Insights from Tableau: Sample Superstore Data Analysis Project

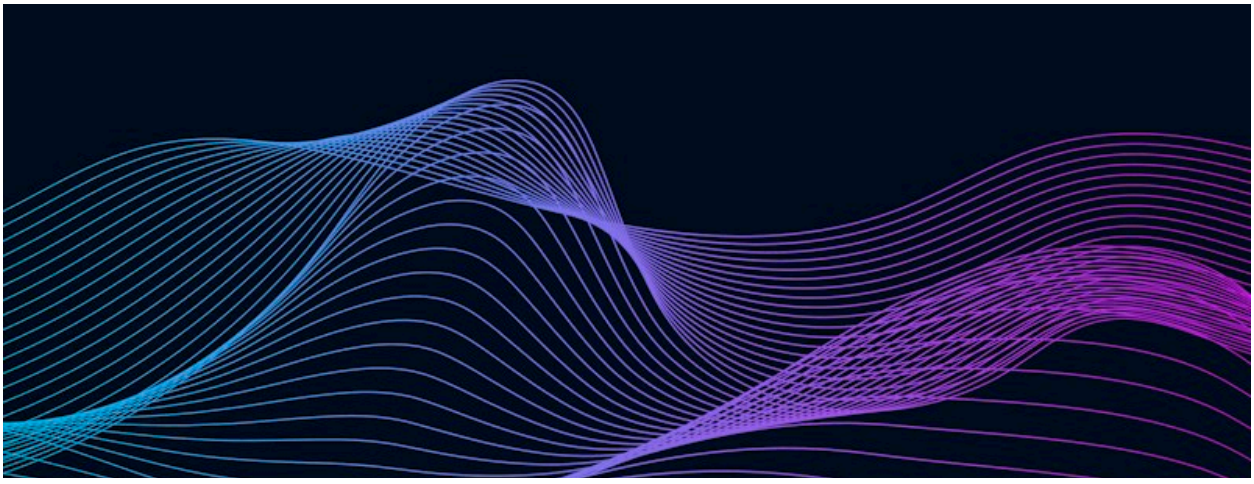
About the Project:

This project involves hands-on exploration of a Sample Superstore dataset, tackling 30 scenario-based questions using data visualization and analysis. The aim is to adeptly choose the optimal chart for each question and elucidate the reasoning behind those choices. The outcome of this project will spotlight my proficiency in data visualization, critical thinking, and the ability to convey complex ideas effectively.

Skills Required:

- Proficiency in data visualisation concepts and techniques.
- Familiarity with Tableau or a similar data visualisation tool.
- Strong analytical and problem-solving skills.
- Ability to choose appropriate charts based on data characteristics and question requirements.
- Clear and concise communication skills.

Dataset Link



Questions

---

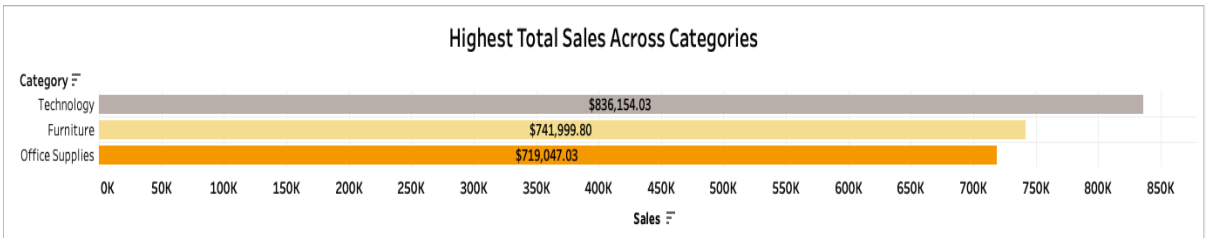
1. Which product categories have the highest total sales in the "Superstore" dataset?

Chart Type - Bar Chart

Reason : Bars rule! They're perfect for comparing data across categories at a glance. Each bar's height shows sales, making it easy to see "*Technology*" dominates at *$836,000*, followed by "*Furniture*" and "*Office Supplies*".

Insight : "*Technology*" tops the chart, raking in over *$836,000*, followed by "*Furniture*" at *$742,000* and "*Office Supplies*" trailing at *$719,000*. This bar chart highlights the highest sales across categories, possibly for comparison or showcasing top performers. It's clear: tech reigns supreme in sales!

Highest Total Sales Across Categories



---

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

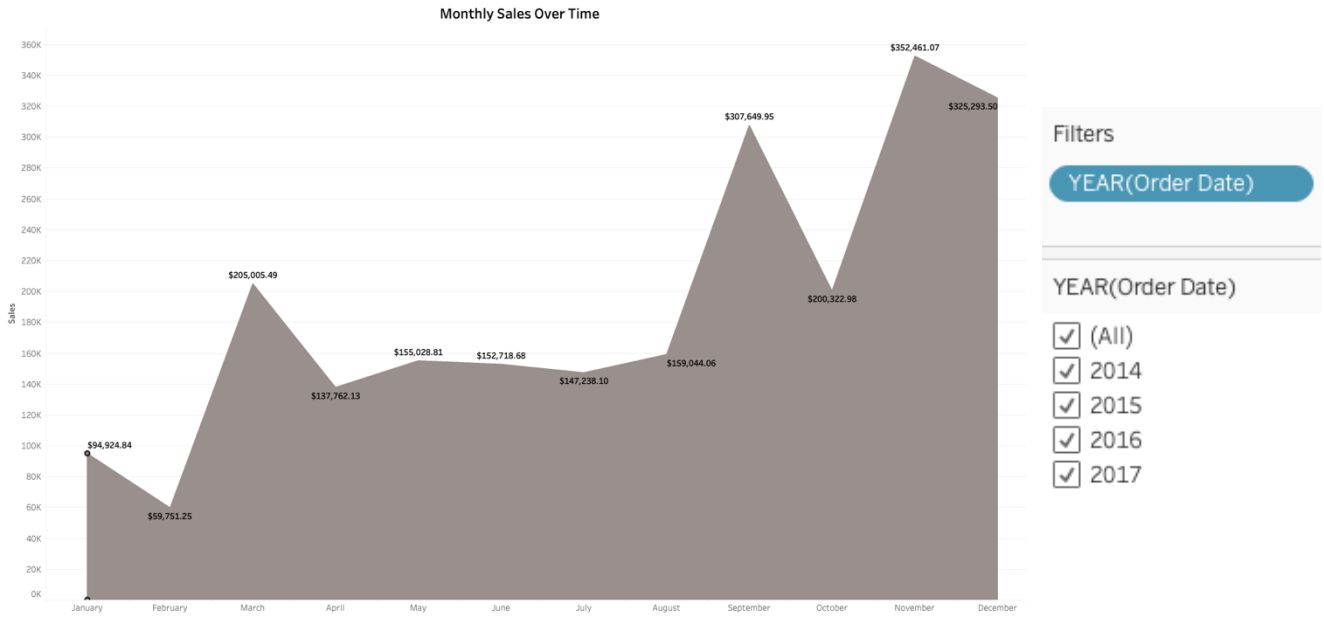
Chart Type – Area Chart

Reason: Trend Over Time! To show change over time it is best suited to use Area chart or Line chart, I selected Area chart for in-depth understanding and easy on the eyes.

Insight: A noticeable trend in the monthly sales data reveals a consistent increase throughout the year, culminating in the highest total sales observed in November. Despite a sharp decline in October subsequent to the peak in September, the sales figures still surpassed the previous low.

Note: We can use Year(Order Date) in Filter Option on panel to get insights from individual year.

Monthly Sales Over Time



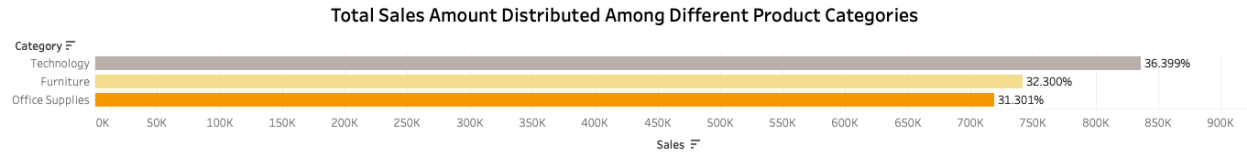
---

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

Chart Type – Bar Chart
Reason: Bars rule! They're perfect for comparing data across categories at a glance.

Insight: It is observed that "*Technology"* from product category has the highest percentage of *36.4%.* Whereas "*Furniture*" and "*Office Supplies*" have difference of *1%.*

Sales vs Category in (%)



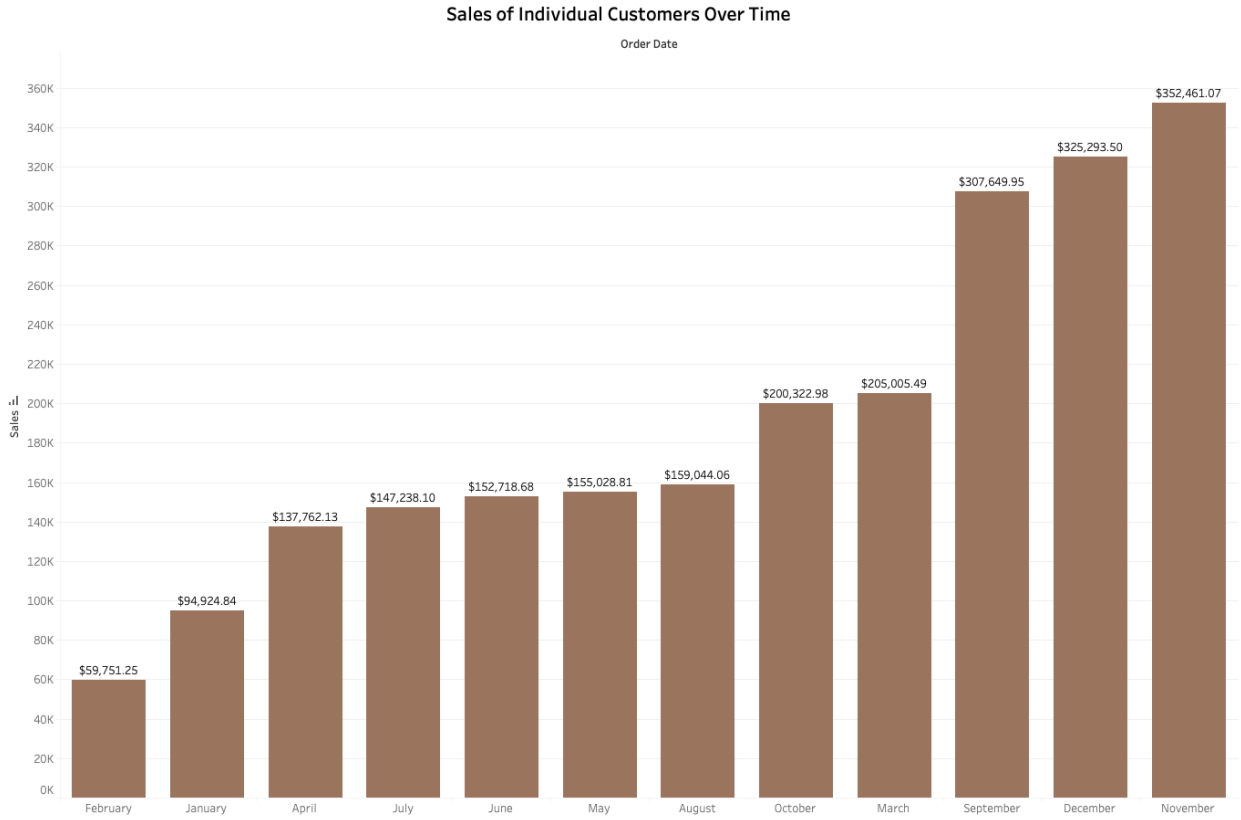
---

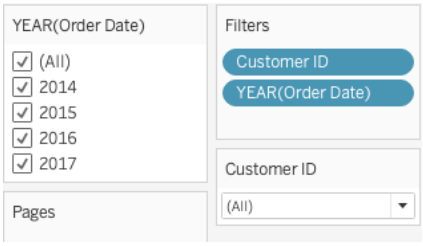
4. Can we analyze the sales performance of individual customers over time?

Chart Type – Column Chart

Reason: To get better understanding how individual customer spends over time, I used column chart to track their spending habit across months. To get performance of single person I used Filter option on panel consisting Year and Customer ID.

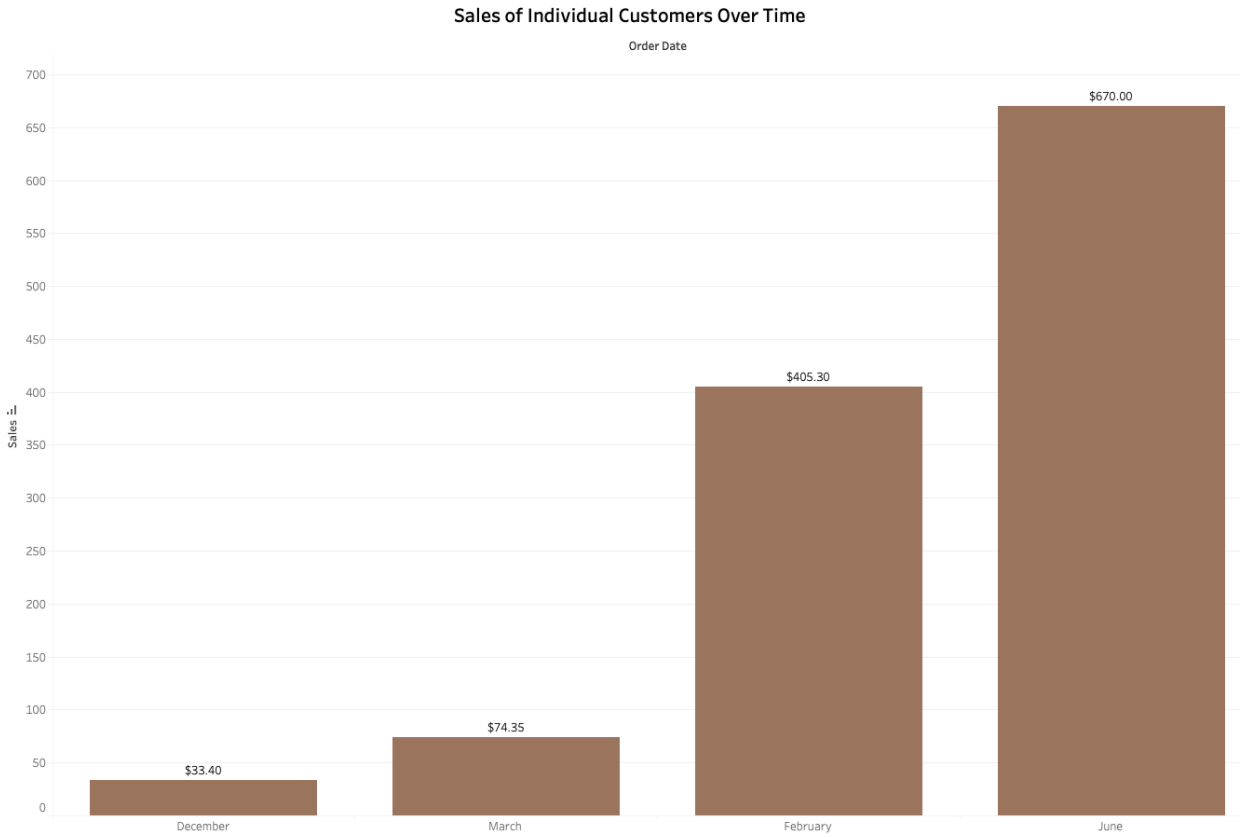
Sales of Individual Customer over Time





Filter Option Containing YEAR(Order Date) and Customer-ID

***Example: Performance of a customer whose Customer ID**: BV-11245 **and Year:** 2015 and 2016*



Insight: We can get insight of every individual in each year from *2014 to 2017.* Just by selecting Customer ID and Year. We can track spending habit of a person like when did he/she spend? And how much did he/she spend?

---

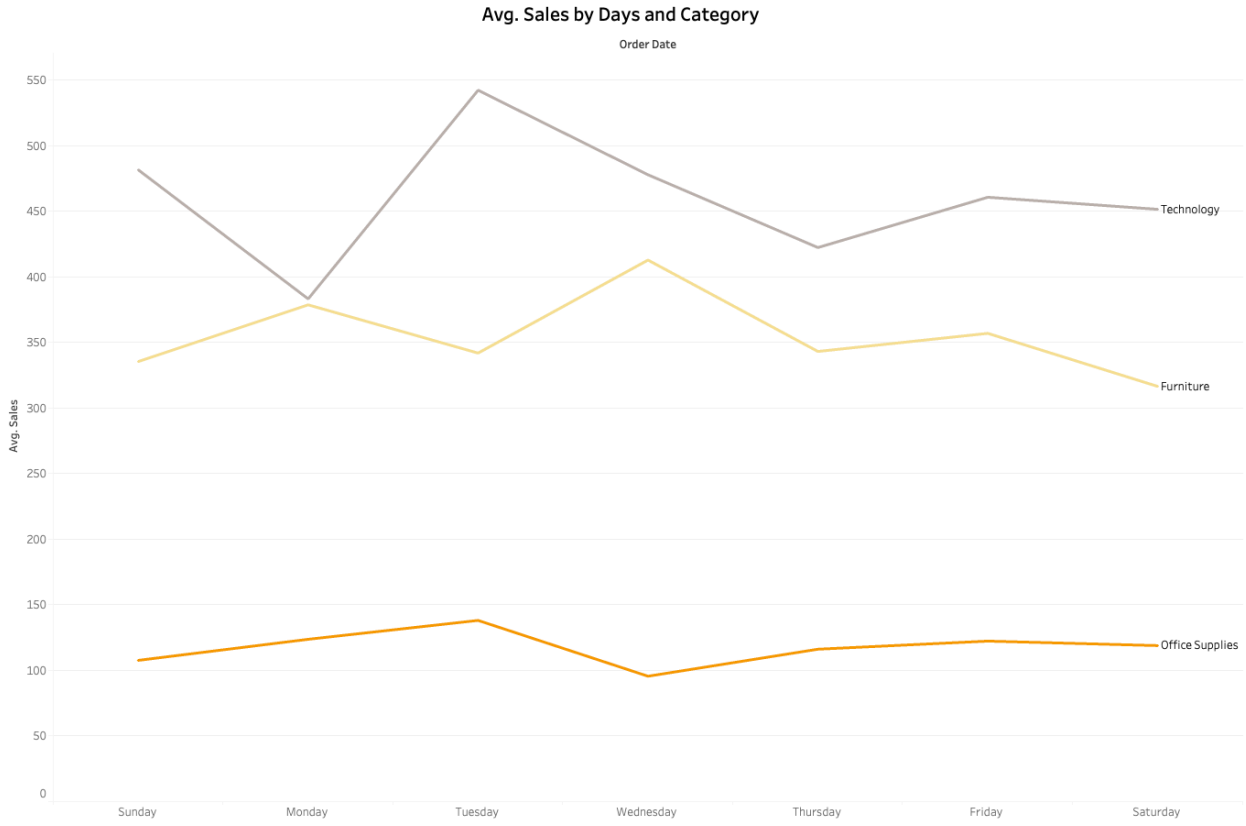
5. How do sales vary based on different days of the week and product categories?

Chart Type – Line Chart

Reason: Utilizing a line chart proves effective in visually examining data trends as they unfold over time. Specifically, in this instance, the line chart is employed to portray the cumulative trend in average daily sales across multiple years. Each distinct line on the chart represents the average daily sales trend for various categories.

Insights: The analysis reveals that, typically, the highest average sales occur on Tuesdays for Office Supplies and Technology, and on Wednesdays for Furniture. Furthermore, it is noticeable that the average sales for Office Supplies remain relatively consistent throughout the week, with only a slight fluctuation observed on Tuesdays and Wednesdays

Average Sales by Week across Category



---

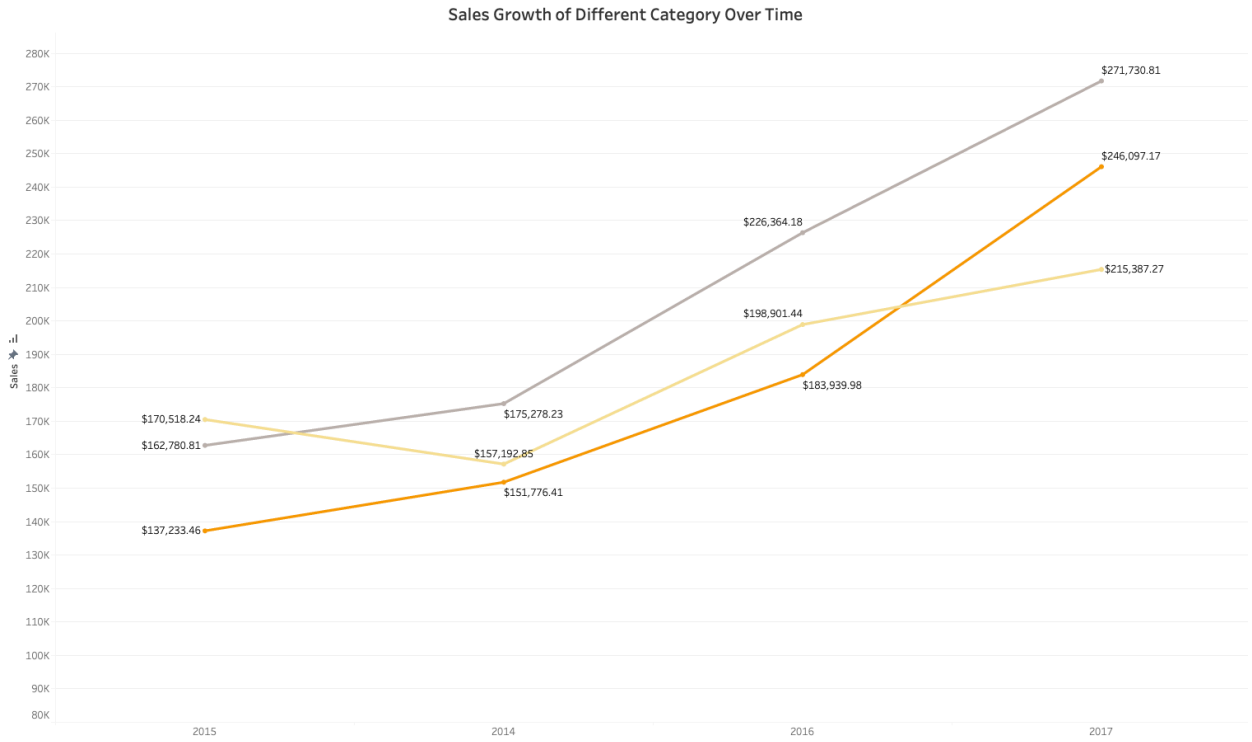
6. Can we visualise the sales growth of different product categories over time?

Chart Type – Line Chart

Reason: A line chart is used for visualising and analysing data trends over time.

Insight: It is evident that the overall sales have shown a consistent upward trend throughout the years. The technology category experienced the highest growth in sales, followed by office supplies and furniture. In 2015, both technology and office supplies recorded their lowest sales, while furniture had its lowest sales in 2014. The peak total sales for all product categories occurred in 2017.

Growth of Sales in Category vs Years



---

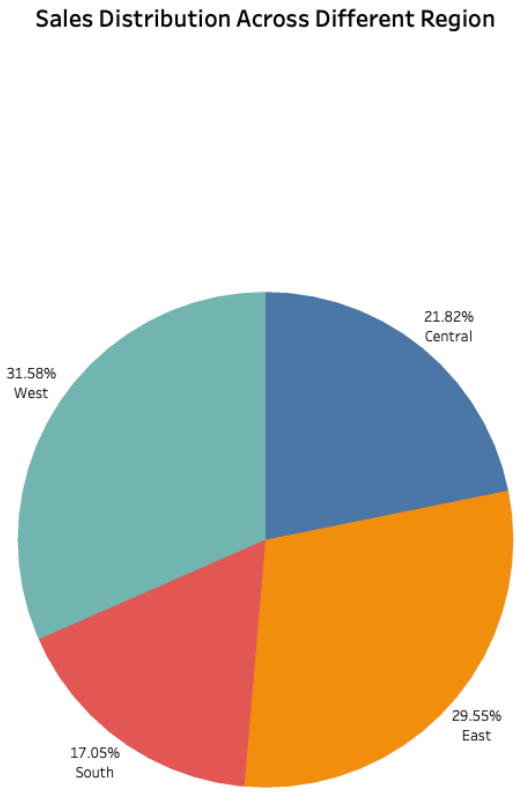
7. How does the sales distribution vary across different regions in the "Superstore" dataset?

Chart Type – Pie Chart

Reason: A pie chart is commonly used for visualizing the distribution of parts or components within a whole. It is particularly effective when you want to represent the proportional relationship between different categories or segments in a dataset.

Insight: A pie chart is useful for illustrating the variation in sales distribution across different regions in the "Superstore" dataset. Each slice of the pie represents a specific region, and the size of each slice reflects the proportion of sales attributed to that region. *"West region"* has the highest distribution with *31.58%.* Whereas, "*South region"* has the lowest among all which is *17.05%.* On the other-hand*, Centra region* and *East region* has *21.80%* and *29.60%* respectively.

Sales Distribution across different Region



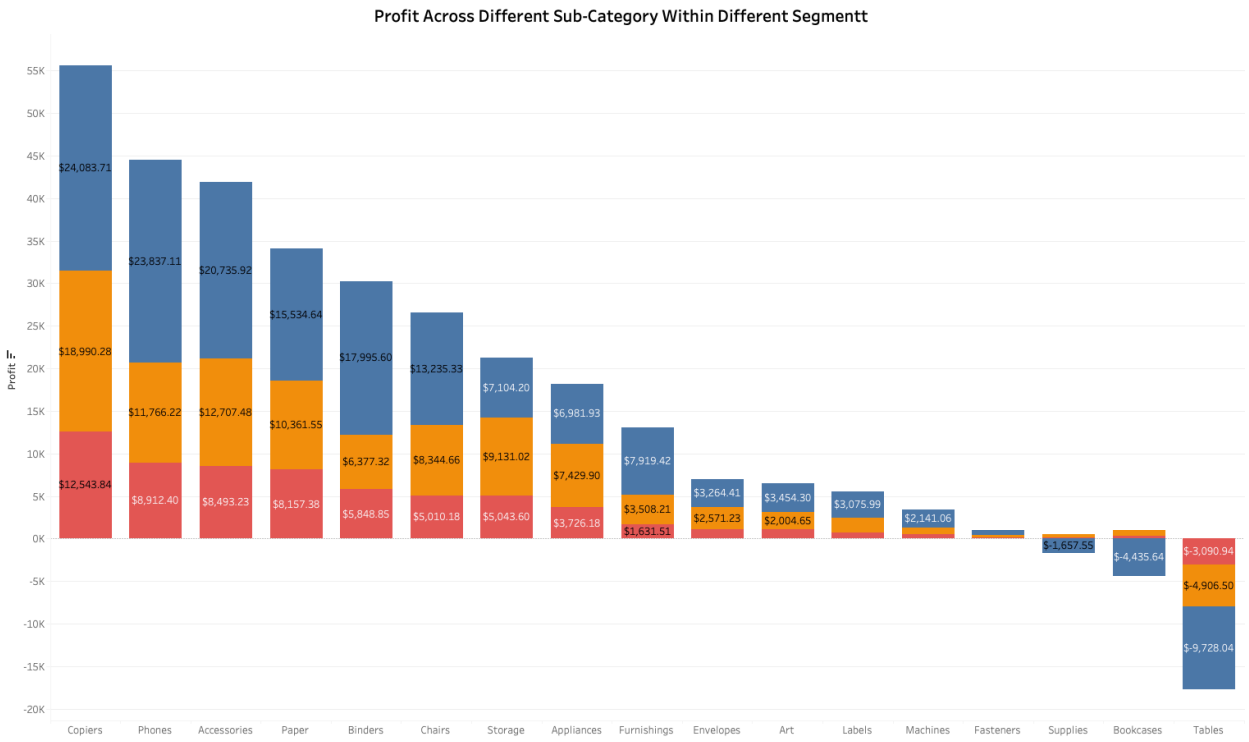
---

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

Chart Type - Stacked Column Chart

Reason: Stacked column charts prove to be a versatile visual aid for understanding and contrasting the makeup, proportions, and trends of various subcategories within categories. In this specific case, a stacked column chart is employed to illustrate the breakdown of profits among different subcategories for distinct customer segments. The x-axis showcases the diverse subcategories, and each bar represents the profit distribution among different customer segments within those subcategories.
Insight: The analysis reveals that copiers stand out as the most lucrative sub-category, whereas tables, supplies, and bookcases have registered losses, with tables being the least profitable product. Within each sub-category, the consumer segment proves to be the most financially rewarding, followed by corporate and then home office.

Profit across different Sub-category within different Segment



---

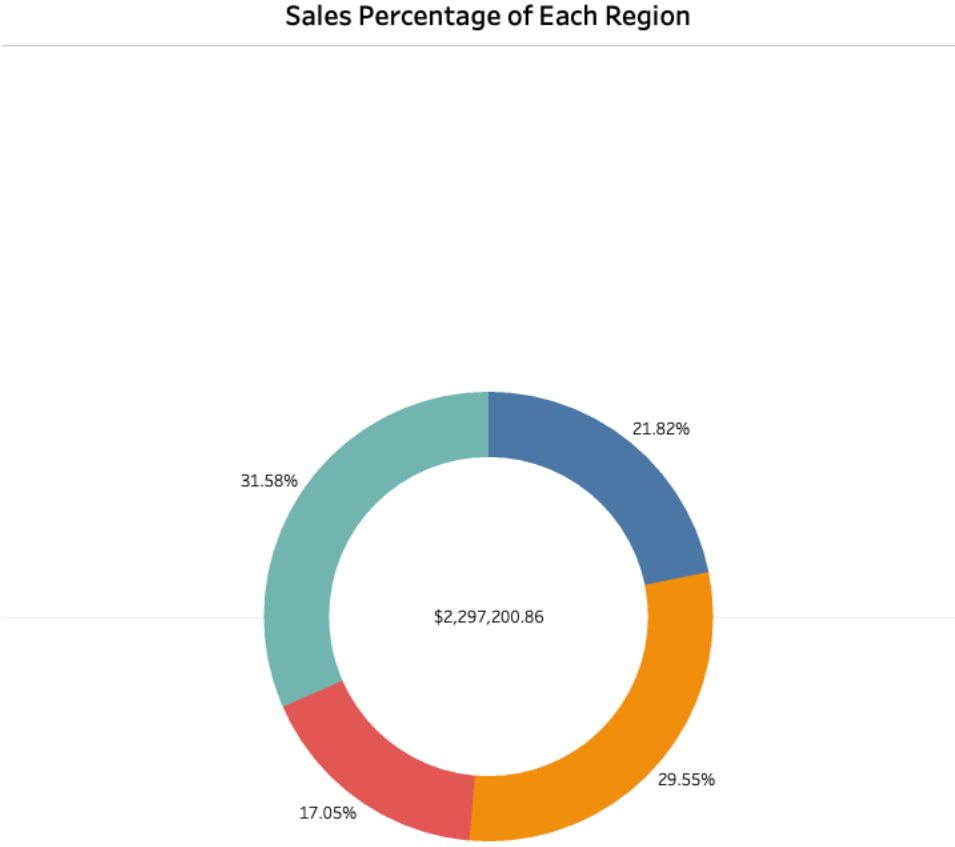
9. What is the percentage contribution of each region to the overall sales?

Chart Type – Donut Chart

Reason: Donut chart or Pie chart is a useful visualisation tool for representing the percentage distribution of categorical data.

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

Sales Percentage of Each Region



---

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

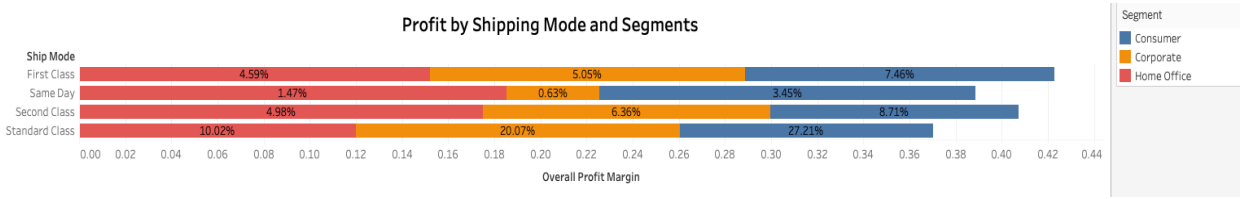
Chart Type – Stacked Bar Chart

Reason: Stacked bar charts prove to be a versatile visual aid for understanding and contrasting the makeup, proportions, and trends of various subcategories within categories. In this specific case, a stacked column chart is employed to illustrate the breakdown of profits among different subcategories for distinct customer segments. The x-axis showcases the diverse subcategories, and each bar represents the profit distribution among different customer segments within those subcategories.

Insight: The analysis reveals that first-class shipping mode has the highest profit margins, followed by second class, same day, and then standard class. In all shipping

modes, except standard class, the home office segment stands out as the most profitable. Under standard class, the corporate segment is the most profitable and least profitable in the same day mode.

Profit percentage by shipping mode and segments



Note: Total Profit Margin: Determining the overall profit margin for various segments is achieved by dividing the sum of profits by the sum of sales.

---

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

Chart Type - Stacked Bar Chart

Reason: Stacked bar charts prove to be a versatile visual aid for understanding and contrasting the makeup, proportions, and trends of various subcategories within categories. In this specific case, a stacked column chart is employed to illustrate the breakdown of profits among different subcategories for distinct customer segments. The x-axis showcases the diverse subcategories, and each bar represents the profit distribution among different customer segments within those subcategories.

Insight: It is noticeable that the average order processing time is approximately equal for various product categories, specifically around 4 days.

Average Shipping days for Product Category

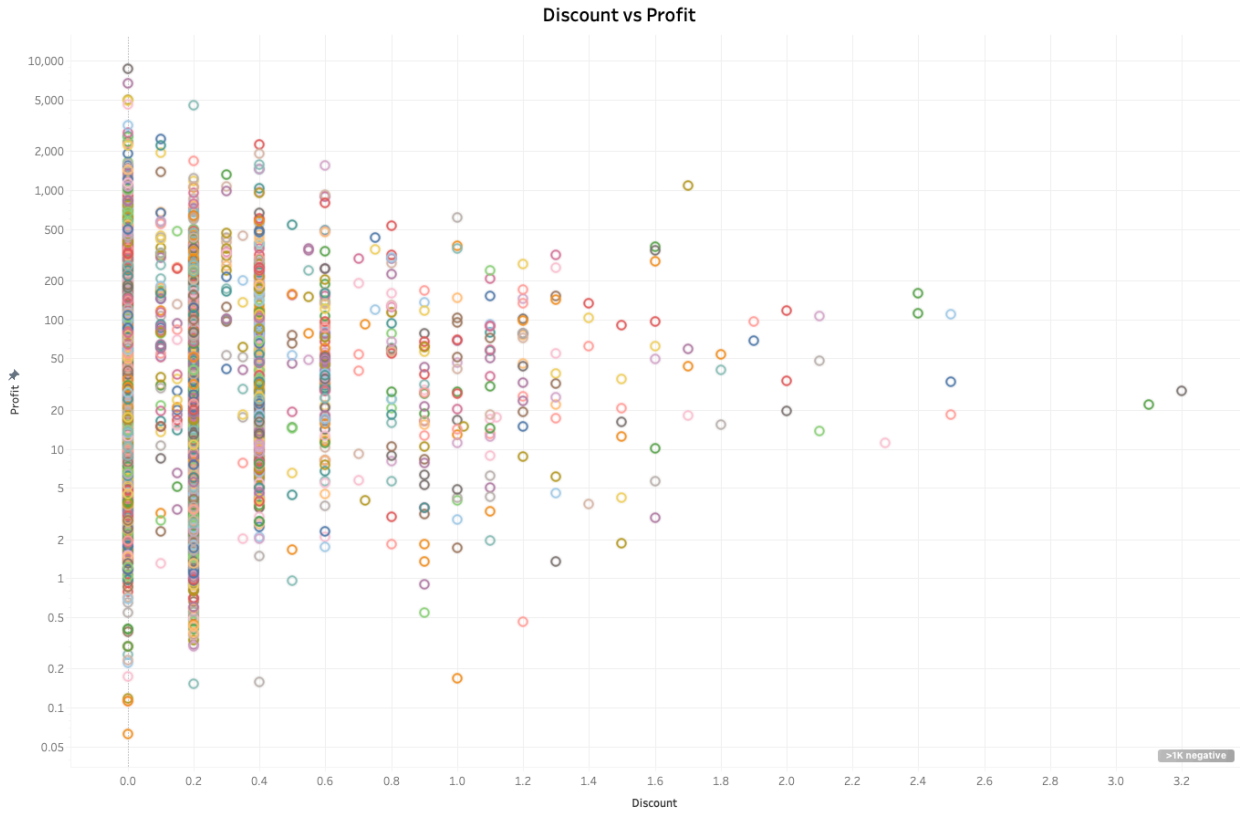
12. How do discounts affect overall profit?

Chart type – Scatter Plot

Reason: Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the relationship between "*discount"* and *"profit"*, we use a scatter plot.

Insight: There is no Linear relationship between discount and profit.

Discount vs Profit



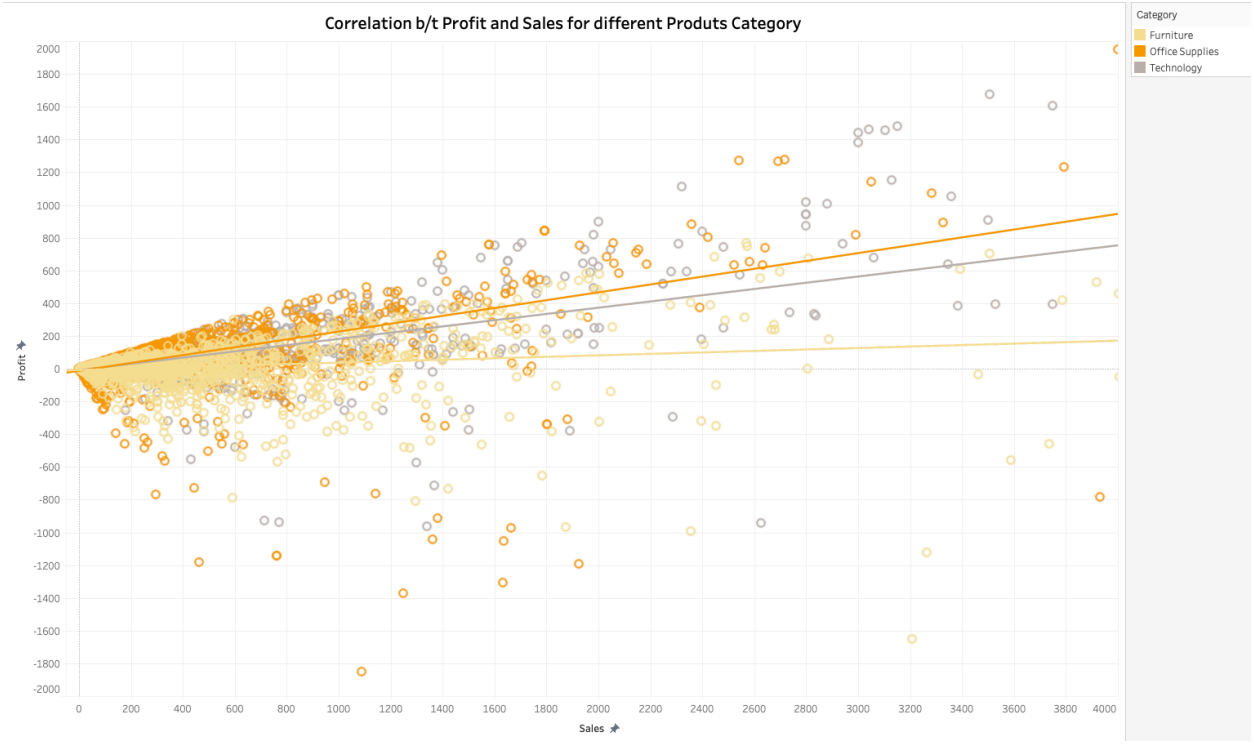
13. Can we visualise the relationship between product sales and profitability for different product categories?

Chart Type – Scatter Plot

Reason: Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the relationship between *Sales* and *Profit*, we use a scatter plot. *Sales* graph is plotted from 0 to 4000 on X-axis. Whereas, on Y-axis it is plotted from -2000 to 2000 for *Profit.*

Insight: As the p-value < 0.0001 for each category, Which means that the independent variable actually has an effect on the dependent variable. But R-Squared value for Furniture is 0.030, for Office Supplies is 0.314, and, Technology is 0.253. Which explains the variance in the dependent variable. *"a lower r-squared value tells you that your model explains less of the variance" .*

Relationship B/T Sales and Profit



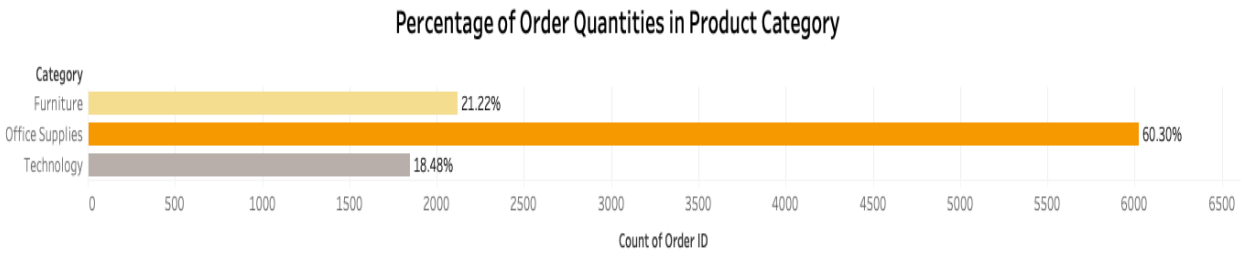
14. What is the distribution of order quantities for products in the dataset?

Chart Type – Bar Chart

Reason: Bars rule! They're perfect for comparing data across categories at a glance. And in this chart we will use bar chart to show distribution in percentage.

Insight: As we can see "*Office Supplies"* is more the double in comparison to other two Category, i.e *60.3%.* whereas, "*Furniture*" has around 21% and "*Technology*" has around 19% of total distribution.

Percentage of Order Quantities in Product Category



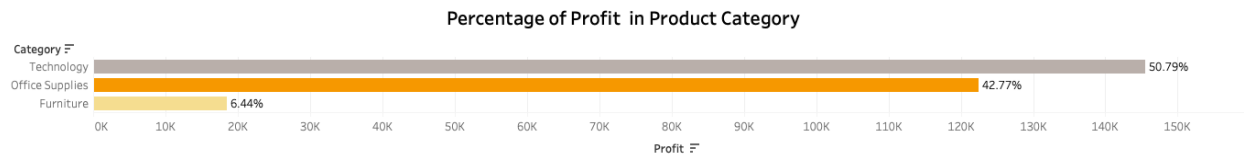
15. How do the profit distributions vary across different product categories?

Chart Type – Bar Chart

Reason: Bars rule! They're perfect for comparing data across categories at a glance. And in this chart we will use bar chart to show distribution in percentage for category.

Insight: As we can see "*Technology*" is just *1%* more than *50%* is almost twice then "*Office Supplies*" and "*Furniture*" that are *43%* and *7%.*

Percentage of Profit in Product Category



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

Chart Type – Bar Chart

Reason: Bars rule! They're perfect for comparing data across categories at a glance.

Insight: It can observed that same day orders are delivered on that day exactly. Whereas, Second class and First class take 3 and 2 days on average to deliver the package. But Standard class take at-least 5 days on average to deliver the courier.

Note: Adding calculated table name Shipping days which tells us how much days it take to deliver the package using this formula (ROUND([Ship Date]-[Order Date],0))

Average Shipping Days vs Shipping Modes

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

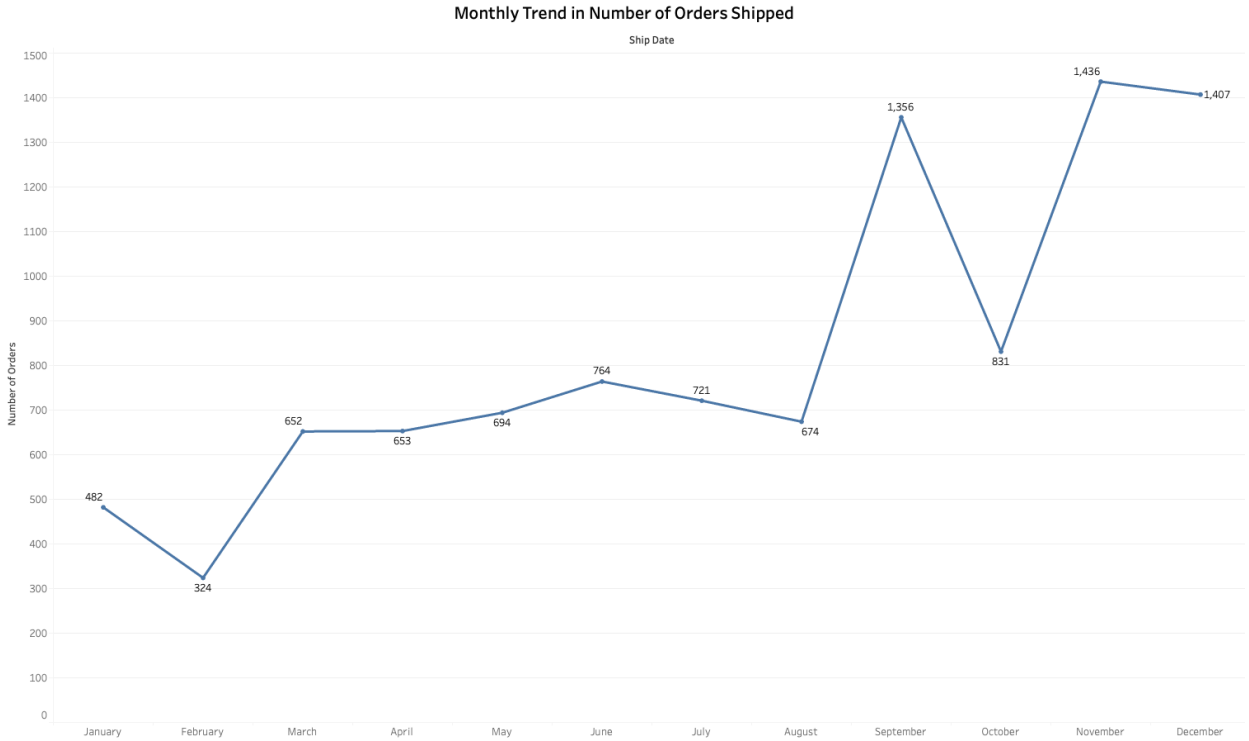
Chart Type – Line Chart

Reason: A line chart is used for visualising and analysing data trends over time.

Line chart is therefore used here to display the monthly trend in the number of orders Shipped.

Insight: We can observe that as the year comes to end Trend shows a positive inclination from January to December. However, in the month of September(1356) the orders were almost doubled in quantities in comparison to August month(674), But it can be observed that orders placed was declined rapidly in October month and gain peeked in November month and stabled by the December month around 1407 orders.

Monthly trend in number of Ordered Shipped



On the other-hand average order place were around 693 from March to August.

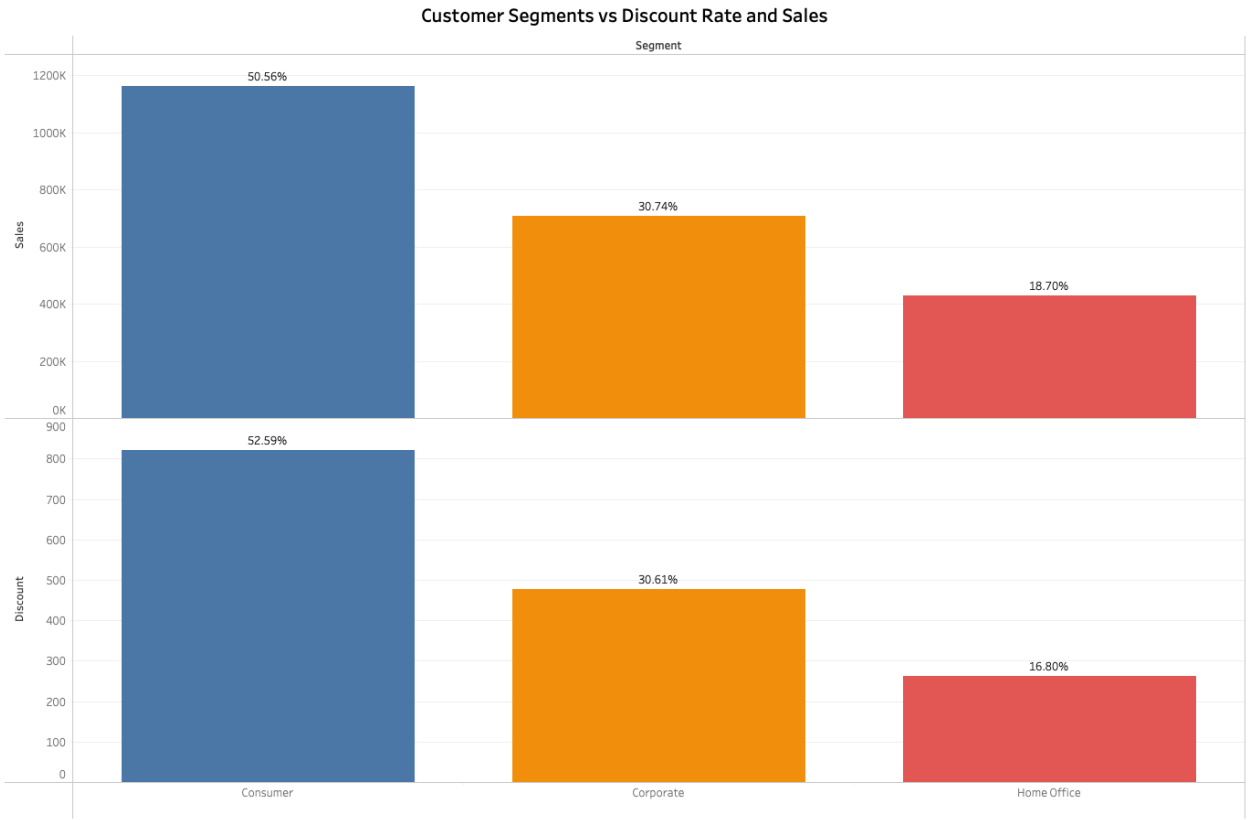
18. How do different customer segments perform in terms of sales and discount rates?

Chart Type – Column Chart

Reason: A column chart is a useful visualisation tool for representing the distribution of categorical data. Here I am going to use two column chart discount and sales in different segments

Insight: It can be observed that Home Office contributes the lowest amounts of sales (*19%*) followed by Corporate-segment which is (*31%*), The highest contributor is Consumer segment nearly contributing around (*50%*), which is half of total contributions. Whereas, if we talk about Consumer, Corporate, and Home-Office segments in Discount Rate *52.59%, 30.61%,* and*, 16.80%* is the contribution respectively.

Customer Segments vs Discount and Sales



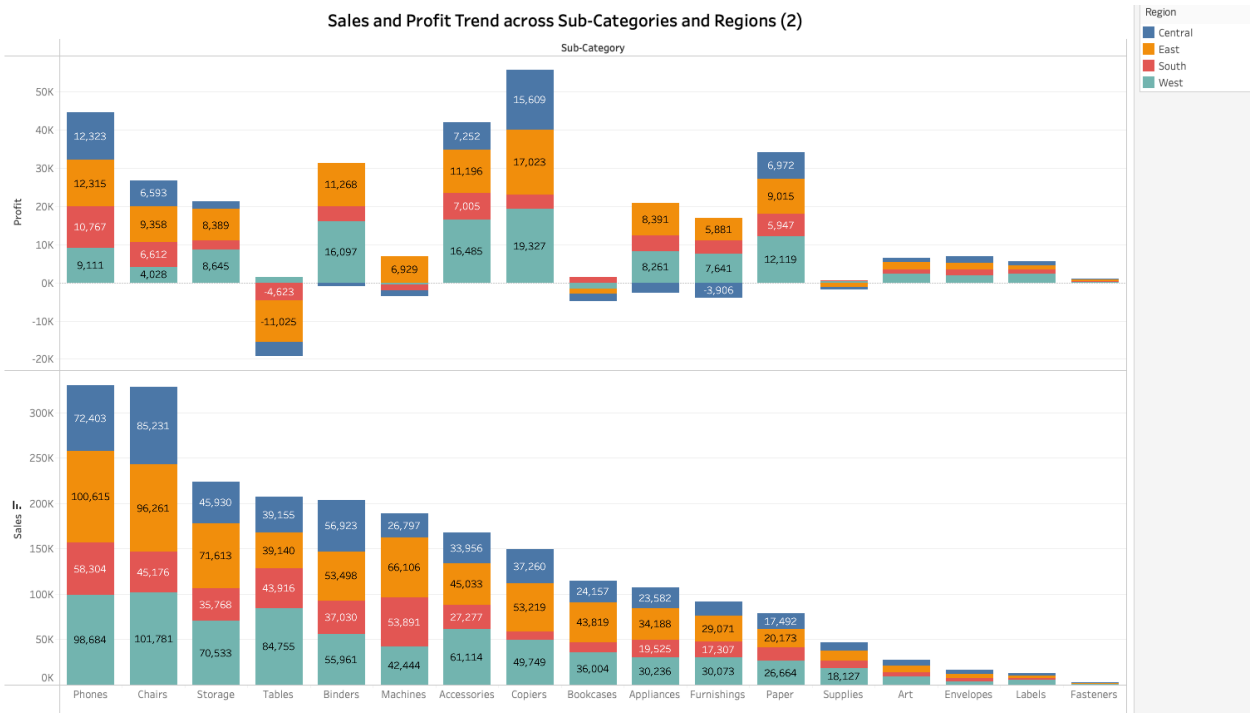
19. What are the sales and profit trends across different product subcategories and regions in the Superstore dataset?

Chart Type – Stacked Column Chart Sub-Plot

Reason: Stacked column chart for overall understanding to sales and profit across different region and individual contribution in profit and sales. Whereas, Sub-plot for in-depth analysis in four different regions like how much profit they are generating in one region as compared to other regions and is it necessary to restock them in those regions.

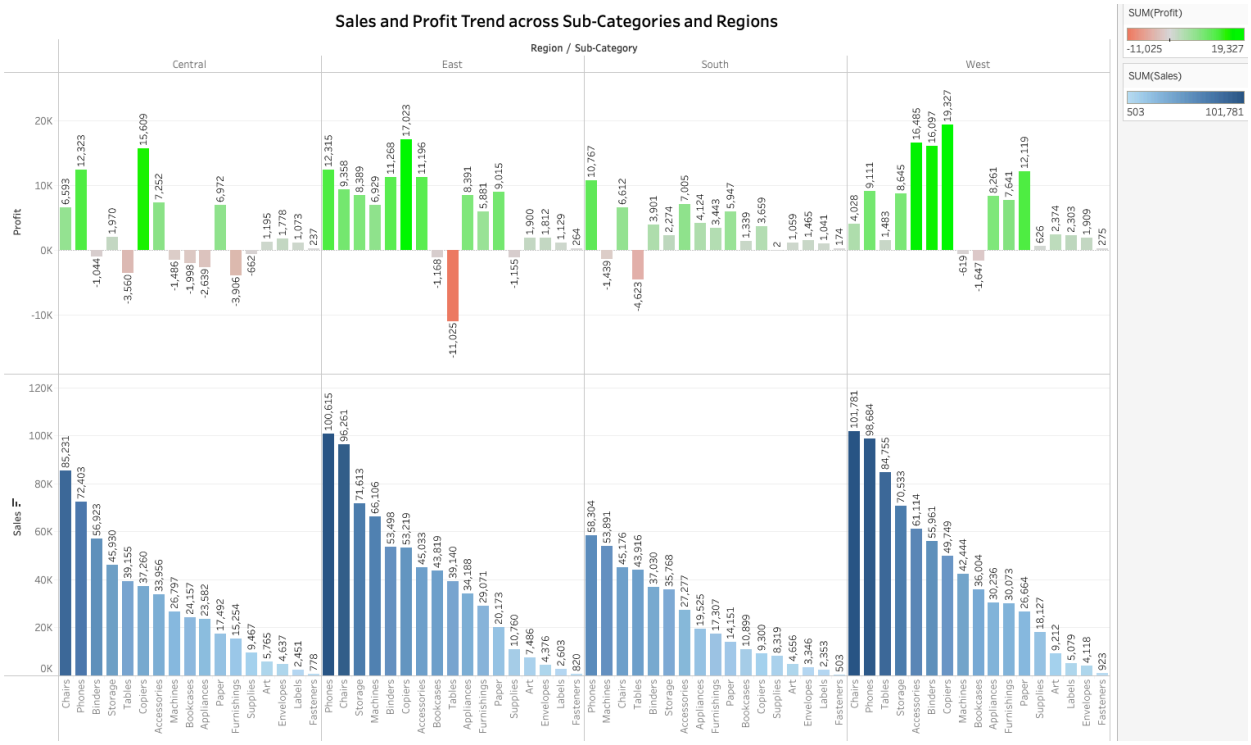
Insight: In Stacked column chart we can see the sale of Phones and Chairs are high as compared to other products but they are generating less profit than Copiers and Accessories. On the other-hand Machines are only profitable in East Region. Tables and Bookcases are profitable in West and South Region. We might only restock them in those region which are generating profit for the company.

Sales and Profit Trend across Sub-Categories and Regions



In sub-plot we can observe that in Central Region there are "*seven" Sub-categories that are responsible for loss generation that are (Binders, Tables, Machines, Bookcases, Appliances, Furnishing, and, Supplies) and highest of them is generating $3,560(Tables) in loss in Central Region. Whereas, in East Region highest recorded loss is of $11,025 in Tables category apart then this in east region Copiers are generating highest profit of $17,023 alone, and the highest sold product is Phone in this region generating around 12% of its total sales value in profit. If we talk about South Region and West Region tables are still in loss in south but are in little profit in west. Phones is generating highest profit in south. whereas, Chairs is highest in west and central region. We can get many more in-depth data from this sub-plots.*

Sales and Profit Trend across Sub-Categories and Regions



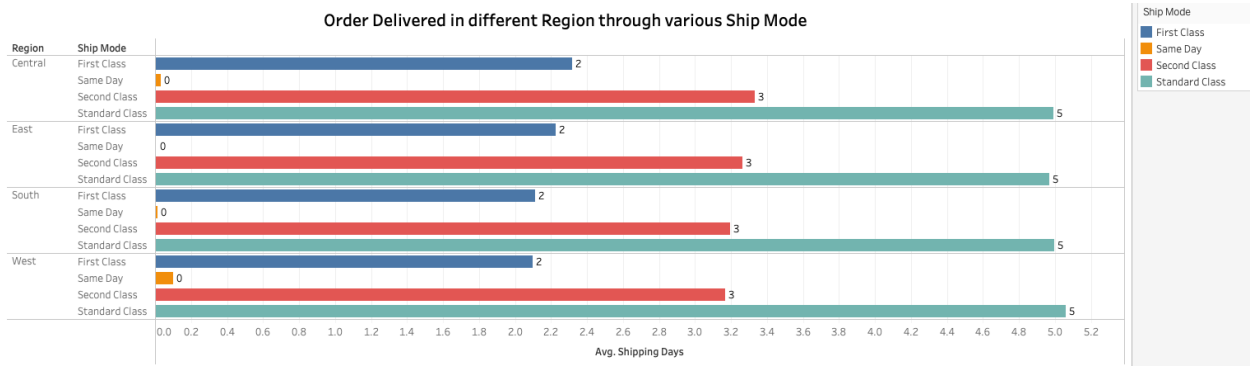
---

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

Chart Type – Bar Chart

Reason: Bars rule! They're perfect for comparing data across categories at a glance. To get better understanding of ship mode within regions

Order delivered in different Region through various Ship Mode



Note: Adding calculated table name Shipping days which tells us how much days it take to deliver the package using this formula (ROUND([Ship Date]-[Order Date],0))

Insight: Average delivery time of product by First Class is 2 days which is same for each region. Similarly Second Class and Standard Class take 3 days and 5 days on average to deliver order respectively in each region. Individuals choosing same day ship mode get there product by end of that day in each region.

---

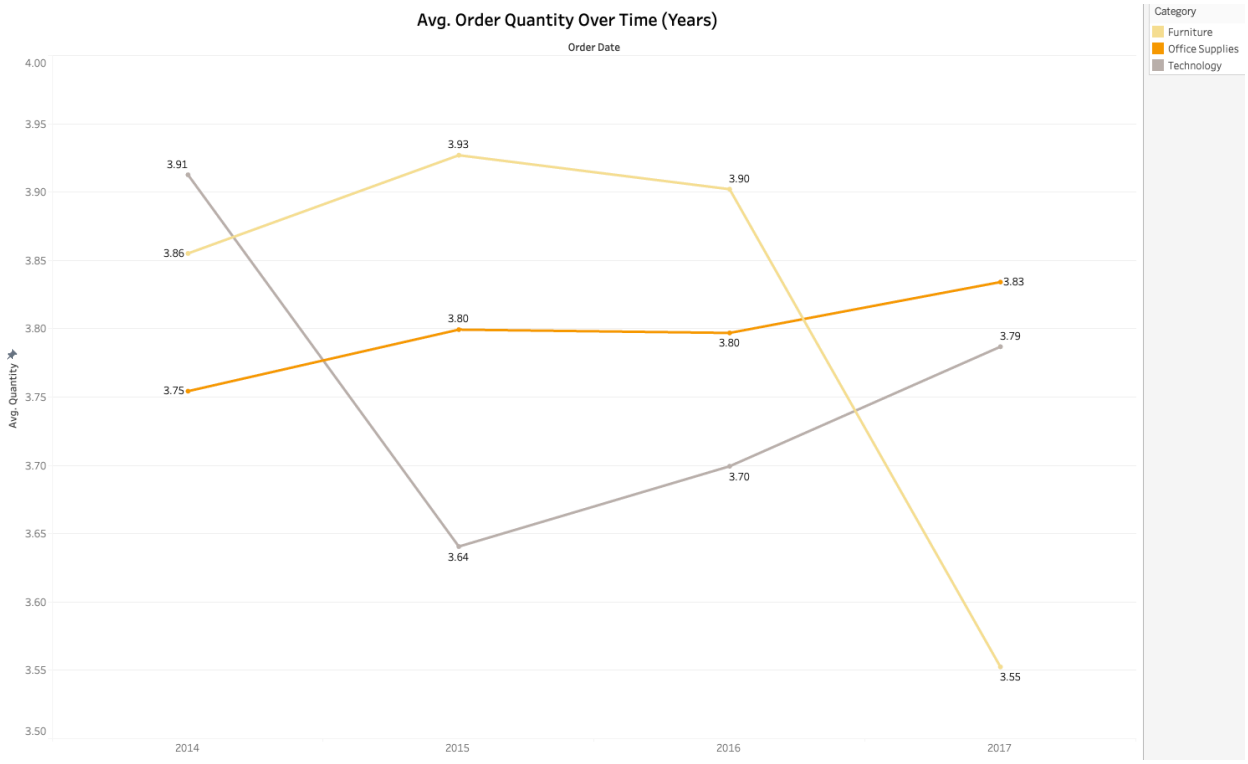
21. How has the average order quantity changed over the years for various product categories?

Chart Type – Line Chart

Reason: We have to show average change in quantities over the years, and to show change over time. Line chart is best option for this.

Insight: The average order quantity showed distinct variations over the years. In 2014, technology topped the list with the highest average order quantity, succeeded by furniture and then office supplies. However, this sequence changed in 2015 and 2016, as furniture took the lead, followed by office supplies and then technology. Another shift occurred in 2017, where office supplies claimed the highest average order quantity, followed by technology and then furniture. Despite these fluctuations, the overall average order quantity for all categories typically fell within the range of 3 to 4.

Average Order Quantity over Time(Years)



---

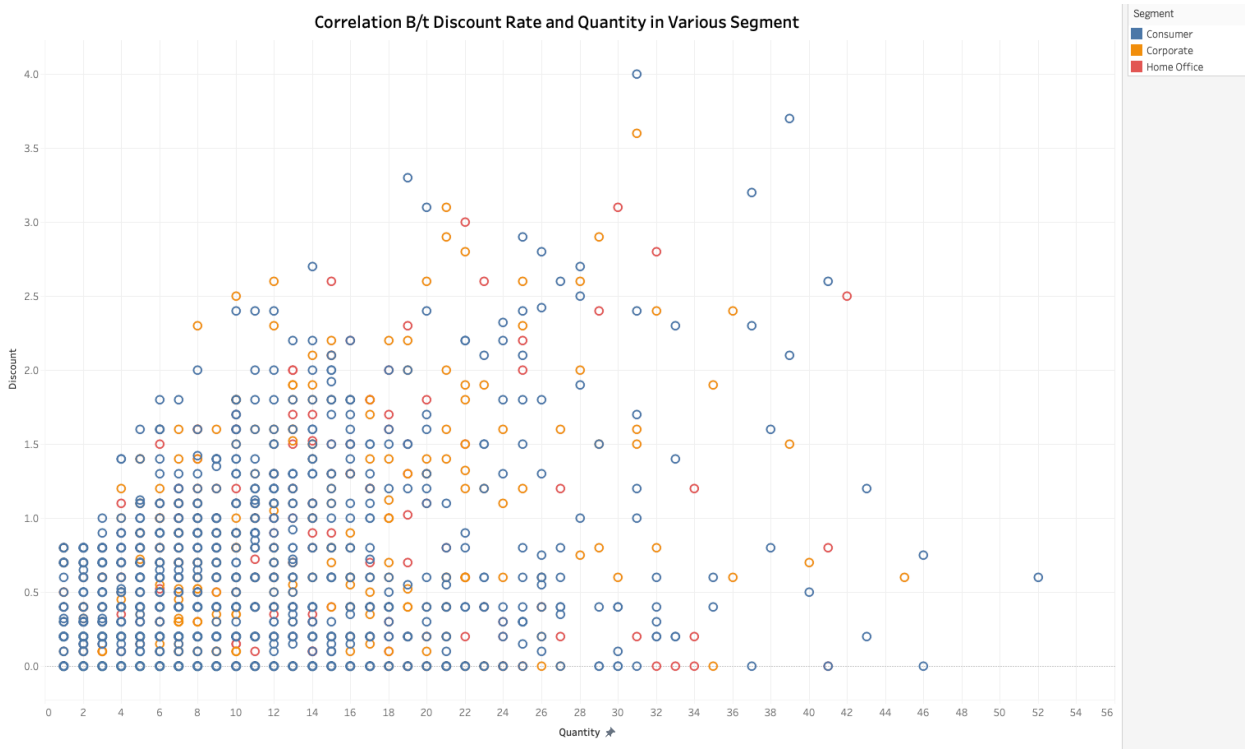
22. Can we visualise the correlation between discount rates and order quantities for different customer segments?

Chart Type – Scatter Plot

Reason: Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the relationship between "*Discount"* and *"Quantity"*, we use a scatter plot.

Insight: There is no relationship between Discount rate and Quantity throughout the different segments. Apart then this the range of X-axis's is from 0 to 56, and range of Y-axis's is from 0 to 4.

Correlation b/t Discount Rate and Quantity in Various Segment



---

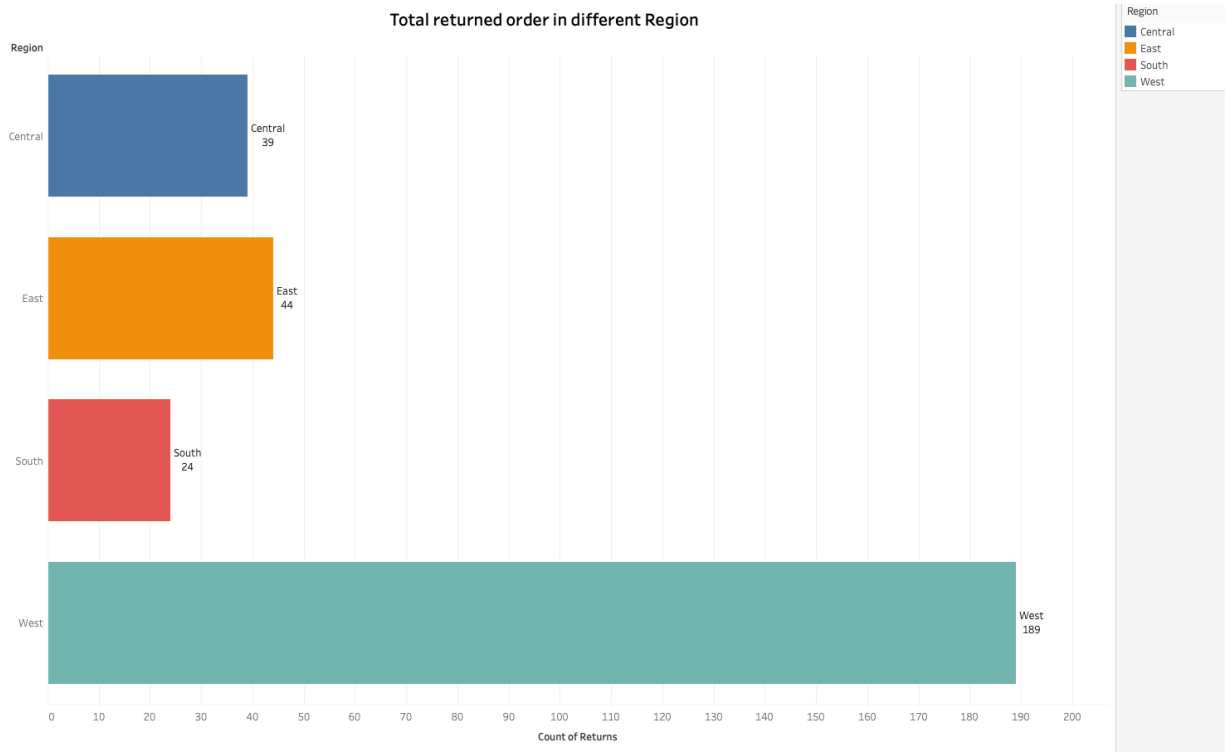
23. What is the proportion of orders returned in each region within the Superstore dataset?

Chart Type – Bar Chart

Reason: Easy to understand and much simpler to decode.

Insight: As we can see most returned orders are from West Region (189) followed by East (44) then Central (39) and on the least number of returned are from South Region (24).

Total Number of Returned Order in Different Regions



---

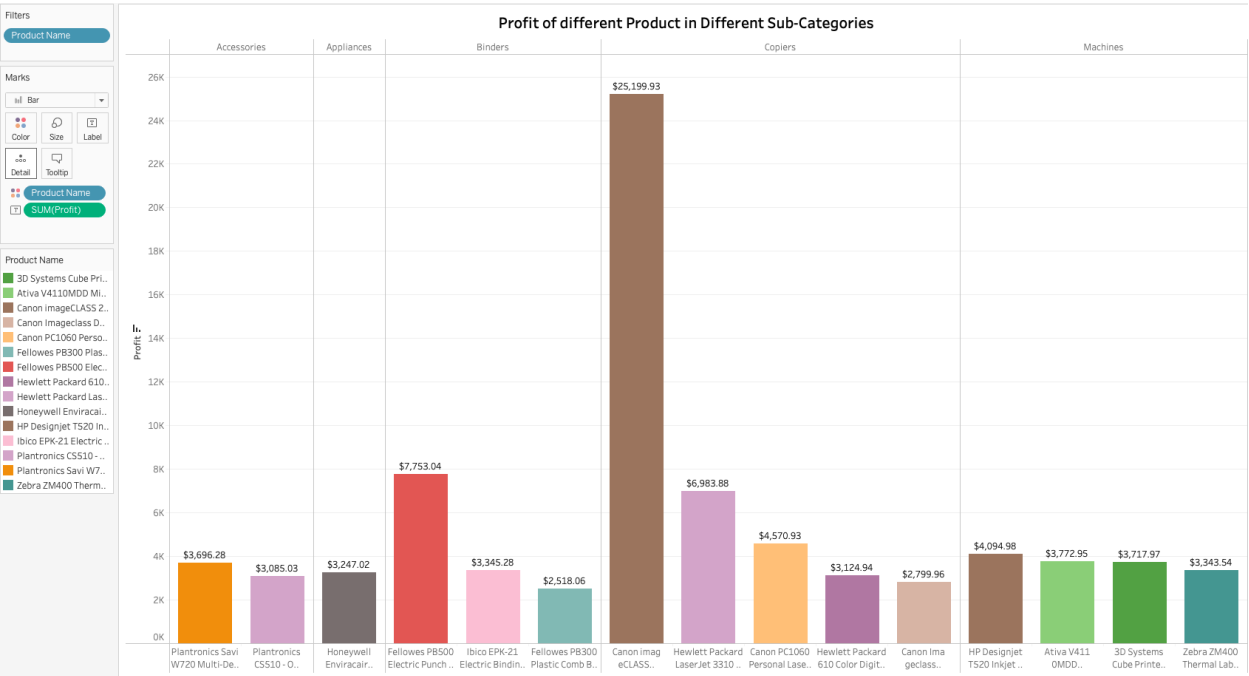
24. Can you compare the profit of different products for different subcategories?

Chart Type – Column Chart

Reason: They're perfect for comparing data across categories at a glance. To get good understanding of which Product is more profitable in Sub-Categories. This chart will help us to analyse which product is generating how much profit.

Insight: As there are ample number of products in Sub-Category, we will use filter option on Product Name and will only use 15 product for now, we can change this to maximum products in sub category. As we observed from data "*Canon imageCLASS 2200 Advance Copier*" is the highest profitable product in the entire sub-category generating around $25.1K in US Dollar, then there are "*Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind"* and "*Hewlett Packard LaserJet 3310 Copier*" generating around $7.7K and $7k in "*Binders"* and "*Copiers"* Sub-Category. And rest of the product fall under the range of $4.5k – $2.5k US Dollar.

Profit of different Products in different Sub-Categories



---

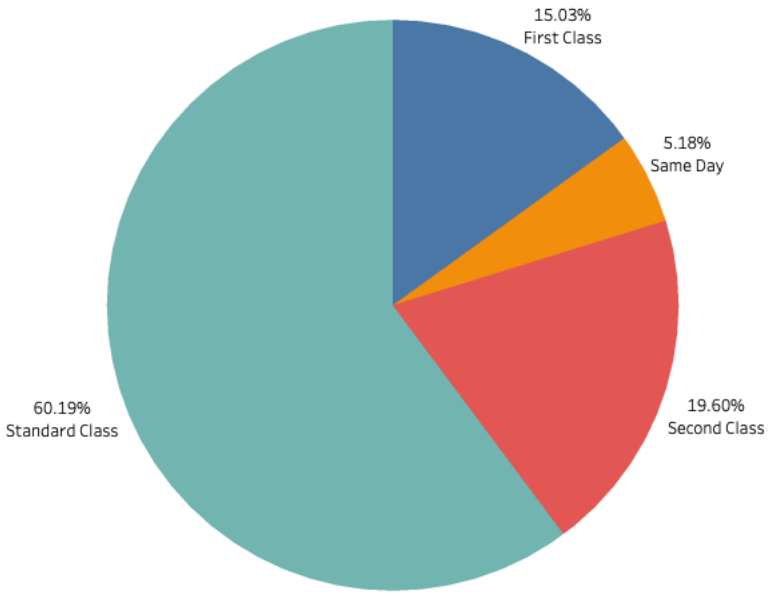
25. Which shipping mode is the most commonly used in the Sample Superstore dataset?

Chart Type – Pie Chart

Reason: To show the most common Shipping modes we can use pie chart, Which are an excellent chart to show distribution using a proportion.

Insight: As we can observe that "*Standard Class*" is the mostly used shipping mode among the users, it is used by *60%* of the population. Furthermore*, 20%* and *15%* users use "*Second Class"* and "*First Class*" shipping modes respectively. There is minority of 5% users choosing "*Same Day"* delivery mode.

Proportion of Shipping Modes



---

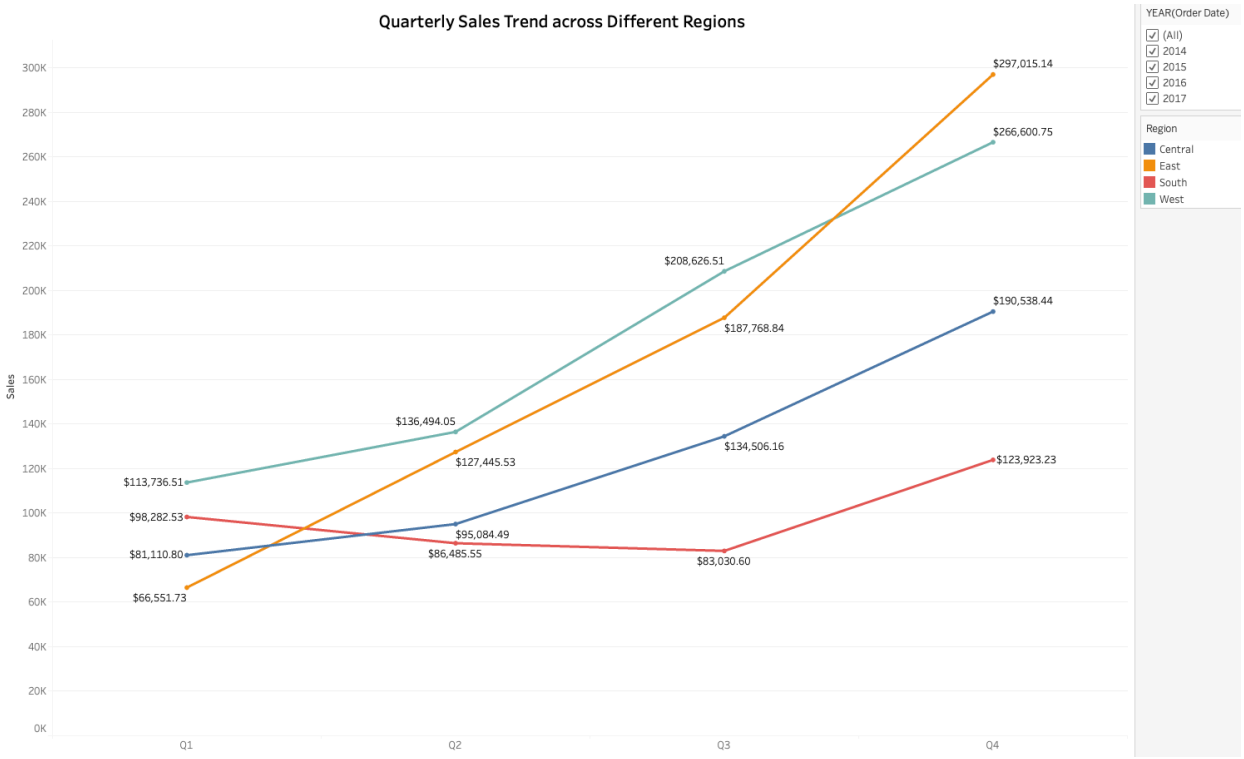
26. How does the sales performance of different regions evolve throughout the quarters of a year?

Chart Type – Line Chart

Reason: Trend over Time! There is no better option then Line chart to show data over time. Line chart is used to show change over the years here we will use this chart to give insights of quarterly sales performance in different regions.

Insight: Here we are looking at overall data of last four years (We can also select individual year or multiple years as per our requirements). By looking at the chart we can see that East region has made lowest sales in Q1 and by the end of Q4 it had made around 4.5 time of Q1 in last four years. Whereas, every region is showing a positive trend line but if we talk about South Region in Q1 it is at second highest in terms of sales but in Q2, Q3, and Q4 it is constantly at the lowest for rest of the quarters. West Region show Exactly opposite trend of South region it was consecutively on top for three quarters but in Q4 it is at second position.

Quarterly Sales Trend across Different Regions



---

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

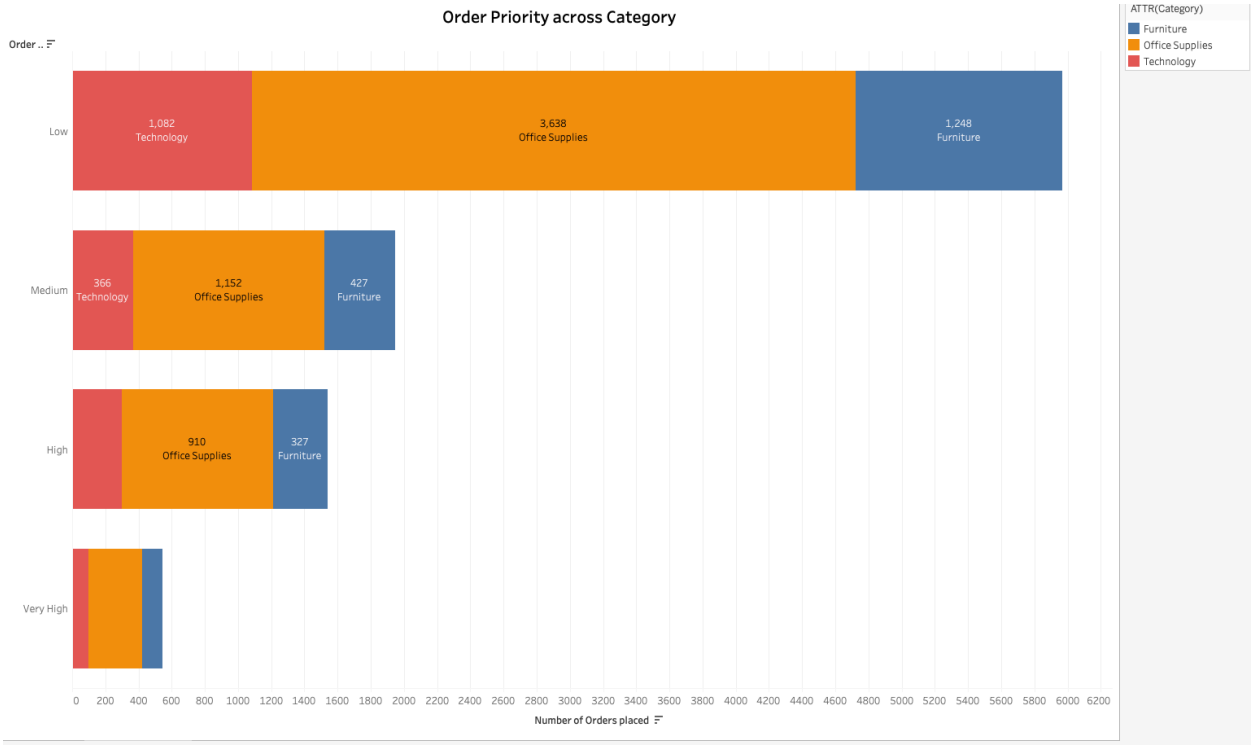
Chart Type – Stacked Bar Chart

Reason: To show one numerical and two categorical data into visualization Stacked Bar Chart is best. Here we have Order Priority, Category and Number of Orders to display, So we will use this chart.
Note: Here End user selecting shipping mode which tells us about their order priority. So, we use formula to convert shipping mode into order priority.
Formula for Order Priority Column.

```
CASE [Ship Mode]

WHEN 'Same Day' THEN 'Very High'

WHEN 'First Class' THEN 'High'

WHEN 'Second Class' THEN 'Medium'

WHEN 'Standard Class' THEN 'Low'

ELSE 'Other'

END
```

Insight: Maximum users (5968) comes under "*Low Priority*" and Minimum numbers of users (543) comes under "*Very High Priority*". (1945) are given "*Medium Priority*" and (1538) are given "*High Priority*". If we take a good look we can see that "Office Supplies" is the largest segment which have highest order placed in every order priority list.

Order Priority across Category



---

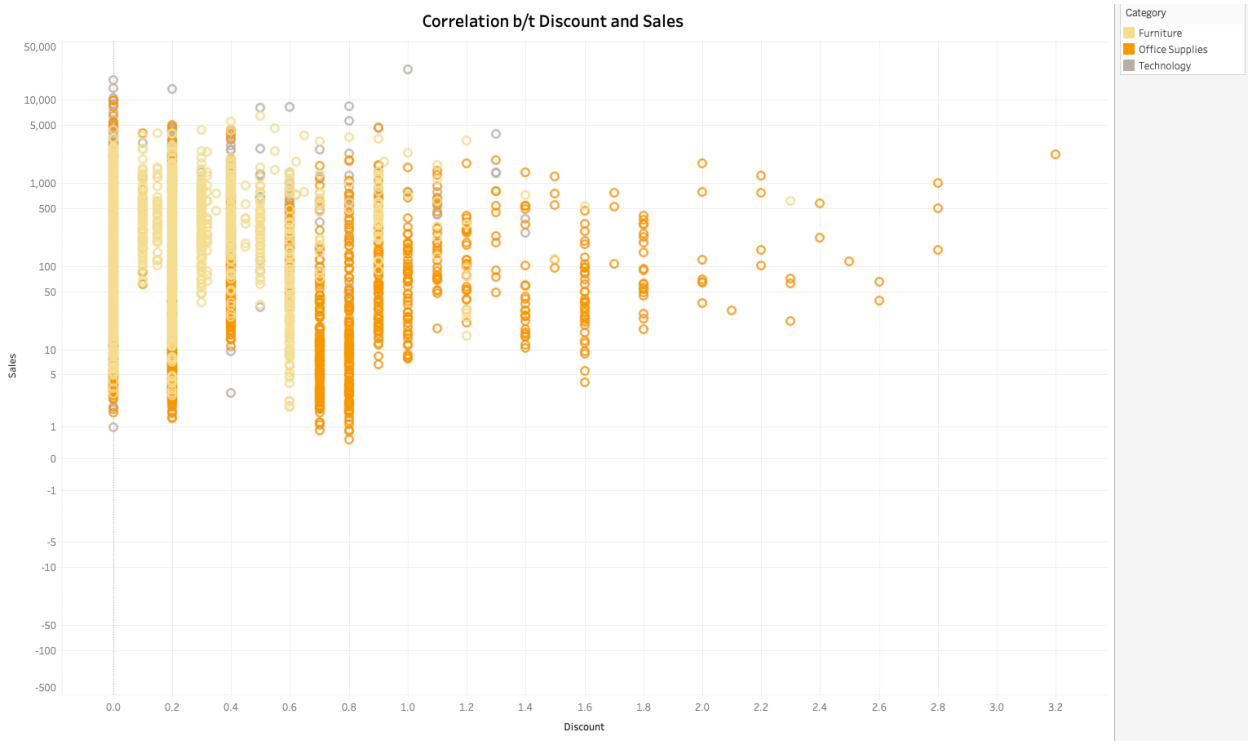
28. What is the relationship between discounts and sales?

Chart Type – Scatter Plot

Reason: Scatter plots are useful for displaying the relationship between two numerical variables. Here, since we want to visualise the relationship between "*Sales*" and "*Discount".* we use a scatter plot.

Insight: There is no relationship between Sales and Discount Rate throughout the different Category. Apart then this the range of X-axis's is from 0 to 3.2, and range of Y-axis's is from -500 to 50,000. There is no major impact on discount on increase in sale.

Correlation b/t Discount Rate and Sales



---

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

Chart Type – Bar Chart

Reason: Bars rule! They're perfect for comparing data across categories at a glance. To get better understanding of average amount spend by new and existing customers.

Note: To identify which customers are new or old we will use this formula used in New or Old Customer Column, Which will help us to Visualize the data.

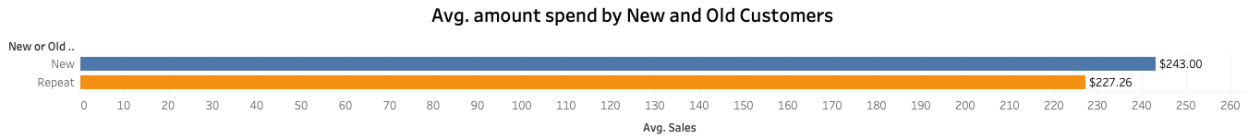
***Formula**: IF { FIXED [Customer ID] : MIN([Order Date]) } = [Order Date] THEN 'New'*

ELSE 'Repeat'

END

Insight: Average amount spend by "*New Customer*" is $243. Whereas, "*Old Customer*" spend around $228 every time the buy something new. Clearly New Customer spends more amount then Existing Customers.

Average Sales vs Customer Type



---

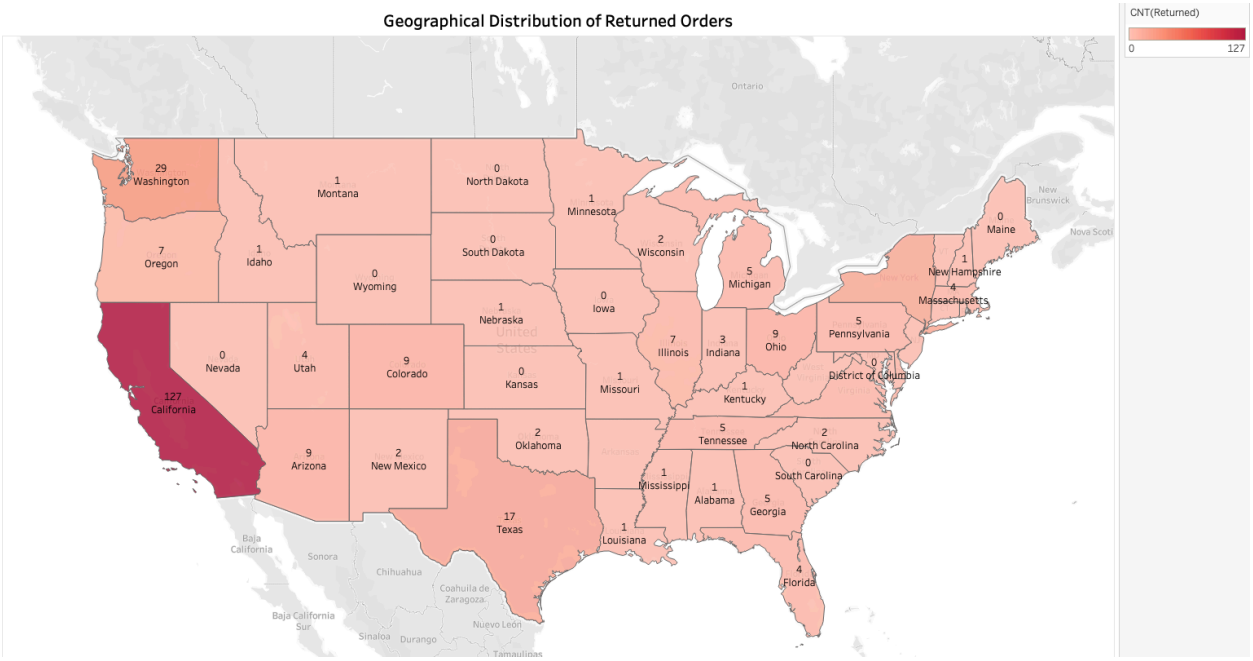
30. What is the geographical distribution of returns and its impact on overall profitability?

Chart Type – Map and Text(Table)

Reason: Map is first choice to show geographical distribution od returned orders. And Table chart is used to show simple Text format table displaying Profit margin before and after returning orders.

Insight: Maximum order returned are from State "*California*" 127 returned order, followed by "*Washington"* and "*Texas*" 29 and 17 respectively. Whereas, "*North Dakota, Wyoming, South Dakota, Lowa, Kansas, Nevada, Arkansas, Maine, District of Columbia, South Carolina, Connecticut, Vermont, and West Virginia"* are with zero order returned.

Geographical Distribution of Returned Orders



Its impact on Overall Profitability

We have used some formula to get the overall profit margin and profit margin (after neglecting returned orders).

1. Overall Profit Margin = ROUND((SUM([Profit]/SUM[Sales])),4)
2. Profit on Returned items = CASE [Returned] WHEN 'Yes' THEN [Profit] ELSE 0 END
3. Sale on Returned Items = CASE [Returned] WHEN 'Yes' THEN [Sales] ELSE 0 END

4. Profit Margin (After Neglecting Returned Orders) = ROUND((SUM([Profit]) - SUM([Profit on Returned])) / (SUM([Sales]) - SUM([Sale on Returned])), 4)

Insight: It is noticeable that even with returned orders, the overall profitability remains relatively stable. The profit margin only experienced a slight decrease of 0.04% as a result of returns.

Impact of Returned Orders on Overall Profitability

Overall Profit Margin	12.47%
Profit Margin (After Neglecting Returned Orders)	12.43%

Skills Earned

Through the exploration and practice of Tableau, a diverse set of skills has been acquired in creating various type of visualizations. The proficiency gained includes:

Line Chart:

- Mastery in representing trends and patterns over time through the use of line chart.

Bar Chart:

- Competence in effectively displaying and comparing data values using bar charts.

Stacked Column Chart:

- Proficiency in showcasing the composition and cumulative values of multiple data categories with stacked column charts.

Pie Chart:

- Skilfulness in illustrating proportional relationship and part-to-whole comparison using pie charts.

Donut Chart:

- Capability in enhancing visual appeal and readability by creating insightful donut charts for data representation.

Scatter Chart:

- Expertise in revealing relationships and correlations between two variables through the creation of scatter charts

Map Chart:

- Adeptness in utilizing geographical data to create informative map charts for spatial analysis and representation.

These acquired skills not only contribute to a comprehensive understanding of data visualization but also empower effective communication of insights across diverse datasets.

Created by: ***Pawan Rana***

LinkedIn : Pawan Rana

Github : Rana-Pawan