**DEPI Microsoft Power BI**

**Graduation project**

**Superstore**

**Instructor: *Eslam El Saghir***

**Team Members**

1. ***Mariam Ghoneim***
2. ***Hager Mahmoud***
3. ***Yasmine Eid***

**Introduction**

This is an Exploratory Data Analysis (EDA) and visualization project about an e-commerce platform known as a superstore

Power BI Skills/Concepts applied:

* Data Cleaning/Validation in Power Query
* Data Visualization
* DAX Concepts: Calculated Measures, Calculated Columns.
* Filters and slicers

**Data Understanding:**

SuperStore dataset describes the sales data of the superstore Ecommerce website, it contains an amount of data collected for the Years (2015-2016-2017-2018). the raw data contain one table (Superstore Sales) which has 18 columns it includes information about the order:

(order ID /order date/ ship date/ ship mode/ customer ID/ customer name/segment/ country /city state/ postal code/ region/ product category/ product subcategory/ product name/ product ID and sales)

**Business Questions & KPI**

The following business questions have been presented to enable the superstore team to gain better insights into their e-commerce platform and optimize available growth opportunities. Our analysis is targeting the sales department.

**Analysis goals:**

Increase sales and revenue

Enhance product and category performance

increase the Expanding sales in different region.

**KPI:**

* sales
* RPC repeat purchase customers
* AOV average order value

**Questions:**

* Trends in sales over time?
* How do sales differ across each segment?
* Does the shipping mode vary on Sales?
* Which product subcategories contribute the most to overall sales performance?
* Average order value
* Top 5 customers based on actual purchase volume?
* What is the top-selling products?
* What percentage of customers are repeat buyers?
* Which state has the highest and lowest profit margins?

Data Transformation

1. Use the first row as the header
2. change the type column sales from text to a ‘decimal number’
3. split order date (DD, MM, YY) and reorder order date (MM, DD, YY) to fit the system and change type column order date to ‘date’
4. split column ship date (DD, MM, YY) and reorder column ship date (MM, DD, YY) to fit the system and change type column ship date to ‘date’
5. there are 11 blank values in the postal code column replace them conditionally and remove the old postal code by using M language code

// Ensure the Postal Code is text

    ChangedType = Table.TransformColumnTypes(ConvertShipDateToDate, {{"Postal Code", type text}}),

    // Replace the blank value in 'Postal Code' with "05401" if 'State' is "Vermont" and 'City' is "Burlington"

    ReplaceBlankConditionally = Table.AddColumn(ChangedType, "New Postal Code", each

        if (Text.Trim(Record.Field(, "Postal Code")) = "" or Record.Field(, "Postal Code") = null) and

           (Record.Field(\_, "State") = "Vermont") and

           (Record.Field(\_, "City") = "Burlington")

        then "05401" else Record.Field(\_, "Postal Code")

    ),

    // Remove the original Postal Code column and rename the new one

    FinalTable = Table.RemoveColumns(ReplaceBlankConditionally, {"Postal Code"}),

    RenamedFinalTable = Table.RenameColumns(FinalTable, {{"New Postal Code", "Postal Code"}}),

    #"Reordered Columns" = Table.ReorderColumns(RenamedFinalTable,{"Row ID", "Order ID", "Order **Date", "Ship Date", "Ship Mode", "Customer ID", "Customer Name", "Segment", "Country", "City", "Postal Code", "State", "Region", "Product ID", "Category", "Sub-Category", "Product Name", "Sales"})**

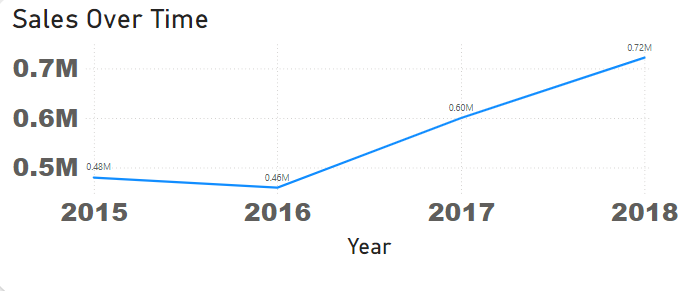
**in**

**#"Reordered Columns"**

**Data Exploration and Visualization in Power BI**

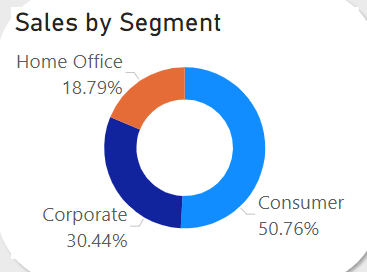
1. Which states are in the top ten for highest sales performance?
2. Trends in sales over time

This figure shows the total sales over time. This was achieved by getting the relation between the sum of sales and the order date

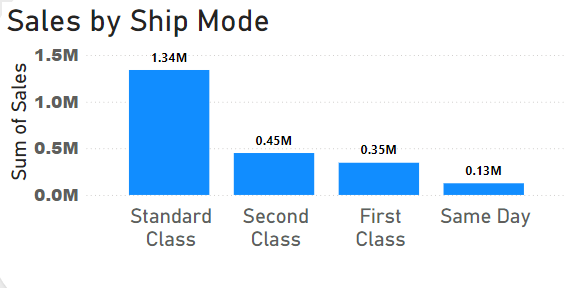
****

1. How do sales differ across each segment?

there are three segments in the superstore (Home office, consumer, and corporate) consumer segment achieved the highest sales with a percentage of 50.76% this figure achieved by the relation between segment and sum of sales

****

1. Does the shipping mode vary depending on the region?

****

1. Does the shipping mode vary depending on the product?
2. What is the difference between the order date and the ship date based on the shipping mode or product availability?
3. Which product subcategories contribute the most to overall sales performance?
4. Average order value

the average order value in the superstore is $459.48. This was achieved by dividing the Total sales  by the total number of orders using the DAX calculated measure below:

Avg Order Value = SUM('Superstore Sales Dataset'[Sales]) / DISTINCTCOUNT('Superstore Sales Dataset'[Order ID])



1. repeat customer

Repeat Customers =

CALCULATE(

    COUNTROWS(

        FILTER(

            VALUES('Superstore Sales Dataset'[Customer ID]),

            CALCULATE(

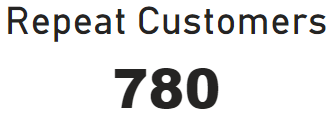
                DISTINCTCOUNT('Superstore Sales Dataset'[Order ID])

            ) > 1

        )

    )

)

****

1. Who are the 10  top customers based on actual purchase volume or revenue?

the top 10 customers achieved  by this DAX measure

TopNCustomers =

CALCULATETABLE(

    TOPN(

        'TopN Customer'[TopN Customer Value],  -- Parameter for Top N selection

        ADDCOLUMNS(

            SUMMARIZE(

                'Superstore Sales Dataset',

                'Superstore Sales Dataset'[Customer ID]  -- Group by Customer ID

            ),

            "CustomerSales", SUMX(

                FILTER('Superstore Sales Dataset', 'Superstore Sales Dataset'[Customer ID] = EARLIER('Superstore Sales Dataset'[Customer ID])),

                'Superstore Sales Dataset'[Sales]  -- Calculate total sales per customer

            )

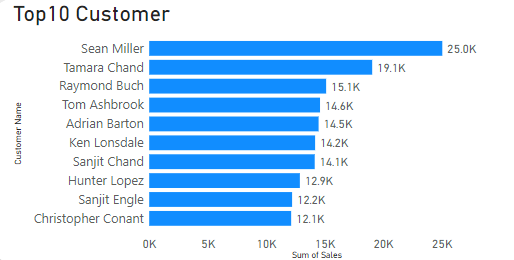
        ),

        [CustomerSales],  -- Sort by the calculated CustomerSales

        DESC

    )

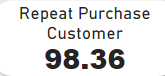
)

****

1. What are the top-selling products during peak seasons or holidays?
2. What percentage of customers are repeat buyers?

percentage repeat customer achieved by this DAX

persentage Repeat customer = [Repeat Customers]/[distinct customer]\*100



**Recommendations:**

* Although the data showed that the sales from the three product categories were close to each other, the technology category was the dominant one. Also, the top products in sales were from the technology category. Therefore, we recommend investing in this category.
* Focus on consumers especially in the states California, New York and Texas.
* Most of our customers order from us again, we can focus on making new customers in the suggested states California, New York and Texas by making offers to our customers that push them to recommend our platform to others.