GLOBAL SUPERSTORE ANALYSIS

Milestone 2– Infosys Internship

TOPIC: Global Superstore Analysis using Power BI

Introduction:

The **Global Super Store Dataset** is a popular sample dataset often used in data analysis, visualization, and machine learning projects. It simulates a retail store's operational data and provides a rich context for exploring business metrics, customer behavior, and supply chain performance.

Key Features of the Dataset

1. **Sales Data**: Includes records of transactions made by the store across various categories and regions.
2. **Product Information**: Details about the products sold, including categories and subcategories.
3. **Customer Information**: Data on customers such as their names, locations, and customer segments (e.g., Consumer, Corporate, Home Office).
4. **Order Information**: Data on order IDs, order dates, shipping details, and priorities.
5. **Geographical Data**: Provides information on regions, states, and countries, enabling geographical analysis.
6. **Performance Metrics**: Fields like sales revenue, profit, and discounts allow for analyzing financial performance.
7. **Delivery and Logistics**: Includes data on shipping modes and delivery durations.

Common Columns in the Dataset

- **Order ID**: Unique identifier for each order.
- **Customer Name**: The name of the customer who placed the order.
- **Segment**: The customer segment (e.g., Consumer, Corporate, Home Office).
- **Region**: The geographical area where the transaction occurred.
- **Category/Sub-Category**: Classification of the products sold.
- **Product Name**: The specific product involved in the transaction.
- **Sales**: The revenue generated from the sale.
- **Quantity**: The number of units sold.

- **Discount**: The discount applied to the transaction.

- **Ship Mode**: The method of shipping (e.g., Standard Class, First Class).

Use Cases

- **Data Visualization**: Explore trends using tools like Tableau or Power BI.
- **Statistical Analysis**: Perform descriptive and inferential statistics to understand sales trends and customer behaviors.
- **Machine Learning**: Develop predictive models for forecasting sales or optimizing inventory.
- **Geospatial Analysis**: Map sales performance across different regions.

This dataset serves as an excellent starting point for learning about data processing, business intelligence, and analytics.

Data Preprocessing:

1. **Loading Data:** Import from sources like Excel, SQL Server, or CSV using **Get Data**.

2. **Data Cleaning:**

   - Remove duplicates.

   - Handle missing values (replace or remove).

   - **Trim spaces, standar**dize text, and correct errors.

3. **Data Transformation:**

   - Rename columns/tables and change data types.

   - Split/merge columns and reshape data with pivot/unpivot.

4. **Data Filtering:** Include/exclude rows or columns based on conditions.

5. **Data Enrichment:**

- Add calculated columns.

- Combine datasets using joins or append.

6. **Grouping & Aggregation:** Summarize data using sums, averages, or counts.

7. **Remove Unnecessary Data:** Delete irrelevant rows/columns for efficiency.

DAX Measures and Calculations:

1. **Average Discount**

    AverageDiscount = AVERAGE (Orders [Discount])

It calculates the average Discount from the Orders Table Discount column in the global Super store   Performance   Table..

2. **Average order Amount**

    AverageOrderAmount Per Customer = AVERAGEX(
    VALUES(orders[Customer ID]), CALCULATE(SUM(Orders [Sales])))
 The formula calculates the **average order amount per customer** using DAX in Power BI. It leverages the AVERAGEX function, which iterates over a table to compute an average value. Here, the table being iterated is generated by VALUES(orders[Customer ID]), which returns a list of unique customer IDs from the Orders table. For each customer ID in this list, the formula calculates the total sales for that specific customer using CALCULATE(SUM(Orders[Sales])).

3. **Profit Per Customer:**

    Profit PerCustomer = CALCULATE (SUM(Orders[Profit]),

     ALLEXCEPT (Orders, Orders [Customer ID]))
The formula calculates the **total profit per customer** using DAX in Power BI. It employs the CALCULATE function, which modifies the context of the calculation to ensure that the sum of profits (SUM(Orders[Profit])) is computed specifically for each customer. The ALLEXCEPT function plays a key role here; it clears all filters on the Orders table except for

the filter on Orders[Customer ID]. This means the calculation ignores any other filters (such as date, region, or product) applied in the current context and considers only the customer-specific filter.

4. **Sales Contribution:**

   Salescontribution=DIVIDE(

   SUM(Orders[Sales]),CALCULATE(SUM(Orders[Sales]),ALL(Orders))

The formula calculates the **sales contribution** of a specific context (such as a customer, region, or product) as a percentage of total sales using DAX in Power BI. It uses the DIVIDE function to perform the division safely, handling cases where the denominator might be zero. By dividing the sales in the current context by the overall total sales, this formula outputs a ratio that represents the proportion of the total sales contributed by the current context. This metric is particularly useful in understanding the relative performance

5. **Total Order Amount For Returns:**

   TotalOrderAmountForReturns = CALCULATE(

   SUM(orders[sales]

   FILTER (Orders, Orders [Order ID] IN VALUES (Returns [Order ID]))
)

The formula calculates the **total order amount for returned items** using DAX in Power BI. It uses CALCULATE to adjust the calculation context and sum up the sales (SUM(Orders[Sales])) specifically for orders that are marked as returned. The key part of the formula is the FILTER function, which filters the Orders table to include only those rows where the Order ID matches an Order ID found in the Returns table. The VALUES(Returns[Order ID]) generates a list of unique order IDs from the Returns table, and the IN operator checks if each Order ID in the Orders table is present in that list.

6. **Total Revenue:**

   TotalRevenue IncludingShipping = SUMX(Orders,Orders [Sales]+ Orders [Shipping

   Cost])

The formula calculates the **total revenue, including shipping costs**, using DAX in Power BI. It employs the SUMX function, which iterates through each row in the Orders table and evaluates an expression for every row. The SUMX function then sums up these computed values across all rows in the Orders table to produce the total revenue, which combines both the sales revenue and the shipping charges collected.

# Understanding Relationships in Power BI:
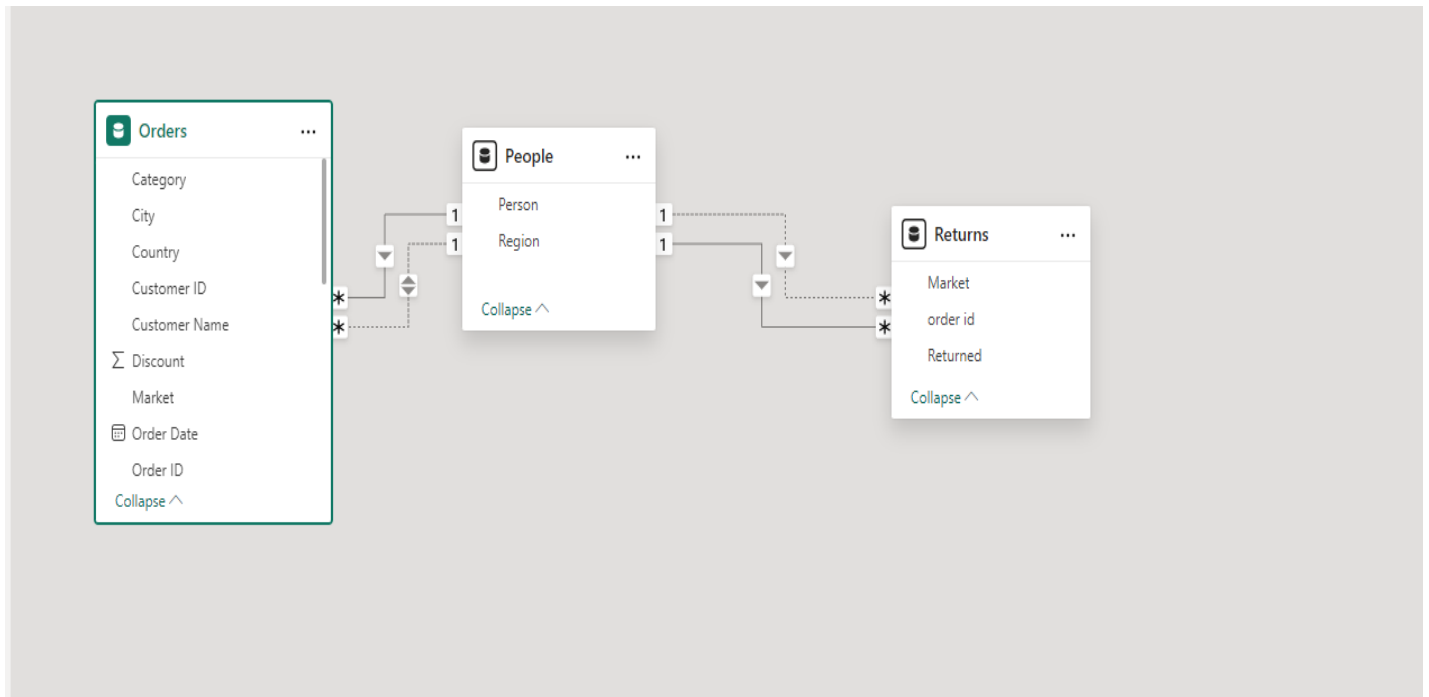
Many-to-One Relationship:
The orders table is having Many to one relation with people table person column each category is related with person in the people table. And the people table is also having many to one relationship returns Table .every Region in people Table is connected to Returns Table Market column. And the customer name in the orders table is having Many to one Relationship with People's table Region. Every region is related to the customer name. And also the Returns Table is having the many to one Relationship with Peoples Table. Every person is related to Returned status column in Returns Table. These are the relations the I have created in the model view according to the Global superstore dataset columns.

 In PowerBI we have many types of Relationships like one to one, many to one, one to many, many to many. Among all the most common type is many to one relationship which is easily understandable by the users for finding the realtion among the different data tables.

Report Views:

Report Image-1:
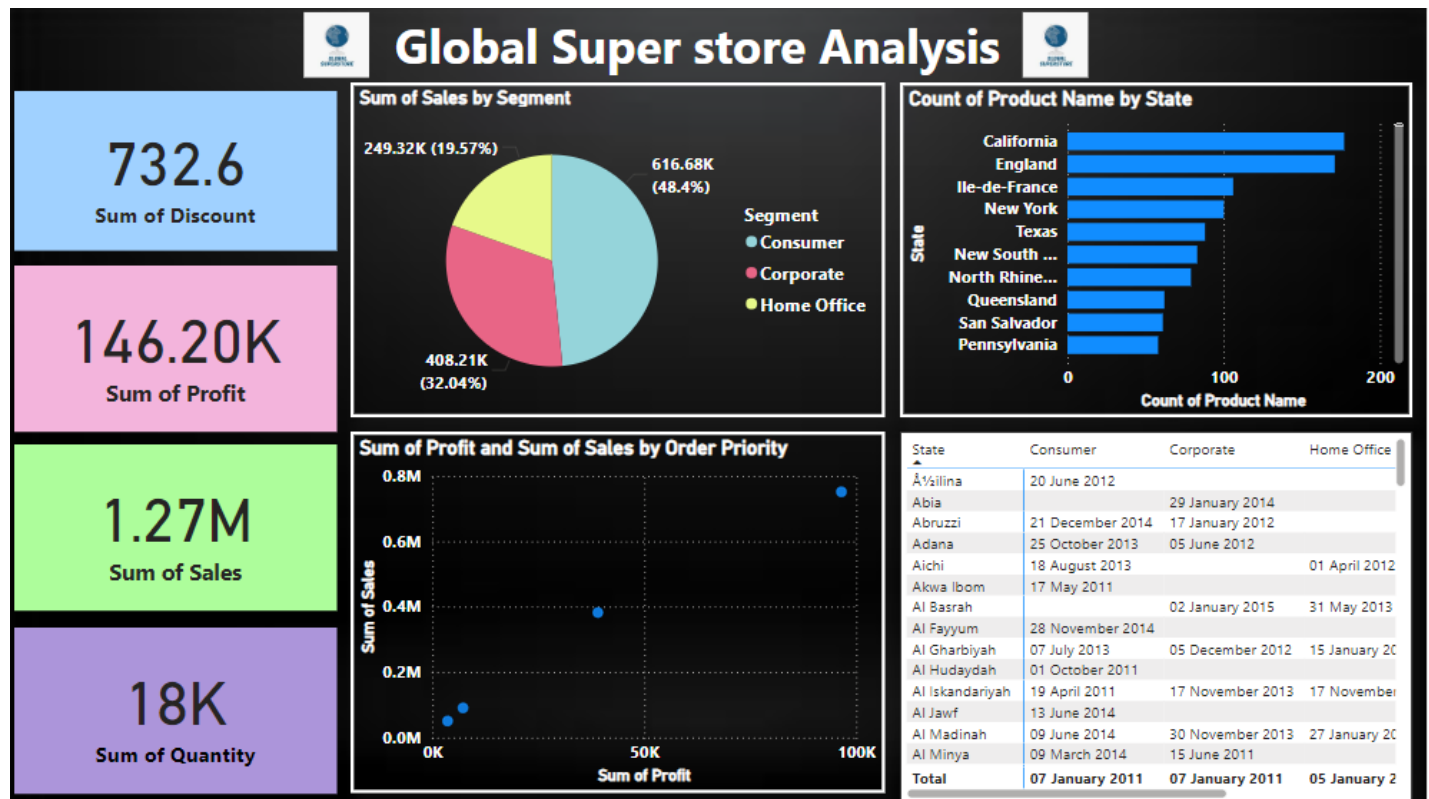
This dashboard provides insights into a global superstore data, breaking down various metrics like sum of Discount, sum of profit ,sum of sales , sum of Quantity.   The " Sum of Discount shows " shows a total of 732.6, The total profit is 146.20 thousand ,next comes to the sum of sales is 1.27 million ,and the sum of Quantity is 18 thousands . The Pie chart represents the Sum of sales by Segment contains three legends like Consumer, Corporate, Home Office. The clustered bar chart represents the count of product by State. The Scatter chart represents the Sum of profit and Sum of Sales By order priority. And a table the represents the information like state, Consumer, Corporate, Home office.

Report Image - 2:
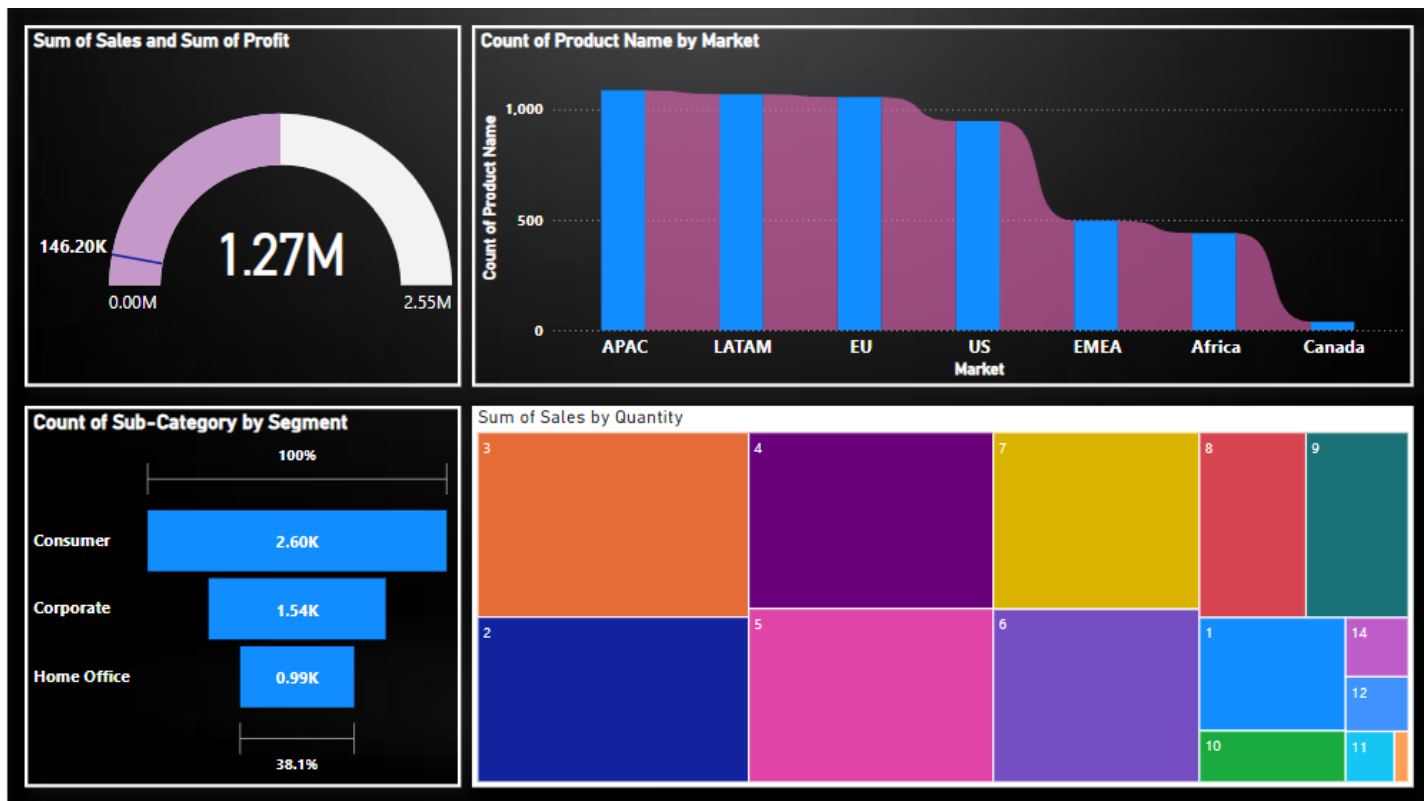
This dashboard provides insights into sales distribution based on Profit .The Gauge meter shows the sum of Sales and sum of Profit. And the second Visualization is waterfall chart the Represents the Count of product Name By Market. And the next Visualization Represents the Count of Sub category by segment and segments are Consumer ,Corporate, Home office. Sum of sales by quantity is Represented by Tree map.

Conclusion:

Analyzing the Global Super Store dataset provides valuable insights into the performance and operations of a fictional retail business. The dataset encompasses various aspects of business operations, such as sales, customer demographics, product categories, and shipping methods, offering a comprehensive view of the store's activities.