**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.

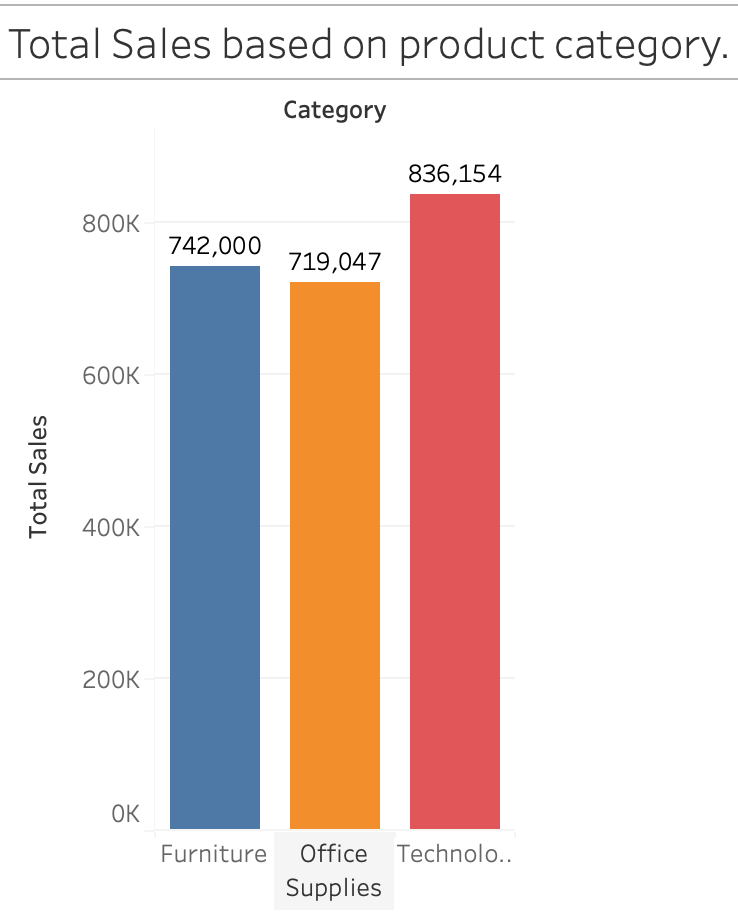
**Good luck!**

**Questions:**

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

**Answer: Technology** had the highest total sales at ₹836,154. **Furniture** followed with total sales of ₹742,000. **Office Supplies** recorded the lowest sales among the three categories at ₹719,047. There's a noticeable difference in sales between Technology and the other two categories. The sales for Furniture and Office Supplies are relatively close to each other.

**Visualization:**



**Insights:**

* **Product Category Sales:** Technology leads significantly, while Furniture and Office Supplies show similar, lower sales.
* **Month-wise Sales:** Sales peak in November/December and are lowest in February, indicating strong year-end performance and mid-year dips. Technology's strength likely contributes to overall high sales.

**Reason for Choosing Column Chart:**

* A bar chart effectively compares total sales across different product categories, making it easy to visualize and interpret the data differences.

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

**Answer: December** had the highest sales (₹340,517), while **February** had the lowest (₹68,546). Sales were generally higher in the latter part of the year, peaking in the final two months.

**Visualization:**

A graph of sales

AI-generated content may be incorrect.

**Insights from the Bar Plot:**

* + **Sales Trend**: Sales initially increased from January to June before peaking in October (**204,727**) and November (**330,682**), showcasing a significant seasonal growth in the latter part of the year.
  + **Sales Decline**: February showed the lowest sales at **68,546**, highlighting a drop early in the year.
  + **Mid-Year Stability**: Sales remained relatively stable through mid-year, with values consistently around **140,000 to 160,000** from April to August.

**Reason for Choosing Bar Plot:**

* Clear Comparisons: A bar plot effectively visualizes the differences in sales across each month.
* Easy Interpretation: The horizontal format allows for straightforward reading and comparison of values.

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

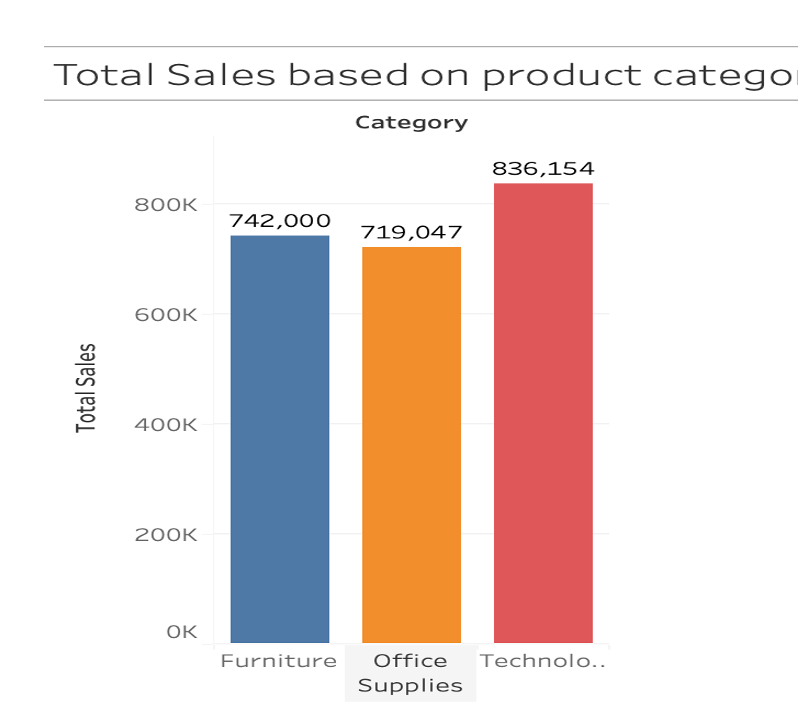
**Answer:**

 **Technology** has the highest total sales at **₹836,154**.

 **Furniture** has the second-highest total sales at **₹742,000**.

 **Office Supplies** has the lowest total sales at **₹719,047**.

**Visualization:**



**Insights:**

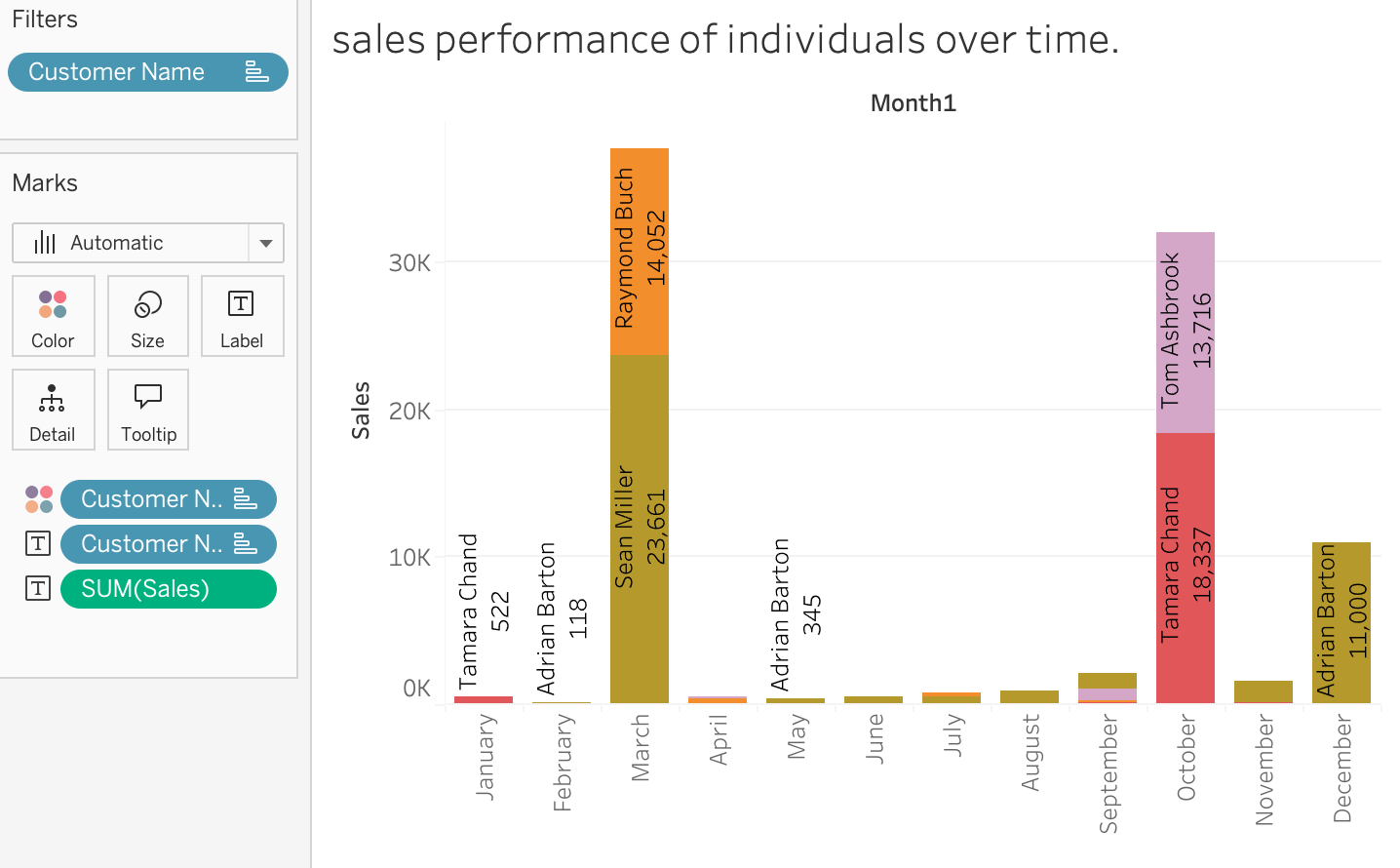
* Highest Sales in Technology: The Technology category achieved the highest total sales, amounting to $836,154.
* Strong Performance in Furniture: Furniture sales follow closely behind, with total sales reaching $742,000, indicating a robust market segment.
* Office Supplies Lagging: The Office Supplies category has the lowest total sales at $719,047, suggesting potential areas for growth or a need for strategic improvement.
* Proximity in Sales Figures: The sales figures for Furniture and Office Supplies are relatively close, with only a difference of $22,953 between them, highlighting a competitive market.
* Overall Sales Observation: The cumulative total sales across all three categories reflect a healthy performance, with total sales exceeding $2.3million collectively.

**Reason for Choosing Column Plot:** The column chart effectively illustrates the total sales across different- different product categories, highlighting variations and trends for informed decision-making.

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

**Answer**: Yes, we can analyse the sales performance of individual customers over time by using various types of visualizations. Among these, the **line chart** is one of the best options for showing changes and trends over time. So, let’s visualize the sales performance of individual customers using a line chart to gain insights into their purchasing behaviour and sales trends.

**Visualization:**



**Insights:**

 **Sean Miller's Strong Start:** Sean Miller shows a significantly high sales performance in March with 23,661, indicating a strong start to the observed period.

 **Tom Ashbrook's Consistent Performance:** Tom Ashbrook consistently generates sales, particularly in October (18,337) and December (11,000), suggesting a reliable sales contribution.

 **Variable Performance:** The sales performance of individuals varies greatly month to month, with some months showing minimal or no sales for certain individuals.

 **December Surge:** December shows sales activity from multiple individuals, highlighting a potential end-of-year sales push or seasonal trend.

 **Focus on Key Performers:** Sean Miller and Tom Ashbrook appear to be key sales drivers, with high sales in specific months. Further analysis may reveal what contributes to their success.

**Reason for choosing column chart:** The column chart is an effective way to represent values over time, making it ideal for our needs. In this case, we aim to visualize the total sales performance of individuals over a specific period.

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

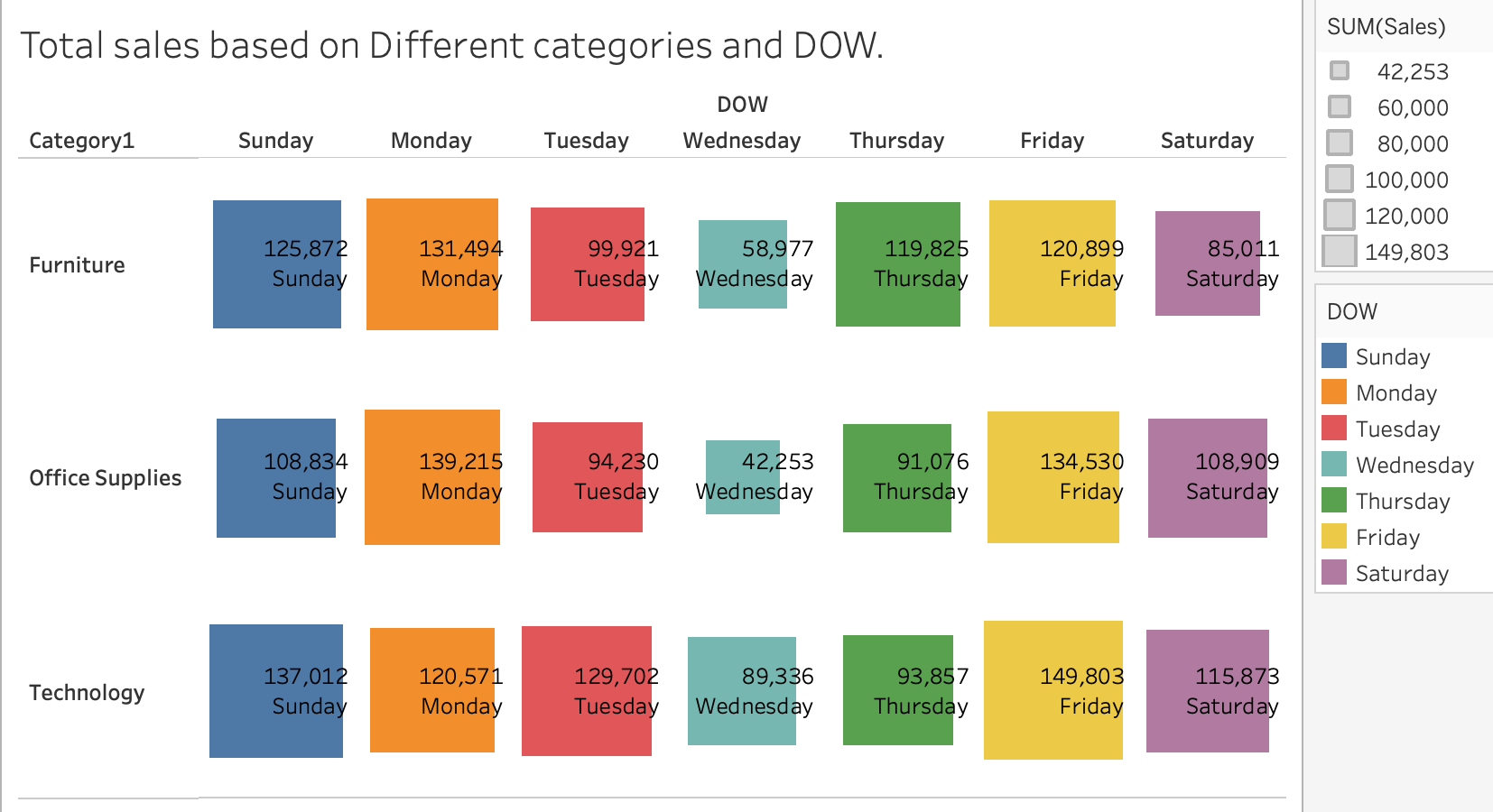
**Answer:**

**Sales Categories:** Displays total sales in three categories: Furniture, Office Supplies, and Technology.

**Days of the Week (DOW):** Presents sales data across seven days: Sunday to Saturday.

**Sales Values:** Numerical values in the cells represent total sales for each category on each day.

**Visualization:**



**Insights:**

**Peak Sales:** Technology had the highest sales on Friday (149,803), making it the top-performing category.

**Friday** is also promising, particularly for Technology.

**Lowest Sales:** Office Supplies had the lowest sales on Wednesday (42,253), indicating potential need for marketing strategies on this day.

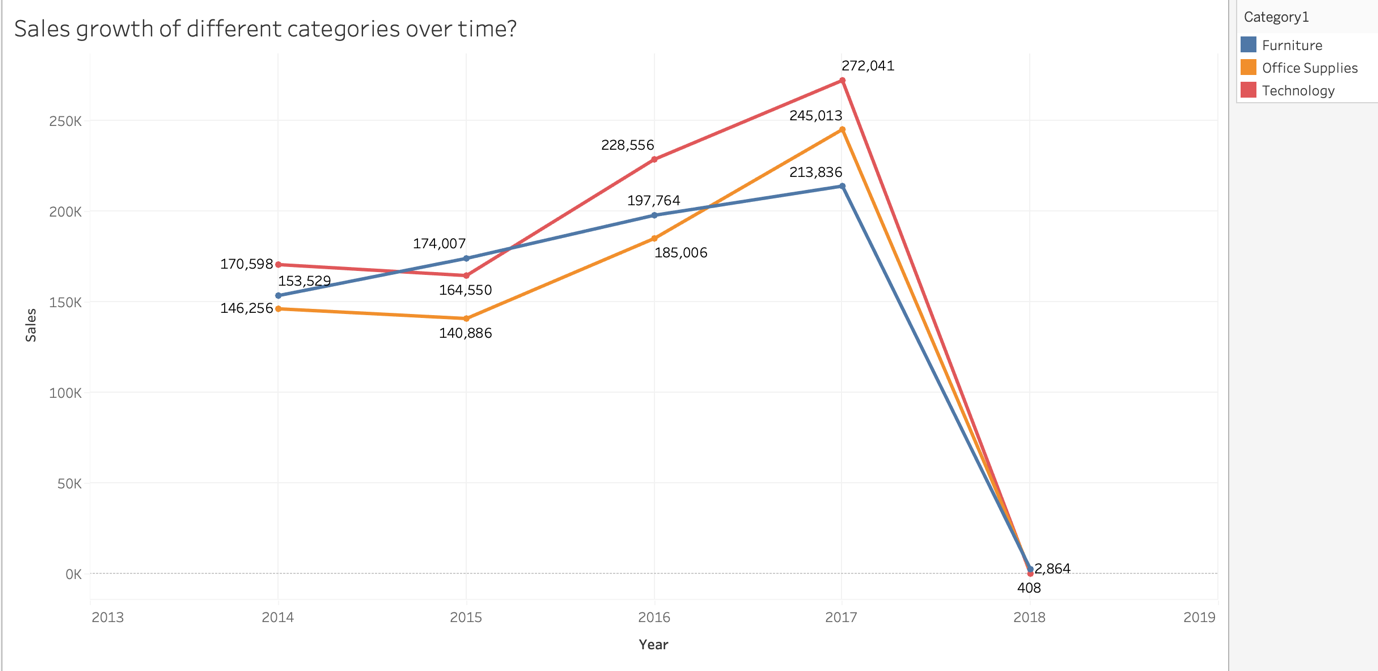
**Consistent Performers:** Both Furniture and Technology show consistently high sales on Monday, indicating a strong start to the week.

**Reasons for Choosing Heatmap Chart:** It effectively visualizes complex data, making it easier to comprehend sales performance at glance.

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

**Answer:** Yes, we can visualise the sales growth of different-different product category over the time.

**Visualization:**



**Insights:**

**Overall Growth:**

* All categories increased sales from 2013/14 to 2017/18.
* Example: The highest-performing category grew from roughly 400 in 2013/14 to about 700 in 2017/18.

**Variable Performance:**

* Categories have different starting and ending sales values.
* Example: In 2015/16, the sales range spanned from approximately 250 (lowest category) to 600 (highest category).
* Growth rates also vary: Some lines are steeper than others.

**Growth Spurts/Lags:**

* Specific years show rapid or slow sales changes.
* Example: The middle category showed a significant increase between 2014/15 and 2015/16 (a steep upward slope).
* Example: The lowest category's growth appears to slow down after 2016/17.

**Convergence/Divergence:**

* Sales gaps between categories change over time.
* Example: Initially, in 2013/14, the gap between the top and middle categories was smaller than the gap in 2017/18, suggesting divergence.

**Trend-Based Projections:**

* Recent trends indicate potential future sales.
* Example: If the highest category maintains its current trend, it could exceed 750 in the next period.
* Example: The lowest category's trend suggests potentially slower growth compared to other categories.

**Reasons for Choosing Line Chart:** Line charts are ideal for visualizing sales growth over time because they clearly show trends and changes in continuous data, making it easy to compare the performance of different product categories simultaneously across the years.

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

**Answer:** The **West** region leads in total sales with ₹725,458, followed closely by the **East** region at ₹678,781. The **Central** region shows sales of ₹501,240, while the **South** region has the lowest total sales at ₹391,722. This highlights a significant regional variation in sales performance.

**Visualization:**

**A graph with different colored rectangles

AI-generated content may be incorrect.**

**Insights:**

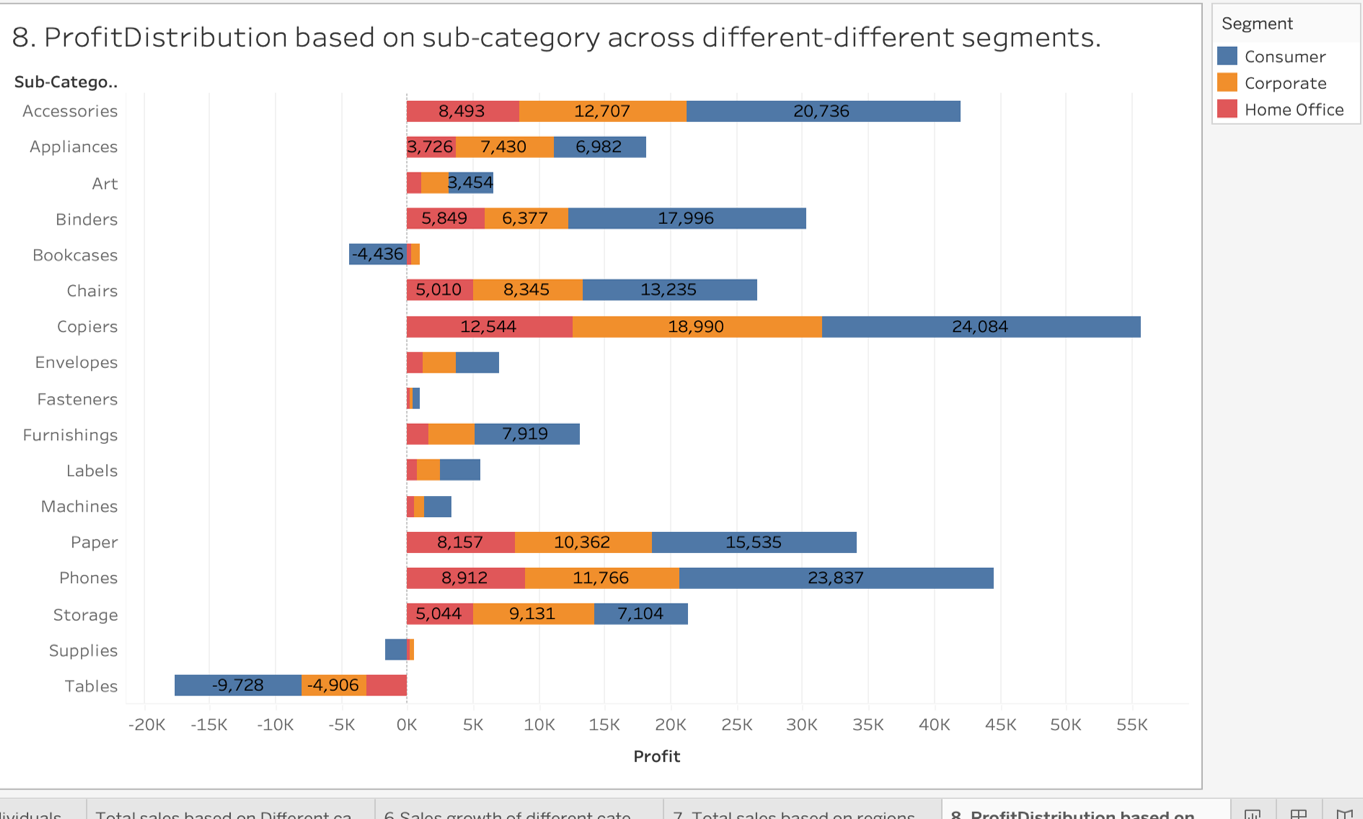
* As we can see in the chart the West region leads the sales and having the most number of sales(7,25,458), it means people from west region they prefer to buy all the products which are offered in the product category.
* After West, East region holds the second most number of sales which is total(6,78,781).
* The Central region's sales (501,240) fall in the middle, suggesting moderate sales activity compared to the top two regions.
* Now when we talk about the less number of sales then South region is at last potion in terms of total sales, it has least total sales(3,91,722).
* **Sales Disparity:** There's a noticeable difference in sales performance across regions. The West's sales are significantly higher than the South's, highlighting an uneven distribution of sales.

**Reasons for this chart:** Given the four categorical regions in our dataset, a column chart provides the optimal visualization for comparing their respective total sales.

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

**Answer:** We can visualize the composition across various subcategories within different customer segments effectively using Tableau. To achieve this, we will place the **Sub-Category** field on the **X-axis** and the **Total Profit** on the **Y-axis**. Customer **Segments** will be distinguished using the **Colour** mark, allowing each segment to be visually differentiated across the subcategories. This approach will clearly highlight the distribution of profits among subcategories for each customer segment, making it easier to identify trends, opportunities, and areas for improvement.

**Visualization:**



**Insights:**

* **Copiers Drive Significant Profit:** The "Copiers" sub-category generates the highest overall profit, with a substantial contribution from all three customer segments (Consumer, Corporate, and Home Office).
* **Tables Incur Losses Across All Segments:** The "Tables" sub-category consistently shows negative profit (losses) across all customer segments, indicating a significant area of concern.
* **Technology Sub-categories are Highly Profitable:** "Copiers" and "Phones" (both likely within the Technology category) show the highest profit margins, suggesting this product area is a key profit driver.
* **Consumer Segment Often Dominates Profit:** For many sub-categories (e.g., Accessories, Binders, Chairs, Paper, Phones, Storage), the "Consumer" segment contributes the largest share of the profit compared to the "Corporate" and "Home Office" segments.
* **Varying Segment Contribution by Sub-category:** The profit contribution of each customer segment varies significantly depending on the sub-category. For example, "Appliances" see a more balanced contribution from all segments, while "Bookcases" profit is primarily driven by the "Corporate" segment.

**Reasons for Choosing column Chart:** A bar plot is suitable here because it effectively displays the profit contribution of different customer segments (categorical data) for each sub-category, allowing for easy comparison of profit amounts and segment composition within each sub-category. The stacked format further enables visualization of the total profit and the relative contribution of each segment to that total.

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

**Answer:** The **West** region contributes the largest percentage to overall sales at 31.58%, closely followed by the **East** region with 29.55%. The **Central** region accounts for 21.82% of the total sales, while the **South** region has the smallest contribution at 17.05%. This highlights the West and East as the primary revenue-generating regions.

**Visualization:**

**A pie chart with numbers and a few different colored circles

AI-generated content may be incorrect.**

**Insights:**

* **Highest Contribution:** The **West region** contributes the largest share at **31.58%** of overall sales.
* **Significant Contributions:** Following the West, the **East region** has a substantial share of **29.55%**, making it the second-largest contributor.
* **Moderate Contribution:** The **Central region** accounts for **21.82%**, placing it in the mid-range of contributions among the regions.
* **Lowest Contribution:** The **South region** shows the smallest contribution at **17.05%**, indicating it has the least impact on overall sales.
* **Overall Distribution:** The combined contribution of the West and East regions (approximately **61.13%**) suggests a strong performance from these areas compared to the Central and South regions.

**Reason for the Pie Chart:** The pie chart is the best way to represent the percentage contribution of each categorical value in any variable, which is why we use a pie chart.

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

**Answer:** We can visualize the total profit margins associated with different shipping modes and customer segments.

**X-axis:** Total profit margin

**Y-axis:** Types of ship modes

**Colour Representation:** Different customer segments

**Visualization:**

****

**Insights:**

* Highest Profit Margin: Standard Class: The Standard Class shipping mode shows the highest profit margin, with a total of **77,924**.
* Significant Difference from Other Modes: There is a notable difference in profit margins between Standard Class and other shipping modes like Second Class (**24,947**) and First Class (**21,374**).
* Second Class vs. First Class: Second Class has a slightly higher profit margin than First Class, with **24,947** versus **21,374**, indicating that it could be a more cost-effective choice.
* Low Profit for Same Day Shipping: Same Day shipping has the lowest profit margin among all modes, reflecting its limited profitability compared to others.
* Revenue Distribution Insights: Within each shipping mode, there's a clear distribution of profit margins across different customer segments, suggesting a diverse customer base with varying preferences for shipping speed and cost.

**Reason for the Column Chart:**

* Clearly displays profit margins across shipping modes (Standard Class, Second Class, First Class, Same Day).
* Colours represent various customer segments for easy distinction and analysis of their contributions.
* Vertical bars highlight the scale of profit margins, showing which modes are most profitable.
* Combines data categories, facilitating analysis of relationships between shipping modes and customer segments.

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

**Answer:**

* **Furniture** takes about **24.1 days** to ship after ordering.
* **Office Supplies** takes about **22.3 days**.
* **Technology** takes about **22.1 days**.

Thus, **Furniture** has the **longest processing time**, while **Technology** has the **fastest** among the three product categories.

**Visualization:**

A screenshot of a graph

AI-generated content may be incorrect.

**Insights:**

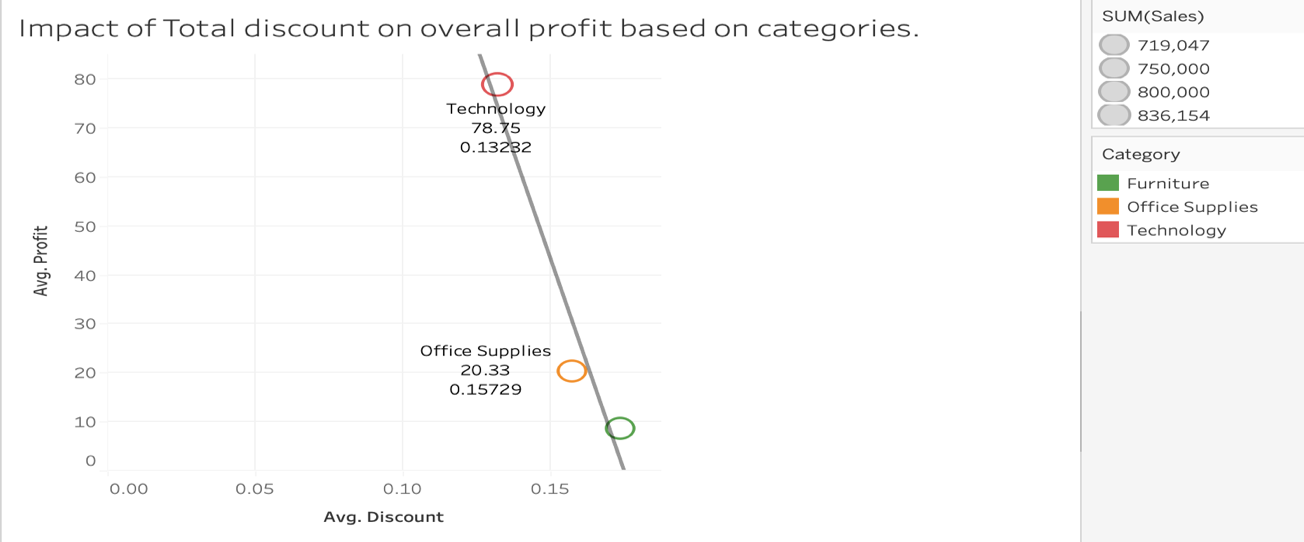
* Technology ships fastest: Technology items are smaller and standardized, leading to quicker processing.
* Office Supplies are in between: Shipping times are moderate, suggesting average handling complexity.
* Small but important difference: The 2-day gap can still affect customer satisfaction, especially at scale.
* **Improvement focus:** Streamlining Furniture logistics could help reduce processing time and boost competitiveness.

**Reason for the Column Chart:** I chose a bar plot to clearly compare the average processing time across distinct product categories (Furniture, Office Supplies, Technology). The bar heights directly represent the average time for each category, making visual comparison easy. This plot effectively shows the distribution of this metric across these separate groups.

12. How do discounts affect overall profit?.

**Answer:** This scatter plot visualizes the relationship between the average discount applied and the average profit for different product categories (Furniture, Office Supplies, Technology). Each circle represents a category, positioned according to its average discount and average profit. The color of each circle further identifies the category. The downward trendline suggests a negative correlation between discount and profit across these categories.

**Visualization:**



**Insights:**

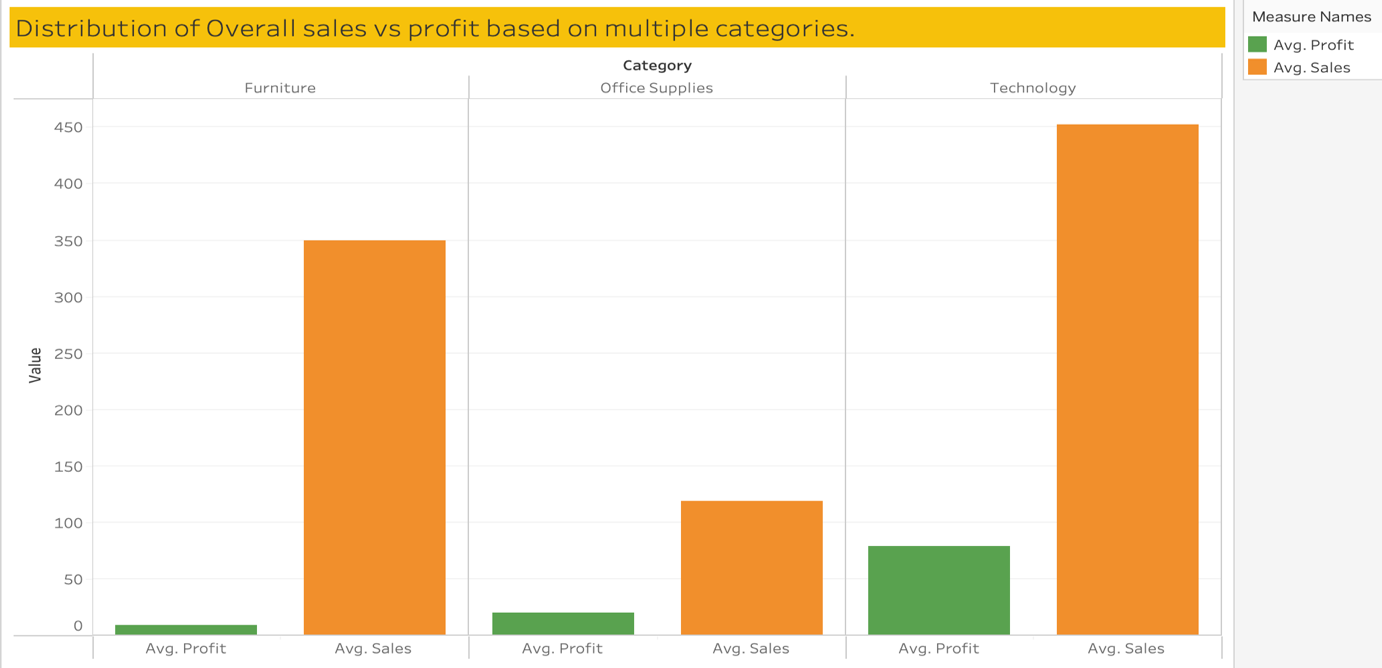
* **Higher discounts usually lead to lower profit:** Categories with bigger discounts tend to earn less profit on average.
* **Technology is the most profitable category:** Even with higher discounts, it brings in the highest average profit, likely due to better margins.
* **Furniture performs poorly:** It has high discounts but very low profit, which may indicate over-discounting.
* **Office Supplies stay profitable without big discounts:** This category offers almost no discounts but still earns decent profit.
* **More sales doesn’t always mean more profit:** Some categories sell more but still earn less profit overall.

**Reason for this chart:** Scatter chart is suitable for visualizing the relationship between two continuous variables – in this case, Average Discount and Average Profit – and how this relationship differs across categories (Furniture, Office Supplies, and Technology). It allows us to see if there's a correlation between the discount offered and the resulting profit margin for each product category.

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

**Answer:** To visualize the data, we first placed 'Category' on the x-axis and both Average Sales and Average Profit on the y-axis. The default visualization was a bubble chart, which wasn't effective for conveying the desired story. We therefore changed the chart type to a side-by-side bar chart for better representation. This allows us to clearly see and compare both the average sales and average profit values for each category.

**Visualization:**

****

**Insights:**

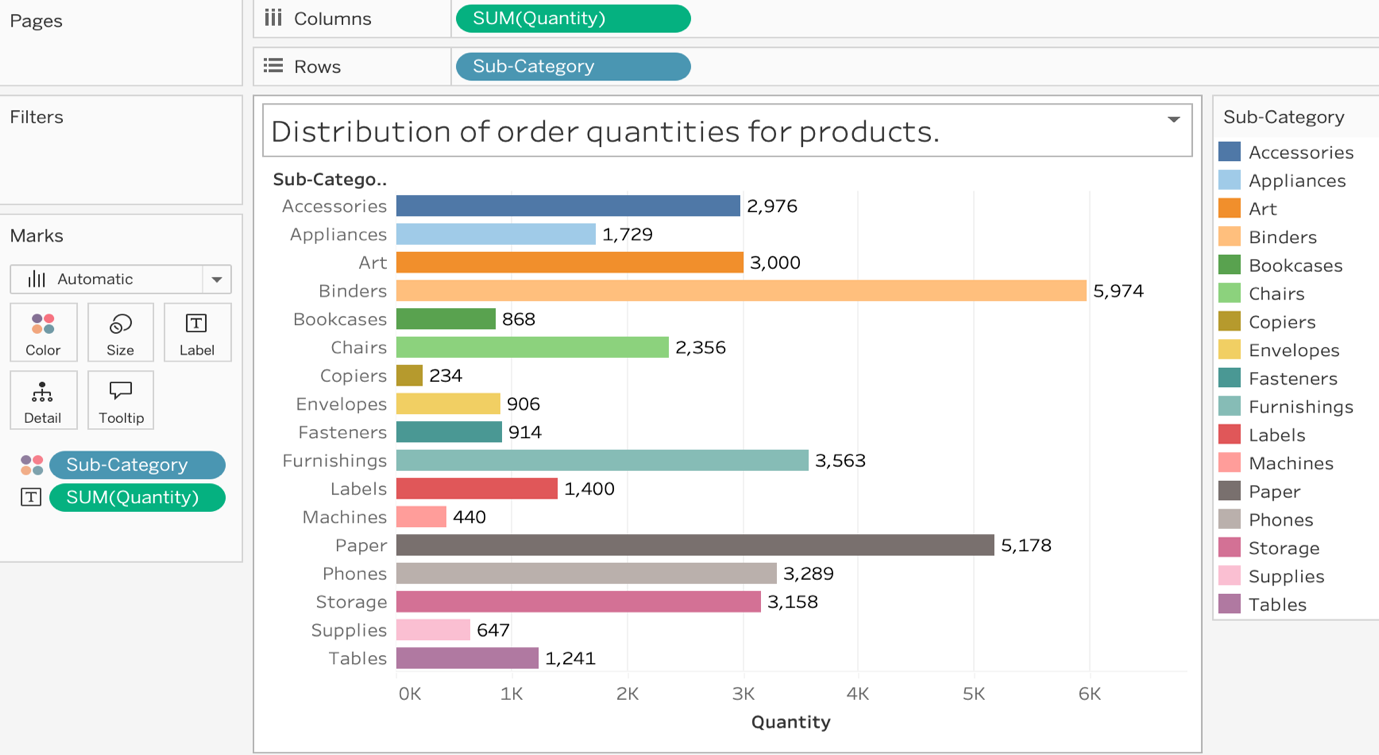
* **Highest Sales Performance:** The Technology category shows the highest average sales. Its sales value significantly surpasses both Furniture and Office Supplies.
* **Top Profit Generator:** Technology also leads in average profit. This category achieves the highest average profit margin among the three.
* **Furniture's Profit Challenge:** Furniture exhibits the lowest average profit. Despite having the second-highest average sales, its profitability is minimal.
* **Office Supplies Balance:** Office Supplies demonstrate modest average sales and profit. This category ranks lowest in average sales but outperforms Furniture in average profit.
* **Sales vs. Profit Disparity:** There isn't a direct correlation between high sales and high profit across all categories.

**Reason for this chart**: We selected a side-by-side column chart to show the distribution of sales and profit margins across different categories. Column charts are suitable for displaying distributions for categorical variables, and the side-by-side format allows for comparing multiple measures (sales and profit margin) for each category.

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

**Answer:** This horizontal bar chart effectively displays the distribution of order quantities across different product sub-categories. Each bar represents a sub-category, and its length corresponds to the total quantity ordered. This visualization allows for a clear comparison of the popularity or demand for each sub-category.

**Visualization:**



**Insights:**

**1. Leading Sub-Category:** Binders emerged as the top-selling product, with a substantial quantity of 5,974 ordered. This indicates a strong demand for binders compared to other product categories.

**2. Art and Furnishings:** The Art sub-category follows closely with 3,000 units sold, highlighting its popularity. Furnishings also show significant orders at 3,563, reflecting a robust interest in office-related décor.

**3. Low Performers:** Copiers demonstrated the lowest order quantity at just 234 units, which may indicate a decline in demand. Machines also had a relatively low performance with only 440 units ordered, suggesting potential overstock or lower interest.

**4. Moderate Sales:** Several categories like Labels (1,400) and Supplies (647) achieved moderate sales figures. This suggests stable but non-exceptional interest in these sub-categories.

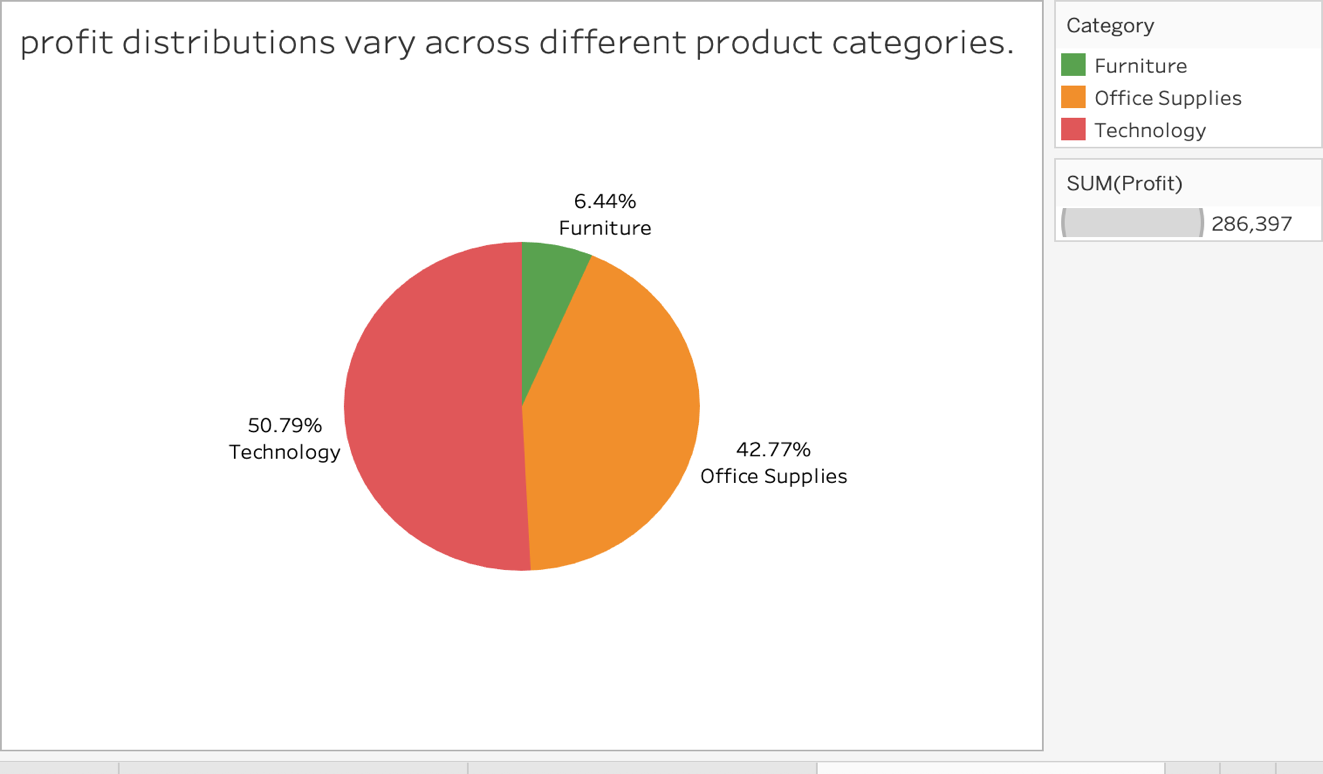
**5. Variation in Demand:** Overall, there is a wide variation in demand across sub-categories, with some like Tables (1,241) showing decent sales, while others lag significantly.

**Reason for this chart**: This horizontal bar chart clearly compares order quantities across 17 sub-categories, making it ideal for analysing product demand.

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

**Answer:** This pie chart illustrates how total profit is distributed across different product categories: Furniture, Office Supplies, and Technology. Each slice represents a category, and its size corresponds to the percentage of the total profit it contributes. The percentages displayed on each slice provide a clear view of the proportional profit contribution of each category.

**Visualization:**

****

**Insights:**

* **Dominant Category**: **Technology** accounts for the largest share of profits at **50.79%**.
* **Significant Contribution**: **Office Supplies** is the second largest category, contributing **42.77%** to the total profits.
* **Minimal Share**: **Furniture** has the smallest representation, making up only **6.44%** of the total profit.
* **Total Profit Summary:** The overall sum of profits across all categories is **$286,397**.
* **Imbalance in Profit Distribution**: There is a notable disparity in profit distribution among the categories, with Technology and Office Supplies together accounting for over **93%** of total profits, indicating a potential focus for business strategy on these categories.

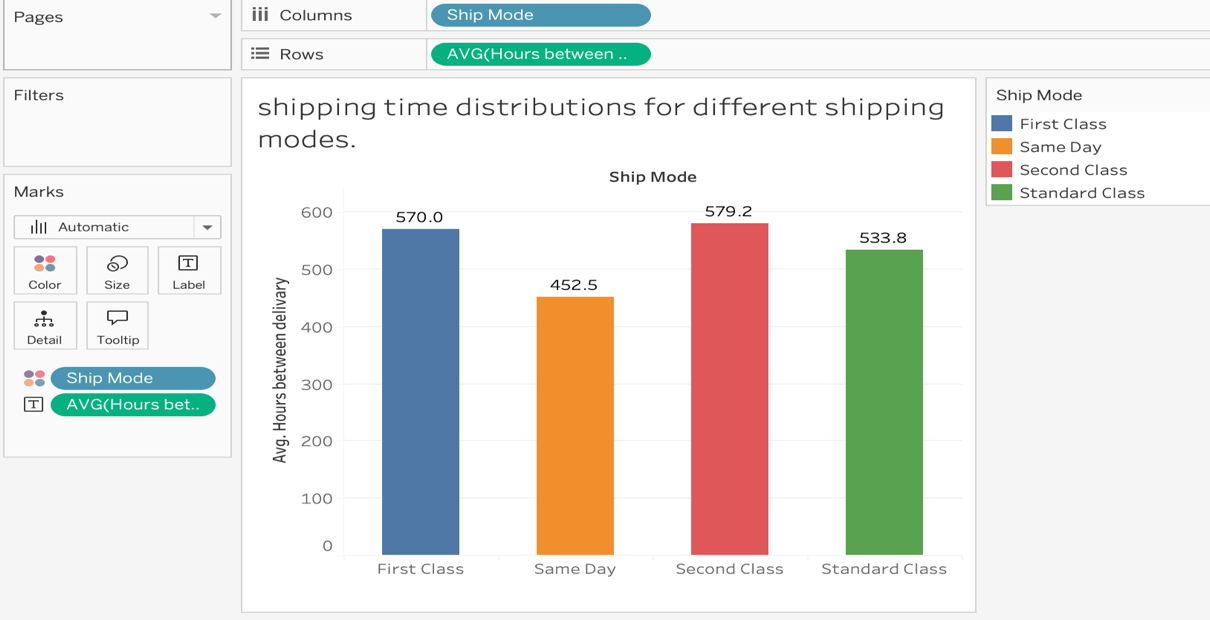
**Reason for this chart type:**

* To compare proportional profit contributions: The pie chart clearly shows how much each product category contributes to the overall profit.
* **T**o identify high and low-performing categories: It instantly highlights that Technology is the top contributor (50.79%), while Furniture lags (6.44%).

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

**Answer:** This chart compares shipping times across different shipping modes. To create it, we first calculated the time difference in hours between the 'Order Date' and 'Shipping Date' using a Tableau calculated field named hour\_between\_delivery (using the DATEDIFF('hour', [Order Date], [Ship Date]) function). We then plotted the [Average/Median - specify if known] hour\_between\_delivery on the y-axis against 'Shipping Mode' on the x-axis. Each shipping mode is distinguished by colour and labels show the specific time difference for each mode.

**Visualization:**



**Insights:**

* **Same Day shipping is the fastest**: It has the lowest average delivery time of approximately **452.5 hours**, which makes it the most efficient option in terms of delivery speed.
* **Second Class is the slowest shipping mode**: With an average delivery time of about **579.2 hours**, Second Class takes the longest among the four shipping options.
* **First Class is not the fastest**: Despite the name, **First Class averages 570 hours**, which is significantly slower than Same Day and even slower than Standard Class.
* **Standard Class performs better than First and Second Class**: At **533.8 hours**, Standard Class ships faster than both First Class and Second Class, making it a better choice than the name might suggest.
* **Naming does not reflect actual performance**: The chart highlights a mismatch between shipping mode names and their actual delivery times, especially with “First Class” and “Second Class” performing worse than “Standard” and “Same Day”.

**Reason for this chart type:**  To effectively visualize the distribution of shipping time (in hours) across different shipping modes, a column chart is an excellent choice. Column charts are well-suited for comparing categorical data, allowing for a clear understanding of how shipping time varies between the different modes.

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

**Answer:**

* X-Axis (Month): Properly represents each month from January to December.
* Y-Axis (Count of Orders): Shows the count of Row ID, which serves as a proxy for number of orders.
* Bar Chart: A good choice to show monthly trends in discrete values.
* Colour Coding: Enhances readability and helps visually differentiate each month.
* Data Labels: The exact count of orders for each month is displayed, making the chart informative.

**Visualization:**

****

**Insights:**

* Lowest orders: February (324 orders)
* Highest orders: November (1,436 orders), followed by December (1,407)
* Gradual increase: Orders generally increase from March through the end of the year, with a notable spike in September, November, and December.

**Reason for this chart type:**  I chose a bar chart to represent the monthly trend in the number of orders shipped because it effectively visualizes changes over time. Since the x-axis represents **months (categorical variable)** and the y-axis shows the **count of orders (quantitative variable)**, a bar chart clearly highlights variations from one month to another. It allows easy comparison, helps identify peak months (like **November and December**), and communicates seasonal trends in shipping volume with clarity and impact.

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

**Answer:** To effectively analyse and visualize how different customer segments perform in terms of total average sales and average discount rates, we use a side-by-side column chart (also called a dual-pane bar chart). This allows us to compare two related but distinct metrics side-by-side across the same dimension Customer Segment.

**Customer Segments vs. Average Sales**

1. **X-axis**: Customer Segments (Consumer, Corporate, Home Office)
2. **Y-axis**: Average Sales
3. Each bar represents the average sales per order for a given segment.
4. **Labels** are added on top of each bar to show exact average sales values for clarity and easy interpretation.

**Customer Segments vs. Average Discounts**

1. **X-axis**: Customer Segments
2. **Y-axis**: Average Discount Rate
3. Each bar displays the average discount percentage given to each customer segment.

**Visualization:**

**A graph of numbers and a number of numbers

AI-generated content may be incorrect.**

**Insights:**

* Home Office segment has the highest average sales: With an average sale of 240.97, the Home Office segment outperforms Corporate (233.82) and Consumer (223.73) in sales value.
* Consumer and Corporate segments receive higher average discounts: Both Consumer (0.15814) and Corporate (0.15823) segments receive slightly higher average discounts compared to Home Office (0.14713).
* Discount rates are relatively similar across all segments: The difference in discount rates between segments is minimal (~0.01), suggesting discount strategy is uniform.
* Higher discounts don’t necessarily lead to higher sales: Despite receiving lower discounts, the Home Office segment generates higher average sales, indicating other factors (e.g., product types or purchasing behaviour) might influence revenue more than discounts.
* Corporate segment balances discounts and sales well: With a moderate discount rate and second-highest sales, the corporate segment appears efficient in generating strong sales without excessive discounting.

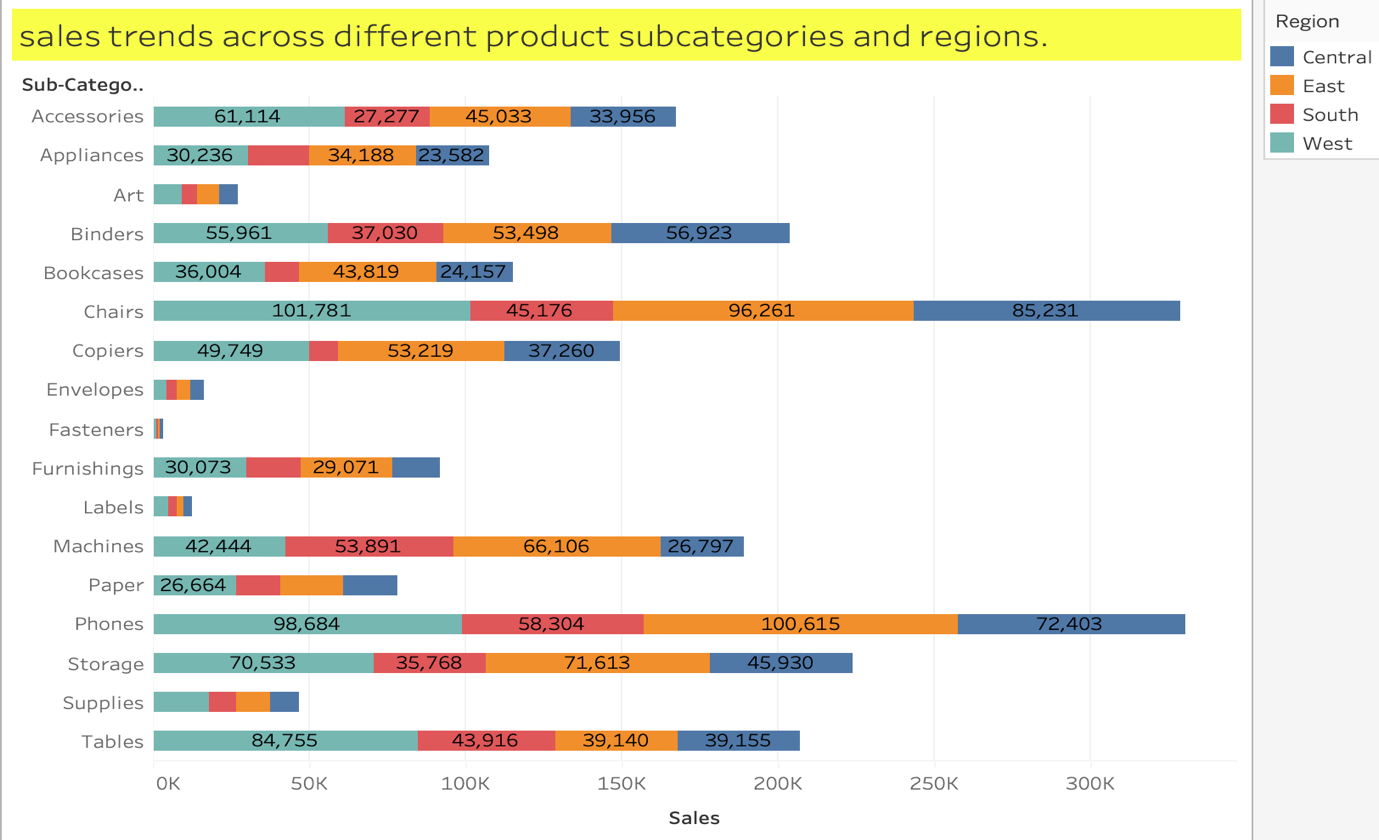
**Reason for this chart type:** We have used two separate column charts to display the distribution of **Average Sales** and **Average Profit** across different customer segments. This is because our goal is to highlight how each segment performs individually in terms of sales and profitability. Since we are comparing a **categorical variable** (Customer Segment) with **continuous measures** (Sales and Profit), bar charts are the most effective and intuitive choice. They clearly show differences in magnitude, making it easier for the audience to interpret and derive insights from the data.

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

**Answer:** This stacked horizontal bar chart effectively illustrates the sales trends for different product sub-categories across various regions (Central, East, South, West). Each bar represents a specific sub-category, and the colored segments within each bar show the sales contribution from each region. The length of the entire bar indicates the total sales for that sub-category. This visualization allows for easy comparison of regional sales performance within each product sub-category and highlights the overall best and worst-performing sub-categories.

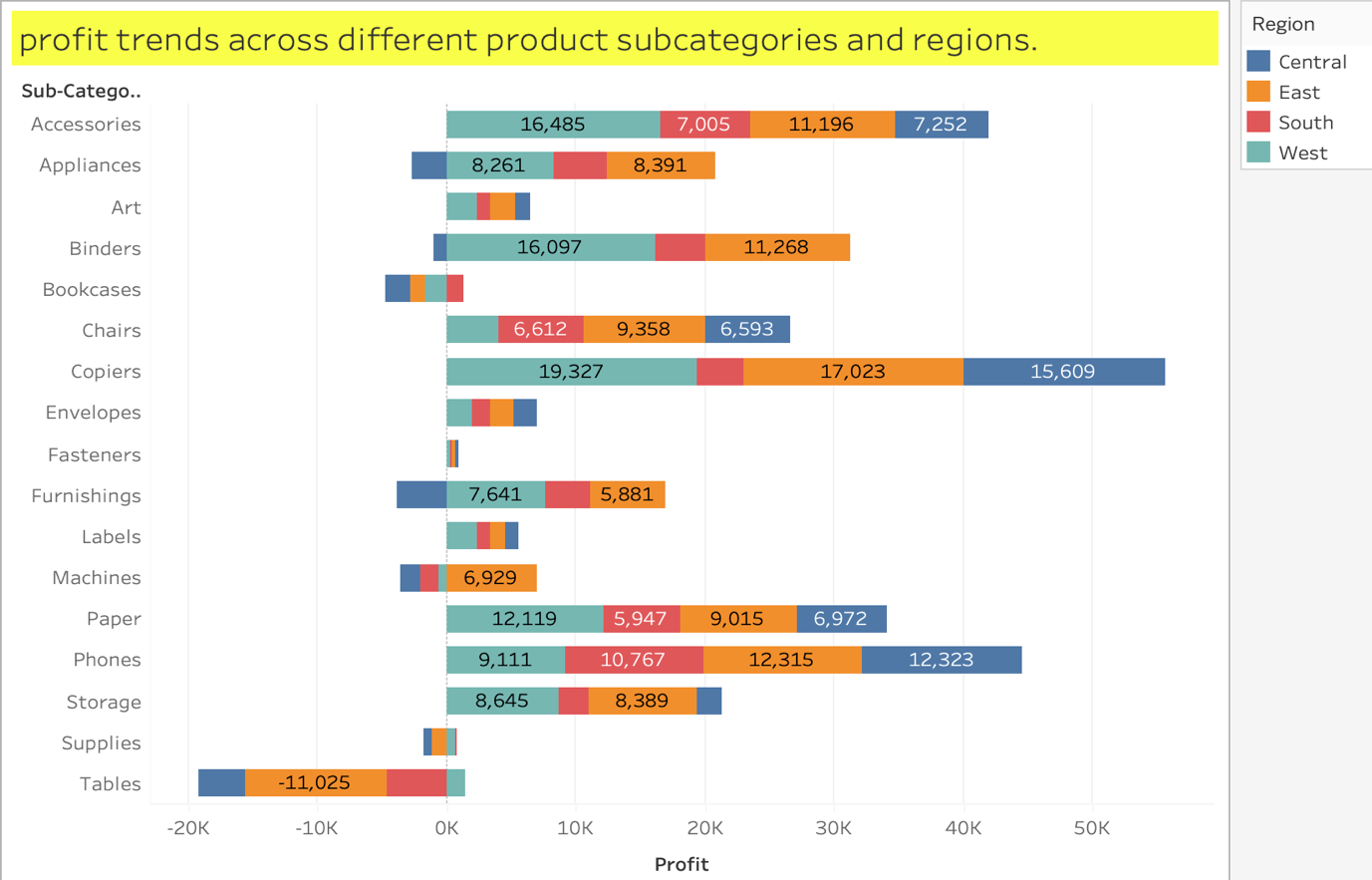
**Visualizations:**

Distribution of Product sub-category vs Total sales:

****

**Insights :**

* + **Chairs and Phones are the top-performing sub-categories overall**, with the **West and East regions contributing the most** to their sales. Chairs alone surpassed 300K in combined regional sales.
  + **The West region consistently dominates sales across most high-performing sub-categories** such as Phones, Chairs, and Tables — indicating stronger demand or market size.
  + **Art, Fasteners, Envelopes, and Labels have the lowest sales** across all regions, suggesting limited popularity or niche usage.
  + **Copiers and Machines show strong performance in the East region**, particularly Machines, where East contributed over **65K** in sales — more than any other region for that item.
  + **Storage and Binders have a balanced demand across all four regions**, making them reliable performers without being heavily skewed toward a single region.



**Insights:**

**Copiers Lead in Profitability**:

* *Total Profit:* West (19,327) + South (17,023) + East (15,609) = **51,959**
* This is the highest among all subcategories, making Copiers the top-performing subcategory across all regions.

**Tables Show Significant Losses**:

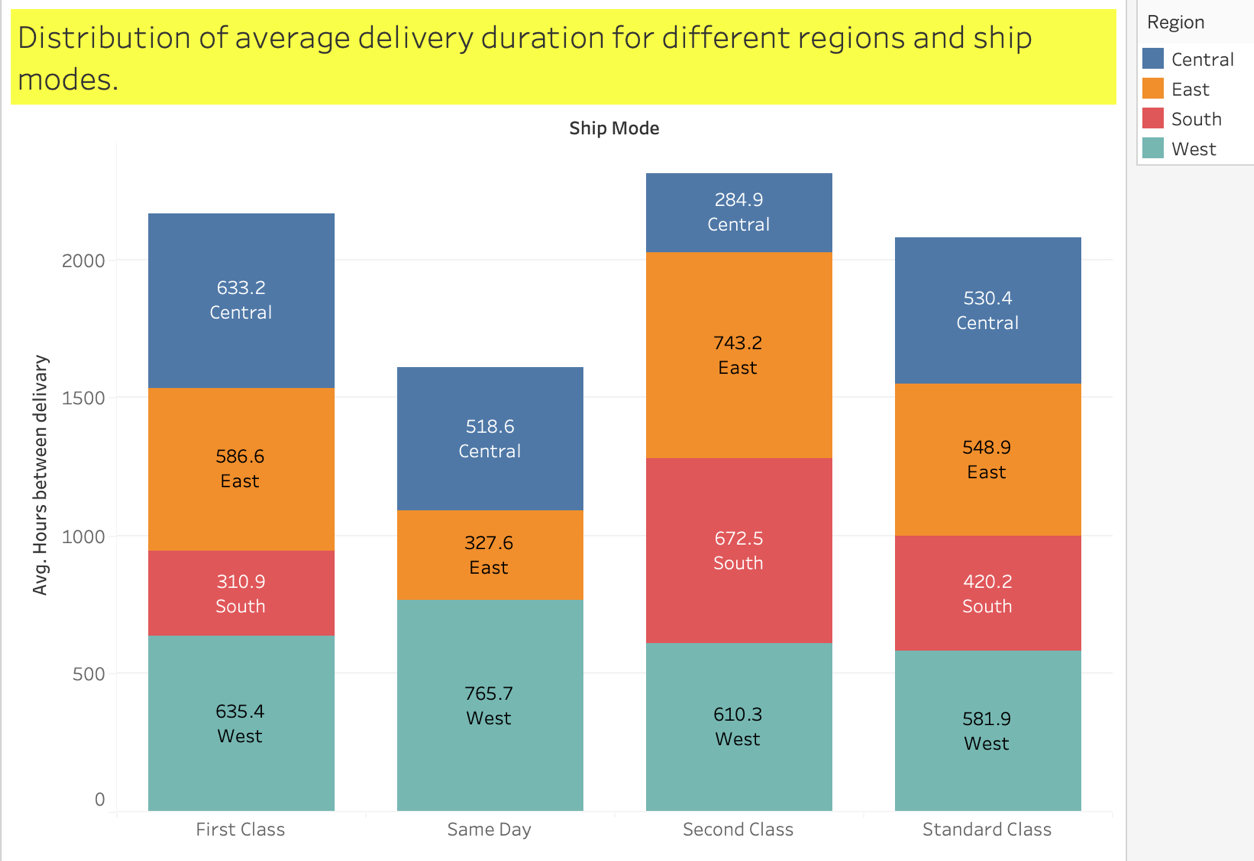
* *Total Loss:* East (-11,025) + South (negative) + Central (negative)
* Tables are the only subcategory with **consistent and large negative profits** across almost all regions.
* **Phones Generate High Profits Across All Regions**: *Notable Contributions:* East (12,315), West (9,111), South (10,767), Central (12,323).
* Phones are among the top profit earners and show a **balanced** positive trend across all four regions.
* **Accessories and Binders Are Strong in the West**: Accessories (West: 16,485), Binders (West: 16,097). These subcategories demonstrate **strong performance in the West**, contributing significantly to total profits.
* **Art, Labels, and Fasteners Have Low Impact**: These subcategories contribute **marginal profits** (below ~5,000 per region) and may be considered lower-priority for profit-driven strategies.

**Reason for this chart type:** I chose stacked bar charts to compare profits and sales across subcategories while also showing regional contributions. This format clearly highlights both total values and regional patterns in one view.

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

**Answer:** For 'First Class' shipments, the average delivery duration is highest in the 'Central' region (around 633 hours) and lowest in the 'South' region (around 311 hours). Similarly, 'Same Day' shipments show the longest average delivery time in the 'West' region (around 766 hours).

**Visualization:**

****

**Insights:**

* **Slowest Mode**: Second Class has the highest average delivery times, especially in East (743.2 hrs).
* **Same Day Delay**: West has unusually high Same Day delivery time (765.7 hrs).
* **First Class Stable**: First Class times are consistent across all regions (~586–635 hrs).
* **Standard Class Efficient**: Standard Class outperforms faster modes in South and West.

**Reason for this chart type:**  I chose a stacked bar chart to compare average delivery durations across ship modes while showing regional contributions clearly. It efficiently highlights differences and anomalies across multiple dimensions in one view.

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

**Answer:** The average order quantity for Furniture and Office Supplies generally decreased over the years shown. In contrast, the average order quantity for Technology fluctuated but showed a significant increase in 2018.

**Visualization:**

****

**Insights:**

* **Technology category** shows a sharp rise in average order quantity in 2018, peaking at **4.33 units**.
* **Furniture** experienced a steady decline in average quantity after 2016, dropping to **3.42 units** in 2018.
* **Office Supplies** remained relatively stable until 2017, then saw a noticeable dip in 2018.
* In **2015**, Furniture had the highest average quantity among all categories.
* From **2016 to 2018**, only the Technology category showed consistent growth in order quantity.

**Reason for this chart:** A line chart is indeed the most effective way to visualise the trend of a continuous variable over time. In this case, it clearly illustrates how the average order quantity for each product category has evolved across the years.

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

**Answer:** Yes we can visualise by following below steps.

* Drag **Discount** to **Columns**.
* Drag **Quantity** to **Rows**.
* Set the **Marks type** to Circle
* Drag **Segment** to the **Columns shelf** → this will create one plot per segment.
* we’ll now see 3 scatter plots side by side (Consumer, Corporate, Home Office)
* Drag Segment to the **Colour shelf** on the Marks card
* Drag **Trend Line** into the view.

**Visualization:**

**A graph with numbers and a line

AI-generated content may be incorrect.**

**Insights:**

* **No strong correlation exists** between Discount and Quantity across all segments: The trend lines are mostly flat, indicating discounts don’t consistently drive higher order volumes.
* **Consumer segment shows a mild positive trend:** Slight increase in quantity with higher discounts, but the effect is weak.
* **Corporate segment shows scattered behaviour:** Quantity ordered doesn’t seem influenced by discount levels at all.
* **Home Office purchases cluster around fixed discount levels:** Common discounts (0%, 20%, 50%) are frequent, but don’t correspond to larger quantities.
* **High discounts don’t lead to large quantities:** Even with 40–70% discounts, order quantities often remain low.

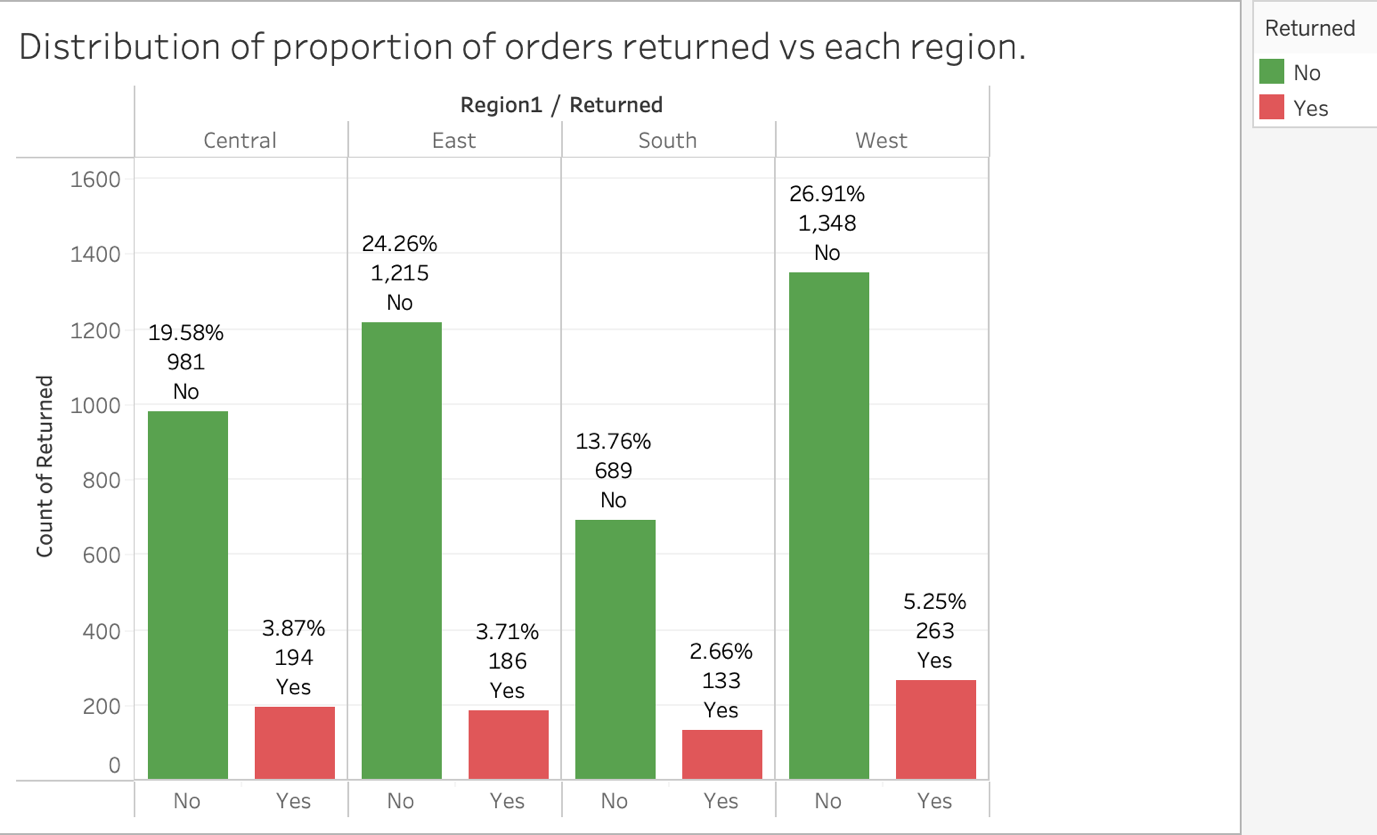
**Reason for this chart:**  The scatter chart with a trend line is particularly effective for illustrating the relationship between average discounts and average quantities across various segments that’s why we selected the scatter chart with trend line.

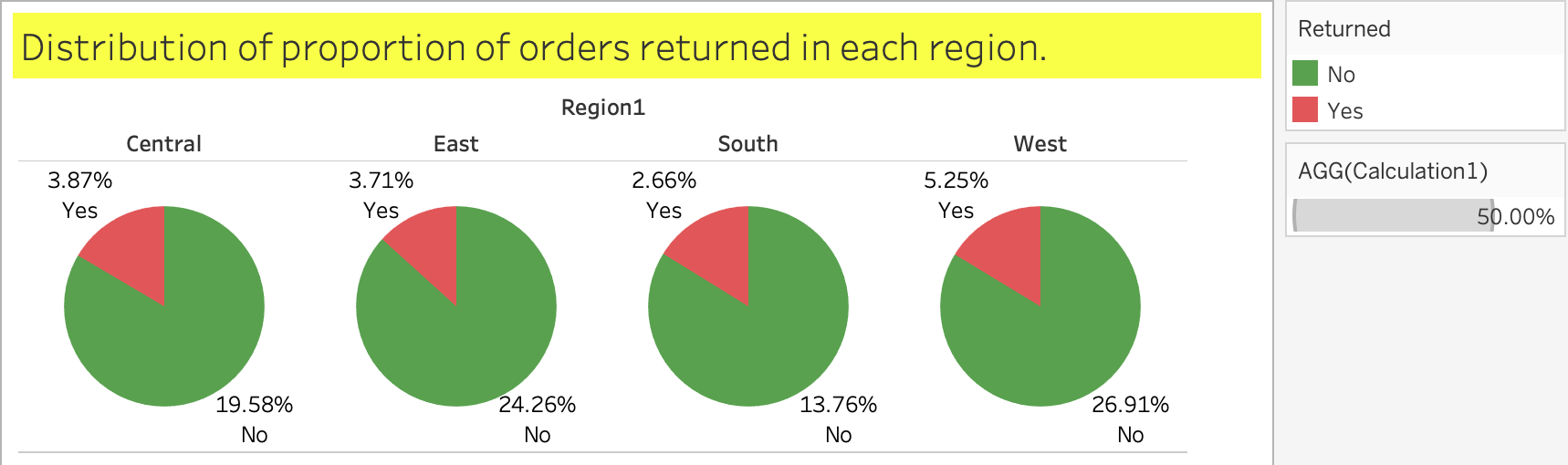
23. What is the proportion of orders returned in each region within the Superstore dataset?.

**Answer:** To analyse order returns, we noted the Superstore dataset lacked a return status column. We imported an external "Returns" dataset and joined it with the main dataset using Order ID. A calculated field was created to identify returned vs. non-returned orders. Based on the analysis, the return proportions are:

Central: 3.87%, East: 3.71%, South: 2.66%, West: 5.25%.

**Visualization:**

****



**Insights:**

* **West has the highest return rate**, with 5.25% of orders returned, making it a region of concern for product issues or customer dissatisfaction.
* **South has the lowest return rate**, at just 2.66%, indicating relatively fewer post-sale issues.
* **All regions show a consistent trend** where non-returned orders significantly outnumber returned orders.
* **East and Central regions have similar return rates**, 3.71% and 3.87% respectively, showing moderate return behaviour.
* **The bar chart provides better clarity** on both proportions and absolute counts, while the pie chart quickly highlights the return rate share in each region.

**Reason for this chart:**

* Pie charts clearly show the proportion of returned vs. non-returned orders in each region, making it easy to compare return rates visually.
* Bar charts display both percentages and actual counts, helping to understand the scale and impact of returns across regions.
* Using both ensures a complete view, combining relative comparison (pie) with quantitative insight (bar) for better decision-making.

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

**Answer:** This chart visualizes the average profit across various sub-categories for the top five products.

**Visualization:**

A screenshot of a graph

AI-generated content may be incorrect.

**Insights:**

* **Profit Dominance by Subcategory:** The **Copiers** subcategory shows a significantly higher average profit compared to **Machines**.
* **Visual Profit Comparison:** The bar representing **Copiers** is substantially taller than those for **Machines**, indicating much greater profitability.
* **Product Performance:** Specific products within the **Copiers** subcategory likely contribute more to overall profit, given its prominence in the average profit representation.
* **Variation in Profit Across Products:** The **Machines** subcategory displays variation in profit levels, but none reach the high profit levels seen in the **Copiers** category.
* **Potential for Growth in Machines:** Given the lower average profit in **Machines**, there may be opportunities for enhancing profitability through better marketing strategies or product adjustments.

**Reason for this chart:**  Here we have selected the bar chart to show the distribution of Average profit vs Different -different sub-categories for top 5 products because bar chart is best way to represent categorical variable with continuous.

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

**Answer:** To identify the most commonly used shipping mode by customers, we first visualize all shipping modes on the X-axis and plot the count of their usage on the Y-axis. A column chart is used to represent this distribution clearly.

**Visualization:**

**A graph with different colored squares

AI-generated content may be incorrect.**

**Insights:**

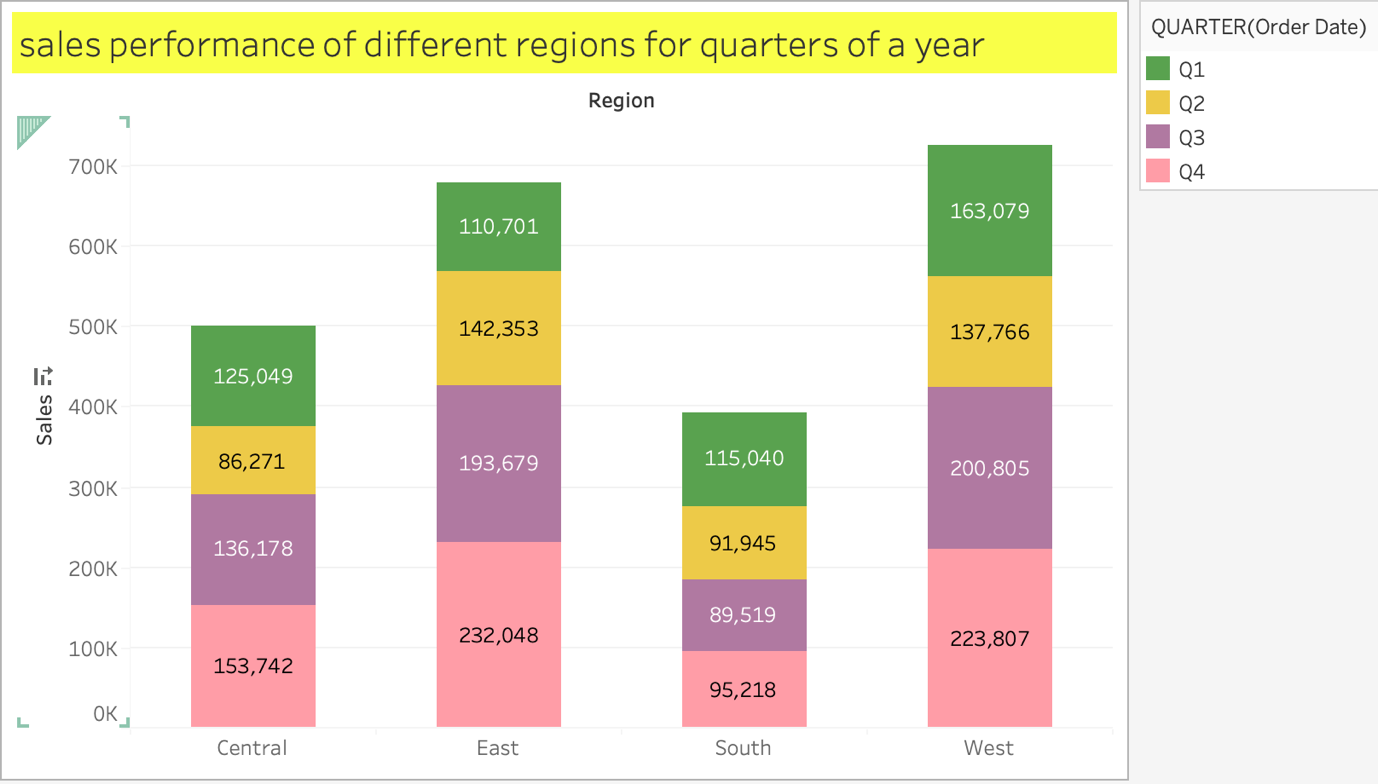
* **Standard Class is the Most Used Shipping Mode**  
  – With **5,968 orders**, *Standard Class* is by far the most frequently chosen shipping option, indicating it’s likely the default or most cost-effective method.
* **Same Day Shipping is Least Used**  
  – *Same Day* shipping has the lowest usage at **543 orders**, suggesting it may be less popular due to higher cost or limited availability.
* **Mid-Range Preference for Second and First Class**  
  – *Second Class* (1,945 orders) and *First Class* (1,538 orders) show moderate usage, likely preferred in certain scenarios needing faster delivery but not at the cost of Same Day shipping.
* **Significant Gap Between Standard and Others**  
  – The number of Standard Class orders is **over 3 times** higher than any other mode, emphasizing its dominance.
* **Customer Behaviour Suggests Cost > Speed**  
  – The overall trend suggests that customers prioritize *cost-effectiveness* over *delivery speed*, given the low usage of faster (and likely more expensive) options.

**Reason for this chart:** To determine the most commonly used shipping mode by customers, it is appropriate to use a bar chart for visualization. Since Shipping Mode is a categorical variable and the number of times it is used is a continuous (quantitative) variable, a bar or column chart effectively displays this relationship. This type of chart allows for a clear comparison across categories, making it easier to identify which shipping mode is used most frequently.

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

**Answer:** Sales performance varies across quarters in all regions, with Q4 generally showing the highest sales.  
The West region leads consistently across quarters, peaking in Q4.  
The East also performs strongly, especially in Q3 and Q4.  
The South has the lowest sales overall, with modest growth in Q1 and Q2.

**Visualization:**

****

**Insights:**

 **West Region**: This region consistently outperforms all others across all quarters. It sees its **highest sales in Q4**, indicating a strong year-end boost—possibly due to seasonal demand or holiday shopping.

 **East Region**: The East follows closely behind the West, with **Q4 and Q3 being its strongest quarters**. It shows steady growth throughout the year, suggesting consistent customer engagement.

 **Central Region**: The Central region shows moderate performance. Sales gradually increase across the quarters, with **Q4 being the peak** and **Q2 being the lowest**. This could reflect mid-year slumps or regional buying patterns.

 **South Region**: The South lags behind the other regions in total sales. Though it shows relatively stable numbers, it has the **lowest figures in Q3**, and only a slight improvement in Q4, suggesting limited growth potential or market size.

In summary, Q4 is the best-performing quarter across all regions, likely due to end-of-year purchasing behaviour. The West and East regions dominate, while the South presents the weakest overall performance.

**Reason for this chart:** The chart shows how sales vary by quarter across four regions using a stacked bar format.It clearly compares total and quarterly sales within and between regions.Color-coded segments and labels make quarterly performance easy to understand.

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

**Answer:**

**Low Priority Orders Dominate:**

* Across all three categories (Furniture, Office Supplies, Technology), **low priority** orders have the highest count.
* **Office Supplies** has the highest number of low priority orders at **5,727**, followed by **Furniture (1,654)** and **Technology (1,451)**.

**High Priority Orders Are Least Frequent:** High priority orders are consistently low across categories:

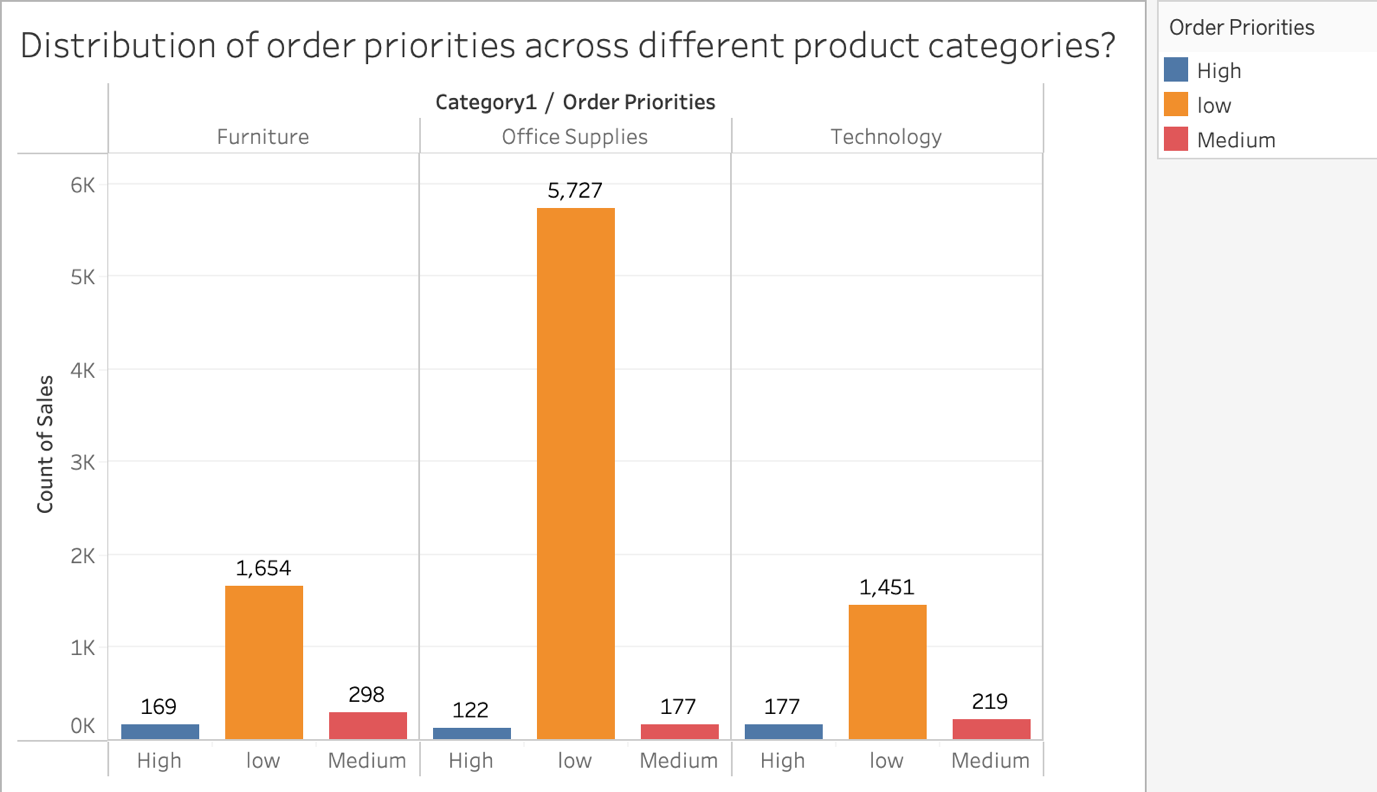
* **Furniture:** 169
* **Office Supplies:** 122
* **Technology:** 177

**Office Supplies Category Is the Most Ordered Overall:** It has the highest total count of sales, driven mainly by the huge number of **low priority** orders.

**Medium Priority Orders Are Moderate:**

* Technology: 219
* Furniture: 298
* Office Supplies: 177
* This priority level is neither dominant nor insignificant.

**Visualization:**

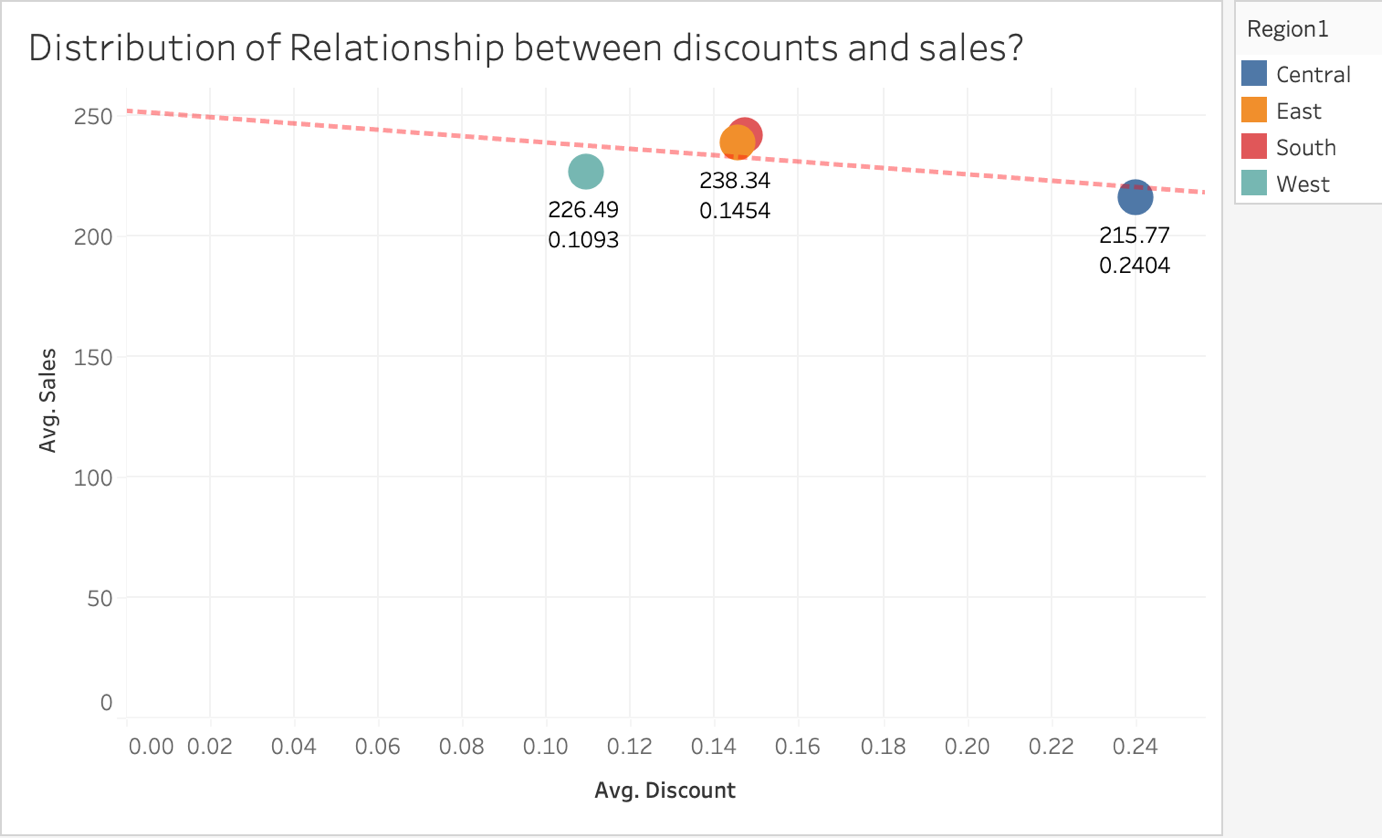
****

**Reason for this chart:** It helps identify which categories require faster fulfilment (if high-priority orders are high) or if most demand is regular/low urgency — supporting inventory planning, logistics optimization, and customer service strategies.

28. What is the relationship between discounts and sales?

**Answer: Negative Correlation Between Discount and Sales:** The dotted red trend line slopes downward, indicating a **negative correlation** — as average discounts increase, average sales tend to **decrease**.

**Visualization:**

****

**Insights:**

1. **Negative Correlation Between Discount and Sales:**
   * The dotted red trend line slopes downward, indicating a **negative correlation** as average discounts increase, average sales tend to **decrease**.
2. **Regional Insights:**
   * **East & South**:
     + Similar average discounts (~0.1454) and highest average sales (~238.34).
     + These regions appear more balanced in discount vs. sales performance.
   * **West**: Lowest average discount (~0.1093) but decent sales (~226.49), suggesting sales are achieved with minimal discounting.
   * **Central**: Highest average discount (~0.2404) but lowest average sales (~215.77), indicating discounts are **not effectively driving higher sales**.

**Reason for this chart:** For effectively distribution we selected theBubble chart with a trend line, showing average **discounts** (X-axis) vs average **sales** (Y-axis) across four regions (Central, East, South, West).

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

**Answer:** Based on the chart, **new customers have a higher average order value** (AOV) compared to repeat customers.

* **New Customers:** Avg. Sales = **248.07**
* **Repeat Customers:** Avg. Sales = **226.28**

This indicates that **first-time buyers tend to spend more per order** than those who have made multiple purchases.

**Visualization:**

**A graph of a number of people

AI-generated content may be incorrect.**

**Reason For this chart:** A **column chart** is perfect for this visualization because it clearly compares the **average order values** between two distinct categories — new and repeat customers — using straightforward vertical bars. This format makes it easy to visually assess the difference in magnitude at a glance.

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

**Answer:**

 **West region** has the highest average profit at **₹33.85**, with a return rate of **49.19%**.

 **East region** follows with an average profit of **₹32.14** and a return rate of **49.12%**.

 **South region** shows **₹28.86** average profit and a slightly higher return rate of **50.74%**.

 **Central region** has the lowest profit at **₹17.09**, despite a return rate close to others at **50.58%**.

 The chart shows that **higher return rates do not always reduce profitability**, as seen in the West.

**Visualization:**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**Insights:**

* **High Returns ≠ Low Profitability**: Despite having nearly **50% return rates**, regions like **West (₹33.85)** and **East (₹32.14)** still achieve the **highest average profits**, indicating strong underlying sales performance or high-margin products.
* **Central Region Struggles**: The **Central region** shows the **lowest average profit (₹17.09)** even though its return rate (**50.58%**) is similar to other regions—highlighting a potential issue with operational efficiency or product mix.
* **West is Leading**: **West region** stands out as the most profitable, showing that high return rates are not necessarily detrimental if balanced with high-value sales.
* **Return Rates Are Consistent**: All regions exhibit **return rates between ~49% and 51%**, suggesting returns are a **universal challenge** across geographies rather than region-specific.
* **Profit Gaps Are Significant**: There's a noticeable **₹16.76 gap between the most and least profitable regions**, pointing to a need for strategic attention in underperforming areas like **Central**.

**Reason For this Chart:**

* **A circle (scatter plot) chart is ideal for visualizing the relationship between two continuous measures**—in this case, *Return Rate* (x-axis) and *Average Profit* (y-axis). Each circle represents a region, and using different colours helps compare multiple categories (regions) in one view.
* It allows quick visual detection of trends, which would be difficult with bar charts or line graphs.