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.

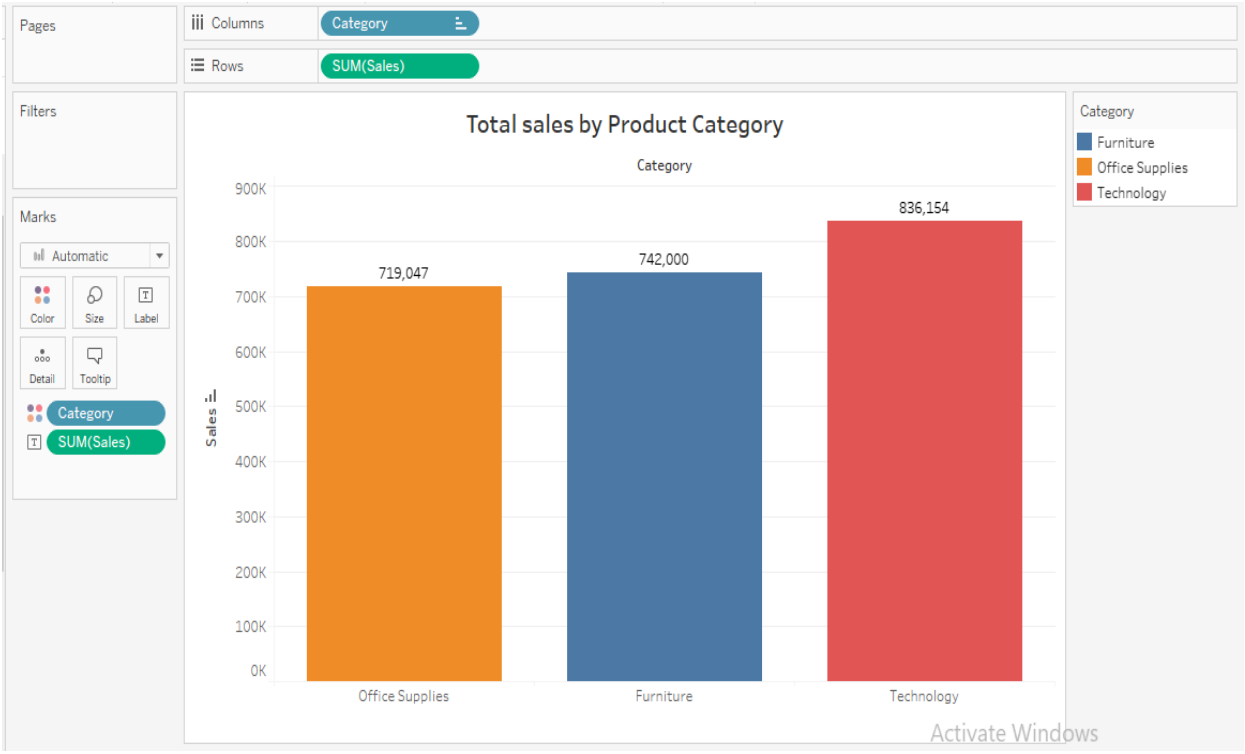
By completing this task effectively, you'll not only demonstrate your proficiency in data visualisation and analysis but also showcase your ability to effectively communicate complex concepts through both text and charts.

Good luck!

Questions:

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

Ans:

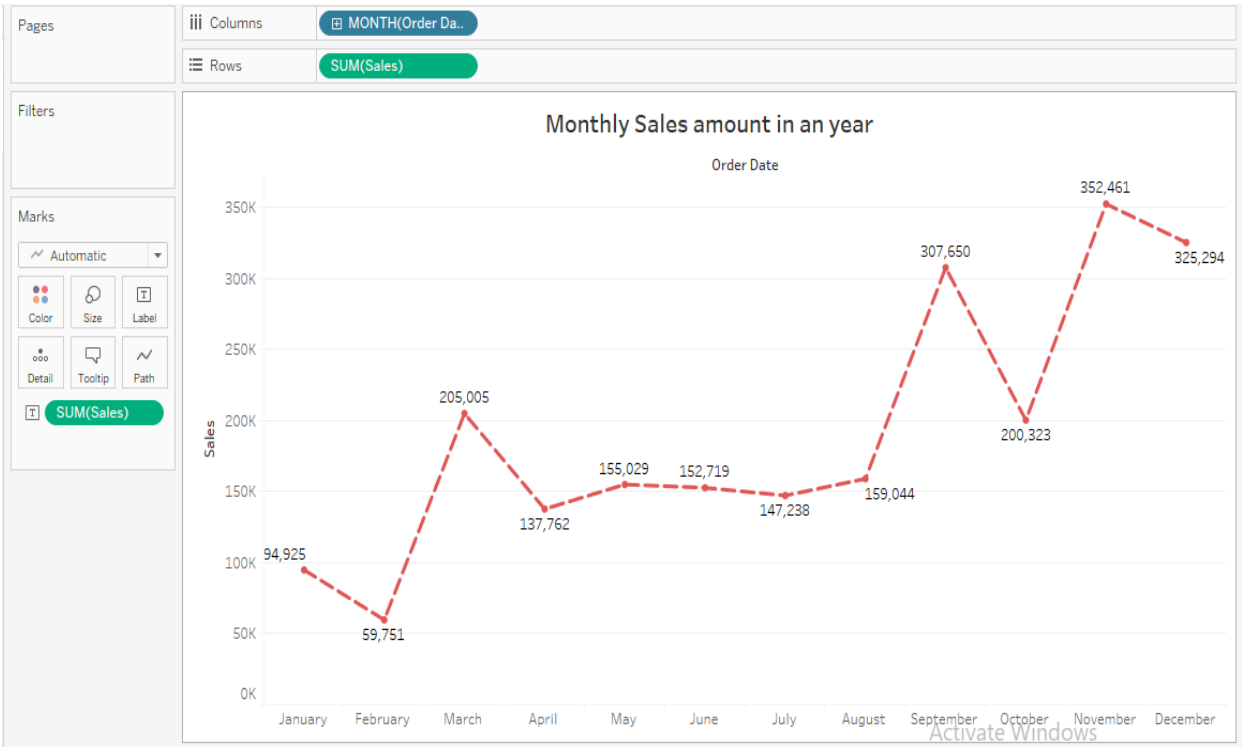


When the category labels are lengthy, as they are in this instance (Technology, Furniture, Office Supplies) the bar chart is quite helpful. Compared to a vertical column chart, where the category titles are more difficult to see this arrangement.

Viewers may easily identify whether a category has the highest or lowest sales by ranking the categories according to length. It's a natural method of representing ordinal data or data with a distinct order, and it makes it clear right away that sales of technology are highest, followed by those of furniture and office supplies. The chart can also be scaled to accommodate more data points and a variety of display formats thanks to its scalability.

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

Ans:

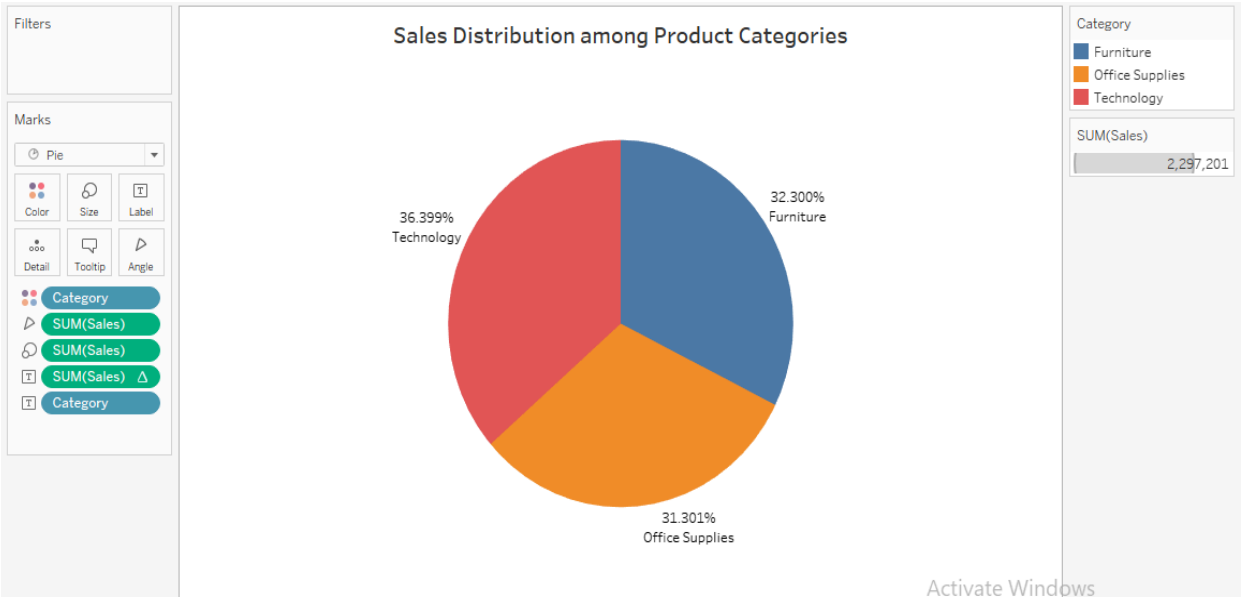


The line chart clearly illustrates the notable trends and swings in the data making it the perfect visualization for demonstrating how monthly sales quantities change over the course of a year. It gives a concise account of sales performance, demonstrating that sales begin slowly in the early months peak in March, descend into a summertime trough, and then increase again to their highest peak in September. As the year comes to a finish there is a decline followed by another increase.

This kind of graphic is very good at showing these temporal shifts providing a visual narrative of periods with high and low sales activity and verifying the dynamic nature of monthly sales throughout the course of the year.

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

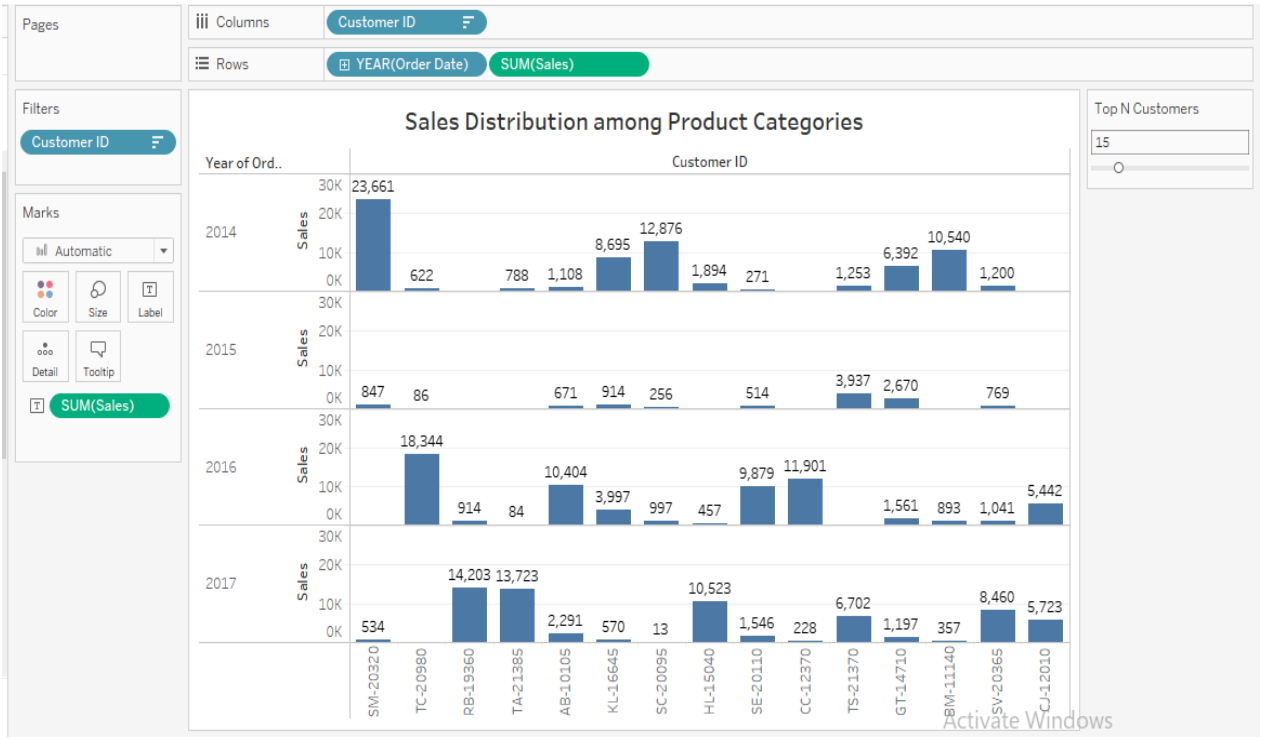
Ans:



Pie charts are a useful visual aid for displaying how sales are distributed among different product categories because they give an instant impression of the market share that each category has in relation to the overall volume of sales. Technology is represented in this figure as the greatest proportion at 36.4%, followed closely by office supplies at 31.3% and furniture at 32.3%. When communicating the relative size of each category as a whole this visual approach works exceptionally well enabling viewers to compare the segments in a comprehensible manner.

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

Ans:

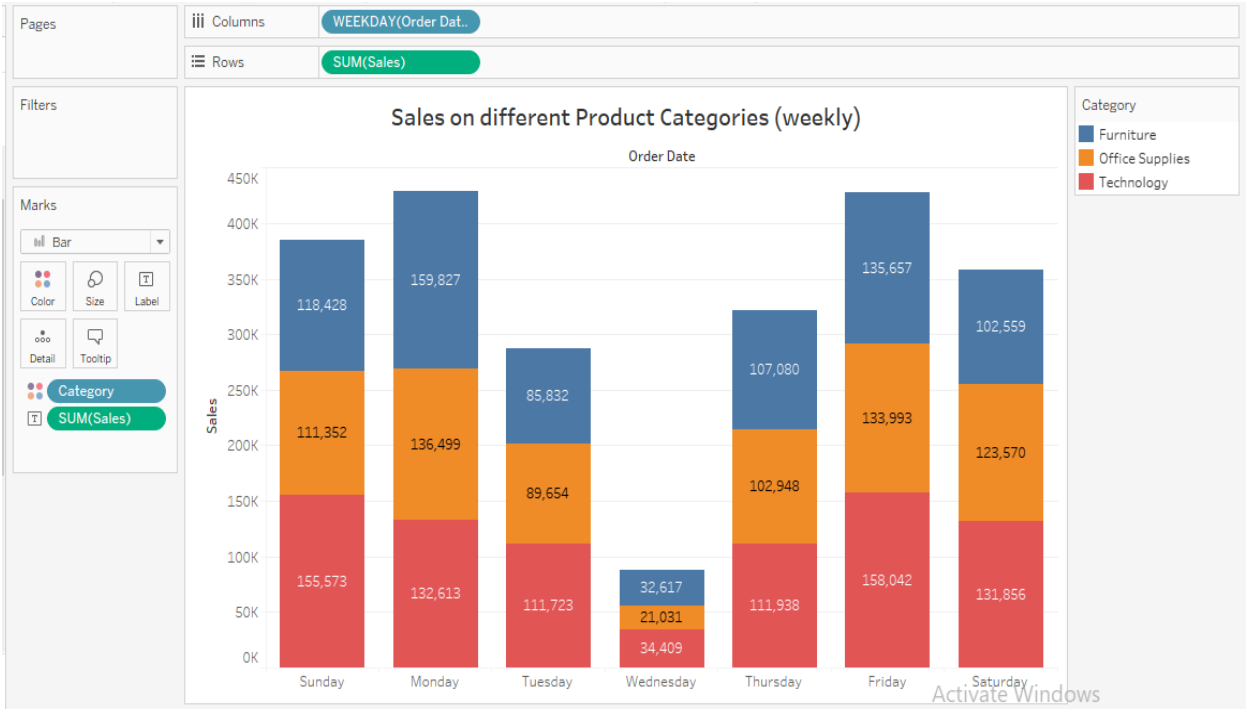


There are several benefits to using a bar chart to show sales performance for specific clients, especially when tracking this information over time. Because of its style which is typically easier to read than horizontal bars it excels at providing clear comparisons of sales numbers between various clients.

The chart becomes much more interactive with the addition of a 'Show Top N' parameter that uses 'Customer ID' as a filter. This allows users to dynamically focus on top-performing customers allowing them to tailor the analysis to specific needs without clutter and making the chart both informative and adaptable. The ideal option for this kind of data is a horizontal bar chart because of its clarity, readability and interactivity.

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

Ans:

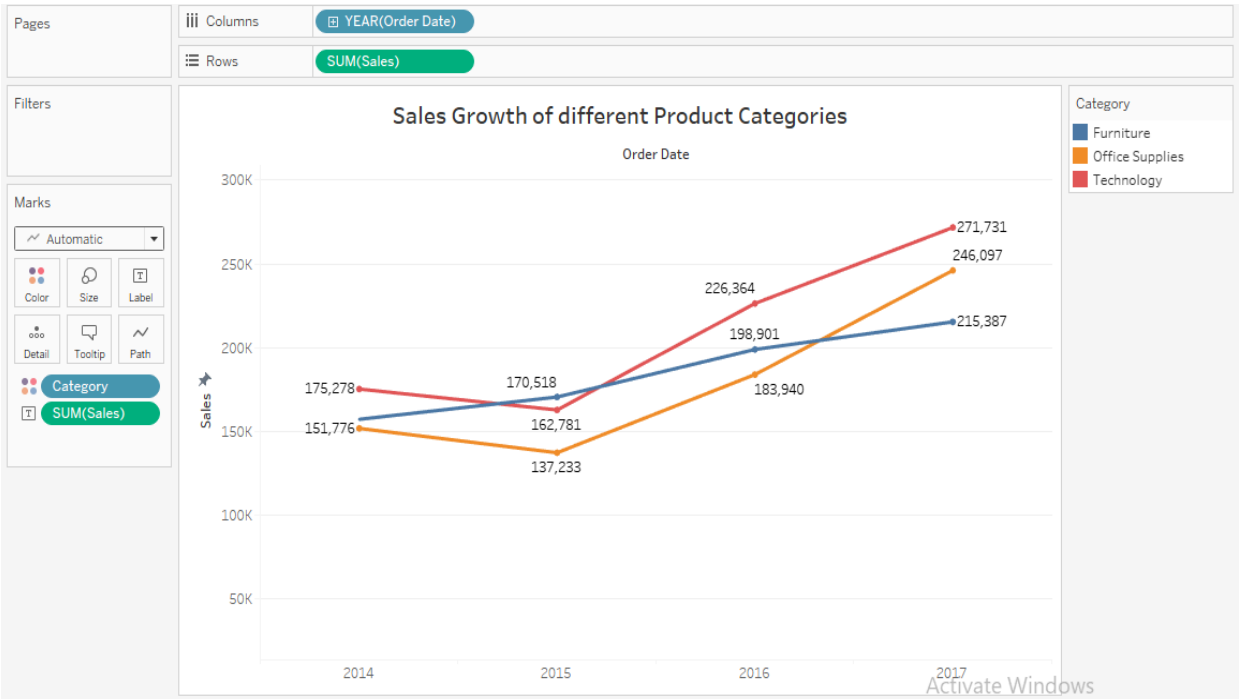


The ability to compare total sales as well as category-specific performance within the same chart makes the stacked bar chart a useful visual aid for analyzing differences in sales over the course of the week and between product categories.

This two-tiered method makes it possible to quickly evaluate both general and category-specific trends such as pinpointing days with the highest sales or figuring out which product categories are more popular on particular days. Each category's color coding improves the visual distinction between them, making it simpler to tell them apart. It shows that performance was below average on Tuesday, Wednesday, and Thursday and far above average on the remaining days especially on Monday and Friday. This kind of graphic is especially good for displaying how various segments relate to the total in both a temporal and category sense.

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

Ans:

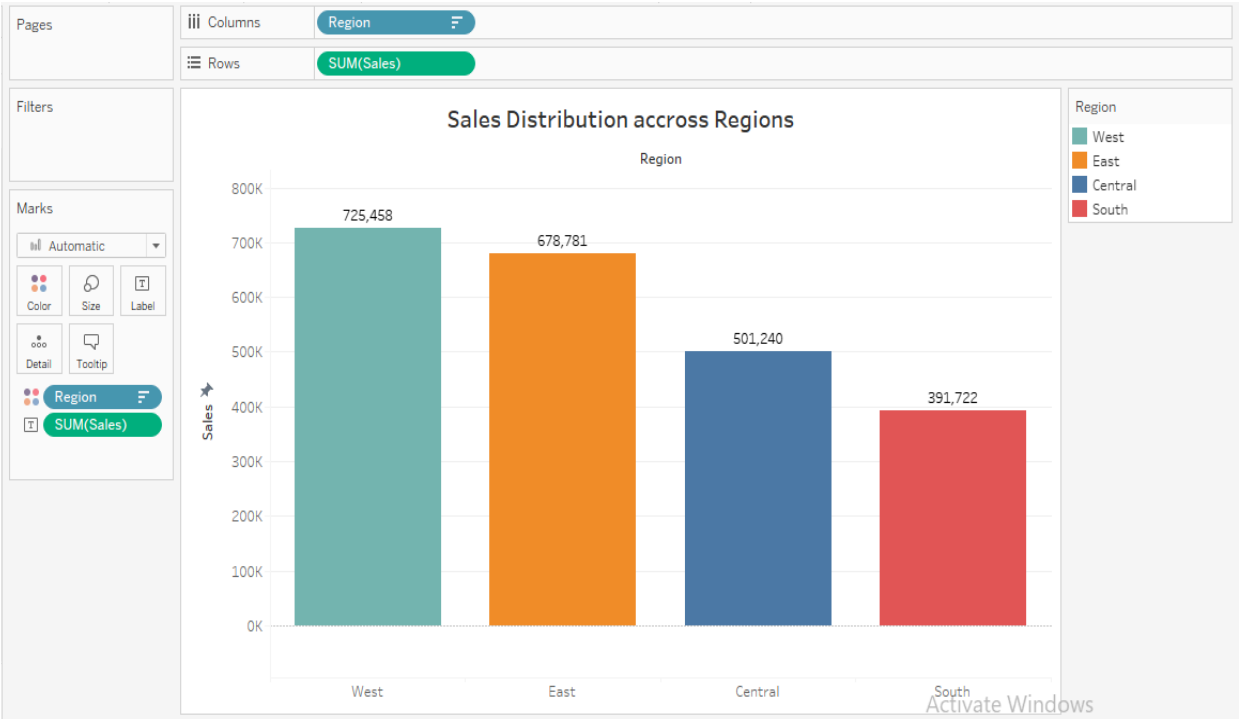


The line chart is the best option for displaying the sales growth of various product categories over time since it presents patterns and the rate of change in an understandable and straightforward manner. Every line signifies a certain product category and the evolution throughout time demonstrates the rise or fall in sales. When analyzing time-series data this kind of graphic works especially well since it helps show how a number changes over time.

This graph shows a significant rise in sales for the categories of technology, office supplies, and furniture between 2014 and 2017. Notably, technology is in the lead with a notable increase, particularly from 2016 to 2017. This suggests that demand is increasing. Office supply is growing steadily and consistently which suggests that market demand is stable. Furniture, on the other hand, shows more sluggish growth which may indicate that strategic actions are needed to improve its sales performance.

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

Ans:

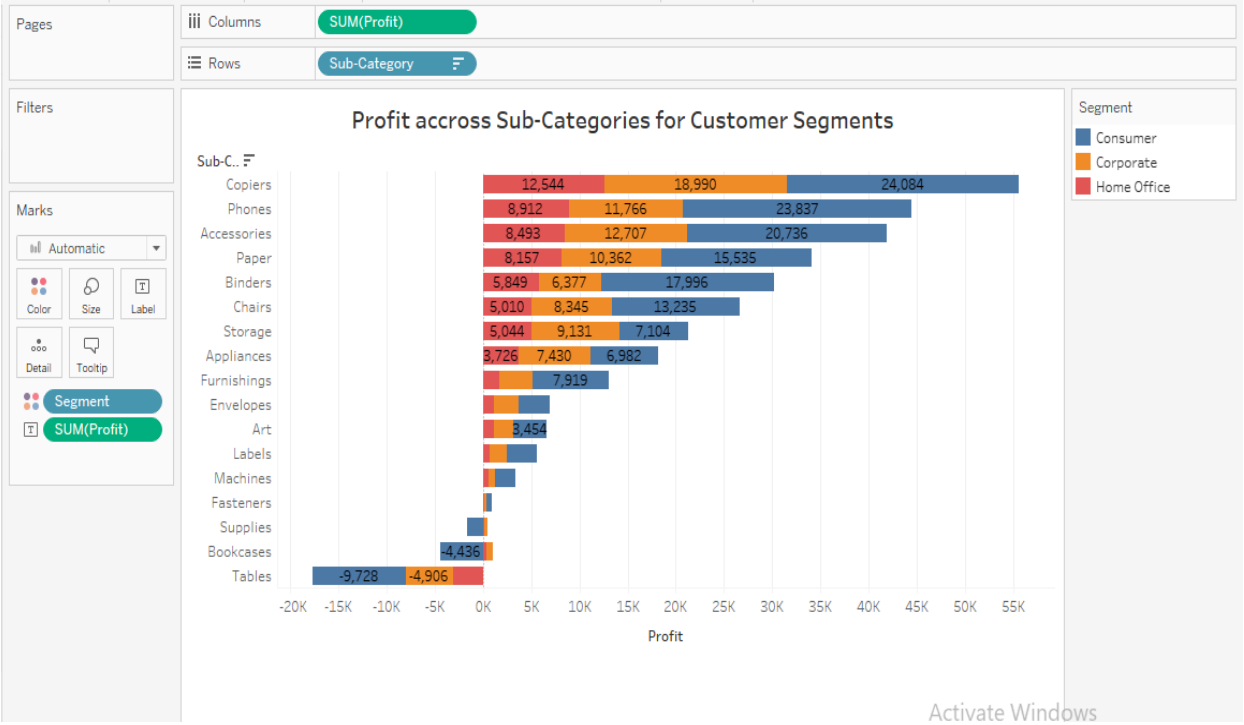


The bar chart's simple form which matches how people read text naturally-from top to bottom and left to right-makes it a great option for visualizing sales distribution across various locations. Furthermore, a clear hierarchy of performance can be shown by the bars' falling or ascending order which can immediately express the ranking of regions based on sales volume.

This kind of chart is the greatest choice for efficiently viewing the sales distribution because it works well for comparing a few different categories. With the greatest sales, estimated at $725,458, the West region is in the lead, followed by the East at roughly $678,781. With sales of nearly $501,240, the Central region comes in second, while the South region has the lowest sales, approximately $391,722.

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

Ans:

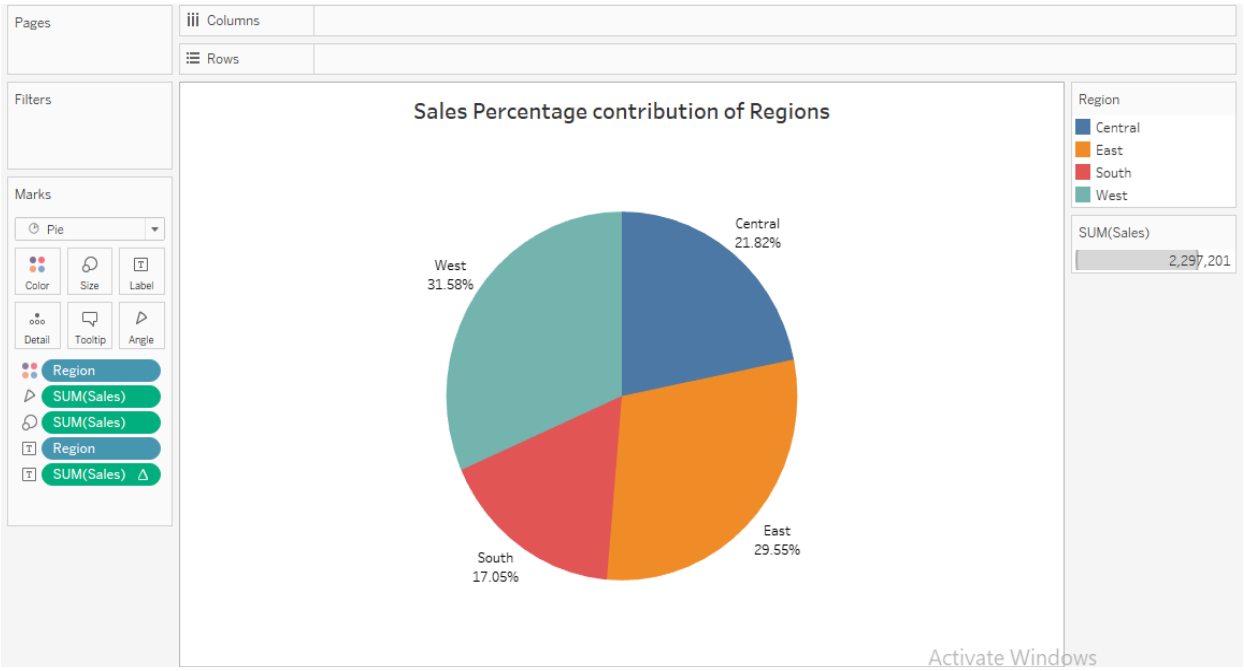


The ability of the horizontal stacked bar chart to divide and contrast profits by subcategory and client segment makes it a useful visualization option. by designating different hues for every group of customers. The design facilitates comparative analysis by displaying the overall profits for each subcategory and the contribution of each segment. It also handles negative values with skill, showing any subcategories that result in losses with bars that extend leftward from the zero line. To help with finding the most and least profitable sectors quickly, the subcategories are also rated according to profitability.

The graph shows that while "Tables" and "Bookcases" lose money in every category, "Copiers" are quite profitable, especially in the Corporate section. The consumer segment exhibits a wide range of profitability, particularly in the "Phones" and "Accessories" categories, indicating a healthy market. The Home Office category, on the other hand, makes significant profits in the "Chairs" and "Storage" categories, indicating a need for home office furniture. These patterns draw attention to the necessity of addressing the underlying problems in the underperforming subcategories as well as the possibility for customized marketing approaches.

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

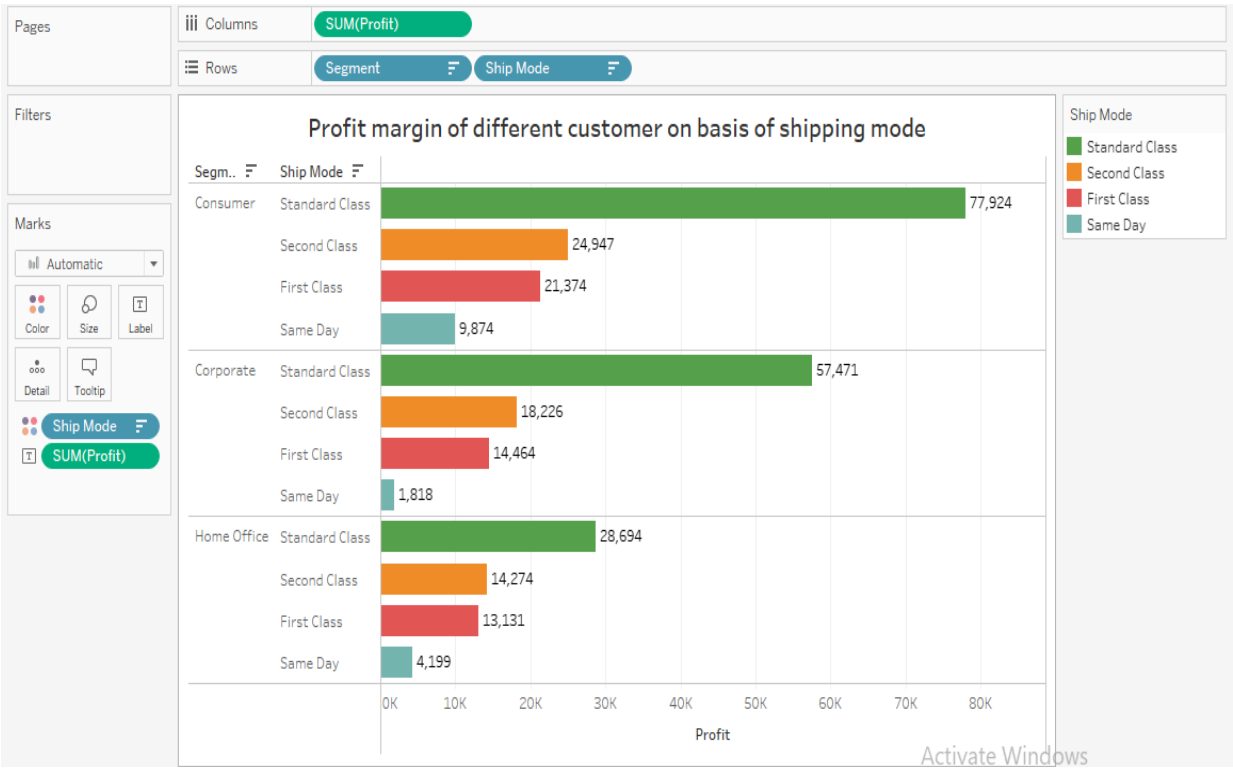
Ans:



Because it naturally depicts proportions, a pie chart is a powerful visual aid for displaying the percentage contribution of each region to total sales. This works especially well for quickly grasping the distribution and sales supremacy among various locations. Furthermore the explicit labeling of percentages on the chart strengthens the visual depiction and makes it possible to quickly identify the share of each location. This pie chart which shows the West as the leader at 31.6%, the East at 29.5%, the Central at 21.8%, and the South at 17.1%, clearly illustrates the percentage contributions made by each region to overall sales.

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

Ans:



Within different customer segments, the grouped horizontal bar chart clearly shows profit margins by shipping manner. The several shipping modes are represented by colored divisions in each horizontal bar which represents a distinct section and makes it simple to determine how much each method contributes to overall revenues. The horizontal arrangement of the chart which allows for lengthy labels and several categories without compromising readability, improves its clarity. The graphic makes clear that Standard Class shipping is the most lucrative of all the divisions with the largest margins being seen in the Consumer group. On the other hand Same Day shipment is the least profitable indicating possible problems with cost-effectiveness or demand that should be resolved to maximize profitability.

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

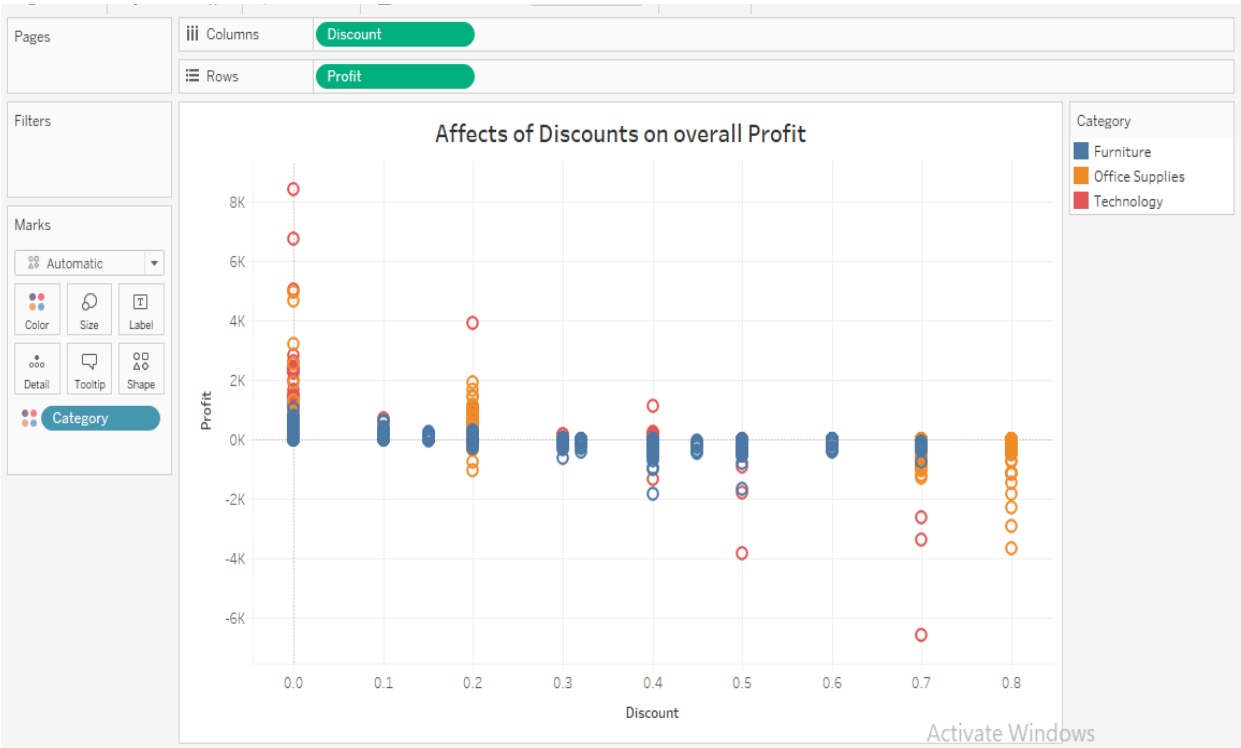
Ans:

The average order processing time for several product categories can be effectively displayed using a vertical bar chart where the height of each bar represents the order to shipment time. By using the computed variable 'Order Processing Time' which is the number of days that separate the 'Order Date' and 'Ship Date' the data is kept accurate and pertinent to the processing timeframe. This type of chart effectively communicates the average timings allowing for clear comparisons between the Technology, Office Supply and Furniture categories.

The average shipping times for furniture, office supplies, and technology are all somewhat under four days as this figure illustrates. This consistency points to the same order processing mechanism for all categories. Office Supplies may have less room for process optimization especially in the Office Supplies category to achieve even better efficiency, since they take an average of 3.98 days compared to 3.92 days for Furniture and Technology.

12. How do discounts affect overall profit?

Ans:

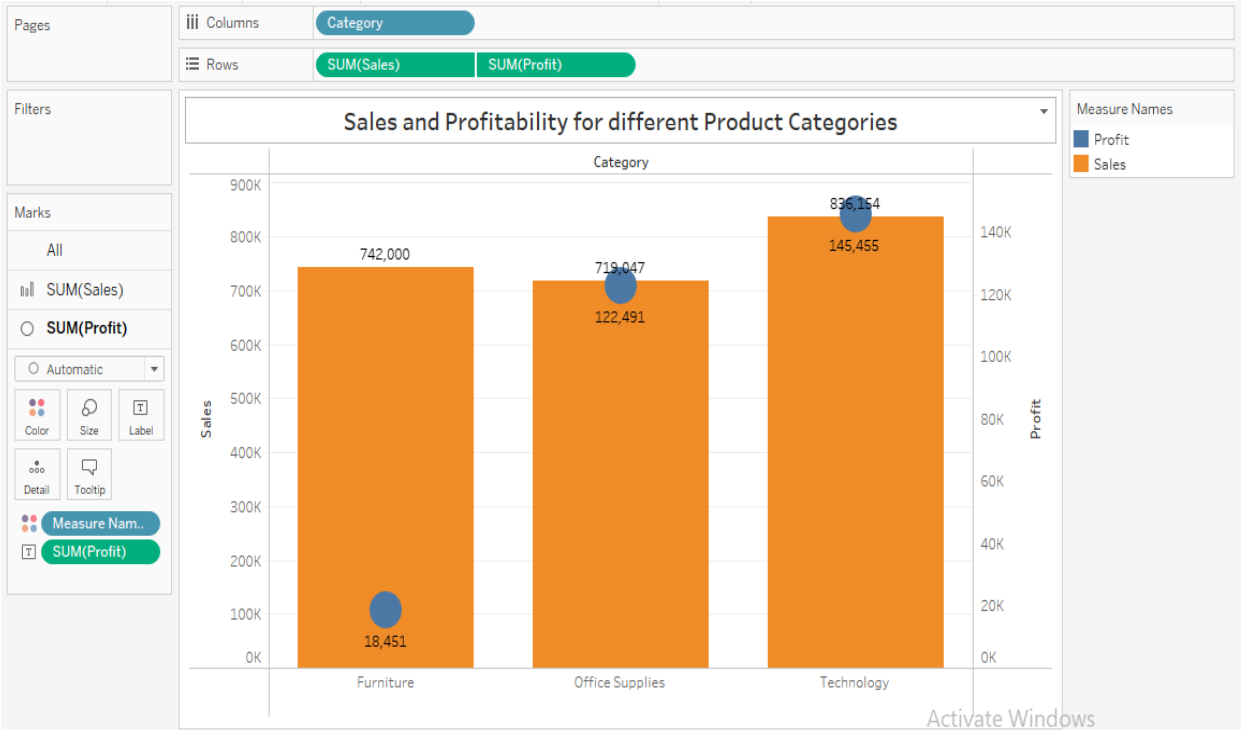


A scatterplot is used to visualize the relationship between two quantitative variables, allowing for the detection of patterns, trends, and correlations. It helps identify whether an increase in one variable corresponds to an increase or decrease in another making it ideal for examining relationships, spotting outliers, and comparing multiple data sets.

Here the comparison between two numerical data are being plotted ie. 'Discount' and 'Profit' are divided into categories where they are denoted by different colors. The spread of small circles described the relation how discount affects the earning profit of different parameters here. Clearly we can see that in the case of 'Technology' the more discounts are affecting the profit margin, somehow similar to 'Office Supplies' but with 'Furniture' having a different case it is maintaining a constant pace between discount and profit.

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

Ans:

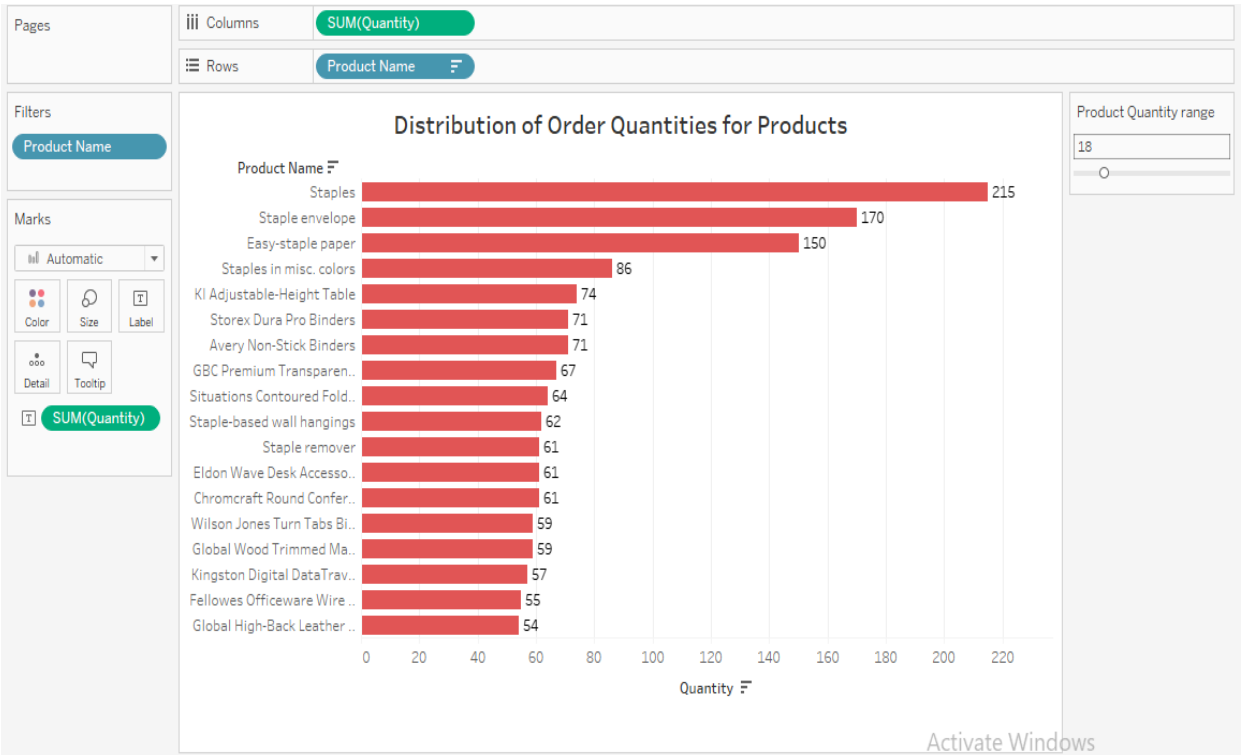


The dual-axis bar chart, which compares two variables on one chart, is very useful for illustrating the relationship between sales and profitability across product categories. The main sales bars present a clear comparative picture of the revenue that each product category generates, emphasizing the categories that have the biggest impact on the bottom line. A direct comparison of the bottom line for each category is made easier by the secondary circles for profit which are all aligned on the same axis and demonstrate how revenue is converted into actual profit.

Technology does well in both areas since it has the highest profitability and the largest volume of sales. Furniture sales volume is high but profitability is much lower, indicating a discrepancy between sales success and profit generating. Office Supplies exhibits a moderate sales volume and profitability, indicating a balanced relationship between sales and profit. These metrics appear compatible with one another.

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

Ans:

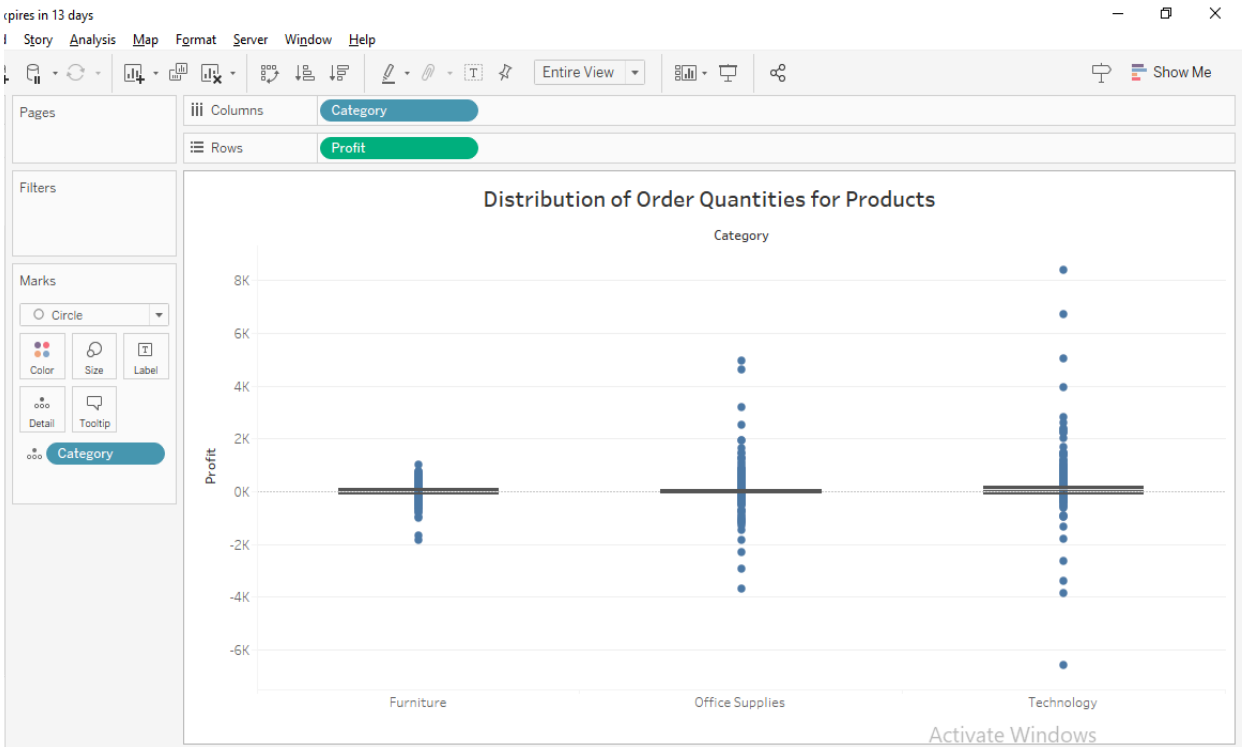


The horizontal bar chart's straightforward design, which improves readability for lengthy product names and makes it simple to compare data points makes it perfect for illustrating the distribution of order numbers across products. While the interactivity provided by the 'Product Quality Range' parameter allows for a concentrated examination on the most significant goods, optimizing both space and interpretative clarity its scalability makes it ideal for managing massive datasets.

The top ordered products in this chart are ‘staples’, ‘staple envelopes’, and ‘easy-staple paper’, which suggests a strong demand or turnover. The graphic highlights the importance of strategic inventory management particularly for high-volume commodities as there is a significant variation in order quantities among products.

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

Ans:

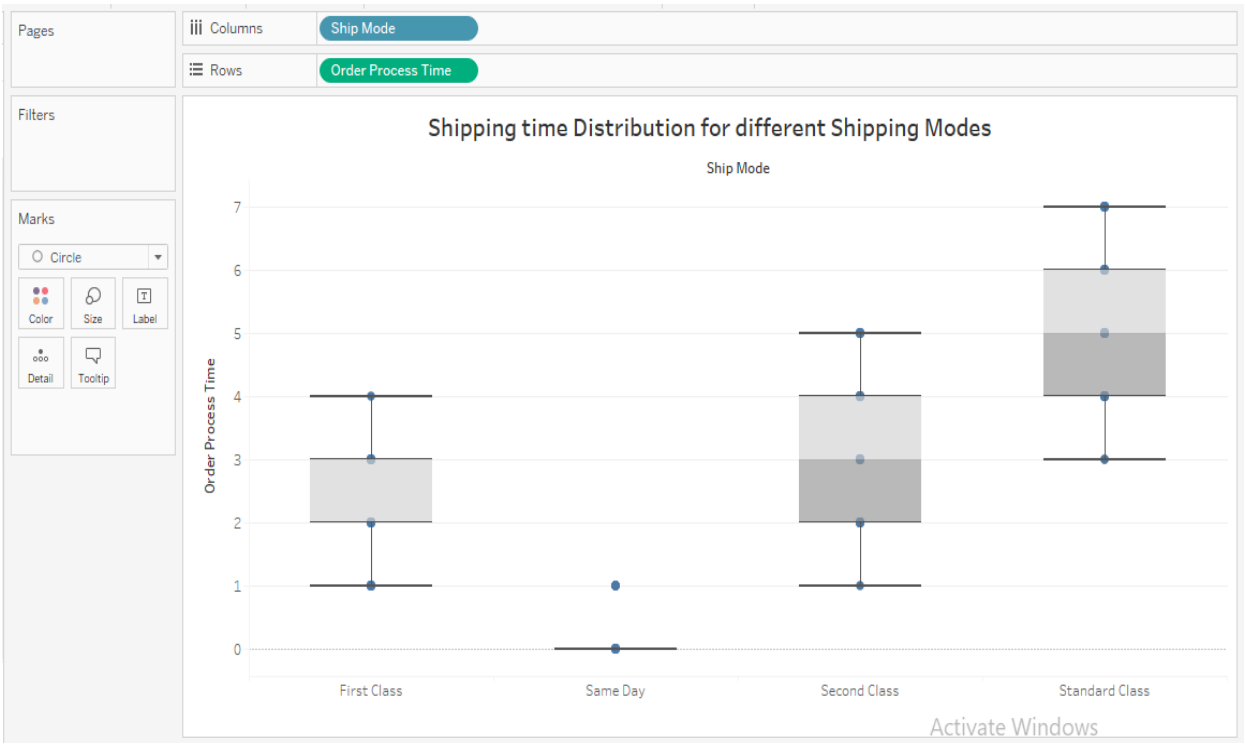


A great option for showing the earnings distributions across several product categories is the box-and-whisker plot. In order to provide an idea of the distribution's spread it additionally shows the interquartile range (IQR) which draws attention to the middle 50% of the data. The 'whiskers' represent the data's range and the spots outside of them denote outliers.

Different profit characteristics for technology, office supplies, and furniture are shown in this box plot. Office Supplies display a more consistent profit range with a median profit lower than Technology but with fewer losses suggesting a stable but less lucrative market; Furniture shows potential for losses as indicated by outliers below the profit line; Technology exhibits a broad range of profits with significant outliers suggesting high variability and the potential for large gains. Within each product group focused methods for optimizing gains and reducing losses can be informed by the variation in profit distribution among different categories.

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

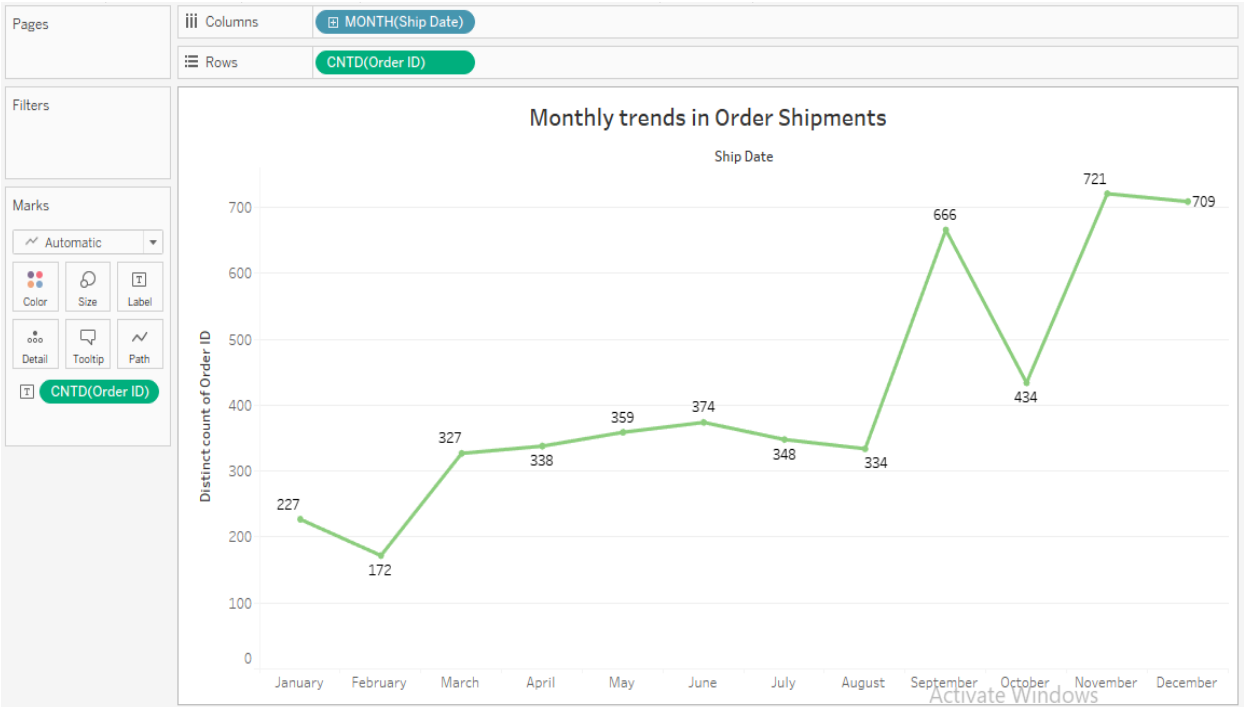
Ans:



It shows both regular and exceptional shipment durations and offers a thorough comparison of the central shipping time tendencies and the diversity within those timings. With a narrow interquartile range suggesting consistent delivery times, this plot shows that First Class has a comparatively short shipping time. As anticipated Same Day delivery times are the fastest with a median near zero. An outlier below zero however is most likely the result of an anomaly or a mistake in data entry as negative shipping times are not possible. The interquartile range of Second Class is wider than that of First Class indicating a wider variety of shipping timeframes. Compared to First Class the median shipping time is longer, suggesting slower arrivals. The interquartile range of Standard Class is the widest indicating the greatest degree of uncertainty in shipment timeframes. It is typically the slowest option as seen by its highest median shipment time. There are outliers that show certain occurrences of exceptionally fast and sluggish deliveries both above and below the box.

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

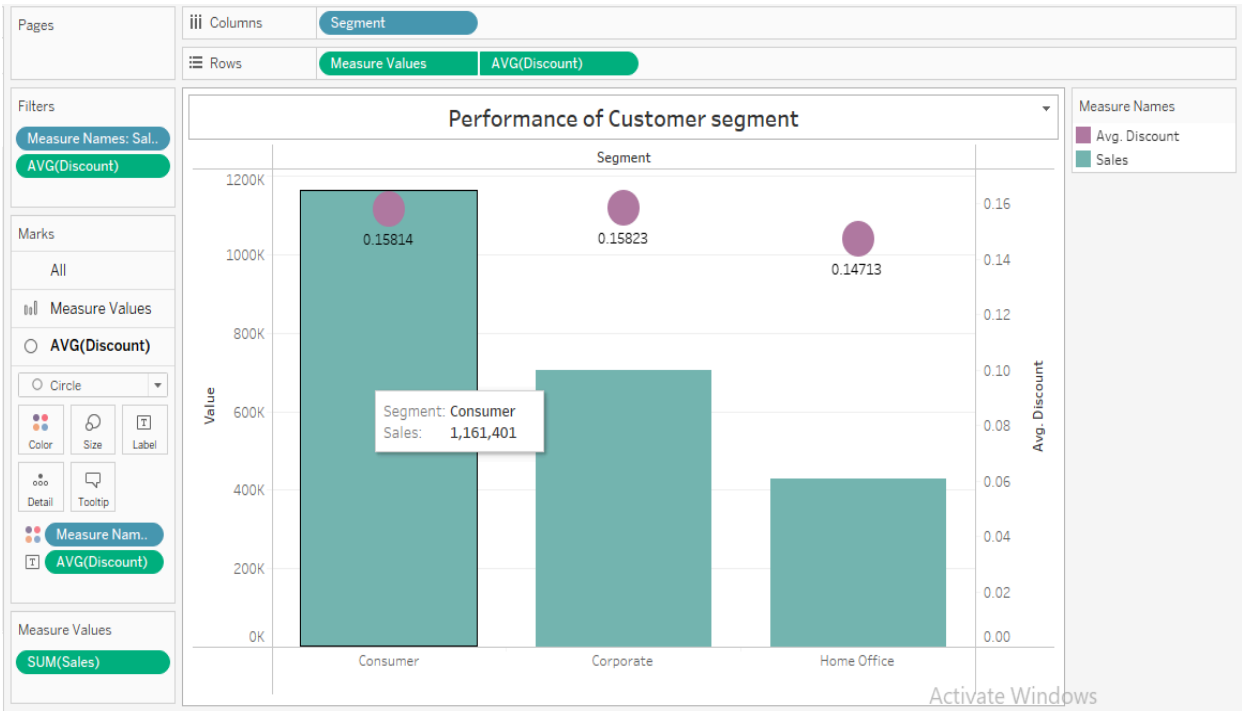
Ans:



Since it shows how the order volume changes over time, the line chart is a great option for viewing the monthly trend in the number of orders shipped. In this instance a clear trend in order shipments from January to March is shown on a monthly basis by this line chart indicating an increase in orders with the start of the new year. There is a noticeable decline in February possibly because the month is shorter or because of seasonal variations in the market. June is the peak, and the summer months that follow have a decline which could be explained by seasonal patterns. Due to holiday shopping, the chart peaks in November and then modestly declines in December.

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

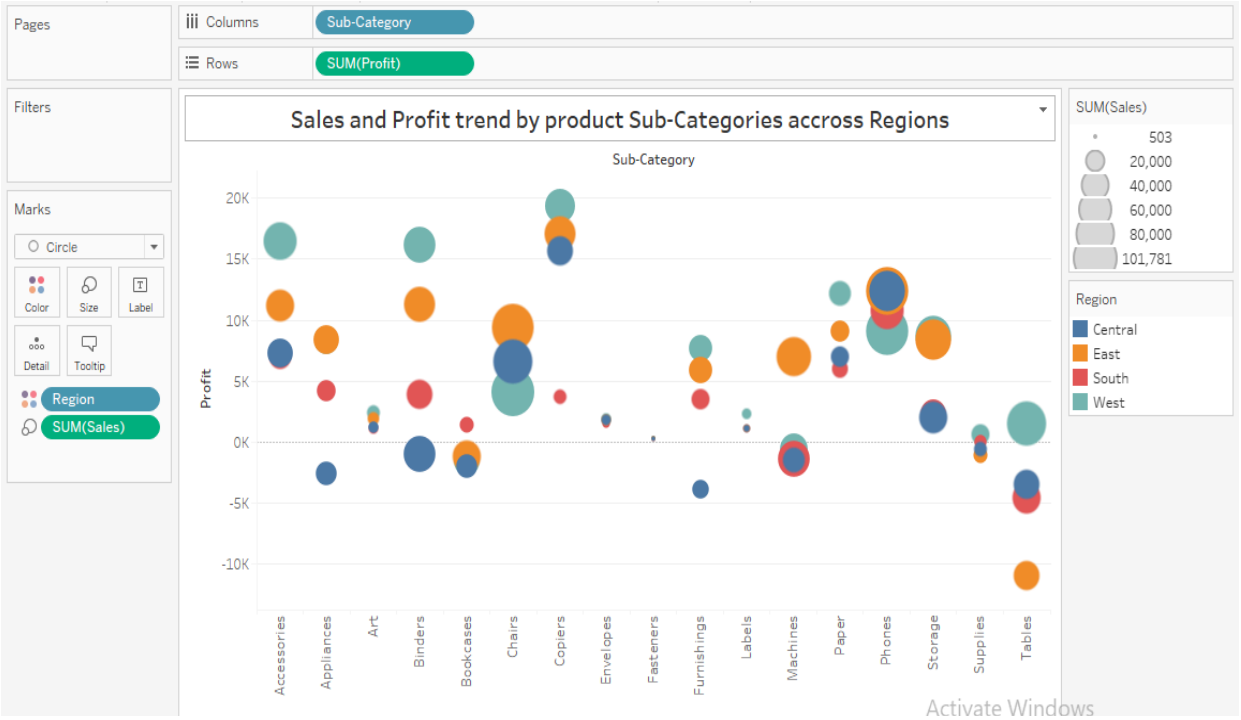
Ans:



Due to its ability to display two linked metrics on one graph, the dual-axis bar chart is a useful tool for comparing and analyzing the sales and discount rates of various consumer segments. The main bar charts show the total sales making it simple to compare the absolute sales figures across customer segments. The circles that are superimposed on the sales figures represent the average discount rates which put the sales figures in context by illustrating the amount of discount that corresponds to each segment's sales. The Home Office segment is an exception to this pattern retaining lower discount rates despite lower sales volumes. Overall the graphic indicates a correlation between sales volumes and discount rates with higher sales volumes being associated with higher discounts.

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

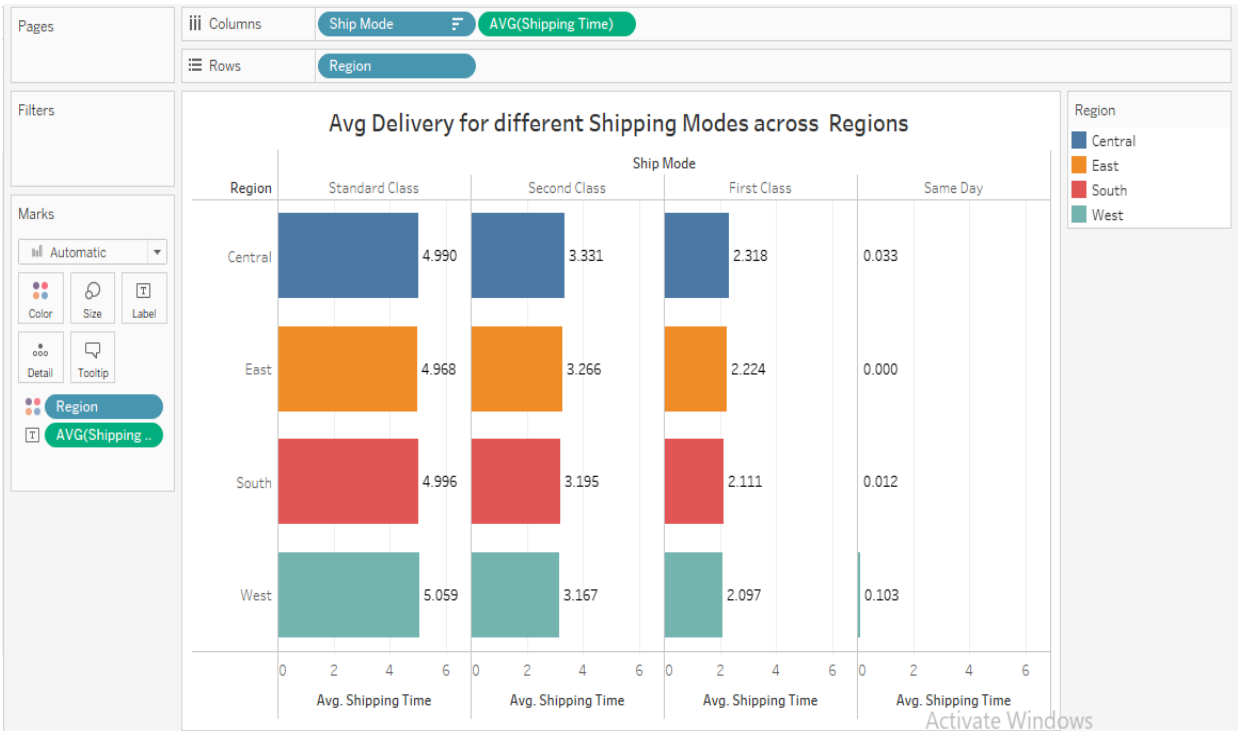
Ans:



The scatter plot, which enables multidimensional analysis inside a single visualization, is a great option for illustrating sales and profit patterns across several product subcategories and geographical areas. The size of the bubbles in this chart indicates the amount of sales, and their placement on the y-axis gives quick insight into the profitability of each subcategory within an area by clearly differentiating profitable from unprofitable goods. It is simple to separate and compare regional performance within the same subcategory thanks to the region-specific color coding. The scatter distribution illustrates the variation in performance between regions, indicating potential areas for strategy modification. This figure demonstrates that, although phone sales are strong in all regions, carrier profits are high even in the face of decreased sales volumes showing larger margins. Losses are consistently displayed in all regions by tables suggesting possible problems with expenses or pricing.

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

Ans:

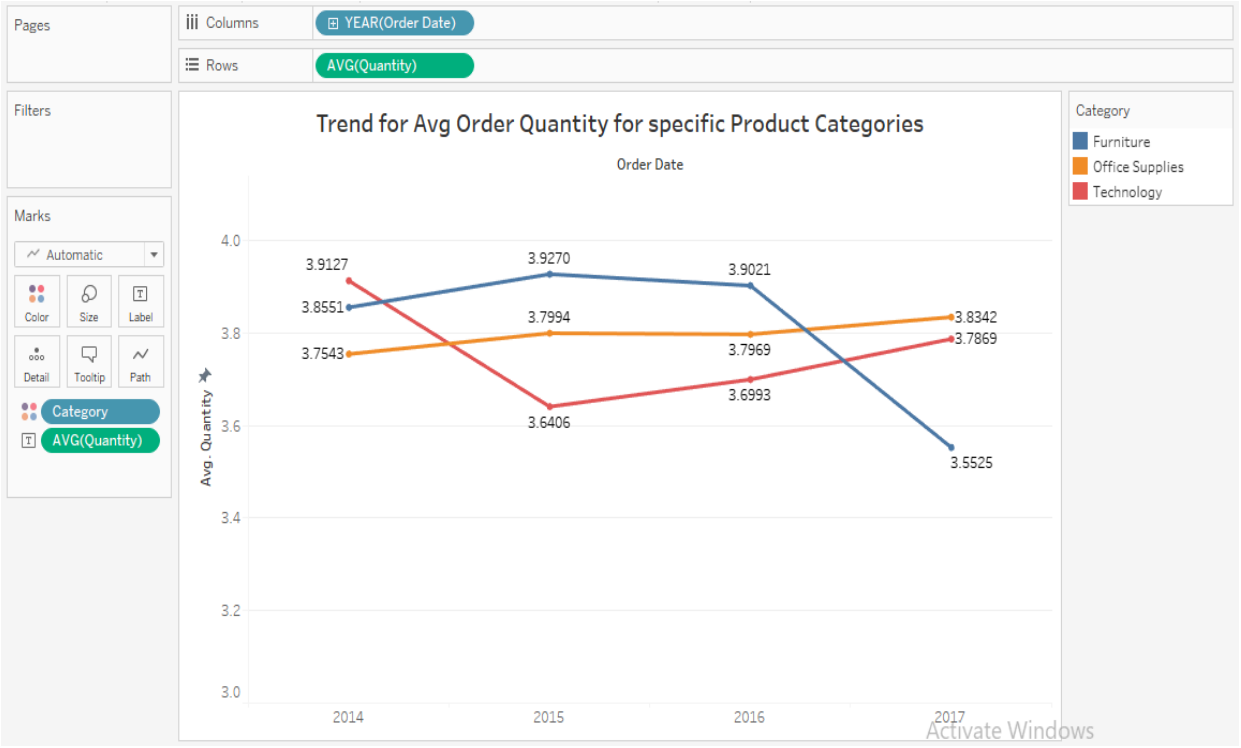


The average delivery time by location and ship mode is effectively visualized by the horizontal bar chart with divided view by color intensity, allowing for quick pattern recognition and comparative analysis. This format is an effective option for presenting delivery metrics in a thorough and easily comprehensible way since it offers a clear and instantaneous understanding of the data.

It indicates that the same day has the most efficient way of shipping , followed by first class with an average between 2.09 to 2.31, then comes second class with an average between 3.16 to 3.33, and the slowest being the standard one. If we bifurcate it region wise it has been seen that for the central region generally taking more time to ship items and the east region being the least when we took the average of all.

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

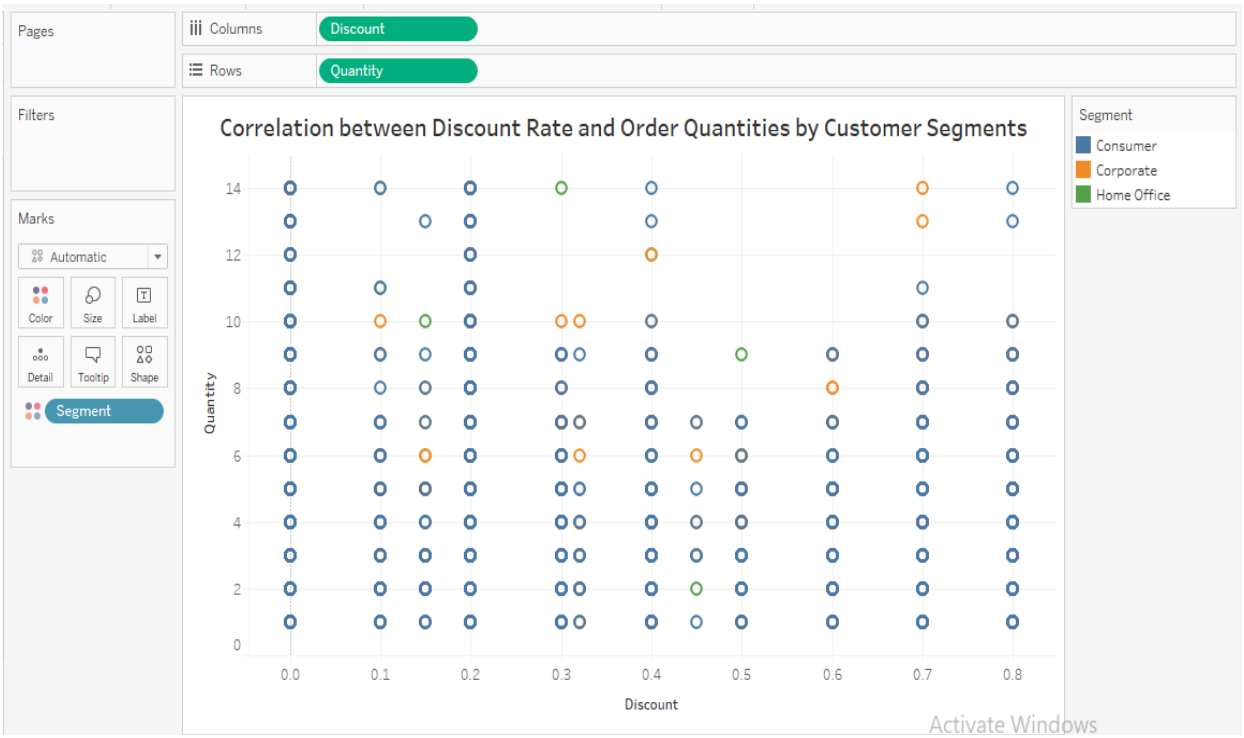
Ans:



The line chart is a good option for illustrating the evolution of the average order quantity over time since it makes comparisons simple and exhibits patterns clearly throughout several product categories. It is easy to see rises, reductions, or consistent patterns in the data because of its capacity to represent continuous quantitative data throughout a time series. According to this line graph between 2014 and 2017 the average order quantity for furniture decreased, reaching a peak in 2015; the average order quantity for office supplies increased slightly overall remaining relatively stable; and the average order quantity for technology showed volatility declining sharply in 2015 before partially recovering in 2017.

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

Ans:

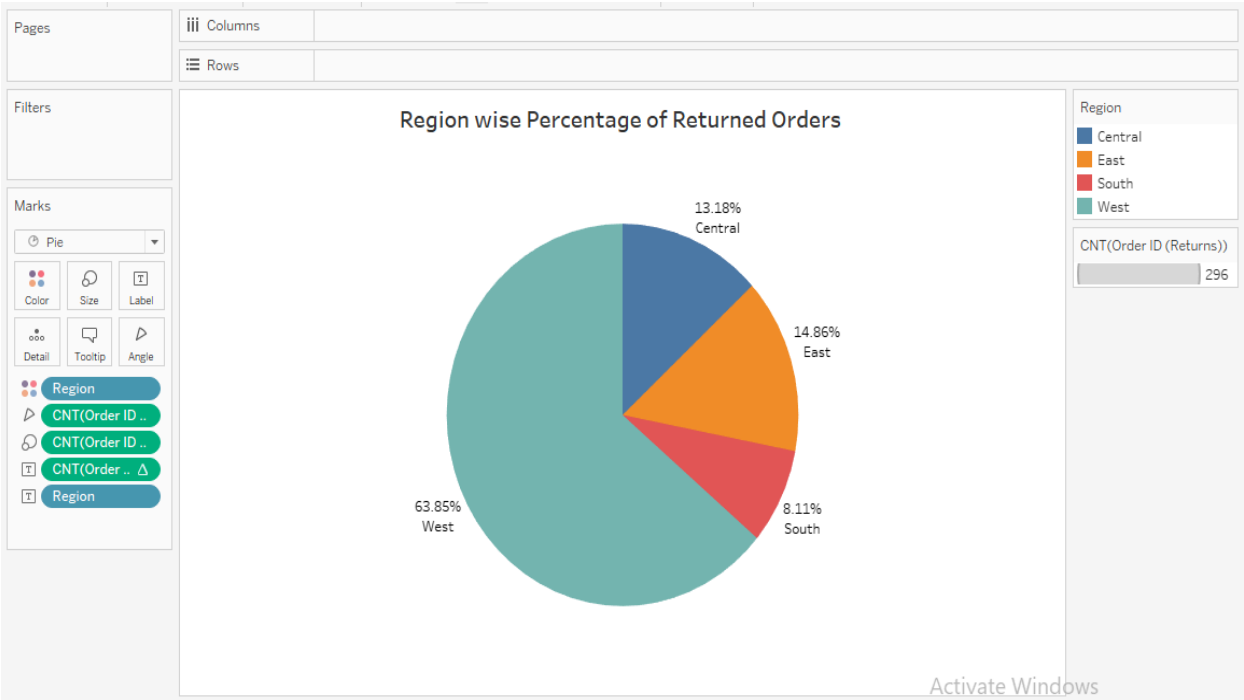


A scatter plot is a great tool for illustrating how different consumer segments' discount rates and order amounts could be correlated. It enables the plotting of individual data points to display the distribution and correlation between the two variables. Because each client segment is represented by a different hue it is simple to identify them and spot trends or behaviors unique to that segment.

The consumer segment shows the strongest trend indicating their higher sensitivity to discounts. The corporate segment shows a moderate correlation suggesting a balanced responsiveness to discount changes. In contrast, the home office segment exhibits the weakest correlation indicating a lower sensitivity to discounts and possibly a focus on other factors when making purchase decisions. The positive correlation across all customer segments suggests that higher discounts tend to incentivize larger orders.

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

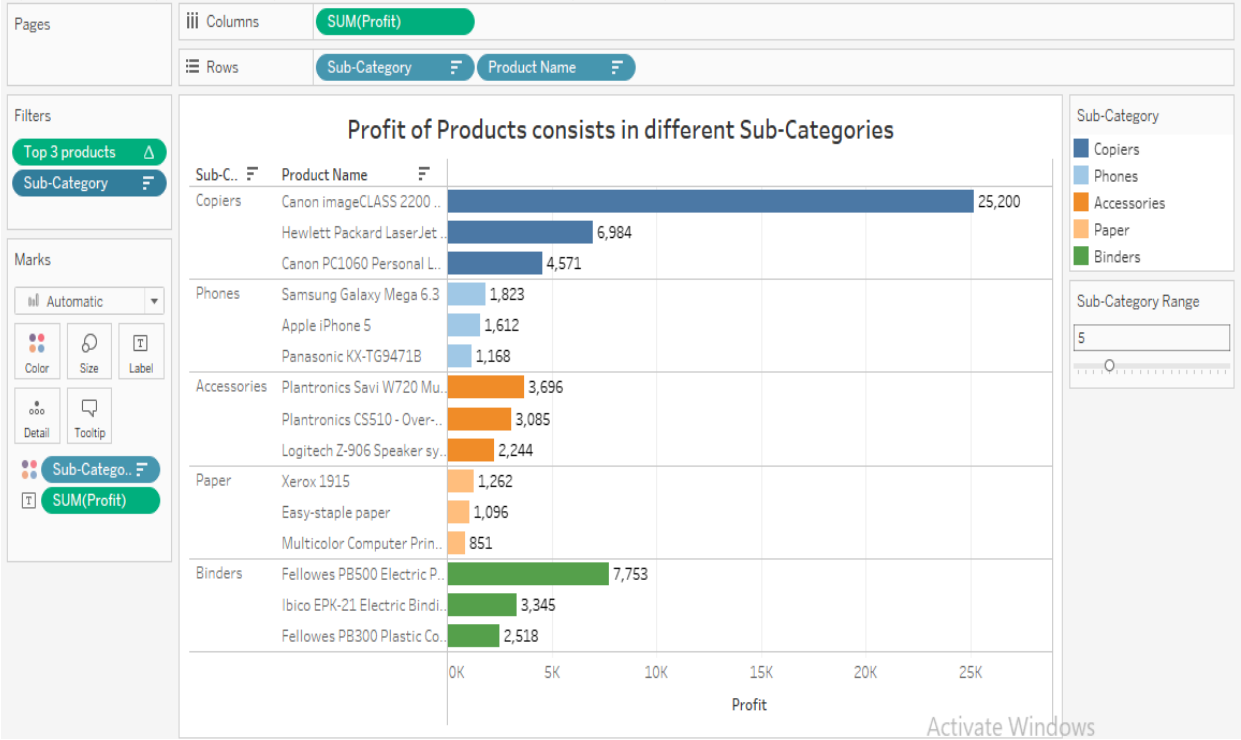
Ans:



Pie charts are a useful visualization tool for showing the percentage of returned orders by area in the Superstore dataset because they provide an intuitive way to see parts of a whole. The chart is easy to read and understand because of this structure which is especially helpful when there are few categories as is the case with the regions in this example. The relative magnitude of returns for each location is immediately conveyed by the size of each slice indicating which regions have higher or lower return rates. The West region has the largest percentage of returns 63.85%, followed by the East (14.86%), the Central region (13.18%), and the South (8.11%) with the lowest share of all returned orders according to the pie chart created using this information.

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

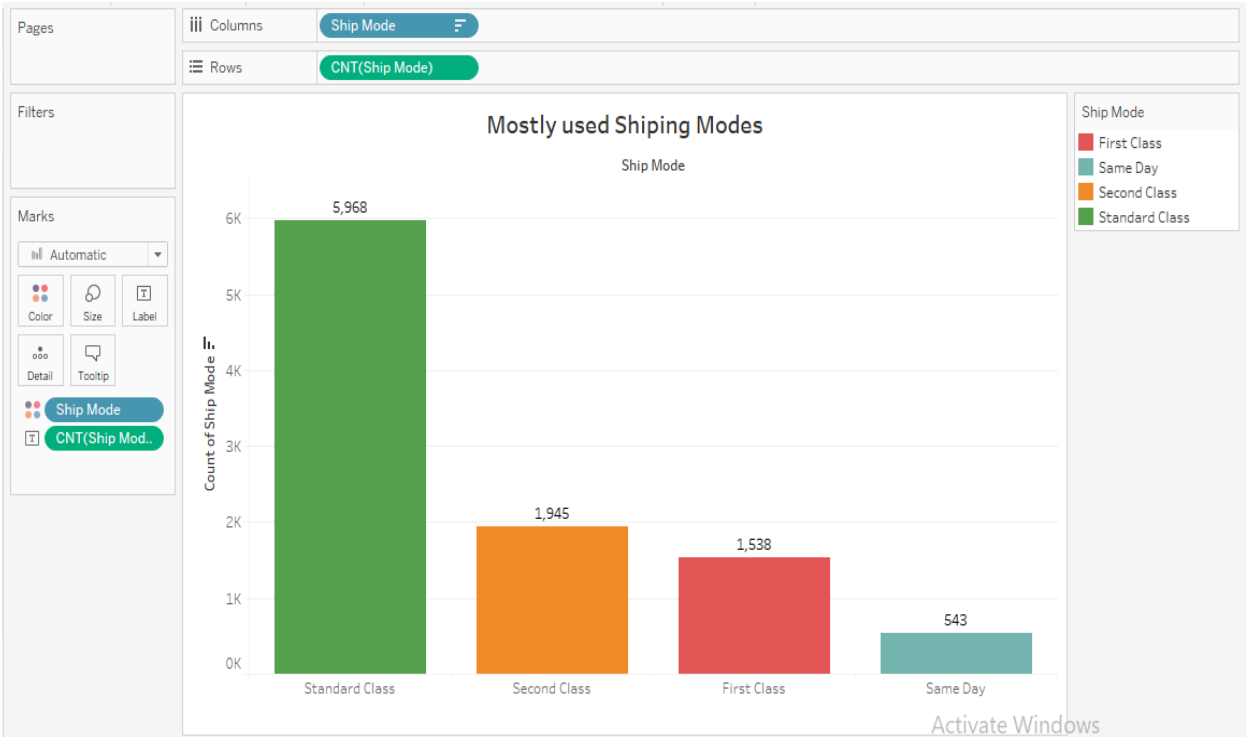
Ans:



Because of its logical and well-organized presentation, the horizontal bar chart is a great option for comparing product earnings across subcategories. Because the bar lengths naturally convey profit values, this sort of chart efficiently displays rankings and facilitates rapid visual evaluation. Longer names are easier to read thanks to the horizontal layout's ability to incorporate text labels without becoming crowded. We can adjust the view to focus on more or fewer groupings, increasing flexibility, by using a parameter to alter the number of visible subcategories. Furthermore, preventing congestion and emphasizing the greatest profit-generating goods is achieved by filtering and displaying only the top three products using a computed field.

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

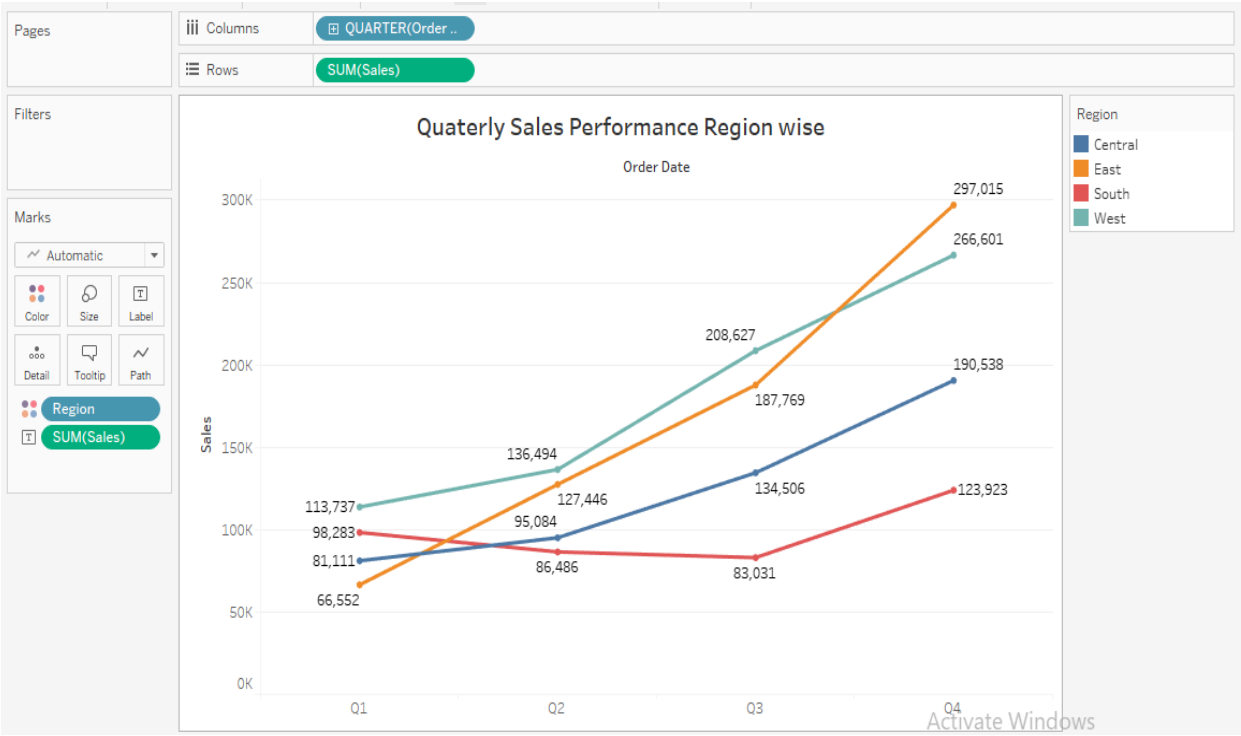
Ans:



The dataset's most popular shipping options are graphically represented in a vertical bar chart that arranges them in descending order of use. The space used up by each shipment mode reflects how prevalent it is; 'Standard Class' is the most visible suggesting it is the largest shipping mode; 'Second Class', 'First Class', and 'Same Day' shipping modes on the other hand indicate they are less common. Without requiring numerical interpretation this visual method is especially useful for emphasizing the distribution of categories at a glance. The huge 'Standard Class' bar's instant visual effect conveys the category's dominance in shipping mode preference.

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

Ans:

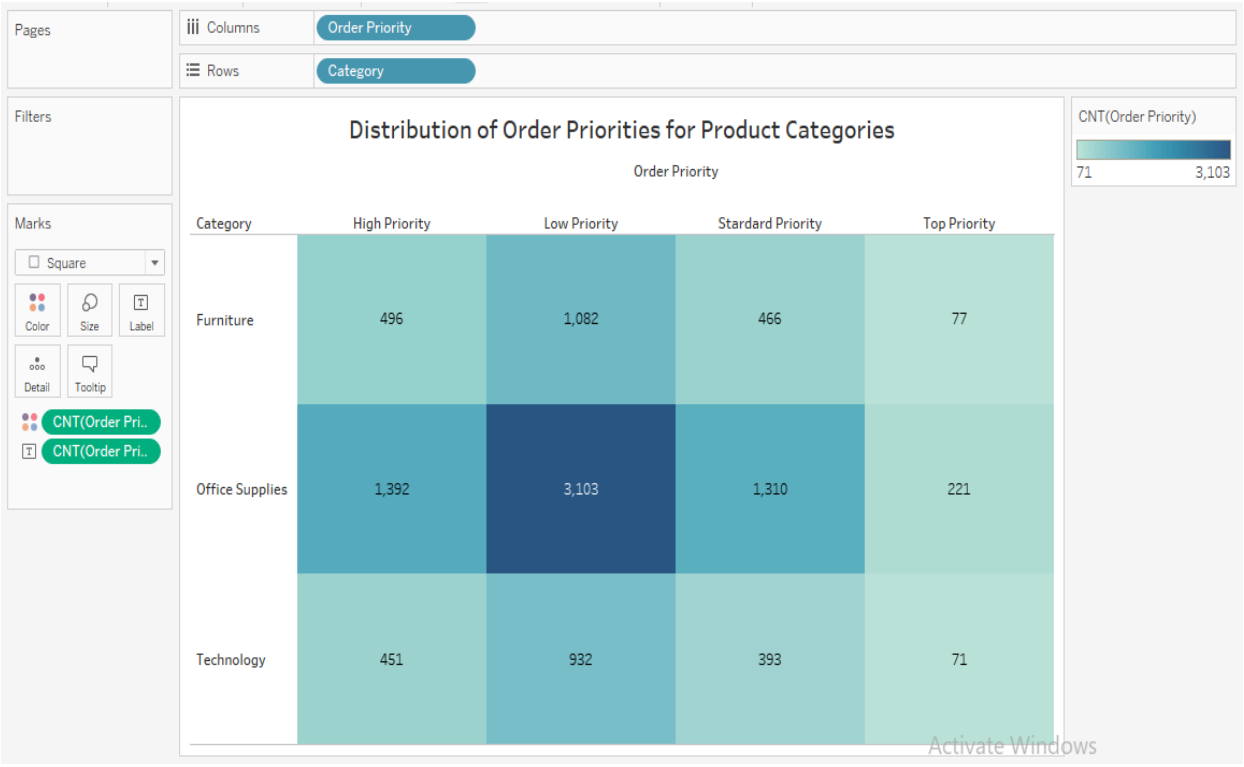


Because it can clearly show patterns over time the line chart is a useful tool for displaying sales performance across different areas over the course of a year's quarters.

The East region's sales increased steadily throughout the year reaching their highest point in the fourth quarter both the South and the Central regions saw growth however by the fourth quarter the Central region had slightly outperformed the South leaving the South with the lowest sales position for the year, the West region which had the highest sales at the beginning of the year continued its upward trend and finished just behind the East.

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

Ans:

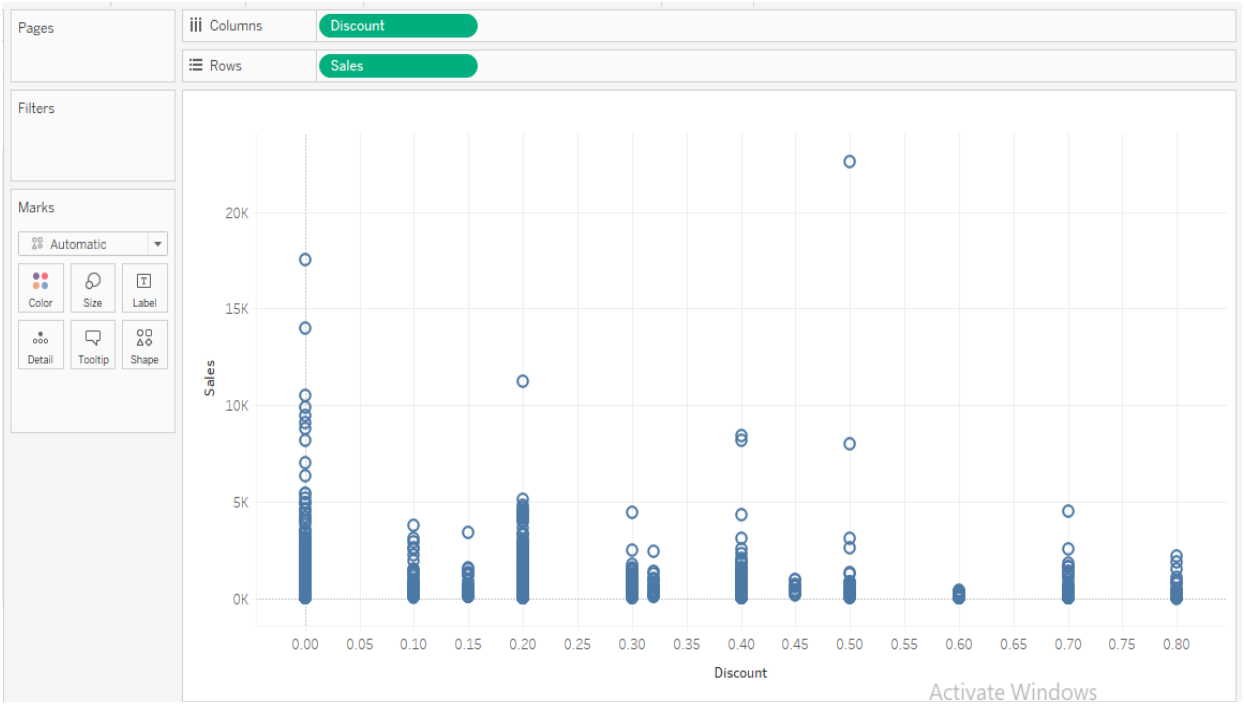


The highlight table chart combines the visual strength of color coding with the simplicity of a table making it a useful option for visualizing the distribution of order priorities across various product categories. This format makes it possible to quickly see trends and make comparisons across categories and priority since richer color tones instantly identify higher numbers.

The 'Order Priority' computed field reclassifies the delivery modes according to a hierarchy of urgency: 'First Class' is classified as 'High Priority', 'Second Class' as 'Standard Priority', 'Same Day' as 'Top Priority', and 'Low Priority' for all other shipping modes. This graph shows that, across all product categories, 'Low Priority' is the most often chosen delivery option. Office supplies are the category with the most orders, especially under 'Low Priority'. Customers in all categories least like the 'Standard Priority' and 'Top Priority' selections, indicating a decreased need for faster shipping. The category with the fewest 'Top Priority' orders is technology, suggesting that there isn't much of a requirement for quick delivery in this market. Furniture, on the other hand, follows the general trend and prefers 'Low Priority', but it has fewer orders than office supplies.

28. What is the relationship between discounts and sales?

Ans:

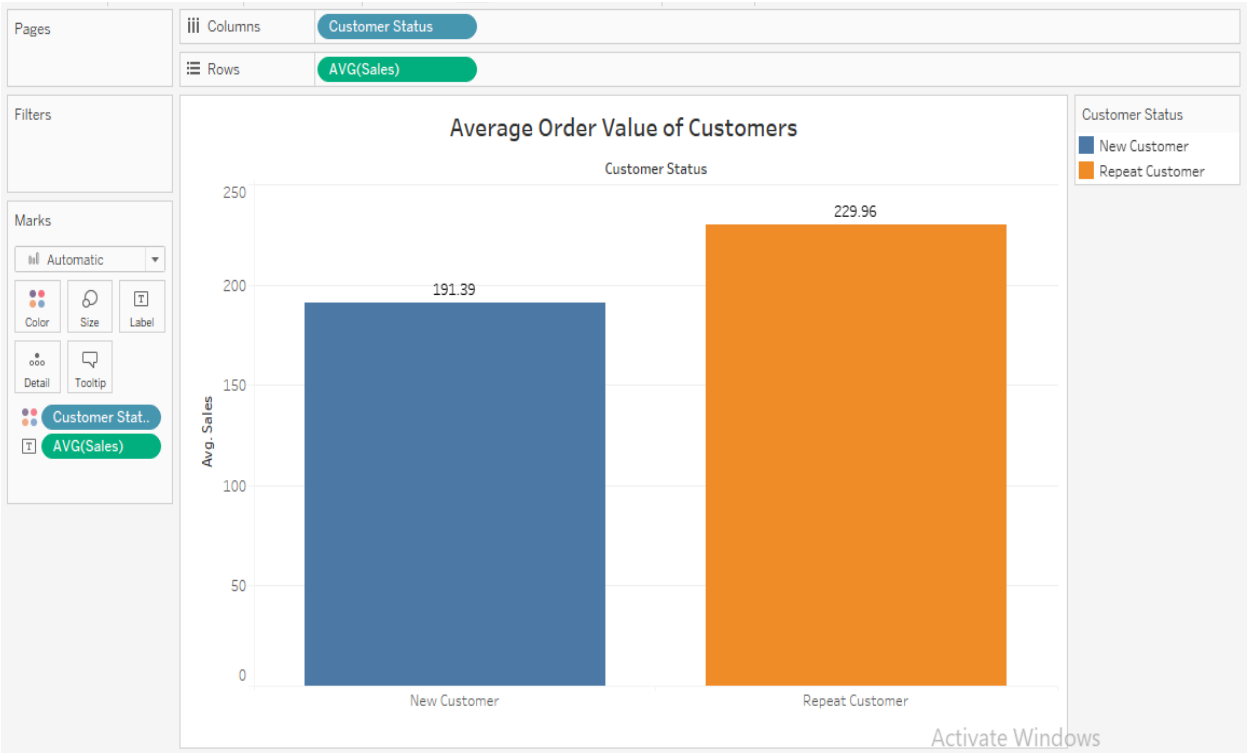


Because it can accurately and easily show the relationship between two continuous variables, such sales and discounts, a scatter plot is the preferred method. It is excellent in displaying the distribution and density of data points as well as finding patterns, outliers, and relationships.

The scatter plot suggests that there isn't much of a relationship between the overall sales and the amount of the discount. Sales are larger at lower discount levels, but there are also some sizable sales at about 0.5 discount level, indicating that even a tiny discount can have a positive impact on sales. Maybe you don't need to provide a big discount to boost sales.

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

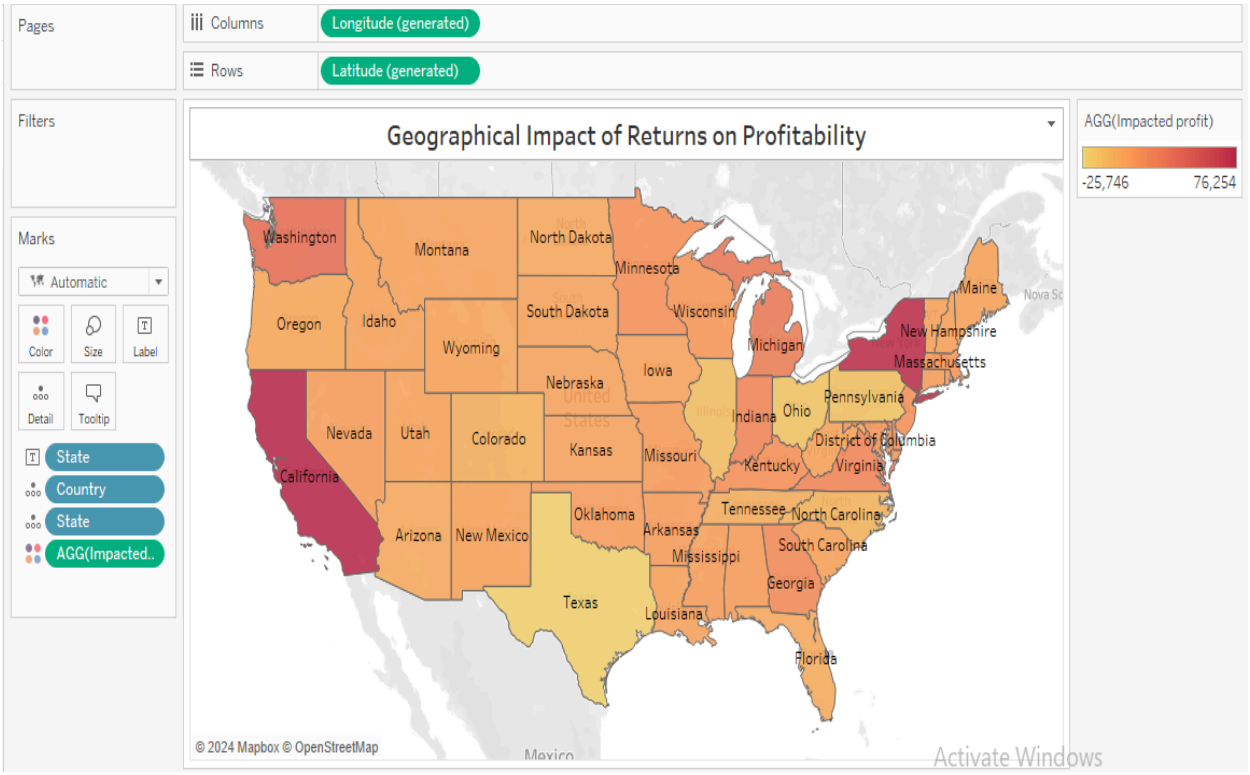
Ans:



Because it presents data in a quantitative and categorical manner, the bar chart is a great option for illustrating the comparison between the average order values of new and returning clients. 'New Customers' against 'Repeat Customers' is a good example of how bar charts work well to compare discrete groups and provide a fast visual assessment of differences. The 'Customer Status' calculated field is designed to classify a customer as a 'New Customer' if their order date is the same as the first order that has been logged for that specific customer ID. Orders that follow from the same customer ID are then labeled as 'Repeat Customers'. It indicates that new customers have a lower average order value ($191.39) compared to repeat customers ($229.96), suggesting initial purchases are more substantial. This could imply that first-time buyers might be incentivized by marketing strategies or introductory offers.

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

Ans:



Because it provides an instantly recognizable spatial context, a map chart is perfect for illustrating the regional distribution of returns and their impact on profitability. Stakeholders may easily see how geography relates to financial performance and which regions are most impacted by returns. Color gradients make it easier to quickly identify areas that require additional attention by indicating different levels of profit impact. The field labeled 'Return Impact on Profit' highlights the precise impact of returns by separating profit from returned products. With the exception of New York and Virginia, the map shows that return impacts are larger in the northern and western regions of the United States and lower in the southern and eastern regions.

Created by: Vishal Sinha Roy

Email: vishalsingha.r05@gmail.com