Bharatiya Vidya Bhavan's

SARDAR PATEL INSTITUTE OF TECHNOLOGY

Advanced Data Visualization

Experiment no. 1

Submitted To

Prof. Pranav Nerurkar

Submitted By

Name: Aakriti Pathak

UID: 2021300094

Batch: BE Comps (Batch B)

1. Aim:

Create Basics Charts Using Tableau/ Power BI/ R/ Python/ D3.js to be performed on the dataset of E-Commerce Field

- Complete all plots on practice dataset and reproduce on e-commerce dataset.
- Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot.
- Calculate Product wise sales, region wise sales.
- Write observations from each chart.

2. Procedure Description:

Step-1: Dataset:

You can view the dataset from this link.

Step-2: Description:

The sample Dataset includes data for the Sales of multiple products sold by the store along with subsequent information related to geography, Product categories, and subcategories, sales, and profits, segmentation amongst the consumers, etc. This sample Dataset presents a common use case, from which I could gather useful insights from the Sales data in order to improve the Marketing and Sales strategies. I can learn about various operations and elements using this sample Dataset and come up with better strategies to improve and grow the business more.

Step-3: MetaData:

- **Row ID:** Unique ID for each row.
- **Order ID:** Unique Order ID for each customer.
- **Order Date:** Date when the order was placed.
- **Ship Date:** Date when the product was shipped.
- **Ship Mode:** Shipping mode chosen by the customer.
- **Customer ID:** Unique ID to identify each customer.
- **Customer Name:** Name of the customer.
- **Segment:** The segment to which the customer belongs.
- **Country:** Country of residence of the customer.

• **City:** City of residence of the customer.
• **State:** State of residence of the customer.
• **Postal Code:** Postal code of the customer's address.
• **Region:** Region where the customer belongs.
• **Product ID:** Unique ID of the product.
• **Category:** Category of the product ordered.
• **Sub-Category:** Subcategory of the product ordered.
• **Product Name:** Name of the product.
• **Sales:** Sales amount of the product.
• **Quantity:** Quantity of the product ordered.
• **Discount:** Discount provided on the product.
• **Profit:** Profit or loss incurred from the sale of the product.

Step-4: Data Modeling - Star Schema:

1. Overview of the Star Schema

The star schema is a widely used data modeling approach in business intelligence for organizing data in a way that supports efficient querying and reporting. In this model, data is divided into fact and dimension tables:

- **Fact Table:** The central table containing quantitative data (metrics) related to business processes, such as sales, profit, quantity, and discount.
- **Dimension Tables:** Surrounding tables that store descriptive attributes related to the facts, such as customer details, product information, and dates.

2. Data Preprocessing

During the data modeling process, it was identified that the dataset contained duplicate records, particularly within the dimension tables. These duplicates could have skewed the analysis and caused issues in establishing relationships between the tables. Therefore, all duplicate records were carefully removed from the dimension tables before proceeding with the analysis. This ensured the integrity and accuracy of the data model.

3. Explanation of Tables

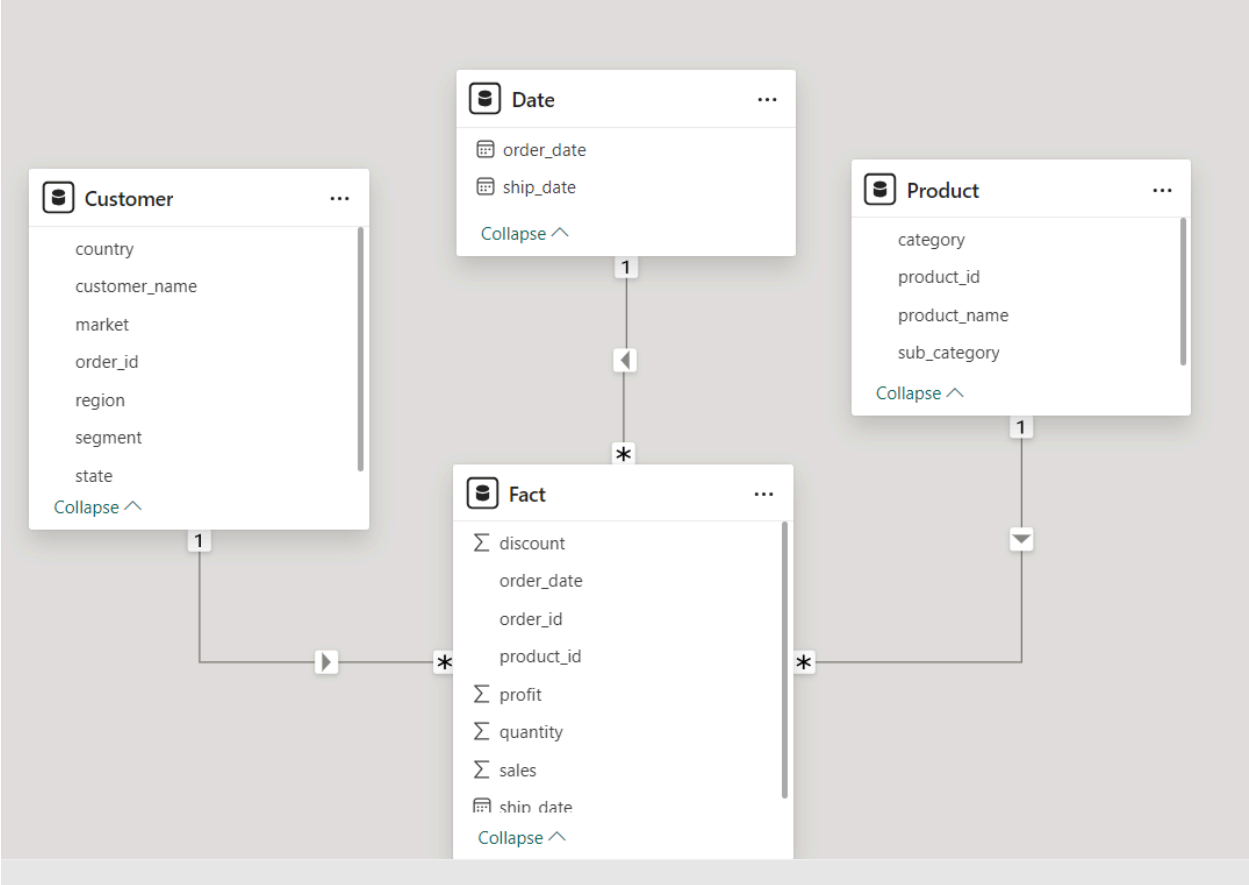
- **Fact Table:**
  - The 'Fact' table includes metrics like sales, profit, quantity, and discount, which are aggregated and analyzed.
- **Dimension Tables:**
  - **Customer Dimension:** Contains details about customers, including their country, name, market segment, region, and state.
  - **Product Dimension:** Holds product-related information, such as product ID, name, category, and sub-category.
  - **Date Dimension:** Manages temporal aspects, linking order and shipping dates to transactions.

4. Relationships

- **Customer Dimension to Fact Table on Customer ID:**
  - **Cardinality:** One-to-Many (One customer can have many orders).
  - **Cross-Filter Direction:** Single.
- **Product Dimension to Fact Table on Product ID:**
  - **Cardinality:** One-to-Many (One product can appear in many orders).
  - **Cross-Filter Direction:** Single.
- **Date Dimension to Fact Table on Order Date:**
  - **Cardinality:** One-to-Many (One date can be linked to many orders).
  - **Cross-Filter Direction:** Single.

5. Star Schema Diagram

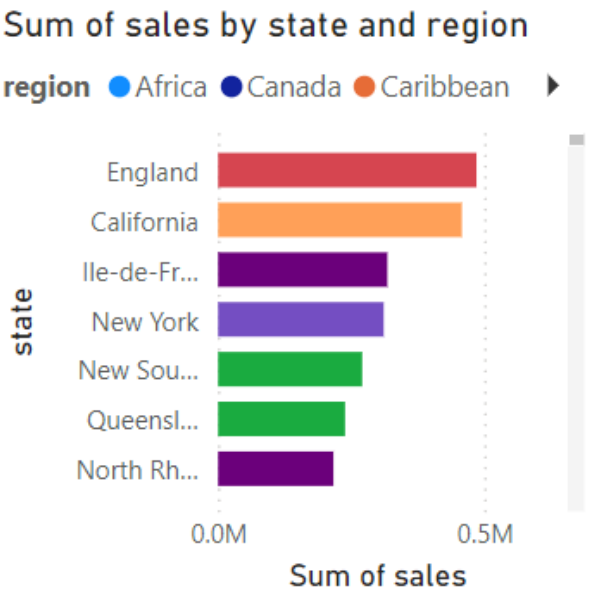
Analysis of Relationships: The relationships between the tables ensure that filtering or aggregating data in any of the dimension tables correctly reflects in the fact table, allowing for accurate and dynamic reporting.



Step-5: Data Visualization Analysis:

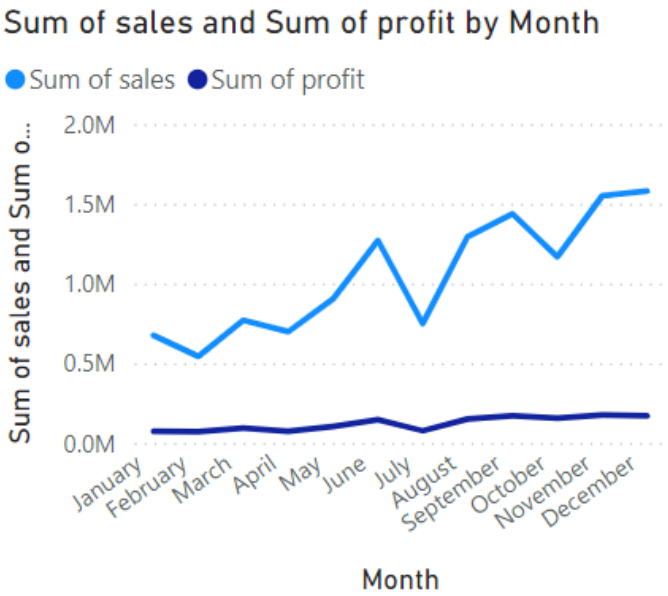
1. Bar Chart: Sum of Sales by State and Region

- **Description:** This bar chart visualizes the total sales across different states and regions.
- **Analysis:**
    - The chart reveals that certain states within specific regions, such as California in the USA, contribute significantly to the overall sales. Understanding regional performance can help target marketing efforts more effectively.



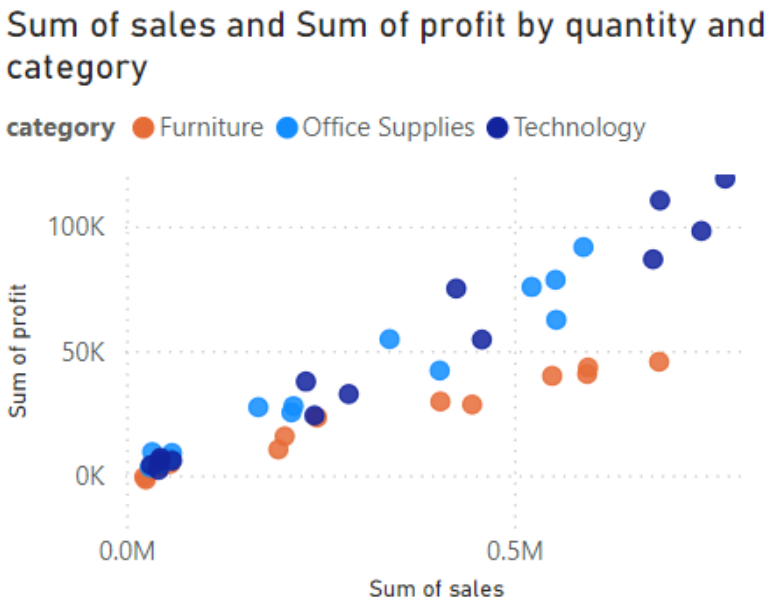
2. Line Chart: Sum of Sales and Profit by Month

- **Description:** This line chart shows the trend of sales and profit over the months.
- **Analysis:**
  - The chart highlights that sales tend to peak towards the end of the year, while profit margins remain relatively stable. Identifying seasonal trends can help in inventory planning and promotional strategies.



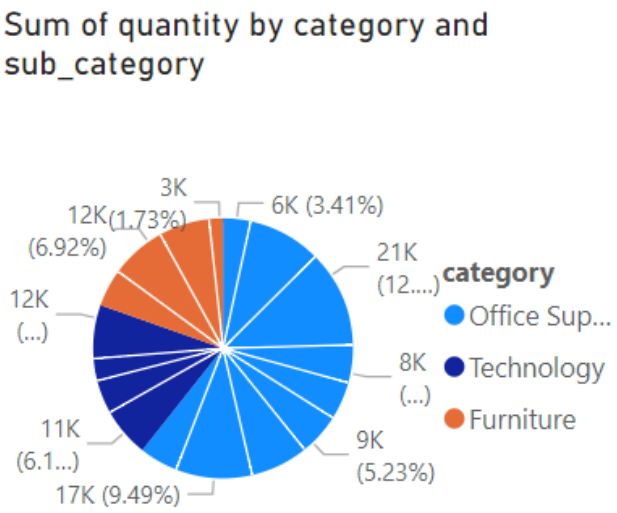
3. Scatter Plot: Sum of Sales and Profit by Quantity and Category

- **Description:** This scatter plot correlates sales and profit with quantity across different product categories.
- **Analysis:**
    - The scatter plot shows that some categories, like Technology, have higher profit margins relative to sales volume. This insight can guide decisions on which product lines to expand or promote.



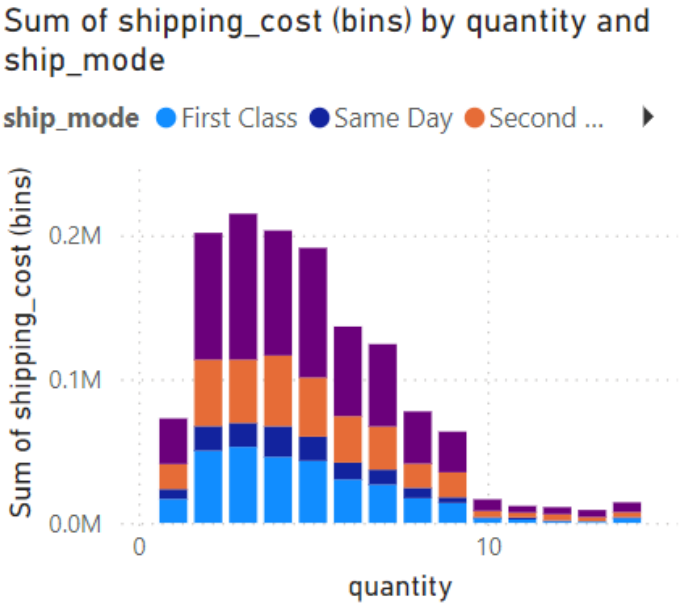
4. Pie Chart: Sum of Quantity by Category and Sub-category

- **Description:** This pie chart depicts the distribution of quantities sold by category and sub-category.
- **Analysis:**
    - Office Supplies appear to dominate in terms of quantity sold, which may indicate a high turnover for lower-cost items. Businesses can use this information to optimize stock levels.



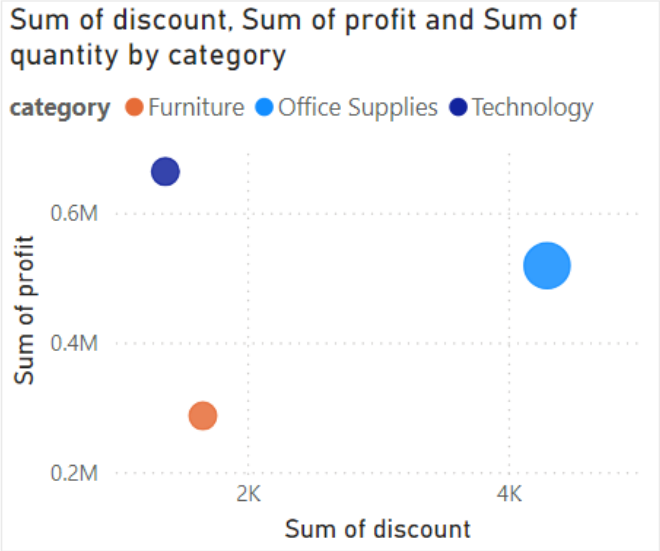
5. Histogram: Sum of Shipping Cost by Quantity and Ship Mode

- **Description:** This histogram displays shipping costs based on the quantity and mode of shipment.
- **Analysis:**
  - Shipping costs vary significantly across different shipment modes, with "Same Day" and "First Class" being more expensive. This can help in optimizing shipping strategies to reduce costs.



6. Bubble Plot: Sum of Discount, Profit, and Quantity by Category

- **Description:** A **Bubble Chart** showing the sum of discount, profit, and quantity across different product categories. The size of the bubbles represents the quantity.
- **Analysis:**
    - This visualization helps to analyze the trade-offs between discounts offered, profits earned, and the quantity sold within each category. Larger bubbles indicate higher quantities sold.



Step-6: Summary of Key Insights:

- **Regional Analysis:** Certain states and regions are key contributors to overall sales.
- **Seasonal Trends:** Sales tend to peak towards the end of the year, which should be factored into inventory and marketing planning.
- **Product Performance:** Technology products yield higher profits, suggesting a focus on these categories could be beneficial.
- **Shipping Costs:** Shipping modes greatly impact costs, highlighting opportunities for cost optimization.
- **Discount Strategy:** High discounts do not necessarily lead to higher profits, particularly in specific categories like Furniture.