



MSc Computer Science

School of Computing, Engineering and Digital technologies

Big Data and Business Intelligence

Development and Stakeholder Report for Global Superstore Sales

Name: Yaswanth Sai Chinthakayala

Student ID: W9640628

Development Report

Introduction

This is a development report for Power BI application. In this report, the detailed analysis of the Global Superstore Sales is done using advanced features like DAX and M language and building dashboards. This dataset consists various columns that are related to sales. The important information related to orders, customer profiles, product categories, and regions are specified. The data analysis on this dataset will help us to gain insights such as stakeholder decisions, improving business processes and enhance overall operational efficiency.

Data Source

This dataset is from the Global Superstores Sales System, which keeps track of all transactions across various markets and regions. The details of this data are provided below.

Link: <https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset/data>

Database Name: The data is available in two formats. Global_Superstore2.csv and Global_Superstore2.xlsx

Columns:

1. Row ID	Distinct Id for each row.
2. Order ID	Distinct Id for each order.
3. Order Date	The date when the order was placed.
4. Ship Date	The date when the order was shipped.
5. Ship Mode	The mode of shipping.
6. Customer ID	A distinct Id for each customer.
7. Customer Name	The name of the customer.
8. Segment	The market segment
9. City	The city where the customer is located.
10. State	Customer's state
11. Country	Customer's country
12. Postal Code	Customer's postal code
13. Market	The market for order
14. Region	The region where the order was placed
15. Product ID	A Distinct Id for each product.
16. Category	Product category
17. Sub-Category	Product sub-category
18. Product Name	The name of the product.
19. Sales	The sales amount for the product.
20. Quantity	The quantity of the products ordered.

21. Discount	Product discount.
22. Profit	Order Profit.
23. Shipping Cost	The cost associated with shipping.
24. Order Priority	The priority of the order.

KPI and Questions

KPI means Key Performance Indicator. There are seven advanced KPI dashboards.

Overview:

- Evaluate sales growth over specific time periods.
- Highlight all the important figures.

Geographic Analysis:

- Sales distribution across different countries and regions.
- Market performance.

Product Analysis:

- Identify top performing products.
- Explore sales for various categories, sub-categories.

Customer Analysis:

- Understand customer demographics.
- Segment based purchases.

Discount and Profit:

- Overall Profit Margin.
- Effect of discounts on profit.

Shipping:

- Average Shipping time for the orders.
- Relation between order priority and shipping costs

AI Visuals:

- Forecasting future sales and factors influencing profit using AI.
- Question and answer feature for stakeholders.

Questions:

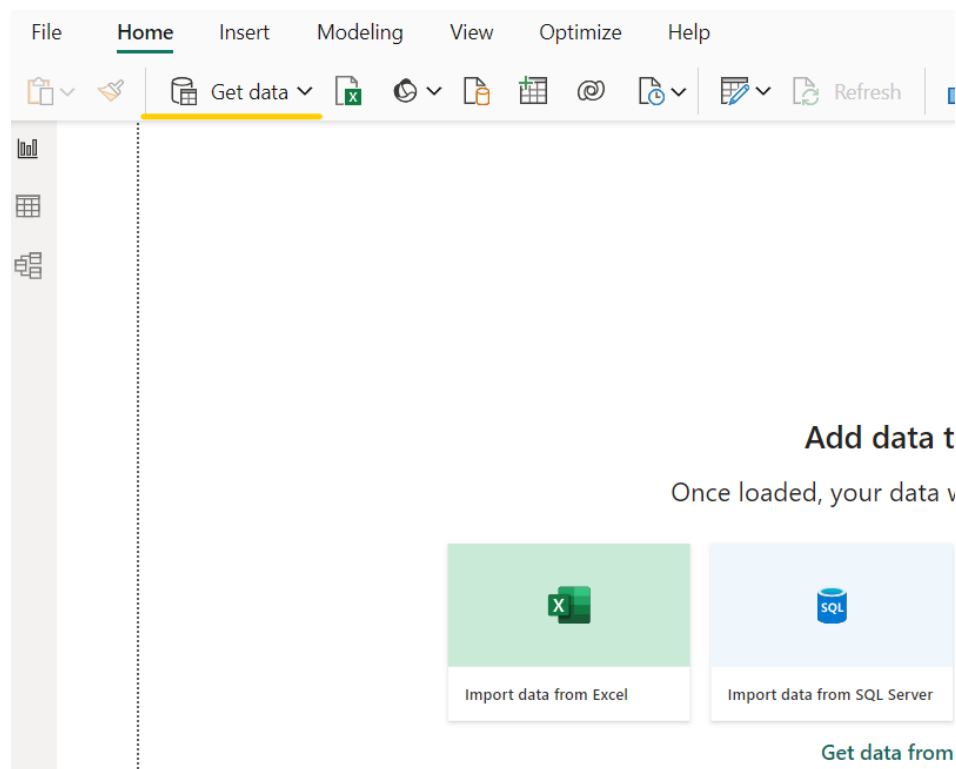
1. What are the total sales for the selected period?
2. What are the top performing locations in terms of sales?
3. What is the sales distribution among the different customer segments?
4. What is the contribution of each product category to the total sales and which sub-categories have the highest growth?
5. What is the average shipping cost for each shipping mode? How does the choice of shipping mode impact shipping costs and order priority?
6. What is the correlation between discount and profit?
7. How Artificial Intelligence helps Global Superstore to increase sales?

Data Processing

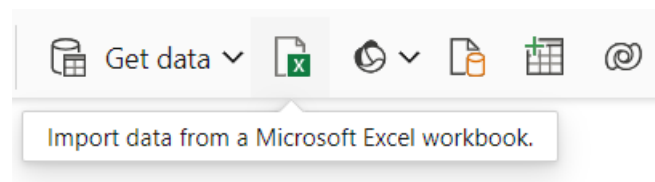
Data Processing is very important step that involves data cleaning and organizing of data in such a way that it will be very helpful to perform data analysis and data visualization.

Loading Data:

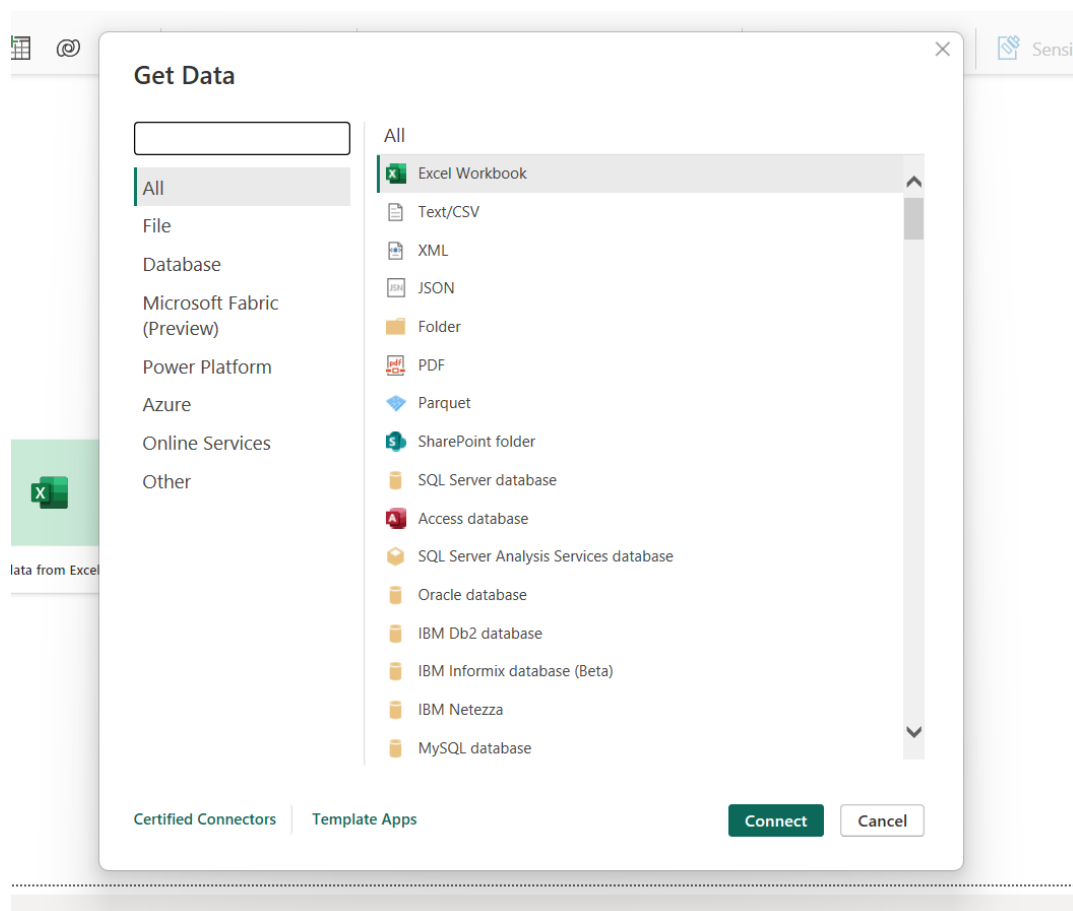
The first step is loading the data. The dataset was loaded by clicking on the get data button as shown here.



If the data file is Excel workbook, the data can be imported directly by clicking on this button below and selecting the file from the file explorer.



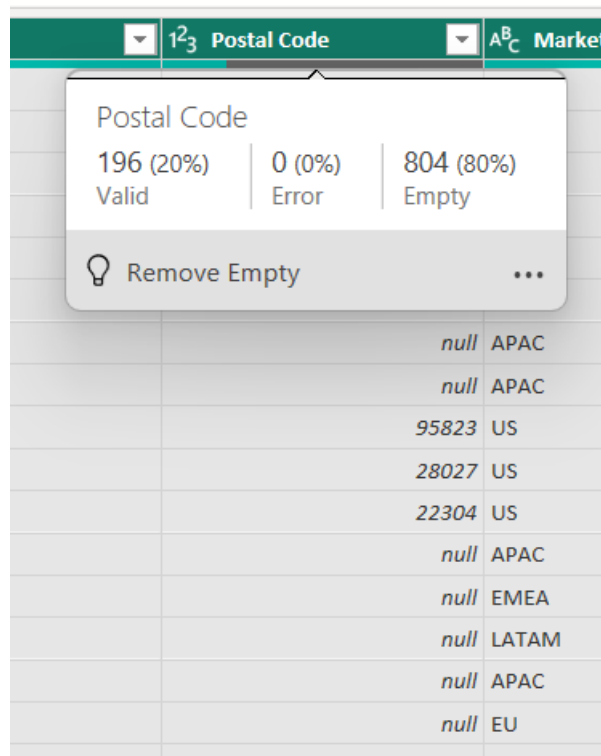
In this window, format of the data file is selected and connect button is clicked. Excel workbook is chosen here because the data is being loaded from the excel workbook.



The data is loaded and displayed here. There may be many excel sheets in the data, Sheet 1 is chosen because all the data is present in this file itself. Next, the data needs to be transformed because there might be some errors or null values.

Dealing with Null Values

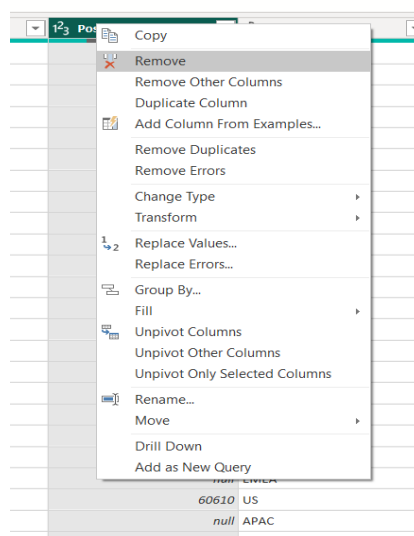
The data is completely examined and most of the data is clean except “Postal Code” column. In this column large number of values are missing. If the mouse is hovered over the column, the details can be seen as shown below.



The screenshot shows a data table with columns 'Postal Code' and 'Market'. A tooltip is displayed over the 'Postal Code' column header, showing a summary of the data: 196 (20%) Valid, 0 (0%) Error, and 804 (80%) Empty. Below the summary, there is a 'Remove Empty' button with a lightbulb icon and a three-dot menu icon. The table data is partially visible below the tooltip.

Postal Code	Market
	APAC
	APAC
95823	US
28027	US
22304	US
	APAC
	EMEA
	LATAM
	APAC
	EU

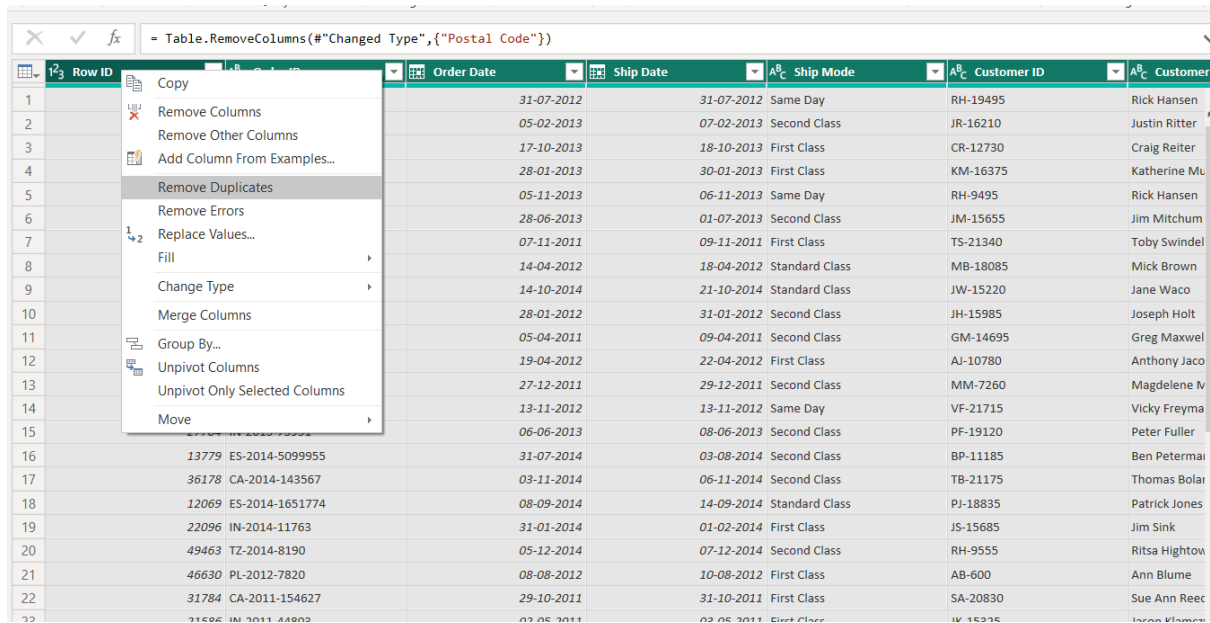
As the null values are 80%, they can't be removed because 80% of the data will be lost. Alternatively, this postal code column can be removed. This is the right choice; this can be done by pressing right click on the column and choosing remove option.



If there are other columns that have fewer null values, it can be dealt with fill down or fill up option, which will replace the null values with the upper or lower row values.

Dealing with Duplicates:

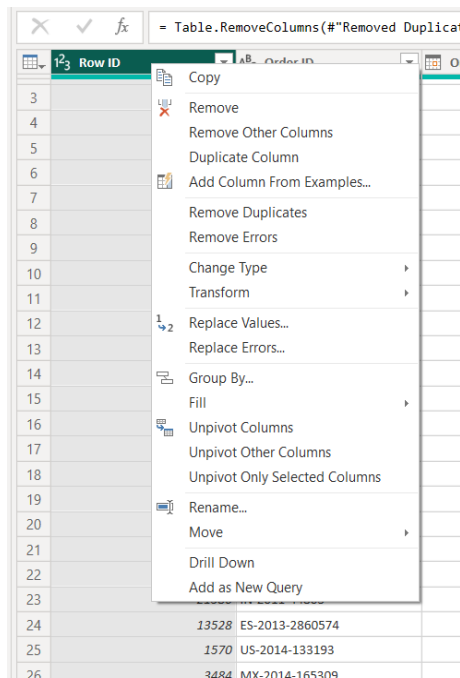
The data is clean, so there might be no duplicates. To be sure, this step needs to be done to avoid future problems. To do this, select all the columns, right click on the column, and choose the remove duplicates option.



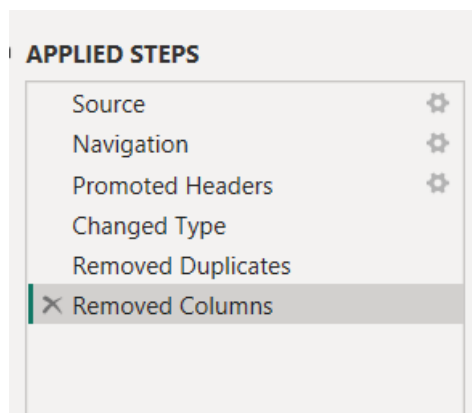
Row ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	31-07-2012	31-07-2012	Same Day	RH-19495	Rick Hansen
2	05-02-2013	07-02-2013	Second Class	JR-16210	Justin Ritter
3	17-10-2013	18-10-2013	First Class	CR-12730	Craig Reiter
4	28-01-2013	30-01-2013	First Class	KM-16375	Katherine Mu
5	05-11-2013	06-11-2013	Same Day	RH-9495	Rick Hansen
6	28-06-2013	01-07-2013	Second Class	JM-15655	Jim Mitchum
7	07-11-2011	09-11-2011	First Class	TS-21340	Toby Swindel
8	14-04-2012	18-04-2012	Standard Class	MB-18085	Mick Brown
9	14-10-2014	21-10-2014	Standard Class	JW-15220	Jane Waco
10	28-01-2012	31-01-2012	Second Class	JH-15985	Joseph Holt
11	05-04-2011	09-04-2011	Second Class	GM-14695	Greg Maxwell
12	19-04-2012	22-04-2012	First Class	AJ-10780	Anthony Jacob
13	27-12-2011	29-12-2011	Second Class	MM-7260	Magdelene Iv
14	13-11-2012	13-11-2012	Same Day	VF-21715	Vicky Freyma
15	06-06-2013	08-06-2013	Second Class	PF-19120	Peter Fuller
16	31-07-2014	03-08-2014	Second Class	BP-11185	Ben Petermai
17	03-11-2014	06-11-2014	Second Class	TB-21175	Thomas Bolai
18	08-09-2014	14-09-2014	Standard Class	PJ-18835	Patrick Jones
19	31-01-2014	01-02-2014	First Class	JS-15685	Jim Sink
20	05-12-2014	07-12-2014	Second Class	RH-9555	Ritsa Hightow
21	08-08-2012	10-08-2012	First Class	AB-600	Ann Blume
22	29-10-2011	31-10-2011	First Class	SA-20830	Sue Ann Reec
23	02-05-2011	02-05-2011	First Class	W-15235	Waco Klamer

Unused columns:

Some columns are redundant as they have no use, or people opt not to use them in data visualization. If these columns are examined, Row Id row is redundant here. This can be removed by pressing right click on the column and choosing the remove option.



All the applied steps can be seen in the right-hand side of the transform data window. These steps can be deleted or changed as for our requirements.



Data Type Adjustments:

Power bi detects the data type of the column, sometimes it may be wrong. In this case, all our data is categorized into correct datatypes.

Data/Feature Engineering:

Feature engineering is a process to add new feature to the dataset, new columns can be created or extracted using the current features. New columns are added to the dataset like, Absolute profit, overall Loss, Day, Month and Year Columns using DAX and M Language. This will be discussed in detail in DAX and M language Section.

Outliers:

Outliers are values which are completely in different range from the normal data points. One more important step is to deal with the outliers. This data is clean, and it has no outliers.

Data Modelling and Star Schema:

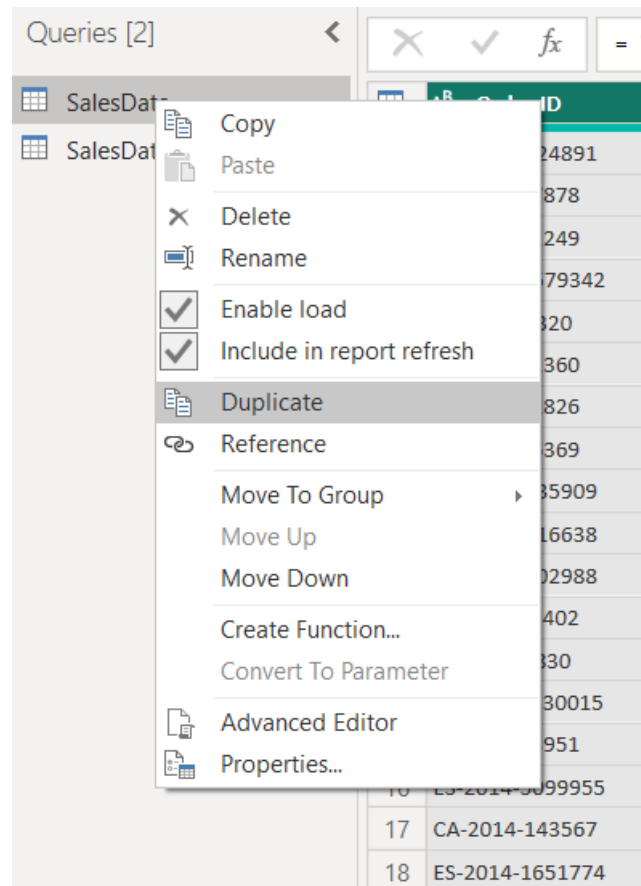
Data modelling is very significant in creating a power bi application. Data modelling is nothing but organizing our data, structuring in a way that it improves efficiency in the database. The advantages of data modelling are simplified data files with unnecessary data, easy data retrieval and it occupies less space which will be very helpful for larger databases.

This dataset has a lot of columns, to keep the data more organized, the data can be split into smaller chunks and can be linked to each other by creating relationships. There are various details in the dataset like product information, customer information, and ship mode.

Star schema is one of the database schemas where there will be one fact table and many dimension tables. The arrangement looks like star as fact table is in centre and dimension tables surrounding it.

Creating a Star Schema

Let us create customer details dimension table now. Firstly, the current table is renamed as SalesData and duplicated.



Now the new table was renamed to CustomerDetails. After that, only the customer related columns were selected, and all the other columns were deleted as shown here.

Close

New Query

Data Sources

Parameters

Query

Manage Columns

Reduce Rows

Sort

Transform

Combine

Queries [2]

</

After this, a relationship between the dimension table and fact table needs to be established. The relation should be many to one, it can't be many to many as other problems may arise in the future. To make relationships many to one, the column that we connect to the dimension table needs to be duplicate free. This was done in our dataset as well.

Here, the Customer ID column is selected, and duplicates are removed. In addition, the column other than the foreign key can be deleted because they are redundant.

= Table.SelectColumns(#"Removed Columns",{ "Cus	
Customer	Segment
1 RH-19495	Consumer
2 JR-16210	Corporate
3 CR-12730	Consumer
4 KM-16375	Home Office
5 RH-9495	Consumer
6 JM-15655	Corporate
7 TS-21340	Consumer
8 MB-18085	Consumer
9 JW-15220	Corporate
10 JH-15985	Consumer
11 GM-14695	Corporate
12 AJ-10780	Corporate
13 MM-7260	Consumer
14 VF-21715	Home Office
15 PF-19120	Consumer
16 BP-11185	Corporate
17 TB-21175	Corporate
18 PJ-18835	Corporate
19 JS-15685	Corporate
20 RH-9555	Consumer
21 AB-600	Corporate
22 SA-20830	Consumer
23 JK-15325	Corporate
Jason Klamczynski	

The customer name and customer segment columns are selected and removed. Only the customer id column is common between the CustomerDetails and SalesData table.

Customer Name	Segment
Rick Hansen	Consumer
Justin Ritter	Corporate
Craig Reiter	Consumer
Katherine Murray	Home Office
Rick Hansen	Consumer
Jim Mitchum	Corporate
Toby Swindell	Consumer
Mick Brown	Consumer
Jane Waco	Corporate
Joseph Holt	Consumer
Greg Maxwell	Corporate
Anthony Jacobs	Corporate
Magdelene Morse	Consumer
Vicky Freymann	Home Office
Peter Fuller	Consumer
Ben Peterman	Corporate
Thomas Boland	Corporate
Patrick Jones	Corporate
Jim Sink	Corporate
Ritsa Hightower	Consumer

- Copy
- Remove Columns
- Remove Other Columns
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Replace Values...
- Fill
- Change Type
- Transform
- Merge Columns
- Group By...
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Move

Similarly, product info can be separated in the same way. This dimension table can be connected to fact table using the Product ID column.

Table.Distinct("#Removed Other Columns", {"Product ID"})

Product ID

Category

Sub-Category

Product Name

1	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wireless Headset System
2	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black
3	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID
4	TEC-PH-10004583	Technology	Phones	Motorola Smart Phone, Cordless
5	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed
6	TEC-PH-10000030	Technology	Phones	Samsung Smart Phone, with Caller ID
7	FUR-CH-10004050	Furniture	Chairs	Novimex Executive Leather Armchair, Adjustable
8	FUR-TA-10002958	Furniture	Tables	Chromcraft Conference Table, Fully Assembled
9	OFF-BI-10003527	Office Supplies	Binders	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Ma...
10	FUR-TA-10000198	Furniture	Tables	Chromcraft Bull-Nose Wood Oval Conference Tables & Bases
11	OFF-SU-10002881	Office Supplies	Supplies	Martin Yale Chadless Opener Electric Letter Opener
12	FUR-TA-10001889	Furniture	Tables	Bevis Conference Table, Fully Assembled
13	TEC-CIS-10001717	Technology	Phones	Cisco Smart Phone, with Caller ID
14	FUR-CH-10002033	Furniture	Chairs	Harbour Creations Executive Leather Armchair, Adjustable
15	OFF-AP-10003500	Office Supplies	Appliances	KitchenAid Microwave, White
16	OFF-AP-10000423	Office Supplies	Appliances	Breville Refrigerator, Red

Query Settings

PROPERTIES

Name

Product

All Properties

APPLIED STEPS

Source

Navigation

Promoted Headers

Changed Type

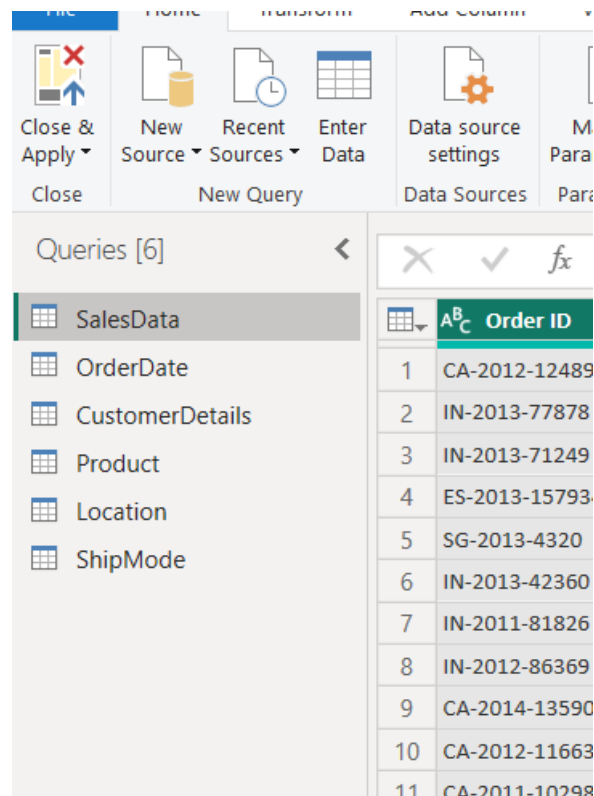
Removed Duplicates

Removed Columns

Removed Other Columns

Removed Duplicates1

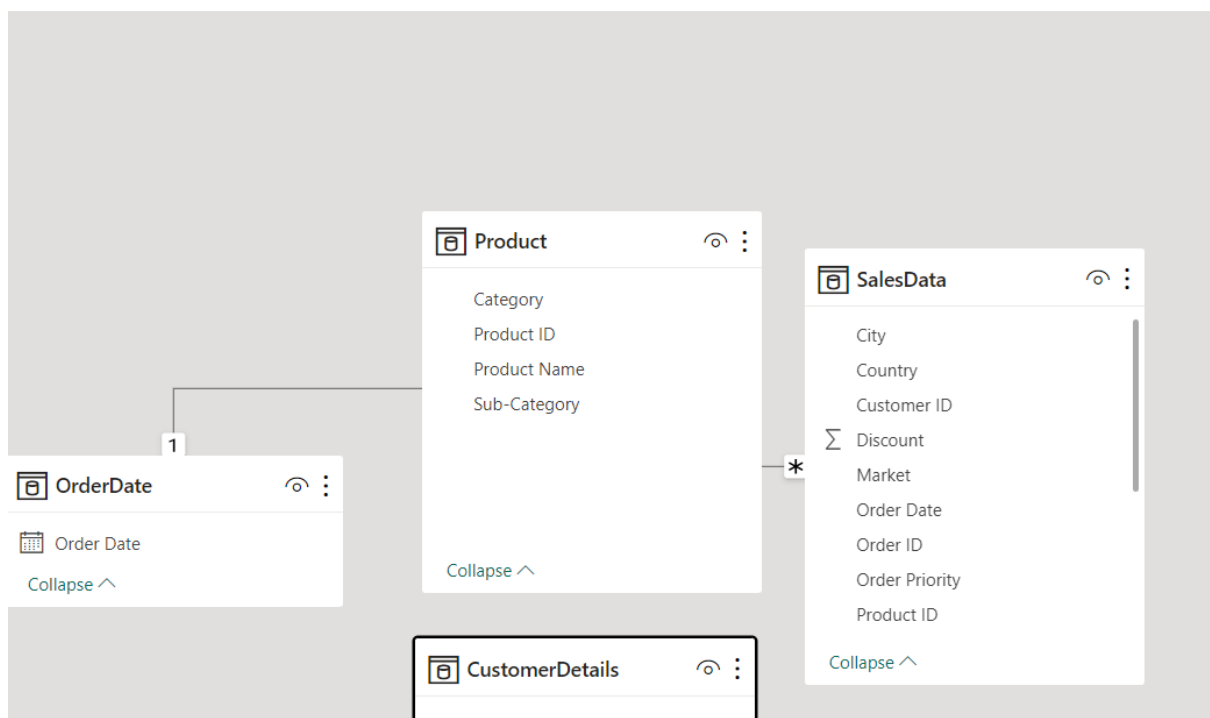
In the same way, all these tables are created. The tables are shown below, the fact table is SalesData, the dimension tables are OrderDate, CustomerDetails, Product, Location and ShipMode.



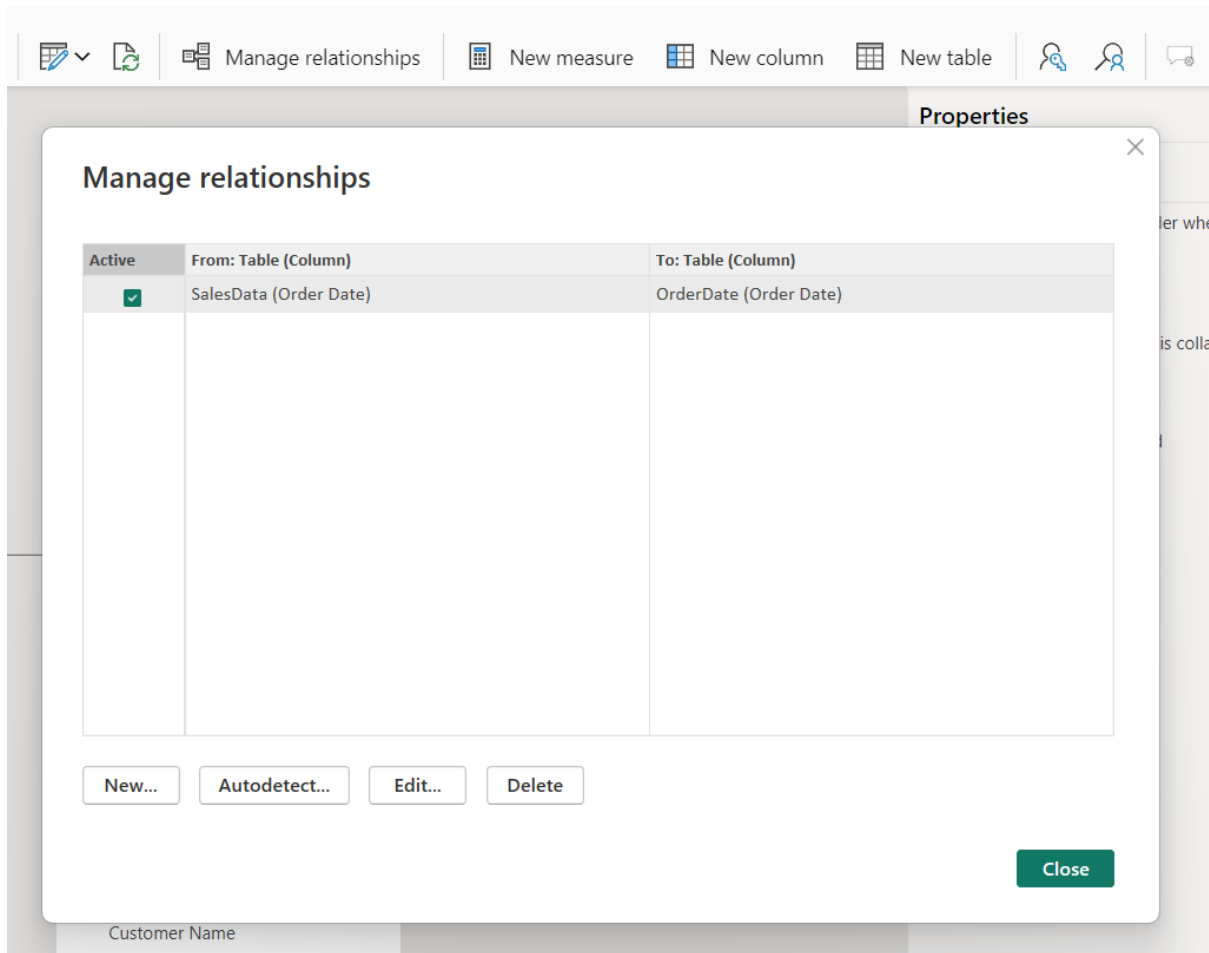
Once all the tables are created, the file changes can be loaded into power bi file by pressing the close and apply button.

Star Schema:

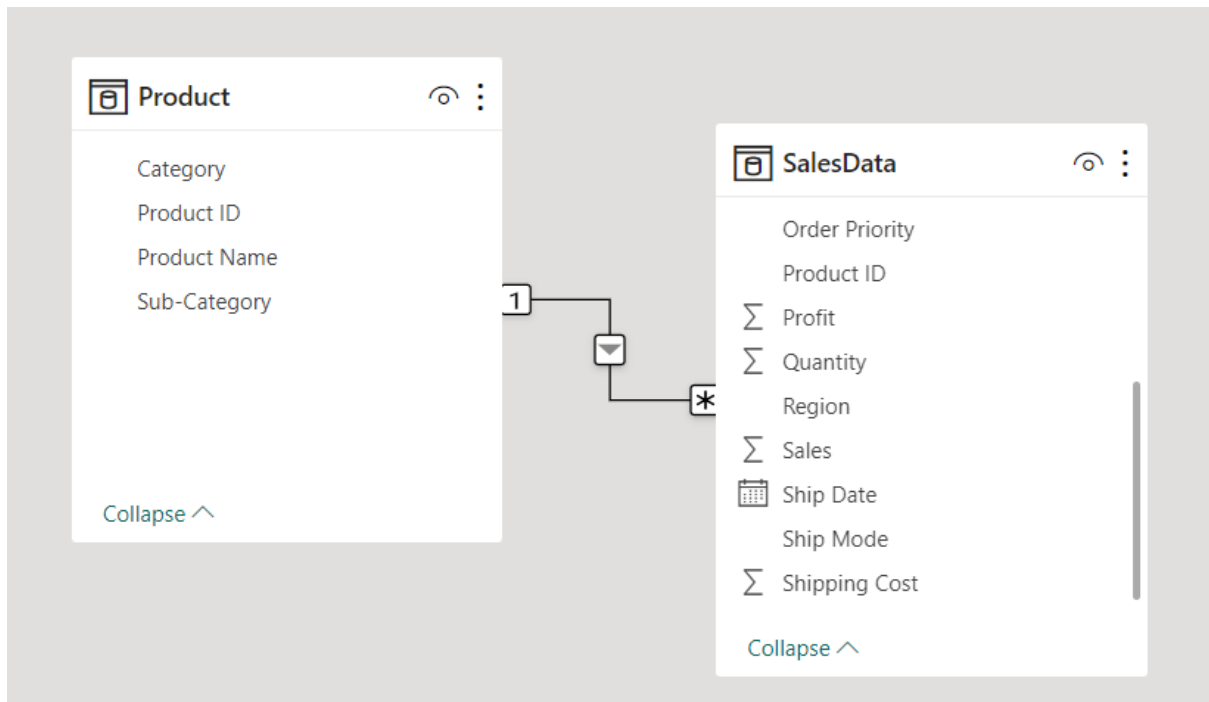
After the changes are applied to the power bi file, some connections are applied by default. These connections may or may not be completely as expected. The best thing to do is to remove all the existing connections.



To remove the existing connections, click on manage relationships, select the existing relation, and click on delete.



To connect new relationship between them, choose the product id, drag it from dimension table and drop onto the fact table.



The connection is established, and it is many to one relationship. To see or edit this relationship, select the option, and check the conditionality, also select the cross-filter detection as both or single depending on the requirements.

Select tables and columns that are related.

SalesData
▼

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Product ID	Sales	Quai
CA-2014-100111	21 September 2014	27 September 2014	Standard Class	SV-20365	TEC-PH-10002885	1299.66	
US-2014-146906	14 March 2014	18 March 2014	Standard Class	MT-17815	TEC-PH-10001809	299.9	
CA-2014-100111	21 September 2014	27 September 2014	Standard Class	SV-20365	TEC-AC-10002647	212.8	

Product
▼

Product ID	Category	Sub-Category	Product Name
OFF-PA-10004385	Office Supplies	Paper	Eaton Parchment Paper, Premium
OFF-PA-10003380	Office Supplies	Paper	Green Bar Memo Slips, Multicolor
OFF-PA-10001385	Office Supplies	Paper	Xerox Memo Slips, 8.5 x 11

Cardinality

Many to one (*:1)
▼

Cross filter direction

Both
▼

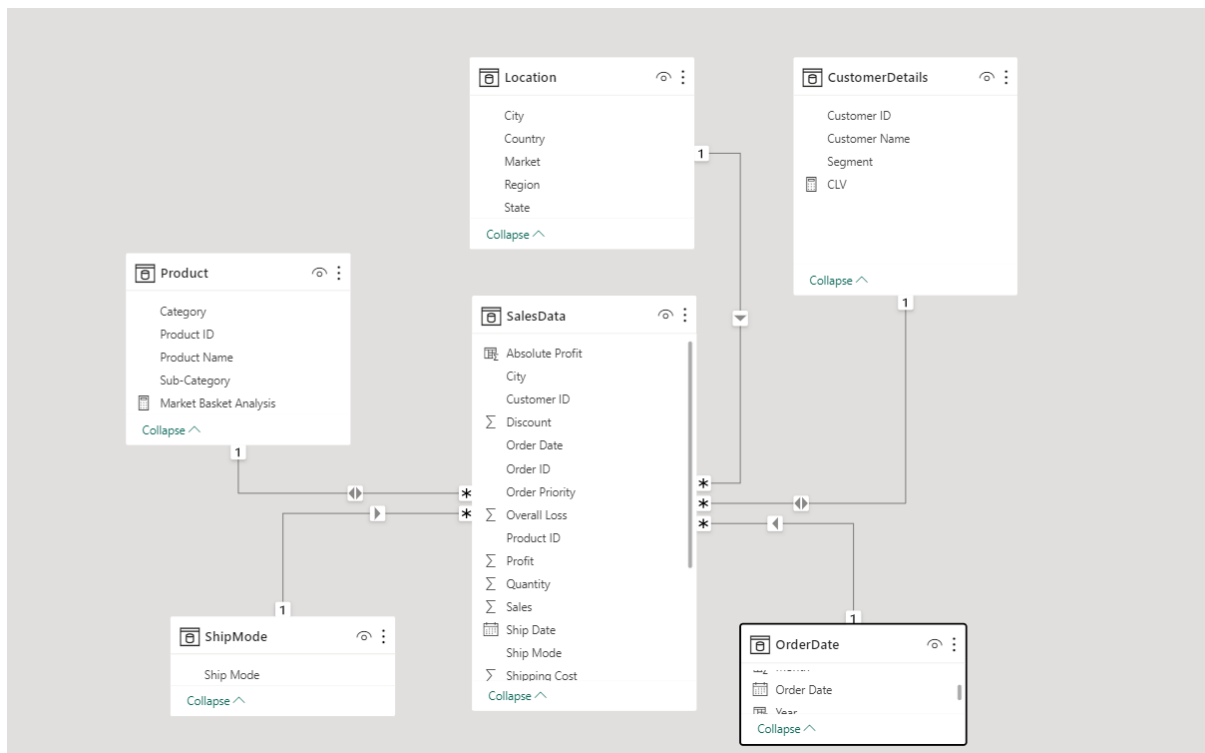
☒ Make this relationship active

☐ Apply security filter in both directions

☐ Assume referential integrity

OK

Cancel



Customer Id in customer table is linked to Customer Id in fact table. Similarly, product ID, City, order date and ship mode are linked. The relationships can be many to one or one to one. They can't be many to many. Here there is no many to many relationships.

This is the completed star schema.

Dax and M language

DAX and M languages are used in power bi to create custom calculations like custom measures, custom columns, and custom tables.

Dax Measures:

All these formulas are used in this section.

COUNTROWS: Counts all the rows of the specific column or table.

AVERAGEX: The AVERAGEX function iterates through each row in the selected table, calculates the query for each row, and then calculates the average of these values.

FILTER: The FILTER function will return a table containing only the rows that meet the specified condition.

Creating Measures/Columns or Tables using Table Tools:

To create a measure, select the table that you want the measure to be created. Click on table tools and select the new measure option.

The screenshot shows the 'Table tools' ribbon with the 'New measure' button selected. Below the ribbon, a table of sales data is displayed. The table has columns: Order ID, Order Date, Ship Date, Ship Mode, Customer ID, City, State, Country, Market, Region, Product ID, Sales, Quantity, Discount, Profit, and Ship Date. A search bar is located on the right side of the table.

Type the formula, in this case “Total Sales” measure is being created. This formula calculates the sum of the sales column in SalesData table and stores it in the variable. This can later be used in the visualizations.

1 Total Sales = SUM(SalesData[Sales])

	Order Date	Ship Date	Ship Mode	Customer ID	City	State	Country
2	18 June 2013	20 June 2013	First Class	MV-18190	Los Angeles	California	United States
12	12 June 2014	12 June 2014	Same Day	DV-13465	Los Angeles	California	United States
15	04 November 2013	09 November 2013	Standard Class	XP-21865	Los Angeles	California	United States
13	29 May 2013	04 June 2013	Standard Class	CL-12565	Los Angeles	California	United States
13	24 September 2014	30 September 2014	Standard Class	DB-13660	Los Angeles	California	United States
11	01 July 2014	06 July 2014	Standard Class	TT-21220	Los Angeles	California	United States
12	13 September 2014	18 September 2014	Second Class	JM-15865	Los Angeles	California	United States
14	11 November 2013	14 November 2013	Second Class	AI-10855	Los Angeles	California	United States

The newly created measures can be seen in the respective tables.

The screenshot shows a table with measures. The 'Product ID' column is highlighted. The measures listed are Profit, Quantity, Region, Sales, Ship Date, Ship Mode, Shipping Cost, State, and Total Sales. The 'Total Sales' measure is highlighted at the bottom.

Measures:

Average Discount rate:

It calculates the average for the entire discount column.

```
1 Avg Discount Rate = AVERAGE(SalesData[Discount])
```

Average Profit Margin:

Profit Margin is the percentage of profit compared to the revenue. It can be calculated by this formula Profit / Sales.

```
1 Avg Profit Margin = AVERAGEX(SalesData, SalesData[Profit] / SalesData[Sales])
```

Average Shipping cost: The mean of the shipping cost, which will show the average value for an order.

```
1 Avg Shipping Cost = AVERAGE(SalesData[Shipping Cost])
```

Customer Lifetime Value:

This is the total amount of money, which the business can expect from a customer.

```
1 CLV =  
2 CALCULATE(  
3     AVERAGEX(  
4         VALUES(SalesData[Customer ID]),  
5         SUMX(  
6             FILTER(SalesData, SalesData[Customer ID] = EARLIER(SalesData[Customer ID])),  
7             SalesData[Profit]  
8         )  
9     )  
10 )
```

Average Fulfilment Time: The average amount of time that took to deliver an order.

```
1 Avg Fulfillment Time =  
2 AVERAGEX(  
3     SalesData,  
4     DATEDIFF(SalesData[Order Date], SalesData[Ship Date], DAY)  
5 )
```

Repeat Customers: This measure shows how many customers purchased more than once.

```
1 Repeat Customers =  
2 COUNTROWS(  
3     FILTER(  
4         SUMMARIZE(  
5             SalesData,  
6             SalesData[Customer ID],  
7             "PurchaseCount", COUNTROWS(SalesData)  
8         ),  
9         [PurchaseCount] > 1  
10    )  
11 )
```

Calculated Columns:

These are the three calculated columns created by Dax formulas, these formulas separate the day, month, and year from the date. For instance, if the date is 26/12/2023, the day = 26, month = 12 and year = 2023

```
Day = DAY(OrderDate[Order Date])
```

```
Month = MONTH(OrderDate[Order Date])
```

```
Year = YEAR(OrderDate[Order Date])
```

This column separates only the positive profit from the profit column.

```
Absolute Profit = IF(SalesData[Profit] > 0, ABS(SalesData[Profit]), 0)
```

Calculated Tables:

This formula is used to create a new table which will show average sales, average profit, and average quantity for each segment, which is important information.

```
1 Customer Segment Analysis =  
2 SUMMARIZE(  
3     SalesData,  
4     CustomerDetails[Segment],  
5     "Avg Sales", AVERAGE(SalesData[Sales]),  
6     "Avg Profit", AVERAGE(SalesData[Profit]),  
7     "Avg Order Quantity", AVERAGE(SalesData[Quantity])  
8 )
```

M Language:

To create a custom column using M language, in the transform data, choose the custom column option in column tab. This window will be opened where the formula can be written.

Overall Loss Column:

This is a custom column which only takes negative values of the profit and stores in overall loss.

Custom Column

Add a column that is computed from the other columns.

New column name

Overall Loss

Custom column formula ⓘ

```
= if [Profit] < 0  
  then -[Profit]  
  else 0
```

Available columns

Order ID
Order Date
Ship Date
Ship Mode
Customer ID
City
Product ID

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

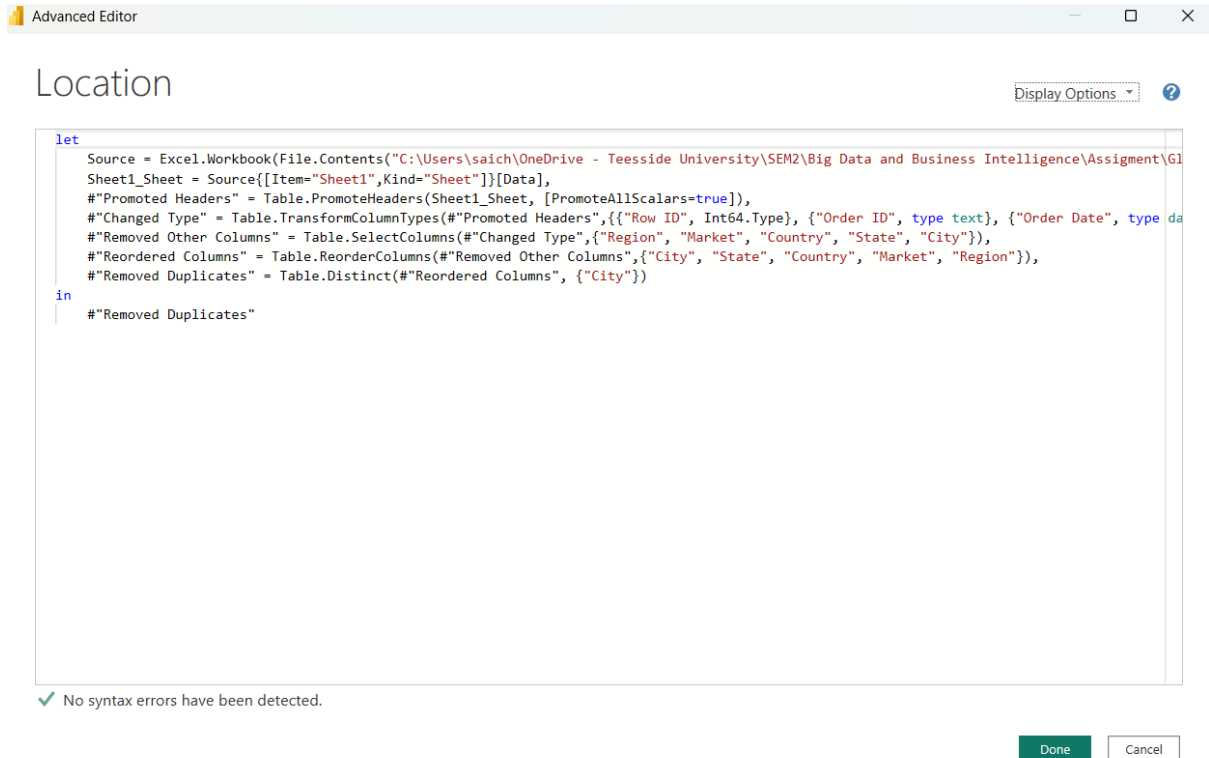
OK

Cancel

Location Table modifications using M:

All the operations such as deleting duplicates, null values and reordering the columns. Everything can be done by M language. This is an example of location table where these steps are followed.

1. Headers are promoted.
2. Datatypes are adjusted.
3. Removed other columns.
4. Reordered columns.
5. Removed Duplicates.



Appendix

Data source Configurations:

- A single excel workbook file is used instead of comma separated file, as it is more storage efficient and easier for power bi to read.
- Only three external visuals are used they are: Globe Map, Animated bar chart, and Infographics. All the rest are built in power bi visuals.

Data Sources:

Link is <https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset/data>

Glossary:

DAX: Data Analysis Expressions

M Language: Power Query language

APAC: Asia-Pacific Market

Stake Holders Report

Executive Summary

Introduction

This summary is an overview of all the findings in Global Superstore Sales Dataset. It will explore sales trends, profit trends, shipping modes, loyal customers, and top products. This analysis helps the stakeholders understand the current trends, in-depth analysis of their customers and products, the top locations. This will aid the stakeholders to make informed decisions and implementing their future strategies.

This report answers all these questions:

1. What are the total sales and profit for the selected period?
2. What are the top performing locations in terms of sales?
3. What is the sales distribution among the different customer segments?
4. What is the contribution of each product category to the total sales and which sub-categories have the highest growth?
5. What is the average shipping cost for each shipping mode? How does the choice of shipping mode impact shipping costs and order priority?
6. What is the correlation between discount and profit?
7. How Artificial Intelligence helps in Global Superstore Sales?

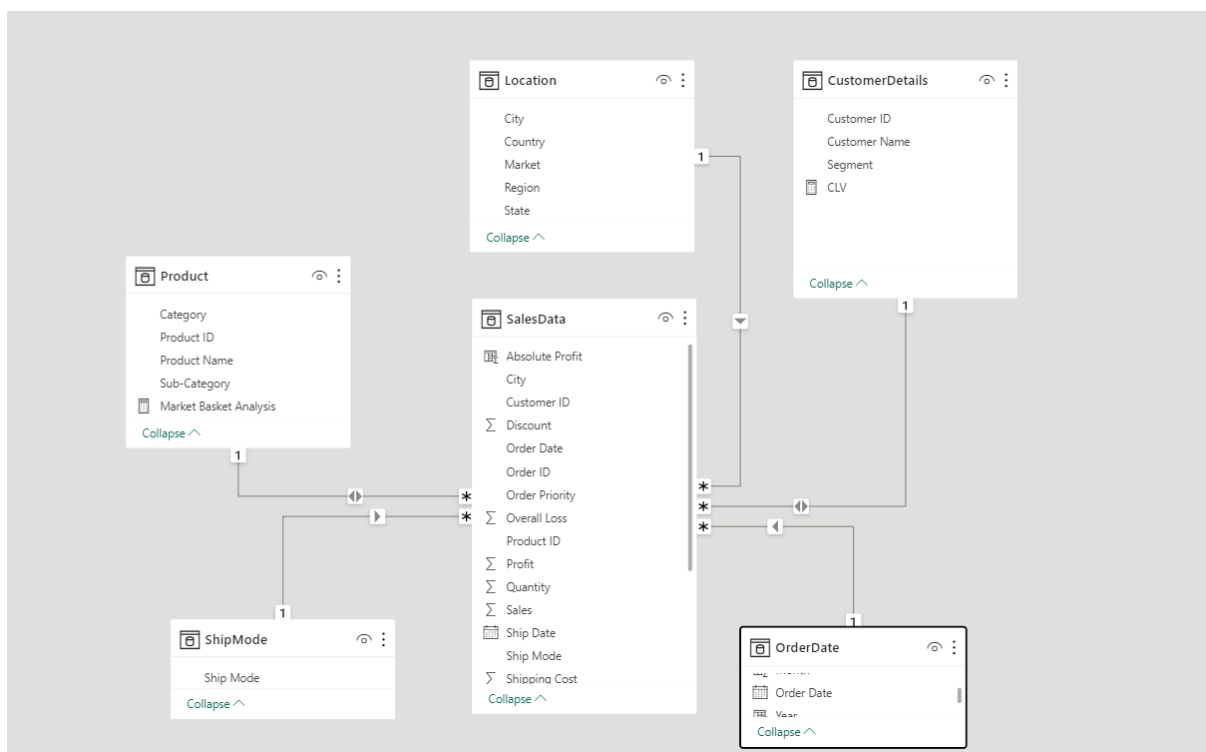
Key Findings

- The company has seen consistent growth over the years. In the next few years, the similar growth will be seen in all aspects including sales, profit, orders, and customers.
- United States is the top performing country in terms of sales and profit. US market and Asia markets are top markets with almost same sales.
- Top three cities in terms of sales are New York, Seattle, and Los Angeles noticeably all are from US. The central regions perform well in all the markets.
- This business can expect 923 dollars from each customer as long as the business relationship exists (Customer Lifetime Value). Almost all the customers are making repeated purchases from this store.
- The highest sales and profit are by consumer customer segment and the least sales are by home office. But, on the average they are almost same.
- The leading product category is Technology, the main sub-categories with top sales in this category are phones and copiers. Among top five highest selling products four are phones.

- All the categories are seeing a growth in the past years. Comparatively the technology is seeing higher growth in terms and sales and profit.
- It takes a shipping cost of 26.38 dollars and 4 days to deliver the order on an average. The total amount spent on shipping is 1.35 million dollars which is better for 25,000 orders.
- The highest sales came from standard mode shipping and least from same day shipping. On average, it takes 5 days to ship orders in standard mode. The business is doing a good job here maximizing the sales while minimizing the shipping costs.
- The average discount percentage is 14 percent and average profit margin is 5 percent. It is clearly visible that more discounts are leading to less profits.
- The forecast tells us that in the next two years the company's sales will increase at the constant pace.

Data Model

This is the completed data model. Star schema is used here to merge tables. All dimension tables are smaller chunks of original table. This schema improves the efficiency and enables faster fetching, in addition to neat organization of data.



Findings based on Visualization.

Question 1: What are the total sales and profit for the selected period?

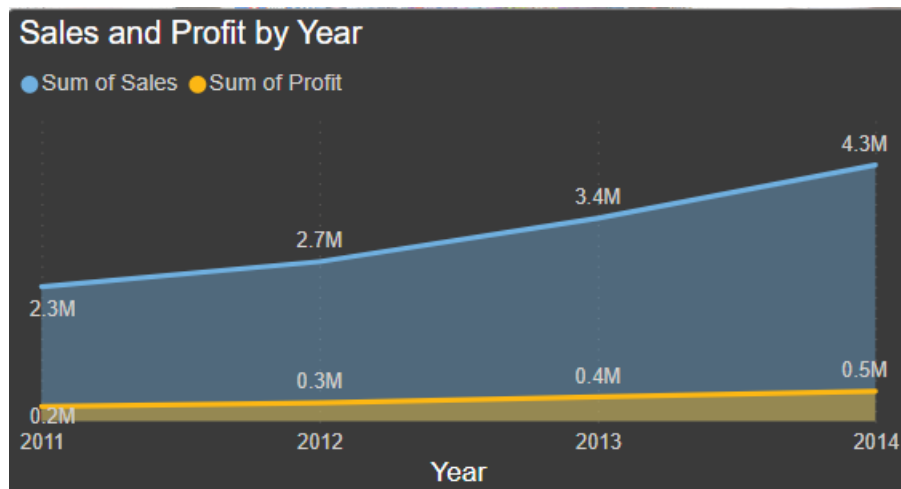
This is the overview dashboard of this application.



A slicer is created for selecting a year from the dropdown. The sales and absolute profit for the selected period will be displayed.



The trend of sales and profit can be observed by a line chart.



This is good for all the stakeholders; the sales and profit are increasing consistently across the year line.

Question 2: What are the top performing locations in terms of sales?



This dashboard shows entire sales data related based on the location. There is a country slicer. Select the slicer and lot of details will be shown regarding that country.

The animated globe rotates to United states. The top 10 cities from the country are shown by animated bar chart. All the markets and regions are shown by Tree map and total sales are clearly shown by Card and donut chart.

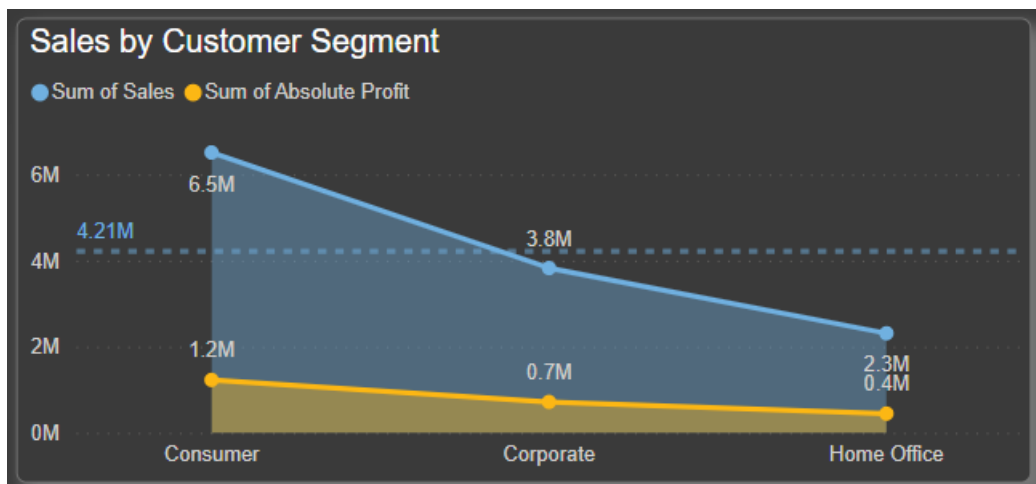


Overall, United States is the top performing country with 2.32 million. The majority sales contribution is from New York City, Seattle, and Los Angeles. All three are from the US. Stakeholders are getting lot of sales from US. They can focus on introducing new stores in this country and can increase the profits.

Question 3: What is the sales distribution among the different customer segments?



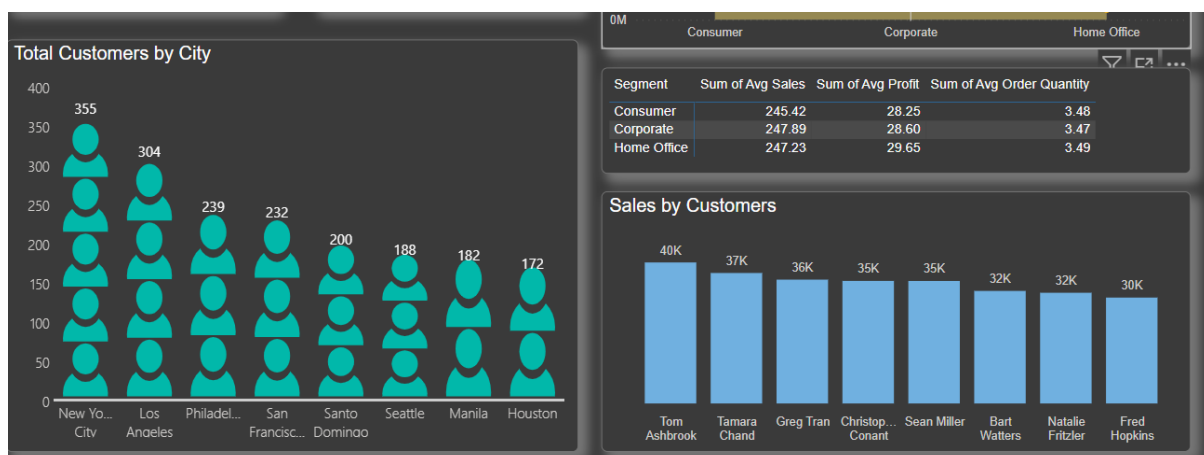
This is customer dashboard where all the detailed information of customers and sales are provided in detail.



There are three customer segments consumer, corporate and Home Office. The highest sales and profit are from consumer segments which equals 6.5M and 1.2M respectively. The second and third highest sales are Corporate and Home Office. The average is shown by a line which is 4.21M.

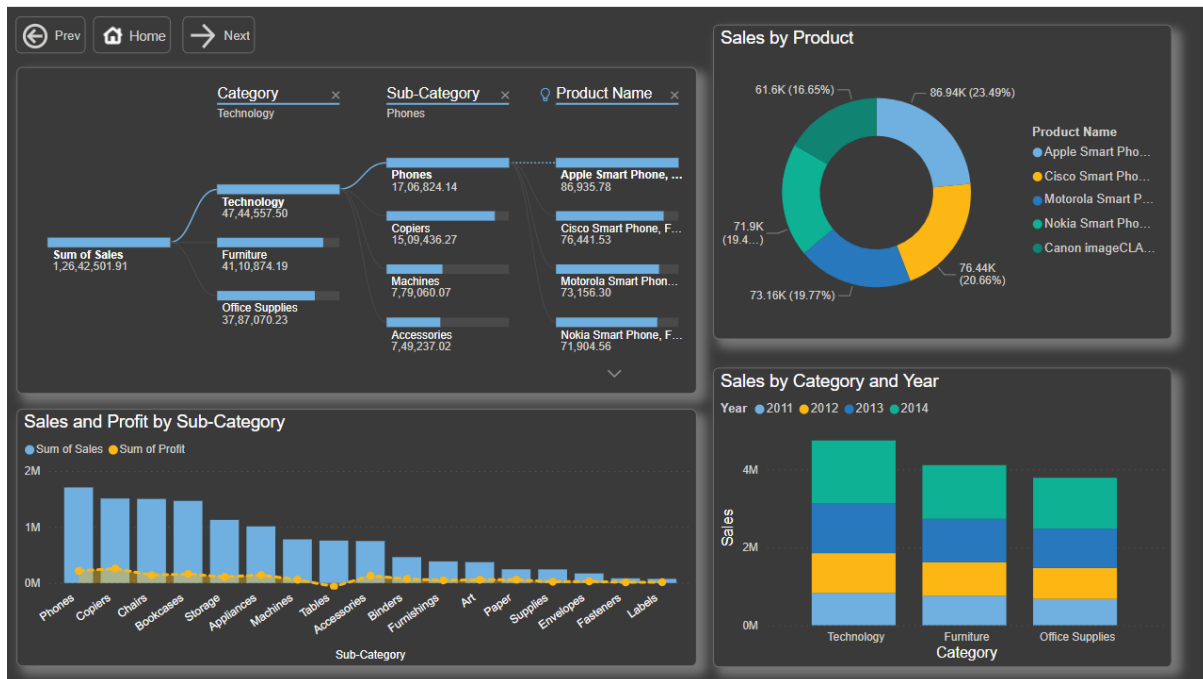
Segment	Sum of Avg Sales	Sum of Avg Profit	Sum of Avg Order Quantity
Consumer	245.42	28.25	3.48
Corporate	247.89	28.60	3.47
Home Office	247.23	29.65	3.49

On average, all segments perform well, the average sales are around 245 and average profit is around 28.

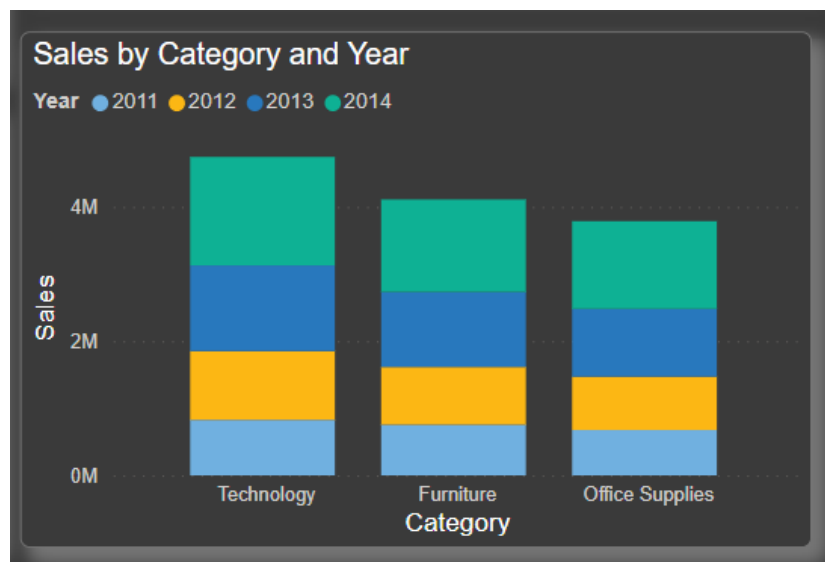


More information on customers like the which city consists high customers and also, all the top buying customers are shown here.

Question 4: What is the contribution of each product category to the total sales and which sub-categories have the highest growth?

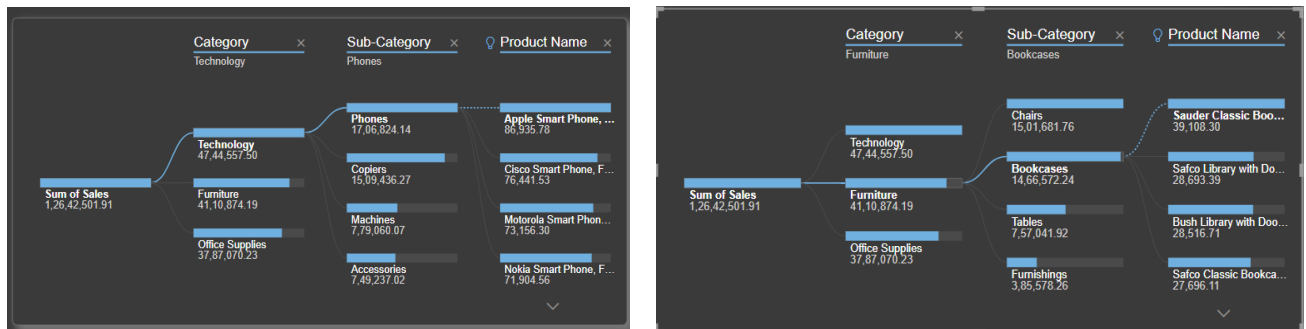


This is the product dashboard where all the categories, sub-categories and product details are shown.

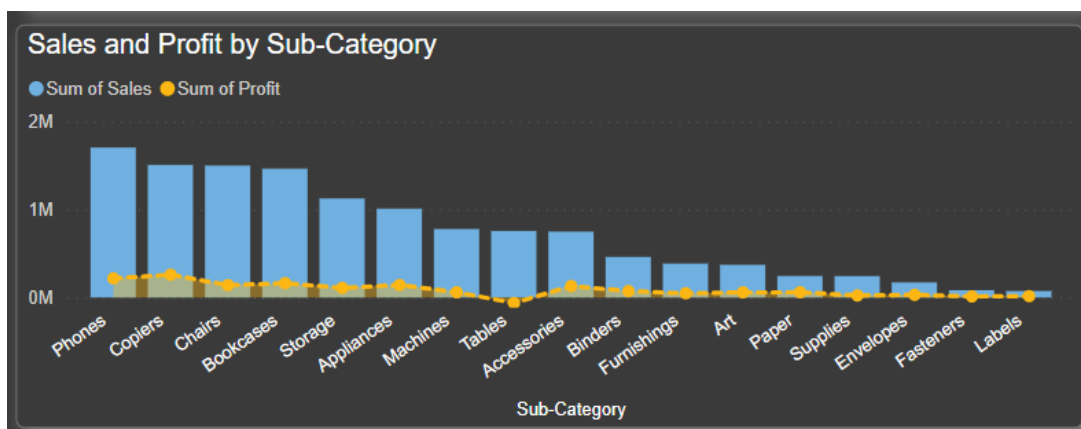


There are three product categories technology, furniture, and office supplies. The highest among them is Technology leading all sales from 2011 to 2014. The second and third are Furniture and Office supplies.

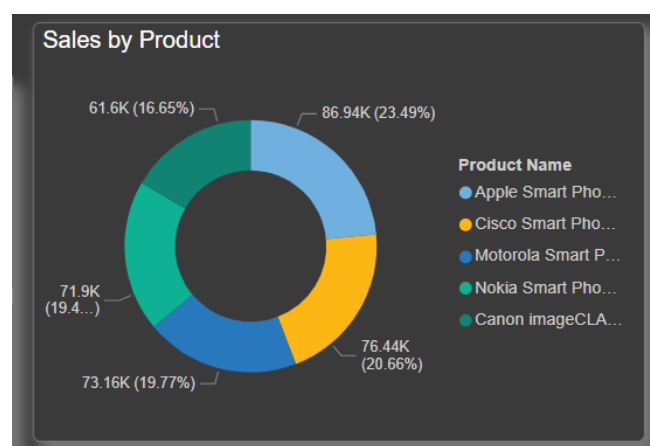
To perform deeper analysis, a decomposition tree is used which shows all the categories, sub-categories and products arranged with sales.



If the stakeholders want to know which bookcases are sold more, then they can find that easily as shown in the above picture.



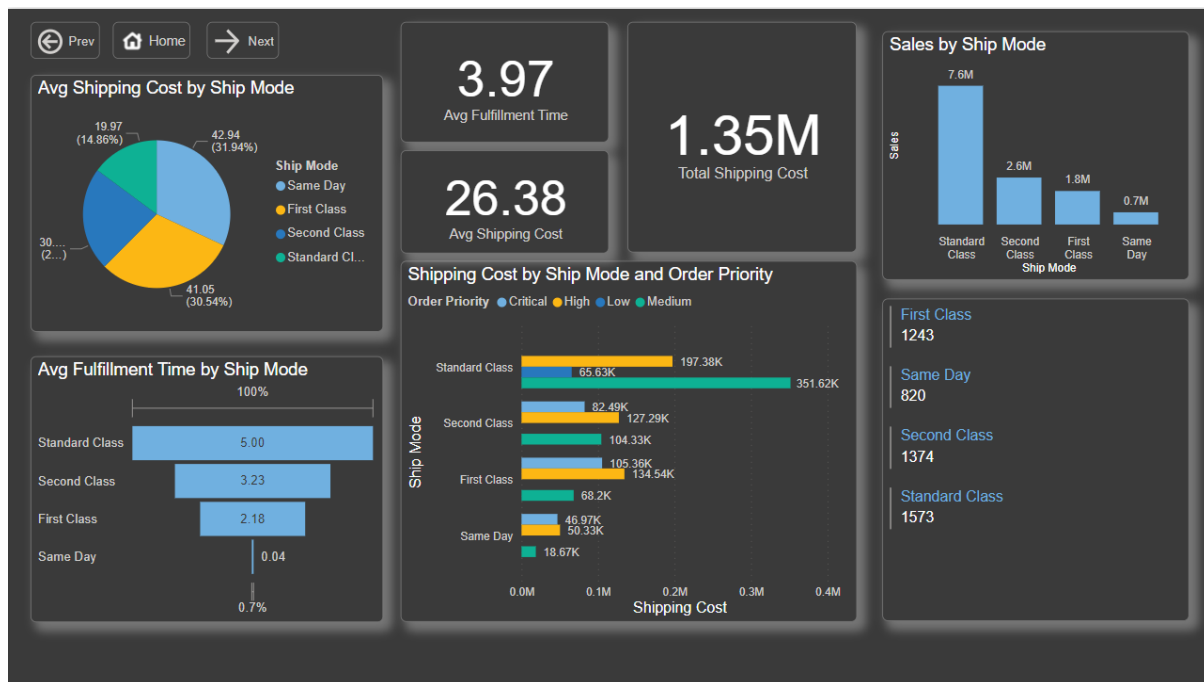
The top selling sub-categories are Phones and the next two are Copiers and bookcases. The same sub-categories are leading in profit as well. Only sub-category which has more sales, and a loss is Tables category.



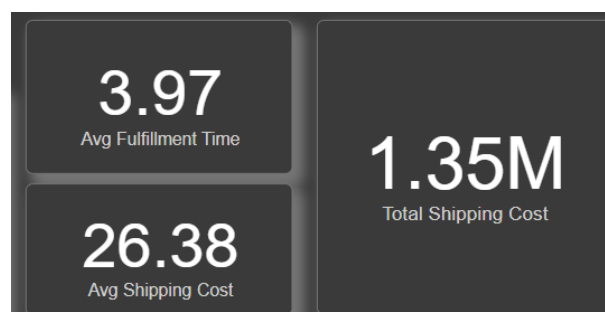
The same applies to the products. As expected, in top 5 selling products 4 are phones and one product is copier. Specifically, Apple phones are leading.

Considering all these, the technology category has the highest sales, and its subcategories phones and copiers are having highest growth, and it will be the same for the future.

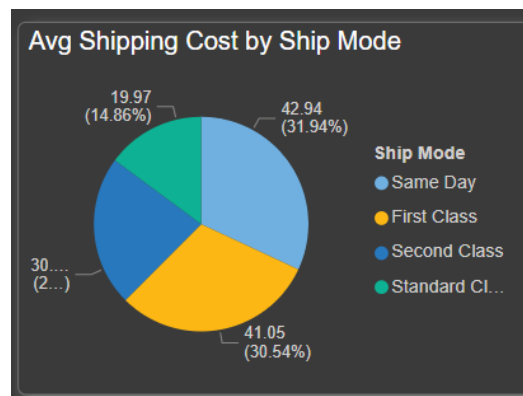
Question 5: What is the average shipping cost for each shipping mode? How does the choice of shipping mode impact shipping costs and order priority?



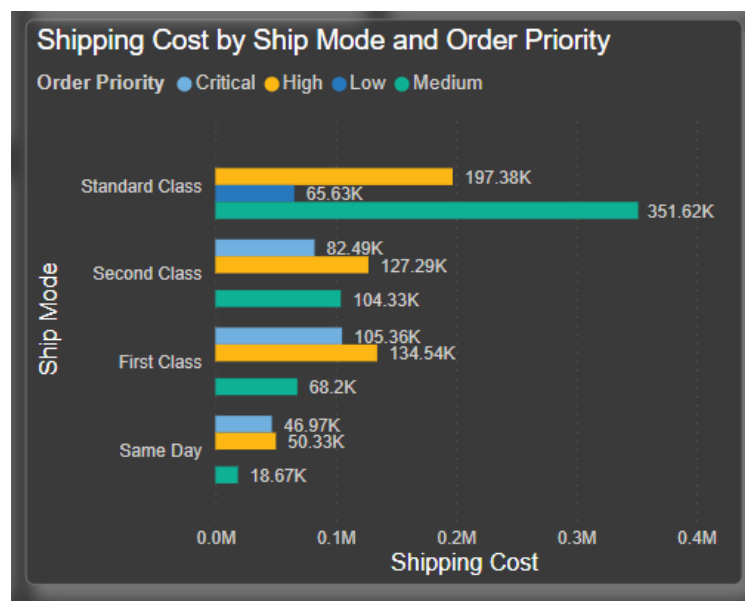
This is the shipping dashboard; it explains all the details about the shipping and mainly about the sale modes.



The average shipping cost is 3.97, It takes around 4 days for every delivery on an average. The total shipping cost is 1.35M and 26.38 cost on an average. Shipping cost measure is used to calculate this. I used the same measure to calculate the cost for all shipping modes.



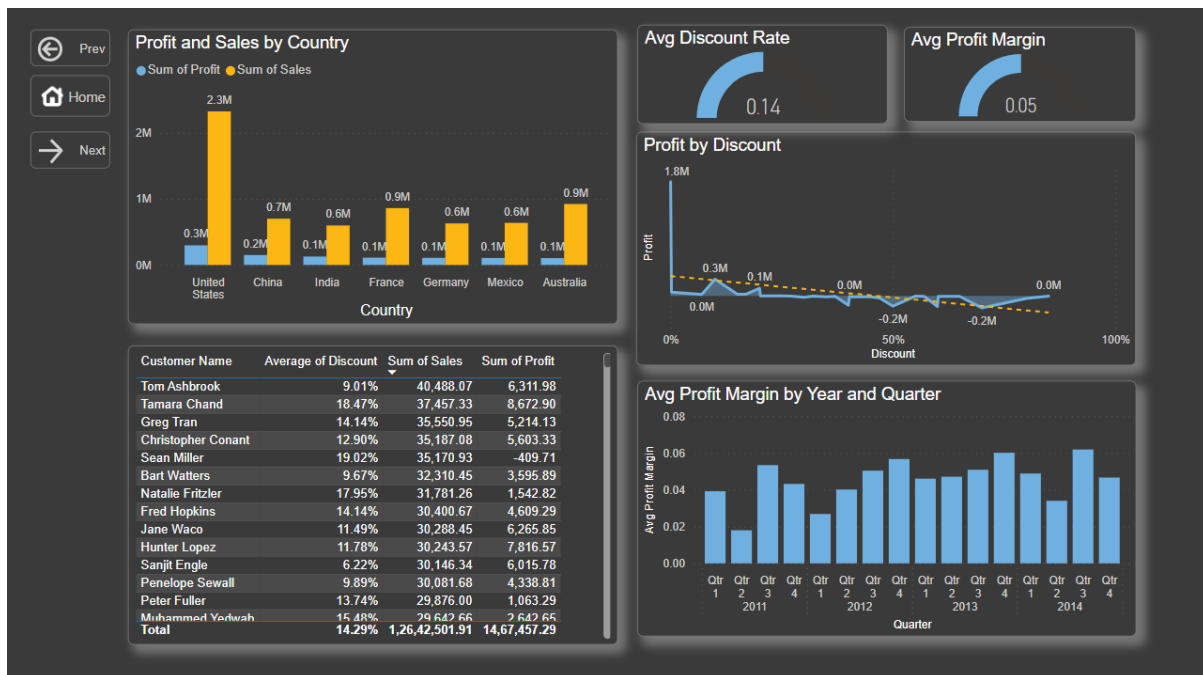
The average shipping costs for all shipping modes are given. The highest and lowest average shipping costs are for Same day with 42.94 dollars and standard delivery with 19.97 dollars respectively.



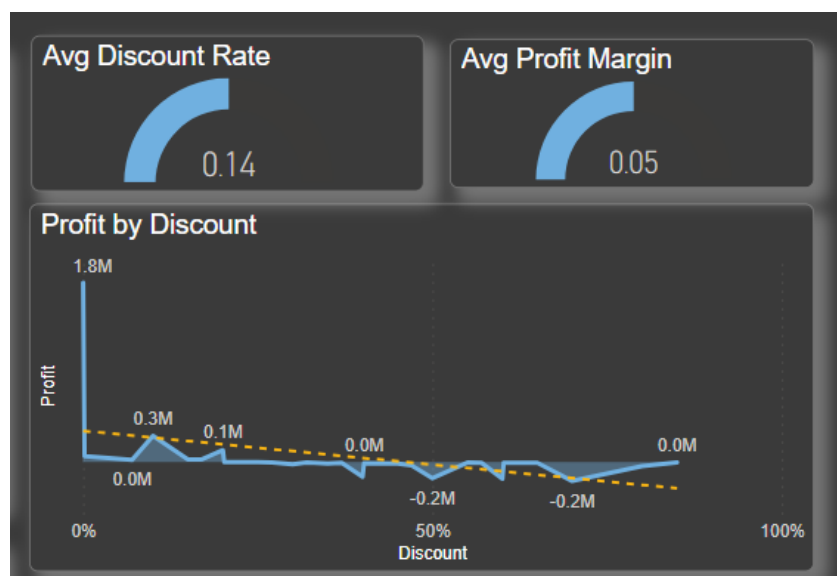
This superstore is doing the good job with the shipping, Standard shipping costs less than all, they are using this shipping method for all priority orders except the critical ones of you see the above graph.

All critical and some emergency orders are being transported using same day deliveries which saves a lot of money for the business.

Question 6: What is the correlation between discount and profit?



This is the discount and profit dashboard, all details are explained here.



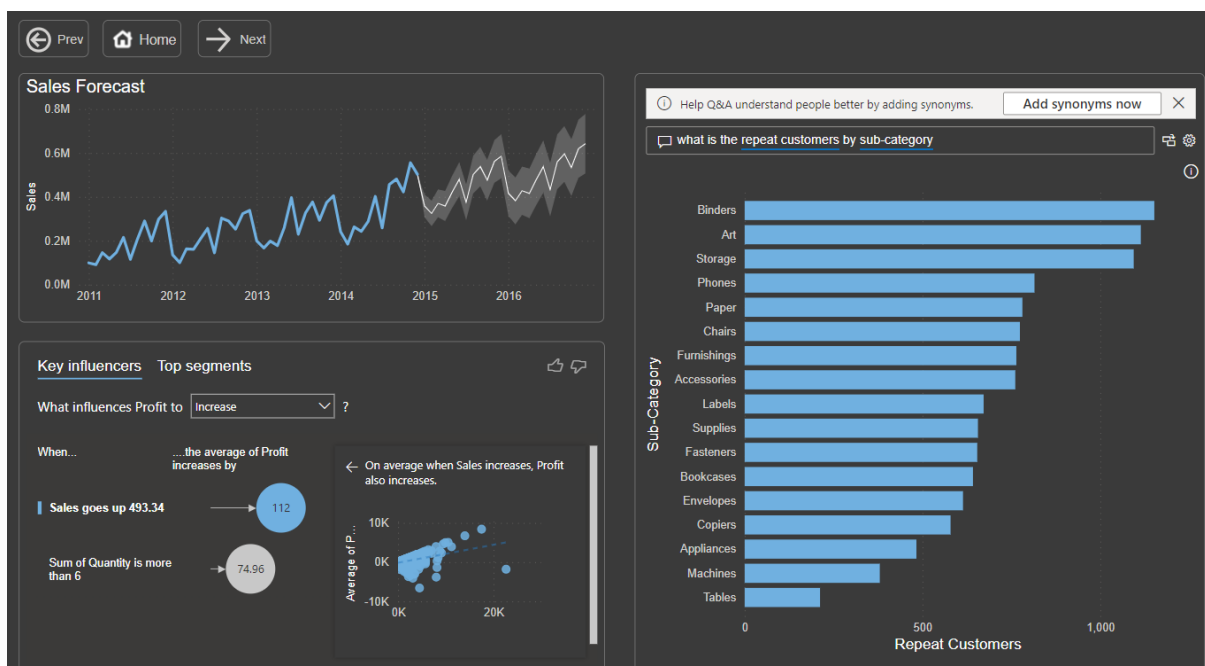
The average discount rate is 0.14 which is equivalent to 14% and average profit margin is 0.05 equal to 5%. Most of the profit is earned at 0% discount, by this it can be analysed that most of the people buy products while there are no offers, and the next phase of profit is earned at 30% discount which is at festive periods. The trend line is, as the discount decreasing the profit is decreasing.

Customer Name	Average of Discount	Sum of Sales	Sum of Profit
Catherine Glotzbach	29.91%	4,115.21	-246.84
Nicole Brennan	27.91%	5,935.08	-691.43
Craig Reiter	27.38%	20,847.54	1,435.44
Bobby Trafton	27.38%	10,396.54	-469.79
Jill Stevenson	27.31%	10,141.33	-161.19
Bill Overfelt	26.05%	8,092.49	415.22
Daniel Lacy	25.79%	12,196.68	1,108.44
Corinna Mitchell	25.76%	15,578.56	-955.06
Joni Sundaresam	25.75%	6,802.11	11.97
Shaun Chance	25.43%	9,606.52	357.02
Carlos Soltero	25.31%	14,937.04	-1,000.08
Maureen Gastineau	25.14%	20,360.12	741.19
Pauline Johnson	24.99%	15,843.54	2,102.49
Darin Van Huff	24.91%	22,746.33	616.32
Total	14.29%	1,26,42,501.91	14,67,457.29

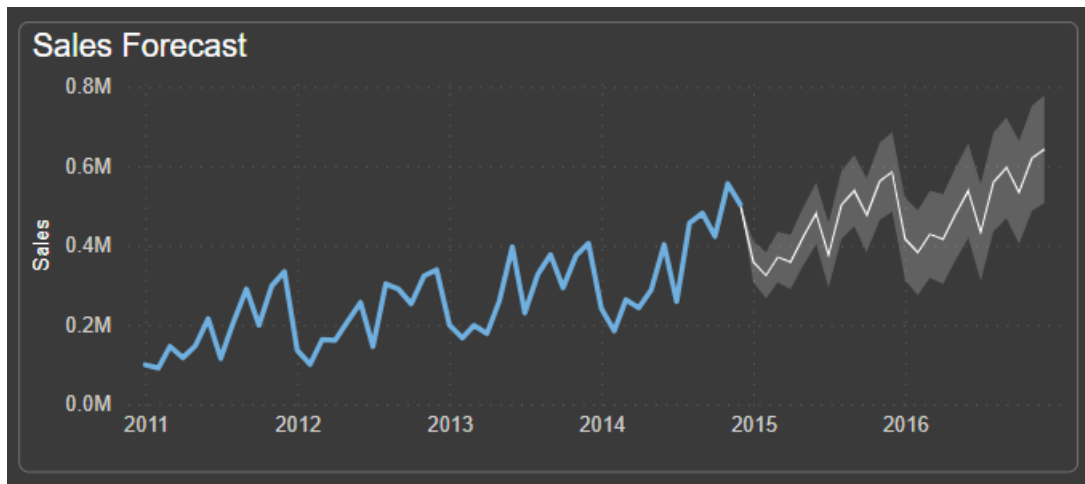
Customer Name	Average of Discount	Sum of Sales	Sum of Profit
Tom Ashbrook	9.01%	40,488.07	6,311.98
Tamara Chand	18.47%	37,457.33	8,672.90
Greg Tran	14.14%	35,550.95	5,214.13
Christopher Conant	12.90%	35,187.08	5,603.33
Sean Miller	19.02%	35,170.93	-409.71
Bart Walters	9.67%	32,310.45	3,595.89
Natalie Fritzler	17.95%	31,781.26	1,542.82
Fred Hopkins	14.14%	30,400.67	4,609.29
Jane Waco	11.49%	30,288.45	6,265.85
Hunter Lopez	11.78%	30,243.57	7,816.57
Sanjit Engle	6.22%	30,146.34	6,015.78
Penelope Sewall	9.89%	30,081.68	4,338.81
Peter Fuller	13.74%	29,876.00	1,063.29
Muhammed Yerdwah	15.48%	29,642.66	2,642.65
Total	14.29%	1,26,42,501.91	14,67,457.29

This table is used to compare discount, sales, and profit. The tables are sorted with sales and profit in descending order. From this it is clear that high discounts don't mean high sales and profit. People with top sales bought more products around 10-20% discounts.

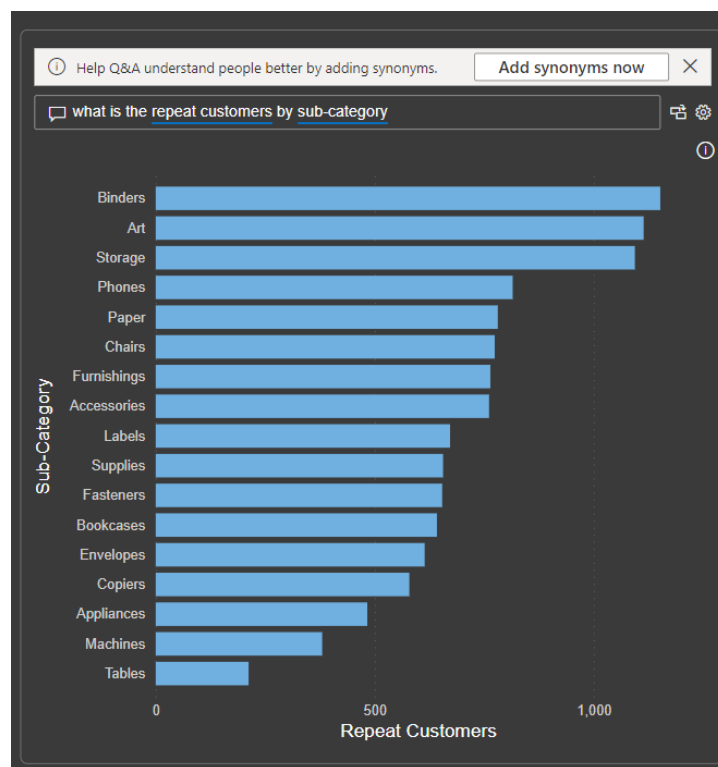
Question 7: How Artificial Intelligence helps Global Superstore to increase Sales?



Artificial Intelligence is the trending technology nowadays. All the major companies are using AI in various technologies, this is the best opportunity to use AI in business models.

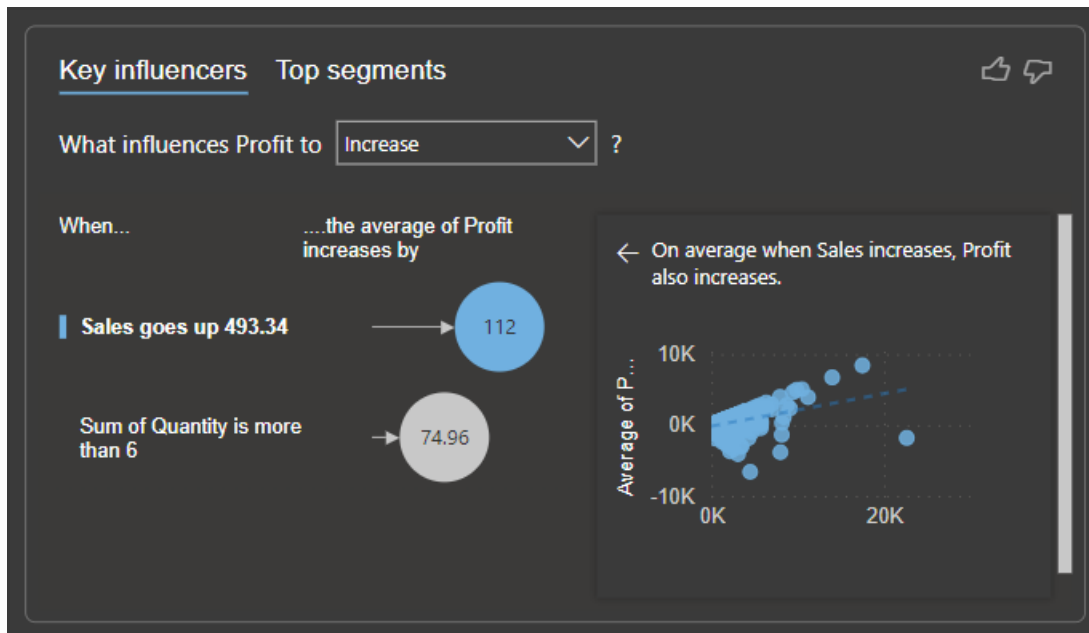


This is the forecast feature in Power bi. For the next two years, the sales trend seems promising. The sales are very high in quarter 3 and quarter 4.



Ask a Question:

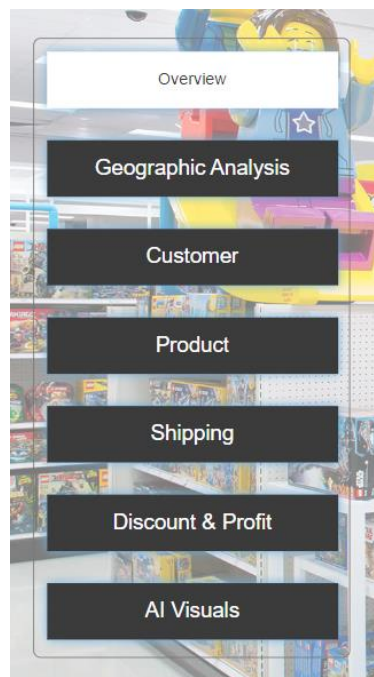
This is one more AI feature, users can ask various questions and the plots will generate by AI.



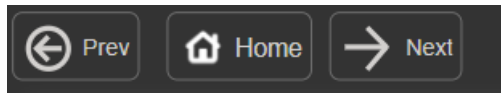
This is key influencers visual. Users can set the analyse to any column and explain sales by other columns. Users can view the detailed analysis, top segments and all the influencing factors.

Navigation

Navigation is made easy for this application. All the dashboards are listed out in the home page.



In every page simple navigation is enabled for improving user experience.



Recommendations

- As the sales are on increasing trend, it will be best for the company to expand their business in various regions. This will still boost the profits with good advertising.
- The Asian-Pacific market (APAC) is second highest in sales and profits, they should focus more on these parts as Asia has larger population, it may overtake US market in future.
- By adding more product categories and sub-categories like fashion and clothing may attract more audience.
- It is better not to give higher discounts over 20% as it may result in loss. The company is getting more sales with 0% and 10% discounts.
- The profit margin and sales are higher in the second half of the year as it is the festive period. It is better to offer festive categories and discounts of 5-10 percent to maximize the sales.

Conclusion

In conclusion, this stakeholder report provides all the necessary information and key findings for the stakeholders. From shipping locally to internationally, providing key trends of sales and profits with period, discount, and locality. This is a great resource for stakeholders to decide and explore deeper. The careful analysis and future strategies will bring the company sales to skyrocket.

References

Power BI on Microsoft Learn. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/training/powerplatform/power-bi>.

Anunaya, S. (2021). *Data Preprocessing in Data Mining -A Hands On Guide*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/>.

Sharma, M. (2018). *Star Schema in Data Warehouse modelling - GeeksforGeeks*. [online] GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/star-schema-in-data-warehouse-modeling/>.

Power BI on Microsoft Learn. [online] learn.microsoft.com. Available at: <https://learn.microsoft.com/en-us/training/paths/implement-advanced-data-visualization-techniques/>

www.youtube.com. (n.d.). *5 DESIGN TRICKS that Make EVERY Power BI Report Look GREAT!* [online] Available at: https://www.youtube.com/watch?v=v6fP8gyCLLc&ab_channel=HowtoPowerBI