



## **FACULTY OF INFORMATION TECHNOLOGY**

**SUBJECT NAME: FUNDAMENTAL OF MODERN DATA**

**SUBJECT CODE: BIT2053**

**LECTURER NAME: NAZMIRUL IZZAD BIN NASSIR**

**PREPARED BY:**

<b>No.</b>	<b>NAME</b>	<b>ID</b>	<b>PROGRAMME</b>
<b>1.</b>	<b>AUDRIE ANNE JEFFRY</b>	<b>202309010419</b>	<b>BIT</b>
<b>2.</b>	<b>EDDY BIN LINUS</b>	<b>202309010409</b>	<b>BIT</b>
<b>3.</b>	<b>GRACE MICHELLE LYNN MAJILIN</b>	<b>202309010414</b>	<b>BIT</b>
<b>4.</b>	<b>STANLEY CLARENCE STEFANUS</b>	<b>202309010396</b>	<b>BIT</b>
<b>5</b>	<b>LYON CARLOS ESTEBAN</b>	<b>202309010424</b>	<b>BIT</b>

**FOR OFFICE USE ONLY**

**RECEIVED BY :**

**DATE :**

## **TABLE OF CONTENT**

INTRODUCTION.....	2
BUSINESS SCENARIO.....	2
DATASET.....	3-6
METHODOLOGY.....	7-8
DATA ANALYSIS AND VISUALISATION.....	9
RESULT.....	10
RECOMMENDATION.....	11
CONCLUSION.....	12

## 1.0 INTRODUCTION

In this age of business competitiveness, companies are more reliant on data-driven decision making as a way of ensuring profitability and sustainability. Retail firms in particular have a lot of transactional data which can be used to discover patterns in customer behavior and product and regional market trends. BI tools have the ability to refrain raw data into workable knowledge by way of dashboards, visualizations and statistical modeling.

This project will use BI principles on a real-world dataset, Global Superstore, a data set that simulates retail data in terms of the various types of regions, product categories as well as customer segments with detailed sales transactions. The main goal is to replicate the real-life business situation and explore the trends in sales and profitability, outline key performance drivers, and deliver recommendations that can be utilized by the business to improve business performance.

The data analysis will start by undertaking preprocessing to eliminate duplicates, take care of missing values and data consistency. The cleaned spreadsheet is then modeled as a star schema with Power BI giving access to filter, slice it and use KPIs. Three important business questions were set to drive the analysis and these are on sales performance by region, product profitability and customer buying behavior. The outcomes of this have been provided in the form of a dashboard consisting of descriptive statistics with a visual analytics tool as written below followed by strategic actions to be taken.

## 2.0 BUSINESS SCENARIO

Global Superstore is a multinational retail company that operates in various regions around the world. It sells a wide range of products, including office supplies, furniture, and technology. Like many retail organizations, the company faces challenges in understanding which regions are most profitable, which products drive sales growth, and how different customer segments contribute to revenue.

The management team wants to leverage Business Intelligence (BI) to support decision-making. By analyzing historical sales transactions, they hope to answer critical business questions such as:

- Which regions and markets generate the highest revenue and profit margins?
  - ❖ Helps allocate resources and marketing budgets more effectively.
- Which product categories or sub-categories contribute the most to profitability?
  - ❖ Supports inventory planning and pricing strategies.
- How do different customer segments behave in terms of purchase volume and profitability?
  - ❖ Provides insights for customer relationship management and targeted promotions.

## 3.0 DATASET

### 3.1 DATASET DESCRIPTION

The dataset used in this study is the *Global Superstore* dataset which is provided from Kaggle, a widely adopted benchmark in business intelligence and analytics. It represents transactional data from a global retail company that sells consumer goods across different countries and regions. The dataset contains over 51,000 sales records with 24 attributes, making it suitable for analyzing business performance across dimensions such as customer behavior, product categories, and regional markets.

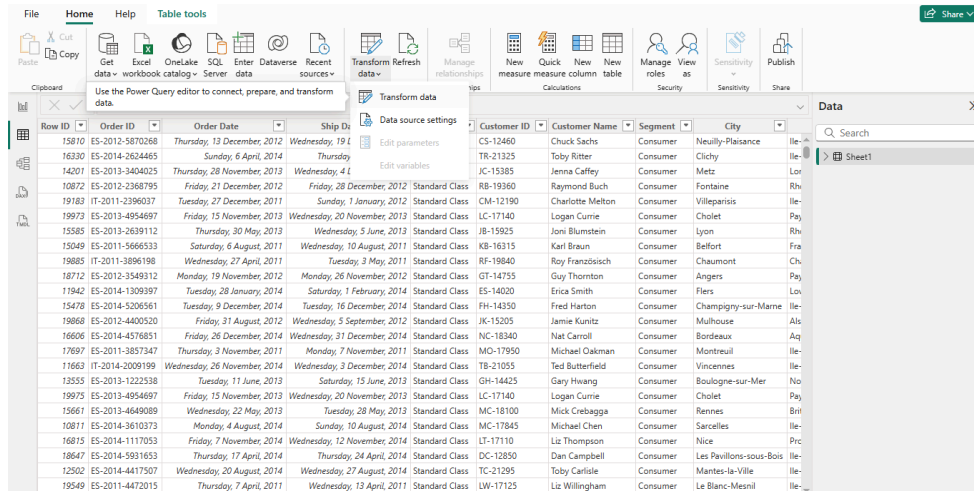
The dataset captures four main types of information. First, sales transaction data includes order date, ship date, sales revenue, profit, discount, and quantity sold. Second, customer information provides customer ID, name, segment (Consumer, Corporate, or Home Office), and geographic attributes such as country, city, region, and postal code. Third, product information records product ID, product name, category (Furniture, Office Supplies, Technology), and sub-category. Finally, order and shipping details include order ID, ship mode, and shipping cost.

During preprocessing, duplicate records were detected and removed to ensure consistency and accuracy of the analysis. In addition, missing values were found in the *Postal Code* column, and date fields such as *Order Date* and *Ship Date* were converted into proper date formats to support time-series analysis. After cleaning, the dataset provides a reliable foundation for descriptive and diagnostic business intelligence, enabling insights into profitability, customer behavior, and product performance.

## 3.2 PREPROCESSING

### 3.2.1 Remove Duplicate

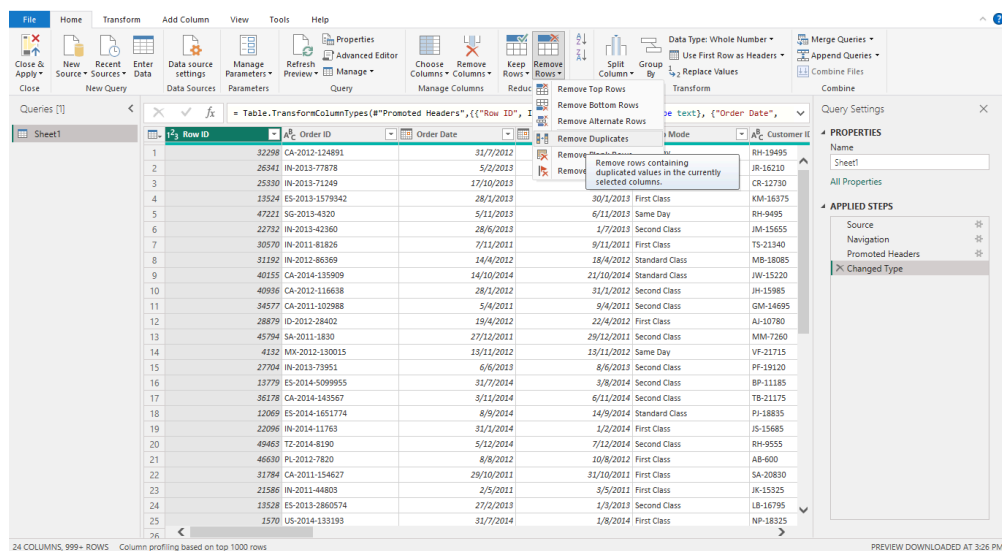
To remove duplicate, we need to open the Power Query Editor by selecting “Transform Data” in Home page



The screenshot shows the Power Query Editor interface. The 'Home' tab is active in the ribbon, and the 'Transform data' button is highlighted. Below the ribbon, a data table is visible with columns: Row ID, Order ID, Order Date, Ship Date, Customer ID, Customer Name, Segment, and City. The table contains 24 rows of data.

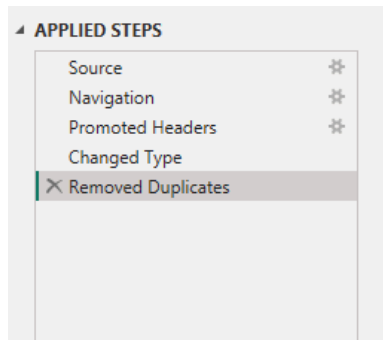
Row ID	Order ID	Order Date	Ship Date	Customer ID	Customer Name	Segment	City
15810	ES-2013-5870368	Thursday, 13 December, 2012	Wednesday, 19 December, 2012	CS-12460	Chuck Sachs	Consumer	Neuilly-Plaisance
16320	ES-2014-2624465	Sunday, 6 April, 2014	Thursday, 19 April, 2014	TR-21325	Toby Ritter	Consumer	Clichy
14201	ES-2013-3404025	Thursday, 28 November, 2013	Wednesday, 4 December, 2013	JC-15385	Jenna Caffey	Consumer	Metz
10872	ES-2012-2368795	Friday, 21 December, 2012	Friday, 28 December, 2012	RB-19360	Raymond Buch	Consumer	Fontaine
19183	IT-2011-2396037	Tuesday, 27 December, 2011	Sunday, 1 January, 2012	CM-12190	Charlotte Melton	Consumer	Villeparisis
19973	ES-2013-4954697	Friday, 15 November, 2013	Wednesday, 20 November, 2013	LC-17140	Logan Currie	Consumer	Cholet
15585	ES-2013-2639112	Tuesday, 30 May, 2013	Wednesday, 5 June, 2013	JB-15925	Joni Blumstein	Consumer	Lyon
15049	ES-2011-5666533	Saturday, 6 August, 2011	Wednesday, 10 August, 2011	KB-16315	Karl Braun	Consumer	Belfort
19883	IT-2011-3896198	Wednesday, 27 April, 2011	Tuesday, 3 May, 2011	RF-19840	Roy Franzosich	Consumer	Chaumont
18712	ES-2012-3249312	Monday, 19 November, 2012	Monday, 26 November, 2012	GT-14755	Guy Thornton	Consumer	Angers
11942	ES-2014-1309397	Tuesday, 28 January, 2014	Saturday, 1 February, 2014	ES-14020	Erica Smith	Consumer	Fiers
15478	ES-2014-5206561	Tuesday, 9 December, 2014	Tuesday, 16 December, 2014	FH-14350	Fred Harton	Consumer	Champigny-sur-Marne
19866	ES-2012-4400520	Friday, 31 August, 2012	Wednesday, 5 September, 2012	JK-15205	Jamie Kunitz	Consumer	Mulhouse
16606	ES-2014-4576851	Friday, 26 December, 2014	Wednesday, 31 December, 2014	NC-18340	Nat Carroll	Consumer	Bordeaux
17697	ES-2011-3857347	Thursday, 3 November, 2011	Monday, 7 November, 2011	MC-17990	Michael Oakman	Consumer	Montreuil
11663	IT-2014-2009199	Wednesday, 26 November, 2014	Wednesday, 3 December, 2014	TB-21055	Teal Butterfield	Consumer	Vincennes
12555	ES-2013-1222536	Sunday, 11 June, 2013	Saturday, 15 June, 2013	GH-14425	Gary Hwang	Consumer	Boulogne-sur-Mer
19975	ES-2013-4954697	Friday, 15 November, 2013	Wednesday, 20 November, 2013	LC-17140	Logan Currie	Consumer	Cholet
15661	ES-2013-4640989	Wednesday, 22 May, 2013	Tuesday, 28 May, 2013	MC-18100	Mick Crebagga	Consumer	Rennes
10811	ES-2014-3610373	Monday, 4 August, 2014	Sunday, 10 August, 2014	MC-17845	Michael Chen	Consumer	Sarcelles
16815	ES-2014-1117053	Friday, 7 November, 2014	Wednesday, 12 November, 2014	LT-17110	Liz Thompson	Consumer	Nice
18647	ES-2014-5931653	Thursday, 17 April, 2014	Thursday, 24 April, 2014	DC-12850	Dan Campbell	Consumer	Les Pavillons-sous-Bois
12502	ES-2014-4417507	Wednesday, 20 August, 2014	Wednesday, 27 August, 2014	TC-21295	Toby Carlisle	Consumer	Mantes-la-Ville
19549	ES-2011-4472015	Thursday, 7 April, 2011	Wednesday, 13 April, 2011	LW-17125	Liz Willingham	Consumer	Le Blanc-Mesnil

After opening the Power Query Editor, in the Home page, select “Remove Row” and then select “Remove Duplicate” and it will remove all duplicate data automatically.



The screenshot shows the Power Query Editor interface with the 'Remove Duplicates' dialog box open. The dialog box has options to 'Remove Top Rows', 'Remove Bottom Rows', 'Remove Alternate Rows', and 'Remove Duplicates'. The 'Remove Duplicates' option is selected. The 'Columns to keep' list includes 'Row ID', 'Order ID', 'Order Date', and 'Customer ID'. The 'Columns to remove' list is empty. The 'Remove' button is highlighted. Below the dialog box, a data table is visible with columns: Row ID, Order ID, Order Date, and Customer ID. The table contains 24 rows of data.

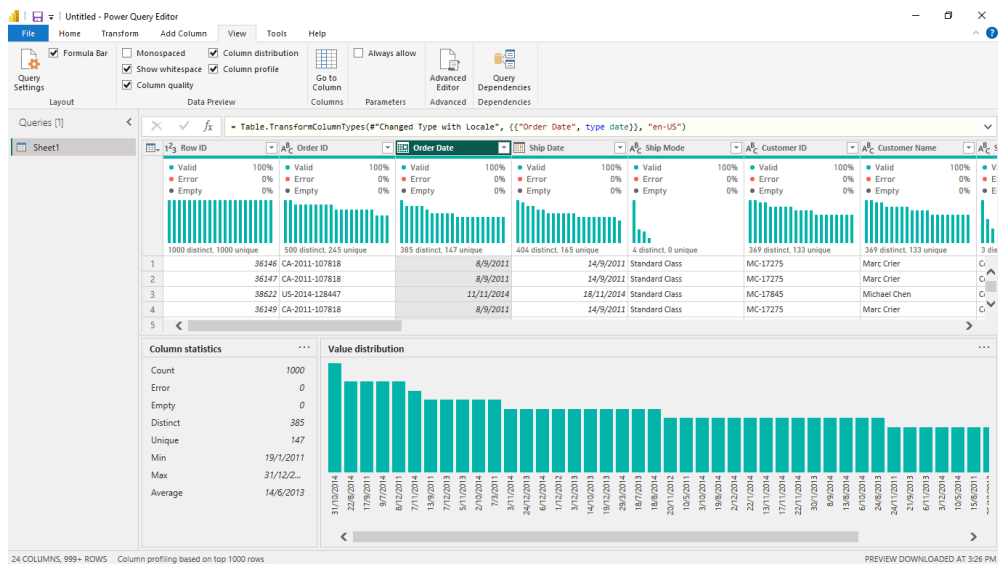
Row ID	Order ID	Order Date	Customer ID
1	32298	CA-2012-124891	31/7/2012
2	26341	IN-2013-77878	5/2/2013
3	25330	IN-2013-71249	17/10/2013
4	13524	ES-2013-1579342	28/1/2013
5	47221	SG-2013-4320	5/11/2013
6	22732	IN-2013-42360	28/6/2013
7	30570	IN-2011-81826	7/11/2011
8	31192	IN-2012-86369	14/4/2012
9	40155	CA-2014-135909	14/10/2014
10	40936	CA-2012-116608	28/1/2012
11	34577	CA-2011-102988	5/4/2011
12	28879	IO-2012-28402	19/4/2012
13	45794	SA-2011-1830	27/12/2011
14	4132	MX-2012-130015	13/11/2012
15	27704	IN-2013-73951	6/6/2013
16	13779	ES-2014-5099955	31/7/2014
17	36178	CA-2014-143567	3/11/2014
18	12069	ES-2014-1651774	8/9/2014
19	22096	IN-2014-11763	31/1/2014
20	49463	TZ-2014-8190	5/12/2014
21	46630	PL-2012-7820	8/8/2012
22	31784	CA-2011-154627	29/10/2011
23	21586	IN-2011-44803	2/5/2011
24	13528	ES-2013-2860574	27/2/2013
25	1570	US-2014-133193	31/7/2014



To ensure the system removes all duplicate data, we can check at “Applied Step” which is a record of action or step that we perform in the Power Query Editor.

### 3.2.2 Data Validation

To validate data, in Power Query Editor, select “Column Quality” and “Column Profile” and it will show the “Valid”, “Error” and “Empty” or “Null (no value)” data

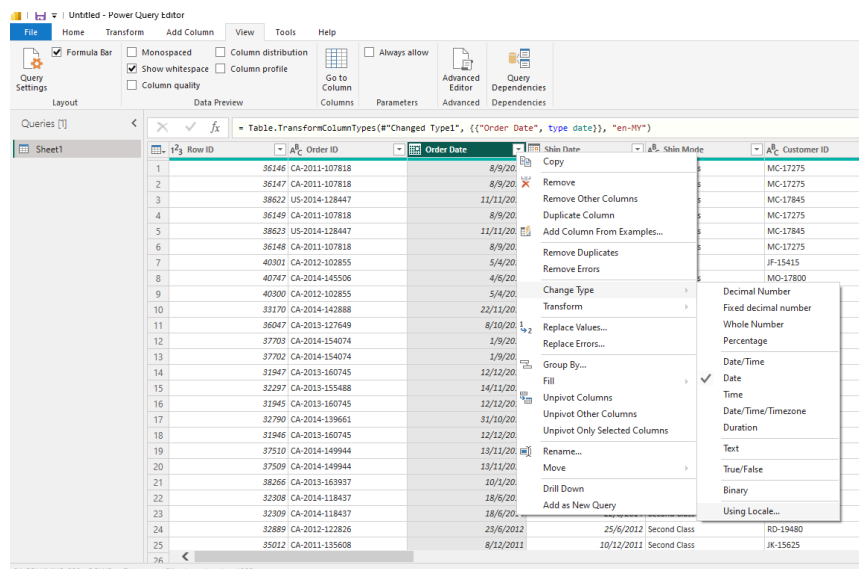


### 3.2.3 Change Format

In our project, we will choose “Order Date” to change the format from “D/M/YYYY” to “M/D/YYYY” for the date

	Row ID	Order ID	Order Date
1	36146	CA-2011-107818	8/9/2011
2	36147	CA-2011-107818	8/9/2011
3	38622	US-2014-128447	11/11/2014
4	36149	CA-2011-107818	8/9/2011
5	38623	US-2014-128447	11/11/2014
6	36148	CA-2011-107818	8/9/2011
7	40301	CA-2012-102855	5/4/2012
8	40747	CA-2014-145506	4/6/2014
9	40300	CA-2012-102855	5/4/2012

In Power Editor Query, Right-Click at “Order Date” row, go to “Change Type” and select “Using Locale....”



After open “using Locale...”, choose “date” as the data type, and the the locale from “D/M/YYYY (English {Malaysia})” format change to “M/D/YYYY” (English {United State}) format.

#### Change Type with Locale

Change the data type and select the locale of origin.

Data Type  
Date

Locale  
English (United States)

Sample input values:

3/29/2016  
Tuesday, March 29, 2016  
March 29  
March 2016

OK

Cancel

## **4.0 METHODOLOGY**

### **4.1 Data Preparation**

The project can be characterized by the sequential approach to the methodology, where the Global Superstore dataset provided by Kaggle was selected. The given dataset was selected because it covers the broad scope of sales transactions in various regions, product segments, and customer groups and is quite appropriate to use as the reflection of real online retail business. The business case was determined based on a multinational retailing company that was interested in maximizing their sales, profitability, and the level of customer engagement through implementation of Business Intelligence tools.

The data preprocessing operations were performed in Power BI, where duplicate entries were filtered out, data type standardization was provided, and data integrity of primary fields (dates, numerical measures) was ensured. A star schema was then created so that efficient analysis will be supported, with a central fact table (Sales) which connected with dimension tables including Products, Customers, Regions, and Date. Scalability and flexibility of reporting was made possible with this structure.

Subsequently, a DAX analysis was undertaken by deriving key performance indicators such as total sales, total profit, total orders, and average discount by using DAX measures. The visualization of these measures was performed on the dashboard in Power BI which consisted of bar charts, line charts, maps and KPI cards. Interactive slicers of region and product category as well as year were enabled so as to explore the data dynamically. Lastly, insights based on the visualizations were generated, ushering recommendations on how to improve performance in the region, on specific product categories with the highest margins, and to develop better customer targeting strategies.



## 4.2 Schema

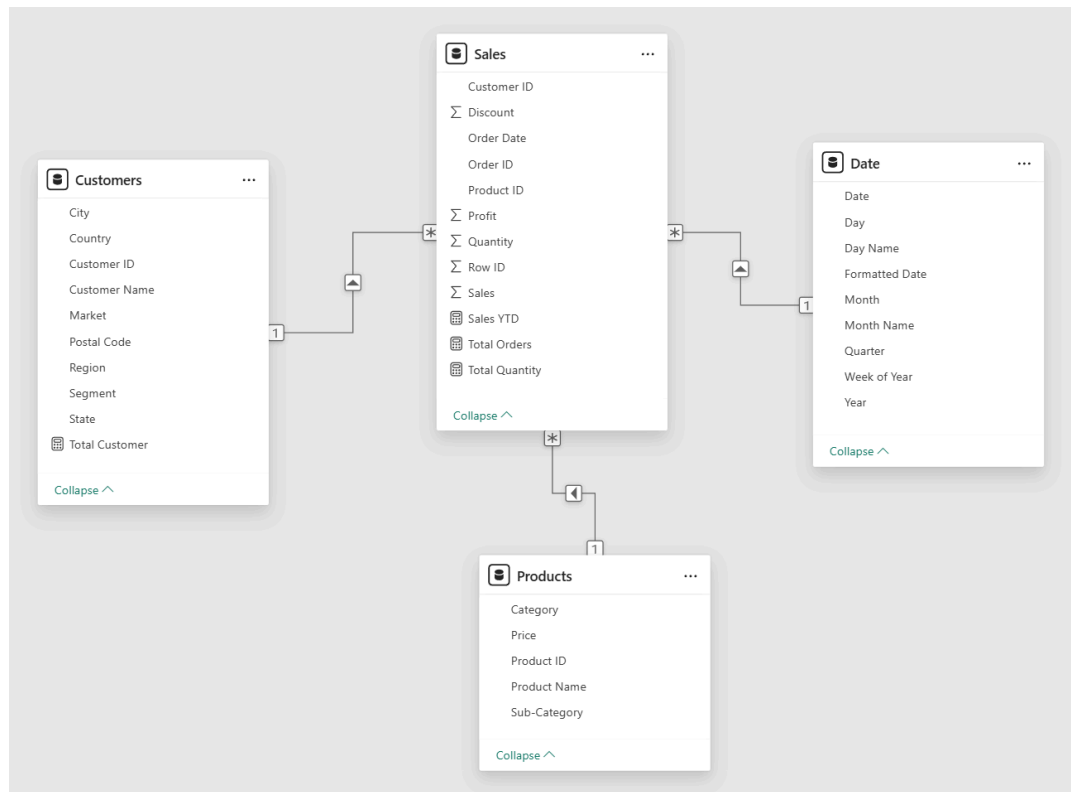


Table (Column)	Relationship	Table (Column)
Sales (Customer ID)	Many to One	Customers (Customer ID)
Sales (Order Date)	Many to One	Date (Formatted Date)
Sales (Product ID)	Many to One	Product (Product ID)

### Star Schema Design

- **Fact Table:** Sales (Order ID, Sales, Quantity, Discount, Profit)
- **Dimension Table:**
  - Customer (Customer ID, Customer Name, Segment, Region)
  - Product (Product ID, Category, Sub-category)
  - Date (Year, Month, Quarter)

## 5.0 DATA ANALYSIS AND VISUALISATION

### 5.1 DATA ANALYSIS

The Global Superstore dataset was analyzed to understand sales performance, profitability, and customer behavior across different dimensions. Descriptive statistics were applied to summarize key performance indicators (KPIs).

After finishing setting up the dashboard, the dataset revealed:

- Total Sales: the sum of all order revenue.
- Total Customers: the overall sum of customers that ordered
- Total Quantity Sold: aggregate number of units sold.
- Total Orders: total unique transactions recorded.

The analysis also highlighted important relationships. For example, regions such as the United States and Western Europe contributed disproportionately higher sales compared to other areas, while certain product categories like Technology consistently generated higher profit margins than categories such as Furniture. Additionally, customer segmentation revealed that Corporate and Consumer segments drove the majority of sales, whereas the Home Office segment lagged behind. These findings provided the foundation for visualization in Power BI, where data could be explored interactively across time, geography, and product categories.

### 5.2 VISUALISATION

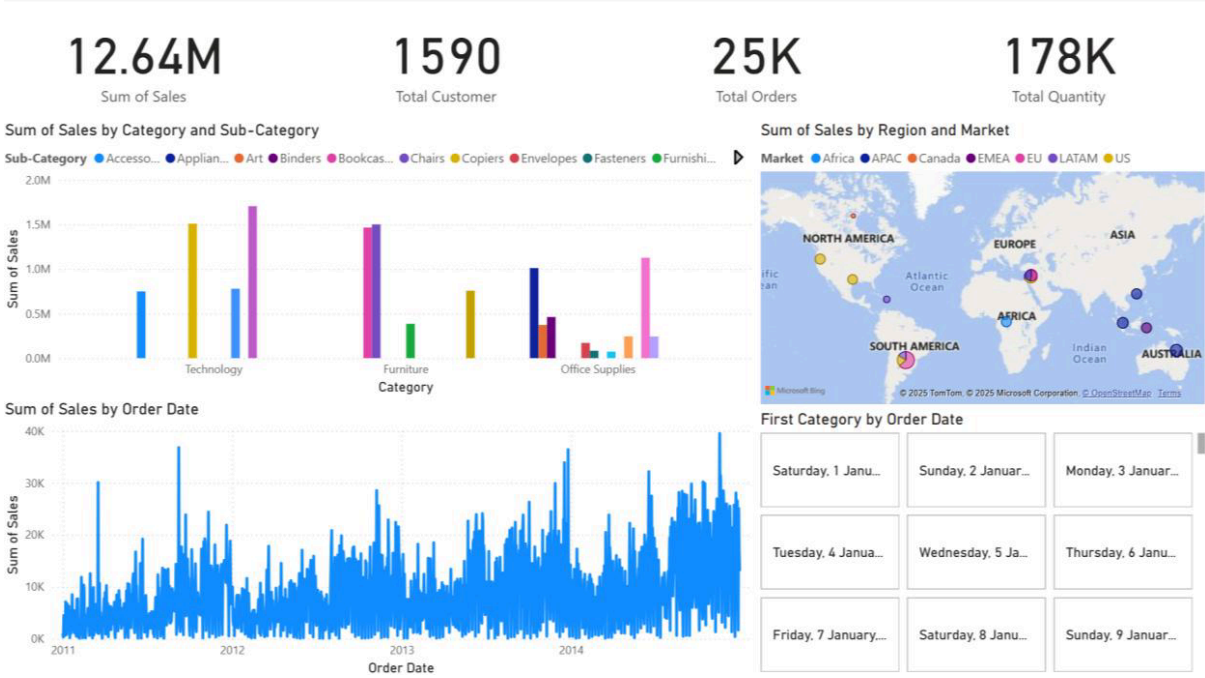
The dashboard provided a comprehensive view of the company's sales operations, making it possible to identify top-performing regions, profitable product lines, and customer behaviors in real time. To improve interactivity, slicers and filters were added for Region, Product Category, Year, and Customer Segment. These enabled users to drill down into specific dimensions and generate ad-hoc insights based on their business needs. To better illustrate trends and enable decision-making, an interactive dashboard was created in Power BI.

The dashboard combined five main types of visuals:

1. KPI Cards: Displayed overall metrics such as *Total Sales*, *Total Profit*, *Number of Orders*, and *Average Discount*.
2. Bar/Column Chart: Compared sales and profit across product categories and sub-categories, highlighting areas of strength and weakness.
3. Line Chart (Time-Series): Tracked sales performance over time, allowing seasonal patterns and year-on-year growth to be observed.
4. Map Visualization: Displayed sales distribution across regions and countries, showing geographical opportunities and underperforming areas.
5. Button Slicer: Enabled users to dynamically filter the dashboard by Region, Product Category, and Customer Segment. This feature provided interactivity, allowing users to explore customized views and focus on areas of interest

6.0 RESULT

6.1 Dashboard



6.2 KPI Cards

Total Sales	12.64M
Total Customer	1,590
Total Order	25K
Total Quantity	178K

## **7.0 RECOMMENDATION**

### **1) Strengthen Corporate Segment Engagement**

Corporate customers generate higher margins and have strong potential for repeat purchases. Loyalty programs with bulk discounts, dedicated account managers, and subscription-based ordering can improve retention and secure steady revenue.

### **2) Expand the Technology Category**

Technology is the top-performing category, led by Copiers and Accessories. Increasing marketing efforts, creating bundled promotions, and partnering with major suppliers will strengthen this segment's growth and profitability.

### **3) Enhance Furniture Profitability**

Furniture sales are stable but less profitable. Reviewing supply chain costs, introducing premium/custom lines, and refining discount strategies can improve profit margins without reducing sales.

### **4) Target Emerging APAC Markets**

APAC shows strong growth potential compared to the saturated US market. Localized pricing, targeted promotions, and partnerships with regional distributors can boost sales and reduce operational costs.

### **5) Leverage Seasonal Trends for Promotions**

Sales peaks in late 2012 and mid-2014 suggest seasonal influence. Planning mid-year and year-end campaigns, combined with early bird offers, can maximize revenue during these periods.

### **6) Focus on High-Value Customers**

A small group of repeat buyers contributes significantly to sales. Personalized offers, loyalty rewards, and predictive analytics can increase their lifetime value and encourage further purchases.

### **7) Optimize Discount Strategies**

Excessive discounting can reduce profits. A tiered discount system with continuous ROI monitoring will ensure discounts drive sales effectively without harming margins.

## 8.0 CONCLUSION

In conclusion, this project successfully demonstrated the use of Power BI as an effective tool for decision making and business intelligence through the analysis of the Global Super Store dataset. By transforming raw data into a star schema model and applying data cleaning, DAX calculations, and dashboard visualizations, meaningful insights were uncovered regarding customer segmentation, product profitability, and global sales trends. The results not only provided a clearer picture of historical sales patterns but also highlighted opportunities for business growth, such as capitalizing on the APAC market and focusing on high-margin product categories like technology. The use of Power BI facilitated intuitive data exploration and decision support through interactive dashboards, proving its value in academic and professional contexts.

Further development of the project could be enhanced by integrating real-time data streams, extending the analysis to additional performance factors such as delivery times and shipping costs, and applying predictive modelling techniques to forecast future sales trends. Overall, the project underscores the value of data-driven decision-making and reinforces the importance of modern BI tools in guiding business strategies.