BIG DATA AND BUSINESS INTELLIGENCE

NAME: REBECCA AGBOLADE

STUDENT ID: D3077427

PROJECT TOPIC: SUPERSTORE DATA ANALYSIS

Contents

TASK 2- TECHNICAL REPORT 4
1. Introduction 4
1.1 **Introduction** 4
1.2 **Data Set Source and Description** 4
1.2.1 Source 4
1.2.2 Rationale for Choosing Dataset 5
1.3 **Objectives** 6
1.4 **Scope of Work and Possible Insight Questions** 6
2. Data Pre-Processing and Cleaning 7
2.1. Loading the Data 7
CHECKING COLUMN QUALITY 9
REMOVING COLUMNS: 9
Changing Data Type 11
REMOVE ERRORS 11
3. DATA MODELLING 12
Duplicating the superstore Dataset 12
Data Modelling 14
Data Model Showing Star Schema After Relationship was built 15
4. DATA ANALYSIS EXPRESSION(DAX) 15
5. DASHBOARD 20
5.1 Introduction 20
5.2. DAX MEASURES AND COLUMN CREATED/Dashboard Overview 22
SALES ANALYSIS DASHBOARD OVERVIEW 25
PRODUCT ANALYSIS DASHBOARD OVERVIEW 29
CUSTOMER ANALYSIS DASHBOARD OVERVIEW 32
GEOGRAPHICAL ANALYSIS DASHBOARD OVERVIEW 35
MARKETING ANALYSIS DASHBOARD OVERVIEW 38
KPI ANALYSIS DASHBOARD OVERVIEW 40
FORECAST ANALYSIS DASHBOARD OVERVIEW 43
6.Technical Details 44
Task 3 – Stakeholder Report 44
7.Executive Summary 44
7.1. Relevance of Power BI in sales Data 44
1. Data Source Description and BI Requirements 45
1.1 Introduction 45
1.2 Data Set Source and Description 45

1.2.1 Source 45
BI KPI Requirement and Questions 46
Business Intelligence Questions 47
2. Methodology 49
Why Power BI? 49
2.1 Data Collection and Preparation 49
Key Findings 50
SALES ANALYSIS DASHBOARD OVERVIEW 52
PRODUCT ANALYSIS DASHBOARD OVERVIEW 54
CUSTOMER ANALYSIS DASHBOARD OVERVIEW 55
GEOGRAPHICAL ANALYSIS DASHBOARD OVERVIEW 56
MARKETING ANALYSIS DASHBOARD OVERVIEW 56
KPI ANALYSIS DASHBOARD OVERVIEW 57
FORECAST ANALYSIS DASHBOARD OVERVIEW 58
3. Challenges and Limitations 58
4. Conclusions and Recommendations 58
Summary: 58
Courses of Action for Stakeholders 59
5. Reference 60

TASK 2- TECHNICAL REPORT
BUSINESS INTELLIGENCE MODEL

1. Introduction

1.1 Introduction

While exploring the dynamic of commercial context, a superstore big company was used in testing my expertise and knowledge to deduce vital sales information based on this dataset by identifying the performance of products, geographically defined dynamics, and a consumer preference which is crucial in the current context of more prominent customer and market demands. When one looks closer, one finds that using predictive model is a strategic tool that can help the business forecast sells and maximize profitability. Tailored plans are driven by this through analysis, which helps uncover revenue-generating possibilities and potential problems. Gaining useful insights from sales data is essential for staying ahead of the competition and making well-informed, strategic decisions in the highly competitive market as the company environment changes.

This BI project seeks to analyse data on Superstore using a supermarket in USA as a case study in order to gain insight on the sales, profitability and best time to shop for goods and how this factor can be driven to generate sales, profit etc.

1.2 Data Set Source and Description

1.2.1 Source

The dataset used is a Kaggle dataset which contains information related to Sales, Profits and other interesting facts of a superstore giant dataset.

Kaggle: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final.

This dataset contains 21 columns, which contains information that best gives an overall knowledge about what the dataset encompasses and how they will contribute to generating insights.

S/N	Column Name	Description
1	Row ID	Unique ID for each row
2	Order ID	Unique order ID for each Customer
3	Order Date	Order Date of the product
4	Ship Date	Shipping date of the product
5	Ship Mode	Shipping mode specified by the customer
6	Customer ID	Unique ID to identify each customer
7	Customer Name	Name of the customer
8	Segment	The segment where the customer belongs
9	Country	Country of residence of the customer

10	City	City of residence of the customer
11	State	State of residence of the customer
12	Postal Code	Postal Code of every customer
13	Category	Category of the product ordered
14	Subcategory	Sub-category of the product ordered
15	Product Name	Name of the product
16	Sales	Sales of the product
17	Quantity	Quantity of the product
18	Discount	Discount Provided
19	Profit	Profit/loss incurred

Table 1: Data Description

Snapshot of the dataset in excel is shown below:

Row ID	Order ID	Order Dat	Ship Date	Ship Mod	Customer	Customer	Segment	Country	City	State	Postal Co	Region	Product ID	Category	Sub-Categ	Product N	Sales	Quantity	Discount	Profit
1	CA-2016-1	########	########	Second Cl	CG-12520	Claire Gut	Consumer	United Sta	Henderso	Kentucky	42420	South	FUR-BO-1(	Furniture	Bookcases	Bush Som	261.96	2	0	41.9136
2	CA-2016-1	########	########	Second Cl	CG-12520	Claire Gut	Consumer	United Sta	Henderso	Kentucky	42420	South	FUR-CH-1(	Furniture	Chairs	Hon Delux	731.94	3	0	219.582
3	CA-2016-1	########	6/16/2016	Second Cl	DV-13045	Darrin Van	Corporate	United Sta	Los Angel	California	90036	West	OFF-LA-1(	Office Sup	Labels	Self-Adhe	14.62	2	0	6.8714
4	US-2015-1	########	10/18/201	Standard (	SO-20335	Sean O'Dc	Consumer	United Sta	Fort Laude	Florida	33311	South	FUR-TA-1(	Furniture	Tables	Bretford C	957.5775	5	0.45	-383.031
5	US-2015-1	########	10/18/201	Standard (	SO-20335	Sean O'Dc	Consumer	United Sta	Fort Laude	Florida	33311	South	OFF-ST-10	Office Sup	Storage	Eldon Fold	22.368	2	0.2	2.5164
6	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	FUR-FU-1(	Furniture	Furnishing	Eldon Exp	48.86	7	0	14.1694
7	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-AR-1(	Office Sup	Art	Newell 32	7.28	4	0	1.9656
8	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	TEC-PH-1(	Technolog	Phones	Mitel 532(	907.152	6	0.2	90.7152
9	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-BI-10	Office Sup	Binders	DXL Angle	18.504	3	0.2	5.7825
10	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-AP-1(	Office Sup	Appliance	Belkin F5(	114.9	5	0	34.47
11	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	FUR-TA-1(	Furniture	Tables	Chromcra	1706.184	9	0.2	85.3092
12	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	TEC-PH-1(	Technolog	Phones	Konftel 25	911.424	4	0.2	68.3568
13	CA-2017-1	4/15/2017	4/20/2017	Standard (	AA-10480	Andrew A	Consumer	United Sta	Concord	North Car	28027	South	OFF-PA-1(	Office Sup	Paper	Xerox 196	15.552	3	0.2	5.4432
14	CA-2016-1	########	########	Standard (	IM-15070	Irene Mac	Consumer	United Sta	Seattle	Washingt	98103	West	OFF-BI-10	Office Sup	Binders	Fellowes I	407.976	3	0.2	132.5922
15	US-2015-1	11/22/201	11/26/201	Standard (	HP-14815	Harold Pa	Home Off	United Sta	Fort Worth	Texas	76106	Central	OFF-AP-1(	Office Sup	Appliance	Holmes Re	68.81	5	0.8	-123.858
16	US-2015-1	11/22/201	11/26/201	Standard (	HP-14815	Harold Pa	Home Off	United Sta	Fort Worth	Texas	76106	Central	OFF-BI-10	Office Sup	Binders	Storex Du	2.544	3	0.8	-3.816
17	CA-2014-1	########	11/18/201	Standard (	PK-19075	Pete Kriz	Consumer	United Sta	Madison	Wisconsin	53711	Central	OFF-ST-10	Office Sup	Storage	Stur-D-Stc	665.88	6	0	13.3176
18	CA-2014-1	5/13/2014	5/15/2014	Second Cl	AG-10270	Alejandro	Consumer	United Sta	West Jord	Utah	84084	West	OFF-ST-10	Office Sup	Storage	Fellowes S	55.5	2	0	9.99
19	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	OFF-AR-1(	Office Sup	Art	Newell 34	8.56	2	0	2.4824
20	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	TEC-PH-1(	Technolog	Phones	Cisco SPA	213.48	3	0.2	16.011
21	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	OFF-BI-10	Office Sup	Binders	Wilson Jo	22.72	4	0.2	7.384
22	CA-2016-1	########	12/13/201	Standard (	KB-16585	Ken Black	Corporate	United Sta	Fremont	Nebraska	68025	Central	OFF-AR-1(	Office Sup	Art	Newell 31	19.46	7	0	5.0596
23	CA-2016-1	########	12/13/201	Standard (	KB-16585	Ken Black	Corporate	United Sta	Fremont	Nebraska	68025	Central	OFF-AP-1(	Office Sup	Appliance	Acco Six-C	60.34	7	0	15.6884
24	US-2017-1	7/16/2017	7/18/2017	Second Cl	SF-20065	Sandra Fla	Consumer	United Sta	Philadelp	Pennsylva	19140	East	FUR-CH-1(	Furniture	Chairs	Global De	71.372	2	0.3	-1.0196
25	CA-2015-1	9/25/2015	9/30/2015	Standard (	EB-13870	Emily Burt	Consumer	United Sta	Orem	Utah	84057	West	FUR-TA-1(	Furniture	Tables	Bretford C	1044.63	3	0	240.2649
26	CA-2016-1	1/16/2016	1/20/2016	Second Cl	EH-13945	Eric Hoffm	Consumer	United Sta	Los Angel	California	90049	West	OFF-BI-10	Office Sup	Binders	Wilson Jo	11.648	2	0.2	4.2224
27	CA-2016-1	1/16/2016	1/20/2016	Second Cl	EH-13945	Eric Hoffm	Consumer	United Sta	Los Angel	California	90049	West	TEC-AC-1(	Technolog	Accessorie	Imation 8(	90.57	3	0	11.7741
28	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	FUR-BO-1(	Furniture	Bookcases	Riverside	3083.43	7	0.5	-1665.05
29	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	OFF-BI-10	Office Sup	Binders	Avery Rec	9.618	2	0.7	-7.0532
30	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	FUR-FU-1(	Furniture	Furnishing	Howard M	124.2	3	0.2	15.525
31	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	OFF-EN-1(	Office Sup	Envelopes	Poly Strin	3.264	2	0.2	1.1016
32	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	OFF-AR-1(	Office Sup	Art	BOSTON M	86.304	6	0.2	9.7092
33	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	OFF-BI-10	Office Sup	Binders	Acco Pres	6.858	6	0.7	-5.715
34	US-2015-1	9/17/2015	9/21/2015	Standard (	TB-21520	Tracy Blur	Consumer	United Sta	Philadelp	Pennsylva	19140	East	OFF-AR-1(	Office Sup	Art	Lumber Cr	15.76	2	0.2	3.546
35	CA-2017-1	10/19/201	10/23/201	Second Cl	MA-17560	Matt Abel	Home Off	United Sta	Houston	Texas	77095	Central	OFF-PA-1(	Office Sup	Paper	Easy-stapl	29.472	3	0.2	9.9468
36	CA-2016-1	########	########	First Class	GH-14485	Gene Hale	Corporate	United Sta	Richardso	Texas	75080	Central	TEC-PH-1(	Technolog	Phones	GE 30524E	1097.544	7	0.2	123.4737

Fig 1: Dataset overview in excel.

1.2.2 Rationale for Choosing Dataset

This dataset was chosen for several reasons:

1. Entireness of the Data collection: The dataset, comprises 9994 entries and 21 columns of data, is quite extensive. This vast amount of data is vital for conducting a detailed analysis, draw significant trends, and identify patterns that could contribute to a comprehensive understanding of insights regarding sales.
2. Values Uniqueness: It exhibits an extensive assortment of items or operations and an interesting variety of information with 5009 distinct values across the dataset. With this range of data, in-depth studies may be conducted and insights that might not be seen in datasets with fewer unique values might be found.

3. Dataset Sizes and complexity: The sheer size of the dataset—nearly 10,000 entries—offers ample of possibilities to demonstrate and practice significant company abilities. With the guidance of this dataset, it is feasible to show and put into practice many tasks, including data cleaning, the pre-processing modelling, and the development of established columns and measurements.
4. Combating Hurdles: The nature of its complexity, the dataset introduces big data-related barriers, such as probable issues with missing data and formatting changes. A vital component of this research involves tackling the problems, which exhibits one's ability for dealing with the complexities of real-world data.

1.3 Objectives

This dataset's aims involve conducting an in-depth investigation of sales patterns and developments. The purpose of the 21 columns, 9994 rows, and 5009 unique values is to uncover valuable information, find connections, and show that one is proficient in business capabilities. Recognizing economic trends, resolving big data issues, and deploying current information to facilitate rational choices constitutes the main objectives. This dataset is a vital resource for honing analytical methods, building visualizations, and showcasing proficiency in managing the multifaceted nature of everyday problems sales data.

1.4 Scope of Work and Possible Insight Questions.

The superstore dataset at the core of this Power BI project is aimed specifically at the USA as a case study. The project's goal is to explore and highlight key aspects of the superstore's US operations by utilizing the vast amount of information in the dataset. Discovering sales patterns, regional differences, and the performance of products makes up the goals. The project's goal is to deliver practical recommendations for business approach optimization in the context of the US market by effectively employing Power BI capabilities. This project seeks to give insights on these questions.

1. what is the distribution of customers across different categories?
2. How does the sum of profit vary across different regions and year?
3. Which states are the top 5 in terms of Sales?
4. Examine discount rates and their correlation with sales, identifying impact on specific product categories.
5. Determine which subcategories are most important, monitor the progression of profit and identify regions where profitability has changed noticeably over time.

2. Data Pre-Processing and Cleaning

2.1. Loading the Data

The initial phase of this analysis involves loading the dataset into Power BI as the first step in this study. Choosing the 'Get Data' button on the home tab starts this process. Pressing the button displays a dropdown menu with several ways to import data into Power BI (Fig 2).

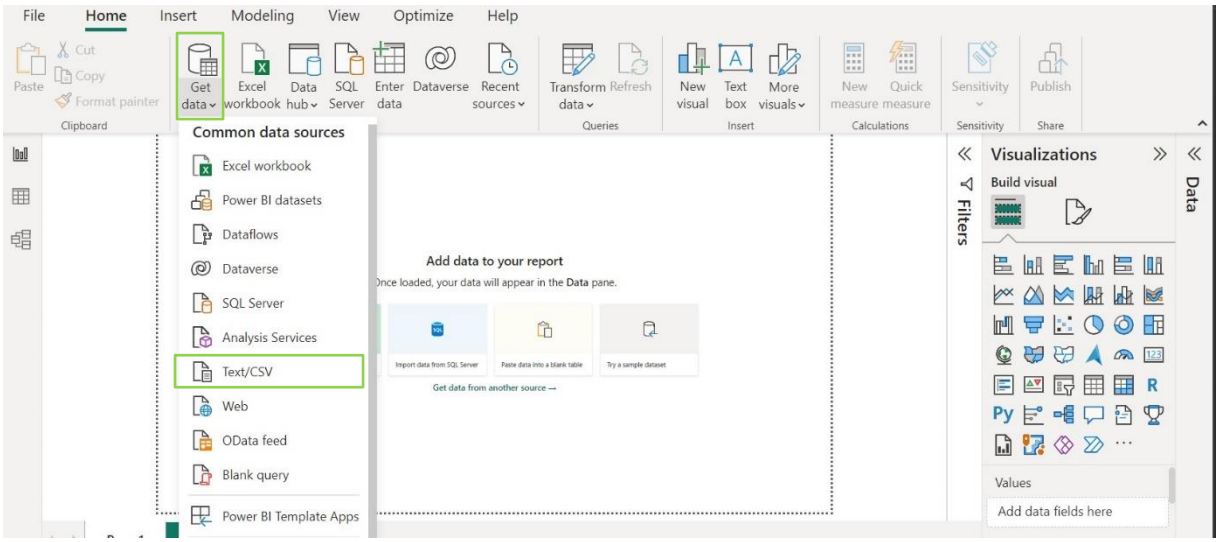
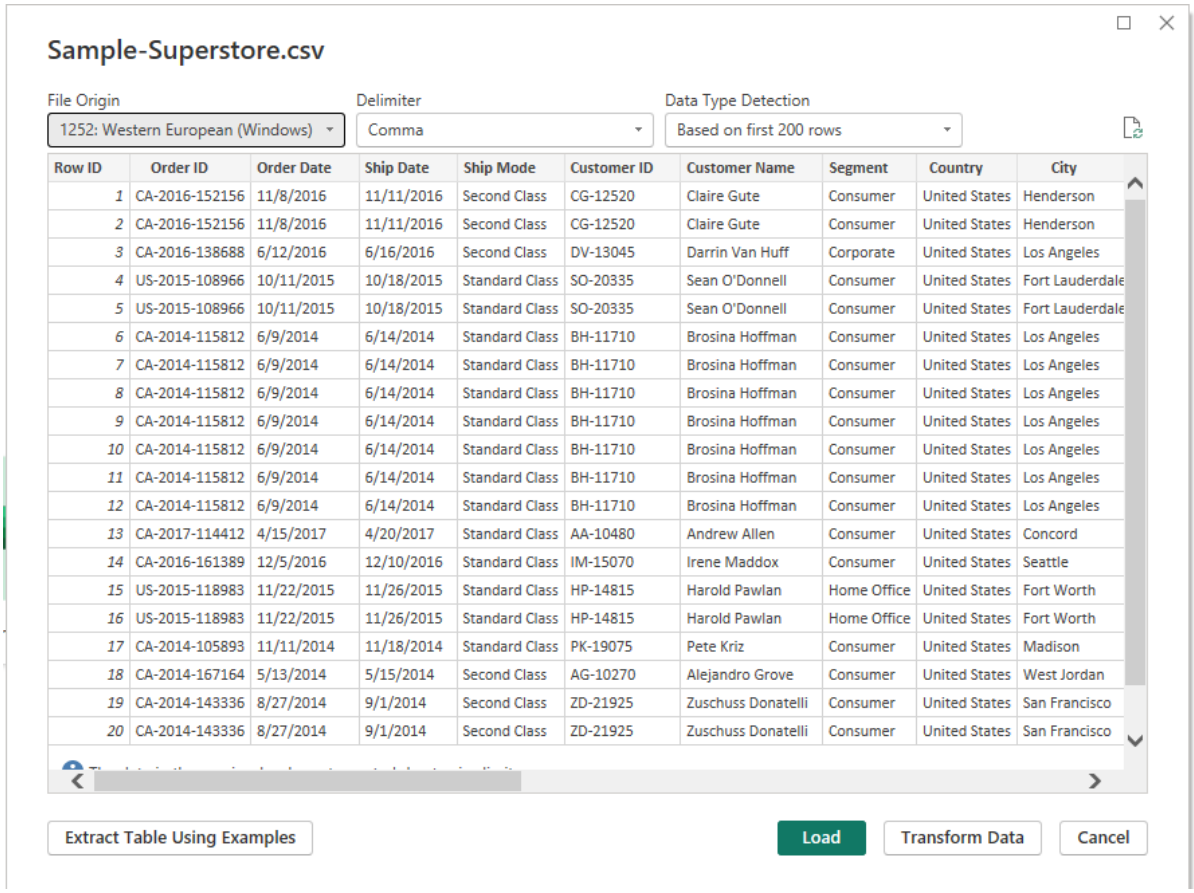


Fig 2: depicts numerous methods to load data into Power BI using the 'Get Data' popup menu.

The dataset, which had been given in CSV format, was loaded using the Text/CSV option to start the analysis. The Superstore Data dataset was selected in the first occurrence. A dialog box displaying the top 20 rows of the data (based on the top 200 rows) was displayed as a result of this operation. As shown in Fig 3, options to load the data, transform it, or abort the procedure were displayed. This essential phase enables decision-making and previewing prior to moving on with additional data processing.



Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City
1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles
4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
6	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
7	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
8	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
9	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
10	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
11	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
12	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
13	CA-2017-114412	4/15/2017	4/20/2017	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord
14	CA-2016-161389	12/5/2016	12/10/2016	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle
15	US-2015-118983	11/22/2015	11/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth
16	US-2015-118983	11/22/2015	11/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth
17	CA-2014-105893	11/11/2014	11/18/2014	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison
18	CA-2014-167164	5/13/2014	5/15/2014	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan
19	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco
20	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco

Fig 3: The dialog box displays the first 20 rows of the dataset along with options for importing.

With the final step of the dataset integration, the Power BI environment is now ready for further analysis, visualization, and exploration. The load option was chosen and the process of loading the Superstore data table was completed. The Superstore data table was successfully imported intoPower BI when the 'Load' option was clicked.

2.2 Data Cleaning

The next phase will consist of pre-processing and data cleansing, which will be handled with Power Query. The 'Transform data' choice on the home tab was chosen to start the process, as shown in Fig. 6. After clicking the 'Transform data' option from this menu, the Power Query Editor opened, as shown in Fig. 7. With comprehensive cleaning and prepping possible within the Power Query environment, this commences the process of data refining.

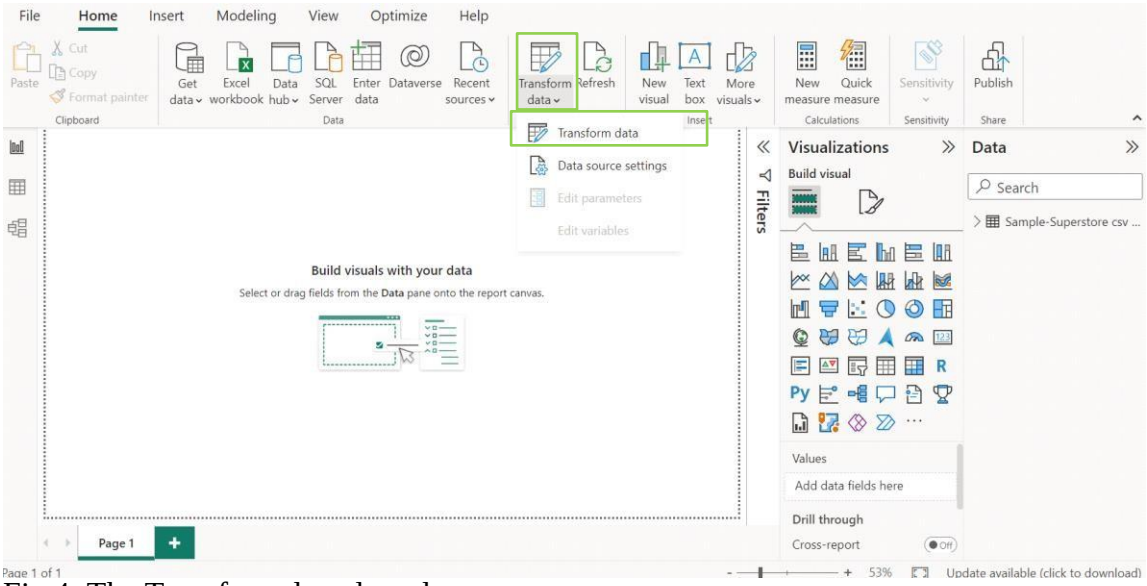


Fig 4: The Transform data drop down.

CHECKING COLUMN QUALITY

To check the for any error of empty column in the dataset, the view option was selected in the ribbon, column quality was then selected to show the quality of each column.

The column quality showed that the dataset had no missing values or error and validity for all columns were 100%.

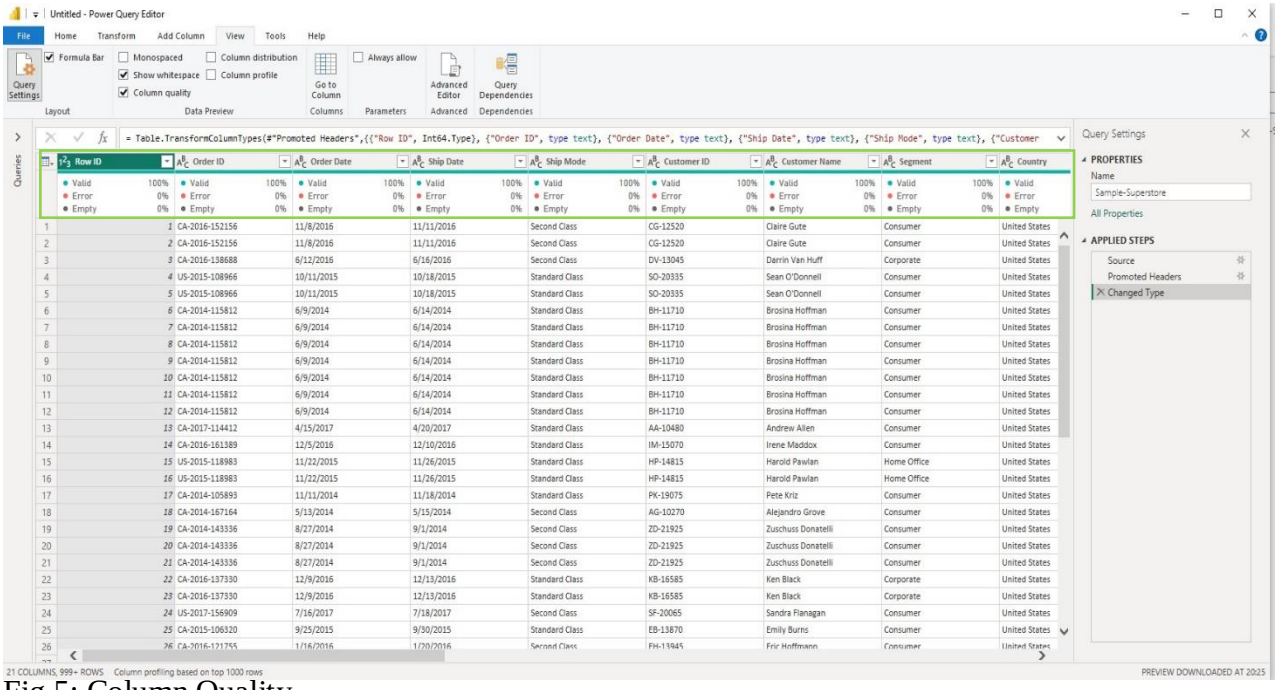


Fig 5: Column Quality.

REMOVING COLUMNS:

I concluded to eliminate the "Postal Code" column because it is redundant information that I already have for the country, state, and city. Since I have no intention to utilize postal codes for analysis on the visualization in my

visualizations, it appears unnecessary to include them. Furthermore, in most cases, keeping the column in place is not likely to produce any more insightful information.

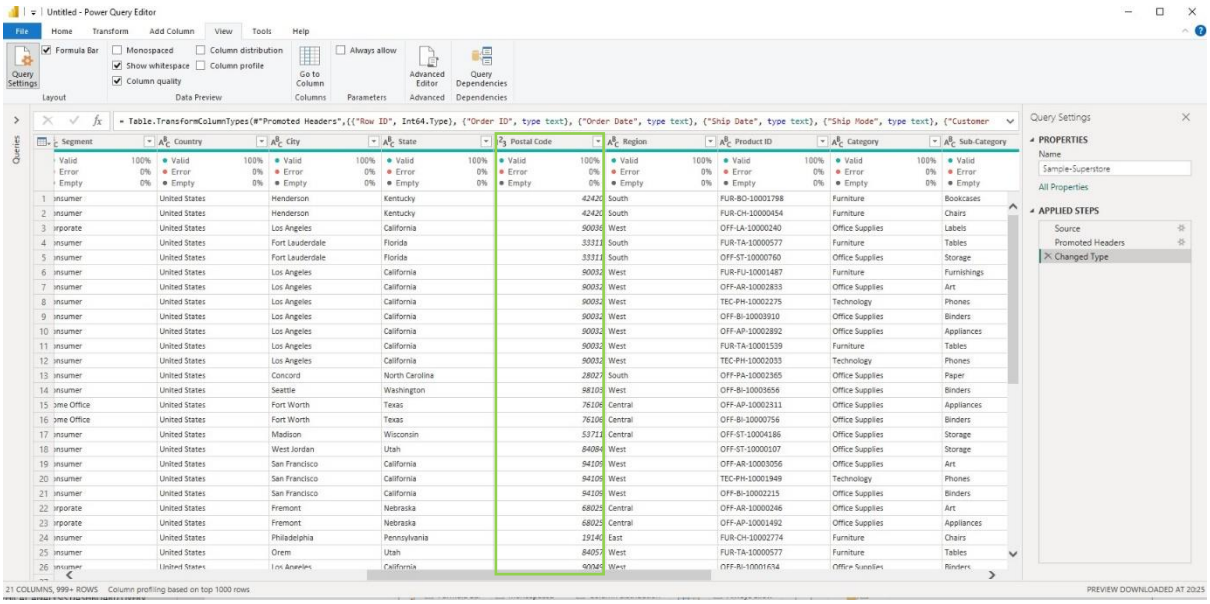


Fig 6: shows the Postal code column before removal.

A functional language created especially for data translation and manipulation within Microsoft Power Query is called the M language. M-Language was used for data cleaning in power query.
The M-language was used to remove the postal code column from the dataset. The M-Language formula used was:
Table.RemoveColumns(#"Changed Type", {"Postal Code"})

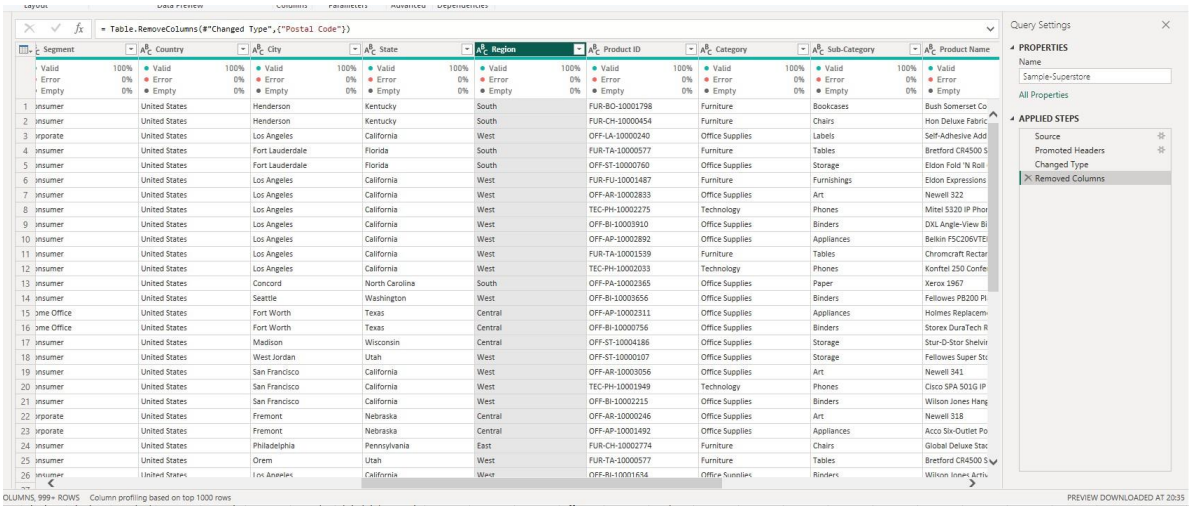


Fig 7: After removal of postal code column

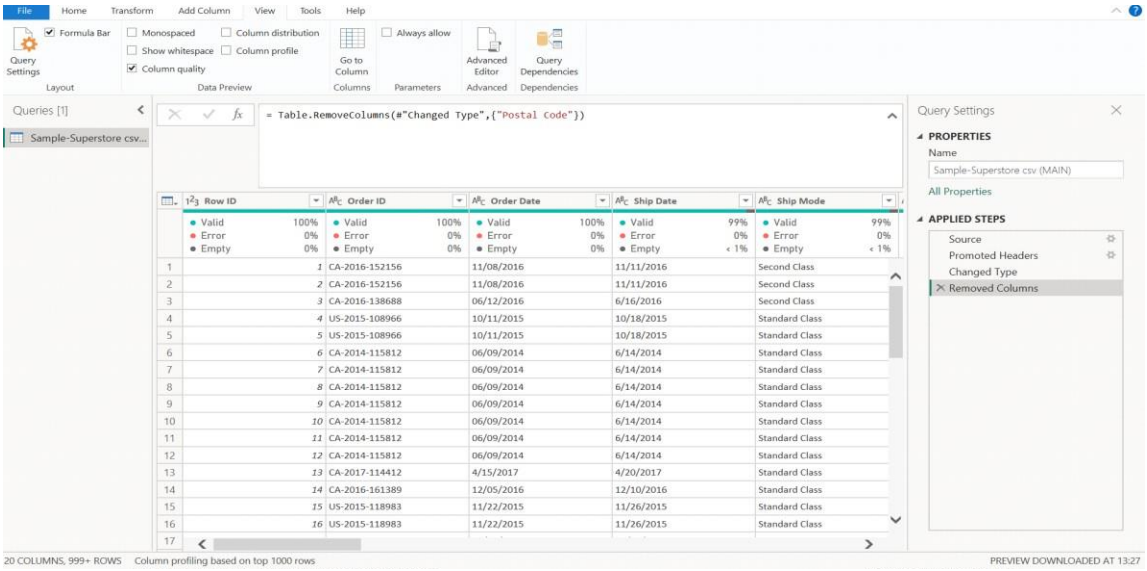


Fig 8: Dataset after postal code column was removed.

Changing Data Type

The "Order Date" and "Ship Date" columns are presently formatted as text types, necessitating a conversion to date types.

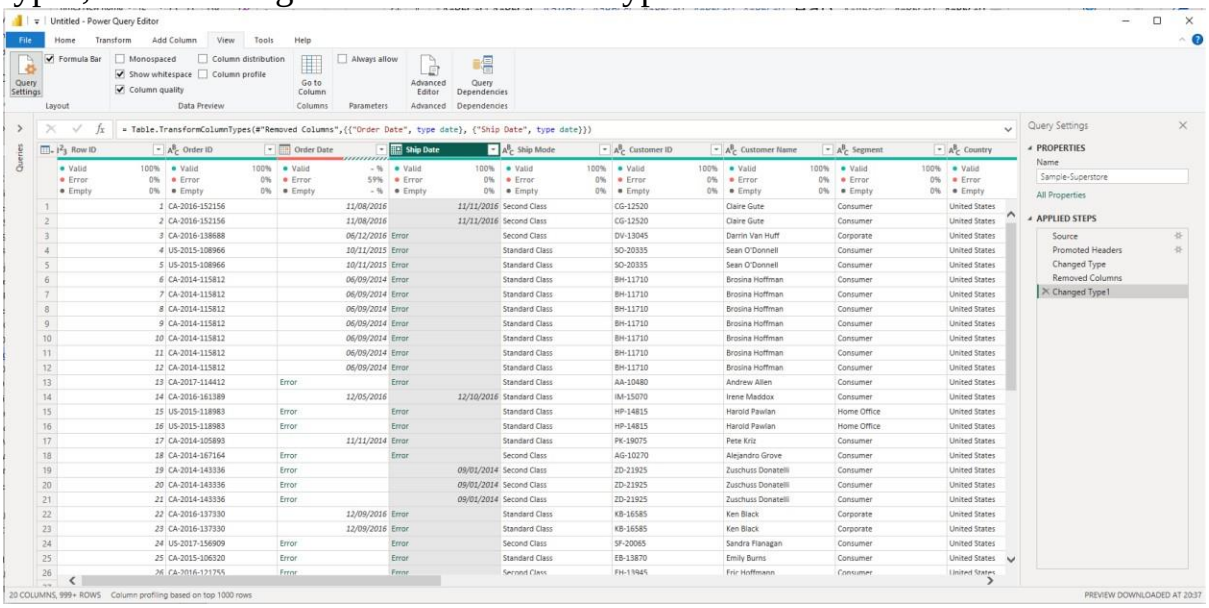


Fig. 9: Examining the Percentage of Empty Rows in Each Column

REMOVE ERRORS:

I addressed errors in the "Order Date" and "Ship Date" by right-clicking on the error-containing columns, selecting "Remove Error," and opting for removal. Leveraging Power Query Editor, I conducted meticulous data cleansing, elevating the dataset's quality and reliability for subsequent analysis.

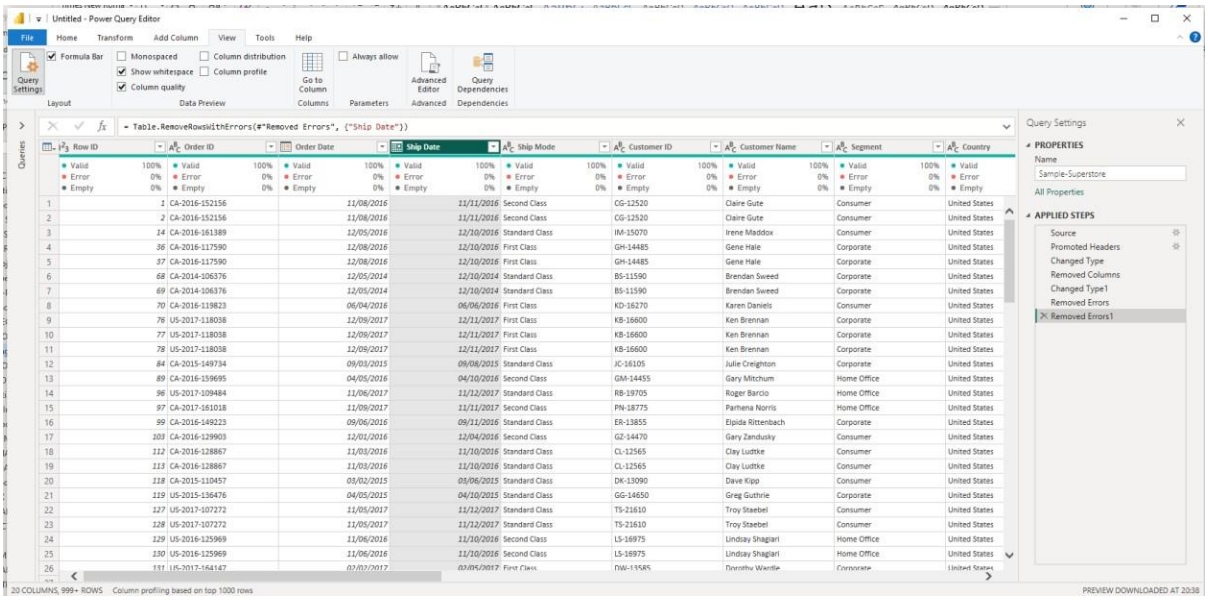


Fig 10: Dataset after removing errors.

3. DATA MODELLING

Following the completion of data cleaning, Fig.11 illustrates the current state of the data model.

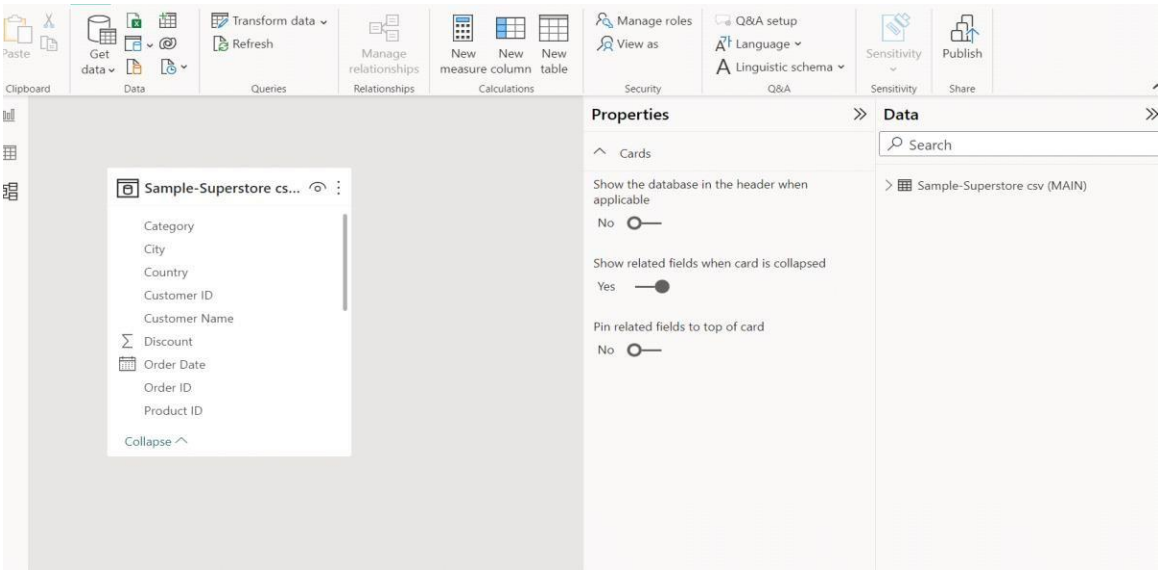


Fig 11: Data Modelling.

A consolidated table (Fact Table) in the present data model must be simplified into smaller tables (Dimension Tables). With the goal to reduce complexity, avoid data that is duplicated, and make analysis and visualization creation more straightforward, this phase is key.

Duplicating the superstore Dataset.

To create the dimension tables, the superstore dataset would be duplicated to create various tables. The duplicated tables would then be renamed to name of choice for the dimension table.

The first-dimension table to be created is the customer table. To create this, in the duplicated table, all columns except Customer ID and Customer Name will be eliminated. To achieve this, select these two columns, right-click, and choose 'Remove Other Columns.

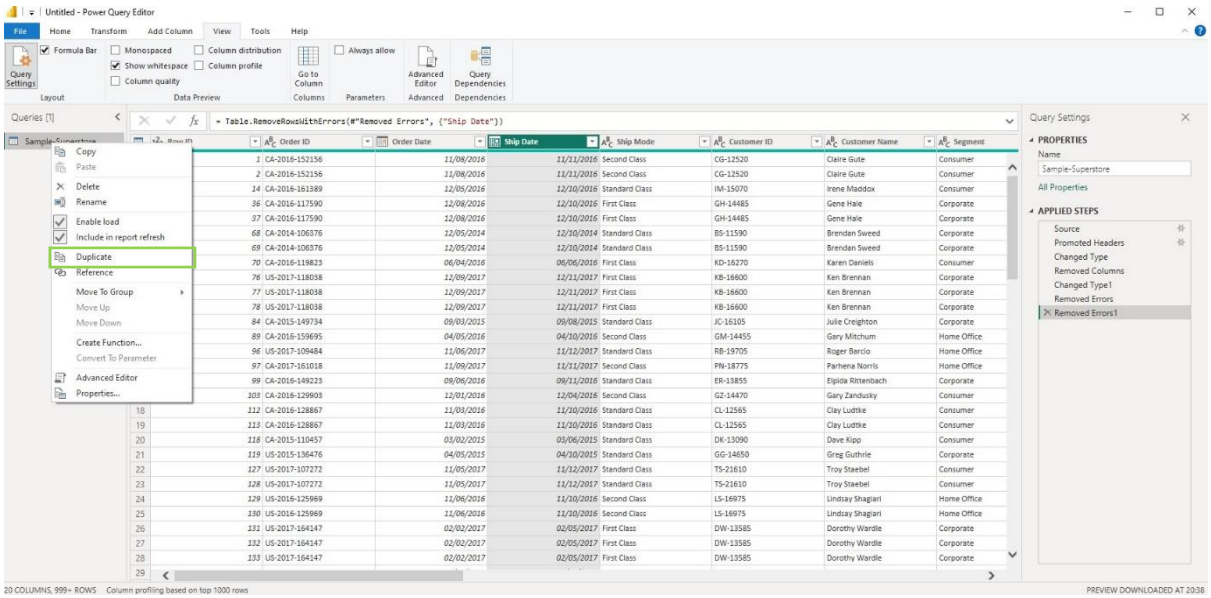


Fig 12: Duplicating the superstore Dataset.

This action yields a table containing only these specified columns. Subsequently, eliminate duplicates by selecting the customer ID column, right-clicking, and choosing 'Remove Duplicates.'

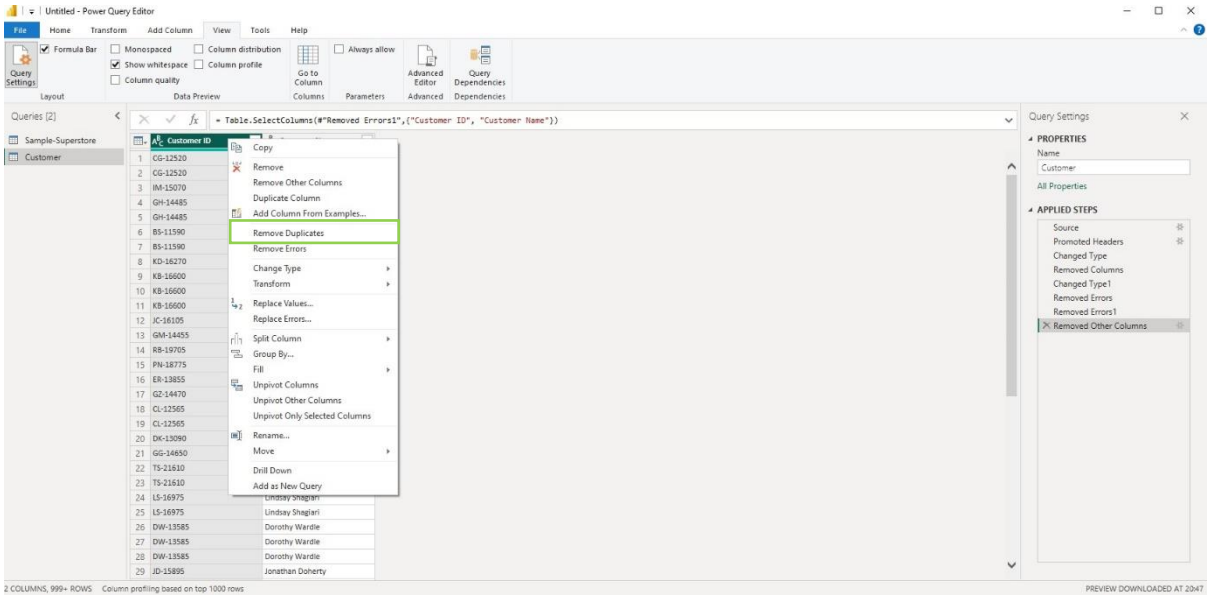


Fig 13: Removing Duplicates.

The same step was carried out to create 9 more-dimension tables which are: Discount, Location, Order, Product, Profit, Quantity, Sales, Segment, Shipping.

Data Modelling

Once all tables have been established, Fig. 34 illustrates the appearance of the data model prior to the creation of relationships.

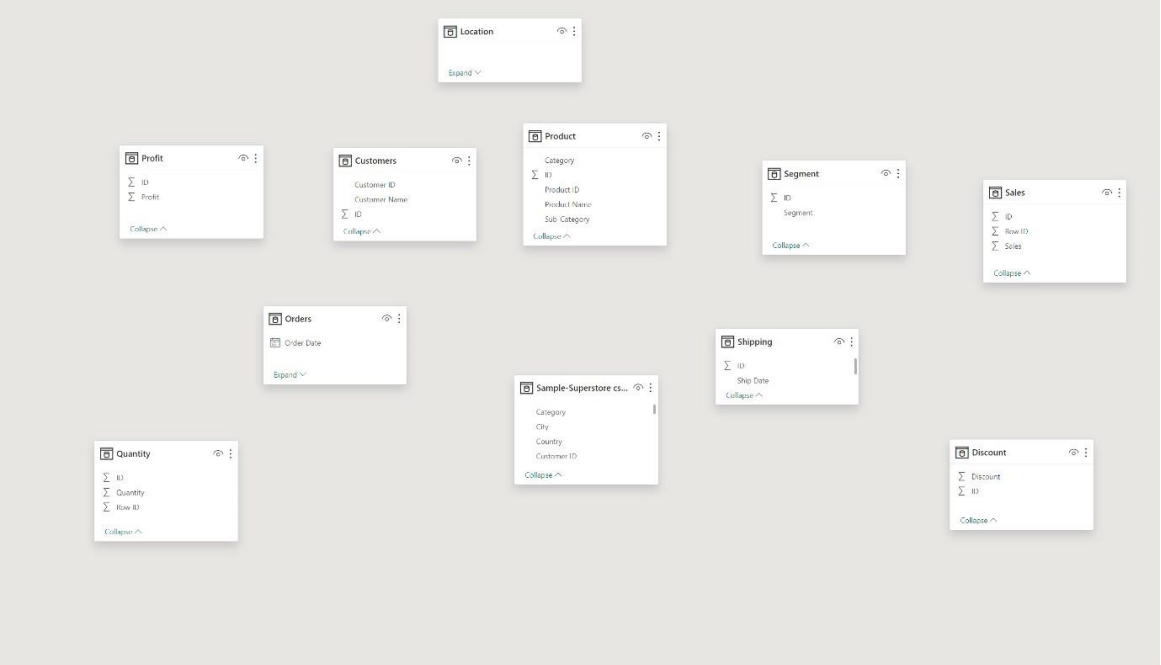


Fig 14: Data Model Before Relationship was created.

Relationships between the tables in our model must be created in order to enable the analysis that is necessary for this project. In the data model area, choose 'Manage relationships' from the main tab. Selecting 'New' will cause a dialog box to appear, allowing you to establish the fact table and dimension table's first relationship using the connecting column 'ROW ID.' With a cross-filter direction set to both, this establishes a One-to-One relationship. For each additional dimension table in the model, repeat these steps.

After all relationships have been created, our model uses a star schema, with multiple dimension tables (such as Order, Customer, Shipping, Sales, Quantity, and more) connected to a single fact table (Superstore) for this project.

Data Model Showing Star Schema After Relationship was built.

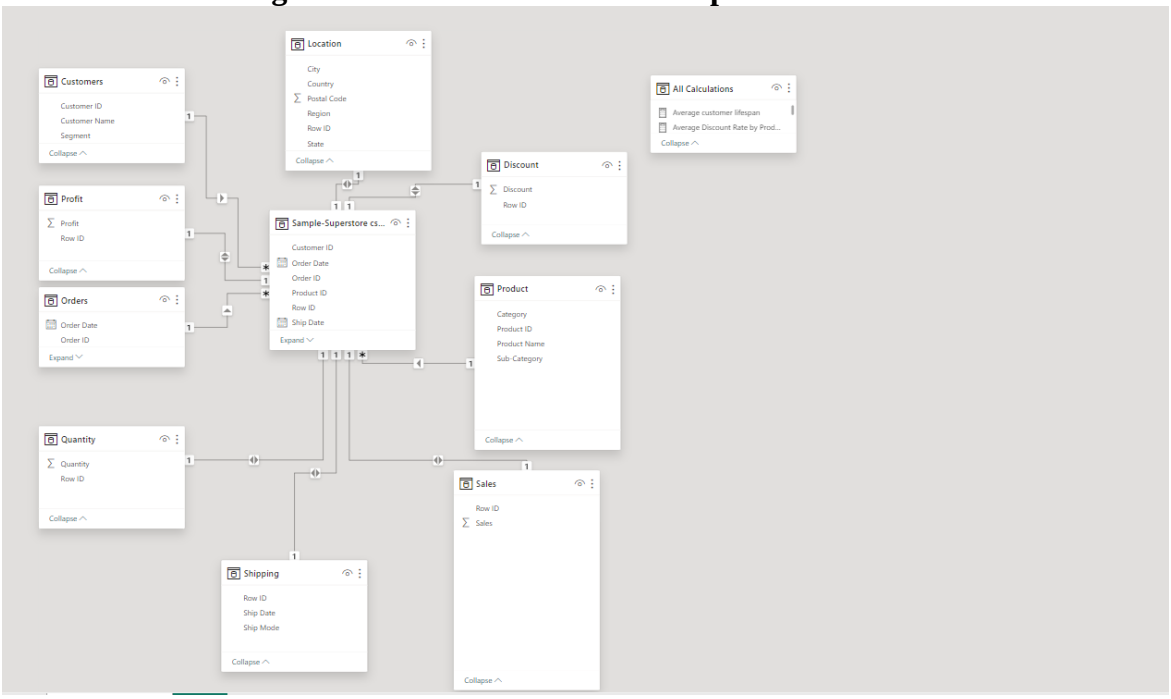


Fig 15: Data Modelling using star schema.

4. DATA ANALYSIS EXPRESSION(DAX)

Microsoft power Bi is a business analytics implementation that employs the formula language known as DAX (Data Analysis Expressions). With Power Bi datasets, it enables consumers to make distinct analyses and accumulation. Advanced data analysis and visualization are made possible by the creation of measures, calculated columns, and tables, all of which require DAX. Because of its adaptability and power Bi integration its essential for deriving valuable insights from data.

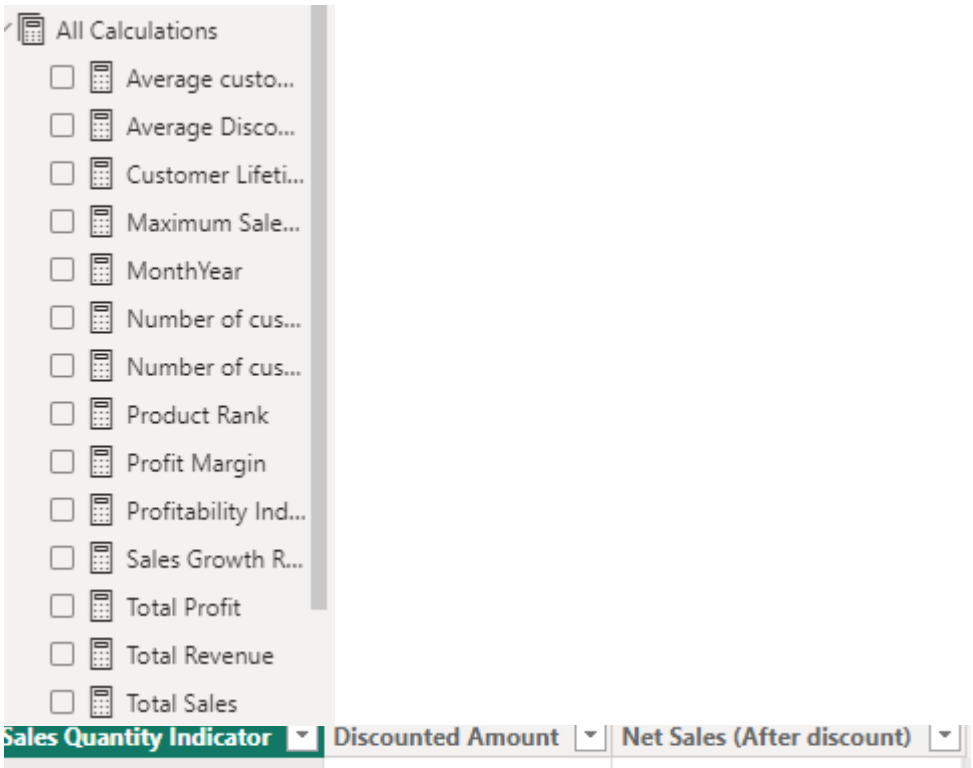


Fig 16: Data Measures created using DAX.

KPI MEASURES	DAX CODE	DAX DESCRIPTION
Average customer lifespan	Average customer lifespan = AVERAGEX (VALUES ('Sample-Superstore csv - Clean'[Customer ID]), DATEDIFF(MIN('Sample-Superstore csv - Clean'[Order Date]),MAX('Sample-Superstore csv - Clean'[Order Date]),DAY))	Average customer lifespan IS THE LENGTH OF TIME A CLIENT INTERACTS WITH A FIRM/BUSINESS BEFORE DISENGAGING
Average Discount Rate by Product	Average Discount Rate by Product = AVERAGE ('Sample-Superstore csv - Clean'[Discount])	Average Discount Rate by Product IS THE MEAN PERCENTAGE SUBTRACTED FROM THE TRANSACTION'S INITIAL PRICE
Customer Lifetime Value	Customer Lifetime Value = ([Total Sales]/ [Number of	Customer Lifetime Value THE APPROXIMATE

	customers]) *[Average customer lifespan]	AMOUNT OF PROFIT A CLIENT IS EXPECTED TO MAKE OVER THE DURATION OF DOING BUSINESS WITH AN ENTERPRISE
Maximum Sales In each Region	Maximum Sales In each Region = MAXX (VALUES ('Sample-Superstore csv - Clean'[Region]), CALCULATE (MAX ('Sample-Superstore csv - Clean'[Sales])))	Maximum Sales In each Region EXPLAINS HOW TO CONTINUOUSLY GET THE BEST REVENUE AND STRATEGIC GROWTH PROSPECTS
Number of customers by City	Number of customers by City = COUNTROWS (VALUES ('Sample-Superstore csv - Clean'[City]))	Number of customers by City ANALYZES THE CUSTOMER DISTRIBUTION BY THE NUMBER OF CUSTOMER IN EACH LOCATION.
Number of customers	Number of customers = COUNTROWS (VALUES ('Sample-Superstore csv - Clean'[Customer ID]))	Number of customers REFERS TO THE TOTAL COUNT OF INDIVIDUALS PURCHASING A PRODUCT/SERVICES
Product Rank	Product Rank = RANKX(ALL('Product'[Sub-Category]), [Profitability Index], DESC)	Product Rank ACCESSES ITEMS BASED ON PERFORMANCE, CRITERIA TO DETERMINE THEIR RELATIVE STANDING
Profit Margin	Profit Margin = [Total Profit] / [Total Sales]	Profit Margin IS THE PERCENTAGE OF REVENUE RETAINED AS PROFIT AFTER

		DEDUCTING COSTS AND EXPENSES
Profitability Index	Profitability Index = [Total Profit] / [Total Sales]	Profitability Index MEASURES PROJECT ATTRCTIVENESS BY COMPARING PRESENT VALUES TO INITIAL INVESTMENTS.
Sales Growth Rate	Sales Growth Rate = DIVIDE ('All Calculations'[Total Sales] - CALCULATE ('All Calculations'[Total Sales], DATEADD('Orders'[Order Date], -1, YEAR)), CALCULATE('All Calculations'[Total Sales], DATEADD('Orders'[Order Date], -1, YEAR))) * 100	Sales Growth Rate MEASURES THE PERCENTAGE INCREMENT IN SALES OVER A PERIOD OF TIME
Total Profit	Total Profit = SUM('Profit'[Profit])	Total Profit IS THE OVERALL PERCENTAGE OF EARNINGS GAINED AFTER DEDUCTING ALL COSTS AND EXPENSES
Total Revenue	Total Revenue = SUM (('Sample-Superstore csv - Clean'[Sales]))	Total Revenue IS THE OVERALL INCOME GENERATED FROM SALES BEFORE DEDUCTION OR EXPENSES
Total Sales	Total Sales = SUM('Sales'[Sales])	Total Sales REPRESENT THE OVERALL REVENUE GENERATED FROM ALL PRODUCTS

		OVER A SPECIFIC PERIOD
Discounted Amount	Discounted Amount = 'Sample-Superstore csv - Clean'[Sales] *'Sample-Superstore csv - Clean'[Discount]	Discounted Amount IS THE REDUCED VALUE AFTER SUBTRACTING A DISCOUNT FROM THE ORIGINAL PRICE.
Net Sales (After discount)	Net Sales (After discount) = 'Sample-Superstore csv - Clean'[Sales] - 'Sample-Superstore csv - Clean'[Discounted Amount]	Net Sales (After discount) TOTTAL REVENUE REMAINING AFTER DEDUCTING APPLIED DISCOUNTS FROM THE INITIAL AMOUNT.
profitability Ratio	profitability Ratio = 'Sample-Superstore csv - Clean'[Profit] /'Sample-Superstore csv - Clean'[Sales]	profitability Ratio ACCESSES A COMPANY LIABAILITY TO GENERARTE EARNINGS IRRESPECT TO ITS COSTS AND INVESTMENTS
Sales Quantity Indicator	Sales Quantity Indicator = IF ('Sample-Superstore csv - Clean'[Quantity] >5, "High", "Low")	Sales Quantity Indicator EXPLAINS THE TOTAL UNITS SOLD REFLECTING PRODUCT DEMAND AND PERFORMANCE OVERTIME.

Table 2: DAX

5. DASHBOARD

5.1 Introduction

For this project, the Superstore data which includes sales, customer, and shipping information is a vital resource for forecasting. Comprehensive investigation is provided with feasible by the interrelated fact and dimension tables, which create relationships inside a Star schema model. The vitality of the dataset originates from its capacity to offer perceptions into consumer conduct, the effectiveness of products, and geographical disparities, so enabling precise projections and tactical decision-making to maximize marketing results.
This report aims to accomplish this by addressing the following inquiries under each category:
The main interface has buttons that are placed in a strategic way to offer direct links for quick access to pages.

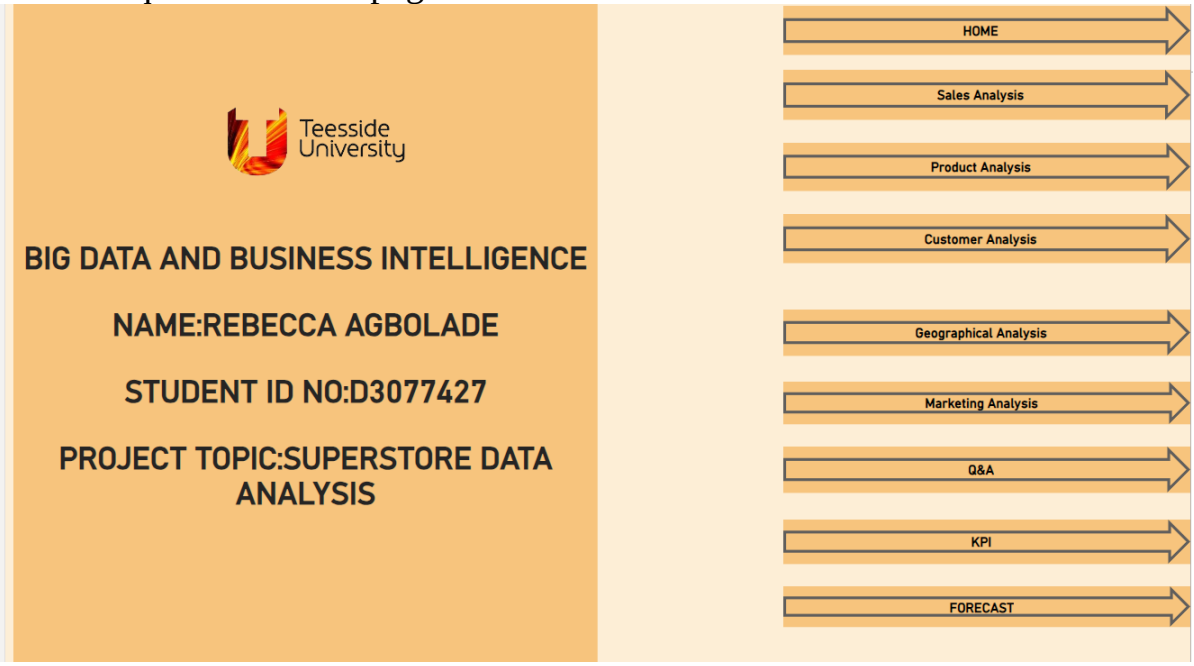


Fig 17: Cover page

Following the Homepage of the dashboard

Fig 18: Home page

A detailed explanation is provided below explaining each.

- ❖ Visuals
- ❖ KPI'S
- ❖ FIELDS
- ❖ BUSINESS QUESTIONS
- ❖ CHARTS
- ❖ DAX MEASURES.

Sales Analysis Questions:

- What is the distribution of customers across different categories?
- How does the sum of profit vary across different region and years?
- What is the contribution of each segment and category to the overall profit?
- Which states are the top 5 in terms of Sales?

Customer Analysis Questions:

- Identify the top-profit product by segments; compare sales quantities and discounts for strategic insights.
- Which year exhibits the highest and lowest overall business performance, considering profit, quantity and quality.
- What is the correlation between product rank and profitability, and how does this relationship vary across different quantity level?

Product Analysis Questions:

- Identify monthly sales variation by category revealing recurring trends or patterns throughout the year?
- Examine the discount rate and their correlation with sales, identifying the impact on specific product categories.
- Determine Which Subcategories are most important, monitor the progression of profit, and identify regions where profitability has changed noticeably over time.

Geographical Analysis Questions:

- Which Region are the highest and least more profitable?
- Which Region has the most Sales?
- Which City and Region has the most customers

Marketing Analysis Questions:

- What was the trend observed over time in terms of sales and profit?
- Which product categories contributes the most to our total sales, profit and discounted amount by category?
- How does the Application of discount impact our sales?

Principal Users Communities:

- CEO's and other management team members are actively involved in spotting important trends that improve company success.
- Marketing staff: Focus on products that are in great demand among a variety of demographics. To increase sales, strategically address products that are not doing so well.
- Operations team (Customer Analysis): Improving order processing efficiency and making the best use of available resources to deliver the greatest possible customer experience.
- Marketing team: Keeping an eye on Revenue, Earnings, and Expenses to help with financial planning.

5.2. DAX MEASURES AND COLUMN CREATED/Dashboard Overview:

I created different columns using DAX.
Here, I will elaborate and clarify the measures i have chosen to implore and why? in creating this Dashboard.
Dashboards Overview
KPI 1: Total Sales
Bi Question: What is the total Sale for the superstore dataset?
DAX Formula: Total Sales = SUM('Sales'[Sales])
Visual: Card

Fig 19: Card

KPI 2: Total Profit
Bi Question: What is the total Profit for the superstore dataset?
DAX Formula: Total Profit = SUM('Profit'[Profit])
Visual: Card

Fig 20: Card

KPI 3: Profitability index
Bi Question: What is the Profitability index for the superstore dataset?
DAX Formula: Profitability Index = [Total Profit] / [Total Sales]
Visual: Card

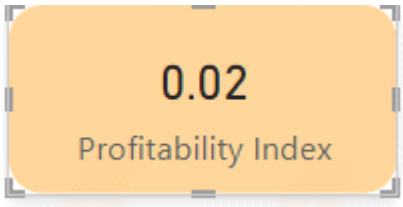


Fig 21: Card
KPI 4: Customer Lifetime Value
Bi Question: What is the Customer Lifetime Value for the superstore dataset?
DAX Formula: Customer Lifetime Value = ([Total Sales]/[Number of customers])*[Average customer lifespan]
Visual: Card

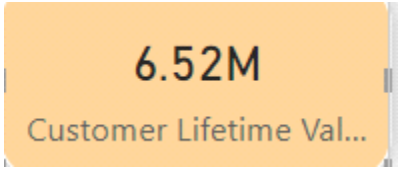


Fig 22: Card

KPI 5: Profit Margin
Bi Question: What is the Profit Margin for the superstore dataset?
DAX Formula: Profit Margin = [Total Profit] / [Total Sales]
Visual: Card

Fig 23: Card

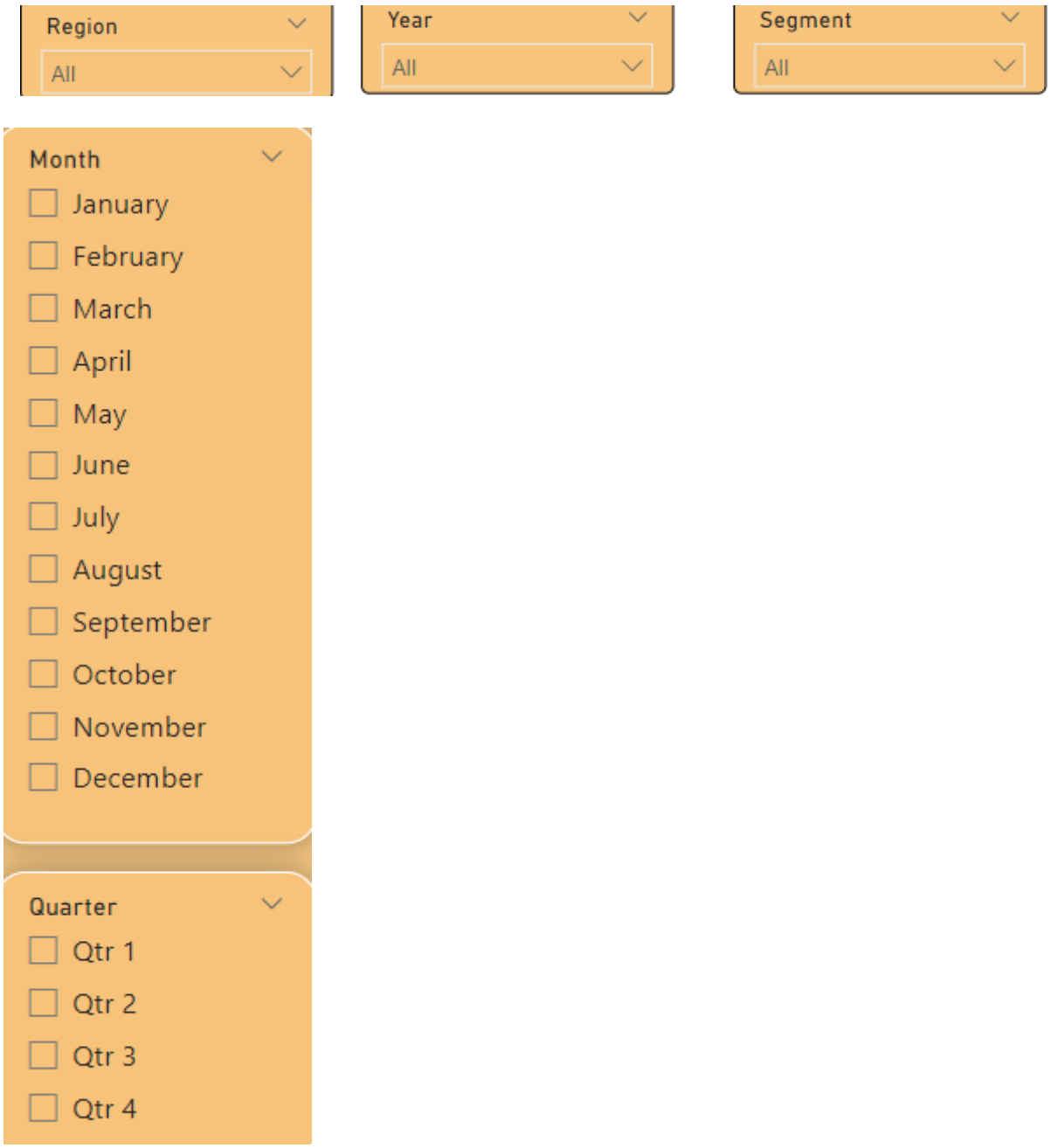


Fig 24: Slicers for Region, Year, and Segment.

SALES ANALYSIS DASHBOARD OVERVIEW

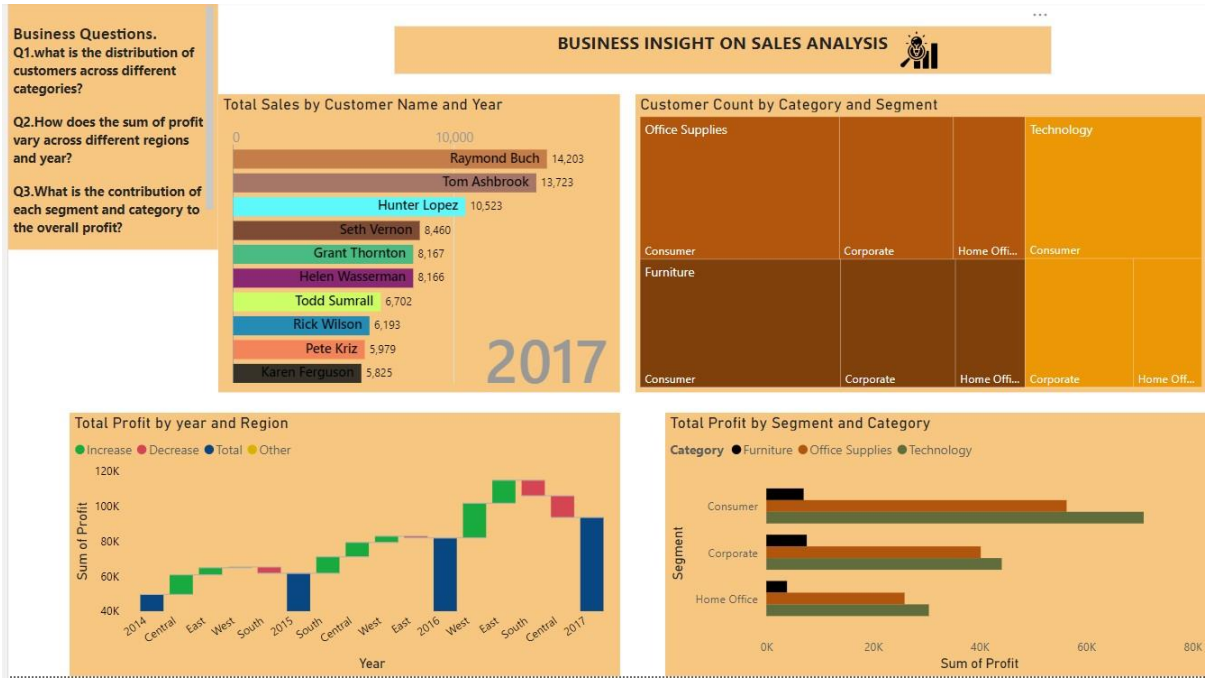


Fig 25: Sales Dashboard.

Sales Analysis Questions:

- How does the profit margin vary across different region and number of customers in each year?
- What is the contribution of each segment and category to the overall profit?
- Which states are the top 5 in terms of Sales?
- Which customer has the highest number of sales?

This page provides a general overview to the questions listed above and the answers to the questions asked.

- 2014
  Profit margin =10.23%
  Total Revenue =484.25k
  No of customer by the regions (central, East, West, North) =268

- 2015
  Profit margin =13.10%
  Total Revenue =470.35k
  No of customer by the regions (central, East, West, North) =275

- 2016
  Profit margin =13.43%
  Total Revenue =609.21k
  No of customer by the regions (central, East, West, North) =320

- 2017
  Profit margin =12.74%
  Total Revenue =733.22k
  No of customer by the regions (central, East, West, North) =350
  Hence, from the explanation given above we can see that the profit margin for 2016 was the highest, and 2017 has the highest total revenue generated. Also,2017 has the highest customer.

  Q2. For each segment we can see that Technology had the highest profit in consumer, corporate and home office.

  Q3. The highest top 5 in terms of sales are California, New York. Texas, Washington Pennsylvania.

  Q4. The customer with the highest number of sales is Raymond Buch made a total of 14,203.

Sales Analysis (Dashboard 1):

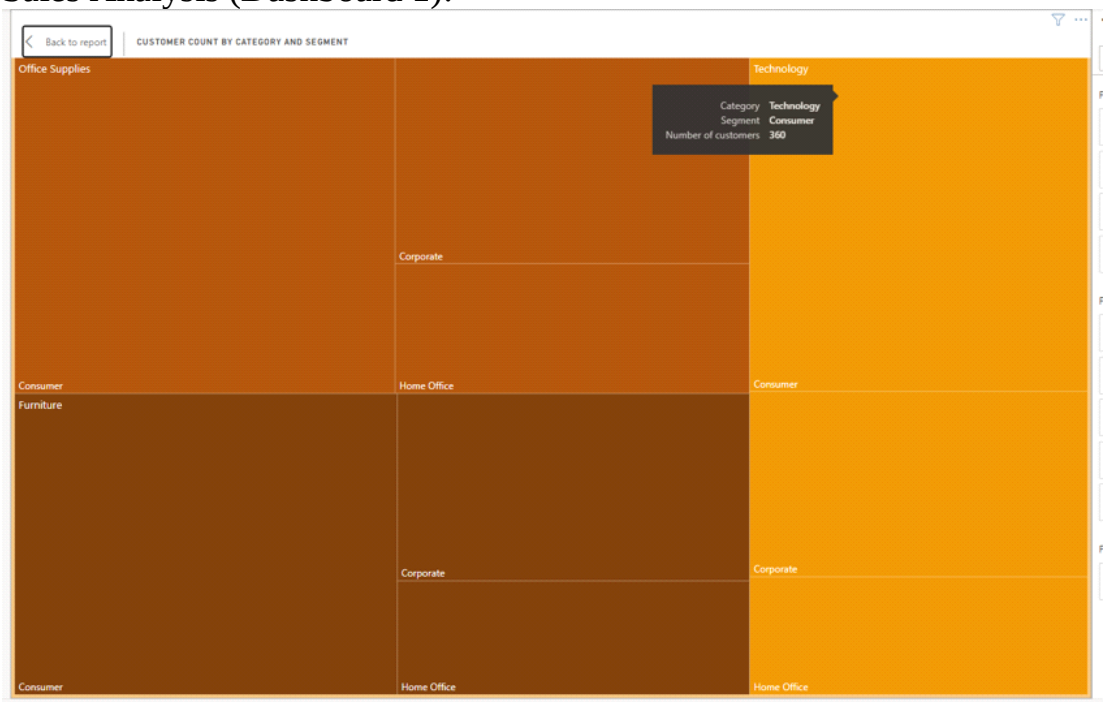


Fig 26: Sales Analysis Dashboard 1

Bi question: This chart explains which category and segment is doing well in terms of customer patronizing/wants on a particular product.
KPI: CUSTOMER COUNT BY CATEGORY AND SEGMENT
Field: Category, segment and number of customers
Visual: Tree map.

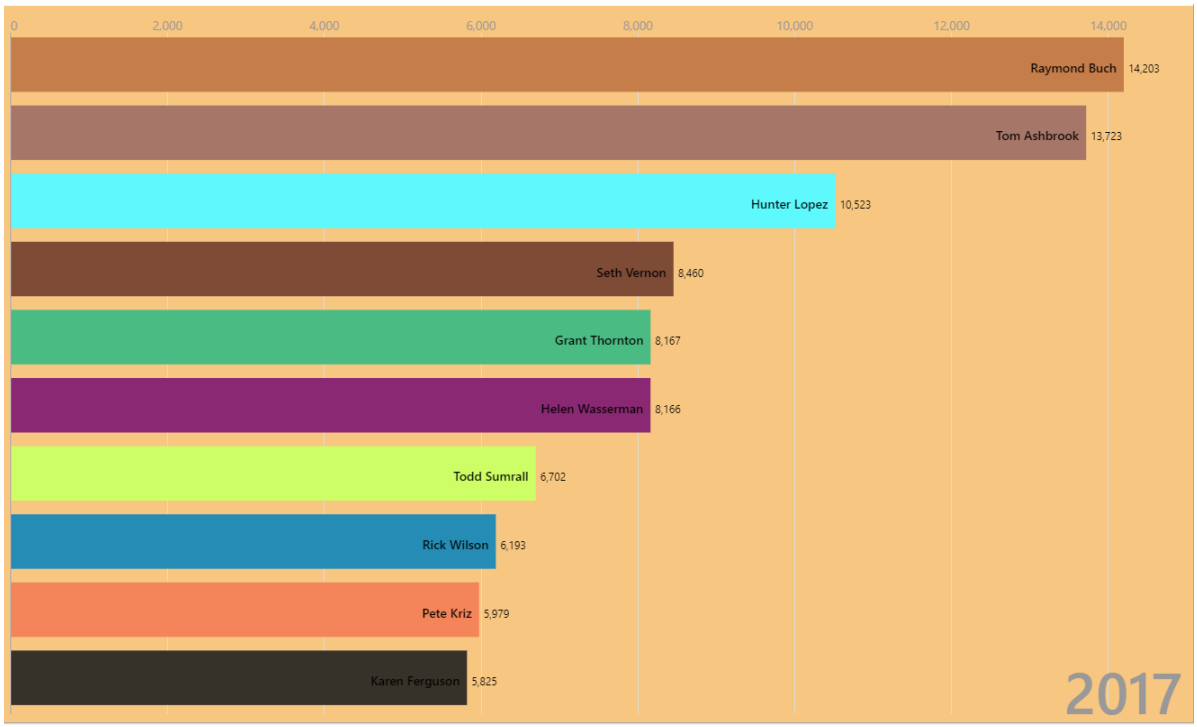


Fig 27: Total Sales by customer per year

Bi question: This chart explains the customer with the highest and lowest sales per year.

KPI: Customer sales per year

Field: Sales, Customer Names and year

Visual: Animated bar chart Race

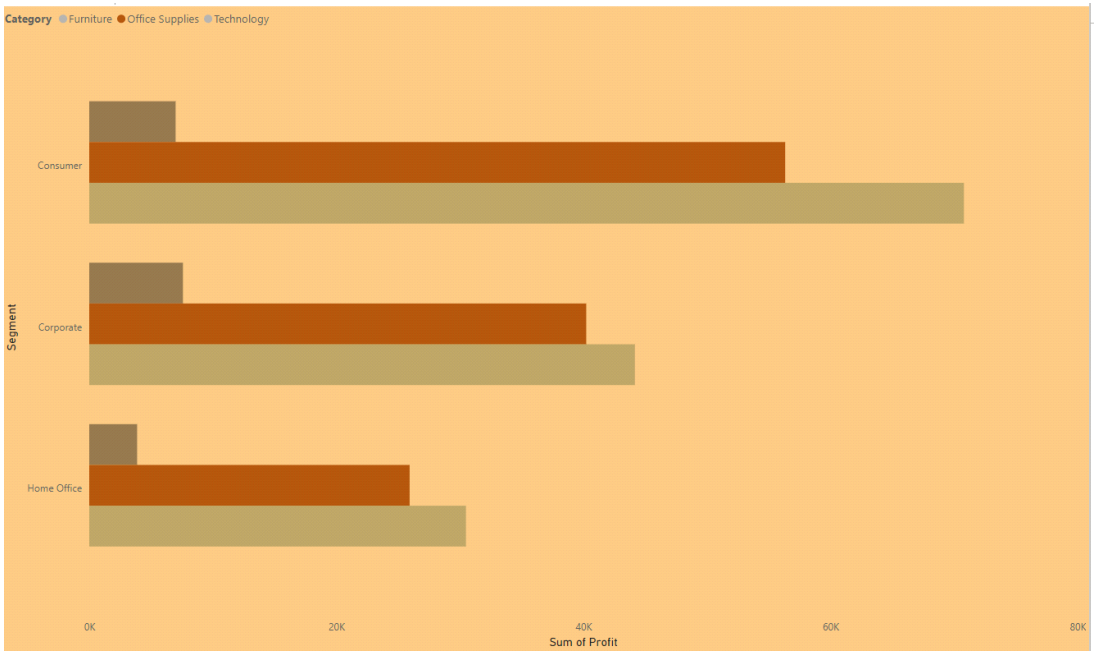


Fig 28: Total Profit by segment and category

Bi question: This chart explains which category and segment is doing well in terms of profit?

KPI: Total Profit by segment and category.

Field: Segment, category and sum of profit.

Visual: Clustered bar chart.

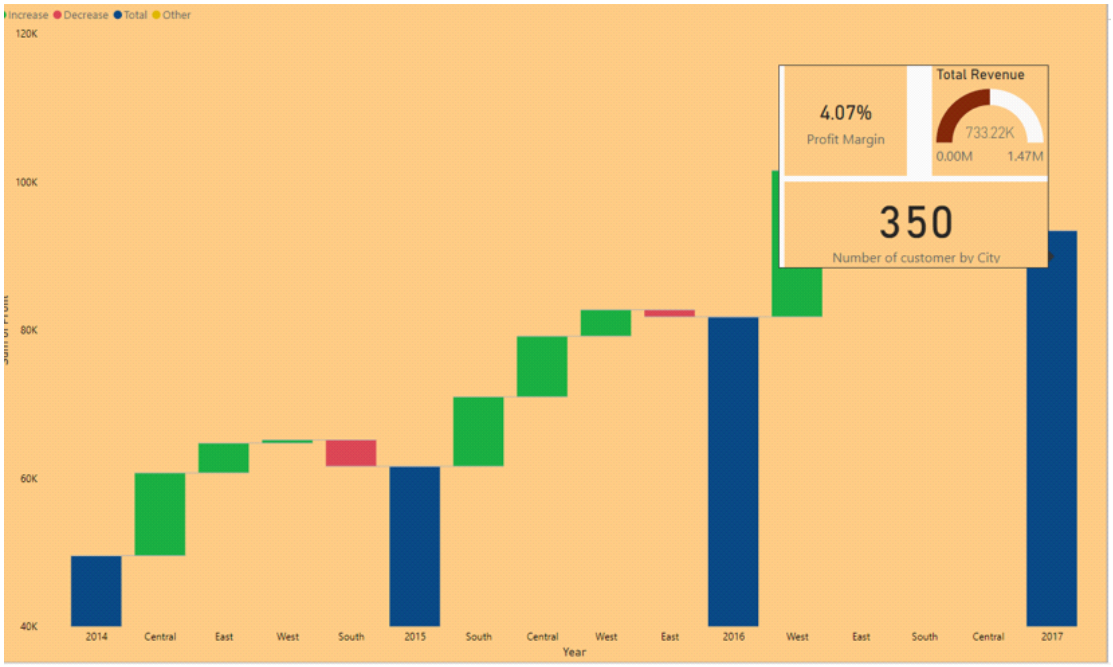


Fig 29: Total Profit by Year and Region

Bi question: This chart explains what year and region is doing well in terms of profit margin, total revenue, Sum of profit change and Number of customers by city?

KPI: Total Profit by Year and Region

Field: Profit, year and Region.

Visual: Waterfall Chart

PRODUCT ANALYSIS DASHBOARD OVERVIEW

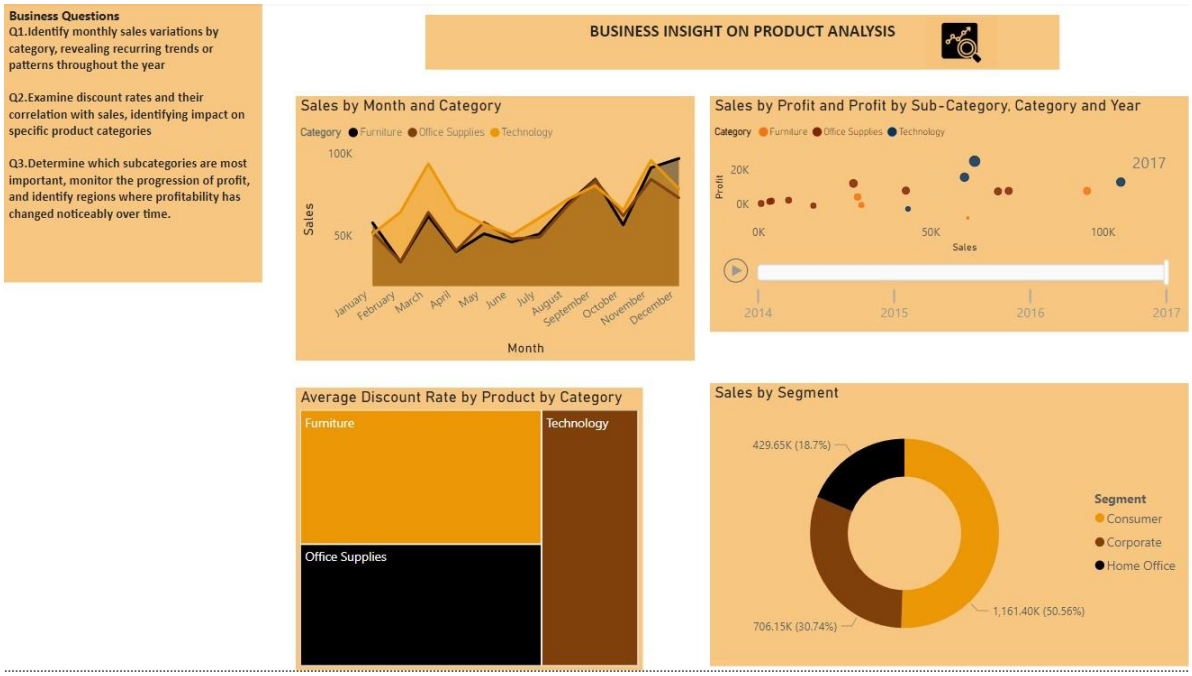


Fig 30: Product Analysis Dashboard.

Product Analysis Questions:

- Identify monthly sales variation by category revealing recurring trends or patterns throughout the year?
- Examine the discount rate and their correlation with sales, identifying the impact on specific product categories.
- Determine Which Subcategories are most important, monitor the progression of profit, and identify regions where profitability has changed noticeably over time.

This page provides a general overview to the questions listed above and the answers to the questions asked.

- November Performed well in terms of sales.
- Technology has the lowest discount and still performed well, my suggestion might be due to the high quality, brand loyalty or high market demand for the product Also, the place of effective marketing strategies cannot be override.
- Copier has improved really well and we can see that in the profit progression.

Product Analysis (Dashboard 2):

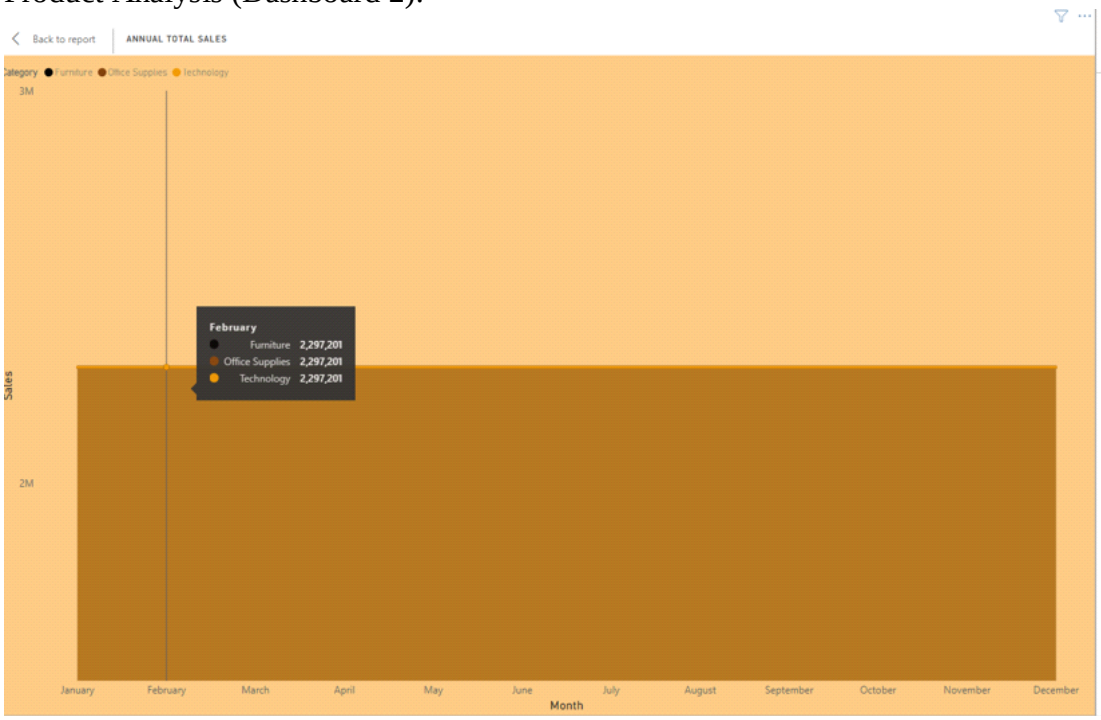


Fig 31: Annual Total Sales By month.

Bi Question: Annual Total Sales in term of Month specifying for each category.

KPI: Category, Month and Sales

Visual: Area Chart.

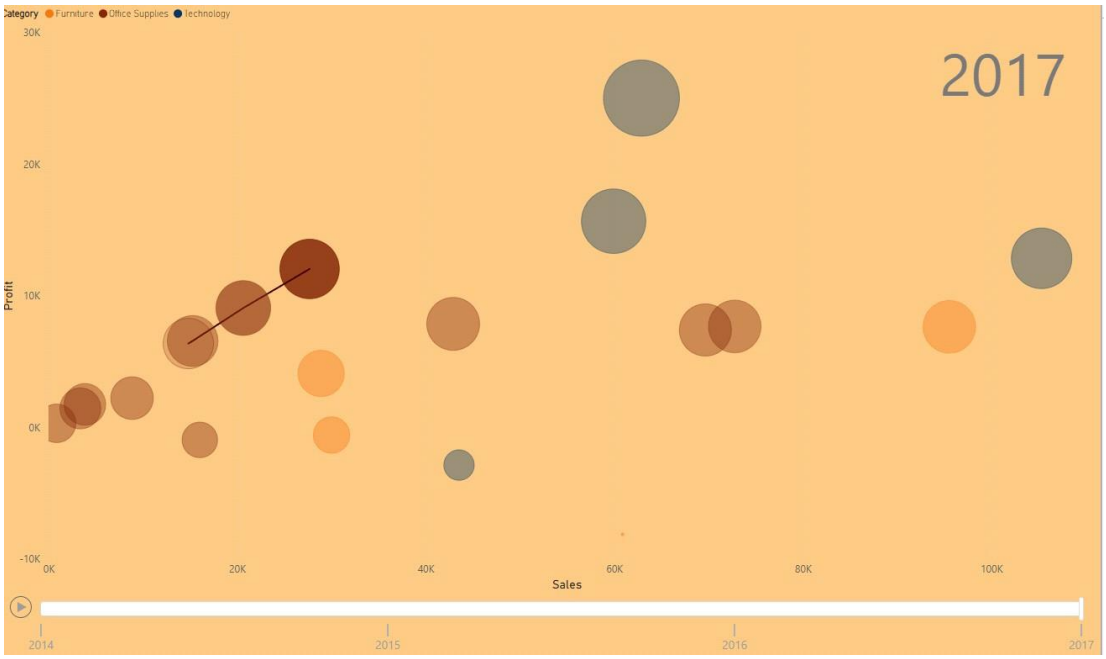


Fig 32: Sales by profit, sub-category, category and year

Bi Question: Use the scatter chart to display and compare the values in terms f category and the profit made per year.

KPI: Category, Profit and year

Visual: Scatter Chart.

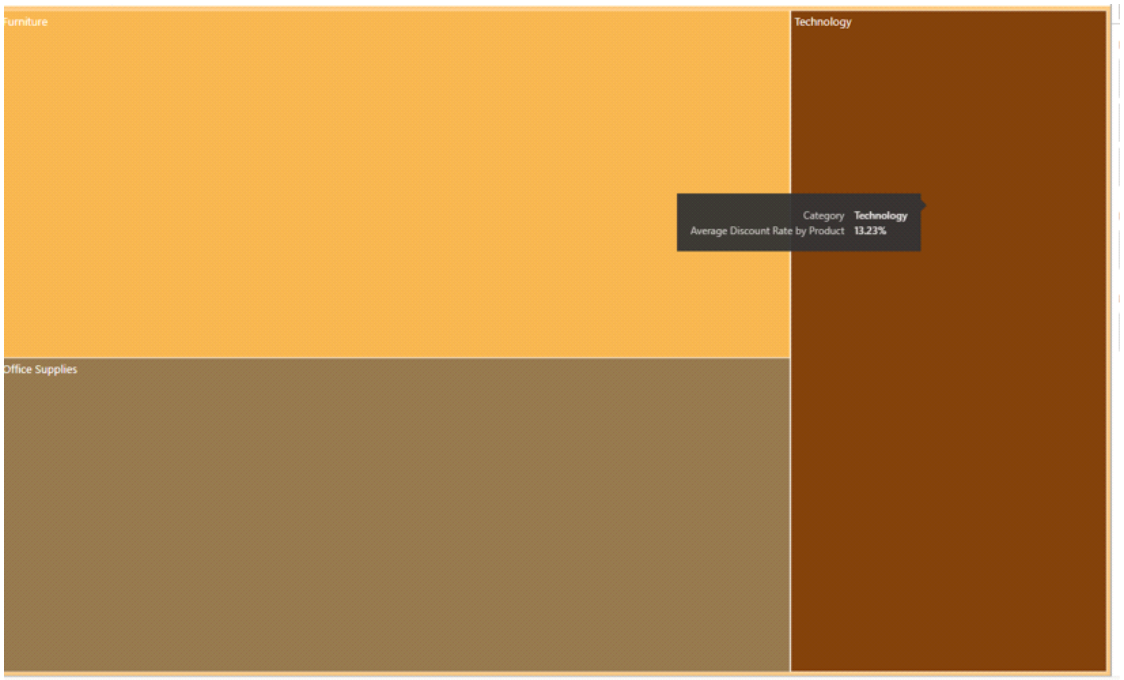


Fig 33: Average Discount Rate by Product

Bi Question: Use the Tree map to show the average discount on each product by category.

KPI: Average Discount Rate by Product and Category

DAX: Average Discount Rate by Product = AVERAGE ('Sample-Superstore csv - Clean'[Discount])
Visual: Tree Map.

CUSTOMER ANALYSIS DASHBOARD OVERVIEW

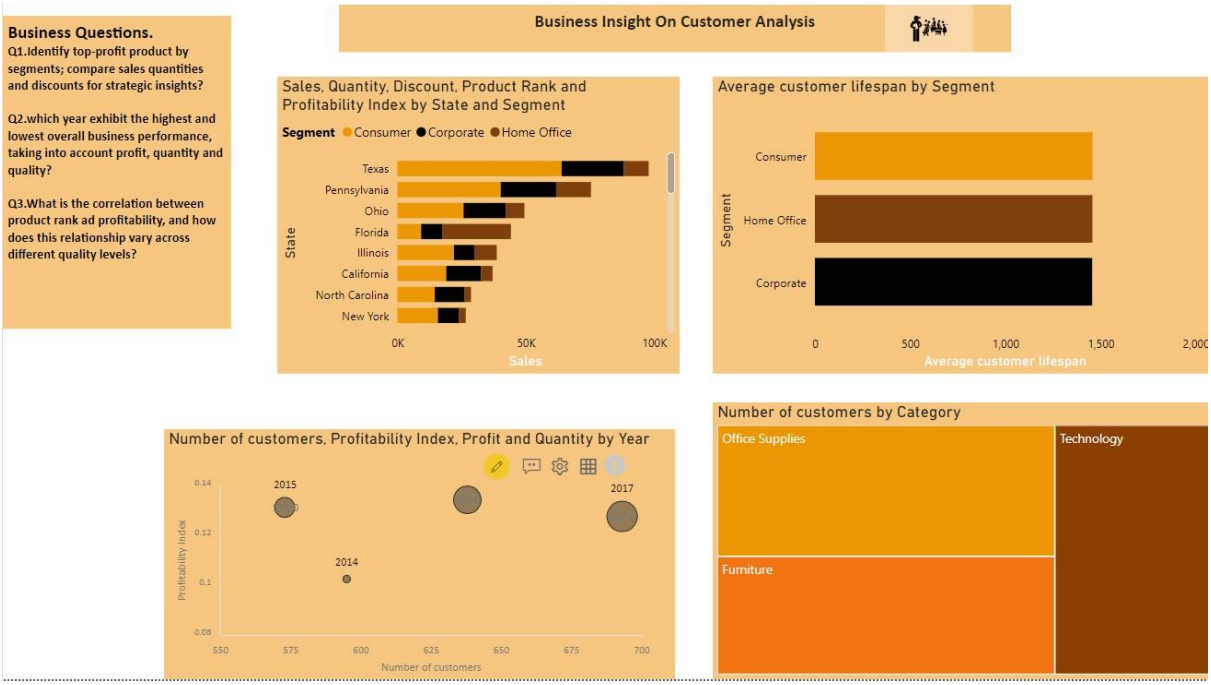


Fig 34: Customer Analysis Dashboard

- Identify the top-profit product by segments; compare sales quantities and discounts for strategic insights.
- Which year exhibits the highest and lowest overall business performance, considering profit, quantity and quality.
- What is the number of customers per category?

Texas
2017 was the highest and 2014 was the lowest.

Customer Analysis (Dashboard 3):

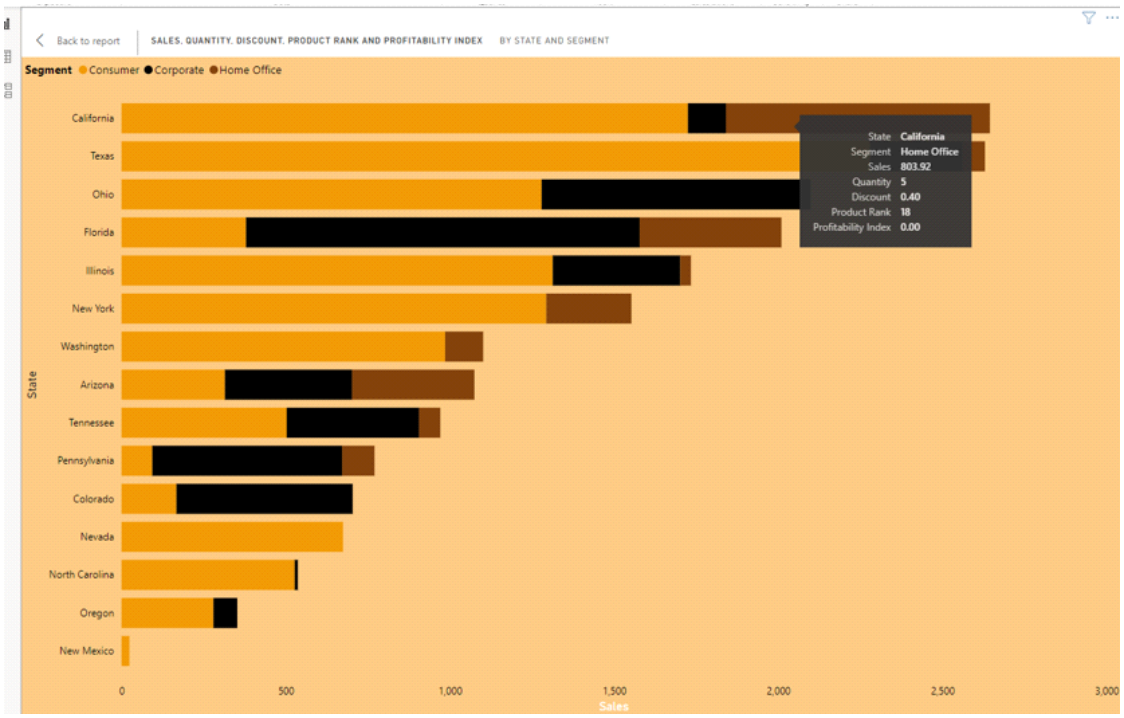


Fig 35: Sales, Quantity, Discount, Product Rank and Profitability Index by State and Segment

Bi Question: Use the Stacked bar chart to show the state, segment, sales, quantity, discount, product rank, and profitability index by state and segment.

KPI: Sales, Quantity, Discount, Product Rank and Profitability Index by State and Segment

DAX: Average Discount Rate by Product = AVERAGE ('Sample-Superstore csv - Clean'[Discount])

Product Rank = RANKX(ALL('Product'[Sub-Category]), [Profitability Index], , DESC)

Profitability Index = [Total Profit] / [Total Sales]

Visual: Stacked bar chart.

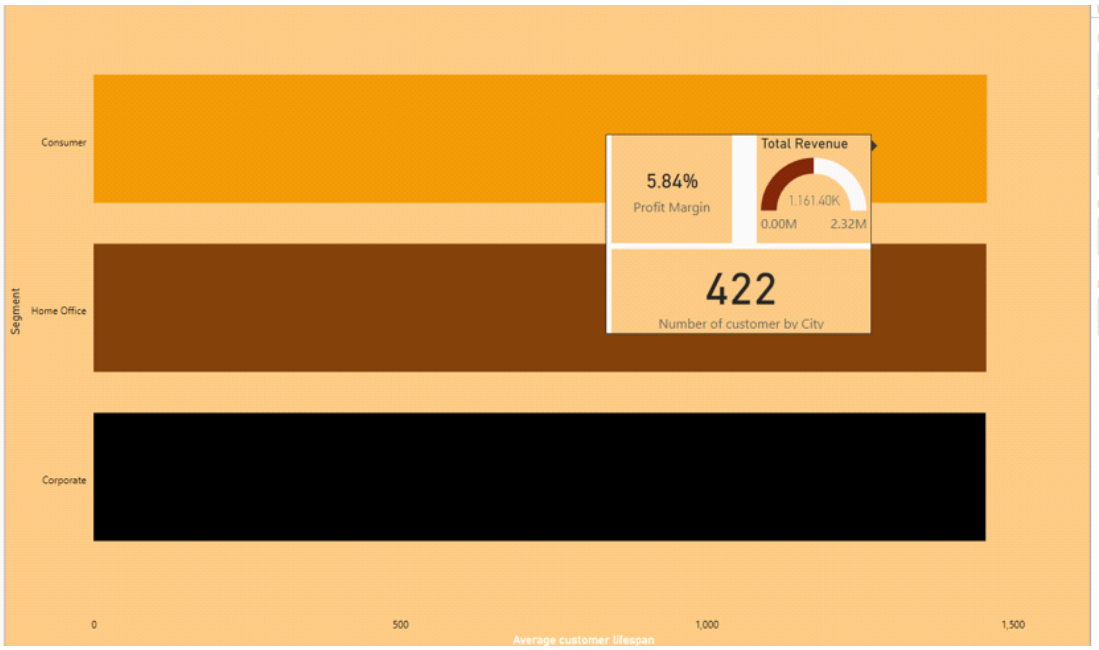


Fig 36 : Average Customer Lifespan

Bi Question: Find the average customer lifetime by segment.

KPI: Find the average customer lifetime and segment.

DAX: Average customer lifespan = AVERAGEX (VALUES ('Sample-Superstore csv - Clean'[Customer ID]), DATEDIFF(MIN('Sample-Superstore csv - Clean'[Order Date]),MAX('Sample-Superstore csv - Clean'[Order Date]),DAY))

Visual: Stacked bar chart.

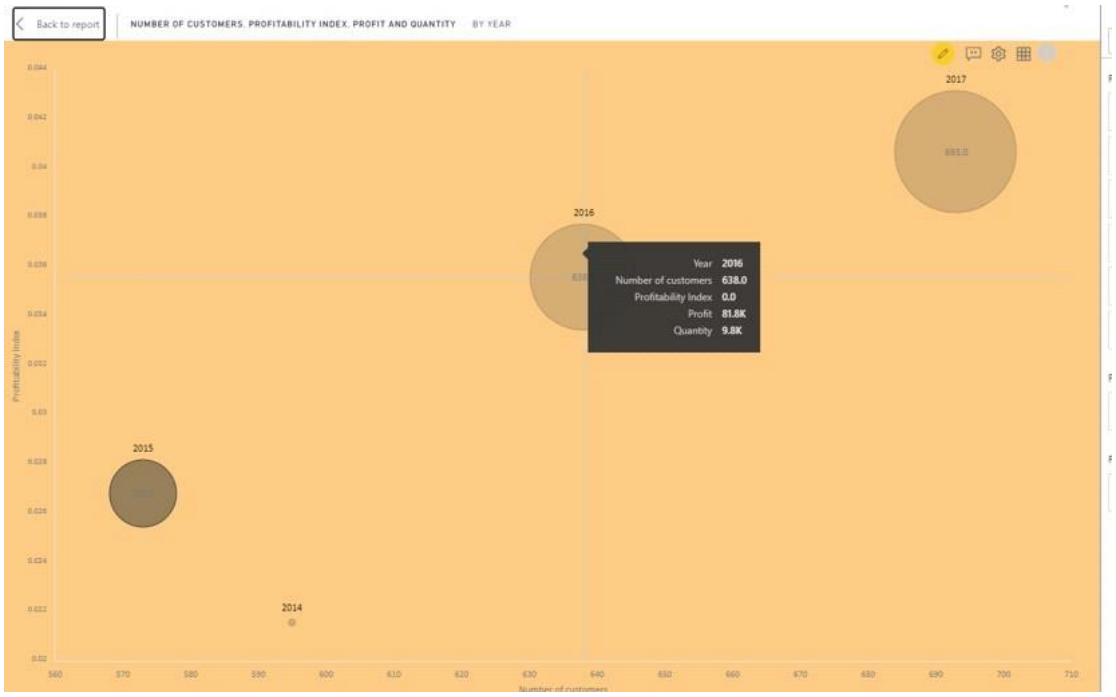


Fig 37: Number of customers, Profitability Index, Profit and Quantity by Year

Bi Question: Find Number of customers, Profitability Index, Profit and Quantity by Year
KPI: Number of customers, Profitability index, profit, quantity and year
DAX: Number of customers = COUNTROWS (VALUES ('Sample-Superstore csv - Clean'[Customer ID]))
DAX 2: Profitability Index = [Total Profit] / [Total Sales]
Visual: Bubble/stacked chart viz 3.0.1

GEOGRAPHICAL ANALYSIS DASHBOARD OVERVIEW

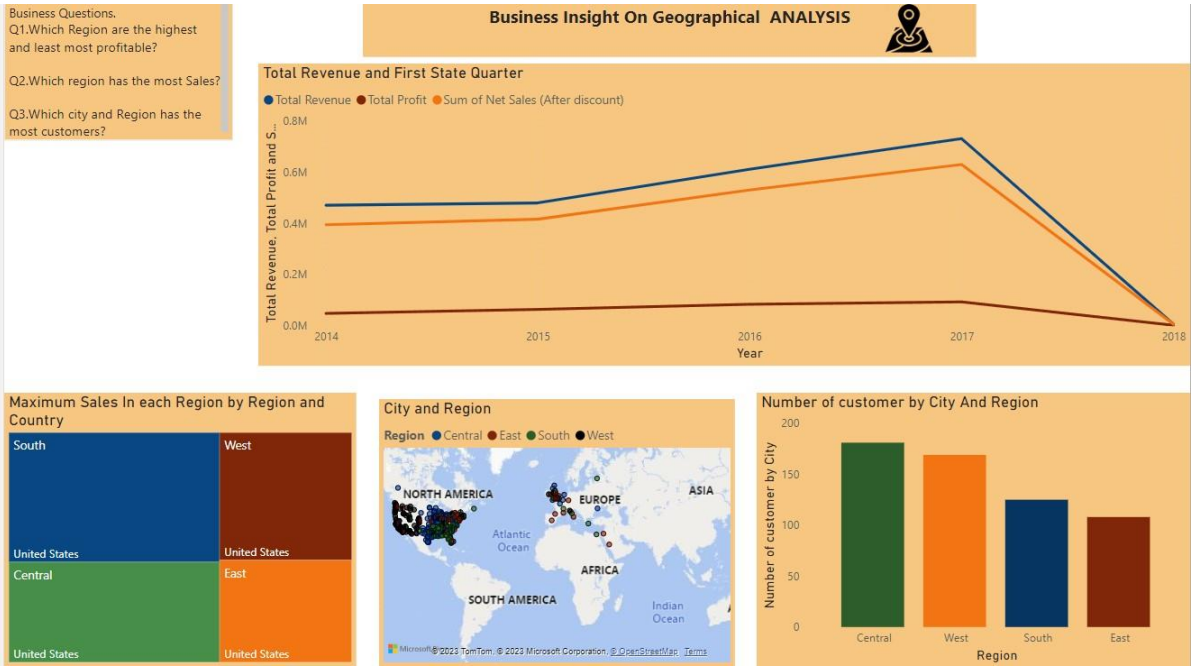


Fig 38: Geographical Analysis Dashboard

Geographical Analysis Questions:

- Which Region are the highest and least more profitable?
- Which Region has the most Sales?
- Which City and Region has the most customers

South had the highest and the East has the lowest.
The East
Central had the most customer.

Geographical Analysis (Dashboard 4):

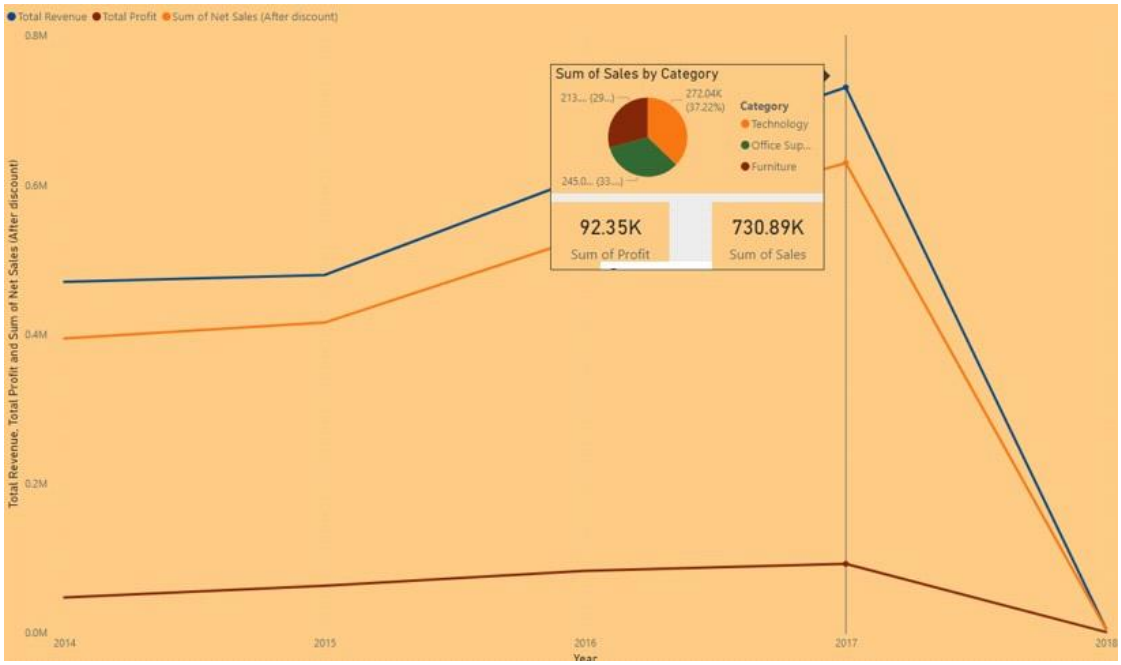


Fig 39: Total Revenue in first quarter of the year

Bi Question: Find the total Revenue in the first quarter of the year by profit and sales.

KPI: Total Revenue

DAX: Total Revenue = SUM (('Sample-Superstore csv - Clean'[Sales]))

Visual: Line Chart.

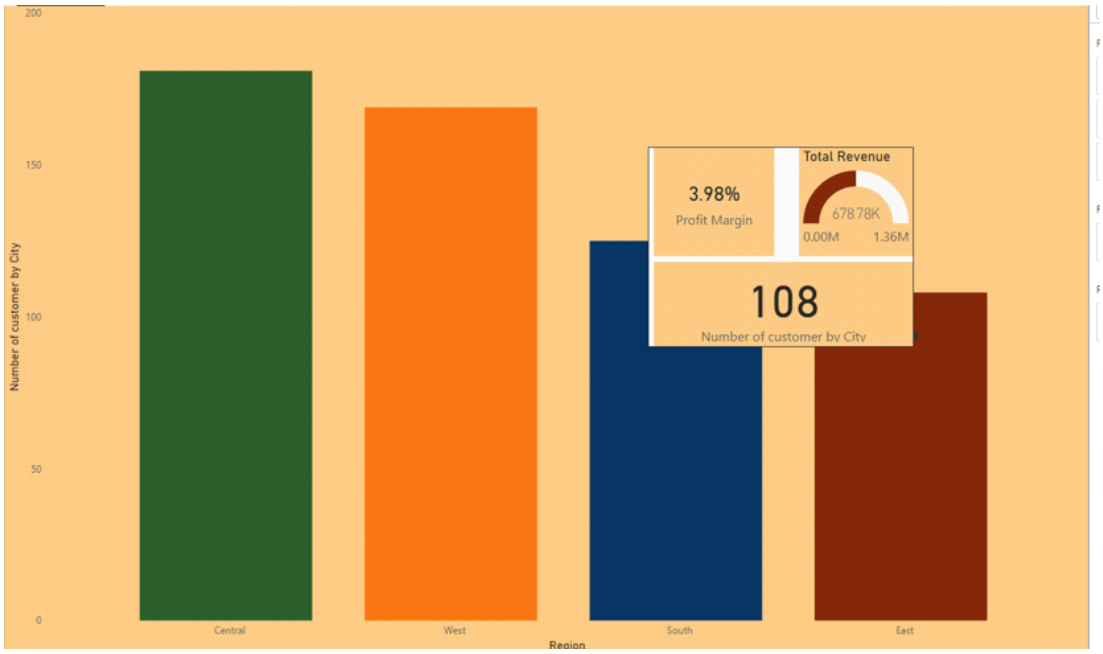


Fig 40: Number of customers by region

Bi Question: Find Total customer by region using profit margin, total revenue and number of customers by city.

KPI: Number of customers by city

DAX: Number of customer by City = COUNTROWS (VALUES ('Sample-Superstore csv - Clean'[City]))

Visual:100% stacked bar chart.

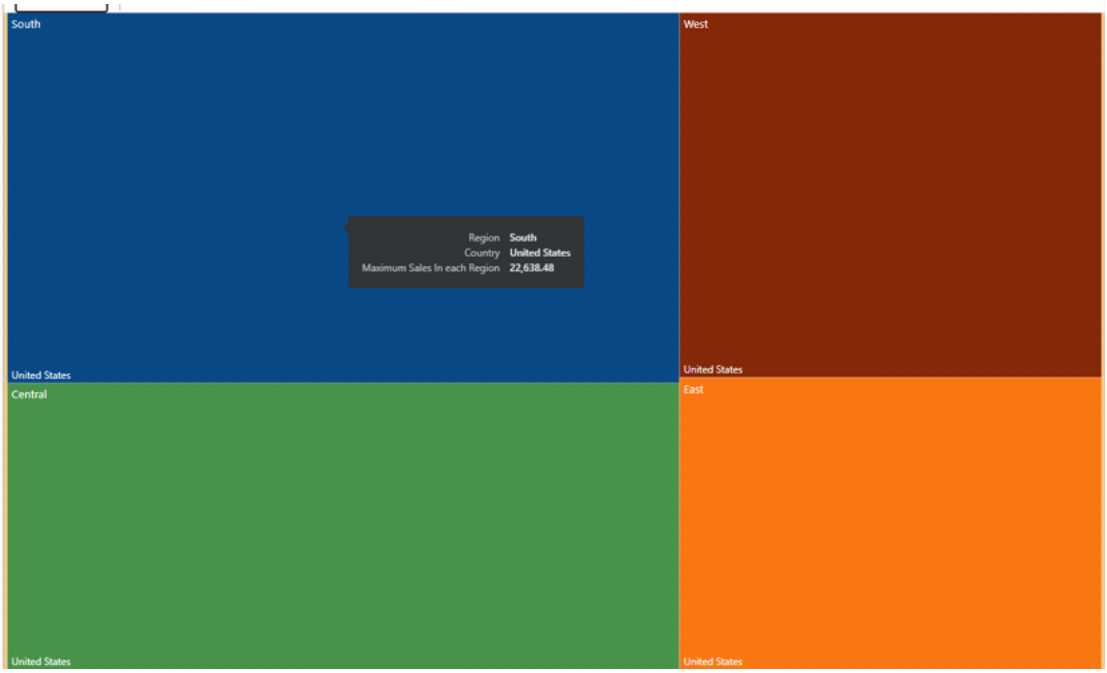


Fig 41: Maximum Sales In each Region by Country

Bi Question: Find the maximum sales in each Region.

KPI: Maximum sales in each region

DAX: Maximum Sales In each Region = MAXX (VALUES ('Sample-Superstore csv - Clean'[Region]), CALCULATE (MAX ('Sample-Superstore csv - Clean'[Sales])))
Visual: Tree map

MARKETING ANALYSIS DASHBOARD OVERVIEW

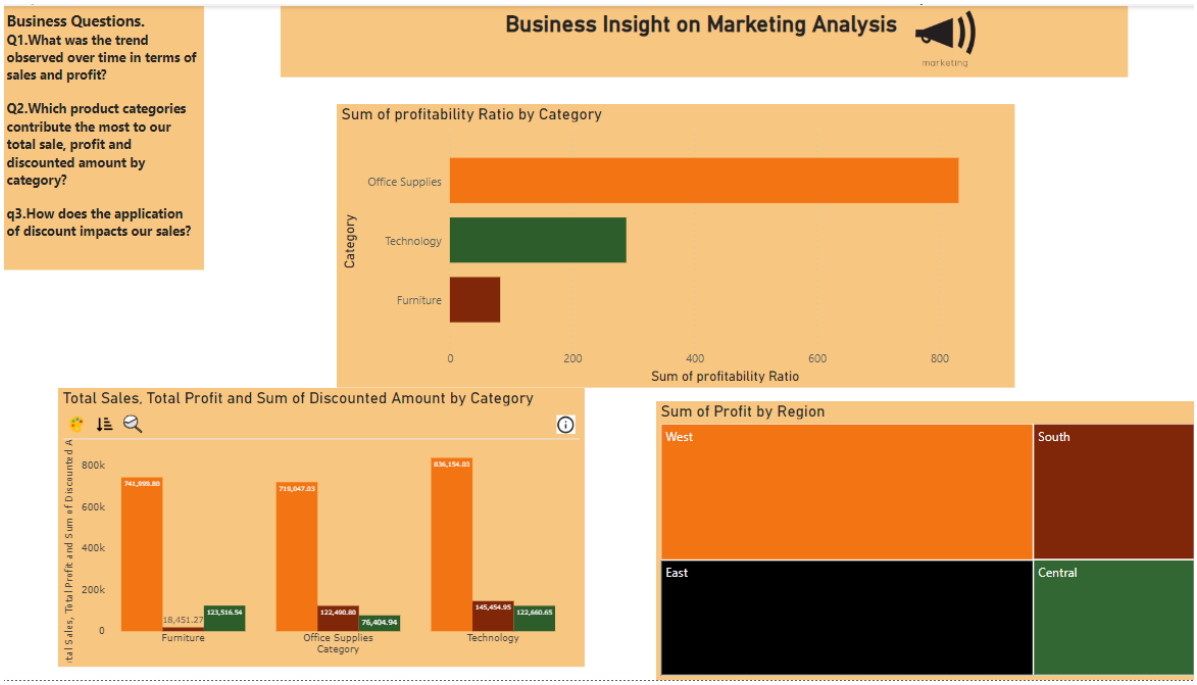


Fig 42: Marketing Analysis Dashboard.

Marketing Analysis Questions:

- What was the trend observed over time in terms of sales and profit?
- Which product categories contributes the most to our total sales, profit and discounted amount by category?
- How does the Application of discount impact our sales?

- Technology had the highest sales and highest profit.
- Technology
- In this dataset not much, because technology had the lowest discount but still sold the most.

Marketing Analysis (Dashboard 4):

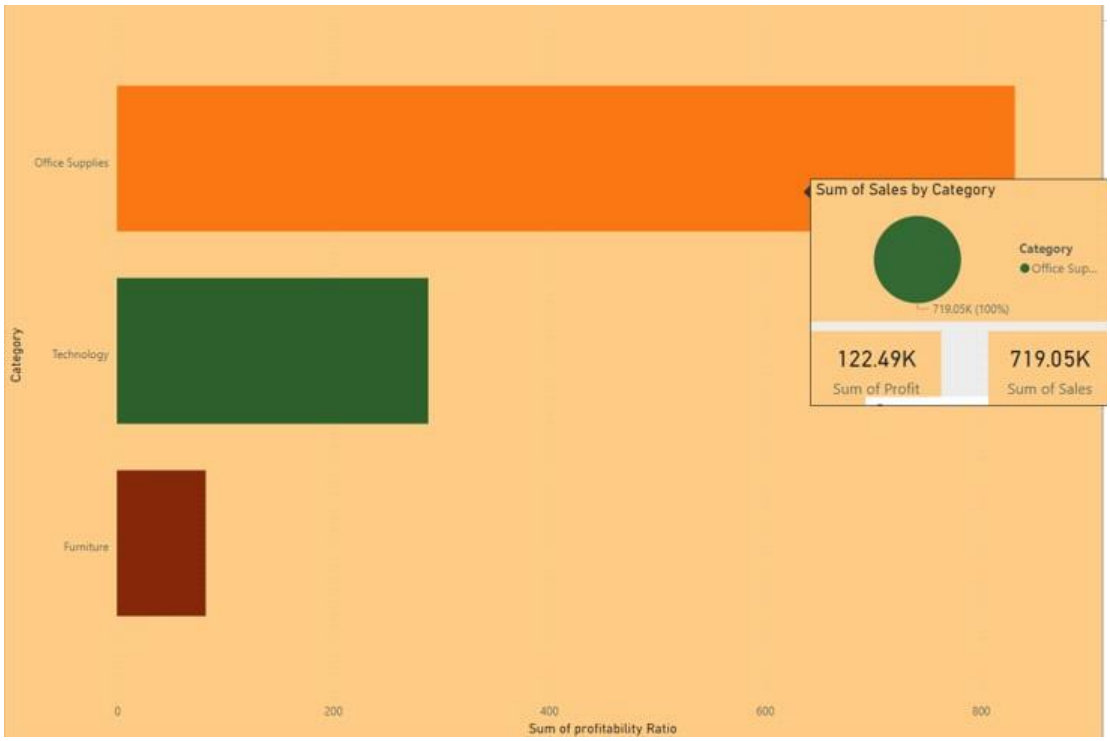


Fig 43: Maximum Sales In each Region by Country

Bi Question: Find the maximum sales in each Region.

KPI: Maximum sales in each region

Visual: Tree map

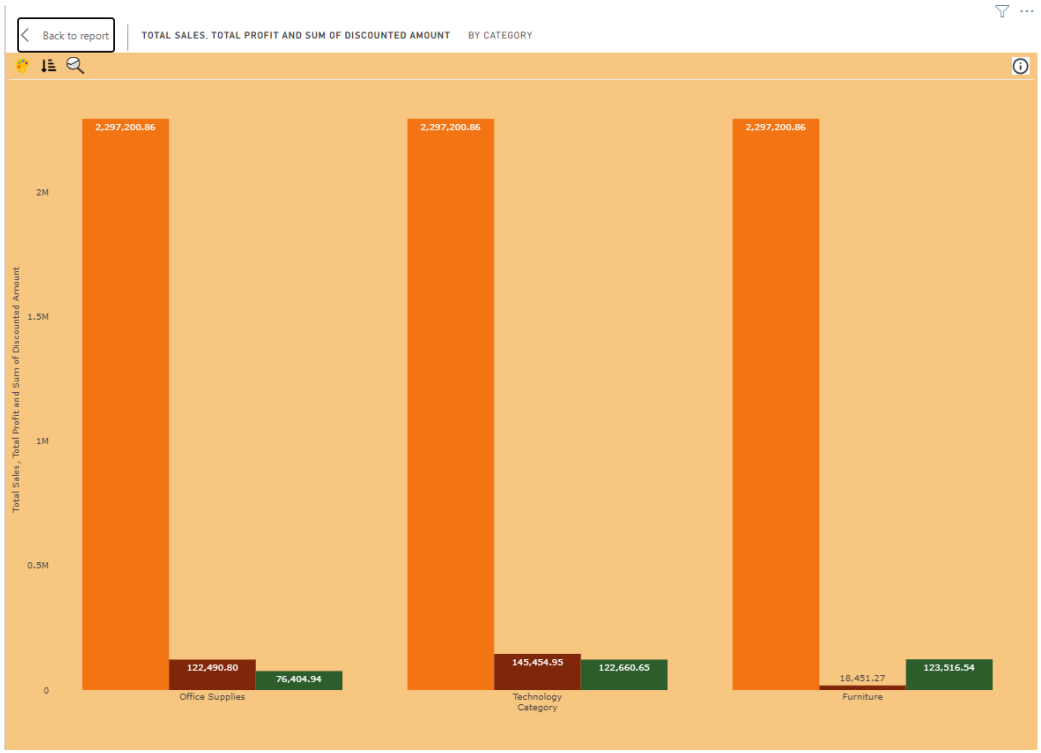


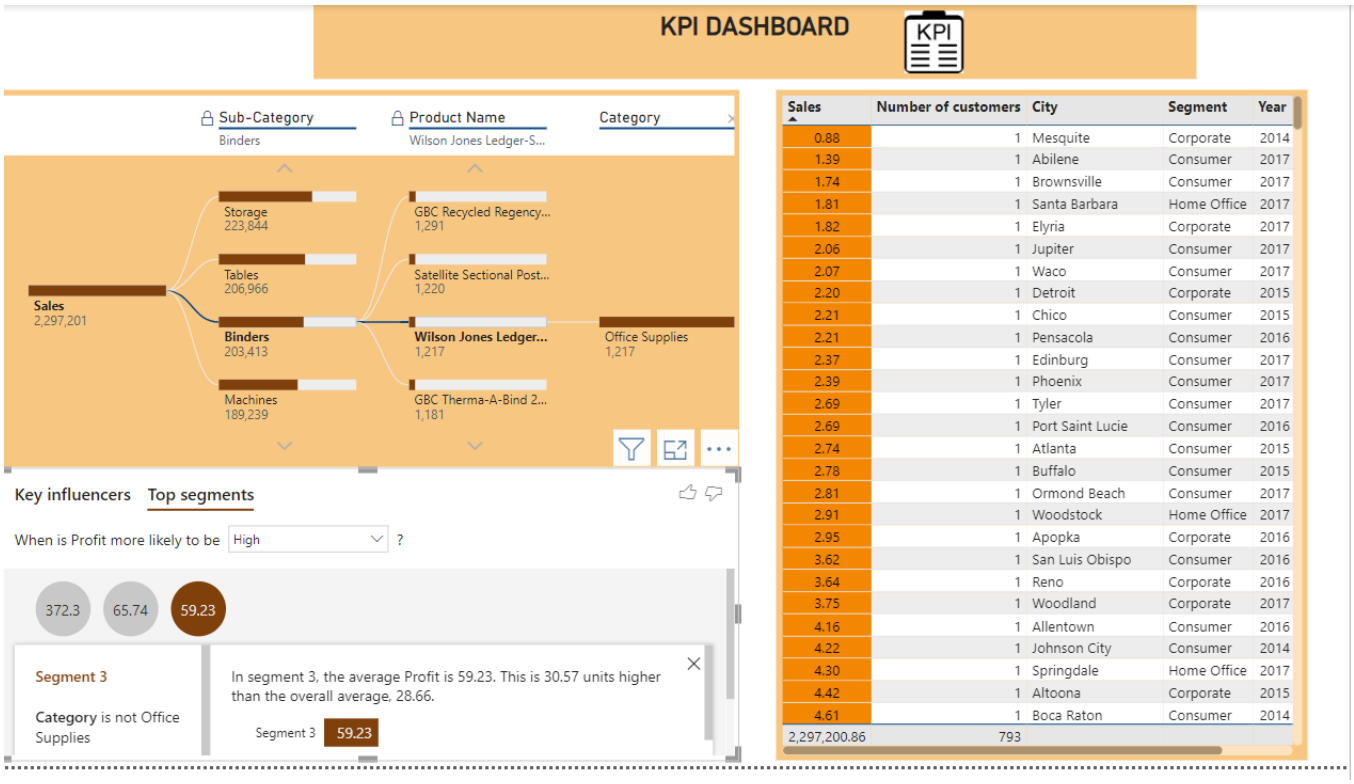
Fig 44:

Bi Question: Find the Total Sales, Total Profit and Sum of Discounted Amount by Category

KPI: Total sales, profit by category

Visual: Dual y axis combo chart.

KPI ANALYSIS DASHBOARD OVERVIEW



Sales	Number of customers	City	Segment	Year
0.88	1	Mesquite	Corporate	2014
1.39	1	Abilene	Consumer	2017
1.74	1	Brownsville	Consumer	2017
1.81	1	Santa Barbara	Home Office	2017
1.82	1	Elyria	Corporate	2017
2.06	1	Jupiter	Consumer	2017
2.07	1	Waco	Consumer	2017
2.20	1	Detroit	Corporate	2015
2.21	1	Chico	Consumer	2015
2.21	1	Pensacola	Consumer	2016
2.37	1	Edinburg	Consumer	2017
2.39	1	Phoenix	Consumer	2017
2.69	1	Tyler	Consumer	2017
2.69	1	Port Saint Lucie	Consumer	2016
2.74	1	Atlanta	Consumer	2015
2.78	1	Buffalo	Consumer	2015
2.81	1	Ormond Beach	Consumer	2017
2.91	1	Woodstock	Home Office	2017
2.95	1	Apopka	Corporate	2016
3.62	1	San Luis Obispo	Consumer	2016
3.64	1	Reno	Corporate	2016
3.75	1	Woodland	Corporate	2017
4.16	1	Allentown	Consumer	2016
4.22	1	Johnson City	Consumer	2014
4.30	1	Springdale	Home Office	2017
4.42	1	Altoona	Corporate	2015
4.61	1	Boca Raton	Consumer	2014
2,297,200.86	793			

Fig 45: KPI ANALYSIS (Dashboard 5)

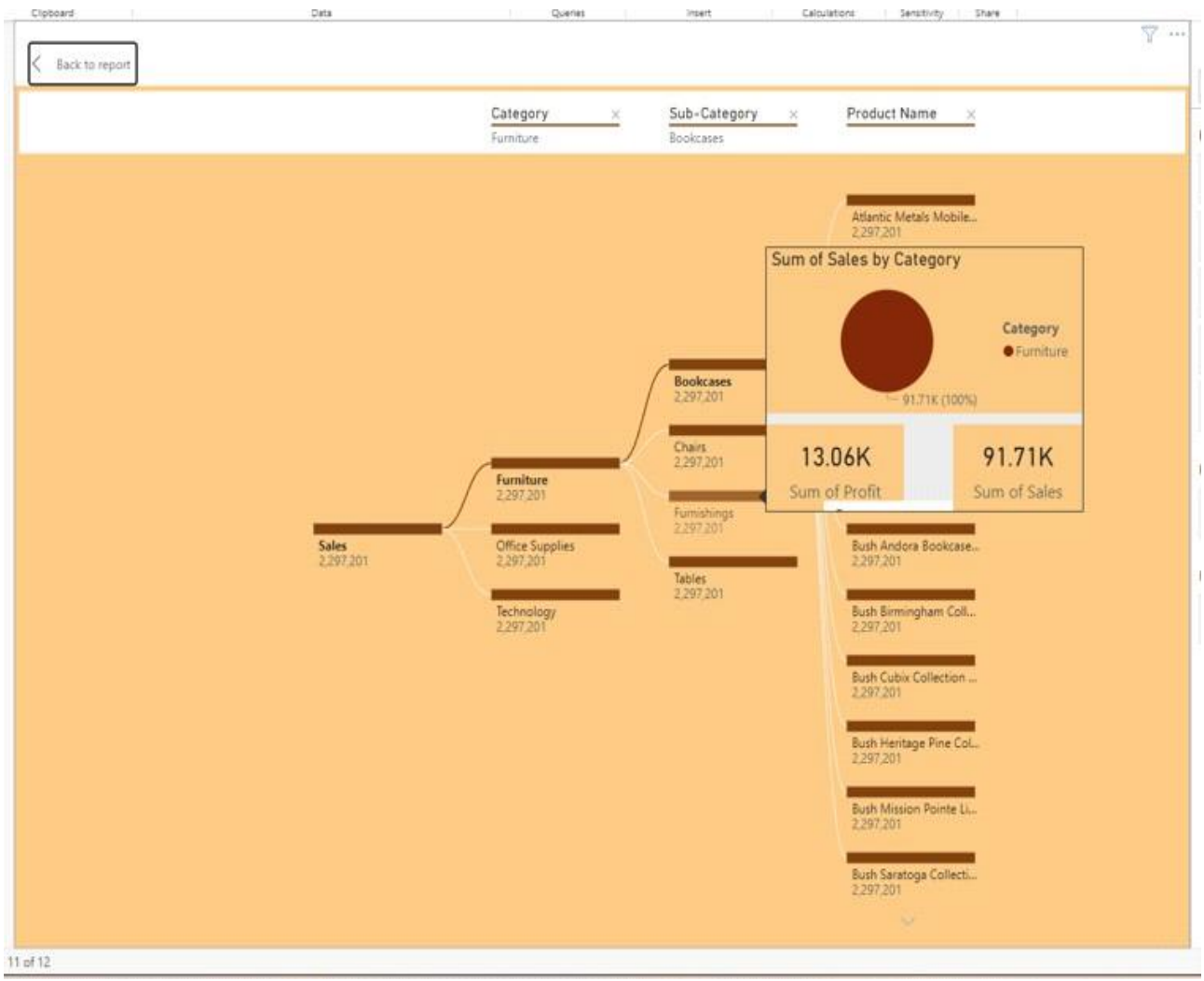


Fig 46: Top sold category/sub-category, product Name Using Decomposition Tree visual.

Bi Question: Top 10 product Name by profit and sales

KPI: Profit and Sales

Created a tooltip that when hover on each product name displays the sum of sales, profit and category.

Sales	Number of customers	City	Segment	Year	Month
23,752.28	2	Jacksonville	Home Office	2014	March
18,336.74	1	Lafayette	Corporate	2016	February
14,593.15	9	New York City	Consumer	2015	September
14,142.18	3	Seattle	Consumer	2017	March
13,741.01	2	New York City	Home Office	2017	October
11,135.52	4	New York City	Consumer	2014	December
10,738.17	3	Newark	Consumer	2017	November
10,539.90	1	San Antonio	Consumer	2014	August
10,461.18	4	Philadelphia	Consumer	2016	May
10,329.81	8	New York City	Consumer	2014	September
9,929.07	5	San Francisco	Consumer	2014	July
9,900.19	1	Minneapolis	Consumer	2014	September
9,892.74	1	Detroit	Consumer	2016	December
9,421.92	6	New York City	Consumer	2017	September
9,135.19	1	Lakewood	Corporate	2016	April
8,812.74	2	Arlington	Consumer	2016	February
8,167.42	1	Burlington	Corporate	2017	April
8,086.99	4	Seattle	Corporate	2017	August
8,086.72	7	New York City	Corporate	2017	November
8,000.95	2	Seattle	Corporate	2015	March
7,558.26	2	Los Angeles	Home Office	2016	October
6,726.27	6	Seattle	Consumer	2017	November
6,542.82	4	Los Angeles	Consumer	2015	January
6,412.77	1	Atlanta	Consumer	2015	March
6,411.24	6	New York City	Consumer	2015	July
6,358.08	3	San Francisco	Corporate	2016	March
6,271.99	6	Los Angeles	Consumer	2016	November
6,258.73	7	Philadelphia	Consumer	2017	September
6,240.50	9	Los Angeles	Consumer	2015	December
6,210.99	3	New York City	Corporate	2015	August
6,160.62	1	Buffalo	Consumer	2014	November
6,125.82	1	Yonkers	Consumer	2016	January
5,865.54	6	Seattle	Home Office	2017	March
5,843.66	2	New York City	Corporate	2017	May
5,820.30	8	New York City	Consumer	2016	December
5,802.70	1	Jackson	Consumer	2017	January
5,579.94	1	Houston	Corporate	2014	September
2,297,200.86	793				

Fig 47: A table showing the sales, number of customers, city, segment, year and month in a table.

Bi Question: Show the following information in a tabular format.

KPI: table showing the sales, number of customers, city, segment, year and month in a table.

Visual: Table

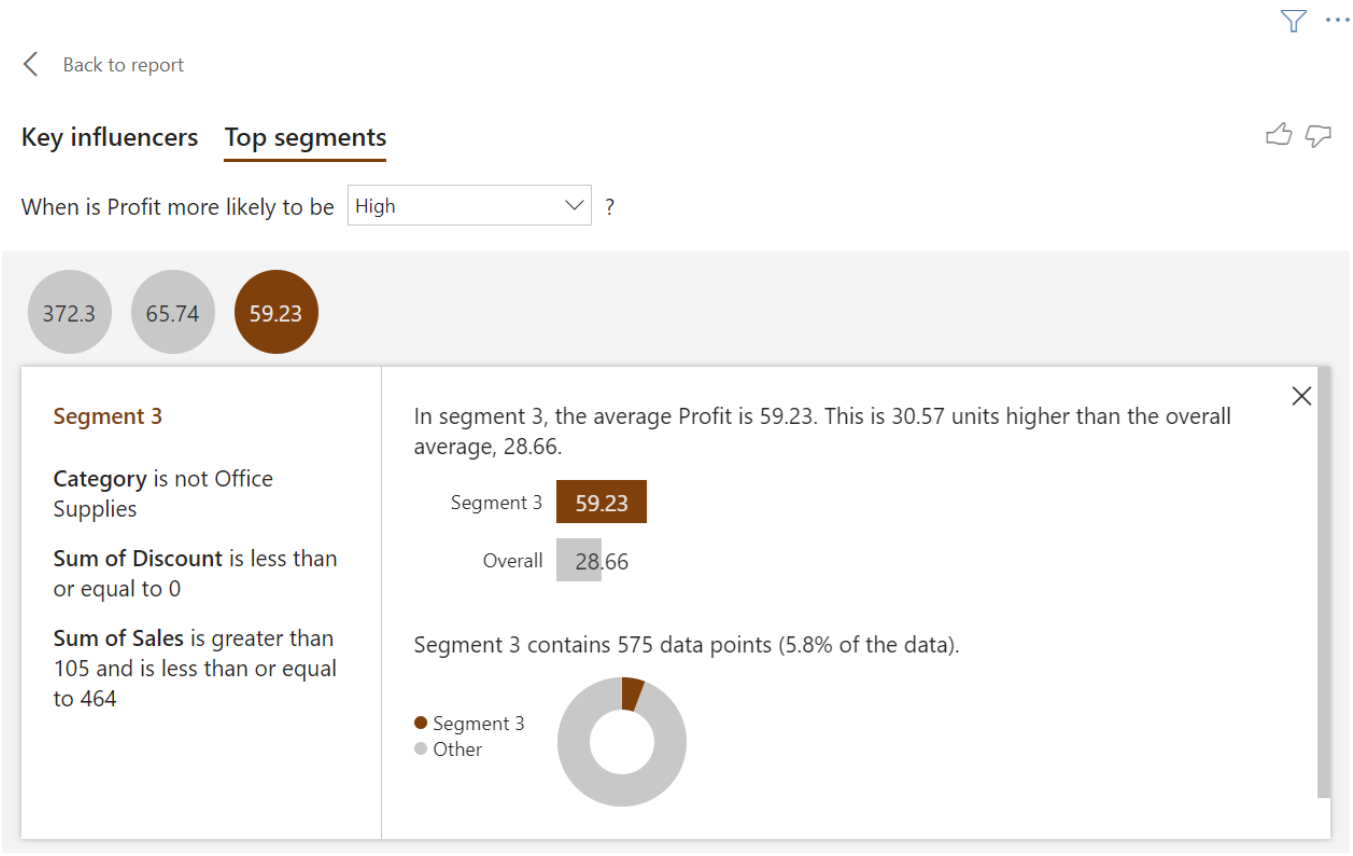


Fig 48: Key influencer are used to identify the most significant drivers behind trends or outcomes in this scenario.

Bi Question: Show the following information using a key influencer.

KPI: table showing the sales, number of customers, city, segment, year and month in a table.

Visual: Key Influencer.

FORECAST ANALYSIS DASHBOARD OVERVIEW

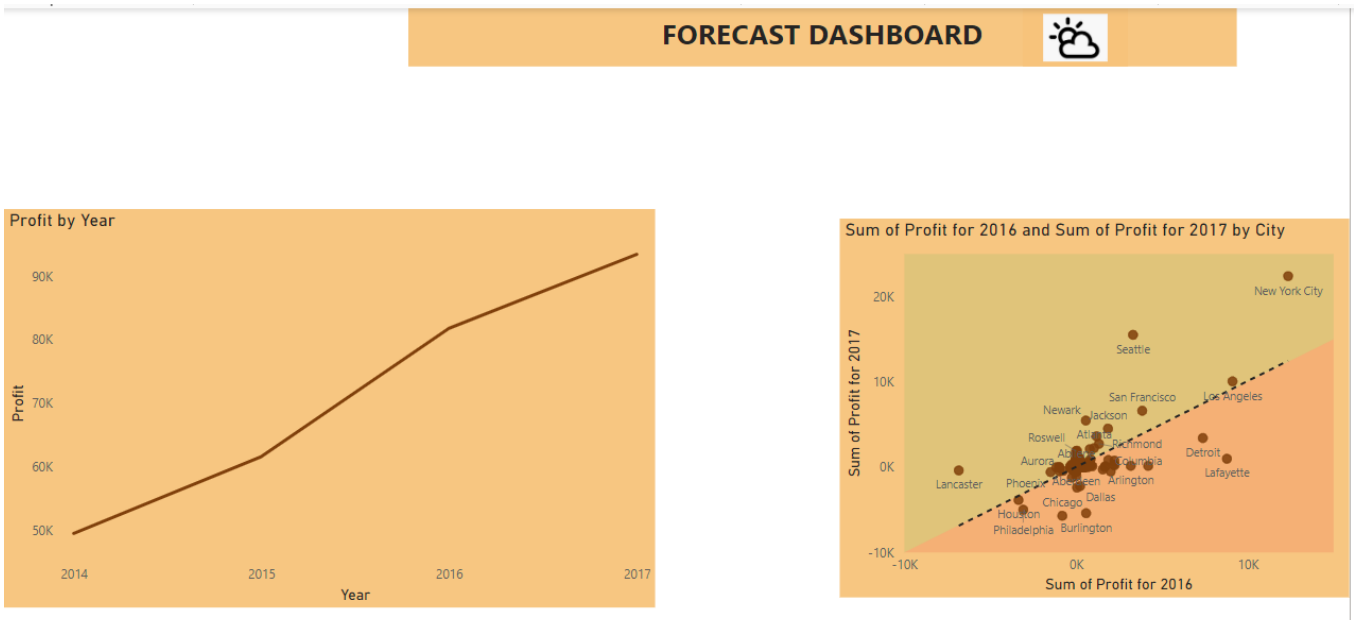


Fig 49: The line chart was implored to forecast how sales was used to predict the growth and actions that need to be taken.
Bi Question: Use the line chart to forecast for hoe each product will do.
KPI: Line chart showing profit by year in each city.
Visual: Line Chart.

6.Technical Details

Data source was gotten from the Kaggle Repository:
https://www.kaggle.com/datasets/vivek468/superstore-dataset-final
Power BI desktop: Version 2.123.742.0 64-BIT (November 2023)

Task 3 – Stakeholder Report

7.Executive Summary

The primary objective of this research is to deliver a thorough analysis of the superstore industry, with particular emphasis on important indicators related to products, Customers, Sales, Profit and Purchases. Finding recurring trends and patterns that will provide guiding insights for the business overall improvement is the goal. The report aims to identify strategic opportunities, operational efficiencies, and areas for growth to support continuous growth and success in the superstore's overall performance through a thorough analysis of these important factors.

7.1. Relevance of Power BI in sales Data

In the global marketplace of sales data, Power BI serves as vital for an array of reasons. First, the programme provides strong data visualisation features that let salespeople transform intricate datasets into aesthetically pleasing graphs and charts. This facilitates making informed choices by facilitating in the quick comprehension of trends and patterns. Sales teams can track important performance metrics in real time courtesy to the interactive dashboards on the platform, which enable them to react quickly to changes in the market. In addition, merging data from numerous locations is made simpler by Power BI's data integration capabilities. This is crucial for sales because there are many numerous sources from which data comes. All things considered, Power BI advances data analysis, backs analytics that are predictive, and encourages flexible, data-driven approaches in the continuously shifting sector of sales.

1. Data Source Description and BI Requirements

1.1 Introduction

While exploring the dynamic of commercial context, a superstore big company was used in testing my expertise and knowledge to deduce vital sales information based on this dataset by identifying the performance of products, geographically defined dynamics, and a consumer preference which is crucial in the current context of more prominent customer and market demands. When one looks closer, one finds that using predictive model is a strategic tool that can help the business forecast sells and maximize profitability. Tailored plans are driven by this through analysis, which helps uncover revenue-generating possibilities and potential problems. Gaining useful insights from sales data is essential for staying ahead of the competition and making well-informed, strategic decisions in the highly competitive market as the company environment changes.

This BI project seeks to analyse data on Superstore using a supermarket in USA as a case study in order to gain insight on the sales, profitability and best time to shop for goods and how this factor can be driven to generate sales, profit etc.

1.2 Data Set Source and Description

1.2.1 Source

The dataset used is a Kaggle dataset which contains information related to Sales, Profits and other interesting facts of a superstore giant dataset.

Kaggle: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final.

This dataset contains 21 columns, which contains information that best gives an overall knowledge about what the dataset encompasses and how they will contribute to generating insights.

S/N	Column Name	Description
1	Row ID	Unique ID for each row
2	Order ID	Unique order ID for each Customer
3	Order Date	Order Date of the product
4	Ship Date	Shipping date of the product
5	Ship Mode	Shipping mode specified by the customer
6	Customer ID	Unique ID to identify each customer
7	Customer Name	Name of the customer
8	Segment	The segment where the customer belongs
9	Country	Country of residence of the customer
10	City	City of residence of the customer

11	State	State of residence of the customer
12	Postal Code	Postal Code of every customer
13	Category	Category of the product ordered
14	Subcategory	Sub-category of the product ordered
15	Product Name	Name of the product
16	Sales	Sales of the product
17	Quantity	Quantity of the product
18	Discount	Discount Provided
19	Profit	Profit/loss incurred

Table 3: Data Description

BI KPI Requirement and Questions

Key Performance Indicator, or KPI, is a measurable metric that shows how well an organisation is doing at accomplishing its main goals. KPIs are essential indicators for assessing performance, monitoring advancement, and making defensible choices that guarantee alignment with strategic objectives and promote achievement in a range of corporate operations, the KPI used in my case includes.

Average customer lifespan
Average Discount Rate by Product
Customer Lifetime Value
Maximum Sales In each Region
Number of customers by City
Number of customers
Product Rank
Profit Margin
Profitability Index
Sales Growth Rate
Total Profit
Total Revenue
Total Sales

Discounted Amount
Net Sales (After discount)
profitability Ratio
Sales Quantity Indicator

Table 4: DAX

For analytics and visuals, high-level scope analysis involves defining project goals, outlining the scope of data analysis, identifying important stakeholders, compiling baseline requirements, and assessing potential hazards. It seeks to create a clear knowledge of the objectives, limitations, and data visualisation. By managing stakeholder expectations and coordinating objectives with organisational goals, this preliminary analysis lays the groundwork for successful analytics programmes.

Business Intelligence Questions

Fig 50: Home Page Dashboard.

A detailed explanation is provided below explaining each.

- ❖ Visuals
- ❖ KPI'S
- ❖ FIELDS
- ❖ BUSINESS QUESTIONS
- ❖ CHARTS

❖ DAX MEASURES.

Sales Analysis Questions:

- What is the distribution of customers across different categories?
- How does the sum of profit vary across different region and years?
- What is the contribution of each segment and category to the overall profit?
- Which states are the top 5 in terms of Sales?

Customer Analysis Questions:

- Identify the top-profit product by segments; compare sales quantities and discounts for strategic insights.
- Which year exhibits the highest and lowest overall business performance, considering profit, quantity and quality.
- What is the correlation between product rank and profitability, and how does this relationship vary across different quantity level?

Product Analysis Questions:

- Identify monthly sales variation by category revealing recurring trends or patterns throughout the year?
- Examine the discount rate and their correlation with sales, identifying the impact on specific product categories.
- Determine Which Subcategories are most important, monitor the progression of profit, and identify regions where profitability has changed noticeably over time.

Geographical Analysis Questions:

- Which Region are the highest and least more profitable?
- Which Region has the most Sales?
- Which City and Region has the most customers

Marketing Analysis Questions:

- What was the trend observed over time in terms of sales and profit?
- Which product categories contributes the most to our total sales, profit and discounted amount by category?
- How does the Application of discount impact our sales?

Principal Users Communities:

- CEO's and other management team members are actively involved in spotting important trends that improve company success.

- Marketing staff: Focus on products that are in great demand among a variety of demographics. To increase sales, strategically address products that are not doing so well.
- Operations team (Customer Analysis): Improving order processing efficiency and making the best use of available resources to deliver the greatest possible customer experience.
- Marketing team: Keeping an eye on Revenue, Earnings, and Expenses to help with financial planning.

2. Methodology

Defining goals, assessing current data sources, modelling data, choosing suitable tools, creating reports and dashboards, educating users, putting the solution into practice, guaranteeing security, and ongoing testing and improvement are all part of implementing a business intelligence (BI) solution. Important components include adaptability, incremental improvements, and stakeholder collaboration. Through the process, BI projects are intended to be in line with organisational objectives, delivering useful information for strategic planning and well-informed decision making.

Why Power BI?

Power BI was selected because of its powerful information visualisation, ease of use, and integration possibilities. It turns unstructured data into useful insights using interactive dashboards and advanced analytics. Power BI is the ideal choice for companies looking for robust and user-friendly BI solutions because of its reasonable price, flexibility, and connection with a variety of data sources.

2.1 Data Collection and Preparation

Data source was gotten from the Kaggle Repository:
https://www.kaggle.com/datasets/vivek468/superstore-dataset-final.
The first step in the Power BI ETL (Extract, Transform, Load) process is data extraction from multiple sources which mine was gotten from Kaggle, including files and databases. Users can convert raw data by applying filters, merging tables, and correcting discrepancies with Power Query, a potent ETL tool within Power BI. By enhancing the accuracy of the data, this transformation prepares it ready for analysis that is insightful. After that, the data that has been cleaned is added to Power BI's data model, which serves as the starting point for building intelligent dashboards and reports. The whole data preparation procedure is made simpler by Power BI's ETL capabilities, enabling a smooth and effective procedure for users to extract insightful information from

Key Findings:

For this project, the Superstore data which includes sales, customer, and shipping information is a vital resource for forecasting. Comprehensive investigation is provided with feasible by the interrelated fact and dimension tables, which create relationships inside a Star schema model. The vitality of the dataset originates from its capacity to offer perceptions into consumer conduct, the effectiveness of products, and geographical disparities, so enabling precise projections and tactical decision-making to maximize marketing results.
This report aims to accomplish this by addressing the following inquiries under each category:
The main interface has buttons that are placed in a strategic way to offer direct links for quick access to pages.

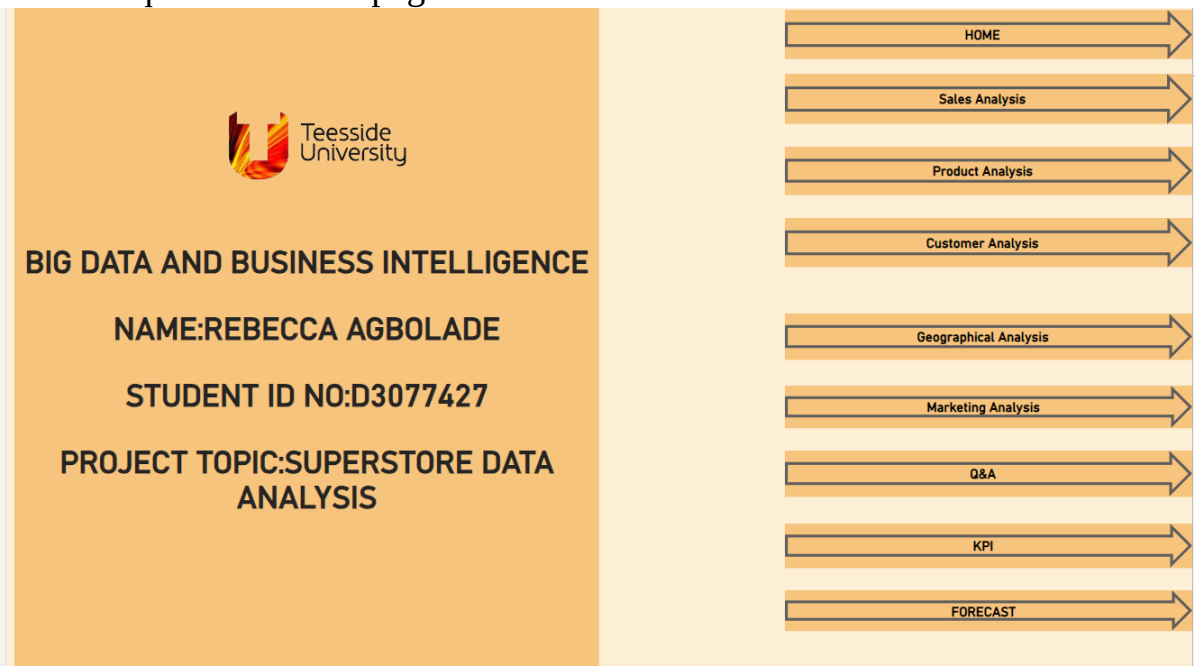


Fig 51: Cover page

Following the Homepage of the dashboard

Fig 52: Home page Dashboard.

A detailed explanation is provided below explaining each.

- ❖ Visuals
- ❖ KPI'S
- ❖ FIELDS
- ❖ BUSINESS QUESTIONS
- ❖ CHARTS
- ❖ DAX MEASURES.

Sales Analysis Questions:

- What is the distribution of customers across different categories?
- How does the sum of profit vary across different region and years?
- What is the contribution of each segment and category to the overall profit?
- Which states are the top 5 in terms of Sales?

Customer Analysis Questions:

- Identify the top-profit product by segments; compare sales quantities and discounts for strategic insights.
- Which year exhibits the highest and lowest overall business performance, considering profit, quantity and quality.
- What is the correlation between product rank and profitability, and how does this relationship vary across different quantity level?

Product Analysis Questions:

- Identify monthly sales variation by category revealing recurring trends or patterns throughout the year?
- Examine the discount rate and their correlation with sales, identifying the impact on specific product categories.
- Determine Which Subcategories are most important, monitor the progression of profit, and identify regions where profitability has changed noticeably over time.

Geographical Analysis Questions:

- Which Region are the highest and least more profitable?
- Which Region has the most Sales?
- Which City and Region has the most customers

Marketing Analysis Questions:

- What was the trend observed over time in terms of sales and profit?
- Which product categories contributes the most to our total sales, profit and discounted amount by category?
- How does the Application of discount impact our sales?

Principal Users Communities:

- CEO's and other management team members are actively involved in spotting important trends that improve company success.
- Marketing staff: Focus on products that are in great demand among a variety of demographics. To increase sales, strategically address products that are not doing so well.
- Operations team (Customer Analysis): Improving order processing efficiency and making the best use of available resources to deliver the greatest possible customer experience.
- Marketing team: Keeping an eye on Revenue, Earnings, and Expenses to help with financial planning.

SALES ANALYSIS DASHBOARD OVERVIEW

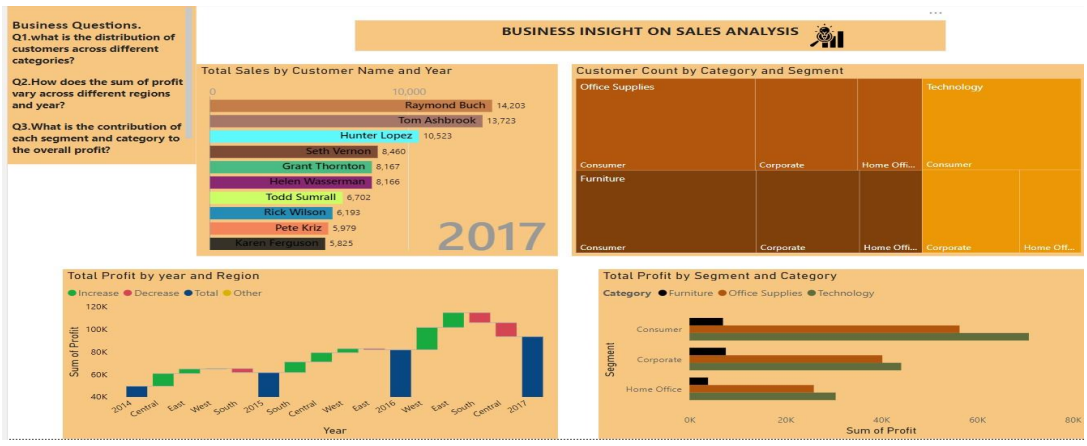


Fig 53: Sales Analysis Dashboard.

Sales Analysis Questions:

- How does the profit margin vary across different region and number of customers in each year?
- What is the contribution of each segment and category to the overall profit?
- Which states are the top 5 in terms of Sales?
- Which customer has the highest number of sales?

This page provides a general overview to the questions listed above and the answers to the questions asked.

- 2014
  Profit margin =10.23%
  Total Revenue =484.25k
  No of customer by the regions (central, East, West, North) =268

- 2015
  Profit margin =13.10%
  Total Revenue =470.35k
  No of customer by the regions (central, East, West, North) =275

- 2016
  Profit margin =13.43%
  Total Revenue =609.21k
  No of customer by the regions (central, East, West, North) =320

- 2017
  Profit margin =12.74%
  Total Revenue =733.22k
  No of customer by the regions (central, East, West, North) =350
  Hence, from the explanation given above we can see that the profit margin for 2016 was the highest, and 2017 has the highest total revenue generated. Also,2017 has the highest customer.

  Q2. For each segment we can see that Technology had the highest profit in consumer, corporate and home office.

  Q3. The highest top 5 in terms of sales are California, New York. Texas, Washington Pennsylvania.

  Q4. The customer with the highest number of sales is Raymond Buch made a total of 14,203.

PRODUCT ANALYSIS DASHBOARD OVERVIEW

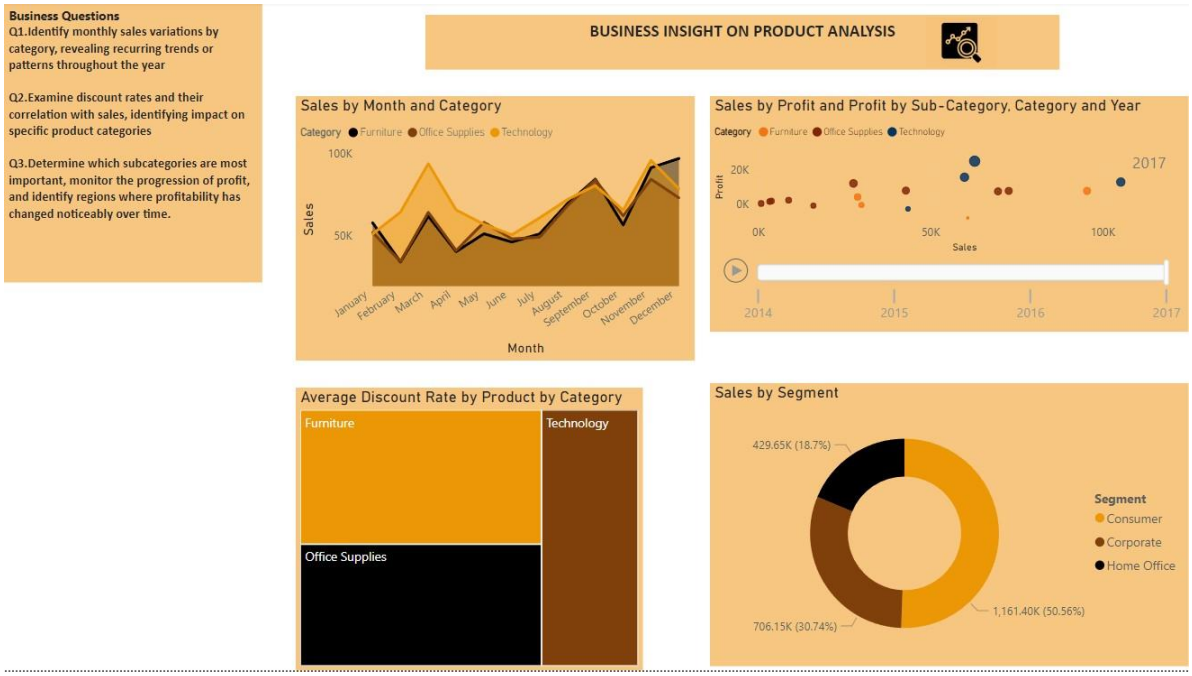


Fig 54: Product Analysis Dashboard.

Product Analysis Questions:

- Identify monthly sales variation by category revealing recurring trends or patterns throughout the year?
- Examine the discount rate and their correlation with sales, identifying the impact on specific product categories.
- Determine Which Subcategories are most important, monitor the progression of profit, and identify regions where profitability has changed noticeably over time.

This page provides a general overview to the questions listed above and the answers to the questions asked.

- November Performed well in terms of sales.
- Technology has the lowest discount and still performed well, my suggestion might be due to the high quality, brand loyalty or high market demand for the product Also, the place of effective marketing strategies cannot be override.
- Copier has improved really well and we can see that in the profit progression.

CUSTOMER ANALYSIS DASHBOARD OVERVIEW

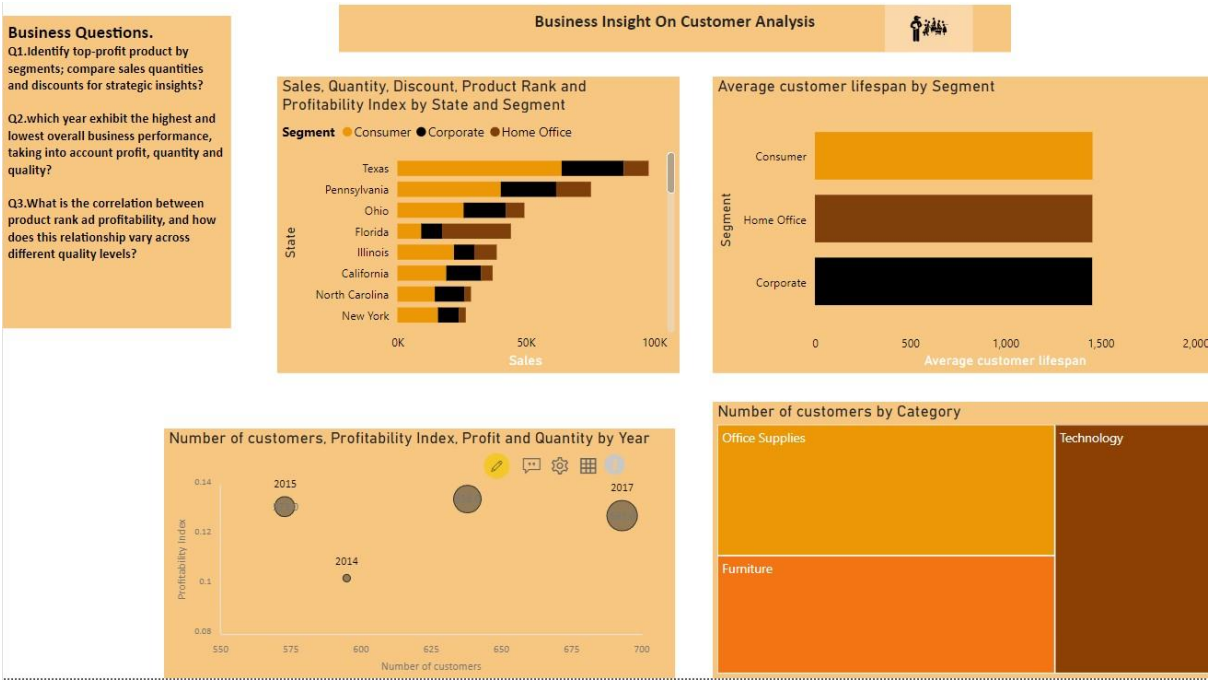


Fig 55: Customer Analysis Dashboard.

- Identify the top-profit product by segments; compare sales quantities and discounts for strategic insights.
- Which year exhibits the highest and lowest overall business performance, considering profit, quantity and quality.
- What is the number of customers per category?

Texas

2017 was the highest and 2014 was the lowest.

GEOGRAPHICAL ANALYSIS DASHBOARD OVERVIEW

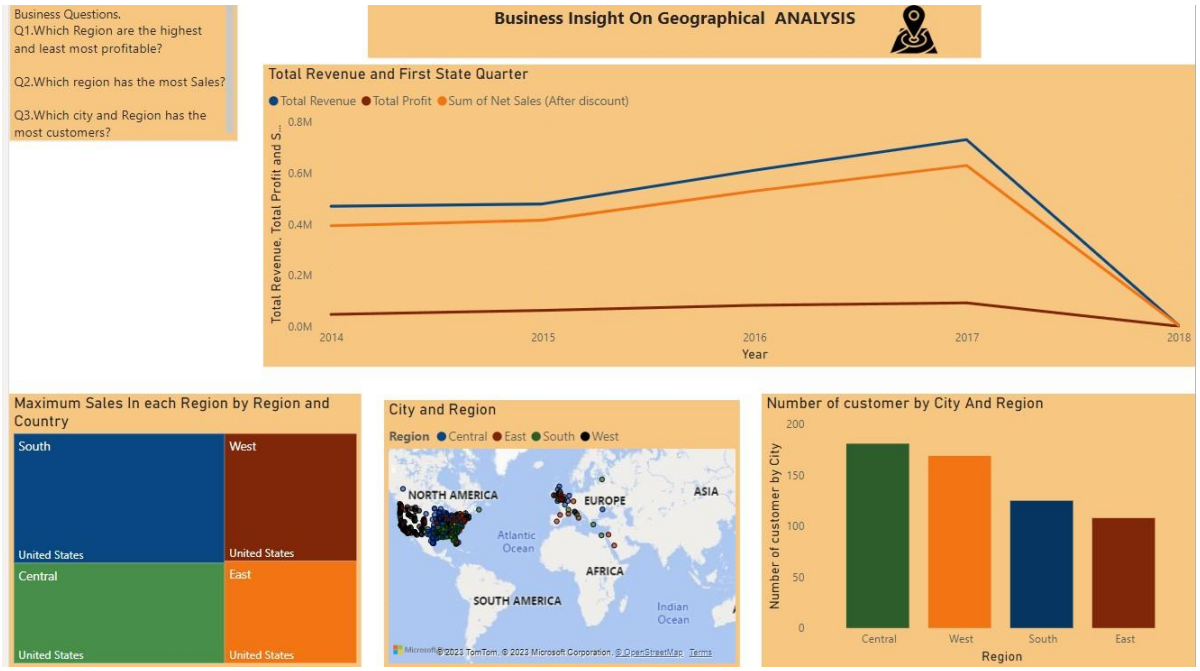


Fig 56: Geographical Analysis Dashboard.

Geographical Analysis Questions:

- Which Region are the highest and least more profitable?
- Which Region has the most Sales?
- Which City and Region has the most customers

South had the highest and the East has the lowest.
The East
Central had the most customer.

MARKETING ANALYSIS DASHBOARD OVERVIEW

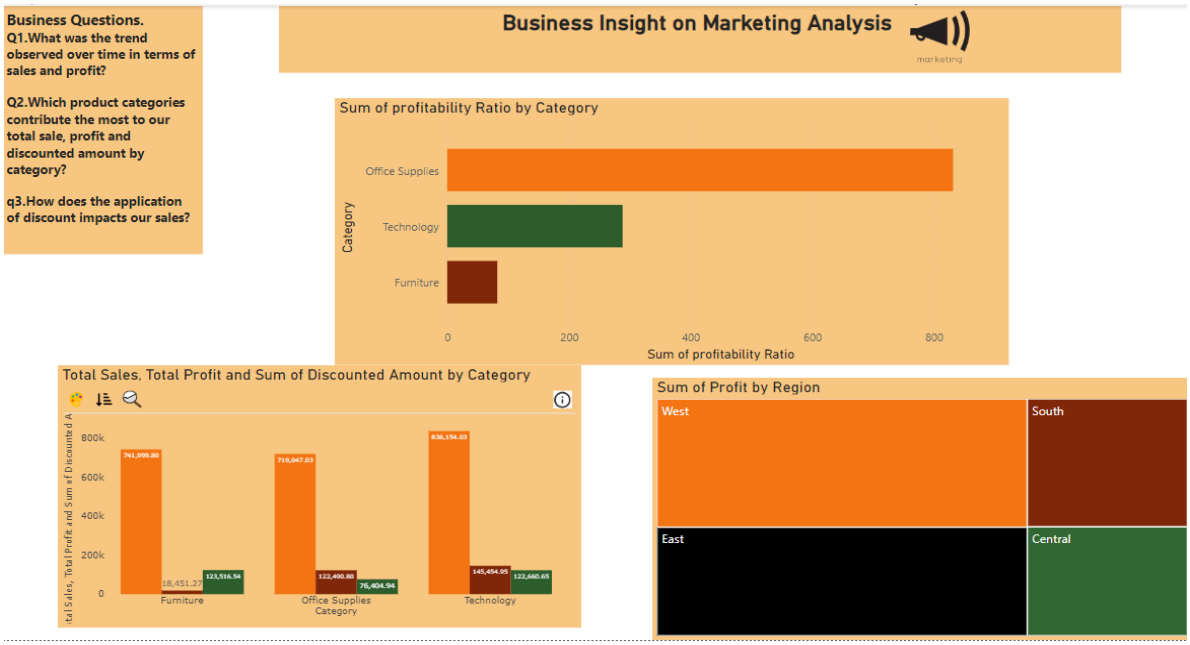


Fig 57: Marketing Analysis Dashboard.

Marketing Analysis Questions:

- What was the trend observed over time in terms of sales and profit?
- Which product categories contributes the most to our total sales, profit and discounted amount by category?
- How does the Application of discount impact our sales?

- Technology had the highest sales and highest profit.
- Technology
- In this dataset not much, because technology had the lowest discount but still sold the most.

KPI ANALYSIS DASHBOARD OVERVIEW

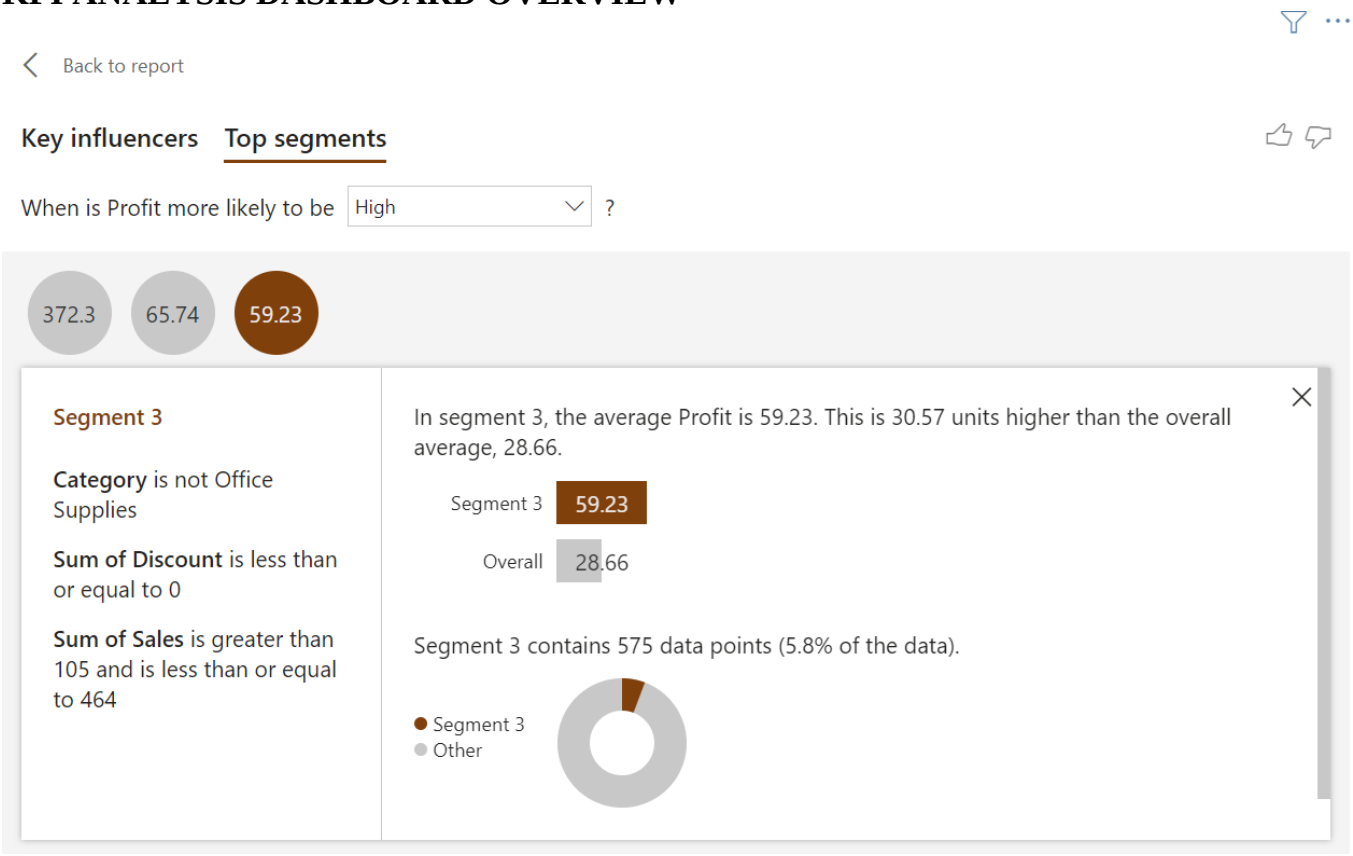


Fig 58: KPI ANALYSIS (Dashboard 5)

FORECAST ANALYSIS DASHBOARD OVERVIEW

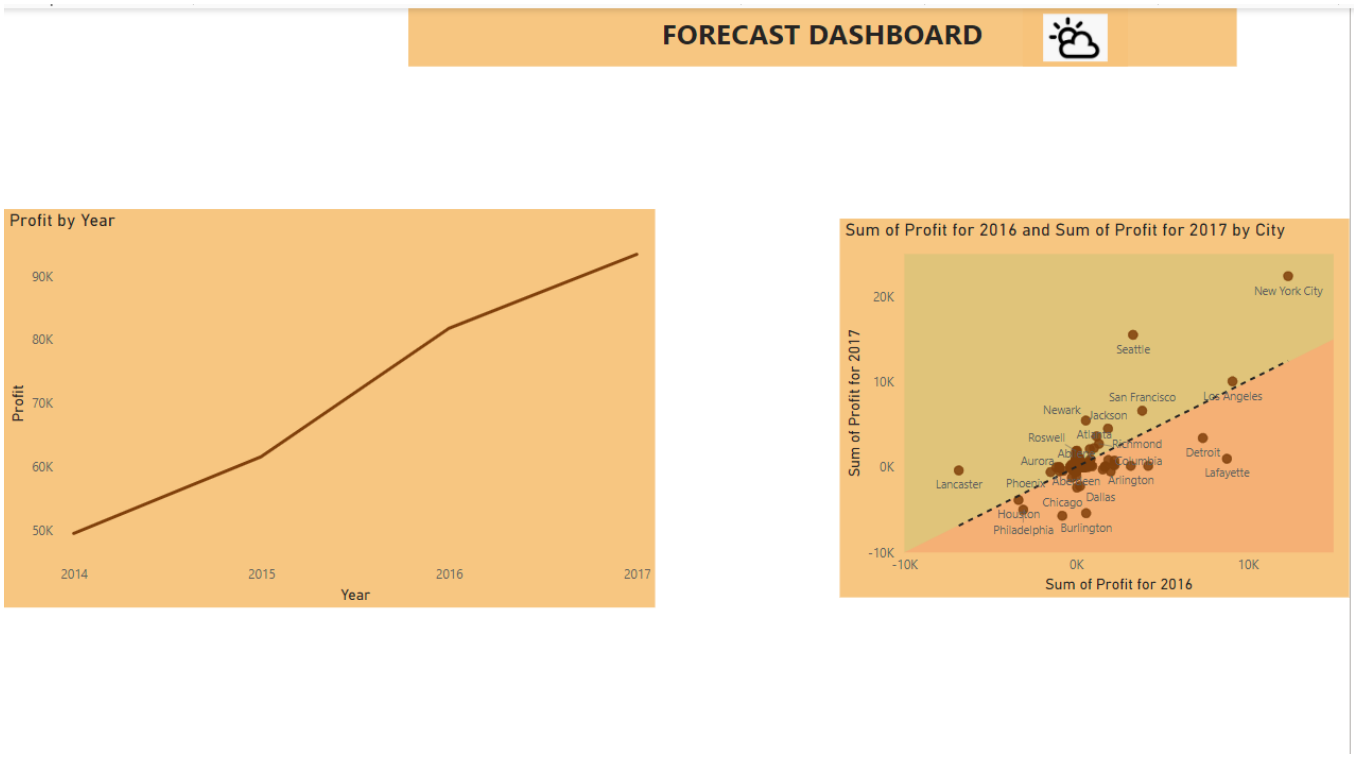


Fig 59: Forecasting Analysis Dashboard.

3. Challenges and Limitations

Handling the complexities of data cleaning and validation, understanding and developing relationships among diverse data dimensions, potential slowdowns when working with large datasets, as well as effectively learning the tool's features are some of the challenges that a student using Power BI with the Superstore dataset may face. Challenges could also include integrating data from many sources, security concerns, and occasional issues with data update. It takes commitment to master Power BI, hone data preparation techniques, and look for assistance to get past these obstacles and have a positive, fulfilling experience analysing the Superstore dataset. Successful analysis can be impeded by data limits, such as poor quality and restricted availability, as well as tool constraints, such as restricted functionality and issues with compatibility. Investing in data quality improvement, investigating innovative approaches, and encouraging an organisational culture of data integration are all necessary to address these issues.

4. Conclusions and Recommendations

Summary:

Sales Analysis:

Customer Distribution: California, New York, Texas, Washington, and Pennsylvania lead in customer distribution.
Profit Variation: 2016 had the highest profit margin; 2017 showed the highest total revenue and customer count in regions (Central, East, West, North).
Top-Profit Product: Technology dominates profits across consumer, corporate, and home office segments.
Yearly Business Performance: November excelled in sales; Technology segment had consistent high profits with low discounts.

Product Analysis:
Monthly Sales Variation: Identify recurring trends in sales by category throughout the year.
Discount Impact: Examine the correlation between discount rates and sales, focusing on specific product categories.

Geographical Analysis:
Profitable Regions: Identify the highest and least profitable regions.
Sales Leaders: Determine regions and cities with the most sales and customers.

Marketing Analysis:
Sales and Profit Trends: Examine trends over time and the contribution of product categories to total sales and profit.
Discount Impact: Evaluate how discount application affects overall sales.

Courses of Action for Stakeholders
Focus on High-Performing Areas: To boost sales and profitability, allocate resources and promotional activities to areas like California and New York.

Optimise the Technology Segment: Use product development and strategic marketing strategies to capitalise on knowledge from the consistently profitable Technology segment.

Strategic Discounting: Examine how discounts affect revenue and profitability and modify plans as necessary to achieve the best possible outcomes.

Engagement with Customers: Pay particular attention to keeping and growing the clientele, particularly in Central and South America.

Constant Monitoring: For quick decision-making, keep an eye on regional performance, product categories, and monthly sales variances.

Investment in Copier Category: Considering the significant uptick in the Copier category's profit growth, you might want to explore making additional investments in this market.

Quality Focus: Take note of the Technology segment's performance with its low discounts; investigate any possible relationships between product quality and customer loyalty.

5. Reference

Kaggle: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final.