



DATA VISUALIZATION AND ANALYSIS PROJECT WITH TABLEAU



About the Project:

This is a Data Visualization and Analysis Project using Tableau and the Superstore Sales Dataset. This project contains 25 scenario-based questions which are solved using data visualization features of Tableau.

Objective:

The objective of this project is to select the best chart for each scenario, explain the reason for the choice, build the chosen chart using Tableau and detail the insights drawn from the visualization.

This project showcases the 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.

Dataset Link:

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

Questions

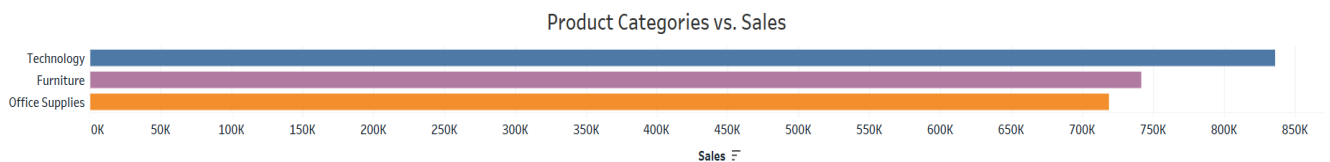
1. Which product categories have the highest total sales in the "Superstore" dataset?

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing categorical data. Bar chart is therefore used here to compare the total sales of different product categories. One can easily identify which product category has contributed to the highest total sales.

Insight: It can be observed that product category 'Technology' has had the highest total sales among the three distinct product categories.

1. Product Categories vs. Sales



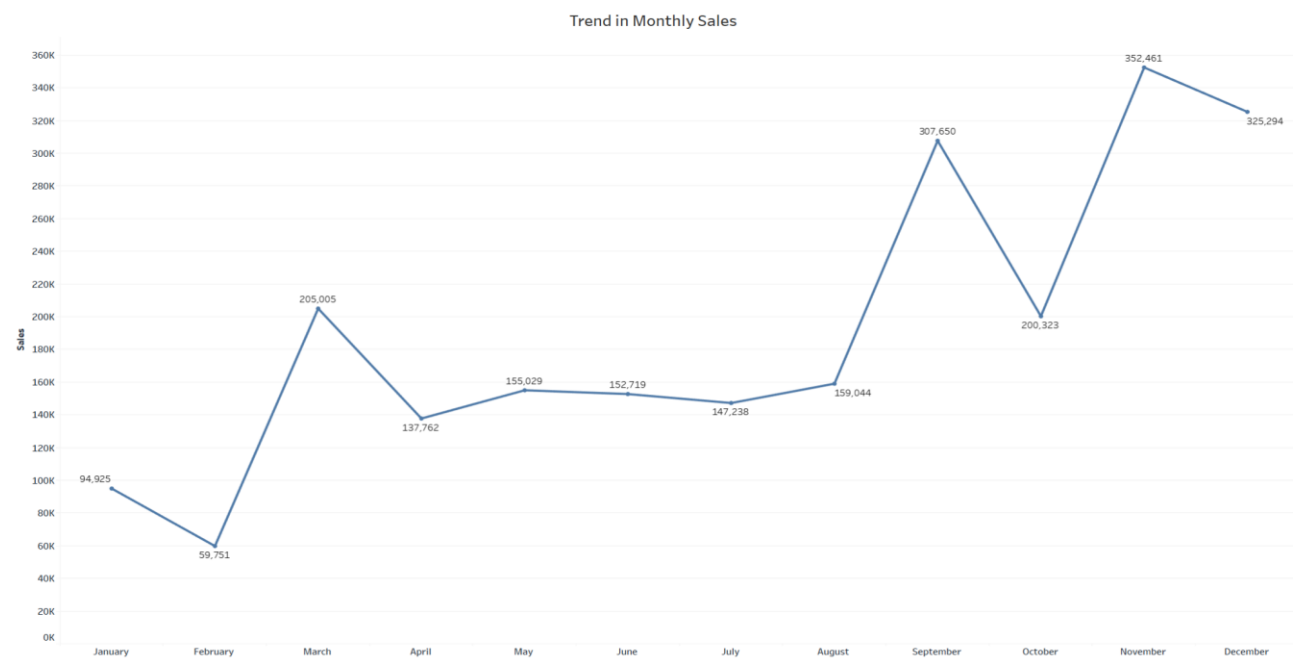
2. How do the monthly sales amounts change over the course of a year?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the trend in the monthly sales over all the years collectively.

Insight: It can be observed that there is generally an upward trend in the monthly sales over the year with total sales peaking in November. After the peak in September, there was a sharp decline in the sales in October, however, it still remained higher than its previous low.

2. Monthly Sales Trend



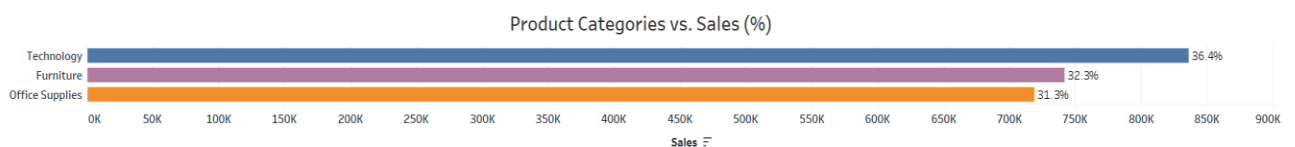
3. How is the total sales amount distributed among different product categories?

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing categorical data. Bar chart is therefore used here to visualise the distribution of the total sales of different product categories. The percentages displayed next to each bar helps to easily visualise the contribution of each product category to total sales.

Insight: It can be observed that product category 'Technology' has the highest total sales among the three distinct product categories.

3. Product Categories vs. Sales (%)



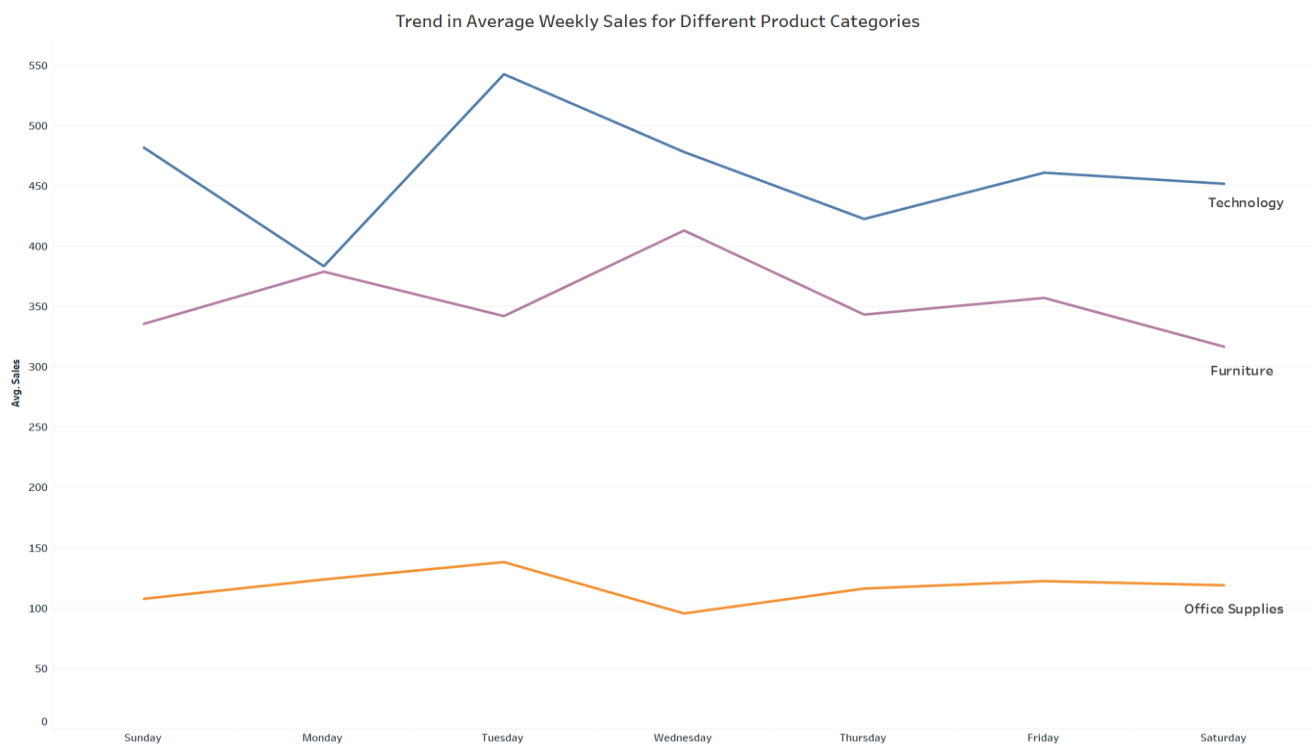
4. How do sales vary based on different days of the week and product categories?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the trend in the average daily sales (cumulatively) across all the years. The different lines on the chart represent the average daily sales trend of different categories.

Insight: It can be observed that the average sales are generally the highest on Tuesday for Office Supplies and Technology and on Wednesday for Furniture. Additionally, it can be observed that the average sales for Office Supplies does not vary much over the week except a slight variation in the sales over Tuesday and Wednesday.

4. Average Weekly Sales Trend across Different Product Categories



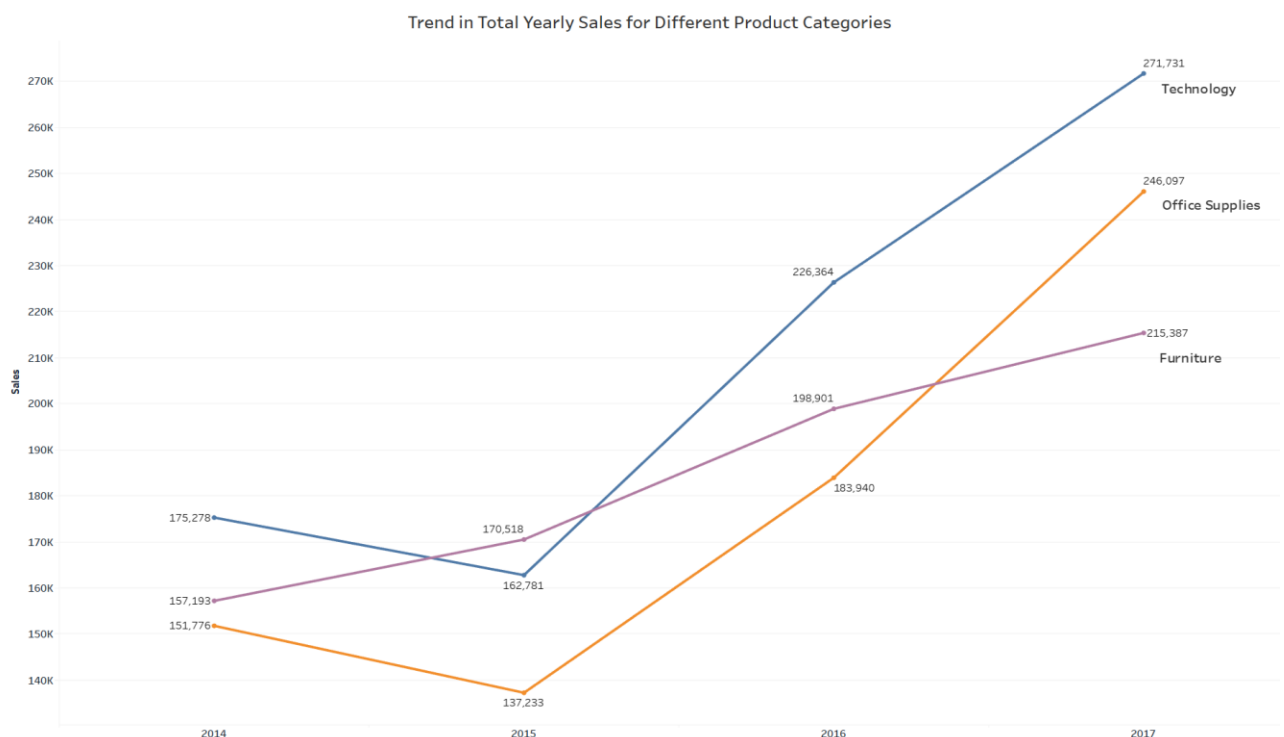
5. Can we visualise the sales growth of different product categories over time?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the trend in the total sales over the years. The different lines on the chart represent the total yearly sales trend of different categories.

Insight: It can be observed that the total sales have generally been in an upward trend over the years. The sales growth of the technology category has been the highest followed by office supplies and then furniture. Year 2015 witnessed the lowest amount of sales in technology and office supplies while furniture had the lowest sales in 2014. The highest amount of total sales for all product categories was in 2017.

5. Yearly Sales Trend across Different Product Categories



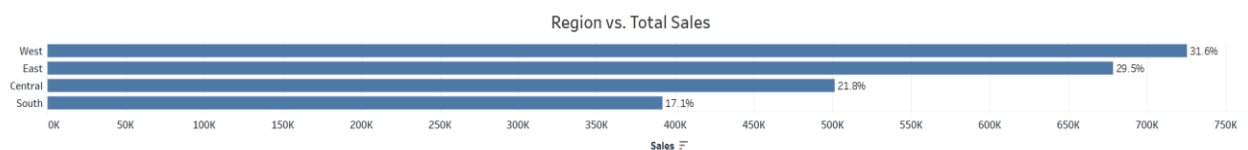
6. How does the sales distribution vary across different regions in the "Superstore" dataset?

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Bar chart is therefore used here to visualise the distribution of total sales over all the years collectively across different regions of the United States of America.

Insight: It can be observed that the highest sales have been in the west region (31.6%), followed by east with not much variation in the total sales contribution (29.5%), then central (21.8%) and the least total sales in the south region (17.1%).

6. Region vs. Total Sales



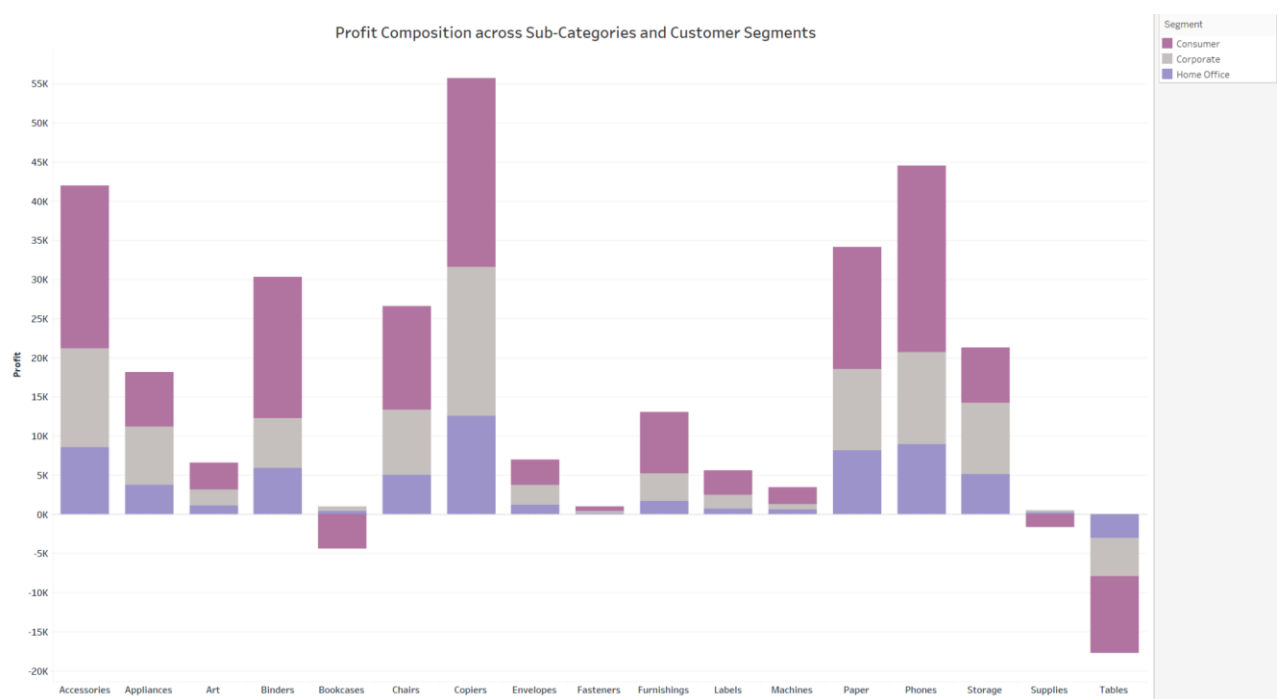
7. Can we visualise the composition of profits across various subcategories within different customer segments?

Chart type - Stacked Column Chart

Why Stacked Column chart? Stacked column charts are a versatile tool for visualising and comparing the composition, proportions, and trends of different subcategories within categories. A stacked column chart is therefore used here to visualise the composition of profits across various subcategories for different customer segments. The x-axis displays the different sub-categories and each bar represents the profit composition of different customer segments in each of those sub-categories.

Insight: It can be observed that copiers have been the most profitable sub-category while tables, supplies and bookcases have incurred losses with tables being the highest loss making product. In each of the sub-categories, consumers is the most profitable customer segment followed by corporate and then home office.

7. Profit Composition across Sub-Categories and Customer Segments



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

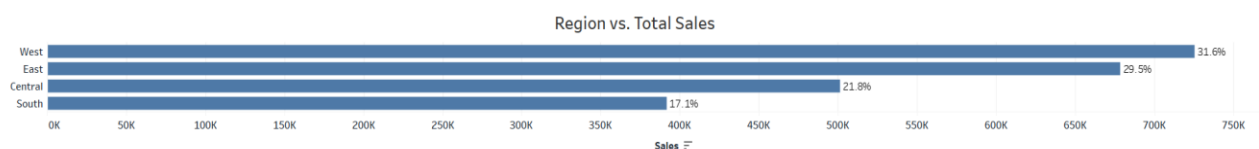
This question is similar to question 7 where a visualisation of the bar chart is displayed to demonstrate the distribution of total sales across different regions which is similar to the percentage contribution of each region to the overall sales as required by this question.

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Bar chart is therefore used here to visualise the distribution of total sales over all the years collectively across different regions of the United States of America.

Insight: It can be observed that the highest sales have been in the west region (31.5%), followed by east with not much variation in the total sales contribution (29.5%), then central (21.8%) and the least total sales in the south region (17.1%).

8. Percentage Contribution of each Region to Total Sales



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

Chart type - Stacked Column Chart

Why Stacked Column chart? Stacked column charts are a versatile tool for visualising and comparing the composition, proportions, and trends of different subcategories within categories. A stacked column chart is therefore used here to visualise the profit margins associated with different shipping modes and customer segments.

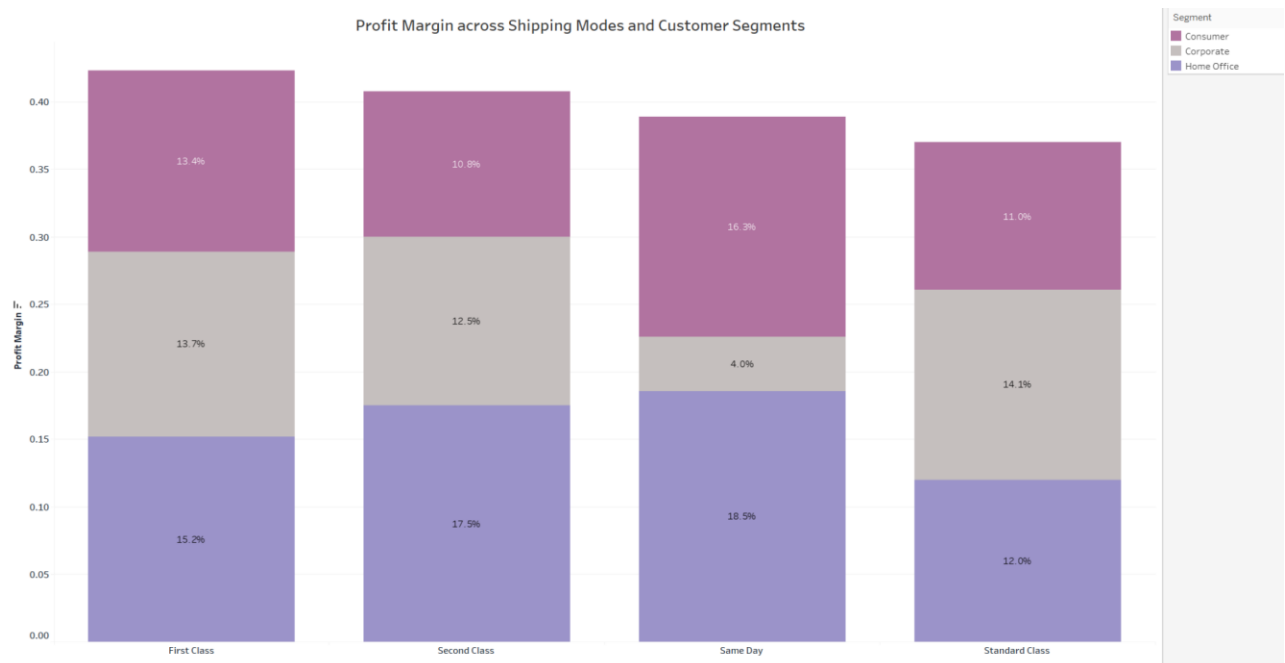
The x-axis displays the different shipping modes and each bar represents the profit margin of different customer segments in each of those shipping modes.

Calculations added:

- Overall Profit Margin: To calculate the profit margin across different segments $\text{SUM}(\text{Profit}) / \text{SUM}(\text{Sales})$

Insight: It can be observed that the profit margins are the highest for the first class shipping mode followed by second class, same day and then standard class. The home office segment is the most profitable in all shipping modes except standard class. Corporate is the most profitable segment under standard class and least profitable in the same day mode.

9. Profit Margins across Shipping Modes and Customer Segments



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

Chart type - Bar Chart

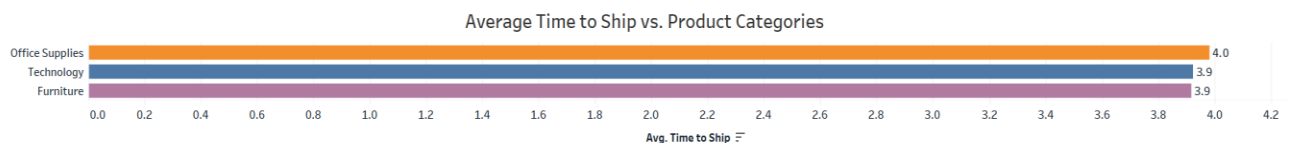
Why Bar chart? A bar chart is a useful visualisation tool for representing categorical data. Bar chart is therefore used here to visualise the average time (in days) it takes to process orders across different product categories.

Calculations added:

- **Time to Ship:** The average time (in days) is calculated as the difference between ship date and order date.

Insight: It can be observed that the average time it takes to process orders across different product categories is nearly the same, i.e. 4 days.

10. Average Shipping Time (days) for Product Categories



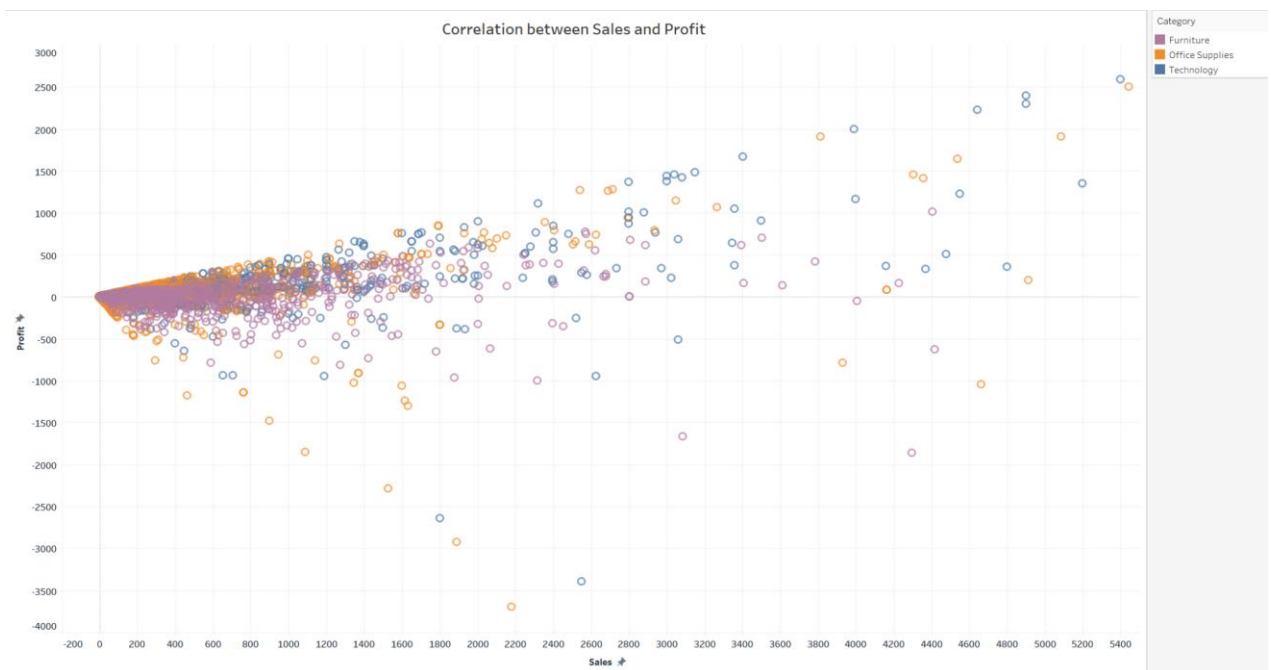
11. Can we visualise the relationship between product sales and profitability for different product categories?

Chart type - Scatter Plot

Why Scatter Plot? Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the relationship between sales and profit, we use a scatter plot. Sales is plotted on the x-axis and Profit on y-axis. Since this data has quite a few outliers, we have considered the range of -200 to 5500 for Sales and -4000 to 3000 for Profit as most of the data is concentrated in this range. This visualisation also shows this relationship across different product categories.

Insight: It can be observed that generally there is a positive correlation between sales and profit, i.e. higher sales = higher profit. However, there are some exceptions to this. For some sales of office supplies, higher sales have given negative returns.

11. Correlation between Total Sales and Total Profit for Different Product Categories



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

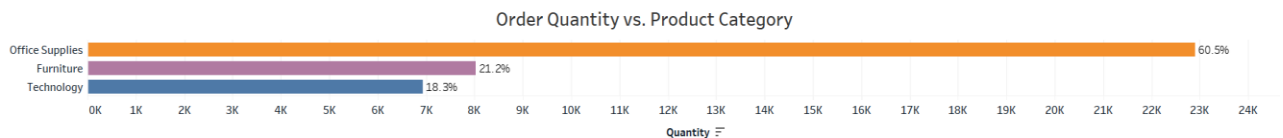
Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Bar chart is therefore used here to visualise the distribution of order quantities across different products.

Insight: It can be observed that the most ordered category is office supplies which constitutes 60.5% of total number of orders, followed by furniture (21.2%) and then technology being the least ordered category (18.3%) in terms of number of orders.

An **interesting observation** can be noticed here related to total sales and order quantity distribution across different categories. Even though office supplies have the highest order quantity (60.5%), its total sales value is the least (31.3%). On the other hand, technology category sales being the least in terms of order quantity (18.3%) has the highest sales value (36.4%). This could be because the individual price of the products in the technology category is much higher than those in the office supplies category.

12. Order Quantity vs. Product Category



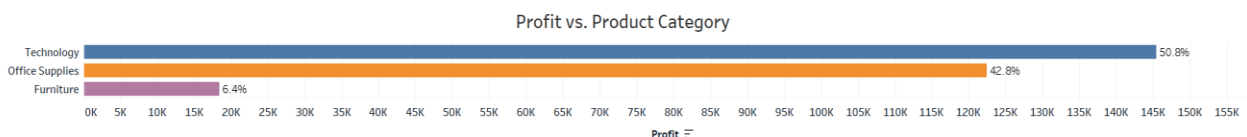
13. How do the profit distributions vary across different product categories?

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Bar chart is therefore used here to visualise the total profit distribution across different product categories.

Insight: It can be observed that the technology category is the most profitable constituting 50.8% of total profits followed by office supplies constituting 42.8% of total profits and then furniture being the least profitable constituting only 6.4% of total profits.

13. Total Profit vs. Product Category



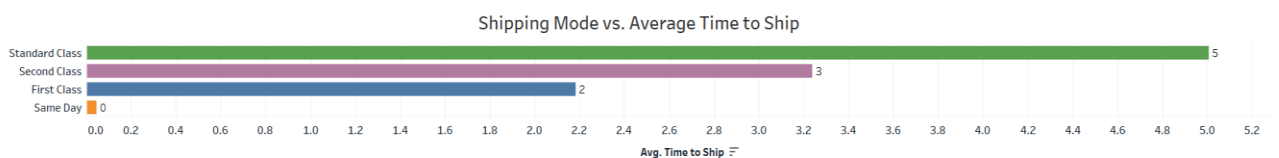
14. Can we compare the shipping time distributions for different shipping modes?

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Bar chart is therefore used here to visualise the average shipping time (in days) for different shipping modes.

Insight: It can be observed that the standard class which is the most common shipping mode (as observed in an earlier visualisation) has the longest shipping time of 5 days on an average followed by the second class shipping mode which takes about 3 days to ship and then the first class shipping mode which takes about 2 days to ship. Same day shipping mode ships the order on the same day as the order date.

14. Average Shipping Time vs. Shipping Mode



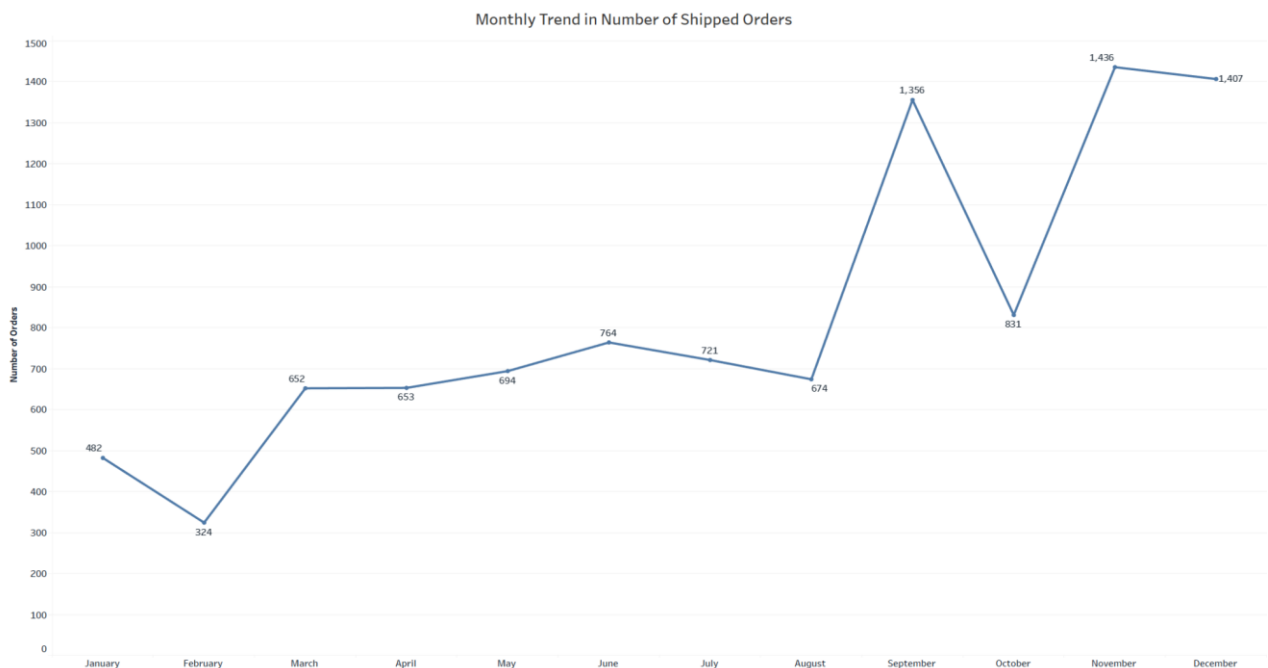
15. What is the monthly trend in the number of orders shipped?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the monthly trend in the number of orders shipped. The shipping month is displayed on the x-axis and the total number of orders on the y-axis.

Insight: It can be observed that generally there is an upward trend in the total number of shipped orders over the months. However, for the months of September, October and November, we can observe an unusual trend. After reaching a new peak in September's total shipped orders, the orders declined sharply in October only to reach a new peak in November after which it remained stable.

15. Monthly Trend in Number of Orders Shipped



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

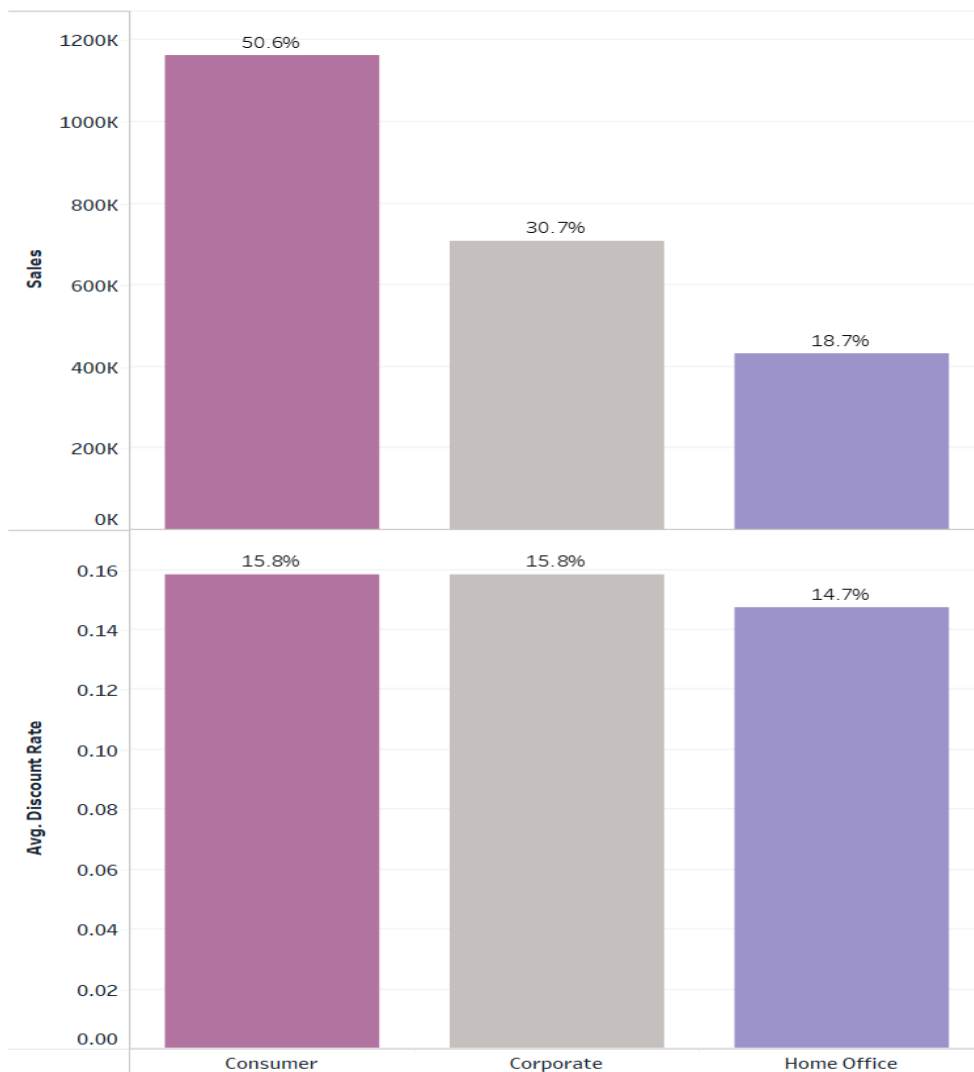
Chart type - Column Chart

Why Column chart? A column chart is a useful visualisation tool for representing the distribution of categorical data. Two column charts are used here to visualise the distribution of customer segments in terms of total sales and the average discount rate given to each customer segment.

Insight: It can be observed that the consumer customer segment contributes to the highest amount of sales (50.6%) followed by corporate (30.7%) and then home office (18.7%). However, the average discount rate given to consumers and corporates is the same (15.8%) despite the variation in their contribution to total sales. Home office customers are given a slightly lower discount of 14.7% as compared to consumers and corporates. Overall there is a wide variation in the total sales contribution of different customer segments but not much variation in their average discount rate.

16. Customer Segment vs. Total Sales and Average Discount Rate

Customer Segments vs. Total Sales and Average Discount Rate



17. How efficiently are different product subcategories being fulfilled in terms of order processing time?

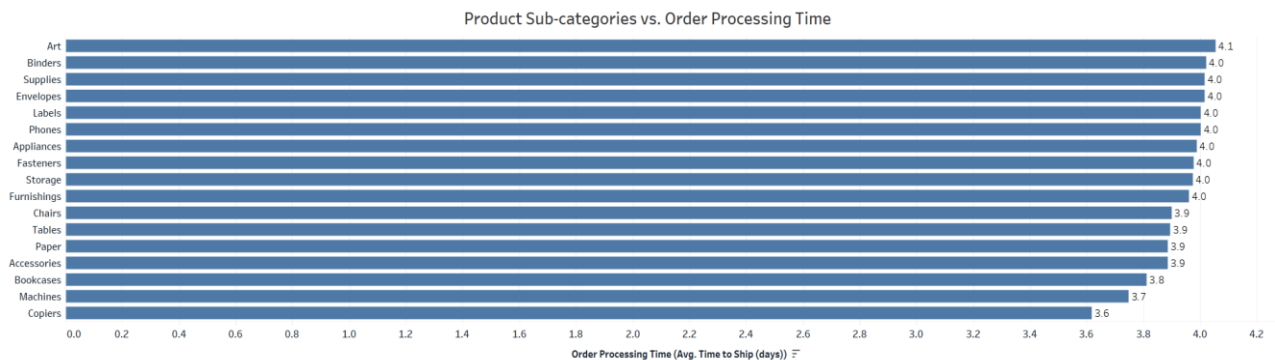
Assumption: The order processing time is assumed to be the difference between order date and ship date.

Chart type - Bar Chart

Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. A bar chart is therefore used here to compare the average order processing time (average time to ship) between different product sub-categories.

Insight: It can be observed that the average order processing time for copiers is the least (3.6 days) followed by machines (3.7 days) and bookcases (3.8 days). For accessories, paper, tables and chairs it is the same (3.9 days) while for art it is the highest (4.1 days). For all the remaining sub-categories, the average order processing time is the same (4 days).

17. Product Subcategories vs. Average Order Processing Time (days)



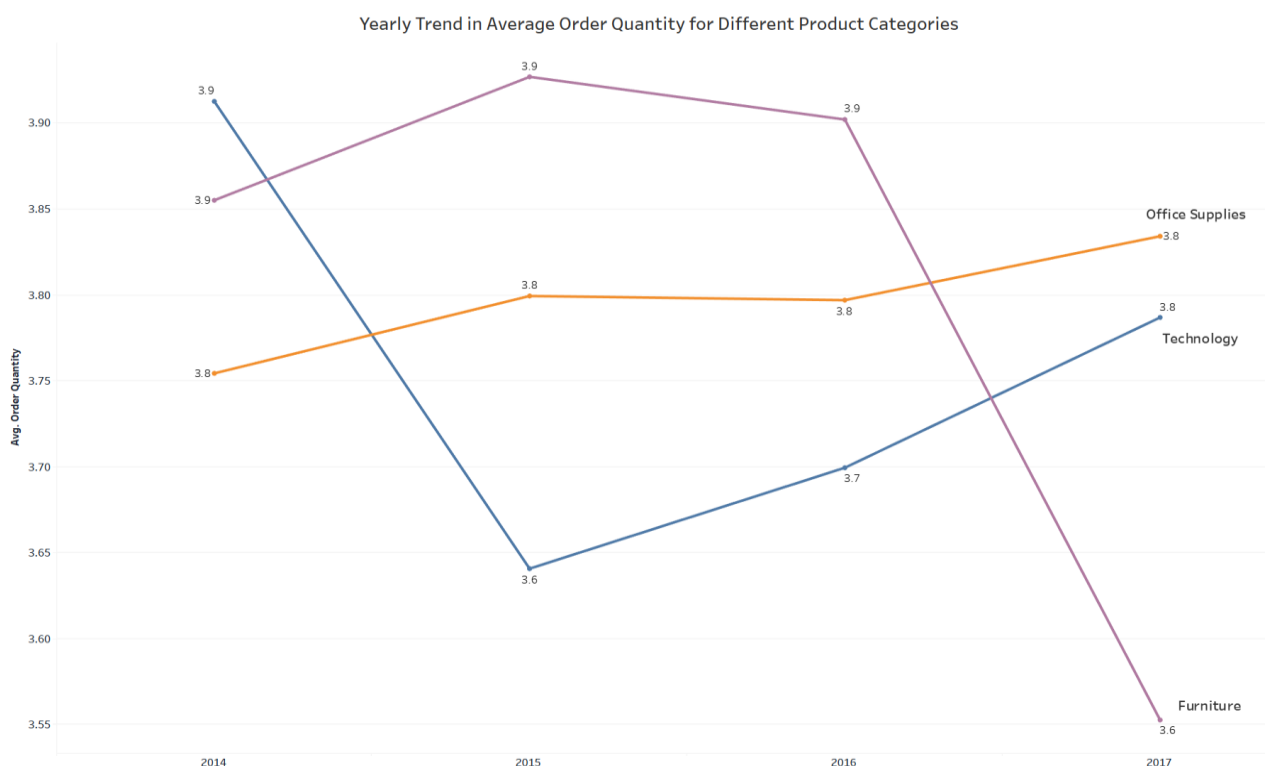
18. How has the average order quantity changed over the years for various product categories?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the yearly trend in the average order quantity. Multiple lines are used here to display this trend across different product categories.

Insight: It can be observed that the average order quantity in 2014 was highest for technology followed by furniture and then office supplies. However, this trend changed in 2015 and 2016 where the average order quantity was the highest for furniture followed by office supplies and then technology. In 2017 there was again a change in trend when the average order quantity for office supplies was the highest followed by technology and then furniture. Overall the average order quantity is in the range of 3 to 4 for all categories.

18. Yearly Trend in Average Order Quantity for Different Product Categories



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

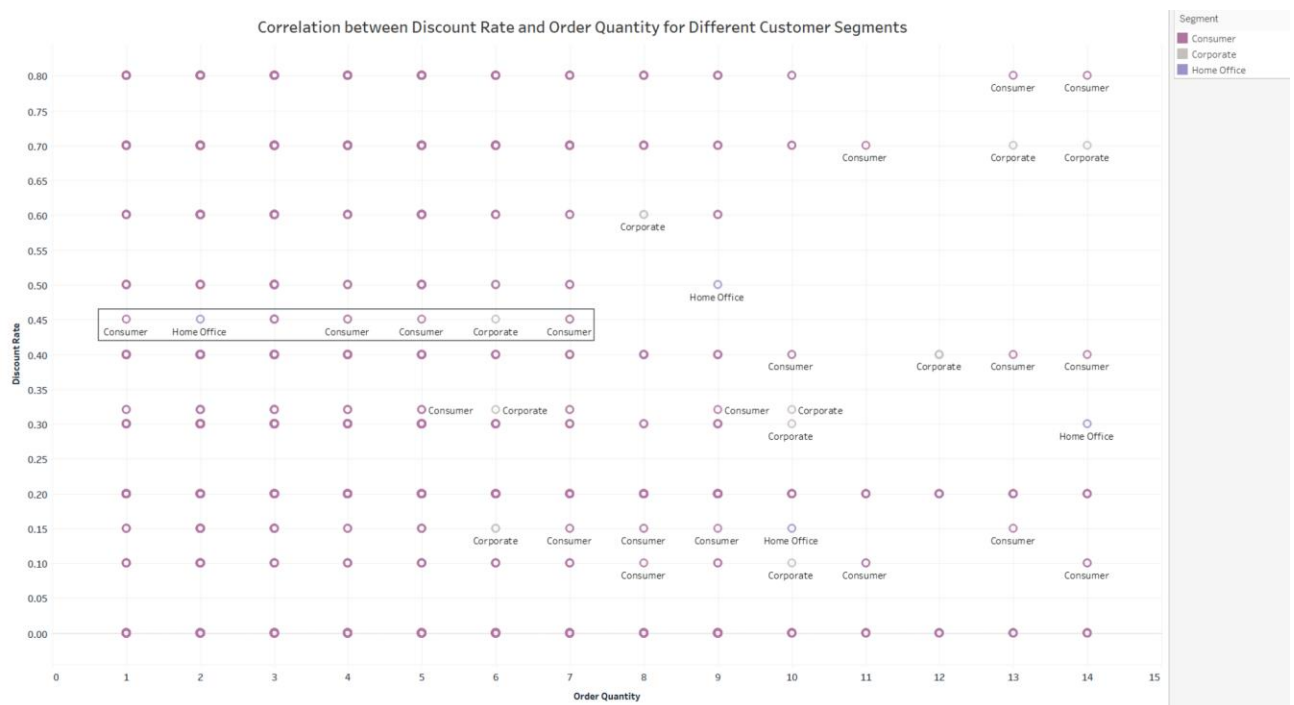
Chart type - Scatter Plot

Why Scatter Plot? Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the correlation between discount rates and order quantities, we use a scatter plot. Order quantity is plotted on the x-axis and discount rate on y-axis. This visualisation shows this correlation for different customer segments .

Insight: It can be observed that there is no correlation between the discount rate and order quantity for any customer segment.

Please refer to the boxed portion of this visualisation as an example. It can be seen that a 45% discount rate is given to all customer segments be it home office, consumer or corporate. So there is no preference for discount rates for different customer segments. Also, the same discount rate of 45% is given for order quantities of either 1 or 7. This observation reiterates the fact that there is no correlation between the two variables.

19. Correlation between Discount Rate and Order Quantity for Different Customer Segments



20. What is the trend of returns and refunds across different regions and product categories?

Note: The data for Returns is given in a separate sheet in the dataset. Hence, that data is joined with the Orders data to generate the following visualisation.

Assumption: For the orders returned, the entire sales amount is refunded.

Chart type - Stacked Column Chart

Why Stacked Column chart? Stacked column charts are a versatile tool for visualising and comparing the composition, proportions, and trends of different subcategories within categories. Two stacked column charts are therefore used here to visualise the number of returned orders and the refunds across various regions for different product categories.

The first one displays the distribution of the number of returned orders across different regions and product categories and the second one displays the distribution of refunds across the same fields.

Calculations added:

- Returned Sales or Refunds: For calculating the sales value of orders that were returned.

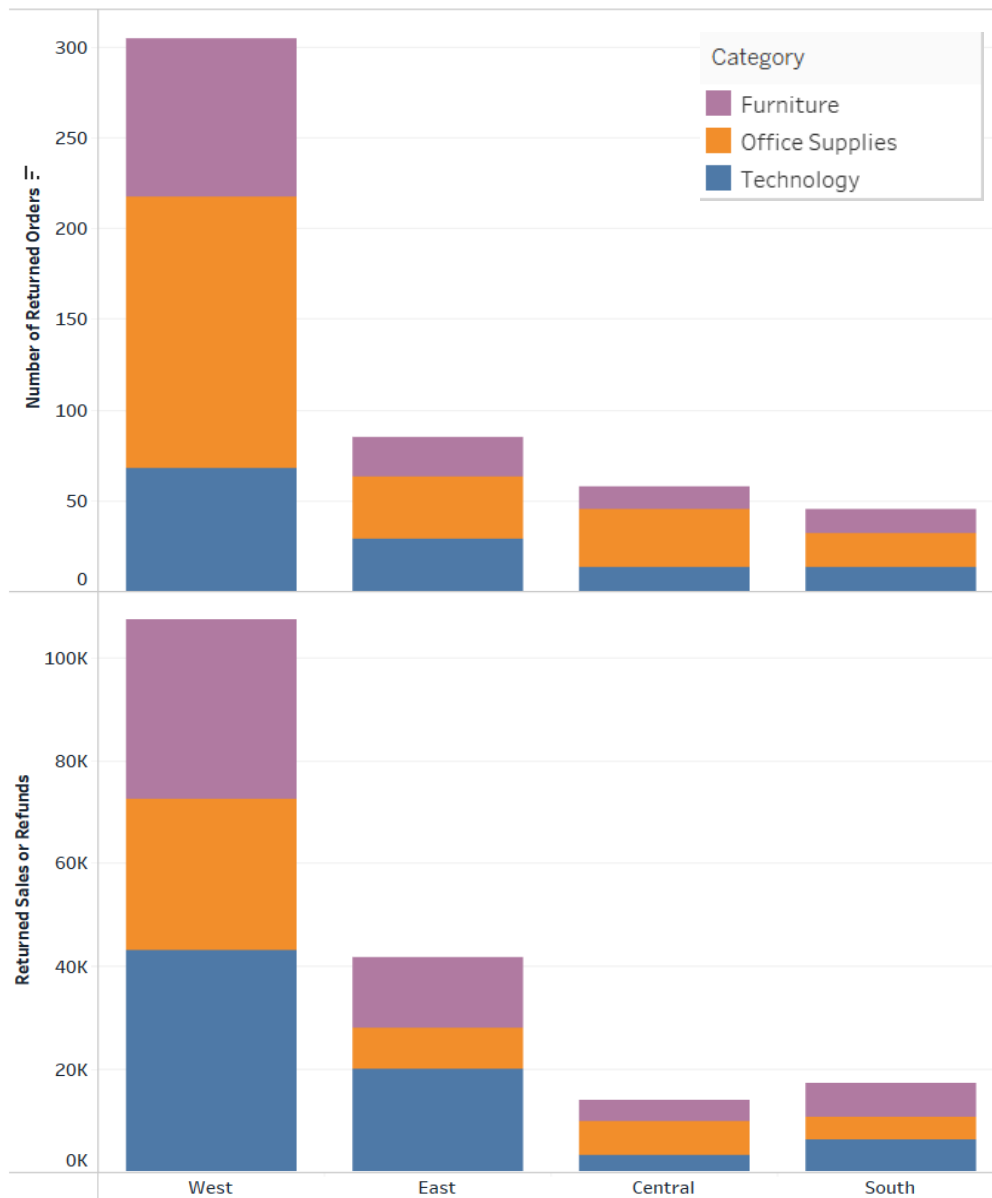
Insight: It can be observed that the highest number of returns have been in the west region, followed by east, central and then south which is also the ranking of the regions in terms of total sales. However, the number of returns and refunds in the west region are much higher than any other regions.

Regarding the categories, it can be observed that the highest number of returns pertain to the office supplies category while the number of returns in the furniture and technology category are nearly equal across all the regions.

Regarding the refund value, it can be observed that in the west and east regions, the largest refunds are of technology followed by furniture and then office supplies. In the central region, the largest refunds are of office supplies and furniture and technology are nearly the same. In the south region office supplies refunds are the least and furniture and technology are nearly the same.

20. Returns and Refunds across Different Regions and Product Categories

Refunds and Returns across Different Regions and Product Categories



21. How do the sales of high-profit products compare with low-profit products over time?

Assumption: It is assumed that high profit products are those with a profit margin of greater than or equal to 20% and low profit products are those with profit margin of less than 20%.

Chart type - Line Chart

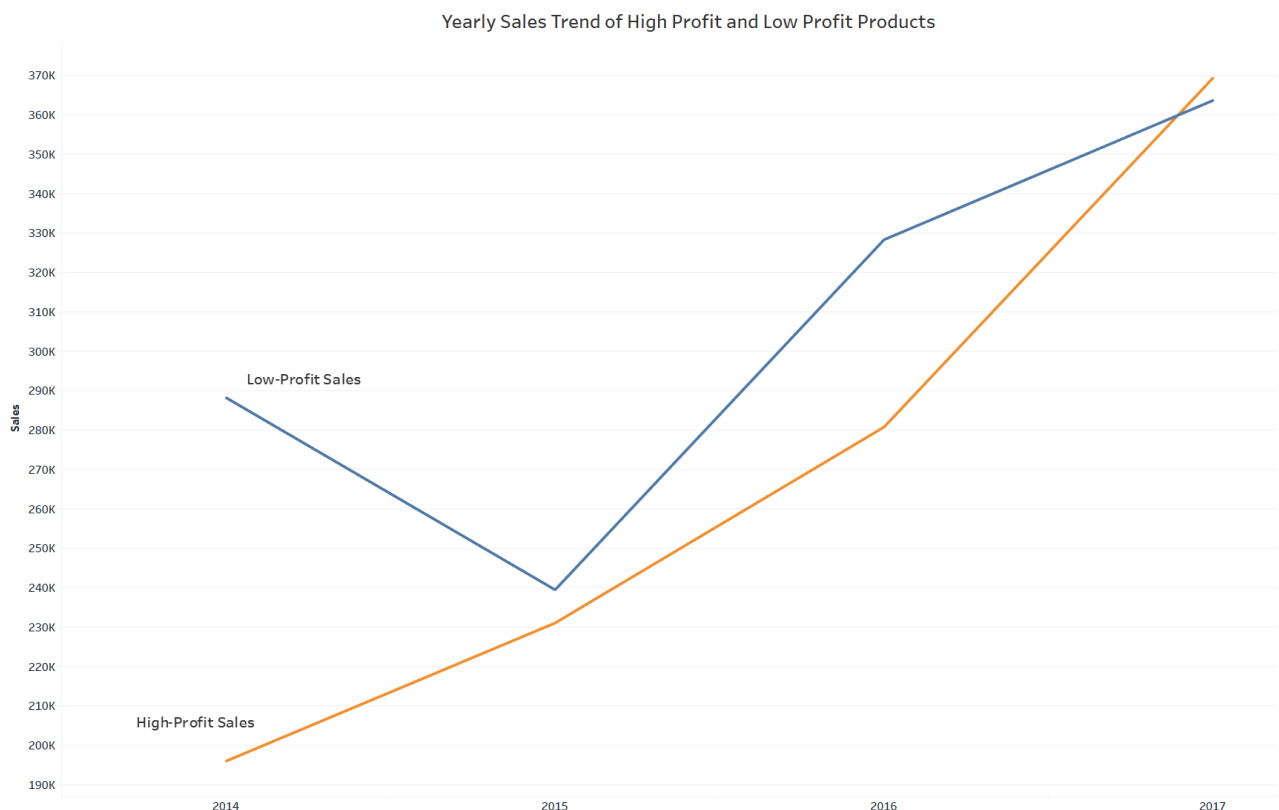
Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the yearly trend in the total sales of high profit and low profit products.

Calculations added:

- Profit Margin: To calculate the profit margin of each product i.e. profit / sales.
- High Profit Margin: To calculate the boolean value of the profit margin field, i.e. True if Profit Margin ≥ 0.20 (high profit products) and False (low profit products) if Profit Margin < 0.20 .

Insight: It can be observed that the sales of low profit products have generally been higher than high profit products over the years with 2017 being an exception to this. However, the sales of high profit products have been in a continuous upward trend over the years. The sales of low profit products do not seem to follow any particular trend as the sales were high in 2014, then a decline in 2015, again a rise in 2016 and another decline in 2017.

21. Yearly Sales Trend of High-Profit and Low-Profit Products



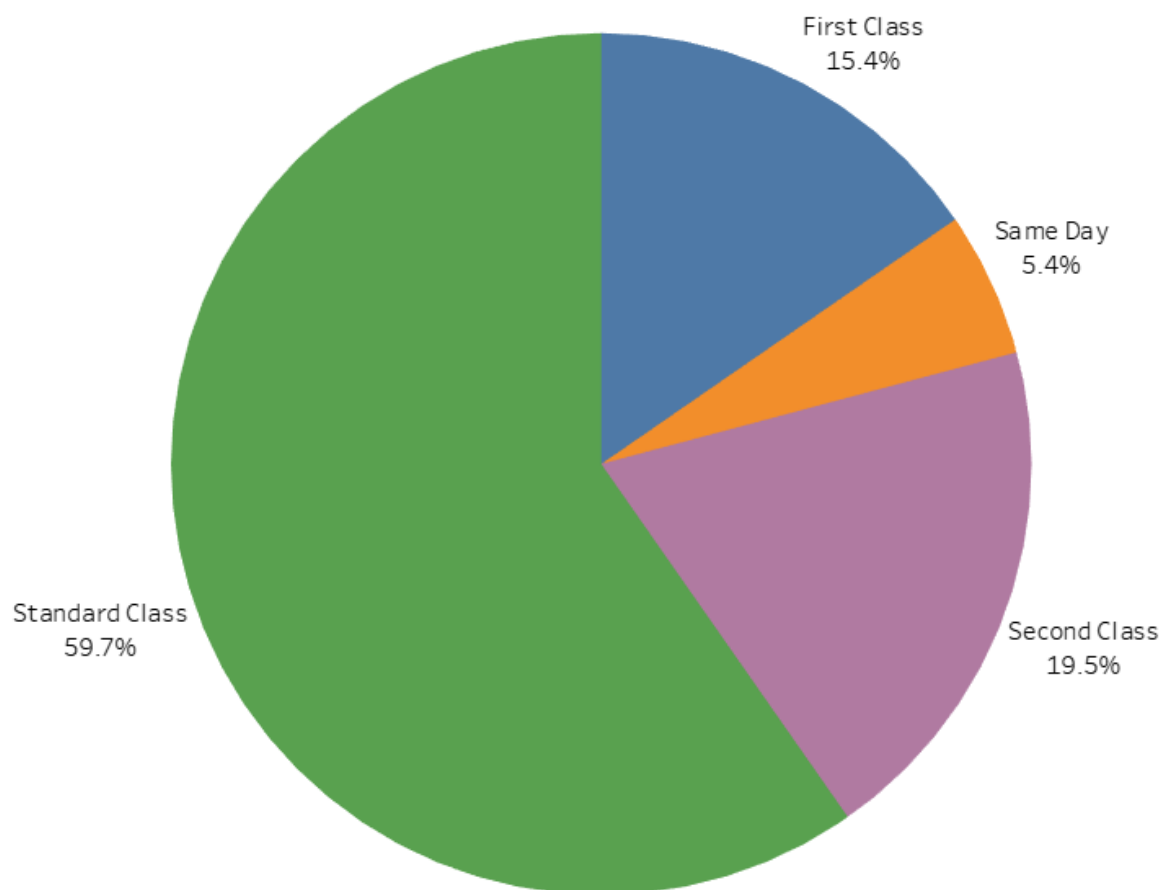
22. Which shipping mode is the most commonly used in the Sample Superstore dataset?

Chart type - Pie Chart

Why Pie Chart? Pie charts are useful to display the proportion of the whole. To identify which is the most common shipping mode, we use a pie chart which clearly displays this information in a simple visualisation.

Insight: It can be observed that the 'Standard Class' shipping mode is the most common as 60% of total orders are shipped using this mode alone.

22. Proportion of Shipping Modes



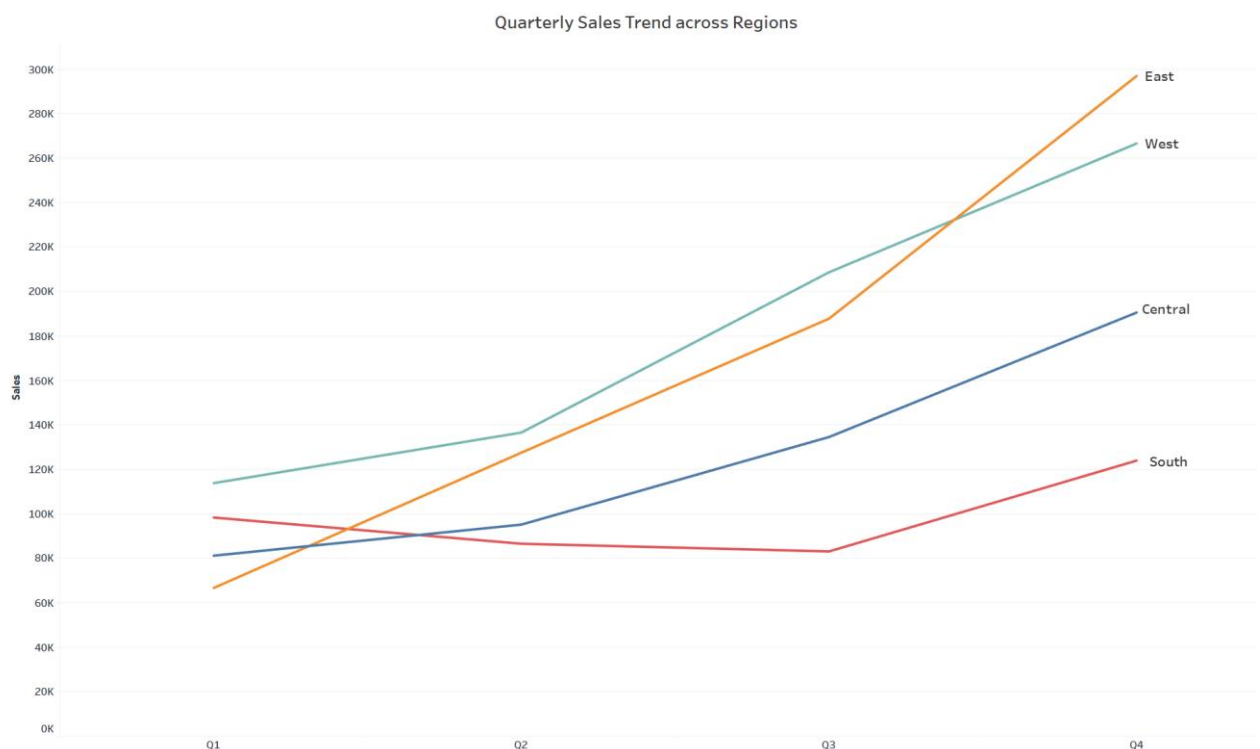
23. How does the sales performance of different regions evolve throughout the quarters of a year?

Chart type - Line Chart

Why Line chart? A line chart is used for visualising and analysing data trends over time. Line chart is therefore used here to display the quarterly trend in the total sales over all the years collectively for all the regions.

Insight: It can be observed that the sales are generally in an upward trend over all the four quarters of the year across all the regions. Additionally it can also be observed that the sales in the west region have been the highest in Q1, Q2 and Q3 but in Q4 the sales in the east region are the highest.

23. Quarterly Sales Trend across Different Regions



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

Chart type - Bar Chart

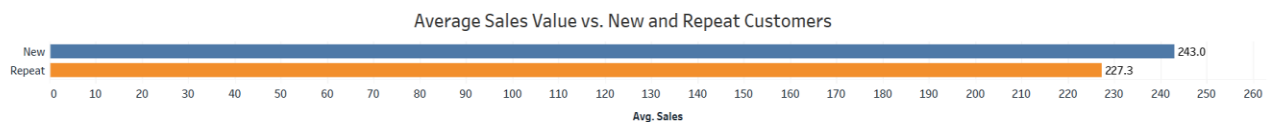
Why Bar chart? A bar chart is a useful visualisation tool for representing the distribution of categorical data. Here, multiple calculated fields are added to the data to calculate the number of repeat and new customers. This represents categorical data and hence a bar chart is used.

Calculations added:

- Date of First Purchase: To calculate the MIN(Order Date) for each customer.
- New or Repeat Customers: If Order Date = Date of first purchase, then “New” else “Repeat”.

Insight: It can be observed that the average sales or order value of new customers is higher than repeat customers.

24. Average Order Value for New and Repeat Customers



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

Assumption: For the orders returned, the entire sales amount is refunded.

Chart type -Map

Why Map? A map is a useful visualisation to display the distribution of geographical data. Therefore here a map is used to show the number of returned orders geographically.

Insight: It can be observed that the highest order returns are from the state of California.

25. Geographical Distribution of Returned Orders

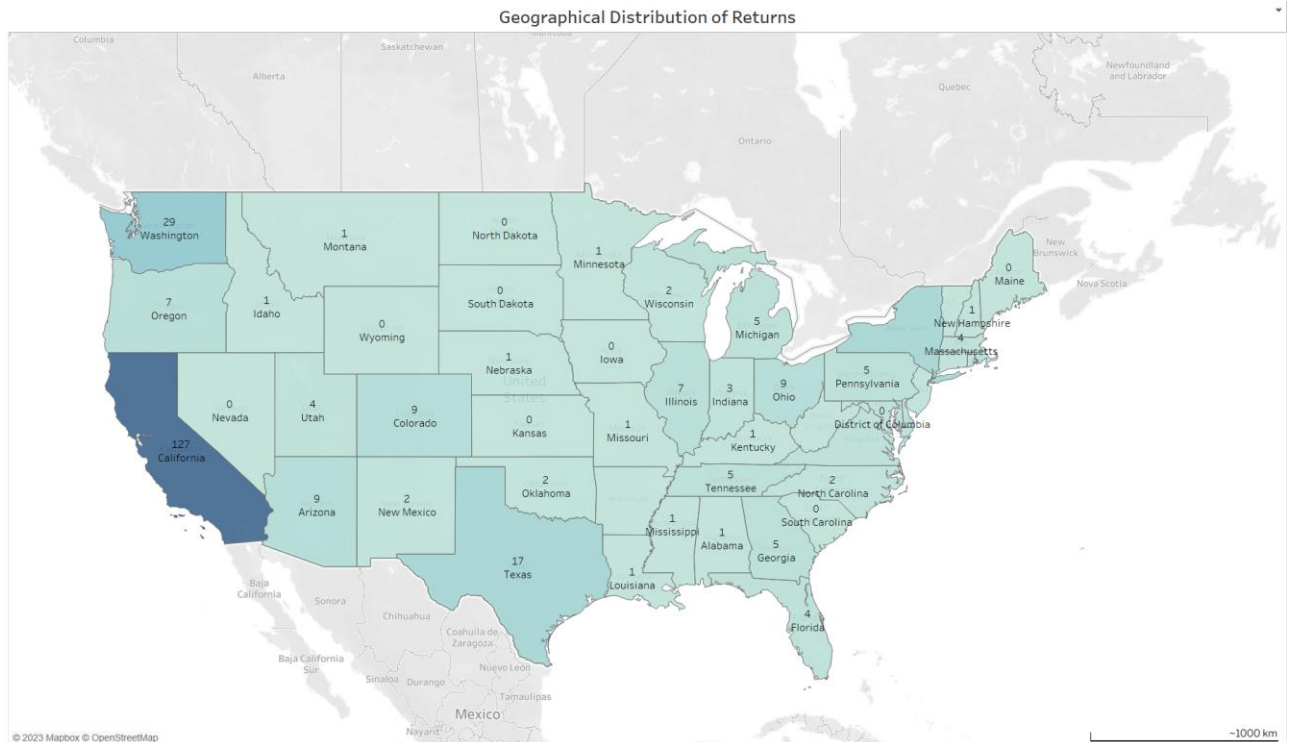


Chart type - Bar Chart

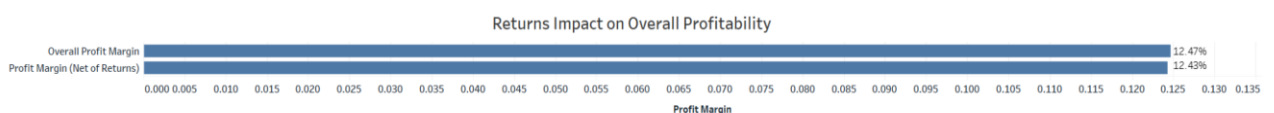
Why Bar chart? A bar chart is a useful visualisation tool for comparing values. The bar chart is therefore used here to visualise the impact of returns on overall profitability where the measures of overall profitability and profitability (net of returns) can be easily visualised.

Calculations added:

- Profit on Returned Orders: For calculating the profit on orders that were returned.
- Value of Returned Orders: For calculating the sales value of orders that were returned.
- Overall Profit Margin: $\text{Total Profit} / \text{Total Sales}$
- Profit Margin (Net of Returns): $(\text{Total Profit} - \text{Profit on Returned Orders}) / (\text{Total Sales} - \text{Value of Returned Orders})$

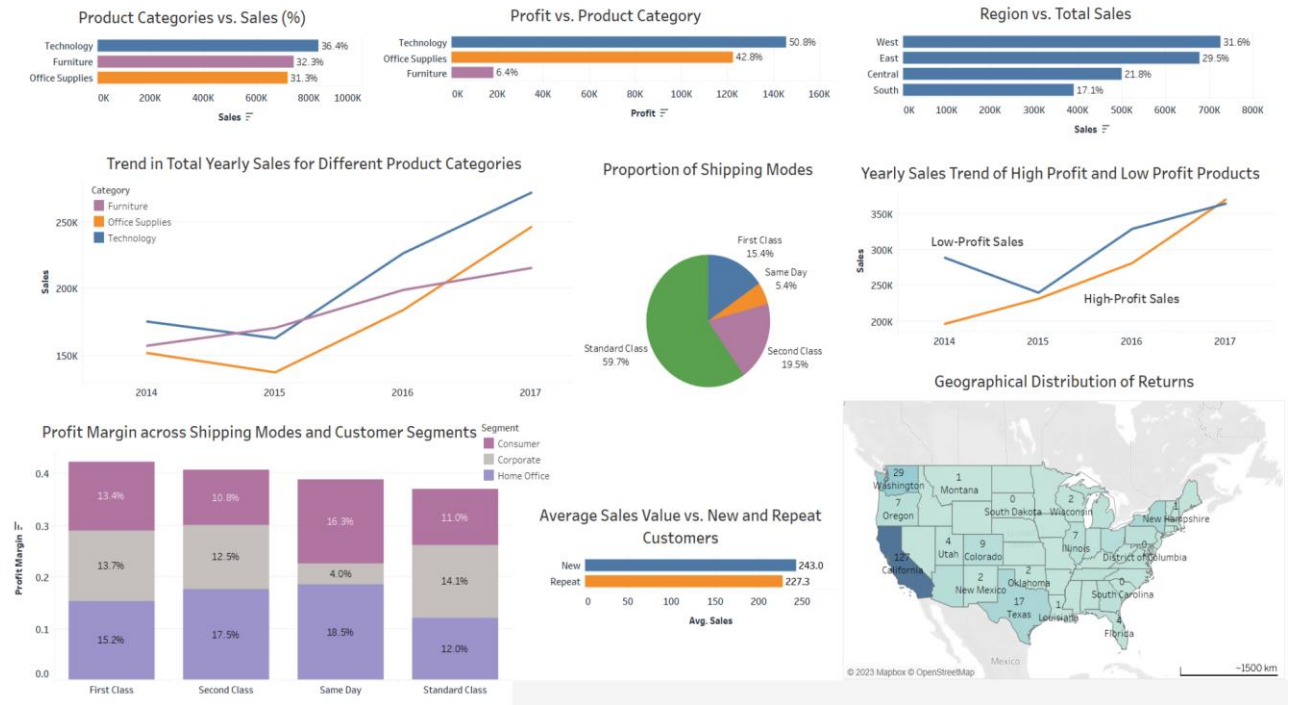
Insight: It can be observed that despite returned orders, there is not much change in the overall profitability. The profit margin reduced by a mere 0.04% due to returns.

25. Returns Impact on Overall Profitability



Superstore Sales Dashboard

(Highlighting some of the important observations of the analysis)



Conclusion:

The Superstore Sales dataset contains a wide variety of information about the sales of different products in the United States of America. The key insights from the analysis can be used to make important business decisions for improving the sales and profit margins across different regions, product categories and customer segments.

Created by: Aiman Sahay