Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

NAME: VARADA KHADAKE **ROLL NO :2021300059**

AIM:Create basic charts using Tableau / Power BI / R / Python / D3.js to be performed on the dataset of Ecommerce field

- (Optional) 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 or any other reports
- Write observations from each chart

DATASET:

Amazon Sales Report: **https://www.kaggle.com/datasets/mdsazzatsardar/amazonsalesreport**

Dataset Overview:

The dataset represents an Amazon sales report, providing detailed information about customer orders, product details, shipping, and fulfillment. It appears to be structured to track individual sales transactions, their status, and logistical details. This dataset is valuable for analyzing sales performance, fulfillment efficiency, and customer service quality.

Column Descriptions:

1. Index:

A unique identifier for each row in the dataset, typically used for referencing and sorting.

2. Order ID:

A unique alphanumeric code assigned to each customer order. This identifier helps in tracking and managing individual orders.

3. Date:

The date on which the order was placed or processed. It can be used for analyzing sales trends over time.

4. Status:

The current status of the order, such as "Pending," "Shipped," "Delivered," "Cancelled," etc. This column helps in understanding the stage of each order.

5. Fulfillment:

Information about how the order is being fulfilled, such as by Amazon (FBA) or by the seller. This helps in analyzing the efficiency and effectiveness of different fulfillment methods.

6. Sales Channel:

The platform through which the sale was made, such as Amazon.com, Amazon.in, etc. Useful for identifying the performance of various sales channels.

7. Ship-Service-Level:

The level of shipping service selected for the order, such as Standard, Expedited, or Priority. This can be used to analyze customer preferences and the cost associated with different shipping options.

8. Style:

The style or variation of the product, like color or design. This column helps in tracking the popularity of different styles.

9. SKU (Stock Keeping Unit):

A unique identifier for each product in the seller's inventory. SKUs are essential for inventory management and tracking specific items.

10. Category:

The category or department under which the product is listed, such as Electronics, Clothing, etc. This helps in segmenting sales data by product categories.

11. Size:

The size attribute of the product, if applicable. It is particularly relevant for clothing, footwear, and other size-specific products.

12. ASIN (Amazon Standard Identification Number):

A unique identifier assigned by Amazon to each product. ASINs are critical for product identification and comparison across the Amazon platform.

13. Courier Status:

The status of the courier service responsible for delivering the order, such as "In Transit," "Delivered," etc. This column provides insight into the shipping process.

14. Qty (Quantity):

The number of units ordered for a particular product. This is useful for understanding order sizes and inventory management.

15. Currency:

The currency in which the order was placed, such as USD, INR, etc. It is essential for financial analysis and reporting.

16. Amount:

The total monetary value of the order in the specified currency. This column is key for revenue tracking and financial analysis.

17. Ship-City:

The city where the order is being shipped. It helps in geographic analysis of sales.

18. Ship-State:

The state or region where the order is being shipped. It complements the geographic analysis by providing regional insights.

19. Ship-Postal-Code:

The postal code of the shipping address. This is useful for detailed geographic analysis and logistics planning.

20. Ship-Country:

The country where the order is being shipped. It provides a macro-level view of sales distribution across different countries.

21. Promotion-IDs:

The IDs of any promotions applied to the order. This helps in analyzing the effectiveness of marketing campaigns and discounts.

22. B2B:

A flag indicating whether the order is a Business-to-Business (B2B) transaction. This column is useful for segmenting sales into B2B and B2C (Business-to-Consumer) categories.

23. Fulfilled-By:

Indicates who fulfilled the order, such as "Amazon" or "Seller." This provides further insight into the fulfillment process.

This dataset can be used for a variety of analyses, such as tracking sales performance, understanding customer preferences, optimizing fulfillment processes, and improving logistical efficiency.

Star Schema Structure:

1. **Fact Table:**
    - **Fact_Sales**: This table contains the quantitative data, primarily numerical facts or metrics that are subject to analysis, such as sales amounts and quantities.
2. **Dimension Tables:**
    - **Dim_Date**: Contains details about the date of the transaction.
    - **Dim_Order**: Contains order-specific information.
    - **Dim_Product**: Contains details about the products sold.
    - **Dim_Customer**: Contains customer-related details such as shipping information.
    - **Dim_Fulfillment**: Contains details about how orders are fulfilled and shipped.
    - **Dim_Promotion**: Contains information about promotions and discounts applied.

Tables and Columns:

1. Fact Table: Fact_Sales

- **Sales_Key** (Primary Key)
- **Order_Key** (Foreign Key to Dim_Order)
- **Date_Key** (Foreign Key to Dim_Date)
- **Product_Key** (Foreign Key to Dim_Product)
- **Customer_Key** (Foreign Key to Dim_Customer)
- **Fulfillment_Key** (Foreign Key to Dim_Fulfillment)
- **Promotion_Key** (Foreign Key to Dim_Promotion)
- **Qty**: Number of units ordered.
- **Amount**: Total amount of the sale.
- **Currency**: The currency in which the sale was made.

2. Dimension Table: Dim_Date

- **Date_Key** (Primary Key)
- **Date**: The date of the transaction.

- **Day**: Day of the month.
- **Month**: Month of the year.
- **Quarter**: Quarter of the year.
- **Year**: Year of the transaction.

3. Dimension Table: Dim_Order

- **Order_Key** (Primary Key)
- **Order_ID**: Unique identifier for the order.
- **Status**: The status of the order (e.g., Shipped, Delivered, Cancelled).
- **Sales_Channel**: The platform through which the sale was made (e.g., Amazon.com).
- **B2B**: Indicates if the order is a B2B transaction.
- **Fulfilled_By**: Specifies who fulfilled the order (e.g., Amazon, Seller).

4. Dimension Table: Dim_Product

- **Product_Key** (Primary Key)
- **ASIN**: Amazon Standard Identification Number.
- **SKU**: Stock Keeping Unit.
- **Style**: Product style or variation.
- **Category**: Category of the product.
- **Size**: Size attribute of the product.

5. Dimension Table: Dim_Customer

- **Customer_Key** (Primary Key)
- **Ship_City**: City where the order was shipped.
- **Ship_State**: State where the order was shipped.
- **Ship_Postal_Code**: Postal code of the shipping address.
- **Ship_Country**: Country where the order was shipped.

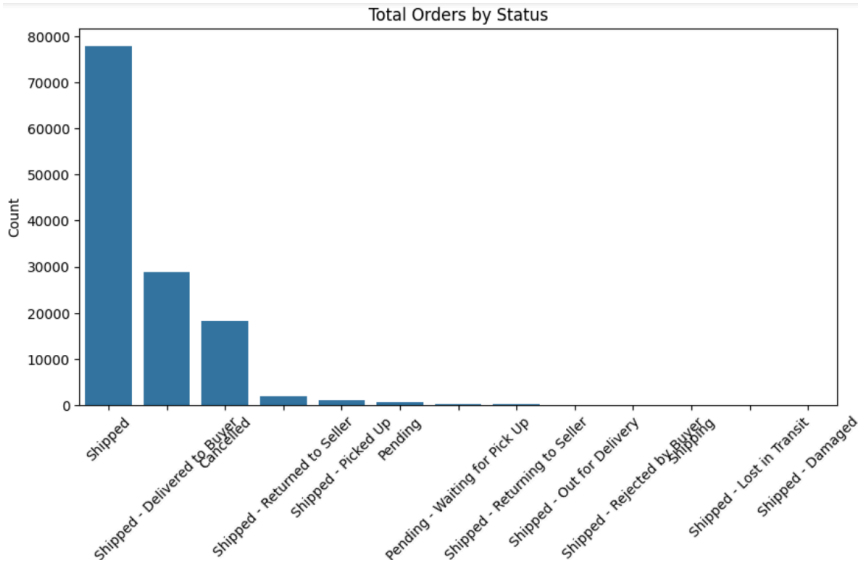
6. Dimension Table: Dim_Fulfillment

- **Fulfillment_Key** (Primary Key)
- **Fulfillment**: Fulfillment method (e.g., FBA, Seller Fulfilled).
- **Ship_Service_Level**: The level of shipping service (e.g., Standard, Expedited).
- **Courier_Status**: The status of the courier service.

7. Dimension Table: Dim_Promotion

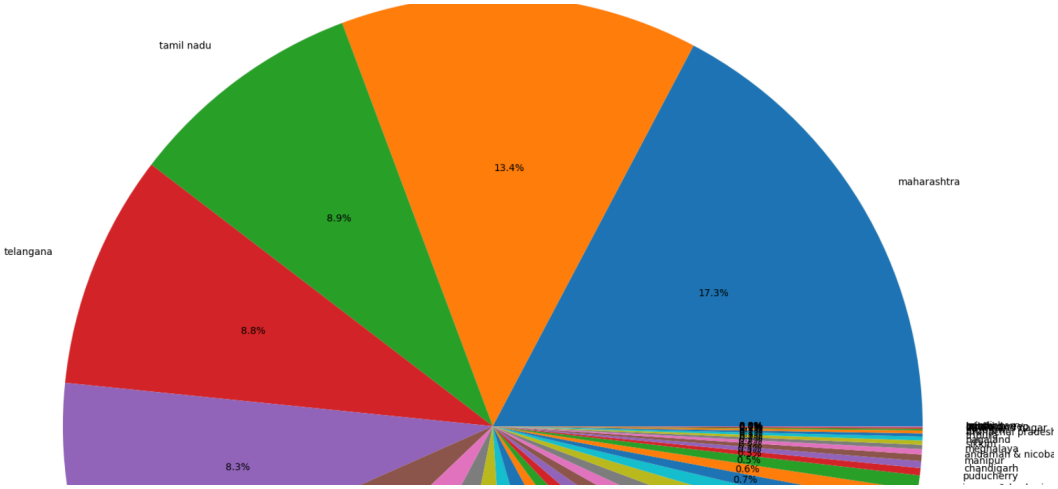
- **Promotion_Key** (Primary Key)
- **Promotion_IDs**: IDs of any promotions applied to the order

Bar Chart



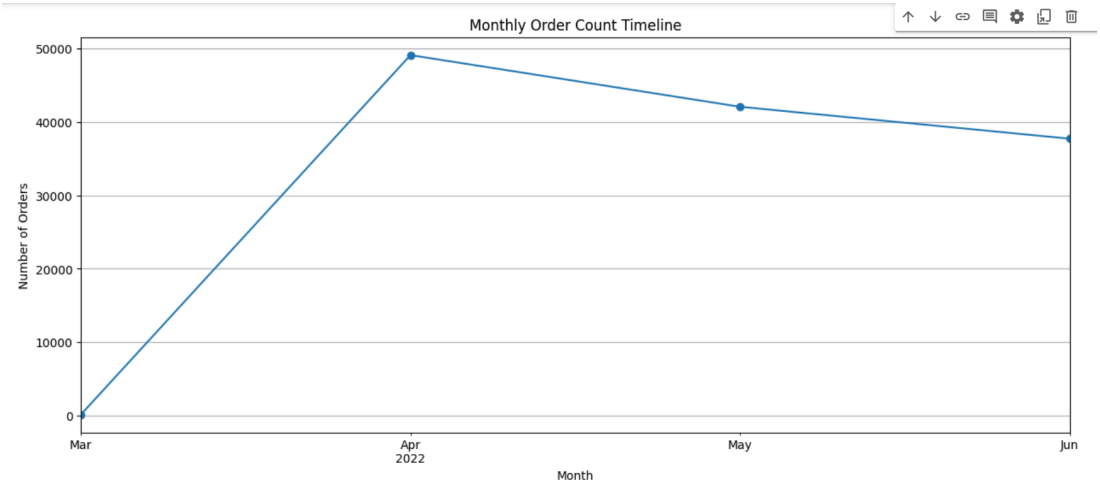
The bar chart titled "Total Orders by Status" shows that the majority of orders are in the "Shipped" status, with a count significantly higher than other statuses, indicating a high fulfillment rate. The next most common statuses are "Delivered to Buyer" and "Cancelled," with "Delivered" being significantly less frequent than "Shipped." Other statuses, like "Returned to Seller," "Picked Up," and "Pending," have much lower counts, suggesting that issues like returns or pending orders are relatively rare in comparison. The distribution highlights that most orders progress smoothly through the fulfillment process, with fewer encountering issues.

Pie Chart



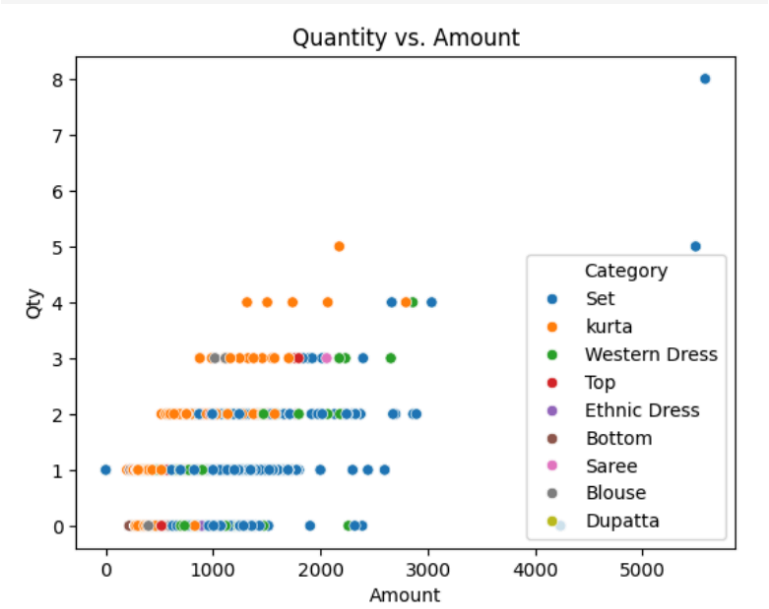
The pie chart for "Total Orders by Ship State" shows that Maharashtra leads with 17.3% of the total orders, followed by Karnataka at 13.4%. Tamil Nadu, Telangana, and Uttar Pradesh contribute 8.9%, 8.8%, and 8.3% respectively. The remaining states make up a much smaller proportion, indicating that a significant majority of orders are concentrated in a few key states, with Maharashtra and Karnataka being the largest contributors to the sales. This distribution suggests that these states are the primary markets for the business, likely due to higher population density or greater consumer demand.

Timeline Chart



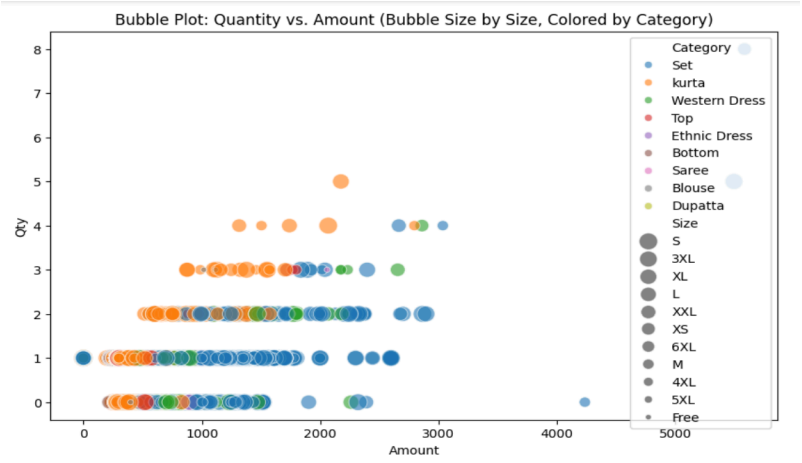
The line graph titled "Monthly Order Count Timeline" illustrates a sharp increase in the number of orders from March to April 2022, where the order count peaks. However, from April to May, there is a noticeable decline in orders, which continues into June, indicating a downward trend after the initial spike. This pattern may suggest a seasonal demand or a particular event in April that led to a significant increase in orders, followed by a gradual decrease.

Scatter Plot



The scatter plot titled "Quantity vs. Amount" displays the relationship between the quantity of items purchased and the amount spent, categorized by different types of clothing. It shows that most transactions involve 1 to 3 items, with a moderate range of spending, particularly for "Sets" and "Kurtas." Higher quantities (4 to 5 items) are less frequent and mostly associated with mid-range amounts. There are a few outliers where a high quantity of items was purchased, leading to higher spending. The plot suggests that customers typically buy small quantities, with "Sets" being the category associated with higher spending, especially when more items are purchased.

Bubble Plot



Analyzing the Bubble Plot: Quantity vs. Amount

Understanding the Visual Elements:

- **X-axis:** Represents the "Amount" of a product.
- **Y-axis:** Represents the "Quantity" of a product.
- **Bubble Size:** Represents the "Size" of the product (S, M, L, XL, etc.).
- **Bubble Color:** Represents the "Category" of the product (Kurta, Western Dress, Top, etc.).

Key Observations:

1. **Quantity vs. Amount Relationship:** There seems to be a general trend where products with higher quantities are also associated with larger amounts. This suggests that bulk purchases or larger orders are more common for certain products.
2. **Category and Quantity:** Some categories, such as "Top" and "Kurta," appear to have a higher overall quantity compared to others. This could indicate that these categories are more popular or in higher demand.
3. **Size and Quantity:** While there are some variations, there doesn't seem to be a strong correlation between the size of a product and its quantity. This suggests that the popularity or demand for different sizes is relatively balanced.
4. **Amount and Size:** There might be a slight trend where larger sizes are associated with higher amounts, but this is not entirely clear. It's possible that larger sizes are more expensive, leading to higher transaction values.