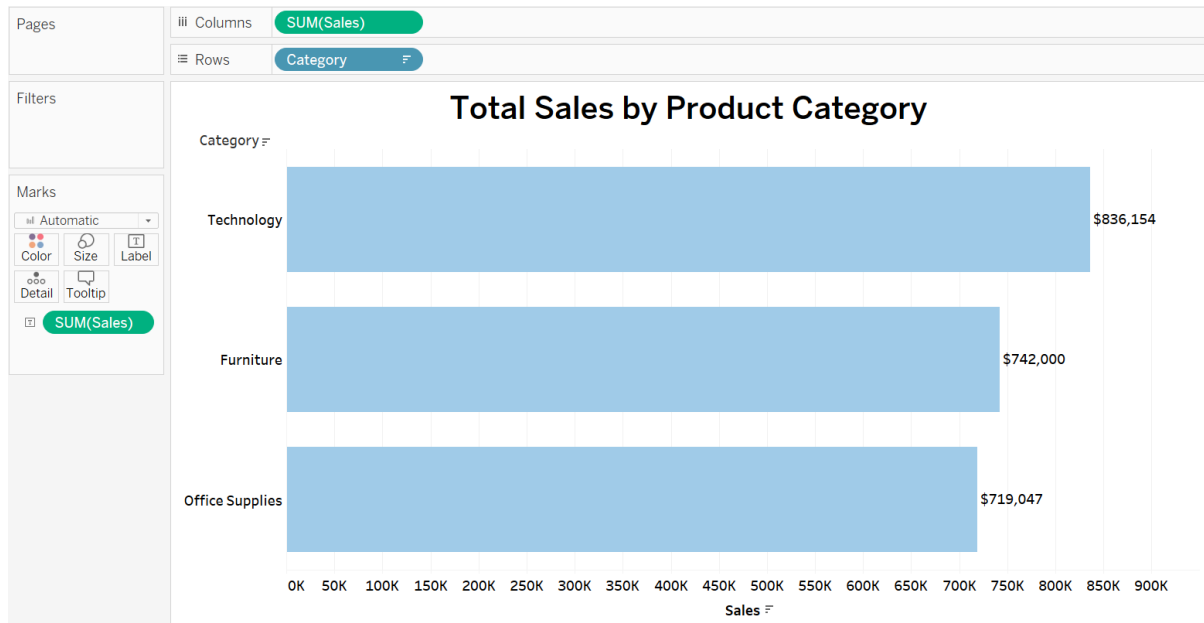
By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

**Good luck!**

## Questions:

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

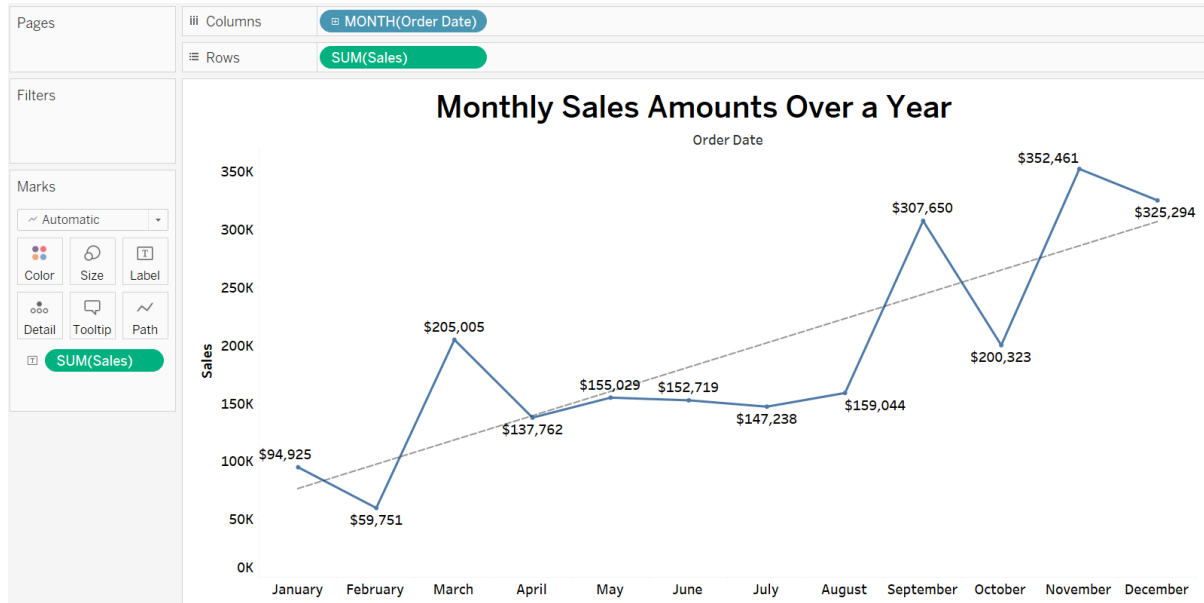
**Ans:**



The horizontal bar chart is particularly useful when the labels for the categories are relatively long, as they are in this case (Technology, Furniture, Office Supplies). This layout makes it easier to read the category names than a vertical column chart, where the names might be condensed or angled, making them less legible. The horizontal orientation allows viewers to quickly rank the categories by length, clearly seeing which category has the highest or lowest sales. It's a natural way to represent ordinal data (data that has a clear order), making it immediately obvious that Technology has the highest sales, followed by Furniture and Office Supplies. Furthermore, the chart's scalability makes it suitable for various display mediums and the inclusion of additional data points.

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

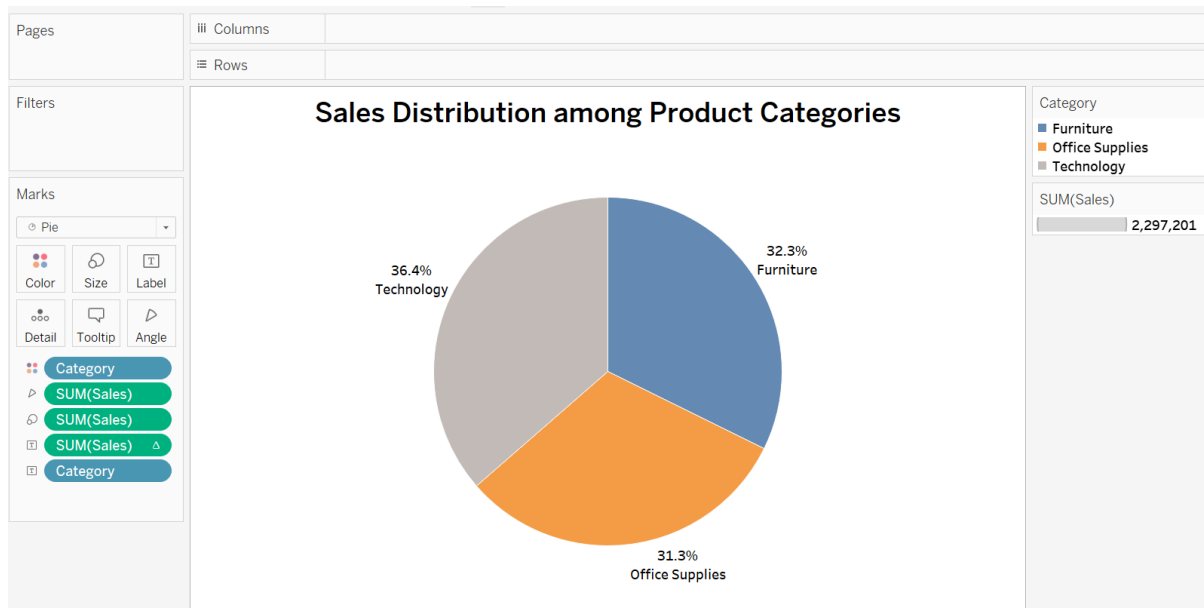
Ans:



The line chart is an ideal visualisation for depicting how monthly sales amounts change over a year, as it effectively displays the significant fluctuations and trends in the data. It provides a clear narrative of sales performance, showing that sales start low in the early months, rise to a peak in March, then fall into a trough during the summer, before reaching the highest peak in September. Subsequently, there is a dip and another rise as the year ends. The trend line indicates an overall increase in sales over the year. This type of chart is particularly adept at illustrating these temporal changes, offering a visual story of high and low sales activity periods, and confirming the dynamic nature of monthly sales over the year.

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

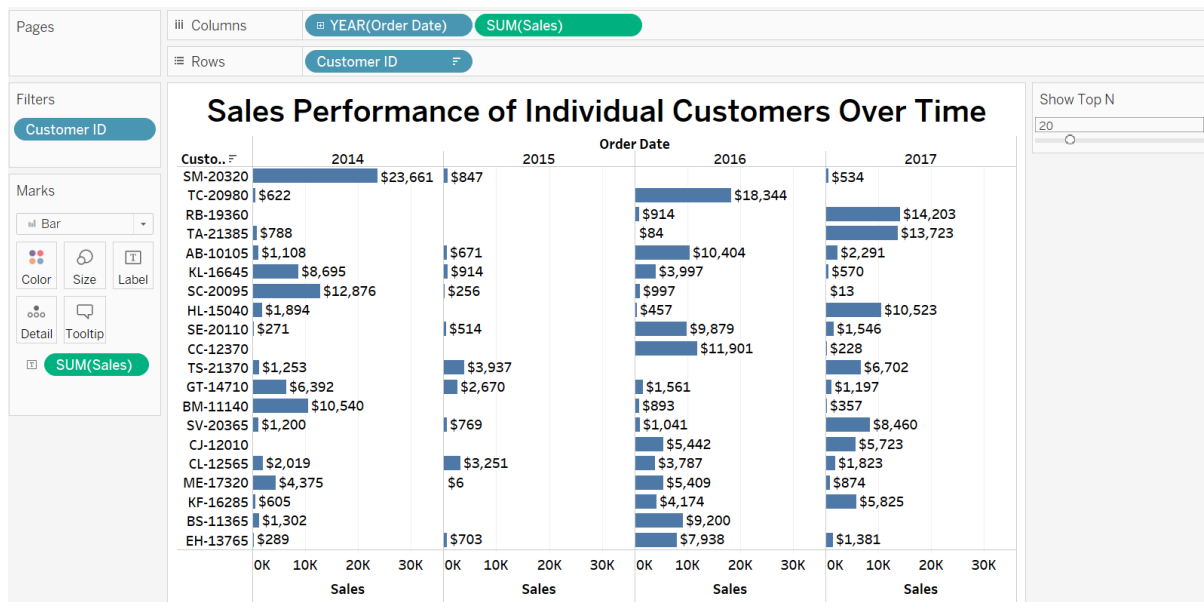
Ans:



The pie chart is an effective visualisation tool for illustrating the sales distribution among product categories, as it provides an immediate sense of the market share each category holds within the total sales volume. In this chart, Technology represents the largest share at 36.4%, with Furniture not far behind at 32.3%, and Office Supplies close as well at 31.3%. This visual format excels when the goal is to convey the relative size of each category in a whole, allowing viewers to compare the segments intuitively. The percentages assigned to each 'slice' give a precise understanding of their contribution, making the pie chart a fitting choice for presenting this type of data.

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

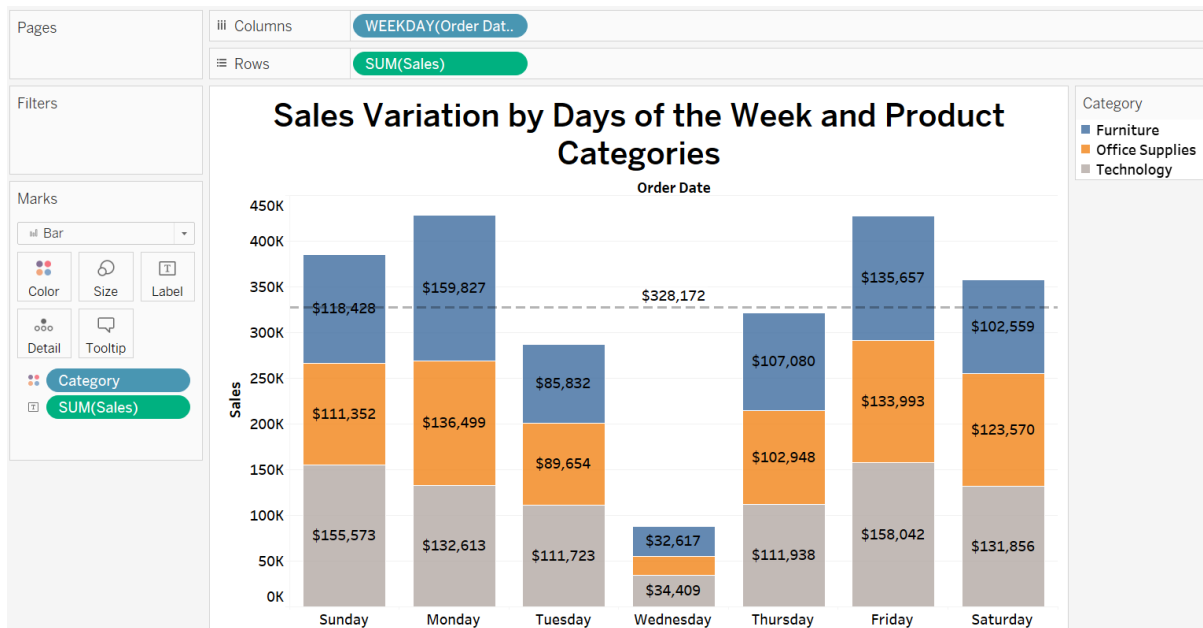
Ans:



The horizontal bar chart holds distinct advantages for displaying sales performance across individual customers, particularly when tracking this data over time. It excels in offering clear comparisons of sales figures between different customers, facilitated by its layout which tends to be more readable than vertical bars, especially when dealing with longer labels such as customer IDs. Moreover, the horizontal layout can effectively represent time progression, with sales trends for each customer becoming easily traceable along the horizontal axis. The addition of a 'Show Top N' parameter, utilising 'Customer ID' as a filter, significantly enhances the chart's interactivity, enabling users to dynamically focus on top-performing customers, thus tailoring the analysis to specific needs without clutter, making the chart both informative and adaptable. This combination of clarity, readability, and interactivity makes the horizontal bar chart the best choice for this type of data.

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

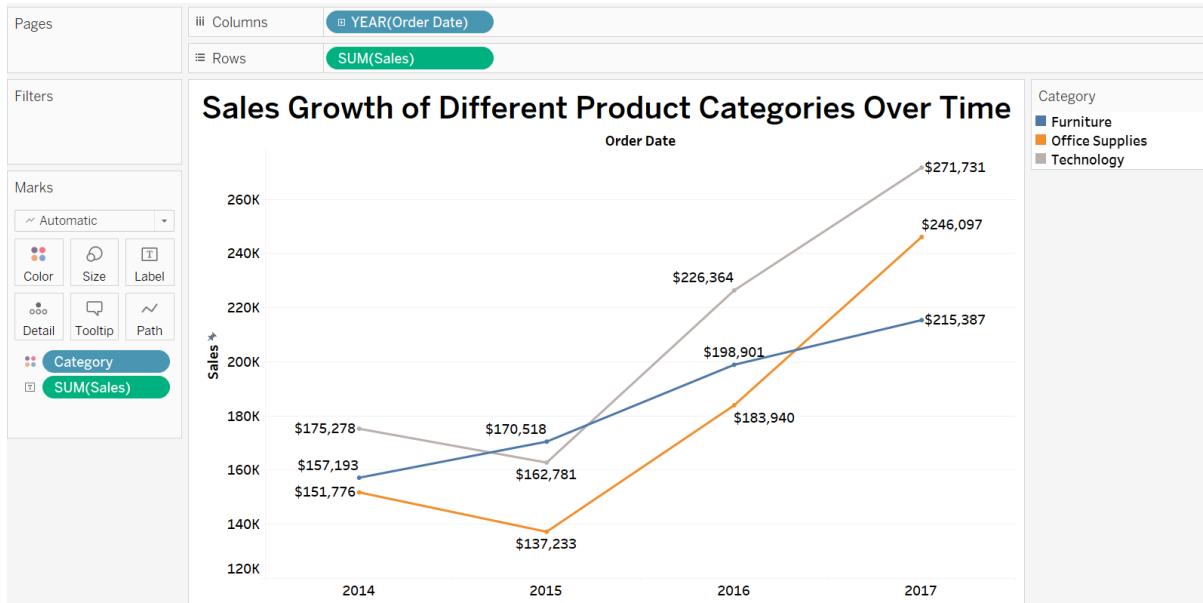
Ans:



The stacked bar chart is an effective visualisation choice for comparing the sales variations across different days of the week and product categories because it allows for a clear and comprehensive comparison of total sales as well as category-specific performance within the same chart. Each bar represents a day of the week, offering a straightforward comparison of total sales per day, while the segments within each bar represent the sales for individual categories, enabling an analysis of each category's contribution to the day's total. This dual-layered approach allows for an at-a-glance assessment of both overall and category-specific trends, such as identifying peak sales days or understanding which product categories dominate on certain days. The colour-coding for each category enhances the visual differentiation, making it easier to distinguish between them. A reference line has been introduced in the chart to show average sales across the week. It reveals that Tuesday, Wednesday, and Thursday performed below average, while the rest of the days performed well above average, particularly on Monday and Friday. This chart type is particularly adept at showing how different segments compose the whole on a categorical and temporal basis, making it a powerful tool for temporal and categorical data analysis.

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

Ans:

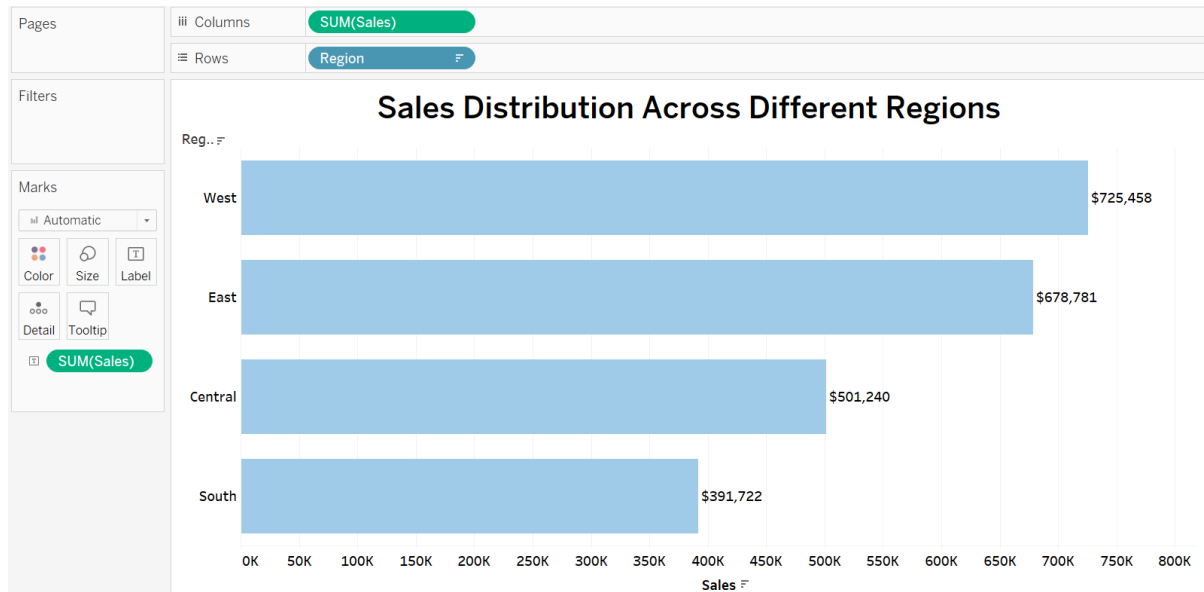


The line chart is an optimal choice for visualising the sales growth of different product categories over time because it clearly illustrates trends and the rate of change in a way that is easy to follow and interpret. Each line represents a product category, with the progression over time showing how sales have increased or decreased. The distinct colours for each category allow for quick comparison of their performance, and the continuous nature of the line effectively communicates the trajectory of growth. This type of chart is particularly suited for time-series data where the focus is on understanding how a value evolves over a period. It is also efficient in highlighting whether there are any significant spikes or dips, corresponding to external factors or seasonal variations, making it invaluable for time-based trend analysis in sales data. This chart depicts a marked increase in sales from 2014 to 2017 across Furniture, Office Supplies, and Technology categories. Notably, Technology leads with a sharp rise, especially between 2016 and 2017, indicating a booming demand. Office Supplies show steady, consistent growth, suggesting stable market demand. In contrast, Furniture demonstrates more modest growth, which could signal a need for strategic initiatives to boost its sales performance. These trends highlight the potential of Technology as a primary revenue driver and suggest a re-evaluation of the Furniture category to enhance its growth trajectory.



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

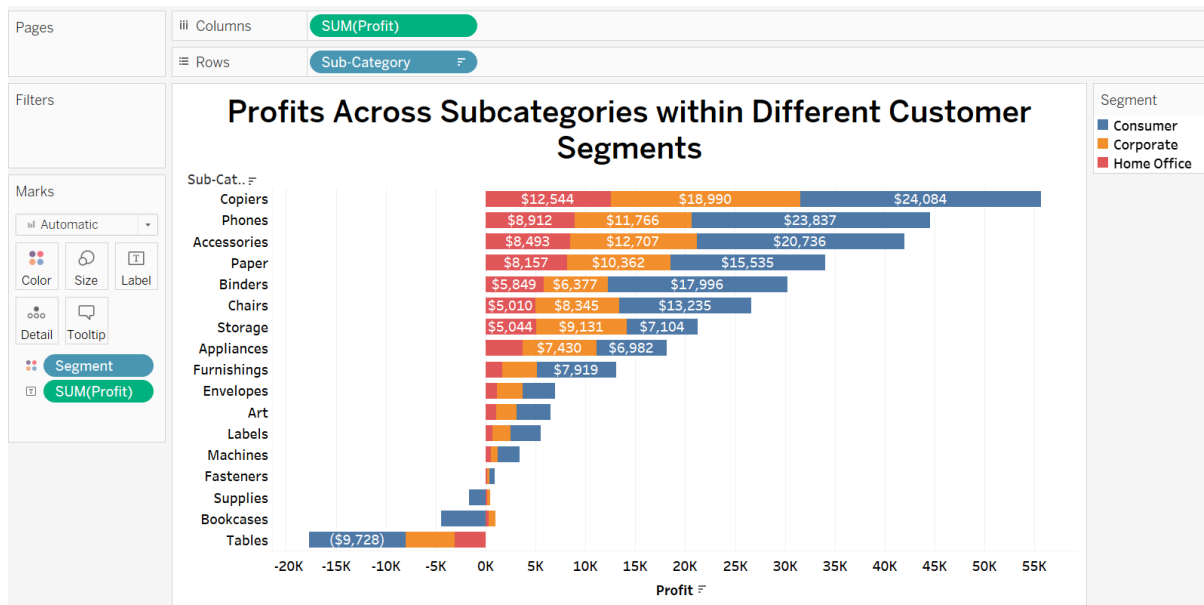
Ans:



The horizontal bar chart is a strong choice for visualising sales distribution across different regions due to its straightforward layout that aligns with how people naturally read text: from top to bottom and left to right. This format allows for easy comparison of sales figures across regions with a quick scan. It's particularly effective when the data labels (such as the names of regions) are longer, as it provides ample space for text without clutter. The horizontal orientation also allows for a longer scale which can display large numbers more comfortably, making it easier to detect differences in sales values. Moreover, the descending or ascending order of the bars can quickly convey the ranking of regions by sales volume, providing a clear hierarchy of performance. This chart type is ideal for comparing a few distinct categories, making it the best option for effectively visualising the sales distribution. It shows that the West region leads with the highest sales at approximately \$725,458, followed by the East with around \$678,781. The Central region comes next with sales of about \$501,240, and the South region has the lowest sales, at roughly \$391,722.

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

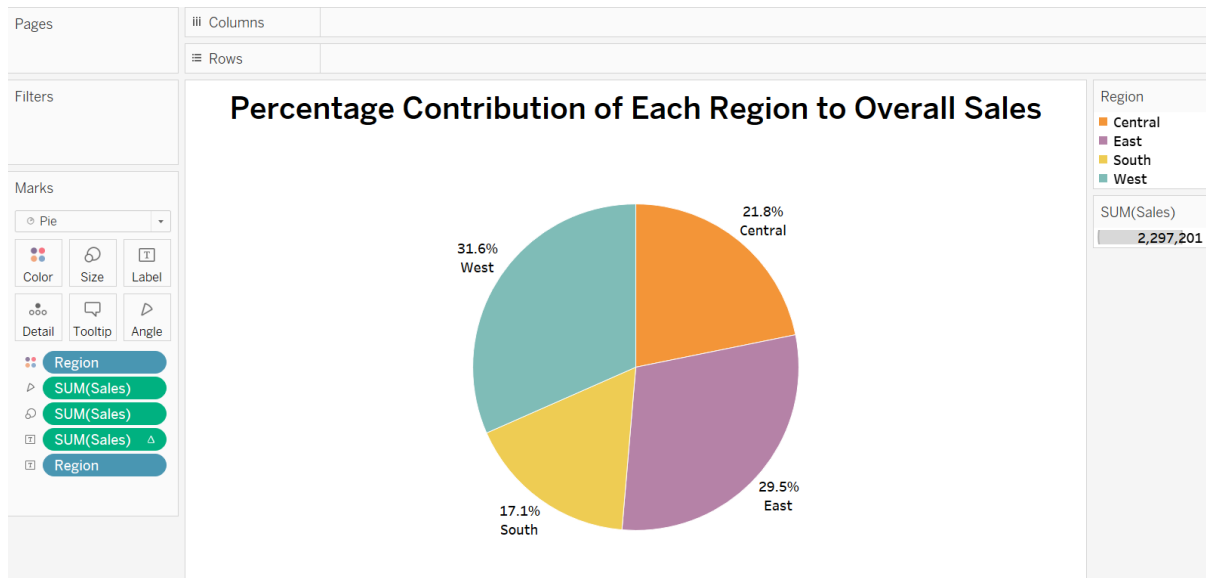
Ans:



The horizontal stacked bar chart is an effective visualisation choice for its ability to segment and compare profits by subcategory and customer segment. By using distinct colours for each customer segment (Consumer, Corporate, Home Office), it allows for detailed analysis of profit contributions within each subcategory. The layout enables comparative analysis, showing the total profits per subcategory and each segment's contribution. It also adeptly handles negative values, clearly displaying any loss-making subcategories with bars extending leftward from the zero line. Additionally, the subcategories are ranked by profitability, facilitating quick identification of the most and least profitable areas. The chart indicates that 'Copiers' are highly profitable, particularly in the Corporate segment, while 'Tables' and 'Bookcases' incur losses across all segments. The Consumer segment shows diverse profitability, especially in 'Phones' and 'Accessories', suggesting a robust market. In contrast, the Home Office segment yields considerable profits in 'Chairs' and 'Storage', pointing to demand for home office furnishings. These trends highlight the potential for tailored marketing strategies and the need to address the underlying issues in the underperforming subcategories.

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

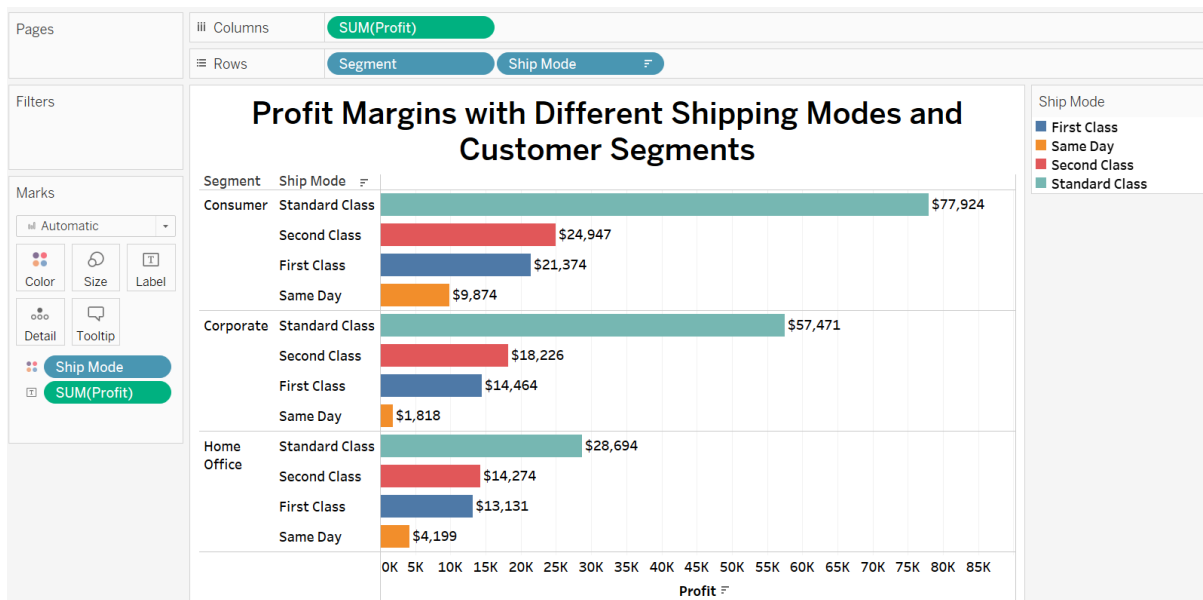
Ans:



The pie chart is a strong visual tool for representing the percentage contribution of each region to overall sales because it intuitively illustrates proportions. Each 'slice' is proportional to the region's sales in relation to the total, offering a clear visual comparison between segments. This is particularly effective for a quick understanding of the distribution and dominance of sales across regions. The pie chart excels when the data comprises a relatively small number of categories, as is the case with the regions here. Additionally, the direct labelling of percentages on the chart reinforces the visual representation, allowing for immediate recognition of each region's share. This pie chart effectively communicates the percentage contribution of each region to overall sales, highlighting the West as the leader at 31.6%, followed by the East at 29.5%, Central at 21.8%, and the South at 17.1%.

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

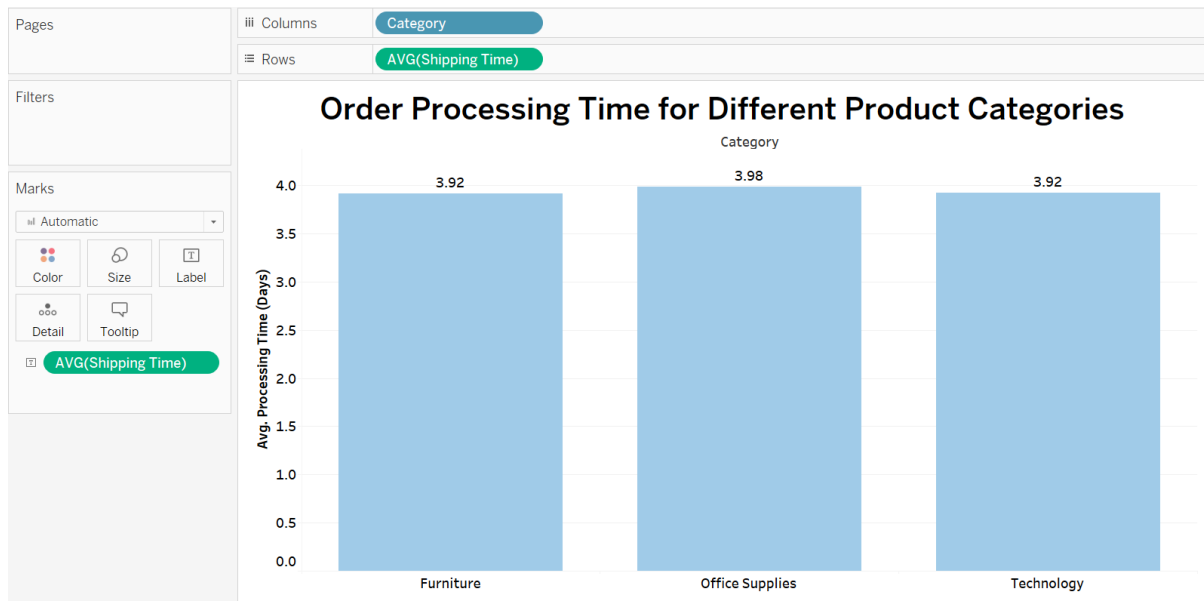
**Ans:**



The grouped horizontal bar chart effectively illustrates profit margins by shipping modes within distinct customer segments. Each horizontal bar, representing a specific segment, is divided by colour to indicate the various shipping modes, making it easy to discern each mode's contribution to overall profits. The chart's clarity is enhanced by its horizontal layout, which accommodates long labels and multiple categories without sacrificing readability. With consistent and distinct colour-coding, the chart facilitates at-a-glance comparisons and swift identification of trends across segments. The visualisation highlights that Standard Class shipping is the most profitable across all segments, particularly within the Consumer segment which shows the highest margins. Conversely, Same Day shipping is the least profitable, suggesting potential issues with demand or cost-efficiency that could be addressed to optimise profitability.

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

**Ans:**



The vertical bar chart is well-suited for displaying the average order processing time across different product categories, with each bar's height indicating the time taken from order to shipment. The utilisation of the calculated field 'Shipping Time', representing the day difference between 'Order Date' and 'Ship Date', ensures that the data is specific and relevant to the processing period. This chart type efficiently conveys the average times, facilitating direct comparisons between the categories of Furniture, Office Supplies, and Technology. The clarity and simplicity of the bar chart, combined with precise numeric labels, make it an effective visualisation tool for conveying the nuances of order processing times within the dataset. This chart shows a negligible difference in average shipping times among Furniture, Office Supplies, and Technology, all hovering just under four days. This uniformity suggests a standardised order processing system across categories. With Office Supplies taking a slightly longer average time (3.98 days) compared to Furniture and Technology (both at 3.92 days), there may be a minimal scope for process optimization, particularly in the Office Supplies category, to achieve even greater efficiency.

**12. How does the performance of different salespeople compare in terms of actual sales, and profitability?**

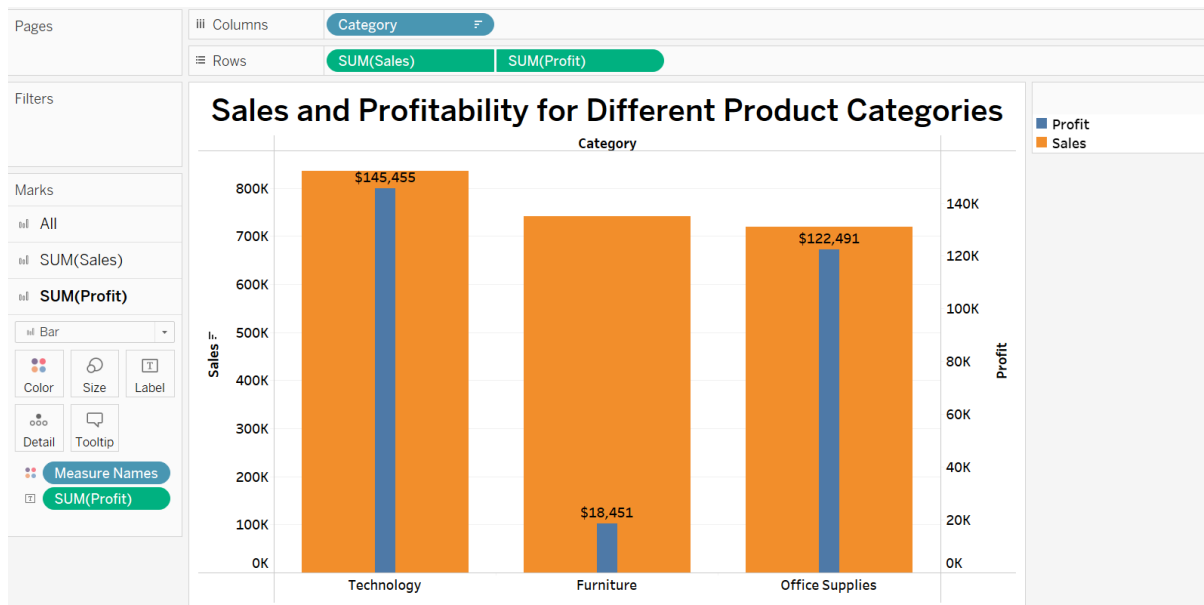
**Ans:**



The dual axis bar chart uses colour differentiation and varying widths to enhance comprehension and contrast between the two metrics. The wider bars represent total sales, typically using a more subdued colour, while the narrower bars, superimposed on the sales bars and in a contrasting colour, indicate profitability. This colour contrast draws the viewer's attention to profitability within the context of sales. The varying widths ensure that both metrics are easily distinguishable, facilitating a quick visual assessment of how each salesperson's profitability stacks up against their sales volumes. Anna Andreadi excels with the highest sales and profitability, showcasing her efficiency and strong profit margins. Chuck Magee follows with robust sales, though his profit ratio suggests a marginally lower conversion efficiency. Kelly Williams, while having higher sales, is at the bottom in terms of profitability, suggesting that despite a higher volume, her transactions may be less profitable. Conversely, Cassandra Brandow, with the least sales, does not have the lowest profitability, indicating that her sales strategy may be more cost-effective, yielding better profit margins relative to the volume of sales.

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

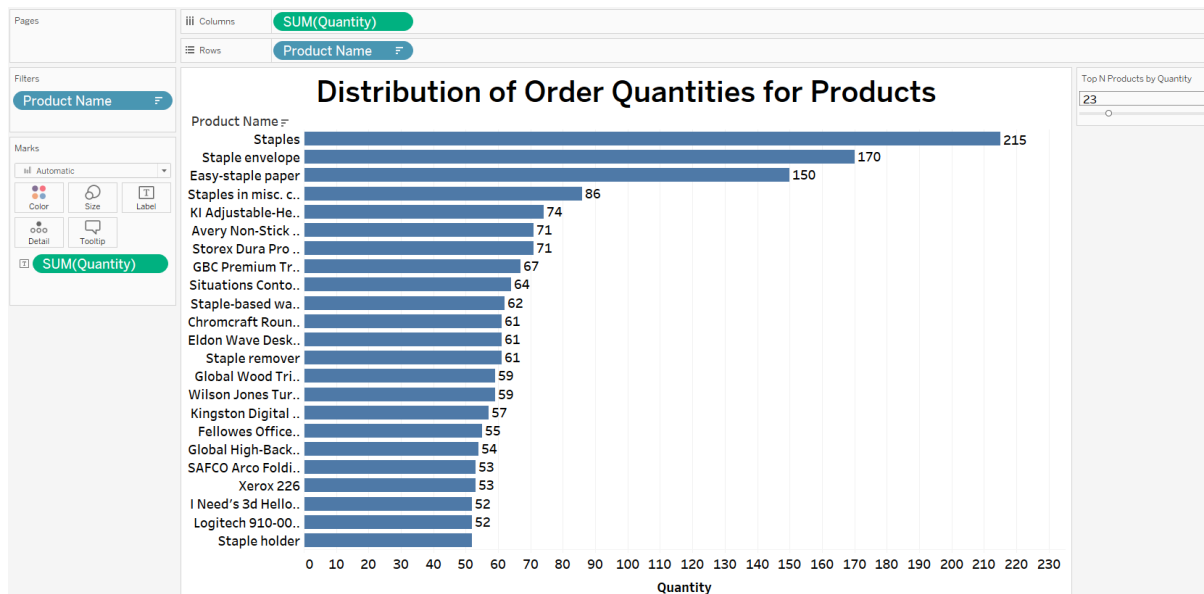
Ans:



The dual-axis bar chart is particularly effective for visualising the relationship between sales and profitability across product categories because it allows for the comparison of two metrics on the same chart. The primary bars for sales give a clear, comparative view of the revenue generated by each product category, highlighting which categories contribute most to the top line. The secondary bars for profit, aligned on the same axis, facilitate a direct comparison of the bottom line for each category, showing how revenue translates into actual profit. The differing bar widths prevent overlap and confusion, ensuring each metric stands out distinctly. For Technology, a high level of sales is accompanied by the highest profitability, indicating a strong performance in both aspects. Furniture shows a considerable volume of sales but with significantly lower profitability, suggesting a disparity between sales success and profit generation. Office Supplies has a moderate sales volume and profitability, which seems consistent with each other, showing a balanced relationship between sales and profit.

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

**Ans:**

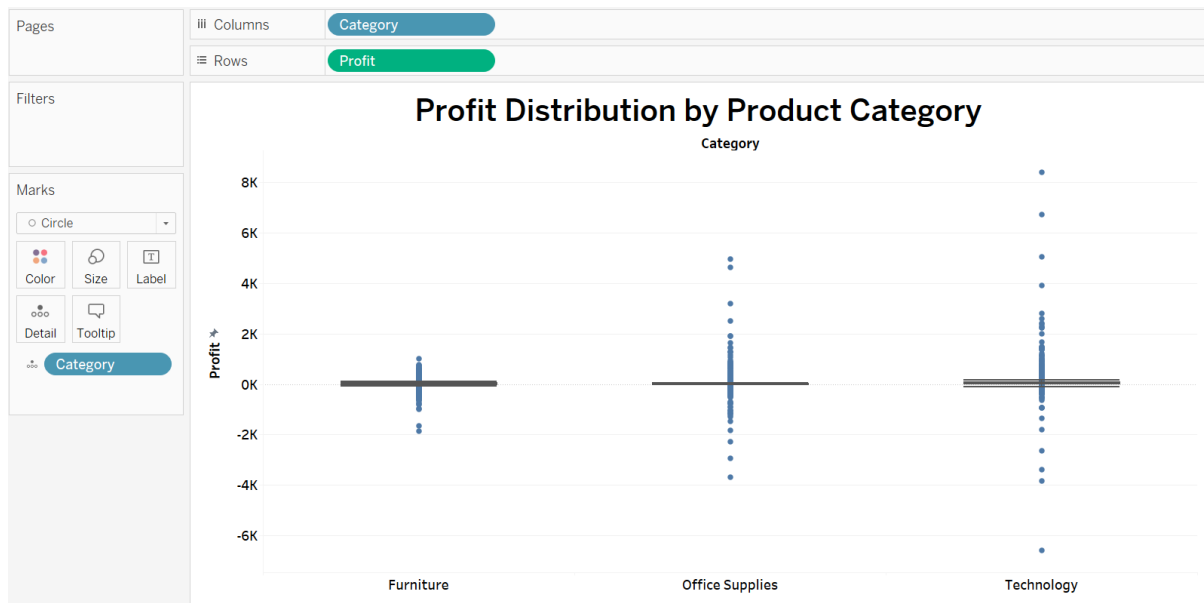


The horizontal bar chart is ideally suited for visualising the distribution of order quantities across products due to its clear layout, which enhances readability for long product names and facilitates easy comparison of data points. Its adaptability allows for displaying a wide range of values, making it perfect for handling large datasets, while the interactivity introduced by the 'Top N Products by Quantity' parameter enables a focused analysis on the most significant products, optimising both space and interpretive clarity. This chart highlights that 'Staples', 'Staple envelope' and 'Easy-staple paper' are the most ordered products, suggesting high demand or turnover. With a wide variance in order quantities among products, the chart underscores the need for strategic inventory management, especially for high-volume items.

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

**Ans:**

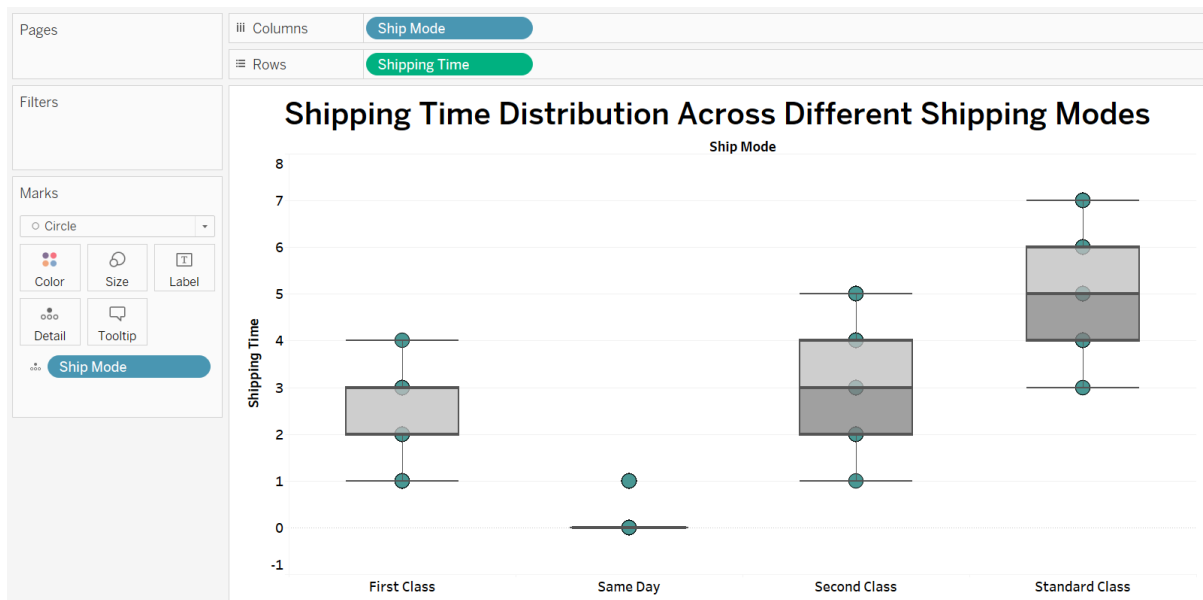




The box-and-whisker plot is an excellent choice for visualising the profit distributions across different product categories because it concisely represents multiple statistical measures in one visualisation. This chart type shows the median, which provides a better sense of the central tendency for skewed distributions than the mean. It also displays the interquartile range (IQR), which highlights the middle 50% of the data, giving a sense of the spread of the distribution. The 'whiskers' extend to show the range of the data, while points outside the whiskers indicate outliers. This box plot indicates distinct profit characteristics across Furniture, Office Supplies, and Technology. Technology exhibits a broad range of profits with significant outliers suggesting high variability and the potential for large gains. Furniture shows potential for losses, as indicated by outliers below the profit line. Office Supplies display a more consistent profit range with a median profit lower than Technology but with fewer losses, suggesting a stable but less lucrative market. The variability in profit distribution across these categories can inform targeted strategies for maximising gains and minimising losses within each product segment.

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

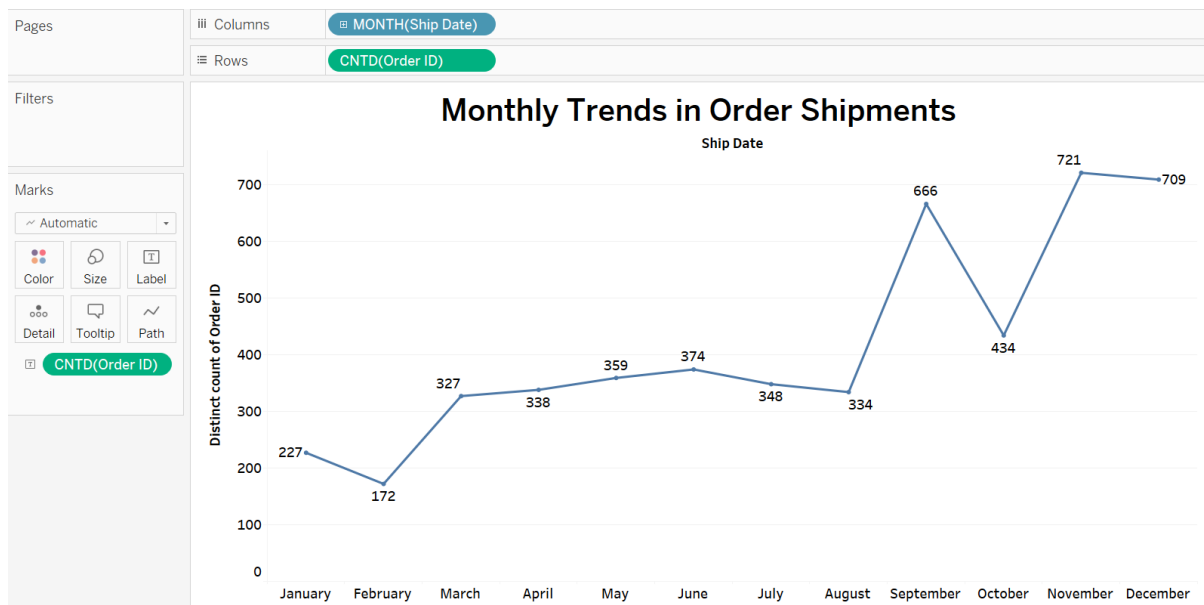
Ans:



The box-and-whisker plot is well-suited for comparing shipping time distributions, as it effectively displays the median, interquartile range, and outliers for each shipping mode in a single view. It provides a comprehensive comparison of the central shipping time tendencies and the variability within those times, illustrating both typical and extreme shipping durations. This plot displays that First Class exhibits a relatively quick shipping time with a tight interquartile range indicating consistent shipping times. There are a few outliers that suggest occasional delays beyond the typical shipping time frame. Same Day presents the quickest shipping times, as expected, with a median close to zero. However, the presence of an outlier below zero is likely an anomaly or data entry error since negative shipping time is not feasible. Second Class shows a broader range of shipping times compared to First Class, as indicated by a wider interquartile range. The median shipping time is higher than First Class, implying slower deliveries. Standard Class has the broadest interquartile range, suggesting the most variability in shipping times. It has the highest median shipping time, indicating it is generally the slowest option. Outliers exist both below and above the box, highlighting some instances of unusually quick and slow deliveries.

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

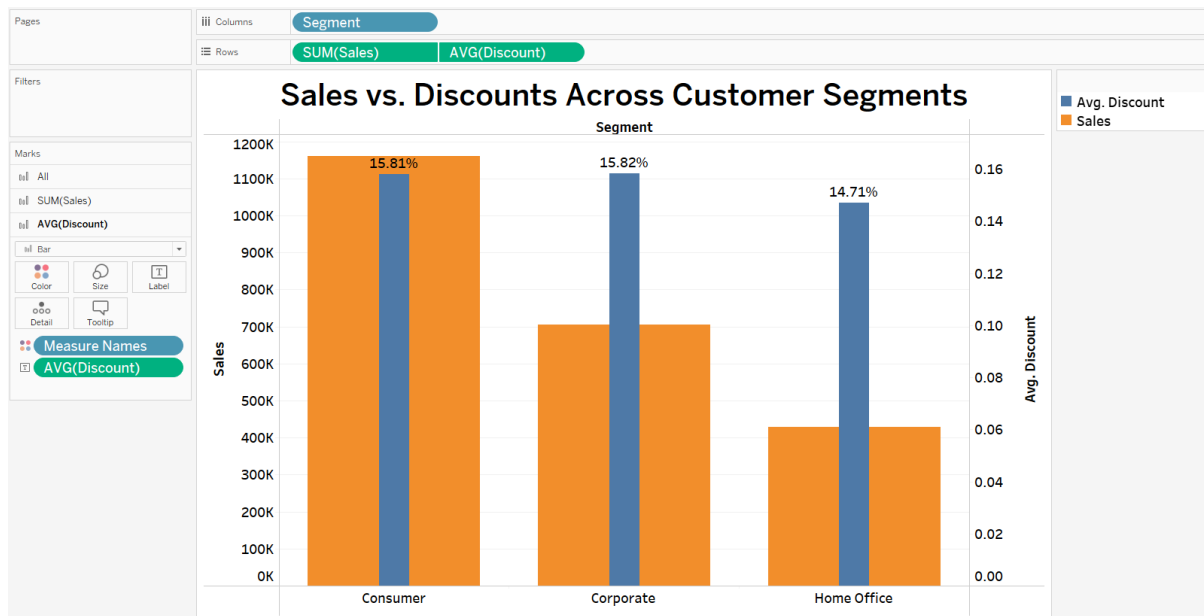
**Ans:**



The line chart is an ideal choice for visualising the monthly trend in the number of orders shipped because it clearly illustrates how the order volume changes over time. Line charts are particularly effective for time series data where the sequence of values is important and we want to track increases or decreases across intervals, in this case, monthly. This line chart presents a distinct pattern in order shipments, showing an initial increase from January to March, suggesting growth in orders with the new year. February experiences a notable decrease, potentially due to its shorter duration or seasonal market changes. There's a high point in June, with a downturn in the following summer months, possibly reflecting seasonal behaviour. The chart peaks in November, likely influenced by holiday shopping, and then dips slightly in December.

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

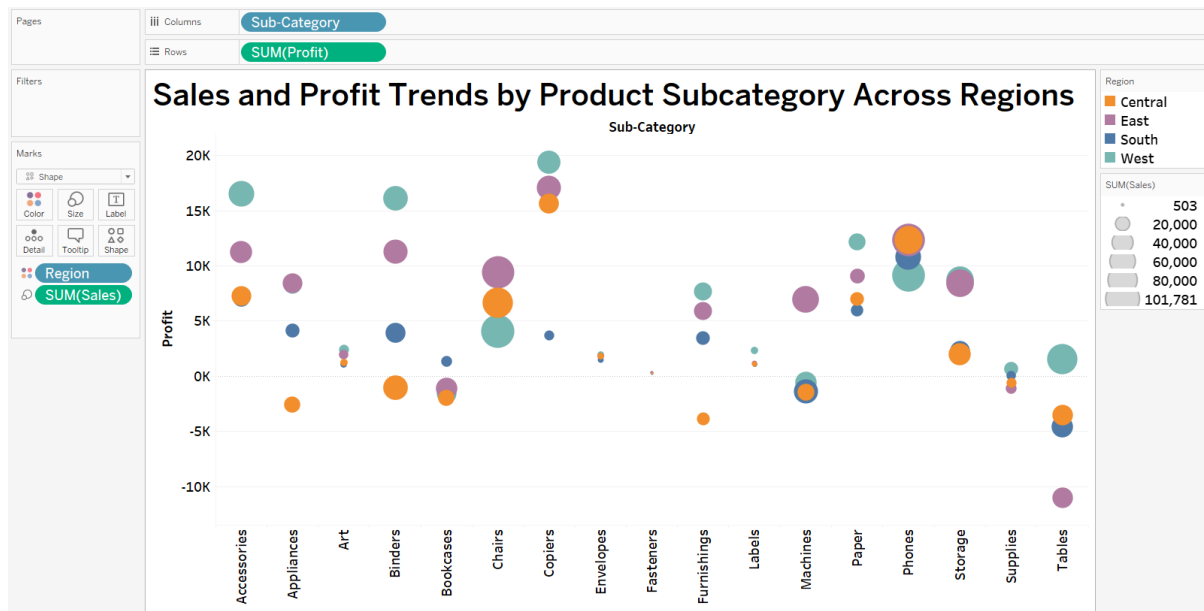
**Ans:**



The dual-axis bar chart is well-suited for visualising and comparing the performance of different customer segments in terms of sales and discount rates because it presents two related metrics on the same graph. The primary bar charts display the total sales, allowing for an easy comparison of the absolute sales figures across customer segments. Superimposed on these are the narrower bars representing the average discount rates, which provide context to the sales figures by showing the discount level associated with each segment's sales. Overall, the chart suggests a correlation between sales volumes and discount rates, where higher sales volumes are associated with higher discounts. However, the Home Office segment presents an exception to this pattern, maintaining lower discount rates despite lower sales volumes.

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

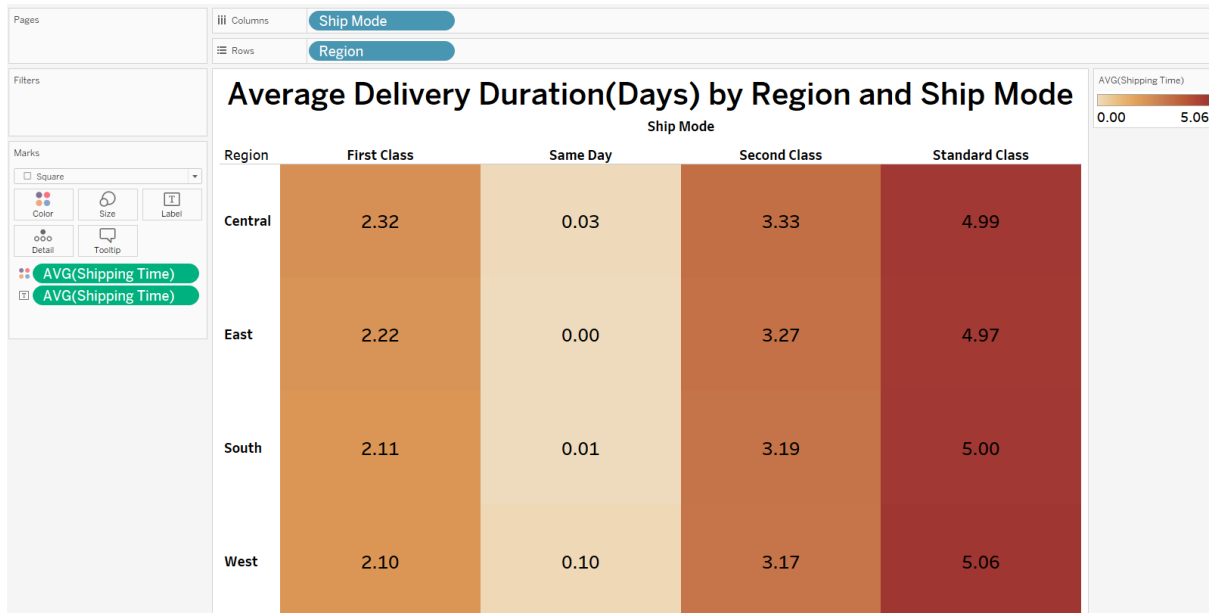
Ans:



The scatter plot is an excellent choice for displaying sales and profit trends across different product subcategories and regions because it allows for multidimensional analysis within a single visualisation. In this chart the position of each bubble on the y-axis provides immediate insight into the profitability of each subcategory within a region, clearly separating profitable from unprofitable items. The size of the bubbles represents the volume of sales, enabling a quick visual assessment of which subcategories and regions are driving the most revenue. The colour coding by region allows for easy segmentation and comparison of regional performance within the same subcategory. The scatter distribution highlights the variance in performance across regions, showing where strategies may need to be adjusted. This plot shows that while Phones have a high sales volume across regions, Copiers yield high profitability despite lower sales volumes, indicating higher margins. Tables consistently show losses in all regions, pointing to potential issues in pricing or costs.

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

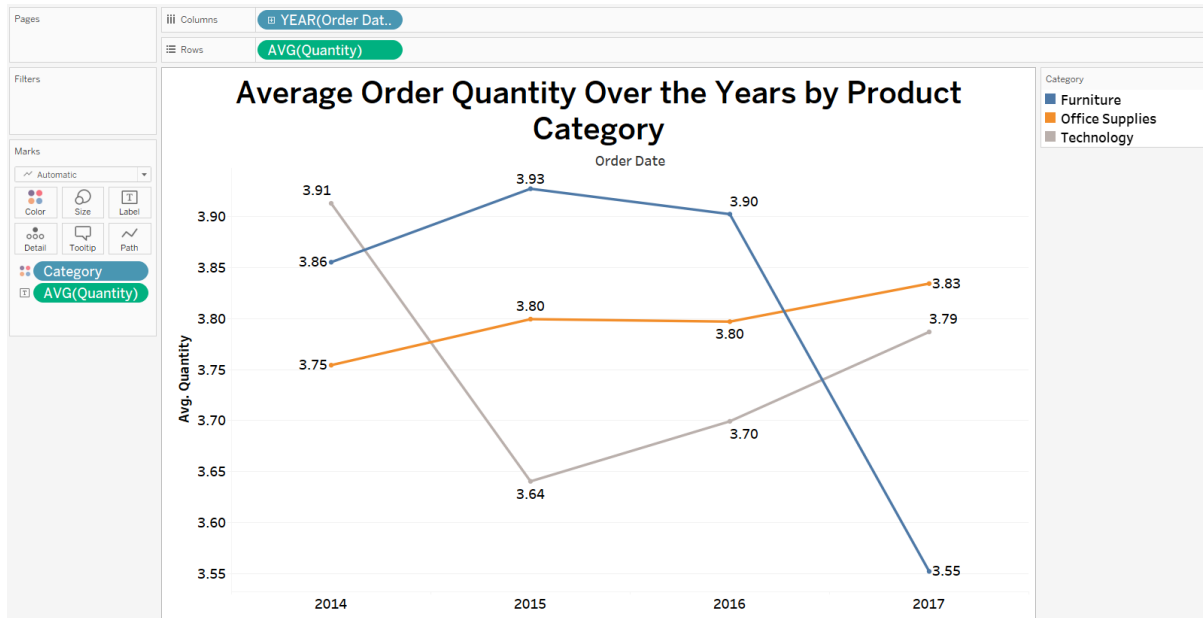
**Ans:**



The highlight table chart efficiently visualises the average delivery duration by region and ship mode through colour intensity, enabling instant pattern recognition and comparative analysis. It offers a space-efficient, intuitive matrix that contrasts delivery times across multiple categories, with colour shades quickly conveying duration darker for longer and lighter for shorter. This format provides a clear and immediate understanding of the data, making it a compelling choice for displaying delivery metrics in a comprehensive and accessible manner. This highlight table chart indicates that First Class deliveries average between 2.10 and 2.32 days, with Central being the slowest and West the quickest. Same Day shipping is nearly immediate across all regions, barring West at 0.10 days. Second Class deliveries average 3.17 to 3.33 days, with Central experiencing the longest delivery times. Standard Class shipping takes the longest, with durations from 4.97 to 5.06 days, and West has the lengthiest delivery times in this category.

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

**Ans:**



The line chart is a strong choice for visualising how the average order quantity has changed over time because it clearly displays trends and allows for easy comparison across different product categories. Its ability to represent continuous quantitative data over a time series makes it simple to observe increases, decreases, or stable patterns in the data. This line chart indicates that from 2014 to 2017, Furniture experienced a decline in average order quantity with a peak in 2015, Office Supplies remained relatively stable with a slight overall increase, and Technology showed volatility with a sharp drop in 2015 followed by a partial recovery in 2017.

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

**Ans:**

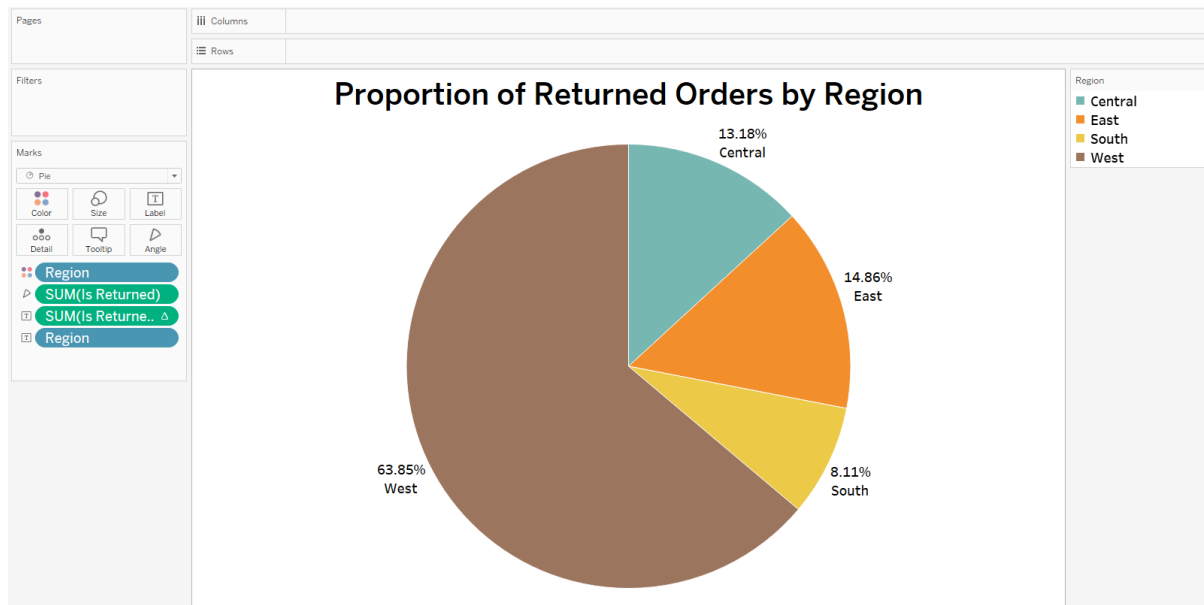


The scatter plot is an excellent choice for visualising the potential correlation between discount rates and order quantities among different customer segments. It allows individual data points to be plotted, showing the distribution and relationship between the two variables. This chart type is particularly effective for identifying patterns, trends, or clusters, as well as any outliers that may exist. Each customer segment is represented by a different colour, which makes it easy to distinguish between them and observe segment-specific trends or behaviours. The positive correlation across all customer segments suggests that higher discounts tend to incentivize larger orders, with this trend being most pronounced in the Consumer segment, indicating their higher sensitivity to discounts. The Corporate segment shows a moderate correlation, suggesting a balanced responsiveness to discount changes. In contrast, the Home Office segment exhibits the weakest correlation, indicating a lower sensitivity to discounts and perhaps a focus on other factors when making purchase decisions.

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

**Ans:**

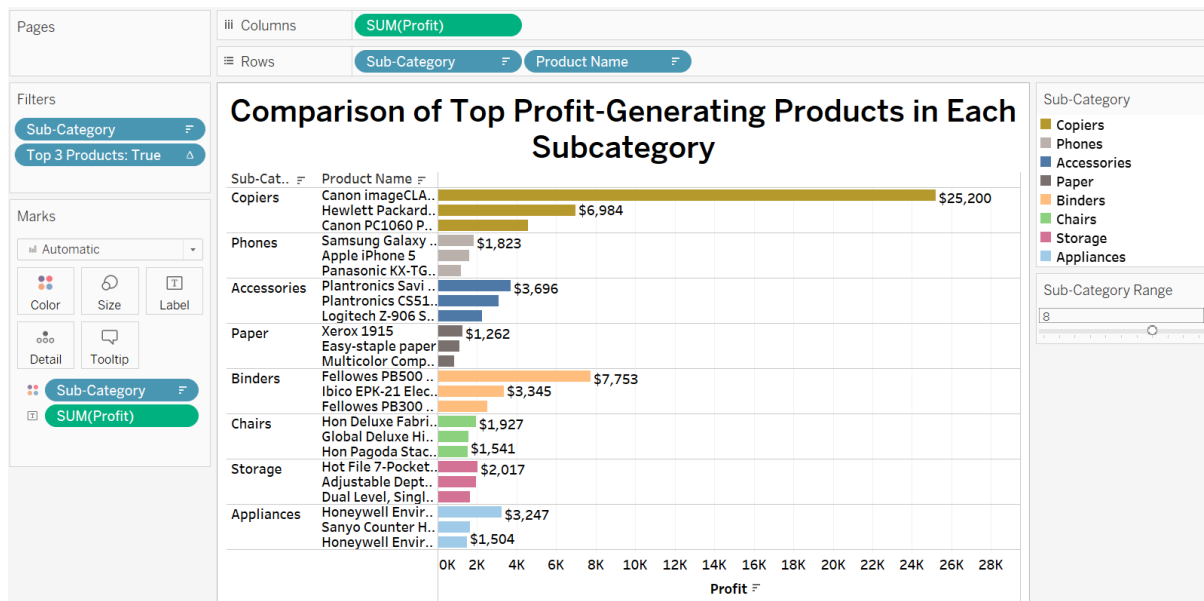




The pie chart is an effective visualisation tool for displaying the proportion of returned orders by region within the Superstore dataset because it illustrates parts of a whole in an intuitive manner. Each slice of the pie represents the percentage of the total returns that are attributable to a specific region, allowing for an immediate visual comparison. This format is particularly useful when the number of categories is limited, as is the case with the regions here, which makes the chart easy to read and interpret. The size of each slice quickly conveys the relative magnitude of returns for each region, making it clear which regions have higher or lower return rates. The calculated field named "Is Returned" effectively distinguishes returned orders within the dataset, where it is set to 1 for each returned order and 0 otherwise. The pie chart using this field reveals that the West region has the highest proportion of returns at 63.85%, the East follows with 14.86%, the Central region has 13.18%, and the South holds the smallest share with 8.11% of the total returned orders.

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

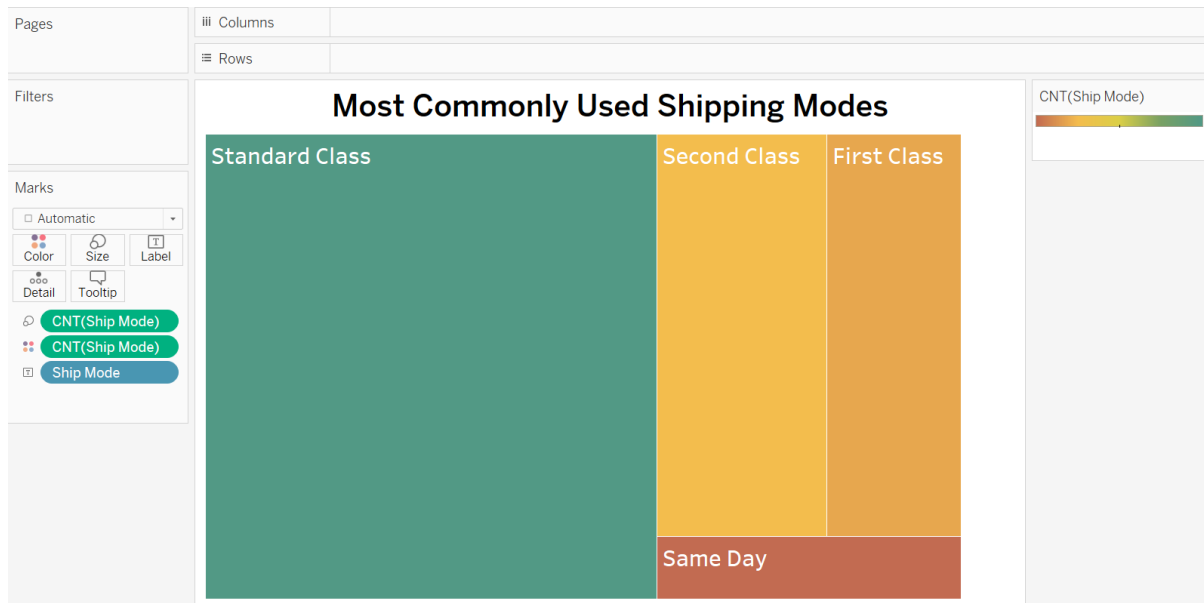
**Ans:**



The horizontal bar chart is an excellent choice for comparing profits of different products across subcategories due to its clear and organised display. This type of chart is particularly effective for showing rankings or comparisons across multiple products, as the length of each bar intuitively indicates the value of the profit, allowing for quick visual assessment. Additionally, the horizontal layout accommodates the text labels for the product names without clutter, making it easy to read even when dealing with longer names. By using a parameter to control the number of subcategories displayed, we can customise the view to focus on more or fewer groups as needed, enhancing the chart's flexibility. Furthermore, the use of a calculated field to filter and display only the top three products ensures that the visualisation is not overcrowded and focuses attention on the highest profit-generating items.

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

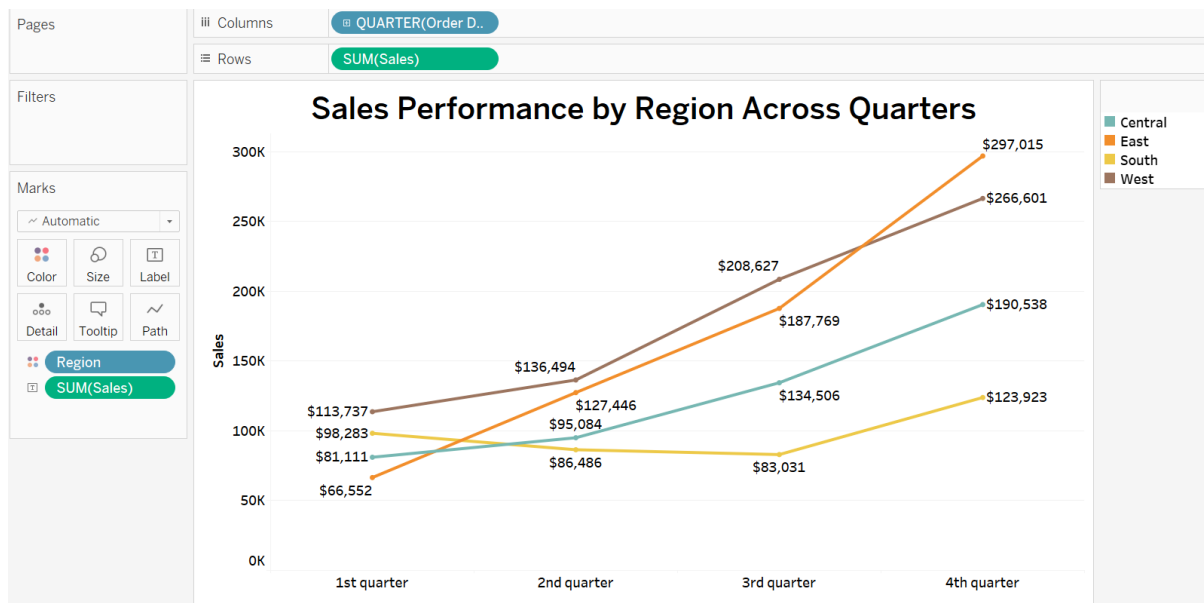
**Ans:**



A treemap effectively showcases the most used shipping modes in the dataset by allocating larger rectangles to more frequently used modes, which enables a quick visual comparison of their relative usage. The area occupied by each shipping mode directly corresponds to its prevalence, with "Standard Class" being the most prominent, indicating it is the most used mode. The smaller rectangles for "Second Class," "First Class," and "Same Day" shipping modes convey that they are less common. This visual approach is particularly beneficial for highlighting the distribution of categories at a single glance, without the need for numerical interpretation, and the immediate visual impact of the large "Standard Class" rectangle communicates its dominance in shipping mode preference.

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

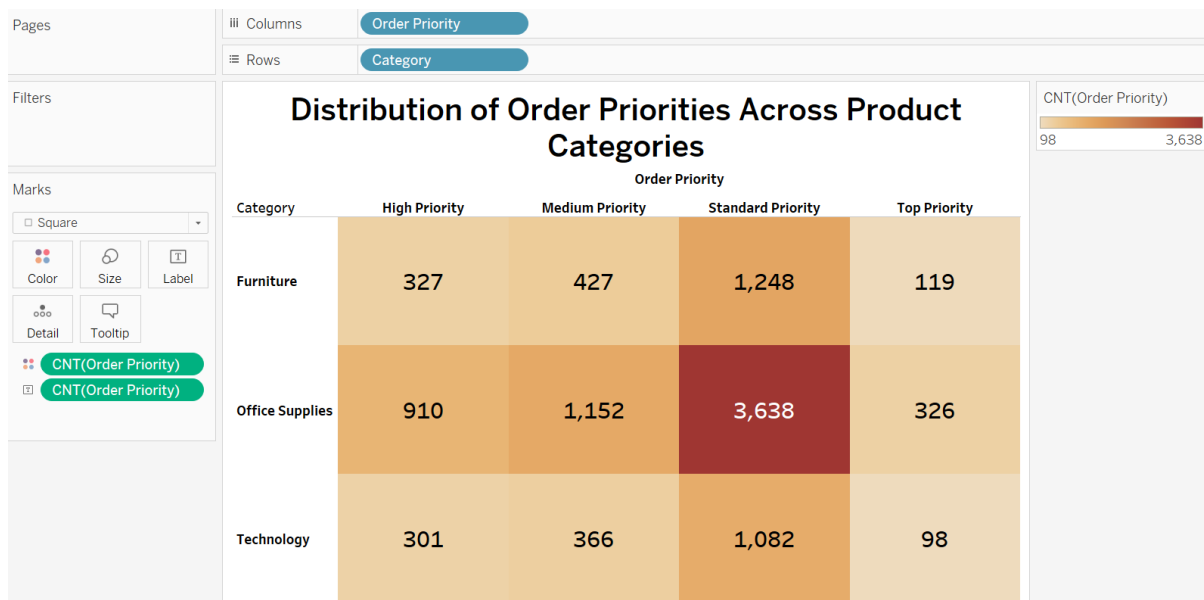
**Ans:**



The line chart is well-suited for visualising the sales performance of different regions across the quarters of a year due to its ability to clearly display trends over time. In the chart, the temporal progression is easily followed along the x-axis, with each line's upward or downward trajectory illustrating the change in sales over time for each region. Sales in the East region consistently rose over the year, culminating with the peak sales in the 4th quarter. The West region, beginning the year with the highest sales, maintained an upward trend and finished just behind the East in the final quarter. Both the South and Central regions experienced growth, but the Central region edged out with slightly higher sales than the South by the 4th quarter, leaving the South at the lowest sales position at year's end.

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

Ans:



The highlight table chart is an effective choice for visualising the distribution of order priorities across different product categories because it combines the simplicity of a table with the visual power of colour coding. This format allows for immediate recognition of patterns and comparisons: higher quantities are instantly identifiable by deeper colour shades, enabling quick comparison across categories and priorities. The 'Order Priority' calculated field reclassifies shipping modes into a hierarchy of urgency: 'Same Day' as 'Top Priority', 'First Class' as 'High Priority', 'Second Class' as 'Medium Priority', and all other shipping modes as 'Standard Priority'. This chart illustrates that 'Standard Priority' is the most common shipping option selected across all product categories, with Office Supplies registering the highest number of orders, particularly under 'Standard Priority'. 'High Priority' and 'Top Priority' options are the least favoured among customers in all categories, suggesting a lower demand for expedited shipping. The Technology category has the fewest 'Top Priority' orders, indicating a minimal need for rapid delivery in this segment, while Furniture follows the overall trend, preferring 'Standard Priority' but with fewer orders than Office Supplies.

## 28. What is the relationship between discounts and sales?

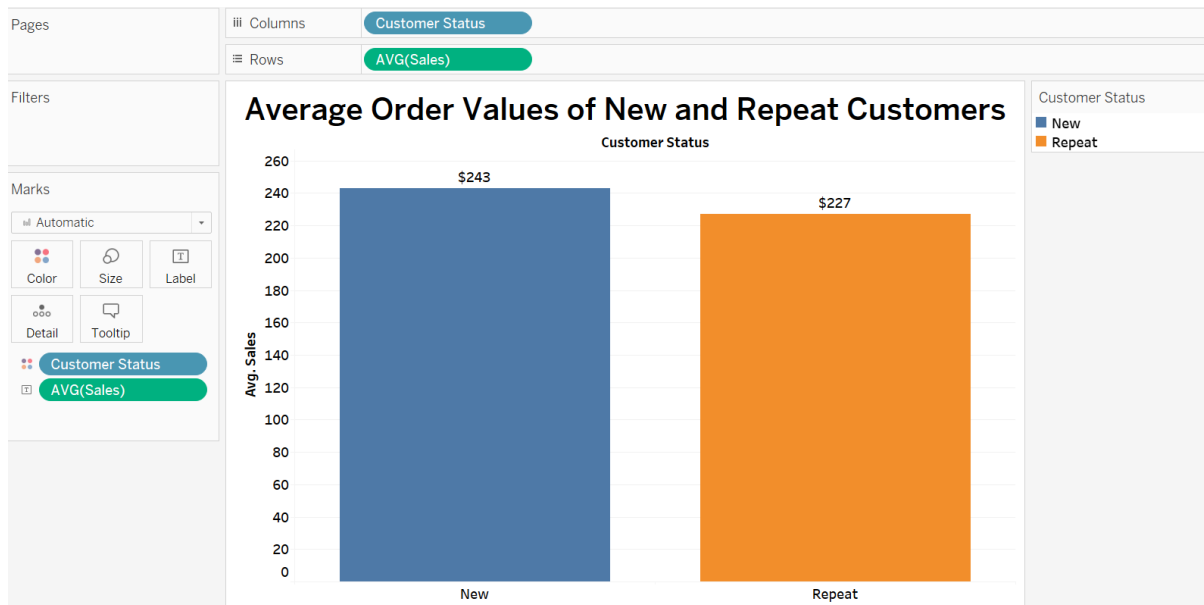
Ans:



A scatter plot is chosen for its ability to display the relationship clearly and precisely between two continuous variables, such as discounts and sales. It excels in revealing the density and distribution of data points, identifying correlations, trends, and outliers. This scatter plot indicates that there isn't a strong correlation between the level of discount and the total sales. There are higher sales at lower discount levels, but also some significant sales at around 0.5 discount level which means a small discount can be effective in increasing sales. A large discount may not be necessary to increase sales. Discounts can be more effective for some products than others.

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

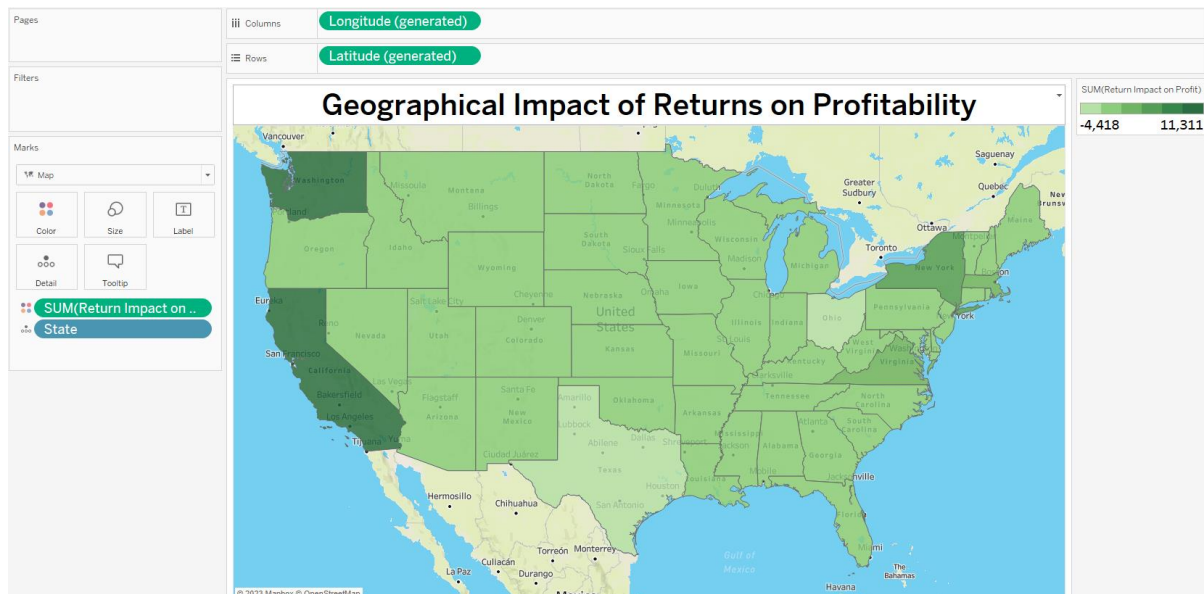
**Ans:**



The bar chart is an excellent choice for visualising the comparison between the average order values of new and repeat customers because it displays data categorically and quantitatively. Bar charts are particularly effective for comparing discrete categories, in this case, 'New' versus 'Repeat' customers, allowing for a quick visual assessment of differences. The average values are immediately apparent, and the chart is inherently simple and easy to interpret, which makes it accessible to a wide audience. The methodology behind the 'Customer Status' calculated field is to label a customer as 'New' when their order date corresponds to the first recorded order for that particular customer ID. Orders that follow from the same customer ID are then labelled as 'Repeat'. It indicates that new customers have a higher average order value (\$243) compared to repeat customers (\$227), suggesting initial purchases are more substantial. This could imply that first-time buyers might be incentivized by marketing strategies or introductory offers.

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

**Ans:**



The map chart is an optimal choice for visualising the geographical distribution of returns and their impact on profitability because it provides a spatial context that is instantly recognizable. It allows stakeholders to see at a glance which regions are most affected by returns, offering a clear visual correlation between location and financial performance. The use of colour gradients to represent varying degrees of profit impact further enhances the chart by conveying the severity of returns in different areas, allowing for a quick assessment of regions that may require more attention. The calculated field 'Return Impact on Profit' conditionally isolates the profit only on returned items, which allows us to focus on the profitability impact of returns specifically. This map shows that there is a general trend of higher returns on profitability in the northern and western parts of the United States, and lower returns on profitability in the southern and eastern parts of the United States with some exceptions like New York and Virginia.

Analysis By: Rohit Paul

Email: [rohitpaul09@gmail.com](mailto:rohitpaul09@gmail.com)

Workbook:

<https://drive.google.com/file/d/11gHYH5bWdbCJ5dtVnK2IyjRsN0SftG0a/view?usp=sharing>