ICA

BIG DATA AND BUSINESS INTELLIGENCE

NAME: AYOOLUWA DORCAS BABALOLA

STUDENT ID: B1202001

Table of Contents

SECTION 1

BUSINESS INTELLIGENCE DESIGN

1 BI Source Description and Questions

1.1 Data Source Description

The dataset used in this project is called "superstore_data" and is available for free and public use on Kaggle. The following is a link to the dataset:

superstore_data | Kaggle

The superstore data collection has about 52000 instances with 24 attributes enclosed in three tables. It is a customer-centric data set that includes information on all orders placed through various vendors and markets from 2011 to 2015 in different countries.

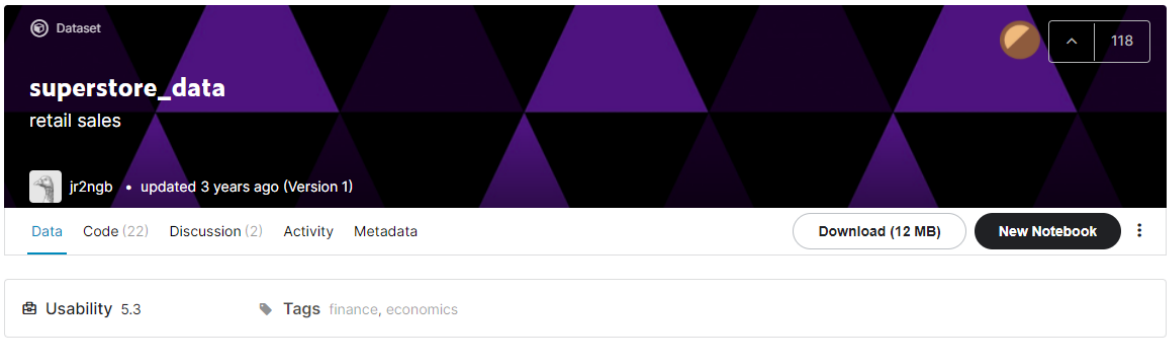


Fig: Pictorial representation of the data source

1.2 Objectives and Approach

A superstore is a major industry that relies on cutting-edge technology and data analytics to maintain its competitive advantage. It's critical that they enlist strategic tools to deliver meaningful insights into their sales in their numerous centres, broken down by product category and subcategory.

Power BI and its report were utilised in this BI project to deliver insights and the analysis focuses on the following aspects:

- predicting the superstore's sales and profit trends
- to assess the product performance
- to assess the store performance based on location

- to calculate the order cycle of orders i.e. product delivery time
- to understand customers' preferences as regards category of products
- to determine customer's most preferred ship mode

1.3 Scope

This project follows the described strategy, except in section 1 it focuses on data pre-processing and data modelling. The major goal of section 2 is to provide business insights and analytical reports utilising interactive dashboards to enhance performance evaluation based on the questions above.

1.4 Rationale

For two reasons, this dataset was chosen for this study. For starters, Power BI appears to be the best business intelligence solution for this dataset because it is both user-friendly and cost-effective. It's also the ideal tool for creating sophisticated visualisations for the superstore's sales performance analysis. The dashboards and report can be customised to meet the needs of the store and its future. Second, the dataset is commercial and contains genuine data that may be utilised to develop analytical applications in a real-world setting. This is an excellent opportunity for me to hone my business analytics skills by putting them to use while preparing for future positions.

1.5 Outcomes

These skills would have been fully shown by the end of this project:

- Data pre-processing and cleaning such as importing data, generating new columns, altering data types, and deleting null values.
- Creating and managing relationships as a method of data modelling
- Power query with M language
- Use DAX to carry out computations
- Use of various visualisations and Artificial Intelligence tools to do critical analysis
- Using the data from the report to create an interactive dashboard
- Report design and publication

1.6 Description of the Dataset

The dataset has three tables, and all of them were used. In tabular form, the tables and their attributes are summarised below.

Column	Definition
Row ID	Unique ID of each row
Order ID	Unique identifier of an order
Order Date	Date when the customer placed their order
Ship Date	Actual shipping date to the customer
Ship Mode	Mode of shipping of each order
Customer ID	Customer ID number
Customer Name	Names of individual customers
Segment	Segment to which each product belongs
Postal Code	Post code of each customer
City	The city where the customer lives
State	State where the customer lives
Country	Country name of each customer
Region	Region where each customer is based
Market	Market to which the customer belongs
Product ID	Unique identifier of a product
Sub-Category	Sub-category to which each product belongs
Product Name	Unique name of each product

Sales	Sales value of each transaction
Profit	Gain value of each transaction
Quantity	The quantities of each product per transaction
Discount	Deduction on each transaction
Order Priority	Priority of each unique order
Shipping cost	Cost of shipping each unique order

The dataset is contained in a CSV file and the figure below shows a screenshot of the dataset

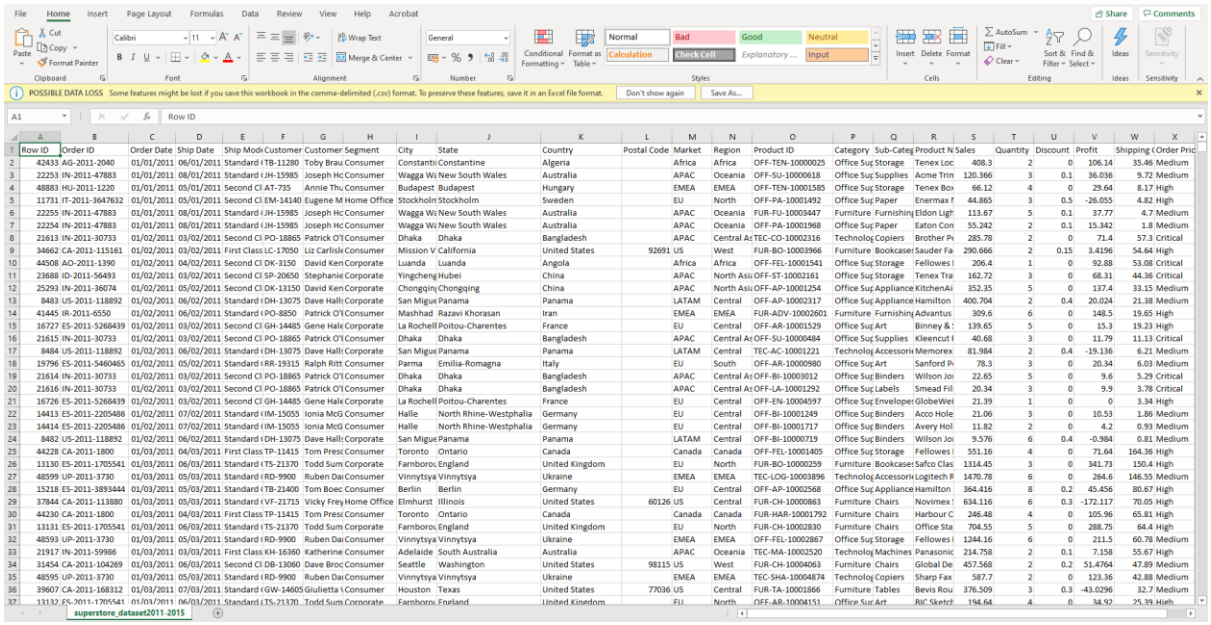


Fig: Pictorial illustration of the data file

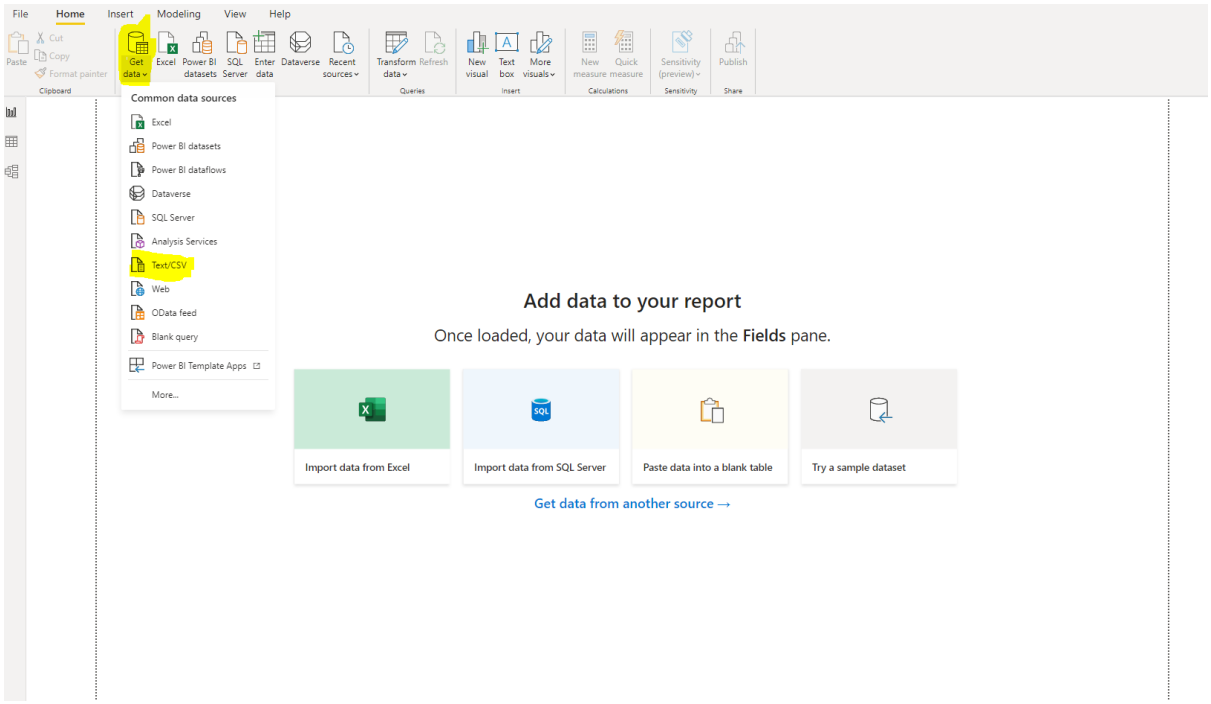
2 BI Data Pre-Processing and Data Cleansing

2.1 Loading the Data

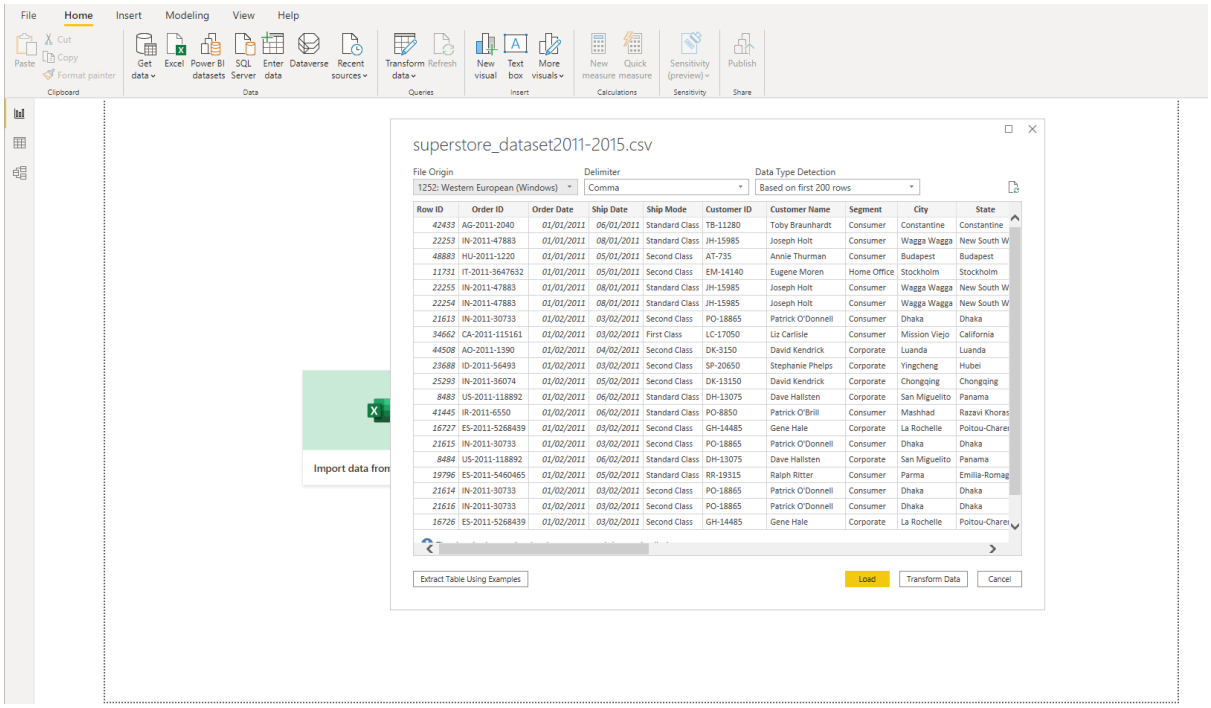
The initial step in performing data analysis and developing business intelligence is to load the raw data. Excel, csv, space separated format, the online API,

databases, and a variety of other sources and files are supported by Power BI to import datasets.

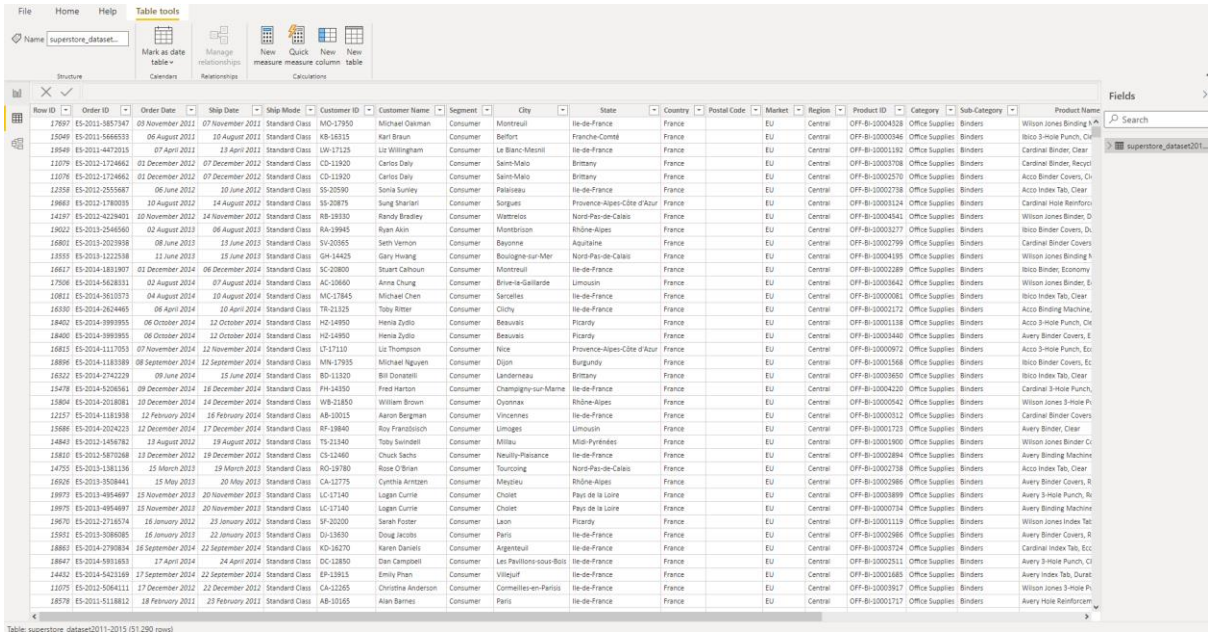
When the application is launched, a dialogue box appears, allowing you to load an existing project or use the '**Get data**' button on the '**Home**' ribbon. The diagram below depicts the framework through which we can import data and how to load it from multiple sources. The data for this dataset came from a csv file.



The data is then loaded into Power BI after the dataset has been imported. The loading of data is depicted in the diagram below.



Then the data is fully loaded



2.2 Data Pre-processing and Cleaning

Following the loading of the data, cleaning and pre-processing begins, which includes removing columns, removing null values, promoting column names,

altering data type, and more. We'll use the '**Transform Data**' button in the Home ribbon for this project, as seen in the figure below.

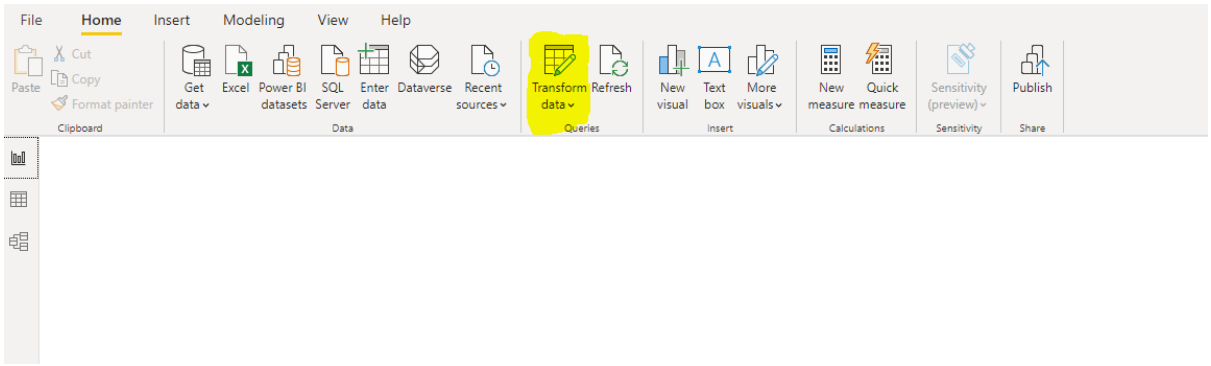
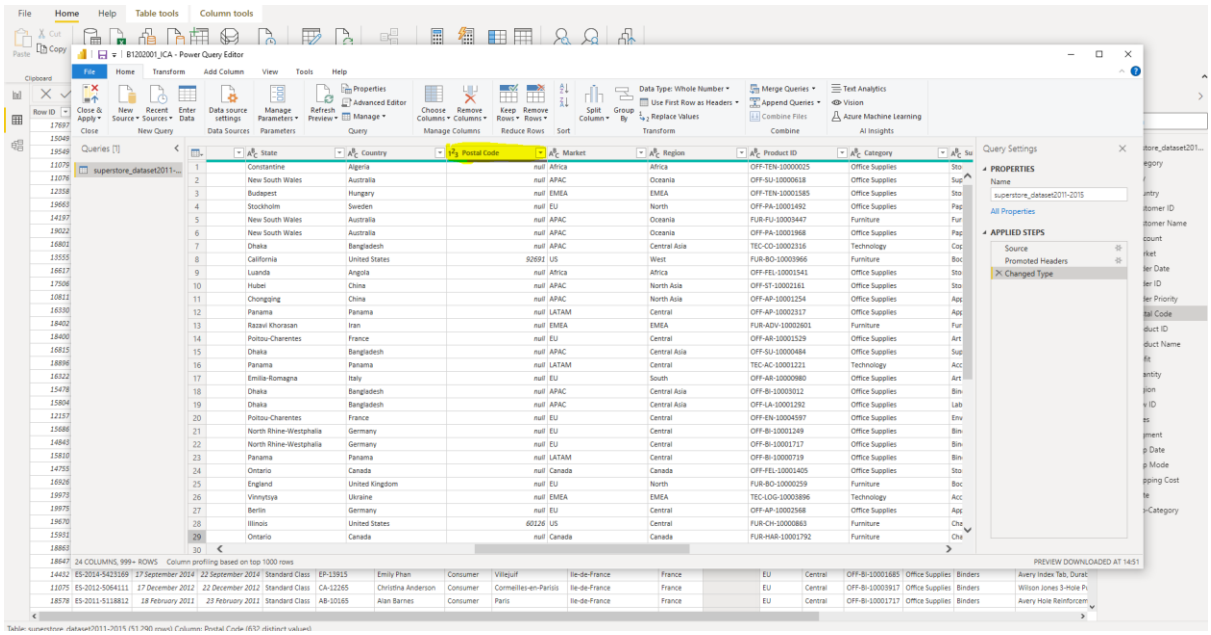


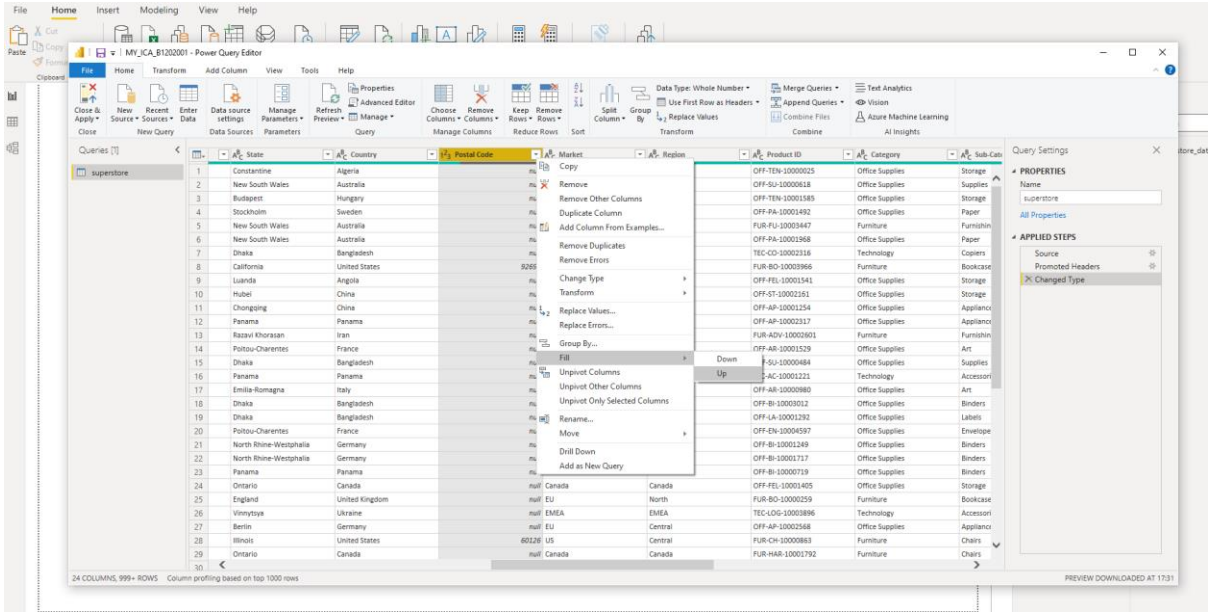
Fig: Selecting the Transform Data option

2.2.1 Removing Null Values

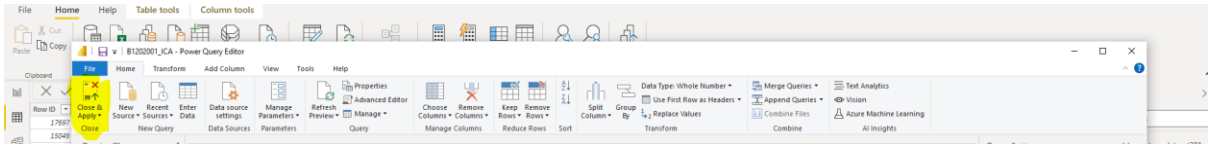
Firstly, we notice that the 'Postal code' column has many null values.



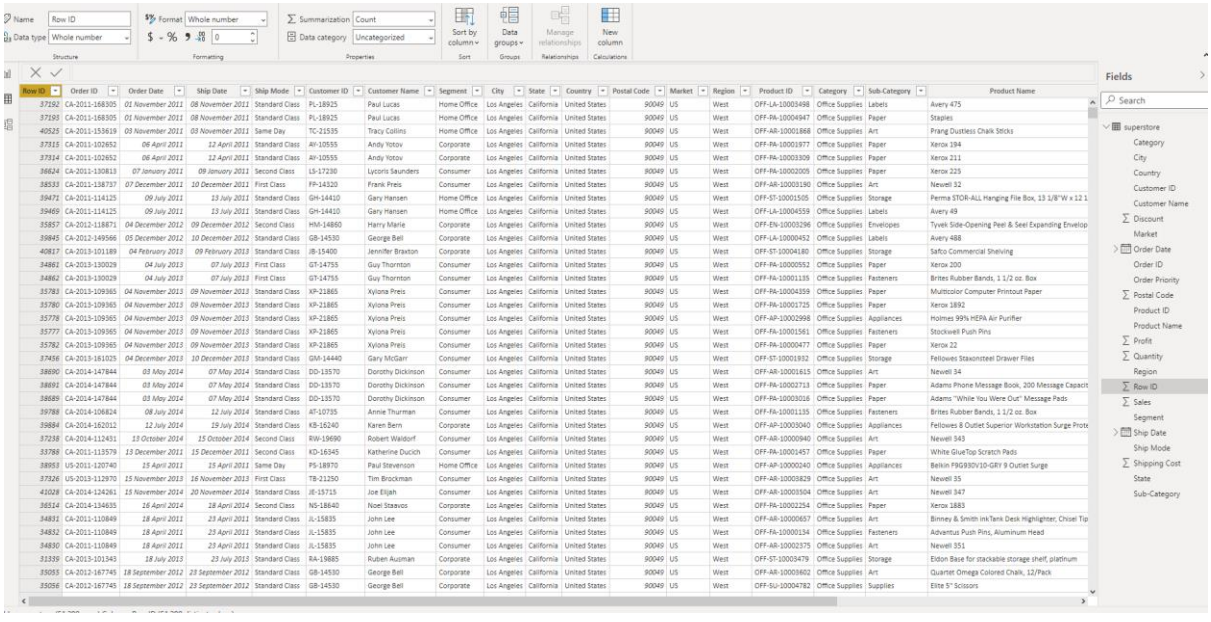
We then remove the null values by right clicking and using the '**fill up**' option as shown below



After which we **close and apply** to see the effect

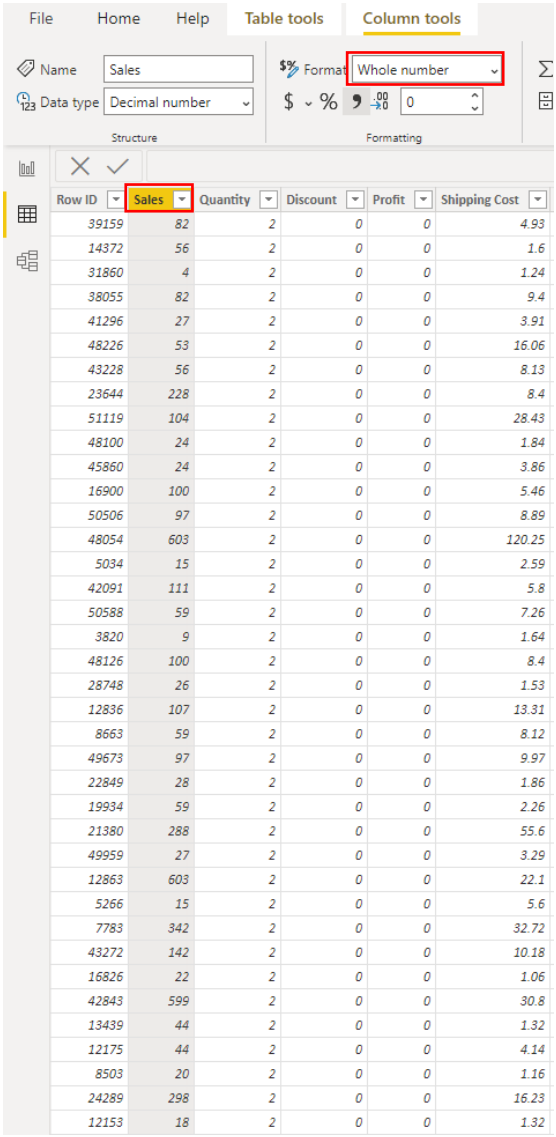


The figure below shows the data after filling up the null values



2.2.2 Changing Data Type

Changing the data type is one of the pre-processing steps that some tables require. It is critical to update the data type of currency-related columns in this dataset from just whole number. The images below show how to change the data type.



Row ID	Sales	Quantity	Discount	Profit	Shipping Cost
39159	82	2	0	0	4.93
14372	56	2	0	0	1.6
31860	4	2	0	0	1.24
38055	82	2	0	0	9.4
41296	27	2	0	0	3.91
48226	53	2	0	0	16.06
43228	56	2	0	0	8.13
23644	228	2	0	0	8.4
51119	104	2	0	0	28.43
48100	24	2	0	0	1.84
45860	24	2	0	0	3.86
16900	100	2	0	0	5.46
50506	97	2	0	0	8.89
48054	603	2	0	0	120.25
5034	15	2	0	0	2.59
42091	111	2	0	0	5.8
50588	59	2	0	0	7.26
3820	9	2	0	0	1.64
48126	100	2	0	0	8.4
28748	26	2	0	0	1.53
12836	107	2	0	0	13.31
8663	59	2	0	0	8.12
49673	97	2	0	0	9.97
22849	28	2	0	0	1.86
19934	59	2	0	0	2.26
21380	288	2	0	0	55.6
49959	27	2	0	0	3.29
12863	603	2	0	0	22.1
5266	15	2	0	0	5.6
7783	342	2	0	0	32.72
43272	142	2	0	0	10.18
16826	22	2	0	0	1.06
42843	599	2	0	0	30.8
13439	44	2	0	0	1.32
12175	44	2	0	0	4.14
8503	20	2	0	0	1.16
24289	298	2	0	0	16.23
12153	18	2	0	0	1.32

Fig: Data type of sales before changing it

We select the drop-down arrow in the format and select currency since the sales is related to it and would be better for the analysis.

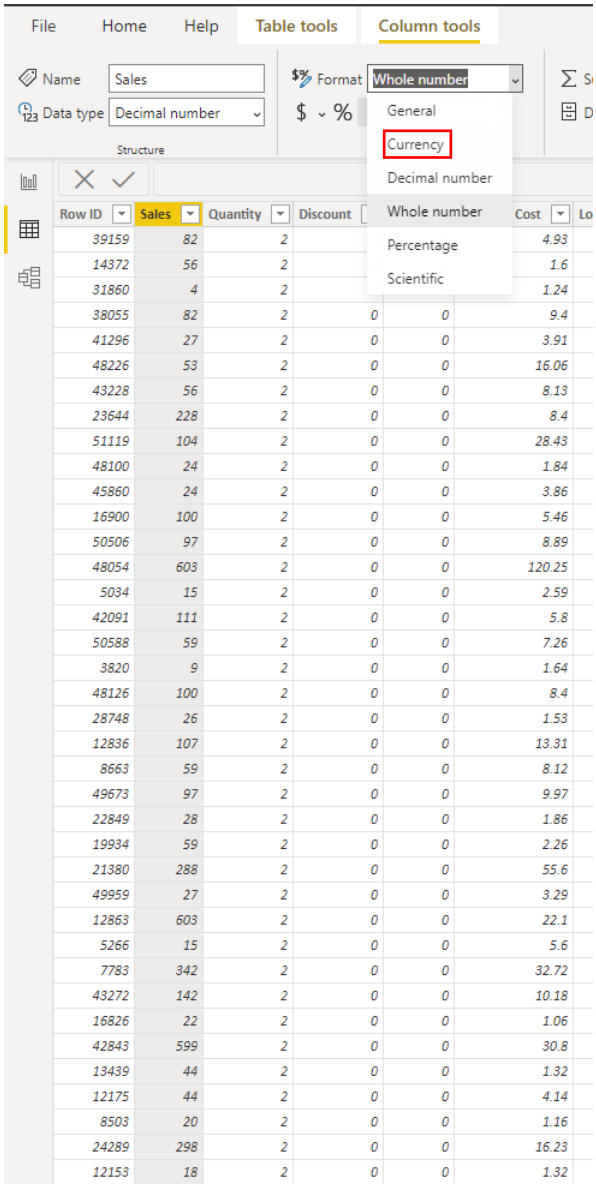
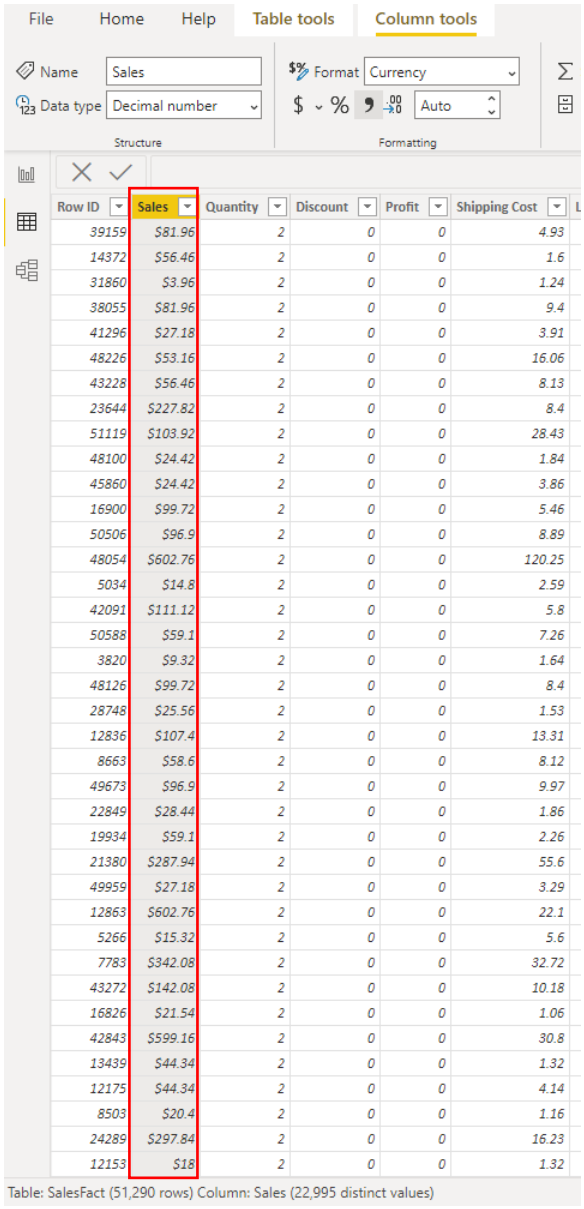


Fig: Changing the type of sales column to currency



Row ID	Sales	Quantity	Discount	Profit	Shipping Cost
39159	$81.96	2	0	0	4.93
14372	$56.46	2	0	0	1.6
31860	$3.96	2	0	0	1.24
38055	$81.96	2	0	0	9.4
41296	$27.18	2	0	0	3.91
48226	$53.16	2	0	0	16.06
43228	$56.46	2	0	0	8.13
23644	$227.82	2	0	0	8.4
51119	$103.92	2	0	0	28.43
48100	$24.42	2	0	0	1.84
45860	$24.42	2	0	0	3.86
16900	$99.72	2	0	0	5.46
50506	$96.9	2	0	0	8.89
48054	$602.76	2	0	0	120.25
5034	$14.8	2	0	0	2.59
42091	$111.12	2	0	0	5.8
50588	$59.1	2	0	0	7.26
3820	$9.32	2	0	0	1.64
48126	$99.72	2	0	0	8.4
28748	$25.56	2	0	0	1.53
12836	$107.4	2	0	0	13.31
8663	$58.6	2	0	0	8.12
49673	$96.9	2	0	0	9.97
22849	$28.44	2	0	0	1.86
19934	$59.1	2	0	0	2.26
21380	$287.94	2	0	0	55.6
49959	$27.18	2	0	0	3.29
12863	$602.76	2	0	0	22.1
5266	$15.32	2	0	0	5.6
7783	$342.08	2	0	0	32.72
43272	$142.08	2	0	0	10.18
16826	$21.54	2	0	0	1.06
42843	$599.16	2	0	0	30.8
13439	$44.34	2	0	0	1.32
12175	$44.34	2	0	0	4.14
8503	$20.4	2	0	0	1.16
24289	$297.84	2	0	0	16.23
12153	$18	2	0	0	1.32

Table: SalesFact (51,290 rows) Column: Sales (22,995 distinct values)

Fig: screenshot of the output of the new data type of sales

This method was used to change the data type of profit, shipping cost and discount too.

Row ID	Sales	Quantity	Discount	Profit	Shipping Cost
32964	$74.352	3	$0.20	$23.24	$22.49
32965	$14.04	3	$0.20	$1.58	$3.69
36997	$40.752	3	$0.20	$15.28	$2.49
32962	$4.344	3	$0.20	$0.87	$1.42
32356	$4.416	3	$0.20	$1.60	$0.31
31850	$10.824	3	$0.20	$2.57	$0.88
32555	$540.048	3	$0.20	($47.25)	$65.74
36445	$842.376	3	$0.20	$105.30	$70.22
35917	$90.48	3	$0.20	$33.93	$15.95
32803	$17.88	3	$0.20	$5.59	$1.07
40152	$21.312	3	$0.20	$7.99	$1.07
34612	$14.376	3	$0.20	$4.85	$0.98
32827	$74.352	3	$0.20	$23.24	$5.16
36401	$16.776	3	$0.20	$5.03	$2.4
33173	$470.376	3	$0.20	$47.04	$20.65
33911	$230.28	3	$0.20	$23.03	$34.42
39127	$311.976	3	$0.20	$39.00	$21.99
39125	$63.936	3	$0.20	$6.39	$4.7
39128	$50.352	3	$0.20	$17.62	$3.33
1701	$271.488	3	$0.20	$84.83	$14.2
38185	$73.344	3	$0.20	$27.50	$4.08
38036	$102.624	3	$0.20	$7.70	$7.26
38038	$13.392	3	$0.20	$3.18	$1.02
18547	$315.432	3	$0.20	($7.94)	$52.44
40424	$50.136	3	$0.20	($11.28)	$1.65
38446	$971.88	3	$0.20	$109.34	$95.63
2337	$852.48	3	$0.20	$213.12	$149.86
33291	$11.304	3	$0.20	($2.12)	$0.56
35135	$674.352	3	$0.20	($109.58)	$45.54
33640	$123.552	3	$0.20	($29.34)	$9.62
40110	$266.352	3	$0.20	$13.32	$25.36
38731	$9.144	3	$0.20	$3.09	$0.5
9179	$223.2	3	$0.20	($39.06)	$28.28
38080	$37.752	3	$0.20	$4.25	$2.8
36765	$585.552	3	$0.20	$73.19	$179.73
33839	$143.952	3	$0.20	$14.40	$7.77
7146	$57.36	3	$0.20	($11.52)	$5.66
39023	$331.536	3	$0.20	($82.88)	$33.17

2.3 Generating Fact and Dimension Tables

Superstore_data is a single flat table with 24 columns and 51290 rows. Order ID, customer ID, customer name, nationality, product name, sales, discount, profit, category, and others are among the records in this table.

This table is renamed by right clicking in the transform data option to **superstore** before generating the fact and dimension tables.

The superstore table is now our flat table. It will be difficult to handle, maintain, and create business-related questions if all the data is contained in a single flat file or table. The table is also unusually huge because it contains duplicated records and information. We therefore split it into fact and dimension tables.

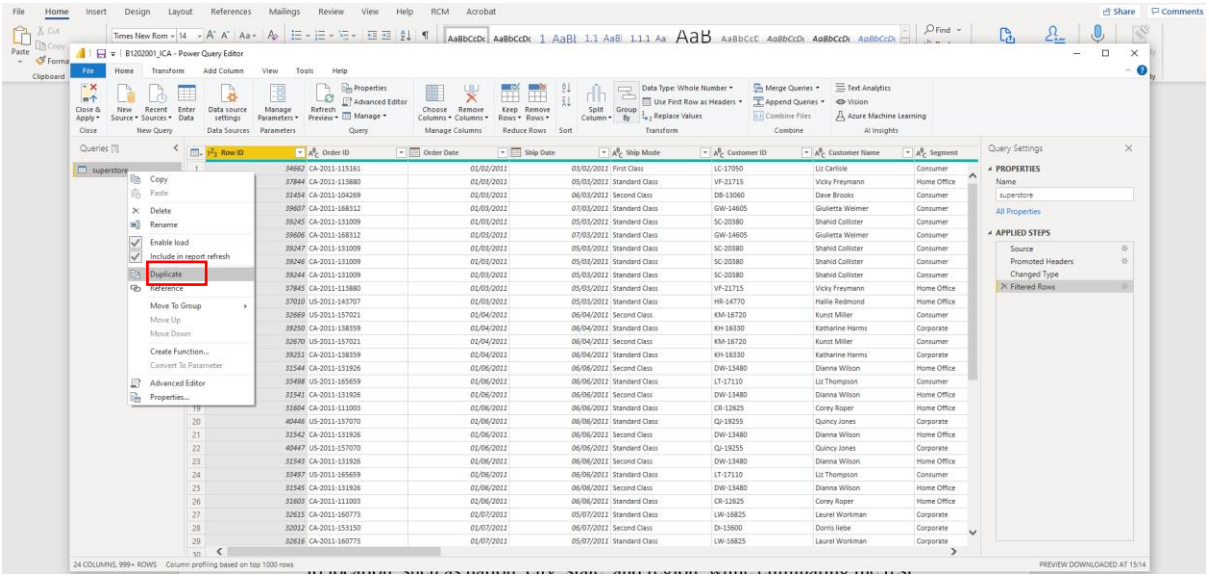


Fig: Creating duplicate table

Right-clicking on the cloned table and selecting the '**Rename**' option from the context menu renamed it into **'SalesFact'** which is the fact table for this model.

Then, we right click on superstore and delete while continuing the work with the fact table that would be formed.

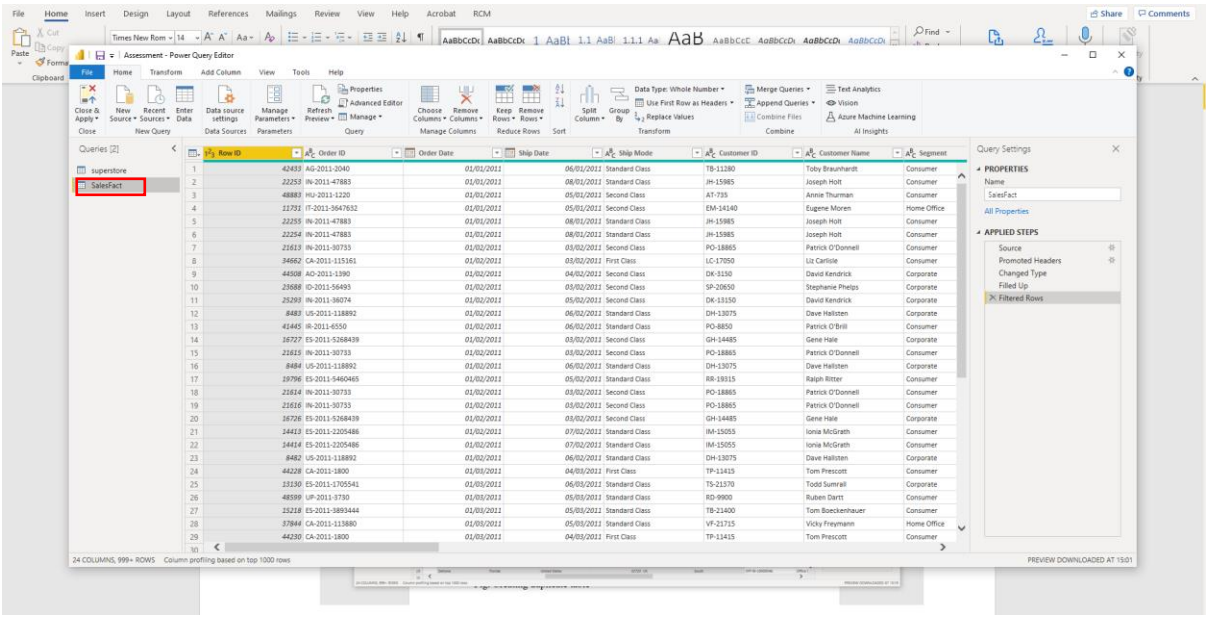


Fig: Generating the fact table

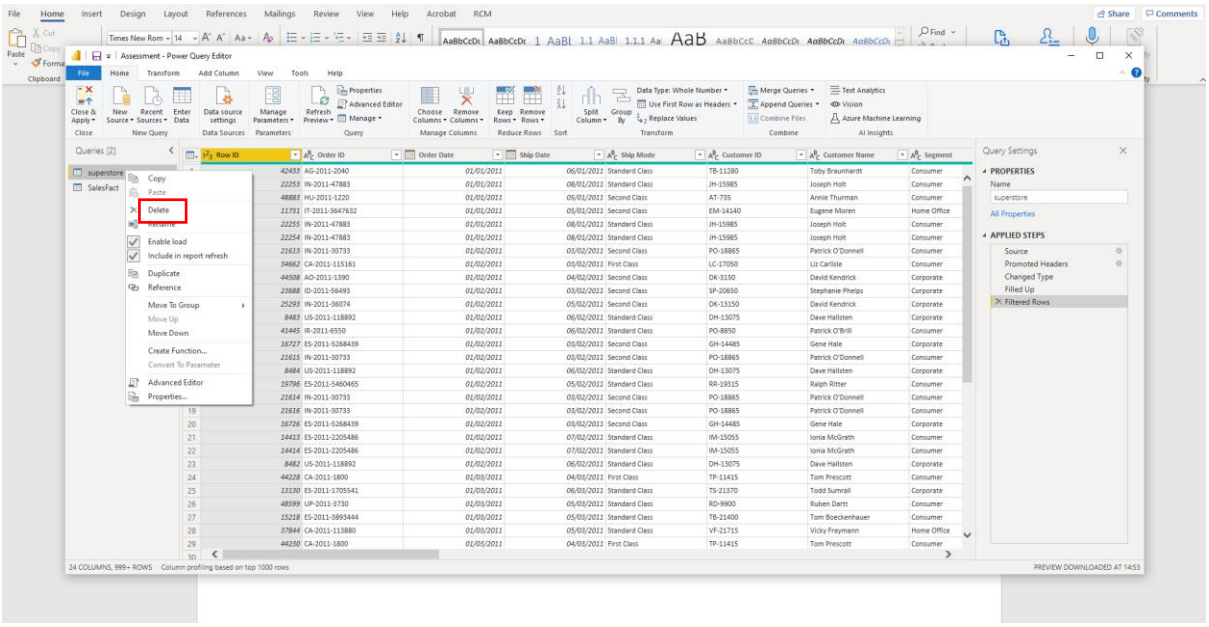


Fig: Remove the superstore table

In generating our dimension tables, we transform data and make duplicates of the fact table to select columns that would be essential for each dimension table. We go ahead to generate the first dimension table. We right-click the SalesFact table in the query and select **'Duplicate'**.

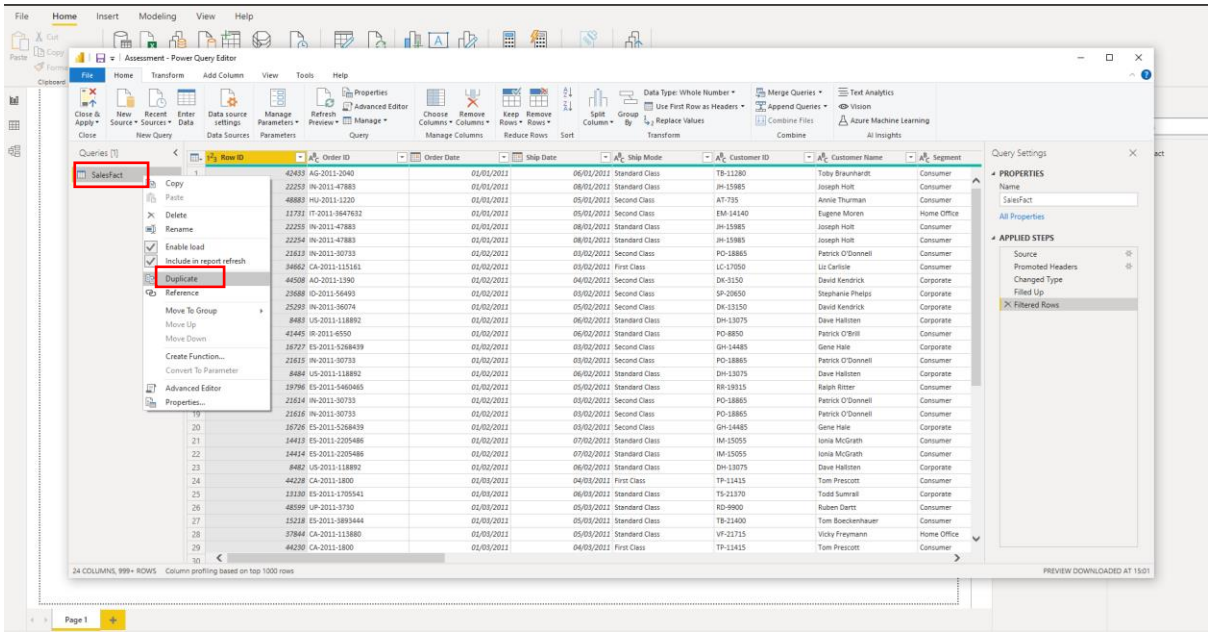


Fig: Duplicate option in the query

2.3.1 Creating Location Dimension Table

A duplicate of the original flat table was generated by right clicking the '**Duplicate**' option and rename as shown below to construct the '**LocationDim**' dimension table. Columns relating to location such as county, region, state, city and market were left in this table while other columns were removed.

To minimise ambiguity, it is critical that the dimension tables have unique rows. We selected all the columns, then right clicked and selected '**Remove Duplicates**' which removed duplicate records from the table.

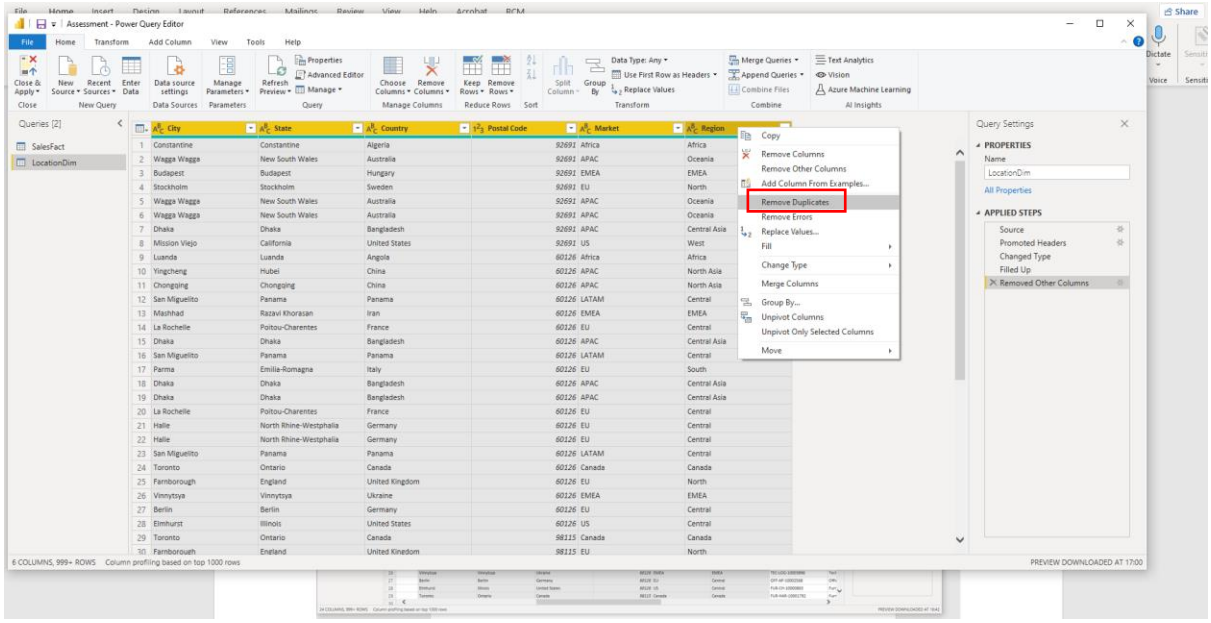


Fig: Removing duplicates from the location dimension table

There is a need for the dimension table to relate to the fact table, we there go to **'Add column'** ribbon and select index from 1 which would correspond the columns in the location dimension and have a connection with the fact table.

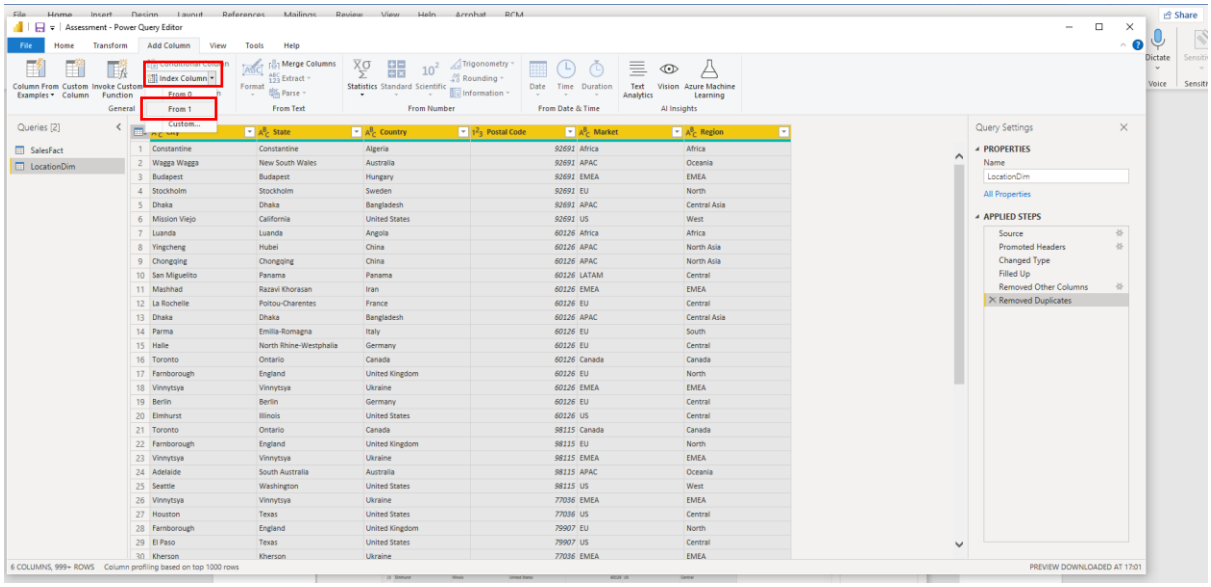


Fig: Adding a new column to the location dimension table

We go further by renaming the index column to 'LocationInd' standing for Location Index. This was regarded as the table's primary key.

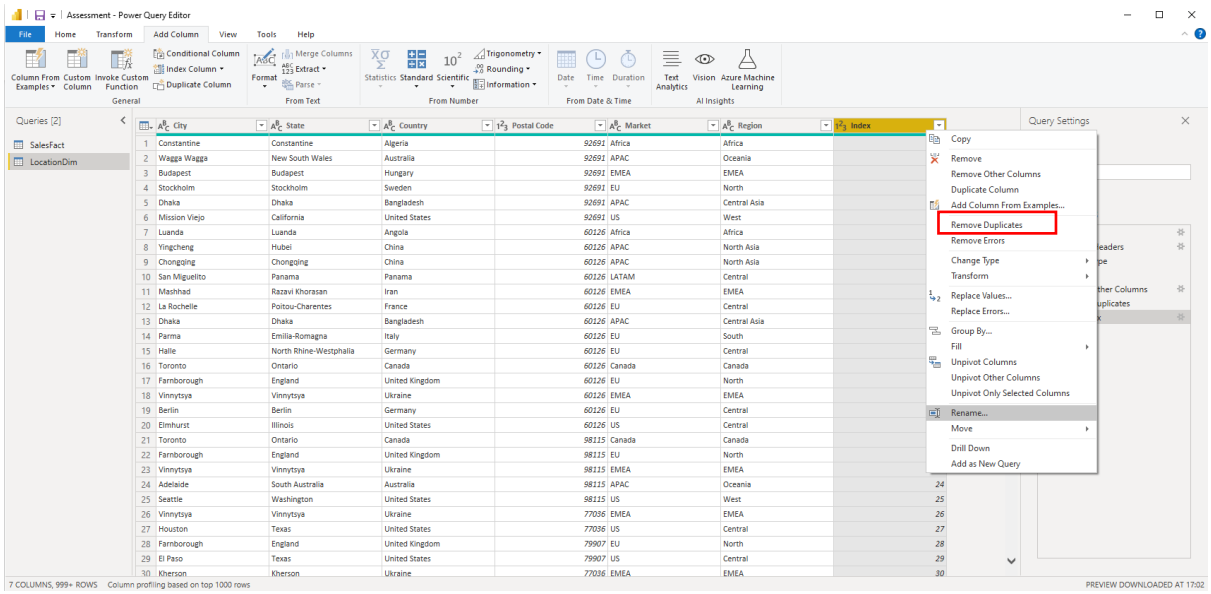


Fig: Rename the index column to LocationInd

Furthermore, we go back to the SalesFact table and merge the new dimension table 'LocationDim' to it using the **'merge queries'** option. We then select the common columns and select 'OK' to merge.

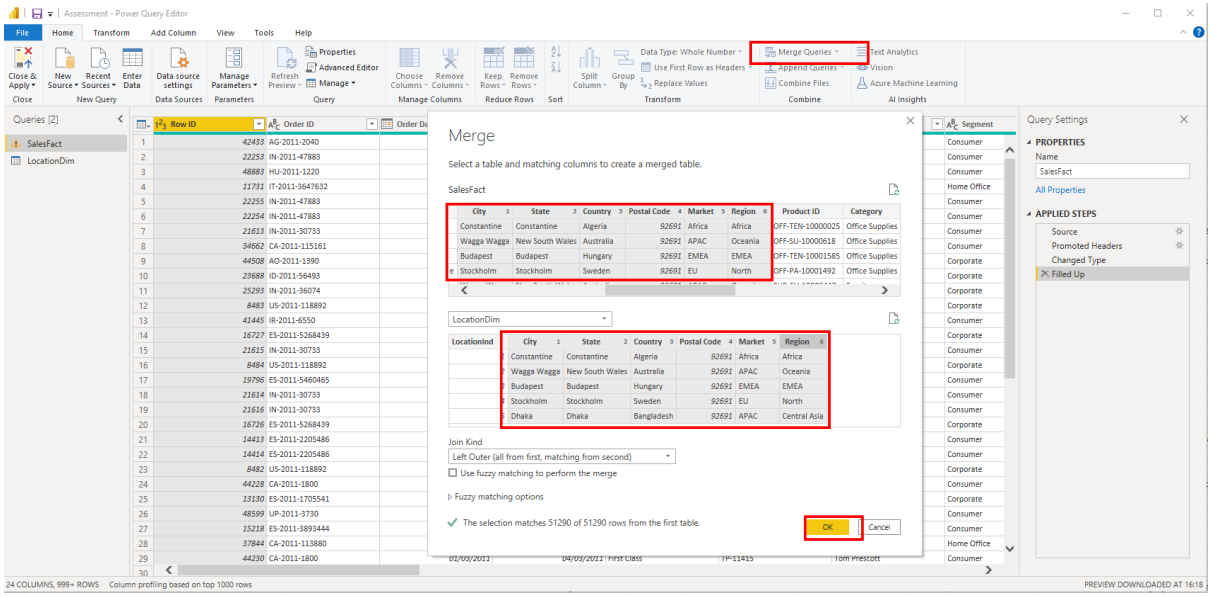
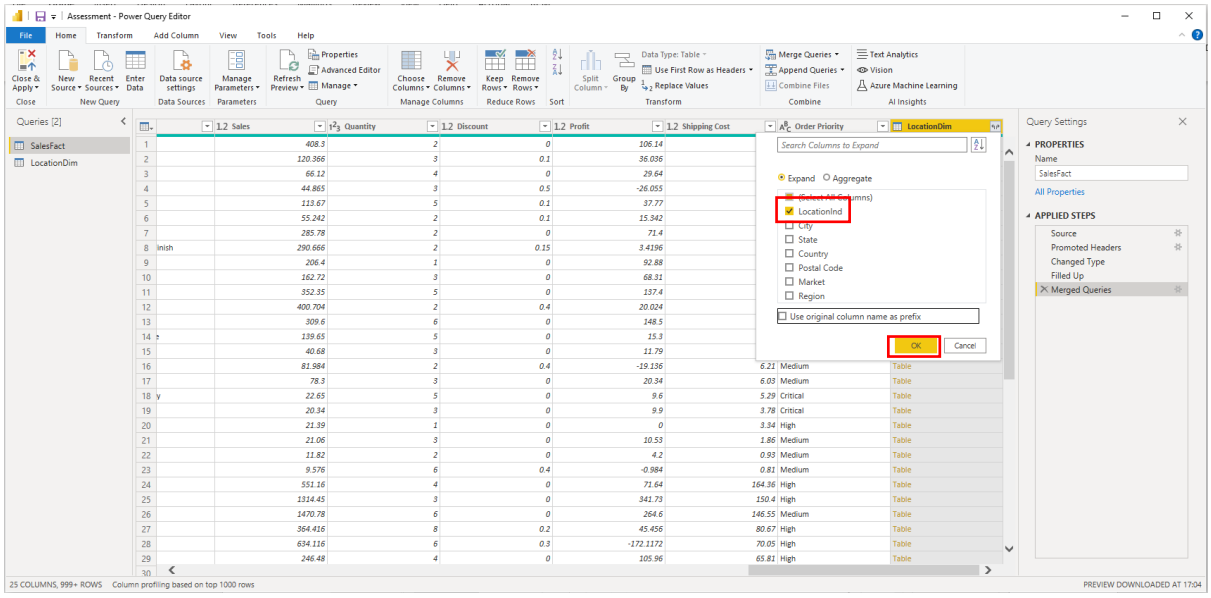


Fig: Matching columns between the sales fact table and the location dimension table.

Merging the location dimension table to the sales fact table reduces the model's complexity and creates a relationship between them. A table is formed by merging and we collapse the table by expanding with the **'LocationInd'** we have created. A table is created in another column and when expanded, only the primary key **'LocationInd'** was selected to represent the other attributes. The columns **'country, state, city, postal code, market and region'** were then deleted from the **Orders** table.



We then remove the columns that has been merged from the fact table

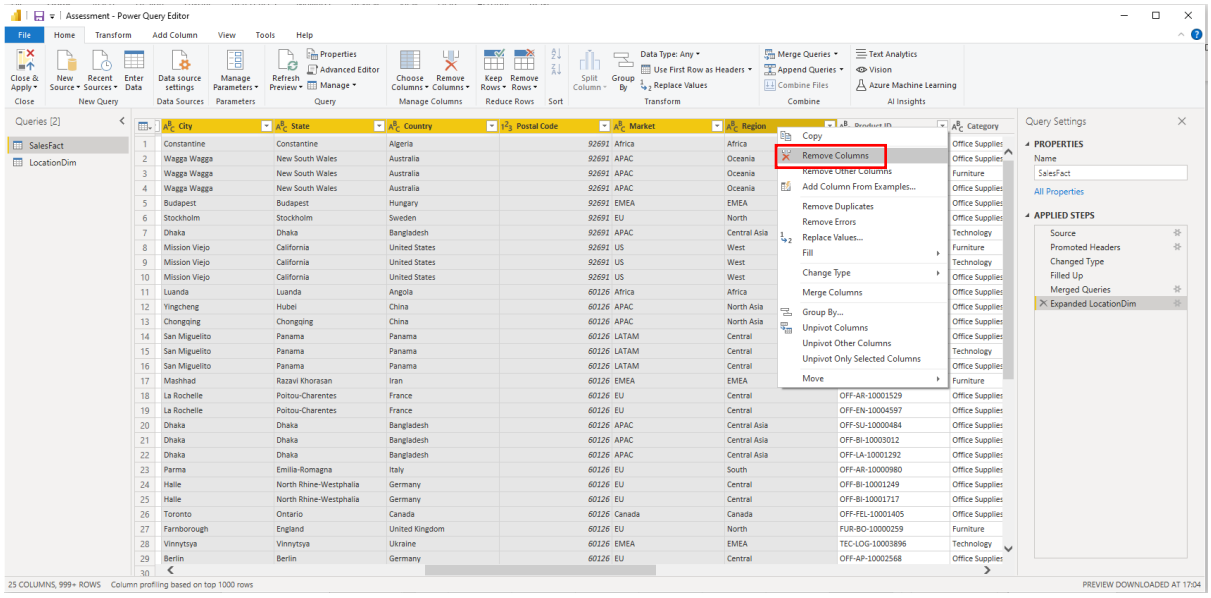
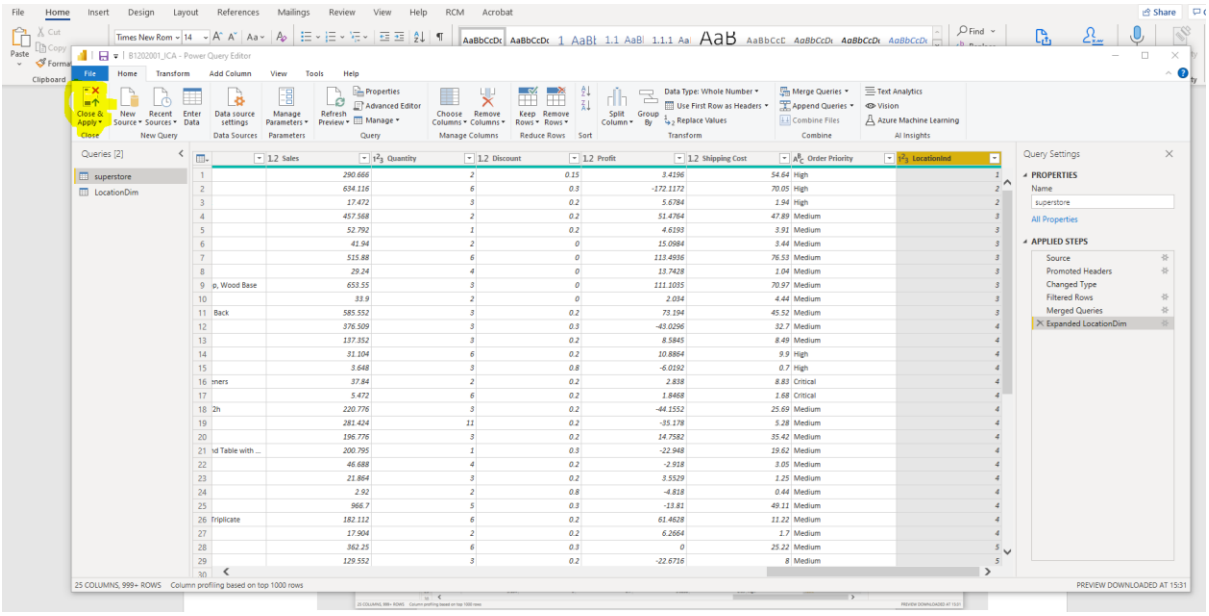


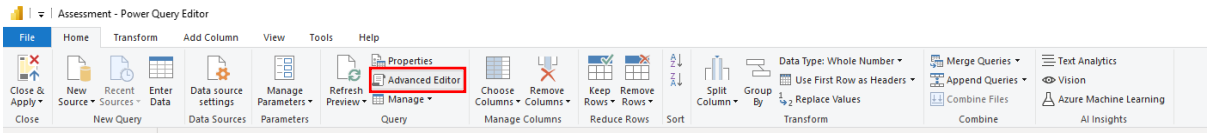
Fig: Selecting the merge queries option to merge LocationDim to superstore



We then close and apply

Similar approaches were performed to generate the other dimension tables: **DateDim, CustomerDim and ProductDim.** As a result, a star schema was created by normalising the flat table into several dimension tables and one fact table.

For the **OrderDim**, make use of **M Language** to remove the columns that have been merged through indexing to the fact table, we go click on '**Advanced Editor'** in the home ribbon



The columns Order ID, Order Date, Ship Date, Ship Mode and Order Priority were removed from the columns through the M language below

Fig: M language to remove columns

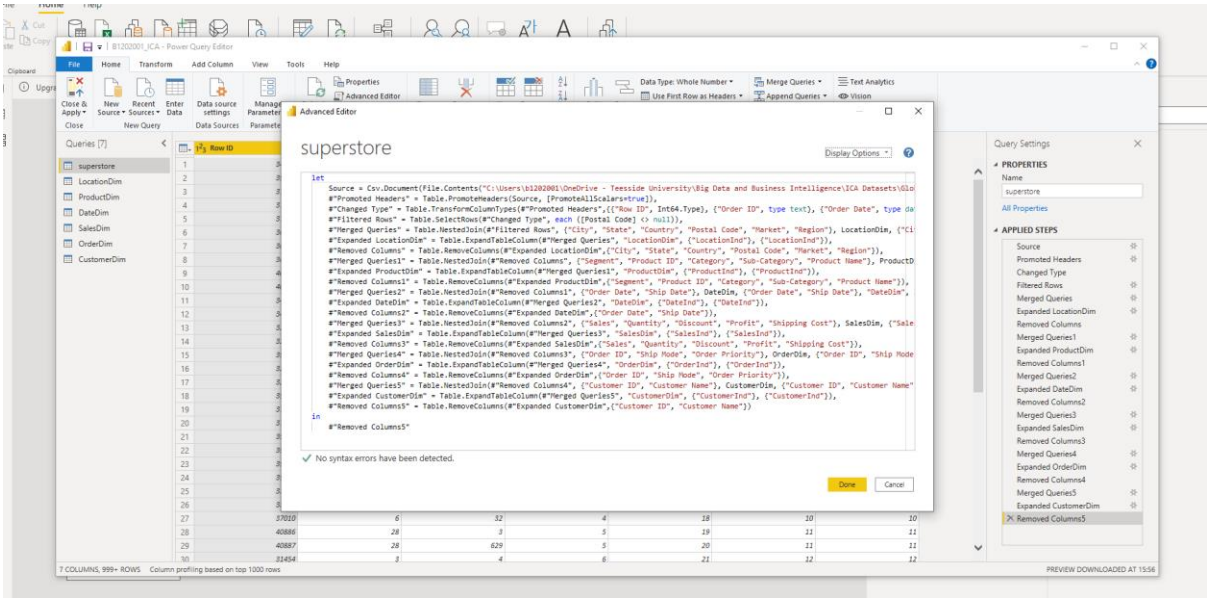


Fig: M language showing how all the dimension tables were generated

2.3.2 Further Pre-processing

In the DateDim table, we computed three new columns called ‘Day’, ‘Month’ and ‘Year’ by clicking on the new column button and writing a DAX formula

to show the corresponding days, months and years of the order date as depicted below.

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

```
Month = FORMAT(DateDim[Order Date], "MMM")
```

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

Fig: DAX expression to extract day, month and year of the order

Additionally, the order date and ship date in the date dimension table was used to generate the number of days the delivery took using the **custom column option**. This is illustrated in the figure below

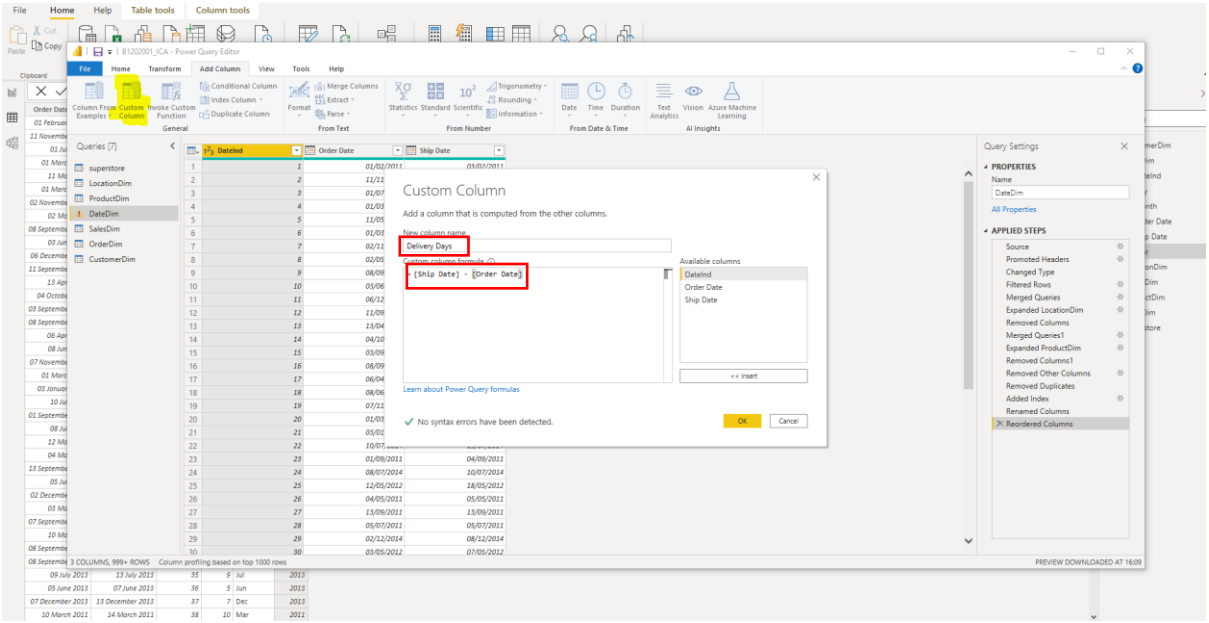
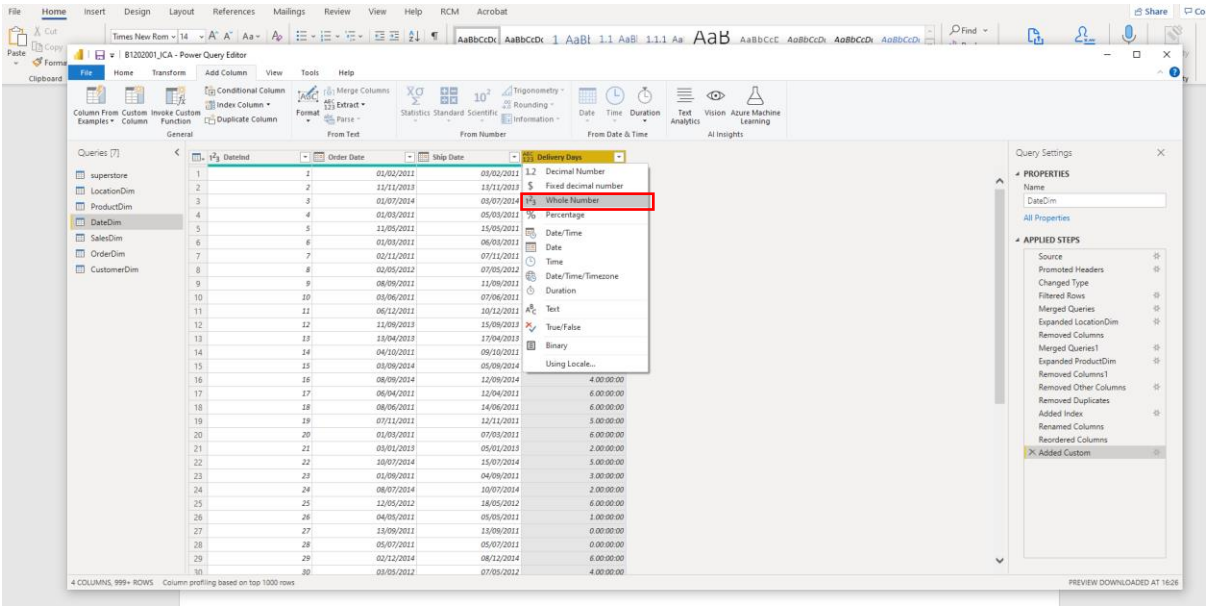
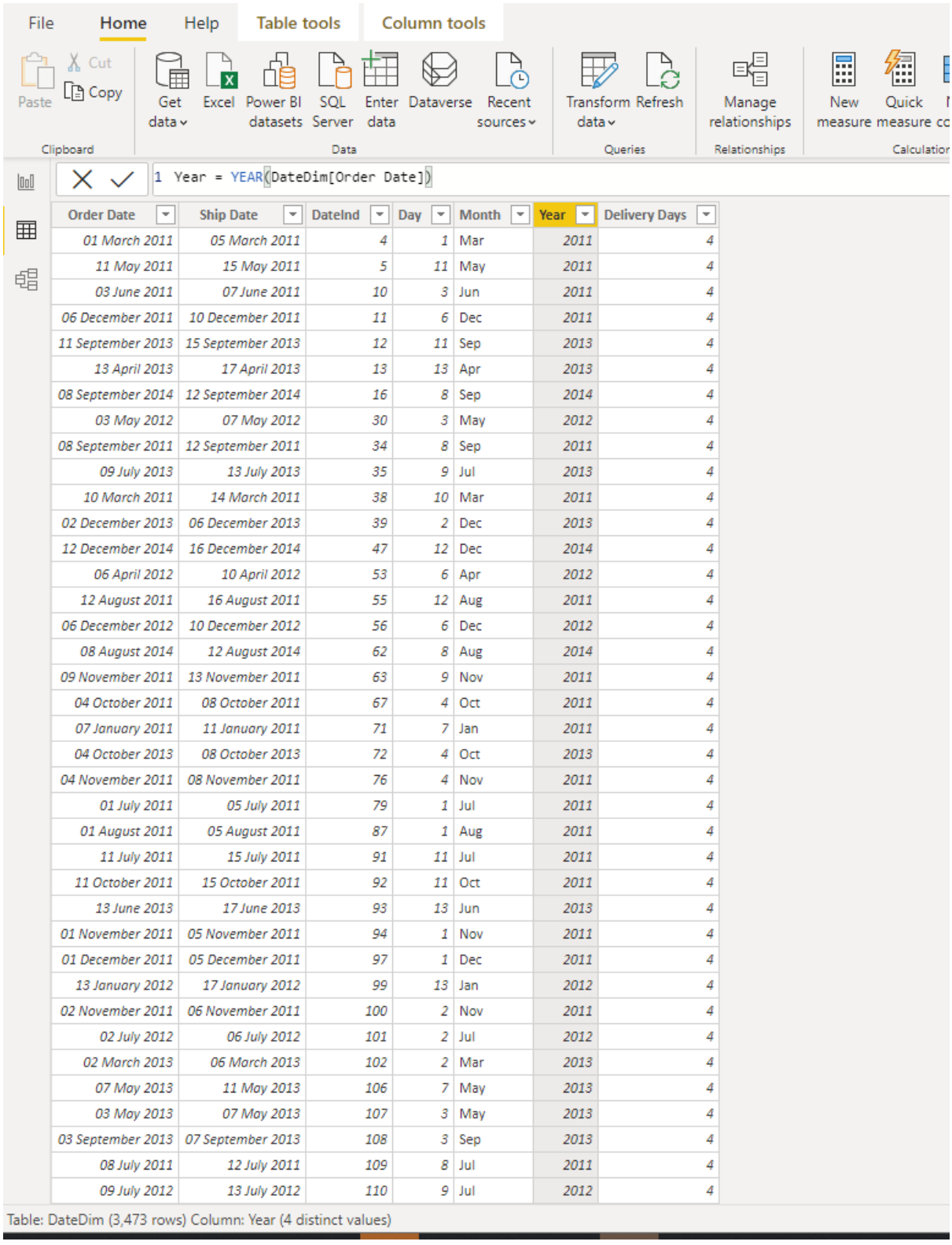


Fig: Creating a new column using power query formulas

After generating the **'Delivery Days'** column, we then change the type to whole numbers by clicking the little type box.



The final **DateDim** table:



Order Date	Ship Date	DateInd	Day	Month	Year	Delivery Days
01 March 2011	05 March 2011	4	1	Mar	2011	4
11 May 2011	15 May 2011	5	11	May	2011	4
03 June 2011	07 June 2011	10	3	Jun	2011	4
06 December 2011	10 December 2011	11	6	Dec	2011	4
11 September 2013	15 September 2013	12	11	Sep	2013	4
13 April 2013	17 April 2013	13	13	Apr	2013	4
08 September 2014	12 September 2014	16	8	Sep	2014	4
03 May 2012	07 May 2012	30	3	May	2012	4
08 September 2011	12 September 2011	34	8	Sep	2011	4
09 July 2013	13 July 2013	35	9	Jul	2013	4
10 March 2011	14 March 2011	38	10	Mar	2011	4
02 December 2013	06 December 2013	39	2	Dec	2013	4
12 December 2014	16 December 2014	47	12	Dec	2014	4
06 April 2012	10 April 2012	53	6	Apr	2012	4
12 August 2011	16 August 2011	55	12	Aug	2011	4
06 December 2012	10 December 2012	56	6	Dec	2012	4
08 August 2014	12 August 2014	62	8	Aug	2014	4
09 November 2011	13 November 2011	63	9	Nov	2011	4
04 October 2011	08 October 2011	67	4	Oct	2011	4
07 January 2011	11 January 2011	71	7	Jan	2011	4
04 October 2013	08 October 2013	72	4	Oct	2013	4
04 November 2011	08 November 2011	76	4	Nov	2011	4
01 July 2011	05 July 2011	79	1	Jul	2011	4
01 August 2011	05 August 2011	87	1	Aug	2011	4
11 July 2011	15 July 2011	91	11	Jul	2011	4
11 October 2011	15 October 2011	92	11	Oct	2011	4
13 June 2013	17 June 2013	93	13	Jun	2013	4
01 November 2011	05 November 2011	94	1	Nov	2011	4
01 December 2011	05 December 2011	97	1	Dec	2011	4
13 January 2012	17 January 2012	99	13	Jan	2012	4
02 November 2011	06 November 2011	100	2	Nov	2011	4
02 July 2012	06 July 2012	101	2	Jul	2012	4
02 March 2013	06 March 2013	102	2	Mar	2013	4
07 May 2013	11 May 2013	106	7	May	2013	4
03 May 2013	07 May 2013	107	3	May	2013	4
03 September 2013	07 September 2013	108	3	Sep	2013	4
08 July 2011	12 July 2011	109	8	Jul	2011	4
09 July 2012	13 July 2012	110	9	Jul	2012	4

Furthermore, three new measures were created, Total Sales, Total Profit, Profit Ratio and Sales by City using the **DAX expressions**.

'**Total Sales'** was generated using the function SUM on Sales table

```
Total Sales = SUM(SalesFact[Sales])
```

'**Total Profit'** was generated using the function SUM on Profit table

```
Total Profit = SUM(SalesFact[Profit])
```

'**Profit Ratio'** was generated using the DIVIDE function on Total Profit by Total Sales

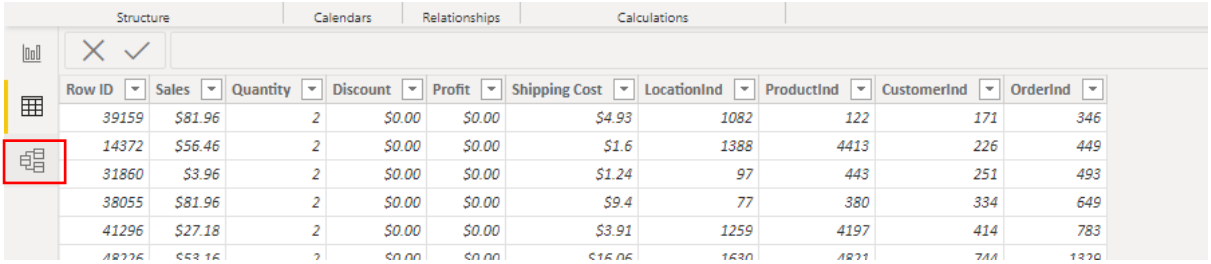
```
Profit Ratio = DIVIDE([Total Profit],[Total Sales],0)
```

'**Sales by City**' was generated with the top 5 cities using City in the LocationDim table and Total Sales.

```
1 Sales by City =
2 CALCULATE([Total Sales],
3 TOPN(5,LocationDim,LocationDim[City],ASC))
```

3 Data Modelling

Power BI is capable of detecting relationships between the models automatically as long as we have created the fact and dimension tables relationally. For the sake of this video, we would create the relationship between the fact and dimension tables manually.



Row ID	Sales	Quantity	Discount	Profit	Shipping Cost	LocationInd	ProductInd	CustomerInd	OrderInd
39159	$81.96	2	$0.00	$0.00	$4.93	1082	122	171	346
14372	$56.46	2	$0.00	$0.00	$1.6	1388	4413	226	449
31860	$3.96	2	$0.00	$0.00	$1.24	97	443	251	493
38055	$81.96	2	$0.00	$0.00	$9.4	77	380	334	649
41296	$27.18	2	$0.00	$0.00	$3.91	1259	4197	414	783

Fig: Screenshot of how to assess the model

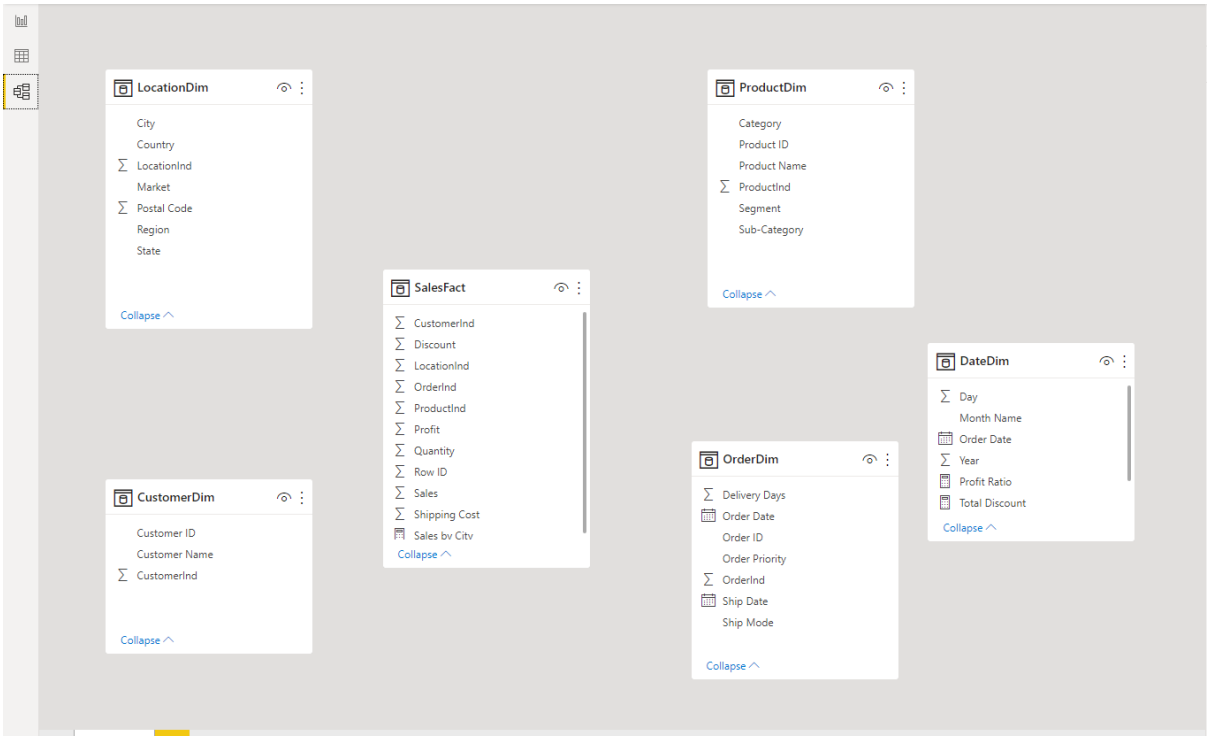
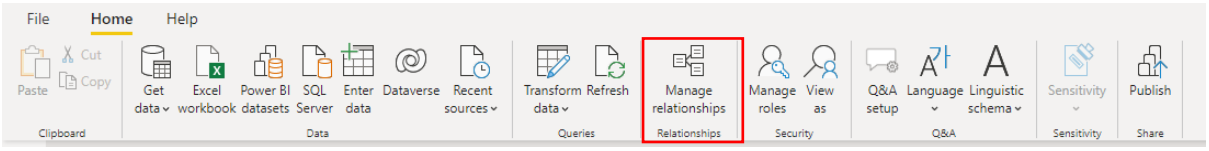
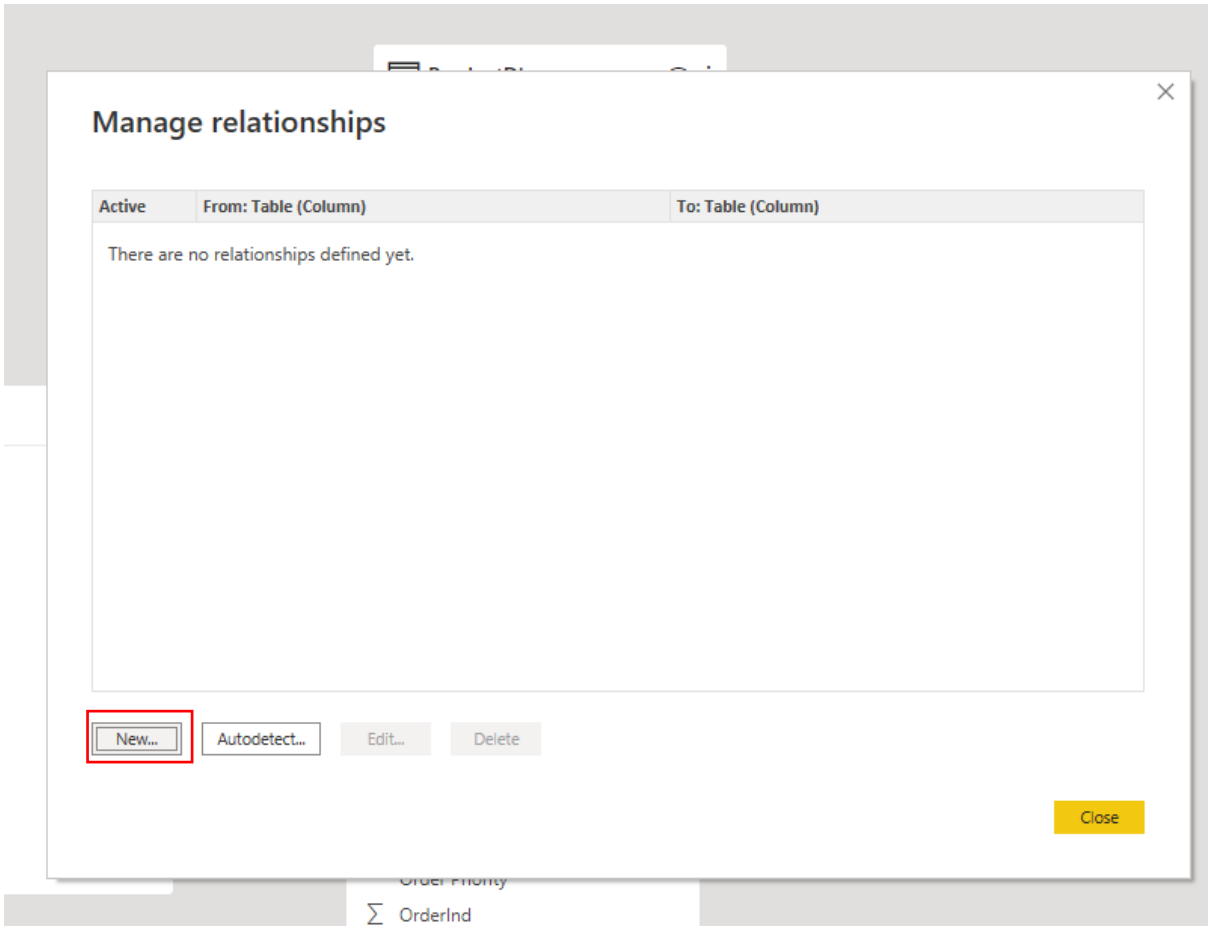
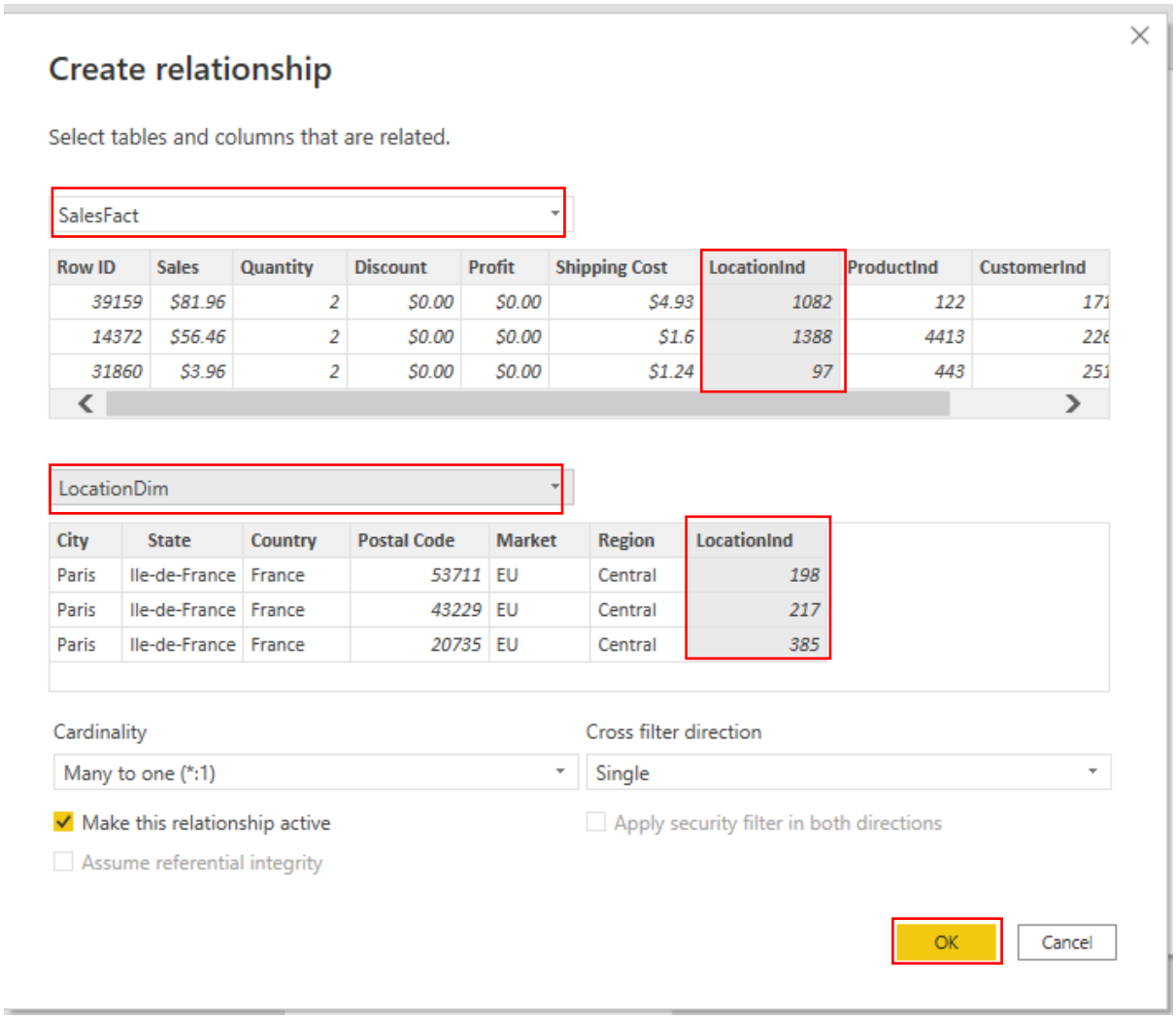


Fig: Screenshot of what the model looks like without relationships

There are three ways by which we can create relationships between these tables, we can either **'drag and drop'** the relational columns or **'manage relationship'** in the ribbon by auto-detecting or creating new. For this sake of this project, we would go with the latter method.




Manage relationships
Active
From: Table (Column)
To: Table (Column)
There are no relationships defined yet.
New...
Autodetect...
Edit...
Delete
Close
OrderInd


Create relationship
Select tables and columns that are related.
SalesFact
Row ID | Sales | Quantity | Discount | Profit | Shipping Cost | LocationInd | ProductInd | CustomerInd
39159 | $81.96 | 2 | $0.00 | $0.00 | $4.93 | 1082 | 122 | 171
14372 | $56.46 | 2 | $0.00 | $0.00 | $1.6 | 1388 | 4413 | 226
31860 | $3.96 | 2 | $0.00 | $0.00 | $1.24 | 97 | 443 | 251
LocationDim
City | State | Country | Postal Code | Market | Region | LocationInd
Paris | Ile-de-France | France | 53711 | EU | Central | 198
Paris | Ile-de-France | France | 43229 | EU | Central | 217
Paris | Ile-de-France | France | 20735 | EU | Central | 385
Cardinality
Many to one (*:1)
Cross filter direction
Single
Make this relationship active
Apply security filter in both directions
Assume referential integrity
OK
Cancel

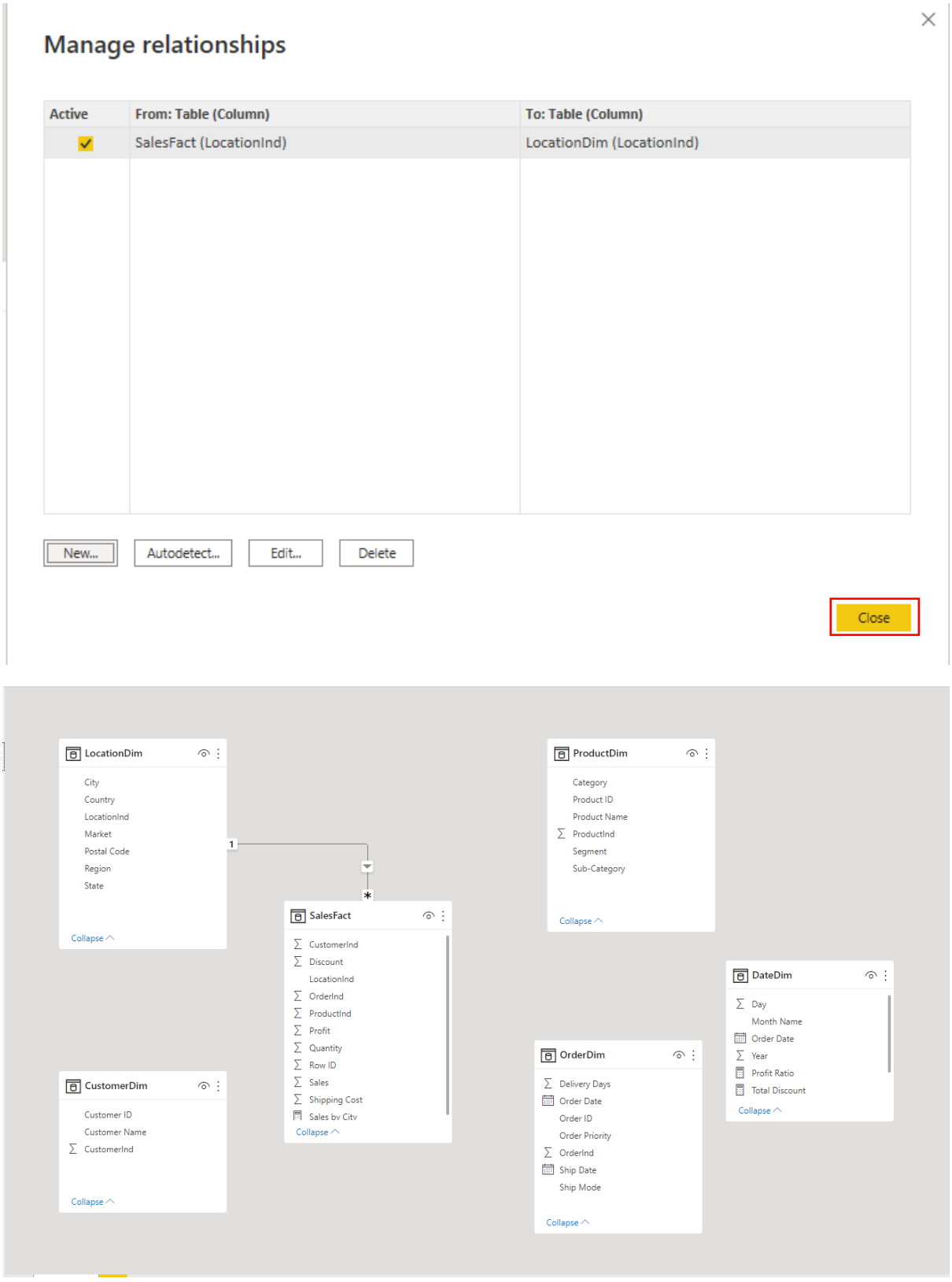


Fig: the first relationship is formed between **SalesFact** and **LocationDim**

We do through these methods to create a relationship between SalesFact and OrderDim, CustomerDim, OrderDim and ProductDim, then OrderDim and DateDim. Below is the data model after creating the relationships.

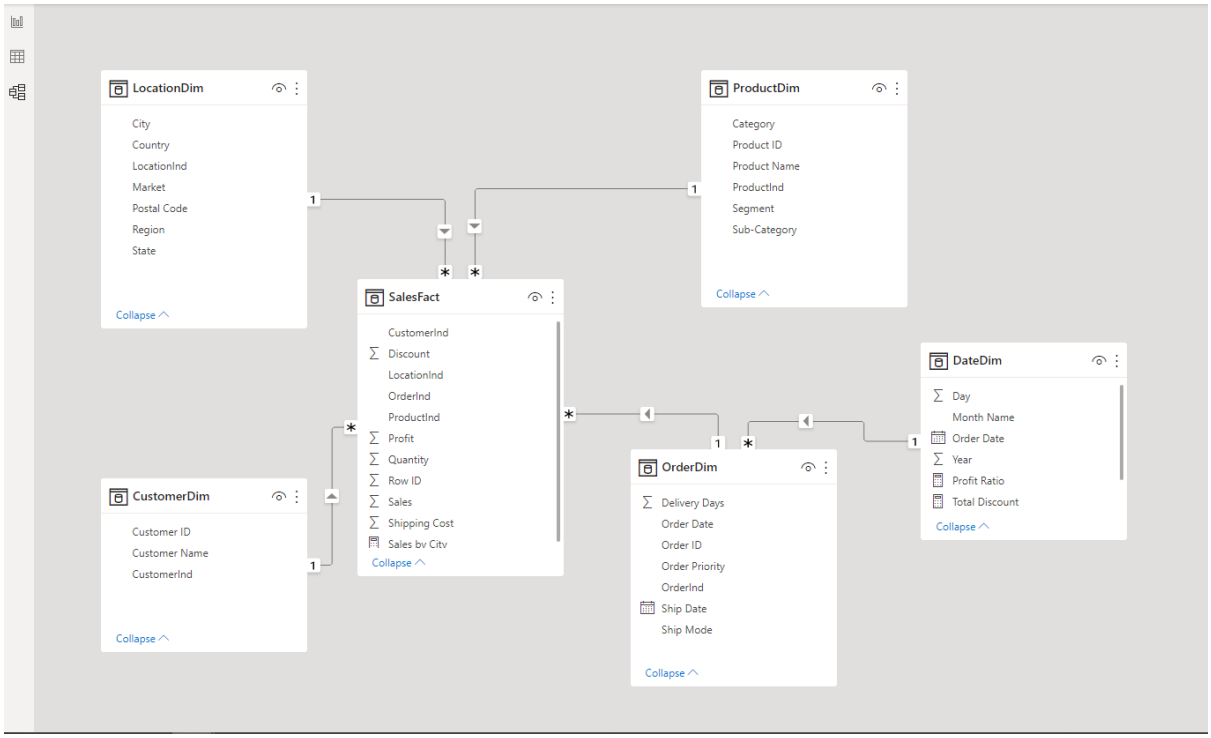


Fig: Data Modelling showing the fact and dimension tables

In the diagram above, the data modelling is a **Snowflake Schema**. The snowflake schema has one or more fact tables linked to numerous dimension tables; however, the dimension table has additional dimensions too. The SalesFact table is linked to other dimension tables in this model, including CustomerDim, OrderDim, ProductDim, and LocationDim. The OrderDim is then linked to the Date dimension, which refers to the Order Date. This model includes one fact table, four dimension tables, and a fifth dimension table linked to one of the preceding dimension tables.

SECTION 2

BUSINESS REPORT

SUPERSTORE SALES ANALYSIS

NAME: AYOOLUWA DORCAS BABALOLA

STUDENT ID: B1202001

1 EXECUTIVE SUMMARY

Superstores are a multi-faceted enterprise that provides a wide range of items in enormous quantities. A superstore is a collection of large stores that sell food, pharmaceuticals, cosmetics, health products, games and toys, furniture, appliances, and a variety of other items. Because of the nature of the industry, most of these products are available in big and bulk quantities. In addition, superstores have locations in other countries and places. The dataset used in this study includes information on client orders from 147 countries and 7 markets. To find insights in superstore sales, many analytical and visualisation approaches are applied.

This report gives insight on the following business questions:

- The trend of sales over the years and the forecast of sales in the next wo years
- The market in which the most and least profit was made
- Reasons for increase and decrease in sales by country
- Category that generated the most revenue
- The time of the year that has the most sale
- Cities in which the most sales were made
- Top customers by sales and profit

Findings and Conclusions - A Power BI dashboard was created to critically analyse the report's questions.

- As the year progresses, sales and profit rise.
- The average number of days it takes to deliver the products is four.
- Technological products are the best-selling products over the years.
- There are usually more sales in the last quarter of the year.
- The more discounts each product group received, the more profit they made. In retrospect, the consumer segment products received the highest discount of $3.8k and, as a result, the highest profit of $750k.
- Customers who contributed the most to superstore sales are not the reason for superstore profit because they are the ones who order the most discounted items.

Recommendations for Improving Performance

- Reducing the number of delivery days to two days would increase customer satisfaction and possibly the sales too.
- To increase profitability, EMEA, African, and Canadian markets should be targeted.
- Customers should be permitted to rate products because this may assist the superstore choose which products to focus on more and which products are the best sellers.
- Customers would have better shopping experiences and sales if they are segmented based on their profile and references.
- Some of the clients who contribute the most to profit aren't regulars. To improve profits, several tactics such as introducing membership cards and promotions can be used to keep customers.

2 INTRODUCTION

A major industry that relies on cutting-edge technology and data analytics to preserve its competitive advantage is the superstore. This initiative intends to boost the superstore's sales and profitability across all products, categories, and countries. This report also examines sales performance from 2011 to 2014 in order to assess effectiveness and identify areas for improvement. The report was created using the dataset specified in Section 1, and the results of the data model are shown below.

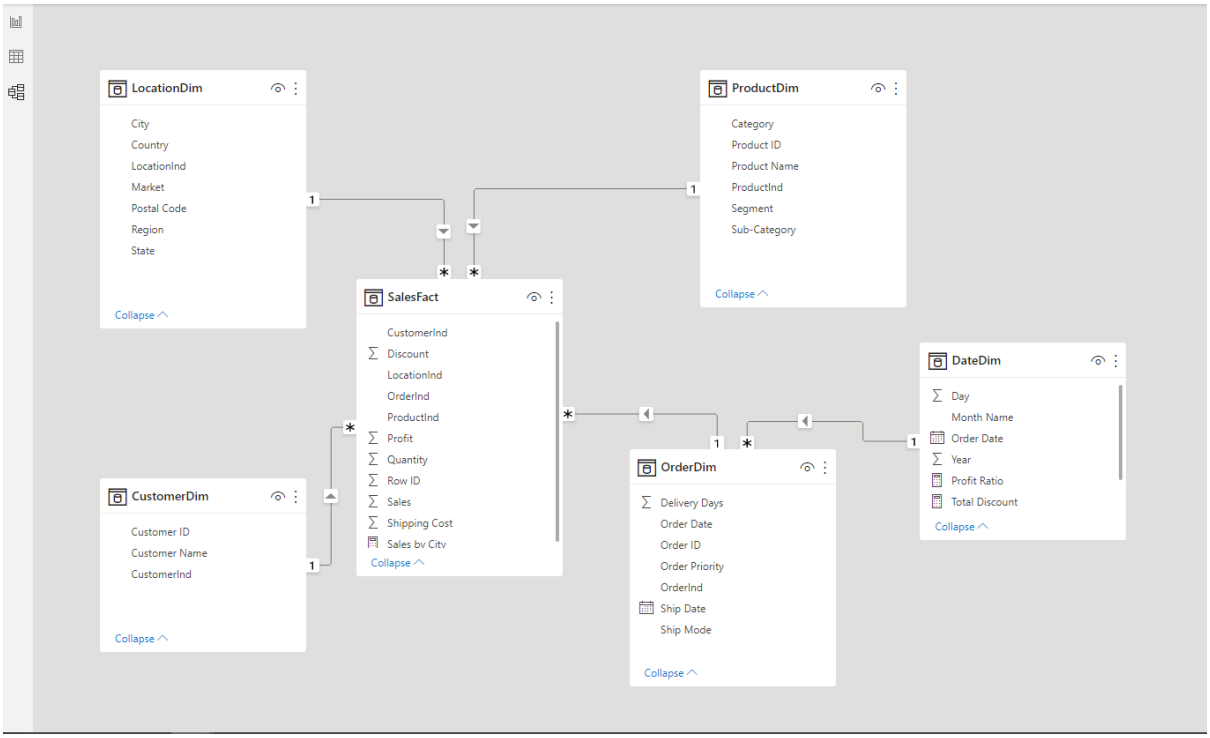


Fig: Data modelling

This model includes six tables named SalesFact, OrderDim, ProductDim, LocationDim, CustomerDim and Date. The data in these tables give detailed information about the orders, sales, profit, products, customers, order and ship dates of the superstore in four years.

3 NEW VISUALIZATION USED IN THIS PROJECT:

According to the project's specifications, a new visualisation tool was imported from the Power BI Visuals package. The **'Scroller'** visual tool aids in the visualisation of data as an animation scrolling text. Its pace, size, status, indicator, and colouring, among other things, can all be adjusted. It can also be used to scroll through personalised messages, notification changes, and receive updates. It was utilised in this report to illustrate the sales of each individual sub-category. Below is an illustration of how it was imported:

First, we go to the visualization pane and click on the 3 dots '…' to see more options, we then select '**Get more visuals**'

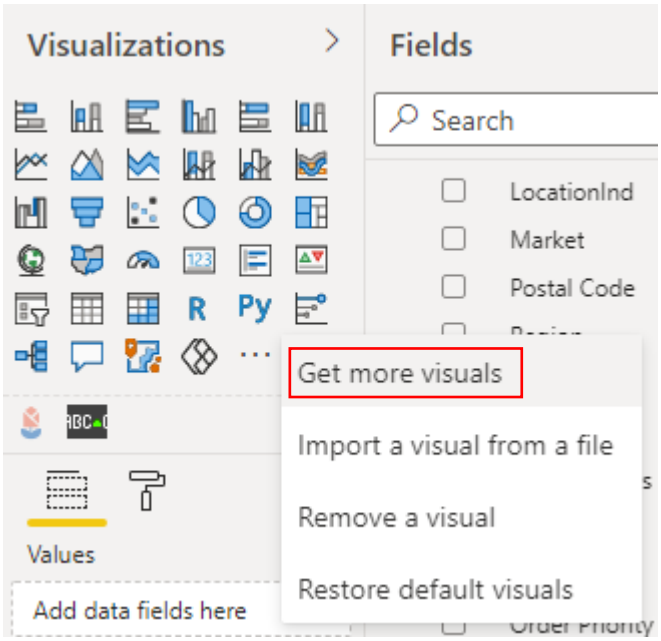


Fig: How to get more visuals

This then takes us to Power BI Visuals page which contains many different visualisation tools, for the sake of this project, we search for ‘Scroller’ and click ‘Add’ to add it to the visualization pane automatically.

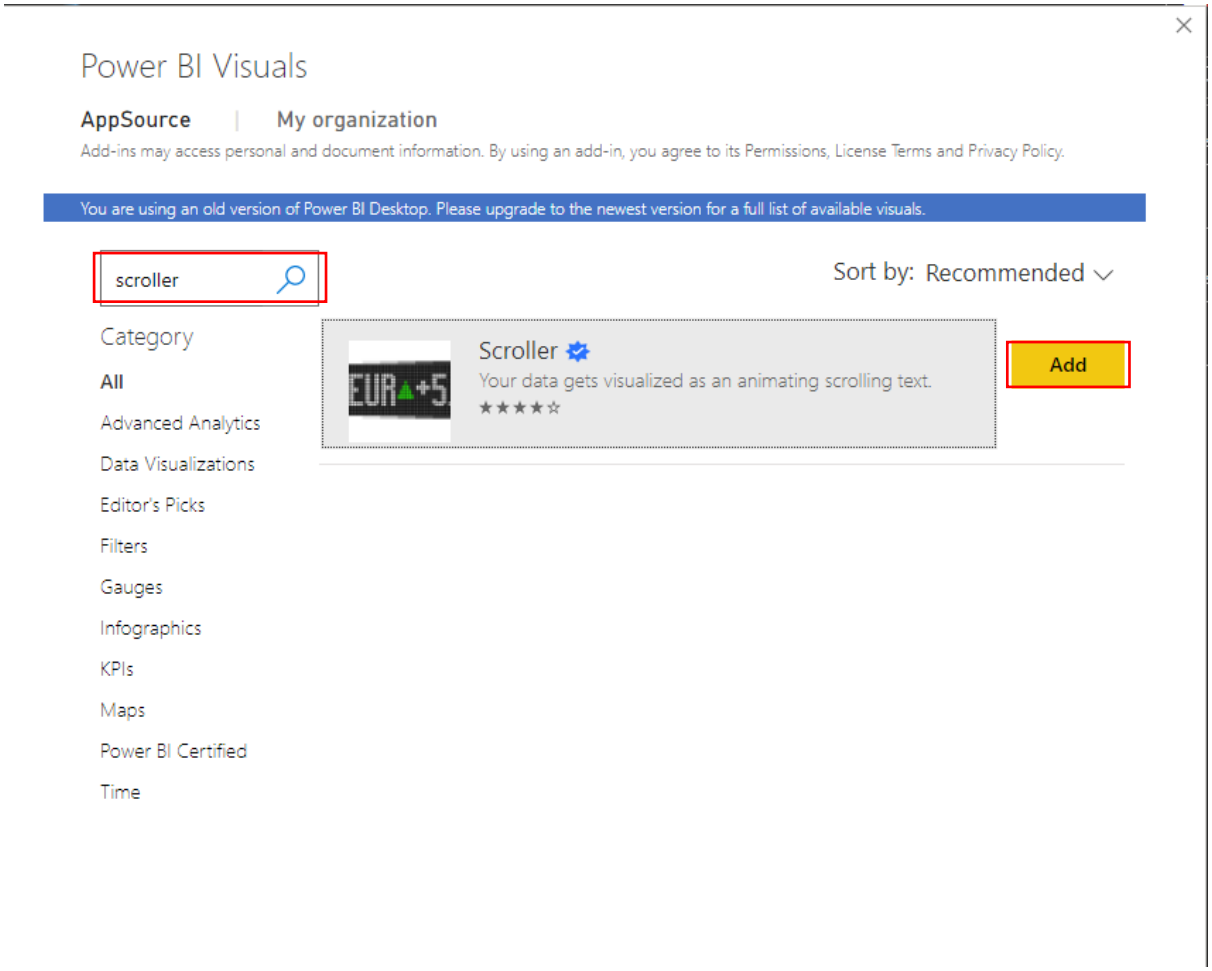
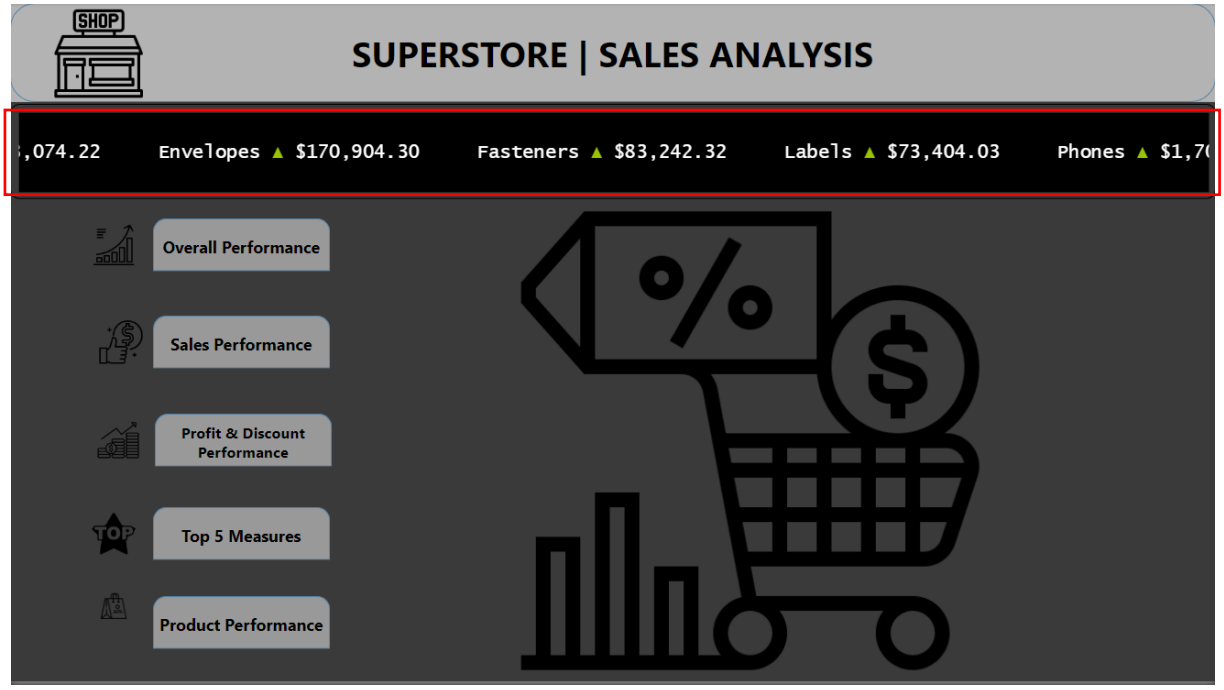


Fig: Importing Scroller as our new visualisation tool

In this project, we can find the scroller being used on the 'Home Page' dashboard.



4 KEY FINDINGS

i. One important crucial finding we can observe on the dashboard over the years is the increase in sales as the year progresses. The overall sales increased by more than 50%. Plotting a clustered column chart and using the trend line in the analytics tab were used to examine the pattern of sales by year. **Total sales** (set to sum) were used as a measure, and **Year** was used as an axis. The superstore began with a total sale of $2.3 million in 2011, and by the end of 2014, it had increased to $4.3 million.
Providing a broader selection of items, establishing new domestic markets, improving customer service, participating in marketing efforts, and many other factors will all contribute to increased sales over time. In the diagram below, the trend line provides extra information. A simple DAX formula was used to calculate the total sales of the given data.

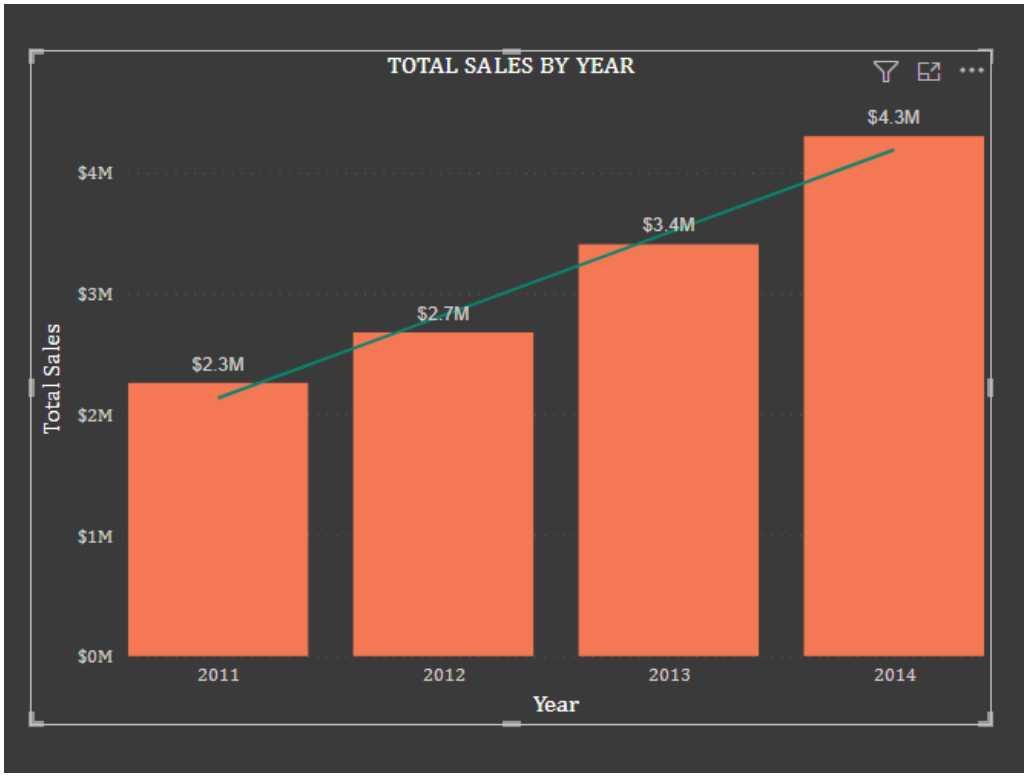


Fig: showing the trend line of sales over the years

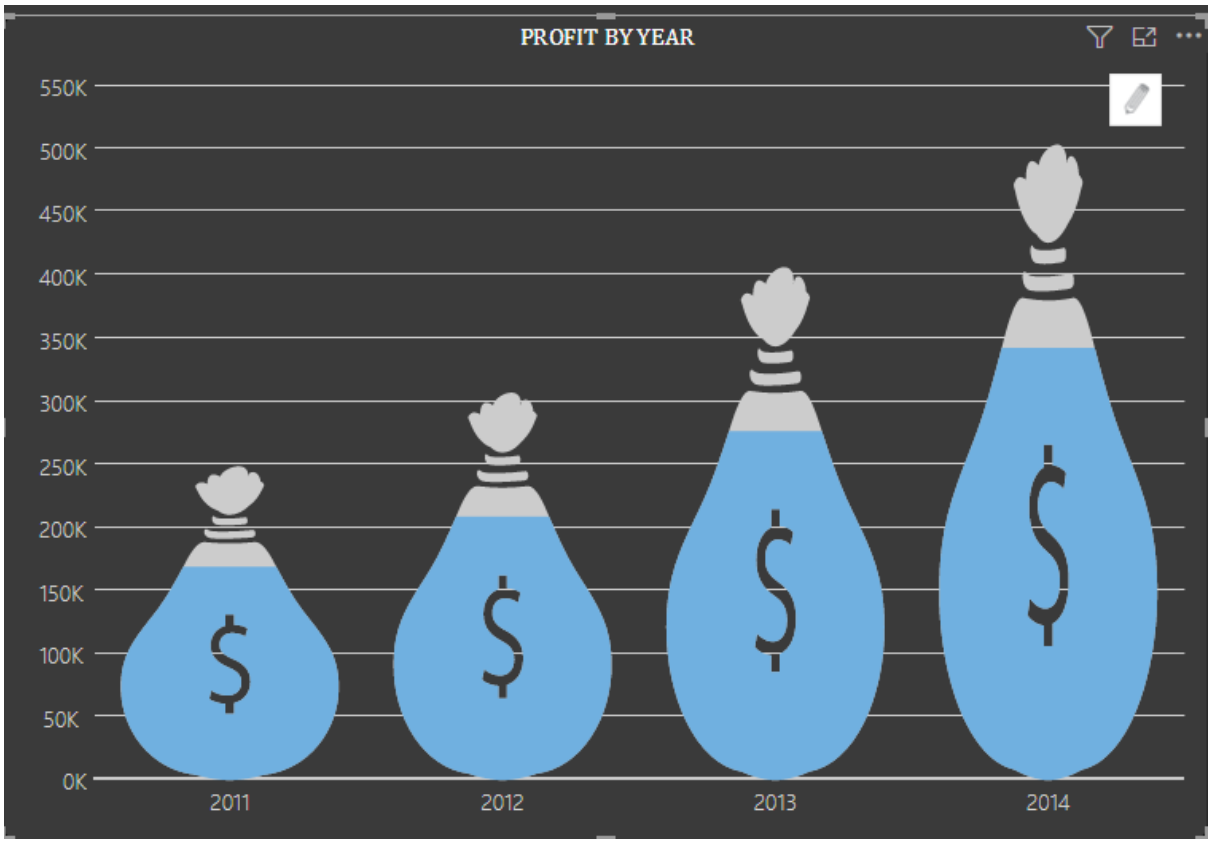


Fig: The profit also progresses with the year

ii. Due to a strong trend in total sales over the years, a prognosis was created for the next two years, with total sales increasing to roughly $10 million by the end of 2016, which is beneficial for the superstore. This is illustrated in the figure below where the **Sales** was plotted against the **Order Date**.

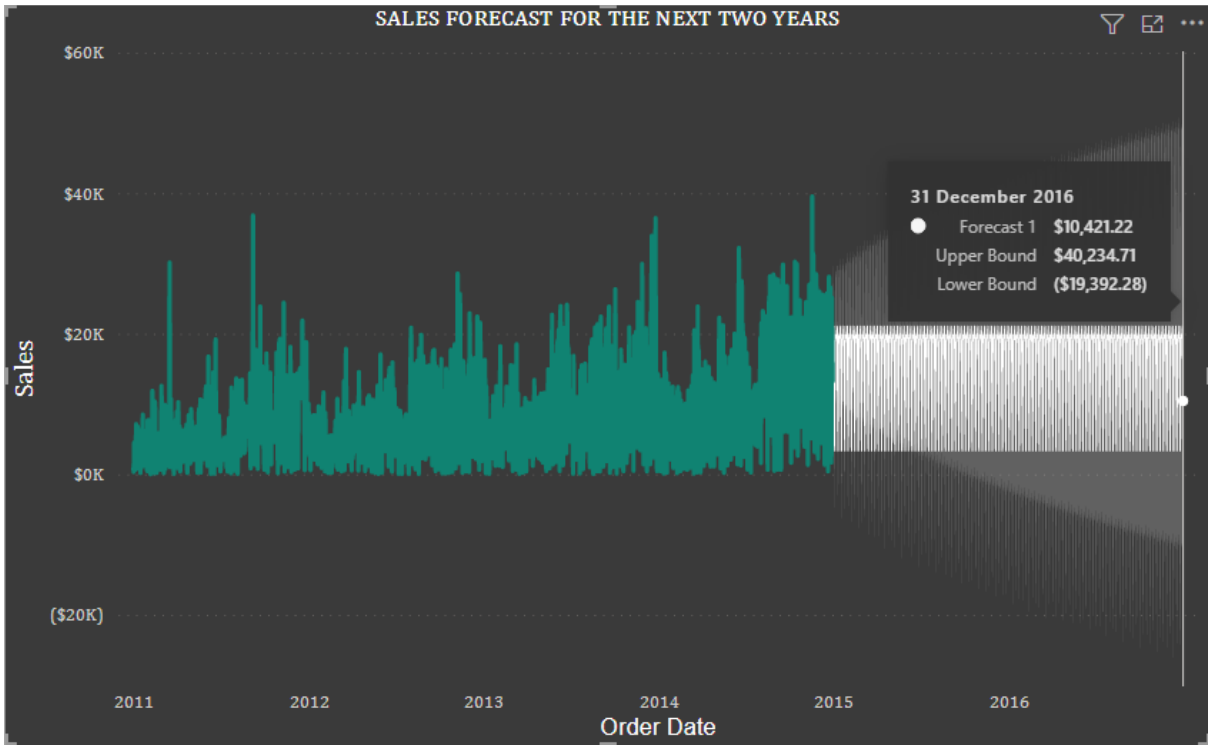


Fig: Showing a forecast of sales by 2016

iii. More profit was made in the APAC market (Asian Pacific Market) while less profit was made in the Canadian market. While concentrating on the market that brings on the most profit, measures should be taken to expand the profit in Canada. As illustrated in the picture below, a waterfall chart was utilised to have a deeper look at the profit by market. **Profit** was used in the SalesFact table, and **Market** was used in the LocationDim table to construct the chart. The order of markets that made more profit too are the EU, US, LATAM (Latin America), Africa and EMEA (Europe, the Middle East and Africa) respectively.

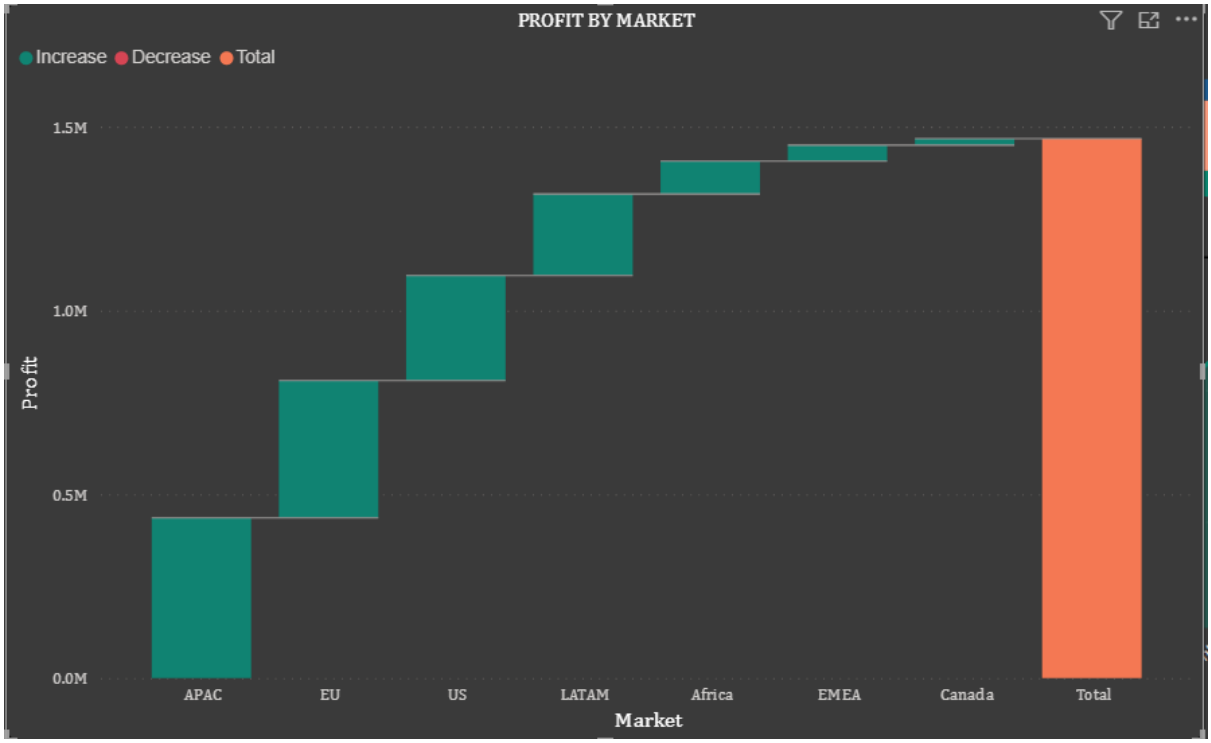


Fig: Profit by Market

iv. The total sales by month varied all round the year however the lowest sales were made between August and September and the highest sale was made between October and November. A line chart was used to analyse the sales through the eyes of the month and quarter, as seen in the image below. The month served as the axis, and the total sales served as the values. Further down, the significant drop and spike were analysed in relation to the countries.

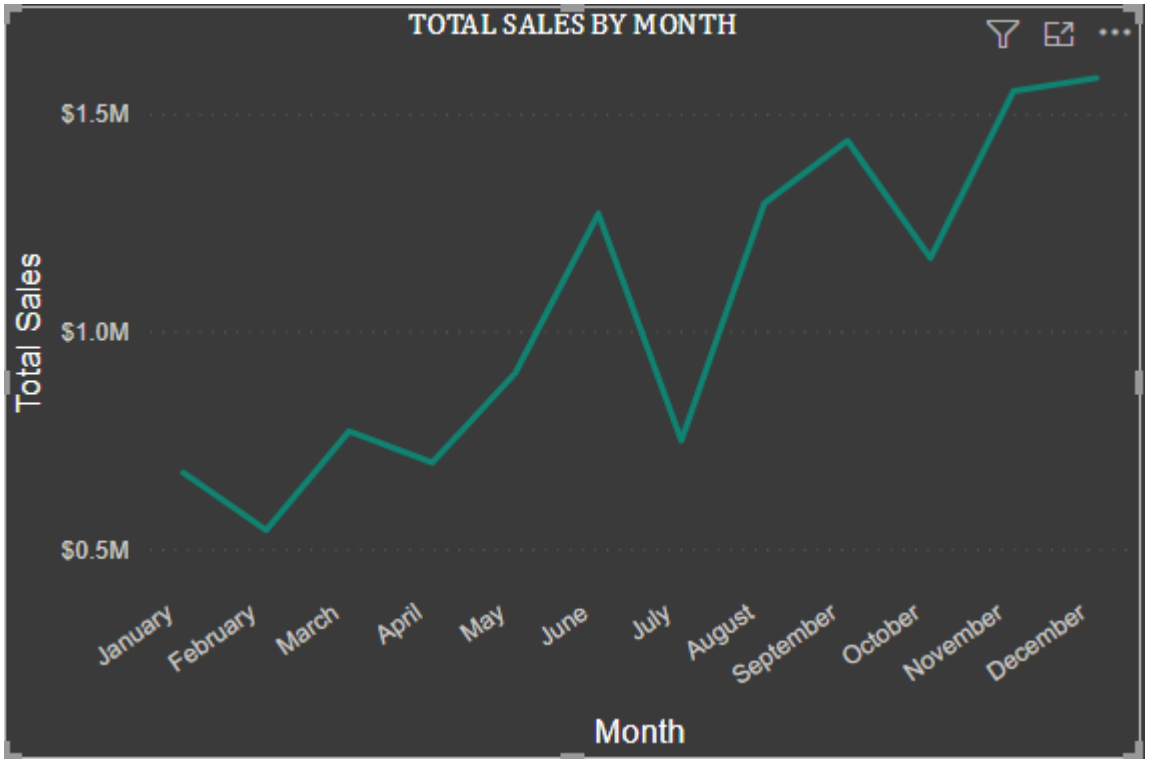


Fig: Total sales by month

Using waterfall charts depending on countries, the explanation for the growth and decline at these times was studied further. Between August and September, the United States, Mexico, and Nicaragua saw the biggest growth among the countries, offsetting China's decline. Furthermore, the United States', China's, and France's relative contributions changed the greatest. This is depicted in the graph below.

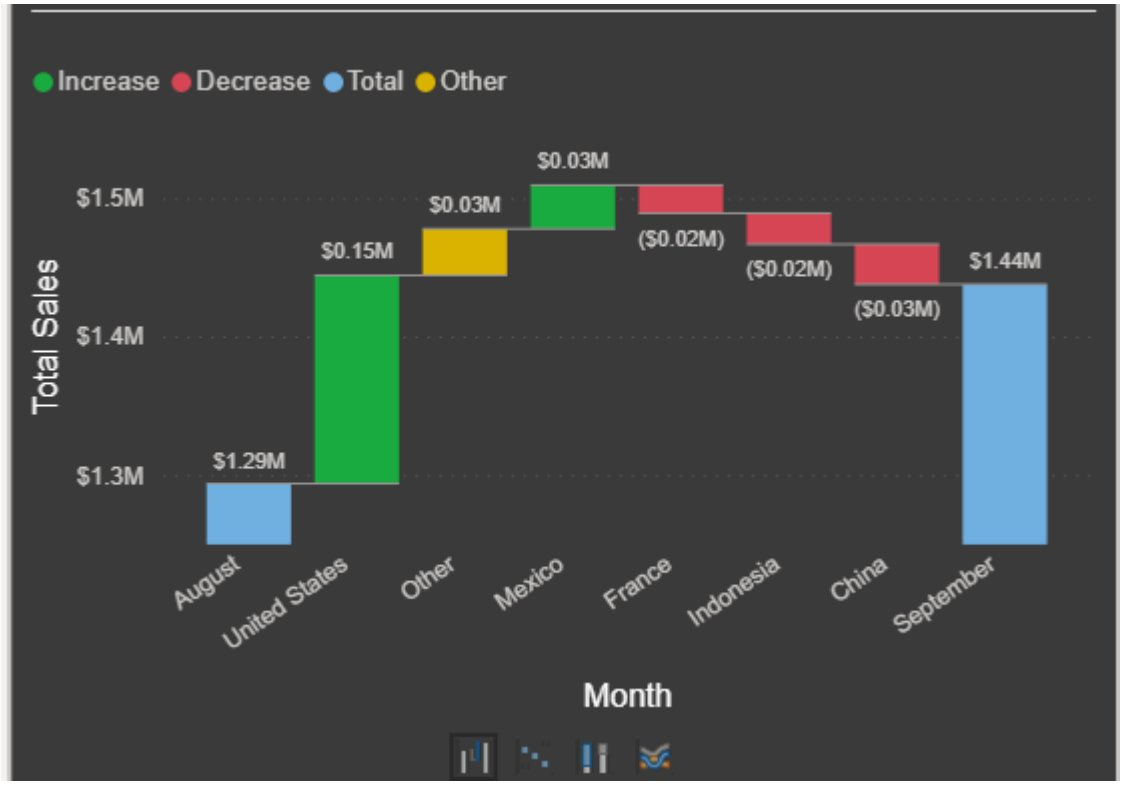


Fig: Increase in sales among countries

Also, Australia, France and United Kingdom had the largest decrease among the countries. The relative contribution made by the United States, Australia and United Kingdom changed the most. This is illustrated in the chart below.

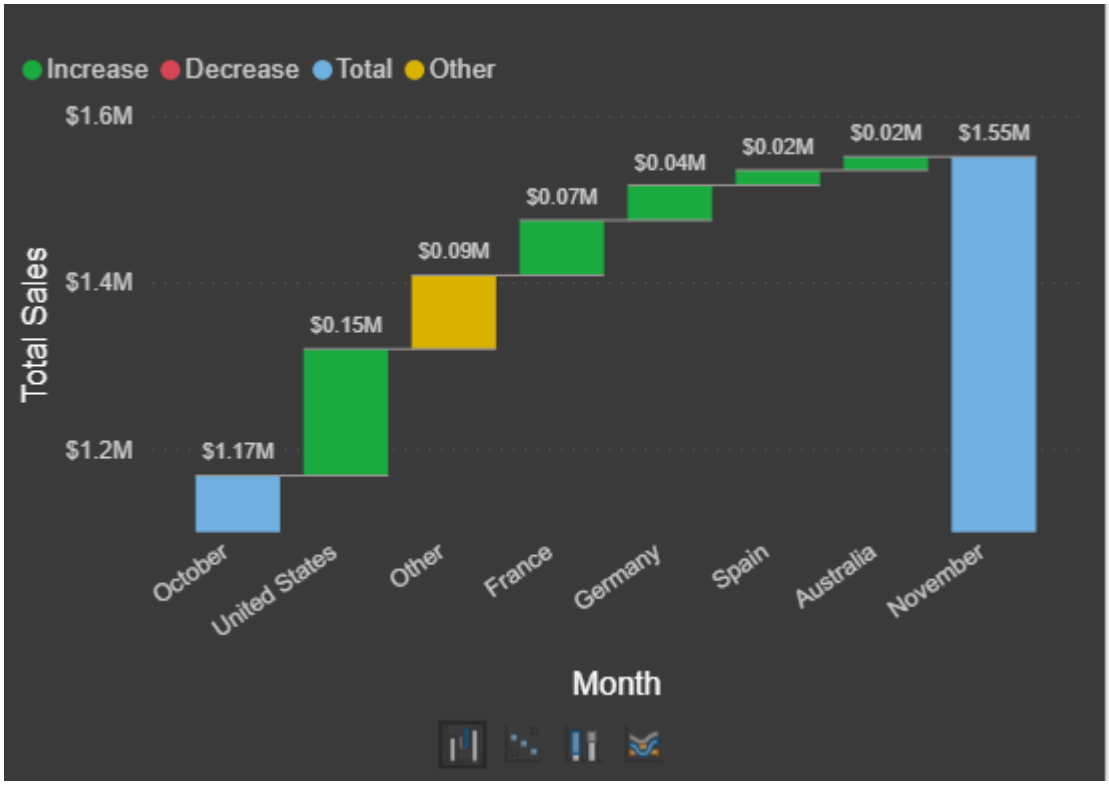


Fig: Decrease in total sales by country

v. The technology category had the highest sales and, as a result, the most profit. Even though office supplies had the lowest sales, they made more money than furniture. As a result, in order to enhance revenues, the store should focus more on furnishings too. A clustered bar chart was used to analyse and visualize this chart with the category as axis and profit and sales as values.

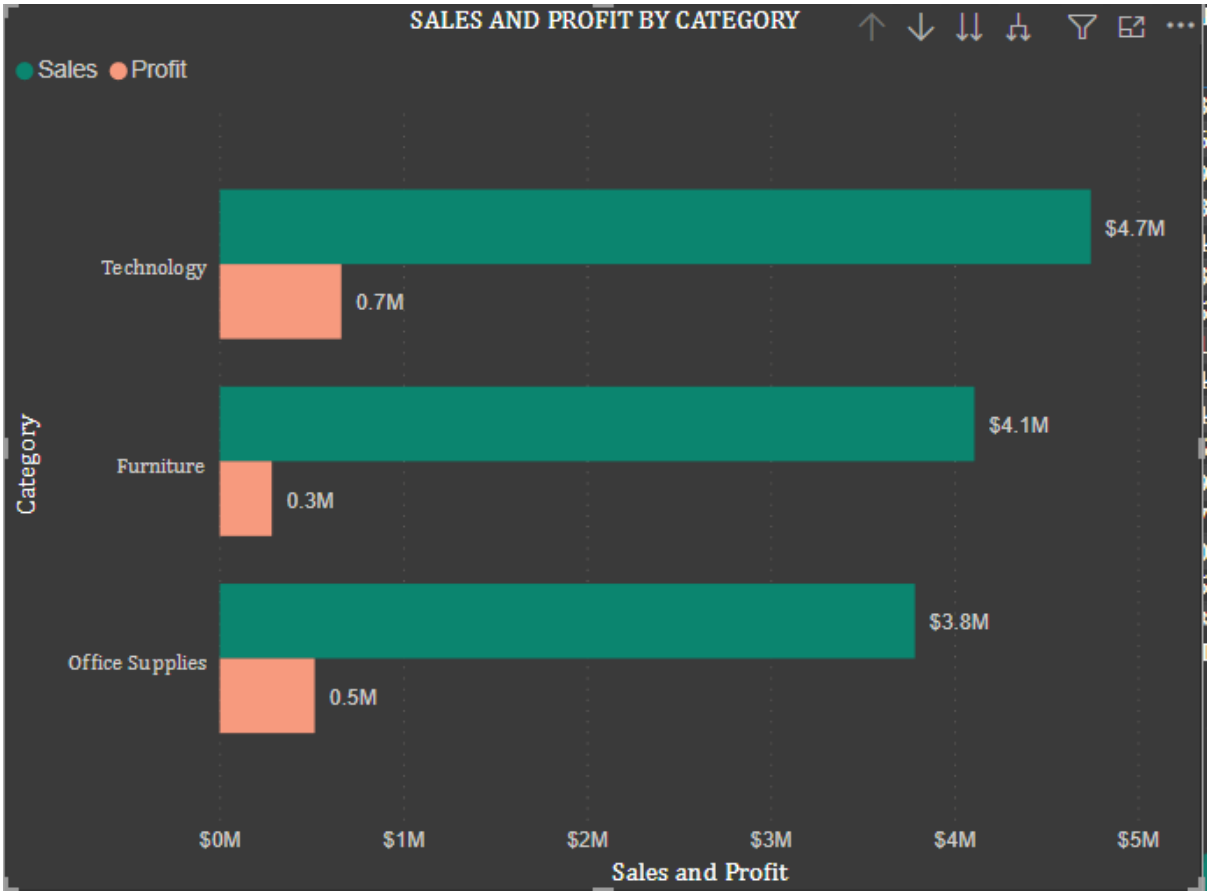
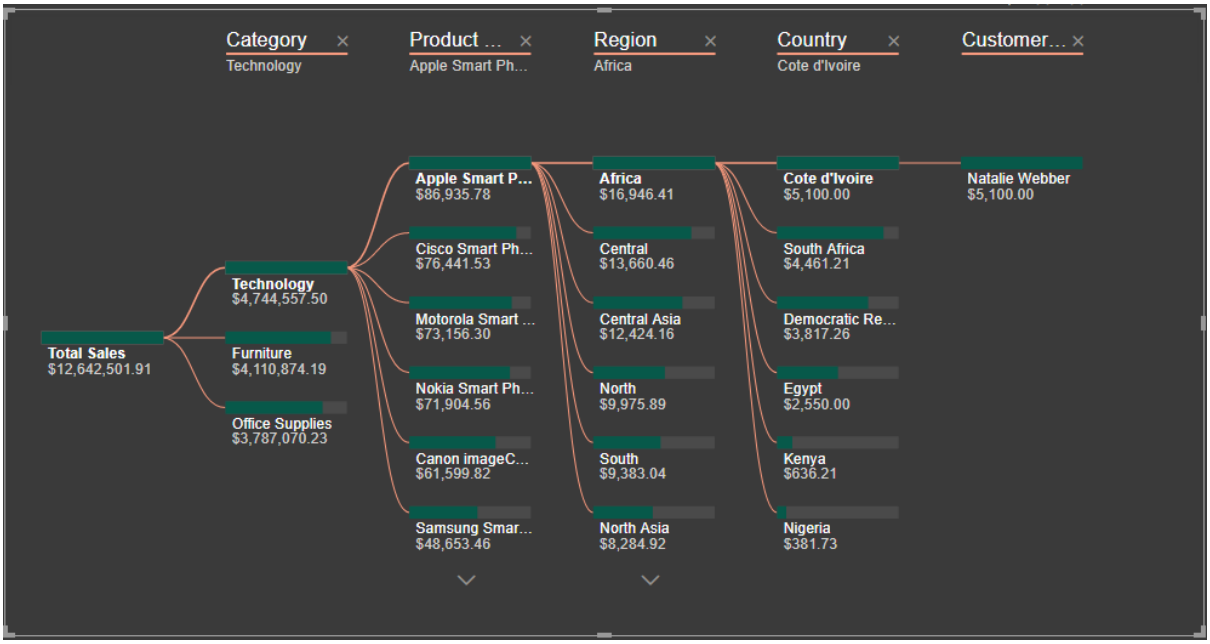


Fig: Sales and Profit by Category

Furthermore, a decomposition tree has been used to buttress the total sales and it was expanded by the category, product, region, country and customer with Technology being the category with the most sales, an Apple smart phone as the highest selling product in Cote d'lvoire in Africa and Natalie Webber as the customer.



vi. It can be seen from the figure below that profit is higher in the fourth quarter of the year. This could be due to a variety of factors, but further investigation reveals that accessories, appliances, and paper generate more earnings.

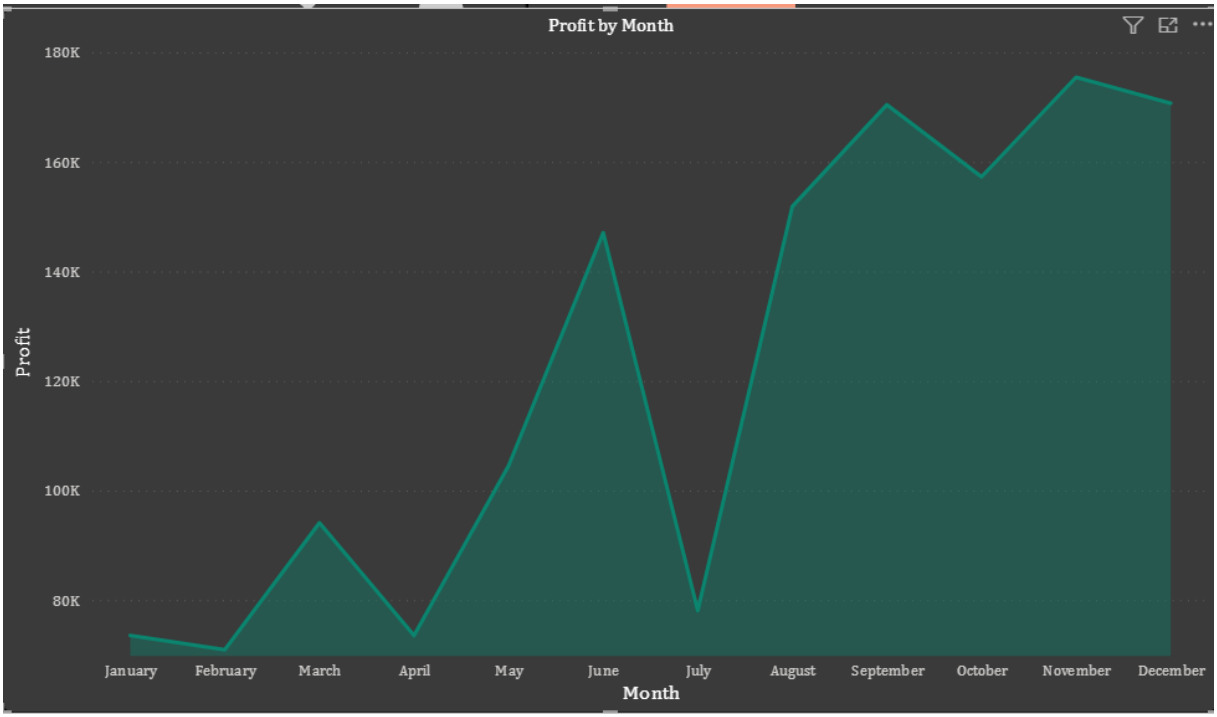


Fig: Trend of profit by month

vii. While analysing the top sales by cities, it was discovered that the top 5 cities with the most sales are all in the United States, or more precisely,

four of the top five cities with the highest sales are all in the United States, with Manilla being the exception. To increase revenue, the superstore must ensure that the standard is maintained while also aiming to expand its locations in other cities.

We first create a measure using the DAX formula called Sales by city by generating the Top 5 cities in ascending order after calculating the Total Sales.

```
1 Sales by City =
2 CALCULATE([Total Sales],
3 TOPN(5,LocationDim,LocationDim[City],ASC))
```

Moving forward, a map was used in the visualization pane by using the city as the location and legend and the Sales by City measure in sizes. This is illustrated in the figure below.

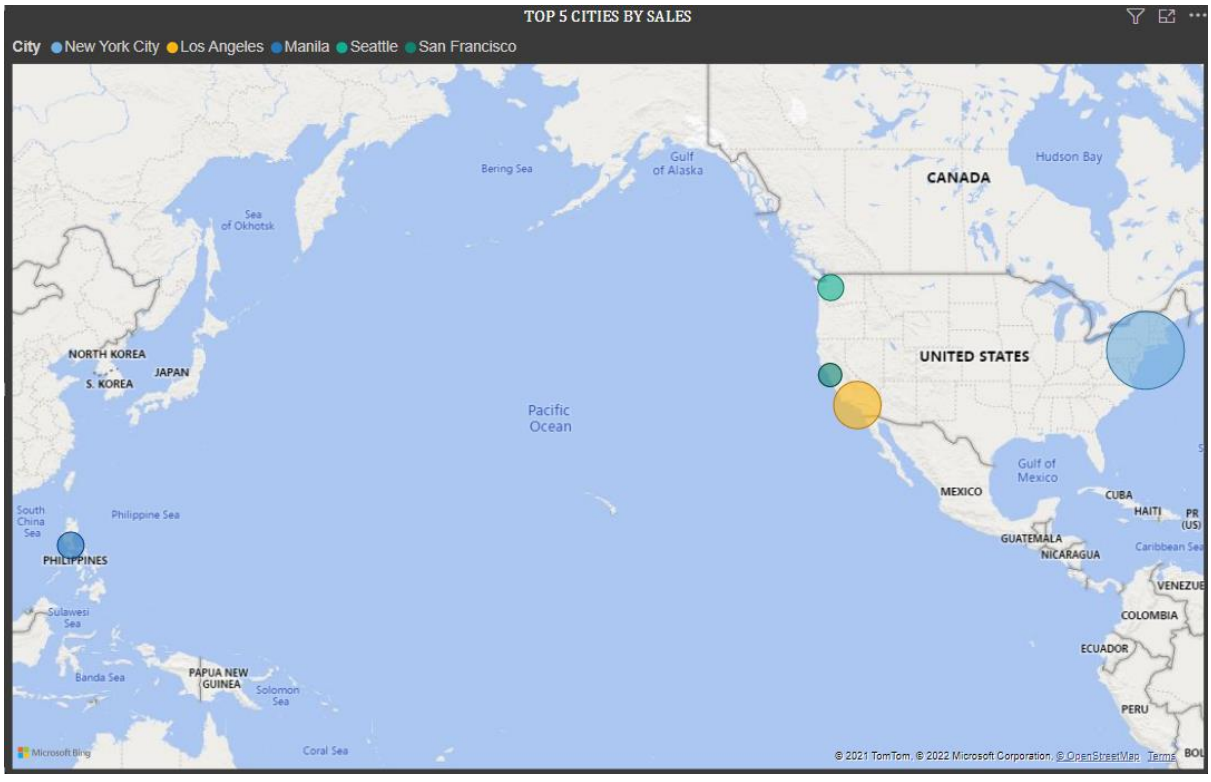


Fig: Top 5 cities with the most sales

viii. Using a clustered column chart to visualize the number of delivery days by the order priority, we see that the low priority order took the most days to deliver, and in that light, critical orders were delivered faster which took

about 2 days. We also use a **'Trend line'** in the analytics tab to find the average number of delivery days.

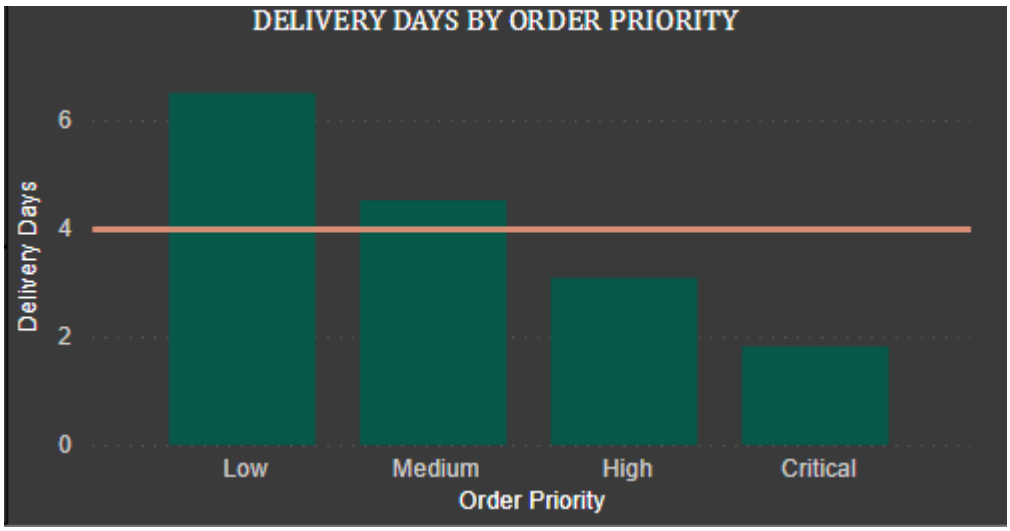
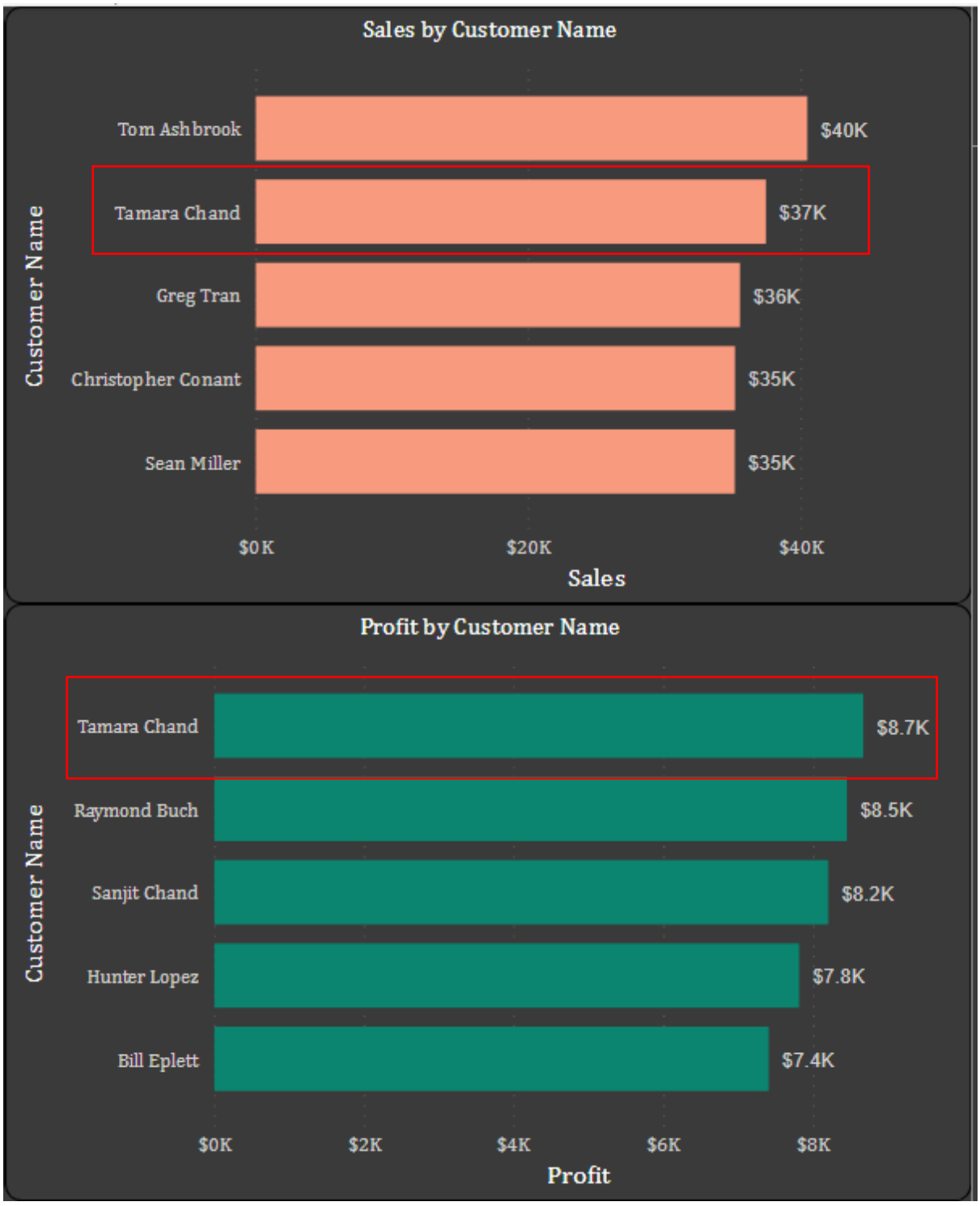
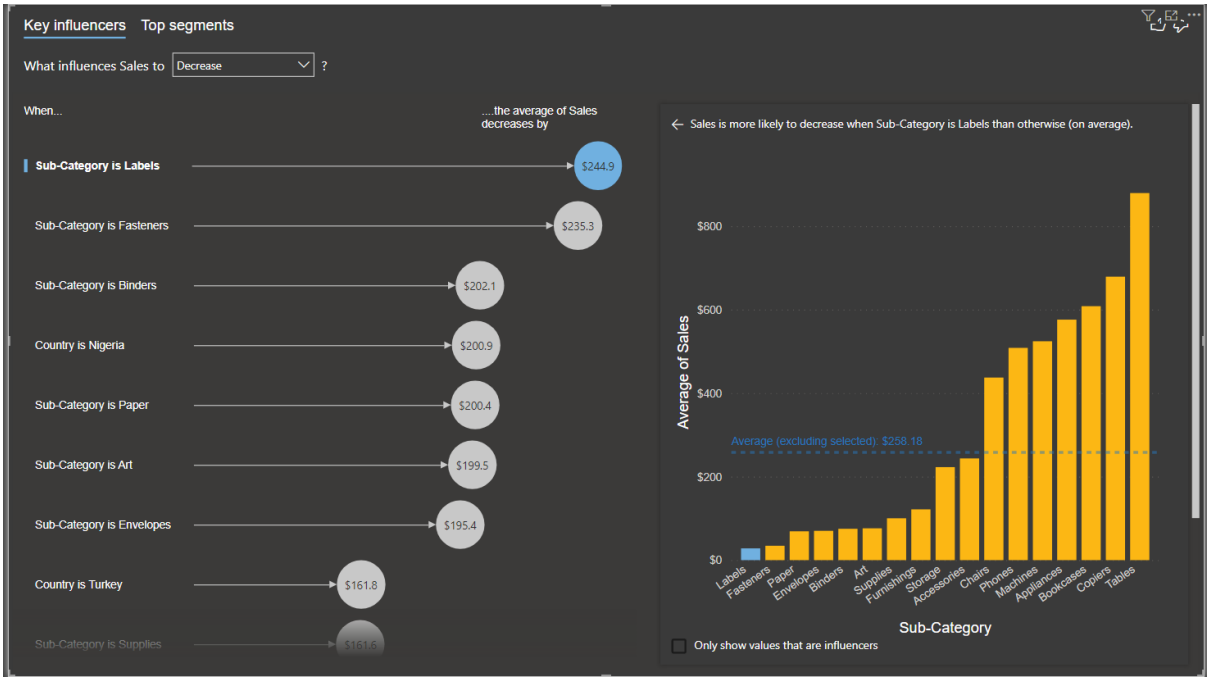


Fig: The average delivery days by order priority

To analyse sales by country, delivery days, sub-category, ship mode, and segment, a **'Key Influencer'** visualisation was used for its description in the image below. When we look at what causes sales to be low or drop, we notice that the sub-category **Labels** has the most impact, reducing sales by $244.9 more than any other aspect. In the right panel, we will find a more detailed graphic about the sales situation. It goes into greater depth on how each item in the subcategory influences sales and causes them to fall. The second most important factor causing a decline in sales is the sub-category of **Fasteners**, which shows a $235.3 decrease in sales. This indicates that the product is not in high demand among buyers, and as a result, fewer sales are generated. Apart from the sub-categories having a significant impact on sales, **Nigeria, Iraq, and Turkey** do not generate many sales. Understanding these important elements will aid in determining the best way to boost sales performance.

ix. The top customers who bring in more sales are not necessarily the reason for the superstore profit. The other customers who order often and bring in more profit are not regular customers. From the charts below, we see that only **Tamara Chand** orders more product and is more profitable to the superstore. A stacked bar chart was used to visualize the top 5 customers by sales and by profits.


Sales by Customer Name
Customer Name
Tom Ashbrook
$40K
Tamara Chand
$37K
Greg Tran
$36K
Christopher Conant
$35K
Sean Miller
$35K
$0K
$20K
$40K
Sales
Profit by Customer Name
Customer Name
Tamara Chand
$8.7K
Raymond Buch
$8.5K
Sanjit Chand
$8.2K
Hunter Lopez
$7.8K
Bill Eplett
$7.4K
$0K
$2K
$4K
$6K
$8K
Profit

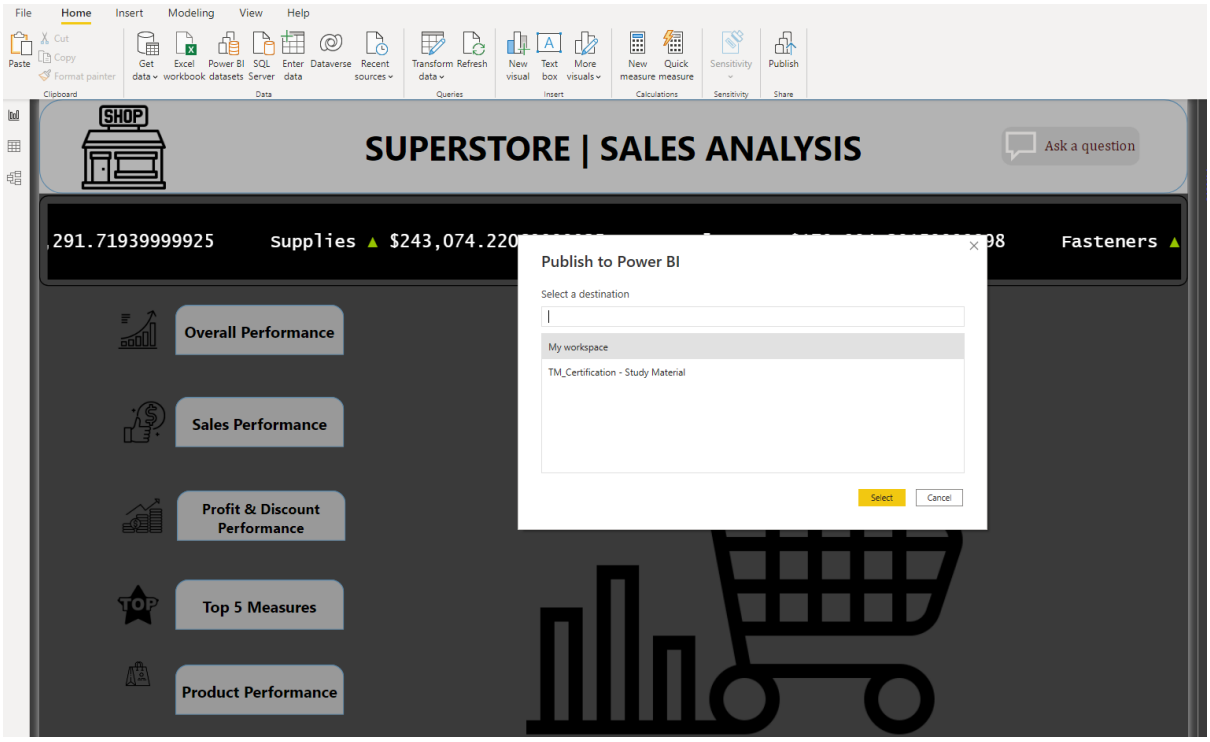


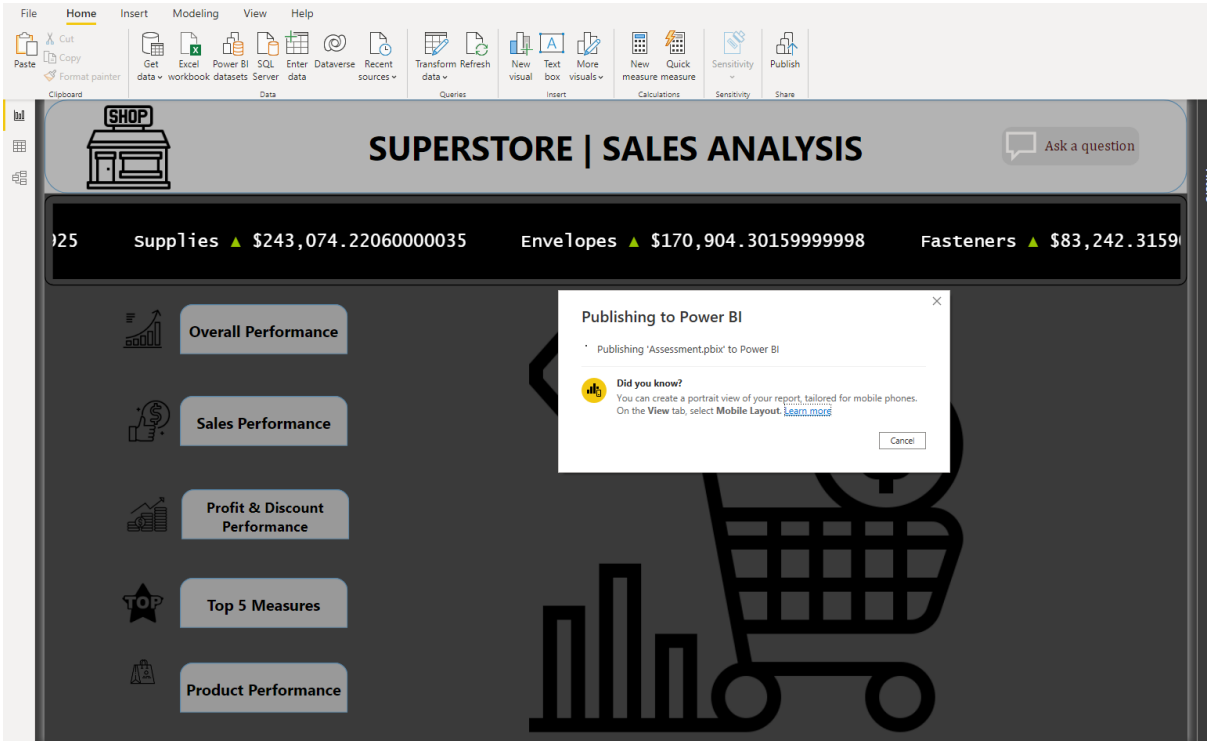
Figs: Key Influencers to decrease in sales

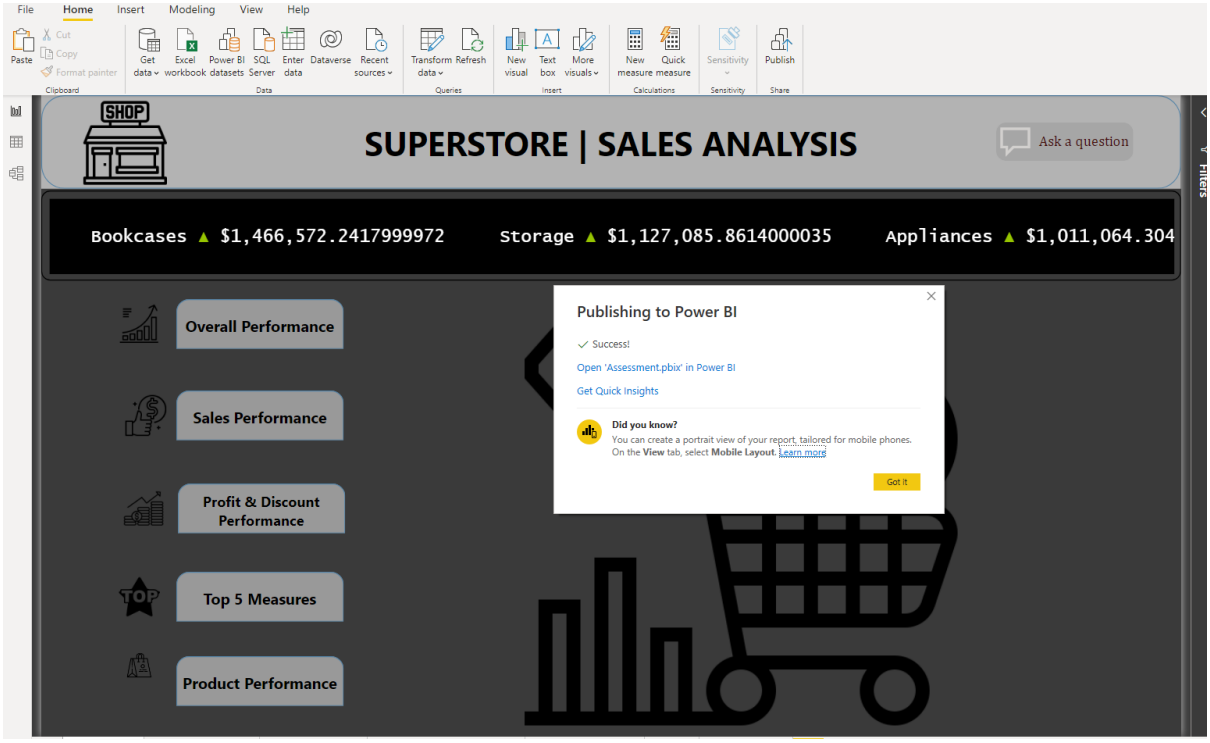
5 REPORT SUPPORT INFORMATION

This report includes a Power BI report (*.pbix) that contains the analysis. This report is divided into several pages, each of which focuses on a different area of sales performance, such as product, location, customer, profit, and discount. The report is published using the '**Publish**' button in

the **Home** ribbon, as shown in the figure below, which illustrates the report's processing and successful completion.

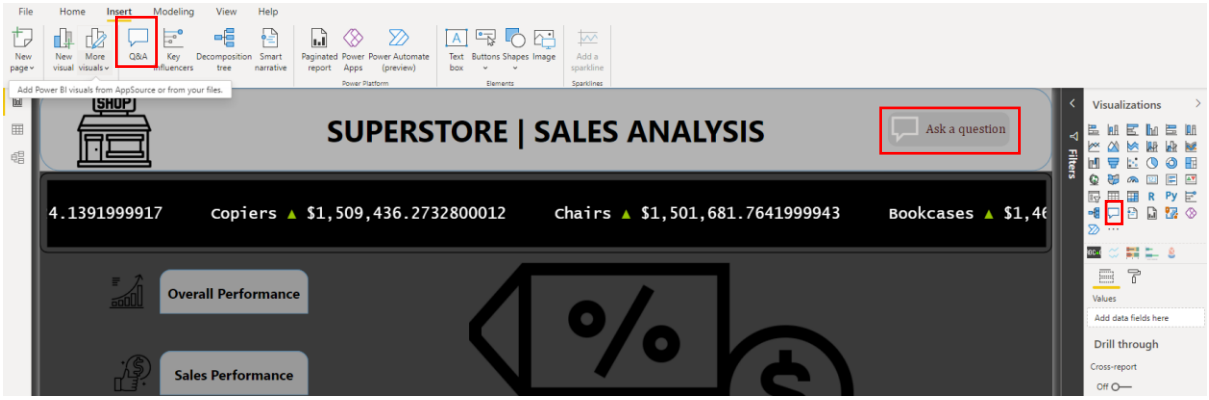






Figs: Publishing a report

The report and dashboard are also supported to provide a viewers to 'ask a question about the data using the 'QnA' button and this gives a prompt answer to data-related questions.



Power BI also proposes some data-related questions. If we wish to know the top countries by sales by city, Power BI will quickly present the answer to the inquiry, as shown below, using the most appropriate visualisation tool:

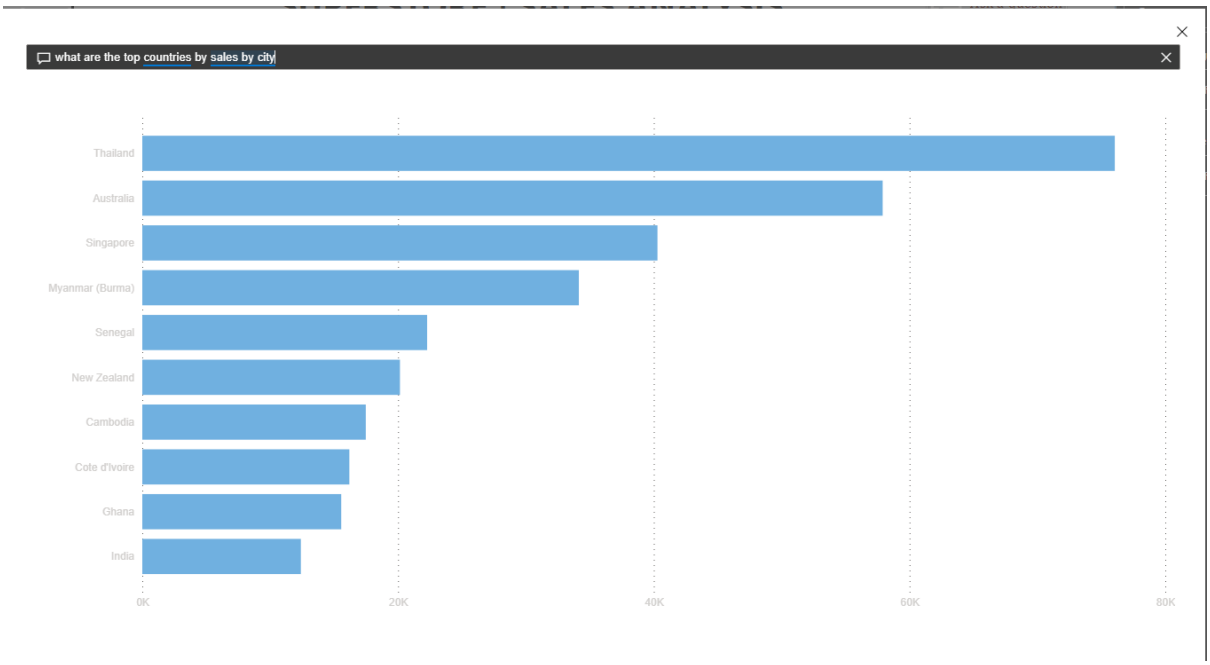


Fig: Power BI suggested question using QnA

Conclusion

- As the year progresses, sales and profit rise.
- The average number of days it takes to deliver the products is four.
- Technological products are the best-selling products over the years.
- There are usually more sales in the last quarter of the year.
- The more discounts each product group received, the more profit they n retrospect, the consumer segment products received the highest discount of $3.8k and, as a result, the highest profit of $750k.
- Customers who contributed the most to superstore sales are not the reason for superstore profit because they are the ones who order the most discounted items.

Recommendations

- Reducing the number of delivery days to two days would increase customer satisfaction and possibly the sales too.
- To increase profitability, EMEA, African, and Canadian markets should be targeted.

- Customers should be permitted to rate products because this may assist the superstore choose which products to focus on more and which products are the best sellers.
- Customers would have better shopping experiences and sales if they are segmented based on their profile and references.
- Some of the clients who contribute the most to profit aren't regulars. To improve profits, several tactics such as introducing membership cards and promotions can be used to keep customers.