Data Visualization and Analysis Project

About the Project:

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

Skills Gained/Required:

- Data Visualization Proficiency: Mastery in understanding and applying data visualization concepts and techniques to represent data effectively and meaningfully.
- Tableau (or Similar Tool) Familiarity: Ability to utilize Tableau or similar datavisualization tools proficiently to create insightful visualizations.
- Analytical and Problem-Solving Skills: Strong capability in analyzing data to derive insights and solve complex problems, ensuring the visualizations address key questions effectively.
- Chart Selection Expertise: Skill in selecting appropriate charts based on data characteristics and specific requirements, enhancing the clarity and impact of the visual representation.
- Communication Skills: Clear and concise communication abilities to articulate the insights derived from visualizations, making complex data easily understandable to a diverse audience.

Dataset Link:

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

The first step in this project is the preparation and exploration of the data preceding the analysis of 30 questions that have been provided. My goal is to make sure the cleanlinessof the dataset is properly executed to avoid discrepancies and/or non-inclusion of someinformation which may result in incorrect interpretation.

Questions:

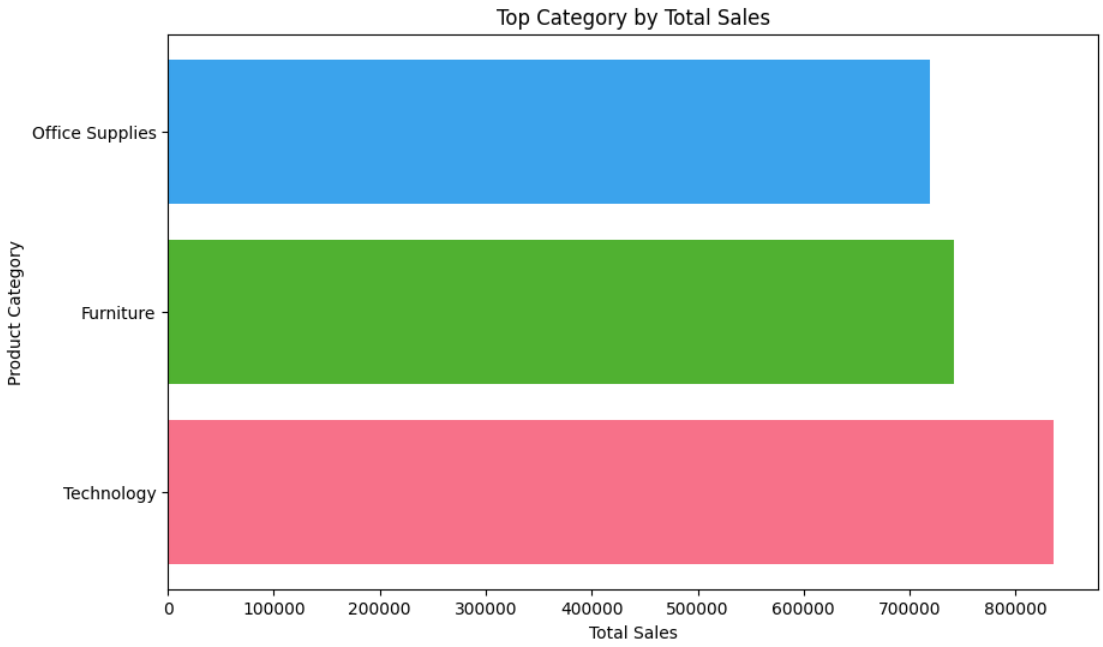
1. **Which product categories have the highest total sales in the "Superstore" dataset?**

Ans-

Why did I choose this chart?

The choice of a horizontal bar chart is based on its suitability for comparing categorical data, especially when the goal is to highlight differences in values, such as total sales across different product categories.

- Comparison of Categories
- Ordered Representation

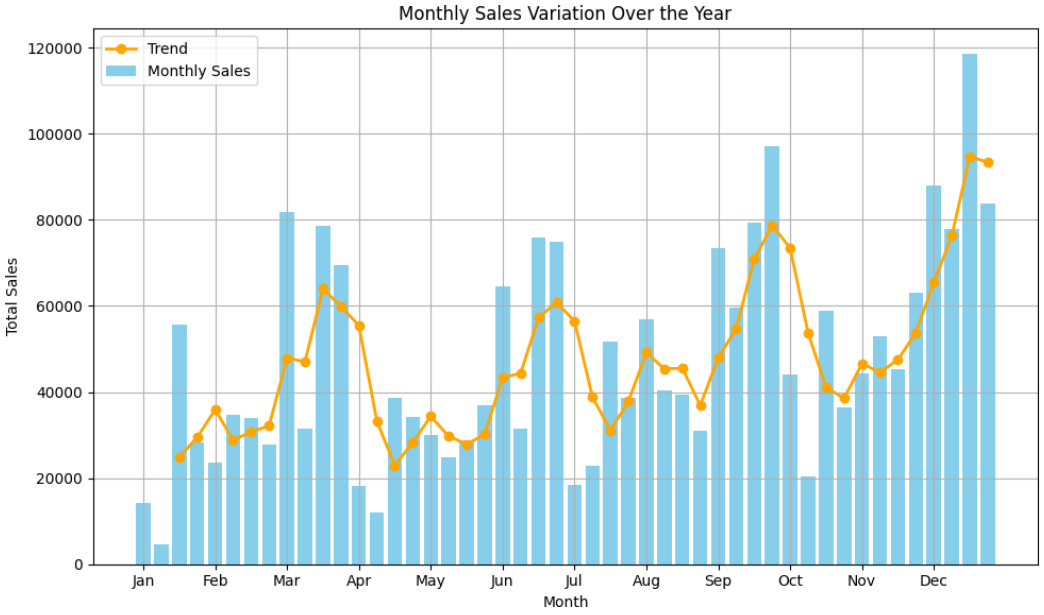


2. **How do the monthly sales amounts change over the course of a year?**
   Ans –

   Why did I choose this chart?

   The combined bar chart with a connecting trend line was chosen for its ability to provide a comprehensive view of the monthly sales variations over the course of a year.

   - Visualizing Trends
   - Ease of Interpretation

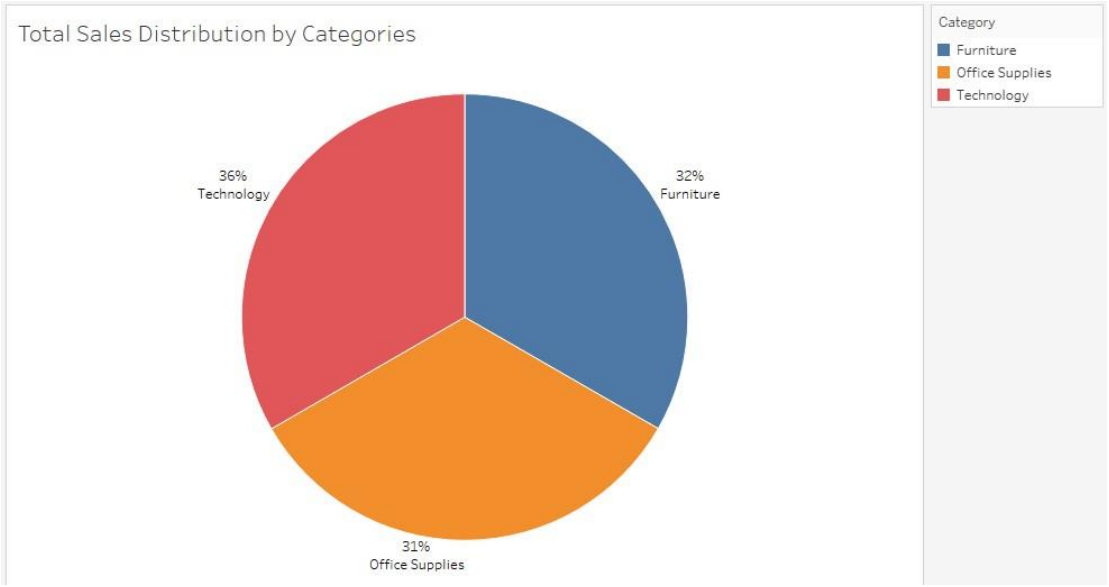


3. **How is the total sales amount distributed among different product categories?**
   Ans-

   Why did I choose this chart?

   To visualize the distribution of total sales among different product categories, a suitable chart choice would be a pie chart. A pie chart effectively displays the contribution of each category to the total sales, allowing for easy comparison of proportions.

   - Comparative Representation
   - Clear Division of Parts



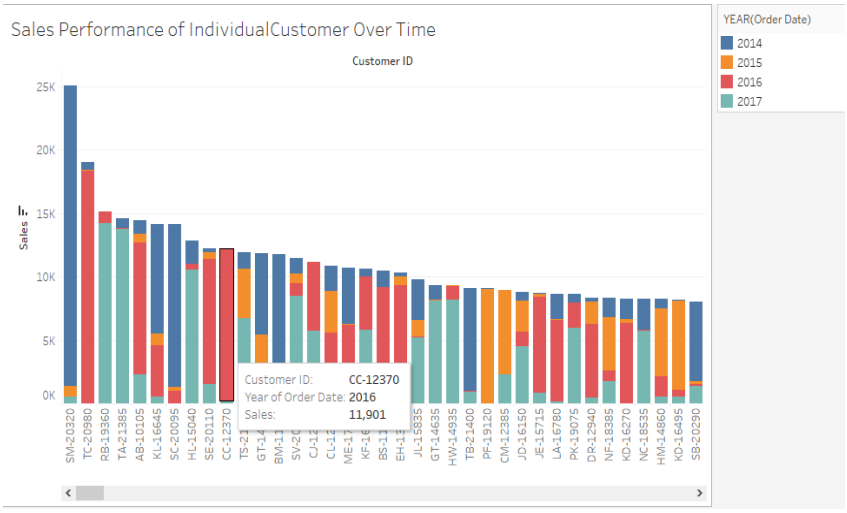
4. **Can we analyze the sales performance of individual customers over time?**

Ans –

Why did I choose this chart?

I choose this chart because it offers a comprehensive and clear visualization of the sales performance. It enables me to

- Analyze sales performance for each year
- Compare the sales performance of individual customers across different years
- Identify trends and patterns in the sales performance of customers over time



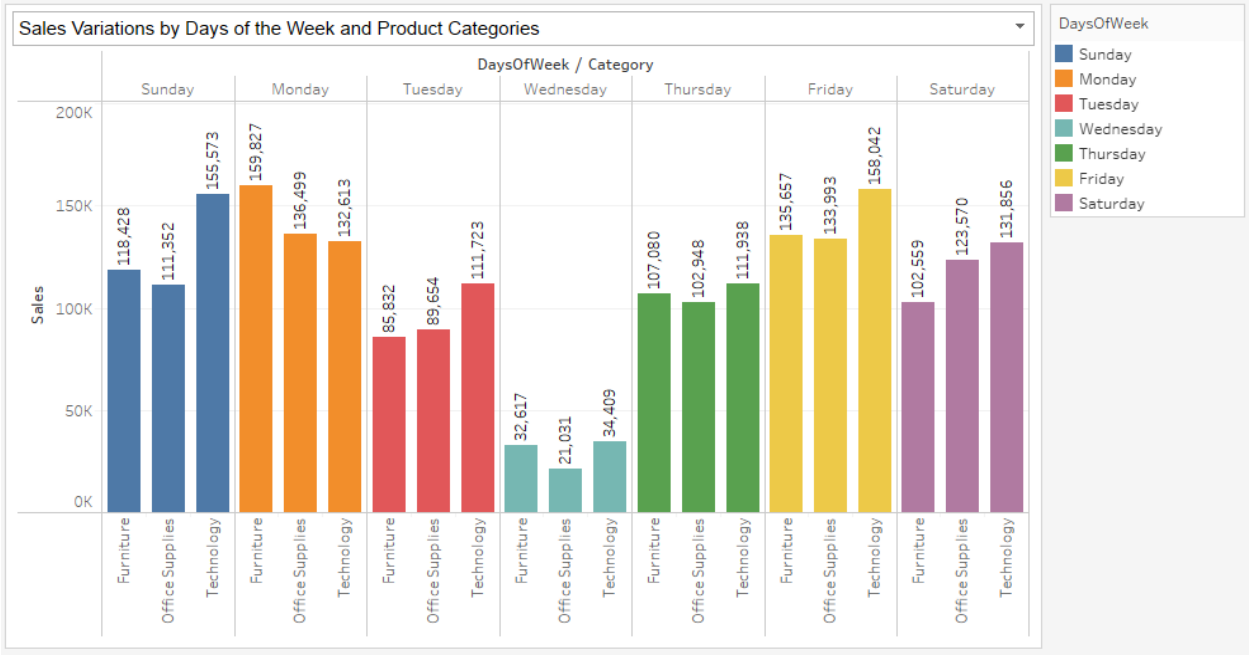
5. **How do sales vary based on different days of the week and product categories?**
   Ans –

   Why did I choose this chart?

   Why the Clustered Bar Chart is Suitable

   It allows for a direct comparison of sales among different product categories for each day of the week through clustered bars, making it straightforward to observe variations. Each day of the week is visually distinct, aiding in the comparison of sales patterns across different days.

   - Comparative View
   - Separate Representation



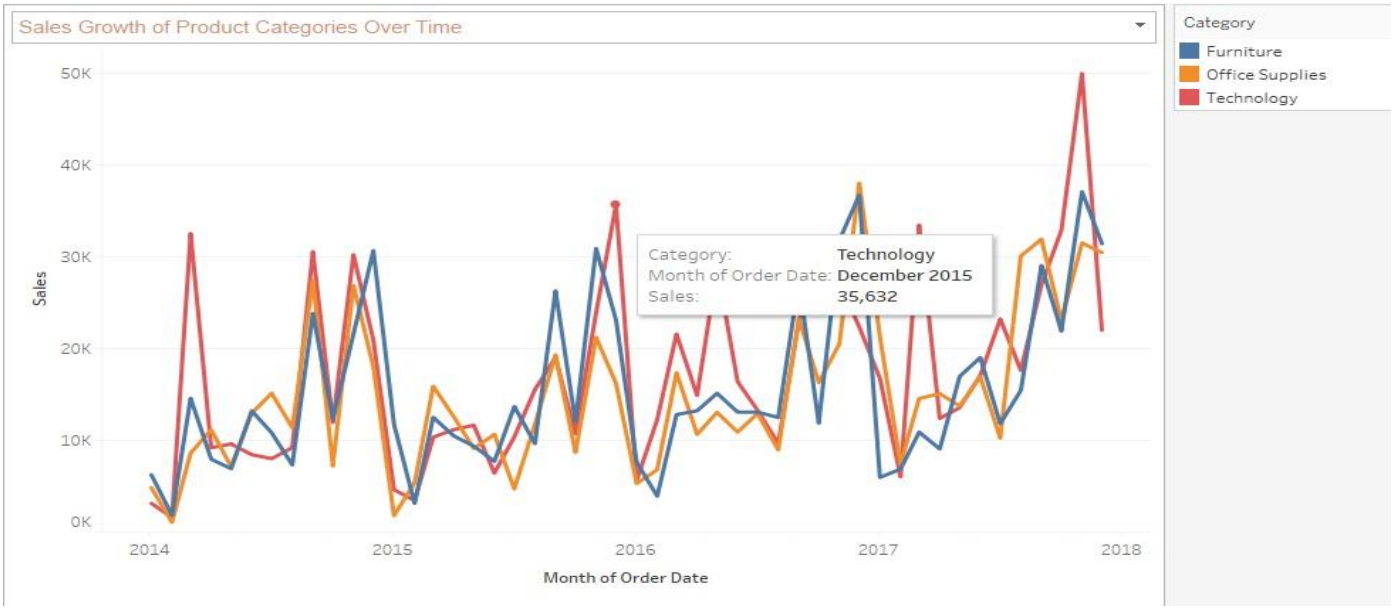
6. **Can we visualise the sales growth of different product categories over time?**
   Ans –

   Why did I choose this chart?

Why the Line Chart is Suitable

This line chart displays the sales growth of different product categories over time. Each line represents the sales growth trajectory for a specific product category

- Trend Visualization
- Comparative Analysis



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

Ans –

Why did I choose this chart?

For visualizing sales distribution across different regions a treemap could be a suitable choice. A treemap allows for the hierarchical display of data, showing the contribution of each region to the total sales while also representing sub-regions within each main region.

- Hierarchical Representation
- Size Encodings

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

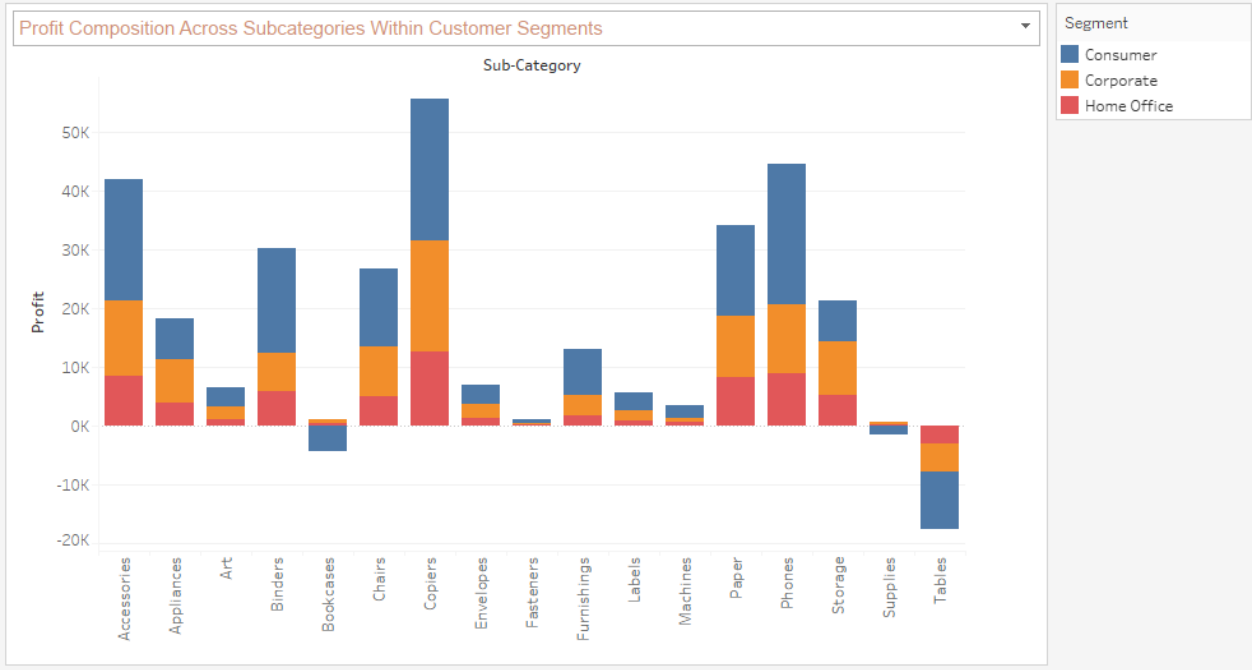
Ans –

Why did I choose this chart?

Why the Grouped Bar Chart is Suitable

Grouped bars enable a side-by-side comparison of profits across subcategories within different customer segments. It allows a clear comparison between different customer segments, showcasing the profits contributed by each subcategory.

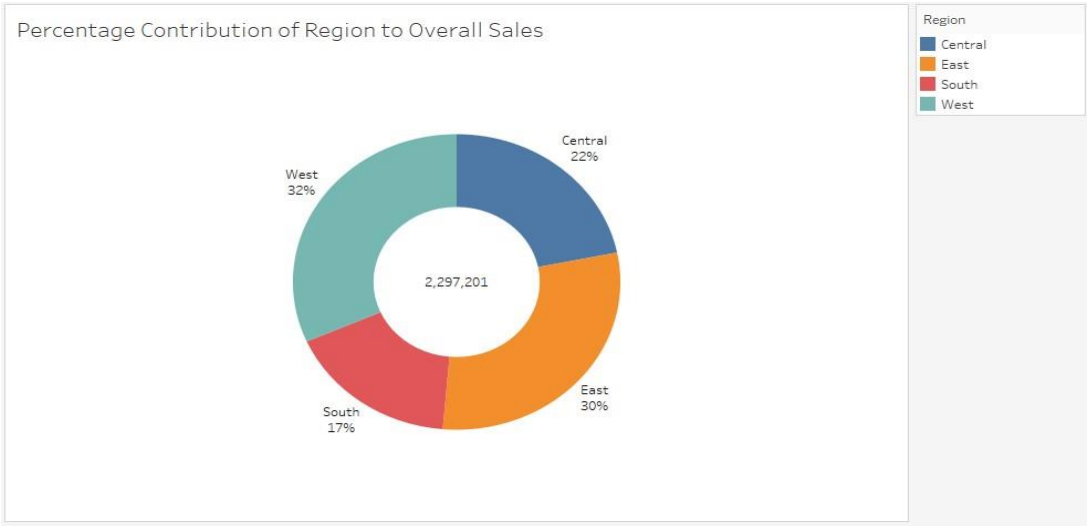
- Comparative View
- Segmented Comparisons



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

   Why did I choose this chart?

   A “Donut chart” can effectively represent the percentage contribution of each region to the overall sales because **donut charts** display data in a circular format, making it easy to visualize the proportion each region contributes to the total sales as segments of a whole.



- Enhanced Readability
- Clear Percentage Display

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

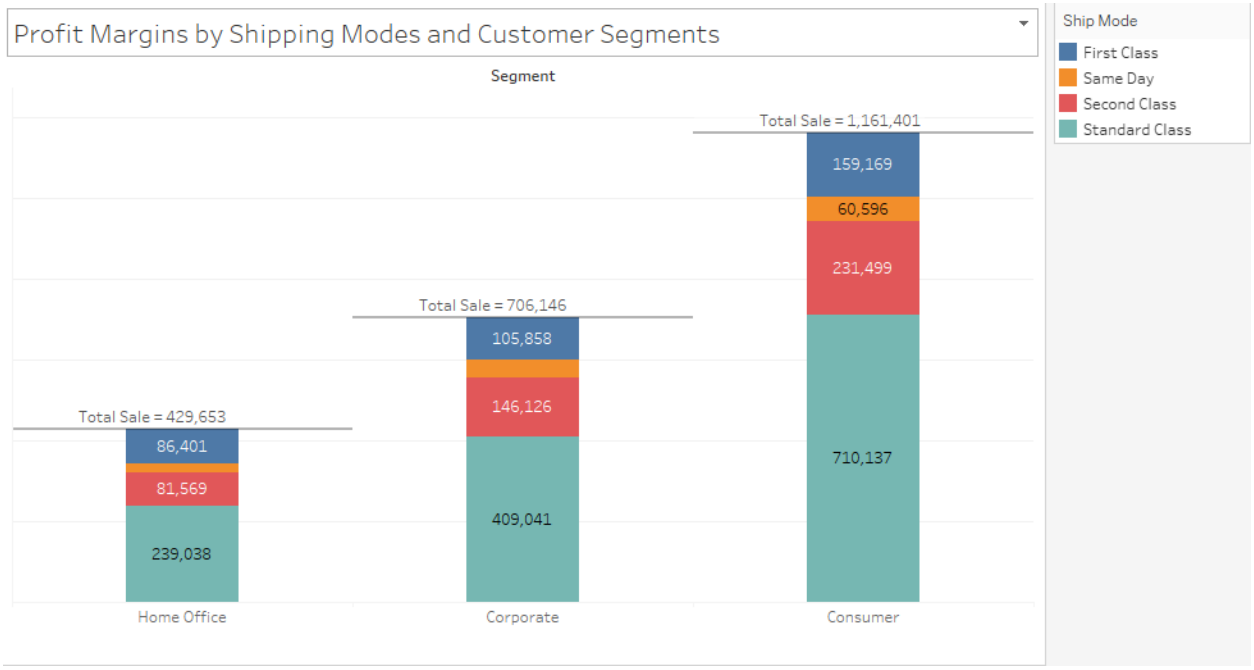
Ans –

Why did I choose this chart?

Why the Stacked Bar Chart is Suitable

Stacked bars represent the total profit margins for each customer segment, visually broken down by shipping modes. It showcases the contribution of each shipping mode to the overall profit margins within each customer segment.

- Composition Display
- Component Comparison



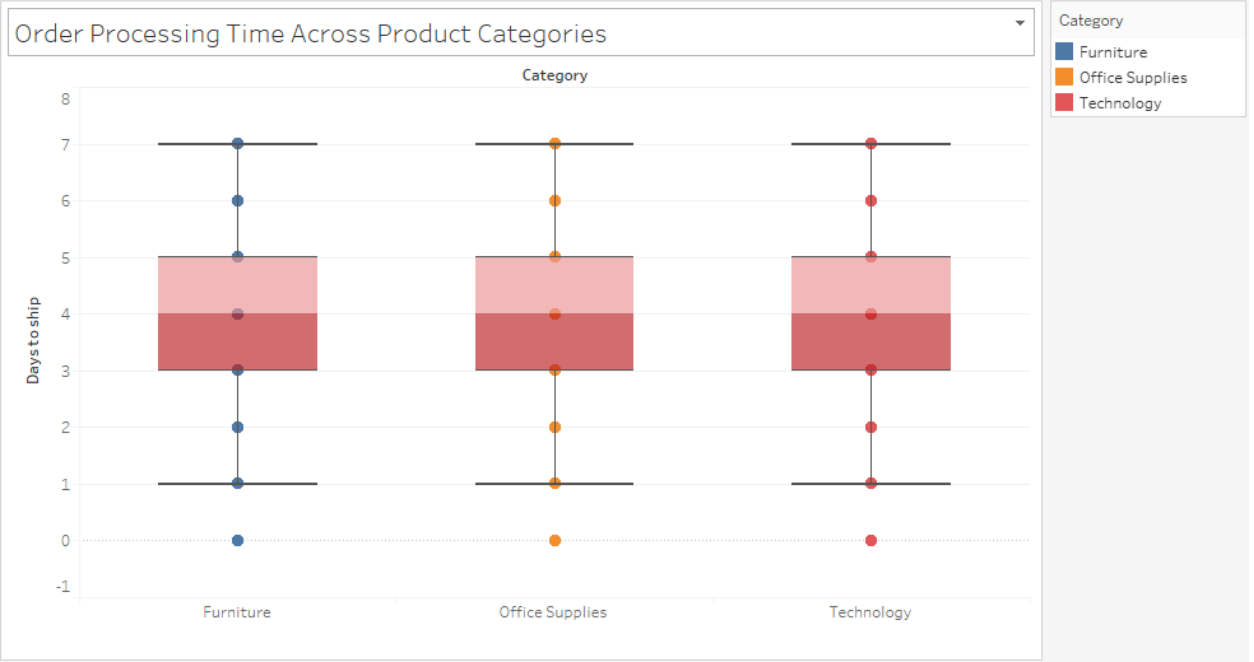
11. How long does it take to process orders for different product categories?

Ans –

Why did I choose this chart?

Why the Box Plot is Suitable: *Comparative View*, *Visualizing Spread*

Box plots show the distribution of processing times for each product category, allowing for easy comparison of median, quartiles, and potential outliers. They display the spread and variability of processing times across different product categories effectively.



12. How do discounts affect overall profit?

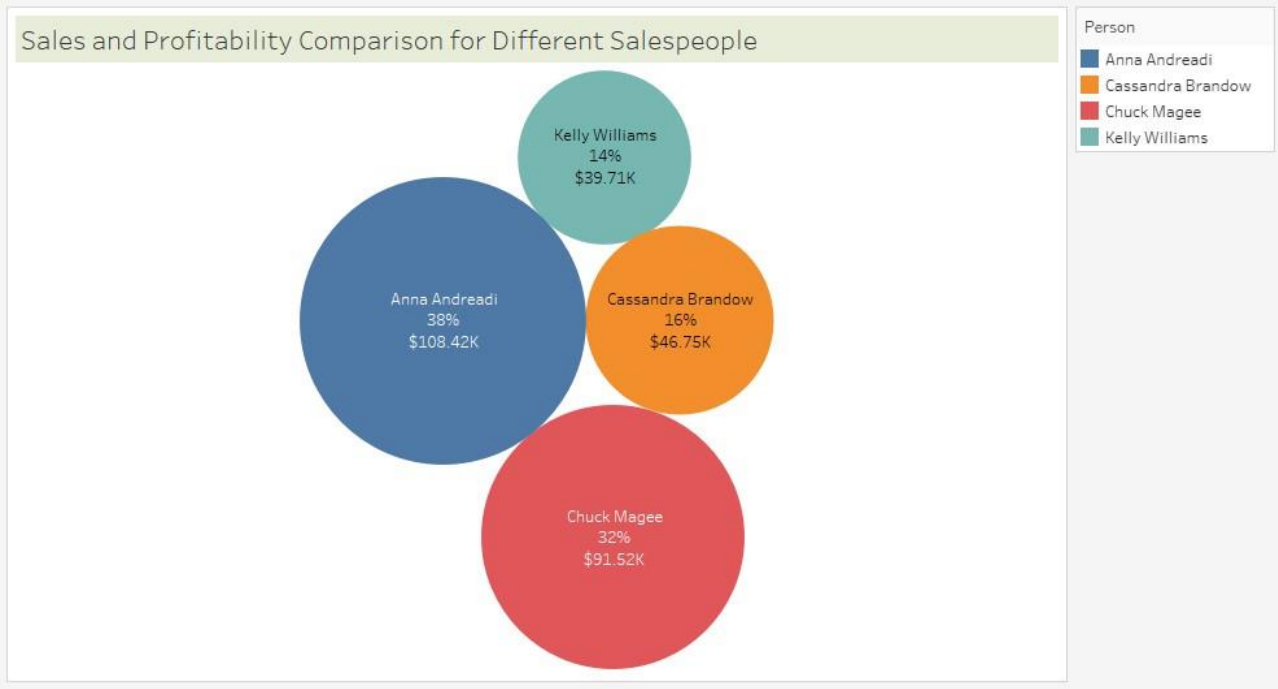
Ans –

Why did I choose this chart?

Packed Bubbles Chart: ***Comparison of Multiple Metrics, Individual Salesperson Analysis***

It allows for the simultaneous comparison of two metrics (such as sales and profitability) for each salesperson through bubble size and color. Each bubble represents a salesperson, showcasing their

performance in sales and profitability relative to others.



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

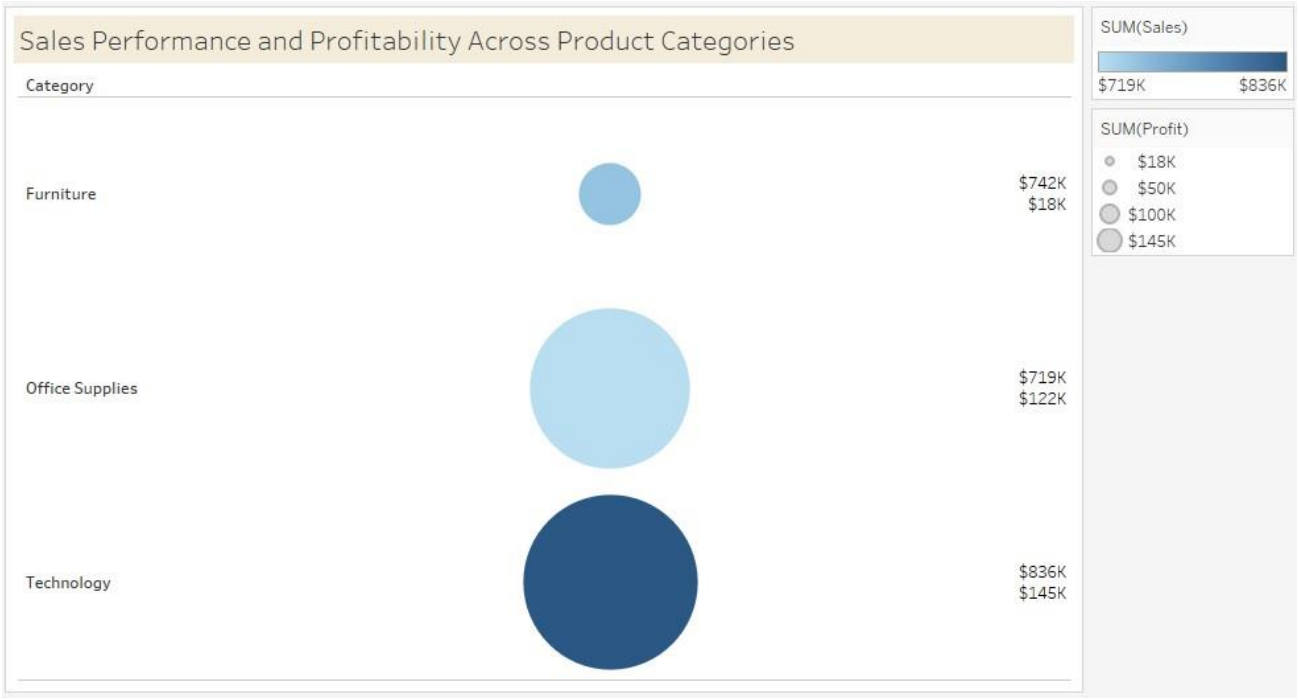
Ans –

Why did I choose this chart?

Customize map Chart Analysis:Product Sales vs. Profitability Across categories Insights from the chart:

Top Performers : the technology category stands out , boasting the highest sales ($836K) and Probability ($145K), Followed by furniture ($742K sales, $18K profit) and office supplies($719K sales,$122K profit).

Visual Comparison : These charts serve well for categorical comparisons, allowing a clear assessment of different category sizes. In this context, they vividly compare sales and profits across product categories.



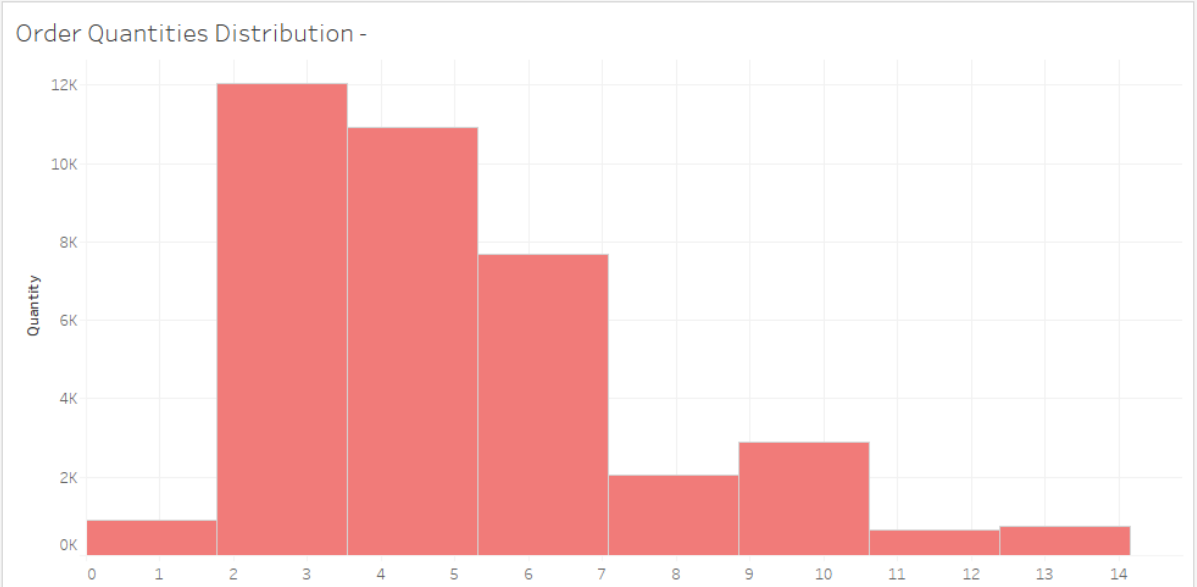
14. What is the distribution of order quantities for products in the dataset?

Ans –

Why did I choose this chart?

Why Histogram is Suitable: **Easy to understand** , **Scalable**

The distribution of order quantities for products in the dataset is skewed to the left, meaning that there are more small orders than large orders. The chart displays this type of data because it allows us to easily see the frequency of each order quantity.



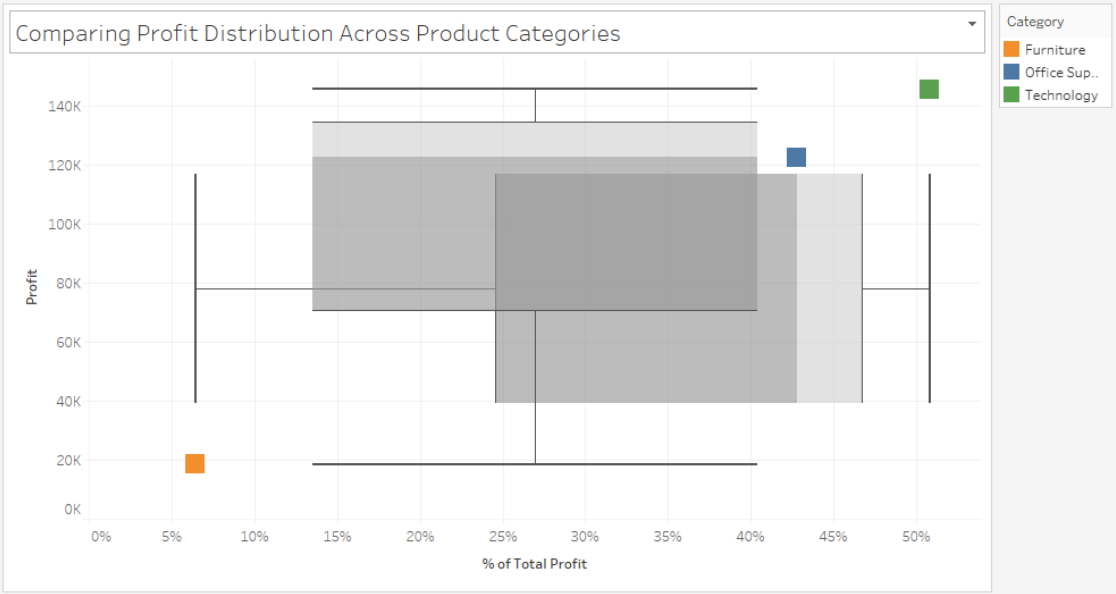
15. How do the profit distributions vary across different product categories?

Ans –

Why did I choose this chart?

A Double Box Plot Analysis displays the distribution of profits within each product category, showcasingthe range, median, quartiles. Enables easy comparison of profit distributions across multiple categories in a single visualization, highlighting variations and central tendencies.

- Display Distribution
- Comparative View
- Identification of Outliers



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

Ans –

Why did I choose this chart?

A Gantt Bar chart typically represents tasks or activities along a timeline.

To utilize the Gantt Bar chart for showing the distribution of shipment durations categorized as "Early," "On-Time," "Soon," and "Wait," I assigned durations to each category and plot them on a timeline

- Timeline Representation
- Comparative Visualization
- Utilization of Time Axis

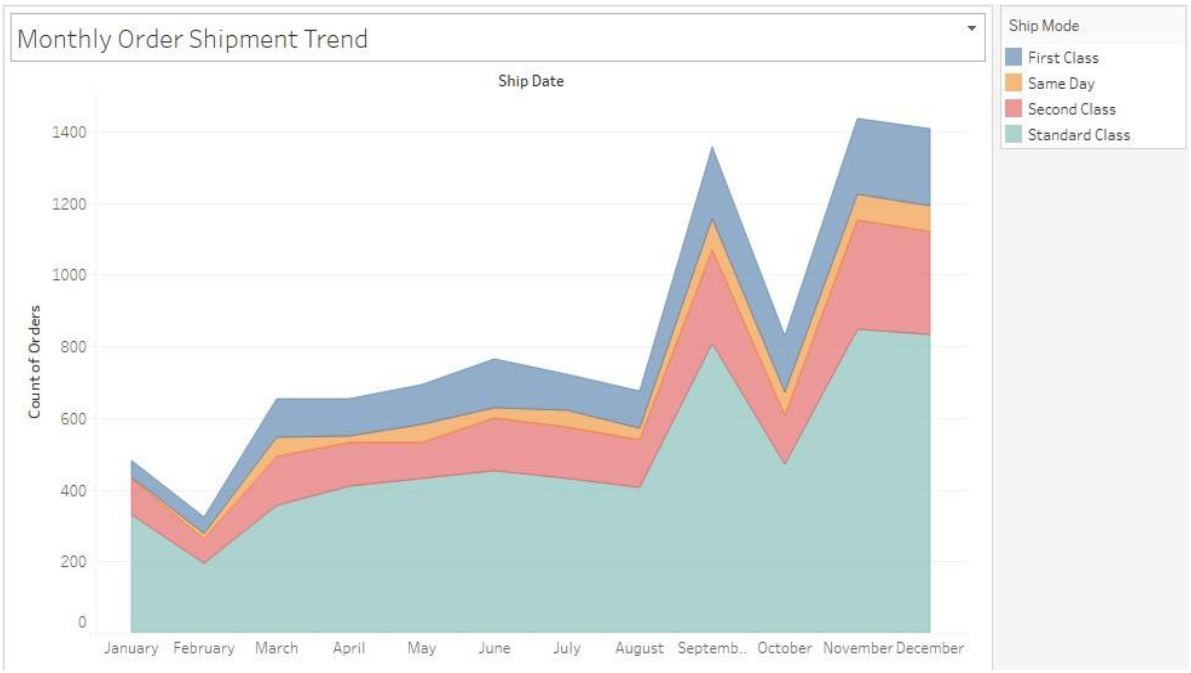
17. What is the monthly trend in the number of orders shipped?

Ans –

Why did I choose this chart?

An area chart is suitable for showcasing the monthly trend in order shipments. This chart is effective for displaying trends over time, making them suitable for illustrating the fluctuation in order shipment numbers across months.

- Visualizing Trends
- Encapsulating Volume Changes
- Clarity in Trends



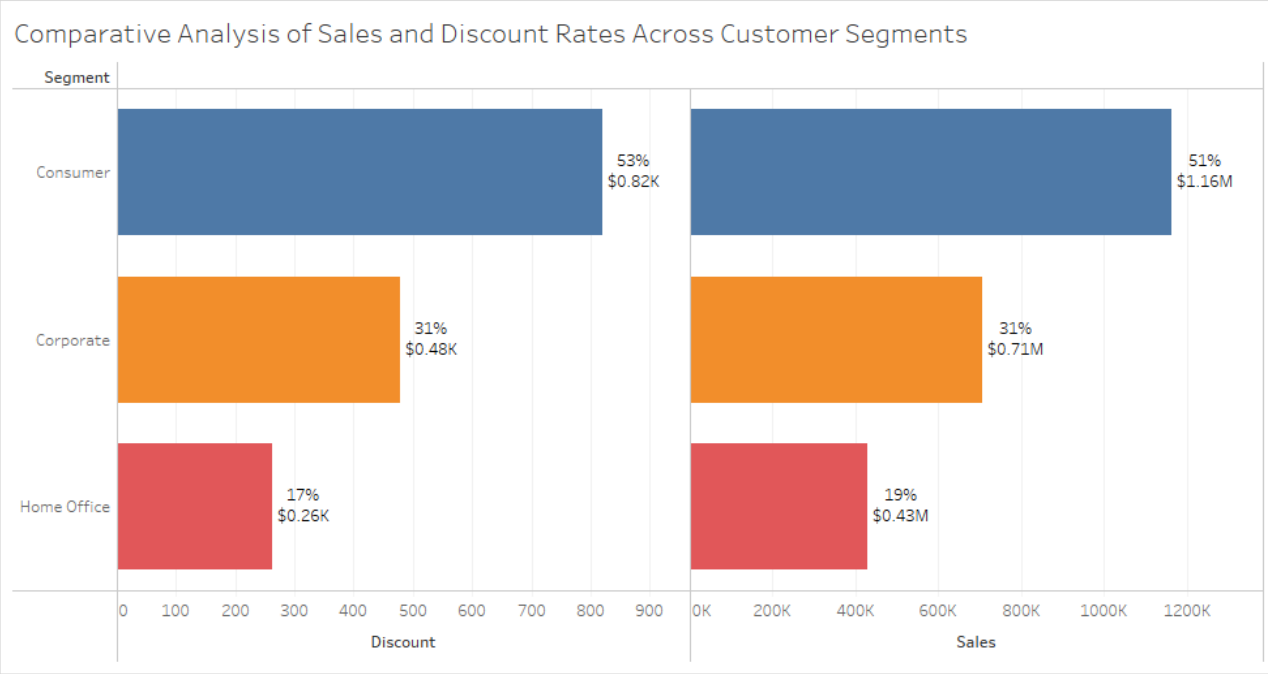
18. How do different customer segments perform in terms of sales and discount rates?

Ans –

Why did I choose this chart?

A horizontal bar chart is suitable for comparing how different customer segments perform in terms of sales and discount rates. These horizontal bar charts allow easy comparison of performance metrics (sales and discount rates in this case) across multiple customer segments placed side by side.

- Comparison of Categories
- Facilitates Ranking
- Clarity in Comparison



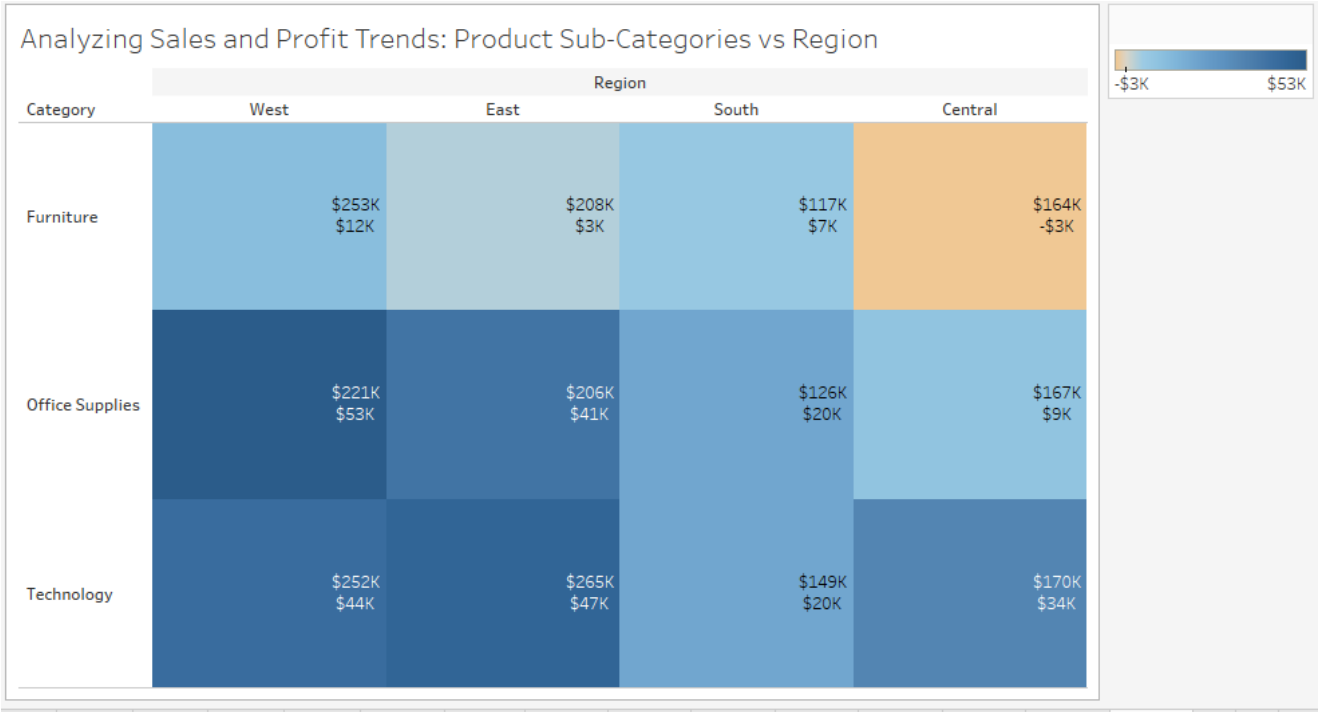
19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?

Ans –

Why did I choose this chart?

A heatmap is suitable for analyzing sales and profit trends across different product subcategories and regions due to these reasons. It allows simultaneous comparison of multiple variables (sales and profits) across both product subcategories and regions. Each cell represents a combination of a subcategory and a region, making it efficient for comparing these metrics.

- Dual Comparison
- Color Representation
- Compact Visualization



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

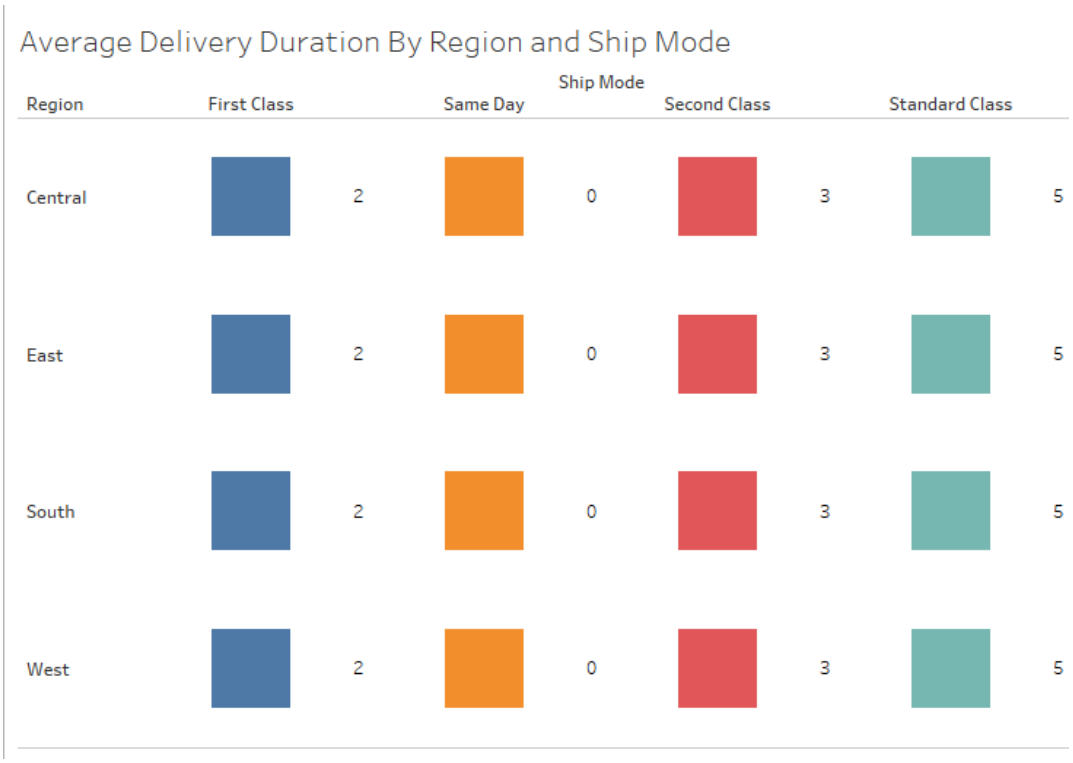
Ans –

Why did I choose this chart?

I am using the **Highlight table**, which is suitable for showcasing the average delivery duration for different regions and ship modes for these reasons:

Highlighting Variations: Tables highlighting differences in data by applying colors show where there are regions or ship modes whose average delivery durations are shorter or longer respectively within just one's glance.

Comparative Analysis: The comparative bar chart, for instance, allows a side-by-side comparison of mean lead times across the region and ship, which quickly makes the outliers evident.



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

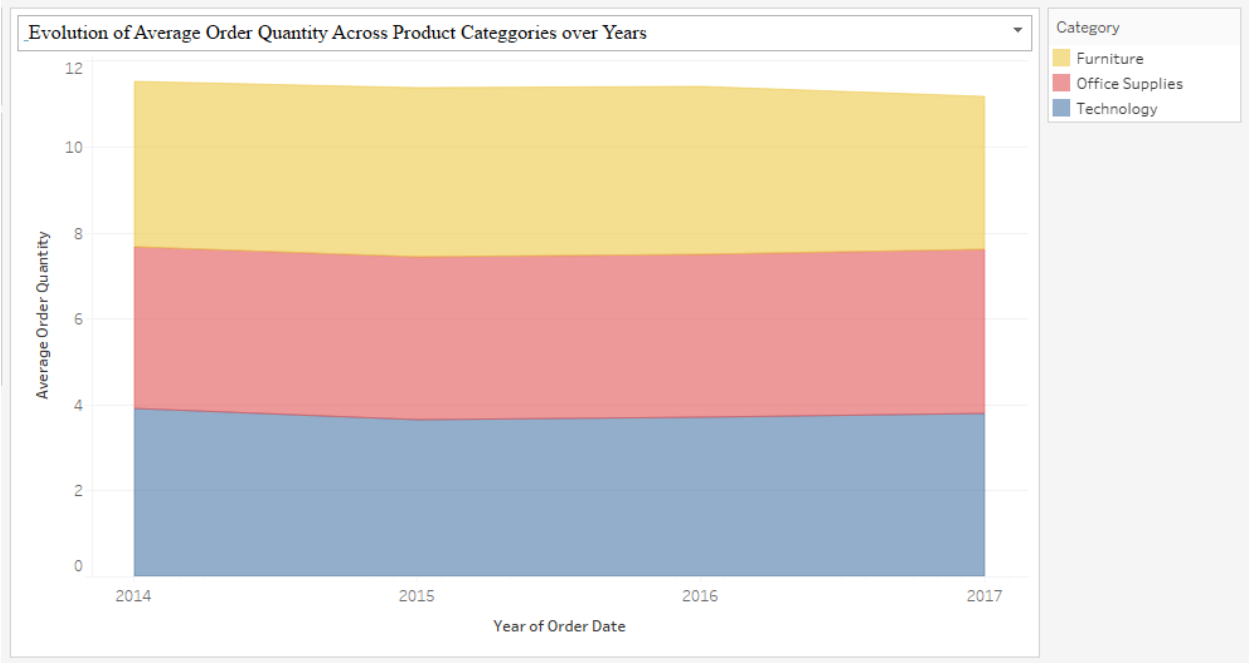
Ans –

Why did I choose this chart?

An area chart is suitable for showcasing how the average order quantity has changed over the years for various product categories due to these reasons:

Trend Visualization: Area charts effectively display trends over time, allowing for the visualization of changes in average order quantity across different product categories over the years.

Magnitude Representation: The filled areas beneath the lines in an area chart can depict the magnitude of changes in average order quantities, providing a sense of volume or scale in the variations.

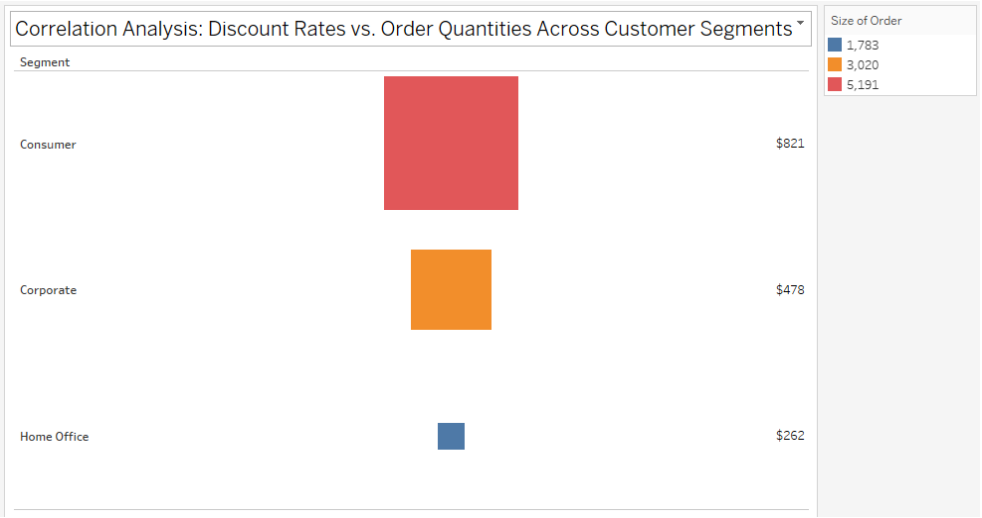


22. Can we visualise the correlation between discount rates and order quantities for different customer segments?

Ans –

Why did I choose this chart?

We can visualize the correlation between discount rates and order quantities for different customer segments using this heat map.This heat map would be suitable because it would allow us to quickly and easily see the correlation between discount rates and order quantities across customer segments.



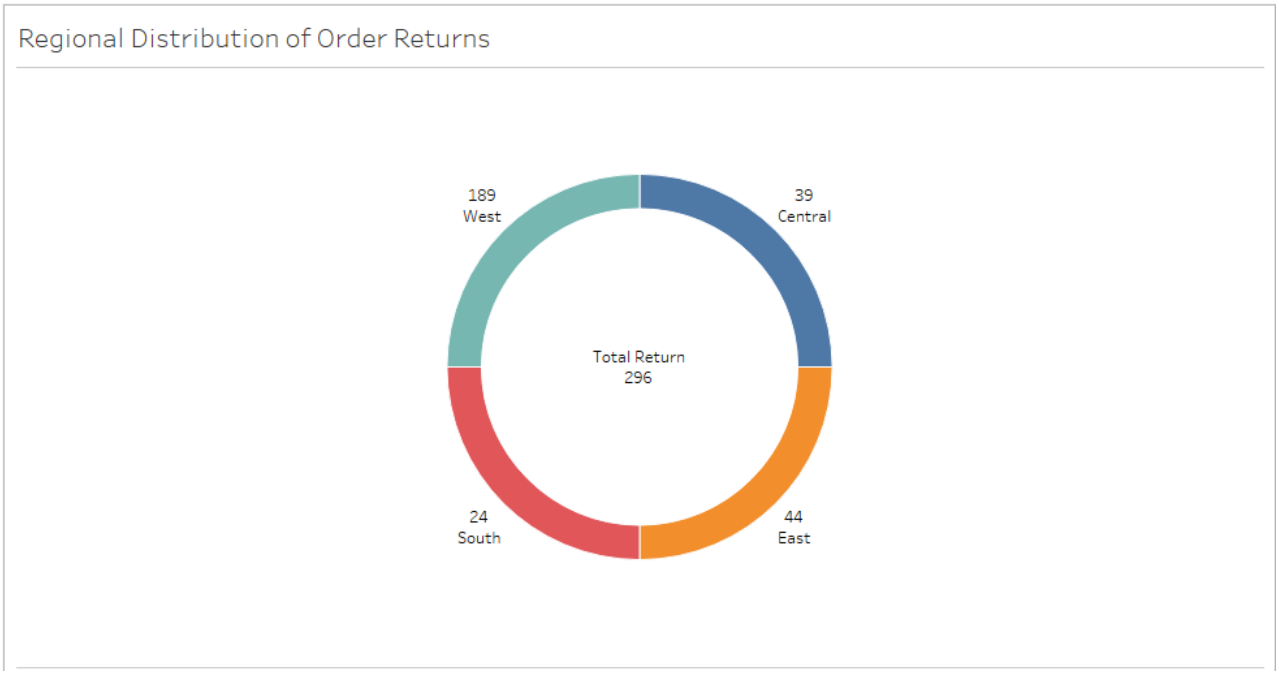
23. What is the proportion of orders returned in each region within the Superstore dataset?

Ans –

Why did I choose this chart?

Donut charts effectively illustrate the contribution of each region to the total orders returned, showing the proportion of returns about the whole dataset. Each segment of the donut chart represents a specific region, making it easy to distinguish and compare the return proportions across different regions.

- Clear Segmentation
- Concise Representation
- Easy Interpretation



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

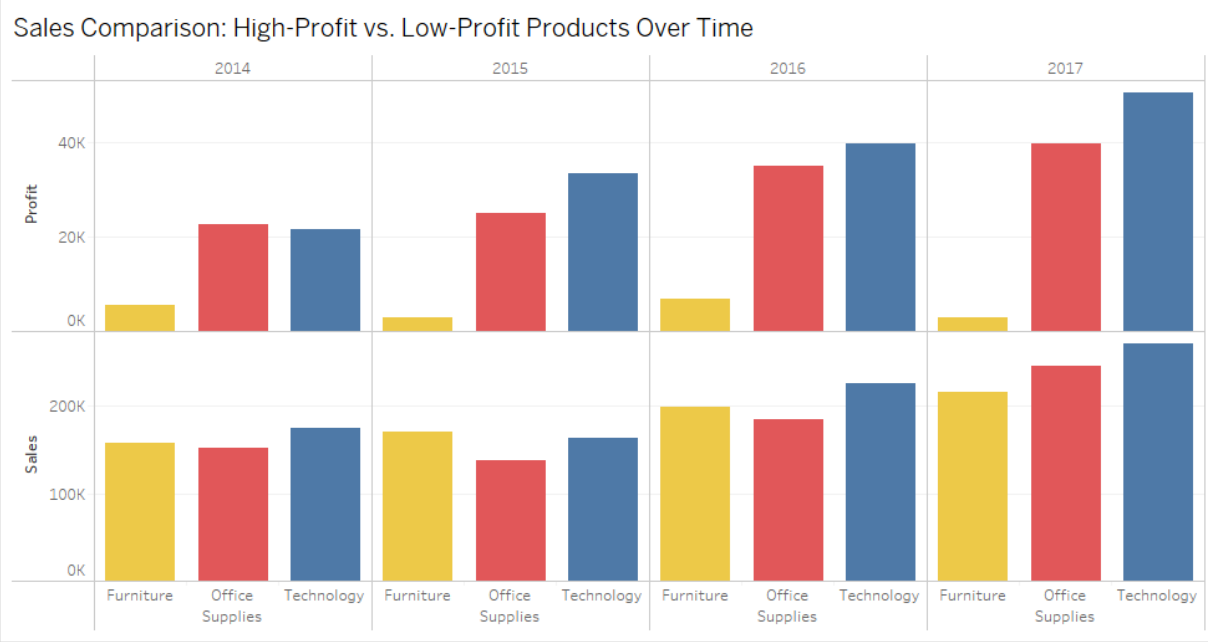
Ans –

Why did I choose this chart?

Stacked bar charts allow the comparison of multiple categories (high-profit

and low-profit products in this case) across different time periods, showing their individual contributions to the total sales.

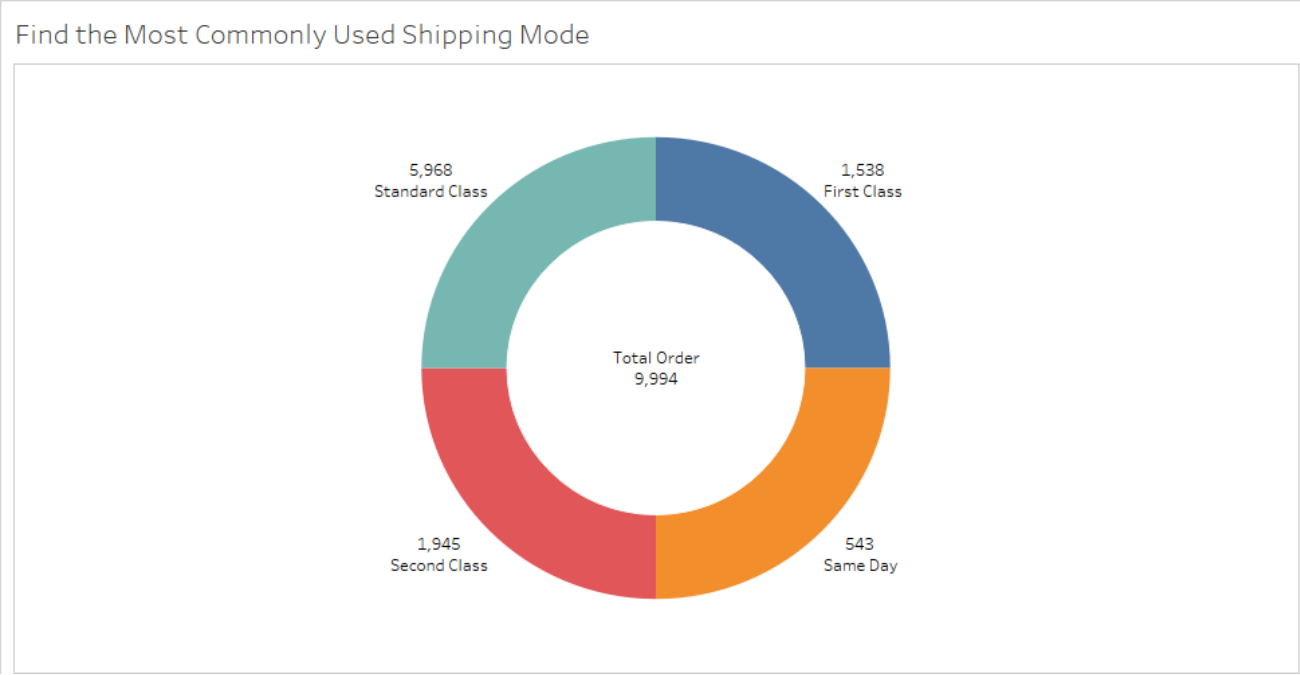
- Comparison of Categories
- Temporal Trends
- Visual Representation



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

Why did I choose this chart?

Donut charts effectively display proportions or percentages, making it easy to visualize the distribution of different shipping modes and identify the most prevalent one.



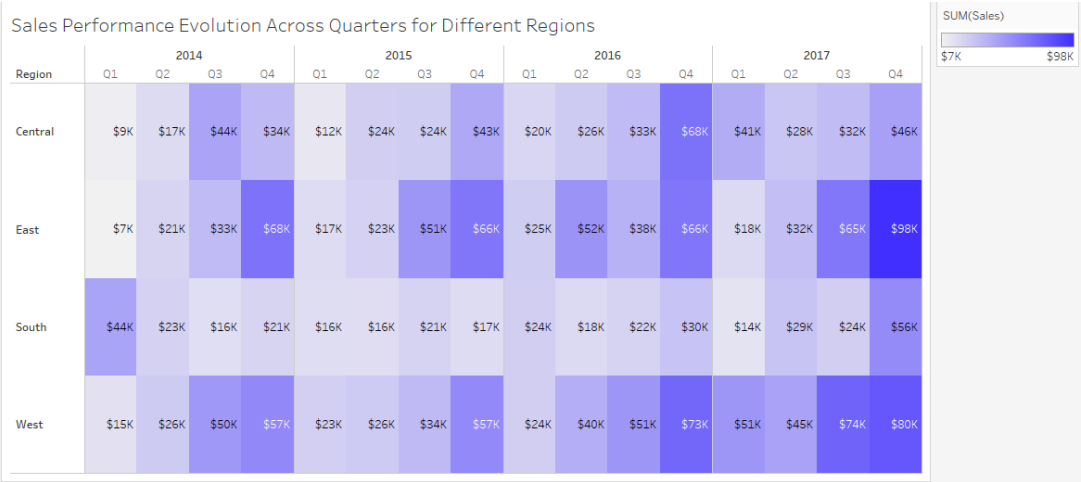
26. How does the sales performance of different regions evolve throughout the quarters of a year?

Ans –

Why did I choose this chart?

Heatmaps effectively display data trends over time (quarters in this case) and across different regions. They allow for a quick comparison of sales performance across multiple categories (regions) simultaneously. It helps to present a large amount of data in a compact format, making it easier to visualize sales performance changes across multiple regions and quarters without overwhelming the viewer.

- Compact Display
- Color-Coded Representation



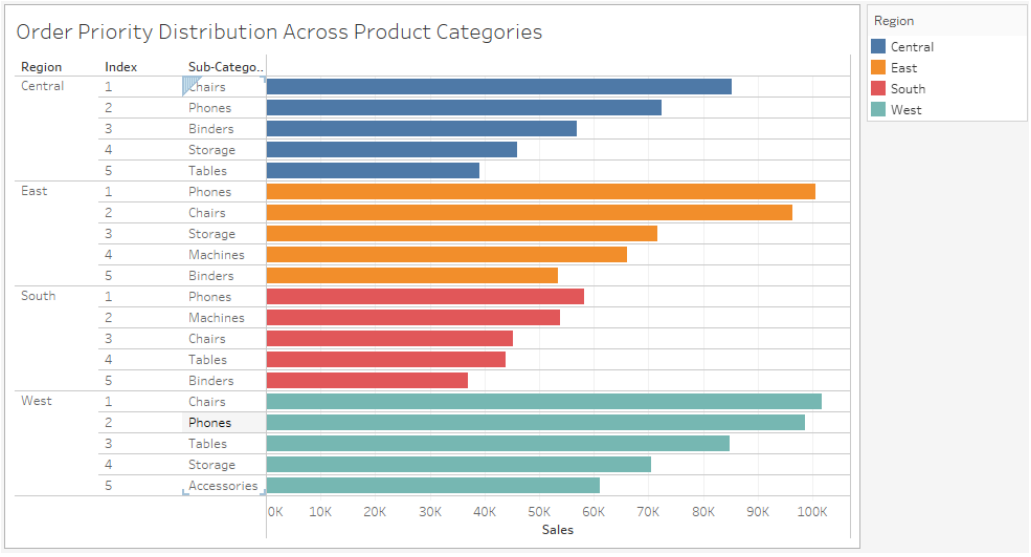
27. What is the distribution of order priorities across different product categories?

Ans-

Why did I choose this chart?

Horizontal bar charts allow for easy comparison of different categories (product categories in this case) by length, making it simple to compare the distribution of order priorities across categories.

- Order and Priority Visualization
- Space Efficiency
- Emphasis on Priority Distribution



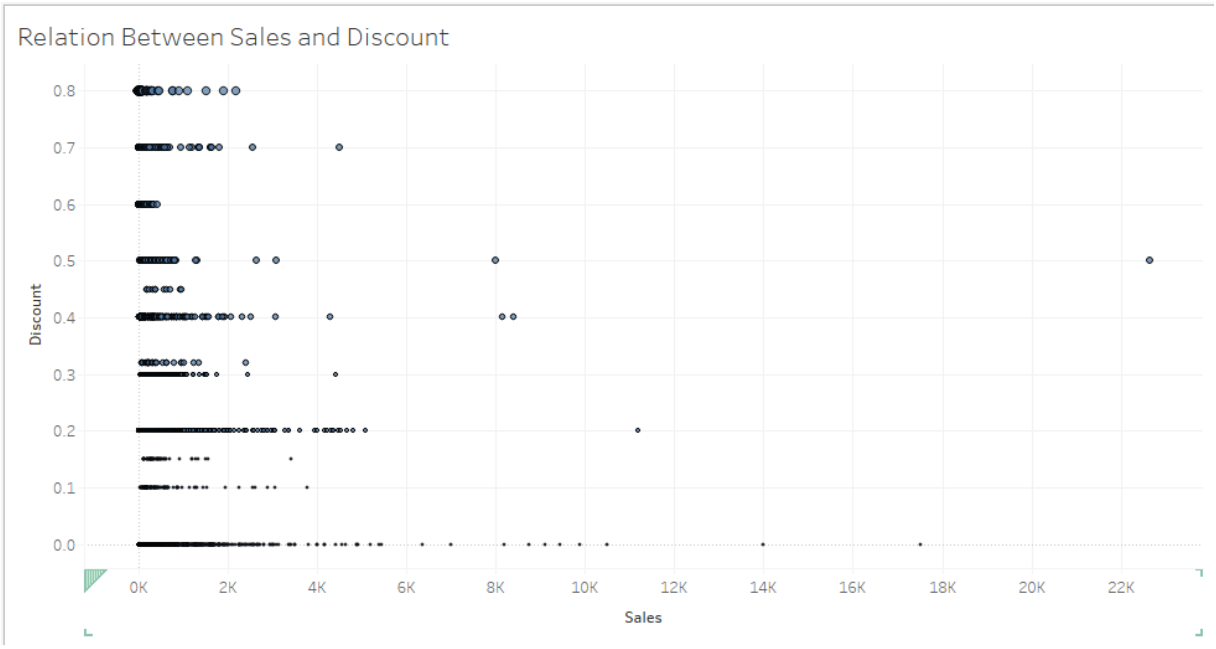
28. What is the relationship between discounts and sales?

Ans –

Why did I choose this chart?

Scatter plots are used to display the relationship between two different variables.

The choice of a scatter plot in this context is appropriate because it effectively shows the correlation between sales and discount rates.



Additionally, a scatter plot with multiple variables can also be used to demonstrate trends and patterns.In this case, it allows the user to visualize how changes in the discount rate affect sales. ***By loofiing aĒĒhe scaĒĒer ploĒ, iĒ is clear ĒhaĒ Ēhere is a general downward Ērend as Ēhe discounĒ raĒe increases. This indicaĒes ĒhaĒ sales generally decrease as Ēhe discounĒ raĒe increases.***

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

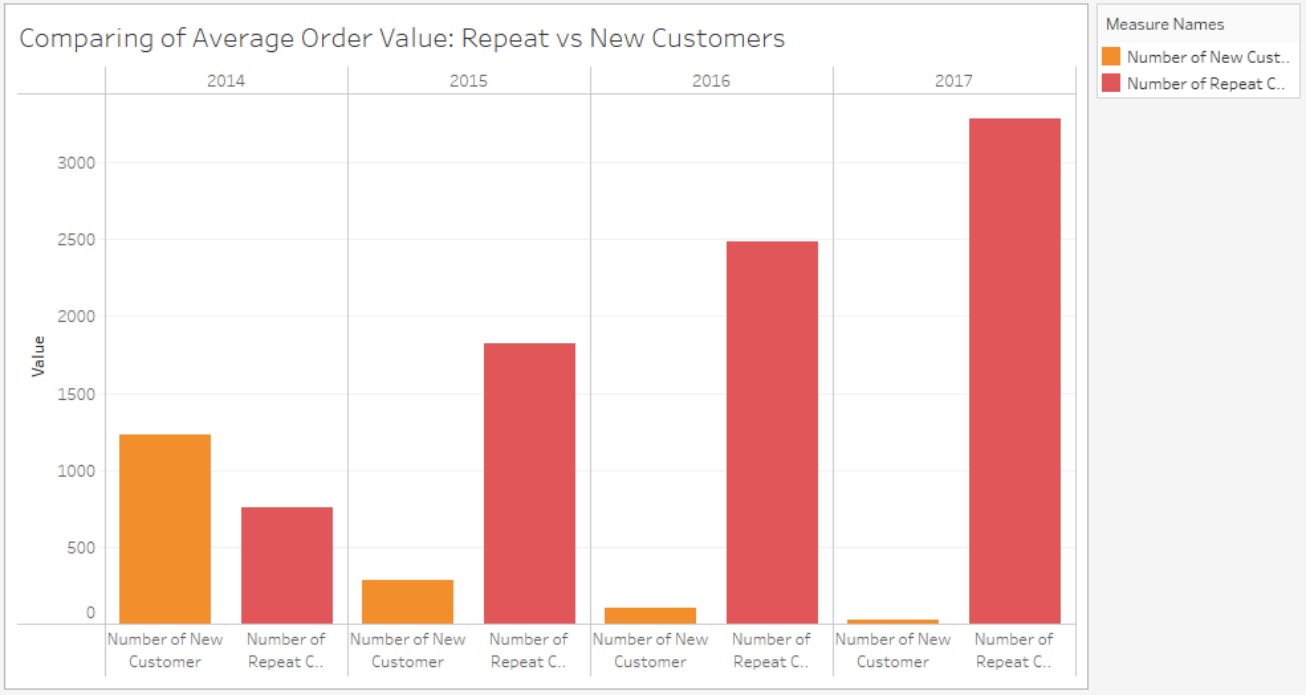
Ans –

Why did I choose this chart?

Side-by-side bar charts effectively display and compare data between two distinct categories (repeat and new customers) by placing their

average order values side by side. The chart highlights the contrast between repeat and new customers' average order values, allowing viewers to quickly understand how these values differ.

- Comparison of Categories
- Emphasis on Contrasts
- Visual Impact



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

Ans –

Why did I choose this chart?

It succinctly encapsulates the main theme—analyzing the impact of returns on profitability across different geographical locations—providing a clear understanding of the chart's focus. It directly addresses the geographical aspect of returns and their influence on overall profitability, aligning with the specific inquiry regarding the

distribution of returns across locations and their impact on profitability.

- Clarity
- Relevance
- Engagement