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:

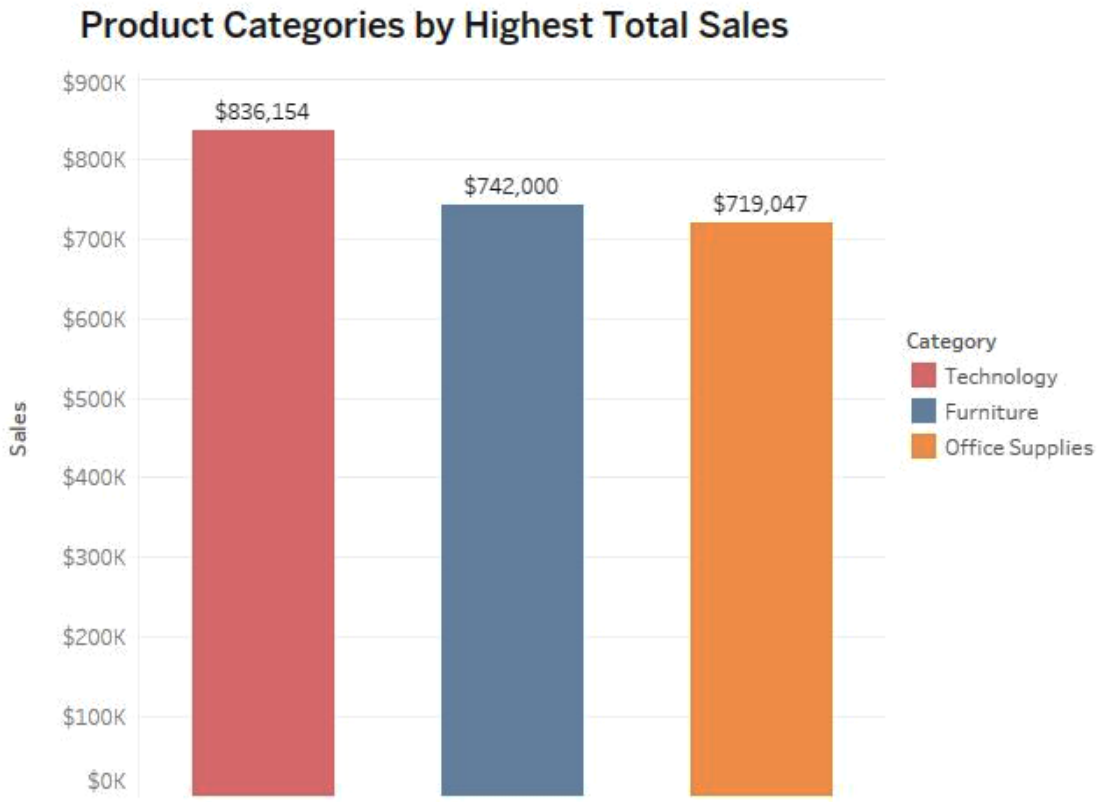
1. **Answer Completion**: Ensure that you furnish answers for all 30 questions and build charts for them.
2. **Encouraged Creativity**: Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

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?



- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because the bar chart is effective for comparing sales values among different categories, with each category represented by a bar and the height of the bar corresponding to the total sales.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  The product categories with the highest total sales were derived from the chart. The category 'Technology' stands out with the highest sales of $836,154, indicating that products in this category are in high demand among customers and generate the highest sales.

  Categories ‘Furniture’ and ‘Office Supplies’ followed by 'Technology' have almost similar sales values, suggesting a competitive market landscape among these segments. This underscores the significance of these categories in driving overall sales performance and highlights potential areas for further analysis and strategic focus.

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

- **Why did you pick the specific chart?**

  A line chart was chosen for this visualization because it effectively shows the sales trend over the years. The lines clearly depict the overall change in monthly sales over the years, making it easy to identify patterns.

- **What is/are the insight(s) found from the chart?**

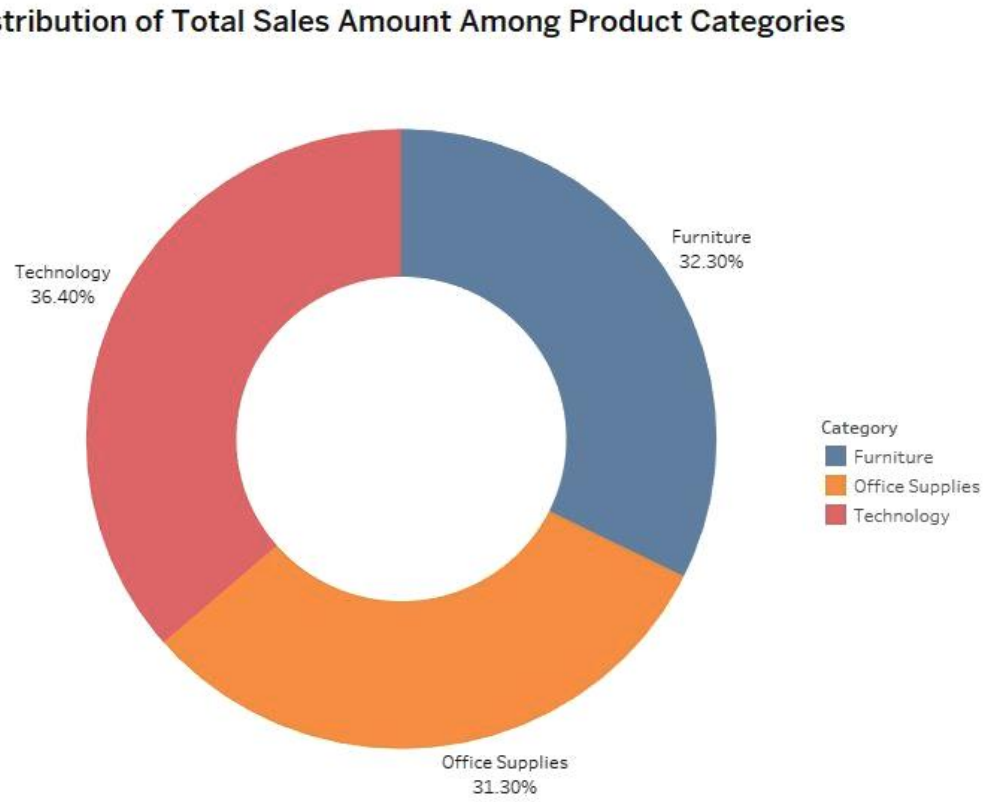
  The key insights derived from the chart are as follows as:

  The chart clearly shows consistent improvement in sales, indicating increased demand for products over the years.

  Major spikes are evident in the month of September and in between November & December, suggesting high demand for products during these periods of the year because of some factors such as holidays, festivals or special events. For instance, September may witness increased sales due to end of season sales and November & December typically experience sales due to the winter holidays, festive season and new year celebrations.

  Conversely, February, August and October are the months with the least demand for products because they may represent periods of lower consumer spending, absence of major holidays or seasonal downturns in certain industries.

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



- **Why did you pick the specific chart?**

  A pie chart was chosen for this visualization because it effectively represents the proportions of each category relative to the whole. The 'donut' effect provides a clean and visually appealing representation, making it easy to observe the proportions.

- **What is/are the insight(s) found from the chart?**

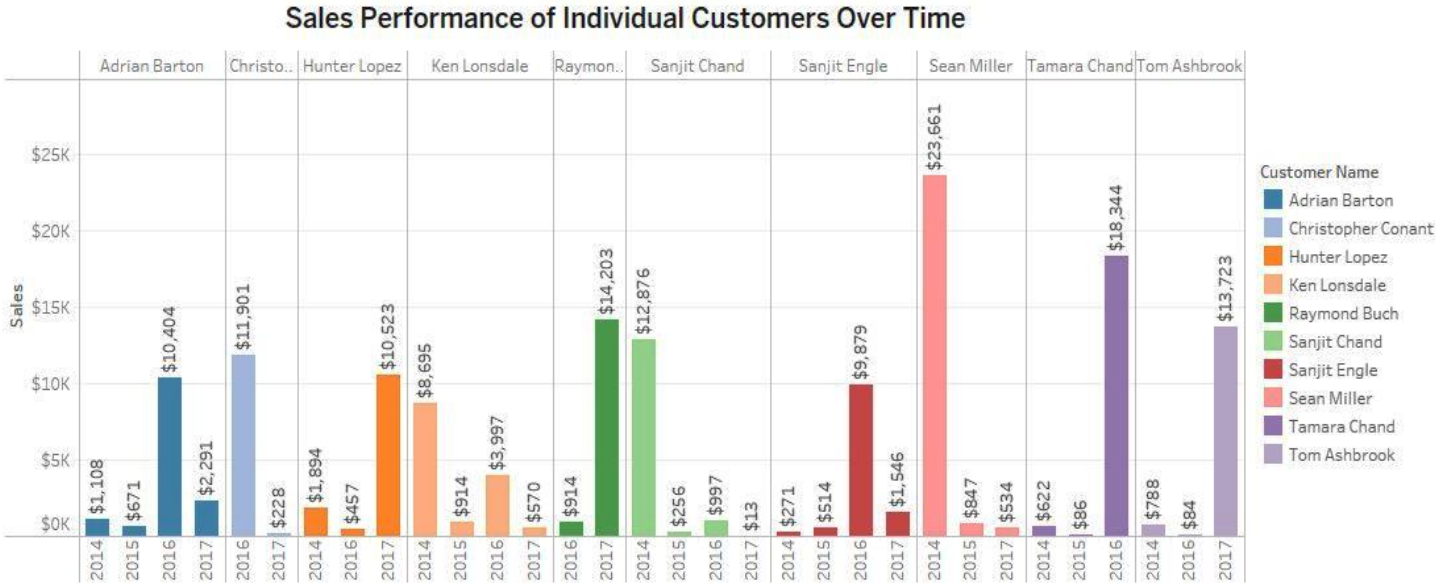
  The key insights derived from the chart are as follows as:

  'Technology' is the product category with the highest total sales amount, accounting for 36.40% of the total, indicating that it has the biggest share of the total sales. This highlights the strong demand for technological products and the significant role they play in driving revenue.

  'Furniture' follows closely behind, making up 32.30% of the total sales, suggesting a significant portion of the market demand.

  'Office supplies' have the lowest total sales amount at 31.30%, implying a smaller share of the overall sales compared to the other categories.

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



- **Why did you pick the specific chart?**

  A bar chart is an appropriate choice for this visualization because it effectively compares discrete categories (individual customers in this case) based on a single data point (total sales in a specific period).

- **What is/are the insight(s) found from the chart?**

  Since the dataset was too huge to analyze the sales performance of individual customers in a single screenshot, therefore data for top 10 customers was taken.
  The key insights derived from the chart are as follows as:

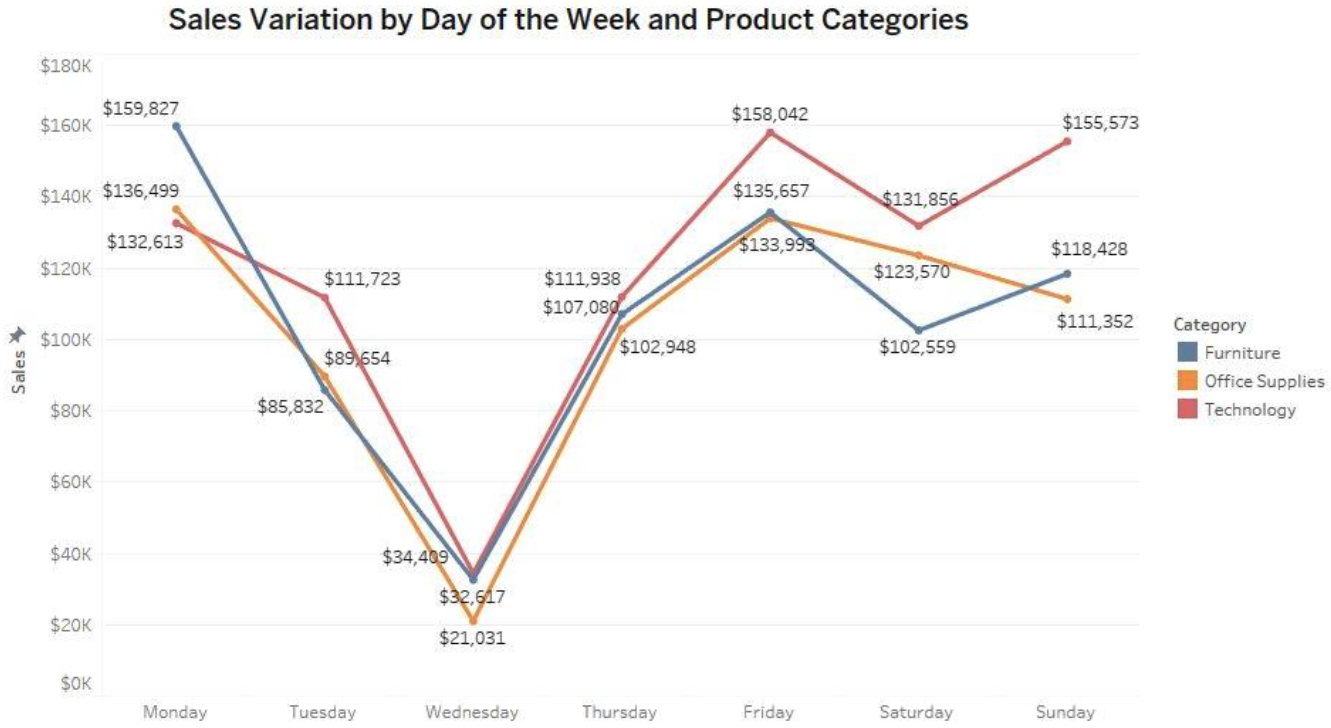
  The heights of the bars differ considerably, indicating a diverse range of sales performance among customers.

  The chart readily shows the customers with the highest and lowest total sales within the depicted timeframe. Wherein, the highest sales of $23,661 was recorded for a customer named Sean Miller in the year 2014.

  Interestingly, 2016 and 2017 are the years where the highest sales were recorded for individual customers. This could be due to various factors such as improved economic conditions, changes in consumer behaviour or the introduction of new products or services that resonated well with customers during those years.

  Grouping customers by factors like customer segment, product category or location could reveal interesting patterns into the underlying factors driving the increase in individual customer sales.

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



- **Why did you pick the specific chart?**

  A line chart was chosen for this visualization because it effectively shows the variation in sales of different product categories over day of week. A line plot is suitable for showing the progression of a numeric variable (sales) across a continuous axis (weekday). Additionally, annotating each data point with the actual sales amount provides further context and clarity.

- **What is/are the insight(s) found from the chart?**

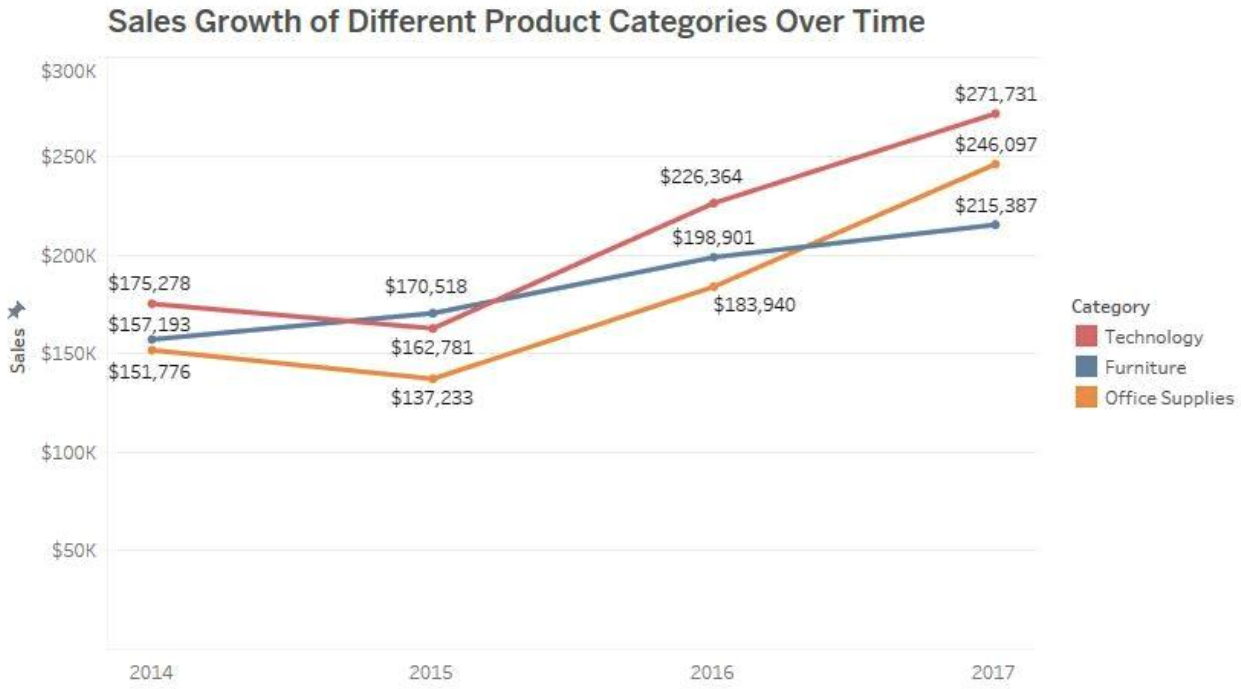
  The key insights derived from the chart are as follows as:

  The line clearly indicates higher sales occurring on Monday, Friday, and Sunday for most product categories. This suggests that these days generally have higher sales, coinciding with the start, end and weekends when consumers may have more time for shopping or leisure activities.

  Interestingly, Wednesdays exhibit the lowest sales across all categories. This could be attributed to mid-week downturn in consumer activity, as individuals prioritise work or other weekday responsibilities, leading to a decrease in shopping outings and purchases during this period.

  Products related to category ‘Technology’ seem to have good demand throughout the week. This consistent demand may reflect the rapid pace of technological advancements among customers.

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



- **Why did you pick the specific chart?**

  A line chart was chosen for this visualization because it effectively shows the sales growth of different product categories over time. A line plot is suitable for showing the progression of a numeric variable (sales) across a continuous axis (years). Additionally, annotating each data point with the actual sales amount provides further context and clarity.

- **What is/are the insight(s) found from the chart?**

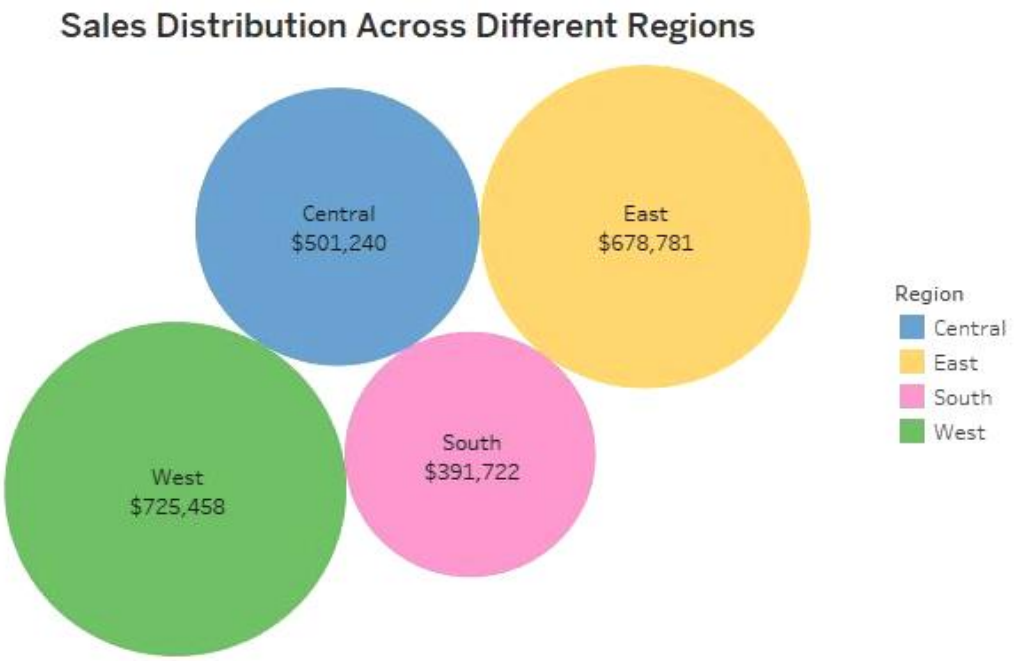
  The key insights derived from the chart are as follows as:

  The upward trend in the sales line for the product category 'Furniture' indicates a positive growth trajectory. Conversely, both the 'Technology' and 'Office Supplies' categories experienced a decline in sales in the year 2015, followed by subsequent increases in sales over the years.

  The product category 'Office Supplies' experienced a sudden increase in sales after the year 2015, which can be caused by several factors such as strategic marketing campaigns, product innovations or changes in consumer behaviour.

  On the other hand, the 'Technology' category consistently maintains the highest sales compared to other categories over time, which indicates sustained consumer demand for technology-related products. This may be attributed to advancements in technology, increased reliance on electronic devices in both personal and professional settings.

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



- **Why did you pick the specific chart?**

  A packed bubble chart was chosen for this visualization because it offers an intuitive visual representation of data distribution. The size of each bubble in the chart corresponds to the magnitude of sales in each region. This allows for easy comparison of sales performance visually.

- **What is/are the insight(s) found from the chart?**

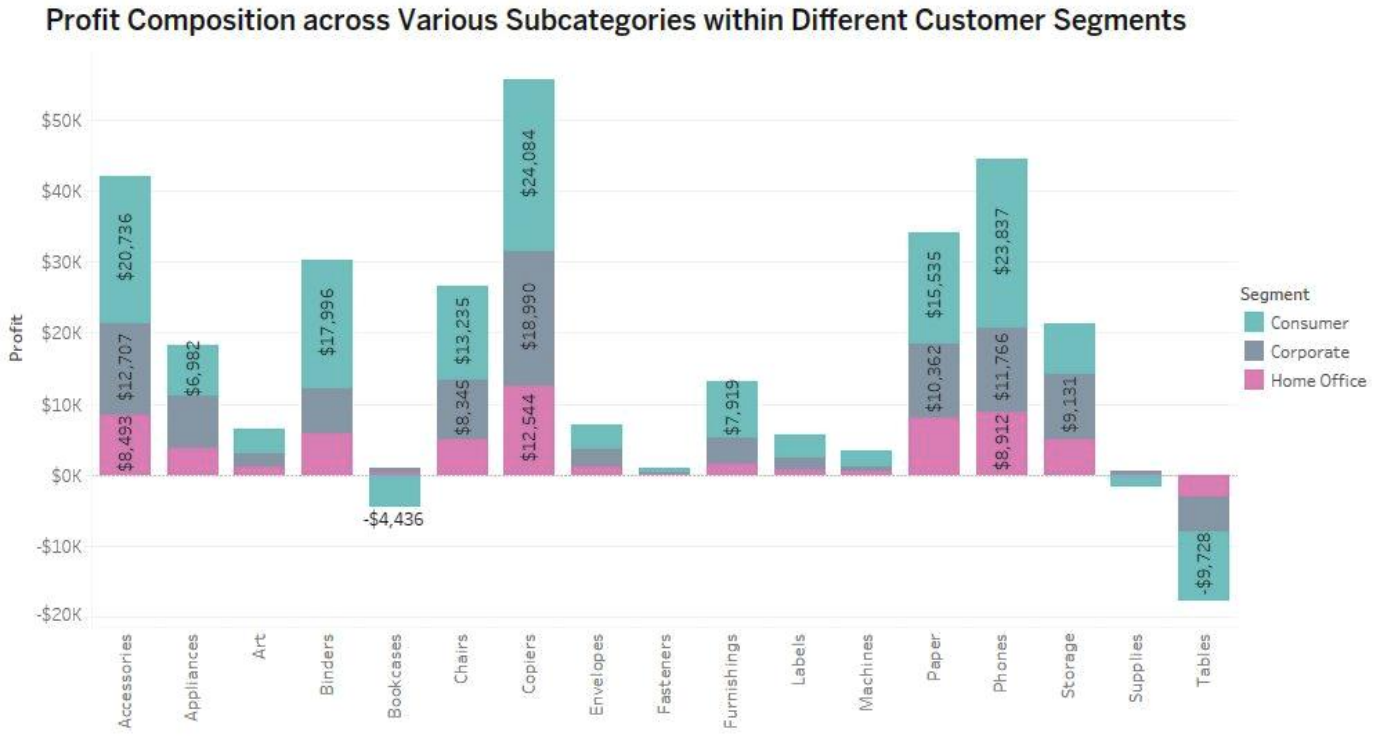
  The key insights derived from the chart are as follows as:

  The West region has the highest sales distribution of $725,458, indicating strong market demand and significant revenue generation from this region. This may suggest potential opportunities for further expansion or targeted marketing efforts to capitalize on the already established sales performance.

  On the other hand, the South region has the lowest sales distribution of $391,722, highlighting potential areas for improvement or marketing strategies to boost sales in this region. Understanding the factors contributing to the lower sales in the South region could help identify opportunities for optimization and growth, such as refining product offerings or adjusting pricing strategies to better align with customer preferences and market dynamics.

  Central and East regions maintain a good sales distribution, suggesting stable market performance and consistent revenue generation in these areas.

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



- **Why did you pick the specific chart?**

  A stacked bar chart was chosen for this visualization because it effectively compares categorical data such as sub-category & customer segment against profit and allows easy visual comparisons between them. In this case, we can see how profit varies across different sub-categories for each customer segment.

- **What is/are the insight(s) found from the chart?**

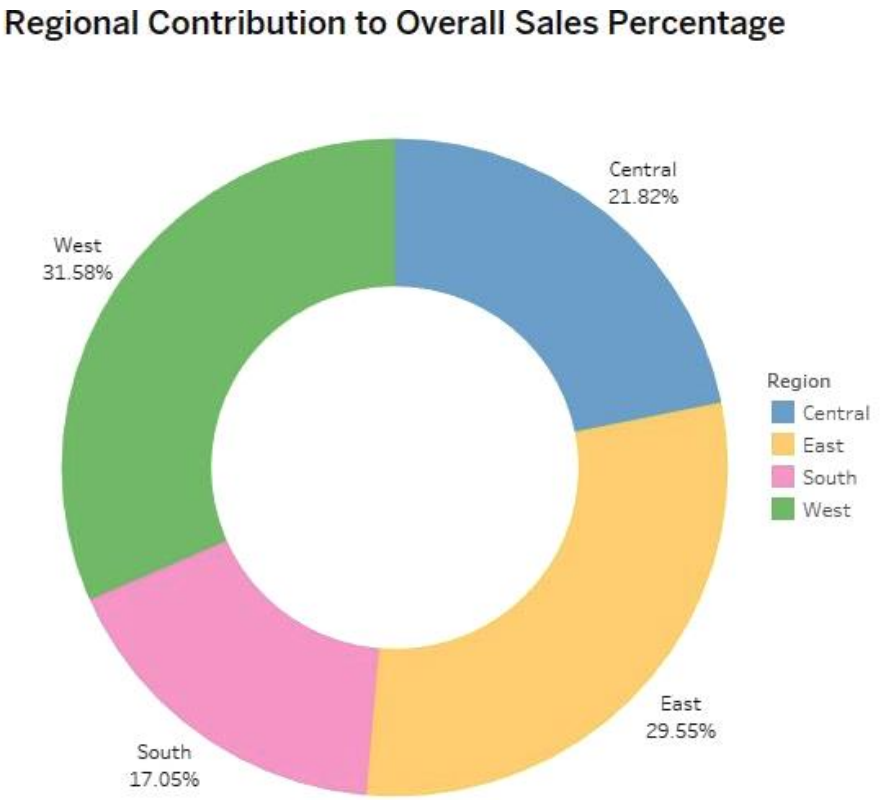
  The key insights derived from the chart are as follows as:

  Sub-categories like Accessories, Binders, Copiers and Phones have the highest profit across all consumer segments, reflecting strong demand and profitability within these product categories.

  Conversely, sub-categories like Bookcases, Fasteners and Supplies have the lowest profit across all customer segments, indicating potential challenges or lower demand for these particular products.

  Interestingly, the sub-category Table has no profit across any customer segments, suggesting potential issues such as high production costs, pricing strategies or lower market demand for this specific type of product. Further analysis is needed to identify the root causes and potential solutions to address the negative profitability of Tables.

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



- **Why did you pick the specific chart?**

  A pie chart was chosen for this visualization because pie charts are effective for visually representing the percentage or proportions of the whole. The 'donut' effect provides a clean and visually appealing representation, making it easy to observe the proportions.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  The West region has the highest sales contribution of 31.58%, indicating robust market demand and significant revenue generation from this region.

  On the other hand, the South region has the lowest sales contribution of 17.05%, highlighting potential areas for improvement or targeted marketing strategies to boost sales in this region.

  East regions maintain a good sales contribution of 29.55% followed by the West region, showing a strong presence and solid market performance in both regions.

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

- **Why did you pick the specific chart?**

  A horizontal stacked bar chart was chosen for this visualization because it effectively compares categorical data such as shipping mode and customer segment against profit and allows easy visual comparisons between them. In this case, we can see how profit varies across different shipping modes for each customer segment.

- **What is/are the insight(s) found from the chart?**

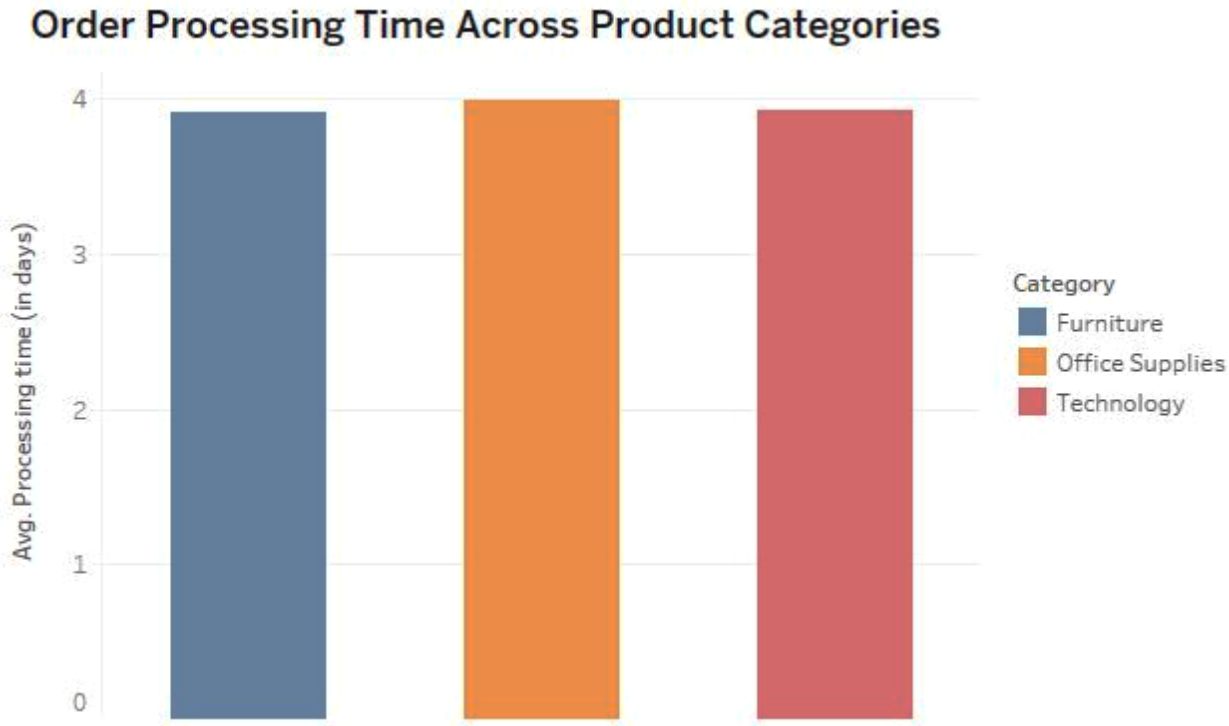
  The key insights derived from the chart are as follows as:

  Standard class shipping mode shows the highest profit margin across all customer segments, indicating its efficiency and cost-effectiveness in delivering products while maintaining profitability.

  Conversely, Same day shipping mode shows the lowest profit margin across all customer segments, likely due to its expedited nature and higher associated costs.

  Interestingly, First class and Second class shipping modes show almost similar profit margins across all customer segments, suggesting comparable levels of service and profitability between these two shipping options.

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



❖ **How to calculate processing time?**

Processing time was calculated by creating a calculated field using the formula as follows:

DATEDIFF('day', [Order Date], [Ship Date])

- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because it effectively compares order processing time among different categories, with each category represented by a bar and the height of the bar corresponding to the processing time.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  Average order processing time among all the categories is nearly 4 days, which means that orders typically take approximately four days from the order date to the ship date. This insight provides valuable information about the efficiency of order processing within the business operations and can help identify areas for improvement in streamlining the fulfilment process to enhance customer satisfaction and optimize resource allocation.

12. How do discounts affect overall profit?

❖ **How to calculate discount over profit?**

Discount over profit was calculated by creating a calculated field using the formula as follows:

IF [Discount] = 0 THEN "No Discount"

ELSEIF [Discount] > 0 THEN "Discount"

END

- **Why did you pick the specific chart?**

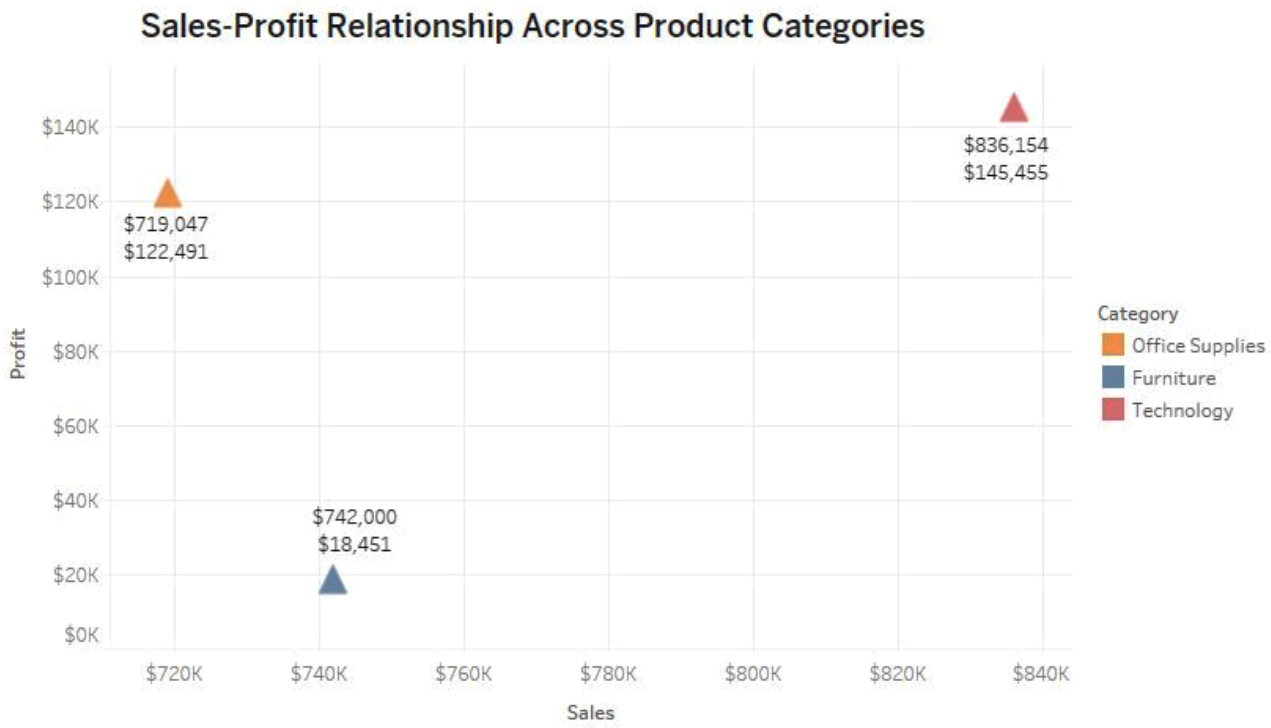
  A bar chart was chosen for this visualization because the bar chart is ideal for comparing profit values with and without discounts or comparing profitability. The length or height of bars makes it easy to visually compare the magnitude of discounts and their impact on overall profit.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  A discount of $34,591 was given on the overall profit of $320,988, resulting in a net profit of $286,397. This indicates that the discount accounted for approximately 10.77% of the total profit. Identifying the most effective discounting strategies can lead to improved customer satisfaction and increased sales while maintaining healthy profit margins.

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



- **Why did you pick the specific chart?**

  A scatter plot was chosen for this visualization because it effectively reveals the relationship between two continuous variables: Sales and Profit. Each data point represents a category, enabling us to observe how sales values correspond to profit.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  The overall sales for ‘Office Supplies’ amounted to $719,047, with a profit generated of $122,491, which accounts for approximately 17.04% of total sales. This indicates a healthy profit margin relative to the sales volume within the 'Office Supplies' category.

  Similarly, the overall sales for ‘Furniture’ amounted to $742,000, with a profit generated of $18,451, constituting approximately 2.49% of total sales. This suggests a comparatively lower profit margin relative to the sales volume in the 'Furniture' category.

  Lastly, the overall sales for ‘Technology’ amounted to $836,154, yielding a profit of $145,455, which accounts for approximately 17.40% of total sales. This highlights a robust profit margin relative to the sales volume in the 'Technology' category, indicating strong profitability within this segment.

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

- **Why did you pick the specific chart?**

  The histogram was chosen for this visualization because it allows a clear visual representation of the distribution of order quantities, enabling easy comparison of frequency across different quantity ranges. This comparison helps in understanding the spread and concentration of order quantities.

- **What is/are the insight(s) found from the chart?**

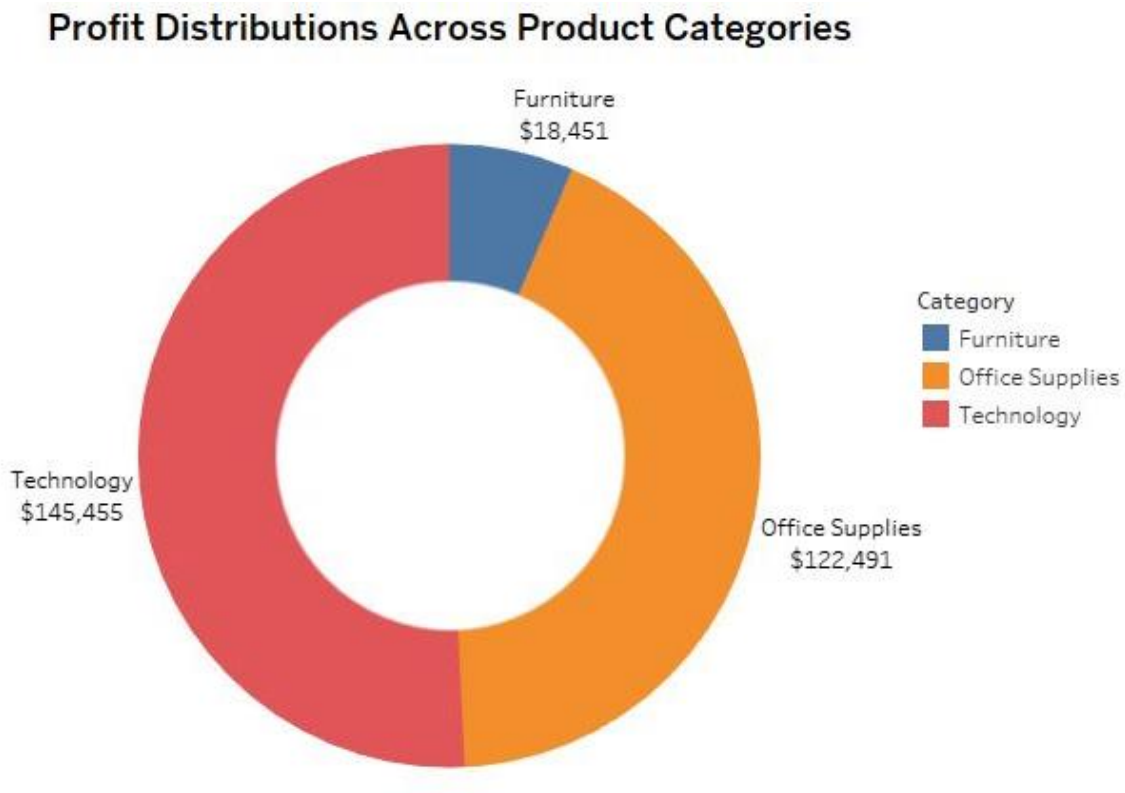
  The key insights derived from the chart are as follows as:

  The highest bin shows 4811 orders where 2 to 3 items are ordered in every single order, indicating a common trend of ordering small quantities or a combination of items.

  There are orders where more than 10 items were ordered, suggesting bulk purchases or large-scale orders, which may be indicative of business or corporate clients.

  Interestingly, single quantity orders are only 899, highlighting a relatively small proportion of orders consisting of a single item. This may imply a preference for purchasing multiple items or bundled products among customers.

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



- **Why did you pick the specific chart?**

  A pie chart was chosen for this visualization because pie charts are effective for visually representing the distribution or proportions of the whole. The 'donut' effect provides a clean and visually appealing representation, making it easy to observe the proportions.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  Product category ‘Technology’ has the highest profit of $145,455, showcasing the robust profitability and strong performance of this category in generating revenue and contributing to overall profit margins.

  Category ‘Furniture’ has the lowest profit of $18,451, indicating that this category may face challenges or lower demand compared to other categories, potentially requiring strategic adjustments or marketing efforts to improve profitability.

  Category ‘Office Supplies’ has a profit of $122,491, suggesting that this category is a significant contributor to overall profits, reflecting steady demand and effective sales strategies within this segment.

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

❖ **How to calculate shipping time?**

Shipping time was calculated by creating a calculated field using the formula as follows:

[Ship Date] - [Order Date]

- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because it effectively compares average shipping time across shipping modes, with each ship mode represented by a bar and the height of the bar corresponding to the shipping time.

- **What is/are the insight(s) found from the chart?**

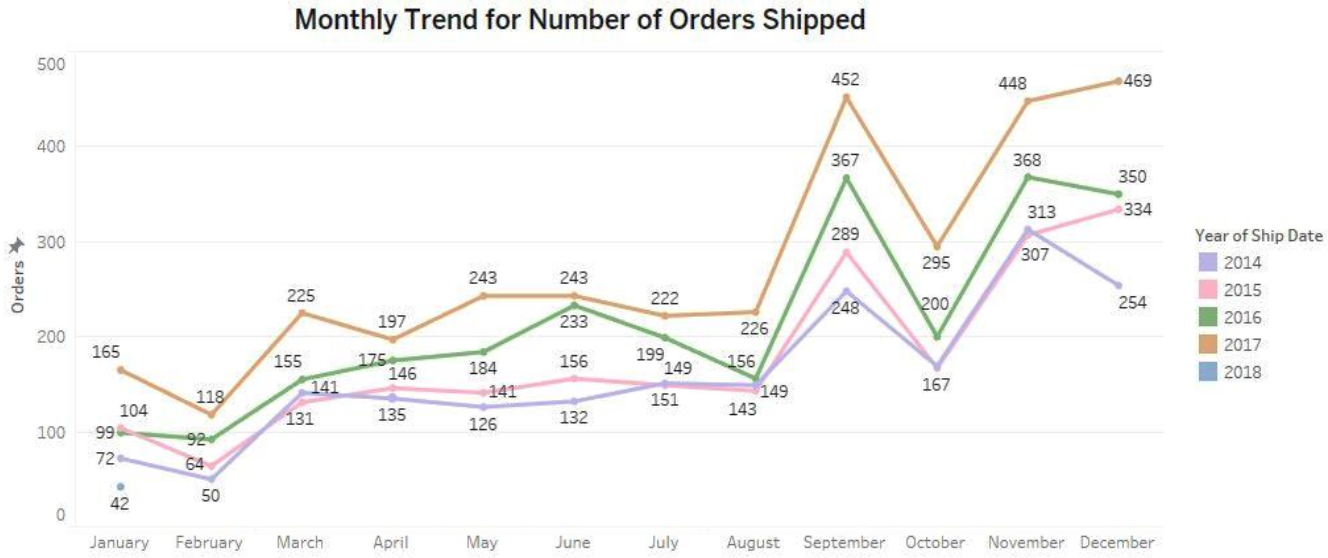
  The key insights derived from the chart are as follows as:

  First class ship mode takes an average of 2 days to ship the order to customers, indicating efficient and expedited delivery service tailored to meet the expectations of customers who prioritize speed and reliability in their shipping preferences.

  Second class ship mode takes an average of 3 days to ship the order to customers, suggesting a slightly longer but still prompt delivery process suitable for customers seeking a balance between speed and cost-effectiveness.

  Standard class ship mode takes an average of 5 days to ship the order to customers, indicating a standard delivery service that may appeal to customers who prioritize cost savings over expedited shipping.

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



- **Why did you pick the specific chart?**

  A line chart was chosen for this visualization because it effectively shows the monthly trend for the number of orders shipped over the years. The lines clearly depicts the overall change in monthly orders over the years, making it easy to identify patterns.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  The chart clearly shows consistent improvement in order numbers, indicating increased demand for products over the years.

  Furthermore, the spikes in order numbers during September, as well as between November and December, could be influenced by promotional events, marketing campaigns or product launches aimed at capitalizing on heightened consumer interest during these periods.

  Conversely, the lower demand for products in February, August, and October may coincide with traditional periods of reduced consumer activity, as individuals may be recovering from holiday spending, focusing on back-to-school preparations, or experiencing seasonal lulls in certain industries.

  Unfortunately, the data was insufficient for the year 2018, so I could not analyze the trends and provide insights for that specific period. However, based on the available data, the observed patterns in order numbers across different months and years offer valuable insights into seasonal variations and consumer behaviour, enabling businesses to make informed decisions and strategic adjustments to their operations and marketing efforts.

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

- **Why did you pick the specific chart?**

  A scatter plot was chosen for this visualization because it effectively reveals the relationship between sales and discount rates across customer segments. Each data point represents a customer segment, enabling us to observe how sales values correspond to discounts.

- **What is/are the insight(s) found from the chart?**

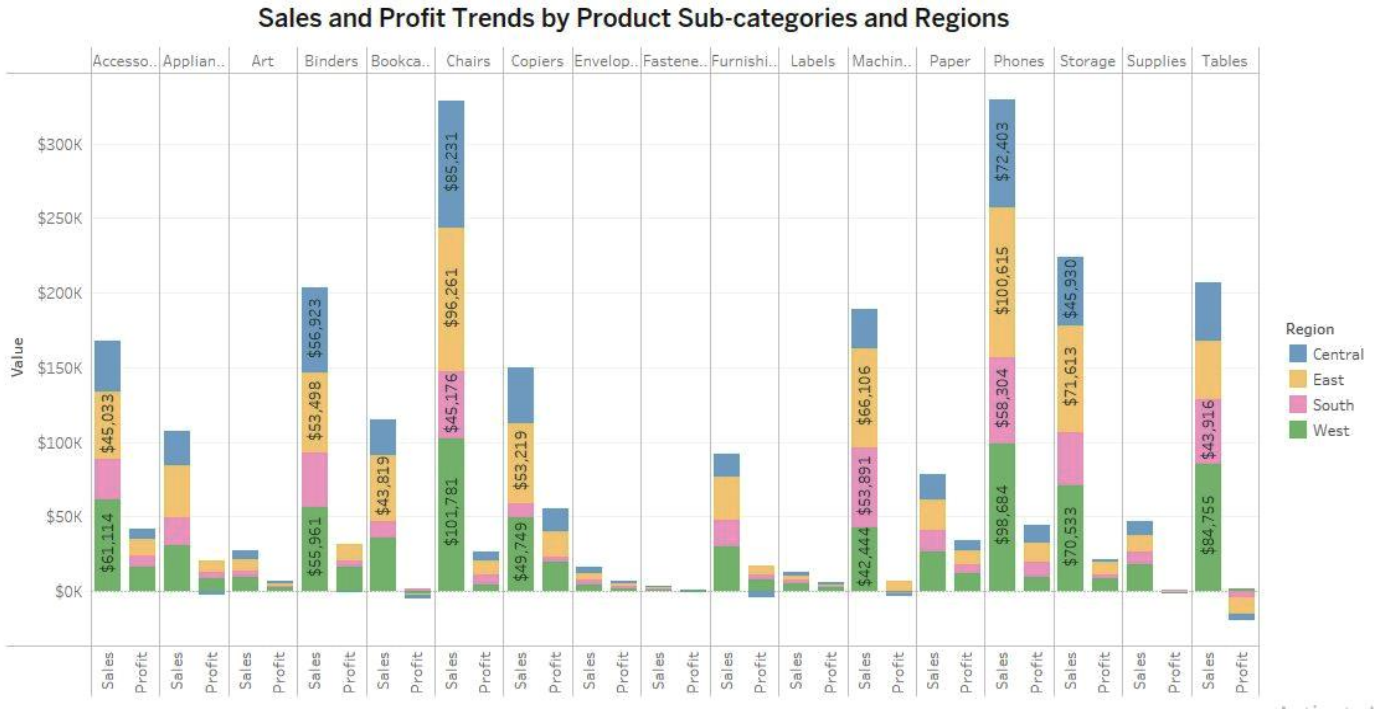
  The key insights derived from the chart are as follows as:

  The Consumer segment has a sale rate of more than $1M and a discount rate of $821, suggesting robust sales performance and effective discount management within this segment, indicative of strong customer engagement and purchasing behaviour.

  The Corporate segment has a sale rate of $706,146 and a discount rate of $478, reflecting steady sales performance and moderate discounting strategies tailored to meet the needs of corporate clients, ensuring competitiveness in the market while maintaining profitability.

  The Home Office segment has a sale rate of $429,653 and a discount rate of $262, indicating steady sales performance and focused discounting strategies aimed at encouraging purchases among home office customers, thereby contributing to continuous revenue generation and fostering customer loyalty.

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



- **Why did you pick the specific chart?**

  A stacked bar chart was chosen for this visualization because it effectively compares categorical data such as sub-categories and regions and allows easy visual comparisons between them. In this case, we can see how sales and profit varies across different product sub-categories for each region.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

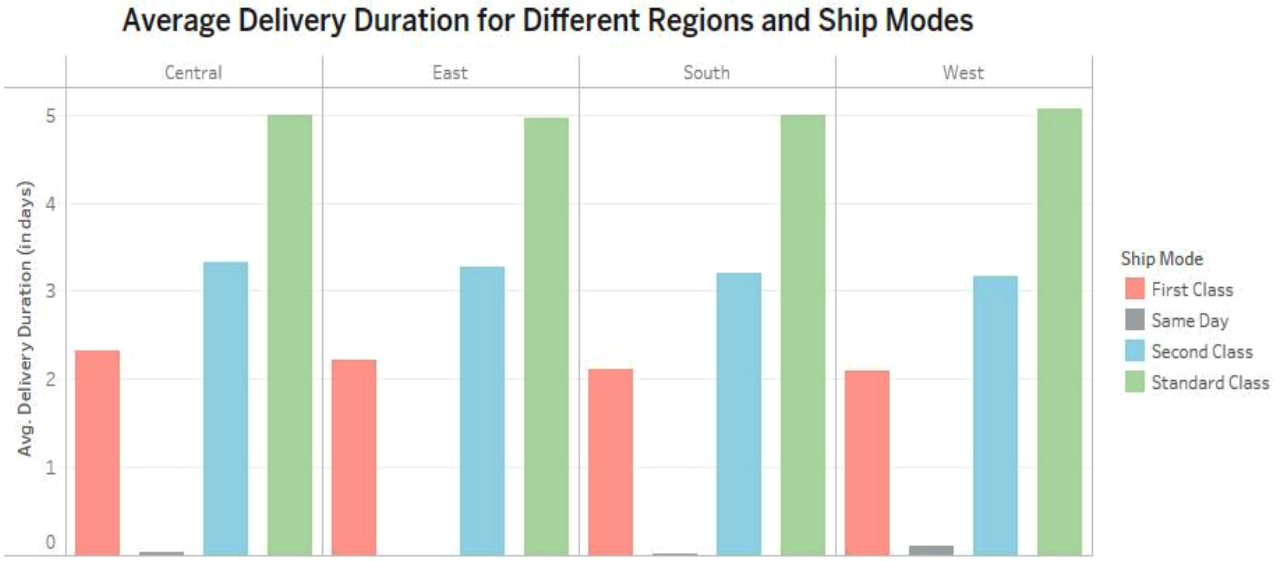
  Product sub-categories like Accessories, Chairs, Copiers and Phones have a high sales-profit ratio across all regions, indicating strong profitability relative to their sales figures. This suggests that these sub-categories are highly efficient in converting sales into profits, contributing positively to overall revenue.

  Some sub-categories like Art, Envelopes, Fasteners, Labels and Paper have low sales across all regions, but the profit ratio is comparatively high. This discrepancy could be attributed to the higher margins associated with these products or efficient cost management strategies, enabling them to generate significant profits despite lower sales volumes.

  Sub-categories like Bookcases, Machines, Supplies and Tables have an average amount of sales but show positive profits across one or two regions. This indicates that these sub-categories are able to generate profitable returns possibly due to targeted marketing efforts or localized demand in the specific regions.

Sub-categories like Appliances, Binders, Furnishing and Machines also have an average amount of sales but show negative profit ratios (loss) specifically in the central region. This suggests that despite moderate sales figures, these sub-categories face challenges in achieving profitability in the central region, indicating potential areas for optimization or cost reduction strategies.

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



- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because it effectively compares average shipping time across shipping modes, with each ship mode represented by a bar and the height of the bar corresponding to the shipping time.

- **What is/are the insight(s) found from the chart?**

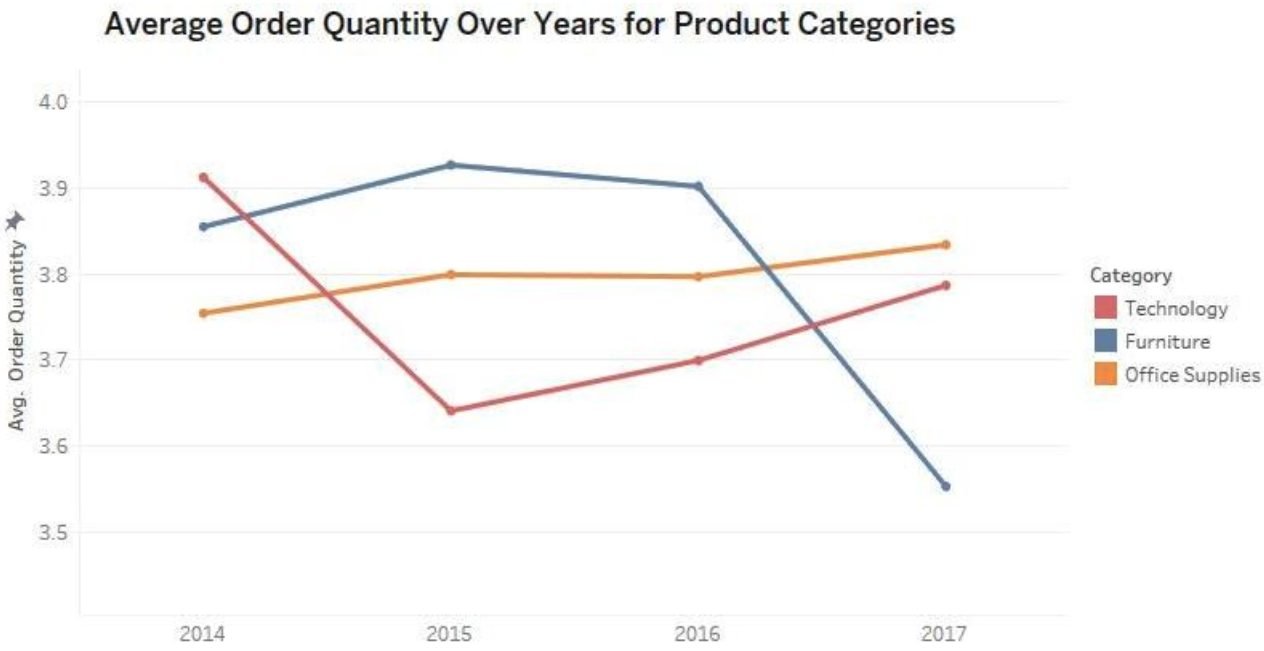
  The key insights derived from the chart are as follows as:

  First class ship mode takes an average of 2 days to ship the order to customers across all regions, indicating efficient and expedited delivery service tailored to meet the expectations of customers who prioritize swift shipping and reliable service.

  Second class ship mode takes an average of 3 days to ship the order to customers across all regions, suggesting a slightly longer but still prompt delivery process. This suggests that customers using second class shipping mode are likely to receive their orders within a reasonable timeframe.

  Standard class ship mode takes an average of 5 days to ship the order to customers across all regions, indicating a standard delivery service. While it may take slightly longer than other shipping modes, standard class shipping still offers reliable and predictable delivery times for customers who prioritize cost-effectiveness over speed.

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



- **Why did you pick the specific chart?**

  A line graph was chosen for this visualization because it effectively shows the change in average order quantity over the time for various product categories. A line plot is suitable for showing the progression of a numeric variable (Avg. Order Quantity) across a continuous axis (years).

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  For the 'Technology' category, a decrease in the order quantity was observed in 2015, but it increased after that, which shows a potential recovery in demand for technological products following a temporary decline, possibly due to market trends or product innovations.

  Interestingly, 'Office Supplies' has maintained a consistent trend over the years and is following an increasing route, which can be attributed to factors such as steady consumer demand, effective marketing strategies, or the introduction of new office supply products.

  Whereas, category 'Furniture' initially started with similar order quantities as other product categories. It remained stable for 2015 and 2016, but then gradually decreased with time. The reason could be attributed to shifts in consumer preferences, changes in market dynamics, or increased competition in the furniture industry, prompting a decline in demand over time.

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

- **Why did you pick the specific chart?**

  A scatter plot was chosen for this visualization because it effectively shows the relationship between order quantity and discount rates across customer segments. Each data point represents a customer segment, enabling us to observe how discount rates correlated to the ordered quantities.

- **What is/are the insight(s) found from the chart?**

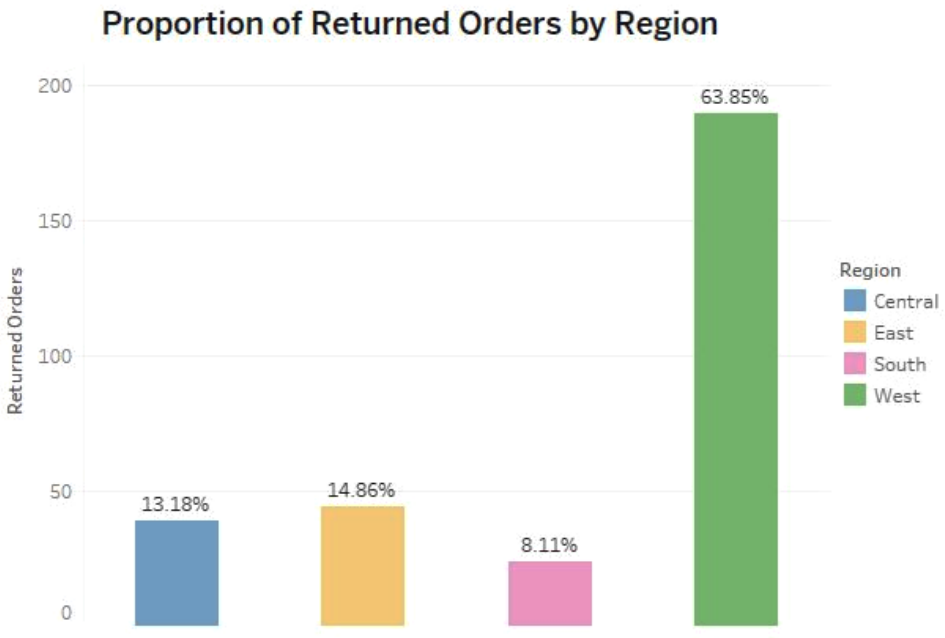
  The key insights derived from the chart are as follows as:

  A discount of $262 was given for the 6,744 order quantities in the customer segment, which indicates a relatively moderate discount rate compared to the total number of orders. This suggests that discounts in this segment may be selectively applied or relatively small in magnitude compared to other segments.

  A discount of $478 was given for the 11,608 order quantities in the corporate segment, highlighting a moderate discount rate applied to a substantial number of orders. This suggests that discounts in this segment may be tailored to meet the needs of corporate clients while maintaining profitability.

  A discount of $821 was given for the 19,521 order quantities in the consumer segment, showcasing a significant discount rate applied to a large volume of orders within this segment, highlighting efforts to attract and retain individual consumers through competitive pricing strategies.

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



❖ **How to calculate returned orders?**

Returned orders were calculated by creating a calculated field using the formula as follows:

IF [Returned] = 'Yes' THEN "1"

ELSEIF [Returned] > 'No' THEN "0"

END

- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because it effectively compares the number of orders returned across different regions, with each region represented by a bar and the height of the bar corresponding to the returned orders.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows:

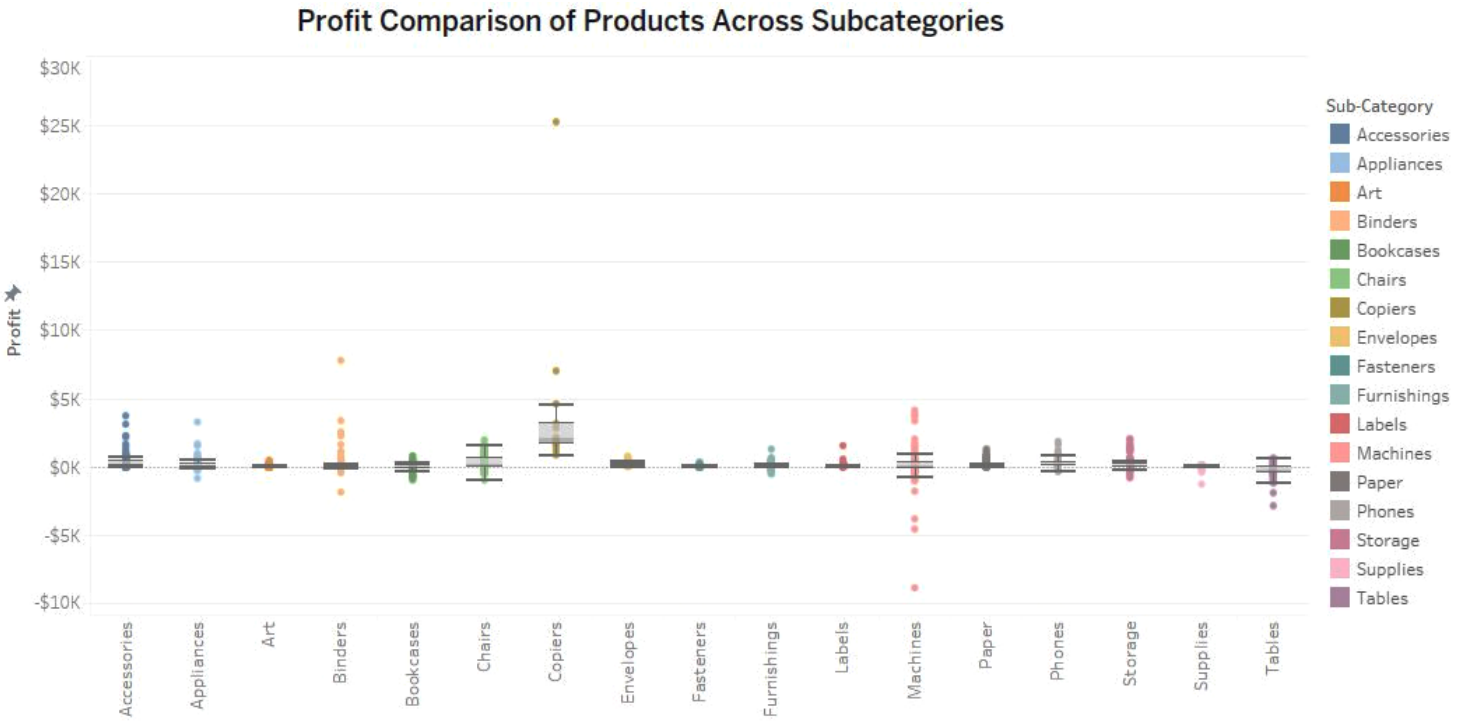
  In the Central region, 39 returned orders are observed, contributing to 13.18% of the total orders returned. This indicates a moderate level of returns in this region, suggesting potential areas for improvement or investigation into customer satisfaction and product quality.

  In the East region, 44 returned orders are observed, contributing to 14.86% of the total orders returned. This indicates a similar trend of returns as observed in the Central region, highlighting the need for attention to customer experience and order fulfilment processes.

In the South region, 24 returned orders are observed, contributing to 8.11% of the total orders returned. This indicates a relatively lower proportion of orders being returned within the South region compared to other regions, which may indicate more favourable customer experiences or operational efficiencies in this area.

In the Central region, 189 returned orders are observed, contributing to 63.85% of the total orders returned. This indicates a significant portion of returns originating from the Central region, signalling potential issues with product quality, shipping, or customer service that require thorough investigation.

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



- **Why did you pick the specific chart?**

  A box chart was chosen for this visualization because the box plot is particularly effective for comparison of the profit distributions for different products within each sub-category. Box plots readily highlight outliers or extreme values within the profit distribution, which could indicate exceptional performance or potential issues affecting profitability.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

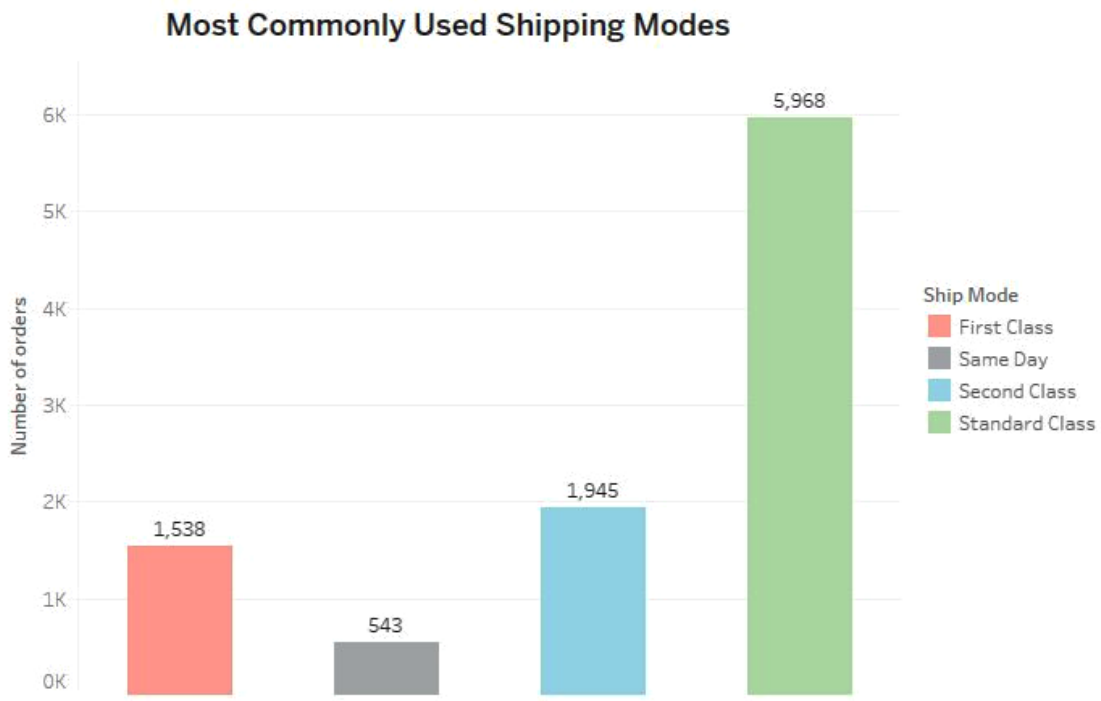
  Subcategories like Binders, Copiers, and Machines exhibit outliers, showcasing products with exceptionally high profit and loss. This facilitates the identification of exceptional performers and areas for potential improvement.

Products in subcategories like Accessories, Appliances, and Storage consistently yield profits ranging from $3k - $5k. This signifies stable and reliable revenue generation across these product categories.

Subcategories including Bookcases, Chairs, Furnishings, Labels, Paper, and Tables generate profits averaging between $1k - $2k. This indicates moderate profitability and suggests areas where optimization strategies could enhance returns.

On the contrary, products in subcategories such as Art, Envelopes, Fasteners, and Supplies have notably low profits, typically below $1k. This underscores challenges in revenue generation and highlights the need for reassessment or strategic adjustments.

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



- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because the bar chart allows for a clear and straightforward comparison of the frequency or count of each shipping mode. The length of each bar in the chart directly corresponds to the frequency of the shipping mode, providing a visual representation of the distribution.

- **What is/are the insight(s) found from the chart?**

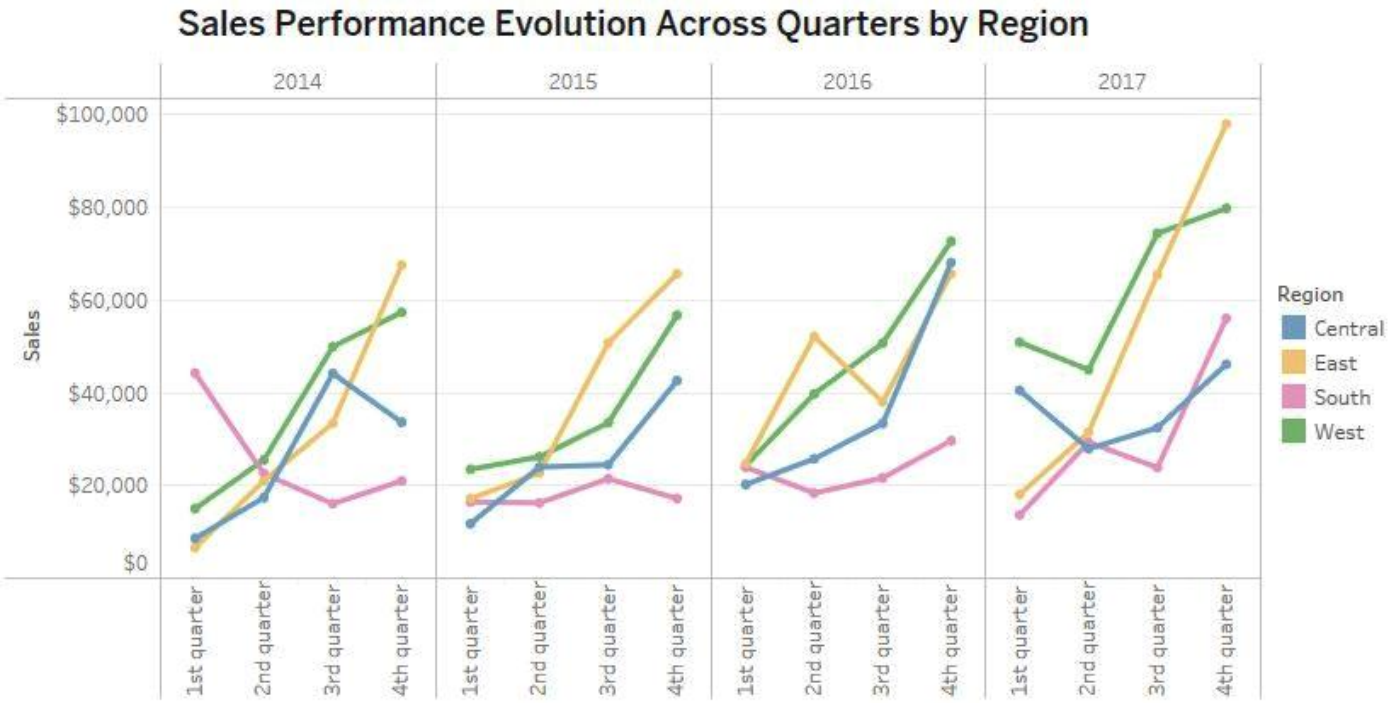
  The key insights derived from the chart are as follows as:

The chart clearly shows the Standard class shipping mode is the most preferable shipping option among customers, with a total of 5,968 orders shipped, indicating most of the customers follow standard delivery options and prioritize cost savings over expedited shipping.

First class and Second class shipping modes, followed by Standard class have nearly equivalent numbers of shipped orders, suggesting that some customers choose timely and reliable delivery options.

In contrast, the Same day shipping mode records the fewest number of orders, indicating that customers may prioritize cost or convenience over immediate delivery. This could be due to the higher costs associated with same day delivery or the perception among customers that same day shipping involves unnecessary expenses.

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



- **Why did you pick the specific chart?**

  A line graph was chosen for this visualization because it effectively shows the change in sales over the time for various regions. A line plot is suitable for showing the progression of a numeric variable (Sales) across a continuous axis (Quarters of a years).

- **What is/are the insight(s) found from the chart?**

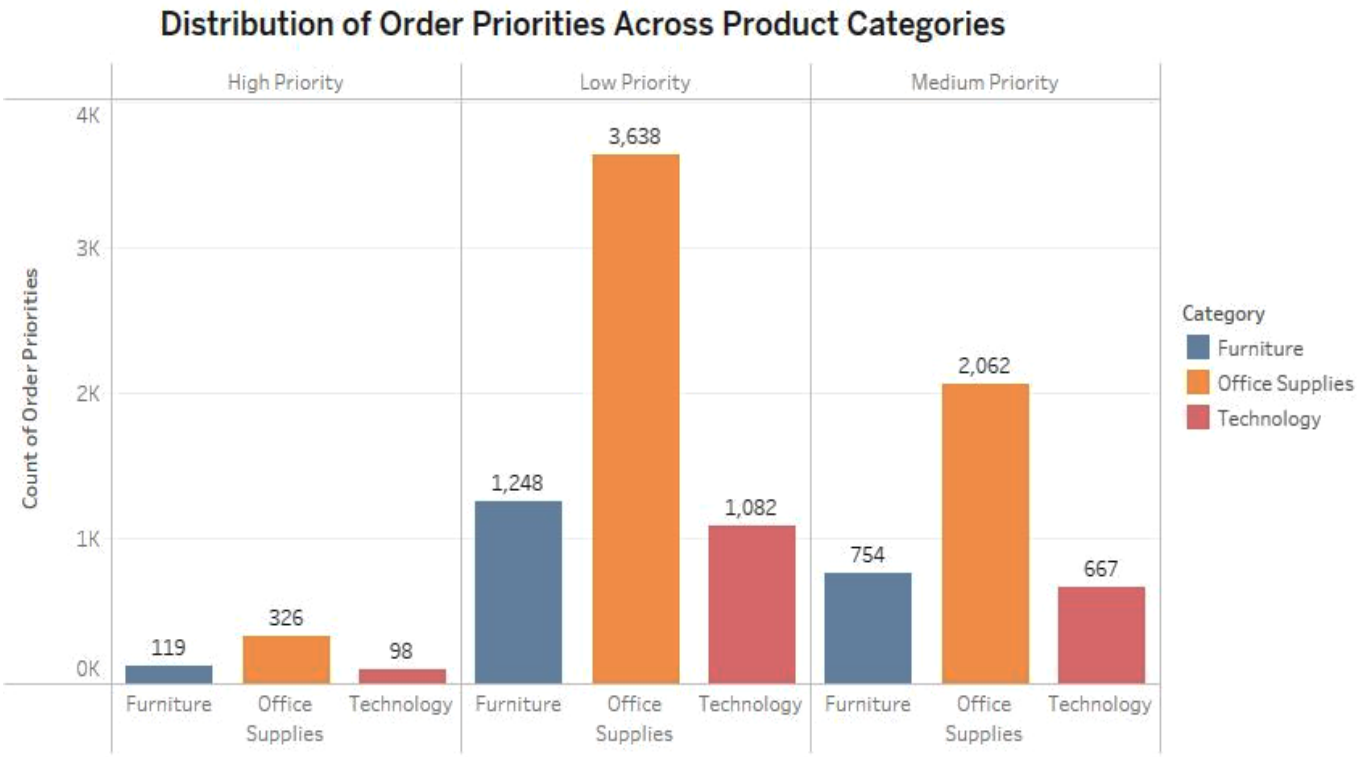
The key insights derived from the chart are as follows as:

Surprisingly, sales in the East region performed exceptionally well and followed an increasing trend for all quarters from 2014 to 2017. This indicates a strong and sustained demand for products in the East region, possibly driven by factors such as economic development or effective marketing strategies in the region.

Sales in the South region started with strong sales in the first quarter of 2014, followed by a gradual decline until the third quarter of 2017. However, sales for this region rebounded thereafter and started showing signs of recovery. This suggests that the South region experienced challenges or fluctuations in demand during the specified period but managed to regain momentum through strategic interventions or market adjustments.

Sales in the Central and West regions remained consistently stable for all quarters until 2016, with no significant deviations in performance. However, sales in the West region began outperforming those in the Central region from the first quarter of 2017 onward. This indicates a potential shift in market dynamics or consumer preferences favouring the West region.

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



❖ **How to calculate order priorities?**

Order priorities were calculated by creating a calculated field using the formula as follows:

IF [Ship Mode] = 'Same day' THEN 'High Priority'

ELSEIF [Ship Mode] = 'First class' OR [Ship Mode] = 'Second class' THEN 'Medium Priority'

ELSE 'Low Priority'

END

- **Why did you pick the specific chart?**

  A bar chart was chosen for this visualization because it effectively compares the number of order priorities across product categories, with each category represented by a bar and the height of the bar corresponding to the number of orders priorities.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  In high order priorities, the product category ‘Furniture’ recorded 119, ‘Office Supplies’ had 326, and ‘Technology’ showed 98 order priorities, all indicating a preference for same day shipping. This suggests a need for quick delivery among customers in these categories, possibly driven by urgent requirements or time-sensitive requirements.

  In medium order priorities, ‘Furniture’ accounted for 754, ‘Office Supplies’ had 2062, and ‘Technology’ showed 667 order priorities, reflecting a substantial volume of orders preferring first and second class shipping. This highlights a balance between delivery speed and cost-effectiveness for customers across these categories.

  In low order priorities, ‘Furniture’ and ‘Technology’ categories exhibited similar numbers of order priorities, while ‘Office Supplies’ had a significantly higher count of 5904 order priorities, predominantly opting for standard shipping. This indicates a preference for cost-efficient delivery methods among customers in the ‘Office Supplies’ category compared to 'Furniture' and 'Technology'.

  Overall, the distribution of order priorities reflects the diverse needs and preferences of customers across different product categories, highlighting the importance of offering flexible shipping options to accommodate varying customer requirements.

28. What is the relationship between discounts and sales?

- **Why did you pick the specific chart?**
  A scatter plot was chosen for this visualization because it effectively shows the relationship between discount and sales for each product. Each data point represents products, enabling us to observe how discount rates correlated to sales.

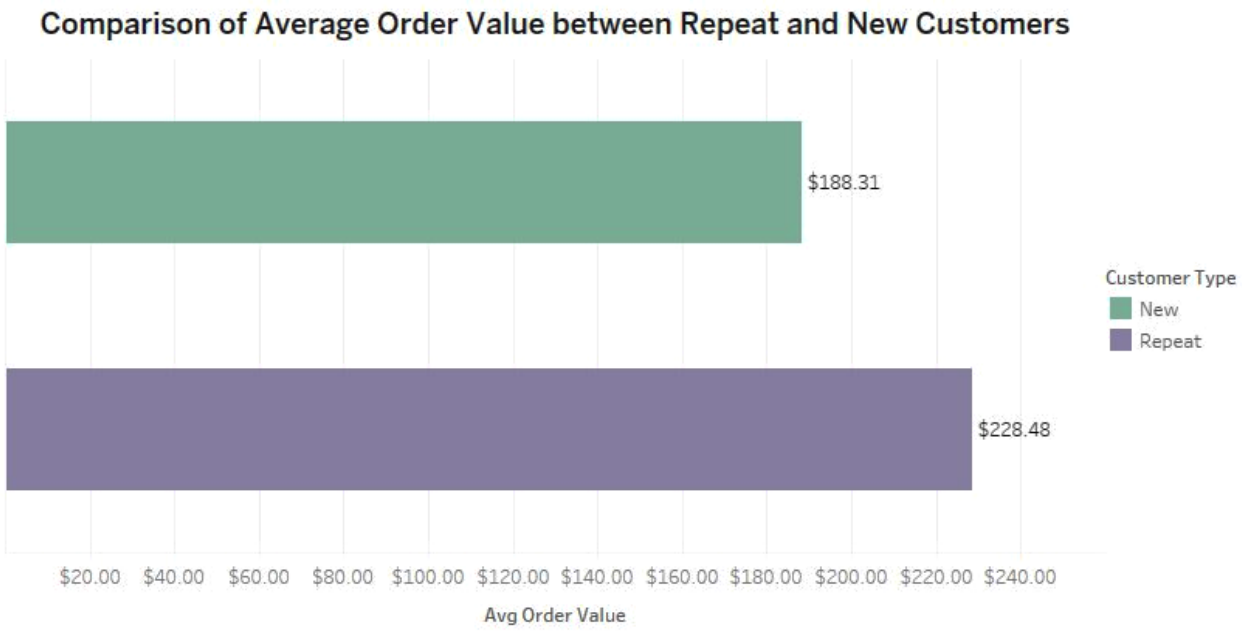
- **What is/are the insight(s) found from the chart?**
  The key insights derived from the chart are as follows as:

  The scattered distribution of data points indicates that most data points cluster between $0.00 and $2.00 on the x-axis (Discount) and between $0.00 and $10.00k on the y-axis (Sales). This suggests a common trend of applying low discounts to products with relatively modest sales volumes.

  There is an outlier showing a total sale of more than $60.00k, indicating an exceptional high-value transaction or an anomaly in the dataset that needs further investigation.

  Many points are widely scattered around the trend lines, suggesting a considerable variability in the relationship between discount rates and sales volumes across different products. This variability may be influenced by factors such as product demand, pricing strategies, or customer preferences.

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



- **How to analyze customer type?**
  Customer type were analyzed by creating a calculated field using the formula as follows:

```
IF { FIXED [Customer ID] : COUNTD([Order ID]) > 1 }
THEN "Repeat"
ELSE "New"
END
```

- **How to calculate average order value?**
  Average order value was calculated by creating a calculated field using the formula as follows:

```
AVG(
  IF [Customer Type] = "Repeat" THEN
    { FIXED [Customer ID] : AVG([Sales]) }
  ELSE
    { FIXED [Customer ID] : AVG(IIF([Customer Type] = "New", [Sales], NULL)) }
  END
```

)

- **Why did you pick the specific chart?**

  A horizontal bar chart was chosen for this visualization because the horizontal bars facilitate comparison of average order value across repeat and new customers, allowing for easy visual interpretation and comparison of the data. Additionally, the horizontal layout provides ample space for displaying detailed labels and annotations, enhancing the clarity of the visualization.

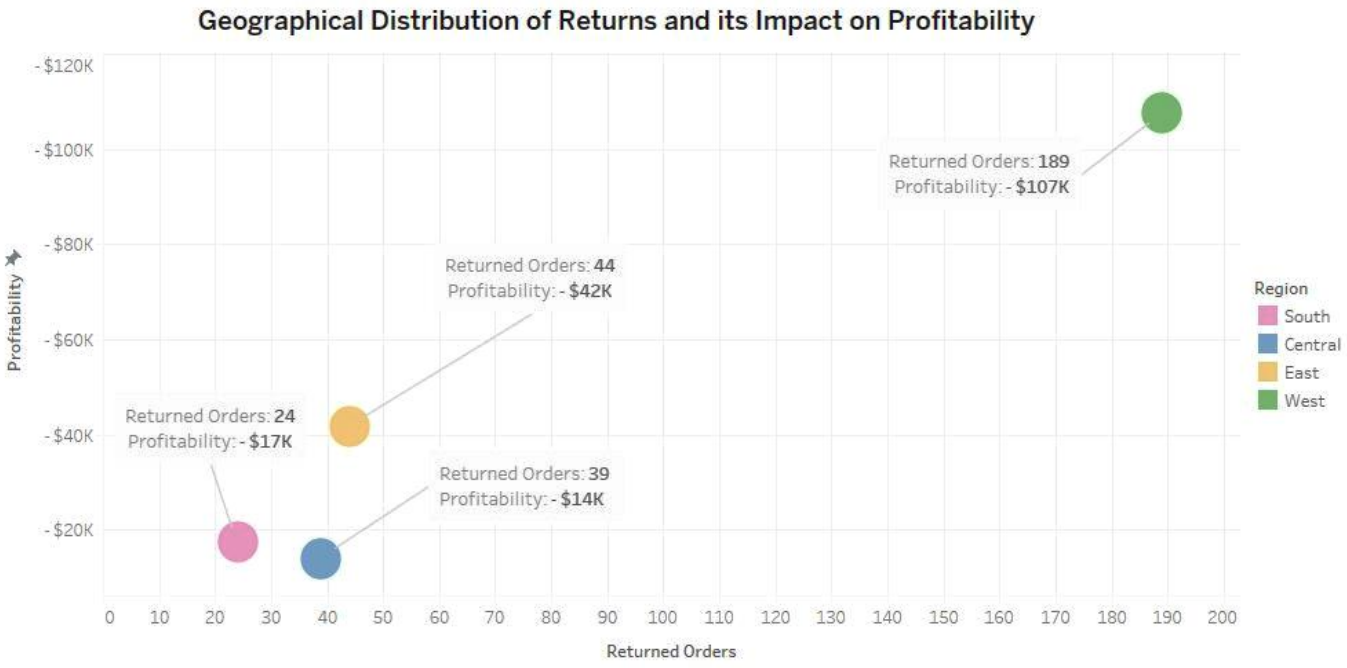
- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  An average order value of $188.31 is observed for new customers, suggesting that they tend to make relatively smaller purchases or may be more price-sensitive in their buying behaviour. This could potentially indicate that new customers are exploring the product offerings or are making initial, low value purchases.

  On the other hand, an average order value of $228.48 is observed for repeat customers, indicating that they tend to make larger purchases or have higher spending habits compared to new customers. This higher average order value may suggest that repeat customers have established trust and are more likely to engage in repeat transactions or make bigger purchases over time.

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



- **How to calculate profitability impact?**

Profitability impact of returned orders was calculated by creating a calculated field using the formula as follows:

```
SUM(
    IF [Returned Orders] = 'Yes' THEN -[Sales]
    ELSE [Sales]
END
)
```

- **Why did you pick the specific chart?**

  A scatter plot was chosen for this visualization because it effectively shows the relationship between returned orders and its effect on profitability. Each data point represents regions, enabling us to observe how returned orders correlated to profitability.

- **What is/are the insight(s) found from the chart?**

  The key insights derived from the chart are as follows as:

  South region experienced 24 returned orders, impacting profitability by around $17k, suggesting a moderate effect on revenue in this region. Possible factors contributing to this impact could include product quality issues, shipping delays or customer dissatisfaction with delivered items.

  In the Central region, 39 returned orders were recorded, affecting profitability by approximately $14k, indicating a similar but slightly higher impact on revenue compared to the South region. Possible reasons such as product defects, incorrect orders or bad customer experiences might have contributed to this impact.

  The East region experienced 44 returned orders, affecting profitability by approximately $42k, indicating a more significant impact on revenue compared to other regions. Possible factors contributing to shipping errors, damaged goods etc.

  In the South region, 189 returned orders were observed, affecting profitability by approximately $107k, indicating a substantial effect on revenue, possibly indicating higher return rates or larger order sizes in this region.