

# BIG DATA AND BUSINESS INTELLIGENCE

Module Code: CIS4008-N

## Business Intelligence Solution and Report

NAME: ROHITH VADDEPALLY

STUDENT ID: W9529921



# TABLE OF CONTENTS

<b>SECTION 1: BUSINESS INTELLIGENCE DESIGN.....</b>	<b>3</b>
<b>BI DATA SOURCE DESCRIPTION AND BI REQUIREMENTS .....</b>	<b>4</b>
Data Source .....	4
Description of Dataset .....	4
<b>BI REQUIREMENTS .....</b>	<b>10</b>
BI Questions .....	10
Intent .....	10
Rationale .....	10
Key Users and Key Features .....	11
Scope.....	11
End Result .....	11
<b>BIG DATA PRE-PROCESSING OR DATA CLEANING .....</b>	<b>12</b>
Data Loading.....	12
Data Pre-Processing.....	14
Removing Columns .....	15
Replacing Values.....	16
Renaming Columns .....	17
Splitting Columns.....	19
Changing Datatypes .....	21
Dealing with errors .....	23
Merging Columns.....	24
Dealing with Missing Values.....	26
<b>BIG DATA MODELLING VIA STAR SCHEMA-FACTS AND DIMENSIONS.....</b>	<b>29</b>
Creating Measures and Calculated Columns .....	40
<b>SECTION 2: BUSINESS INTELLIGENCE SOLUTION .....</b>	<b>42</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>43</b>
Key Findings .....	43
Recommendations.....	44
<b>INTRODUCTION .....</b>	<b>43</b>
BI Questions .....	45
Description of the data used in the model .....	46

<b>FINDINGS BASED ON ANALYSIS AND EVALUATION .....</b>	<b>46</b>
Timeline Slicer chart (Chart not covered in lesson).....	65
Buttons.....	67
Screenshots of the final dashboard.....	68
Publishing the Power BI Report.....	72
<b>CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>73</b>
Summary of the Report .....	73
Findings.....	73
Conclusions .....	73
Personal Conclusions .....	74
Recommendations.....	74
References .....	75

# SECTION 1

## BUSINESS INTELLIGENCE DESIGN



# BI DATA SOURCE DESCRIPTION AND BI REQUIREMENTS

## Data Source:

The Dataset chosen for this **BIG DATA AND BUSINESS INTELLIGENCE ICA** is "[Data Co SMART SUPPLY CHAIN FOR BIG DATA ANALYTICS](#)". This Dataset is available in the website **Kaggle.com** for public use and it's free and link to this Dataset is given below.

<https://www.kaggle.com/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

Providing the screenshot of this Dataset below:

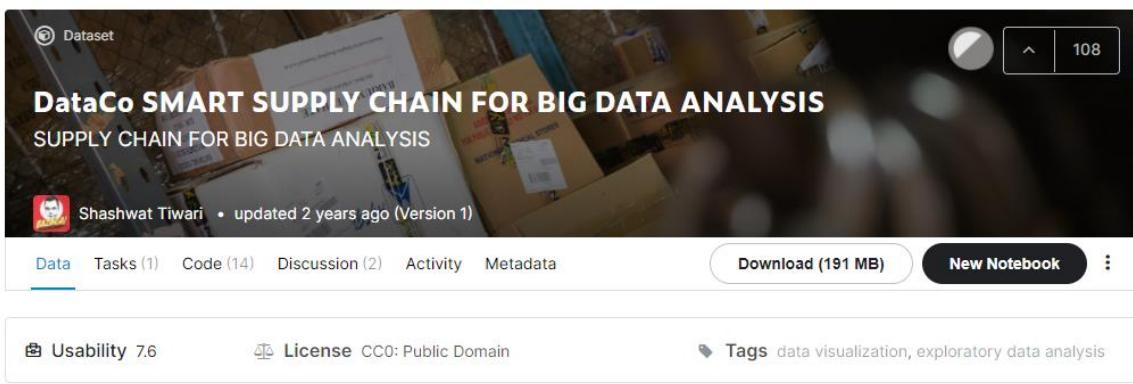


Figure 1: Screenshot of Data source.

## Description of Dataset:

This Dataset was used by a company called Data Co Global for data analysis. It allowed them to use different machine learning algorithms and R software. It is also used to know the customer behaviour and big data analysis as mentioned in the Kaggle website. But for this ICA project it is used for analysing the performance of all the sections like sales, production, shipment of this supply chain management process.

Understanding the Supply chain process:

**Supply chain** is the sequence of processes involved in the production and distribution of a commodity.

SELLER → PRODUCER → CUSTOMER

The Dataset chosen contains a total of 53 columns in which there are some columns like Customer Email, Customer Phone Number etc. contains null values will be deleted (or) removed in Power BI further as the information of those particular columns will not be used in this ICA or those will be not useful in future as well.

As mentioned above the dataset **Data Co SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS** contains a single table of 53 columns and 1,80,519 rows but further those 53 columns will be split into six different tables in power BI. But as of now column name and description will be given below.

S.NO	COLUMN NAME	DESCRIPTION
1	Transaction Type	Type of transaction made through ➤ DEBIT ➤ TRANSFER ➤ PAYMENT ➤ CASH
2	Days for shipping (real)	Actual shipping days of the purchased product
3	Days for shipment (scheduled)	Days of scheduled delivery of the purchased product
4	Benefit per order	Earnings per order placed
5	Sales per customer	Total sales made per customer
6	Delivery Status	Delivery status of orders such as: ➤ Advance shipping ➤ Late delivery ➤ Shipping cancelled ➤ Shipping on time
7	Late_delivery_risk	Categorical variable that indicates if ➤ sending is late (1) ➤ it is not late (0)
8	Category Id	Product category code
9	Category Name	Description of the product category.
10	Customer City	City where the customer made the purchase.
11	Customer Country	Country where the customer made the purchase.
12	Customer Email	Customer's email
13	Customer Fname	Customer first name
14	Customer Id	ID of the Customer
15	Customer Lname	Customer last name
16	Customer Password	Customer's masked key
17	Customer Segment	Types of Customers such as:

		<ul style="list-style-type: none"> <li>➤ Consumer</li> <li>➤ Corporate</li> <li>➤ Home Office</li> </ul>
18	Customer State	State to which the store where the purchase is registered belongs
19	Customer Street	Street to which the store where the purchase is registered belongs
20	Customer Zip-code	Zip-code of customer's address
21	Department Id	Department code of store
22	Department Name	Department name of store
23	Latitude	Latitude corresponding to location of store
24	Longitude	Longitude corresponding to location of store
25	Market	Market to where the order is delivered: <ul style="list-style-type: none"> <li>➤ Africa</li> <li>➤ Europe</li> <li>➤ LATAM</li> <li>➤ Pacific Asia</li> <li>➤ USCA</li> </ul>
26	Order City	Destination city of the order
27	Order Country	Destination country of the order
28	Order Customer Id	Customer order code
29	order date (Date Orders)	Date on which the order is made
30	Order Id	Order code
31	Order Item Card prod Id	Product code generated through the RFID reader
32	Order Item Discount	Order item discount value
33	Order Item Discount Rate	Order item discount percentage
34	Order Item Id	Order item code
35	Order Item Product Price	Price of products without discount
36	Order Item Profit Ratio	Profit Ratio per order item
37	Order Item Quantity	Number of products per order

38	Sales	Value in sales
39	Order Item Total	Total amount per order
40	Order Profit Per Order	Profit from each Order
41	Order Region	<p>Region where the order is delivered:</p> <ul style="list-style-type: none"> <li>➤ Southeast Asia</li> <li>➤ South Asia</li> <li>➤ Oceania</li> <li>➤ Eastern</li> <li>➤ Asia</li> <li>➤ West Asia</li> <li>➤ West of USA</li> <li>➤ US Centre</li> <li>➤ West Africa</li> <li>➤ Central Africa</li> <li>➤ North Africa</li> <li>➤ Western Europe</li> <li>➤ Northern</li> <li>➤ Caribbean</li> <li>➤ South America</li> <li>➤ East Africa</li> <li>➤ Southern Europe</li> <li>➤ East of USA</li> <li>➤ Canada</li> <li>➤ Southern Africa</li> <li>➤ Central Asia</li> <li>➤ Europe</li> <li>➤ Central America</li> <li>➤ Eastern Europe</li> <li>➤ South of USA</li> </ul>
42	Order State	State of the region where the order is delivered
43	Order Status	<p>Status of the order:</p> <ul style="list-style-type: none"> <li>➤ COMPLETE</li> <li>➤ PENDING</li> <li>➤ CLOSED</li> <li>➤ PENDING_PAYMENT</li> <li>➤ CANCELLED</li> <li>➤ PROCESSING</li> <li>➤ SUSPECTED_FRAUD</li> <li>➤ ON_HOLD</li> <li>➤ PAYMENT REVIEW</li> </ul>
44	Product Card Id	Product code
45	Product Category Id	Product category code

46	Product Description	Description of the product
47	Product Image	Link of visit and purchase of the product
48	Product Name	Name of the product
49	Product Price	Price of the product
50	Product Status	Status of the product stock: ➤ 1-product is not available ➤ 0-product is available
51	Shipping date (Date Orders)	Exact date and time of shipment
52	Shipping Mode	The following shipping modes are presented: ➤ Standard Class ➤ First Class ➤ Second Class ➤ Same Day

Entire dataset is represented in Microsoft Excel file and screen shot of that Excel file is given below:

Type	Days for settle	Benefit per sales	Delivery s	Late_deliv	Category	Category	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Customer	Dep
DEBIT	3	4	91.25	314.64	Advance	s	0	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Cally	20755	Holloway	XXXXXXX	Consumer PR	5365 Nobl	725
TRANSFER	5	4	-249.09	311.36	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Irene	19492	Luna	XXXXXXX	Consumer PR	2679 Rusti	725
CASH	4	4	-247.78	309.72	Shipping	c	0	73	Sporting	San Jose	EE. UU.	XXXXXXX	Gillian	19491	Maldonad	XXXXXXX	Consumer CA	8510 Roun	95125
DEBIT	3	4	22.86	304.81	Advance	s	0	73	Sporting	Los Angel	EE. UU.	XXXXXXX	Tana	19490	Tata	XXXXXXX	Home Off CA	3200 Amb	90027
PAYMENT	2	4	134.21	298.25	Advance	s	0	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Orli	19489	Hendricks	XXXXXXX	Corporate PR	8671 Iron	725
TRANSFER	6	4	18.58	294.98	Shipping	c	0	73	Sporting	Tonawanc	EE. UU.	XXXXXXX	Kimberly	19488	Flowers	XXXXXXX	Consumer NV	2122 Hazy	14150
DEBIT	2	1	95.18	288.42	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Constance	19487	Terrell	XXXXXXX	Home Off PR	1879 Gree	725
TRANSFER	2	1	68.43	285.14	Late deliv	i	1	73	Sporting	Miami	EE. UU.	XXXXXXX	Erica	19486	Stevens	XXXXXXX	Corporate FL	7595 Cottc	33162
CASH	3	2	133.72	278.59	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Nichole	19485	Olsen	XXXXXXX	Corporate PR	2051 Dust	725
CASH	2	1	132.15	275.31	Late deliv	i	1	73	Sporting	San Ramo	EE. UU.	XXXXXXX	Oprah	19484	Delacruz	XXXXXXX	Corporate CA	9139 Blue	94583
TRANSFER	6	2	130.58	272.03	Shipping	c	0	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Germane	19483	Short	XXXXXXX	Corporate PR	4058 Quie	725
TRANSFER	5	2	45.69	268.76	Late deliv	i	1	73	Sporting	Freepor	EE. UU.	XXXXXXX	Freya	19482	Robbins	XXXXXXX	Consumer NY	3243 Shad	11520
TRANSFER	4	2	21.76	262.2	Late deliv	i	1	73	Sporting	Salinas	EE. UU.	XXXXXXX	Cassandra	19481	Jensen	XXXXXXX	Corporate CA	131 Sunny	93905
DEBIT	2	1	24.58	245.81	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Natalie	19480	Mcfadden	XXXXXXX	Corporate PR	2531 Wish	725
TRANSFER	2	1	16.39	327.75	Late deliv	i	1	73	Sporting	Peabody	EE. UU.	XXXXXXX	Kimberley	19479	Sharpe	XXXXXXX	Corporate MA	6417 Silve	1960
DEBIT	2	1	-259.58	324.47	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Sade	19478	Lancaster	XXXXXXX	Corporate PR	257 Harve	725
PAYMENT	5	2	-246.36	321.2	Late deliv	i	1	73	Sporting	Canovana	Puerto Rico	XXXXXXX	Brynn	19477	Giles	XXXXXXX	Corporate PR	7342 Hazy	729
CASH	2	1	23.84	317.92	Late deliv	i	1	73	Sporting	Paramour	EE. UU.	XXXXXXX	Clara	19476	Bird	XXXXXXX	Corporate CA	7787 Lazy	90723
DEBIT	2	1	102.26	314.64	Late deliv	i	1	73	Sporting	Caguanas	Puerto Rico	XXXXXXX	Bo	19475	Griffin	XXXXXXX	Consumer PR	5136 Rusti	725
PAYMENT	0	0	87.18	311.36	Shipping	c	0	73	Sporting	Mount Pr	EE. UU.	XXXXXXX	Kim	19474	Simon	XXXXXXX	Consumer IL	1723 Tawr	60056

Figure 2: Screenshot of the dataset in Excel file.

## **BI Requirements:**

This Power BI project is made in the point of view of an Analyst who is to report the performance of sales, profits, orders, shipment to their superiors in the company. Mainly the report is based on the supply chain performance in the FY 2015 and the report also discusses the fluctuations of the sales and predicts the sales and profits after March in the FY 2016.

**BI Questions:** The report of this ICA focus addressing the following questions:

- To provide the number of Total Sales, Cost for making order item product, Total Orders Profits and also calculate these with regarding to 'Fitness' Department Name in FY 2015?
- Give the minimum value of the Order item Product Price in the FY 2015. Find out the number of Order items of completed and pending order status. Give the orders count by different shipping mode. Evaluate the sales by each country in FY 2015?
- Evaluate the Sales KPI and find out the achievement with Order item total and cost of making the product in FY 2015?
- Provide the average number of deliveries happened in FY 2015. Find out what type of delivery speed service is chosen by the customers mostly in FY 2015?
- On which weekday the most sales are happening in FY 2015. To provide the highest and lowest profited category name in FY 2015?
- Analyse the sales and explain it by Customer country, Customer city, Department Name, Category Name, Product Name?
- Forecast the sales for the next 2 years and find out what could be the sales value on 1<sup>st</sup> February 2018?
- Find out why the decrease in sales happened in Jan-Feb period and increase in sales in Nov-Dec period of FY 2015?
- Provide the count of the orders with different range of discount rate?
- Create a Time Slicer chart and use it for monitoring the sales for the Quarter 4 of FY 2015.

## **Intent:**

With the training and involvement in **Big Data and Business Intelligence module, Business Intelligent Design (SECTION-1)** and **Business Intelligence Solution (SECTION-2)** I can say that these BI questions will help to determine the KPI of Sales, Cost of making of products and the profits made out of it over the period of time. This report and these Business Questions are designed to show my development in the business skills.

## **Rationale:**

This dataset is selected to analyse the performance of company's supply chain management process through different visuals and dashboard as this dataset has wide range of columns and rows with various factors of information in it which makes to showcase my skills in Pre-processing, Modelling, Findings, Evaluations and Report writing through **POWER BI**.

### **Key Users and Key Features:**

As mentioned before the Key Users of this dataset are managers and their superiors of the company in different fields of this process. The key features of this ICA project are to show the profits, sales, orders, deliveries etc to them with the use of different charts and infographic visuals.

### **Scope:**

This selected dataset addresses the Big Data problem which is having high number of columns in one table. This is solved by normalizing them into different tables which is clearly mentioned Data Modelling in section-1. Solving this problem gives the results with accuracy and precision. The business process involved with discount rates, sales, cost of making the products and forecasting the sales are the most critical to measure in the report.

### **End Result:**

Outcomes of this project are given below:

- Data Pre-processing or Data cleaning
- Normalisation of data and data modelling with relationships
- Creating calculated columns and measures with DAX formulae
- Usage of different charts and infographics for visualization
- Creating and designing interactive dashboards
- Publishing the report

# BI DATA PRE-PROCESSING OR DATA CLEANING

## Data Loading:

The first step is to load the dataset into the power BI desktop and for that open the Power BI desktop in the system. After opening it the desktop screen will be shown as below.

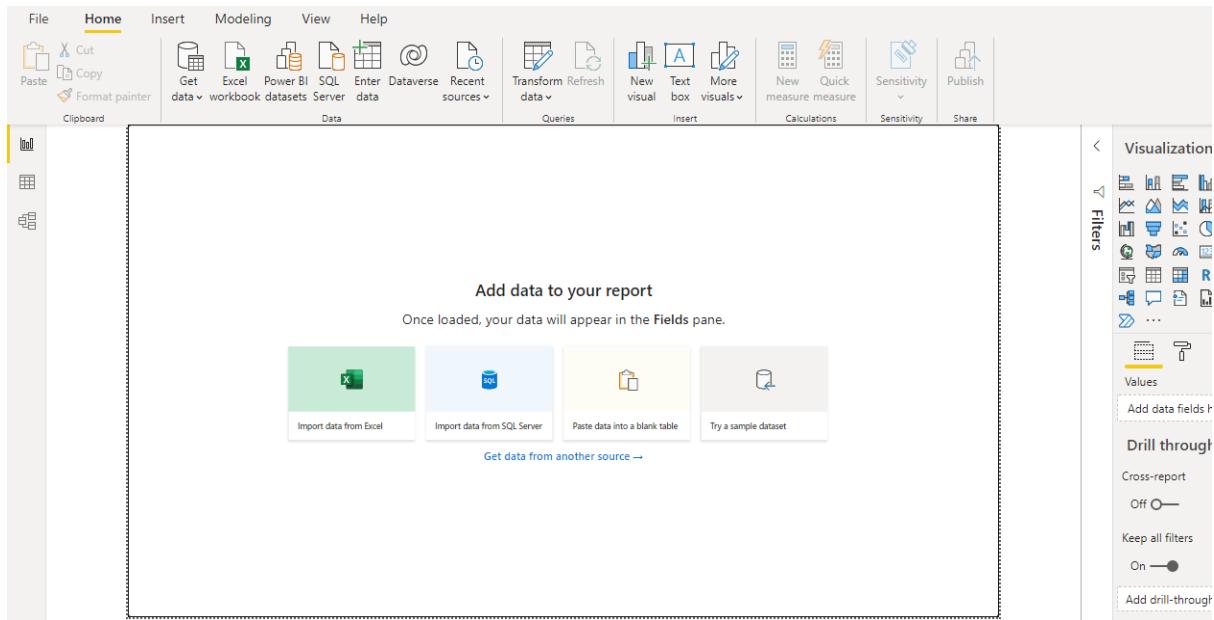


Figure 3: Screenshot of Microsoft power BI after starting it.

Now, click on 'Import data from Excel'

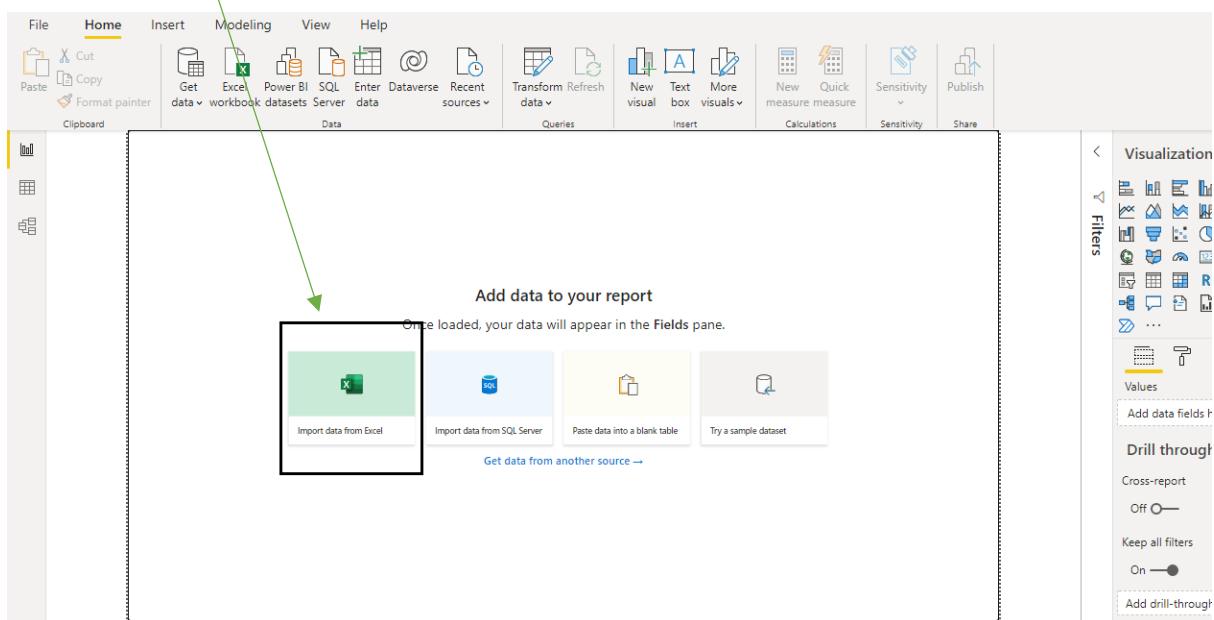


Figure 4: Screenshot of Power BI screen before importing the data.

Select the “**DataCoSupplyChainDataset**” Excel file and then click on open. Then the screen will look like below.

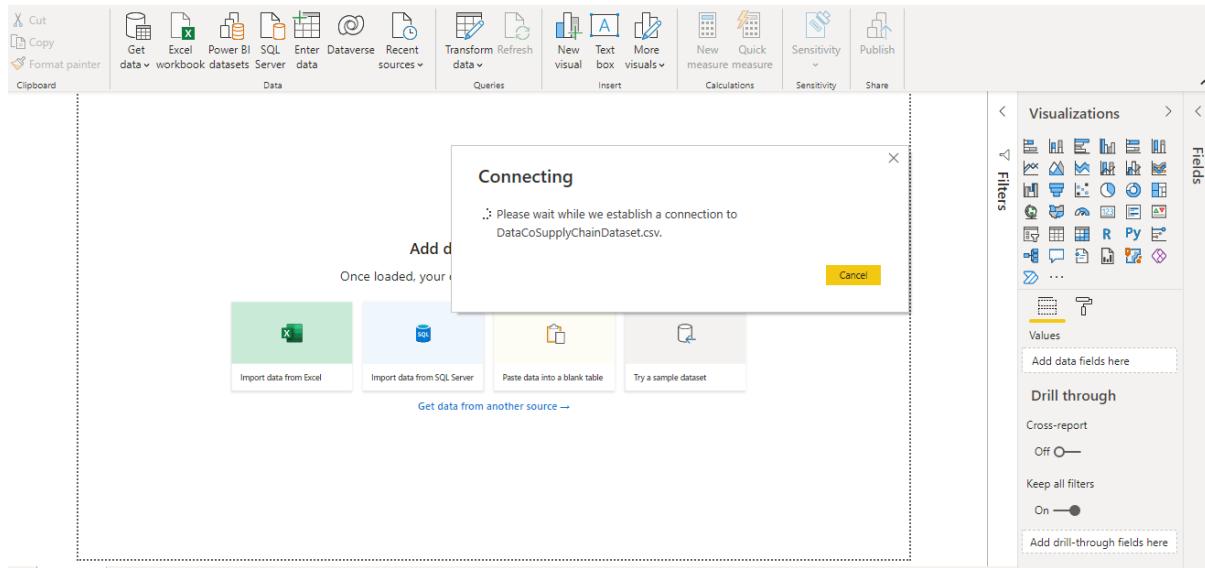


Figure 5: Screenshot of Power BI screen after selecting the dataset.

TransactionType	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery
DEBIT PAYMENT	3	4	91.25	314.6400146	Advance shipping	
BANK TRANSFER	5	4	-249.089963	311.3599854	Late delivery	
CASH	4	4	-247.779988	309.7200012	Shipping on time	
DEBIT PAYMENT	3	4	22.86000061	304.8099976	Advance shipping	
CREDIT PAYMENT	2	4	134.2100067	298.25	Advance shipping	
BANK TRANSFER	6	4	18.57999992	294.9800112	Shipping canceled	
DEBIT PAYMENT	2	1	95.18000031	288.4200134	Late delivery	
BANK TRANSFER	2	1	68.43000031	285.1400148	Late delivery	
CASH	3	2	133.7200012	278.5899963	Late delivery	
CASH	2	1	132.1499939	275.3099976	Late delivery	
BANK TRANSFER	6	2	130.5800018	272.0299984	Shipping canceled	
BANK TRANSFER	5	2	45.68999863	268.7600098	Late delivery	
BANK TRANSFER	4	2	21.7600023	262.2000122	Late delivery	
DEBIT PAYMENT	2	1	24.57999992	245.8099976	Late delivery	
BANK TRANSFER	2	1	16.38999939	327.75	Late delivery	
DEBIT PAYMENT	2	1	-259.5799866	324.4700012	Late delivery	
CREDIT PAYMENT	5	2	-246.3600006	321.2000122	Late delivery	
CASH	2	1	23.84000015	317.9200134	Late delivery	
DEBIT PAYMENT	2	1	102.2600021	314.6400146	Late delivery	
CREDIT PAYMENT	0	0	87.18000031	311.3599854	Shipping on time	

Figure 6: Screenshot of Power BI screen after selecting the dataset.

Now, click on “**Load**”. Then the dataset will be loaded into the power BI.

After that, we can find three tabs on the left-hand side of the screen of the power BI. Those are:

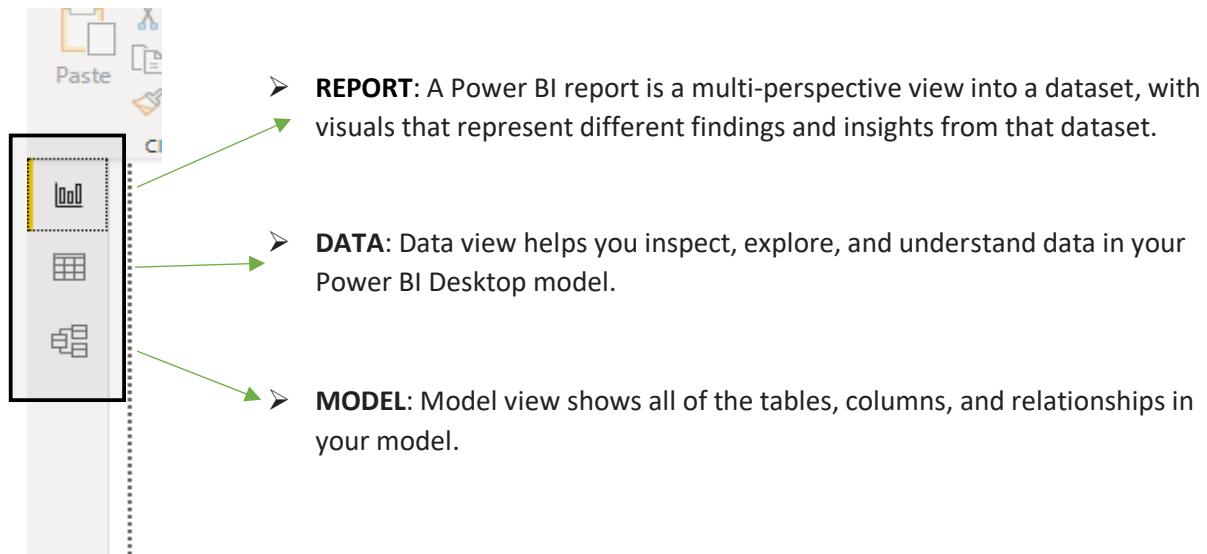


Figure 7: Screenshot of Power BI screen which shows **REPORT**, **DATA**, **MODEL** view.

### **Data Pre-Processing:**

After that, data pre-processing and data cleaning needs to be done to the dataset in Power Query editor within the power BI. For this particular dataset the following are done:

- Removing Columns
- Replacing Values
- Renaming Columns
- Splitting Columns
- Changing Datatypes
- Dealing with errors
- Merging columns
- Dealing with Missing Values

And other than this some data pre-processing is done in section 2.

Click on the ‘Transform data’ in the queries section which can be seen on the top of the power BI screen in report view tab. Screenshot of this is given below.

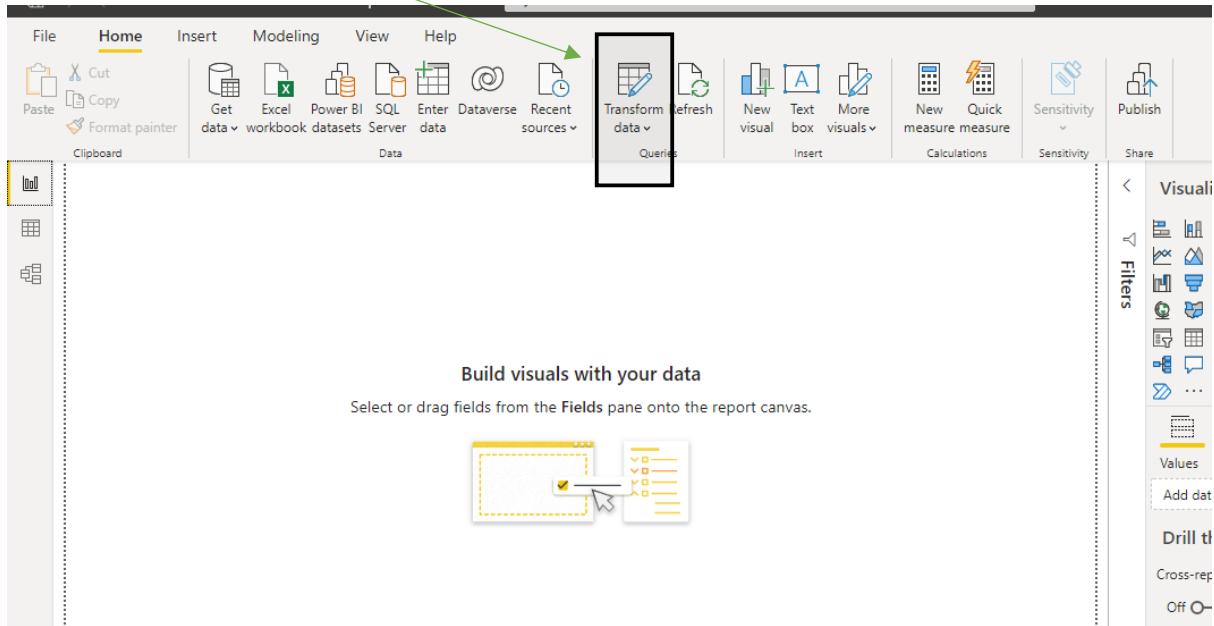


Figure 8: Screenshot of power BI screen before opening the ‘Power Query Editor’.

#### Removing Columns:

After that some of the columns need to be removed as those columns are no use for us to perform Analytics and Business Intelligence questions. Those columns are Customer Email, Customer Password, Product Description, Product Status. The reason why we are removing these particular columns is because

- Customer Email: Contains XXXXX data in all the rows.
- Customer Password: Contains XXXXX data in all the rows.
- Product Description: Contains blank data in all the rows.
- Product Status: Contains value ‘0’ in all the rows.

For removing these columns, we need to select them in Power Query Editor and then right click on any one of the columns and click on ‘Remove Columns’ and that screenshot is given below.

The screenshot shows the Power BI Power Query Editor interface. The 'Transform' tab is active in the ribbon. A context menu is open over three selected columns: 'Customer Email', 'Customer Password', and 'Product Description'. The 'Remove Columns' option is highlighted in the menu. The 'APPLIED STEPS' pane on the right shows the steps taken: 'Source', 'Promoted Headers', 'Changed Type', and 'Reordered Columns'.

Figure 9: Screenshot of Power BI screen while removing the selected columns.

### Replacing values:

Now, we are replacing the old text to the new text in the Type column. For that we are using Replace values in the transform section.

The text we are changing are:

- DEBIT –> DEBIT PAYEMNT
- TRANSFER -> BANK TRANSFER
- CASH -> CASH PAYMENT
- PAYMENT -> CREDIT PAYMENT

The steps for this are demonstrated below with screenshots. Go to the transform tab and then click on the ‘Replace Values’.

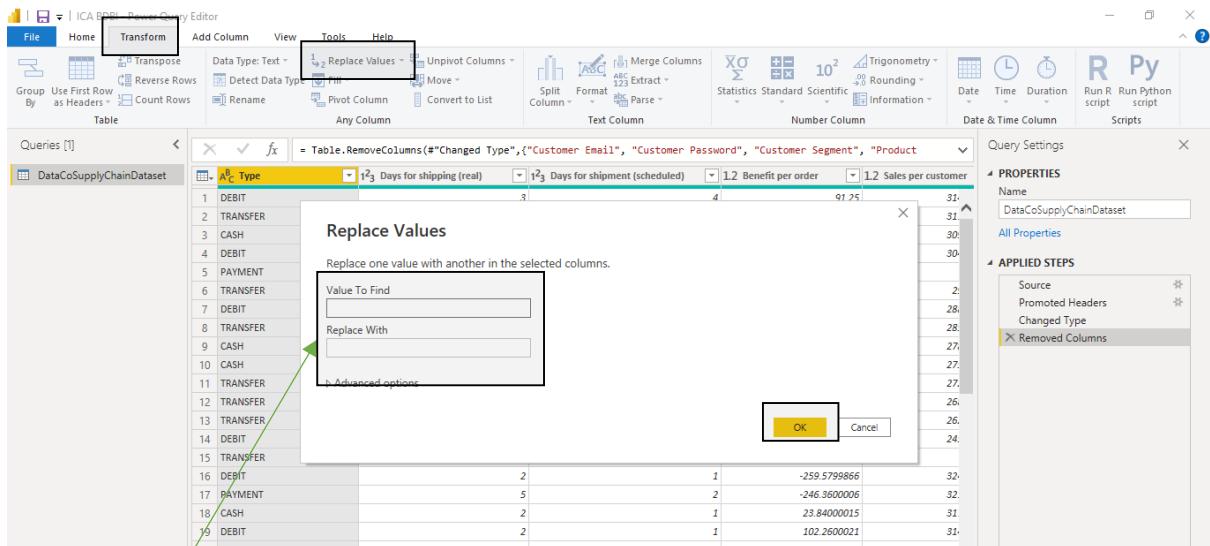


Figure 10: Screenshot of Power BI before replacing the values.

Then enter the values in ‘Values To Find’ and ‘Replace With’ and then click on OK.

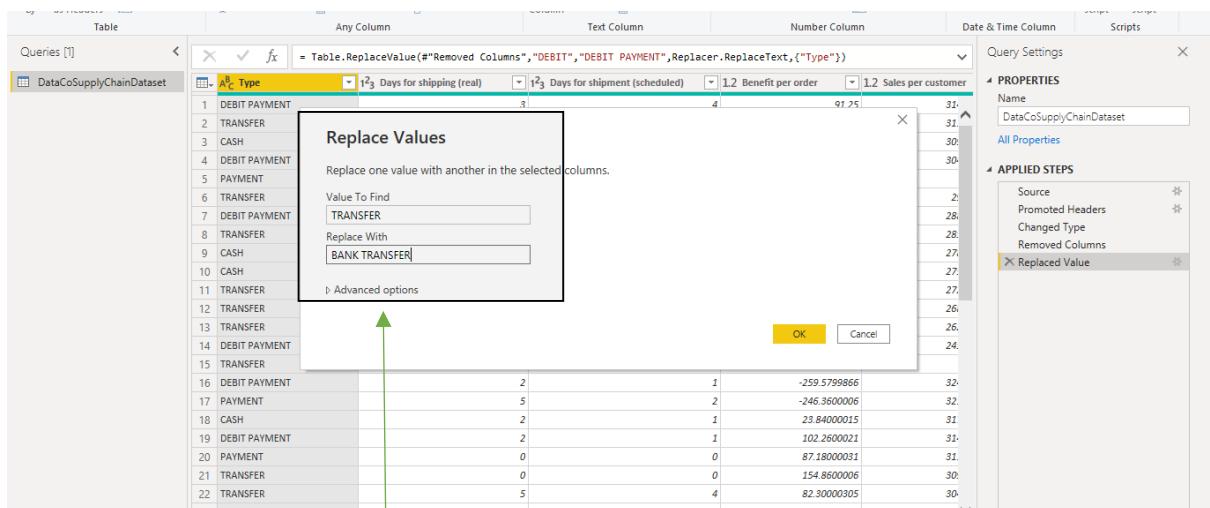


Figure 11: Screenshot of Power BI while replacing the values.

Enter ‘TRANSFER’ in Value To Find and ‘BANK TRANSFER’ in Replace with tab.

After that the Power BI screen will look like below with the new replaced values. Repeat the same for DEBIT, CASH and PAYMENT values.

The screenshot shows the Power BI desktop interface with the 'DataCoSupplyChainDataset' query selected. The 'Transform' tab is active. A context menu is open over the 'Type' column header, showing options like 'Replace Values', 'Unpivot Columns', 'Merge Columns', etc. The 'APPLIED STEPS' pane on the right lists the steps taken so far, including the most recent 'Replaced Value3' step.

Figure 12: Screenshot of Power BI after replacing the values.

### **Renaming Columns:**

Now, we will rename some of the columns in our dataset in the **Power Query Editor**. Steps for renaming the columns are demonstrated below with screenshots.

Double click on the column for which we need to rename it. In this case we are changing 'Type' column name to 'Transaction type'.

The screenshot shows the Power BI desktop interface with the 'DataCoSupplyChainDataset' query selected. The 'Transform' tab is active. A green arrow points from the 'Manage Columns' button in the ribbon to the 'Type' column header in the table, indicating where to click to rename it. The 'APPLIED STEPS' pane on the right lists the steps taken so far, including the most recent 'Replaced Value3' step.

Figure 13: Screenshot of Power BI while renaming the column.

Now change it to ‘Transaction Type’ and the result is shown below screenshot.

The screenshot shows the Power BI interface with the 'Data Co Supply Chain Dataset' query selected. In the ribbon, the 'Transform' tab is active. A green arrow points from the 'Transform' tab to the 'Properties' section of the ribbon, which contains a 'Renamed Columns' button. Below the ribbon, the query editor shows a table with the first column renamed to 'Transaction Type'. The 'APPLIED STEPS' pane on the right lists the steps taken, including 'Renamed Columns'.

Figure 14: Screenshot of Power BI after renaming the column.

Also renaming **Late\_delivery\_risk** to late delivery risk factor, so repeat the same steps for this column too. There will be more renaming columns further in this report.

### Splitting Columns:

After renaming columns, our next step is to split one column into two different columns. With this dataset we are going to split **order date (Date Orders)** into two separate columns and **shipping date (Date Orders)** into two separate columns. After splitting up order date (Date Orders) columns we are going to see one column for date and one for time. Steps for splitting the columns are given below with screenshots.

First select the column ‘**order date (DateOrders)**’ and click on ‘**Split Column**’ on the top of the ‘**Home**’ tab. In ‘**Split Column**’ section click on ‘**By Delimiter**’.

The screenshot shows the Power BI interface with the 'Data Co Supply Chain Dataset' query selected. In the ribbon, the 'Transform' tab is active. A green arrow points from the 'Transform' tab to the 'Split Column' button in the ribbon's 'Transform' section. The 'order date (DateOrders)' column is selected in the query editor. The 'APPLIED STEPS' pane on the right shows the previous renaming steps.

Figure 15: Screenshot of Power BI before splitting the column.

In ‘Split Column By Delimiter’ tab select ‘Space’ in enter delimiter section and then click on ‘OK’.

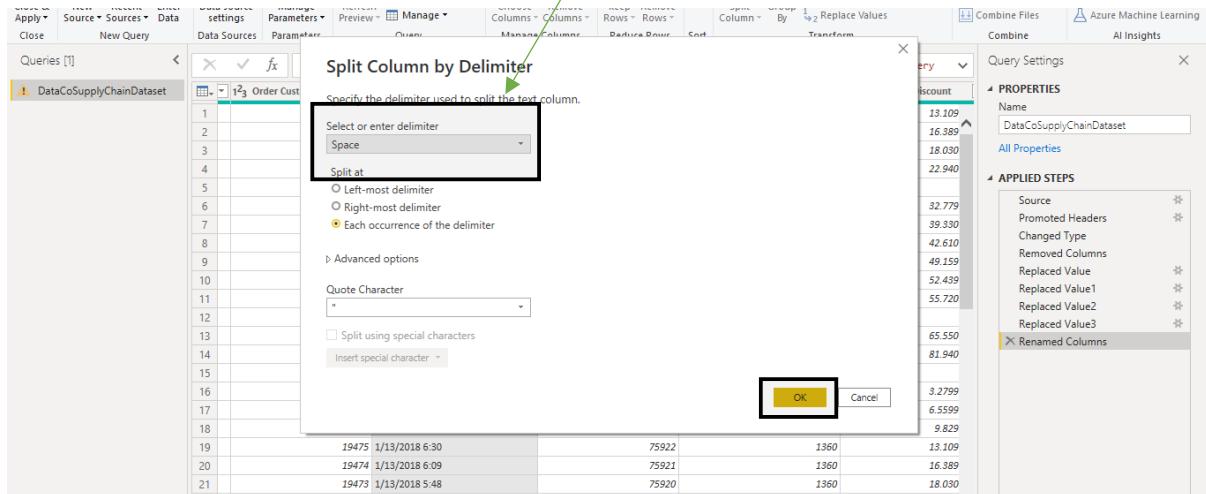


Figure 16: Screenshot of Power BI in ‘Split Column By Delimiter’ section.

After clicking on ‘OK’ the order date (DateOrders) columns will be split into order date (DateOrders).1 and order date (DateOrders).2. This is shown in the below screenshot.

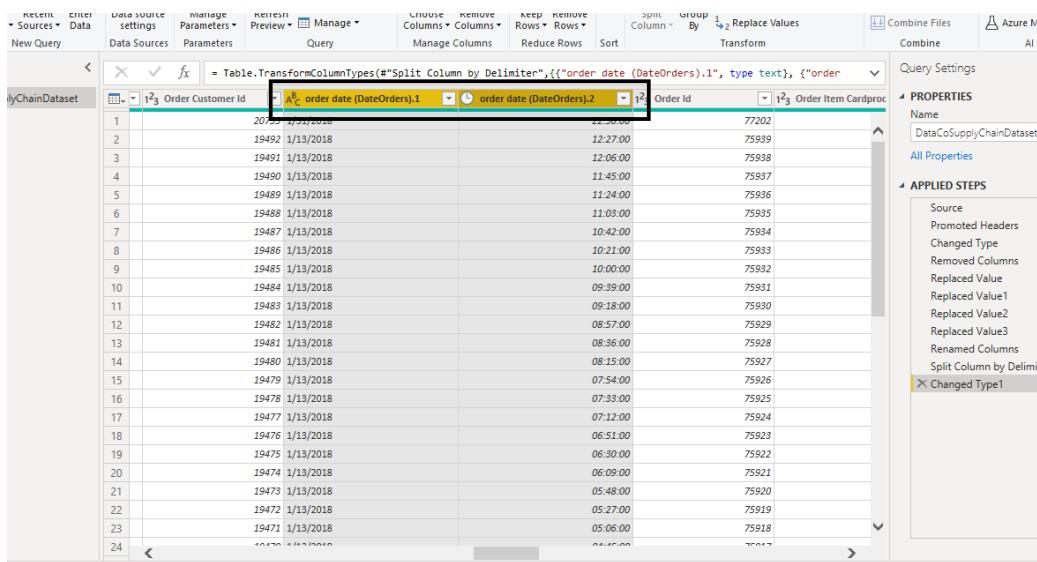


Figure 17: Screenshot of Power BI after splitting the column.

Now the next step is renaming these two columns. Order date (DateOrders).1 is renamed to ‘Orders Date’ and Order date (DateOrders).2 is renamed as ‘Orders Time’. For Renaming the columns detailed steps with screenshots are demonstrated above in this report. When done renaming the column names the next step is to change the datatypes of these two columns.

## Changing Datatypes:

The datatype for these two columns is string (or) text by default. We need to change it to respective datatypes and in order to do that click on the datatypes button which is on ‘Orders Time’ column header that we are going to change and it is shown in the below screenshot.

This screenshot shows the Power BI desktop interface with a query editor open. The 'Orders Date' column is currently set to 'Text'. A context menu is open over the 'Orders Time' column, with 'Time' selected. The 'Orders Time' column now displays time values like 11:45:00, 12:27:00, etc.

Order Customer Id	Orders Date	Orders Time	Order Id	Order Item Cardprod Id
20755	1/31/2018	11:45:00	77202	
19492	1/13/2018	12:27:00	75939	
19491	1/13/2018	12:06:00	75938	
19490	1/13/2018	11:45:00	75937	
19489	1/13/2018	11:24:00	75936	
19488	1/13/2018	11:03:00	75935	
19487	1/13/2018	10:42:00	75934	
19486	1/13/2018	10:21:00	75933	
19485	1/13/2018	10:00:00	75932	
19484	1/13/2018	09:39:00	75931	
19483	1/13/2018	09:18:00	75930	
19482	1/13/2018	08:57:00	75929	
19481	1/13/2018	08:36:00	75928	
19480	1/13/2018	08:15:00	75927	
19479	1/13/2018	07:54:00	75926	
19478	1/13/2018	07:33:00	75925	
19477	1/13/2018	07:12:00	75924	

Figure 18: Screenshot of Power BI while changing the datatypes.

Then click on the ‘Time’ data type. Now the datatype for that column is changed from String to Time. The result is shown as below.

This screenshot shows the Power BI desktop interface with a query editor open. The 'Orders Date' column is now set to 'Text'. A context menu is open over the 'Orders Time' column, with 'Time' selected. The 'Orders Time' column now displays time values like 22:56:00, 12:27:00, etc.

Order Customer Id	Orders Date	Orders Time	Order Id	Order Item Cardprod Id
20755	1/31/2018	22:56:00	77202	
19492	1/13/2018	12:27:00	75939	
19491	1/13/2018	12:06:00	75938	
19490	1/13/2018	11:45:00	75937	
19489	1/13/2018	11:24:00	75936	
19488	1/13/2018	11:03:00	75935	
19487	1/13/2018	10:42:00	75934	
19486	1/13/2018	10:21:00	75933	
19485	1/13/2018	10:00:00	75932	
19484	1/13/2018	09:39:00	75931	
19483	1/13/2018	09:18:00	75930	
19482	1/13/2018	08:57:00	75929	
19481	1/13/2018	08:36:00	75928	
19480	1/13/2018	08:15:00	75927	
19479	1/13/2018	07:54:00	75926	
19478	1/13/2018	07:33:00	75925	
19477	1/13/2018	07:12:00	75924	

Figure 19: Screenshot of Power BI after changing to Time datatype.

**Note:** In the column ‘Orders Time’ now the data is showing as Hour: Minute: Seconds format but seconds is of no use as it is data contains ‘00’ values. So, click on Close and Apply and then in Power BI desktop go to Data view and here also we can see that it is in the Hour: Minute: seconds format.

	Orders Date	Orders Time
1	10 April 2015	14:40:00
2	18 August 2015	21:55:00
3	08 July 2015	10:33:00
4	17 September 2015	23:17:00
5	08 April 2015	00:13:00
6	04 May 2015	06:52:00
7	23 April 2015	18:03:00
8	15 February 2015	03:36:00
9	01 December 2016	12:05:00
10	01 December 2016	10:20:00
11	24 October 2015	23:45:00

Now go to Column Tools tab and drop down the format section. Then select \*13:30 (Short Time). After selecting it Seconds will be removed from it and Format will be changed to Hour: Minute.

Figure 20: Screenshot of Power BI before changing the format of the time.

After changing the format of the 'Orders Time' column looks like below.

Query Id	Product Image	Product Name	Product Price	shipping date (DateOrders)	Shipping Mode	Orders Date	Orders Time
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	10/10/2015 14:40	Second Class	10 April 2015	14:40
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	8/20/2015 21:55	Second Class	18 August 2015	21:55
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	8/13/2015 10:33	Second Class	08 July 2015	10:33
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	9/23/2015 23:17	Second Class	17 September 2015	23:17
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	8/10/2015 0:13	Second Class	08 April 2015	00:13
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	4/11/2015 6:52	Second Class	04 May 2015	06:52
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	4/29/2015 18:03	Second Class	23 April 2015	18:03
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	2/20/2015 3:36	Second Class	15 February 2015	03:36
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	1/16/2016 12:05	Second Class	01 December 2016	12:05
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	1/16/2016 10:20	Second Class	01 December 2016	10:20
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	10/26/2015 23:45	Second Class	24 October 2015	23:45
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	11/21/2015 16:01	Second Class	17 November 2015	16:01
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	11/1/2015 3:47	Second Class	28 October 2015	03:47
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	12/18/2015 21:14	Second Class	13 December 2015	21:14
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	11/30/2015 22:29	Second Class	26 November 2015	22:29
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	1/16/2016 12:05	Second Class	01 December 2016	12:05
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	7/2/2015 13:45	Standard Class	30 June 2015	13:45
45	http://images.acmesports.sports/Field+%26+Stream+Spo	Field & Stream Sportsman 16 Gun Fire Safe	399.980011	8/5/2015 12:18	Standard Class	08 March 2015	12:18

Figure 21: Screenshot of Power BI after changing the format of time.

Now, the next step is to change the type of Orders Date column as Date in Power Query Editor. By while doing it error has occurred and it can be seen in the screenshot below.

Recent Sources	Enter Data	Data settings	Manage Parameters	Refresh Preview	Properties Advanced Editor	Choose Columns	Remove Columns	Keep Rows	Remove Rows	Sort	Data Type: Date	Merge Queries	Text
New Query													
plyChainDataset													

Figure 22: Screenshot of Power BI when error occurred while changing the type.

## Dealing with errors:

To resolve these errors in 'Orders Date' column first we need to undo the step changed type. Then split the 'Orders Date' column in three columns i.e; day, month, year by each occurrence of delimiter '/'.

The screenshot shows the Power BI interface with a query editor. A green arrow points from the 'Columns' tab in the ribbon to the 'Orders Date' column in the table preview. The 'Orders Date' column is highlighted with a black box. The table preview shows 23 rows of data with columns: Order Customer Id, Orders Date.2, Orders Date.1, and Orders Date.3. The 'Orders Date' column has been split into these three integer columns. The 'Orders Time' column is also visible.

Figure 23: Screenshot of Power BI after splitting 'Orders Date' into three different columns.

After splitting up the columns the data type automatically changes to **INTEGER** and then rename those new columns to Day, Month, Year. The steps for renaming the column are explained above in this report and it can be seen in the below screen shot.

The screenshot shows the Power BI interface with a query editor. A black box highlights the 'Day', 'Month', and 'Year' columns in the table preview. The 'Orders Date' column has been renamed to 'Day', 'Month', and 'Year'. The 'Orders Time' column is also visible.

Figure 24: Screenshot of Power BI after renaming the Day, Month, year columns.

Now close and apply and then create a new column in the Table Tools tab with a name 'Orders date' with the formulae:

```
Orders Date =
DATE('DataCoSupplyChainDataset'[Year], 'DataCoSupplyChainDataset'[Month], 'DataCoSupplyChainDataset'[Day])
```

The screenshot shows the Power BI formula bar with the formula `DATE('DataCoSupplyChainDataset'[Year], 'DataCoSupplyChainDataset'[Month], 'DataCoSupplyChainDataset'[Day])` entered.

Figure 25: screenshot of formulae of the date when creating a new column.

After creation of new column Orders date column will look like below:

Day	Month	Year	Orders Date
18	9	2015	18 September 2015
18	9	2015	18 September 2015
1	11	2015	01 November 2015
1	1	2015	01 January 2015
1	1	2015	01 January 2015
12	3	2015	12 March 2015
15	1	2016	15 January 2016
28	11	2015	28 November 2015
5	1	2016	05 January 2016
23	12	2015	23 December 2015
13	9	2015	13 September 2015
30	6	2015	30 June 2015
3	8	2015	03 August 2015
19	9	2015	19 September 2015
23	6	2015	23 June 2015
9	10	2015	09 October 2015
19	8	2015	19 August 2015
13	10	2015	13 October 2015
11	9	2015	11 September 2015
20	10	2015	20 October 2015

Figure 26: Screenshot of Power BI after creating Orders Date column.

Now go to Column Tools and then change Data Type to 'Date' and the format to (dd mmmm yyyy).

Format	14 March 2001 (dd mmmm yyyy)
Summarization	Don't summarize
Date formats	<ul style="list-style-type: none"> <li>*14-03-2001 (Short Date)</li> <li>*14 March 2001 (Long Date)</li> <li>14 March 2001 (dd mmmm yyyy)</li> <li>Wednesday, 14 March, 2001 (dddd, d mmmm, yyyy)</li> <li>14-03-2001 (dd-mm-yyyy)</li> <li>14-03-01 (dd-mm-yy)</li> <li>14-3-01 (d-m-yy)</li> <li>14.3.01 (d.m.yy)</li> <li>2001-03-14 (yyyy-mm-dd)</li> <li>14-03-2001 (dd/mm/yyyy)</li> <li>March, 2001 (mmmm, yyyy)</li> <li>2001-03 (yyyy-mm)</li> <li>14 March (d mmmm)</li> <li>01 (yy)</li> <li>2001 (yyyy)</li> </ul>

Figure 27: Screenshot while changing the data type to Date.

Figure 28: Screenshot of changing the format to (dd mmmm yyyy).

(or)

### Merging Columns:

In the Power Query Editor select the three columns which are now renamed as Orders Day, Orders Month, Orders Year. Then we are going to merge these three columns as a single one. In Transform section click on **Merge columns** and after that a merge columns tab will pop up.

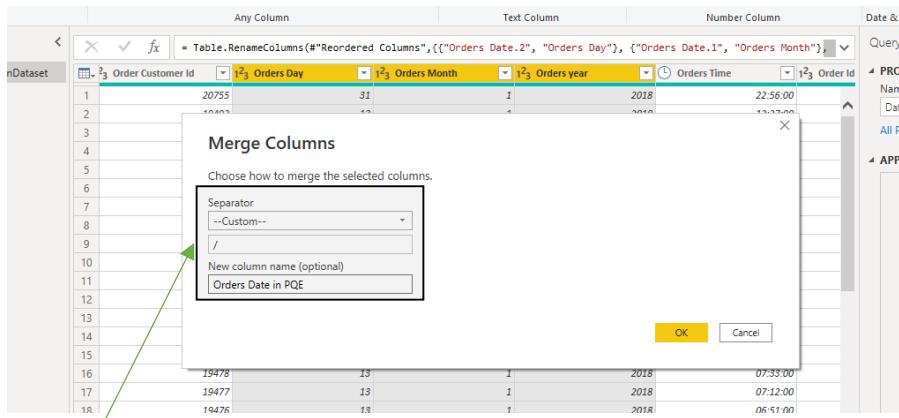


Figure 29: Screenshot of Power BI merging the columns in Power Query Editor.

In 'Separator' select Custom and then select '/' and give column name as Orders Date in PQE as shown in the above screenshot. Click on OK.

Now we can change the datatype of Orders date in PQE column successfully by clicking on type and then 'Date' type.

Figure 30: Screenshot of power BI while changing the type to 'Date'.

After that Orders Date column's date type is now completely changed to Date without any errors.

Customer Id	Orders Date in PQE	Orders Time	Order Id
20755	31-01-2018	22:56:00	
19492	13-01-2018	12:27:00	
19491	15-01-2018	12:06:00	
19490	13-01-2018	11:45:00	
19489	13-01-2018	11:24:00	
19488	13-01-2018	11:03:00	
19487	13-01-2018	10:42:00	
19486	13-01-2018	10:21:00	
19485	13-01-2018	10:00:00	
19484	13-01-2018	09:39:00	
19483	13-01-2018	09:18:00	
19482	13-01-2018	08:57:00	
19481	13-01-2018	08:36:00	
19480	13-01-2018	08:15:00	
19479	13-01-2018	07:54:00	
19478	13-01-2018	07:33:00	
19477	13-01-2018	07:12:00	
19476	13-01-2018	06:51:00	

Figure 31: Screenshot of Power BI after changing the type of Orders Date column.

**Note:** Previously while changing the type of Orders Date there were errors occurring. But we split and reordered that into Day, Month, Year columns for which Power BI automatically detected the data type as Integer. Again we merged those columns as one again and this time we are successful changing the type to 'Date'.

Repeat the same steps for the **Shipping Date (DateOrders)** columns as **orders Date(DateOrders)** like

- Splitting the column into two different columns.
- Renaming them as Shipping Date and Shipping Time.
- Changing their Date types.
- If errors occurred during changing of their datatypes then split the columns and merge them again

At the end the **Shipping Date** and **Shipping Time** column will look like below:

s [ ]	Product ID	Product Name	Product Price	Shipping Date	Shipping Time	Shipping Mode
1	327.5	atch	327.5	03-02-2018	22:56:00	Standard Class
2	327.5	atch	327.5	18-01-2018	12:27:00	Standard Class
3	327.5	atch	327.5	17-01-2018	12:06:00	Standard Class
4	327.5	atch	327.5	16-01-2018	11:45:00	Standard Class
5	327.5	atch	327.5	15-01-2018	11:24:00	Standard Class
6	327.5	atch	327.5	19-01-2018	11:03:00	Standard Class
7	327.5	atch	327.5	15-01-2018	10:42:00	First Class
8	327.5	atch	327.5	15-01-2018	10:21:00	First Class
9	327.5	atch	327.5	16-01-2018	10:00:00	Second Class
10	327.5	atch	327.5	15-01-2018	09:39:00	First Class
11	327.5	atch	327.5	19-01-2018	09:18:00	Second Class
12	327.5	atch	327.5	18-01-2018	08:57:00	Second Class
13	327.5	atch	327.5	17-01-2018	08:36:00	Second Class
14	327.5	atch	327.5	15-01-2018	08:15:00	First Class
15	327.5	atch	327.5	15-01-2018	07:54:00	First Class
16	327.5	atch	327.5	15-01-2018	07:33:00	First Class
17	327.5	atch	327.5	18-01-2018	07:12:00	Second Class
18	327.5	atch	327.5	15-01-2018	06:51:00	First Class
19	327.5	atch	327.5	15-01-2018	06:30:00	First Class
20	327.5	atch	327.5	13-01-2018	18:09:00	Same Day
21	327.5	atch	327.5	13-01-2018	17:48:00	Same Day
22	327.5	atch	327.5	18-01-2018	05:27:00	Standard Class
23	327.5	atch	327.5	17-01-2018	05:06:00	Second Class
24	327.5	atch	327.5	16-01-2018	04:45:00	...

Figure 32: Screenshot of Power BI after splitting up to **Shipping Date** and **Time** columns and changing the datatypes.

### Dealing with Missing Values:

Next comes the step which tells how to deal with the errors. In this dataset there are some missing values in the columns '**Order item Profit ratio**' and '**Order item Total**'. We are now going to see how to deal with the missing values in these two columns. There are four ways to deal with the missing values. They are:

- Elimination
- Identification
- Inspection
- Imputation

As the missing values in these columns is less than 10% we are going to eliminate the rows which contain those missing values. So we are going to use the first way to deal with the missing values which is '**Elimination**'. The percentage of missing values in these columns and missing values can be seen in the below screenshot.

	Order Item Profit Ratio	Order Item Total	Order Profit P
79	Order Item Profit Ratio 998 (99%) Valid	Order Item Total 999 (99%) Valid	Order Profit P
80	0 (0%) Error	0 (0%) Error	
81	2 (< 1% Empty)	1 (< 1% Empty)	
82			
83	Remove Empty	Remove Empty	...
84	...	...	...
85	50 0.159999996	327.75 303.7200012	
86	50 0.479999989	327.75 304.8099976	
87	50 null	327.75 298.25	
88	50 0.10000001	327.75 294.980011	
89	50 0.34999994	327.75 288.4200134	
90	50 0.07999998	327.75 null	
91	50 0.129999995	327.75 278.5899963	
92	0168 0.330000013	327.75 275.3099976	
93	50 0.039999999	327.75 272.0299988	
94	0168 0.289999992	327.75 268.7600098	
95	0168 null	327.75 262.2000122	
96	50 0.219999999	327.75 245.8099976	
97	50 -0.100000001	327.75 327.75	
98	30 0.49000001	327.75 324.4700012	
99	0168 0.059999999	327.75 321.2000122	
100	50 0.100000001	327.75 317.9200134	
101	50 0.300000012	327.75 314.6400146	
102	50 null	327.75 311.3500084	

Figure 33: Screenshots of Power BI in columns ‘Order Item Profit Ratio’ and ‘Order Item Total’ representing percentage of missing values.

Inorder to eliminate the complete row in that particular column click on the filter option of the column to drop it down and click on ‘Remove Empty’.

The screenshot shows the 'Column Quality' dialog for the 'Order Item Profit Ratio' column. The 'Remove Empty' button is highlighted with a green arrow. The dialog also shows other filter options like 'Sort Ascending', 'Sort Descending', and 'Clear Filter'.

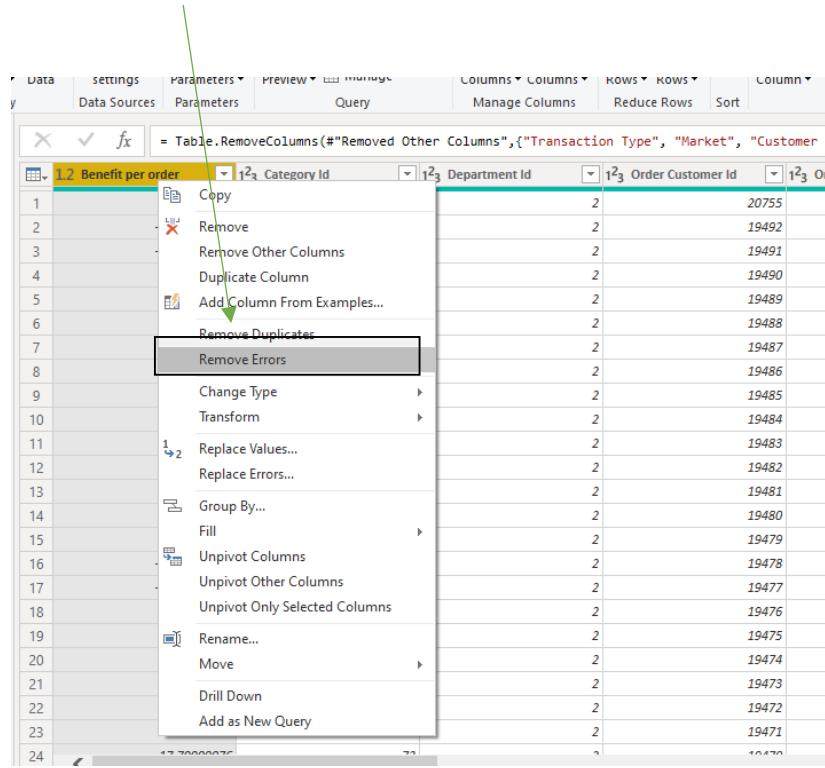
Figure 34: Screenshot of Power BI while removing the empty values in the ‘Order Item Profit ratio’ column.

After clicking on ‘Remove Empty’ the entire row which contains the ‘null’ or ‘missing’ values will be eliminated. Repeat the same with the ‘Order Item Total’ column as well. After performing these steps with these columns all the missing values will be eliminated.

**Note:** In this dataset the column ‘Order Zipcode’ contains null values with 97% and numerical data of 3%. In order to proceed with Data Modelling we should remove the whole column as the left 3% numerical data is of no use for BI questions and analytics. Please refer above to remove the columns in this report. Also there is a point to make here that removing rows will lead to data loss as null values are in higher percentag. So, it’s better to eliminate the whole column.

## Dealing with errors:

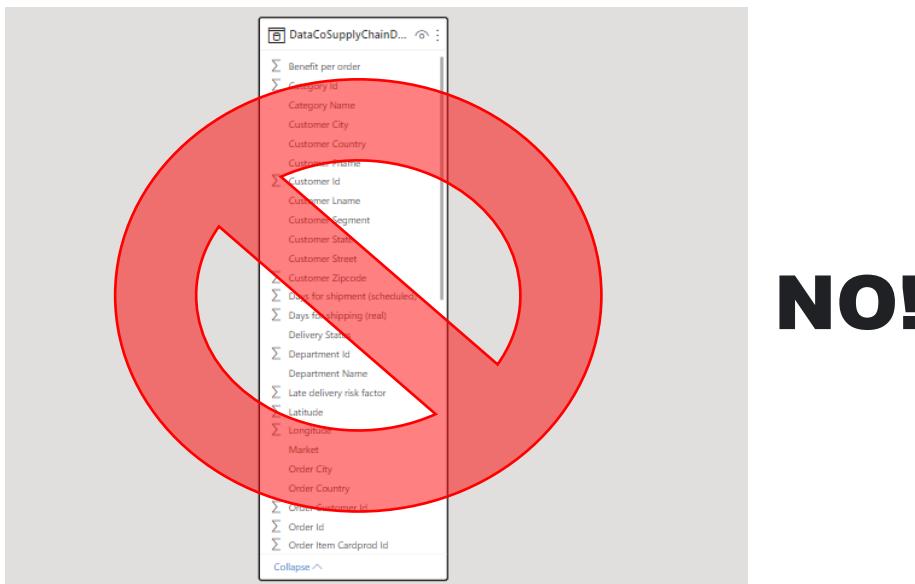
In this step we are going to see how to remove the errors and there are no errors in this dataset. But to explain how to remove the errors in a column for the showcase my skills in Power BI few steps are demonstrated below. Give a right click on the column which contains errors in it and then click on '**Remove Errors**'. By doing this errors will be eliminated from the dataset.



Screenshot of Power BI while removing the errors in the column.

## BI DATA MODELLING VIA STAR SCHEMA - FACTS AND DIMENSIONS

After Data Pre-processing or data cleansing next come Data Modelling which is the most important part of Business analytics and in this ICA too for providing results with accuracy. As we begin to do the modelling of data we could find that there is only one table in this dataset with which we can easily create some charts, graphs, infographics or even a dash board. But creating charts and dashboards with one table will lead to inaccurate results and will also lead to providing wrong information for the reader.



**NO!**

Figure 35: Screenshot of Power BI that not to use single table for dashboard.

As we decided not to proceed with one table, the next step is to create multiple tables from this first table (duplicating) and give some relationships to each table. First click on '**Transform data**' and enter into '**Power Query Editor**'. Give a right click on the table name then these options will pop up.

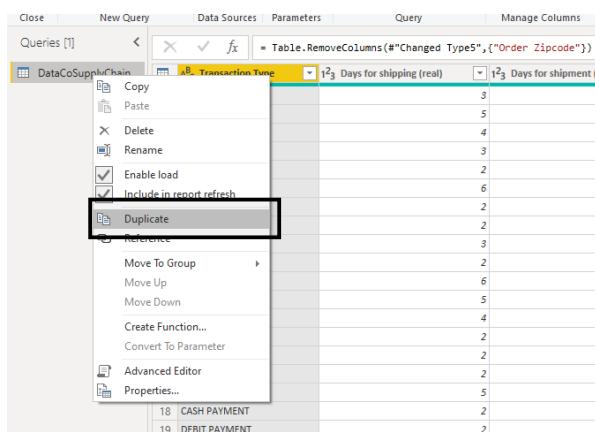


Figure 36: Screenshot of Power BI before duplicating the first table.

After clicking on duplicate another table will be created of the same original table which will be renamed as 'Products'. 'Products' table will be having columns:

- Product Card Id
- Product Category Id
- Product Image
- Product Name
- Product Price

The rest of the columns should be removed and for removing columns please refer above in this report in Data pre-processing section. It is shown in the below screenshot.

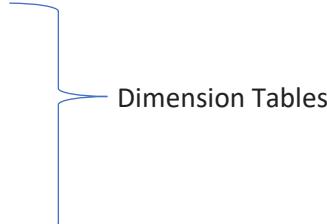
	Product Card Id	Product Category Id	Product Image	Product Name
1	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
2	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
3	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
4	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
5	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
6	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
7	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
8	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
9	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
10	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
11	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
12	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
13	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
14	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
15	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
16	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
17	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
18	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
19	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
20	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
21	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch
22	1360	73	http://images.acmesports.sports/Smart+watch	Smart watch

Figure 37: Screenshot of Power BI after duplicating the 'Products' table.

Repeat the same steps for creating multiple tables that are Category, Department, Customers, Order Details, Shipment, Order Sales.

The tables listed below are:

- Products
- Category
- Department
- Customers
- Order Details
- Shipment



- Order Sales



Category table will have the columns:

- Category Id
- Category Name

Department table will have the columns:

- Department Id
- Department Name

Customers table will have the columns:

- Customer Id
- Sales per customer
- Customer city
- Customer country
- Customer Name
- Customer Segment
- Customer state
- Customer street
- Sales per Customer

Order Details table will have the columns:

- Order Item Id
- Transaction Type
- Market
- Order city
- Order country
- Order region
- Order state
- Order status
- Orders Date
- Orders Time

Shipment table will have the columns:

- Order Id
- Shipping Mode
- Late Delivery risk factor
- Delivery Status
- Days for shipment (Scheduled)
- Days for Shipping (Real)
- Shipping Date
- Shipping Time

Order Sales will have the columns:

- Category Id
- Department Id
- Order Customer Id
- Order Id

- Order Item Card prod Id
- Transaction Type
- Benefit per order
- Order Item Discount
- Order Item discount Rate
- Order Item Product Price
- Order Item Profit Ratio
- Order Item Quantity
- Sales
- Order Item Total
- Order Profit per order

After duplicating the dimension and fact tables in Power BI will look below:

	Category
1	Sporting Goods
2	Sporting Goods
3	Sporting Goods
4	Sporting Goods
5	Sporting Goods
6	Sporting Goods
7	Sporting Goods
8	Sporting Goods
9	Sporting Goods
10	Sporting Goods
11	Sporting Goods
12	Sporting Goods
13	Sporting Goods
14	Sporting Goods

Figure 38: Screenshot of Power Bi after duplicating and creating dimension and fact tables.

Now click on the '**MODEL**' view and new tables can be seen with no relationships below:

Department	Department Id Department Name Collapse ^
Category	Category Id Category Name Collapse ^
Products	Product Card Id Product Category Id Product Image Product Name Product Price Collapse ^
Customers	Customer City Customer Country Customer Name Customer Id Customer Lname Customer Segment Customer Street Customer Zipcode Collapse ^
Order Sales	Benefit per order Category Id Department Id Market Order Customer Id Order Id Order Item Cardprod Id Order Item Discount Order Item Discount Rate Order Item Product Price Order Item Profit Ratio Order Item Quantity Order Item Total Order Profit Per Order Sales Shipping Mode Transaction Type Collapse ^
Shipment	Days for shipment (scheduled) Days for shipping (real) Delivery Status Late delivery risk factor Order Id Shipping Date Shipping Mode Shipping Time Collapse ^
Order Details	Market Order City Order Country Order Item Id Order Region Order State Order Status Orders Date Orders Time Collapse ^

Figure 39: Screenshot of Power BI in '**MODEL**' view without relationships.

Next step is to create relationship between these tables. For this click on '**Manage relationships**' in '**Home**' tab. Manage relationships tab will pop up on the screen, then click on '**New**'. After that we will get an opportunity to create a relationship with two tables.

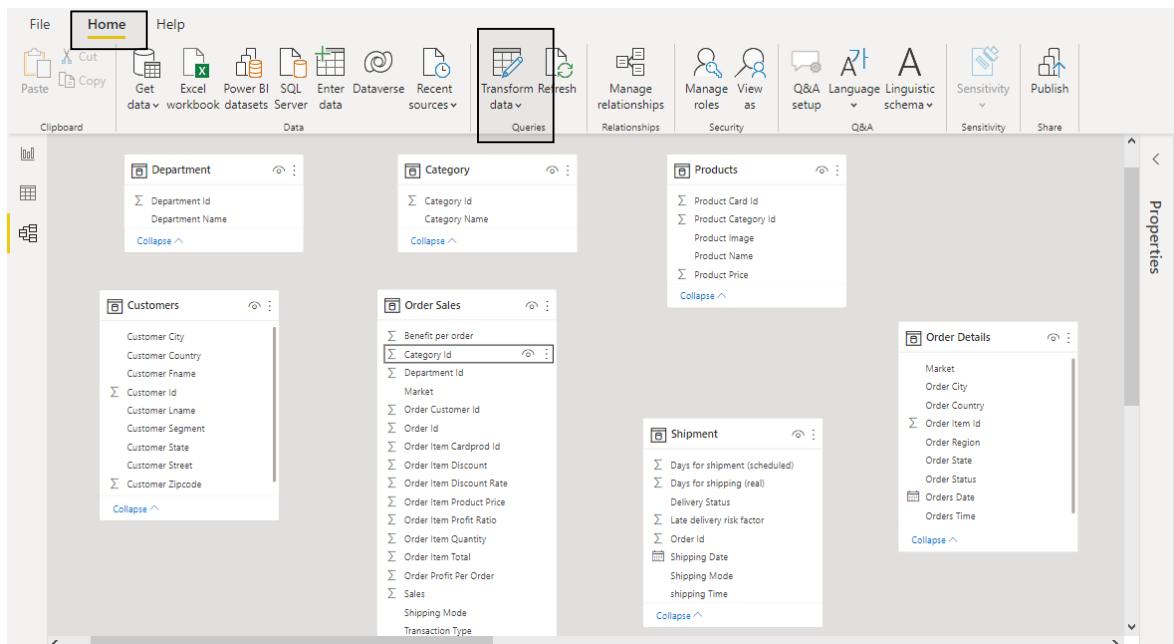


Figure 40: Screenshot of Power BI explaining the steps to create the relationships.

In Manage relationships click on ‘New’.

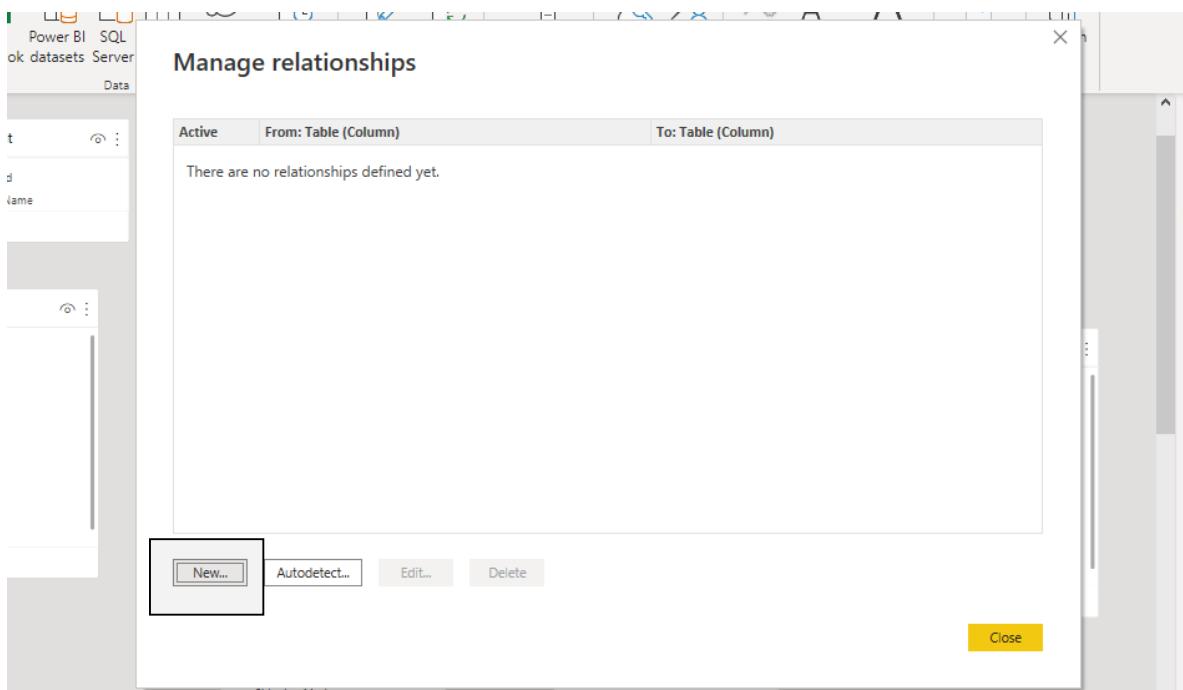


Figure 41: Screenshot of Power BI explaining the steps to create the relationships.

Now let’s try to create a relationship between the tables ‘Category’ and ‘Order Sales’. Select category and order sales in **Create relationship** tab.

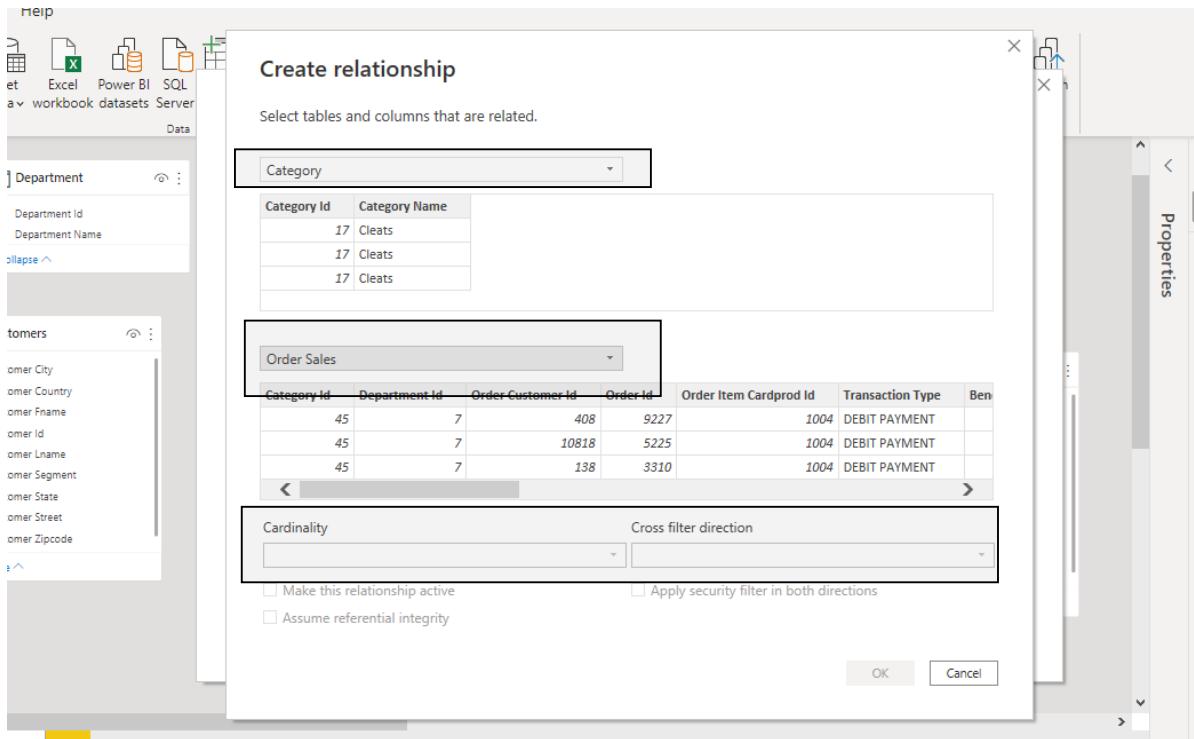


Figure 42: Screenshot of Power BI while creating the relationship.

At first the Power BI does not give us the Cardinality and cross filter direction. For that we have to select the columns in both the tables from which the relationships are created. In '**Category**' select Category Id and do the same in '**Order Sales**' table.

After selecting columns in both the tables Power BI automatically detects the Cardinality as '**Many to many (\*:\*)**' and Cross filter direction as '**Both**'.

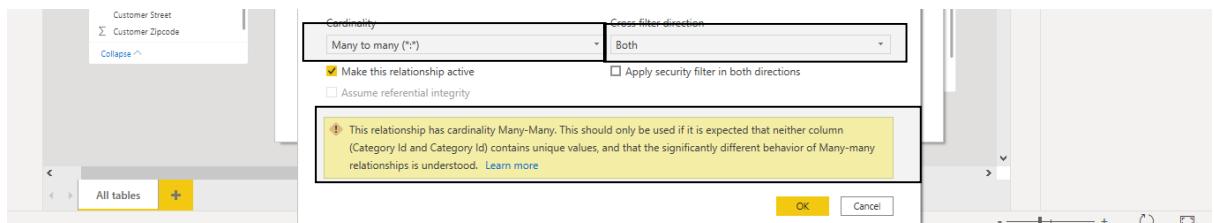


Figure 43: Screenshot of Power BI showing Cardinality and Cross filter direction while creating a relationship.

**NOTE:** When using Many to many relationships, the model won't be flexible to report visuals in our dashboard. Many to many relationships are used very rarely in the field of data science and it needs to be used when the values are unique in all the columns of the table. This cardinality gives the wrong information through charts when they are created in the dashboard. I don't recommend to use Many to many relationships while data modelling unless I absolutely had to because it can introduce some ambiguities and produce various questions in charts.

The above screenshot shows Power BI automatically created Many-to-many cardinality. So, for avoiding this first remove the relationship between the two tables and then remove duplicates in the Category table.

For deleting the relationship, give a right click and then **Delete**.

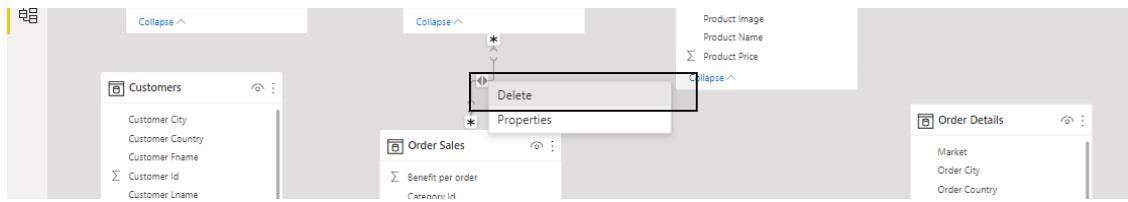


Figure 44: Screenshot of Power BI while deleting the relationship.

After deleting the relationship get into **Power Query Editor** and then in Category table select both the columns. Give a right click and select '**Remove Duplicates**'.

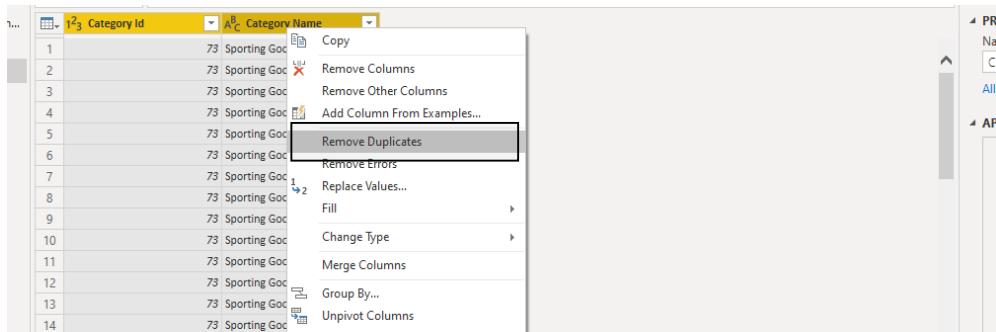


Figure 45: Screenshot of Power BI while removing the duplicates.

After removing duplicates, the Power Query Editor will look like below. Now **close and apply**.

The screenshot shows the Power Query Editor with the 'Category' table loaded. The table now contains unique values in the 'Category Name' column, as indicated by the formula bar above the table: '= Table.Distinct(#"Removed Other Columns")'. The table has two columns: 'Category Id' and 'Category Name'. The 'Category Name' column lists various categories like 'Sporting Goods', 'Cleats', 'Shop By Sport', etc.

Category Id	Category Name
1	Sporting Goods
2	Cleats
3	Shop By Sport
4	Women's Apparel
5	Electronics
6	Boxing & MMA
7	Cardio Equipment
8	Trade-In
9	Electronics
10	Kids' Golf Clubs
11	Hunting & Shooting
12	Baseball & Softball
13	Men's Footwear
14	Camping & Hiking
15	Consumer Electronics
16	Cameras
17	Computers
18	Basketball
19	Soccer
20	Girls' Apparel
21	Accessories
22	Women's Clothing
23	Crafts
24	Men's Clothing

Figure 46: Screenshot of Power BI after removing the duplicates.

After removing duplicates in ‘Category’ table, now again create a relationship between the tables and it will look like below.

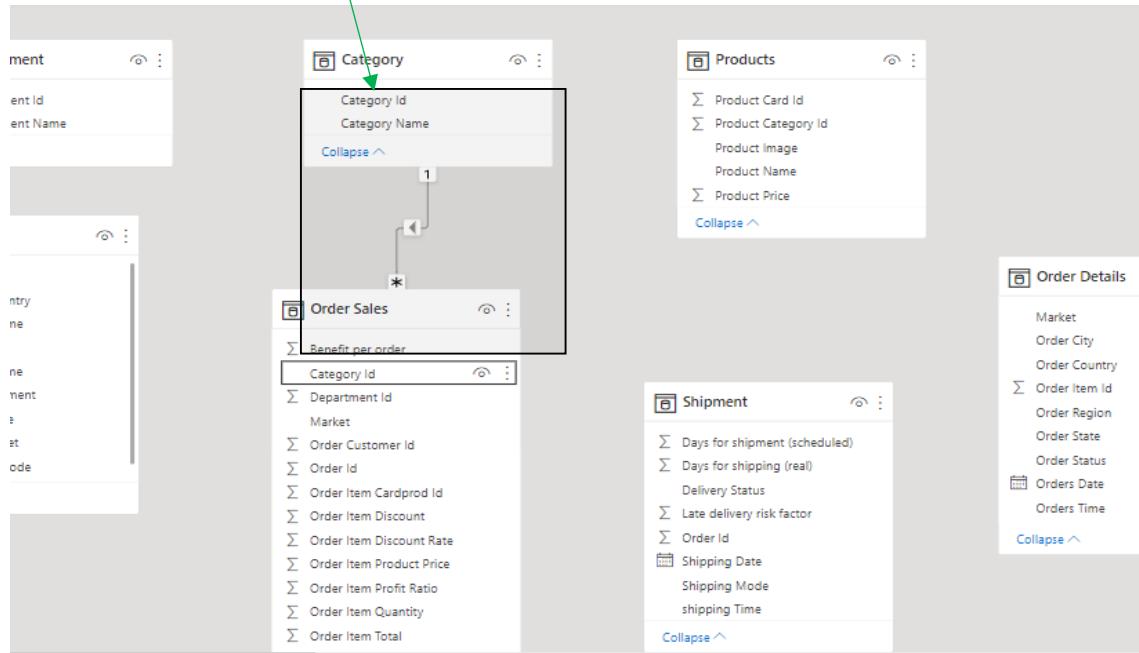


Figure 47: Screenshot of Power BI after creating a One to many (1:\*) relationship between Category and Order Sales.

Repeat the steps above and now create a relation between ‘Order Sales’ and ‘Products’ tables (or just use drag and drop). We can see Many to many relationships forming again between them. So now remove the duplicates in ‘products’ table. Now again create a relationship with ‘Order Item card Prod Id’ column in ‘Order Sales’ table and ‘Product Card Id’ column in ‘Products’ table.

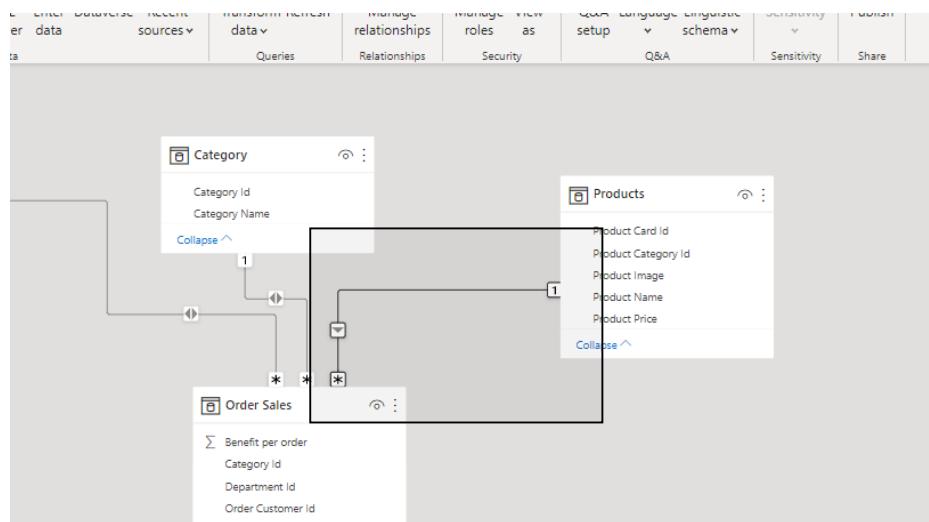


Figure 48: Screenshot of Power BI after creating a One to many (1:\*) relationship between Order Sales and Products.

Next is to change the cross-filter direction to ‘Both’ which means **Bi-directional** flow. Let’s take a moment see if cross-filter direction is set as ‘Single’. Create a Table chart with values ‘Product Name’ and ‘Order Item product price’ and it looks like below one.

**NOTE:** First thought of creating a snowflake schema model with this step but later decided to go with a **Star Schema**. The next two figures were added to understand the Cross-filter direction in Snowflake schema. But to showcase my understanding of directional flow of data the following two steps were kept as it was before.

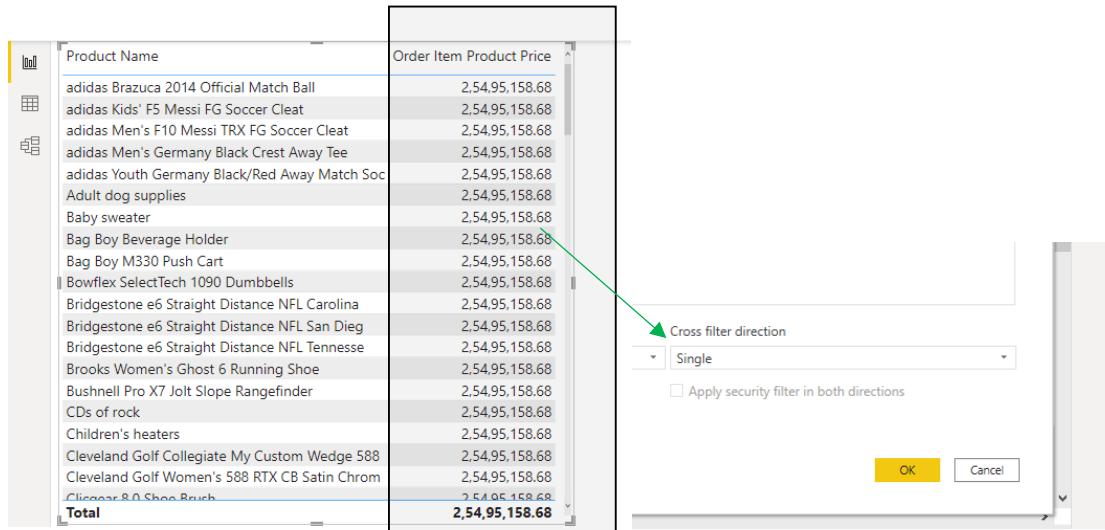


Figure 49: Screenshots of Power BI Screen after creating a Table chart with Product name in 'Products' and order Item Product price in 'Order Sales'.

In the above table chart the same Order item product price is shown to every Product Name as Cross filter direction is 'Single'. There is no data flow between 'products' and the 'Order Sales' table as data is flowing from 'Category' to 'Products' and 'Category' to 'Order Sales'.

Now change it to 'Both' then the data can flow from 'Products' to 'Order Sales' through 'Category'. The result can be seen below.

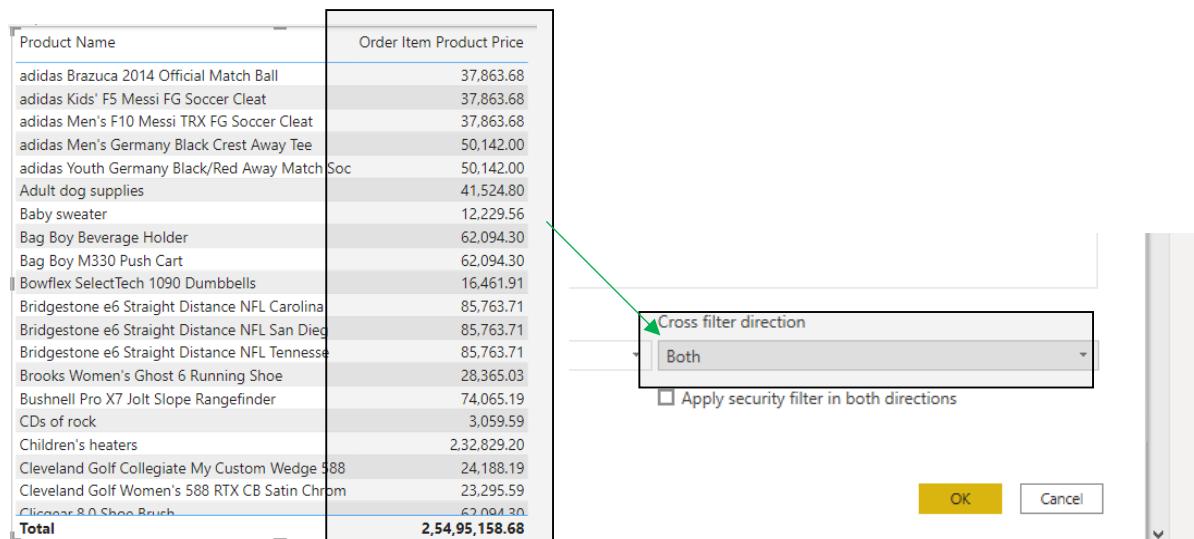


Figure 50: Screenshot of Power BI after changing the direction filter to 'Both'.

Repeat the same steps to create a relationship between ‘Department’ and ‘Order Sales’ tables and it will look like below.

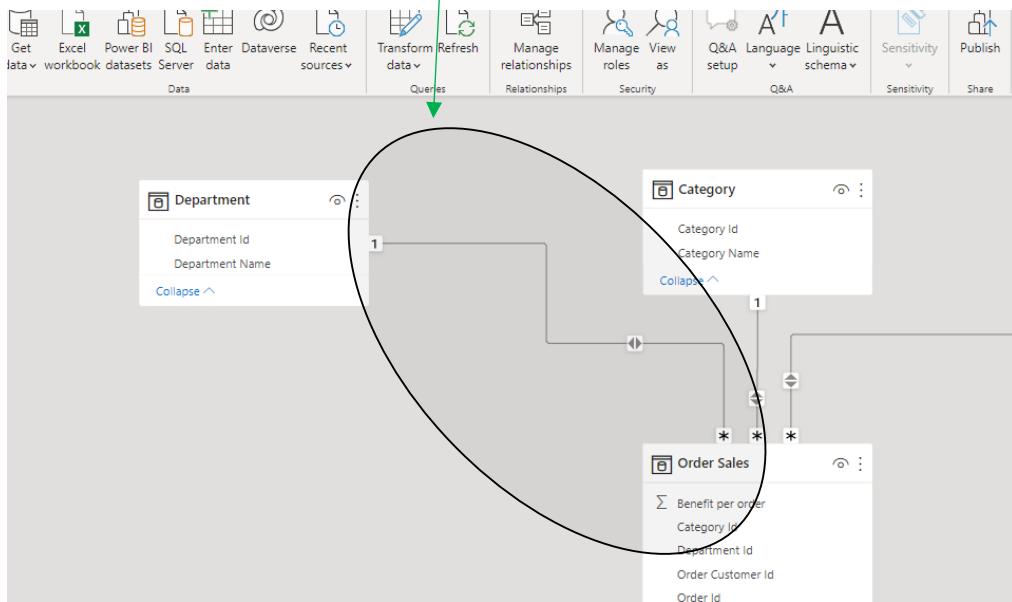


Figure 51: Screenshot of Power BI after creating relationship between ‘Department’ and ‘Order Sales’.

Now let’s create a relation between ‘Order Sales’ and ‘Customers’ tables. After creating a relation’ in both tables a many to many relationship is formed.

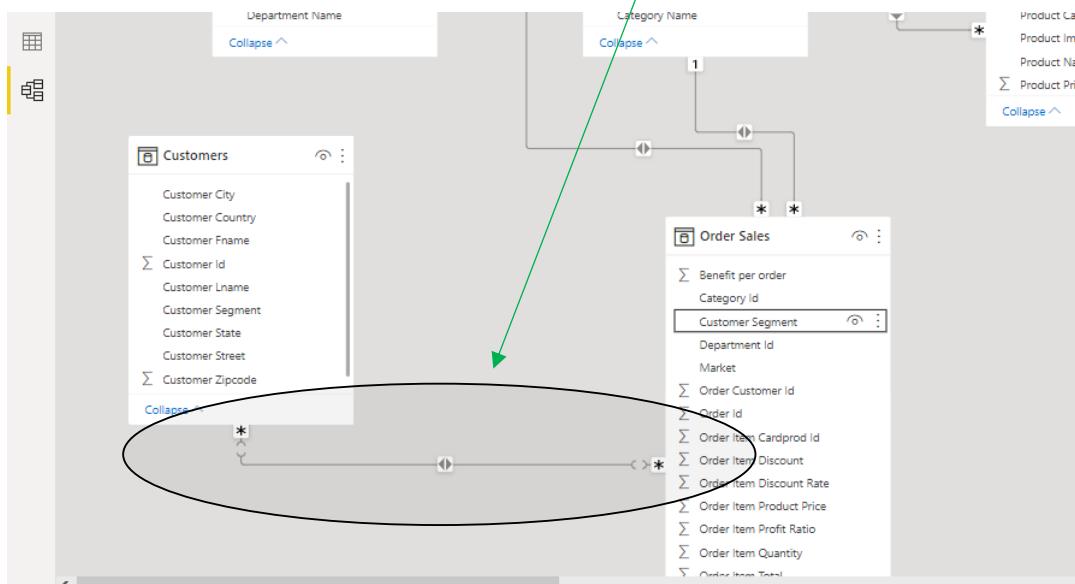


Figure 52: Screenshot of Power BI showing a Many to many relationship between ‘Customers’ and ‘Order Sales’.

In this scenario, to avoid many to many relationships here we need to remove the duplicate values in the ‘Customers’ tables which are causing the many to many relations in the first place. Removing duplicates steps can be seen in this report above.

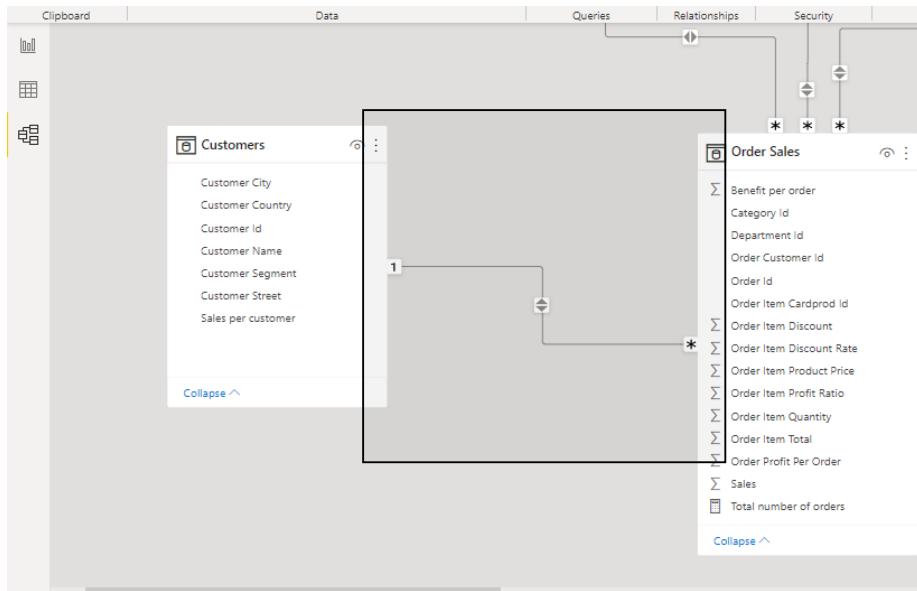


Figure 53: Screenshot of Power BI showing relationship between ‘Order Sales’ and Customers’.

Repeat the same steps and create relationship between ‘Order Sales’ and ‘Order Details’, ‘Order Sales’ to ‘Shipment’. After that Power BI will look like below.

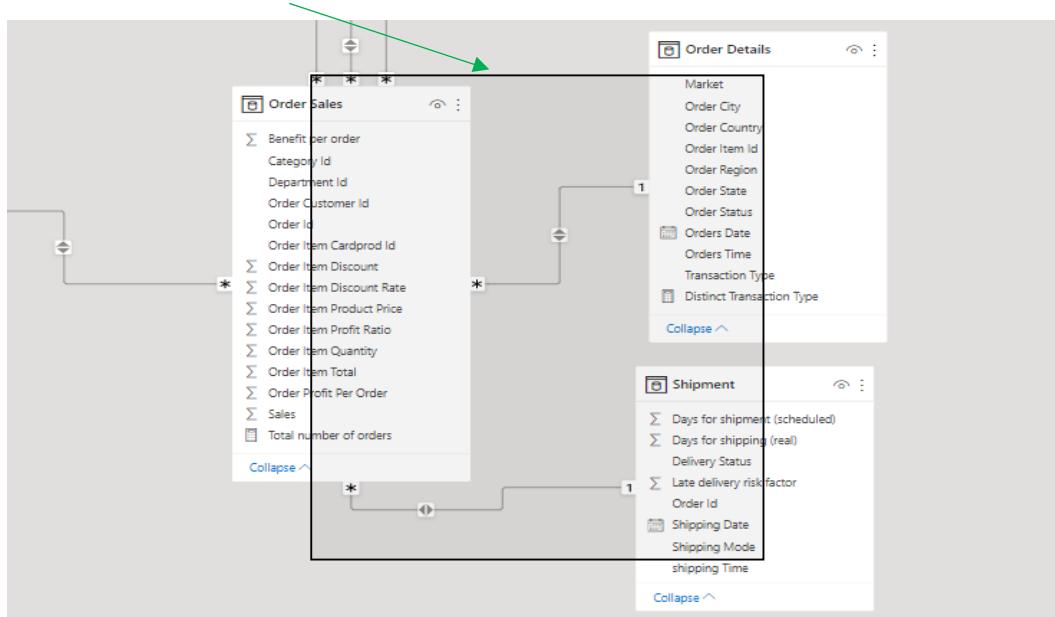


Figure 54: Screenshot of Power BI showing relationship between ‘Order Sales’ and ‘Order Details’, ‘Order Sales’ and ‘Shipment’.

Now, we are done creating relationships for all the tables confirming that the Data Model is a ‘**Star Schema**’ as each dimension table is directly related to the fact table.

Hence,

- ‘Products’, ‘Category’, ‘Department’, ‘Customers’, ‘Order details’, ‘shipment’ are the dimension tables.
- ‘Order Sales’ is a Fact table.

**NOTE:** For even better understanding of dataset some the columns are merged like Fname and Lname to Customer Name, some of the columns were removed in these tables.

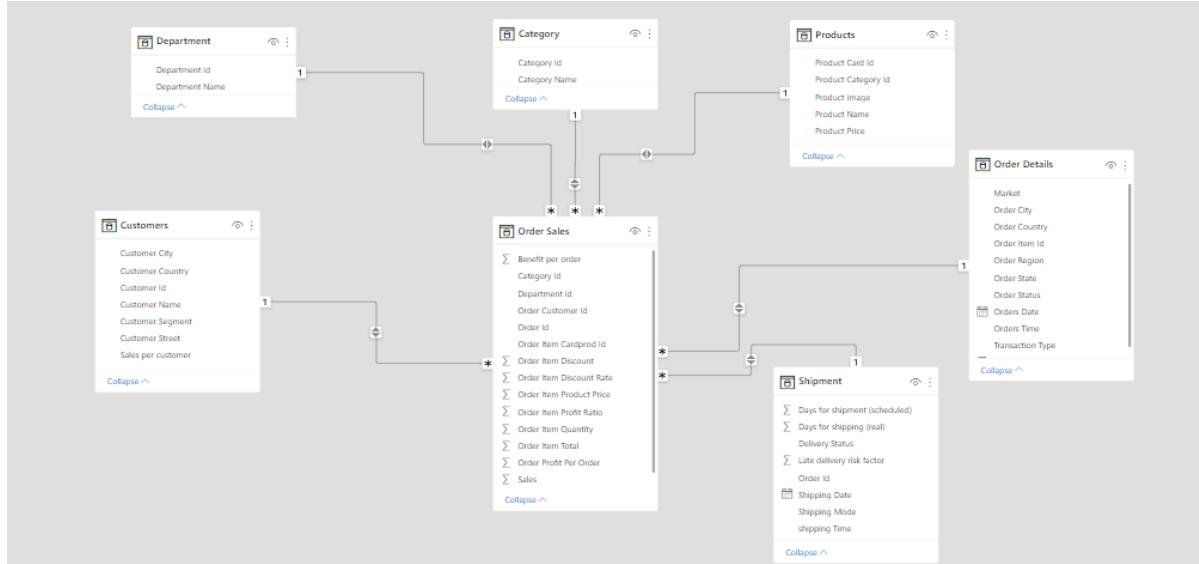


Figure 55: Screenshot of Power BI showing final MODEL view.

### Creating Measures and Calculated Columns:

Before creating measures and calculated columns need to change the number of decimal places for these columns Order Item Discount, Order Item product price, Order Item Profit Ratio, Sales, Order Item Total, Order Profit per Order by 2. Also change the Order Item Discount Rate into percentages.

Column	Format	Properties
Order Item Discount	Decimal number (2)	Summation: Sum, Data category: Uncategorized
Order Item Discount Rate	Percentage	Summation: Sum, Data category: Uncategorized

Figure 56: Screenshots of Power BI for formatting the decimal places adding percentages.

Now let's create calculated column where we can calculate the cost of making of the Order item. Theoretically speaking we have price of the product per Order and Profit made per Order. Subtracting these two in our Order Sales table can give us Cost of making each order on a row-to-row basis as a calculated column.

Click on 'New Column' in the 'Home' tab. Type the required DAX formulae as below:

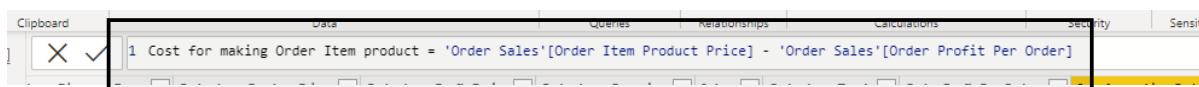


Figure 57: Screenshot of Power BI showing the DAX formulae for cost making of Order item.

Then click on OK. A new column will be created as shown below. Add \$ to the column in the formatting. Add \$ currency to all the amount columns in all the tables.

The screenshot shows the Power BI 'Data' view with a table containing several columns: 'Order Item Profit Ratio', 'Order Item Quantity', 'Sales', 'Order Item Total', 'Order Profit Per Order', and 'Cost for making Order Item product'. The 'Cost for making Order Item product' column contains numerical values such as 76.80, 49.99, 19.99, etc. To the right of the table is a 'Fields' pane where the new column is listed under the 'Order' category. Below the table, a note says 'making Order Item product (46,373 distinct values)'.

Figure 58: Screenshot of Power BI after creating a new column in 'Order Sales'.

Now, let's create a measure calculating total number of orders placed in this dataset. For that click on 'New measure' in 'Home' tab same as new column. Type the DAX formulae as below;

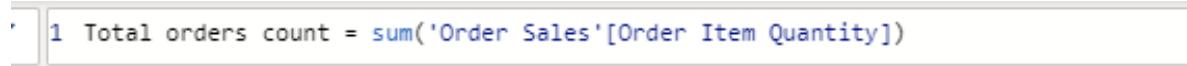


Figure 59: Screenshot of Power BI showing the formulae for Total orders count.

This measure adds the count for quantity of the products and gives us the answer as a single measure as shown below.

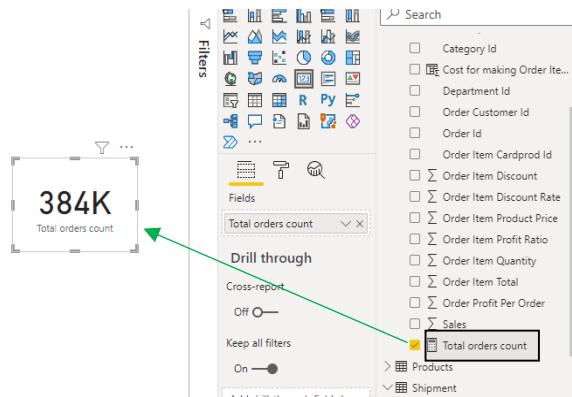


Figure 60: Screenshot of Power BI after creating total orders count measure.

More calculated columns will be created further in this report.

## SECTION 2

# BUSINESS INTELLIGENCE SOLUTION

Analysis of DATACO GLOBAL's Supply  
chain performance in the FY 2015

NAME: ROHITH VADDEPALLY

STUDENT ID: W9529921



## EXECUTIVE SUMMARY

The Business report tells us about the performance of **DATACO GLOBAL** supply chain network. The registered activities of the company are Sales, Production, Provisioning and commercial distribution. Types of products manufactured and sold in this process are Clothing, Sports and Electronic Supplies and many more. How the supply chain works?



The seller sells the quantity of item required for the manufacturing of the product to the producer, the producer manufactures it and sells the product to the customer and makes money out of it. This report analyses the sales performance in the financial year 2015 that is from January 2015 to December 2015. With this analysis it is also discussed what changes in the process need to be made and strategies are according to them. This ICA is done from the point of view of a data analyst.

**Key Findings:** These key findings are addressed in Power BI

- Total Sales made in the financial year 2015 are \$30.99M with the Total Orders placed by the customers are 344K Orders.
- Total Cost for making for all the Order item products in the financial year 2015 is \$17.38M.
- Total Order Profits in the financial year 2015 is \$3.33M.
- In FY 2015 total cost for making of all the order items products is \$17.38M and total Sales made are \$30.99M. So, the profits made as of now are \$13.61M and we have seen that Total Profits made by the producer are \$3.33M, \$10.28M goes to the Seller for the items that they have given to producer for manufacturing.
- There are a total 23K late deliveries in the FY 2015.
- Averagely \$161.52 sales were made to each individual customer in the FY 2015.
- Monday was the weekday in the FY 2015 that the more Sales were happening compared to other weekdays.
- A sales analysis is made with the help of a decomposition tree and found that most sales were happening in United States (EE. UU) compared to Puerto Rico (US Territory in North America). The decomposition tree in Power BI uses AI and Machine Learning techniques and finds out the high and low value with Customer City, Department Name, Category Name, Product Name.
- In FY 2016 from January to March, there is a tremendous downfall in the Orders, sales, profits with the company. Based on the data of FY 2015(January to December)-2016(January to March) forecasting of sales is made in the power BI for the next two years and results will be discussed further in this report.
- There were huge sales fluctuations in FY 2015 that sales going down from January to February and going up from November to December. Insights for these fluctuations are explained in this report.
- Count of Total orders with respective of their discount percentage which are 0-5%,5-10%,10-15%,15-20%,20-25%. Explanation for this is further given in this report.
- Sales of fourth quarter of the FY 2015 in the form a line chart are monitored with a use of a Timeline slicer chart.

## Recommendations:

- As there is 11.38% decrease in sales between January and February in the FY 2015 it is recommended to reduce but not stop completely stop the sales in the states like Parana, Baja California, Santiago de Chile as these states caused decrease in sales. Countries Mexico, Brazil, Colombia accounted for the majority of the decrease in sales. Regions like Central America and South America caused a decrease in Sales. It is recommended to increase the production and sales in Caribbean region but no to completely dissolve the business in the above areas.



Figure 1: Screenshots of waterfall chart in Power BI explaining the decrease in sales between January and February.

- There is also an increase in sales by 5.47% between November and December in the FY15. Reason for this rise in sales in 'Kids Golf Club', Men's Golf Clubs and Fishing with respective to the Category Name. There is also a decrease in sales for the Category 'Electronics'. So, it is recommended to increase the production and sales with these categories in the regions South Asia, Oceania, South East Asia in November and December.



Figure 2: Screenshots of waterfall chart and Scatter chart in Power BI explaining the increase in sales between November and December.

'Fishing' Category had highest number of Order Profits in the FY 2015 of \$677.17K by making the Sales of \$6.28M in total. It is recommended to make more production in Fishing category as we can except more profits from it.



Figure 3: Screenshot of Bar chart with cash infographic in it showing the Orders Profits by category Name.

## INTRODUCTION

### BI Questions:

Questions addressed in this report are:

- To provide the number of Total Sales, Cost for making order item product, Total Orders Profits and also calculate these with regarding to 'Fitness' Department Name in FY 2015?
- Give the minimum value of the Order item Product Price in the FY 2015. Find out the number of Order items of completed and pending order status. Give the orders count by different shipping mode. Evaluate the sales by each country in FY 2015?
- Evaluate the Sales KPI and find out the achievement with Order item total and cost of making the product in FY 2015?
- Provide the average number of deliveries happened in FY 2015. Find out what type of delivery speed service is chosen by the customers mostly in FY 2015?
- On which weekday the most sales are happening in FY 2015. To provide the highest and lowest profited category name in FY 2015?
- Analyse the sales and explain it by Customer country, Customer city, Department Name, Category Name, Product Name?
- Forecast the sales for the next 2 years and find out what could be the sales value on 1<sup>st</sup> February 2018?
- Find out why the decrease in sales happened in Jan-Feb period and increase in sales in Nov-Dec period of FY 2015?
- Provide the count of the orders with different range of discount rate?
- Create a Time Slicer chart and use it monitoring the sales for the Quarter 4 of FY 2015.

### Description of the data used in the model:

There are a total of 7 tables in this data model in which one is fact table and other six are dimension tables. Order Sales is the fact table and the Department, Category, Products, Customers, Order Details, Shipment. These tables are exactly the one which are mentioned in Data Modelling in Section-1. Power BI visuals, charts, graphs and infographics are made from this Data Model. This Data Model is used to analyse the orders, sales, profits, shipping details of the supply chain in FY 2015-2016.

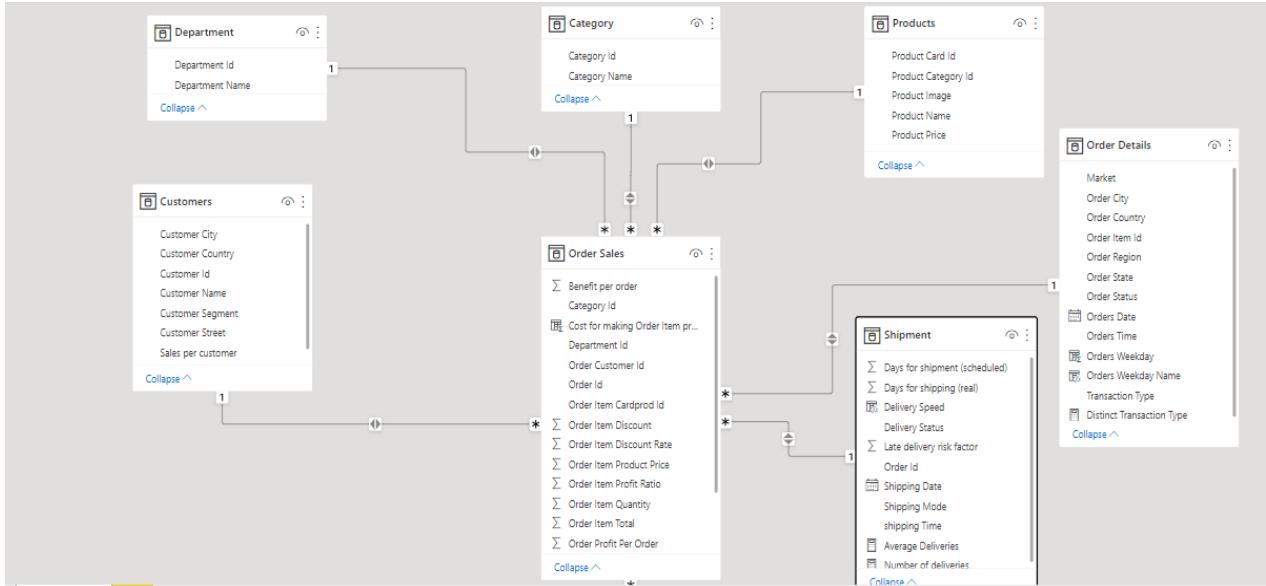


Figure 4: Screenshot of Data Model in Power BI.

## FINDINGS BASED ON ANALYSIS AND EVALUATION

These Power BI visuals and dashboards are created to answer the Business questions.

For providing a number of Total Sales, Total Order Profits we are going to create a card visual on to the power BI and in add Sales in the fields section. When we right click on the fields make sure it is as SUM.

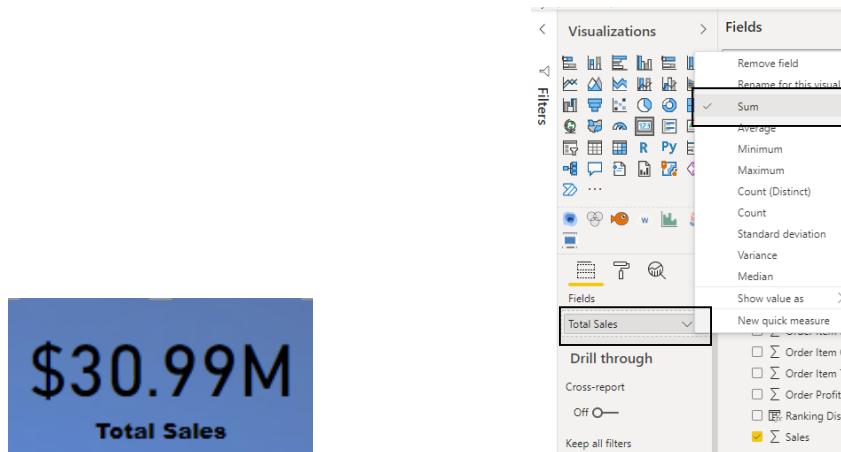


Figure 5: Screenshot of card visual of total number of sales in FY 2015 on it.

Repeat the same steps for ‘Total Orders profits’ in FY 2015 and visual will look like below.



Figure 6: Screenshot of card visual of Total Orders Profits in FY 2015 on it.

For finding out the Cost for making the Order item product in FY 2015, a new calculated column needs to be created in the ‘Order Sales’ table and a DAX formulae needs to be used which is already mentioned in Data Modelling in Section-1.

`1 Cost for making Order Item product = 'Order Sales'[Order Item Product Price] - 'Order Sales'[Order Profit Per Order]`

Subtracting the profit of each Order from each order item price will get us the Cost for making Order item Product. This will create a new calculated in ‘Order Sales’ table

Now repeat the same steps of using the card visual in Power BI and drag the new column in to the Fields section. The new card visual will look like below.



Figure 7: Screenshot of card visual of Cost for making Order Item Product.

Now showing the Sales, Order Profits and cost of making the products in FY 2015 only in the fitness department. For that, added a slicer visual onto the Power BI where in the slicer the Department Name is selected as ‘Fitness’.

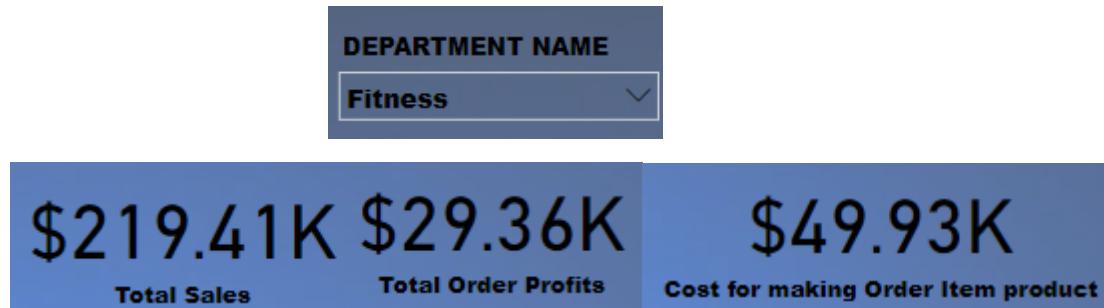


Figure 8: Screenshot of card visuals of Total Sales, Total Order Profits, Cost for making Order Item Product when the department name is ‘Fitness’.

In the ‘Fitness’ department the total Sales were \$219.41K, the total order profits were \$29.36K and cost for making order items were \$49.93K.

For finding the minimum value of the Order item product price we are going to use Infographic Designer visual in which the measure is selected as 'Minimum'. Then click on edit button on the infographic and then select coins in the 'Mark Designer' tab.

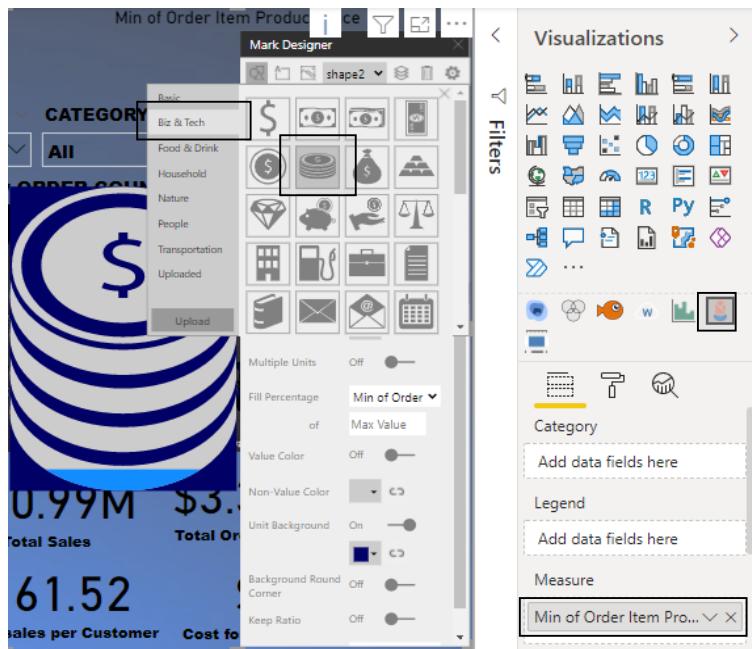


Figure 9: Screenshot while creating an infographic designer with a minimum of Order item product price.

After that the visual will look like below with a minimum value of the Order Item Product price which is \$9.99 in FY 2015.



Figure 10: Screenshot of the infographic designer with minimum value of the Order item product price.

To find out the number of Order item of the ‘Completed’ and ‘Pending’ order status in the FY 2015 we are going to create a clustered column chart in which ‘Order Status’ will be in the axis and ‘Order Item Quantity’ will be in the Values. 111313 is the order item quantity of the Order status ‘COMPLETE’ and 38385 is the order item quantity of the Order status ‘PENDING’.



Figure 11: Screenshot of clustered column chart of Order item quantity by Order status.

To find out the total orders count by different shipping mode in the FY 2015 we are going to create a ‘Donut chart’. For this we should create a measure that sums up all the Order item quantity. Click on ‘New Measure’ in the Home tab and enter the DAX formulae as shown below.

```
1 Total orders count = sum(['Order Sales'[Order Item Quantity]])
```

NOTE: Creating a measure does not create a new column in that particular table only a calculated column will.

Give the measure name as ‘Total orders count’ and it has a one and only unique value which add all the order item quantity. Keep the ‘Total orders count’ from the ‘Orders Sales’ table in the values and ‘Shipping Mode’ from the ‘Shipment’ table in the ‘Legend’. Now we get a donut chart of total orders count by different shipping mode in FY 2015.



Figure 12: Screenshot of Donut chart of total orders count by different shipping mode in FY 2015.

For evaluating the Sales by each country in FY 2015 we need to create a Map chart. In Map chart the ‘Order country’ from ‘Order details’ will be placed in Location and ‘Sales’ from ‘Order Sales’ will be placed in size. Then the Power BI gives us the most sales happening countries in bigger circles and less sales happening countries in smaller circles. Then the visual will look like below.



Figure 13: Screenshot of MAP chart of SALES by ORDER COUNTRY in FY 2015.

For evaluating the Sales KPI in this supply chain with respective to Order Item Total and Cost for making order item product in FY 2015, we need to create two different charts for this. One is ‘KPI chart’ and other one is ‘Gauge chart’.

### **KPI (Key Performance Indicator) → Target**

To evaluate the sales KPI with Order Item Total we need to create a KPI chart. Place ‘Sales’ in indicator, ‘Order item Total’ in target goals and ‘Order Date’ in ‘Trend Axis’. While manufacturing the total order item is \$27.85M and later the total sales happened \$30.99M which means target has been reached as there are more sales happened than the goal. Hence the goal has reached with the percentage of +11.29% in FY 2015. Colour ‘Green’ shows us that target has been reached. This information can be seen in the below KPI chart visual.



Figure 14: Screenshot of KPI chart of Sales and Order Item Total by year.

To evaluate the sales KPI with Cost for making Order Item Product we need to create a Gauge chart. Place ‘Sales’ in Value and ‘cost for making order item product’ in Target value. This Gauge chart clearly explains that profit has been made by having more sales than the cost for making order item

product in the FY 2015. The \$17.38M reader line on the gauge is total cost for making order item and 'Blue' region tells us the \$30.99M sales. This can be seen in the below Gauge chart.



Figure 15: Screenshot of Gauge chart of Sales and Total cost for making order item product.

To provide the number of the average deliveries happened in 2015 we should create a measure that counts total number of average deliveries happened in 2015. Formulae is shown below:

```
1 Average Deliveries = AVERAGE(Shipment[Order Id])
```

Then a measure will be created. Then again add a chart called Infographic designer and then place Average Deliveries measure in the measure section.

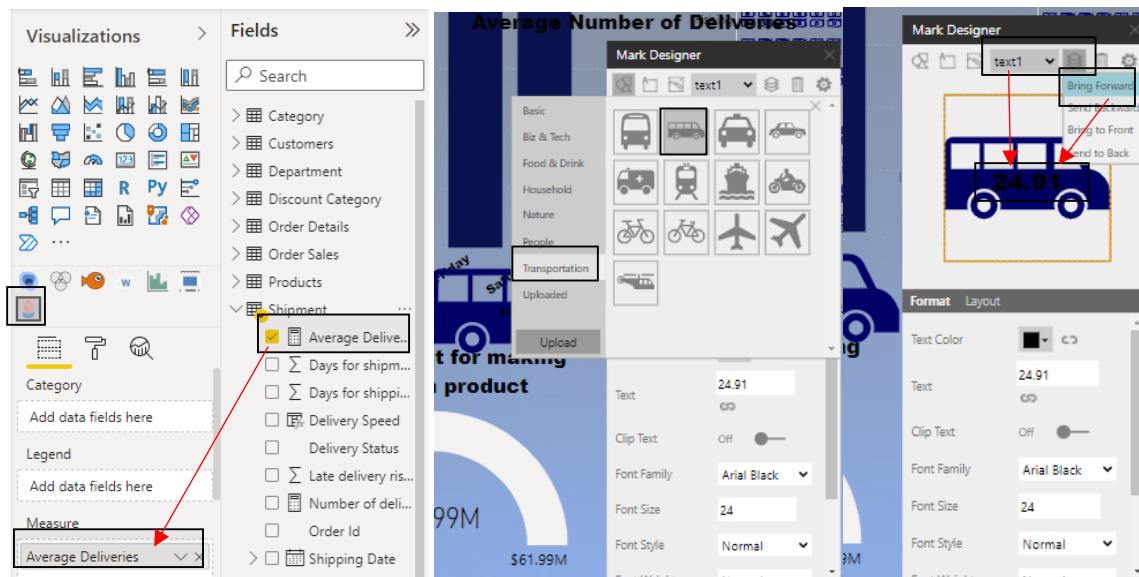


Figure 16: Screenshots of power BI explaining the steps to add the chart 'Infographic designer'.

After adding measure to this chart, click on the edit mark option on the card and in mark designer add the truck visual and text '2 4.91' which is the average deliveries happened in FY 2015. In Mark Designer bring the text1 to forward. Then this infographic chart with a truck symbol with the text on it will look like below.



Figure 17: Screenshot of ‘Infographic Designer’ with a truck infographic and average number of deliveries in FY 2015.

To find out what type of delivery speed service is chosen by the customers we need to create a calculated column with the name of Delivery speed service. The new calculated DAX formulae is given below:

```
X ✓ 1 Delivery Speed Service = if([Shipment[Days for shipping (real)] <=2, "Super fast delivery",
2 if([Shipment[Days for shipping (real)] <=5,"Medium fast delivery", "Slow delivery")]
```

This DAX formulae is calculated with the ‘Days for shipping (real)’ column. If the delivery happens within the 2 days of order date, then it’s called ‘Superfast delivery’. If it is greater than 2 days and less than or equal to 5 days then it’s called ‘Medium fast delivery’. If not both then it is considered as ‘Slow delivery’. With this a new calculated column will be created as shown below:

Delivery Speed Service
Super fast delivery

Figure 18: Screenshot of Power BI showing the new calculated column called ‘Delivery Speed Service’.

Now let’s create a Clustered column chart with ‘Delivery Speed Service’ on the X-Axis and count of deliveries on the Y-Axis measuring each delivery service in 3 bars Superfast delivery, Medium fast delivery, Slow delivery. Place Delivery speed service column in the Axis and place count of days for shipping (real) column in the value, then a Clustered column chart is created like shown below.

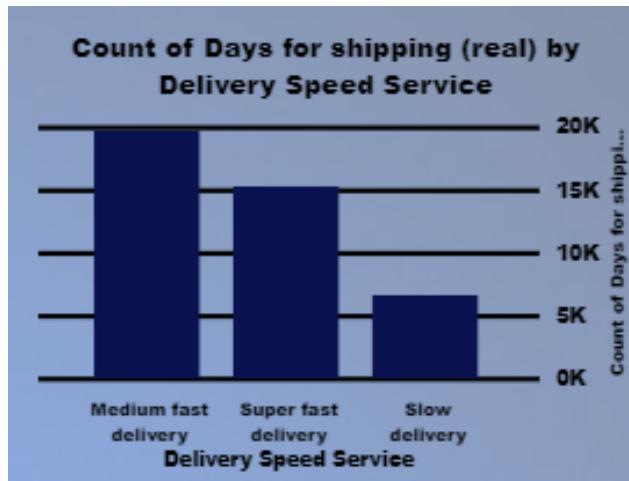


Figure 19: Screenshot of clustered column chart showing Delivery Speed service by count of deliveries in FY 2015.

To find out on which weekday the most sales happened in FY 2015, we need to create a calculated column using DAX formulae which pulls the weekday name from the Orders Date. The formulae for creating new column shown below.

```
  1 Orders Weekday Name = FORMAT('Order Details'[Orders Date], "dddd")
```

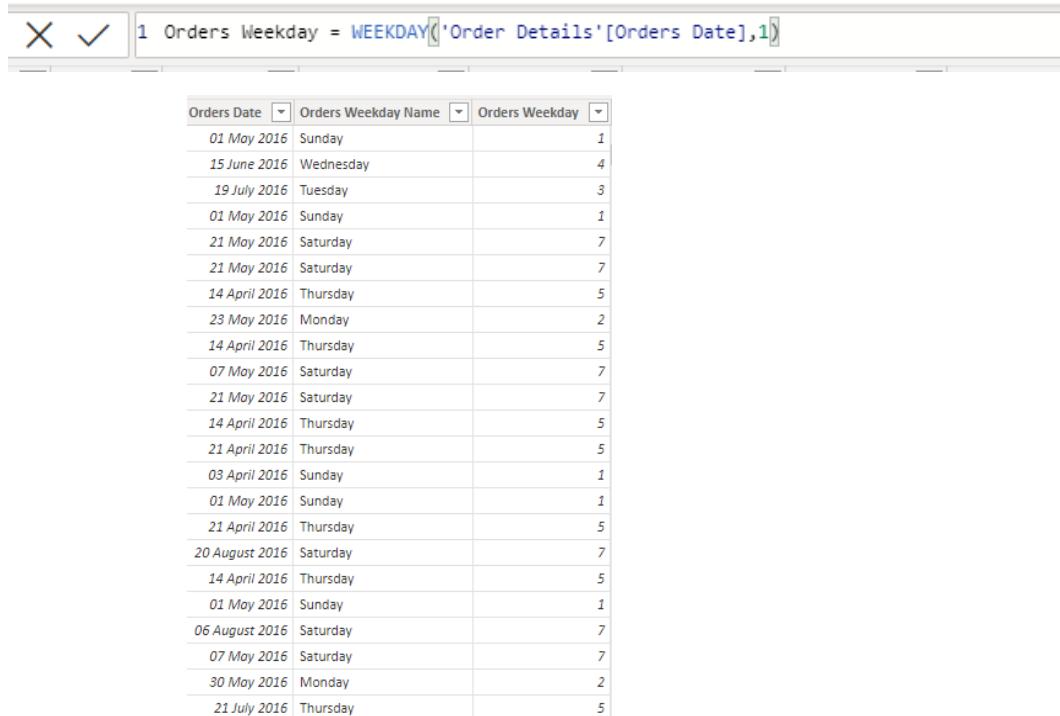
The above formulae pull up the weekday from the Orders Date and creates a new column as shown below.

	Orders Date	Orders Weekday Name
1	01 May 2016	Sunday
15 June 2016	Wednesday	
19 July 2016	Tuesday	
01 May 2016	Sunday	
21 May 2016	Saturday	
21 May 2016	Saturday	
14 April 2016	Thursday	
23 May 2016	Monday	
14 April 2016	Thursday	
07 May 2016	Saturday	
21 May 2016	Saturday	
14 April 2016	Thursday	
21 April 2016	Thursday	
03 April 2016	Sunday	
01 May 2016	Sunday	
21 April 2016	Thursday	
20 August 2016	Saturday	
14 April 2016	Thursday	
01 May 2016	Sunday	
06 August 2016	Saturday	
07 May 2016	Saturday	
30 May 2016	Monday	
21 July 2016	Thursday	

Figure 20: Screenshot of Power BI showing the new calculated column called 'Orders weekday Name'.

But, when we create a Clustered column chart with Sales by Orders Weekday Name the bars in the chart are created with no particular order as Friday, Monday, Tuesday, Saturday, Sunday, Thursday,

Wednesday. To sort this order from Sunday to Saturday we need to create another calculated column with DAX formula as shown below.



The screenshot shows a Power BI interface with a table view. At the top, there is a formula bar with the text: 1 Orders Weekday = WEEKDAY('Order Details'[Orders Date],1). Below the formula bar is a table with three columns: Orders Date, Orders Weekday Name, and Orders Weekday. The table contains approximately 30 rows of data, showing various dates and their corresponding weekday names and weekday numbers (1 for Sunday, 2 for Monday, etc.).

Orders Date	Orders Weekday Name	Orders Weekday
01 May 2016	Sunday	1
15 June 2016	Wednesday	4
19 July 2016	Tuesday	3
01 May 2016	Sunday	1
21 May 2016	Saturday	7
21 May 2016	Saturday	7
14 April 2016	Thursday	5
23 May 2016	Monday	2
14 April 2016	Thursday	5
07 May 2016	Saturday	7
21 May 2016	Saturday	7
14 April 2016	Thursday	5
21 April 2016	Thursday	5
03 April 2016	Sunday	1
01 May 2016	Sunday	1
21 April 2016	Thursday	5
20 August 2016	Saturday	7
14 April 2016	Thursday	5
01 May 2016	Sunday	1
06 August 2016	Saturday	7
07 May 2016	Saturday	7
30 May 2016	Monday	2
21 July 2016	Thursday	5

Figure 21: Screenshot of Power BI showing the new calculated column 'orders Weekday'.

The new column name will be Orders Weekday and by using this formulae order 1 will assigned to Sunday, 2 to Monday, 3 to Tuesday, 4 to Wednesday, 5 to Thursday, 6 to Friday, 7 to Saturday. Now again try to create clustered column chart and the chart will like below.



Figure 22: Screenshot of power BI of a clustered column chart showing sales by orders weekday Name.

From the above chart it is found that 'Thursday' is the weekday that most sales happen with the sales amount of \$45,08,594.13 when compared to other week days in the FY 2015.

To provide the highest and lowest profited category name in the FY 2015 we need to create Infographic Designer chart with ‘Order profits’ on the X-Axis and ‘Category Name’ on the Y-Axis.

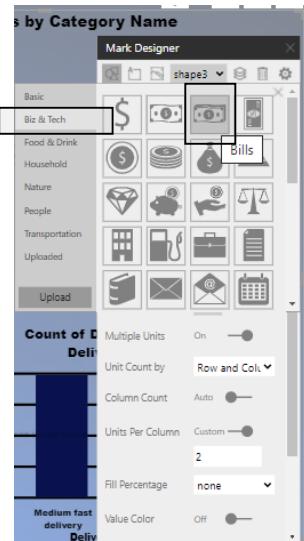


Figure 23: Screenshots of power BI explaining the steps to add the chart ‘Infographic designer’.

Now click on ‘Edit Mark’ button and it will be directed to Mark Designer and add Bills infographic from Biz and Tech to each ‘Category Name’. After that the chart will look like below



Figure 24: Screenshot of ‘Infographic Designer’ with Order Profits by Category Name in FY 2015.

From the above chart we can find that ‘Fishing’ is the most profited category and ‘Electronics’ is the least profited category in the FY 2015.

Now to analyse and explaining the sales it by Customer Country, Customer City, Department Name, Category Name, Product Name, we need to create a ‘Decomposition tree’. After selecting the ‘Decomposition Tree’ place the ‘Sales’ from ‘Order Sales’ in ‘Analyse’ and place ‘Product Name’, ‘Category Name’, ‘Department Name’, ‘Customer City’, ‘Customer Country’ from different tables. It is shown below:

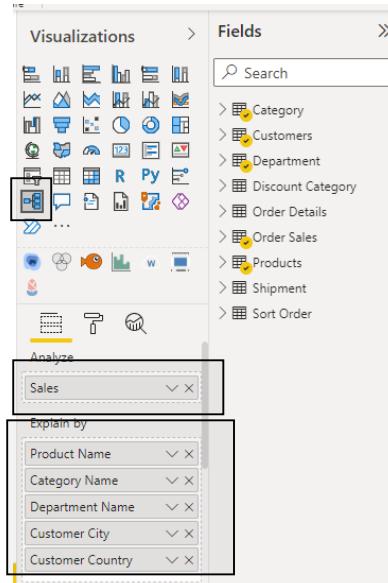


Figure 25: Screenshot of Power BI showing the ‘Visualizations’ and ‘Fields’ tabs for the decomposition tree.

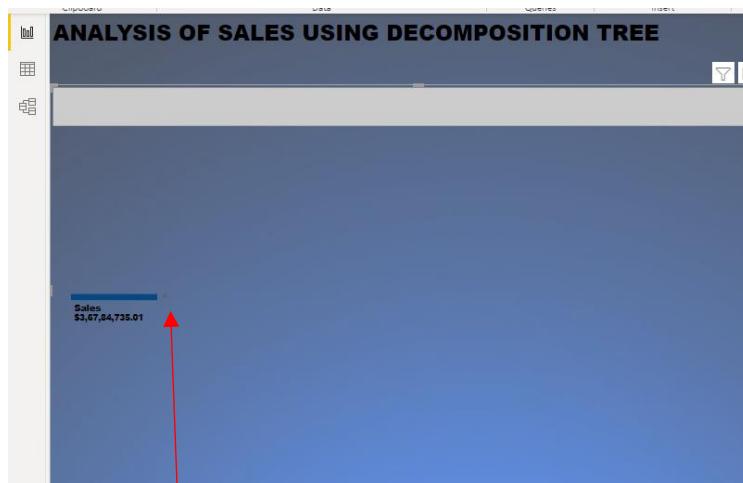
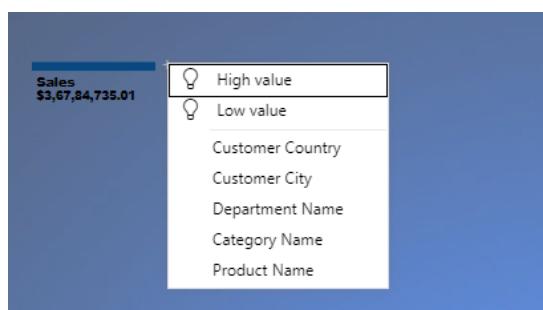


Figure 26: Screenshot of Power BI showing Decomposition before analysing the sales.

Now, click on + symbol then a tab will pop up as shown below.



Now click on the 'High value', then Power BI will find out the highest sale value in all five categories and shows us and in this case, it is 'Customer Country' which is EE.UU. The screenshot of this is given below.

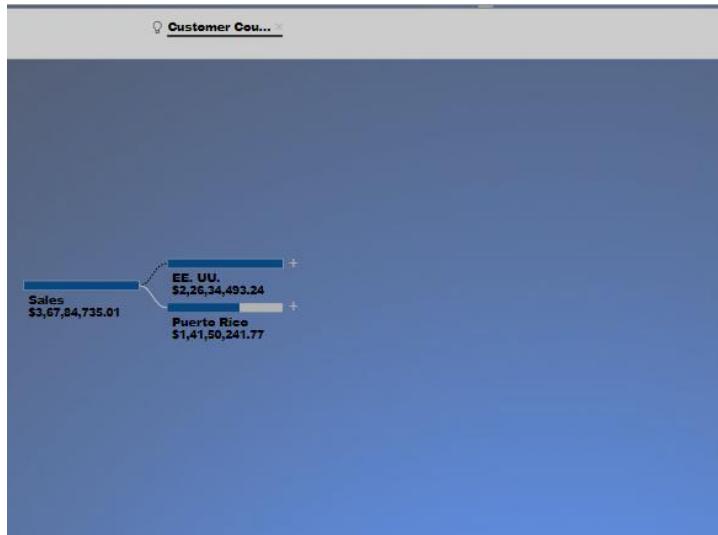


Figure 27: Screenshot of decomposition tree after analysing sales with Customer Country.

Repeat the same steps with 'High value' or 'Low value' from the 'Customer Country' (or) the user has also the freedom of selecting the Category by which the sales need to be explained. After repeating these steps, the tree will look like below.

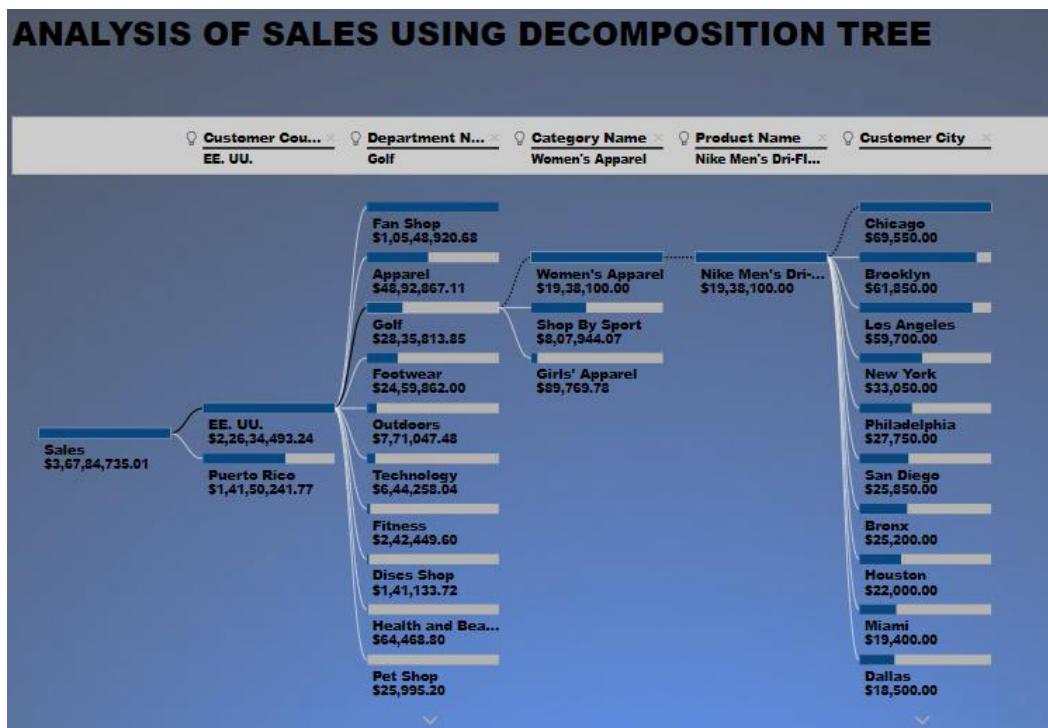


Figure 28: Screenshot of Decomposition tree after analysing the sales with different categories.

There is slicer added to the page which has the FY 2015 and 2016. As the Power BI is interactive selecting the year 2015 the sales values in the tree will also be changed according to the FY 2015.



Figure 29: screenshot of Slicer of the years 2015 and 2016.

To forecast the Sales for the next two years, we need to create a Line chart with the Sales and Orders Date. 'Orders Date' from 'Order Details' will be placed in the Axis on X-Axis and 'Sales' from 'Order Sales' goes to Values on Y-Axis. The line chart will be showing sales information from January 2015 to March 2016. Now select the Analytics tab and then click on forecast line.

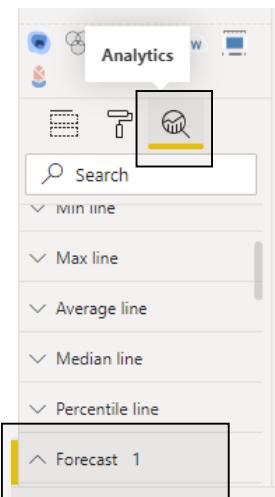


Figure 30: Screenshot of Power BI showing the Analytics tab and Forecast line.

Add a forecast line and rename it as 'Forecasting next 2 years' and give the Forecast length as '2 Years'. The forecasting will happen from 27<sup>th</sup> March 2016 to 28<sup>th</sup> March 2018.

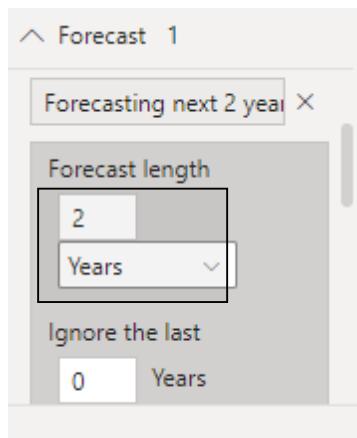


Figure 31: Screenshot of Power BI after selecting forecast length as two.

The forecasting line chart will like below.

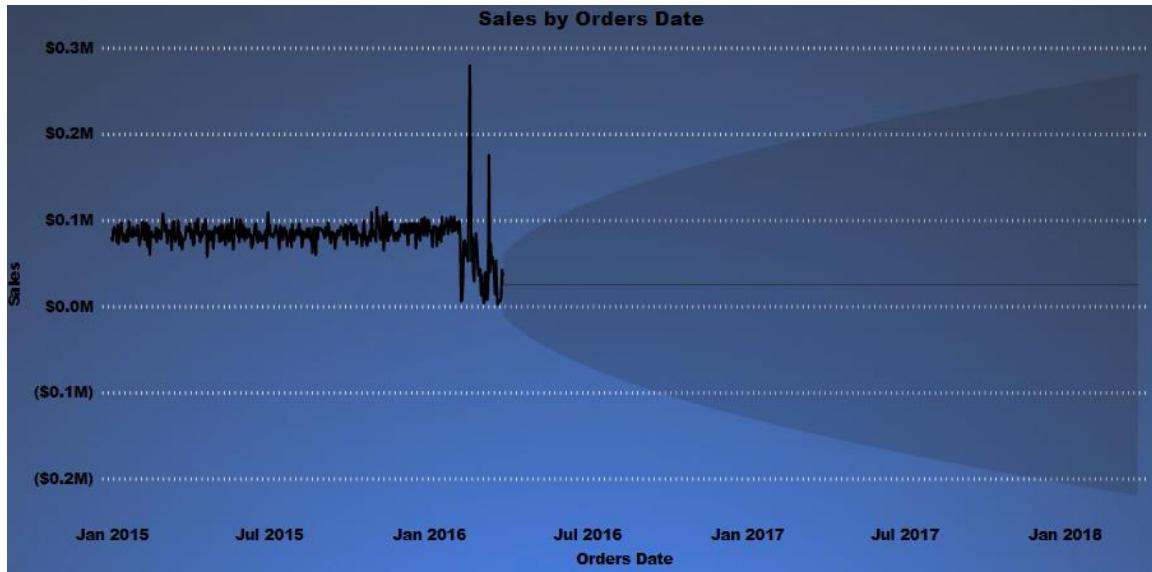


Figure 32: Screenshot of Line chart forecasting the sales for the next two years.

The sale value on the 1<sup>st</sup> February 2018 is \$25,088.42. A slicer of years 2015 and 2016 is added to this page and with this slicer we can know sales in their respective year.

To find out the reasons for sales fluctuations in the period of Jan-Feb and Nov-Dec in the FY 2015, we need to create a Line chart of Sales by Month. The screenshot of the chart is shown below.

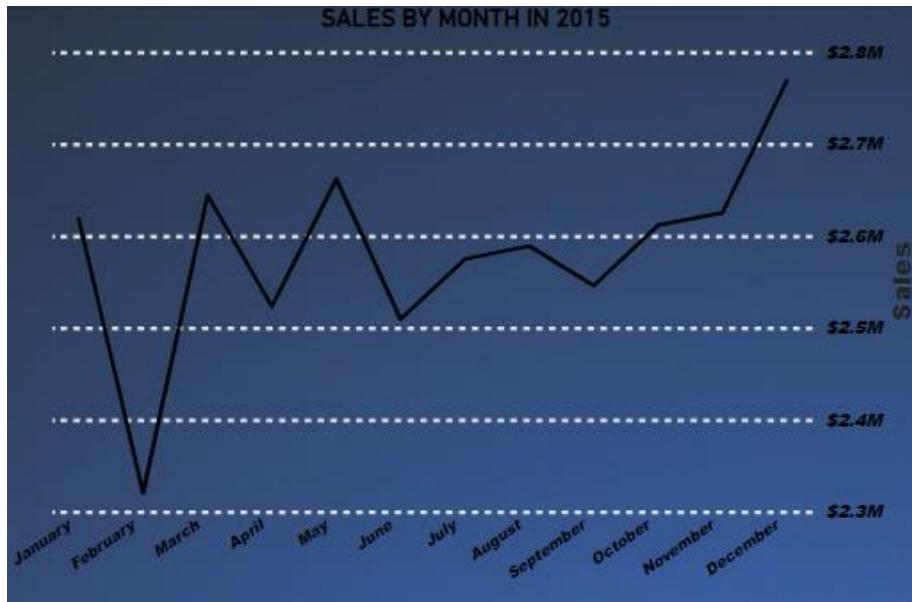


Figure 33: Screenshot of Line chart of sales by month in FY 2015.

As per the line graph there are decrease in sales from the month January and February. Give a right click on the slope and then click on 'Analyse' and then 'Explain the decrease'.

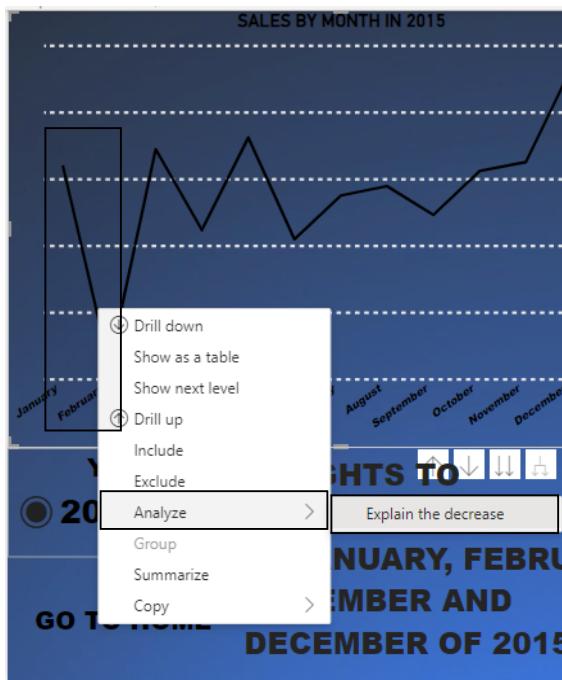


Figure 34: Screenshot of Power BI explaining the decrease in Sales.

This explains the reasons why there is a decrease in sales in the beginning of the FY 2015. This is already explained in the recommendations section above in this report. Screenshots of the reasons are given below.

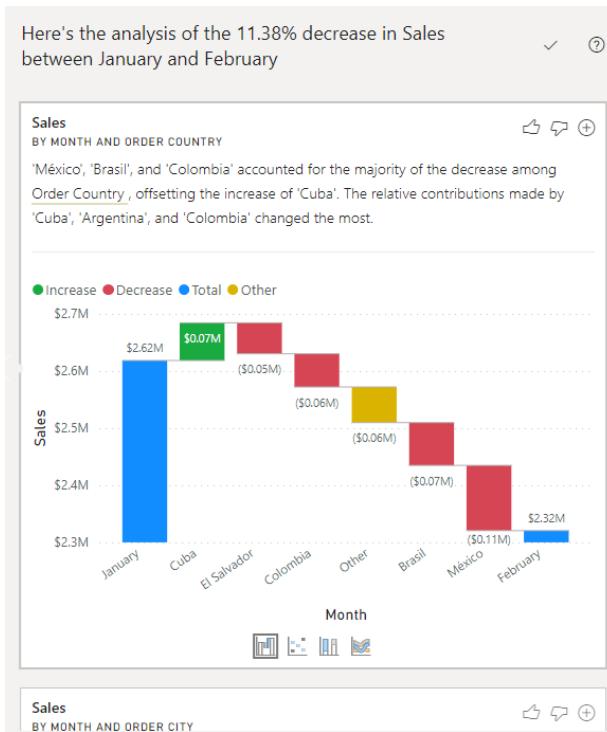


Figure 35: Screenshot of reasons of why there is a decrease in sales (Jan-Feb of FY 2015).

Repeat the same steps for finding out why there is increase in sales from the month November to December in FY 2015.

Finally, the reasons of fluctuations in sales in the beginning of the year and at the end year with a waterfall chart and a scatter chart will look like below.

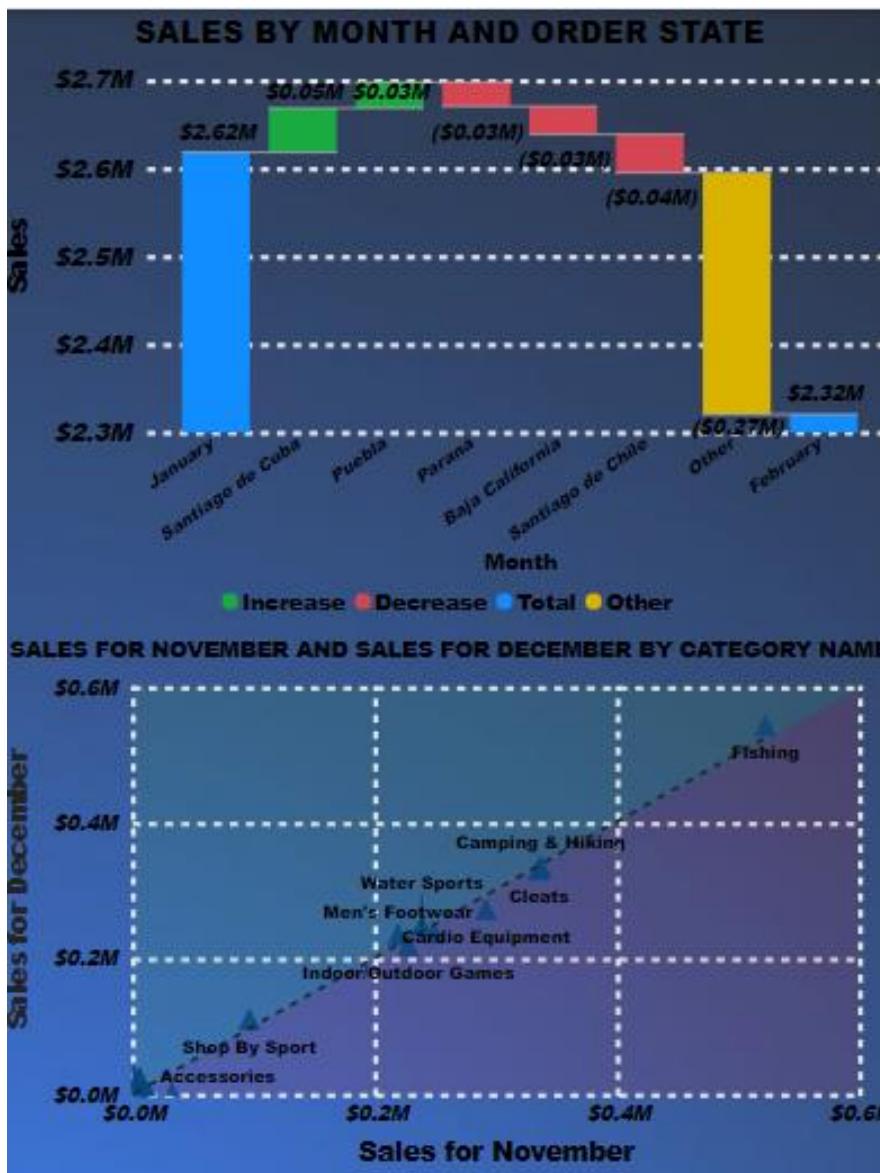


Figure 36: Screenshot of charts showing the fluctuations in sales.

To provide the count of the orders with different range of discount rate we need to divide the Order item discount rate in five different range of categories as shown below.

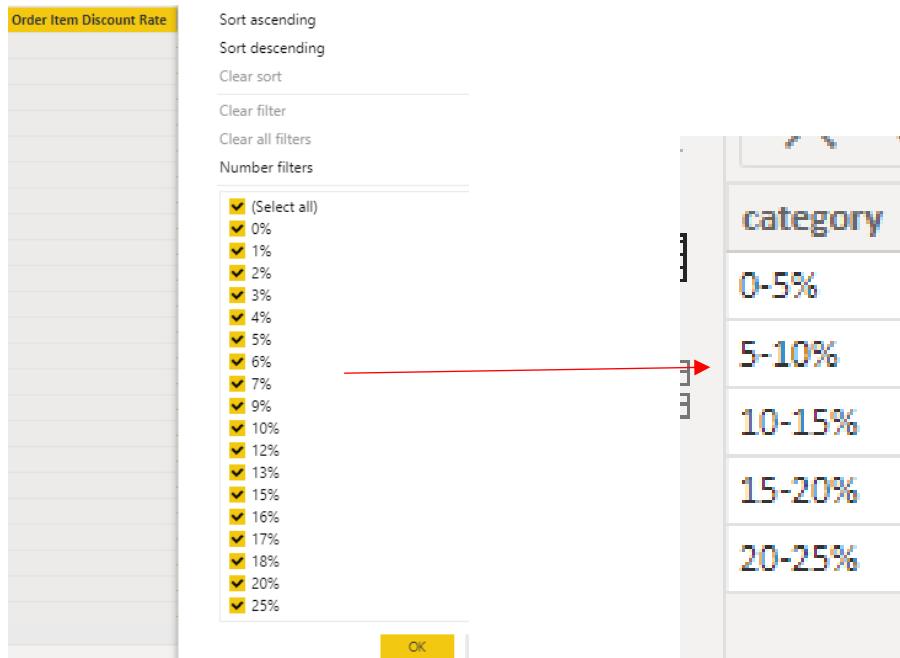


Figure 37: Screenshots of Power BI explaining how to divide the discount rate into category.

First click on 'Blank Query' from the Home tab and then click on 'Advanced Editor' in Power Query Editor as shown below.

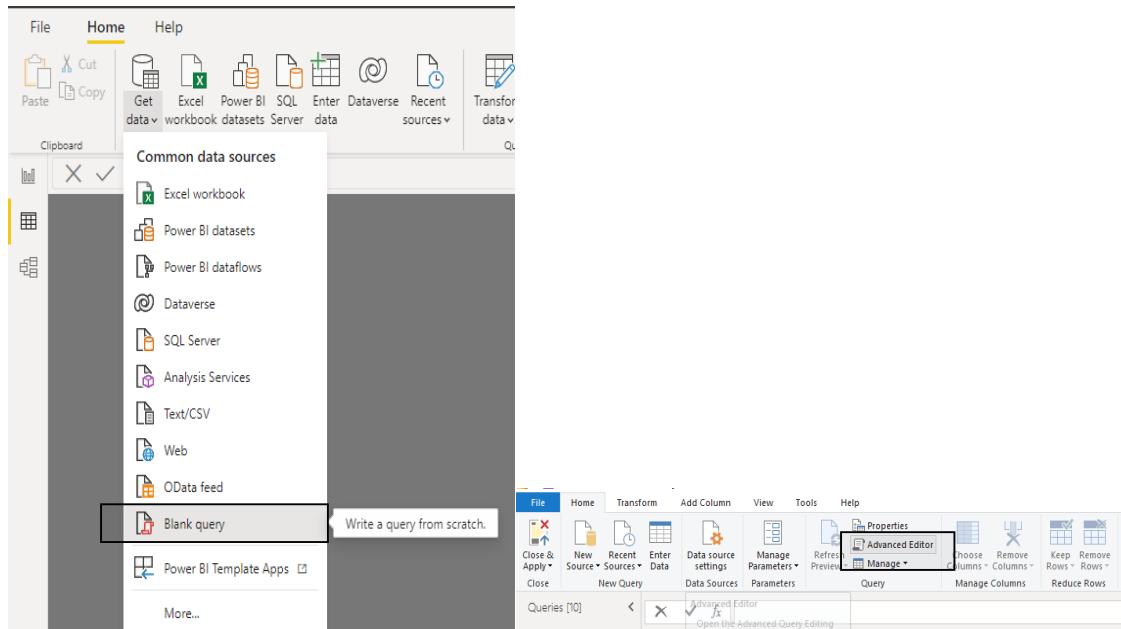
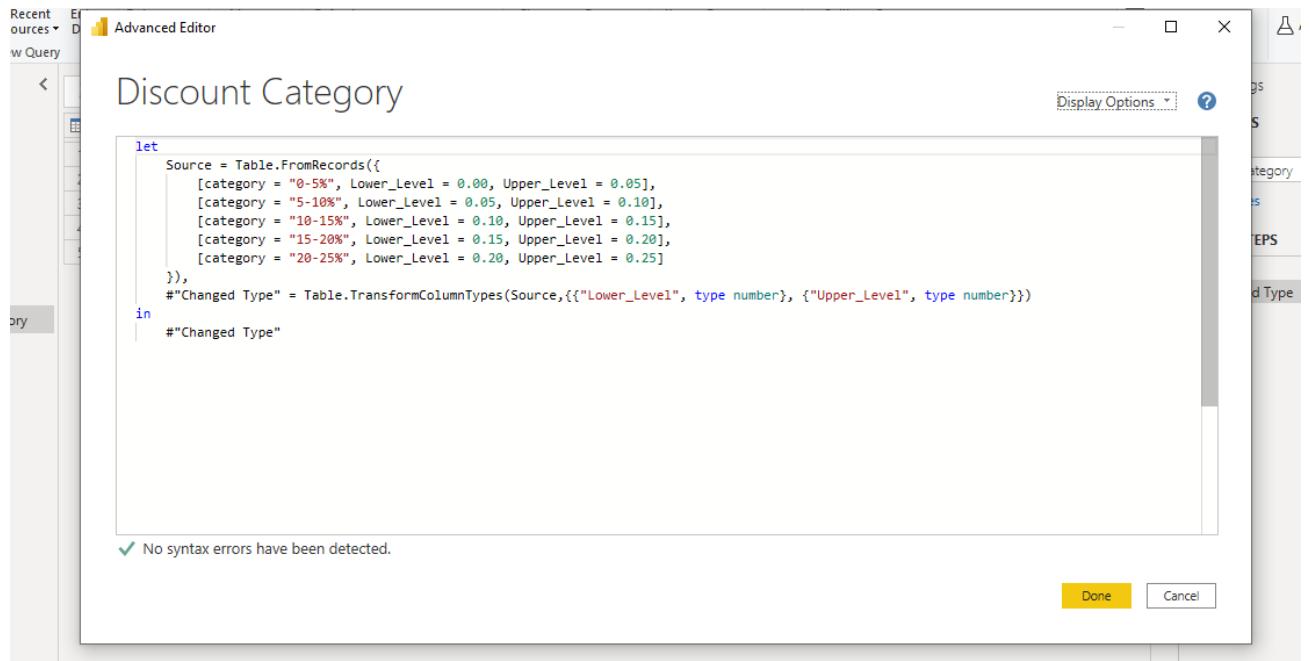


Figure 38: Screenshots of Power BI explaining the steps before creating a new table.

In Advanced editor we need to write code in M language for creating each row with percentages range of 0-5,5-10,10-15,15-20,20-25. Screenshot of the M language is give below.



```

let
    Source = Table.FromRecords({
        [category = "0-5%", Lower_Level = 0.00, Upper_Level = 0.05],
        [category = "5-10%", Lower_Level = 0.05, Upper_Level = 0.10],
        [category = "10-15%", Lower_Level = 0.10, Upper_Level = 0.15],
        [category = "15-20%", Lower_Level = 0.15, Upper_Level = 0.20],
        [category = "20-25%", Lower_Level = 0.20, Upper_Level = 0.25]
    }),
    #"Changed Type" = Table.TransformColumnTypes(Source,{{"Lower_Level", type number}, {"Upper_Level", type number}})
in
    #"Changed Type"

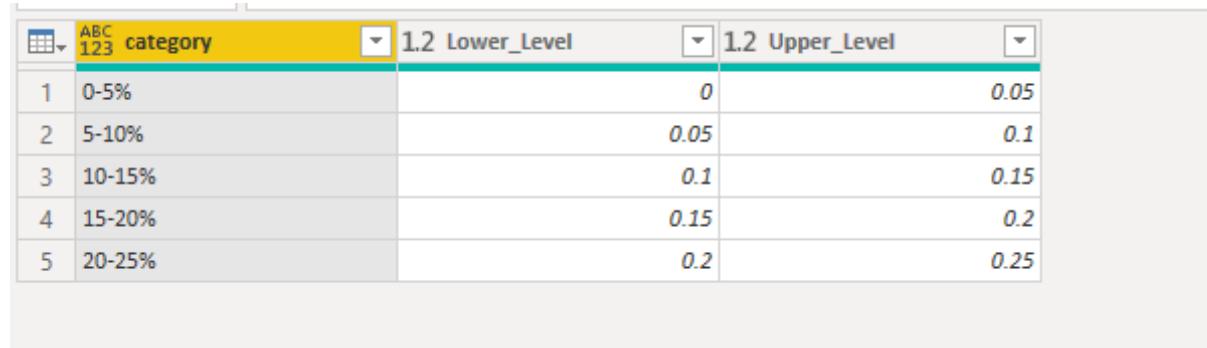
```

No syntax errors have been detected.

Done Cancel

Figure 39: Screenshot of the M language code for creating a Discount Category table.

Then click on done a new table with name ‘Discount Category’ will be created. Change the datatypes for Lower\_Level and Upper\_Level as decimal and it looks like below.



	category	Lower_Level	Upper_Level
1	0-5%	0	0.05
2	5-10%	0.05	0.1
3	10-15%	0.1	0.15
4	15-20%	0.15	0.2
5	20-25%	0.2	0.25

Figure 40: Screenshot of the new table ‘Discount Category’.

Now create a new column in the ‘Order Sales’ table with the name of ‘Ranking discount’. While creating a column the DAX formulae for it is shown below.



```

1 Ranking Discount =
2 VAR CatVar=[Order Item Discount Rate]
3 RETURN
4 CALCULATE (VALUES ('Discount Category'[category]), CatVar>'Discount Category'[Lower_Level],CatVar<='Discount Category'[Upper_Level])

```

Figure 41: Screenshot of a new DAX formulae for the creation of a column in ‘Order Sales’ table.

Using the above DAX formulae in calculated column will sort the Order Item Discount Rate value according to it in this new column which means if the Discount rate is 4% then in that new column Ranking Discount it will be sorting as 0-5%. Screenshot of this is given below.

Ranking Discount
0-5%
5-10%
5-10%
10-15%
10-15%
5-10%
5-10%
15-20%
15-20%
0-5%
0-5%
10-15%
15-20%
15-20%

Figure 42: Screenshot of the new column ‘Ranking discount’.

Now, a clustered column chart is created with ‘Order Item Quantity’ and ‘Ranking Discount’ from ‘Order Sales’ table. But a slight problem as occurred that order of the Ranking Discount is showing as 5-10%, 10-15%, 20-25%, 15-20%, 0-5%. To sort this, we should create new table named ‘Sort table’ with two columns (Ranking Discount & Order) in it which looks like below one.

Ranking Discount	Order
0-5%	1
5-10%	2
10-15%	3
15-20%	4
20-25%	5

Figure 43: Screenshot of the table ‘Sort Table’ showing the two columns in it.

Now let’s create a new column in the ‘Order Sales’ table with the name ‘Sort table’ with the DAX formula

```
X ✓ 1 Sort order = RELATED('Sort Order'[Order])
```

This formula connects ‘Sort Order’ column in ‘Order Sales’ table to ‘Order’ column in ‘Sort Table’ to present the ‘Ranking Discount’ in the order way. Before that select ‘Sort by column’ as ‘Sort order’ for the ‘Ranking Discount’ column.

Sort order
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

Figure 44: Screenshot of creation of the new column 'Sort Order' in 'Order Sales' table.

Now, we can see that 'Ranking Discount' is presented in an orderly manner in the clustered column chart as below.

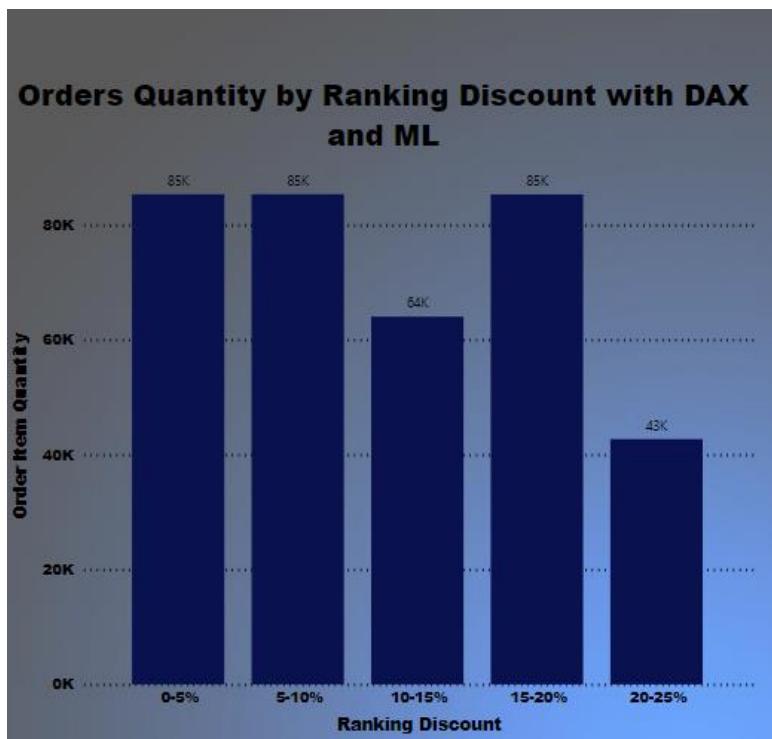


Figure 45: Screenshot of clustered column chart with Order Item Quantity by Ranking Discount.

From the above chart it is found that 85340 Order Items are in the 5-10% discount category, 85333 order items are in 0-5%, 85305 order items are in 15-20%, 64028 order items are in 10-15%, 42706 order items are in 20-25%.

#### Timeline Slicer chart (Chart not covered in lesson):

As part of the ICA, it is needed to show a type of chart which is not included in any of the lesson. For this, based on my Data Model I have chosen chart called Timeline. It is sort of slicer but with time

which is useful for monitoring the sales for the FY 2015-2016. The screenshot of the chart is given below.



Figure 46: Screenshot of the Timeline chart for the FY 2015-2016

This chart is used to monitor the sales in FY 2015-2016 and for that we need to add a line chart of sales by Orders date in the same page. We can set with the timeline slicer and I want to know the sales info of the fourth quarter of FY 2015.

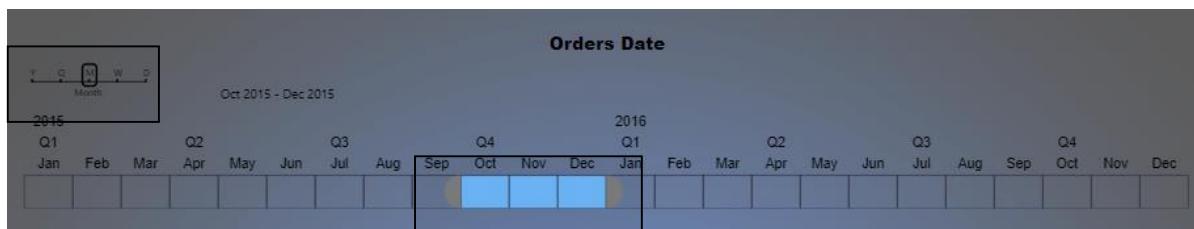


Figure 47: Screenshot of the Timeline Slicer showing the adjustment of Q4 of FY 2015.

We can also set the slicer for Year, Quarter, Week and Day. The result for the Sales in fourth quarter of FY 2015 is shown below.



Figure 48: Screenshot of sales information with a Line chart and a Timeline slicer.

**NOTE:** For all these above charts their dashboard pages have been added with a background and for changing the fonts, fonts size, adding legends, different data colours, title for each chart, changing the background ‘FORMAT’ tab in Visualizations sections has been used.

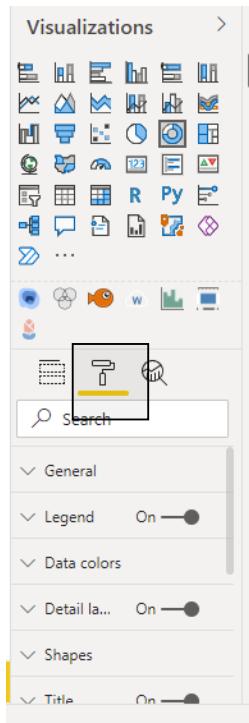
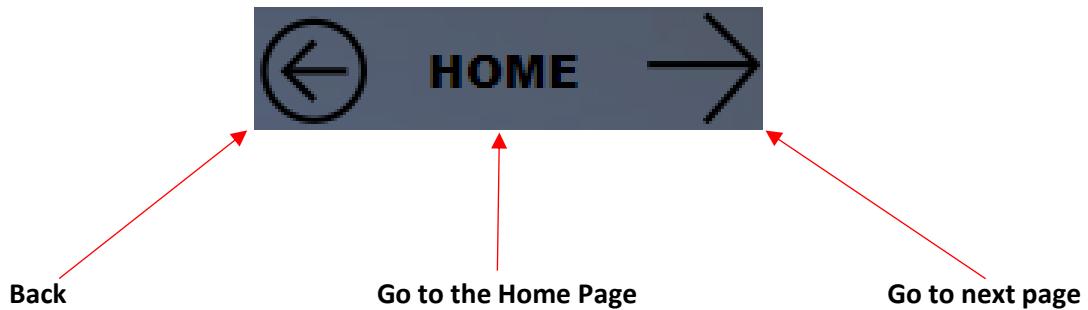


Figure 49: Screenshot of Power BI showing the ‘FORMAT’ tab.

#### Buttons:

Buttons in Power BI are utilized to create reports that behave like applications, and to create an environment where users can hover over, click on, and interact with Power BI content. In this dashboard mainly three buttons are in every page of Power BI file which are Back, Home, Next page.



The above buttons are used for navigating through different pages. The home page has button to enter into many pages and for each dashboard page there is a Back button, Home button, Next page button to go back and forth from that specific page.

Screenshots of the final dashboard:



Figure 50: Screenshot of HOME page in the dashboard.

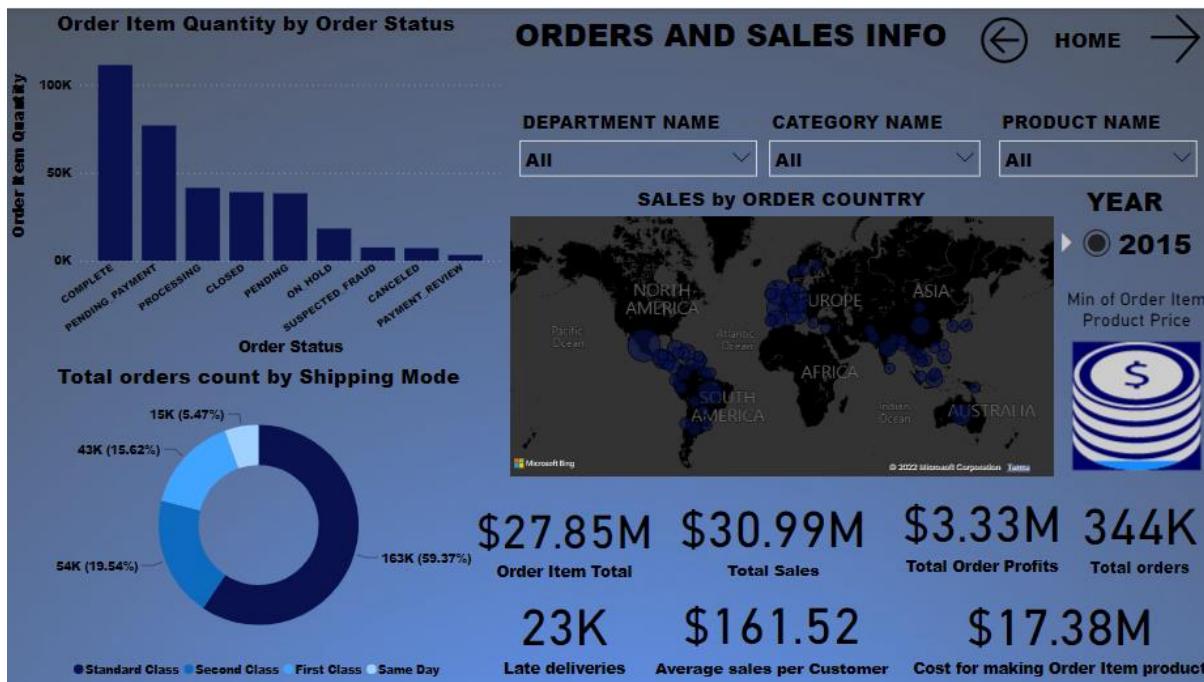


Figure 51: Screenshot of Orders and Sales Info page in the dashboard.

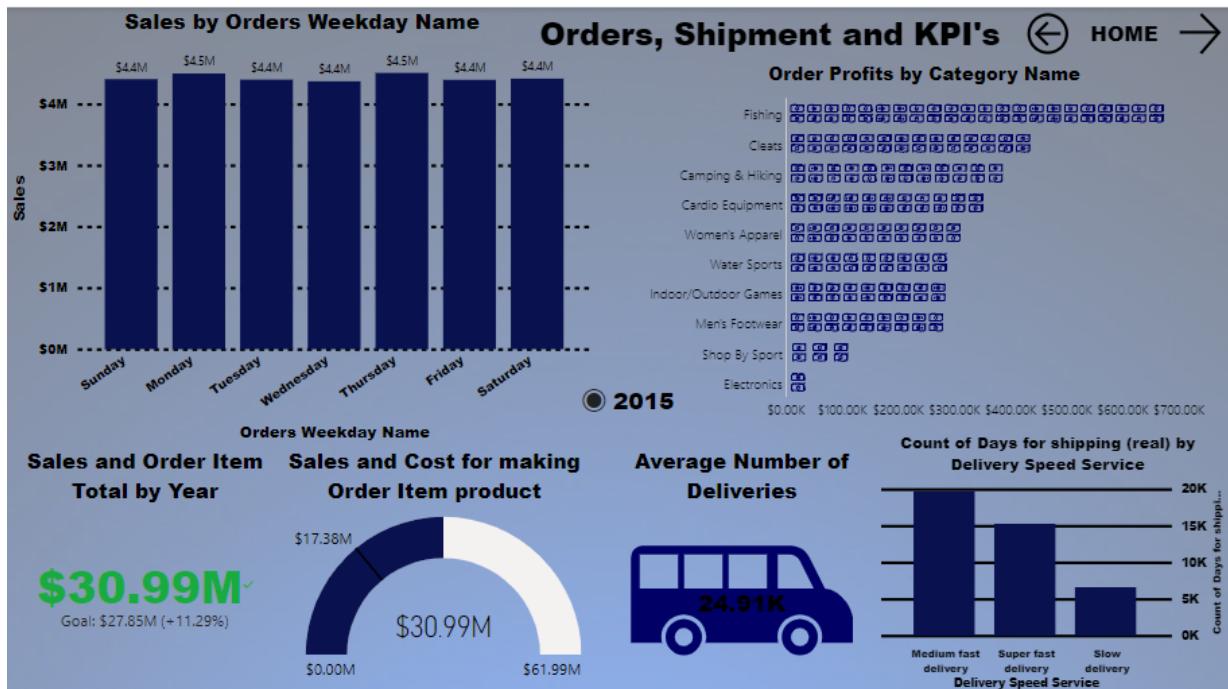


Figure 52: Screenshot of Orders, Shipment and KPI's page in the dashboard.

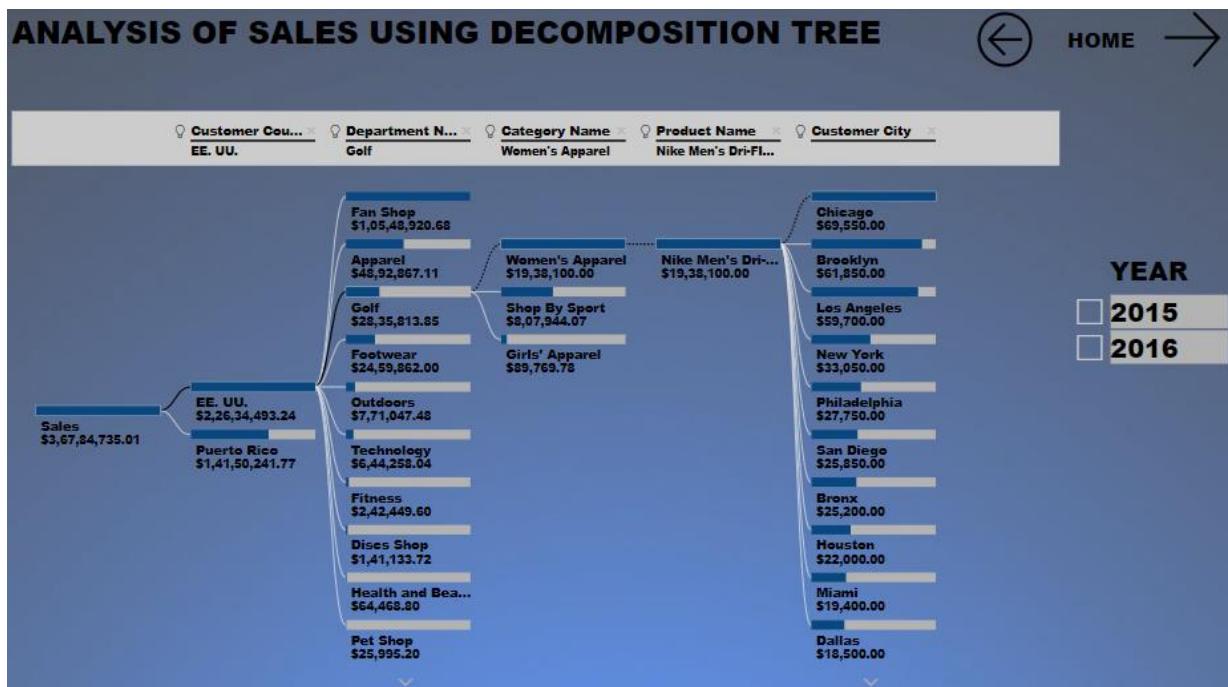


Figure 53: Screenshot of the page 'Analysis of sales using decomposition tree' in the dashboard.

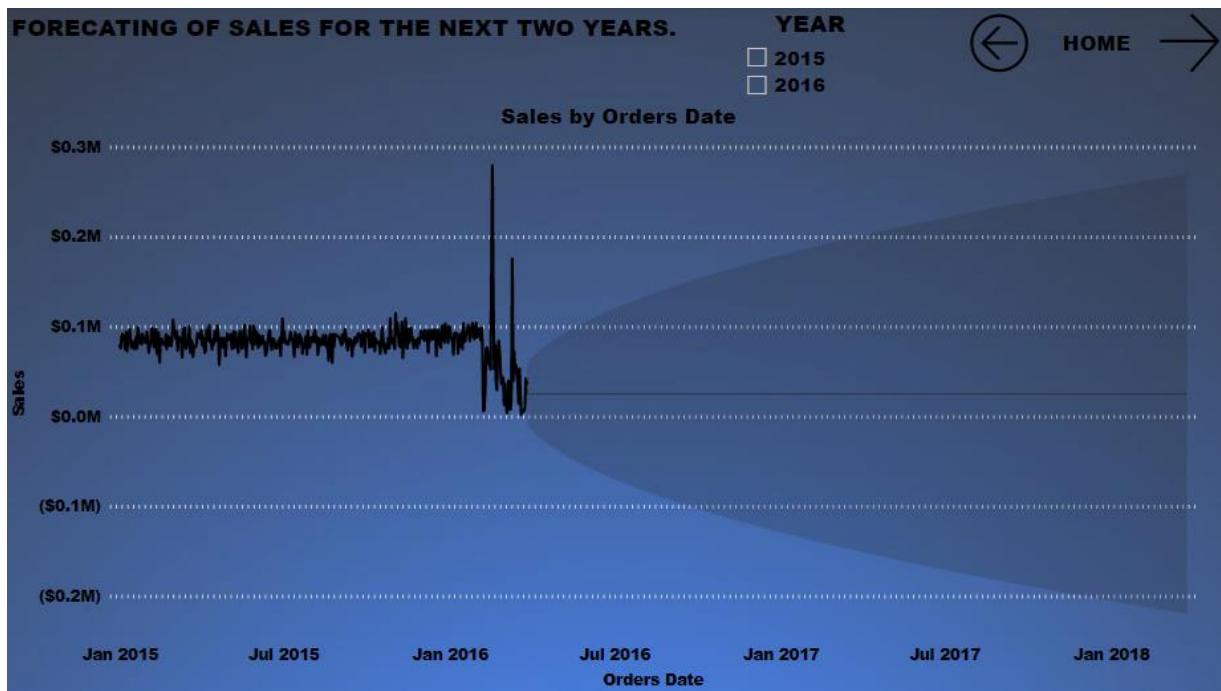


Figure 54: Screenshot of page 'Forecasting of sales for the next two years' in the dashboard.

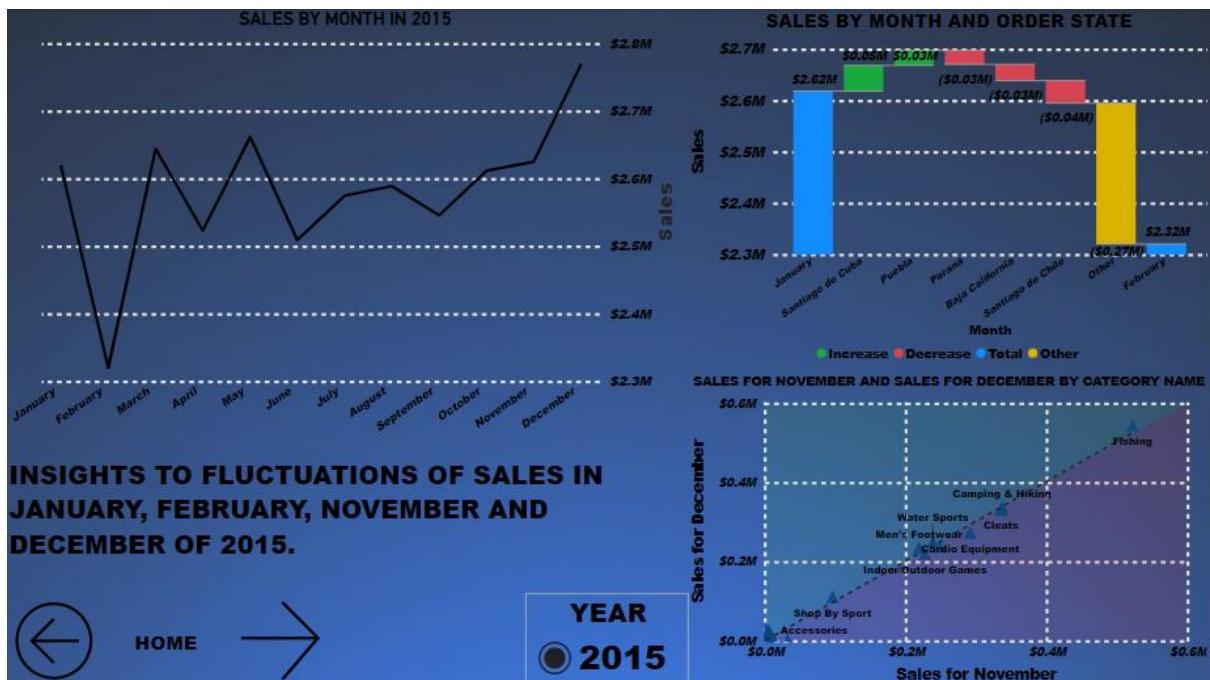


Figure 55: Screenshot of page 'Insights to sales fluctuations in the year 2015' in the dashboard.

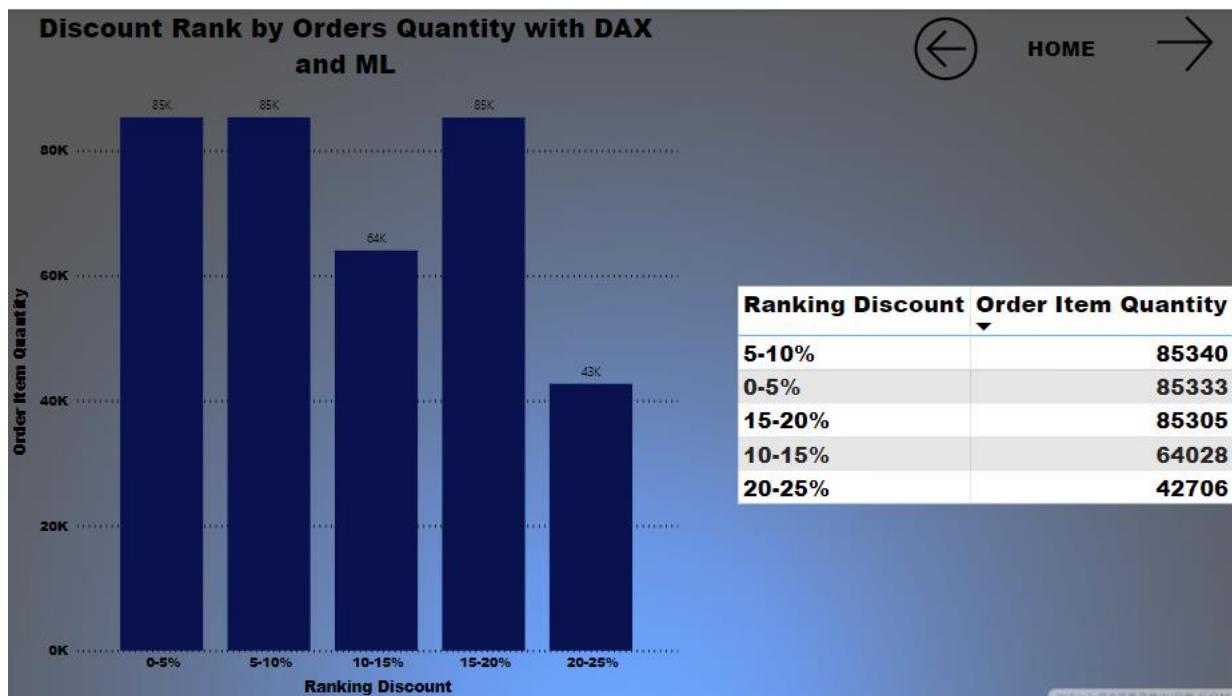


Figure 56: Screenshot of page ‘Discount rank by orders quantity with DAX and ML’ in the dashboard.

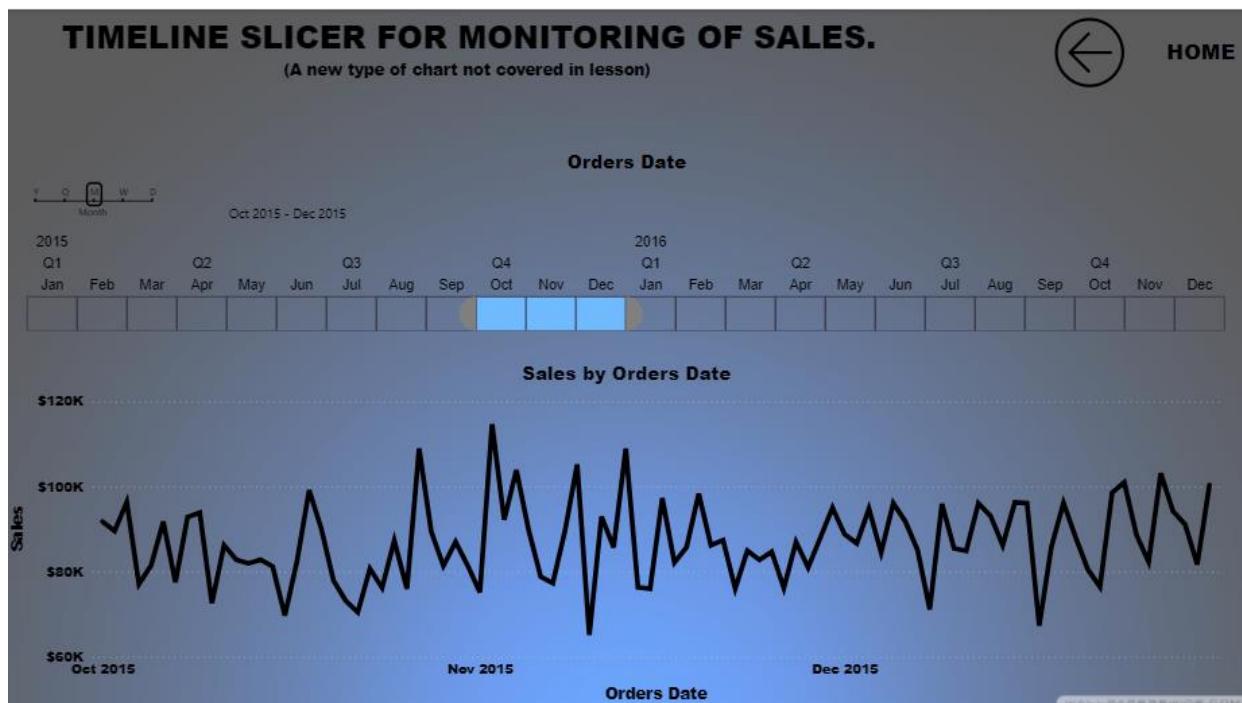


Figure 57: Screenshot of the page ‘Timeline slicer for monitoring of sales’ in the dashboard.

## Publishing the Power BI report:

Once the Power BI report is ready, next step is to publish it into the My Workspace in Power BI. In Home click on ‘Publish’ and then it will look like below screens.

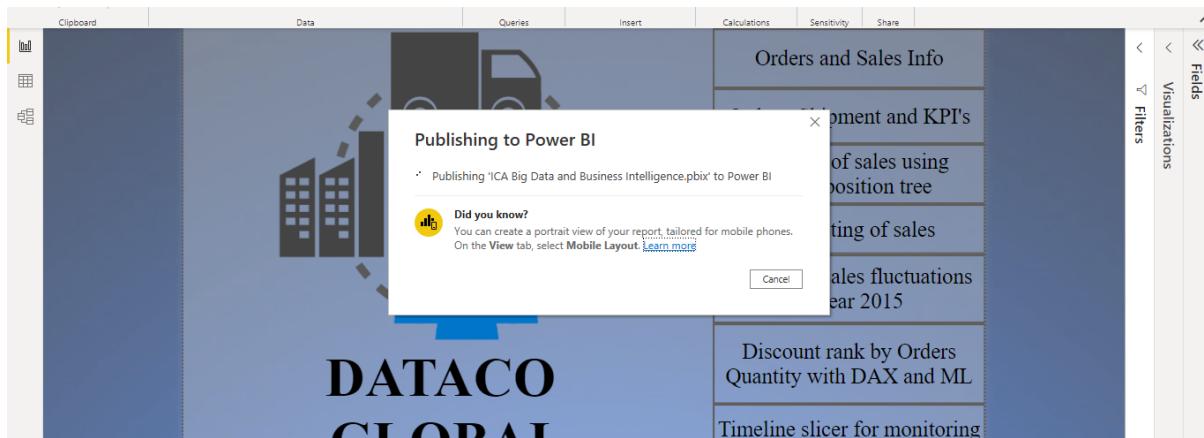


Figure 58: Screenshot of Power BI while publishing the report.

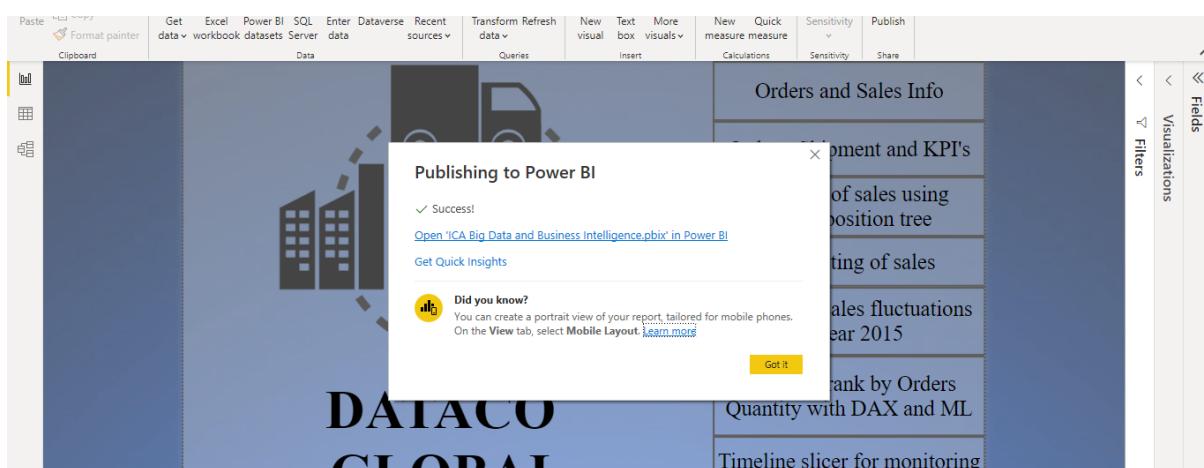


Figure 59: Screenshot of Power BI after the report is published in the ‘My Workspace’.

A **QR code** is generated to open Power BI report directly from the mobile.



## **CONCLUSIONS AND RECOMMENDATIONS**

### **Summary of the Report:**

With the report and the dashboard, the sales, orders, shipment and profits in the FY 2015 can be easily analysed by the superiors of the company Data Co Global and new implementations and improvements can be made.

### **After creating the dashboard and the report, these below findings are drawn:**

- Total Sales made in the financial year 2015 are \$30.99M with the Total Orders placed by the customers are 344K Orders.
- Total Cost for making for all the Order item products in the financial year 2015 is \$17.38M.
- Total Order Profits in the financial year 2015 is \$3.33M.
- In FY 2015 total cost for making of all the order items products is \$17.38M and total Sales made are \$30.99M. So, the profits made as of now are \$13.61M and we have seen that Total Profits made by the producer are \$3.33M, \$10.28M goes to the Seller for the items that they have given to producer for manufacturing.
- There are a total 23K late deliveries in the FY 2015.
- Averagely \$161.52 sales were made to each individual customer in the FY 2015.
- Monday was the weekday in the FY 2015 that the more Sales were happening compared to other weekdays.
- A sales analysis is made with the help of a decomposition tree and found that most sales were happening in United States (EE. UU) compared to Puerto Rico (US Territory in North America). The decomposition tree in Power BI uses AI and Machine Learning techniques and finds out the high and low value with Customer City, Department Name, Category Name, Product Name.
- In FY 2016 from January to March, there is a tremendous downfall in the Orders, sales, profits with the company. Based on the data of FY 2015(January to December)-2016(January to March) forecasting of sales is made in the power BI for the next two years and results will be discussed further in this report.
- There were huge sales fluctuations in FY 2015 that sales going down from January to February and going up from November to December. Insights for these fluctuations are explained in this report.
- Count of Total orders with respective of their discount percentage which are 0-5%,5-10%,10-15%,15-20%,20-25%. Explanation for this is further given in this report.
- Sales of fourth quarter of the FY 2015 in the form a line chart are monitored with a use of a Timeline slicer chart.

### **Based on these findings, some conclusions are proposed which are given below:**

- Customers prefer standard class service than any other shipping mode in the FY 2015 as it is the faster one.
- There are more than 100K completed order status in the FY 2015 but it can be more if the pending order status are less.
- In the FY 2015, Thursday is the most happening weekday in terms of Sales and it is just a midweek day.

- A sales target of \$27.85M is achieved and went beyond it and made it up to \$30.99M (+11.29%) in 2015 which is called a good performance in the supply chain process.
- With the help of the Gauge chart, it is known that Sales are higher than the total cost of making order item product in FY 2015. It is analysed that some profits are made out of it
- Sales for the next two years after 2016 March would not be satisfying as it was in the FY 2015.
- The order purchased quantity for the discount category 20-25% are less because it might be that prices of the products with this discount category are high.

### **Personal Conclusions:**

As a 1<sup>st</sup> year master's student, studying the module Big Data and Business Intelligence was a start in career to become a **Data Scientist**. Thanks to teaching faculty and Teesside University for using Microsoft Power BI throughout the course. Power BI is very effective and useful in creation of various charts and dashboards. It is the best tool to answer our business questions and analytics.

At the beginning of the course, selecting a right dataset was a bigger challenge for performing the necessary things in the Power BI tool. More than 15 columns are required to showcase the skills which are acquired throughout the module. Module leader **Dr. Annalisa Occhipinti** and module instructors are of a great help regarding this.

By doing this ICA, I am very confident that I developed a lot of skills with respect to this module and with the Power BI tool as well. I feel this is a best way to start a career in the field of **Data Science**.

### **Recommendations:**

Based on the findings, some recommendations are proposed:

- Reduce the number of cancelled, pending orders as they are high in the FY 2015.
- Increase the delivery speed of the orders as there would be a chance in the increase in orders.
- It is recommended to reduce but not stop completely stop the sales in the states like Parana, Baja California, Santiago de Chile as these states caused decrease in sales. Countries Mexico, Brazil, Colombia accounted for the majority of the decrease in sales in FY 2015.
- There is also an increase in sales by 5.47% between November and December in the FY15. Reason for this rise in sales in 'Kids Golf Club', Men's Golf Clubs and Fishing with respective to the Category Name. There is also a decrease in sales for the Category 'Electronics'. So, it is recommended to increase the production and sales with these categories in the regions South Asia, Oceania, South East Asia in November and December.
- 'Fishing' Category had highest number of Order Profits in the FY 2015 of \$677.17K by making the Sales of \$6.28M in total. It is recommended to make more production in Fishing category as we can expect more profits from it.
- Last recommendation is that focus on the marketing of the company to expect increase in orders, sales and profits.

**References:**

Guy in a cube. (2019, August 28). 2 ways to reduce your BI dataset size. Retrieved from [https://www.youtube.com/watch?v=c-ZqToc85Yc&t=4s&ab\\_channel=GuyinaCube](https://www.youtube.com/watch?v=c-ZqToc85Yc&t=4s&ab_channel=GuyinaCube)

Kevin, Stratvert. (2021, May 16). Using Power BI DAX. Retrieved from [https://www.youtube.com/watch?v=waG\\_JhBgUpM&t=1847s&ab\\_channel=KevinStratvert](https://www.youtube.com/watch?v=waG_JhBgUpM&t=1847s&ab_channel=KevinStratvert)

Sam, Mckay. (2021). Supply chain management techniques in Power BI. Retrieved from <https://blog.enterpriseDNA.co/supply-chain-management-techniques-in-power-bi/>

Andy, Brown. (2018). Creating Measures using Dax. Retrieved form <https://www.red-gate.com/simple-talk/databases/sql-server/bi-sql-server/creating-measures-using-dax/>