

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!

DRIVE LINK [click here](#)

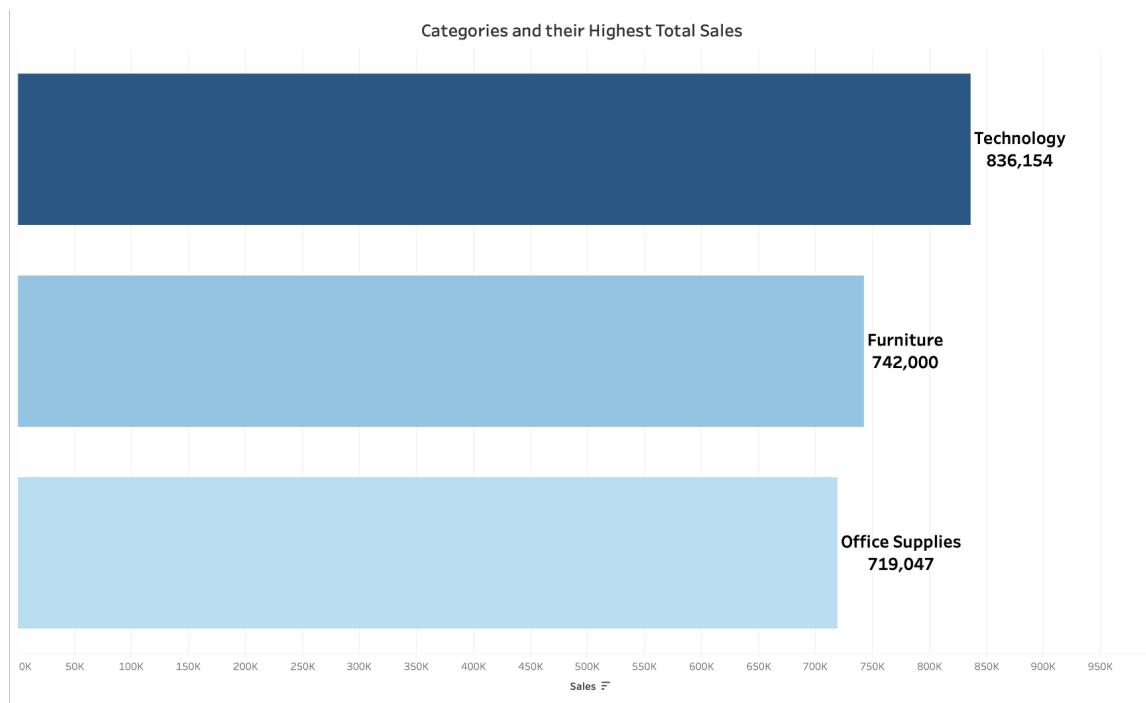
GITHUB LINK [click here](#)

Questions:

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

Ans.

Chart: Bar Chart



Why Bar Chart?

It is ideal for comparing total sales across categories. It allows for easy comparison of values and identifies the categories with the highest sales.

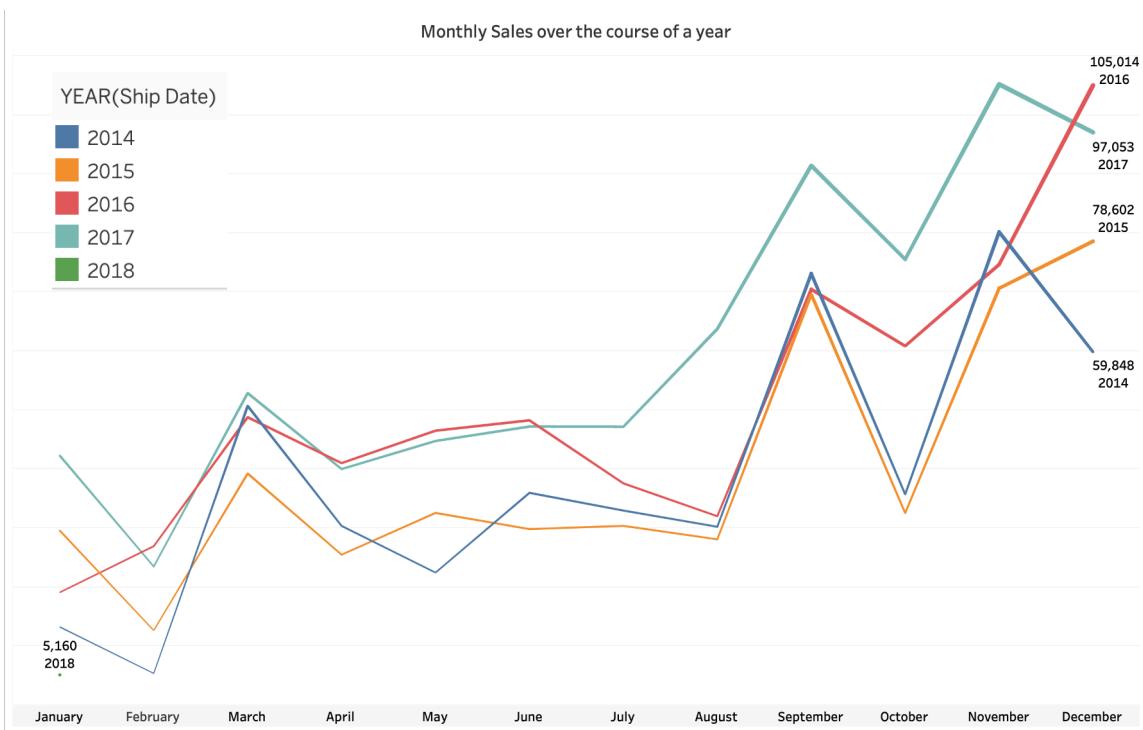
Insight:

From the Graph it is clear that **Technology** is the **highest total sales** category with the value of **8,36,154/-** among other categories.

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

Ans.

Chart: Line Chart



Why line chart?

It Effectively shows changes in **monthly sales over time**. It's easy to observe trends, seasonal trends and variations throughout the year.

Insight:

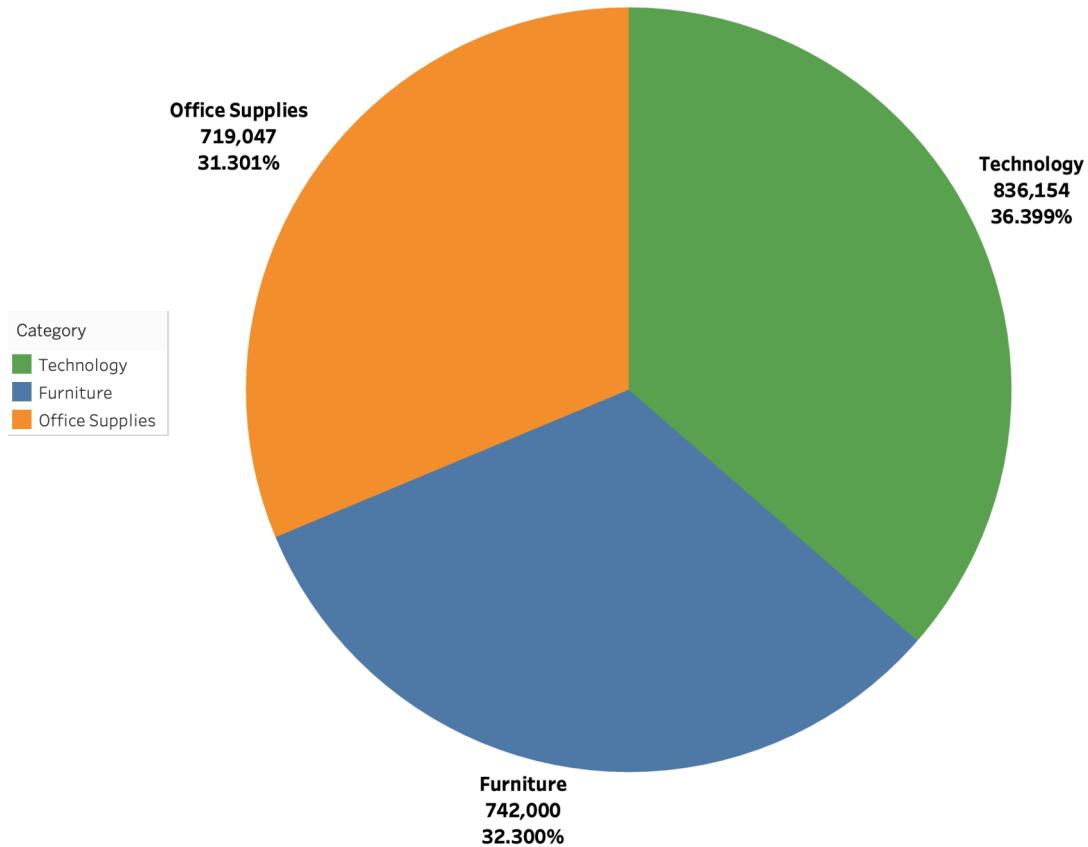
From the multiple line chart, we gain valuable insights into the monthly sales trends throughout the year. We observe a steady increase in sales from **2014 (69,546)** to **2016 (96,999)**. In **2017**, sales saw a significant rise up to **November(118,448)** but experienced a sudden decline by the end of **December(83,829)**.

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

Ans.

Chart: Pie Chart

Sales distribution among different Categories



Why pie chart?

It visually represents the distribution of Sales among different categories. It's useful for showing proportions and the contribution of each category to total Sales.

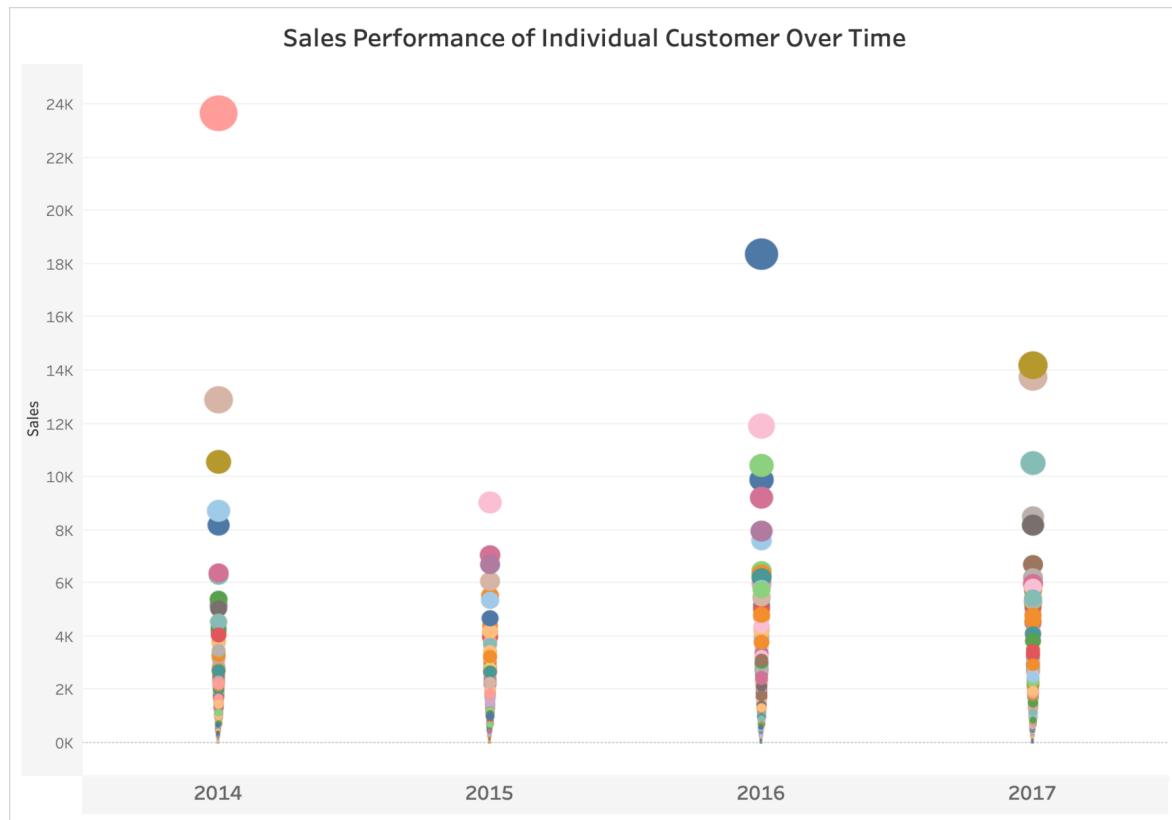
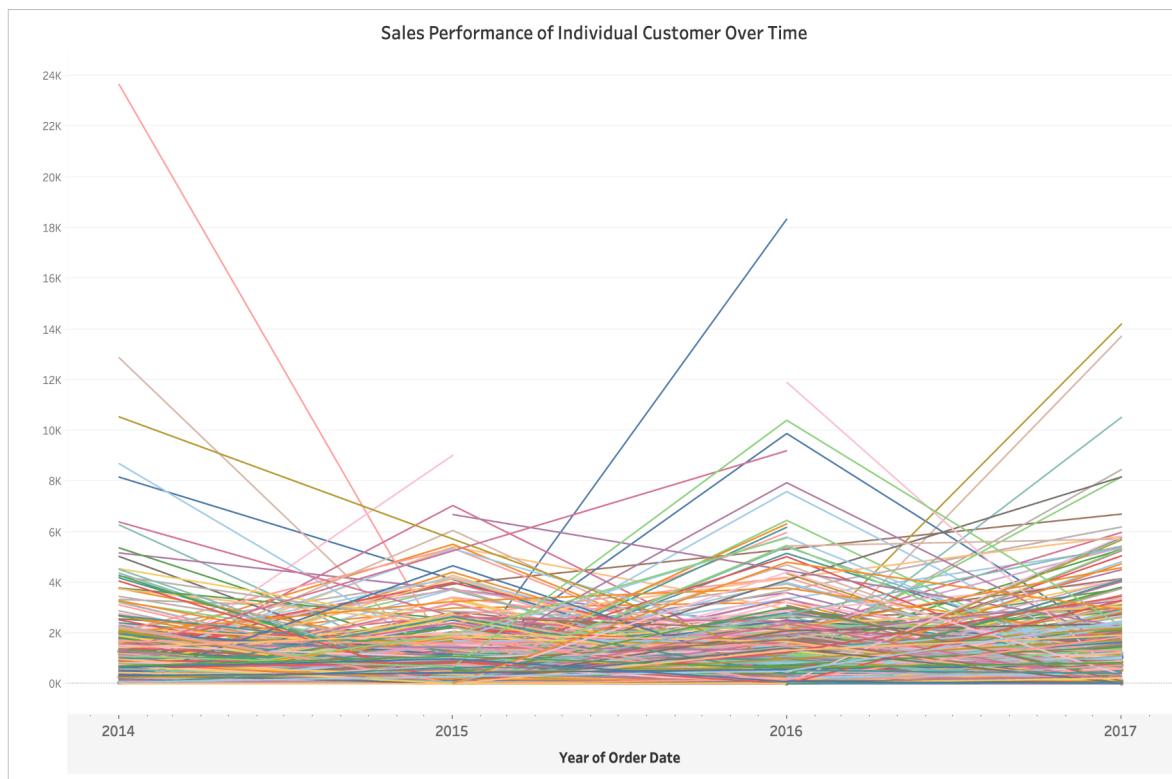
Insight:

From the graph, All the **three categories show almost equal contribution** to total sales. Starting from lowest contributor, **Office Supplies** - contributed **31.30%**, **Furniture** - contributed **32.30%** and **Technology** - contributed **36.40%** which is higher than the other two but still all three lie around the near percentage value.

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

Ans.

Chart: Line Chart or Scatter Plot



Why Line or Scatter chart?

Line chart can show sales trends for individual customers over time, while a scatter plot can show sales volume over time with each point representing a customer's purchase.

Insight:

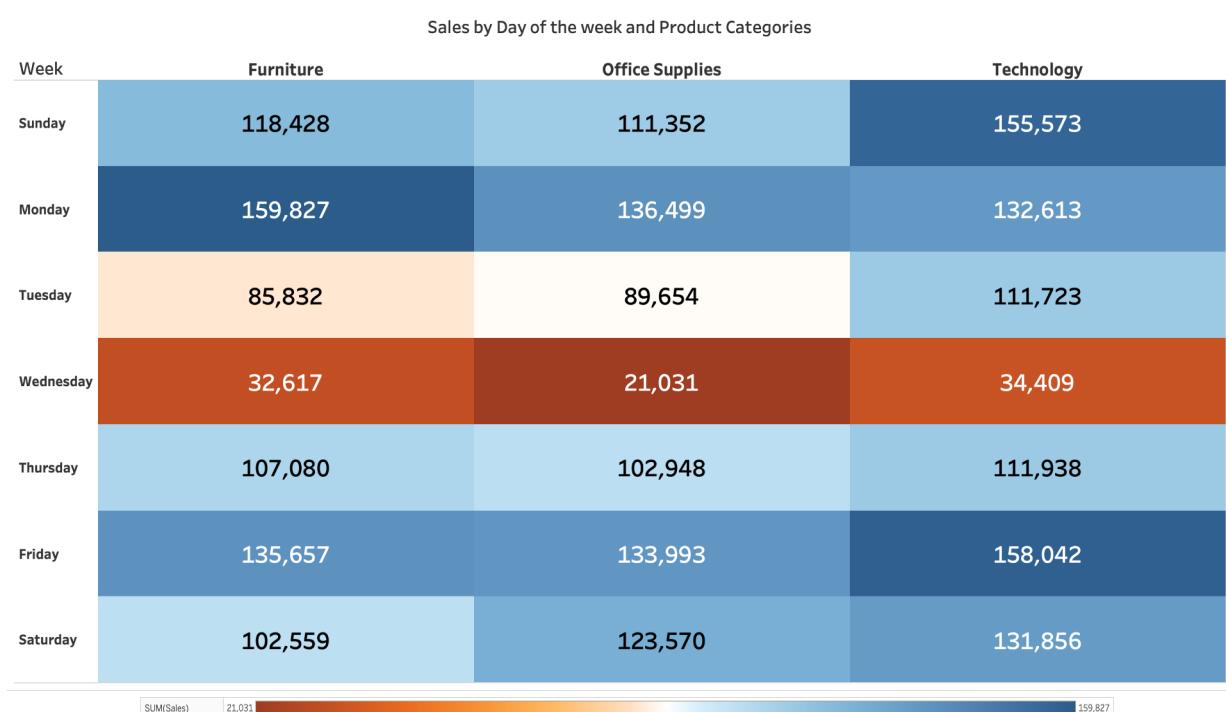
Most of the orders have **seasonality** factors.

Customers base is diverse, With some customer being more active than other others.

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

Ans.

Chart: Heatmap



Why Heatmap chart?

Heatmap is effective for showing variations in sales across days and categories. The color intensity makes it easy to identify high and low sales periods.

Insight:

Stronger Sales on Certain Days: Some product categories might show significantly higher sales on specific days of the week. For example, the "Office Supplies" category has dark blue colors for days (like Monday or Friday), it suggests higher sales for this category during the mid-week period.

Weaker Sales on Certain Days: Conversely, if some product categories show consistently lower sales on certain days (e.g., mid-week (Tuesday and Wednesday)), it indicates potentially lower demand for these products during those days.

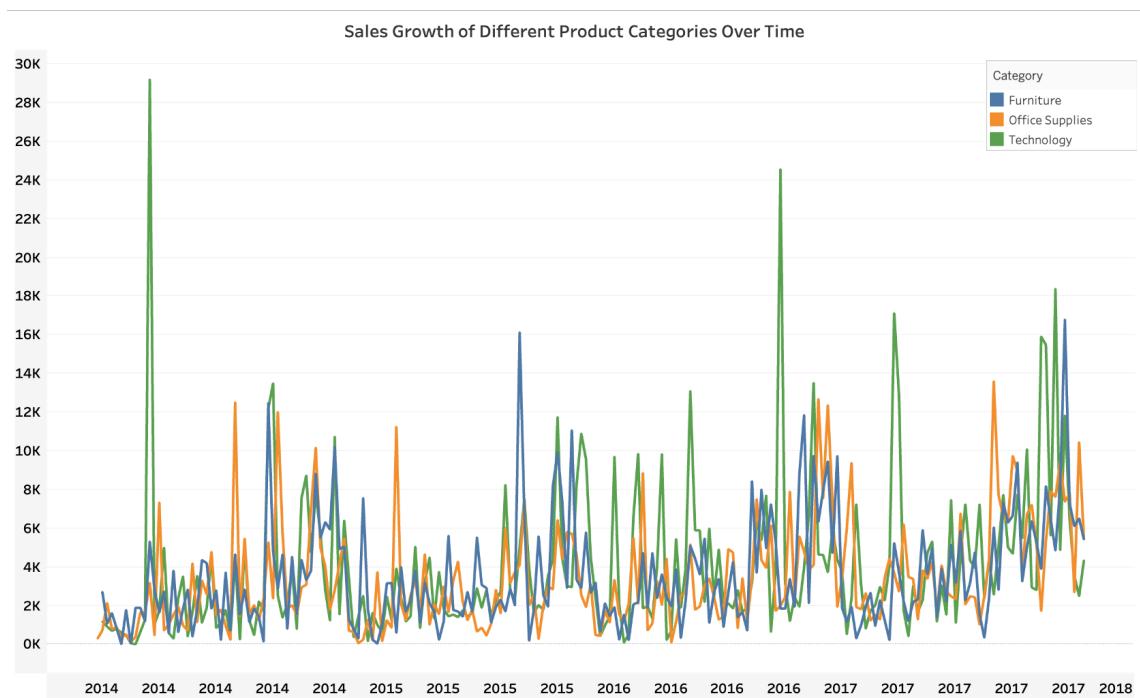
Identifying Patterns: Observing the general color intensity across rows and columns can reveal overall patterns. If, for example, the whole row for "Wednesday" is relatively red color, it implies lower overall sales across all categories during that day of the week.

Category-Specific Trends: It helps to understand whether there are specific categories that perform better or worse on particular days. For instance, the "Technology" category might have higher sales on Mondays as businesses might be more likely to place technology-related orders at the beginning of the week.

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

Ans.

Chart: Line Chart



Why line chart?

It can **display the growth trajectory** of each **category over time**, making it easy to compare trends and identify categories with significant growth.

Insight:

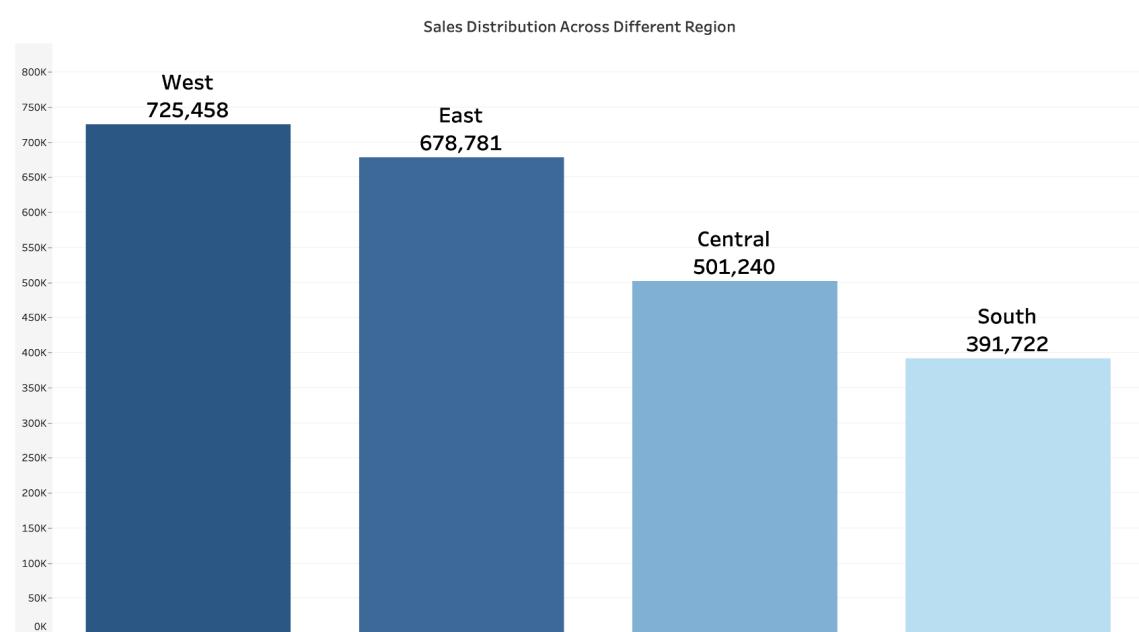
The chart shows that the "**Technology**" category consistently has higher sales and a steeper growth trend compared to "**Furniture**," it suggests that the Technology product line is more successful and potentially has a greater potential for future growth. Also, Thw sales of "**Office Supplies**" show a seasonal spike during back-to-school periods, it could inform future marketing and inventory management strategies for that category.

By analyzing this chart effectively, we can gain a comprehensive understanding of the sales performance of different product categories, identify growth opportunities, optimize inventory levels, and implement targeted marketing campaigns.

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

Ans.

Chart: Bar Chart



Why Bar chart?

It effectively highlights the differences in **total sales across regions**, making it easy to compare values visually. It allows for quick identification of which regions are performing better or worse in terms of sales.

Insight:

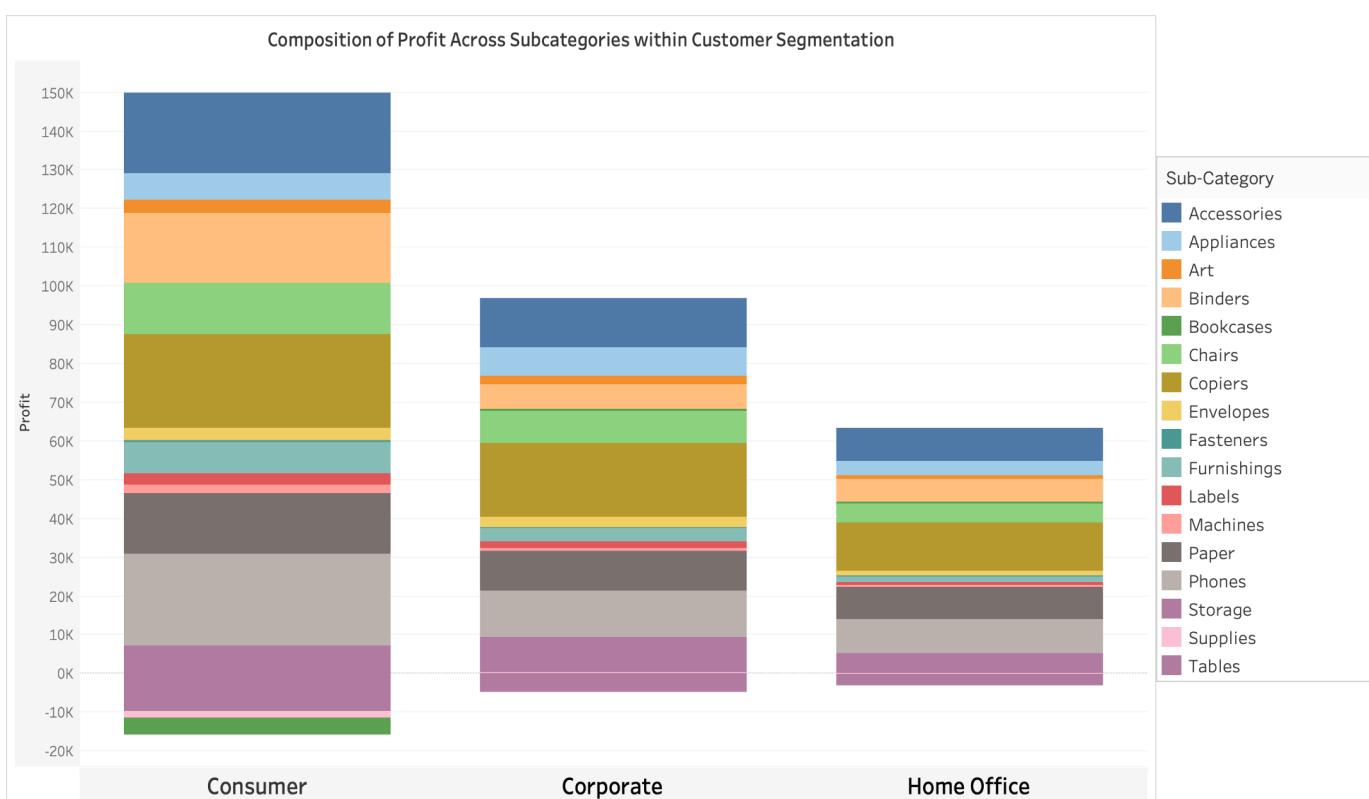
West is doing much better than other region. Might be because of higher population or central state is located in West region.

Region	Sales
Central	501239.89
East	678781.24
South	391721.90
West	725457.82

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

Ans.

Chart: Stacked Bar Chart



Why Stacked Bar chart?

A stacked bar chart shows the composition of profits across subcategories and segments, allowing for easy comparison and assessment of segment profitability.

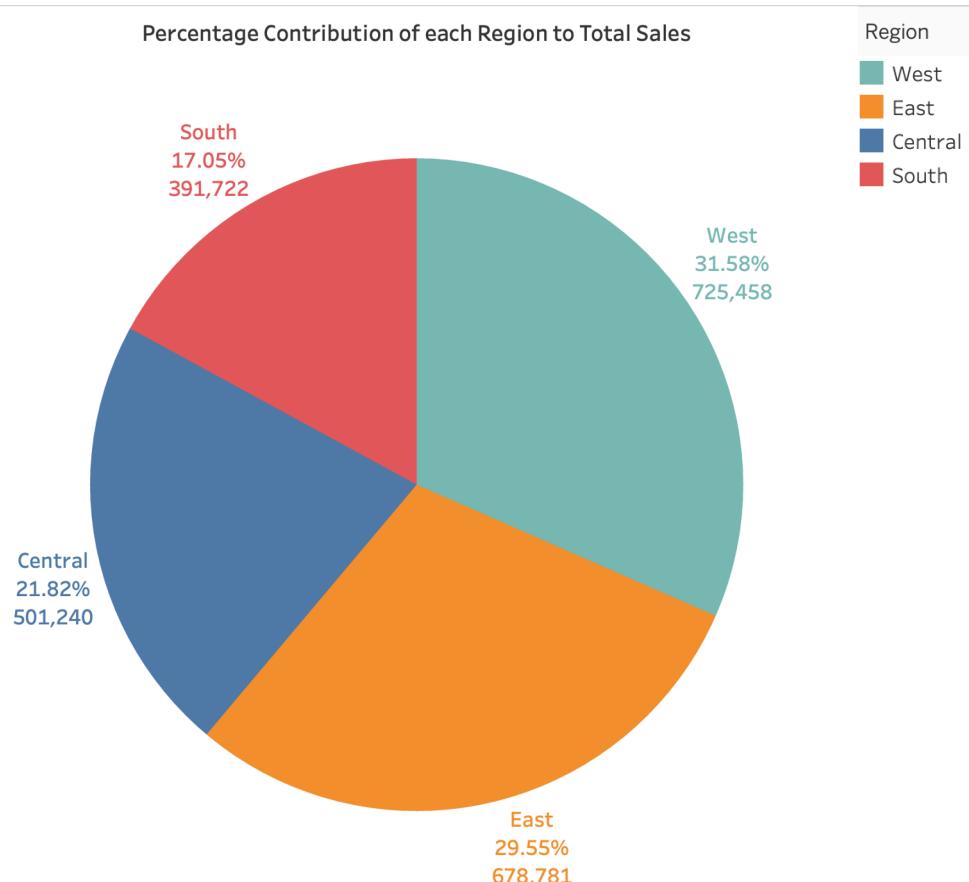
Insight:

The chart clearly highlights which **subcategories** are the **most profitable** for each **customer segment**. You can see which subcategories contribute the largest portion of the total profit for each segment. For instance, you might find that "Copiers" are the most significant profit driver for all segment, while "Phones" are more critical for the "Home Office" segment.

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

Ans.

Chart: Pie Chart



Why pie chart?

These charts effectively depict the contribution of each region to total sales, making it easy to identify the most and least contributing regions

Insight:

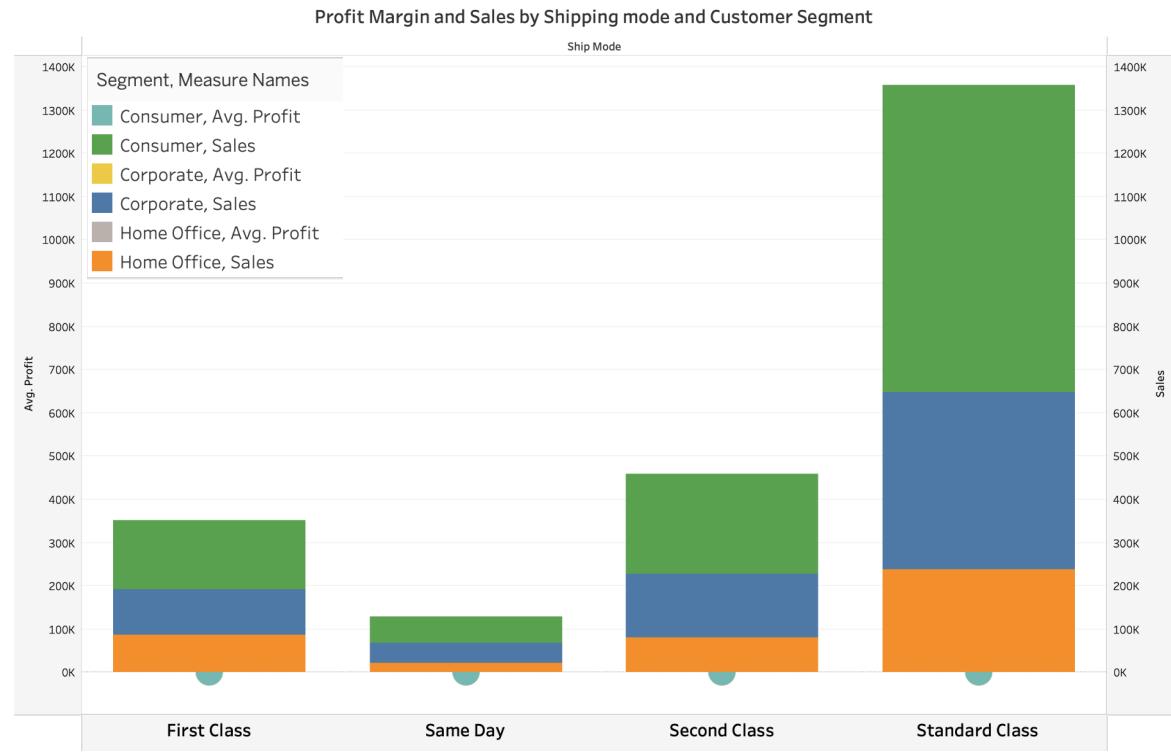
From the graph, It is clear that west has contributed **31.58%** of sales. East has contributed **29.55%** of sales. Central has contributed **21.82%** of sales and South has contributed **17.05%** of sales.

Region	Percent of total Sales
West	31.58%
East	29.55%
Central	21.82%
South	17.05%

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

Ans.

Chart: Bar Chart with Dual Axis



Why bar chart with a dual chart?

It can show profit margins alongside sales figures, allowing for comparison across shipping modes and customer segments.

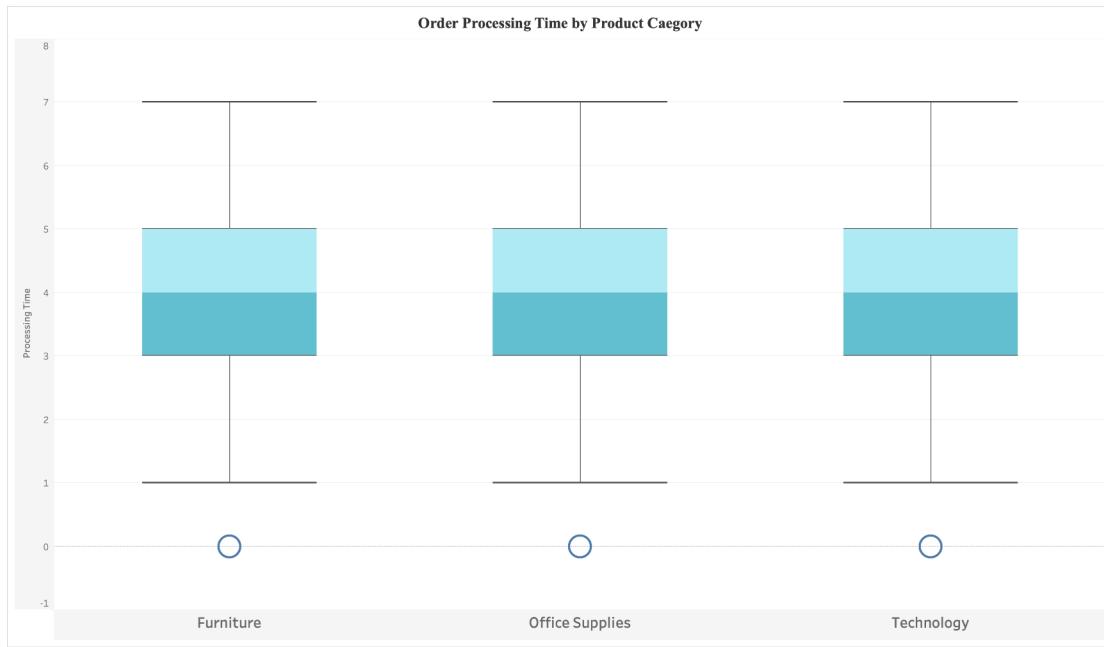
Insight:

The Profit margin of Consumer is more across every shipping mode and in particular the segment of consumer in Standard Class shipping mode is the highest. Meaning that **standard shipping mode** is more preferred than the others.

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

Ans.

Chart: Box Plot



Why Box Plot chart?

A box plot can display the processing time distribution for each category, highlighting the median, quartiles, and any outliers.

access the data from previous cell.

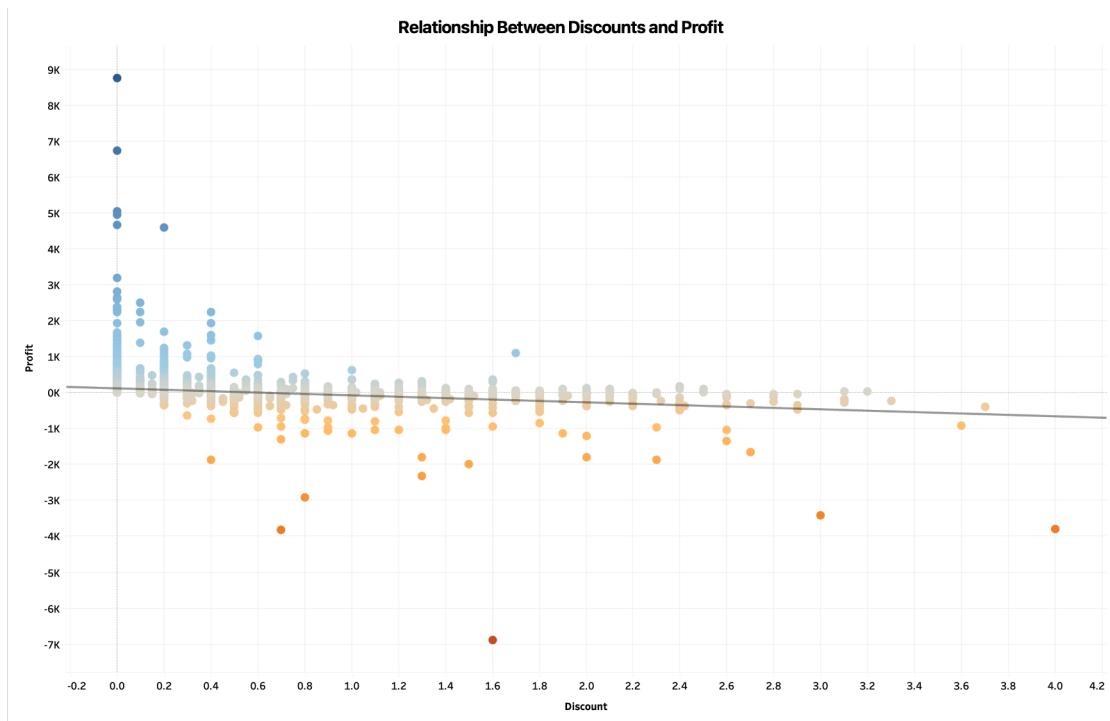
Insight:

From the graph I can see that each categories generally have **same processing time** for order. **0 to 7 represents the week days** and processing time = Ship date – order date. Each box plot whisker has same 0 to 7 range and each categories have Q1, Q2(middle) and Q3 representing the higher whisker density inside box.

12. How do discounts affect overall profit?

Ans.

Chart: Scatter Plot



Why Scatter Plot chart?

A scatter plot can show the relationship between discounts and profit, with each point representing an order. It helps in identifying any trends or negative correlations.

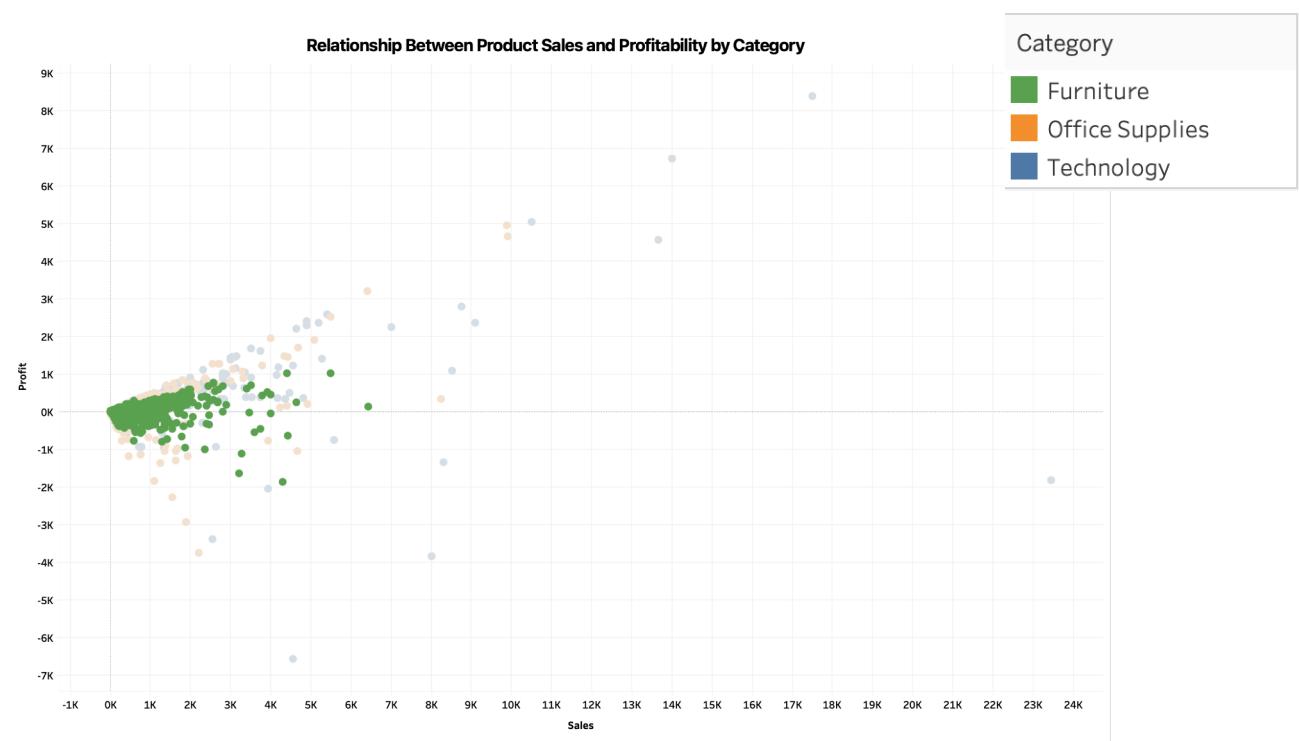
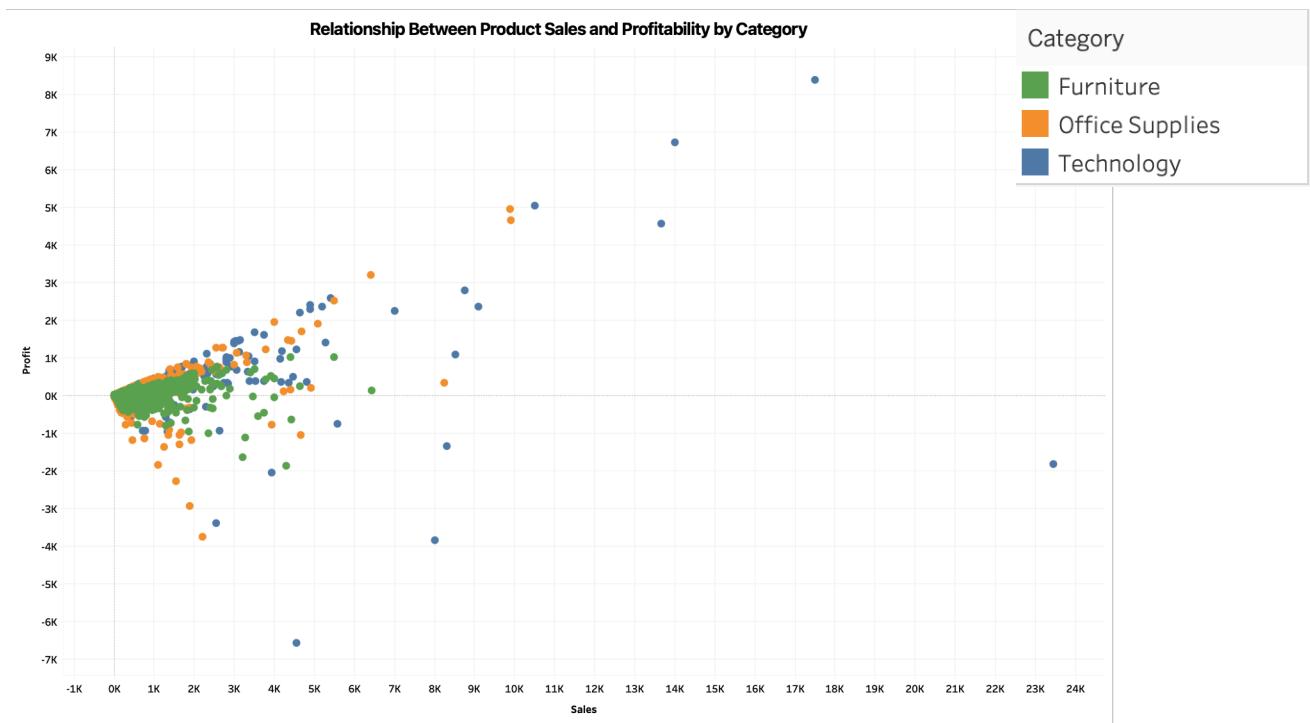
Insight:

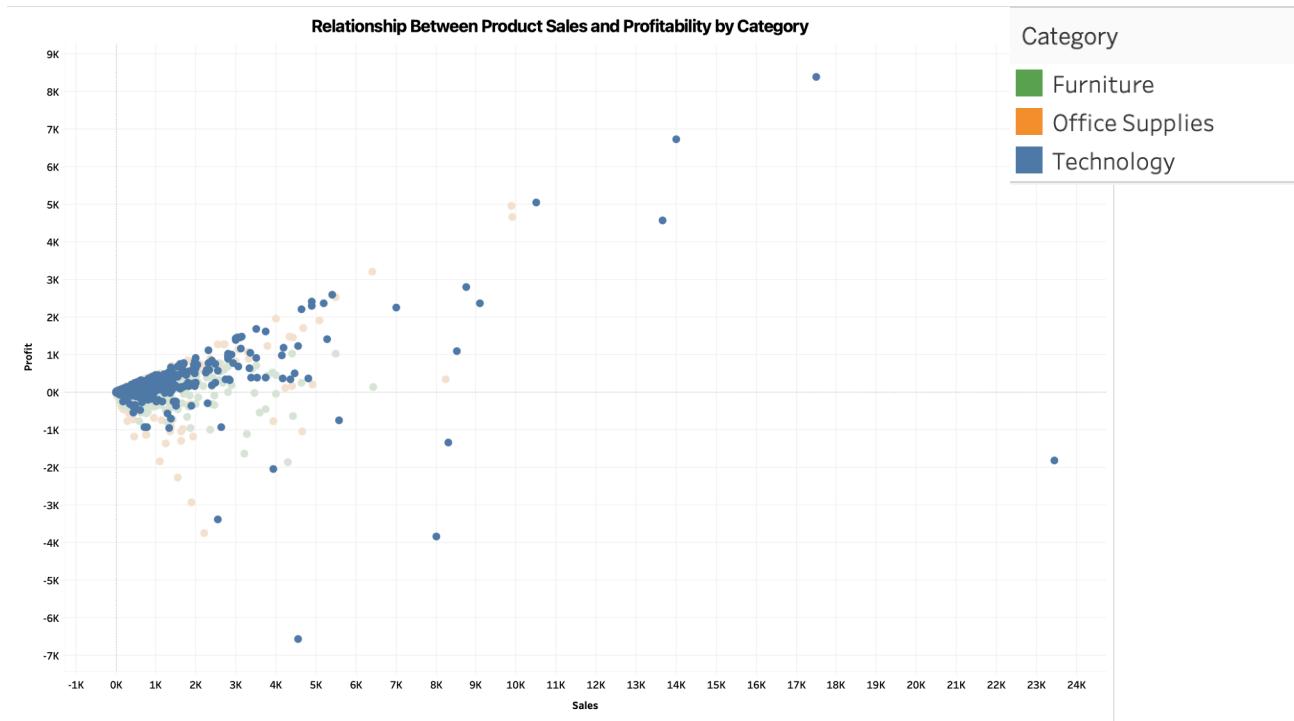
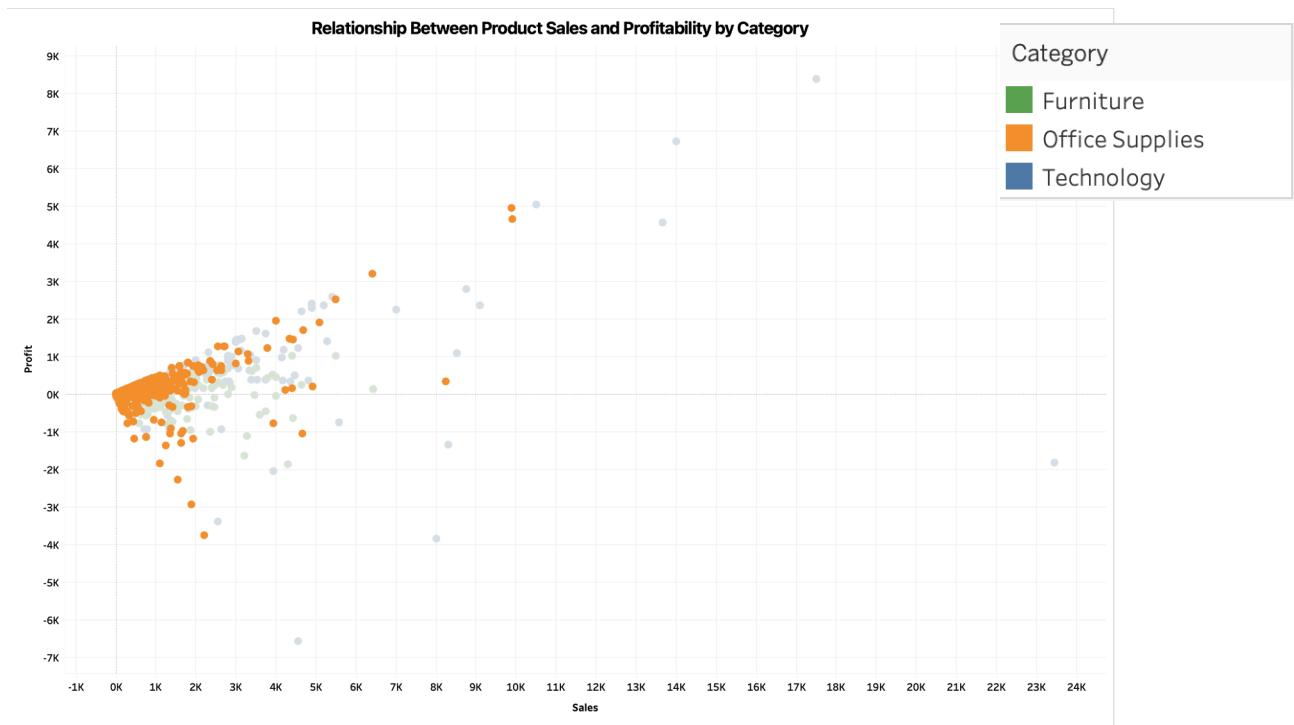
From the chart I can understand that Profit increases when discount decreases And Profit decreases as Discount increases. The trend line indicates negative growth which implies that more discount can lead to **no profitability**.

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

Ans.

Chart: Scatter Plot





Why Scatter Plot chart?

A Scatter Plot effectively visualizes the relationship between sales and profit, making it easy to spot high sales, high profit categories.

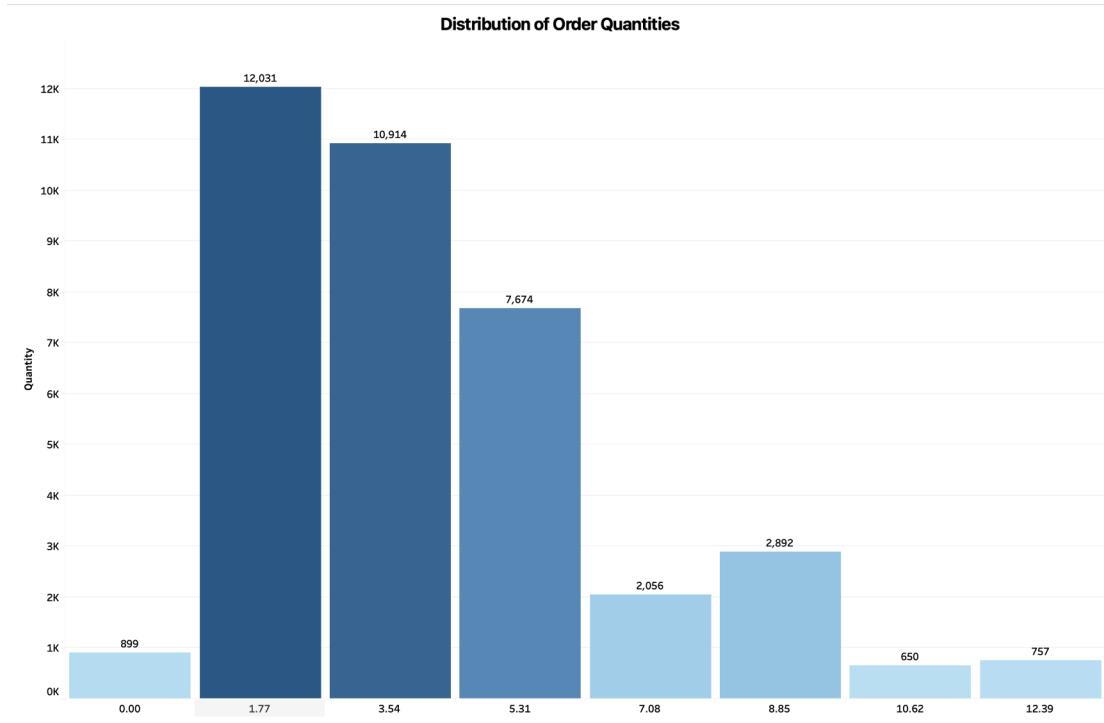
Insight:

From the graph it is clearly visible that each product categories faces profit and sometimes non profit. Where in now days high demand for technology, **Technology** category has good profit margin over non profit.

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

Ans.

Chart: Histogram



Why Histogram chart?

It can show the **frequency distribution of order quantities**, allowing for a clear view of common order sizes and their spread.

Insight:

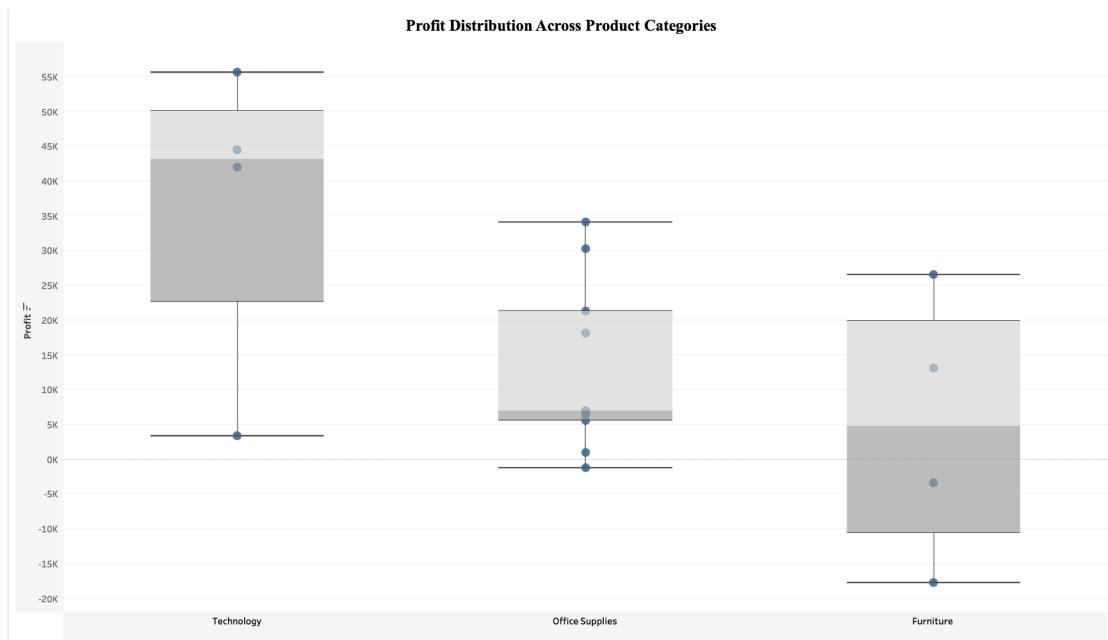
I understand from this chart that most orders fall within the range of 1.77 to 3.54 units, with around 12,000 orders in this range. As the order quantity increases, the number of orders generally

decreases. Orders above 8.85 units are much less common, with fewer than 3,000 occurring in these ranges. The least frequent order sizes are between 10.62 and 12.39, with only around 650 to 757 orders. **Overall, smaller order quantities, especially those below 5.31 units, make up the majority of the orders.**

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

Ans.

Chart: Box Plot



Why Box Plot/chart?

It provides a **visual representation of profit distribution** for each category, highlighting the range, medians and outliers.

Insight:

Based on the box plot, I understand the following insights:

- **Technology** has the highest median profit among all categories and a wide profit range, from losses to profits nearing 55K. There are also a few outliers with unusually high profits.
- **Office Supplies** shows the narrowest profit spread, with a median profit close to zero, meaning most profits are modest.

There are also outliers on both sides, representing significant profits and losses.

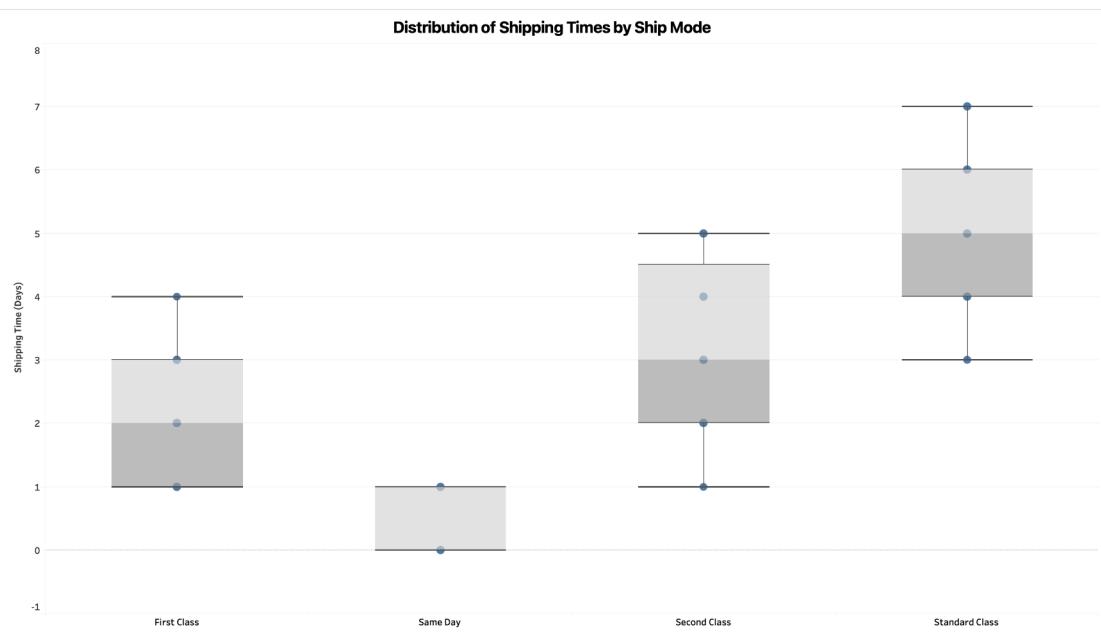
- **Furniture** has a balanced profit distribution with moderate variability. Some losses are seen, but they are not as extreme as in technology.

Overall, **Technology** is the most profitable category, with the highest potential for profits. **Office Supplies** has the least variability and lower profits, while **Furniture** falls in between with moderate profits and variability.

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

Ans.

Chart: Box Plot



Why Box plot/chart?

It's suitable for comparing distributions, showing variation in shipping time for different modes.

Insight:

Based on the shipping data, here's what I understand:

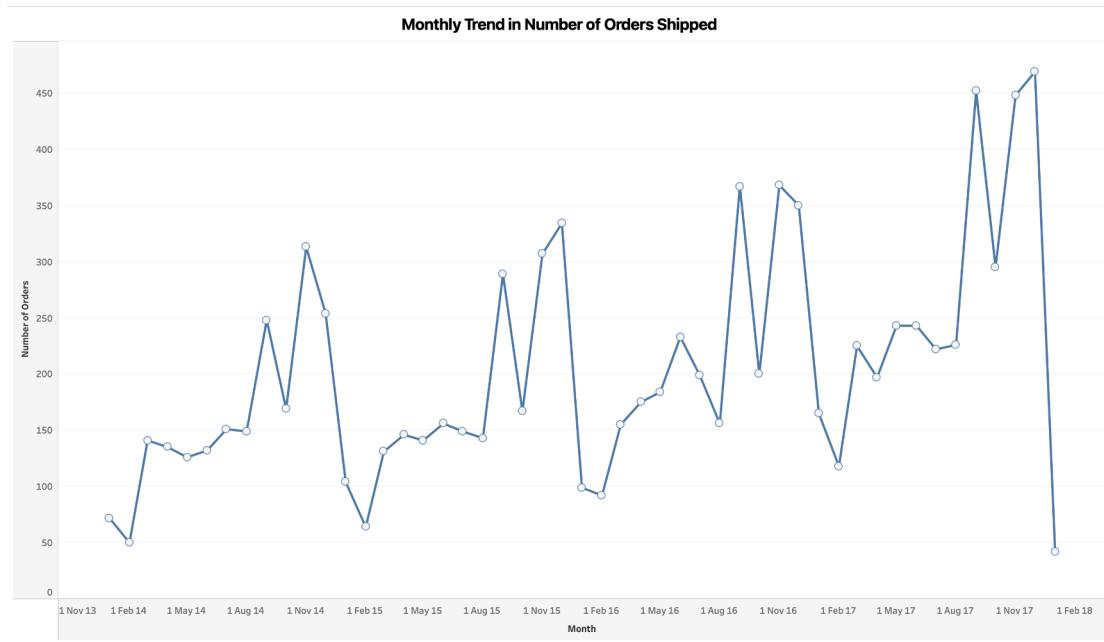
- **Same Day** is the fastest and most reliable, with all deliveries happening within a single day, showing no variability in shipping times.
- **First Class** is the second fastest, with a median shipping time of about 2 days. Most deliveries are made within 1 to 3 days, with a maximum of 4 days.
- **Second Class** has a moderate delivery time, with a median of 3 days. The shipping times range from 2 to around 5 days, showing more variability than "First Class."
- **Standard Class** has the slowest delivery, with a median of about 5 days. The shipping time can vary widely, from 3 days up to 7 days, making it the least consistent option.

Overall, the shipping modes are predictable: faster options like "**Same Day**" and "**First Class**" are more consistent and quicker, while slower options like "**Second Class**" and "**Standard Class**" have more variability and longer delivery times. If speed and reliability are important, "**Same Day**" or "**First Class**" are better choices, but "**Standard Class**" can be a cost-effective option for non-urgent shipments.

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

Ans.

Chart: Line chart



Why Line chart?

It displayed the **monthly trend** in shipping order, helping to identify patterns or seasonality in order volume.

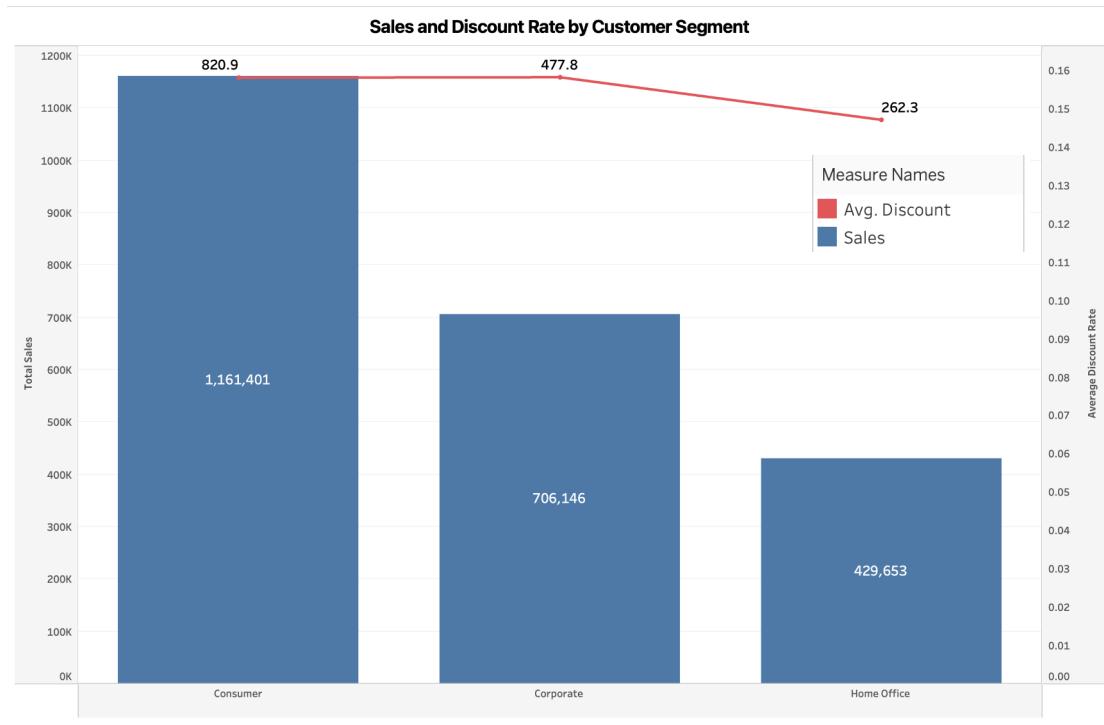
Insight:

This chart shows the monthly trend in the number of orders shipped from November 2013 to February 2018. I understand that the chart reveals significant fluctuations in order quantities, with noticeable **peaks around November each year**, likely due to holiday shopping, and drops around February, indicating a seasonal trend in shipping orders.

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

Ans.

Chart: Bar chart with Dual axis



Why Bar chart with dual axis?

It can **display sale performance along side discount rate**, providing a comparative view of segment behavior.

Insight: I understand that the consumer segment has the highest total sales, paired with a higher average discount rate, while the home office segment has the lowest sales and a lower average discount. **This suggests that offering higher discount may drive more sales**, especially evident in the consumer segment.

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

Ans.

Chart: Heat map



Profit by Subcategory and Region				
Sub-Categ..	Central	East	South	West
Accessories	7,252	11,196	7,005	16,485
Appliances	-2,639	8,391	4,124	8,261
Art	1,195	1,900	1,059	2,374
Binders	-1,044	11,268	3,901	16,097
Bookcases	-1,998	-1,168	1,339	-1,647
Chairs	6,593	9,358	6,612	4,028
Copiers	15,609	17,023	3,659	19,327
Envelopes	1,778	1,812	1,465	1,909
Fasteners	237	264	174	275
Furnishings	-3,906	5,881	3,443	7,641
Labels	1,073	1,129	1,041	2,303
Machines	-1,486	6,929	-1,439	-619
Paper	6,972	9,015	5,947	12,119
Phones	12,323	12,315	10,767	9,111
Storage	1,970	8,389	2,274	8,645
Supplies	-662	-1,155	2	626
Tables	-3,560	-11,025	-4,623	1,483

Sales by Subcategory and Region				
Sub-Categ..	Central	East	South	West
Accessories	33,956	45,033	27,277	61,114
Appliances	23,582	34,188	19,525	30,236
Art	5,765	7,486	4,656	9,212
Binders	56,923	53,498	37,030	55,961
Bookcases	24,157	43,819	10,899	36,004
Chairs	85,231	96,261	45,176	101,781
Copiers	37,260	53,219	9,300	49,749
Envelopes	4,637	4,376	3,346	4,118
Fasteners	778	820	503	923
Furnishings	15,254	29,071	17,307	30,073
Labels	2,451	2,603	2,353	5,079
Machines	26,797	66,106	53,891	42,444
Paper	17,492	20,173	14,151	26,664
Phones	72,403	100,615	58,304	98,684
Storage	45,930	71,613	35,768	70,533
Supplies	9,467	10,760	8,319	18,127
Tables	39,155	39,140	43,916	84,755

Why Heat map/chart?

It can show **sales and profit trends across subcategories and region**, making it easy to spot high-performing areas.

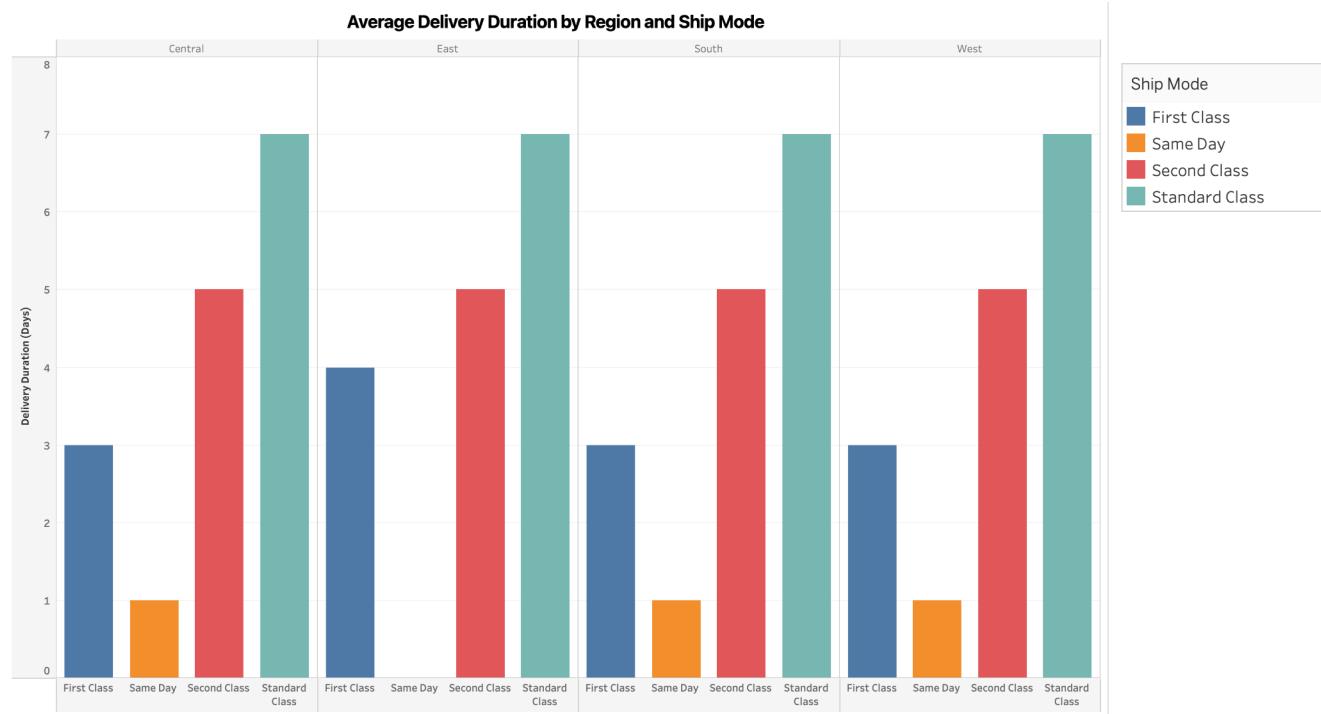
Insight:

I understand that the table shows that the **West and East** regions dominate in sales across most subcategories. Categories like Accessories, Chairs, and Phones have significantly higher sales in these regions, while the Central and South regions often show lower sales. This suggests that the **West and East** regions are key markets for driving sales in the Superstore dataset.

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

Ans.

Chart: Bar Chart



Why Bar chart?

It can compare **average delivery duration across regions and ship modes**, making it easy to identify slow and fast delivery combination.

Insight:

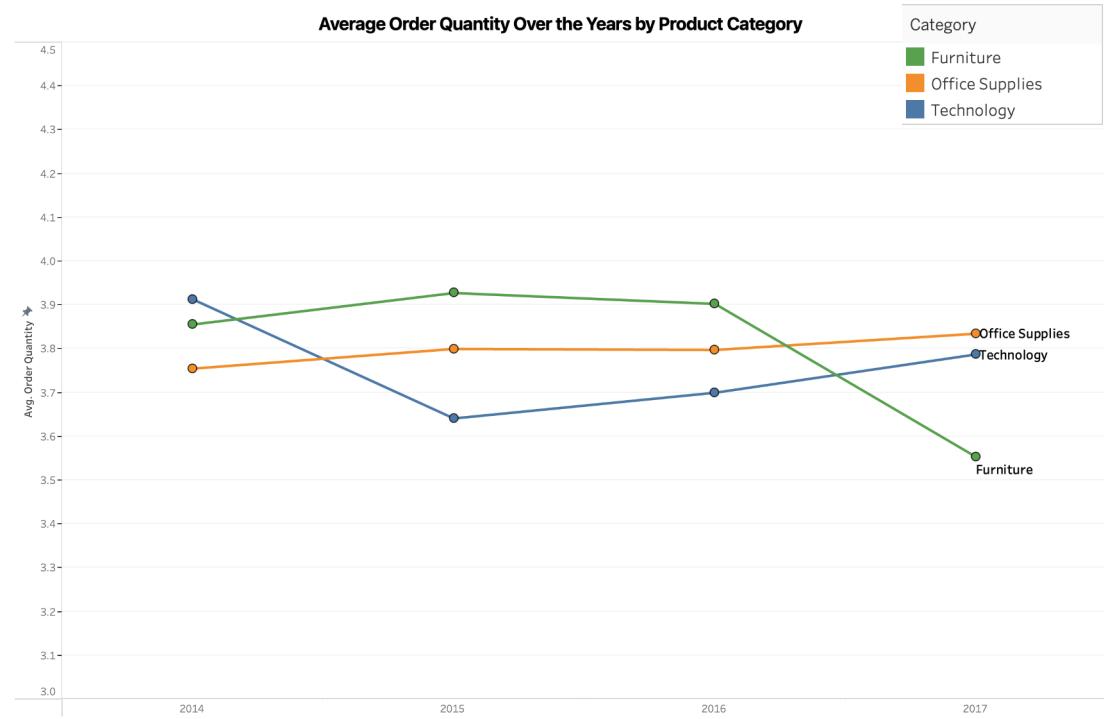
I understand that this chart reveals that the average delivery duration varies by region and ship mode. **Same Day** shipments

consistently have the shortest delivery times across all regions, while **Standard Class** shipments have the longest. **First Class** and **Second Class** fall in between, with slight variations depending on the region. The Central and East regions show longer delivery times for First Class shipments compared to the South and West regions.

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

Ans.

Chart: Line Chart



Why Line chart?

It can show **trends in average order quantities over time**, highlighting changes and patterns across categories.

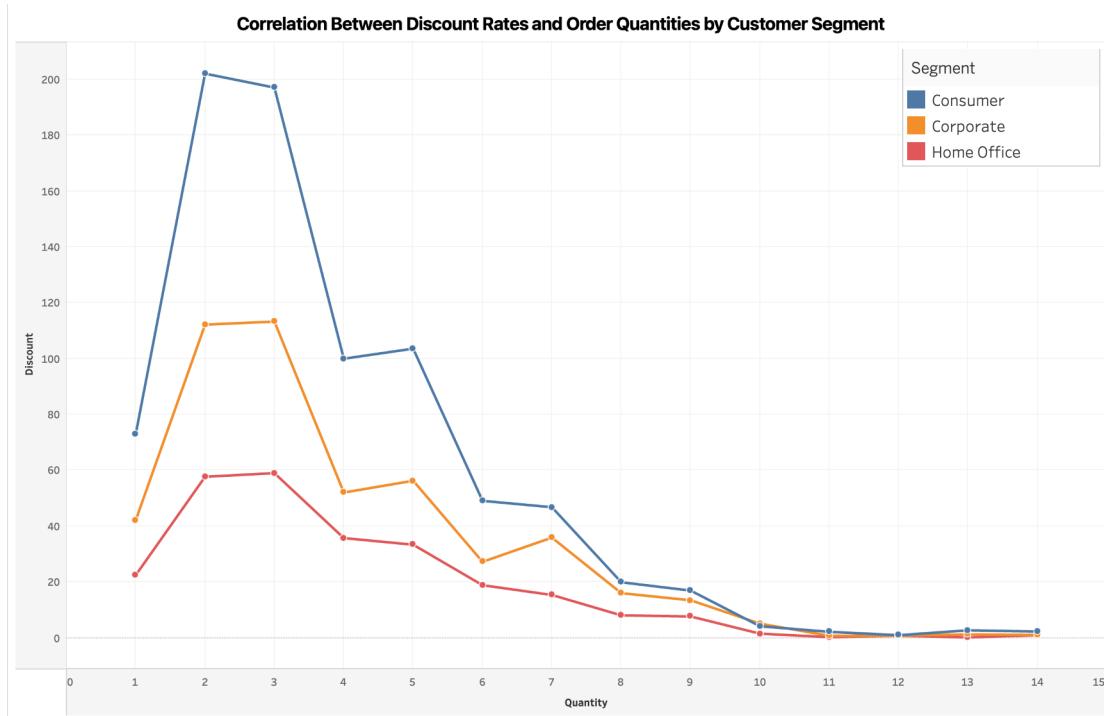
Insight:

I understand that the graph shows that the average order quantity for **Furniture decreased slightly over the years**, while **Technology and Office Supplies remained relatively stable**. Technology saw minor fluctuations, and Office Supplies maintained a steady average order quantity.

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

Ans.

Chart: Line chart



Why Line chart?

It can show to **correlation between discount rate and order quantities**, revealing any patterns specific to customer segments.

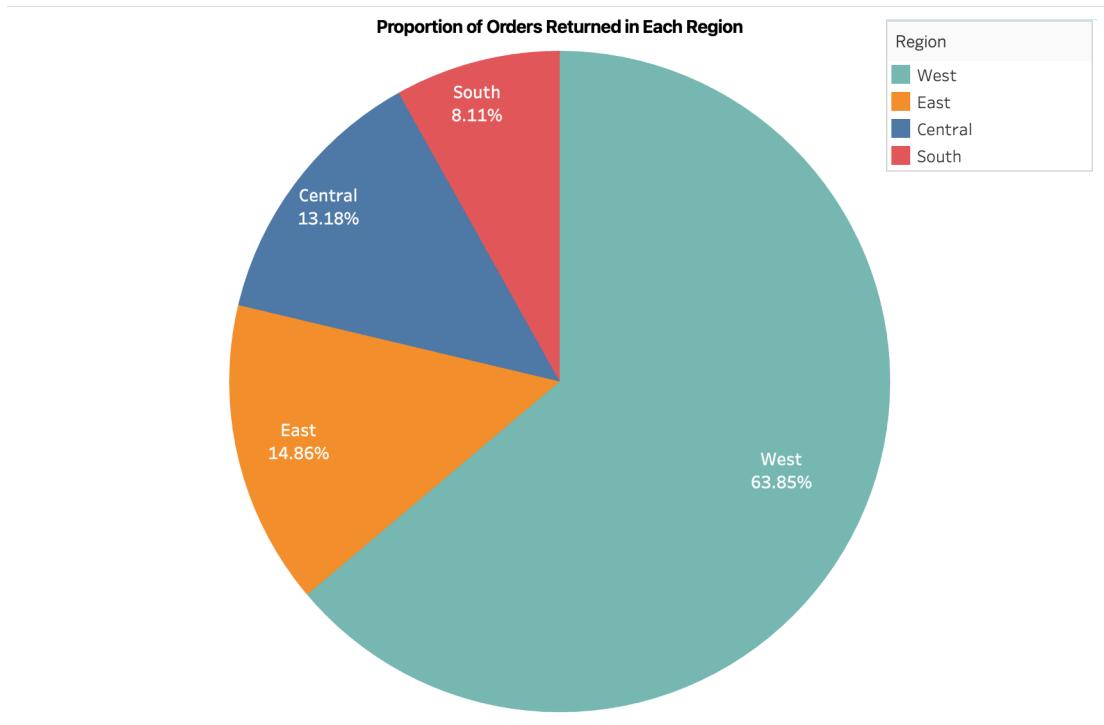
Insight:

I understand that higher discount rates often lead to an increase in order quantities across different customer segments. This trend indicates that **discounts can drive more orders**, particularly in segments that respond well to price reductions.

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

Ans.

Chart: Pie chart



Why Pie chart?

It can show the **proportion of returns**, helping identify regions with high or low return rates.

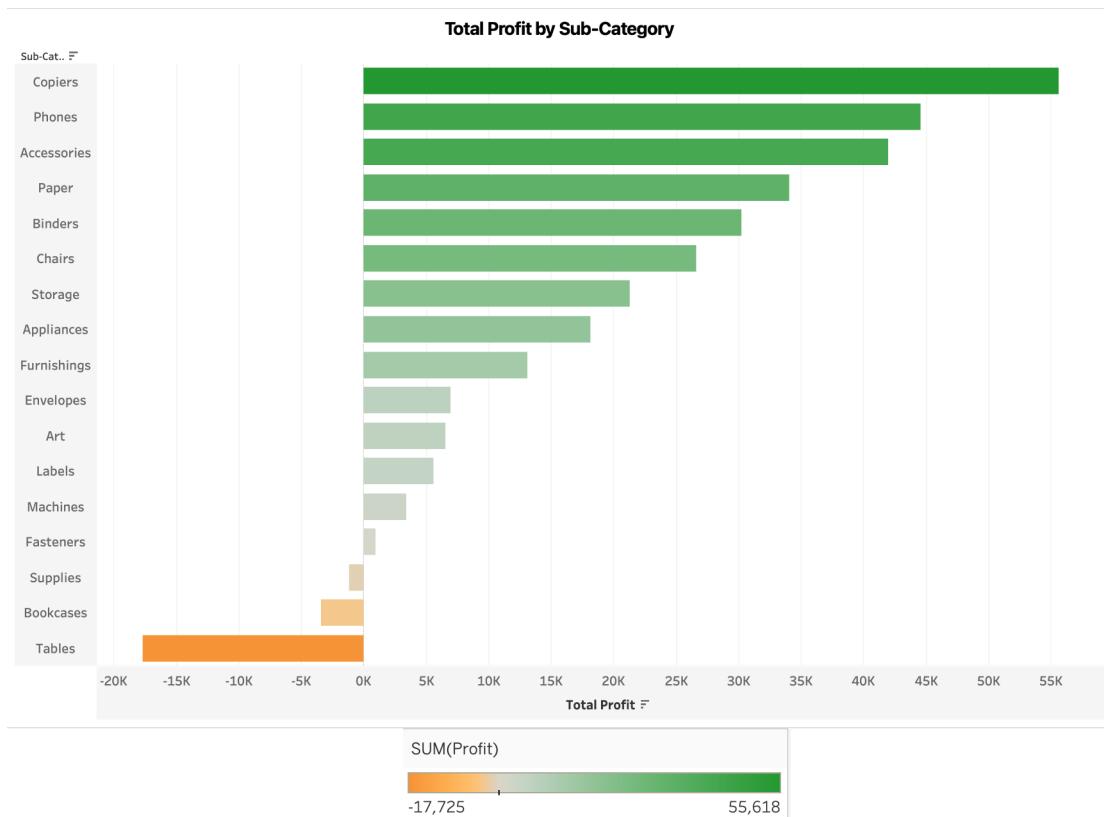
Insight:

I understand that the pie chart shows that the **West region** has the highest proportion of returned orders at **63.85%**, followed by the **East at 14.86%**, the **Central at 13.18%**, and the **South at 8.11%**. This indicates that returns are significantly more common in the West compared to other regions.

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

Ans.

Chart: Bar Chart



Why Bar chart?

It can visually **compare profits across products and subcategories**, highlighting top-performing items.

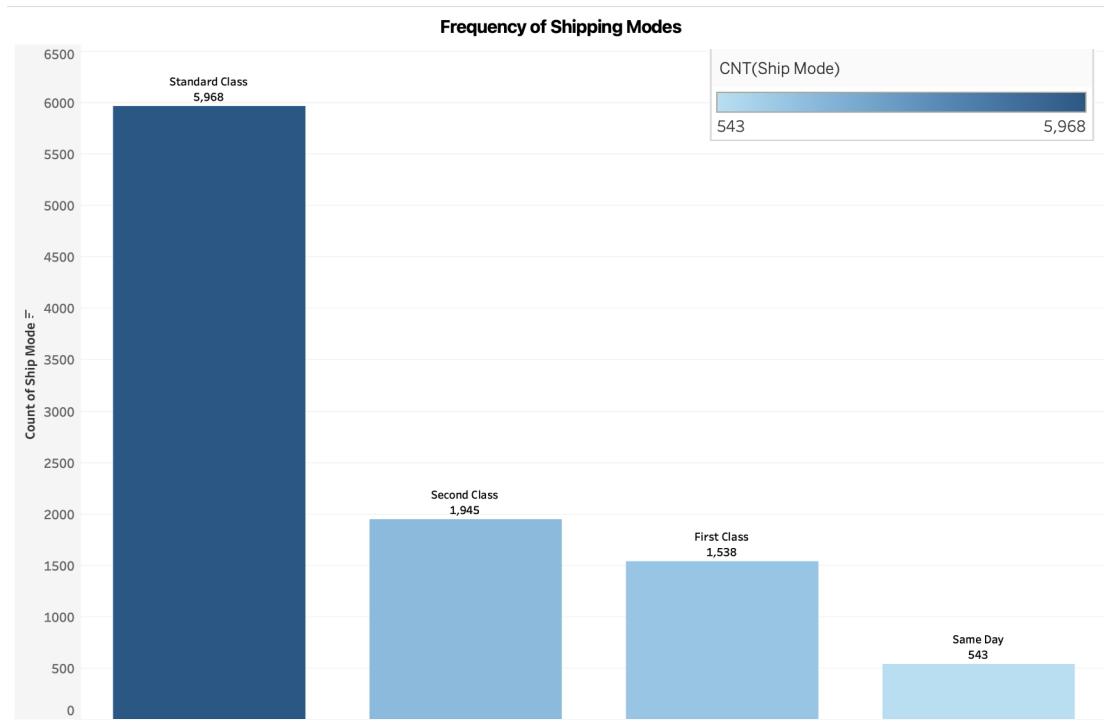
Insight:

I understand that the bar chart shows that Copiers, Phones, and Accessories are the most profitable subcategories, while Tables and Bookcases generate the least profit, even incurring losses. This suggests that focusing on high-profit subcategories like **Copiers and Phones might be beneficial for increasing overall profitability**.

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

Ans.

Chart: Bar Chart



Why Bar chart?

It can show the **frequency of shipping mode** used, making it easy to see which mode is preferred.

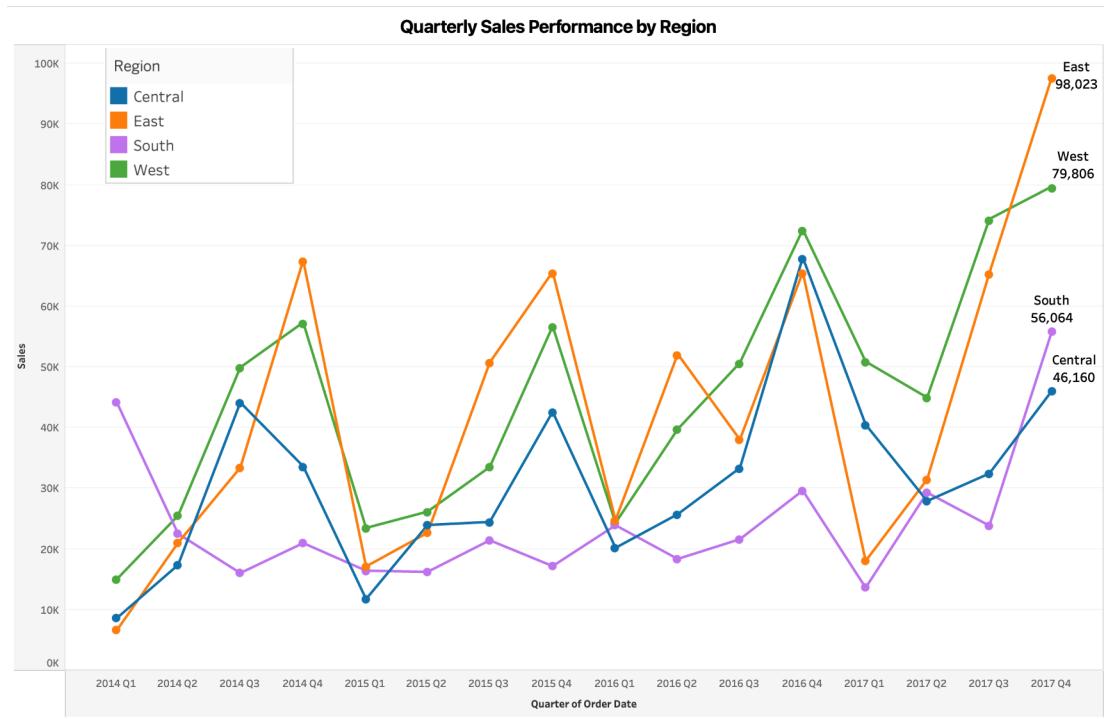
Insight:

I understand that "**Standard Class**" is the most commonly used shipping mode in the Sample Superstore dataset, with the highest count of **5,968**. This indicates a preference for this mode over others like Second Class, First Class, and Same Day.

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

Ans.

Chart: Line Chart



Why Line chart?

It can show **sales trends over quarters**, highlighting region and performance fluctuations throughout the year.

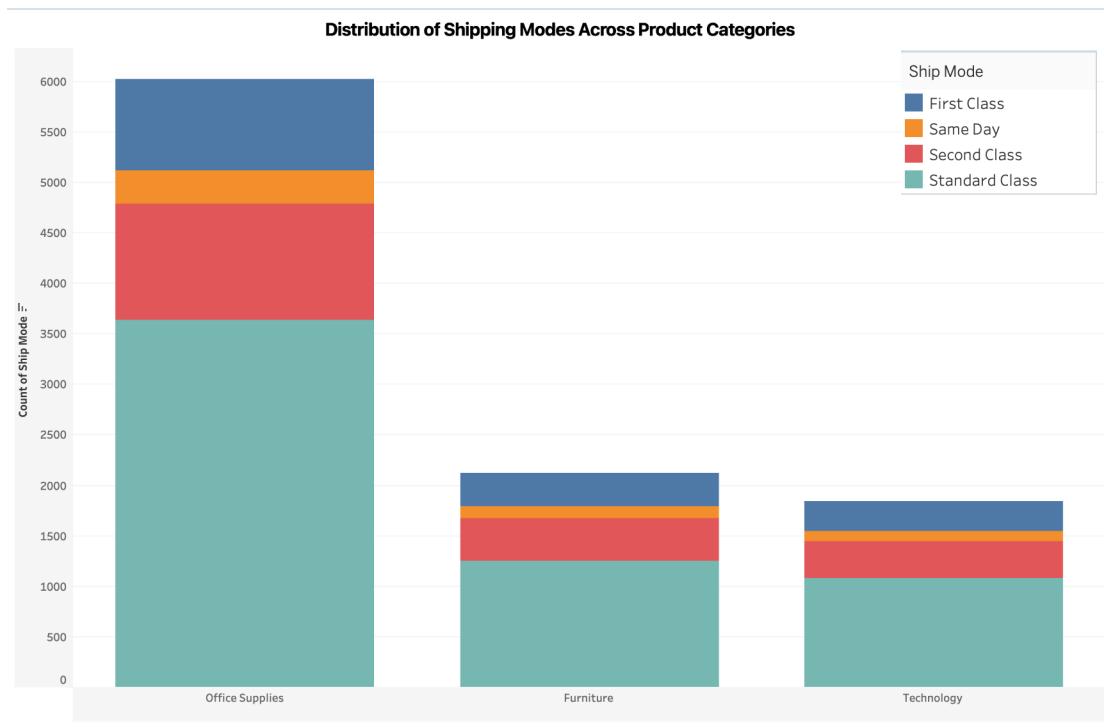
Insight:

I understand that the sales performance chart shows significant growth in the East and West regions, especially in 2017 Q4, where they see the highest peaks. The South region is more volatile but also rises notably in 2017 Q4. Meanwhile, the **Central region maintains a stable performance with a gradual increase towards the end of the period**.

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

Ans.

Chart: Stacked Bar Chart



Why Stacked Bar chart?

It can show the distribution of order priorities, helping understand which categories have high or low urgency.

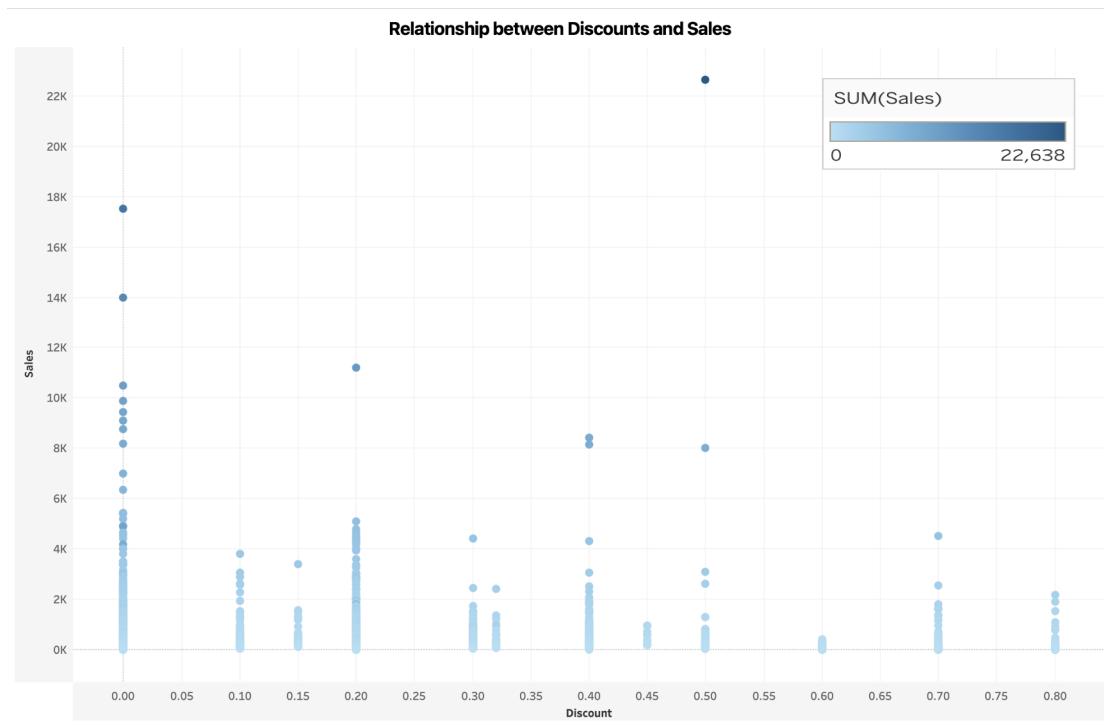
Insight:

I understand that the chart shows Office Supplies have the highest count of shipping modes, dominated by a particular mode indicated in teal. **Furniture and Technology have similar distributions** but with fewer counts, following the same pattern of shipping mode preference. This indicates that shipping preferences are quite consistent across different product categories, with **Office Supplies being the most frequently shipped**.

28. What is the relationship between discounts and sales?

Ans.

Chart: Scatter Plot



Why scatter plot/chart?

It can **visualizes the relationship between discount and sales**, showing how discount impact sales volume.

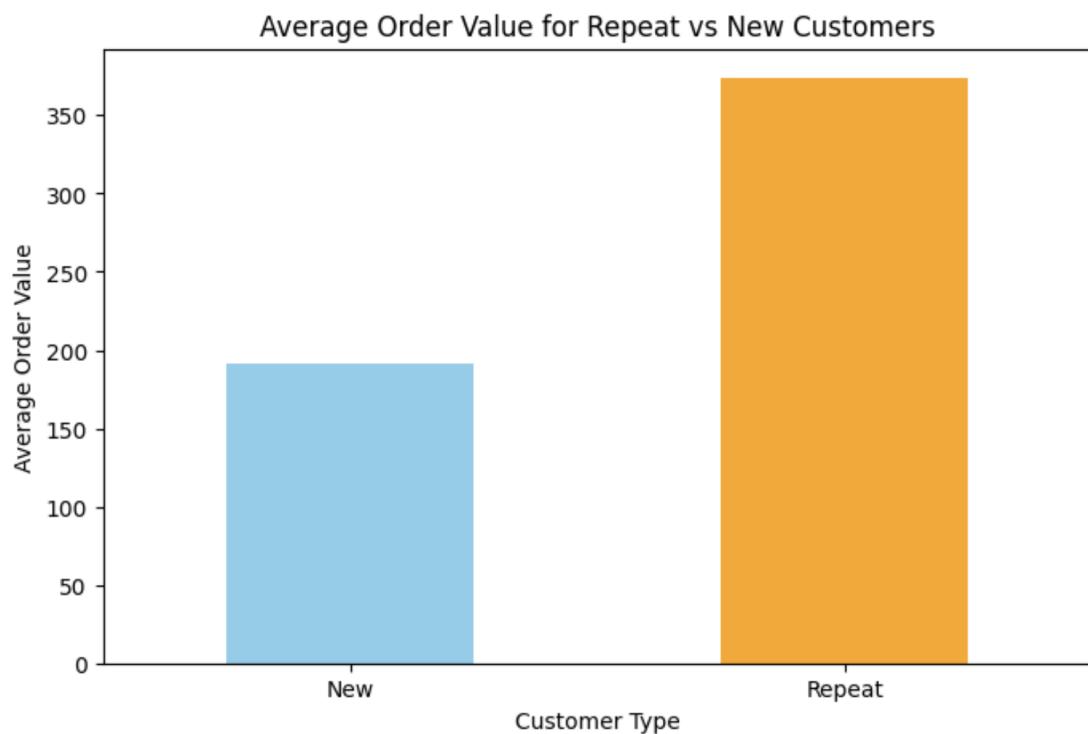
Insight:

I understand that the scatter plot shows sales mostly concentrate around lower discount rates. While high sales values appear at specific discount rates, such as around 0.00, 0.20, and 0.50, there isn't a consistent pattern. **This suggests no clear linear relationship between discounts and sales, meaning other factors likely influence sales as well.**

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

Ans.

Chart: Bar Chart



Why bar chart?

It compares average order values between customer type, clearly showing difference in spending patterns

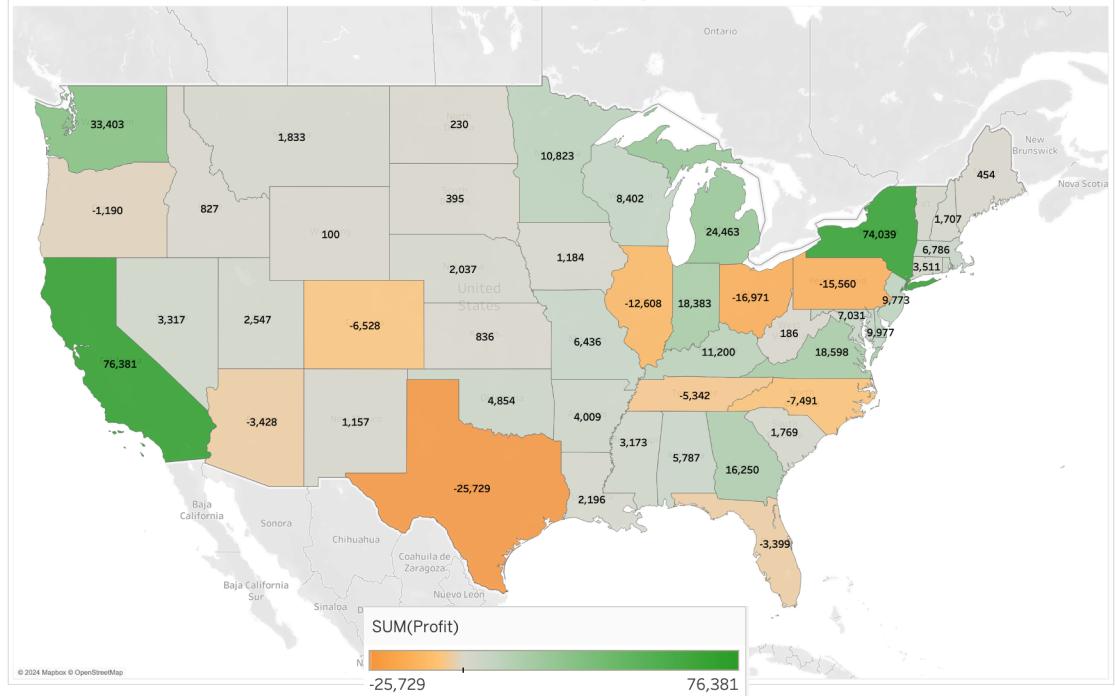
Insight:

I understand that the bar chart shows the average order value for repeat customers is significantly higher, around 350, compared to new customers, who average around 200. **This indicates that repeat customers tend to place higher-value orders than new customers.**

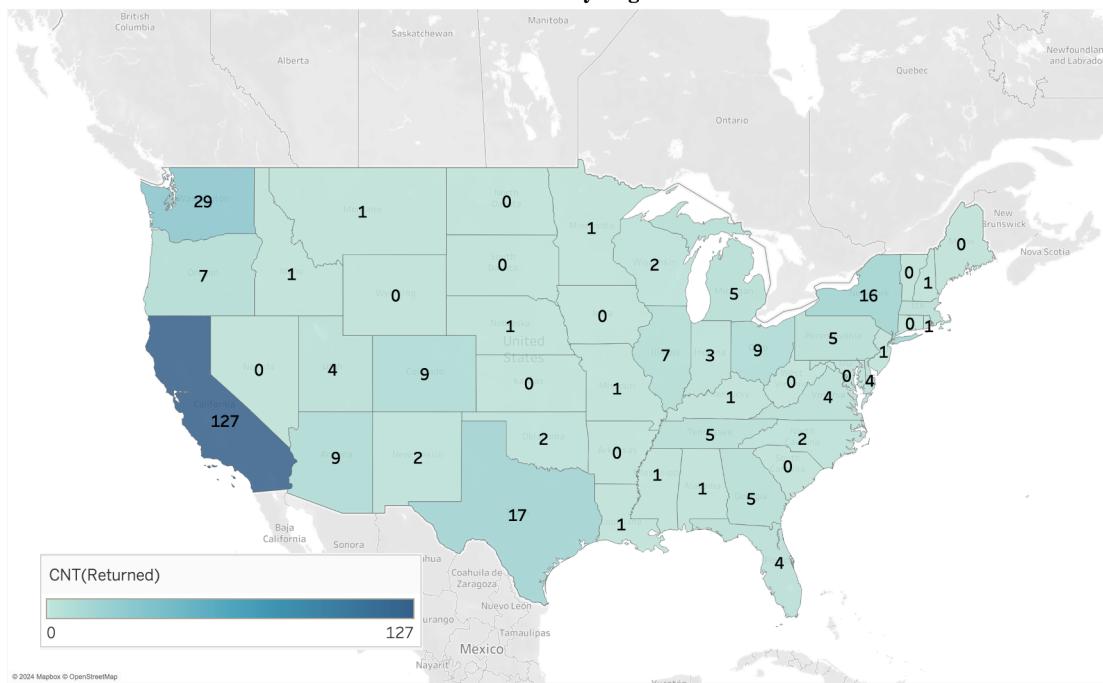
30. What is the geographical distribution of returns and its impact on overall profitability?
Ans.

Chart:

Profit Impact by Region



Return Rate by Region



Why map chart?

Map chart show geographical representation, It provide insights into how returns affect profitability across regions.

Insight:

Overall, I understand that **California and Washington have the highest return rates but maintain positive profits**. Conversely, Texas, Illinois, and Indiana experience significant returns that negatively affect their profitability, highlighting differences in how states manage returns and maintain profitability.
