

Sardar Patel Institute of Technology

SEM VII:ADVANCE DATA VISUALIZATION.

Name	Smruti Sonekar
UID no.	2021700064
Branch	BE CSE DS (BATCH B)
Experiment no.	1

Topic:	Create basic charts using Tableau / Power BI / R / Python / D3.js to be performed on the dataset of the Ecommerce field.
Theory:	<p>Complete all plots on practice dataset and reproduce on e-commerce dataset.</p> <ul style="list-style-type: none">● Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot● Calculate Product wise sales, region wise sales● Write observations from each chart. <p>Dataset: https://www.kaggle.com/datasets/bhanupratapbiswas/superstore-sales</p> <p>"Superstore Sales" generally refers to the sales data and performance of a superstore or a large retail store that offers a wide range of products and merchandise. These superstores are typically known for their extensive inventory, competitive pricing, and one-stop shopping experience.</p> <p>The "Superstore Sales" dataset contains detailed information about sales transactions from a large retail store. It includes the following columns:</p>

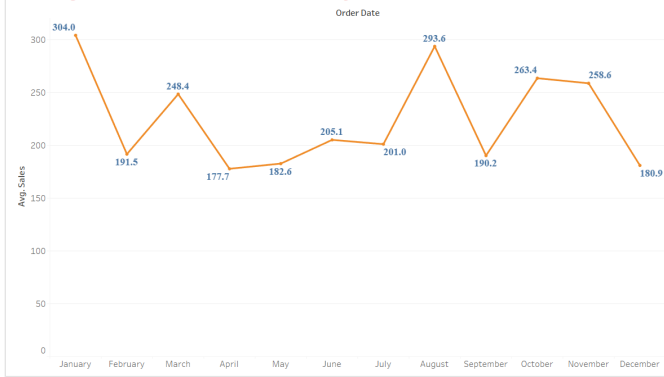
	<ul style="list-style-type: none"> ● Row_ID: Unique identifier for each record. ● Order_ID: Identifier for each order. ● Order_Date: Date when the order was placed. ● Ship_Date: Date when the order was shipped. ● Ship_Mode: Shipping method used for the order. ● Customer_ID: Identifier for each customer. ● Customer_Name: Name of the customer. ● Segment: Customer segment (e.g., consumer, corporate, home office). ● Country: Country where the customer is located. ● City: City where the customer is located. ● State: State where the customer is located. ● Postal_Code: Postal code of the customer's address. ● Region: Region where the customer is located. ● Product_ID: Identifier for each product. ● Category: Product category (e.g., Furniture, Office Supplies, Technology). ● Sub_Category: Specific sub-category of the product. ● Product_Name: Name of the product. ● Sales: Sales amount for the product.
Program:	<p>1) Data Preprocessing:</p> <p>First, we check out the dataset , and find any missing values , postal_code had 11 missing values but for that rows we have city,country and state so we have replaced Nan values with generalized zip code.</p> <p>We have orderDate and Ship Date in string format so converted into datetime format for easier to work in tableau.</p> <pre>missing_postal_code_rows = d1[d1['Postal_Code'].isnull()] d1['Postal_Code'].fillna('05401', inplace=True)</pre>

```
d1['Order_Date']=pd.to_datetime(d1['Order_Date'],format='%d/%m/%Y')
d1['Ship_Date']=pd.to_datetime(d1['Ship_Date'],format='%d/%m/%Y')
```

2) Different Data Visualization:

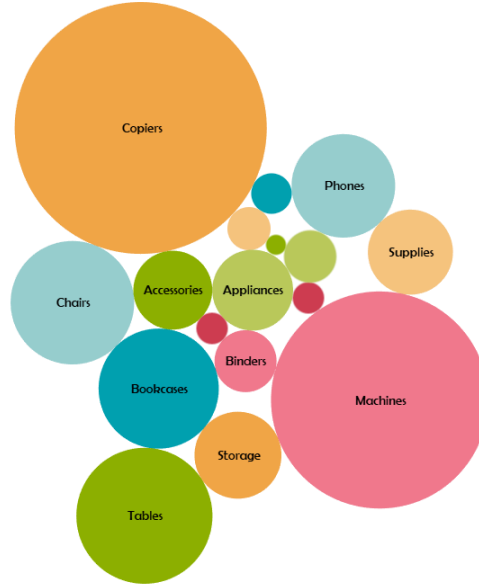
- **Line Plot:** I have data from 2015, 2016, 2017, and 2018, so we can see the average sales for the entire year 2018 below, broken down by month. According to the statistics, January marked the top of the average sales for 2018, suggesting a possible seasonal tendency. Using this representation, you may see trends and decide on the best course of action for your next sales tactics.

Average Sales Distribution over the year 2018(Line Plot)



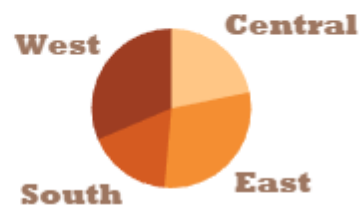
- **Bubble Plot:** A bubble plot in your dataset visually compares the average sales across different sub-categories. The **size of the bubbles** represents the average sales, allowing you to easily identify which sub-categories perform better or worse. Larger bubbles indicate higher average sales, while the plot's positioning shows the relative performance of each sub-category. If color coding is used, it can further differentiate categories, providing a clear and intuitive overview of the data.

Sales distrubtion sub category wise (Bubble Plot)

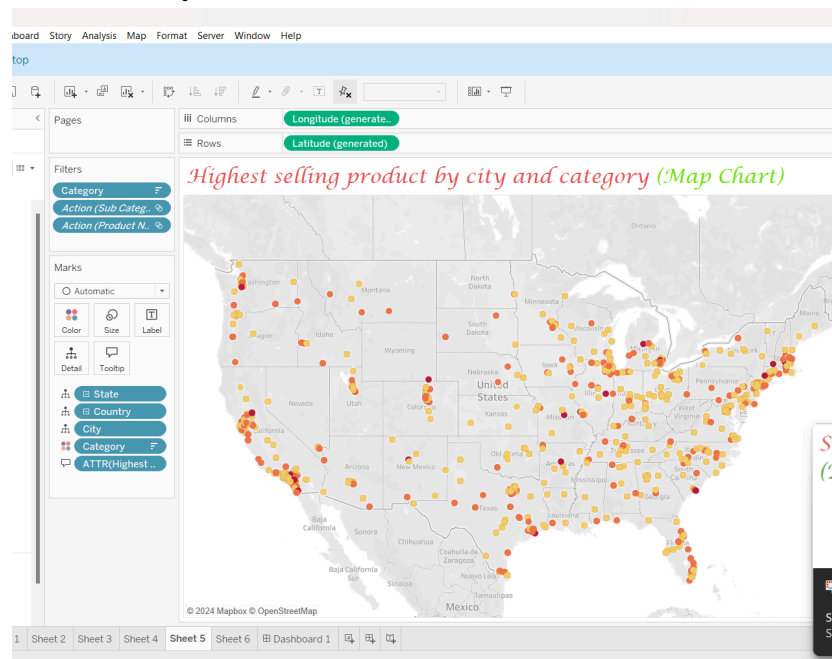


- **Pie Chart:** Showing the total sales for all four years by area (east, west, central, and southern), it can be inferred that the majority of business comes from the West region of the United States while the South region contributes less to the total amount of business. Businesses may maximize revenues by using this information to assist them strategy and concentrate their efforts on areas with higher sales potential. To find areas for expansion, it might also be helpful to investigate the causes of the Southern region's lower sales figures in more detail.

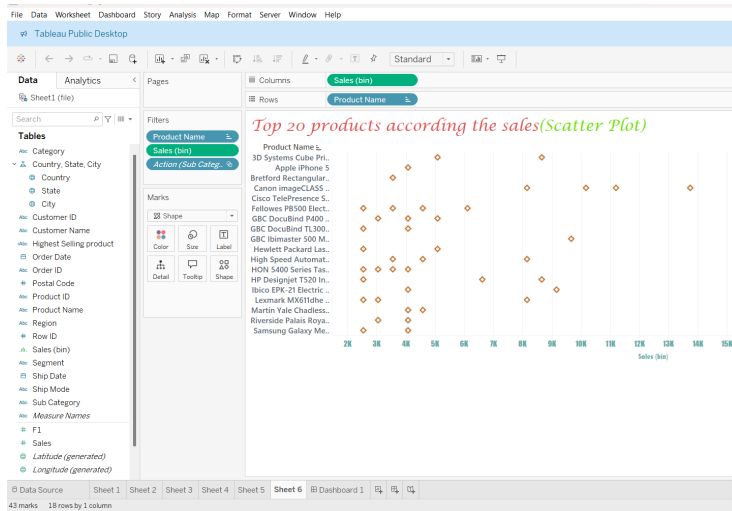
Sum of Sales RegionWise (Pie Chart)



- **Map Chart:** In tableau , for every state count of orders and the highest selling product in every state with respective category is depicted , used calculated formula : IF { FIXED [State]: MAX([Sales]) } = [Sales] THEN [Product_Name] END .To create a comprehensive Tableau map chart, first generate a map with the number of orders, then identify popular categories and the highest-selling products. Sort by Sales and create a calculated field for each state. Enhance tooltips to display order distribution, popular categories, and top-selling products. Adjust colors, sizes, and labels for a visually informative and user-friendly visualization.

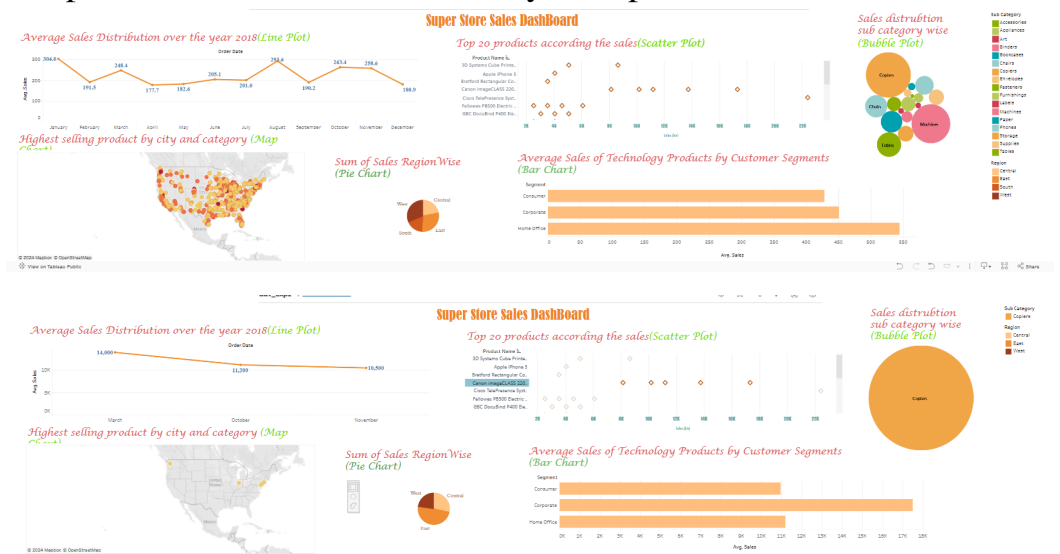


- **Scatter Plot:** Here we have used filter option in tableau the top 20 highest selling products according to the sales and at least the sales has to be 5k This allows us to focus on the most successful products and make data-driven decisions based on their performance. By setting a minimum sales threshold of 5k, we ensure that only significant sales data is included in our analysis.



Result:

For the final dashboard, I combined all the individual charts into a single, cohesive dashboard in Tableau. To enhance interactivity, I made the dashboard dynamically responsive by utilizing the "Use as Filter" option on both the scatter plot and bubble plot. This feature allows users to click on any of the top 20 products, and the entire dashboard will automatically update to display analysis results specifically for that selected product. This approach ensures that all related visualizations, such as sales trends, geographic distribution, and category breakdowns, are filtered to focus on the product of interest, providing a comprehensive and interactive analysis experience.



Conclusion:	<p>The experiment taught me about various methods of data visualization, as well as how to represent and use category data in a way that summarizes superstore sales data. I also learned how to select the best visualization approach for the data being presented, as well as how to effectively express insights and trends via visual representations.</p> <p>Overall, this experiment helped me improve my abilities in data analysis and presenting. I also learnt a lot about Tableau's capabilities, strategies, and options, which helped me construct a dashboard.</p> <p>Here is the link of tableau:</p> <p>https://public.tableau.com/views/adv_exp1_17240835941970/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link</p>
--------------------	--