Guidelines for Data Visualization and Analysis Project

About the Project:

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

Skills Required:

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

Deliverables:

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

Rubrics for Assessment:

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

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

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

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

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

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

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

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

Dataset Link:

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

Please keep in mind:

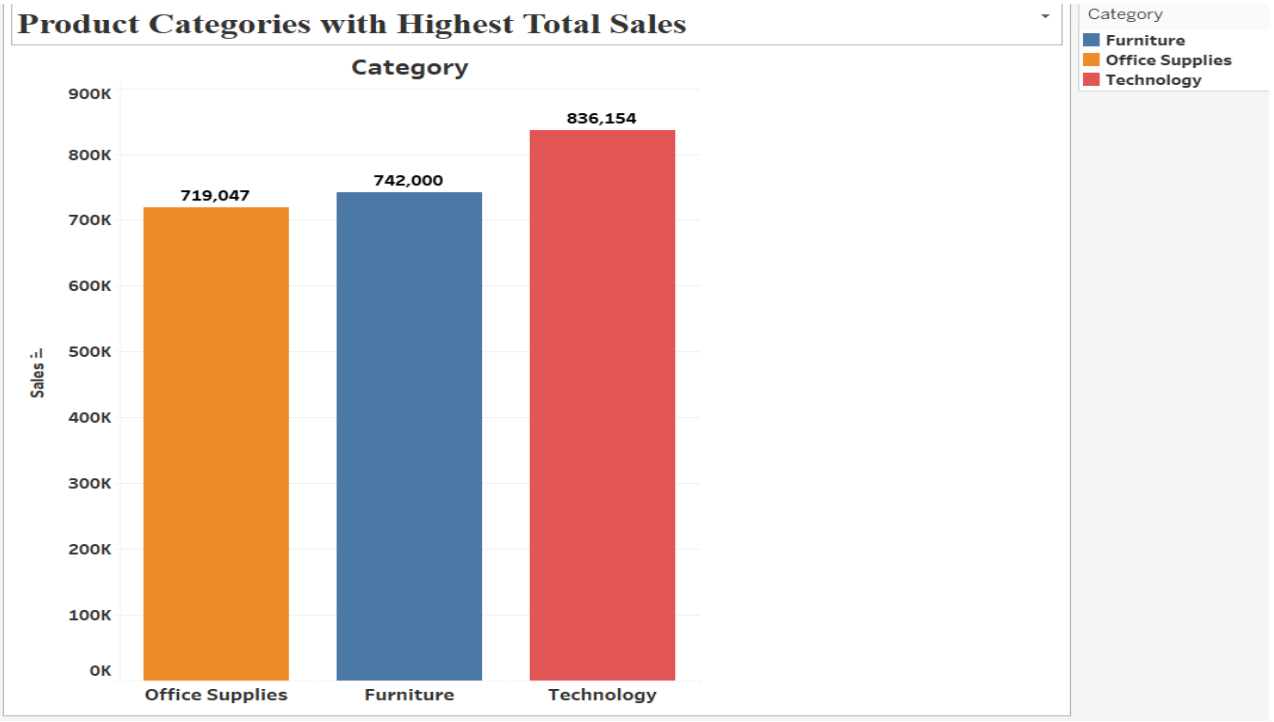
1. **Answer Completion**: Ensure that you furnish answers for all any 30 questions and build charts for them.
2. **Encouraged Creativity**: Don't hesitate to employ visuals, creative elements, or any other innovative approaches to enhance the quality of your responses.

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

Good luck!

Questions:

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



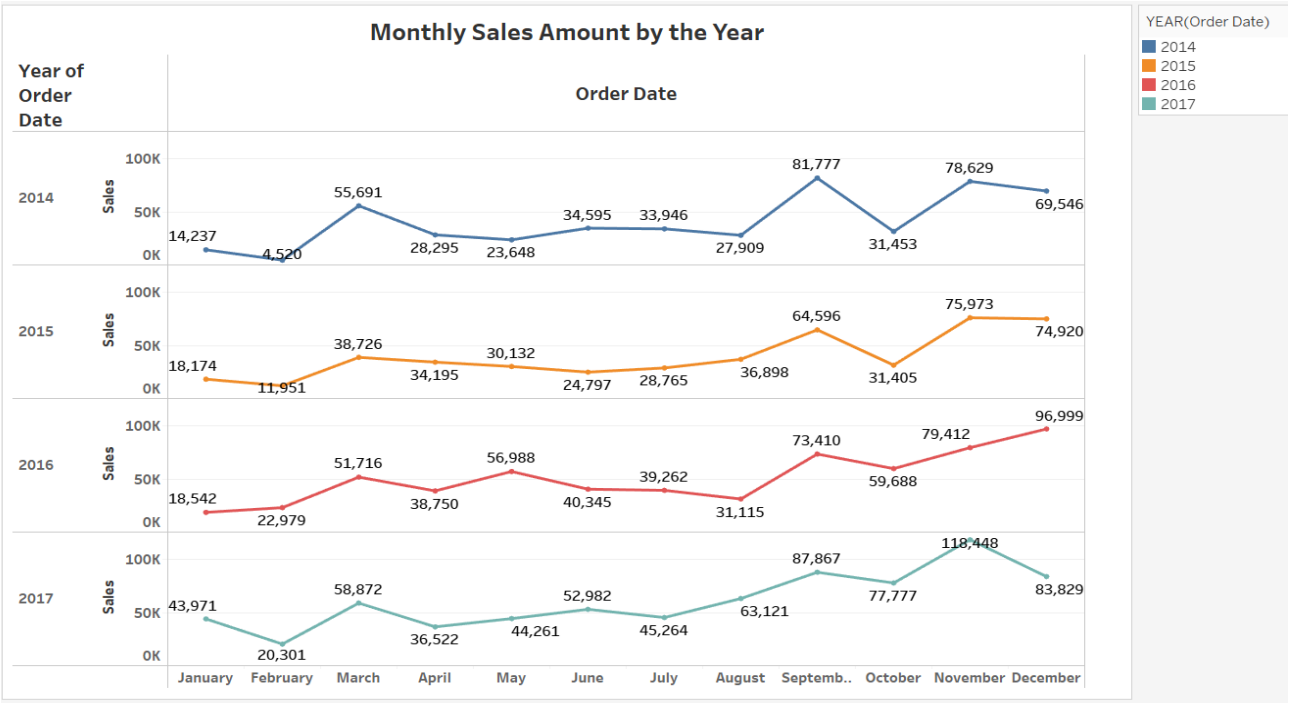
Why to choose this specific chart?

A bar chart was chosen for this visualization because a bar chart is simple, easy to understand and effective for comparing different categories. In this bar chart length of each bar represents the sales value of different categories.

Description:

This chart clearly shows that 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 followed by categories 'Furniture' which shows 742,000 sales and 'Office Supplies' which shows 719,047 sales.

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



Why to choose this specific chart?

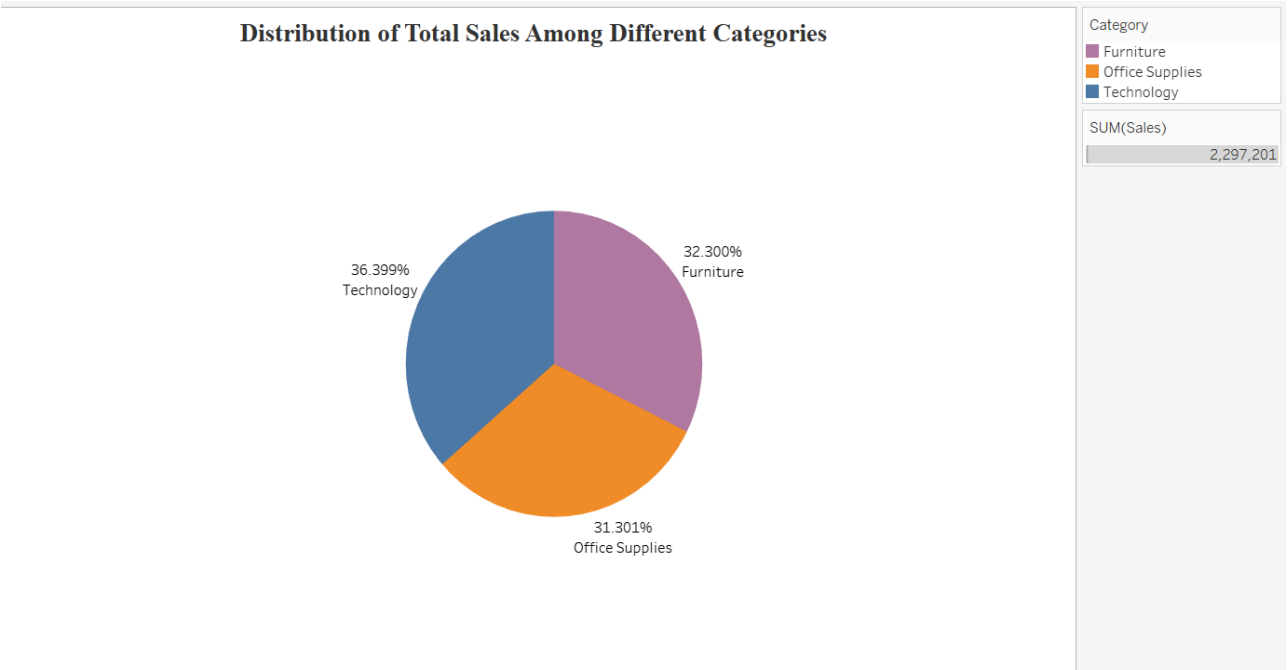
A line chart was chosen for this visualization because line charts effectively show the trends, patterns and fluctuations over time. The lines clearly depicts the overall change in monthly sales over the years, making it easy to identify patterns and trends.

Description:

Major spikes are evident in the months of March, September and November of all the years suggesting high demand for products during these periods because of some factors such as holidays, festivals or special events.

Conversely, February month shows the least demand for products in all the years except 2016 which shows the least demand of products in January month. This may be because they may represent periods of lower consumer activities, absence of major holidays or seasonal downturns in certain industries.

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



Why to choose this specific chart?

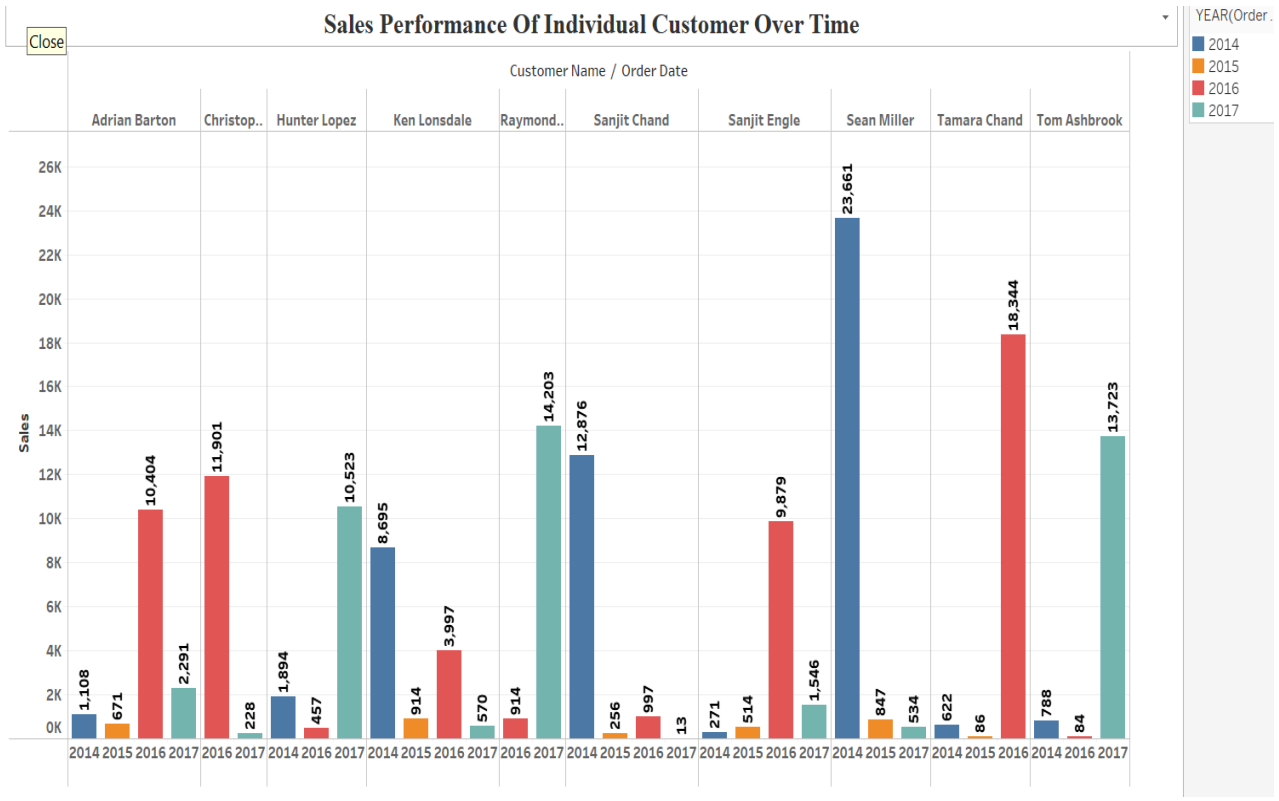
A pie chart was chosen for this visualization because pie charts are excellent for showing the relationship of parts to a whole, are simple and easy to understand. Each slice represents a proportion of the total , making it easy to see how different segments compare to the whole.

Description:

From this pie chart we found that 'Technology' is the product category with the highest total sales amount, accounting for 36.399% of the total, indicating that it has the biggest share of the total sales followed by 'Furniture' making up 32.300% of the total sales and 'Office supplies' have the lowest total sales amount at 31.301%, implying a smaller share of the overall sales compared to the other categories.

This shows that the 'Technology' products are in higher demand and 'Office Supplies' products are in lower demand from the rest.

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



Why to choose this specific chart?

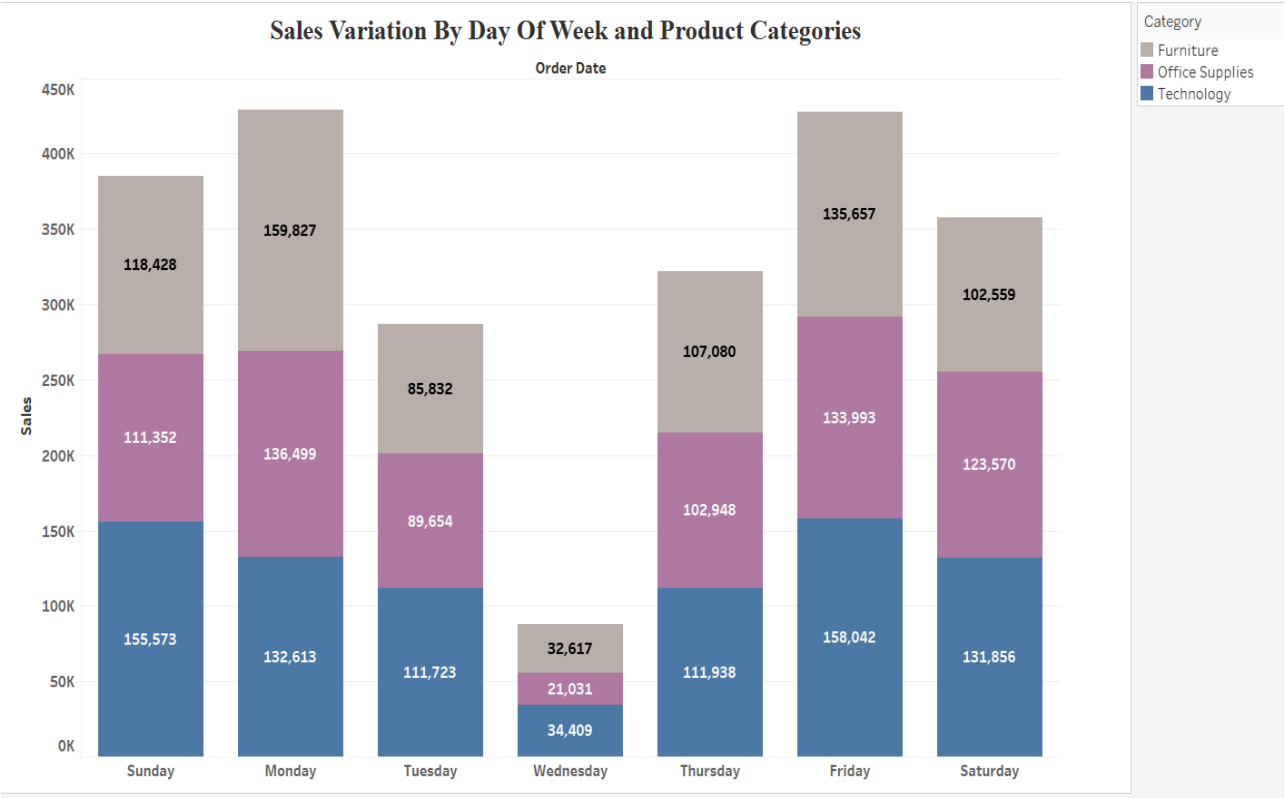
A bar chart is an appropriate choice for this visualization because bar charts are effective for comparing different categories or groups of categories and are easy to understand.

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.

Description:

The heights of the bars differ considerably, indicating a diverse range of sales performance among customers in every year.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 and the lowest sales of 13 was recorded for a customer named 'Sanjit Chand' in the year 2017.

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



Why to choose this specific chart?

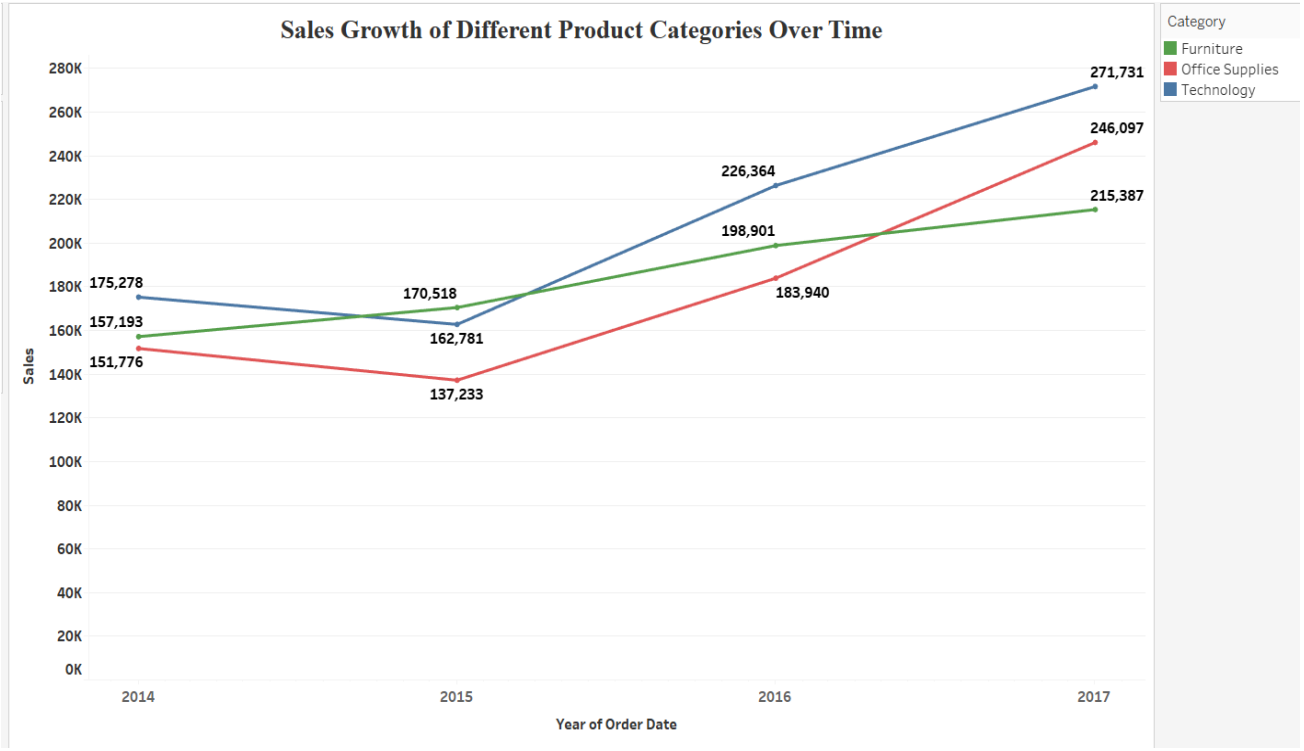
A Stacked bar chart was chosen for this visualisation because it effectively compares the variation in sales of different product categories over day of week.

Description:

The chart clearly indicates higher sales occurring on Monday, Friday, Saturday and Sunday for most product categories. This suggests that days in the starting of the week and ending of the week generally have higher sales, when consumers may have more time for shopping or leisure activities. Wednesday exhibits the lowest sales across all categories followed by Tuesday and Thursday. It might be because of not getting enough time for shopping due to work or maybe they didn't have any plans for shopping at that time.

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 to choose this specific chart?

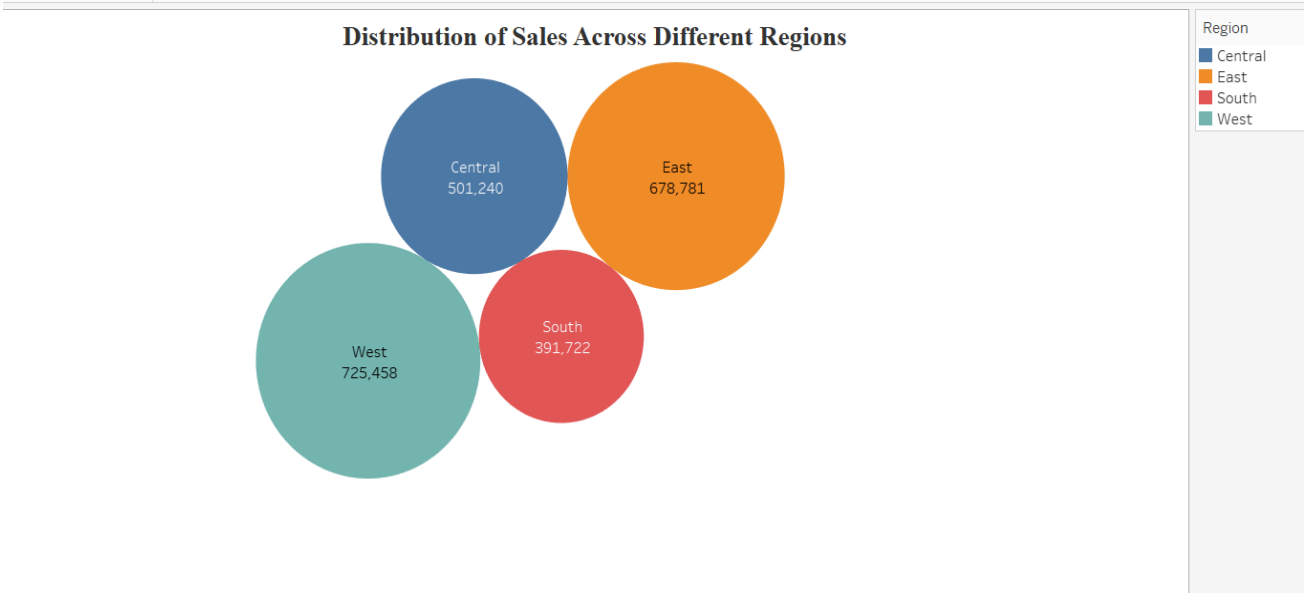
A line chart was chosen for this visualisation because it effectively shows the trends in sales growth of different product categories over time. Each colour on the chart will represent a different product category, enabling easy comparison and identification of growth or decline in sales over different periods.

Description:

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 'Technology' category consistently maintains the highest sales compared to other categories over time, which indicates sustained consumer demand for technology-related products.

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



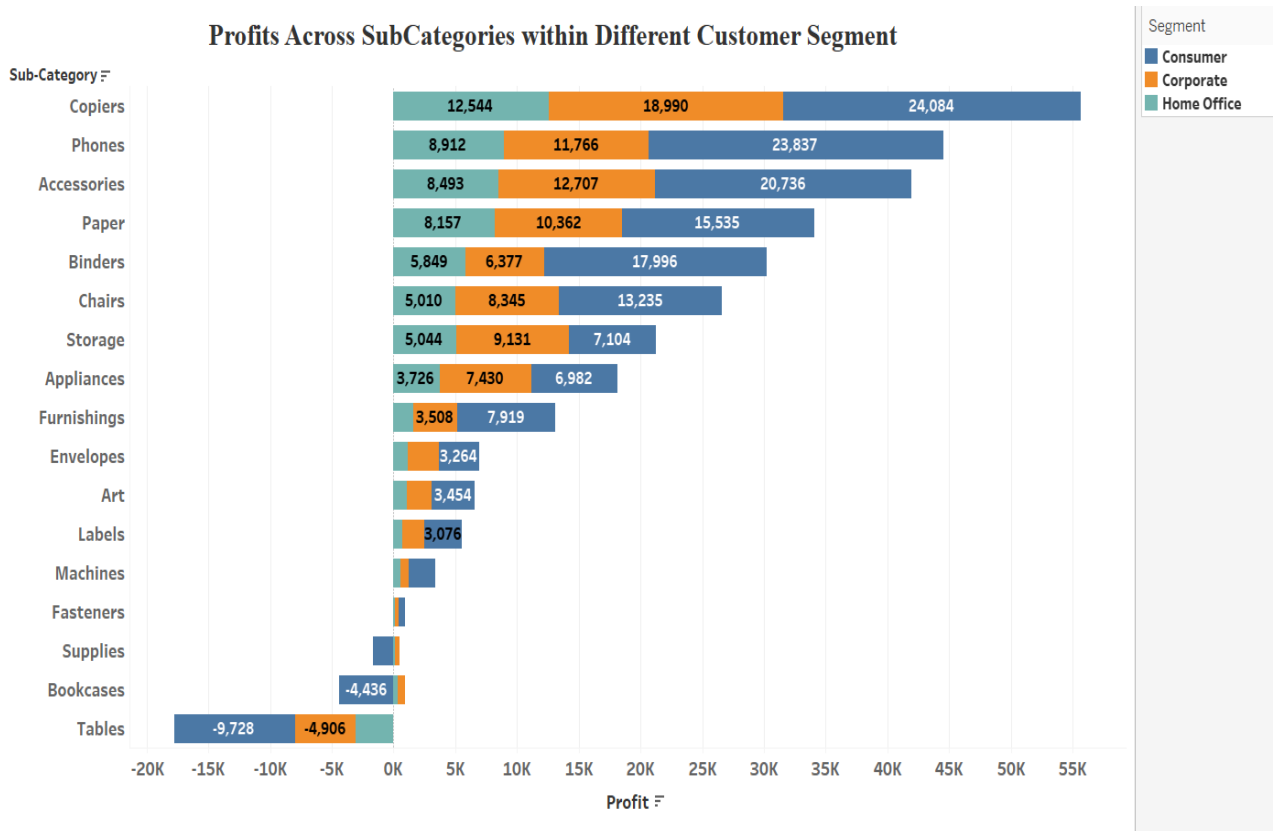
Why to choose this specific chart?

A packed bubble chart was chosen for this visualisation because they are excellent for comparing the relationships between different data points. Here, sizes of each bubble in the chart corresponds to the magnitude of sales in each region. This allows for easy comparison of sales performance visually.

Description:

The West region has the highest sales distribution of 725,458, indicating strong market demand and significant revenue generation from this region. 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. Central and East regions maintain a good sales distribution 501240 and 678781 respectively.

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



Why to choose this specific chart?

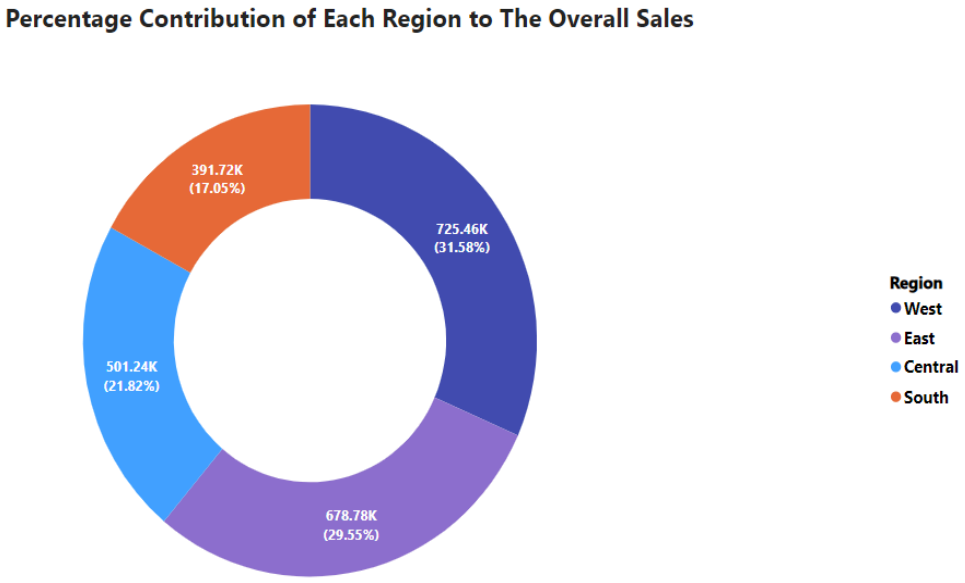
A stacked bar chart was chosen for this visualization because it effectively compares categorical data such as sub-category & customer segment against profit and allows easy visual comparisons between them. In this case, we can see how profit varies across different sub-categories for each customer segment.

Description:

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

Sub-category 'Tables' has the negative profit across all customer segments, indicating potential challenges or lower demand for this particular product. On the other hand Subcategories like Bookcases and Supplies have negative profit for 'Consumer' segment and very low profit range for other customer segments.

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



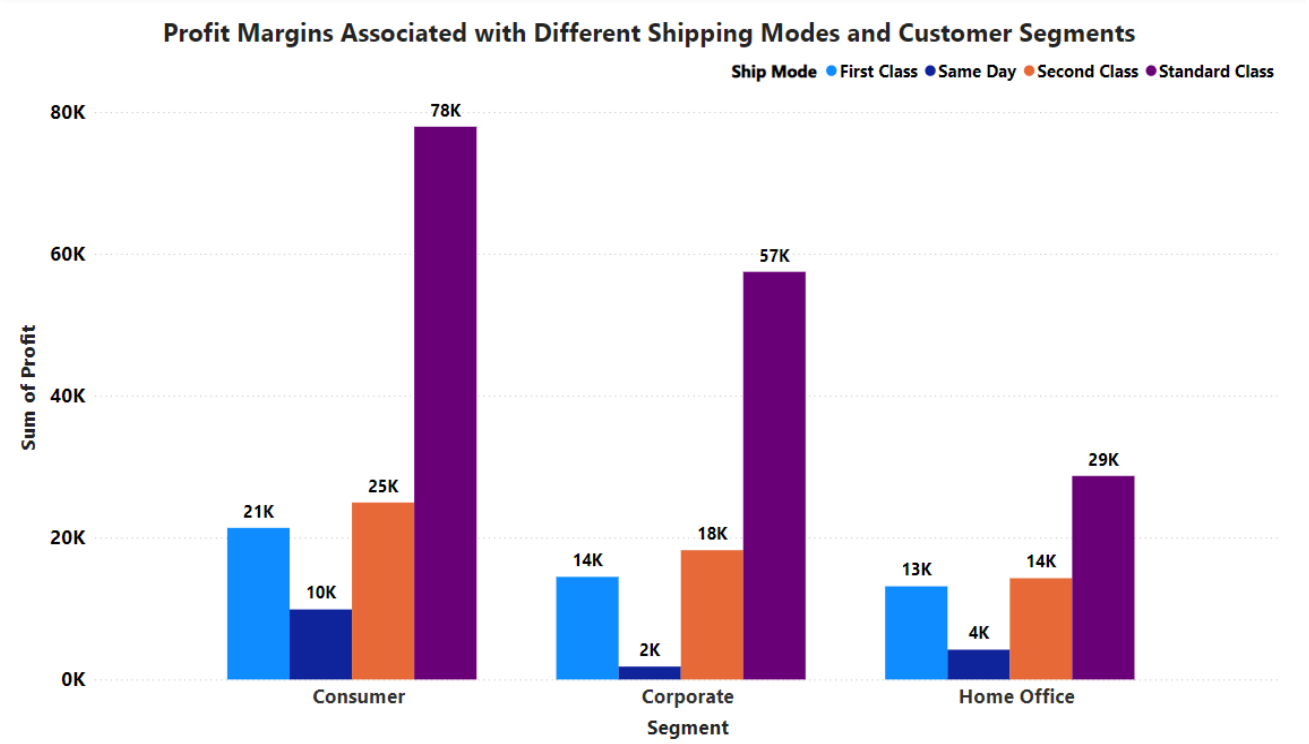
Why to choose this specific chart?

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.

Description:

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. East regions have a good sales contribution of 29.55% followed by the Central region, showing a contribution of 21.82%.

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



Why to choose this specific chart?

A clustered 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.

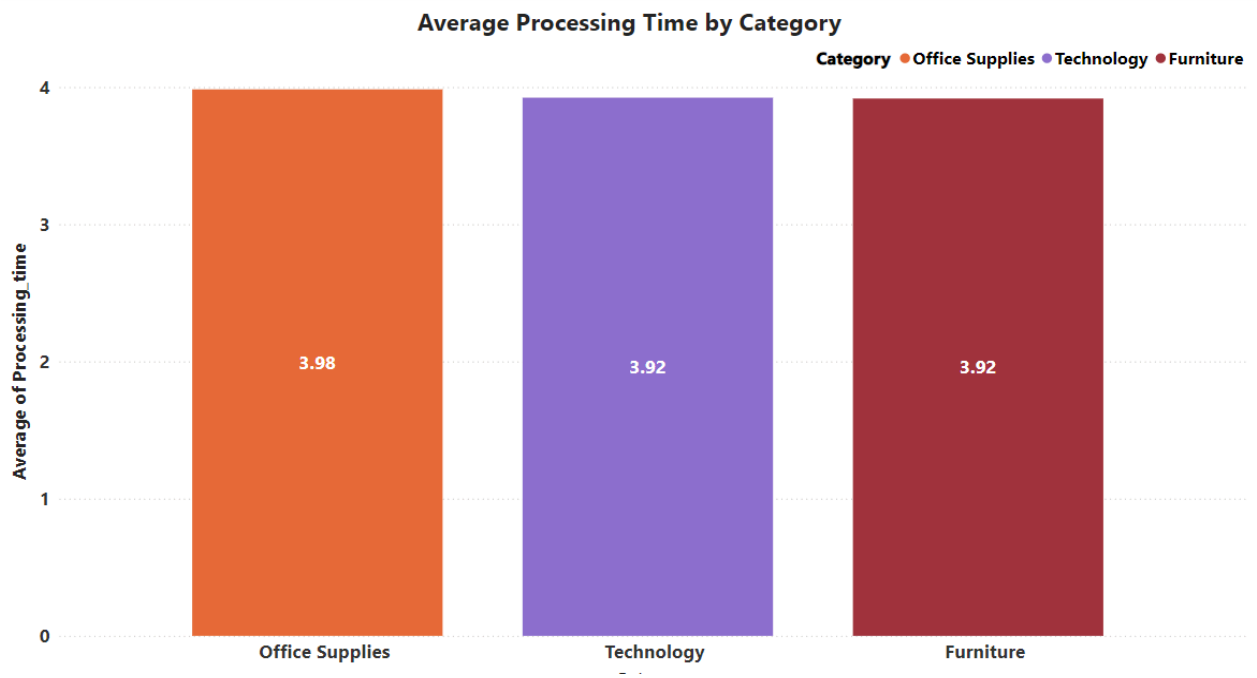
Description:

'Standard class' shipping mode shows the highest profit margin across all customer segments, it may be because of large number of orders had been shipped by this mode.

Conversely, 'Same day' shipping mode shows the lowest profit margin across all customer segments, it may be due to less number of orders had been shipped by this mode.

First class and Second class shipping modes show almost similar profit margins across all customer segments.

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



Why to choose this 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.

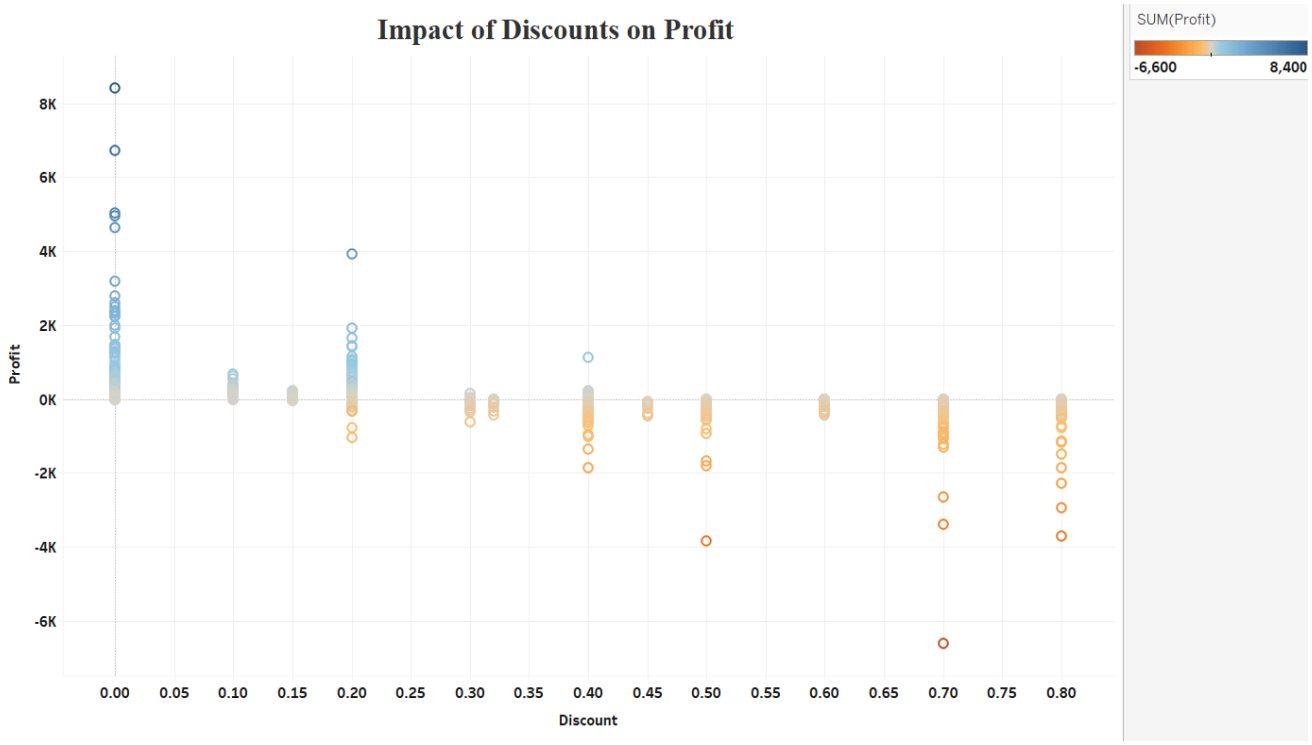
Processing time was calculated by creating a calculated field using the formula as follows:

DATEDIFF(Orders[Order Date], Orders[Ship Date],DAY)

Description:

From this bar chart we found that 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.

12. How do discounts affect overall profit?



Why to choose this specific chart?

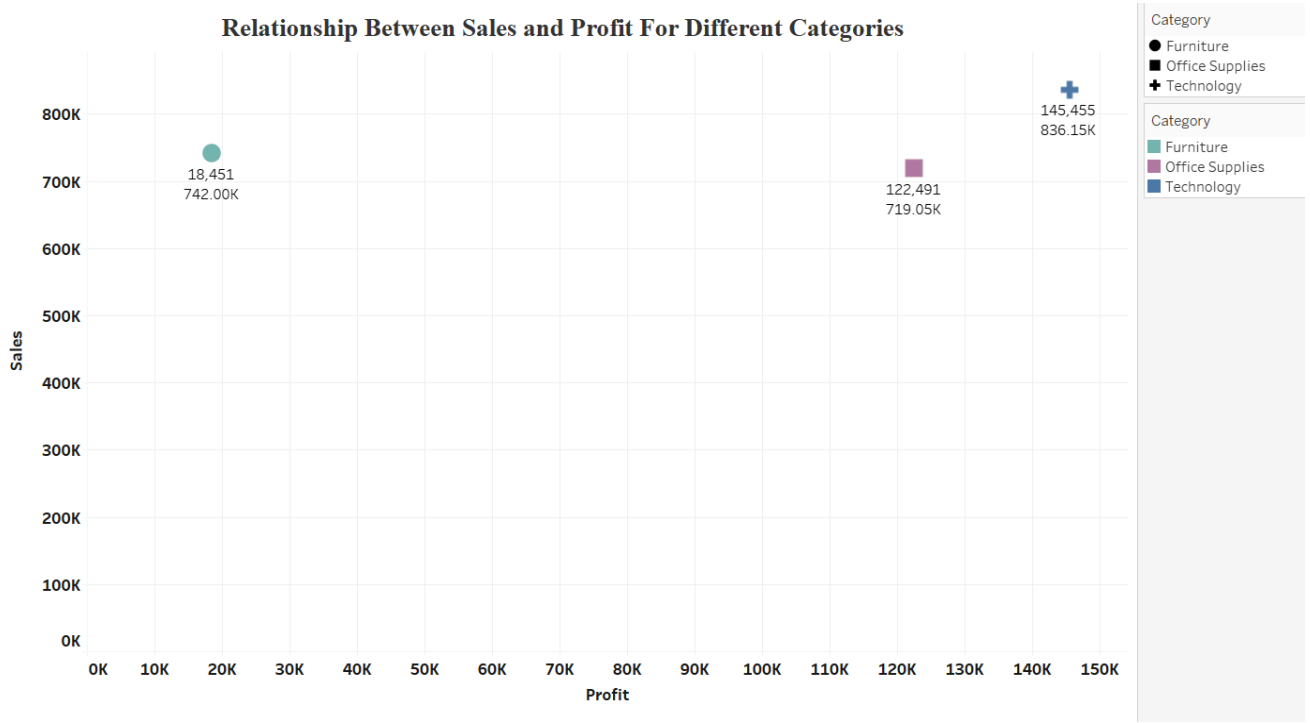
A Scatter Plot has been chosen for the visualisation to show the effect of discount on profit. The Scatter plot is the most suitable chart to show the relationship between two numerical columns. They can help to spot outliers or anomalies in a data.

Description:

Profits at every discount level can be observed with the help of scatter plot, with the help of the visualisation we can observe that with increasing discount the profit keeps on decreasing for many orders.

The highest loss is 6600 at 70% discount, while the highest profit of 8,400 has been observed at 0% discount.

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



Why to choose this specific chart?

A scatter plot was chosen for this visualisation 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.

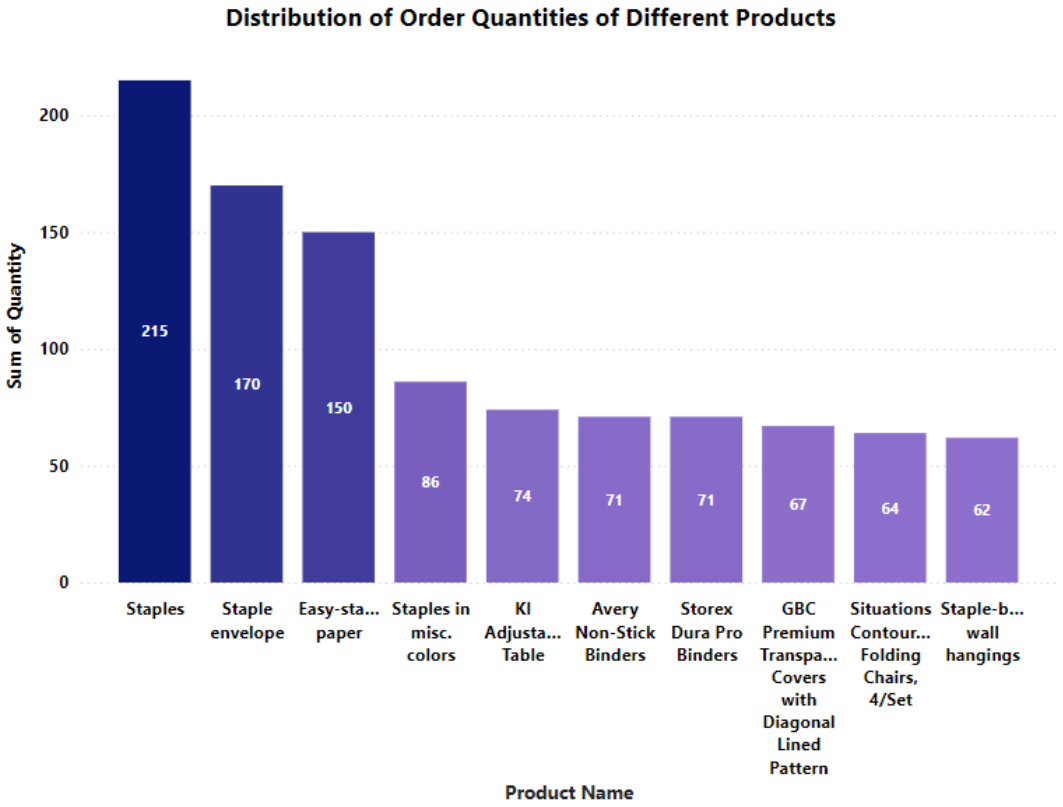
Description:

Similarly, the overall sales for 'Furniture' amounted to 742.00K, with a profit generated of 18,451. This suggests a comparatively lowest profit margin from the rest category.

The overall sales for 'Office Supplies' amounted to 719.05k, with a profit generated of 122,491. This indicates a healthy profit margin relative to the sales volume.

Lastly, the overall sales for 'Technology' amounted to 836.15k, yielding a profit of 145,455. This highlights a robust profit margin relative to the sales volume in the 'Technology' category, indicating strong profitability within the customer segment.

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



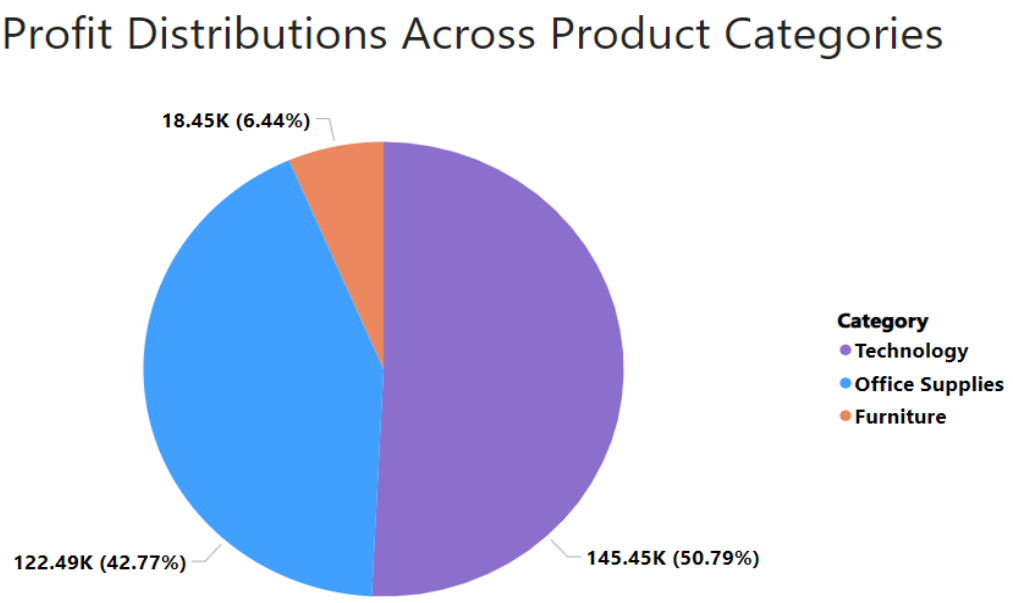
Why to choose this specific chart?(needs attention)

The bar chart was chosen for this visualisation because it allows a clear visual representation of the distribution of order quantities, enabling easy comparison of frequency across different quantity ranges.

We have taken top 10 products by order quantity , because we have large number of products and we can't clearly visualise all the products in one chart.

The chart shows that there is a product named 'Staples' which shows highest quantity (215), followed by 'Staple Envelop' (170) and 'Easy Staple Paper' which shows quantity (150).

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



Why to choose this 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. A Pie Chart provides a quick and intuitive overview of how profits are distributed among different product categories. Each slice represents a product category's share of the total profits.

Description:

Product category 'Technology' has the highest share of profit which is 50.79% of total, showcasing the robust profitability and strong performance of this category in generating revenue and contributing to overall profit margins.

Category 'Furniture' has the lowest share of profit which is 6.44% of total, indicating that this category may face challenges or lower demand compared to other categories.

Category 'Office Supplies' has a share of 42.77% of total, 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?

Why to choose this 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.

Description:

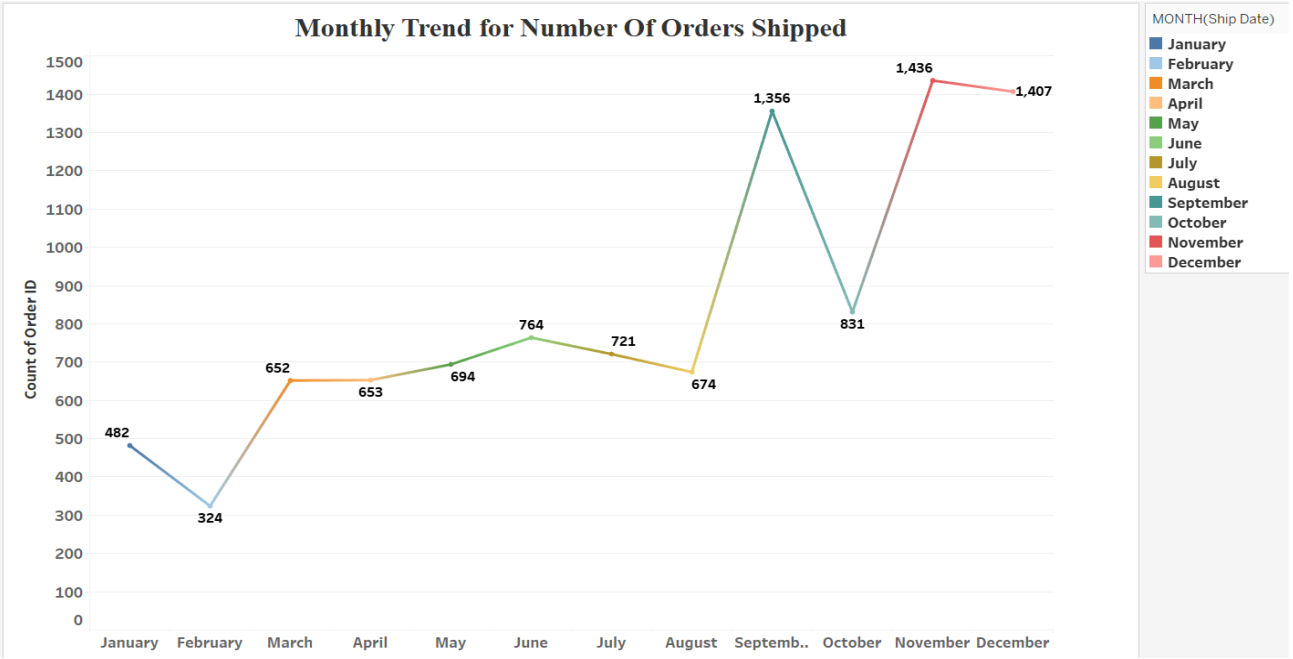
Same Day ship mode is the fastest shipping mode among all to meet the expectations of customers who prioritise speed and reliability in their shipping preferences.

First class ship mode takes an average of 2 days to ship the order to customers, indicating efficient and fast delivery service

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.

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 prioritise cost savings over expedited shipping.

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



Why to choose this specific chart?

A line chart was chosen for this visualization because it effectively shows the monthly trend for the number of orders shipped. A line chart is effective for showcasing trends over time, providing a clear visualisation of how the volume of orders changes on a monthly basis.

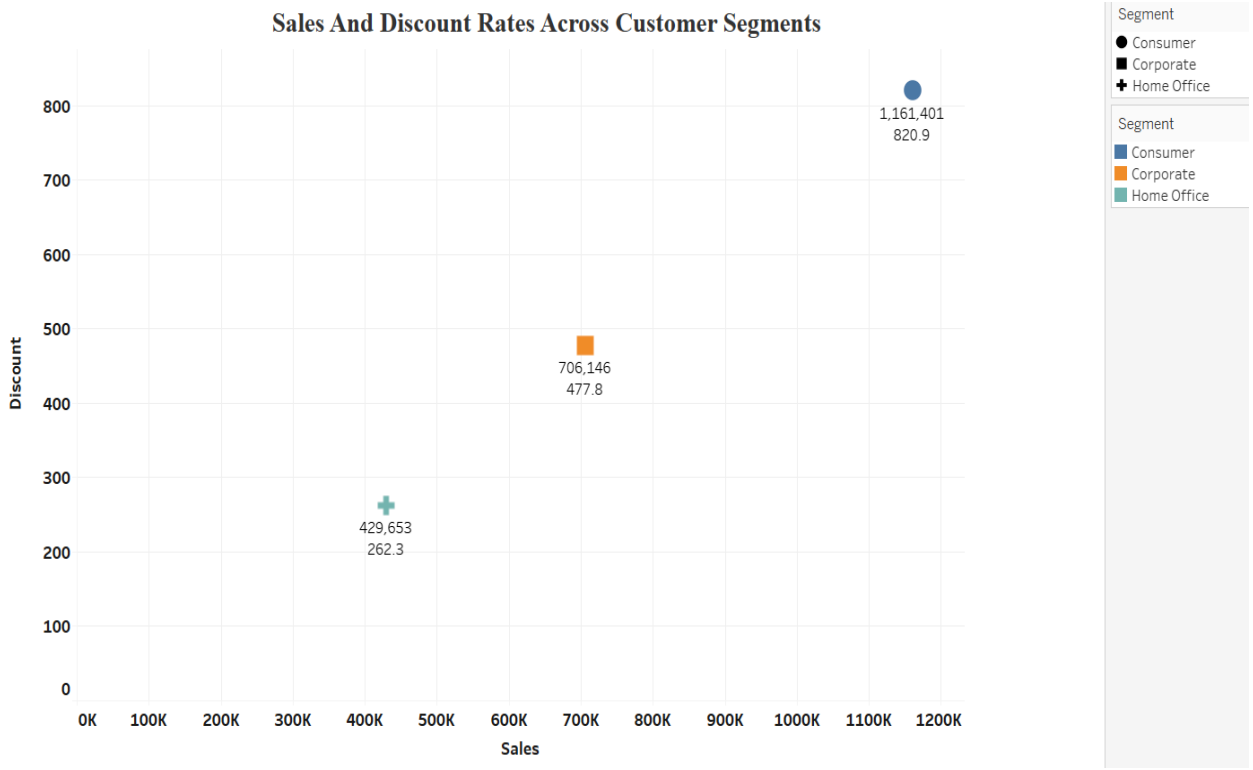
Description:

From the chart we found that there are Lot of fluctuations in the number of orders shipped from January to December.But there is a positive trend over the year, as the number of orders shipped, goes on increasing from January to December.

The spikes in order numbers during September, as well as between November and December are high. It could be influenced by festive seasons, end year or new year sales and product launches during these periods.

Conversely, the lower demand for products in January,February and April may be due to post holiday slowdown, seasonal trends or reduced consumer activity, as individuals may be recovering from holiday spendings.

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



Why to choose this specific chart?

A scatter plot was chosen for this visualisation 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.

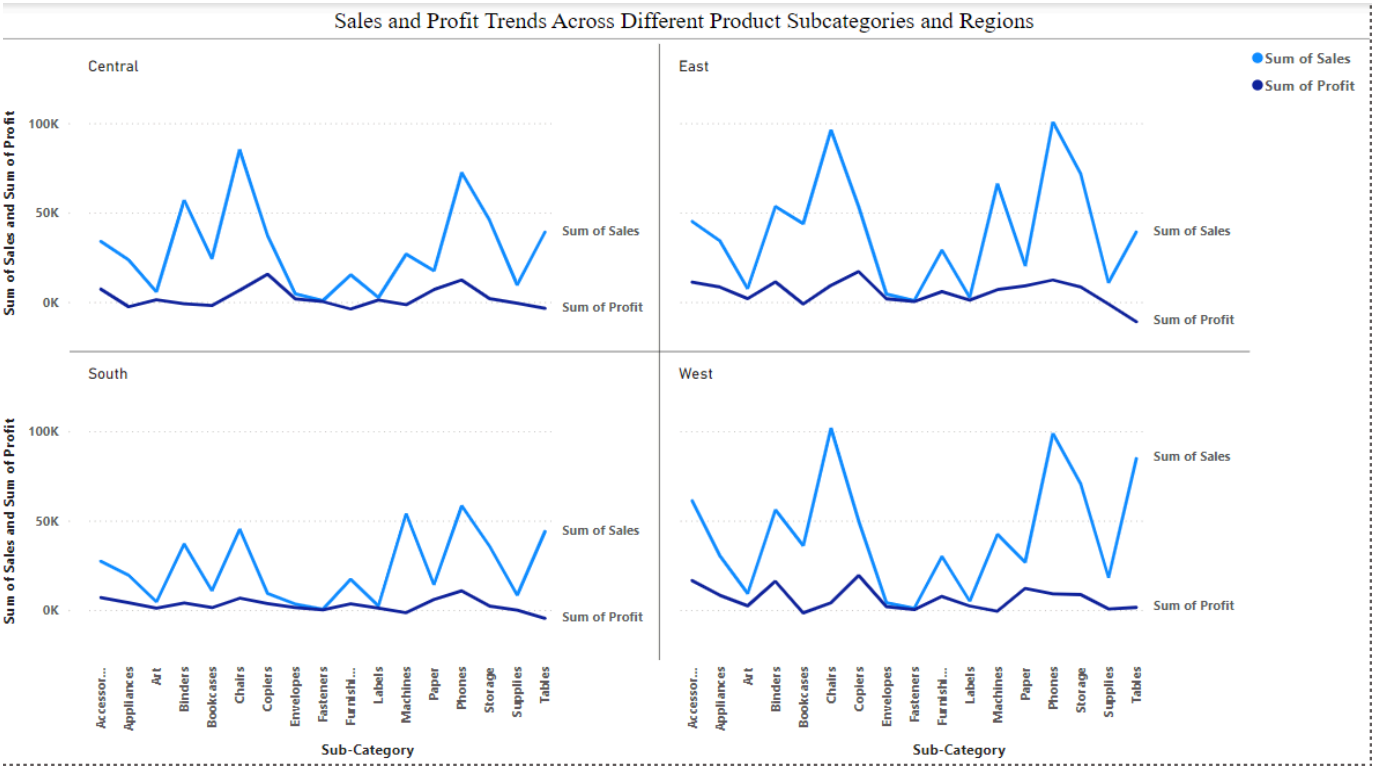
Description:

The Consumer segment has a sale of 1161401 and a discount of 821, maintaining highest sales and discount from other segments, indicative of strong customer engagement and purchasing behaviour.

The Corporate segment has a sale of 706,146 and a discount of 478 indicating low sales and discounts from customer segments.

The Home Office segment has a sale of 429,653 and a discount of 262, indicating low sales and discounts from all other customer segments.

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



Why to choose this specific chart?

A Dual Line Chart would be the most suitable chart to properly visualise the trends of sales and profit over the period of time for different product sub categories and regions. In this case, we can see how sales and profit varies across different product sub-categories for each region over the time.

Description:

Products subcategories like Accessories, Blinders, Chairs, Machines, Phones and Tables show high sales in all regions, contributing positively to overall revenue. Conversely products subcategories like Art, Bookcases, Papers, Fasteners and Labels show low sales in all regions. Some other subcategories like Accessories and Tables have an average amount of sales.

Products subcategories like Chairs, Copiers, Furnishing and Phones show high profit in all regions, contributing positively to overall revenue. Conversely products subcategories like Art, Bookcases, Fasteners and Suppliers show low profit in all regions.

Subcategory Tables, Furnished, Machines and supplies shows negative profit in some regions, contributing negatively to overall revenue.

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

Why to choose this specific chart?

A Highlighted Table was chosen for this visualization because it clearly and effectively shows average shipping time across shipping modes, with each ship mode and average shipping time represented by a different colour.

Description:

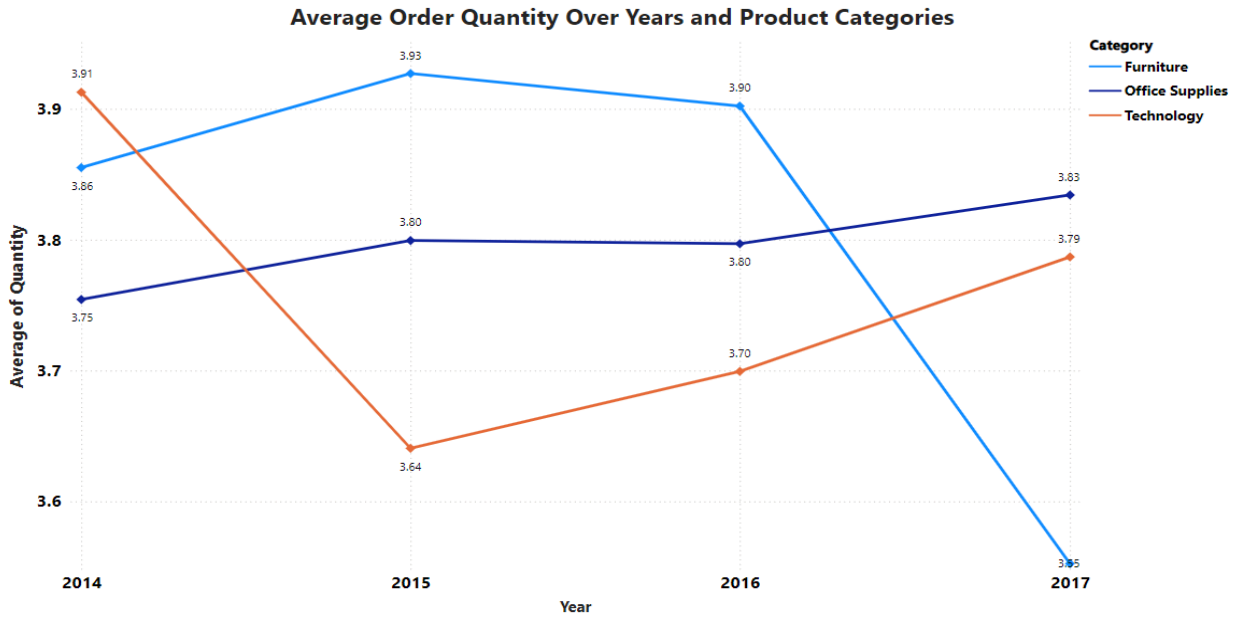
Same day ship mode takes an average ‘0’ (zero) day to ship the orders to customers across all regions, indicating the fastest and most efficient delivery service.

First class ship mode takes an average of ‘2’ days to ship the order to customers across all regions, indicating efficient and fast delivery service tailored to meet the expectations of customers.

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.

Standard class ship mode takes an average of '5' days to ship the order to customers across all regions. While it takes slightly longer than other shipping modes, Yet it offers a standard delivery service.

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



Why to choose this specific chart?

A line graph was chosen for this visualisation because it effectively shows the change in average order quantity over the time for various product categories. A line chart effectively communicates trends over time, providing a clear visualisation of how average order quantities change in all the years.

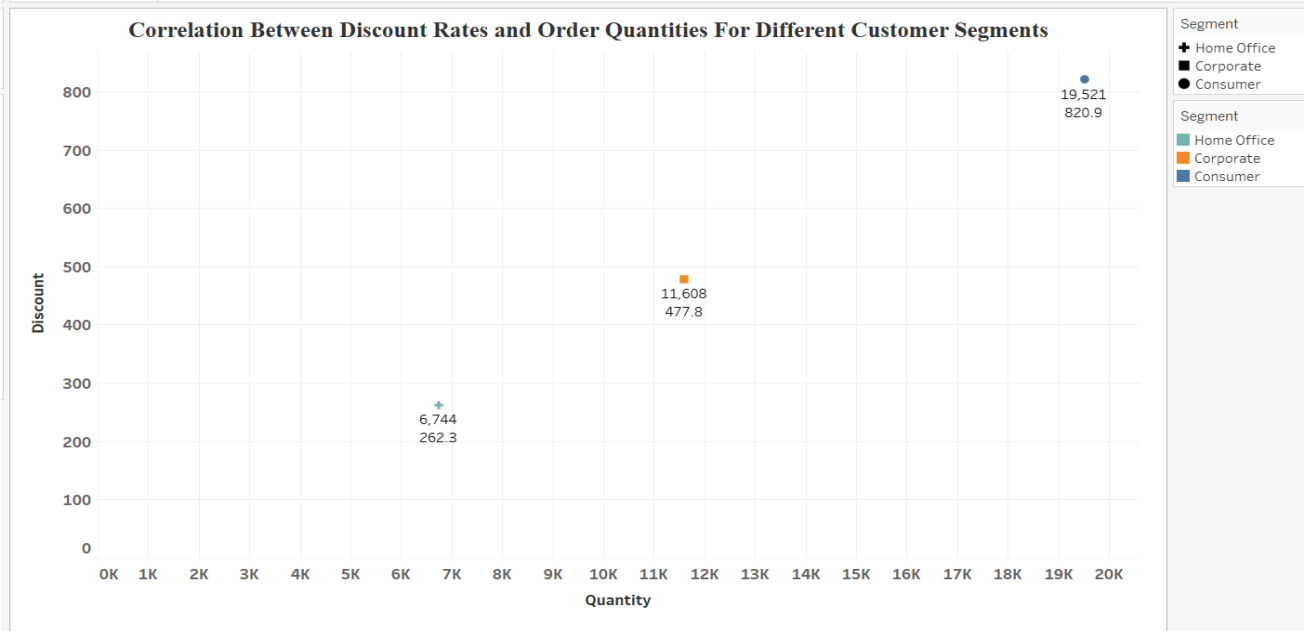
Description:

The Average order quantity for product category 'Furniture' first increases from 2014 to 2015, showing the highest average order quantities in 2015. After 2015 it goes on decreasing till 2017, showing lowest order quantity in 2017.

The average order quantity of product category ‘Office Supplies’ shows positive trend in all the years as it goes on increasing from 2014 to 2017.

The average order quantity of product category ‘Technology’ first decreases from 2014 to 2015 then goes on increasing after 2015 and maintaining the positive trend till 2017.

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



Why to choose this specific chart?

A scatter plot was chosen for this visualisation 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.

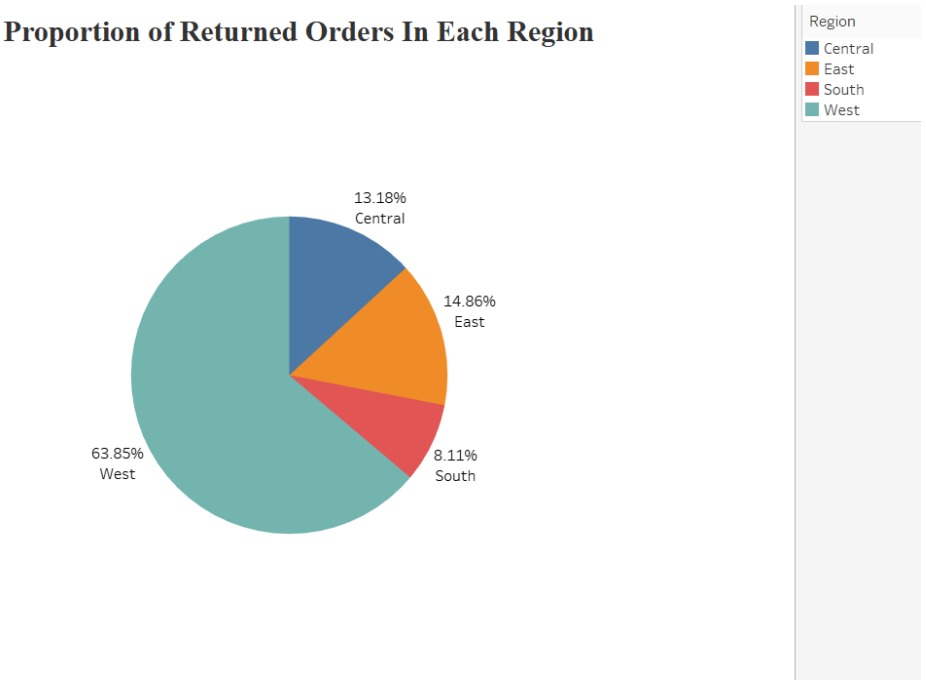
Description:

A discount of 262 was given for the 6,744 order quantities in the ‘Home Office’ customer segment, which indicates a relatively low discount rate and total number of orders from other customer 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.

A discount of 821 was given for the 19,521 order quantities in the consumer segment, showcasing a high discount rate applied to a large volume of orders within this segment.

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



Why to choose this specific chart?

A pie chart was chosen for this visualization because it effectively compares the proportions of different regions, with each region represented by a different colour and the sizes of each slice corresponding to the percentage of returned orders.

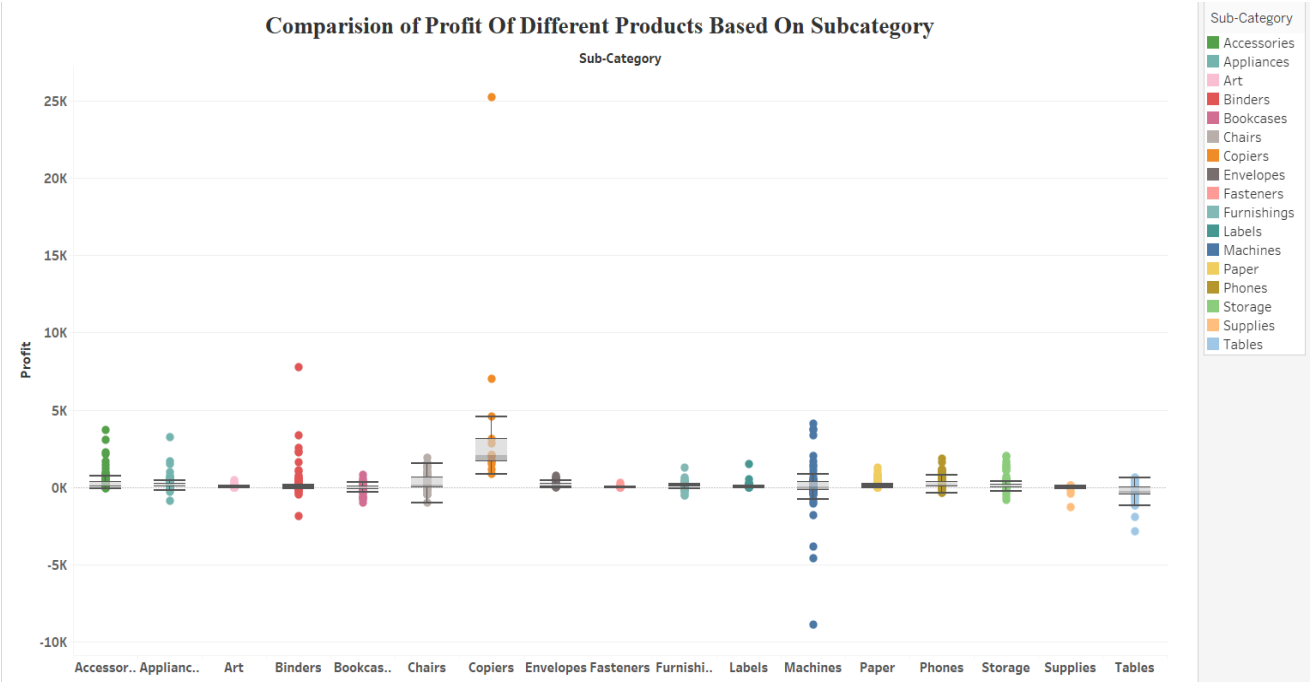
Description:

The Central region contributes to 13.18% of the total orders returned. This indicates a moderate level of returns in this region.
The East region contributes to 14.86% of the total orders returned. This indicates a similar level of returns as observed in the Central region.

The South region contributes to 8.11% of the total orders returned. This shows Lowest proportion of orders being returned as compared to other regions.

The West region contributes to 63.85% of the total orders returned. This shows Highest proportion of orders being returned as compared to other regions.

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



Why to choose this specific chart?

A box plot was chosen for this visualization because the box plot is effective for comparison of the profit distributions for different products based on sub-category, by providing a clear summary of minimum, maximum and outliers. Box plots readily provide a lot of information in a compact form.

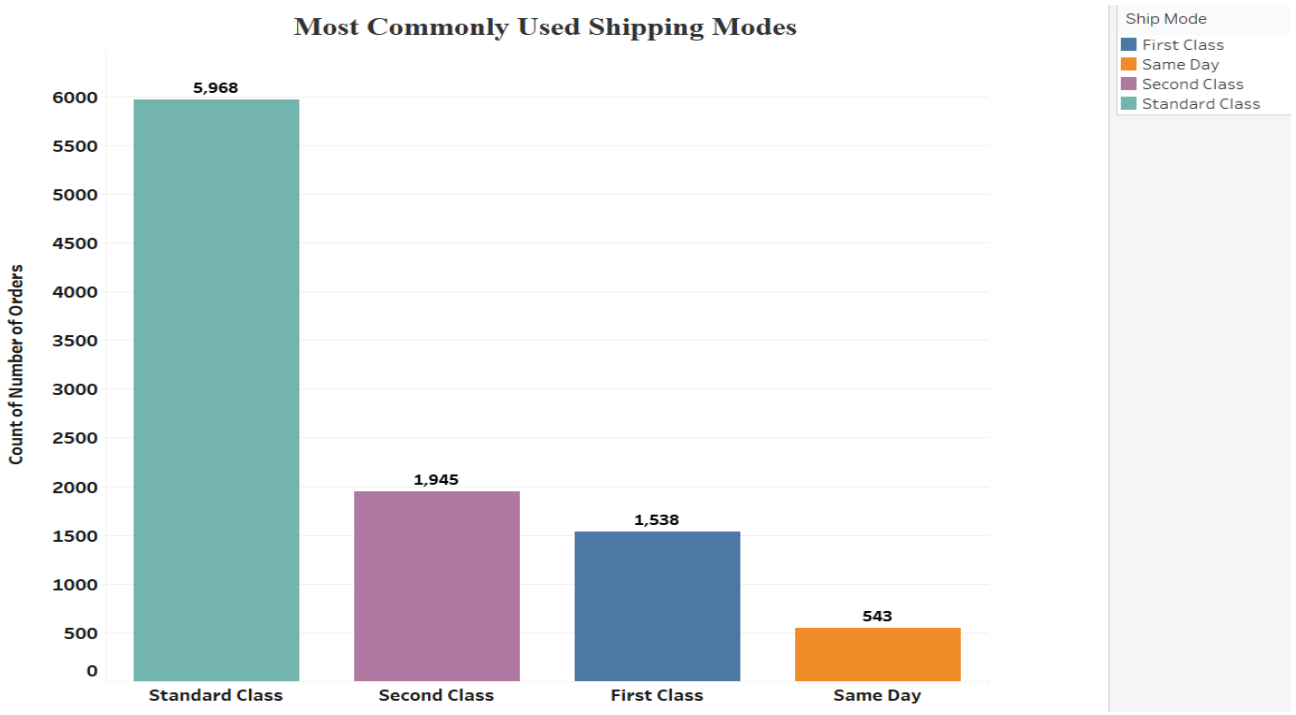
Description:

From the box plot we can clearly see that there are many products under different subcategories which shows positive profit as well as negative profit (called loss).

There is a product under Copiers subcategory showing the highest profit and a product under Machines subcategory showing the highest negative profit or loss.

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

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



Why to choose this specific chart?

A bar chart was chosen for this visualization because the bar chart allows for a clear 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 better comparison of the distributions of shipping mode.

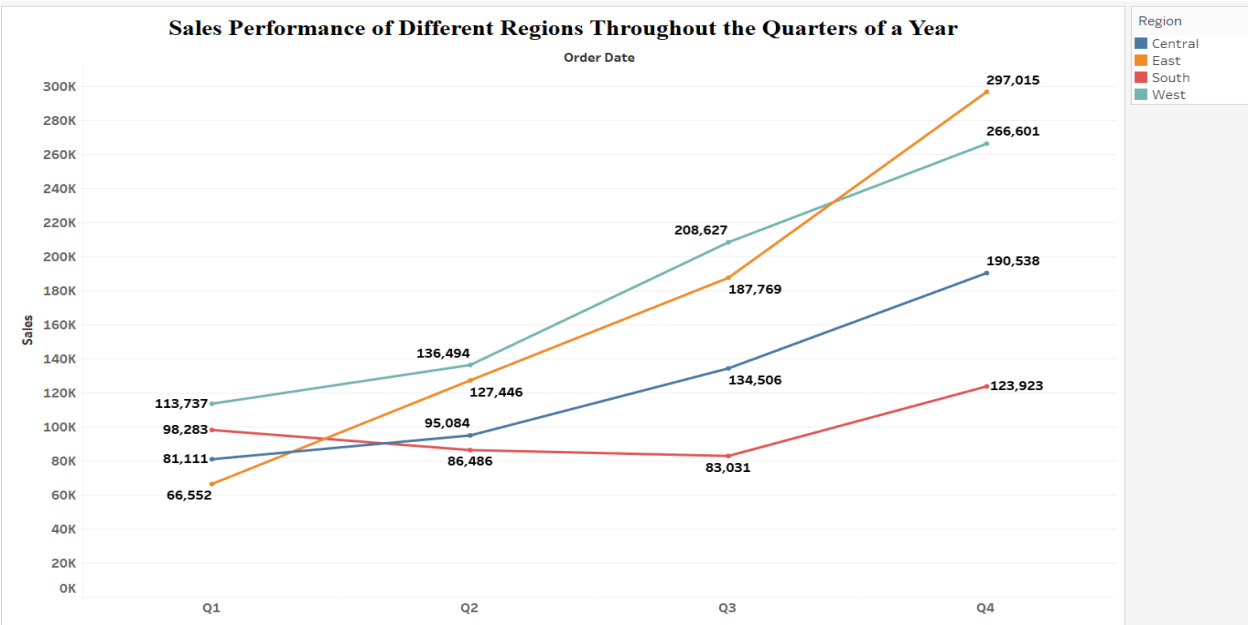
Description:

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.

First class shows 1528 and Second class shows 1945 shipping orders, indicating less chosen shipping mode than standard class.

In contrast, the Same day shipping mode records the lowest number of orders, indicating that less number of orders are shipped in this mode.

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



Why to choose this specific chart?

A line graph was chosen for this visualisation because it effectively shows the sales performance of different regions throughout the quarters of a year. A line chart effectively communicates trends over time, providing a clear visualisation of how sales evolve across different regions each quarter.

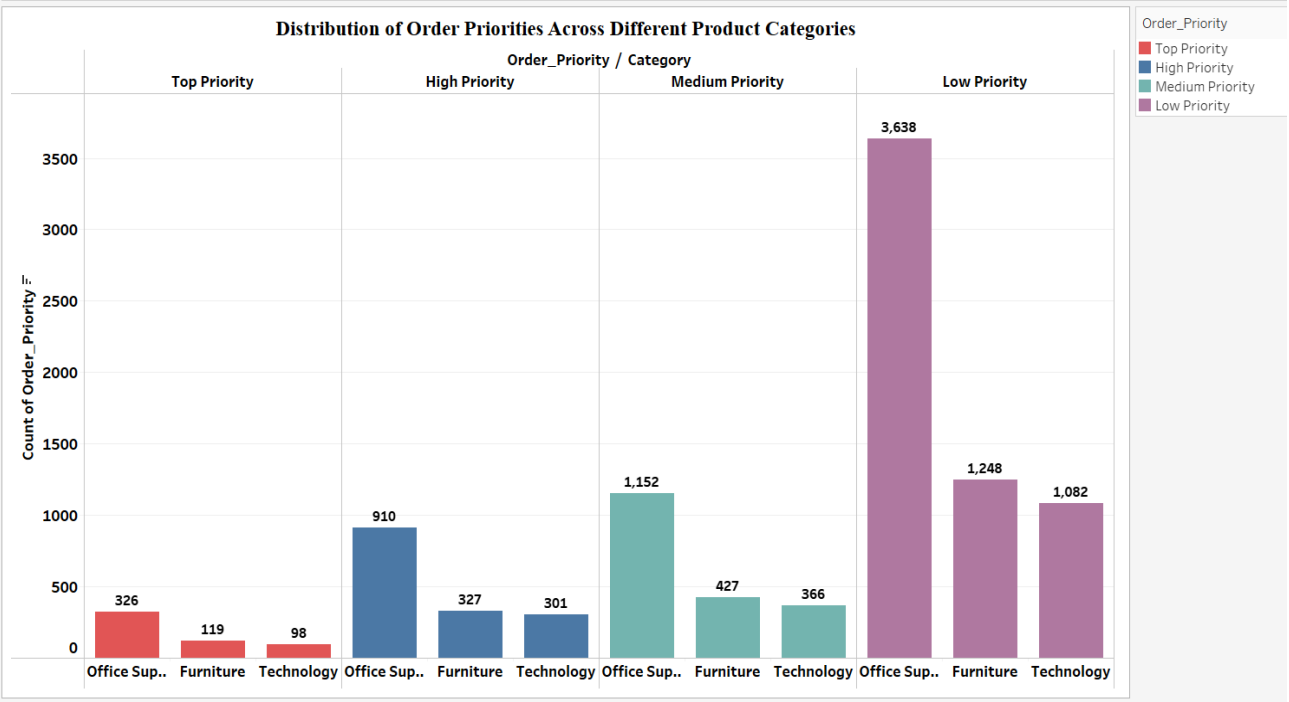
Description:

The chart shows that sales of all the regions shows positive trend going from lowest to highest from quarter 1 to quarter 4 except South region. The sales of South region goes on decreasing from quarter 1 to quarter 3 and then increasing till quarter 4.

The lowest sales of 66,552 occur in quarter 1 in East region.

The highest sales of 2,97,015 occur in quarter 4 in East region.

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



Why to choose this 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.

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

```
IF [Ship Mode] = 'Same day' THEN 'Top Priority'
ELSEIF [Ship Mode] = 'First class' THEN ‘High Priority’
ELSEIF [Ship Mode] = 'Second class' THEN 'Medium Priority'
ELSE 'Low Priority'
END
```

Description:

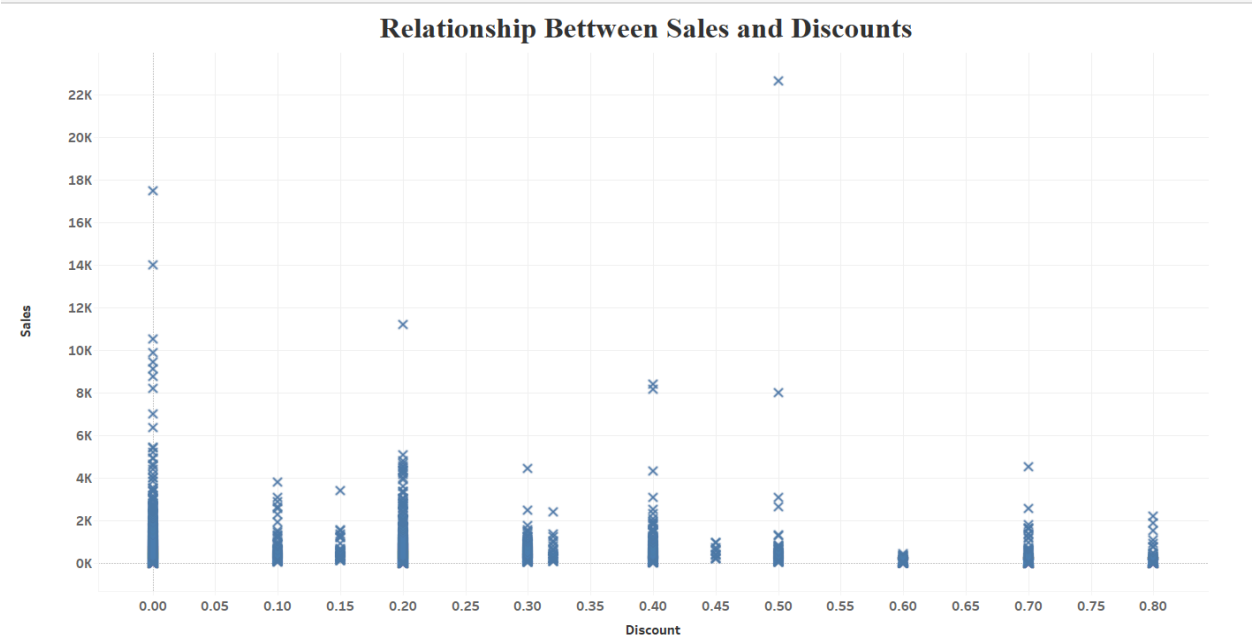
In Top order priorities, the product category ‘Furniture’ recorded 119, ‘Office Supplies’ had 326, and ‘Technology’ showed 98 order priorities.

In High order priorities, ‘Furniture’ accounted for 327, ‘Office Supplies’ had 910, and ‘Technology’ showed 301 order priorities.

In Medium order priorities, ‘Furniture’ recorded 427, ‘office supplies’ has 1152 and ‘Technology’ has 366 orders.

In Low order priorities, ‘Furniture’ recorded 1248, ‘office Supplies’ has 3638 and ‘Technology’ have 1082 orders.

28. What is the relationship between discounts and sales?



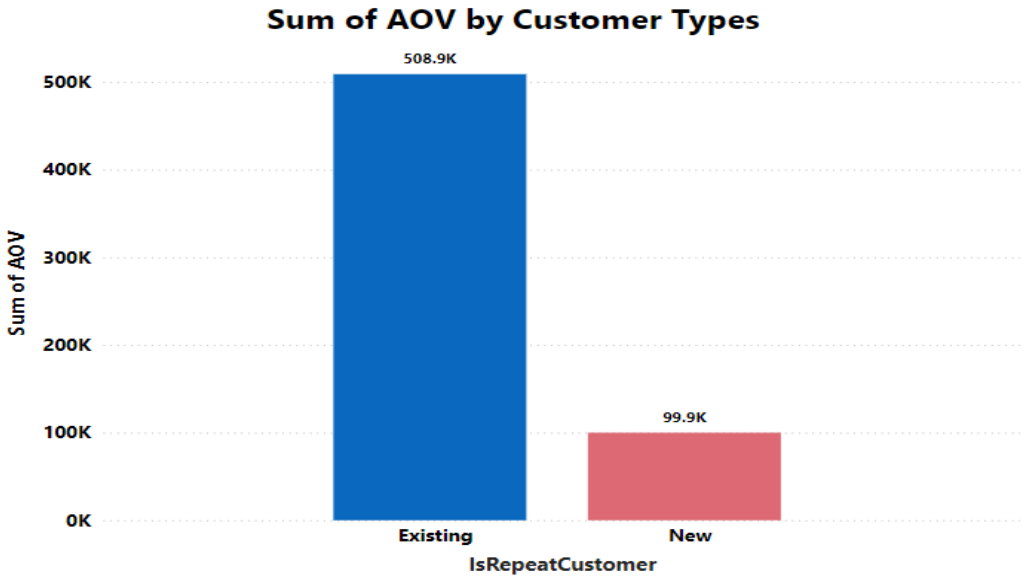
Why to choose this specific chart?

A scatter plot was chosen for this visualisation because it effectively shows the relationship between discount and sales, showcasing whether there is a positive, negative, or no correlation between discounts and sales. Each data point shows us how discount rates correlated to sales.

Description:

The chart shows that higher sales are occurring at lower discount rates. As the discount rates go high the sale amount decreases. Most sales occurred at 0 % and less sales occurred at 60 % .
There is an outlier showing a total sale of more than 22 k at 50 % discount, indicating an exceptional high-value transaction or an anomaly in the dataset that needs further investigation.

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



Why to choose this specific chart?

A bar chart was chosen for this visualisation because it effectively displays the average order value of different customer type. Bar chart allows easy visual interpretation and comparison of the data.

To visualise Average Order Value for Customer Type we create some measures such as :

FirstPurchaseDate :

```
CALCULATE(MIN('Orders'[Order Date]),
ALLEXCEPT('Orders', 'Orders'[Customer ID]))
```

Is Repeat Customer :

```
IF('Orders'[Order Date] > 'Orders'[FirstPurchaseDate], "Existing",
"New")
```

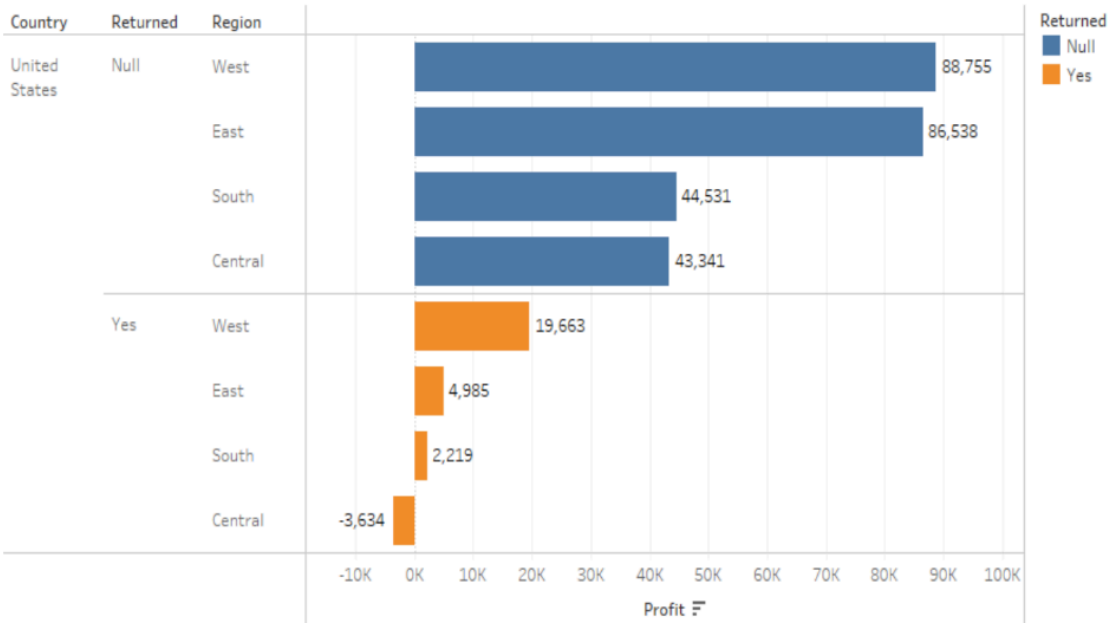
Average order Value :

```
Orders[Sales]/Orders[Quantity]
```

Description:

Total Average Order value for Existing customers are 508.9k , indicating that there are large number of customers who placed order many times.
On other side, Total Average Order Value for New Customers are 99.9k, indicating a decent number of customers.

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



Why to choose this specific chart?

A bar chart is chosen to visualise the distribution of returns and its impact on profit. Bar charts provide a clear visual representation of data across different areas, making it easier to understand profit and returns.

Description:

The chart shows that products with return generally result in lower profitability compared to those that are not returned across all regions.

West region shows 88.7 k profit when return not occured and 19.6 k when return occur, showcasing the highest profitability region among all regions.

On the other side, Central region shows 43.3 k profit when return not occured and -3.6 k when return occured, showcasing the lowest profitability region among all regions.