



# BIG DATA AND BUSINESS INTELLIGENCE IN-COURSE ASSESSMENT

BOWALE ADEDAMOLA  
APPLIED DATA SCIENCE(MSc.)  
A0353496

January 12, 2022

Section 1: Business Intelligence Design

## Table of Contents

SECTION 1 – BUSINESS INTELLIGENCE DESIGN .....	4
1. BI Data Source Description and BI Requirements.....	4
1.1 Dataset.....	4
1.2 Business Requirements .....	5
2. BI Data Splitting and Data Pre-processing .....	6
2.1 Loading Data .....	6
2.2 Splitting Data (Generating Dimension and Fact Tables) .....	9
a. Creating Dimension Table (D_Order).....	9
b. Creating Dimension Table (D_Customer) .....	12
c. Creating Dimension Table (D_Product) .....	15
d. Creating Dimension Table (D_Location) .....	17
e. Creating fact Table (Transaction) .....	19
f. Creating dimension Table (D_Continent) .....	21
g. Creating date Table (D_Calendar).....	24
3. BI Data Modelling.....	29
SECTION 2 – BUSINESS INTELLIGENCE SOLUTION .....	31
1. Executive Summary.....	31
Abstract.....	31
Key Findings .....	31
Recommendations .....	32
2. Introduction .....	32
Data Model .....	32
3. Findings based on analysis and evaluation:.....	33
Table Creation (For Analysis) .....	33
Data analysis steps .....	35
PowerBI Visuals.....	40
Top 5 Company By Sales .....	40
Bottom 5 Company by Sales .....	40
Total Sales By Year .....	41
Total Sales By Month .....	41

Weekday Sales Trend.....	42
Weekend Sales Trend .....	42
Sales By Order Status .....	43
Quantity By Order Status .....	43
Product Line by Sales .....	44
Product Line by Quantity .....	44
Top 5 Countries by Total Sales .....	45
Top 5 City by Total Sales .....	45
Scorecards.....	46
Scroller .....	48
Description of the Power BI dashboard.....	50
Key findings.....	53
Conclusion.....	53

# SECTION 1 – BUSINESS INTELLIGENCE DESIGN

## 1. BI Data Source Description and BI Requirements

### 1.1 Dataset

“Sales Sample Data” is an open-source dataset that contains information about the Sales of some Transportation products (Such as Trains, Classic cars, Vintage cars, Trains, Ships, Motorcycles, Aeroplanes, Buses and Trucks) to different companies that are located in the different parts of the world. The dataset was gotten from Kaggle (<https://www.kaggle.com/kyanyoga/sample-sales-data>). The dataset has 25 Columns and 2,824 rows.

The image below shows the screenshot of the “Sales Sample Data” dataset in CSV format.

The screenshot shows an Excel spreadsheet titled "sales\_data\_sample - Excel". The spreadsheet contains a single table named "Table1". The columns are labeled A through V. The data includes columns for Order Number, Quantity Ordered, Price Each, Order Line Number, Order Date, Status, Quarter ID, Month ID, Year ID, Product Line, MRP, Product ID, Customer ID, Phone Number, Address, City, State, Postal Code, Country, Territory, and Contact. The data spans from row 1 to row 2824, with various product types like Motorcycles, Toys, and Electronics listed across different countries and territories.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTID	CUSTOMERID	PHONE	ADDRESS	ADDRESS2	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY	CONTACT
2	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1	2	2003	Motorcycle	95	S10_1678	Land of Tc	2.13E+09	897 Long Airport Aver	NYC	NY	10022	USA	NA	Young	
3	10121	34	81.35	5	2765.9	5/7/2003 0:00	Shipped	2	5	2003	Motorcycle	95	S10_1678	Reims Coll	26.47.1551	59 rue de l'Abbaye	Reims		51100	France	EMEA	Henric	
4	10134	41	94.74	2	3884.34	7/1/2003 0:00	Shipped	3	7	2003	Motorcycle	95	S10_1678	Lyon Souvi	+33 1 46 6 27 74	77 rue du Colonel Pie	Paris		75508	France	EMEA	Da Cu	
5	10145	45	83.26	6	3746.7	8/25/2003 0:00	Shipped	3	8	2003	Motorcycle	95	S10_1678	ToysGrov	6.27E+09	78934 Hillsde Dr.	Pasadena	CA	90003	USA	NA	Young	
6	10159	49	100	14	5205.27	10/10/2003 0:00	Shipped	4	10	2003	Motorcycle	95	S10_1678	Corporate	6.51E+09	7734 Strong St.	San Franci	CA	94217	USA	NA	Brow	
7	10168	36	96.66	1	3479.76	10/28/2003 0:00	Shipped	4	10	2003	Motorcycle	95	S10_1678	Technics	6.51E+09	9408 Furth Circle	Burlings	CA	94217	USA	NA	Hiran	
8	10180	29	86.13	9	2497.77	11/11/2003 0:00	Shipped	4	11	2003	Motorcycle	95	S10_1678	Daedalus	20.16.1551	184, chausse de Touri Lille			59000	France	EMEA	Rance	
9	10188	48	100	1	5512.32	11/18/2003 0:00	Shipped	4	11	2003	Motorcycle	95	S10_1678	Herkie	47.4472	2267 : Drammen 121, PR 74+	Bergen		N 5804	Norway	EMEA	Oezta	
10	10201	22	98.57	2	2168.54	12/1/2003 0:00	Shipped	4	12	2003	Motorcycle	95	S10_1678	Minil Whee	6.51E+09	5557 North Pendale	S San Franci	CA	75016	France	EMEA	Perr	
11	10211	41	100	14	4708.44	1/15/2004 0:00	Shipped	1	1	2004	Motorcycle	95	S10_1678	Auto Cana	(1) 47.55.625	9, rue Lauriston	Paris		10022	USA	NA	Murpl	
12	10223	37	100	1	3965.66	2/20/2004 0:00	Shipped	1	2	2004	Motorcycle	95	S10_1678	Australian Min	9.50 2020 45.636	St Kild Level 3	Melbourne	Victoria	3004	Australia	APAC	Fergus	
13	10237	23	100	7	2333.12	4/5/2004 0:00	Shipped	2	4	2004	Motorcycle	95	S10_1678	Vitachrom	2.13E+09	2678 Kings Suite 101	NYC	NY	10022	USA	NA	Frick	
14	10251	28	100	2	3188.64	5/18/2004 0:00	Shipped	2	5	2004	Motorcycle	95	S10_1678	Tekni Colle	2.02E+09	7476 Moss Rd.	Newark	NJ	94019	USA	NA	Brow	
15	10263	34	100	2	3676.76	6/28/2004 0:00	Shipped	2	6	2004	Motorcycle	95	S10_1678	Gilt Depot	2.04E+09	25593 South Bay Ln.	Bridgewat	CT	97562	USA	NA	King	
16	10275	45	92.83	1	4177.35	7/23/2004 0:00	Shipped	3	7	2004	Motorcycle	95	S10_1678	Marla's Re	6.18E+09	3932 Spinaker Dr.	Cambridge	MA	44000	France	EMEA	Labru	
17	10285	36	100	4	4099.68	8/27/2004 0:00	Shipped	3	8	2004	Motorcycle	95	S10_1678	Marta's	6.18E+09	9557 North Pendale	S San Franci	CA	51247	USA	NA	Herna	
18	10299	23	100	9	2597.39	9/30/2004 0:00	Shipped	3	9	2004	Motorcycle	95	S10_1678	La Rochell	40.67.8551	67, rue des Cinqante Nantes			21240	Finland	EMEA	Karttu	
19	10309	41	100	5	4394.38	10/15/2004 0:00	Shipped	4	10	2004	Motorcycle	95	S10_1678	Baane Min	07-98 955	Erling Skakkes gate 78	Stavern		4110	Norway	EMEA	Bergu	
20	10318	46	94.74	1	4358.04	11/2/2004 0:00	Shipped	4	11	2004	Motorcycle	95	S10_1678	Diecast Cl	2.16E+09	7586 Pompton St.	Allentown	PA	70267	USA	NA	Yi	
21	10329	42	100	1	4396.14	11/15/2004 0:00	Shipped	4	11	2004	Motorcycle	95	S10_1678	Land of Tc	2.13E+09	997 Long Airport Aver	NYC	NY	10022	USA	NA	Yi	
22	10341	41	100	9	7737.93	11/24/2004 0:00	Shipped	4	11	2004	Motorcycle	95	S10_1678	Salzburg C	6562-9555	Geislweg 14	Salzburg		5020	Austria	EMEA	Pipps	
23	10361	20	72.55	13	1451	12/17/2004 0:00	Shipped	4	12	2004	Motorcycle	95	S10_1678	Souvenirs +61	2 9495	Monitor N/ Level 6	Chatswo	-NSW	2067	Australia	APAC	Huxle	
24	10375	21	34.91	12	733.11	2/3/2005 0:00	Shipped	1	2	2005	Motorcycle	95	S10_1678	La Rochell	40.67.8551	67, rue des Cinqante Nantes			44000	France	EMEA	Labru	
25	10388	42	76.36	4	3207.12	3/3/2005 0:00	Shipped	1	3	2005	Motorcycle	95	S10_1678	FunGifte	5.09E+09	1785 First Street	New Bedf	MA	50553	USA	NA	Benite	
26	10403	24	100	7	2434.56	4/8/2005 0:00	Shipped	2	4	2005	Motorcycle	95	S10_1678	UK Collect	(171) 555	Berkeley Gardens 12	Liverpool		WX1 6LT	UK	EMEA	Devor	
27	10417	66	100	2	7516.08	5/13/2005 0:00	Disputed	2	5	2005	Motorcycle	95	S10_1678	EuroShop	(91) 555 94	C/ Moralzarzal, 86	Madrid		28034	Spain	EMEA	Freyre	
28	10103	26	100	11	5404.62	1/29/2003 0:00	Shipped	1	1	2003	Classic Car	214	S10_1949	Baane Min	07-98 955	Erling Skakkes gate 78	Stavern		4110	Norway	EMEA	Bergu	
29	10110	70	100	1	7300.11	2/18/2003 0:00	Shipped	1	2	2002	Classic Car	214	S10_1040	Volvo Mar	00311.12.21	Bromma, non	Tulu		C 0E9.77	Sweden	EMEA	Bozeni	

Figure 0.1: Screenshot of the CSV file.

The Table below describes the “Sales Sample Data” dataset.

S/N	Column Name	Description
1	ORDERNUMBER	This is a Unique Number per Order
2	QUANTITYORDERED	The quantity of Product that was ordered
3	PRICEEACH	The Price of each product that was ordered
4	ORDERLINENUMBER	The Individual line items in a Sales order
5	SALES	The total amount of the product that was sold per order
6	ORDERDATE	The date the Order was made
7	STATUS	The Status of the Order
8	QTR_ID	The Quarter of the Year in which the transaction falls
9	MONTH_ID	The Month of the Year in which the transaction falls
10	YEAR_ID	The Year in which the transaction happened
11	PRODUCTLINE	The type of Product being sold

12	MSRP	Manufacturer Suggested Retail price
13	PRODUCTCODE	The Unique Product ID
14	CUSTOMERNAME	The Company (Customer) Name that purchased a product
15	PHONE	Phone Number of the Customer
16	ADDRESSLINE1	Customer's Address 1
17	ADDRESSLINE2	Customer's Address 2
18	CITY	Customer's City
19	STATE	Customer's State
20	POSTALCODE	Customer's Postal code
21	COUNTRY	Customer's Country
22	TERRITORY	Customer's Territory
23	CONTACTLASTNAME	Customer's last name
24	CONTACTFIRSTNAME	Customer's first name
25	DEALSIZE	The Size of the Deal/Order

Table 1.1: "Sales Sample Data" dataset description

## 1.2 Business Requirements

The process performance metrics (or KPI) for this business falls majorly under the Financial Perspective (SALES) and a bit under the Internal Process perspective (Order, Order Status, etc.). This analysis will help to answer the business intelligence questions below.

- Customers with highest and lowest sales record
- Year with highest and lowest sales record
- Month with highest and lowest sales record (for weekend and weekdays also)
- Sales record based on the status of the order (Shipped order, Cancelled order, etc.)
- Sales record by unique products
- Sales record by Location

Business intelligence generally helps company to make better and smarter decisions however, the business intelligent questions above helps us to narrow down where exactly to look at. The answer/information it provides will help to Identify ways to increase sales, understand customer's location density for expansion, identify which product to increase/decrease its production, identify the causes of low sales per timeframe/location.

This information is usually needed by Top Managements of the company in order to make the final decision going forward.

## 2. BI Data Splitting and Data Pre-processing

### 2.1 Loading Data

In order to use PowerBI for analysis, the first step is to import the data into PowerBI. This could be through an API or Database connection, Importing a CVS or an Excel file, etc. By clicking the “Get data” button as shown below (Figure 2.1), the “Text/CSV” button was selected to specify the format of the data to be loaded.

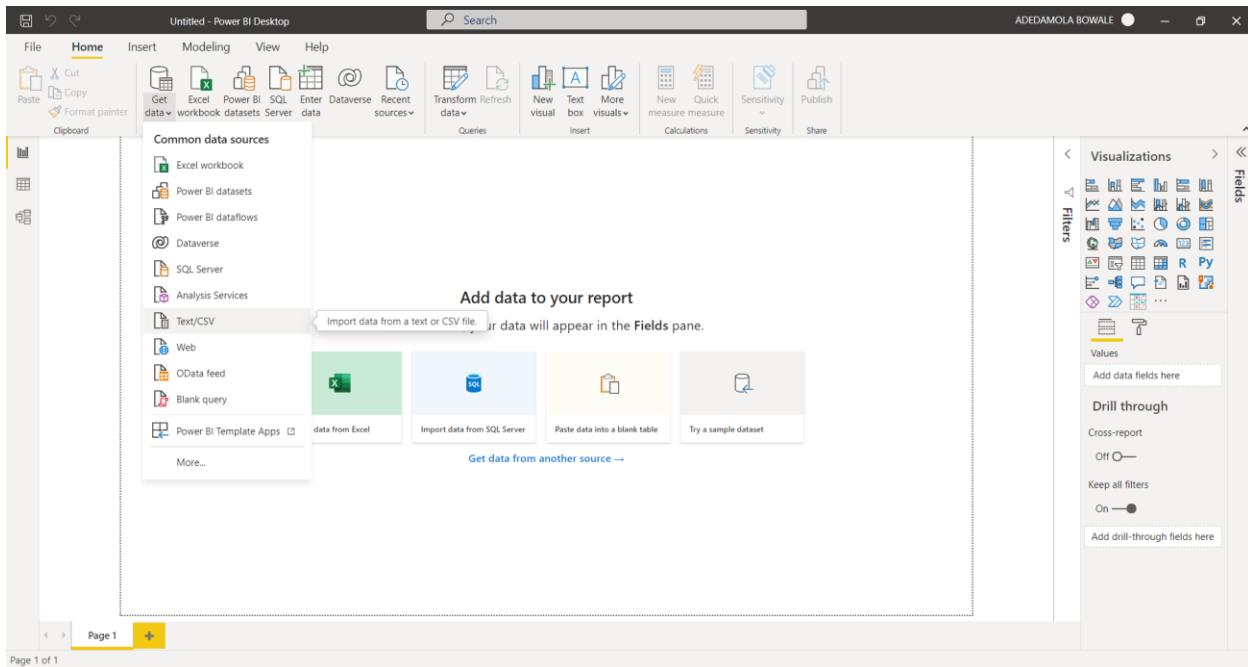


Figure 0.2.1: Loading Data (Before)

After navigating to the location of the dataset and then clicked on, the image below (Figure 2.2) will pop up. Then the “Load” button can be clicked. By doing this, the dataset has been loaded successfully into PowerBI. The table view of the dataset can now be viewed in PowerBI as seen in Figure 2.3.

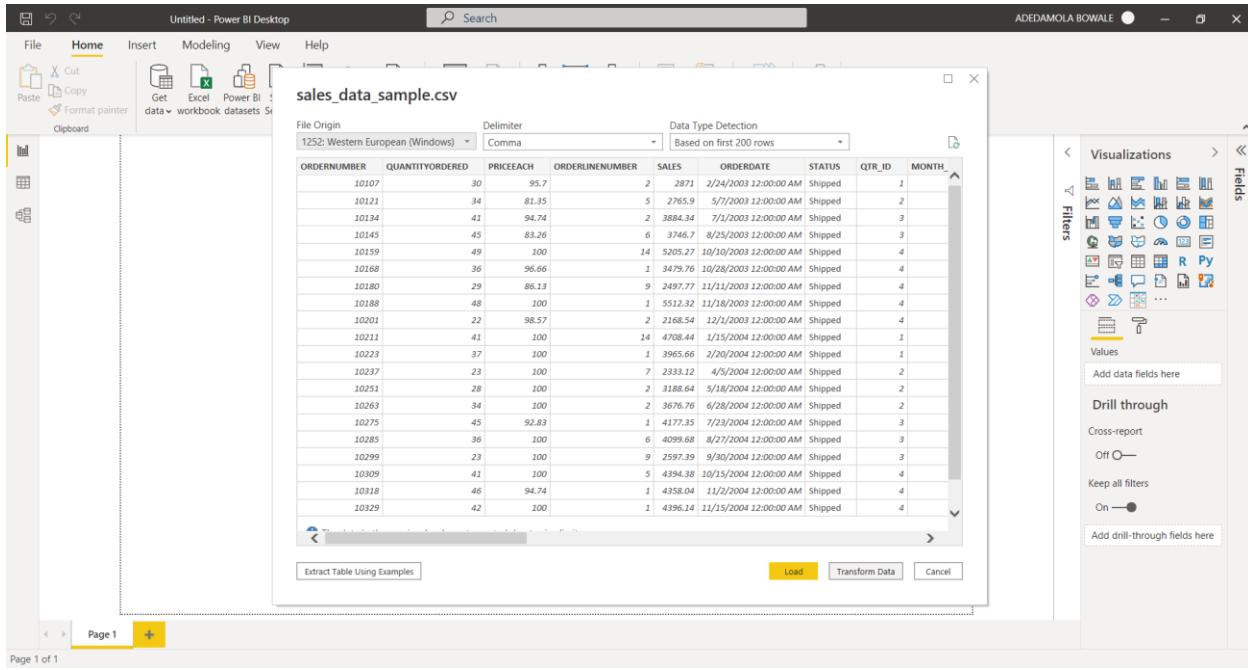


Figure 2.2: Loading Data (During)

ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME
10203	20	100	8	3930.4	12/2/2003 12:00:00 AM	Shipped	4	12	2003 Classic Cars	194	S12_1099	Euro Shopping
10153	20	100	11	4904	9/28/2003 12:00:00 AM	Shipped	3	9	2003 Classic Cars	207	S12_1108	Euro Shopping
10104	34	100	1	5958.5	1/31/2003 12:00:00 AM	Shipped	1	1	2003 Classic Cars	157	S12_3148	Euro Shopping
10153	42	100	12	5393.64	9/28/2003 12:00:00 AM	Shipped	3	9	2003 Classic Cars	157	S12_3148	Euro Shopping
10212	39	100	16	4945.76	1/16/2004 12:00:00 AM	Shipped	2	1	2004 Classic Cars	117	S12_3380	Euro Shopping
10104	41	100	9	4615.78	1/31/2003 12:00:00 AM	Shipped	1	1	2003 Trucks and Buses	118	S12_4473	Euro Shopping
10246	46	100	5	5069.66	5/7/2004 12:00:00 AM	Shipped	2	5	2004 Trucks and Buses	118	S12_4473	Euro Shopping
10412	54	100	5	5951.34	5/3/2005 12:00:00 AM	Shipped	2	5	2005 Trucks and Buses	118	S12_4473	Euro Shopping
10203	47	100	5	5195.85	12/2/2003 12:00:00 AM	Shipped	4	12	2003 Classic Cars	115	S12_4675	Euro Shopping
10212	33	100	15	4180.44	1/16/2004 12:00:00 AM	Shipped	1	1	2004 Classic Cars	115	S12_4675	Euro Shopping
10212	29	100	10	4186.73	1/16/2004 12:00:00 AM	Shipped	1	1	2004 Classic Cars	141	S18_1129	Euro Shopping
10205	36	100	2	3735.72	12/2/2003 12:00:00 AM	Shipped	4	12	2003 Vintage Cars	102	S18_1342	Euro Shopping
10244	40	100	7	4684.8	4/29/2004 12:00:00 AM	Shipped	2	4	2004 Vintage Cars	102	S18_1342	Euro Shopping
10212	38	100	6	4492.36	1/16/2004 12:00:00 AM	Shipped	1	1	2004 Classic Cars	124	S18_1589	Euro Shopping
10379	39	100	2	5399.55	2/10/2003 12:00:00 AM	Shipped	1	2	2005 Vintage Cars	170	S18_1749	Euro Shopping
10212	41	100	9	4840.87	1/16/2004 12:00:00 AM	Shipped	1	1	2004 Classic Cars	142	S18_1984	Euro Shopping
10204	24	100	8	3457.92	1/31/2003 12:00:00 AM	Shipped	1	1	2003 Classic Cars	161	S18_2238	Euro Shopping
10246	40	100	4	6549.2	5/3/2004 12:00:00 AM	Shipped	2	5	2004 Classic Cars	161	S18_2238	Euro Shopping
10412	41	100	4	6712.93	5/3/2005 12:00:00 AM	Shipped	2	5	2005 Classic Cars	161	S18_2238	Euro Shopping
10104	29	100	12	3772.61	1/31/2003 12:00:00 AM	Shipped	1	1	2003 Trucks and Buses	122	S18_2319	Euro Shopping
10383	27	100	11	3843.99	2/22/2003 12:00:00 AM	Shipped	1	2	2003 Trucks and Buses	122	S18_2319	Euro Shopping
10380	40	100	10	4931.6	2/16/2005 12:00:00 AM	Shipped	1	2	2005 Vintage Cars	127	S18_2325	Euro Shopping
10244	43	100	8	5950.34	4/29/2004 12:00:00 AM	Shipped	2	4	2004 Vintage Cars	160	S18_2795	Euro Shopping
10212	40	100	7	4910.4	1/16/2004 12:00:00 AM	Shipped	1	1	2004 Classic Cars	130	S18_2870	Euro Shopping
10311	43	100	10	5278.68	10/16/2004 12:00:00 AM	Shipped	4	10	2004 Classic Cars	132	S18_2870	Euro Shopping
10379	29	100	5	5127.2	2/10/2005 12:00:00 AM	Shipped	1	2	2005 Classic Cars	132	S18_2870	Euro Shopping
10744	30	100	1	3525.76	4/29/2004 12:00:00 AM	Shipped	?	4	2004 Vintage Cars	101	S18_2949	Finn Shinnin

Table: sales\_data\_sample (2,823 rows)

Figure 2.3: Loaded Data (after)

After looking at the data closely, I noticed the Date related columns (Delete QTR\_ID, MONTH\_ID and YEAR\_ID) and I decided to delete them as I prefer to work with a Date Table (which would be populated later). From the home tab, the “Transform data” button should be clicked in order to carry out this process as seen in figure 2.4.

Untitled - Power BI Desktop

File Home Help Table tools

Clipboard Data Queries Relationships Calculations Roles Security Sensitivity Share

Fields

Table: sales\_data\_sample (2,823 rows)

Figure 2.4: Transforming Data

BowaleICA - Power Query Editor

File Home Transform Add Column View Tools Help

New Source Recent Enter Data Data source settings Manage Parameters Refresh Preview Manage Properties Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Data Type: Whole Number Use First Row as Headers Merge Queries Append Queries Vision Combine Files Azure Machine Learning Transform Combine AI Insights

Queries [1]

sales\_data\_sample

Table.TransformColumnTypes(\*"Promoted Headers", {{"ORDERNUMBER", Int64.Type}, {"QUANTITYORDERED", type number}, {"PRICEEACH", type number}}

Query Settings

Properties Name: sales\_data\_sample All Properties

Applied Steps

Source Promoted Headers Changed Type

Figure 2.5: Selected columns to be deleted

## 2.2 Splitting Data (Generating Dimension and Fact Tables)

One of the best reasons to split information into multiple related tables is so that there is less redundant data and fewer places to update and also, it is easier to create reports by using the broken-down tables. In this section, grouping columns (by duplicating the main table and deleting other columns) that answers questions such as “What ?”, “Who ?”, “where ?”, and “when ?” together helps me to split my dataset and build a model.

### a. Creating Dimension Table (D\_Order)

In order to create this table, the first step is to duplicate the “sales\_data\_sample” table by right clicking on the table and selecting Duplicate (Figure 2.6), and then rename it as “D\_Orders”.

The screenshot shows the Power Query Editor interface with the 'sales\_data\_sample' table selected. A context menu is open, and the 'Duplicate' option is highlighted. Other options visible in the menu include 'Copy', 'Paste', 'Delete', 'Rename', 'Enable load', 'Include in report refresh', 'Reference', 'Move To Group', 'Move Up', 'Move Down', 'Create Function..', 'Convert To Parameter', 'Advanced Editor', and 'Properties...'. The main workspace displays the 'sales\_data\_sample' table with columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER, SALES, ORDERDATE, and STATUS. The 'APPLIED STEPS' pane on the right shows steps like 'Source', 'Promoted Headers', 'Changed Type', 'Removed Column', and 'Merged Columns'. The 'PROPERTIES' pane shows the table's name as 'sales\_data\_sample'. The bottom status bar indicates '23 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

Figure 2.6: Duplicating and renaming table

The next step is to choose the columns that will be used in the D\_Order table and delete the others. We can pick the required columns by holding down the ctrl key and clicking on the desired column, then right-clicking and selecting “Remove other columns”. The “ORDERNUMBER”, “ORDERLINENUMBER”, “STATUS” and “DEALSIZE” column were selected while the other columns were removed. (See Figure 2.7).

The screenshot shows the Power Query Editor interface with a query named 'sales\_data\_sample (2)'. The table contains 28 rows of data with columns: POSTALCODE, COUNTRY, TERRITORY, CONTACTLASTNAME, CONTACTFIRSTNAME, and DEALSIZE. A context menu is open over the 'DEALSIZE' column, with the 'Remove Other Columns' option highlighted.

Figure 2.7: Creating the D\_Order table with required columns

The next step is to remove Duplicates, this process will help to truncate the large number of rows we have by removing duplicate values so as to have unique rows in the table. (See figure 2.8)

The screenshot shows the Power Query Editor interface with a query named 'D\_Orders'. The table contains 28 rows of data with columns: ORDERNUMBER, STATUS, and DEALSIZE. A context menu is open over the 'ORDERNUMBER' column, with the 'Remove Other Columns' option highlighted.

Figure 2.8: Remove duplicate records

The most part of creating the D\_Order dimension table has been carried out. However, the column names are in caps and I decided to change it to lowercase and rename appropriately. The "ORDERNUMBER", "ORDERLINENUMBER", "STATUS" and "DEALSIZE" column were renamed to "OrderID", "OrderLineNumber", "Status" and "DealSize" respectively. Then Close and Apply.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Queries [2]** list: sales\_data\_sample, D\_Orders.
- D\_Orders** query preview table:
 

	OrderID	OrderLineNumber	Status	DealSize
1	10107	2	Shipped	Small
2	10121	5	Shipped	Small
3	10134	2	Shipped	Medium
4	10145	6	Shipped	Medium
5	10159	14	Shipped	Medium
6	10168	1	Shipped	Medium
7	10180	9	Shipped	Small
8	10188	2	Shipped	Medium
9	10201	2	Shipped	Small
10	10211	14	Shipped	Medium
11	10223	1	Shipped	Medium
12	10237	7	Shipped	Small
13	10251	2	Shipped	Medium
14	10263	2	Shipped	Medium
15	10275	1	Shipped	Medium
16	10285	6	Shipped	Medium
17	10299	9	Shipped	Small
18	10309	5	Shipped	Medium
19	10318	1	Shipped	Medium
20	10329	1	Shipped	Medium
21	10341	9	Shipped	Large
22	10361	13	Shipped	Small
23	10375	12	Shipped	Small
24	10388	4	Shipped	Medium
25	10403	7	Shipped	Small
26	10417	2	Disputed	Large
27	10103	11	Shipped	Medium
28	10112	1	Shipped	Large
- Query Settings** pane: Properties (Name: D\_Orders), Applied Steps (Renamed Columns).

Figure 2.9: Renaming Columns

The steps above were simply followed to create the Dimension tables for "D\_Customer", "D\_Product", and "D\_Location". Figures 2.10 to 2. Shows the processes.

b. Creating Dimension Table (D\_Customer)

The screenshot shows the Power Query Editor interface with the 'D\_Orders' table selected. A context menu is open, and 'Duplicate' is highlighted. The 'APPLIED STEPS' pane on the right shows a step named 'Merged Columns'.

Figure 2.10: Duplicating and renaming table

The screenshot shows the Power Query Editor interface with the 'D\_Customers' table selected. A context menu is open, and 'Remove' is highlighted. The 'APPLIED STEPS' pane on the right shows a step named 'Changed Type'.

Figure 2.11: Creating the D\_Customer table with required columns

Only the “CUSTOMERNAME”, “PHONE”, “ADDRESSLINE1”, “ADDRESSLINE2”, “CONTACTLASTNAME” and “CONTACTFIRSTNAME” columns were retained. Other columns were removed.

The screenshot shows the Power Query Editor interface with the following details:

- File**: BowaleICA - Power Query Editor
- Queries [4]**: sales\_data\_sample, D\_Orders, D\_Customers, sales\_data\_sample (2)
- Transform ribbon**: Home, Transform, Add Column, View, Tools, Help
- Transform pane** (selected):
  - Source: D\_Customers
  - Properties: Name: D\_Customers
  - Applied Steps: Removed Other Columns
- Table** (Preview pane):
 

	A <sub>C</sub> CUSTOMERNAME	A <sub>C</sub> PHONE	A <sub>C</sub> ADDRESSLINE1	A <sub>C</sub> ADDRESSLINE2	A <sub>C</sub> CONTACTLASTNAME	A <sub>C</sub> CONTACTFIRSTNAME
1	Land of Toys Inc.	40.67.8555	897 Long Airport Avenue		Yu	Kwai
2	Reims Collectables	26.47.1555	59 rue de l'Abbaye		Henriot	Paul
3	Lyon Souvenirs	+33 1 46 62 7555	27 rue du Colonel Pierre Avia		Da Cunha	Daniel
4	Toys4GrownUps.com	6265557265	78934 Hillside Dr.		Young	Julie
5	Corporate Gift Ideas Co.	6505551386	7734 Strong St.		Brown	Julie
6	Technics Stores Inc.	6505556809	9408 Furth Circle		Hirano	Juri
7	Daedalus Designs Imports	20.16.1555	184, chausse de Tournal		Rance	Martine
8	Herkuu Gifts	+47 2267 3215	Drammen 121, PR 744 Sentrum		Oeztan	Veysel
9	Mini Wheels Co.	6505553787	2678 Kingston Rd.	Level 3	Murphy	Julie
10	Auto Canal Petit	(1) 47.55.6555	5557 North Pendale Street	Suite 101	Perrier	Dominique
11	Australian Collectors Co.	03 9520 4555	25, rue Lauriston		Ferguson	Peter
12	Vitachrome Inc.	2125551500	636 St Kilda Road		Frick	Michael
13	Tekni Collectables Inc.	2015559350	1785 First Street		Brown	William
14	Gift Depot Inc.	2035552570	25939 South Bay Ln.		King	Julie
15	La Rochelle Gifts	40.67.8555	67, rue des Cinquante Otages		Labrune	Janine
16	Marta's Replicas Co.	6175558555	39323 Spinaker Dr.		Hernandez	Marta
17	Toys of Finland, Co.	90-224 8555	Keskuskatu 45		Karttunen	Matti
18	Baane Mini Imports	07-98 9555	Erling Skakkes gate 78		Bergulfesen	Jonas
19	Diecast Classics Inc.	2155551555	7586 Pompton St.		Yu	Kyung
20	Salzburg Collectables	6562-9555	897 Long Airport Avenue		Pipps	Georg
21	Souveniers And Things Co.	+61 2 9495 8555	Geisweg 14	Level 6	Huxley	Adrian
22	FunGiftideas.com	5085552555	Monitor Money Building, 815 Pacific Hwy		Berghuisen	Jonas
23	UK Collectable Ltd.	(171) 555-2282	1785 First Street		Benitez	Violeta
24	Euro Shopping Channel	(91) 555 94 44	Berkeley Gardens 12 Brewery		Devon	Elizabeth
25	Volvo Model Replicas, Co	0921-12 3555	C/ Moralzarzal, 86		Freyre	Diego
26	Corrida Auto Replicas, Ltd	(91) 555 22 82	C/ Aragüil, 67		Berghuisen	Christina
27	Dragon Souveniers, Ltd.	+65 221 7555	Bronz Sok., Bronz Apt. 3/6 Tesvikiye		Sommer	Martin
28					Natividad	Eric

Figure 2.12: Remove duplicate records

The screenshot shows the Power Query Editor interface with the following details:

- File**: BowaleICA - Power Query Editor
- Queries [4]**: sales\_data\_sample, D\_Orders, D\_Customers, sales\_data\_sample (2)
- Transform ribbon**: Home, Transform, Add Column, View, Tools, Help
- Transform pane** (selected):
  - Source: D\_Customers
  - Properties: Name: D\_Customers
  - Applied Steps: Renamed Columns
- Table** (Preview pane):
 

	A <sub>C</sub> CompanyName	A <sub>C</sub> Phone	A <sub>C</sub> Address	A <sub>C</sub> ADDRESSLINE2	A <sub>C</sub> CONTACTLASTNAME	A <sub>C</sub> CONTACTFIRSTNAME
1	Land of Toys Inc.	2125557818	897 Long Airport Avenue		Yu	Kwai
2	Reims Collectables	26.47.1555	59 rue de l'Abbaye		Henriot	Paul
3	Lyon Souvenirs	+33 1 46 62 7555	27 rue du Colonel Pierre Avia		Da Cunha	Daniel
4	Toys4GrownUps.com	6265557265	78934 Hillside Dr.		Young	Julie
5	Corporate Gift Ideas Co.	6505551386	7734 Strong St.		Brown	Julie
6	Technics Stores Inc.	6505556809	9408 Furth Circle		Hirano	Juri
7	Daedalus Designs Imports	20.16.1555	184, chausse de Tournal		Rance	Martine
8	Herkuu Gifts	+47 2267 3215	Drammen 121, PR 744 Sentrum		Oeztan	Veysel
9	Mini Wheels Co.	6505553787	2678 Kingston Rd.	Level 3	Murphy	Julie
10	Auto Canal Petit	(1) 47.55.6555	5557 North Pendale Street	Suite 101	Perrier	Dominique
11	Australian Collectors Co.	03 9520 4555	25, rue Lauriston		Ferguson	Peter
12	Vitachrome Inc.	2125551500	636 St Kilda Road		Frick	Michael
13	Tekni Collectables Inc.	2015559350	1785 First Street		Brown	William
14	Gift Depot Inc.	2035552570	25939 South Bay Ln.		King	Julie
15	La Rochelle Gifts	40.67.8555	67, rue des Cinquante Otages		Labrune	Janine
16	Marta's Replicas Co.	6175558555	39323 Spinaker Dr.		Hernandez	Marta
17	Toys of Finland, Co.	90-224 8555	Keskuskatu 45		Karttunen	Matti
18	Baane Mini Imports	07-98 9555	Erling Skakkes gate 78		Bergulfesen	Jonas
19	Diecast Classics Inc.	2155551555	7586 Pompton St.		Yu	Kyung
20	Salzburg Collectables	6562-9555	897 Long Airport Avenue		Pipps	Georg
21	Souveniers And Things Co.	+61 2 9495 8555	Geisweg 14	Level 6	Huxley	Adrian
22	FunGiftideas.com	5085552555	Monitor Money Building, 815 Pacific Hwy		Benitez	Violeta
23	UK Collectable Ltd.	(171) 555-2282	1785 First Street		Devon	Elizabeth
24	Euro Shopping Channel	(91) 555 94 44	Berkeley Gardens 12 Brewery		Freyre	Diego
25	Volvo Model Replicas, Co	0921-12 3555	C/ Moralzarzal, 86		Berghuisen	Christina
26	Corrida Auto Replicas, Ltd	(91) 555 22 82	C/ Aragüil, 67		Sommer	Martin
27	Dragon Souveniers, Ltd.	+65 221 7555	Bronz Sok., Bronz Apt. 3/6 Tesvikiye		Natividad	Eric
28						

Figure 2.13: Renaming Columns

The “ADDRESSLINE2” column was removed as it has a lot of blank values and may affect analysis.

The screenshot shows the Power Query Editor interface with the 'D\_Customers' query selected. In the main grid, the 'ADDRESSLINE2' column is highlighted with a yellow border. A context menu is open over this column, with the 'Remove' option being selected. The 'APPLIED STEPS' pane on the right shows the step 'Renamed Columns'.

Figure 2.14: Remove unwanted column

The “CONTACTLASTNAME” and “CONTACTFIRSTNAME” Columns were renamed to “LastName” and “FirstName”, reordered, merged and then Renamed to “FullName”.

The screenshot shows the Power Query Editor interface with the 'D\_Customers' query selected. A 'Merge Columns' dialog box is open over the 'CONTACTLASTNAME' and 'CONTACTFIRSTNAME' columns. The 'Separator' dropdown is set to 'Space'. A new column name 'FullName' is entered in the 'New column name (optional)' field. The 'OK' button is highlighted. The 'APPLIED STEPS' pane on the right shows the steps 'Renamed Columns', 'Renamed Columns1', and 'Reordered Columns'.

Figure 2.15: Rename Columns, Merge Columns and Rename merged.

### c. Creating Dimension Table (D\_Product)

The screenshot shows the Power Query Editor interface with the 'sales\_data\_sample' query selected. A context menu is open over the 'PRODUCTLINE' column, listing options such as Copy, Paste, Delete, Rename, Enable load, Duplicate, Reference, Move To Group, Move Up, Move Down, Create Function..., Advanced Editor, and Properties... The main pane displays a table with columns: ORDERNUMBER, QUANTITYORDERED, PRICEACH, ORDERLINENUMBER, SALES, ORDERDATE, and STATUS. The 'APPLIED STEPS' pane on the right shows the 'Promoted Headers' step.

Figure 2.16: Duplicating and renaming table

The “PRODUCTLINE” and the “PRODUCTCODE” columns were retained. The remaining columns were removed.

The screenshot shows the Power Query Editor interface with the 'D\_Product' query selected. A context menu is open over the 'PRODUCTCODE' column, listing options such as Copy, Remove Columns, Remove Other Columns, Add Column From Examples..., Remove Duplicates, Remove Errors, Replace Values..., Fill, Change Type, Transform, Merge Columns, Group By..., Unpivot Columns, Unpivot Other Columns, Unpivot Only Selected Columns, and Move. The main pane displays a table with columns: MONTH\_ID, YEAR\_ID, PRODUCTNAME, MSRP, PRODUCTCODE, CUSTOMERNAME, and PHONE. The 'APPLIED STEPS' pane on the right shows the 'Promoted Headers' step.

Figure 2.17: Creating the D\_Product table with required columns

Power Query Editor - BowaleICA - D\_Product

**APPLIED STEPS**

- Source
- Promoted Headers
- Changed Type
- Removed Other Columns
- Reordered Columns

Figure 2.18: Reorder and remove duplicate records

Power Query Editor - BowaleICA - D\_Product

**APPLIED STEPS**

- Source
- Promoted Headers
- Changed Type
- Removed Other Columns
- Reordered Columns
- Renamed Columns

Figure 2.19: Renaming Columns

d. Creating Dimension Table (D\_Location)

The sales\_data\_sample table was duplicated and renamed to “D\_Location”. The “CITY”, “STATE”, “POSTALCODE”, “COUNTRY” and “TERRITORY” columns were retained and others were removed.

Detailed description: This screenshot shows the Power Query Editor interface. In the center, there is a table with columns labeled ADDRESSLINE2, CITY, STATE, POSTALCODE, COUNTRY, and TERRITORY. The rows contain various location data. On the right side, a context menu is open over the 'POSTALCODE' column. The menu includes options such as 'Remove Columns', 'Remove Other Columns', 'Remove Duplicates', 'Replace Values...', 'Change Type', 'Transform', 'Merge Columns', 'Group By...', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', and 'Move'. The 'Remove Duplicates' option is highlighted. The status bar at the bottom indicates '25 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

Figure 2.20: Creating the D\_Location table with required columns

Detailed description: This screenshot shows the Power Query Editor interface. The table structure has been modified to include only four columns: POSTALCODE, STATE, COUNTRY, and TERRITORY. The rows show a subset of the previous data. A context menu is open over the 'POSTALCODE' column, with the 'Remove Duplicates' option selected. The status bar at the bottom indicates '5 COLUMNS, 74 ROWS' and 'Column profiling based on top 1000 rows'.

Figure 2.21: Reorder and remove duplicate records

After removing duplicates, there was a blank space in the “POSTALCODE”. We will be replacing the Blanks in “POSTALCODE” column with “1”.

The screenshot shows the Power Query Editor interface. A 'Replace Values' dialog box is open, centered over a table of location data. The table has columns: POSTALCODE, CITY, STATE, COUNTRY, and TERRITORY. The 'Replace Values' dialog box has 'Value To Find' set to a blank field and 'Replace With' set to '1'. The 'OK' button is highlighted. The 'APPLIED STEPS' pane on the right shows the step 'Removed Duplicates'.

Figure 2.22: Replace blank fields

The screenshot shows the Power Query Editor interface. A 'Rename Columns' dialog box is open, centered over a table of location data. The table has columns: POSTALCODE, CITY, STATE, COUNTRY, and TERRITORY. The 'Rename Columns' dialog box shows the new column names: PostalCode, City, State, Country, and Territory. The 'OK' button is highlighted. The 'APPLIED STEPS' pane on the right shows the steps 'Renamed Columns' and 'Replaced Value'.

Figure 2.23: Renaming Columns

e. *Creating fact Table (Transaction)*

The sales\_data\_sample table was duplicated and renamed to "Transaction". The "ORDERNUMBER", "QUANTITYORDERED", "PRICEEACH", "SALES", "ORDERDATE", "MSRP", "PRODUCTCODE", "CUSTOMERNAME", "POSTALCODE" columns were retained and others were removed.

Figure 2.24: Creating the Transaction table with required columns

The Columns were renamed from "ORDERNUMBER", "QUANTITYORDERED", "PRICEEACH", "SALES", "ORDERDATE", "MSRP", "PRODUCTCODE", "CUSTOMERNAME", "POSTALCODE" to "OrderID", "Quantity", "Unit Price", "Amount", "OrderDate", "Manufacturer Suggested Price", "ProductID", "Company Name", "PostalCode".

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. In the 'Applied Steps' pane, the 'Renamed Columns' step is highlighted. The main area displays a table with columns: OrderID, Quantity, Unit Price, Amount, OrderDate, and Manufacturer Suggested Price. The 'Properties' pane shows the query name is 'Transaction'. The preview pane at the bottom right shows the first 24 rows of the data.

Figure 2.25: Renaming Columns

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. In the 'Applied Steps' pane, the 'Changed Type' step is highlighted. The main area displays a table with columns: Company Name, PostalCode, Quantity, Unit Price, Amount, and Manufacturer Suggested Price. The 'Properties' pane shows the query name is 'Transacción'. The preview pane at the bottom right shows the first 24 rows of the data.

Figure 2.26: Reordering columns

### f. Creating dimension Table (D\_Continent)

I noticed that the countries in the dataset span across the different continents, therefore, there is a need to create a Continent table and add it to our model. This might be useful for the analysis. This was also helpful and useful to create a customized unique color per Continent. The “D\_Location” table was duplicated; The Country column was retained while other columns were removed.

The screenshot shows the Power Query Editor interface with the following details:

- File Bar:** Home, Transform, Add Column, View, Tools, Help.
- Queries List:** sales\_data\_sample, D\_Orders, D\_Customers, D\_Product, D\_Location, Transaction, D\_Location (2).
- Table View:** A table with columns: PostalCode, City, State, Country, Territory. The 'Country' column is highlighted.
- Context Menu (open over 'Country'):**
  - Copy
  - Remove Columns
  - Remove Other Columns** (selected)
  - Add Column From Examples...
  - Remove Duplicates
  - Remove Errors
  - Replace Values...
  - Fill
  - Change Type
  - Transform
  - Merge Columns
  - Group By...
  - Unpivot Columns
  - Unpivot Other Columns
  - Unpivot Only Selected Columns
  - Move
- Query Settings:** Properties (Name: D\_Location (2)), All Properties.
- Applied Steps:** Removed Other Columns.
- Bottom Status:** PREVIEW DOWNLOADED AT 5:39 PM.

Figure 2.27: Creating the D\_Location table with required columns

The screenshot shows the Power Query Editor interface with the following details:

- File Bar:** Home, Transform, Add Column, View, Tools, Help.
- Queries List:** sales\_data\_sample, D\_Orders, D\_Customers, D\_Product, D\_Location, Transaction, D\_Location (2).
- Table View:** A table with columns: Country. The 'Country' column is highlighted.
- Context Menu (open over 'Country'):**
  - Copy
  - Remove
  - Remove Other Columns
  - Duplicate Column
  - Add Column From Examples...
  - Remove Duplicates** (selected)
  - Remove Errors
  - Change Type
  - Transform
  - Replace Values...
  - Replace Errors...
  - Split Column
  - Group By...
  - Fill
  - Unpivot Columns
  - Unpivot Only Selected Columns
  - Rename...
  - Move
  - Drill Down
  - Add as New Query
- Query Settings:** Properties (Name: D\_Location), All Properties.
- Applied Steps:** Removed Duplicates.
- Bottom Status:** PREVIEW DOWNLOADED AT 5:41 PM.

Figure 2.28: Remove duplicate records

Now that we have unique values in the Country column, the Continent table can now be created. By copying the unique Country column, clicking on the “Enter data” button on the Home tab in PowerBI, Paste the copied data and manually populate the Continent column.

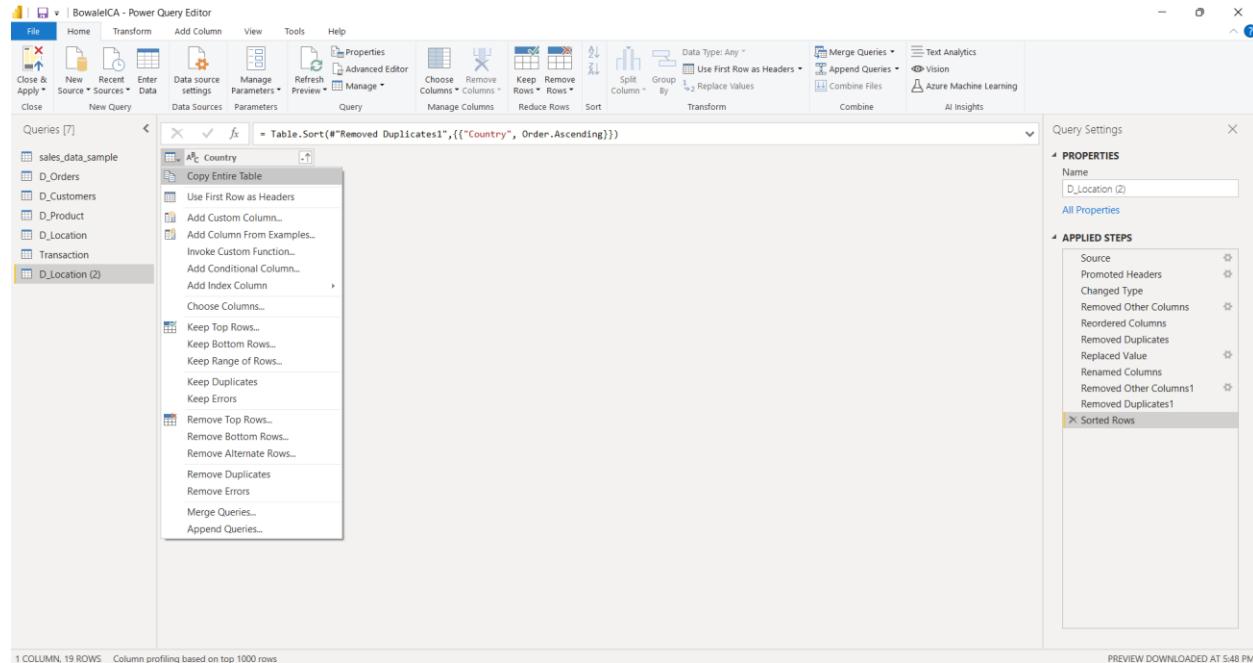


Figure 2.29: Sort Column and Copy table

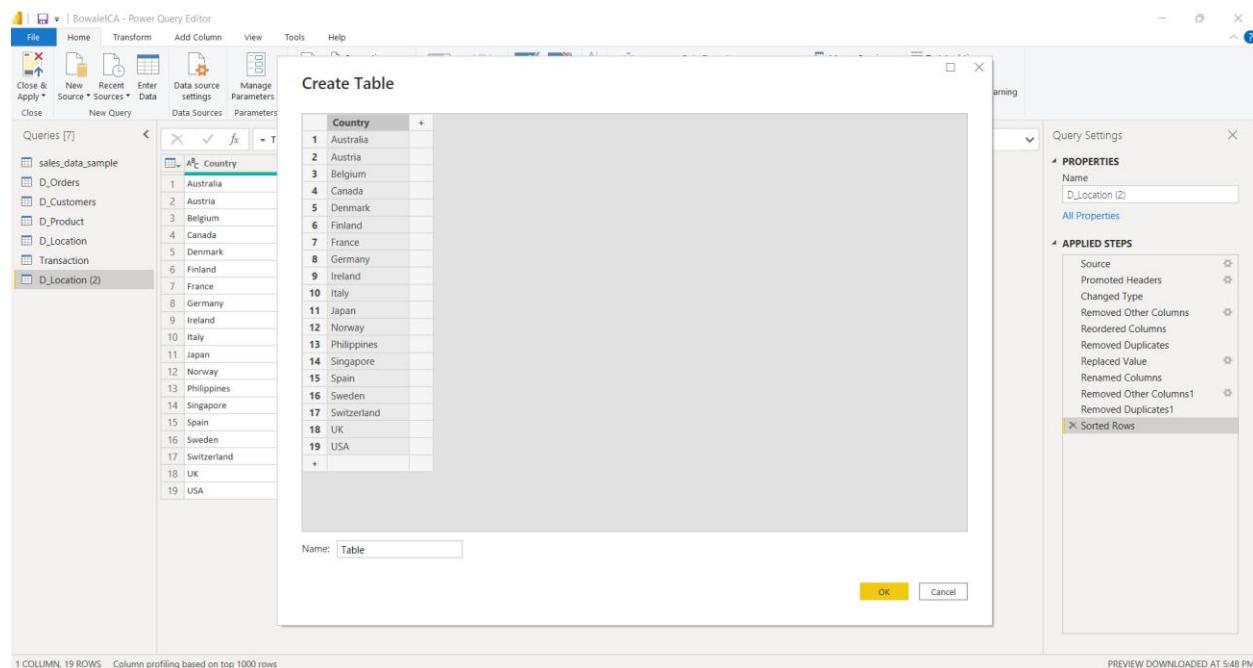


Figure 2.30: Create table and paste data

BowaleICA - Power Query Editor

**Create Table**

Country	Continent
1 Australia	Australia
2 Austria	Europe
3 Belgium	Europe
4 Canada	America
5 Denmark	Europe
6 Finland	Europe
7 France	Europe
8 Germany	Europe
9 Ireland	Europe
10 Italy	Europe
11 Japan	Asia
12 Norway	Europe
13 Philippines	Asia
14 Singapore	Asia
15 Spain	Europe
16 Sweden	Europe
17 Switzerland	Europe
18 UK	Europe
19 USA	America

Name: D\_Location

**Query Settings**

- Properties
  - Name: D\_Location (2)
- All Properties

**APPLIED STEPS**

- Source
- Promoted Headers
- Changed Type
- Removed Other Columns
- Reordered Columns
- Removed Duplicates
- Replaced Value
- Renamed Columns
- Removed Other Columns1
- Removed Duplicates1
- Sorted Rows

Figure 2.31: Rename Table

BowaleICA - Power Query Editor

**Delete Query**

Are you sure you want to delete 'D\_Location (2)'?

**Query Settings**

- Properties
  - Name: D\_Location (2)
- All Properties

**APPLIED STEPS**

- Source
- Promoted Headers
- Changed Type
- Removed Other Columns
- Reordered Columns
- Removed Duplicates
- Replaced Value
- Renamed Columns
- Removed Other Columns1
- Removed Duplicates1
- Sorted Rows

Figure 2.32: Delete Unused table

The screenshot shows the Power Query Editor interface with the following details:

- File**: Home, Transform, Add Column, View, Tools, Help.
- Data Sources**: Close & Apply, New Source, Recent Sources, Enter Data.
- Parameters**: Manage Parameters.
- Properties**: Refresh Preview, Advanced Editor.
- Query**: Manage Columns, Choose Columns, Remove Rows, Keep Rows, Remove Columns, Group By, Replace Values.
- Transform**: Data Type: Whole Number, Use First Row as Headers, Merge Queries, Append Queries, Text Analytics, Split Column, Combine Files, Azure Machine Learning, AI Insights.
- Queries [7]**: sales\_data\_sample, D\_Orders, D\_Customers, D\_Product, D\_Location, Transaction, D\_Continent.
- Table View**: Shows a table with columns: ORDERNUMBER, QUANTITYORDERED, PRICEACH, ORDERLINENUMBER, SALES, ORDERDATE, and STATUS.
- Delete Query**: A modal dialog box is displayed, asking "Are you sure you want to delete 'sales\_data\_sample'?". It contains "Delete" and "Cancel" buttons.
- Properties Panel**: Shows the name as "sales\_data\_sample" and the applied step "Promoted Headers" with "Changed Type".
- Query Settings**: Shows preview settings and download history.

Figure 2.33: Delete sales\_data\_sample table

#### g. Creating date Table (D\_Calendar)

Populating the date table (DAX): Here the Date, Day, Day\_Num, Month, Month\_Num, Quater, Year , weekend, and weekday will be populated in order to analyze the data with the above Metrics

Click on New Table and Enter

The screenshot shows the Power BI Desktop interface with the following details:

- File**: Home, Help.
- Table tools**: Name Table, Mark as date table, Calendars, Manage relationships, Relationships, New, Quick measure, New measure column, New table.
- Structure**: Write a DAX expression to create a new table. The expression is: `i D_Calendar = CALENDAR(AUTO)`.
- Fields**: Search bar, list of available tables: D\_Continent, D\_Customers, D\_Location, D\_Orders, D\_Product, Table (selected), Transaction.
- Table Information**: Shows "Table: Table (1 rows)".

Figure 2.34: Create new table

Click on the New Table and add the expression “D\_Calendar = CALENDARAUTO()”

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A new table named 'D\_Calendar' is being created with the expression 'D\_Calendar = CALENDARAUTO()' in the formula bar. The Fields pane on the right lists various tables and columns, including 'Table' which is currently selected.

Figure 2.34: Write DAX Expression for the table

Add a new column with the expression “Day = FORMAT(D\_Calendar[Date], "ddd")” to Create the Day Column.

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. A new column named 'Day' is being added with the expression 'Day = FORMAT(D\_Calendar[Date], "ddd")'. The Fields pane on the right shows the newly created 'Day' column under the 'D\_Calendar' table.

Figure 2.34: Add new "Day" column

Add a new column with the expression “Day\_Num = FORMAT(D\_Calendar[Date], “d”)” to Create the Day\_Num Column.

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A new column named 'Day\_Num' is being created, defined by the DAX expression: `1 Day_Num = FORMAT(D_Calendar[Date], "d")`. The 'Data type' is set to 'Text'. The Fields pane on the right shows the hierarchy of tables and columns, with 'Day\_Num' highlighted under the 'D\_Calendar' table. The main table view displays the first 27 rows of the calendar data, showing dates from Wednesday, January 1, 2003, to Monday, January 27, 2003, and their corresponding day numbers (1 through 27).

Figure 2.35: Add new "Day\_Num" column

Add a new column with the expression “Month = FORMAT(D\_Calendar[Date], “Mmm”)” to Create the Month Column.

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A new column named 'Month' is being created, defined by the DAX expression: `1 Month = FORMAT(D_Calendar[Date], "Mmm")`. The 'Data type' is set to 'Text'. The Fields pane on the right shows the hierarchy of tables and columns, with 'Month' highlighted under the 'D\_Calendar' table. The main table view displays the first 27 rows of the calendar data, showing dates from Wednesday, January 1, 2003, to Monday, January 27, 2003, and their corresponding month abbreviations (Jan through Jan).

Figure 2.36: Add new "Month" column

Add a new column with the expression “Month\_Num = FORMAT(D\_Calendar[Date], “m”)” to Create the Month\_Num Column.

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A new column named 'Month\_Num' is being created, with the DAX formula `FORMAT(D_Calendar[Date], "m")` entered in the formula bar. The Fields pane on the right shows the newly created 'Month\_Num' column under the 'D\_Calendar' table.

Date	Day	Day_Num	Month	Month_Num
Wednesday, January 1, 2003	Wed	1	Jan	1
Thursday, January 2, 2003	Thu	2	Jan	1
Friday, January 3, 2003	Fri	3	Jan	1
Saturday, January 4, 2003	Sat	4	Jan	1
Sunday, January 5, 2003	Sun	5	Jan	1
Monday, January 6, 2003	Mon	6	Jan	1
Tuesday, January 7, 2003	Tue	7	Jan	1
Wednesday, January 8, 2003	Wed	8	Jan	1
Thursday, January 9, 2003	Thu	9	Jan	1
Friday, January 10, 2003	Fri	10	Jan	1
Saturday, January 11, 2003	Sat	11	Jan	1
Sunday, January 12, 2003	Sun	12	Jan	1
Monday, January 13, 2003	Mon	13	Jan	1
Tuesday, January 14, 2003	Tue	14	Jan	1
Wednesday, January 15, 2003	Wed	15	Jan	1
Thursday, January 16, 2003	Thu	16	Jan	1
Friday, January 17, 2003	Fri	17	Jan	1
Saturday, January 18, 2003	Sat	18	Jan	1
Sunday, January 19, 2003	Sun	19	Jan	1
Monday, January 20, 2003	Mon	20	Jan	1
Tuesday, January 21, 2003	Tue	21	Jan	1
Wednesday, January 22, 2003	Wed	22	Jan	1
Thursday, January 23, 2003	Thu	23	Jan	1
Friday, January 24, 2003	Fri	24	Jan	1
Saturday, January 25, 2003	Sat	25	Jan	1
Sunday, January 26, 2003	Sun	26	Jan	1
Monday, January 27, 2003	Mon	27	Jan	1

Figure 2.37: Add new "Month\_Num" column

Add a new column with the expression “Quarter = FORMAT(D\_Calendar[Date], “\Qq”)” to Create the Quater Column.

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A new column named 'Quarter' is being created, with the DAX formula `FORMAT(D_Calendar[Date], "\Qq")` entered in the formula bar. The Fields pane on the right shows the newly created 'Quarter' column under the 'D\_Calendar' table.

Date	Day	Day_Num	Month	Month_Num	Quarter
Wednesday, January 1, 2003	Wed	1	Jan	1	Q1
Thursday, January 2, 2003	Thu	2	Jan	1	Q1
Friday, January 3, 2003	Fri	3	Jan	1	Q1
Saturday, January 4, 2003	Sat	4	Jan	1	Q1
Sunday, January 5, 2003	Sun	5	Jan	1	Q1
Monday, January 6, 2003	Mon	6	Jan	1	Q1
Tuesday, January 7, 2003	Tue	7	Jan	1	Q1
Wednesday, January 8, 2003	Wed	8	Jan	1	Q1
Thursday, January 9, 2003	Thu	9	Jan	1	Q1
Friday, January 10, 2003	Fri	10	Jan	1	Q1
Saturday, January 11, 2003	Sat	11	Jan	1	Q1
Sunday, January 12, 2003	Sun	12	Jan	1	Q1
Monday, January 13, 2003	Mon	13	Jan	1	Q1
Tuesday, January 14, 2003	Tue	14	Jan	1	Q1
Wednesday, January 15, 2003	Wed	15	Jan	1	Q1
Thursday, January 16, 2003	Thu	16	Jan	1	Q1
Friday, January 17, 2003	Fri	17	Jan	1	Q1
Saturday, January 18, 2003	Sat	18	Jan	1	Q1
Sunday, January 19, 2003	Sun	19	Jan	1	Q1
Monday, January 20, 2003	Mon	20	Jan	1	Q1
Tuesday, January 21, 2003	Tue	21	Jan	1	Q1
Wednesday, January 22, 2003	Wed	22	Jan	1	Q1
Thursday, January 23, 2003	Thu	23	Jan	1	Q1
Friday, January 24, 2003	Fri	24	Jan	1	Q1
Saturday, January 25, 2003	Sat	25	Jan	1	Q1
Sunday, January 26, 2003	Sun	26	Jan	1	Q1
Monday, January 27, 2003	Mon	27	Jan	1	Q1

Figure 2.38: Add new "Quater" column

Add a new column with the expression “Week Cat = `switch(WEEKDAY(D_Calendar[Date],2),`  
`6, "Weekend",`  
`7, "Weekend",`  
`"Weekday")`” to Create the Week Cat Column.

Table: D\_Calendar (1,096 rows) Column: Week Cat (2 distinct values)

Figure 2.39: Add new "weekend and weekday" column

Add a new column with the expression “Year = `FORMAT(D_Calendar[Date], "yyyy")`

” To Create the Year Column. Date Table Created.

Table: D\_Calendar (1,096 rows) Column: Year (3 distinct values)

Figure 2.40: Add new "Year" column

### 3. BI Data Modelling

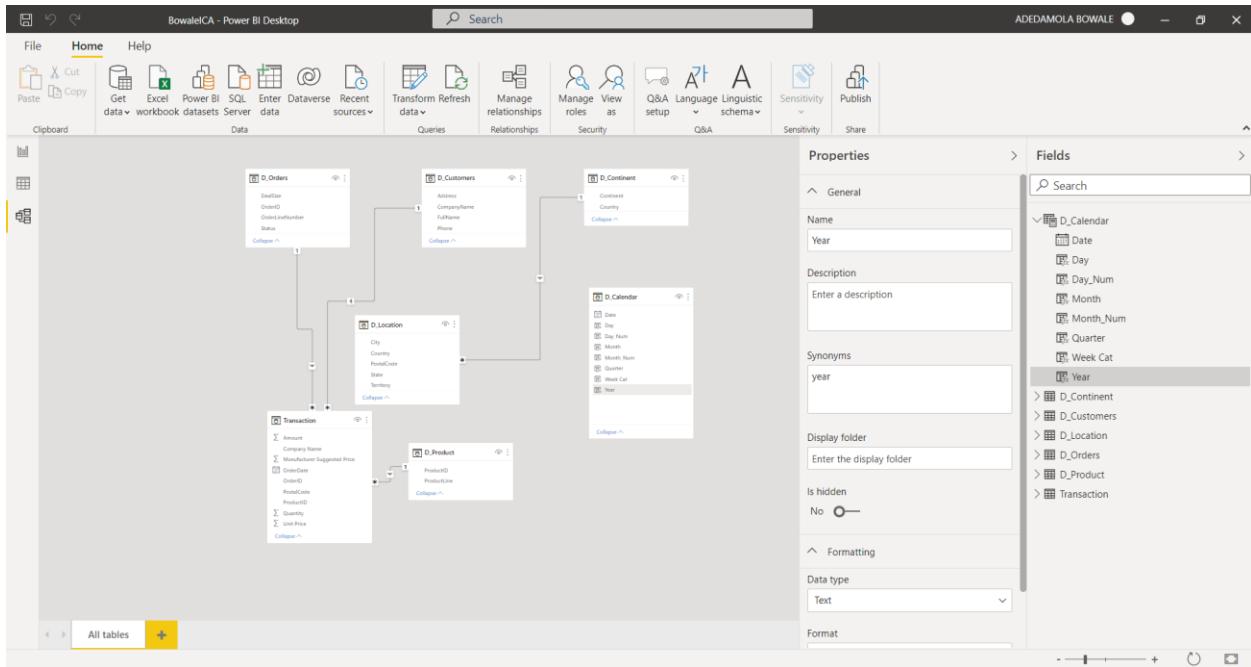


Figure 3.1: Automatic Modelling by PowerBI

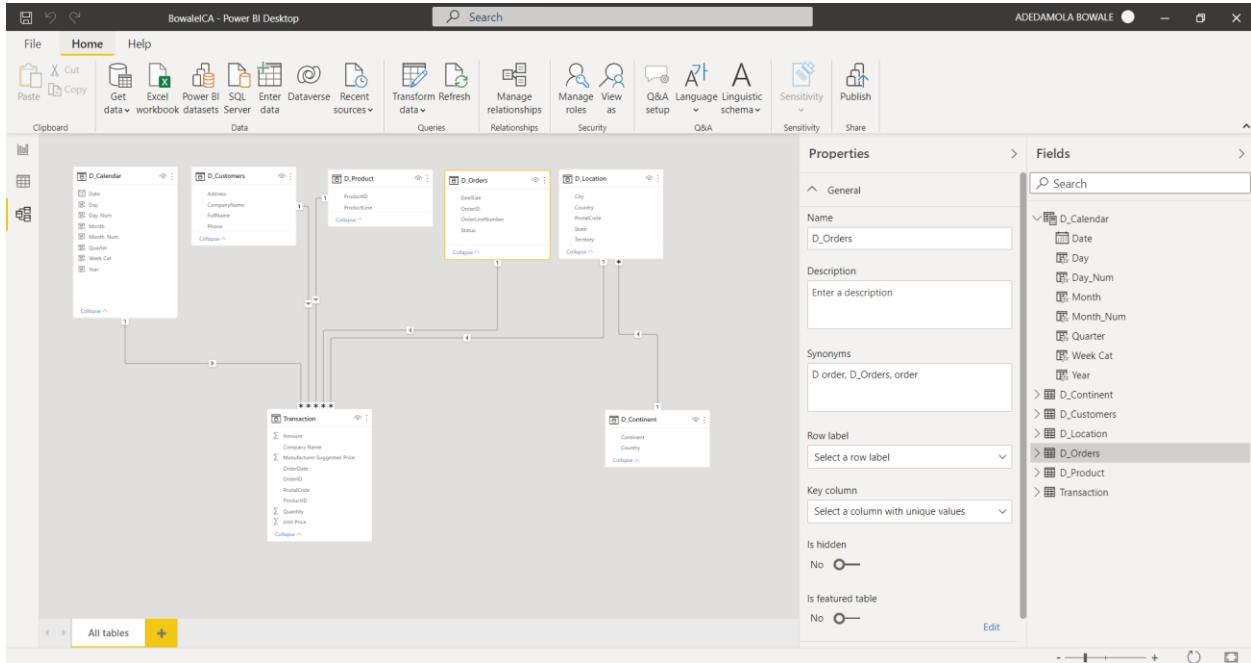


Figure 3.1: Fact and dimension tables with established relationships



Inspiring success

# BIG DATA AND BUSINESS INTELLIGENCE IN-COURSE ASSESSMENT

BOWALE ADEDAMOLA  
APPLIED DATA SCIENCE(MSc.)  
Ao353496

January 12, 2022

Section 2: Business Report

## SECTION 2 – BUSINESS INTELLIGENCE SOLUTION

### 1. Executive Summary

#### Abstract

The purpose of this report is to analyze an open-source data with Sales Data, Order Information, Customer information, Shipping information, location information, etc., that was provided on Kaggle for a vehicle retail company that sells Classic cars, Vintage cars, Airplanes, ships, trains, Buses and trucks.

To build the appropriate data model, the dataset was evaluated and preprocessed. The transformed data model is then used to generate insights and also answers to the aforementioned business questions.

#### Key Findings

- More sales were made in 2004 compared to 2003 and 2005
- More sales were made during weekdays than weekends
- Classic cars are the best-selling products while Trains are the least
- More sales were made in November compared to other months of the year
- More sales were made to Spain compared to other countries.

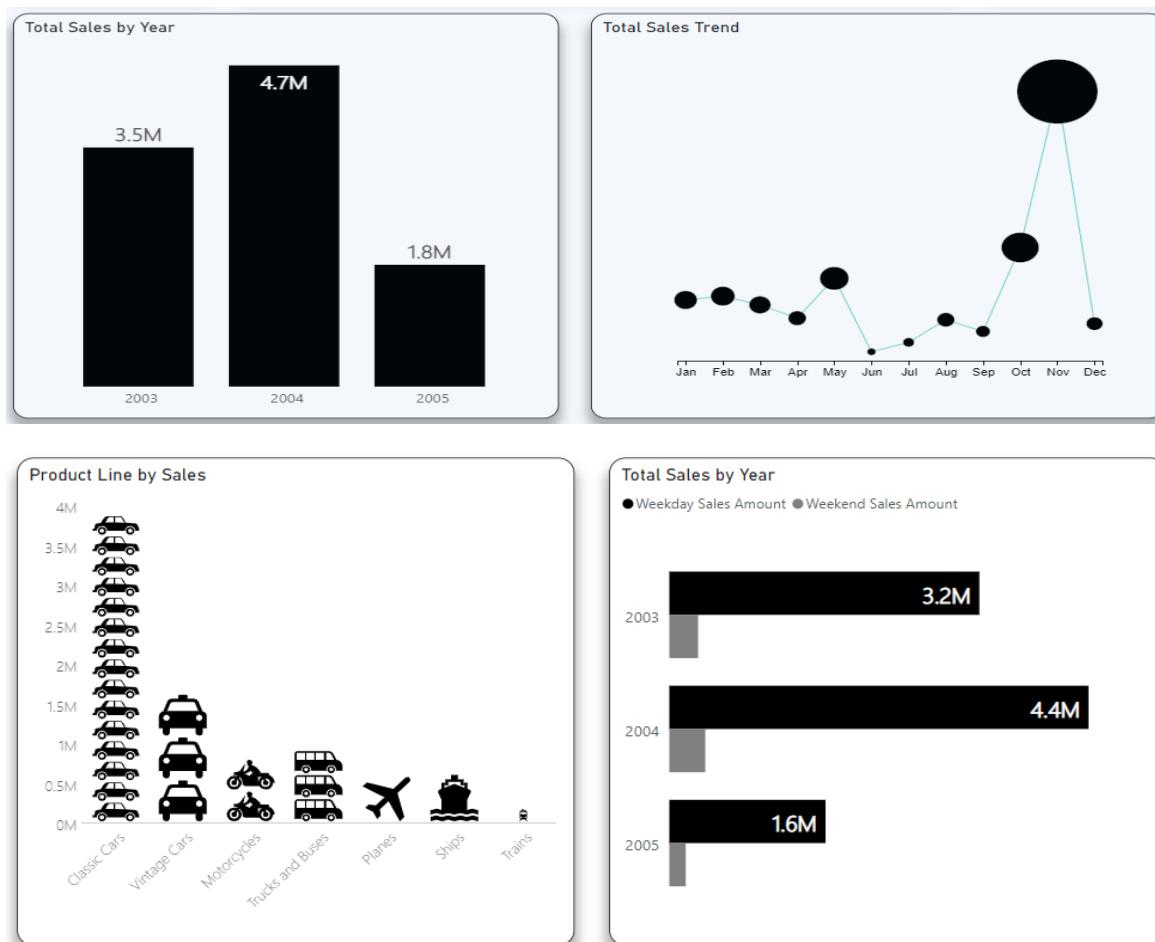


Figure 3. 1: Some findings

## Recommendations

- Investigate the cause of sales drop between 2004 and 2005
- Enhance sales channels and Marketing activities to increase sales
- Improve location relations
- Carry out regular market comparisons and pricing analysis
- Carry out research on unique features that could be added to low selling products to give it a boost

## 2. Introduction

Sales and customer relationships are the primary end goals of a retail company. Also, the ability and necessity to transport goods and people over long or short distances at high speeds in comfort and safety has been an indicator of civilization and technological progress. A vehicle retail company that sells Classic cars, Vintage cars, Airplanes, ships, trains, Buses and trucks aims to sell its products all across the world. This report analyzes the sales Data, Order Information, Customer information, Shipping information, location information for 2003, 2004 and 2005 to evaluate the progress status, generate insights and answer the questions below:

- Customers with highest and lowest sales record
- Year with highest and lowest sales record
- Month with highest and lowest sales record (for weekend and weekdays also)
- Sales record based on the status of the order (Shipped order, Cancelled order, etc.)
- Sales record by unique products
- Sales record by Location

## Data Model

The flat file (“Sales\_data\_sample”) was normalized into several dimension tables and a fact table. The Power BI tool was used to create the data model and the relationships between the tables. All the calculated measures were also arranged under a table called “Measures\_table”, however, no relationship was established as it wasn’t needed. Please, see final model below.

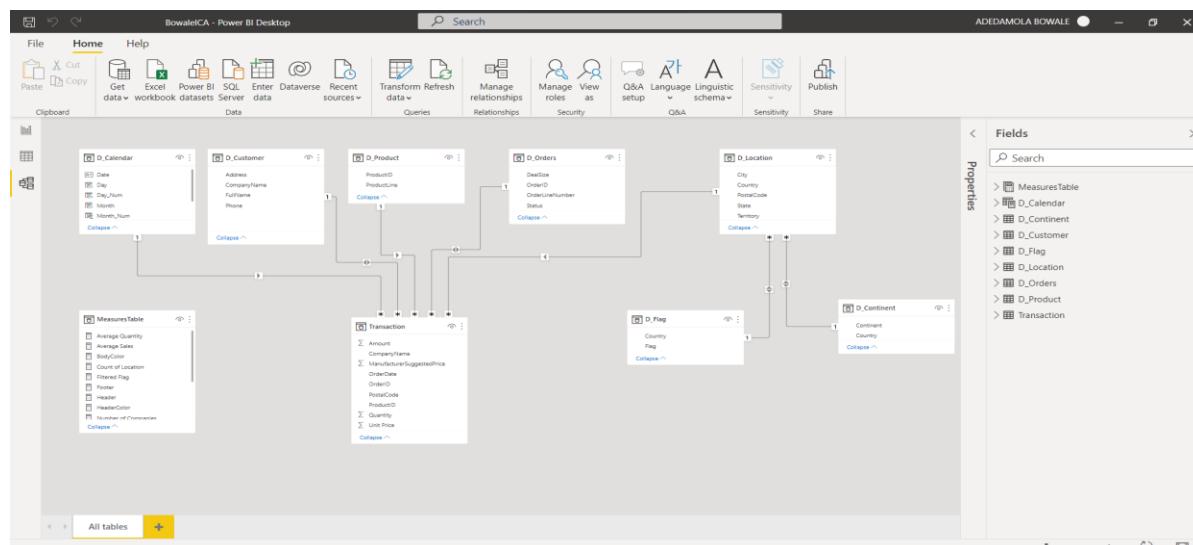


Figure 3. 2: Final Model

### 3. Findings based on analysis and evaluation:

The process of creating a dashboard opens us to the possibility of more analysis. During the dashboard preparation, there were needs for creating another table , some calculated measures and also write some M-queries. This is all to make the dashboard better.

#### Table Creation (For Analysis)

With the presence of a list of Countries in the dataset, there might be a need to add country flags attached to each Country. Therefore, Another Table (D\_Flag) will be populated to accommodate the above idea. The Country flags were populated from a webpage <https://flagpedia.net/>. The D\_flag table could also be created on excel and imported into PowerBI.

Country	Flag
1 Australia	https://flagpedia.net/data/flags/h80/au.png
2 Austria	https://flagpedia.net/data/flags/h80/at.png
3 Belgium	https://flagpedia.net/data/flags/h80/be.png
4 Canada	https://flagpedia.net/data/flags/h80/ca.png
5 Denmark	https://flagpedia.net/data/flags/h80/dk.png
6 Finland	https://flagpedia.net/data/flags/h80/fi.png
7 France	https://flagpedia.net/data/flags/h80/fr.png
8 Germany	https://flagpedia.net/data/flags/h80/de.png
9 Ireland	https://flagpedia.net/data/flags/h80/ie.png
10 Italy	https://flagpedia.net/data/flags/h80/it.png
11 Japan	https://flagpedia.net/data/flags/h80/jp.png
12 Norway	https://flagpedia.net/data/flags/h80/no.png
13 Philippines	https://flagpedia.net/data/flags/h80/ph.png
14 Singapore	https://flagpedia.net/data/flags/h80/sg.png
15 Spain	https://flagpedia.net/data/flags/h80/se.png
16 Sweden	https://flagpedia.net/data/flags/h80/se.png
17 Switzerland	https://flagpedia.net/data/flags/h80/ch.png
18 UK	https://flagpedia.net/data/flags/h80/gb.png
19 USA	https://flagpedia.net/data/flags/h80/us.png

Figure 3. 3: Create D\_flag table for further analysis

Power BI created an automatic one-one relationship between D\_Flag and D\_Continent and this needs to be rerouted.

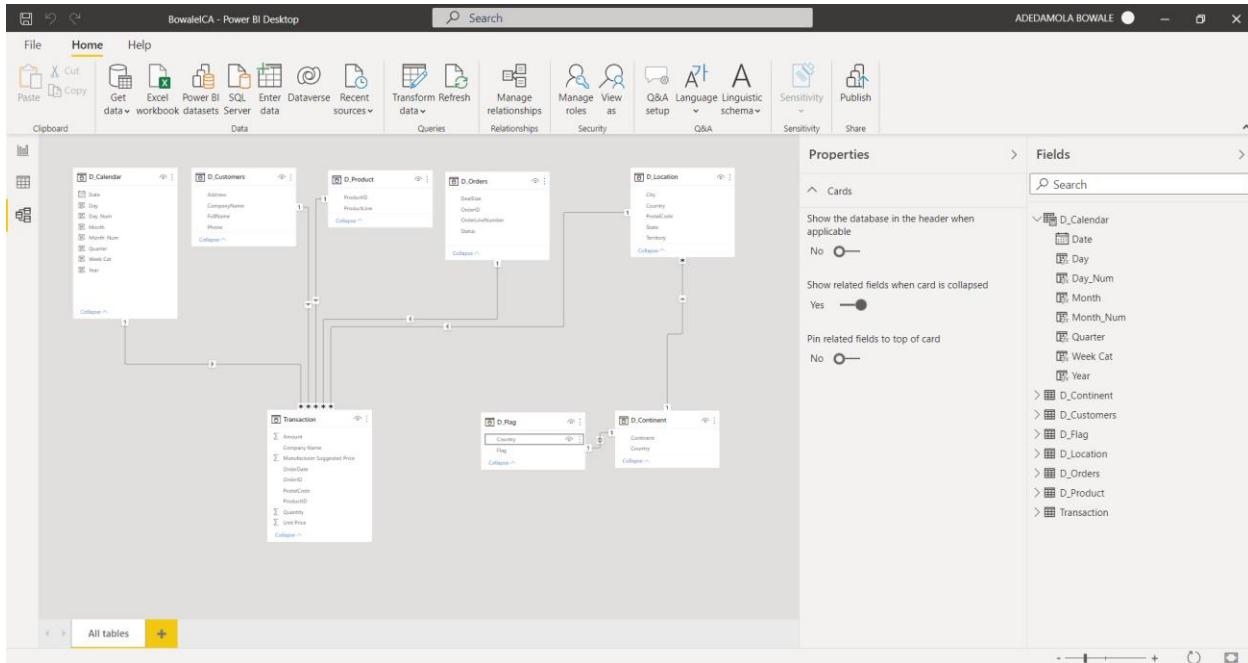


Figure 3. 4: Automatic generated model

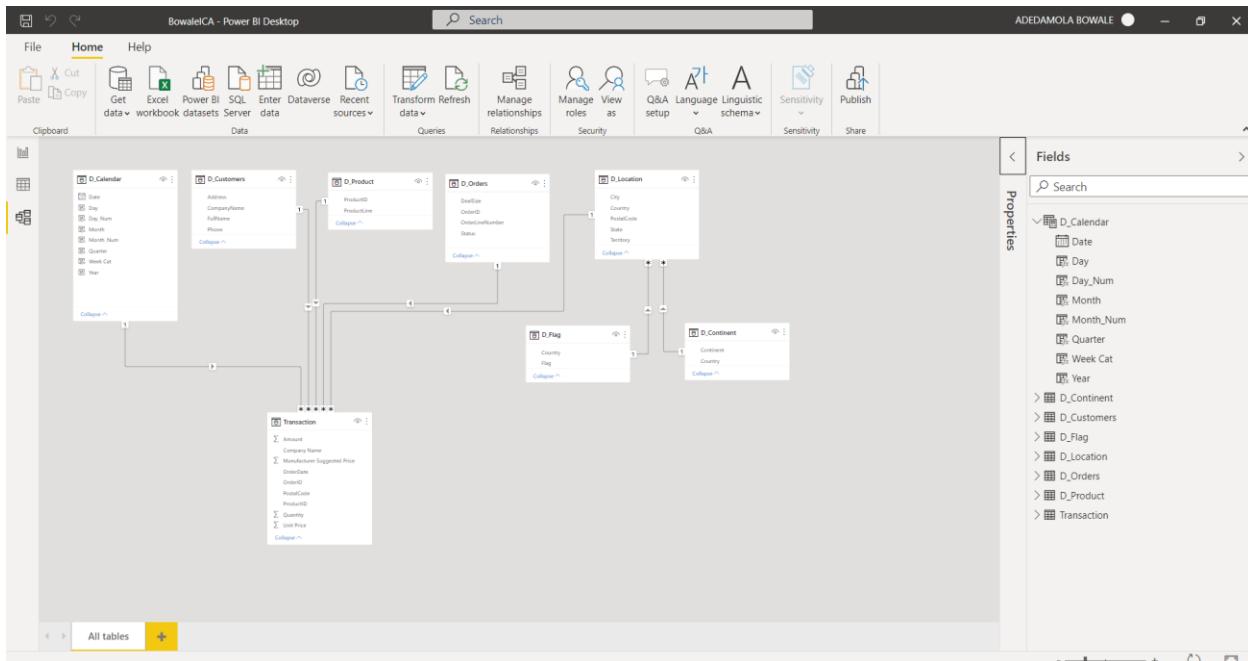


Figure 3. 5: Model update (D\_Flag)

## Data analysis steps

In order to make the Model Tidy and arranged, all the calculated measures will be created inside a single table called “MeasuresTable”. In this section, all the required calculated measures will be created.

Click Enter Data to create the “MeasuresTable”, Click OK, Close and Apply

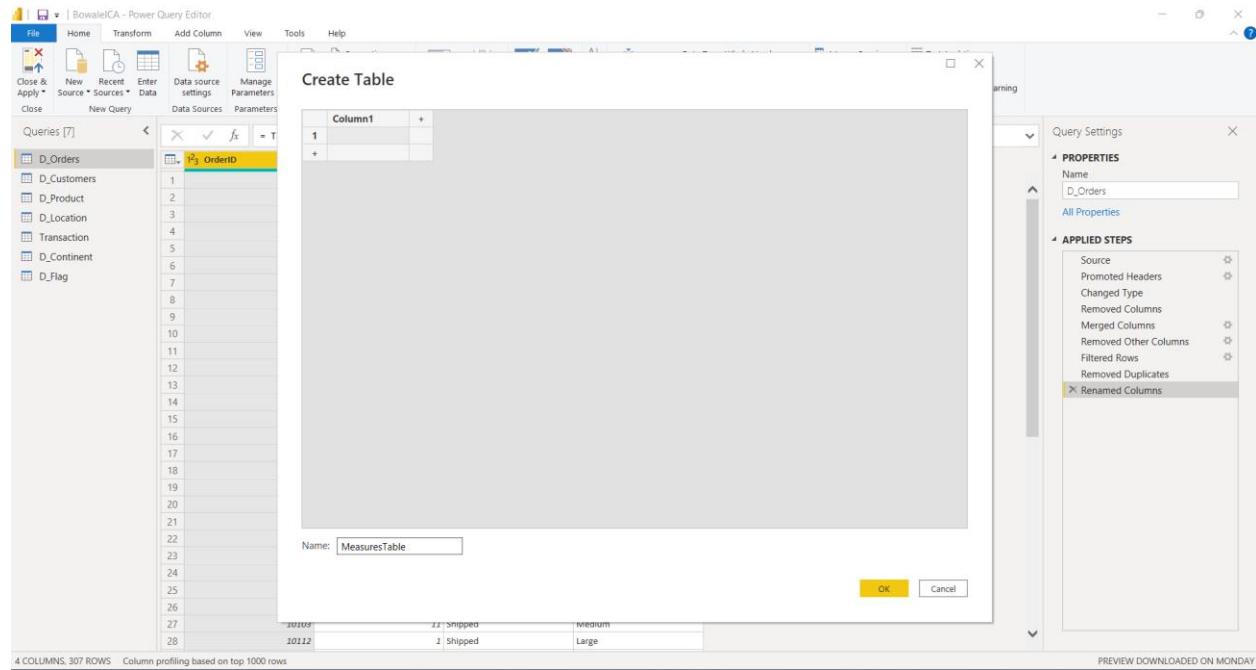


Figure 3. 6: Create Measurestable

Calculating the “TotalSales” measure. Total Sales = `SUM('Transaction'[Amount])`

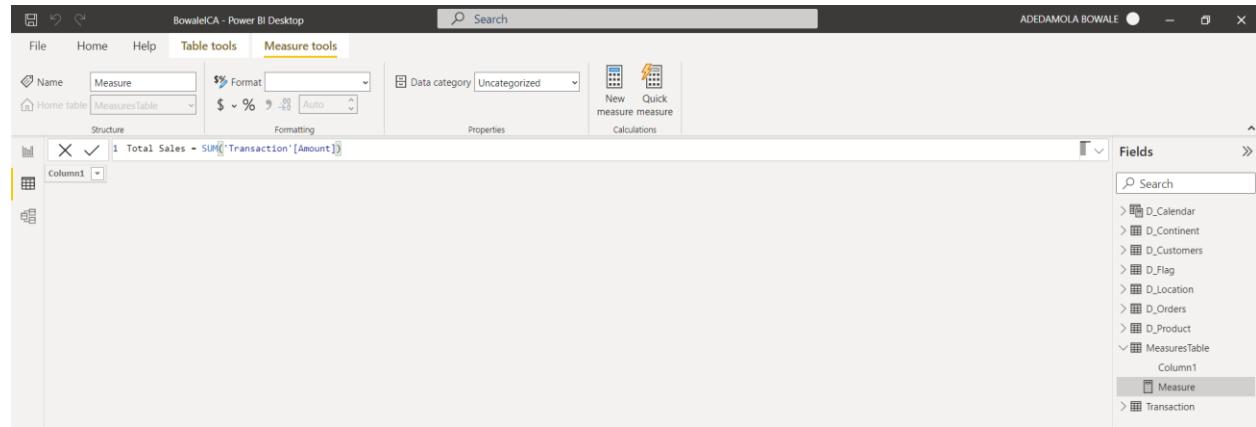


Figure 3. 7:create 'Total Sales' measure

Calculating the “TotalQuantity” measure. Total Quantity =  $\text{SUM}(\text{'Transaction'}[\text{Quantity}])$

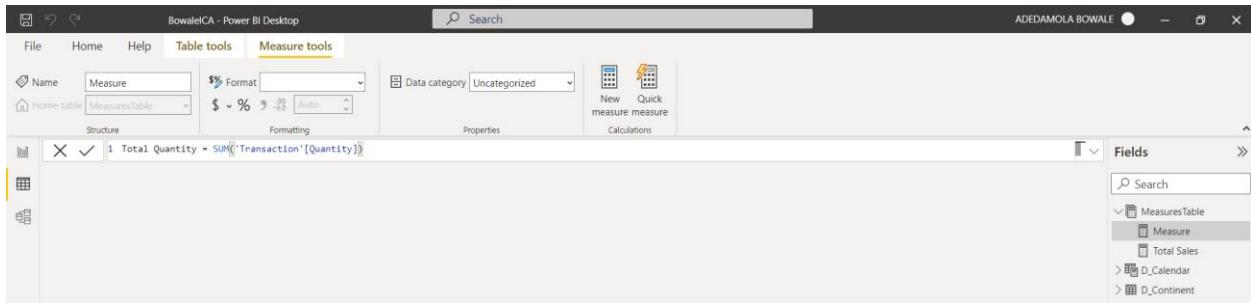


Figure 3. 8: create 'Total Quantity' measure

Calculating the “Number of Products” measure. Number of Products =  
 $\text{DISTINCTCOUNT}(\text{D_Product}[\text{ProductLine}])$

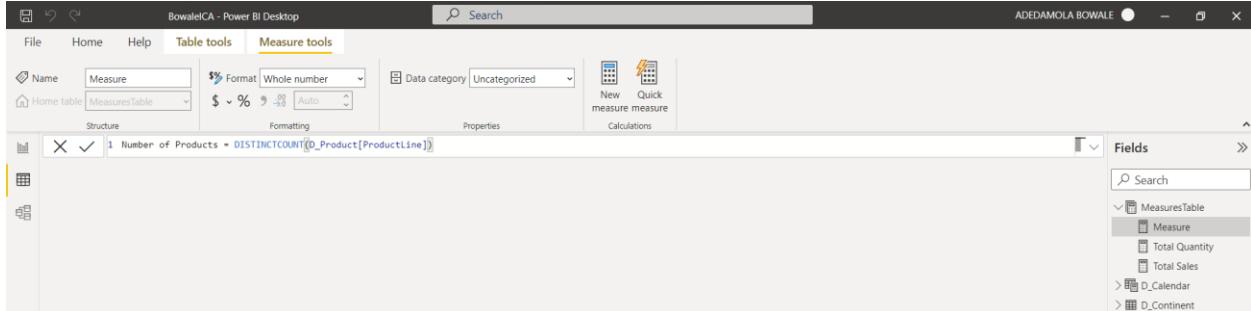


Figure 3. 9: create 'Number of Products' measure

Calculating the “Number of Companies” measure. Number of Companies =  
 $\text{COUNTA}(\text{D_Customer}[\text{CompanyName}])$

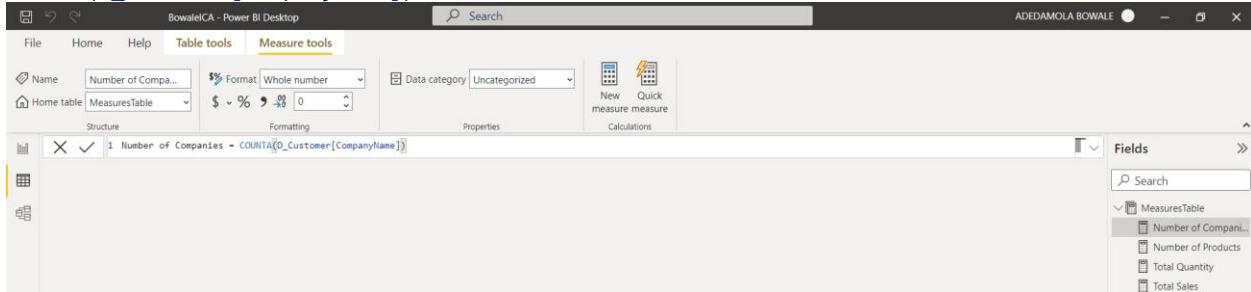


Figure 3. 10: create 'Number of Company' measure

Calculating the “Count of Location” measure. Count of Location = COUNTROWS(D\_Location)

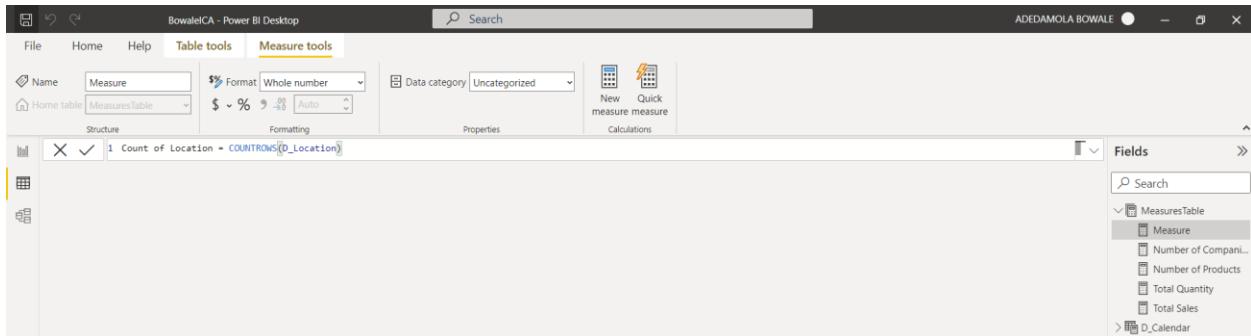


Figure 3. 11: create 'Count of Location' measure

Calculating the “Average Sales” measure. Average Sales = round(AVERAGE('Transaction'[Amount]),2)

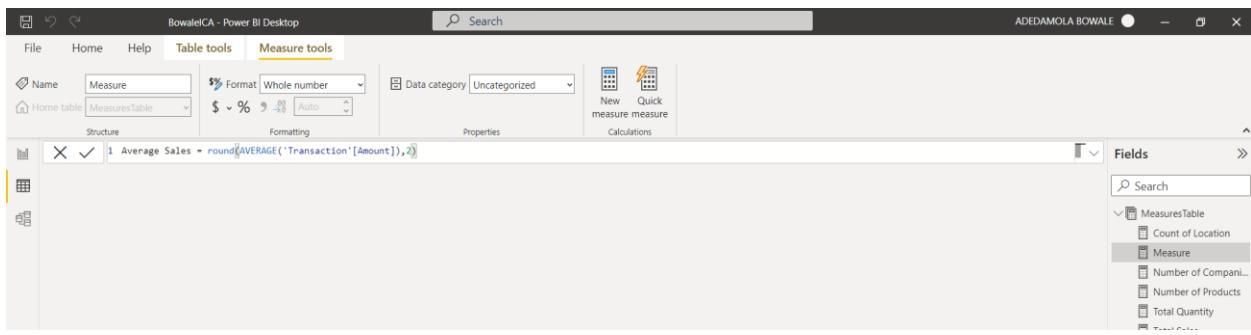


Figure 3. 12: create 'Average Sales' measure

Calculating the “Average Quantity” measure. Average Quantity = round(AVERAGE('Transaction'[Quantity]),2)

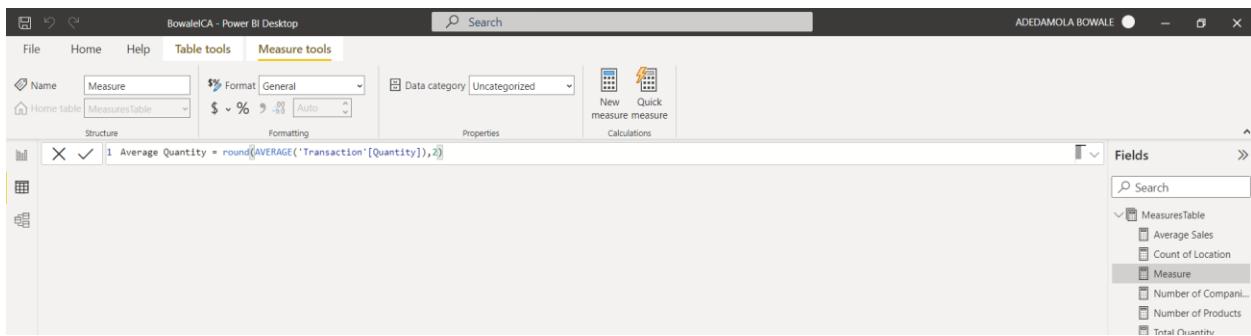


Figure 3. 13: create 'Average Quantity' measure

Calculating the “Weekday Sales Amount” measure. Weekday Sales Amount = `CALCULATE([Total Sales],D_Calendar[Week Cat] = "Weekday")`

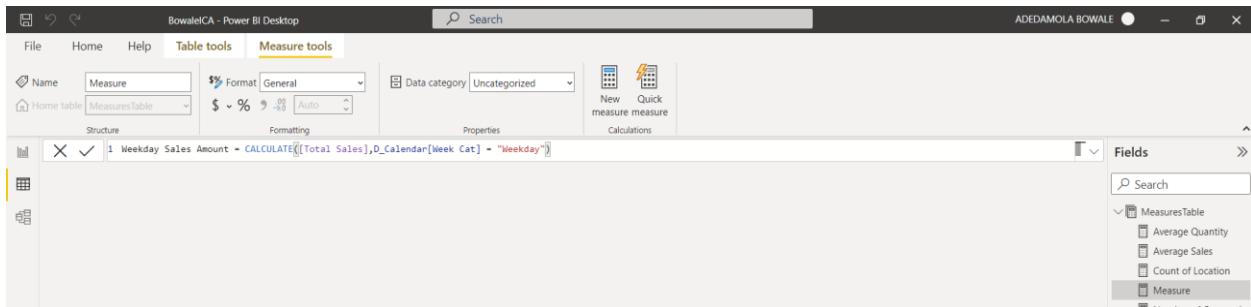


Figure 3. 14: create 'weekday sales amount' measure

Calculating the “Weekend Sales Amount” measure. Weekend Sales Amount = `CALCULATE([Total Sales],D_Calendar[Week Cat] = "Weekend")`

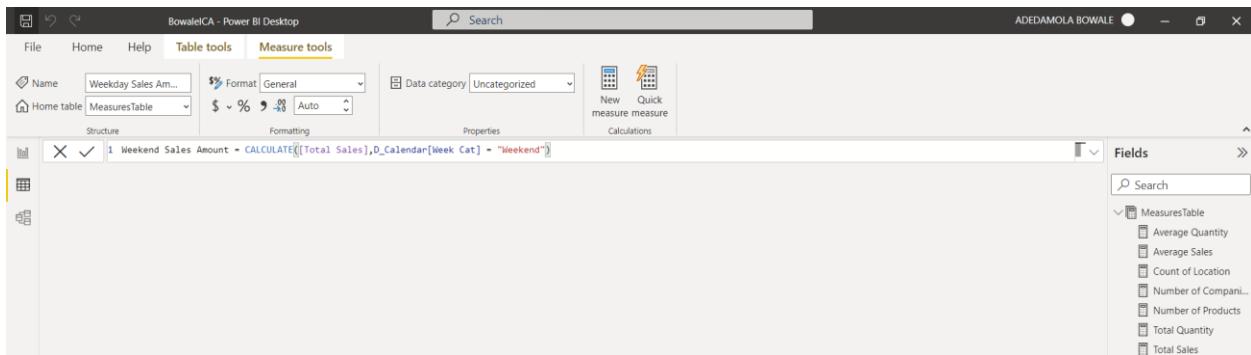


Figure 3. 15: create 'weekend sales amount' measure

Calculating the “Header” measure. This is to customize the Header

```
Header = if(SELECTEDVALUE(D_Continent[Continent])<> BLANK(),
           SELECTEDVALUE(D_Calendar[Year]) & " Sales Report for "
           & SELECTEDVALUE(D_Continent[Continent]), SELECTEDVALUE(D_Calendar[Year])
           & " Sales Report")
```

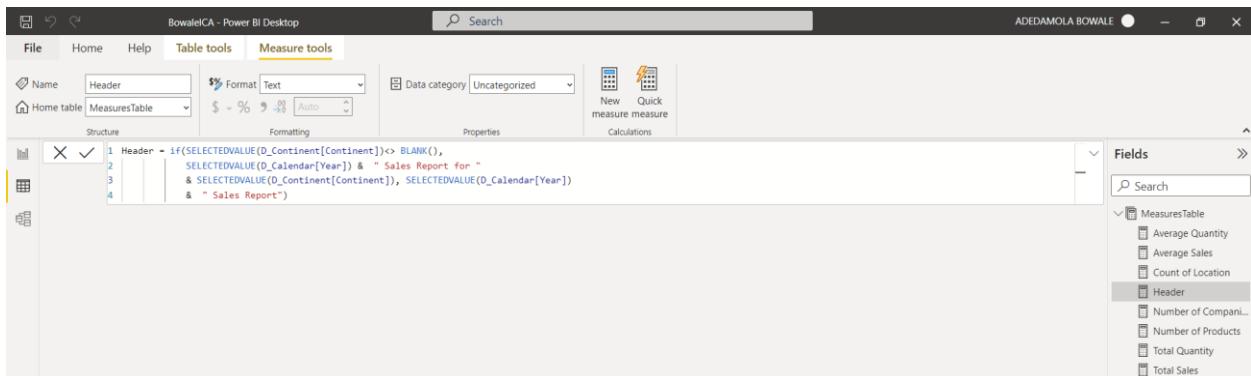


Figure 3. 16: create 'Header' measure

Calculating the “Footer” measure. To customize the footer

```
Footer = "Report as at " & format(MAX(D_Calendar[Date]), "ddd" & "dd")
```

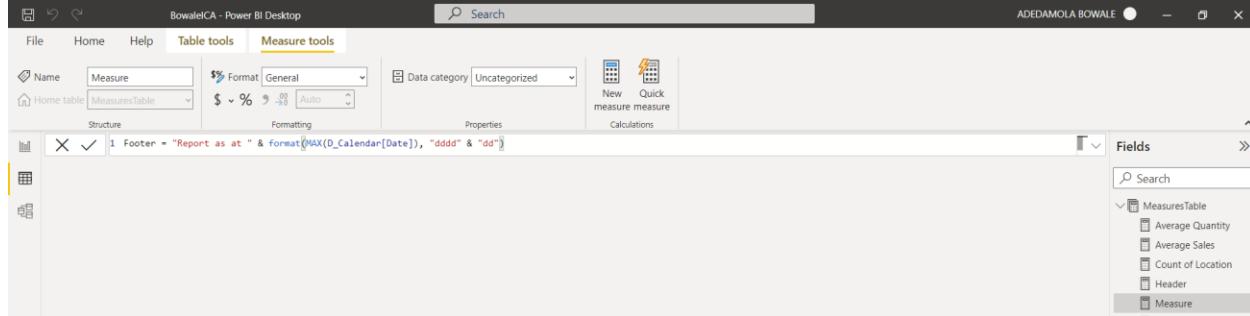


Figure 3. 17: create 'Footer' measure

Calculating the “Report Colour” measure. To customize the report colour

```
Report Colour = SWITCH(
```

```
    SELECTEDVALUE(D_Continent[Continent]),
    "America", "Dark Green",
    "Asia", "Dark Yellow",
    "Australia", "Dark Red",
    "Europe", "Dark Blue",
    "Black")
```

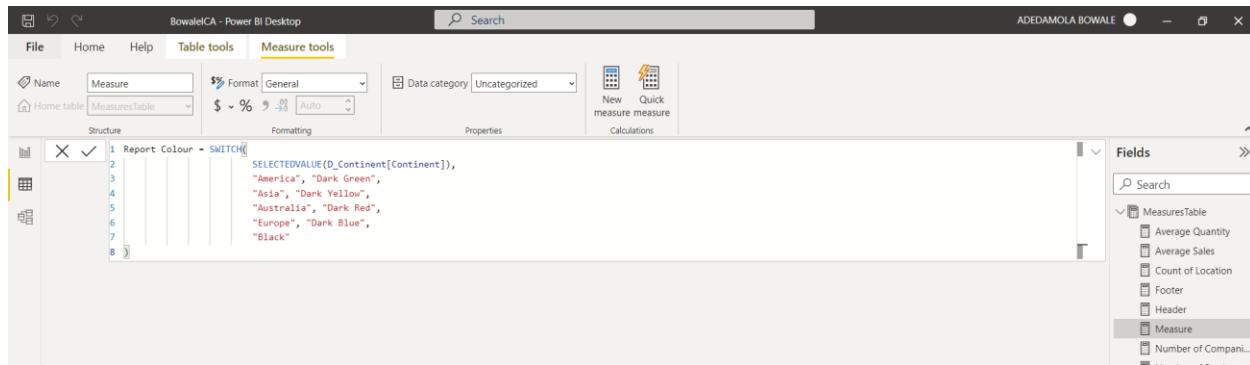


Figure 3. 18: create 'Report Colour' measure

Calculating the “Filtered Flag” measure.

```
Filtered Flag = CALCULATE (IF(ISFILTERED( D_Location[Country]), SELECTEDVALUE(D_Flag[Flag]),
"https://worlddiabetesday.org/wddbrk/wp-content/uploads/2020/10/event_map_image.jpg" ))
```

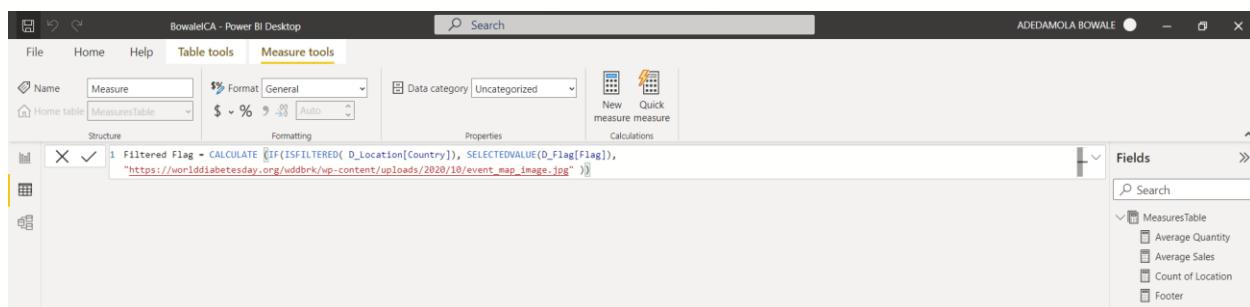


Figure 3. 19: create 'Filtered flag' measure

## PowerBI Visuals

### *Top 5 Company By Sales*

An analysis was carried out to identify the top 5 companies using a stacked bar chart (see Figure 3.20). This graph was created using the “Totalsales” measure as the value and “CompanyName” as the axis. According to Figure 3.20, “Euro Shopping Channel” leads in sales with 0.19million, followed by “Mini Gift Distributions Limited” with 0.65 million.

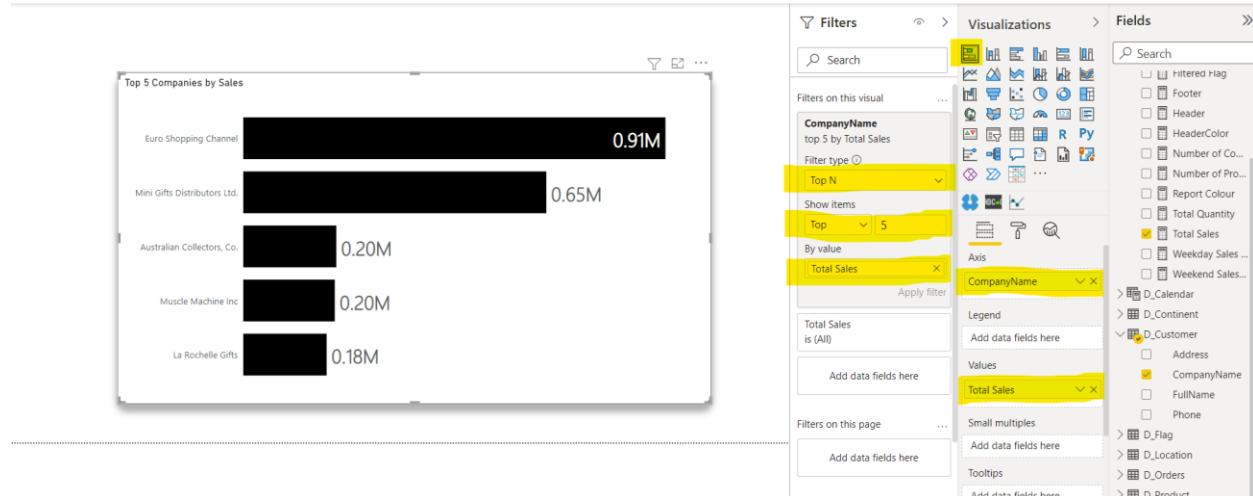


Figure 3. 20: Top 5 company by sales

### *Bottom 5 Company by Sales*

An analysis was carried out to identify the bottom 5 companies using a stacked bar chart (see Figure 3.21). This graph was created using the “Totalsales” measure as the value and “CompanyName” as the axis. According to Figure 3.21, “Boards & Toys Co.” made the lowest sales of 9.1 thousand, followed by “Atelier graphique” 24.2 thousand.

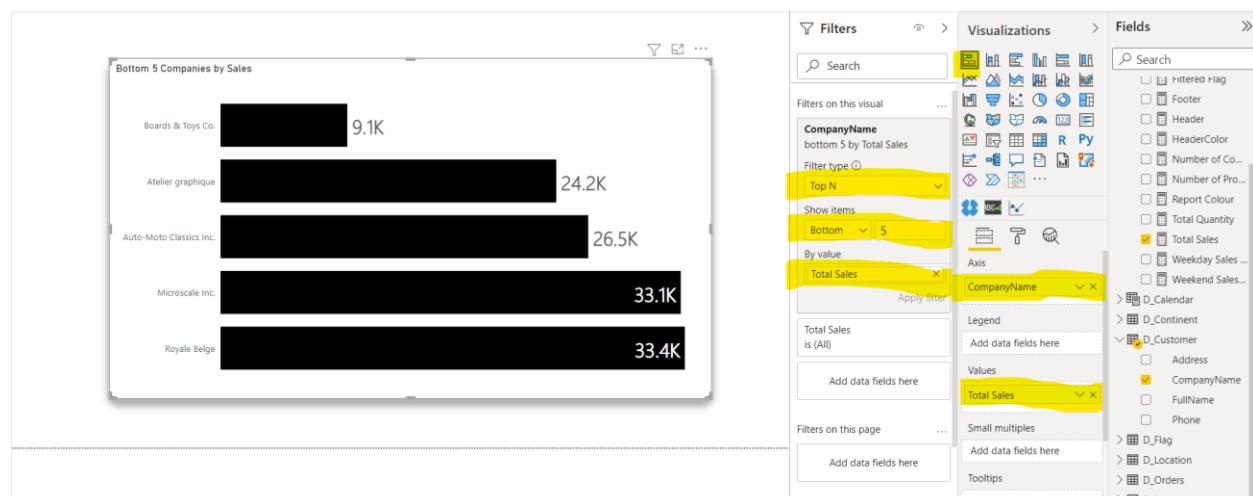


Figure 3. 21: Bottom 5 company by sales

### Total Sales By Year

An analysis was carried out to identify the year that made the highest sales using a stacked column chart (see Figure 3.22). This graph was created using the “Totalsales” measure as the value and “Year” as the axis. According to Figure 3.22, the highest sales was in 2004 and lowest in 2005 with 4.7million and 1.8million respectively.

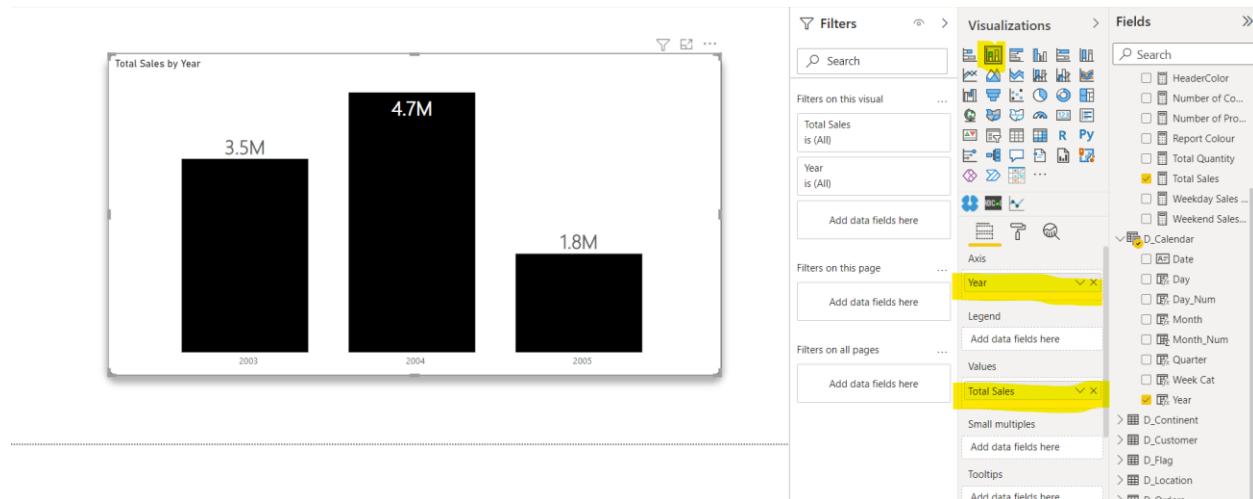


Figure 3. 22: Total sales by Year

### Total Sales By Month

An analysis was carried out to identify the month that made the highest sales using a LineDot chart imported from PowerBi Visual (see Figure 3.23). This graph was created using the “Totalsales” measure as the value and “month” as the Date According to Figure 3.23, the highest sales was in November followed by October.

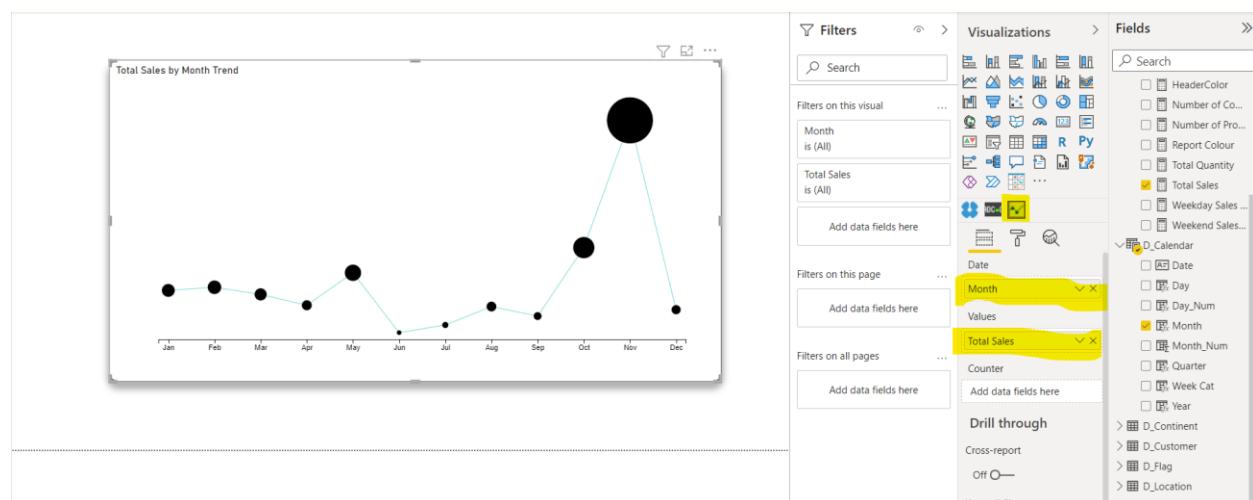


Figure 3. 23: Total Sales By Month

### Weekday Sales Trend

An analysis was carried out to identify the month during weekdays that made the highest sales using a LineDot chart imported from PowerBi Visual (see Figure 3.24). This graph was created using the “Weekday Sales Amount” measure as the value and “month” as the Date. According to Figure 3.24, the highest sales was in November followed by October.

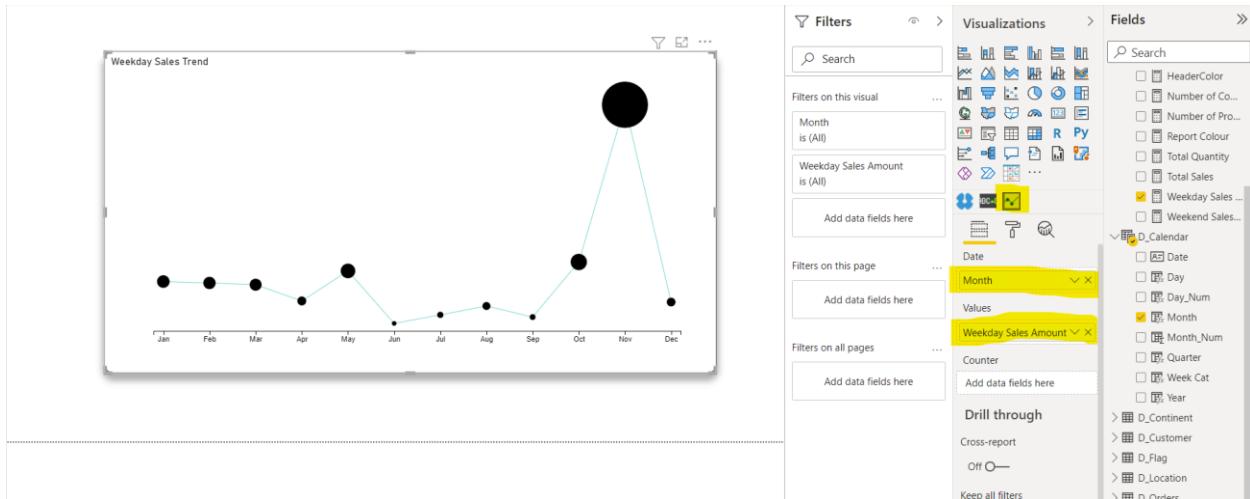


Figure 3. 24: Weekday Sales Trend

### Weekend Sales Trend

An analysis was carried out to identify the month during weekends that made the highest sales using a LineDot chart imported from PowerBi Visual (see Figure 3.25). This graph was created using the “Weekend Sales Amount” measure as the value and “month” as the Date. According to Figure 3.25, the highest sales was in October followed by November.

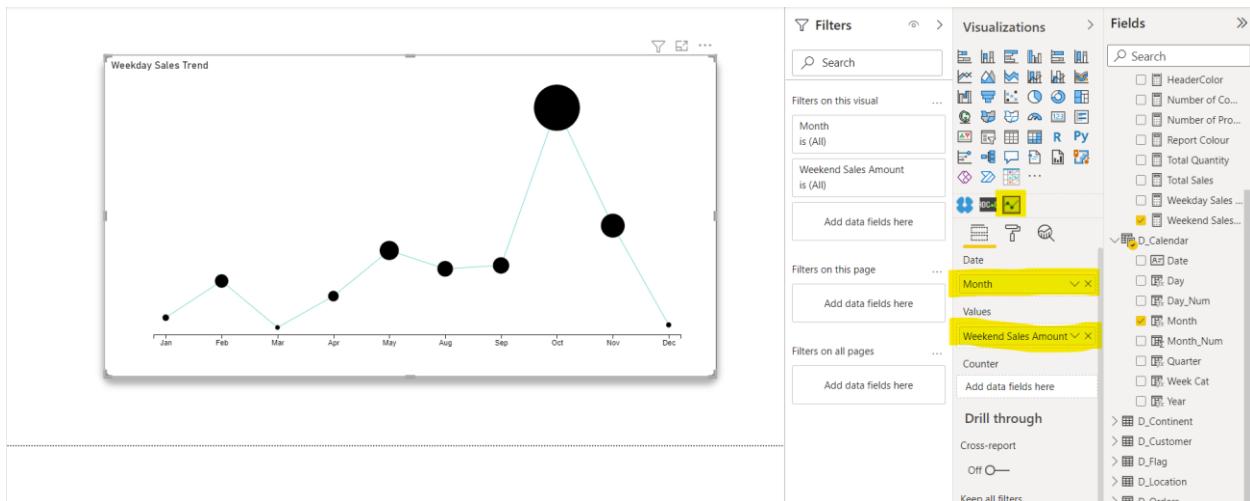


Figure 3. 25: Weekend Sales Trend

### Sales By Order Status

An analysis was carried out to identify the amount the status of an order has made (or worth). i.e. Amount of products that were successfully shipped, cancelled, on hold, Resolved, In-process and disputed using a stacked bar chart (see Figure 3.26). This graph was created using the “Amount” measure as the value and “Status” as the axis. According to Figure 3.26, Shipped products are worth 9.3 million and cancelled product are worth 200 thousand.

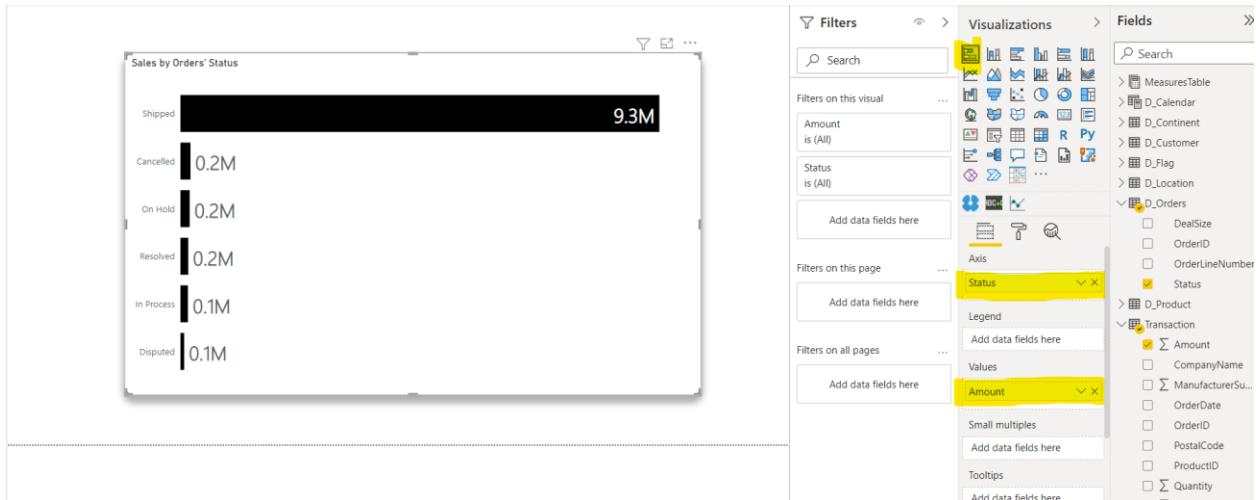


Figure 3. 26: Sales By Order Status

### Quantity By Order Status

An analysis was carried out to identify the order status with the highest quantity. i.e. The quantity of products that were successfully shipped, cancelled, on hold, Resolved, In-process and disputed using a stacked bar chart (see Figure 3.27). This graph was created using the “Quantity” measure as the value and “Status” as the axis. According to Figure 3.27, 91 thousand products were shipped and 2 thousand products were cancelled.

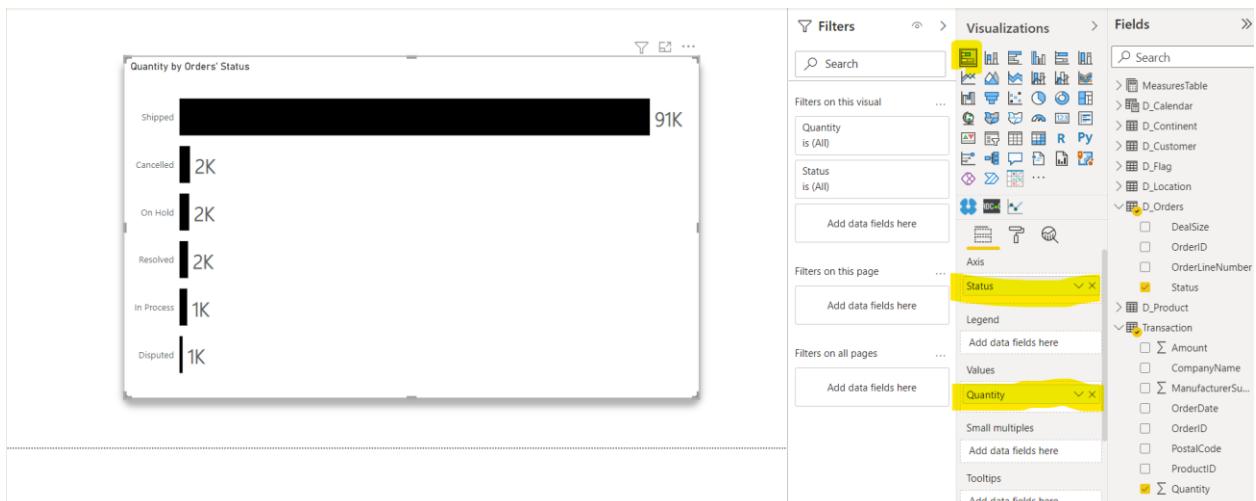


Figure 3. 27: Quantity By Order Status:

### Product Line by Sales

An analysis was carried out to identify the product Line that made the highest sales using Infographic Designer chart imported from PowerBi Visual (see Figure 3.28). This graph was created using the “Total Sales” measure as the value and “ProductLine” as the Category. According to Figure 3.28, Classic Cars made the highest sales with 3.7 million followed by Vintage cars with 1.4 million. The least were Trains with about 300 thousand.



Figure 3. 28: ProductLine by Sales

### Product Line by Quantity

An analysis was carried out to identify the product Line with the highest and lowest quantity sold using Infographic Designer chart imported from PowerBi Visual (see Figure 3.29). This graph was created using the “Quantity” measure as the value and “ProductLine” as the Category. According to Figure 3.29, Classic Cars had the highest quantity of sales compared to the other Products.

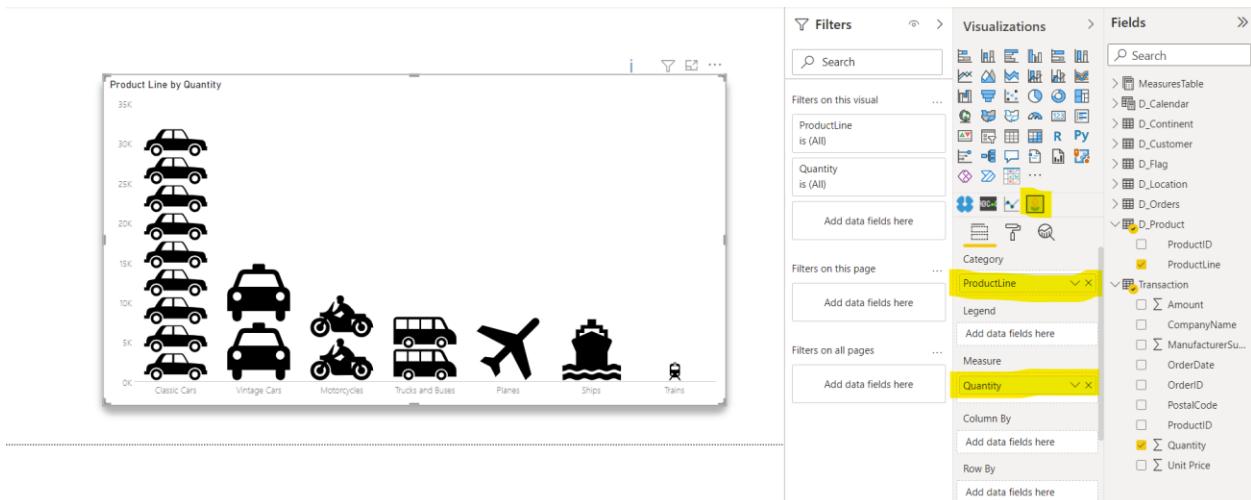


Figure 3. 29: ProductLine by Quantity

### *Top 5 Countries by Total Sales*

An analysis was carried out to identify the Top 5 countries with the highest sales using Stacked bar chart (see Figure 3.30). This graph was created using the “Total Sales” measure as the value and “Country” as the Axis. According to Figure 3.30, Spain had the highest sales followed by the USA.

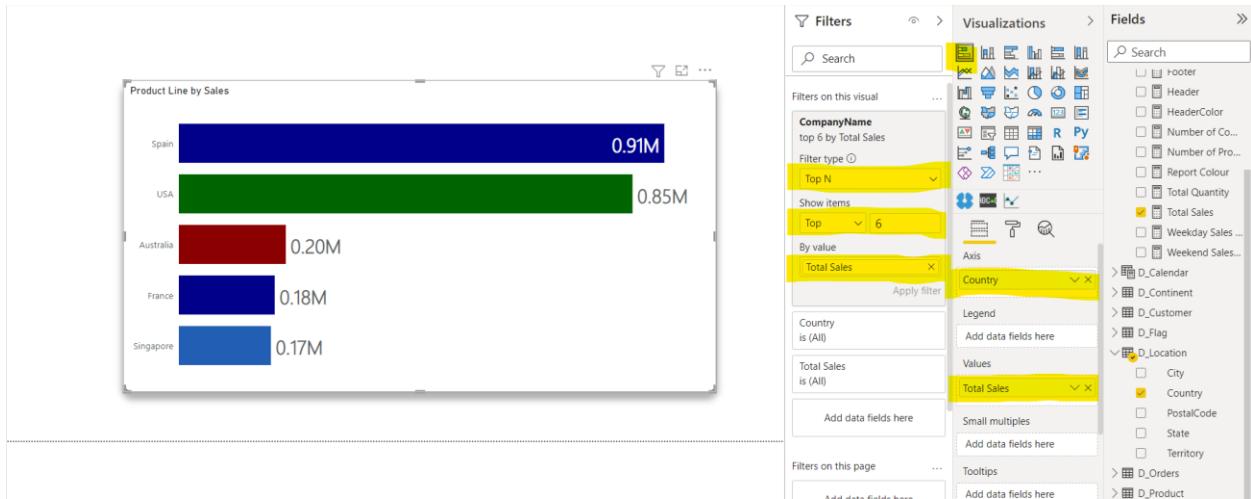


Figure 3. 30: Top 5 Countries by Total Sales

### *Top 5 City by Total Sales*

An analysis was carried out to identify the Top 5 cities with the highest sales using Stacked bar chart (see Figure 3.31). This graph was created using the “Total Sales” measure as the value and “City” as the Axis. According to Figure 3.31, Madrid had the highest sales followed by Bridgewater.

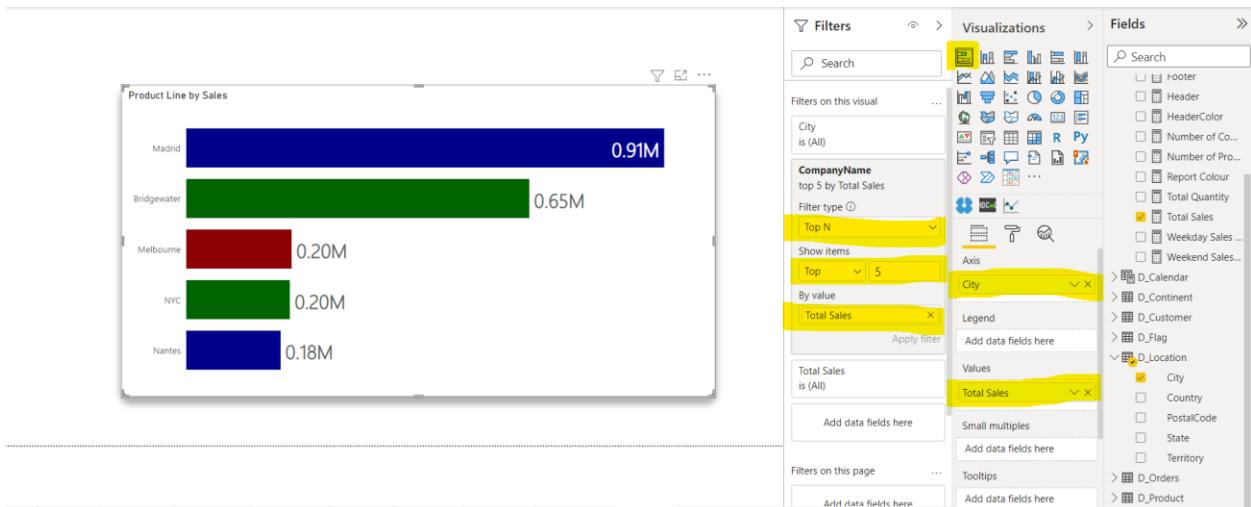


Figure 3. 31: Top 5 City by Total Sales

## Scorecards

Calculated Measures were already created for the score cards. Selecting the score card visual, then dragging the calculated measures into the “Fields” input will give us the visual as seen in Figure 3.32. Scorecards were created for “Total Sales”, “Count of Location”, “Weekend Sales Amount”, “Weekday Sales Amount”, “Total Quantity”, “Number of Products” and “Number of Companies”.

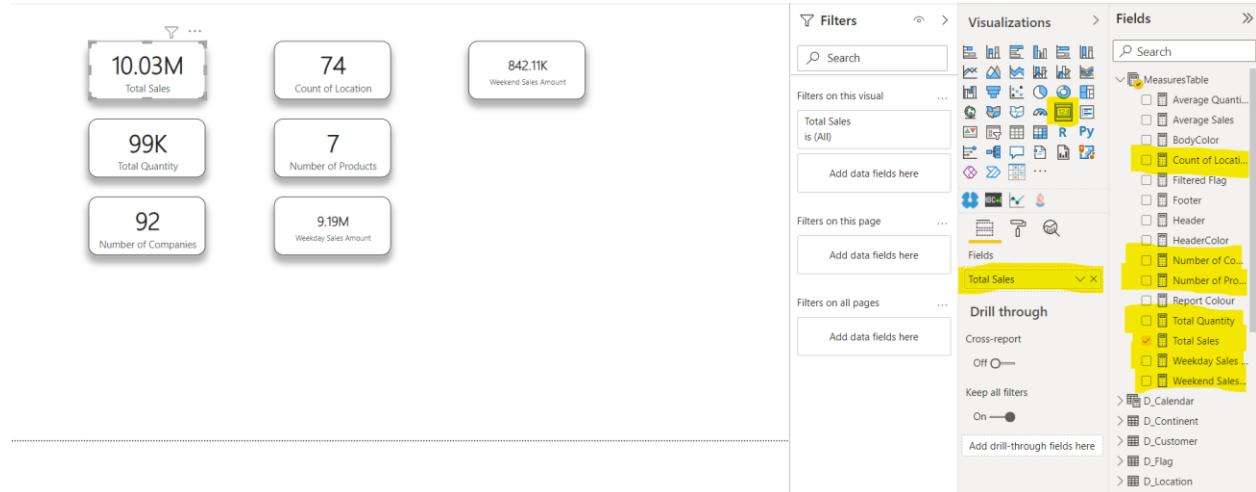


Figure 3. 32: Scorecards

The “Order Count”, “Orderline Count”, “Header” and “Footer” scorecards were also created as seen in Figure 3.33 to 3.36

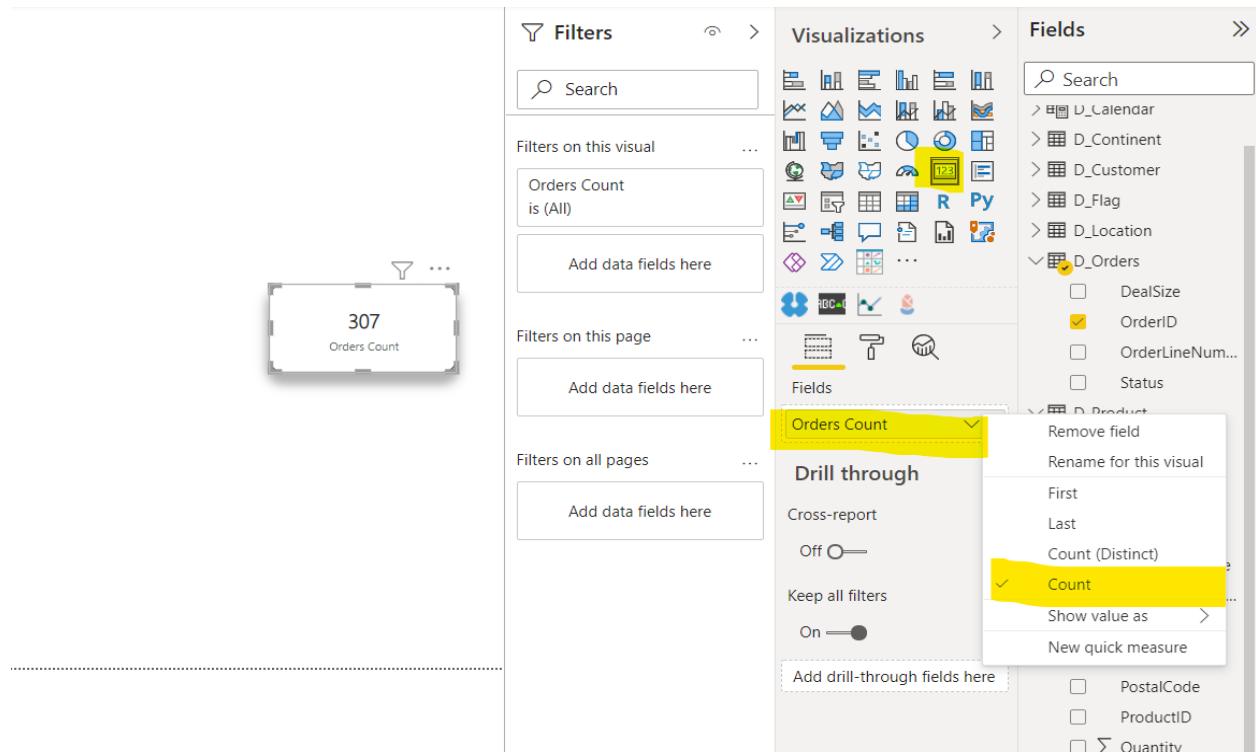


Figure 3. 33: Order Count

## Orderline Count

The screenshot shows the Power BI desktop interface with the 'Orderline Count' visualization selected. The visualization displays the value '17' and the text 'Distinct OrderLineNumber'. The context menu for this visualization is open, with the 'Fields' section highlighted. In the 'Fields' section, the 'D\_Orders' table is expanded, showing fields like 'DealSize', 'OrderID', 'OrderLineNumber' (which has a checked checkbox), and 'Status'. A yellow box highlights the 'Count (Distinct)' option under the 'Count' section of the context menu, which is also highlighted.

Figure 3. 34: Orderline Count

## Header

The screenshot shows the Power BI desktop interface with the 'Header' visualization selected. The visualization displays the text 'Sales Report'. The context menu for this visualization is open, with the 'Fields' section highlighted. In the 'Fields' section, the 'Header' field is selected and highlighted with a yellow box. The 'MeasuresTable' table is expanded, showing various measures like 'Average Quantity', 'BodyColor', 'Count of Locations', 'Filtered Flag', 'Footer', 'Header' (which has a checked checkbox), 'HeaderColor', 'Number of Companies', 'Number of Products', 'Report Colour', 'Total Quantity', 'Total Sales', 'Weekday Sales', and 'Weekend Sales'. A yellow box highlights the 'Header' field in the 'Fields' section of the context menu.

Figure 3. 35: Header

## Footer

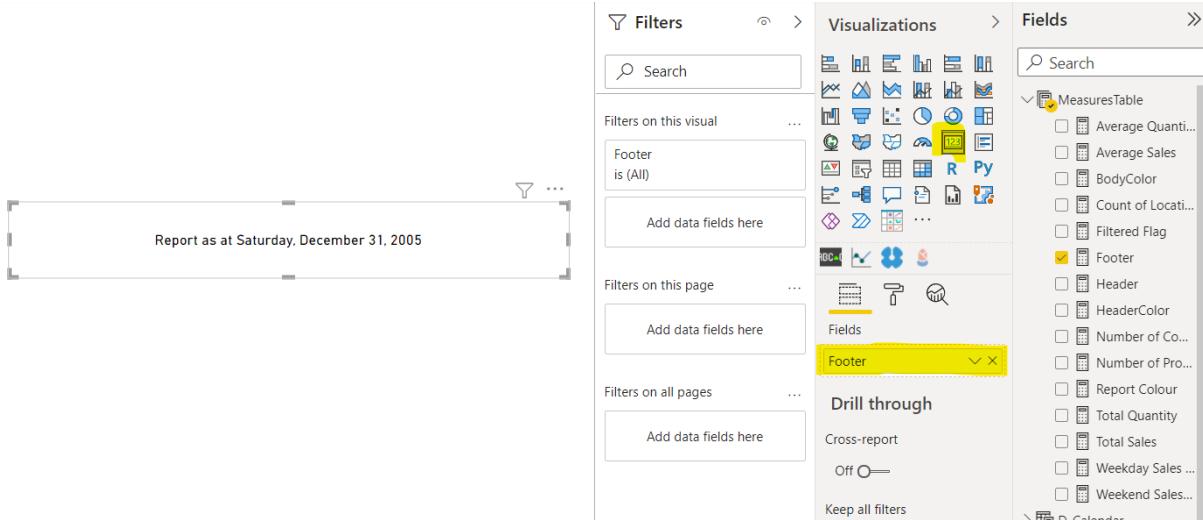


Figure 3. 36: Footer

## Scroller

An analysis was carried out to identify the Average Sales and Average quantity of each Product Line (see Figure 3.37). This visual was created using the “Average Quantity” measure as the Measure Deviation, “Average Sales” measure as the Measure Absolute and “ProductLine” as the Category

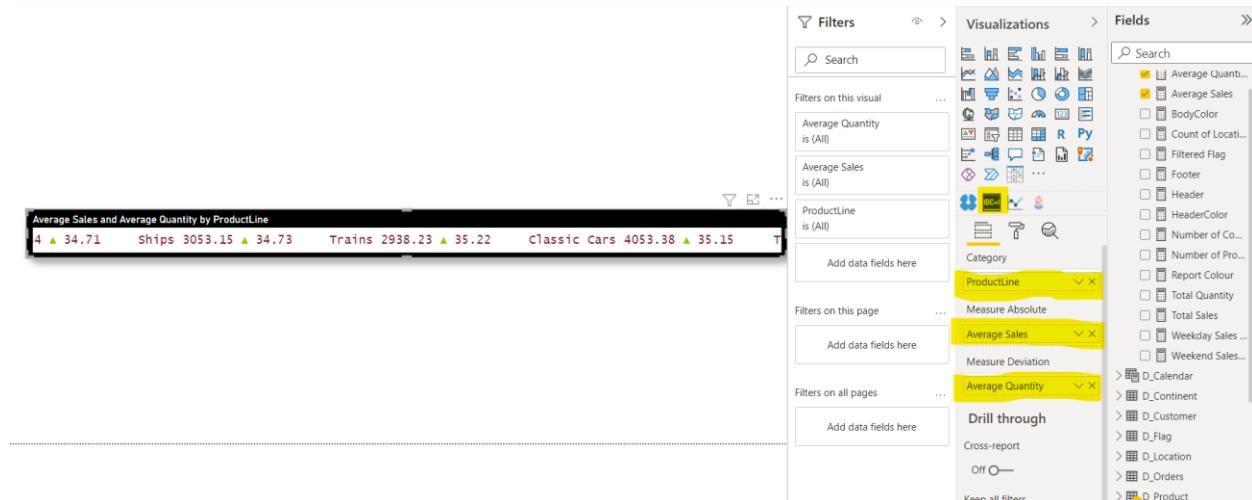


Figure 3. 37: Scroller

Image by Cloudscope: The visuals help us to choose an image, However, the filtered flag calculated measure reconfigures the visuals based on which country was selected. Please, see Example below.

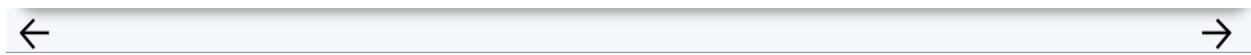
The screenshot shows the Power BI visualization interface. On the left, there is a legend titled "Country" with a list of countries: (Blank), Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Norway, and Philippines. The entry for "France" is highlighted with a yellow box. In the center, there is a visual element displaying the flag of France. On the right, the "Filters" pane shows a search bar and several filter sections: "Filters on this visual" (with "Filtered Flag is (All)" selected), "Filters on this page" (with "Add data fields here"), "Filters on all pages" (with "Add data fields here"), and "Image url" (with "Filtered Flag" selected). The "Fields" pane also has a search bar and lists various measures and fields, with "Filtered Flag" checked and highlighted in yellow.

Figure 3. 38: Image by Cloudscope

Buttons were added to enable easy Navigation for Users.



Navigation Arrows



To view the open-source Dataset



## Description of the Power BI dashboard

- Sales report summary page

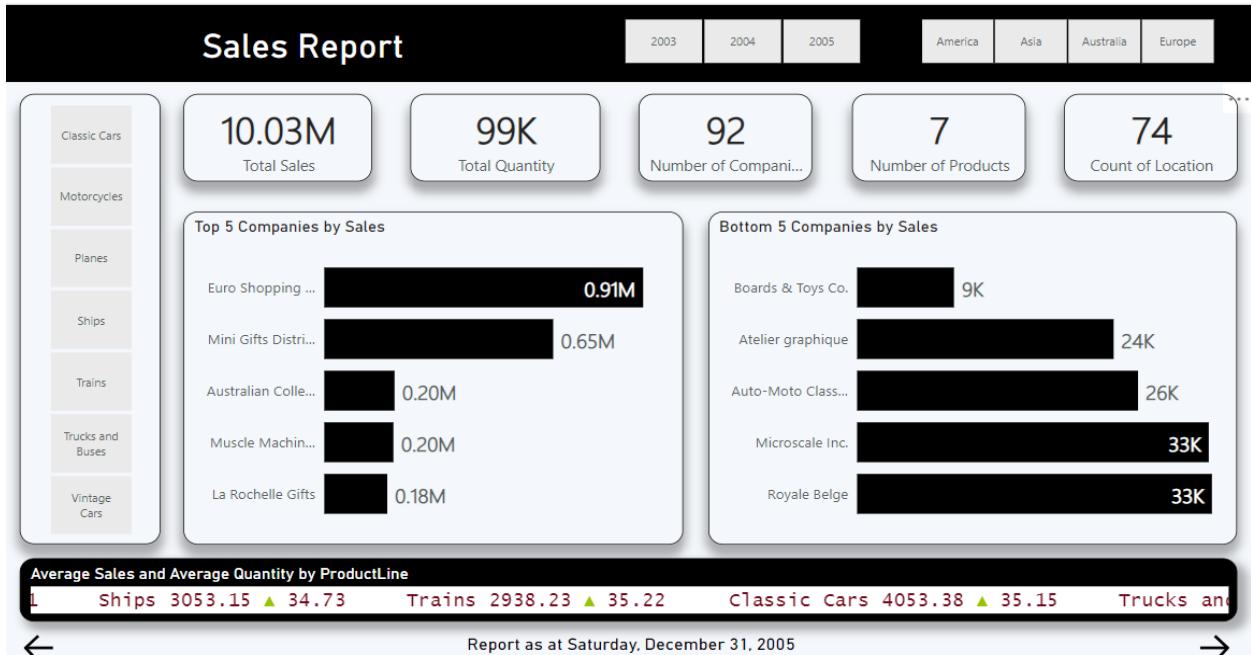


Figure 3. 39: Sales report Summary page

- Sales report by Year and Month



Figure 3. 40: Sales report by Year and Month

- Sales Report by Weekends and Weekdays

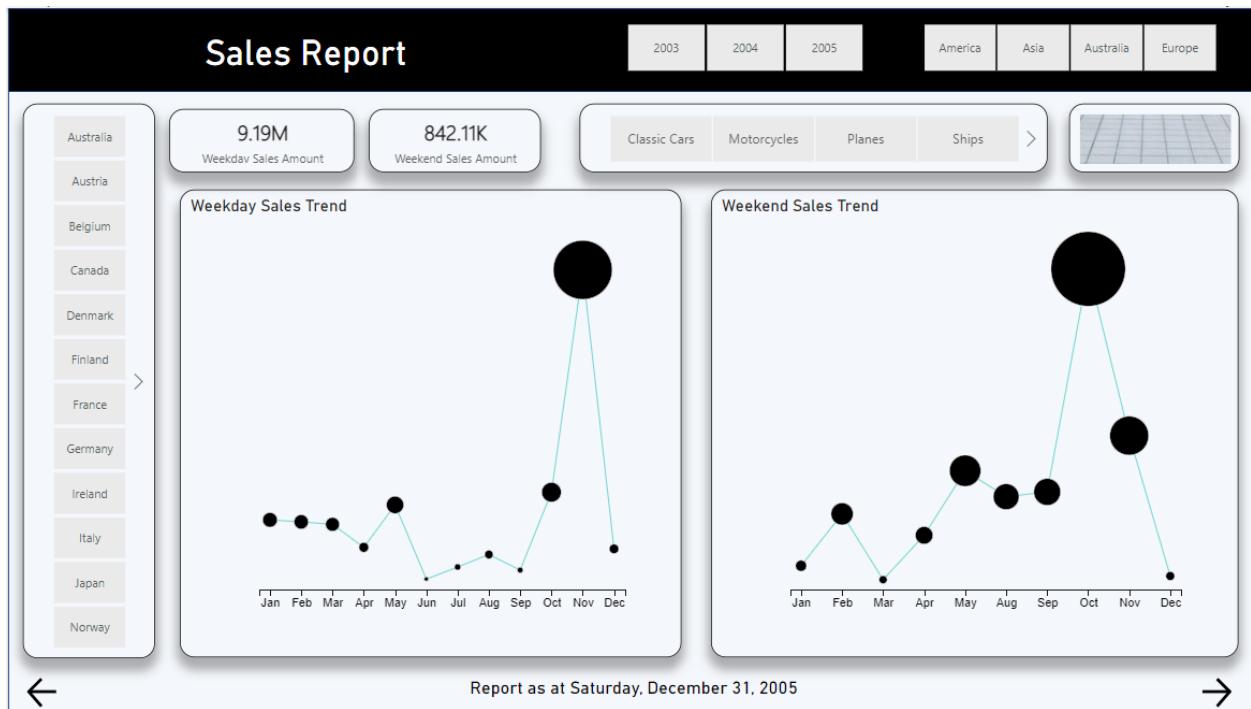


Figure 3. 41: Sales Report by Weekends and Weekdays

- Sales Report by Order Status (Sales and Quantity)

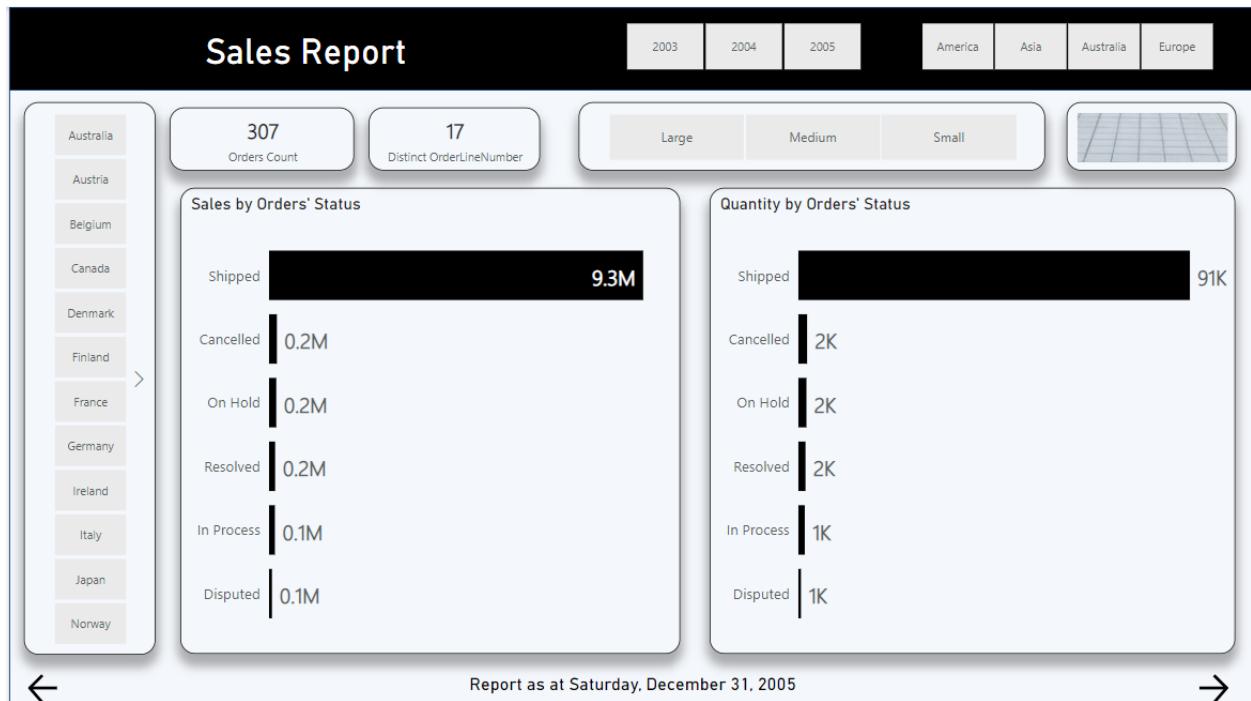


Figure 3. 42: Sales Report by Order Status (Sales and Quantity)

- Sales Report by Product Line(Sales and Quantity)

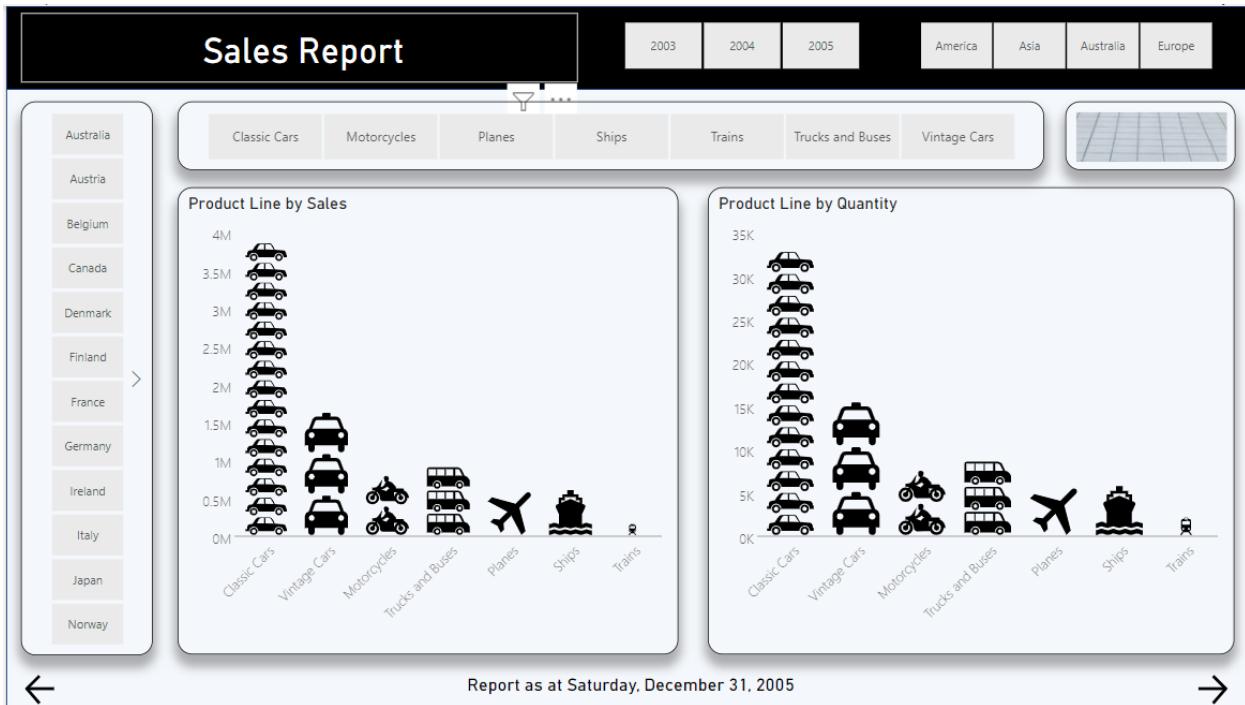


Figure 3. 43: Sales Report by Product Line(Sales and Quantity)

- Sales report by Country (Sales and Quantity)

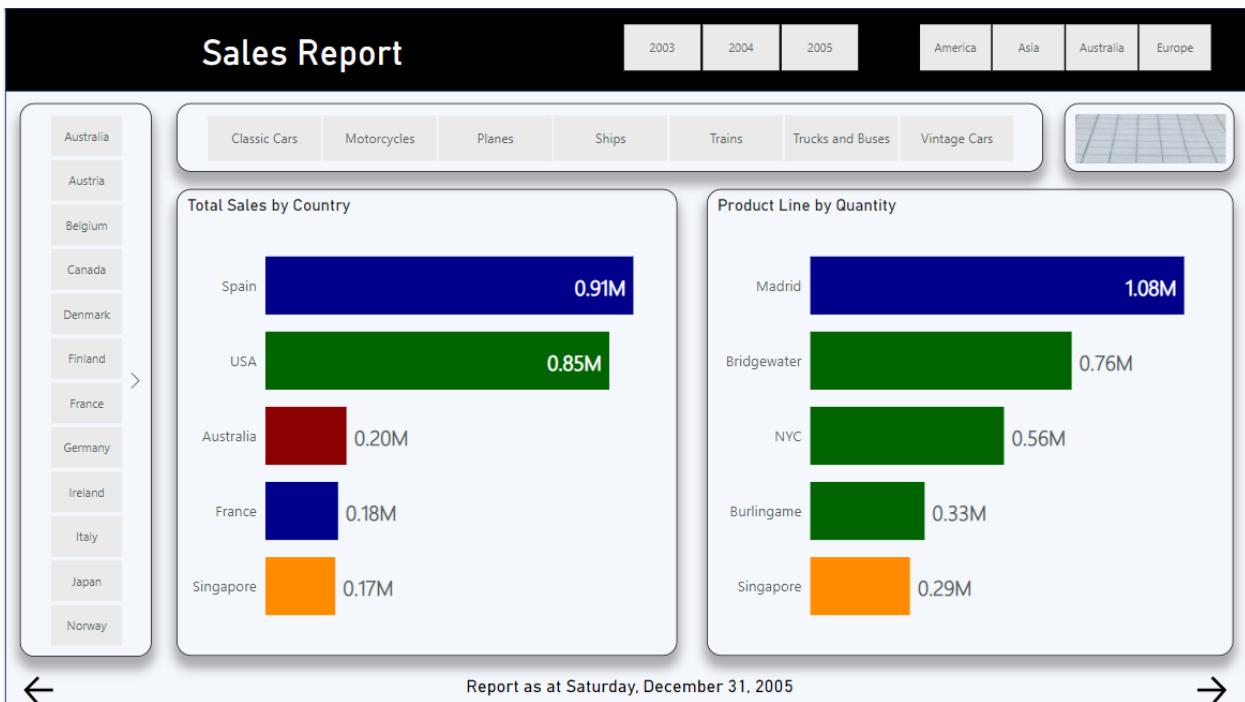


Figure 3. 44:Sales report by Country (Sales and Quantity)

## Key findings

- The “Switch” function was very effective. (Less complicated compared to the “IF” function)
- People are more interested in classic cars compared to vintage cars
- Sales are usually extremely high on November compared to other months
- There are more customers in Europe compared to other continents

## Conclusion

The implementation and design of this solution has empowered us to be able to generate insights and create intuitive reports which eventually helps in taking actions and making smart decisions. There is no doubt that the Vehicle retail company has established their market in Europe with classic cars, However, there is an absolute need for expansion to sell to other parts of the world.