

Big Data and Business Intelligence

Module Code: CIS4008-N

Project Topic

E-Commerce Analysis on Sales Discounts

Name: Chinedu Emmanuel Agwunobi

Student ID: B1555134

Table of Contents

1	Executive Summary.....	4
2	Introduction	7
3	Key Findings Based on Analysis and Evaluation.....	9
3.1	Sales Overview	9
3.1.1	Card Visuals.....	10
3.1.2	Gender Infographic Design	11
3.1.3	Discounts Values using Treemap	11
3.1.4	Hierarchical Representation of the Regions	12
3.1.5	Top 10 Countries and Geographical Locations	12
3.1.6	The Line and Bar Chart for Sales and Profits.....	13
3.2	Sales Performance	17
3.2.1	Card Visuals.....	17
3.2.2	Monthly Performance of Sales.....	19
3.2.3	Product Category Analysis	20
3.2.4	Top 10 Countries by Sales Revenue.....	21
3.2.5	Summary Table	22
3.3	Sales Growth	23
3.3.1	The Waterfall Chart.....	24
3.3.2	Stacked Area Chart.....	25
3.4	Key Influencers.....	27
3.4.1	Card Visuals.....	27
3.4.2	Relationship between Quantity Sold and Product Price.....	28
3.4.3	Relationship between Sales Price and Product Price.....	29
3.4.4	Key Influencers Chart	29
3.4.5	Matrix table.....	32
3.5	Sales Prediction.....	33
3.5.1	Sale Price and Product Category.....	34
3.5.2	Decomposition Tree.....	35
3.5.3	Sales Prediction for the Next 5 Years.....	35
3.5.4	Relationship between Current Sales and Sales Growth.....	36
4	Summary	36
5	Recommendation.....	37
6	Conclusion.....	38
7	Data Pre-Processing and Cleaning	40

7.1	Loading the Data	40
7.2	Data Cleaning	41
7.2.1	Removing Errors, Blanks, NA, and Column	42
7.3	Data Modelling.....	45
7.3.1	Creating Tables.....	45
7.3.2	Star Schema	50
7.4	DAX and M-Language.....	53
7.4.1	DAX.....	53
7.4.2	M-Language	55
7.5	Dashboard.....	56

Section 1

1 Executive Summary

1.1. Introduction

Sales discounts is an eye-catching technique used virtually in every industry, perhaps the most prominent platforms being retail and eCommerce. Promotional pricing adverts like flash, seasonal, and clearance sales are common examples of sales discounts. These techniques are used to attract new customers, project new products to the market, and retain customers, which most times have proven good methods of generating demand and revenue and can be disastrous if not well implemented.

1.2. Key Findings

- Products with 8% discounts have the highest quantity sold with sales revenue of about 4 million dollars.
- The year 2019 has the lowest profit margin(Total Profit / Sales Revenue) compared to other years.
- There was a rapid increase in total profit between 2019 to 2020.
- No specific month retailed a steady increase of sales annually between 2016 to 2020.
- Product Category; Industrial, has the highest demand over the years with an average demand of 839.6 products sold.
- Amount lost by discounts per sales revenue is 5.8%, which is less than 10% obviously the discounts were well implemented.
- Africa had the highest sales revenue with Ivory Coast having the highest demand of 484 products.
- By country, the US being the number one generated about 237 million dollars.
- We recorded a sale growth rate of 2.8%, could this be because of the sales discounts?

- Some products still had a negative growth rate even though we had an overall growth rate.
- There is a linear correlation between product price and quantity sold by discounts.
- Most countries were below average sales monthly even though we had a 2.8% growth rate.
- Note, that when the discount percentage increases, could increase demand but reduces sales revenue and profits.
- There could be an increase in sale revenue in the next five years with little fluctuations.

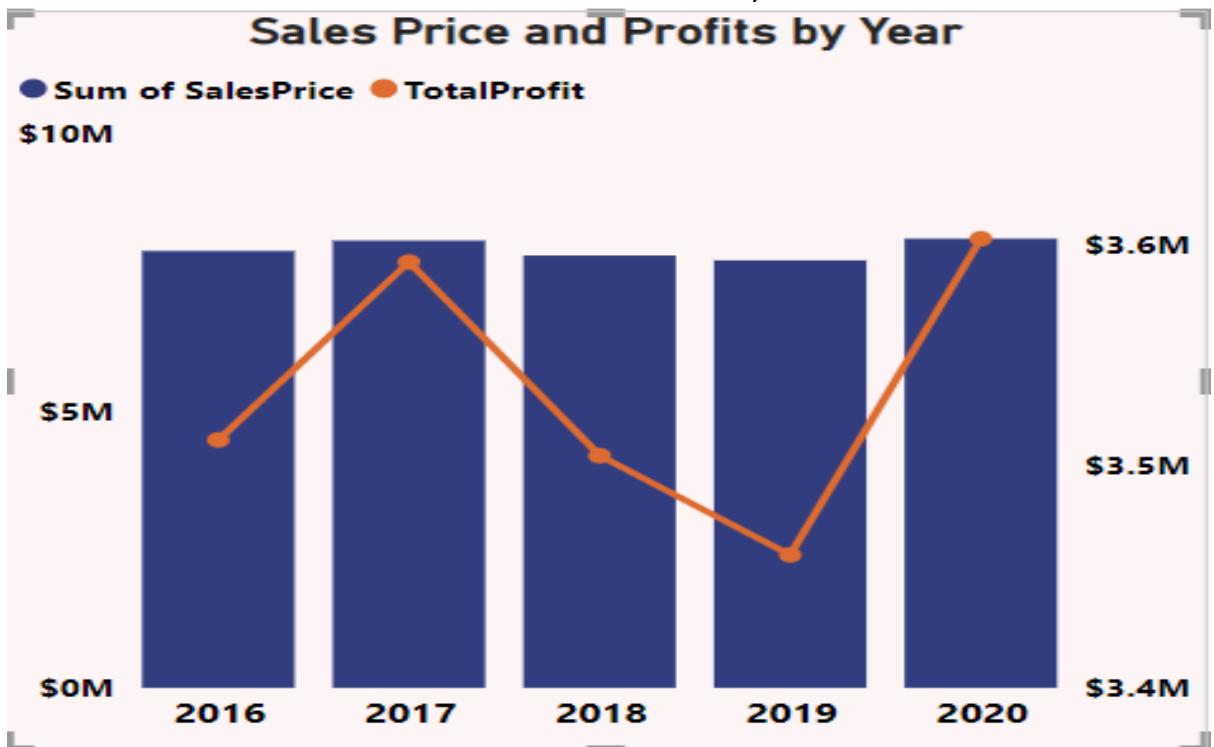


Fig.1. Sales Price and Profits by Year

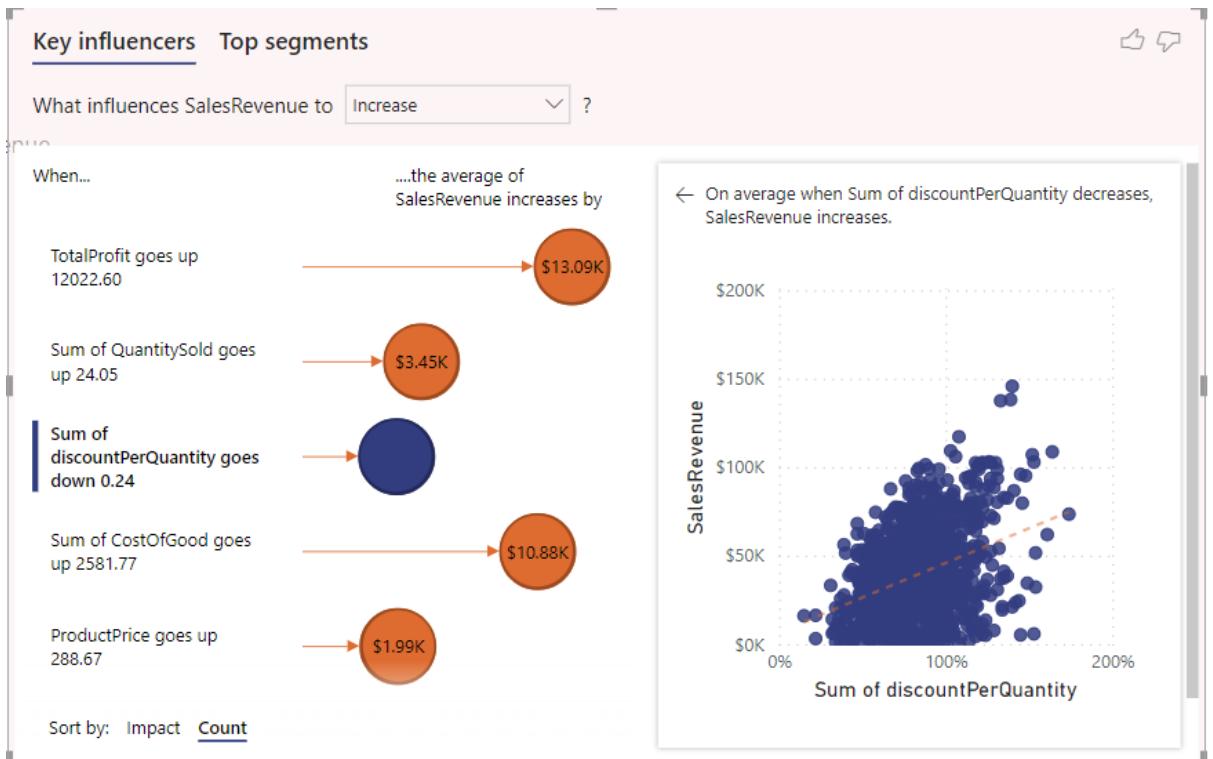


Fig.2. Discounts as a Key Influencer

1.3. Recommendations

- Set a target to reach more of developing countries, especially the African region.
- Offer promotional sales like flash sales to clear old products, monthly, seasonal, or buy one get one free(BOGOF).
- Set discounts for new customers, it acts as an incentive and attracts more customers.
- Have a loyalty program or coupons for loyal customers, for a long business relationship.
- Provide discounts for a certain number of quantities purchased by a customer.
- Make use of percentage discounts than absolute discounts, a 25% discount on \$40 is the same \$10 discount, but the percentage looks higher.
- Check out your competitors before implementing sales discounts.
- Avoid fraudulent promotional sales by inflating product prices before implementing sales discounts.
- Track and monitor the amount lost by discounts, could be fatal for the business.

2 Introduction

Sales discounts have contributed to most of the growth we have seen today on eCommerce, virtually everyone wants a reduced price, whether implied or not. According to Statista, 62% of consumers try a new store or online retailer during the holidays because of better prices, and 44% try a new store because of coupons or discounts.

Technically, most people go online to search for sales coupons, discounts, and promotions before making an order online. Sales discounts today act as incentives to many eCommerce businesses.

Discounts were implemented in all the sales in this data, the following are the questions that the analysis seeks to answer:

- Which region generates the highest revenue and profit?
- Which country generates the highest revenue and profit?
- Which African countries are among the top 10 sales revenue?
- What year had the highest revenue.
- What are the ranges of discounts implemented?
- What discount % has the highest growth rate?
- Is there a discount % with a negative growth rate?
- Was there any month that maintained a steady increase annually?
- Does Discounts increase demand?
- Does discounts influence sales?
- Will sales revenue increase over the next five years?

2.1 Data Source

The dataset used for this analysis is the E-Commerce Dataset.csv which was downloaded as a CSV file from [Kaggle](#). The dataset has one table with 18 columns.

Index	Column name	Description
1	CustomerId	The ID of the customers
2	CustomerName	The name of the customers
3	CustomerEmail	The email address of the customer
4	CustomerGender	The gender of the customer; male or female
5	CustomerIp	The IP location of the customer
6	ProductName	Name of the product
7	ProductPrice	The market price of the product
8	CostOfGood	Initial cost of goods

9	ProductCategory	The group the product belongs to
10	CustomerCountry	The geographic of the customer
11	CustomerContinent	The region of the customer
12	OrderTime	The time of the day at which the product was ordered
13	OrderDay	The day the product was ordered.
14	OrderMonth	The month the product was ordered.
15	QuantitySold	The number of products sold
16	discountPerQuantity	The specific discount on each product
17	salePrice	The price at which the product was sold
18	OrderYear	The year the product was ordered.

The E-Commerce Dataset was later broken down into 5 tables during the data preposing: E-Commerce_Customers, Date_Dim, E-Commerce_Countries/Regions, E-Commerce_Products, E-Commerce_Sales_facts , as shown in Fig.3. This will be discussed later in the chapter.

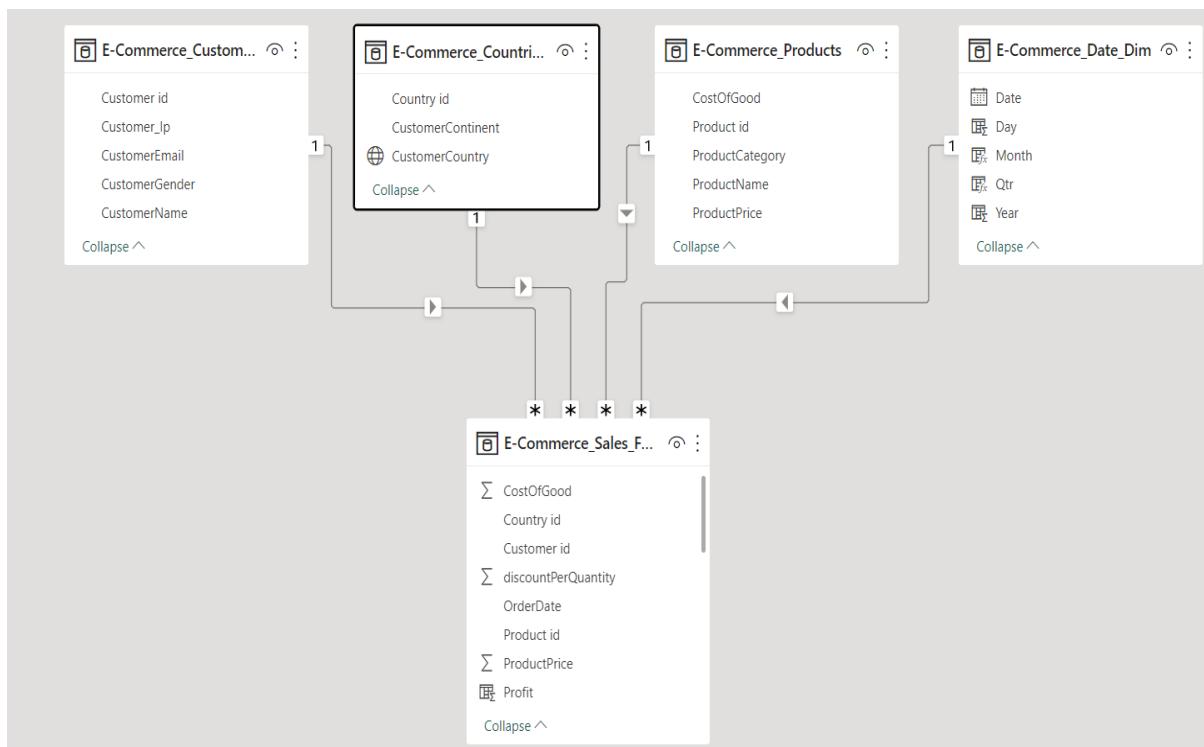


Fig.3. Data Model

3 Key Findings Based on Analysis and Evaluation

Below are various visualisation charts created for this report to answer the above business questions. The dashboards were created by carefully observing the data in this order; Sales Overview, Sales Performance, Sales Growth, Key Influencers, and Sales Prediction.

3.1 Sales Overview

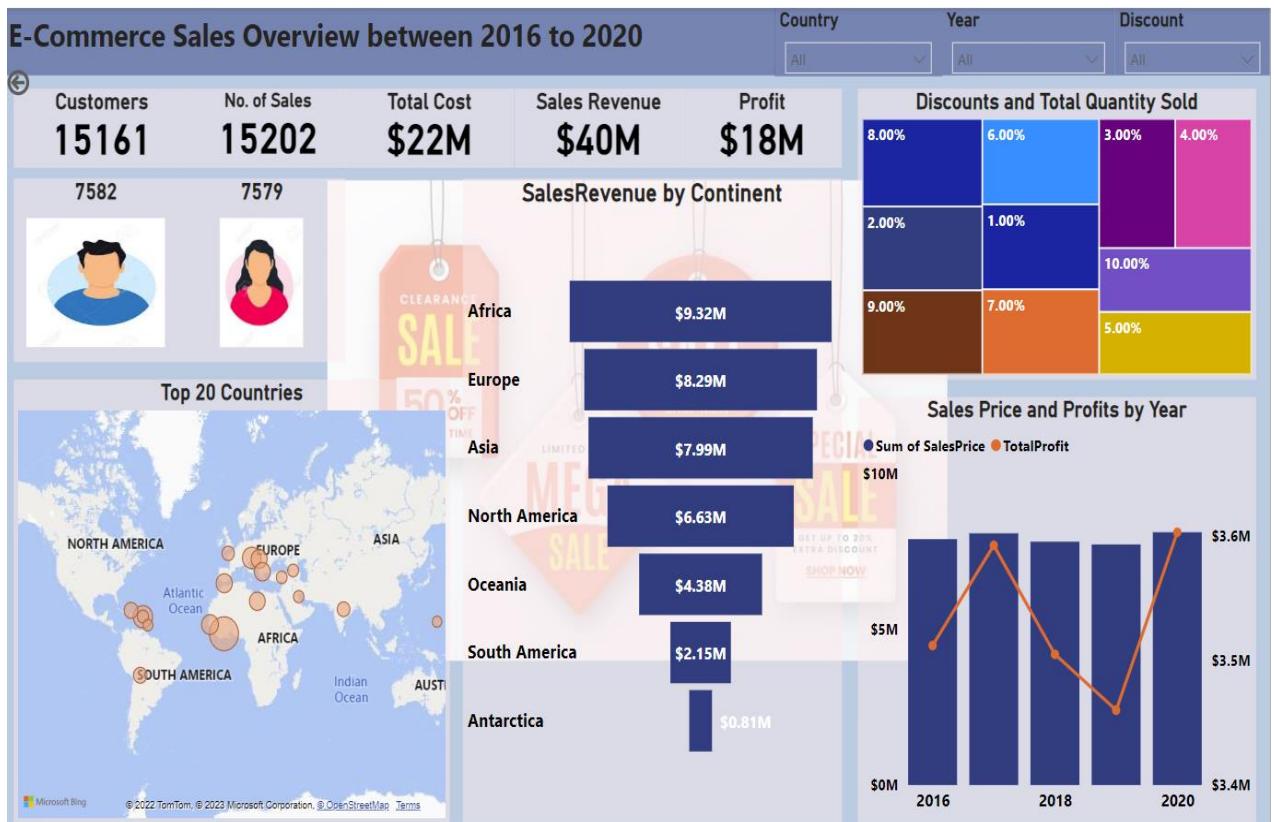


Fig.4. Sales Overview

This dashboard was created to give a concise summary of the dataset, giving the analysis of the customers, genders, number of sales, the ranges of the discounts implemented on each purchase, best performing geographical locations, hierarchical representation of the 7 regions involved, line and bar chart for the total revenue and total profit, and all these happened between 2016 to 2020. At the top right corner are three slicers; country, year, and discounts for easy viewing of data and charts.

3.1.1 Card Visuals

Customers	No. of Sales	Total Cost	Sales Revenue	Profit
15161	15202	\$22M	\$40M	\$18M

Fig.5. Sales values

3.1.1.1 Customers

From the Power Bi report view, it was easy to get the number of customers by using the “count of Customer name”, this is necessary to know the number of people that made purchases. The “distinct count” wasn’t used because I observed during the pre-processing some customers have the same name but different gender.

3.1.1.2 Number of Sales

A good business analyst will like to know the number of sales made. I calculated this by using a simple DAX formular;

```
TotalOrders = DISTINCTCOUNT('E-Commerce'[Sales id])
```

3.1.1.3 Total Cost

This was easily computed because we already have a column with “cost of goods”. Using DAX;

```
TotalCostofGoods = SUMX('E-Commerce_Sales_Facts', 'E-Commerce_Sales_Facts'[CostOfGood] * 'E-Commerce_Sales_Facts'[QuantitySold])
```

SUMX will add up the individual values after the multiplication.

3.1.1.4 Sales Revenue

This has to be calculated to know how well your business is performing. Since I already had the sales price, all I needed to do is to SUM them up.

3.1.1.5 Profit

Calculating the profit let us know how well the business is doing. Using DAX;

```
TotalProfit = 'E-Commerce_Sales_Facts'[SalesRevenue] - 'E-Commerce_Sales_Facts'[TotalCostofGoods]
```

3.1.2 Gender Infographic Design



Fig.6. Gender Infographic

Using the Power Bi Infographic designer, I uploaded a male and female image, did a count on “CustomerGender” to get their individual values. This also informs us that our customer targets are both male and female.

3.1.3 Discounts Values using Treemap

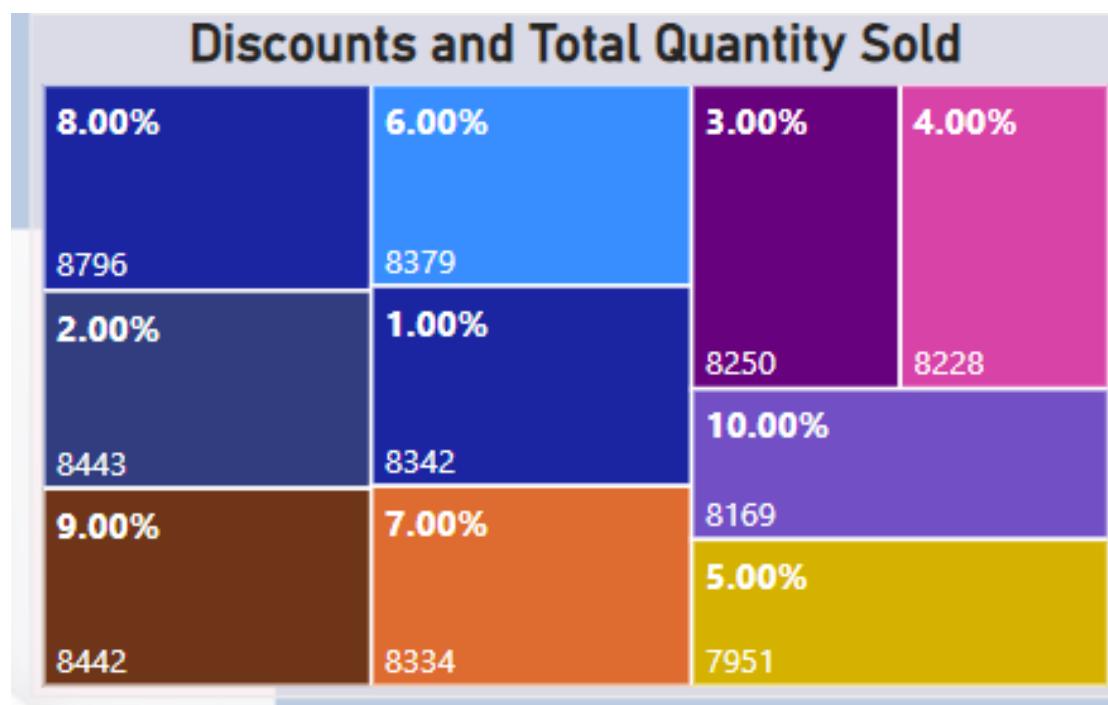


Fig.7.Range of Discounts

Treemap is a hierarchical chart, I used this to show the discount values with the number of quantity of products sold. The chart is visible and precise, plotted by placing the “discountperquantity” on the Category and “Sum of quantitysold” on the values. 8% discount has the highest number of products sold.

3.1.4 Hierarchical Representation of the Regions



Fig.8.Regions

The Power Bi Funnel Chart is great for hierarchical representations. Looking at the above chart will let you know that Africa has the highest sales revenue, this should be one of the target regions. The chart was created by placing “CustomerContinent” on the Category and “SalesRevenue” on the Values.

3.1.5 Top 10 Countries and Geographical Locations



Fig.9.Geographical Locations

The above chart shows the top 10 with the highest sales revenue, precisely looking at Ivory Coast as the highest African revenue generator in the eCommerce dataset, one of the areas to target. The chart was made by placing “CustomerCountry” on the location, “SalesRevenue” on the Bubble Size, “count of SalesPrice” and “Sum of QuantitySold” on the Tooltips.

3.1.6 The Line and Bar Chart for Sales and Profits

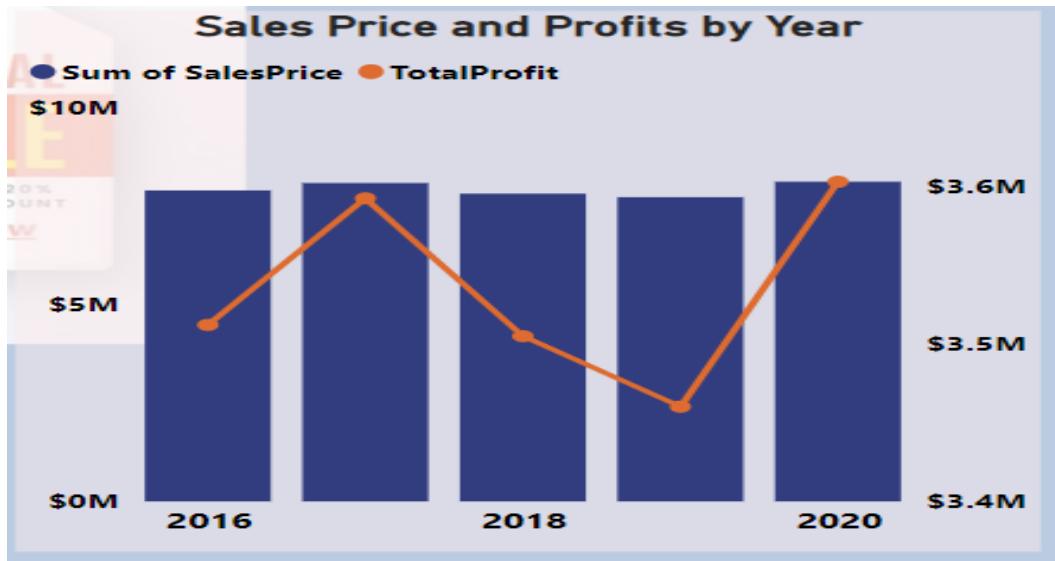
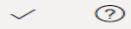


Fig.10. Sales Price and Profits by Year

This is one of the most important charts in this dashboard because it shows us the summary sales revenue and total profits over the years. This chart was created by placing the “Date(year)” on the X-axis, “sum of salesprice” on the Column Y-axis, and the “TotalProfit” on the Line Y-axis. We could see that there is a sharp decrease in total profit in 2019, we can use Power Bi to analyse this decrease by right clicking on the line at the point of decrease and click on analyse. Below charts are the analysis by Power Bi base on the dataset.

Here's the analysis of the 1.28% decrease in TotalProfit between 2018 and 2019



TotalProfit for 2018 and TotalProfit for 2019 BY QTR

'Qtr 1' had the most significant decrease among Qtr.

● TotalProfit for 2018 ● TotalProfit for 2019

\$1.0M

\$0.5M

\$0.0M

Qtr 1

Qtr 2

Qtr 3

Qtr 4

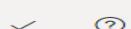
Qtr



TotalProfit for 2018 and TotalProfit for 2019

Fig.11. Power Bi Analysis1

Here's the analysis of the 1.28% decrease in TotalProfit between 2018 and 2019



TotalProfit for 2018 and TotalProfit for 2019 BY CUSTOMERCONTINENT

'Asia' had the most significant decrease among CustomerContinent.

● TotalProfit for 2018 ● TotalProfit for 2019

\$1.0M

\$0.5M

\$0.0M

Africa

Asia

Europe

North America

Oceania

South America

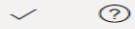
CustomerContinent



TotalProfit for 2018 and TotalProfit for 2019

Fig.12. Power Bi Analysis2

Here's the analysis of the 1.28% decrease in TotalProfit between 2018 and 2019

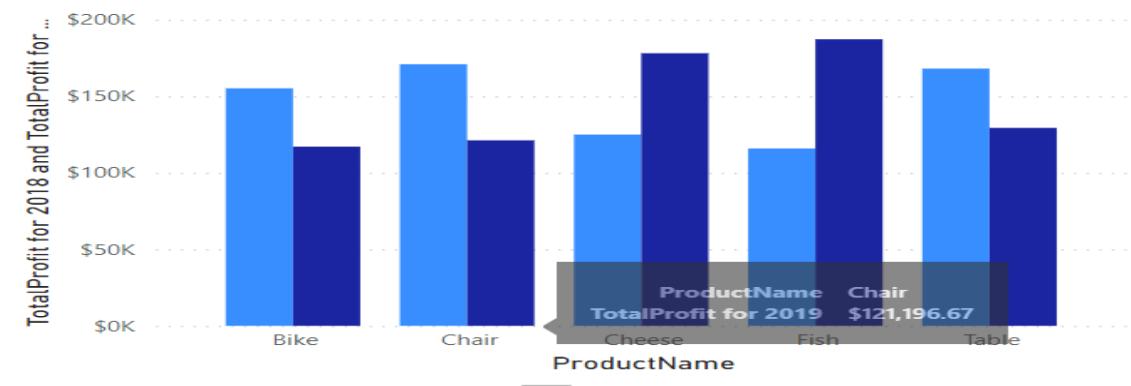


TotalProfit for 2018 and TotalProfit for 2019 BY PRODUCTNAME



'Chair', 'Table', and 'Bike', among others, had the most significant decrease among ProductName .

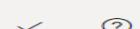
● TotalProfit for 2018 ● TotalProfit for 2019



Show largest changes ⓘ

Fig.13. Power Bi Analysis 3

Here's the analysis of the 1.28% decrease in TotalProfit between 2018 and 2019



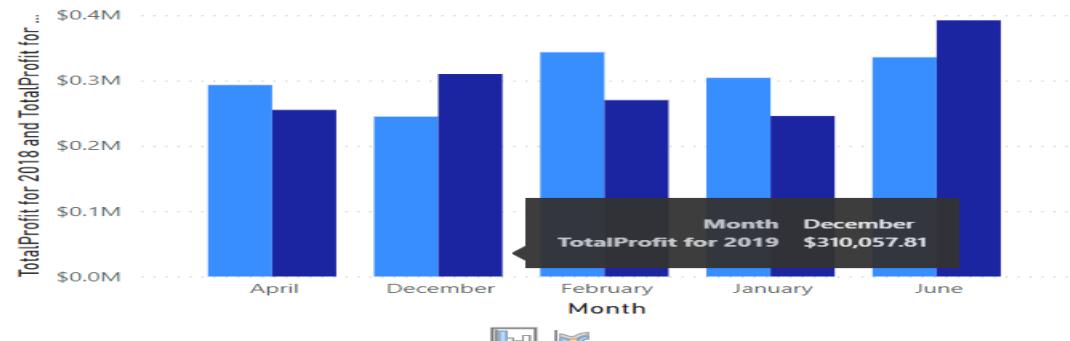
Show largest changes ⓘ

TotalProfit for 2018 and TotalProfit for 2019 BY MONTH



'February', 'January', and 'April', among others, had the most significant decrease among Month .

● TotalProfit for 2018 ● TotalProfit for 2019



Show largest changes ⓘ

Fig.14. Power Bi Analysis 4

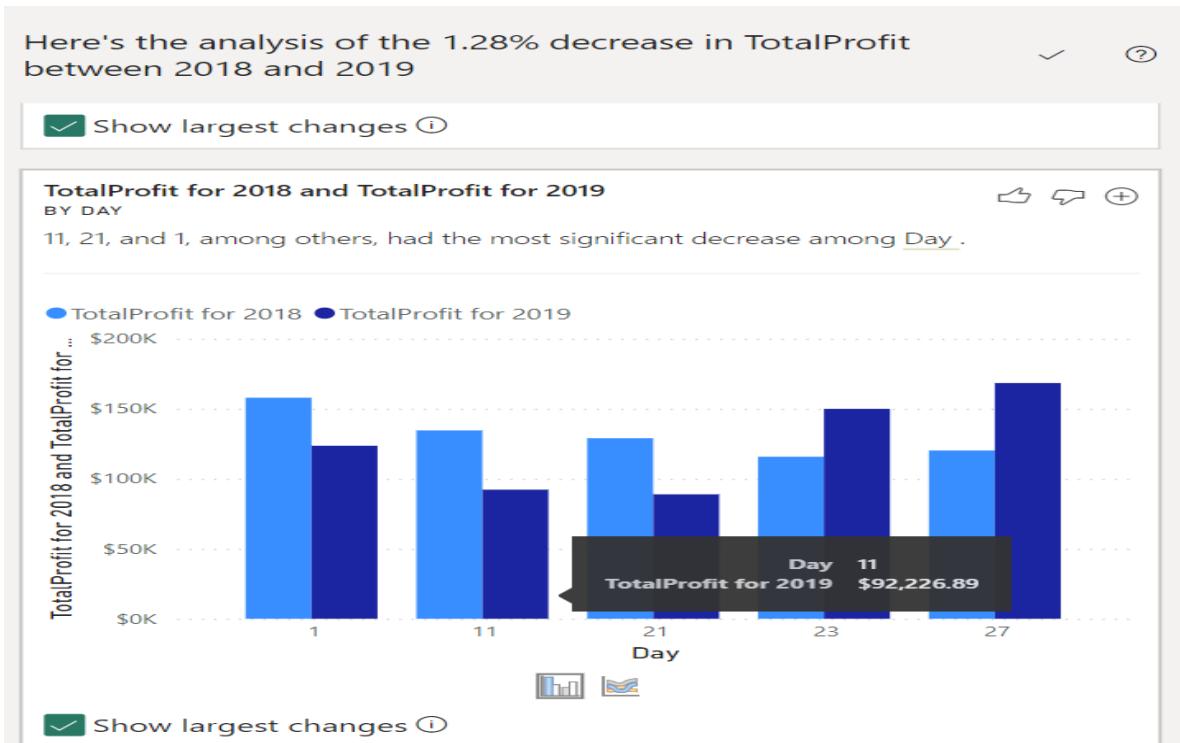


Fig.14. Power Bi Analysis 5

The above 5 charts are the Power Bi analysis based on the dataset, giving reasons for the sharp dip, we can also do the same for the sharp increase in 2020.

3.2 Sales Performance

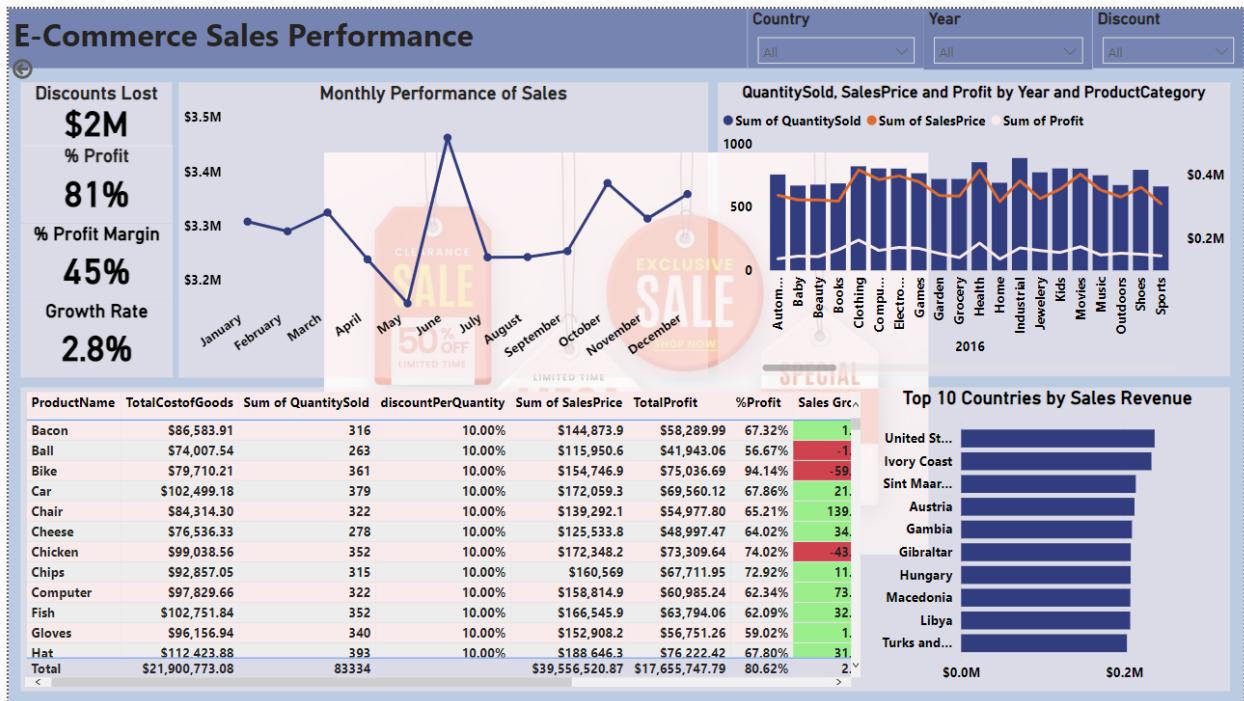


Fig.15. Sales Performance

This dashboard helps to look into a dataset and see what really happened over the years on the monthly by year, sales made, type of products sold, quantity sold, the most performing country, losses, profits, and growth rate. The power of this dashboard lies mostly on the three slicers at the top right corner; the country, year, and discounts.

3.2.1 Card Visuals

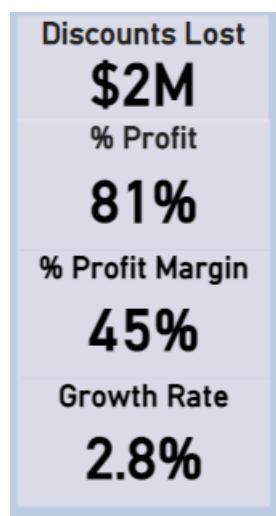


Fig.16. Lost, Profit, and Growth

3.2.1.1 Discounts Lost

As a business analyst, before implementing discounts, create revenue to track and monitor losses incurred due to sales discounts. This is a good way to make sure the business does not fail. Before using the card to show the values, I used DAX to calculate a measure for the value;

```
AmtLostByDiscount = SUMX('E-Commerce_Sales_Facts', 'E-Commerce_Sales_Facts'[SalesPriceWithoutDiscount] - 'E-Commerce_Sales_Facts'[SalesPrice])
```

Observing the value and calculating the ratio by sales revenue to be 5.8%, which is less than 10% shows that the discounts were well implemented.

3.2.1.2 Percentage Profit

81% profit shows that lots of profits were made even with the discounts. This can actually imply that the discounts increased demand and revenue. Using DAX to get the value;

```
%Profit = DIVIDE('E-Commerce_Sales_Facts'[TotalProfit], 'E-Commerce_Sales_Facts'[TotalCostofGoods], 0)
```

3.2.1.3 Percentage Profit Margin

This value is used to gauge how profits are made in businesses. It allows the analyst to see how the business is performing. Using DAX to calculate the measure;

```
%ProfitMargin = DIVIDE('E-Commerce_Sales_Facts'[TotalProfit], 'E-Commerce_Sales_Facts'[SalesRevenue], 0)
```

45% profit margin shows the business is making good profit.

3.2.1.4 Growth Rate

2.8% growth rate shows that there has been growth in sales over the period of 2016 to 2020. Using DAX;

```
1 Sales_Growth =
2 var FirstYear = CALCULATE(MIN('Date Dim'[Year]))
3 var LatestYear = CALCULATE(MAX('Date Dim'[Year]))
4 var GDPFirstYear = CALCULATE('E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = FirstYear)
5 var GDPLatestYear = CALCULATE('E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = LatestYear)
6 var Diff = GDPLatestYear - GDPFirstYear
7 return DIVIDE(Diff, GDPFirstYear, 0)
```

By utilizing the discount slicers on the dashboard, 4%, 5%, 8%, 10% showed negative growth rates.

3.2.2 Monthly Performance of Sales



Fig.17. Monthly Performance

The line graph gives precise trends on the performance of the sales, it can be drilled from years to quarters, months, and days. The monthly view can also be regulated using the slicers on the dashboard. There is a steep between March to May, but a sharp one follows it from May to June. These months; February to March, May to June, September to October, and November to December maintained an average increase over the years. The chart was created by placing the “date” at the X-axis and “sum of the salesprice” at the Y-axis.

We can analyse the sharp incline between May to June using Power Bi analytic tools.

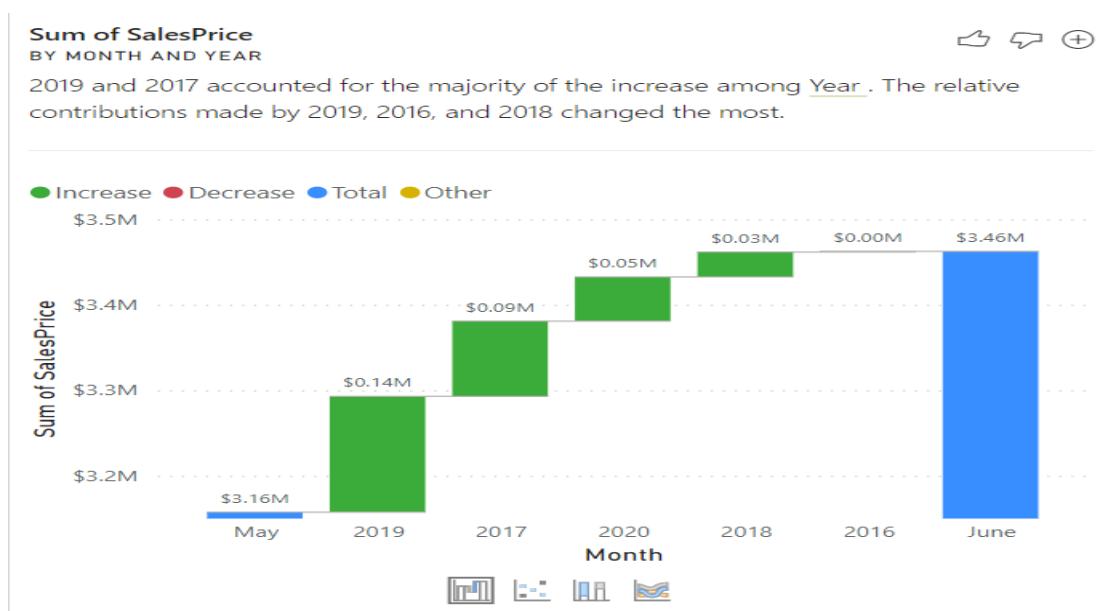


Fig.18. Power Bi Analysis

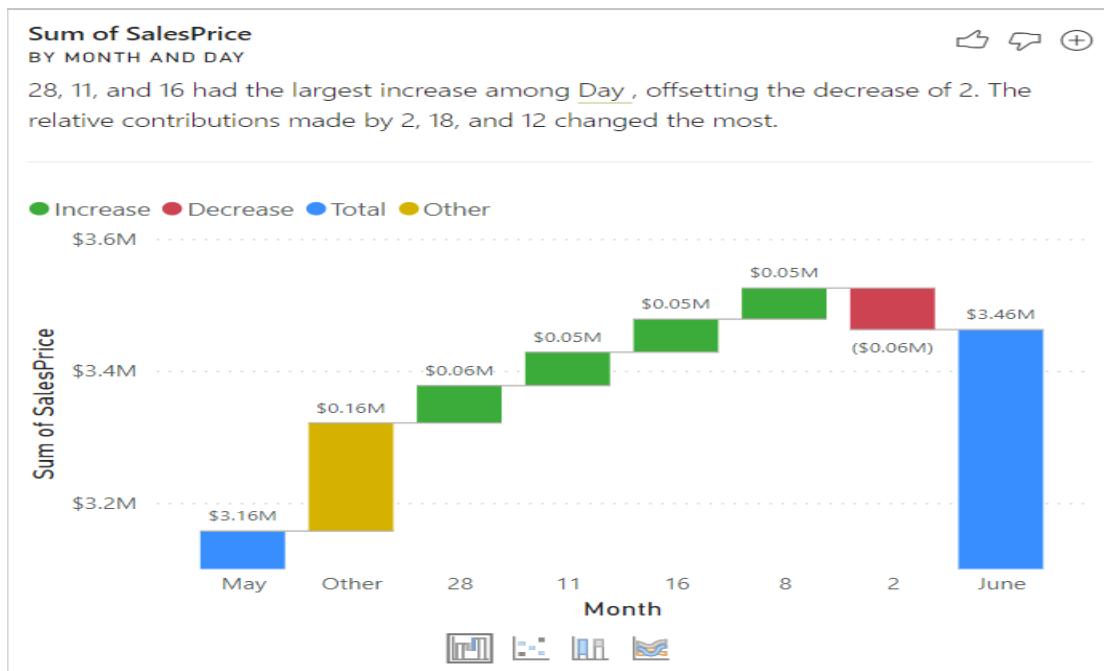


Fig.19. Power Bi Analysis

Power Bi mentions the specific days, months, years and gives reasons why the fluctuations occurred.

3.2.3 Product Category Analysis

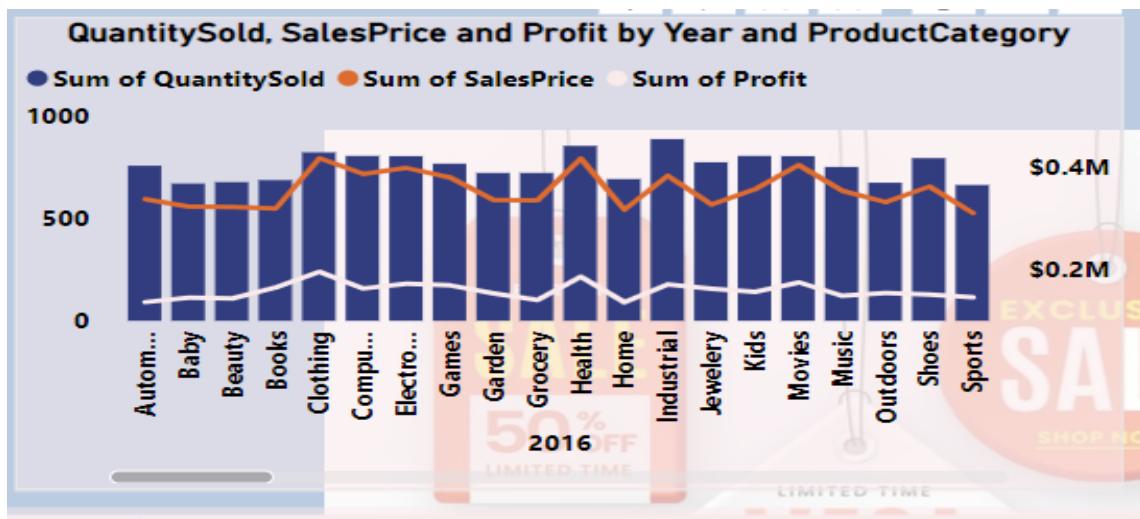


Fig.20. Product Category Analysis

Line and clustered column charts allow you to compare two sets of measures. The chart above shows the product category performance between 2016 to 2020 including the total quantity

sold, total sales, and total profit. We can view the product category with the highest performance for each country, year, and even by the specific discount.

The Industrial category had the highest number of quantities sold by an average of 839.6. Between 2016 to 2020, the highest revenue was by product category Grocery in 2020 by \$470,314.18.

The above was created by placing “Date” and “ProductCategory” on the X-axis, “sum of quantitysold” on column Y-axis, “sum of salesprice” and “sum of profit” on Line Y-axis.

3.2.4 Top 10 Countries by Sales Revenue



Fig.21. Top 10 Countries

A bar chart gives a numerical representation of categorical features. The graph was plotted by placing the “CustomerCountry” on the Y-axis, “SalesRevenue” on X-axis, “Quantitysold” on the Tooltips.

The US had the highest sales revenue of \$237,052.2 and 439 quantities of products.

3.2.5 Summary Table

ProductName	TotalCostofGoods	Sum of QuantitySold	discountPerQuantity	Sum of SalesPrice	TotalProfit	%Profit	Sales Growth	%ProfitMargin	AmtLostByDiscount	%AmtLostByDiscount
Pants	\$138,587.30	499	6.00%	\$249,032.32	\$110,445.02	79.69%	79.91%	44.35%	\$15,895.68	6.38
Towels	\$136,291.61	475	7.00%	\$240,423.6	\$104,131.99	76.40%	107.58%	43.31%	\$18,096.40	7.53
Sausages	\$119,398.78	474	8.00%	\$210,576.04	\$91,177.26	76.36%	-34.16%	43.30%	\$18,310.96	8.70
Cheese	\$129,992.16	471	4.00%	\$238,426.56	\$108,434.40	83.42%	-46.89%	45.48%	\$9,934.44	4.17
Pizza	\$121,894.27	469	7.00%	\$205,555.11	\$83,660.84	68.63%	-6.56%	40.70%	\$15,471.89	7.53
Shoes	\$123,495.66	459	7.00%	\$213,364.32	\$89,868.66	72.77%	-43.71%	42.12%	\$16,059.68	7.53
Pants	\$114,488.70	445	4.00%	\$201,781.44	\$87,292.74	76.25%	151.21%	43.26%	\$8,407.56	4.17
Sausages	\$106,761.26	444	2.00%	\$200,588.36	\$93,827.10	87.88%	-18.84%	46.78%	\$4,093.64	2.04
Cheese	\$109,133.14	441	8.00%	\$192,314.04	\$83,180.90	76.22%	-8.44%	43.25%	\$16,722.96	8.70
Sausages	\$111,562.72	437	10.00%	\$204,902.1	\$93,339.38	83.67%	28.25%	45.55%	\$22,766.90	11.11
Mouse	\$117,128.23	436	3.00%	\$205,255.88	\$88,127.65	75.24%	-34.65%	42.94%	\$6,348.12	3.09
Gloves	\$132,939.30	432	2.00%	\$245,861.42	\$112,922.12	84.94%	-24.83%	45.93%	\$5,017.58	2.04
Total	\$21,900,773.08	83334		\$39,556,520.87	\$17,655,747.79	80.62%	2.80%	44.63%	\$2,310,064.13	5.84

Fig.22. Summary Table

This table summarizes the sales for each of the products, cost, quantity, discounts, profits, and sales growth. I applied advanced conditioning to the “Sales Growth”, green ones are sales growth equal to or above 0%.

Background color - Sales Growth

Format style Apply to

Rules Values only

What field should we base this on?

Sales Growth Reverse color order + New rule

Rules

If value <input type="text" value="≥"/> <input type="text" value="0"/> <input type="button" value="Number"/> and <input type="text" value="≤"/> <input type="text" value="100"/> <input type="button" value="Percent"/> then <input type="color" value="green"/>	<input type="button" value="↑"/> <input type="button" value="↓"/> <input type="button" value="×"/>
If value <input type="text" value="≥"/> <input type="text" value="0"/> <input type="button" value="Percent"/> and <input type="text" value="≤"/> <input type="text" value="0"/> <input type="button" value="Number"/> then <input type="color" value="red"/>	<input type="button" value="↑"/> <input type="button" value="↓"/> <input type="button" value="×"/>

[Learn more about conditional formatting](#) OK Cancel

Fig.23. Conditional Formatting

3.3 Sales Growth



Fig.24. Sales Growth

The waterfall and stacked area charts are good visuals for the representation of sales growth and changes that occur over time. A waterfall chart shows a running total as values are added or subtracted. It's useful for understanding how sales revenue is affected over the years by discounts or quantity of products.

The stacked area chart can be used to show the trends over time. The above charts can easily be used by utilizing the text filter for the countries and the two slicers for the year and discounts.

3.3.1 The Waterfall Chart

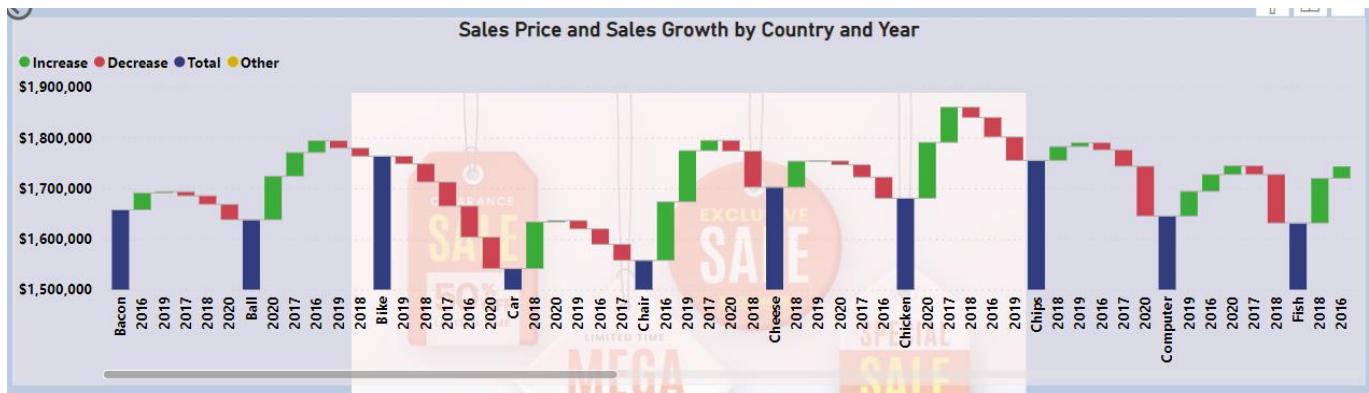


Fig.25. Sales Price and Sales Growth

The chart was plotted by placing “ProductName” on Category, “Date(year)” on Breakdown, “sum of SalesPrice” on Y-Axis.

By clicking on the year slicer, you will see the the changes in sales price for a given product. In 2018, Pizza had the highest negative change of -30% on sales price compared to other products.



Fig.26. Pizza Negative change

3.3.2 Stacked Area Chart



Fig.27. Trends of Sales Price by Year

The chart was plotted by placing the “Date” on the X-axis and “Sum of sales price” on the Y-axis. A lot of drill down can be done on this chart from days to years. We can use the Power Bi analytics to highlight the steepest increase in 2017.

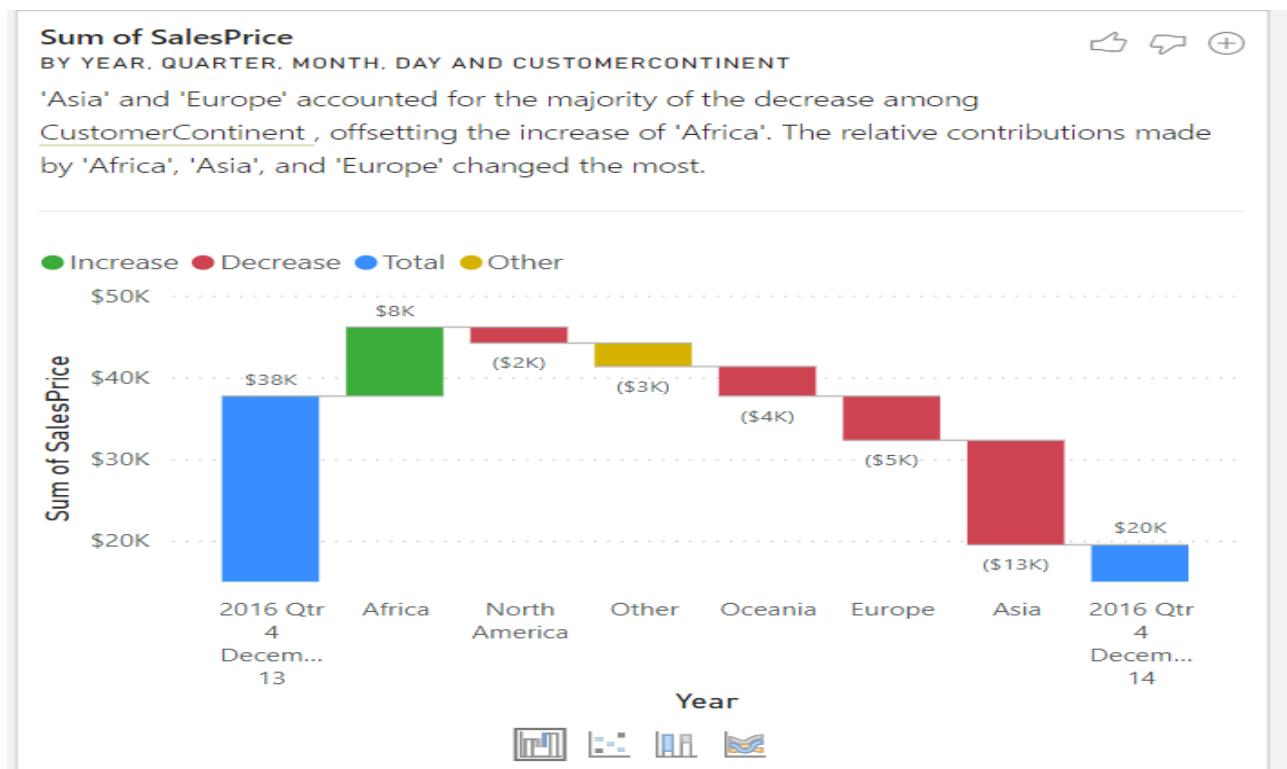


Fig.28. Power Bi Analysis

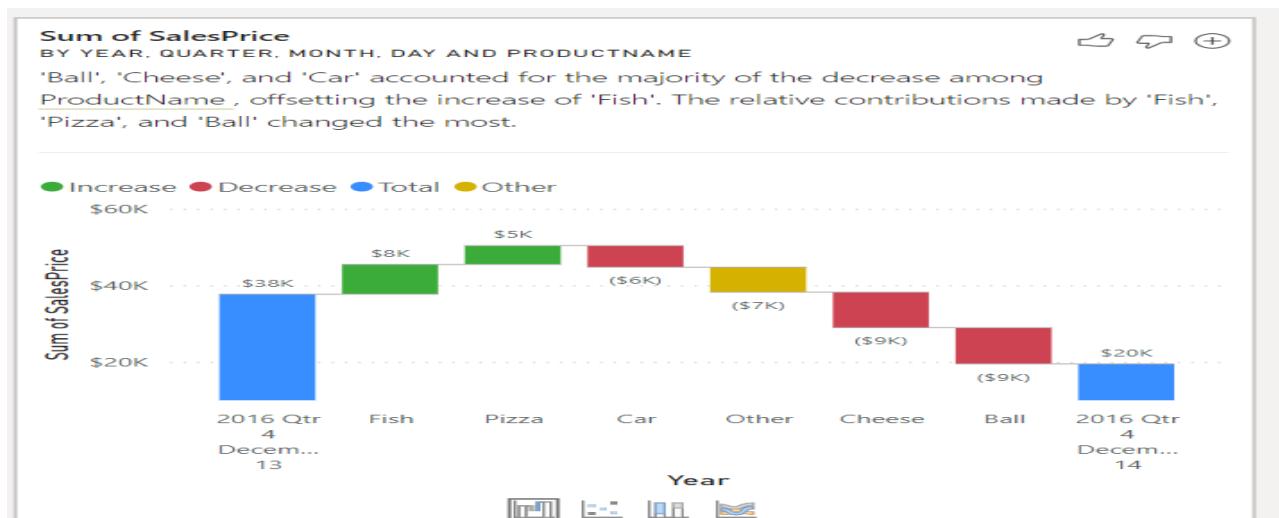


Fig.29. Power Bi Analysis

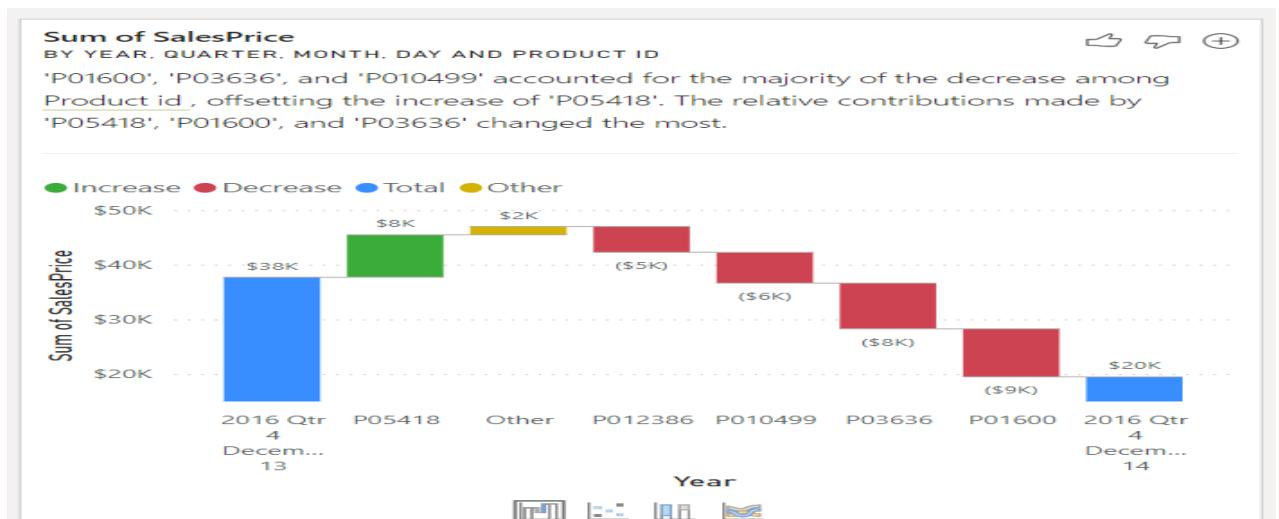


Fig.30. Power Bi Analysis

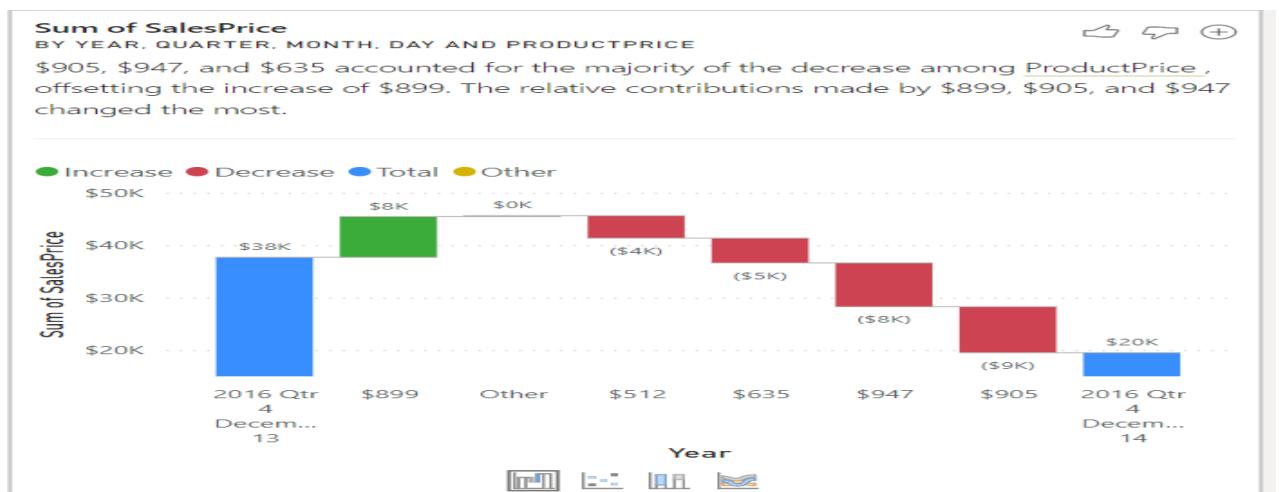


Fig.31. Power Bi Analysis

3.4 Key Influencers

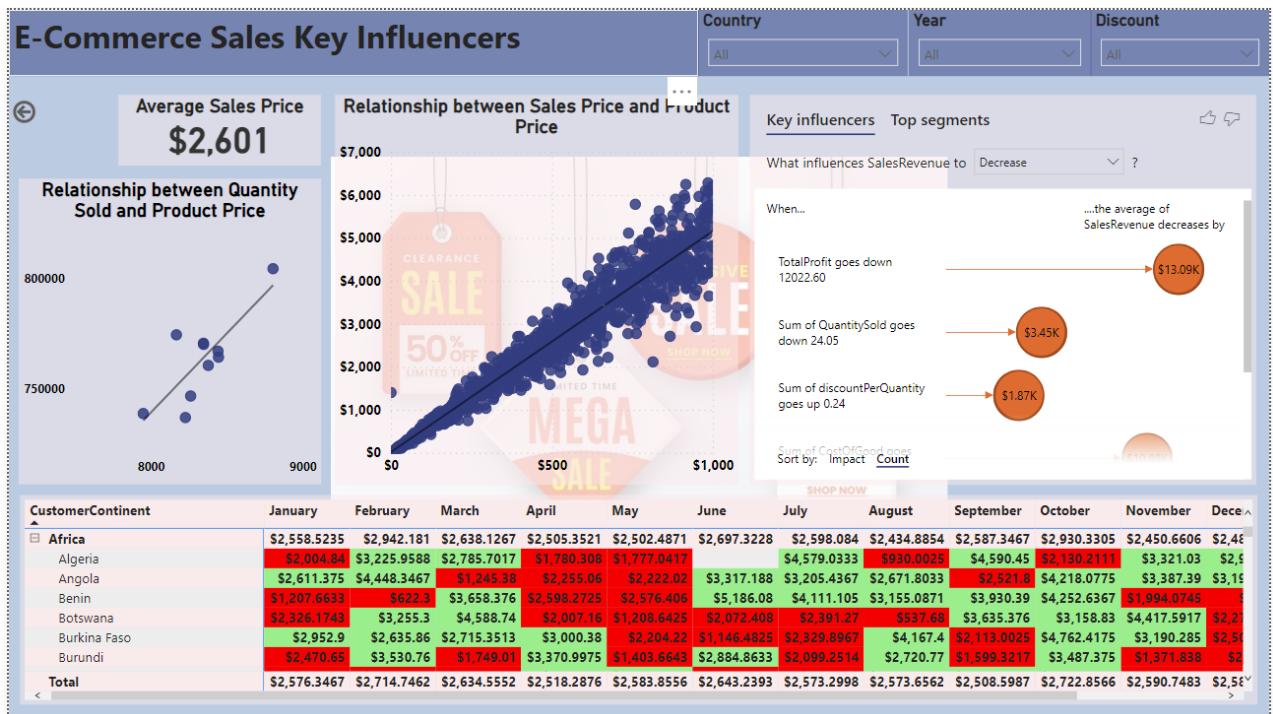


Fig.31. Key Influencers

The dashboard above seeks to find and illustrate key influencers in the eCommerce dataset. This will enable us to get more insights into what will affect future sales.

The three slicers above make the dashboard easier to access and navigate through.

3.4.1 Card Visuals



Fig.32. Average Sales Price

The value was easy to calculate using the card on the report view of the Power Bi dashboard. The average is needed for the advanced formatting done on the matrix table. This calculated by taking the average of the total sales by all the country.

3.4.2 Relationship between Quantity Sold and Product Price



Fig.33. Quantity Sold and Product Price

One of the questions asked was, does discount increase demand? The chart above is a scatter plot used to depict correlations or relationships among variables. The chart plot by placing “discountperquantity” on values, “sum of QuantitySold” on the X-axis, and “sum of ProductPrice” on Y-axis. The chart shows a linear correlation between quantity sold and product price over discounts. Basically, product price will increase if demand increases, this could be the same when discounts are implemented.

3.4.3 Relationship between Sales Price and Product Price



Fig.34. Sales Price and Product Price

The chart above was plotted directly by placing the product price in the X-axis and sales price on the Y-axis. The relationship is linear, which implies that as the product price increase, the sales price increases.

3.4.4 Key Influencers Chart

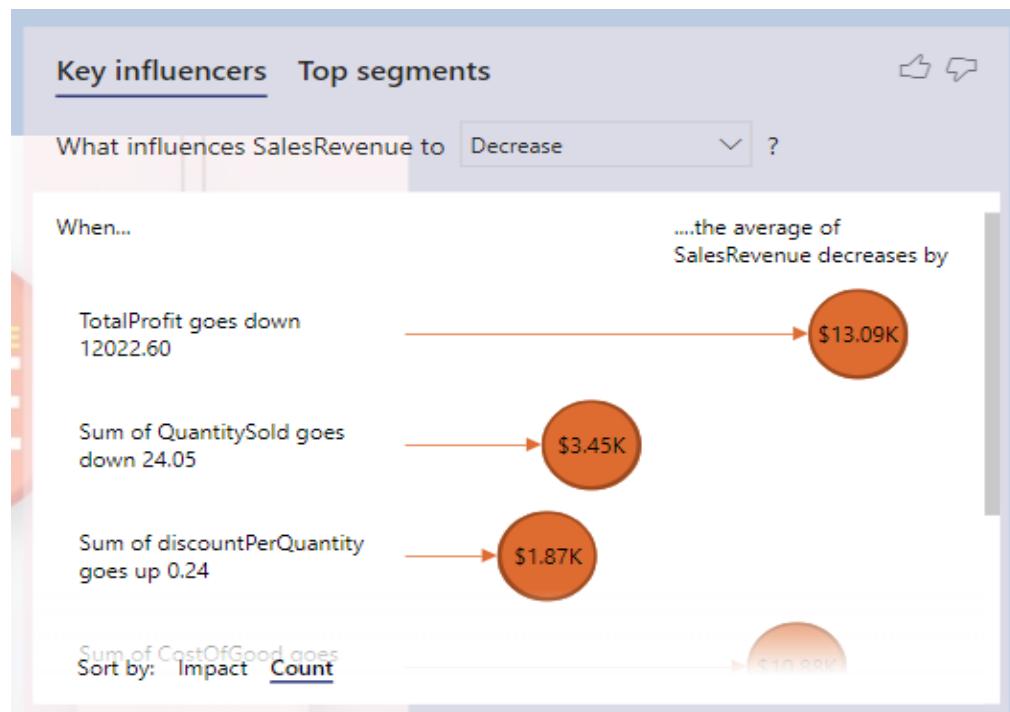


Fig.35. Key Influencers

Key Influencers is a good Power BI analytic tool, can be used to show the major contributors on the measure you want to analyse, it can be done both at the increase or decrease. The chart was created to analyse the “SalesRevenue” by “ProductPrice”, “sum of CostofGoods”, “sum of QuantitySold”, “TotalProfit”, and “sum of discountperquantity”. Below are some charts created and the analyses.

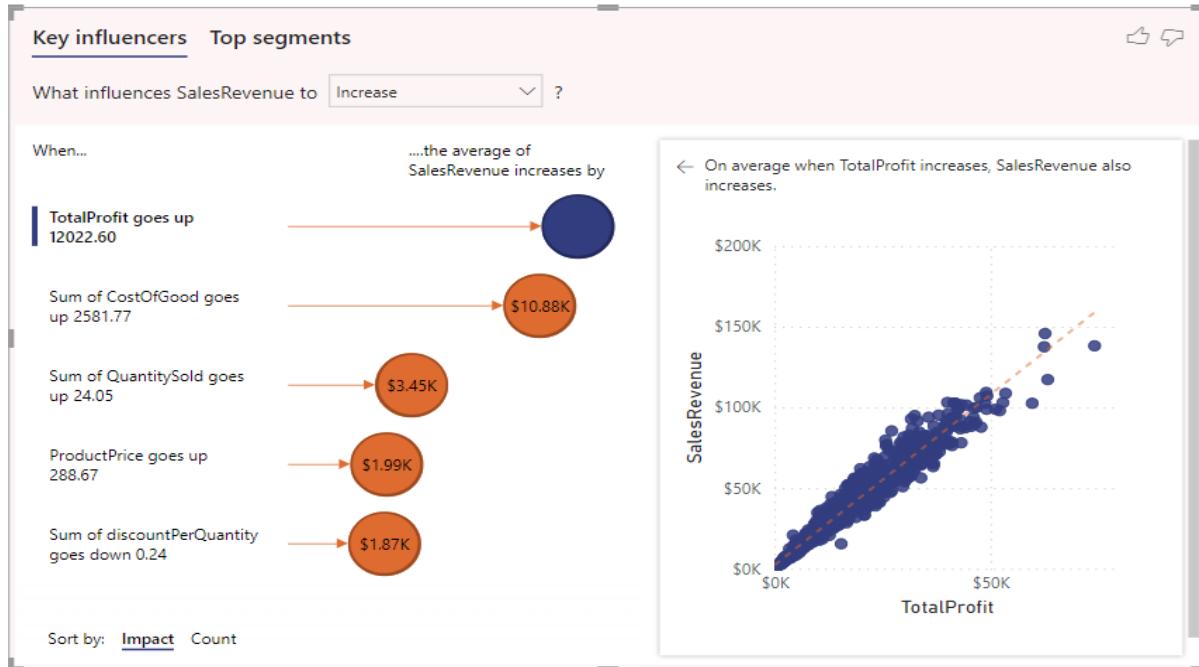


Fig.36. Key Influencers Analysis

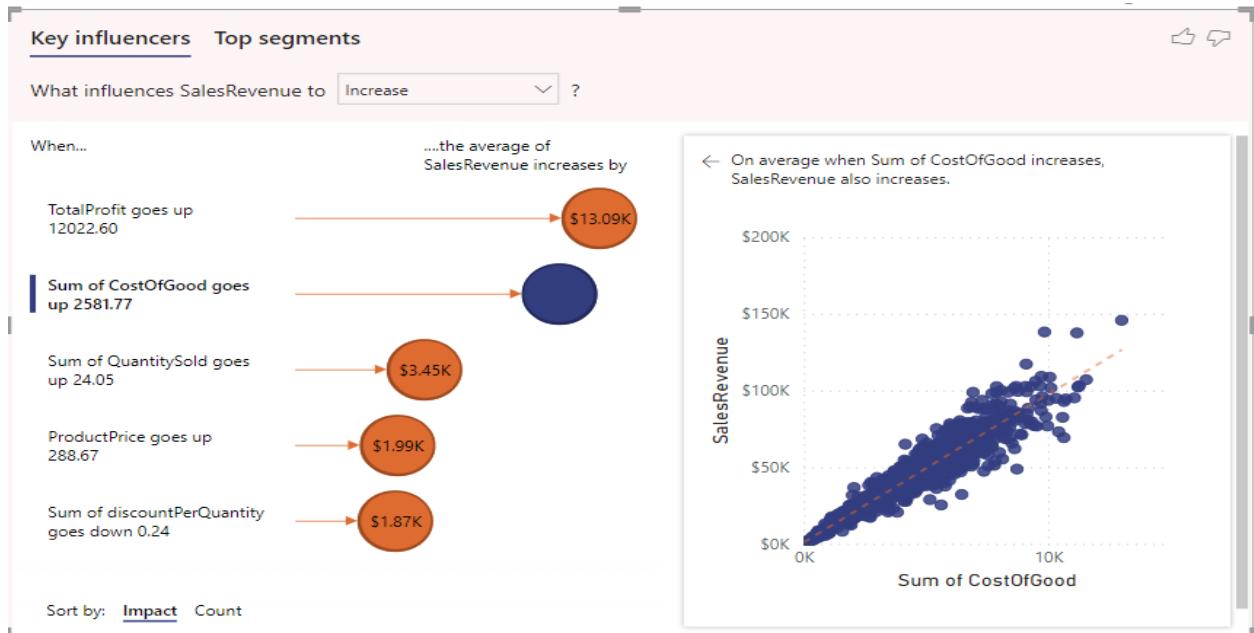


Fig.37. Key Influencers Analysis



Fig.38. Key Influencers Analysis



Fig.39. Key Influencers Analysis

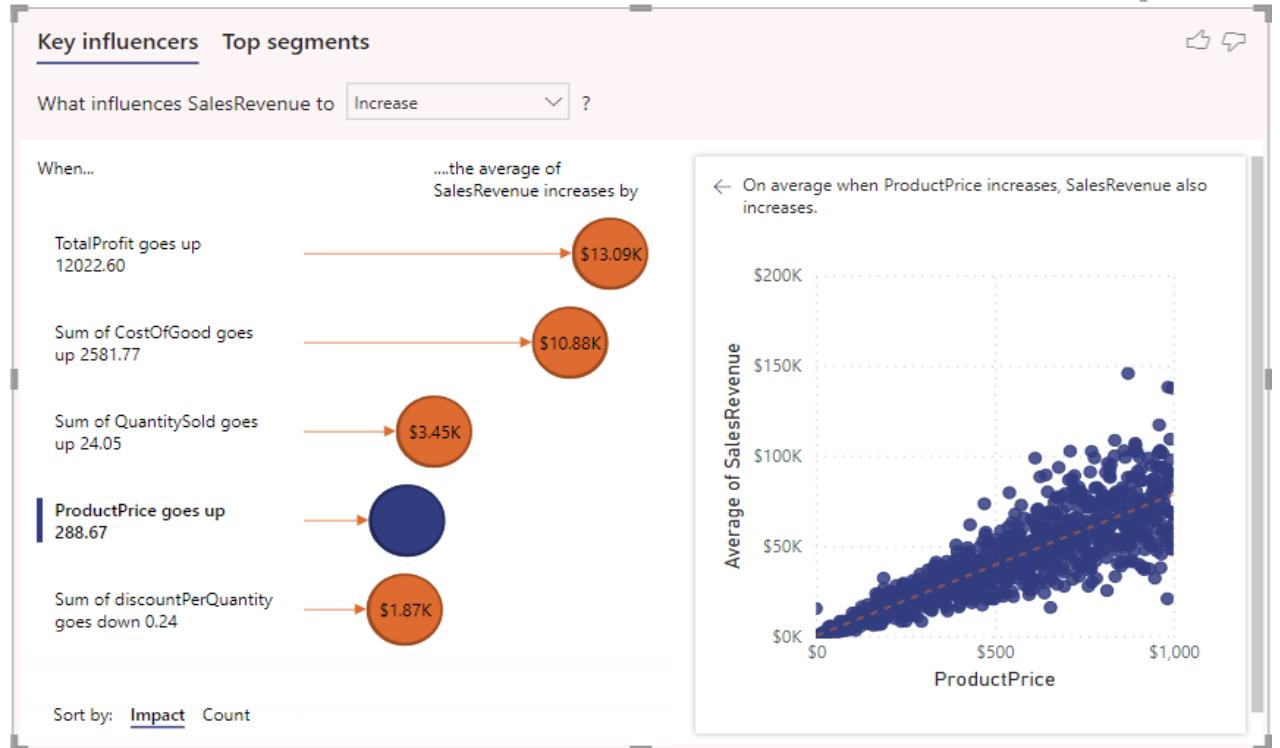


Fig.40. Key Influencers Analysis

3.4.5 Matrix table

CustomerContinent	January	February	March	April	May	June	July	August	September	October	November	December
	\$2,558.5235	\$2,942.181	\$2,638.1267	\$2,505.3521	\$2,502.4871	\$2,697.3228	\$2,598.084	\$2,434.8854	\$2,587.3467	\$2,930.3305	\$2,450.6606	\$2,481.1000
Africa	\$2,004.84	\$3,225.9588	\$2,785.7017	\$1,780.308	\$1,777.0417	\$4,579.0333	\$930.0025	\$4,590.45	\$2,130.2111	\$3,321.03	\$2,9	\$2,481.1000
Algeria	\$2,611.375	\$4,448.3467	\$1,245.38	\$2,255.06	\$2,222.02	\$3,317.188	\$3,205.4367	\$2,671.8033	\$2,521.8	\$4,218.0775	\$3,387.39	\$3,115.0000
Angola	\$1,207.6633	\$622.3	\$3,658.376	\$2,598.2725	\$2,576.406	\$5,186.08	\$4,111.105	\$3,155.0871	\$3,930.39	\$4,252.6367	\$1,994.0745	\$1,994.0745
Benin	\$2,326.1743	\$3,255.3	\$4,588.74	\$2,007.16	\$1,208.6425	\$2,072.408	\$2,391.27	\$537.68	\$3,635.376	\$3,158.83	\$4,417.5917	\$2,271.0000
Botswana	\$2,952.9	\$2,635.86	\$2,715.3513	\$3,000.38	\$2,204.22	\$1,146.4825	\$2,329.8967	\$4,167.4	\$2,113.0025	\$4,762.4175	\$3,190.285	\$2,500.0000
Burkina Faso	\$2,470.65	\$3,530.76	\$1,749.01	\$3,370.9975	\$1,403.6643	\$2,884.8633	\$2,099.2514	\$2,720.77	\$1,599.3217	\$3,487.375	\$1,371.838	\$2,500.0000
Burundi	\$2,576.3467	\$2,714.7462	\$2,634.5552	\$2,518.2876	\$2,583.8556	\$2,643.2393	\$2,573.2998	\$2,573.6562	\$2,508.5987	\$2,722.8566	\$2,590.7483	\$2,581.1000
Total	\$2,576.3467	\$2,714.7462	\$2,634.5552	\$2,518.2876	\$2,583.8556	\$2,643.2393	\$2,573.2998	\$2,573.6562	\$2,508.5987	\$2,722.8566	\$2,590.7483	\$2,581.1000

Fig.41. Average Sales by Country

The matrix table shows a concise analysis on how each country performed against the average sales price monthly. The green shows values above average.

The table was created by placing “CustomerContinent” and “CustomerCountry” on the Rows, “Date” on Columns, and “Average of SalesPrice” on the Values.

Few drill down were made to move from years to quarters and then to months, before formatting the values. To see the different year, use the year slicer on the dashboard.

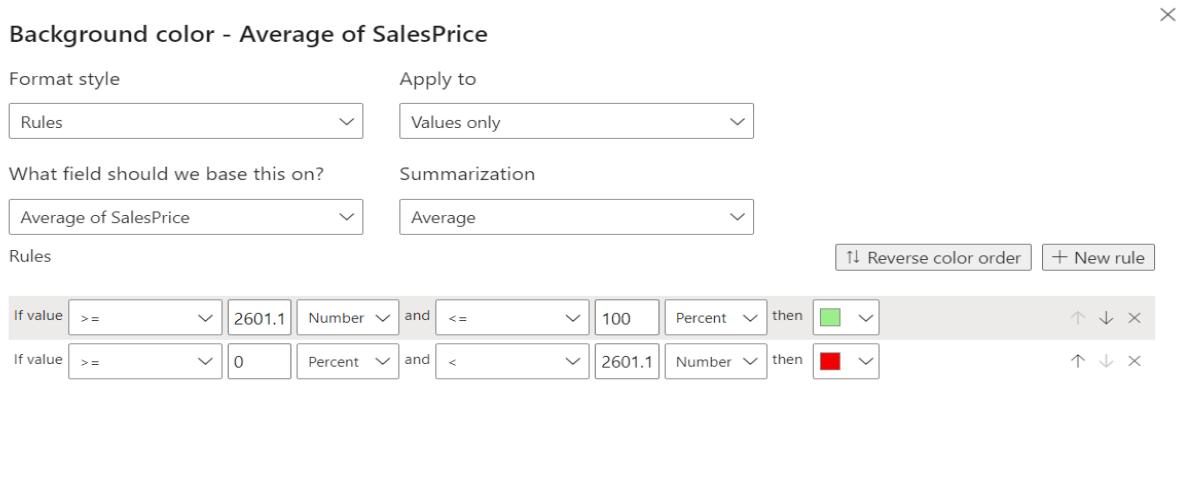


Fig.42. Advanced Formatting

3.5 Sales Prediction

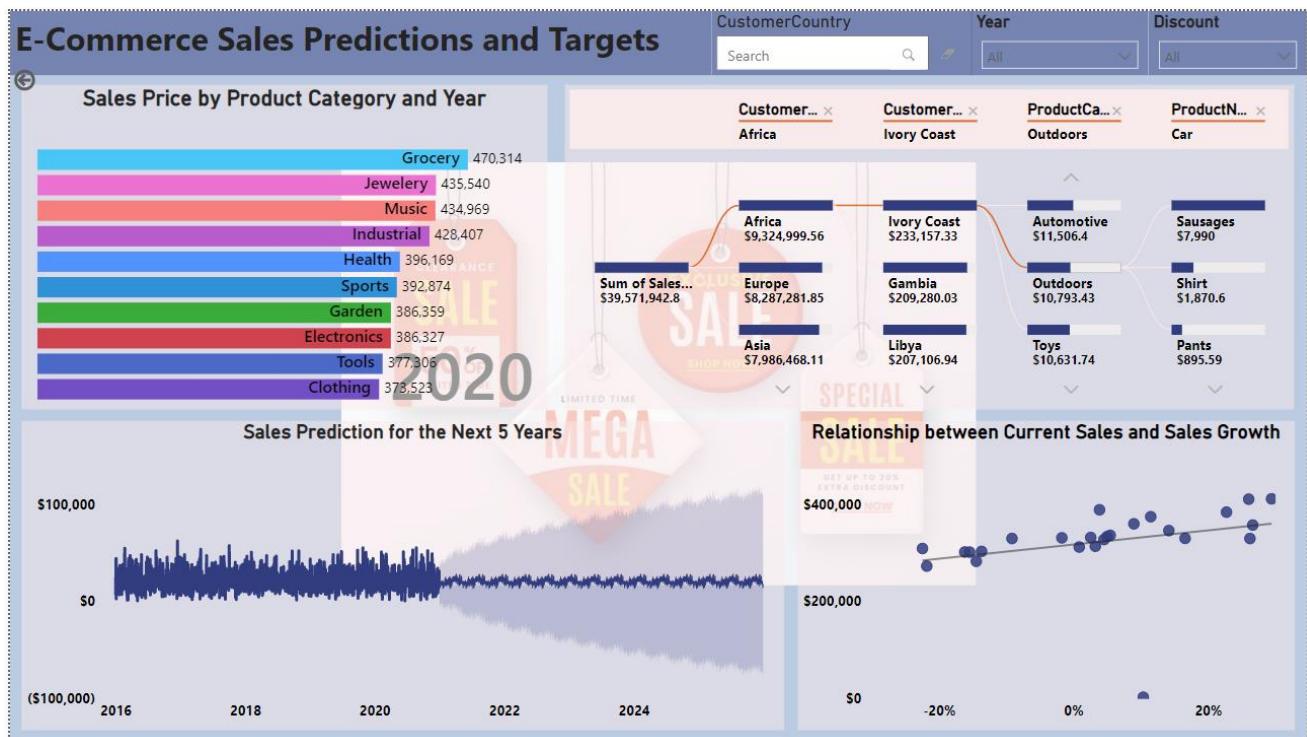


Fig.43. Sales Prediction

One of the major purposes of business analytics is to have good insights of your data, to enable you to make predictions using machine learning tools. A great addition to Power Bi is that it can make these predictions using analytic tools.

3.5.1 Sale Price and Product Category

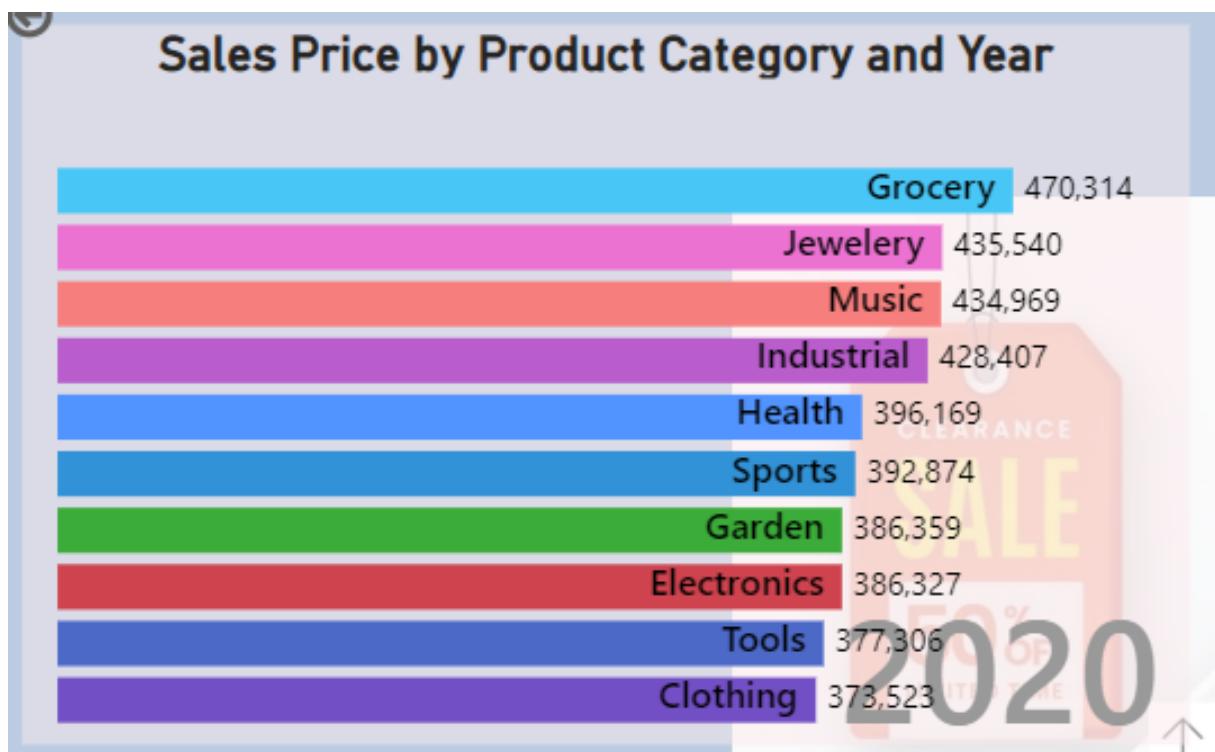


Fig.44. Animated Charts

The animated chart above displays the top 10 sales revenue by product category for a particular year. In 2016, Clothing had the highest revenue of \$417,195, Movies by \$446,172 in 2017, Industrial by \$424,396 in 2018, and 2019 Health at \$399,047. The chart above shows the current year 2020, Grocery by 470,314 which is the highest compared to other years. This could be a result of the Covid-19 Pandemic, movement were restricted, so people were focusing majorly on what to eat.

The chart was created by placing “ProductCategory” on Name, “Sum of SalesPrice” on Value, and “Date(year)” on Period.

3.5.2 Decomposition Tree.

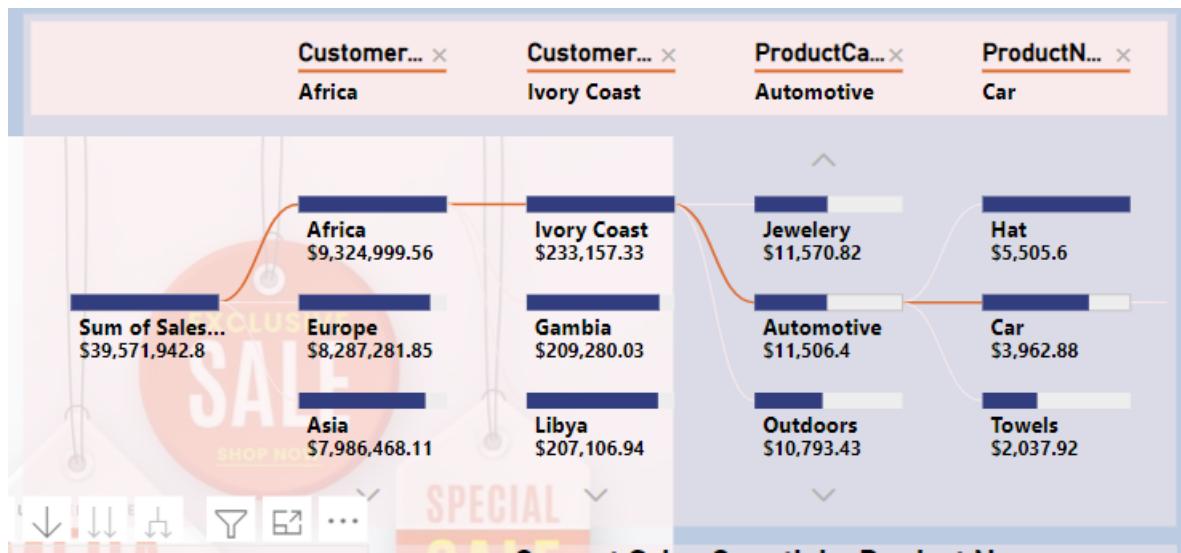


Fig.45. Decomposition Tree.

The chart was created by placing “Sum of SalesPrice” on Analyze, and others on Explain by. This will give you a long drill down of values to the list of the values. This can also be used for comparison when the Relative Analysis type is selected.

The decomposition tree helps you to get more insights about your data by allowing you visualize across multiple dimensions. This is really a great analytic tool for getting more insights about your data.

3.5.3 Sales Prediction for the Next 5 Years



Fig.46. Sales Predictions

The above chart is a line chart with predictive ability. This was created by plotting “Sum of SalesPrice” against “Date”, then turning the “Forecast” tab on and adjusting the Units and Forecast length.

Predictions were made for the next five years, sales price will increase with few fluctuations in the values though guarded by upper and lower bounds.

3.5.4 Relationship between Current Sales and Sales Growth

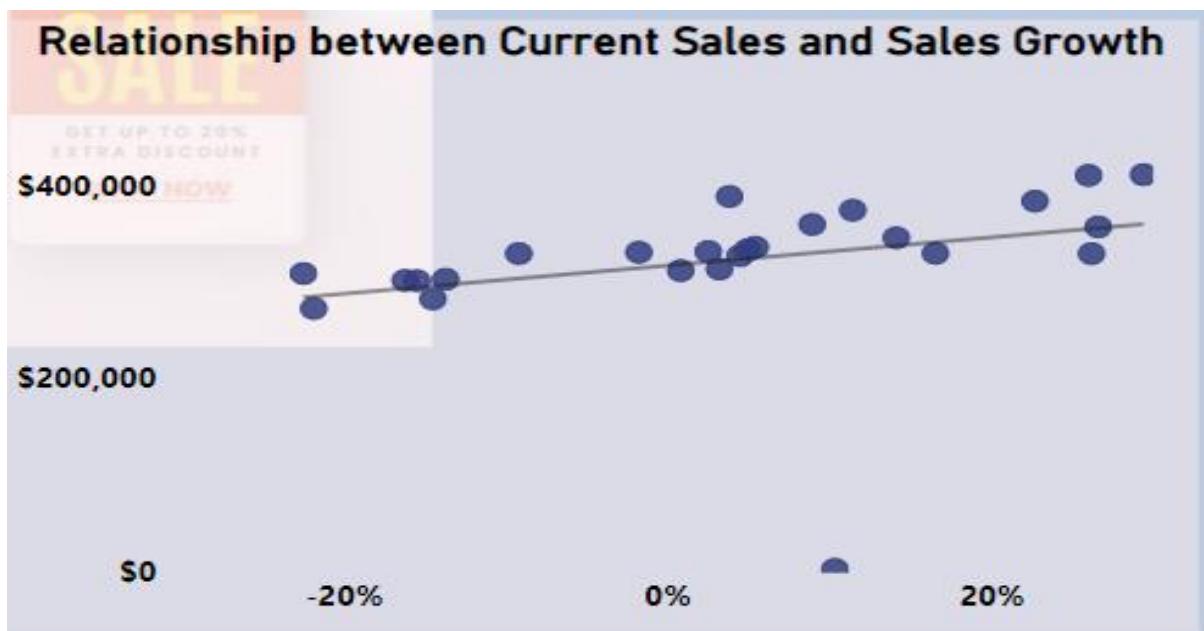


Fig.47. Current Sales vs Sales Growth

The chart was created to visualise the correlation between current sales and sales growth, this could help to make a good forecast of sales for the next year.

The chart has a gradual linear relationship between sales price and sales growth.

4 Summary

- Products with 8% discounts have the highest quantity sold with sales revenue of about 4 million dollars.
- The year 2019 has the lowest profit margin(Total Profit / Sales Revenue) compared to other years.

- There was a rapid increase in total profit between 2019 to 2020.
- No specific month retailed a steady increase of sales annually between 2016 to 2020.
- The Industrial category had the highest number of quantities sold by an average of 839.6. Between 2016 to 2020, the highest revenue was by product category Grocery in 2020 by \$470,314.18.
- Amount lost by discounts per sales revenue is 5.8%, which is less than 10% obviously the discounts were well implemented.
- Africa had the highest sales revenue with Ivory Coast having the highest demand of 484 products.
- By country, the US being the number one generated about 237 million dollars.
- In 2018, Pizza had the highest negative change of -30% on sales price compared to other products.
- The sales recorded a 45% profit margin, which shows the business is booming well.
- We recorded a sale growth rate of 2.8%, could this be because of the sales discounts?
- Some products still had a negative growth rate even though we had an overall growth rate.
- There is a linear correlation between product price and quantity sold by discounts.
- Most countries were below average sales monthly even though we had a 2.8% growth rate.
- Note, that when the discount percentage increases, could increase demand but reduces sales revenue and profits.
- There could be an increase in sale revenue in the next five years with little fluctuations.

5 Recommendation

- Set a target to reach more of developing countries, especially the African region.
- Offer promotional sales like flash sales to clear old products, monthly, seasonal, or buy one get one free(BOGOF).
- Set discounts for new customers, it acts as an incentive and attracts more customers.
- Have a loyalty program or coupons for loyal customers, for a long business relationship.
- Provide discounts for a certain number of quantities purchased by a customer.
- Make use of percentage discounts than absolute discounts, a 25% discount on \$40 is the same \$10 discount, but the percentage looks higher.
- Check out your competitors before implementing sales discounts.
- Avoid fraudulent promotional sales by inflating product prices before implementing sales discounts.

- Track and monitor the amount lost by discounts, could be fatal for the business.

6 Conclusion

Working with Power Bi was an intriguing and challenging experience knowing that this is my first analysis report in this career path. Power Bi is a fascinating and powerful visualisation tool, I want to keep exploring this tool in order to apply this bundle of knowledge I have garnered in this module to the industries and the world at large.

Section 2

Appendix: BI Design

7 Data Pre-Processing and Cleaning

7.1 Loading the Data

The first process when working with Power Bi is loading of the data which is done using the “Get Data” button. Clicking on the button brings a drop down of different methods of loading data in Power Bi.

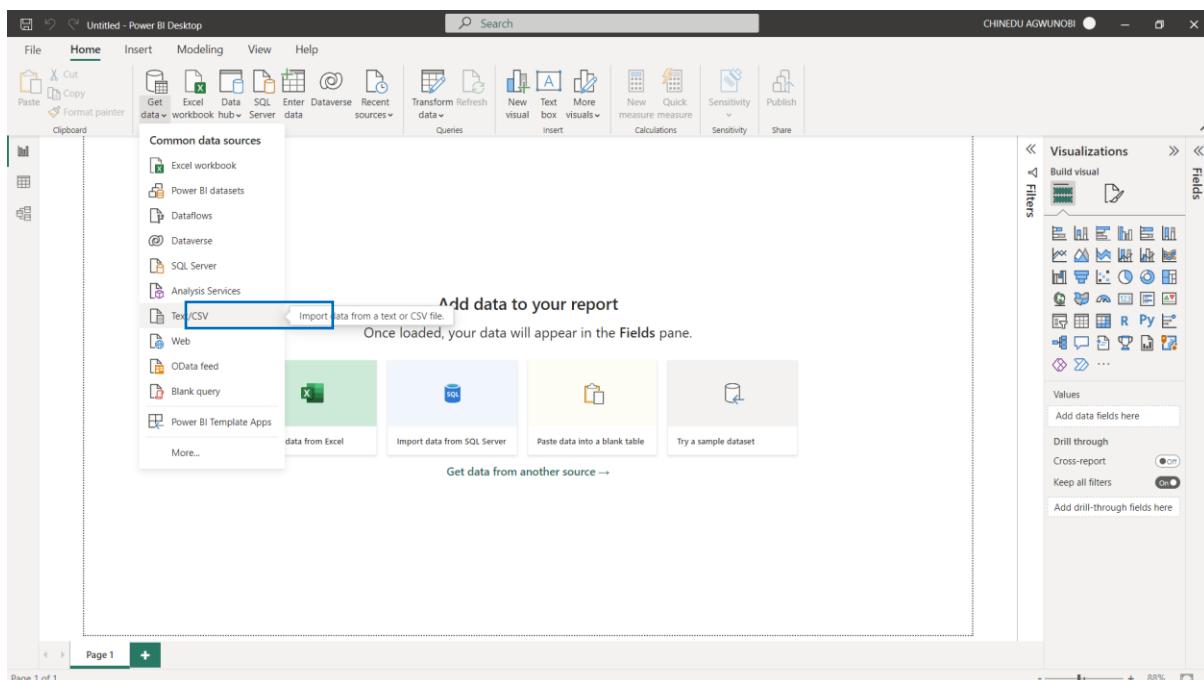


Fig.48. Get data dropdown

The E-Commerce dataset is a CSV file, the Text/CSV option was used to load the dataset into Power Bi. The encoding of the file origin is 1252: Western European (Windows), delimiter is a comma separated file and the first 20 rows were loaded initially. By clicking the Load button, the dataset was fully loaded.

E-Commerce Dataset.csv

File Origin		Delimiter	Data Type Detection					
1252: Western European (Windows)		Comma	Based on first 200 rows					
CustomerId	CustomerName	CustomerEmail	CustomerGender	CustomerIp	ProductName	ProductPrice	CostC	
4070	Roberta Lesch	Rodrick_Kassulke88@gmail.com	Female	138.152.90.54	Hat	713		
2530	Adaline Hoeger	Vicky93@yahoo.com	Female	35.240.213.166	Cheese	60		
2697	Yesenia Hahn	Linda48@gmail.com	Male	180.226.79.181	Chips	310		
2405	Fern Stark	Eriberto5@yahoo.com	Female	227.148.70.138	Chair	468		
4394	Ronaldo Bashirian	Nicole56@hotmail.com	Male	85.197.31.229	Chicken	360		
2013	Moses Runolfsdottir	Polly99@hotmail.com	Male	118.177.129.243	Salad	190		
396	Jared Purdy	Nicholaus77@hotmail.com	Female	6.136.10.26	Keyboard	860		
4151	Minerva O'Kon	Daphne_Kessler@yahoo.com	Female	122.246.199.255	Gloves	95		
2987	Shana Beier	Valentine89@gmail.com	Female	129.204.15.31	Pants	188		
26	Tristin Bogan	Jaylen15@hotmail.com	Female	69.224.113.170	Keyboard	915		
4205	Bernard West	Jakob_Wintheiser@gmail.com	Male	79.237.168.36	Chair	180		
1462	Rodrigo Swift	Jerome.VonRueden30@hotmail.com	Female	211.175.54.230	Keyboard	326		
705	Ena Roberts	Cornelius.Lakin88@hotmail.com	Female	5.100.150.96	Chair	265		
2388	Miles McDermott	JoshuaH_Cruickshank97@hotmail.com	Male	103.91.207.71	Tuna	162		
4303	Rosalia Graham	Emory21@yahoo.com	Female	192.8.48.121	Soap	642		
4269	Olin Blanda	Chyna.Skiles@hotmail.com	Female	82.195.168.140	Towels	196		
2822	Rick Kunze	Reggie.Ruecker@yahoo.com	Female	145.165.117.81	Computer	692		
2271	Minnie Fadel	Cheyanne41@hotmail.com	Male	112.223.174.18	Bacon	642		
235	Vernie Labadie	Ima_Walker@yahoo.com	Male	68.205.51.63	Sausages	570		
2209	Rosina Ernser	Fae_Maggio38@hotmail.com	Male	62.38.157.198	Shirt	235		

Extract Table Using Examples **Load** Transform Data Cancel

Fig.49. Loading view

Errors were detected by Power Bi immediately the data was loaded; this requires some data cleaning.

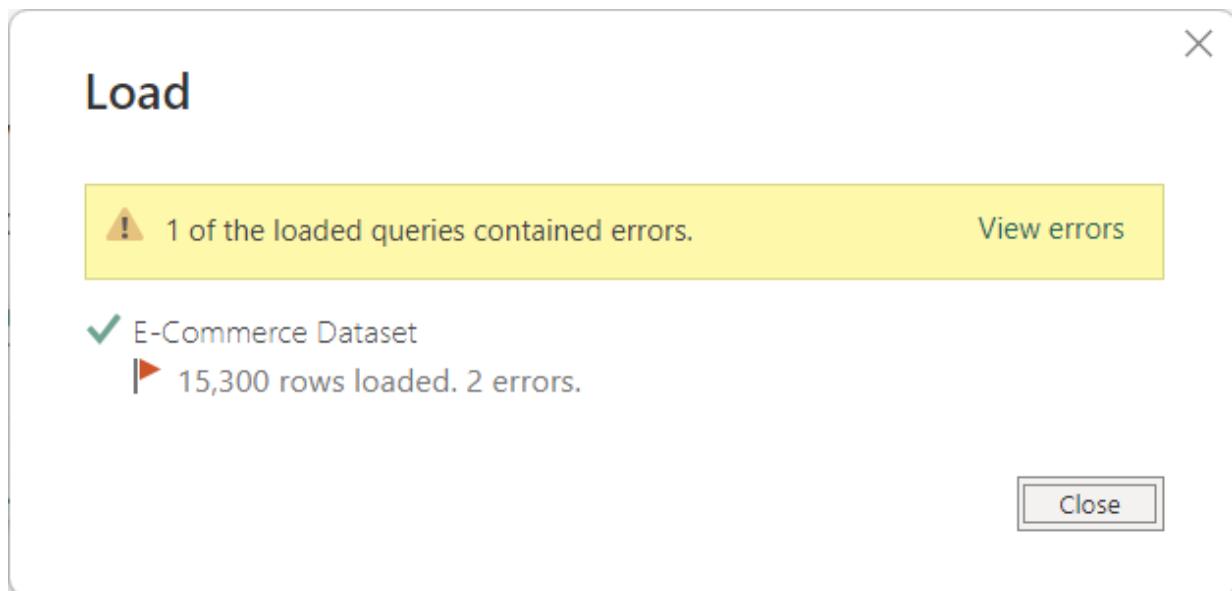


Fig.50. Errors Detected

7.2 Data Cleaning

To clean the data, this will be done using the Power Query. To take us to power query, we use the transform data drop down at the home tap, then click transform data.

Power Query Editor - Untitled - Power Query Editor

Queries [2]

E-Commerce Dataset

CustomerName

Properties

Transform

Add Column

View

Tools

Help

Close & Apply

New Source

Recent Data

Data Sources

Manage Parameters

Refresh Preview

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Split Column

Group By

Replace Values

Data Type: Whole Number

Merge Queries

Text Analytics

Combine Files

Use First Row as Headers

Sort

Transform

Combine

Query Settings

PROPERTIES

Name: E-Commerce Dataset

All Properties

APPLIED STEPS

Source: Promoted Headers

Changed Type: CustomerName

PREVIEW DOWNLOADED AT 12:52

Fig.51. Power Query view

7.2.1 Removing Errors, Blanks, NA, and Column

The “ProductName” had only blanks, “ProductCategory” had errors, blanks, and NA, “QuantitySold” had errors and blanks. Errors were detected on “OrderDay” when the type was changed to whole numbers.

Power Query Editor - Untitled - Power Query Editor

Queries [2]

Errors in E-Commerce Dataset

CustomerName

Properties

Transform

Add Column

View

Tools

Help

Close & Apply

New Source

Recent Data

Data Sources

Manage Parameters

Refresh Preview

Advanced Editor

Choose Columns

Remove Columns

Keep Rows

Remove Rows

Split Column

Group By

Replace Values

Data Type: Decimal Number

Merge Queries

Text Analytics

Combine Files

Use First Row as Headers

Sort

Transform

Combine

Query Settings

PROPERTIES

Name: Errors in E-Commerce Dataset

All Properties

APPLIED STEPS

Source: Detected Type Mismatches

Added Index: Kept Errors

Reordered Columns

PREVIEW DOWNLOADED AT 21:02

Fig.52. Errors, Blanks and NA

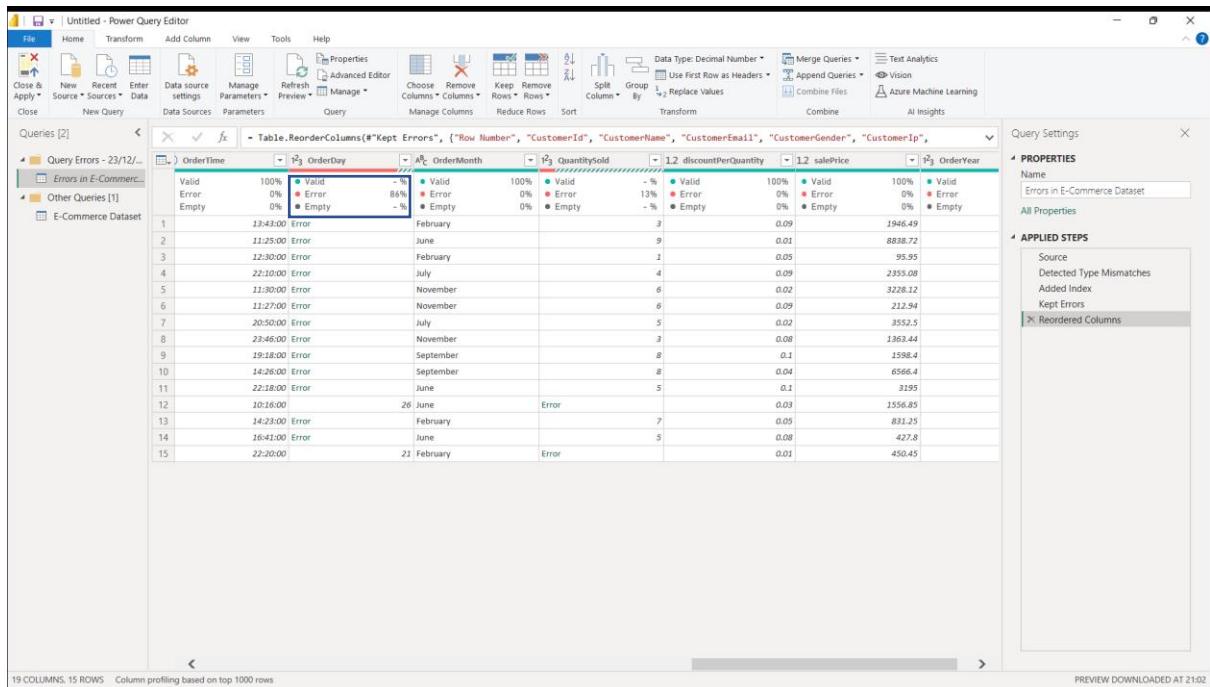


Fig.53. Errors, Blanks and NA

The percentages of the errors, blanks and NA are less than 10% and will not a big effect on the dataset, I chose to remove them.

M-Language used in removing errors = `Table.RemoveRowsWithErrors(#"Removed Errors", {"QuantitySold"}, {"ProductCategory"}, {"OrderDay"})`

M-Language used in filtering columns = `Table.SelectRows(#"Removed Errors1", each ([ProductName] <> "") and ([ProductCategory] <> "NA"))`

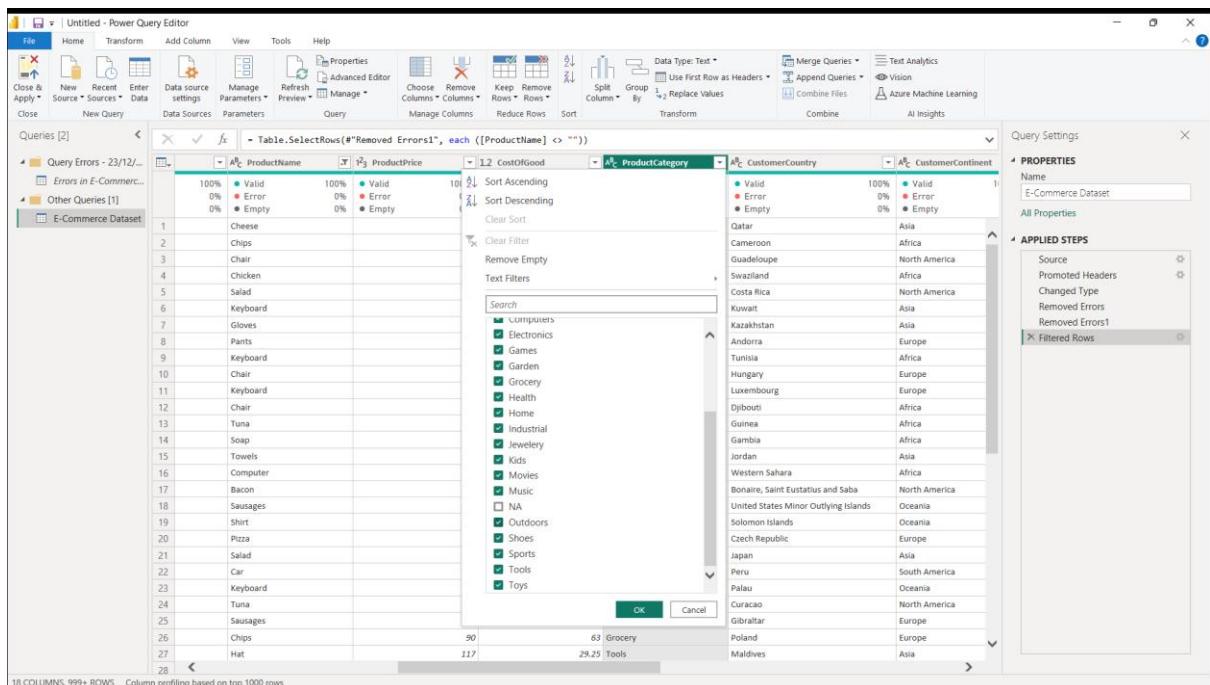


Fig.54. Removing Errors, Blanks and NA

To create an OrderDate column from OrderDay, OrderMonth, OrderYear, I replaced the values of the OrderMonth with numerical representations, changed the type to whole number. The customised column was used to create the OrderDate.

M-Language for the OrderDate is shown below.

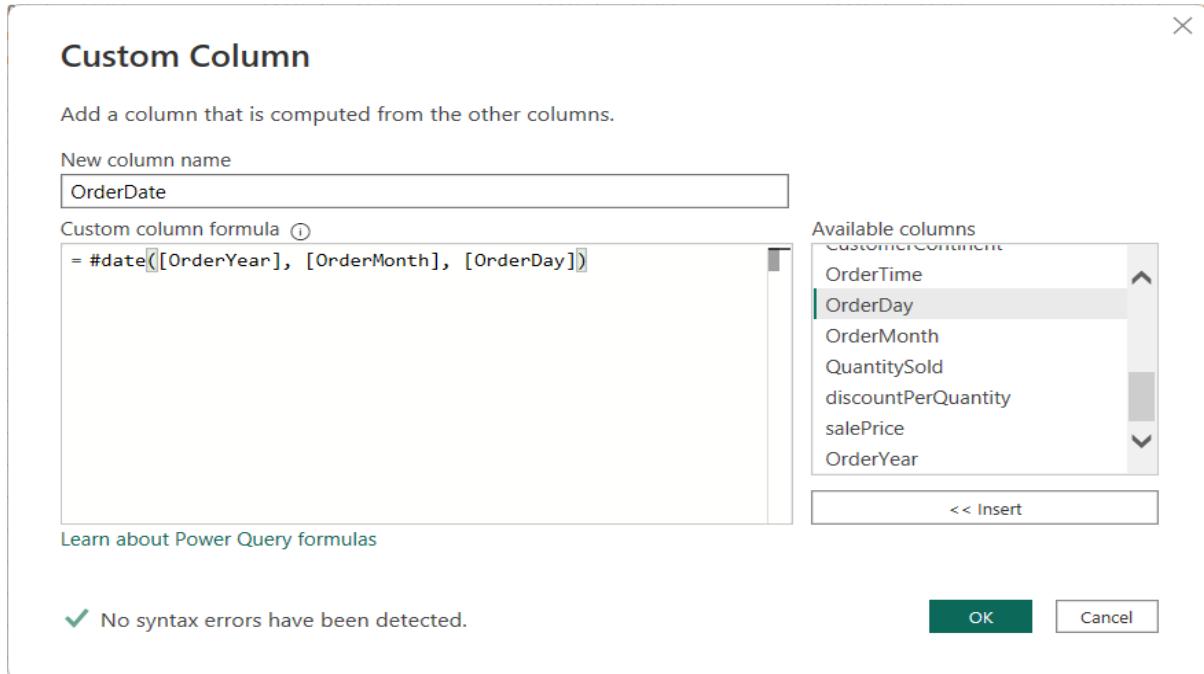


Fig.55. M-Language for OrderDate

Few errors were detected again and were removed

The screenshot shows the Power Query Editor interface with the 'Applied Steps' pane open. The 'APPLIED STEPS' section lists several steps: Promoted Headers, Changed Type, Removed Errors, and so on. The 'Removed Errors' step is highlighted. The main area shows a table with columns: OrderMonth, OrderDay, OrderYear, and OrderDate. The 'OrderDate' column is highlighted. The 'APPLIED STEPS' pane also shows 'Removed Errors2'.

Fig.56. OrderDate Errors Removed

Four columns were removed, "OrderTime", "OrderDay", "OrderMonth", "OrderYear" }.

M-Language for removing the columns = = Table.RemoveColumns(#"Removed Errors2", {"OrderTime", "OrderDay", "OrderMonth", "OrderYear"})

7.3 Data Modelling

After data cleaning, the E-Commerce dataset has one table and 15 columns. For easy visualisation and data modelling, more tables will be created from the dataset and connected to each other with a relationship, it can be one to one, one to many, many to one or many to many under star schema or snowflake schema.

For this modelling, one to many relationships under star schema.

7.3.1 Creating Tables

Three new index columns were added and renamed “countryid”, “productid” and “salesid” before duplicating the table for E-Commerce_Customers table.

7.3.1.1 E-Commerce_Customers

The screenshot shows the Power BI Data Editor interface. The top ribbon includes File, Home, Transform, Add Column, View, Tools, and Help. The main area displays a table titled "Removed Duplicates1" with columns: Customer_id, CustomerName, CustomerGender, CustomerEmail, and Customer_ip. The table contains 28 rows of customer data. To the left is a "Queries [5]" pane listing E-Commerce, E-Commerce_Customers, E-Commerce_Products, E-Commerce_Countries/..., and E-Commerce_Sales_Facts. On the right, the "Query Settings" pane is open, showing the "PROPERTIES" section with the name set to "E-Commerce_Customers" and the "APPLIED STEPS" section listing various modifications made to the query.

	Customer_id	CustomerName	CustomerGender	CustomerEmail	Customer_ip
1	F0010131	Aaliyah Anderson	Female	Joany_Hackett37@hotmail.com	100.236.30.83
2	M004487	Aaliyah Flatley	Male	Aurelia95@yahoo.com	201.133.68.154
3	F0012713	Aaliyah Johns	Female	Lydia_Huels76@hotmail.com	21.235.251.54
4	F0013854	Aaliyah Langworth	Female	Tyrell_Robel@hotmail.com	121.66.6.88
5	M009346	Aaliyah Miller	Male	Joshuah_Kunze17@gmail.com	56.141.45.48
6	M0011849	Aaron Breahan	Male	Lauri_Johnson@gmail.com	78.148.52.115
7	M002834	Aaron Bednar	Male	Junius68@gmail.com	77.247.101.31
8	F009975	Aaron Corkery	Female	Shania_Fay62@gmail.com	29.33.98.9
9	F007929	Aaron Cruickshank	Female	Freddy95@yahoo.com	38.150.78.194
10	M0011512	Aaron Donnelly	Male	Marco95@yahoo.com	62.184.25.127
11	M00773	Aaron Mante	Male	Angelina66@gmail.com	124.83.20.29
12	F009761	Abigail Crona	Female	Mavis9@yahoo.com	44.157.159.243
13	F005611	Abigail Hauck	Female	Katrine49@hotmail.com	122.86.46.68
14	M002133	Abigail Kiehn	Male	Geo_Rohan@gmail.com	20.71.247.47
15	M008846	Abigail Kovacek	Male	Freedab9@yahoo.com	33.6.42.46
16	M003808	Abigail Kunde	Male	Cierra_Schimmel76@yahoo.com	89.6.80.195
17	F009182	Abigail Sawany	Female	Dayana_Goodwin@hotmail.com	255.138.177.23
18	F0011919	Abigail Steuber	Female	Godfrey_Little@yahoo.com	79.145.168.59
19	M009759	Abby Altenwerth	Male	Anne_Runte@gmail.com	58.236.145.209
20	F001192	Abbey Cummings	Female	Gina84@gmail.com	52.89.82.144
21	F003167	Abbey Hickle	Female	Mason_Wiza@gmail.com	107.191.2132.32
22	F009987	Abbey Hilpert	Female	Emrie_Legros@gmail.com	113.152.204.244
23	F008893	Abbey Labadie	Female	Francesca_Fisher56@yahoo.com	206.253.176.29
24	F002672	Abbey Quigley	Female	Kurtis13@gmail.com	197.40.27.176
25	M001303	Abby Waelchi	Male	Preston_Beer@yahoo.com	95.247.231.176
26	M00614	Abby Wiza	Male	Chance_Will49@yahoo.com	144.200.191.107
27	F009758	Abbie Christiansen	Female	Roberto.Collins@gmail.com	4.134.200.247
28	M001556	Abbie Gorczany	Male	Malinda.Raynor@yahoo.com	118.85.210.139

5 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 09:04

Fig.57. E-Commerce_Customers

The following important steps were done before the final sorting of the rows by “CustomerName” and “CustomerGender” and their M-Languages respectively.

Removed Other Columns = Table.SelectColumns(#"Renamed Columns2", {"CustomerName", "CustomerEmail", "CustomerGender", "Customer_Ip"})

Remove Duplicates = Table.Distinct(#"Removed Other Columns", {"CustomerName", "CustomerGender"})

Added Conditional Column = Table.AddColumn(#"Reordered Columns1", "Custom", each if [CustomerGender] = "Male" then "M" else "F")

Added Custom1 = Table.AddColumn(#"Changed Type3", "Customer id", each [Custom] & ("00") & [Index])

The "Custom id" was customized with the "CustomerGender"

7.3.1.2 E-Commerce_Products

Fig.58. E-Commerce_Products

The important steps and M-Language used in formatting the table:

Removed Other Columns = Table.SelectColumns(#"Renamed Columns2", {"ProductName", "ProductPrice", "CostOfGood", "ProductCategory"})

Added Custom1 = Table.AddColumn(#"Changed Type3", "Product id", each ("P0") & [Index])

Sorted Rows1 = Table.Sort(#"Filtered Rows1", {"ProductName", Order.Ascending})

7.3.1.3 E-Commerce_Countries/Regions

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

- Queries [5]:** E-Commerce, E-Commerce_Customers, E-Commerce_Products, E-Commerce_Countries/..., E-Commerce_Sales_Facts.
- Current Query:** A table with columns: CountryId, CustomerContinent, and CustomerCountry.
- Transform ribbon:** Home, Transform, Add Column, View, Tools, Help.
- Transform pane:**
 - Properties: Name = E-Commerce_Countries/Regions.
 - Applied Steps: A list of 30 steps including Reordered Columns, Removed Duplicates, and Sorted Rows.
- Preview pane:** Shows the first 28 rows of the table.

CountryId	CustomerContinent	CustomerCountry
AF00167	Africa	Algeria
AF00214	Africa	Angola
AF00128	Africa	Benin
AF00185	Africa	Botswana
AF0062	Africa	Burkina Faso
AF0070	Africa	Burundi
AF002	Africa	Cameroun
AF00184	Africa	Cape Verde
AF0044	Africa	Central African Republic
AF0043	Africa	Chad
AF00153	Africa	Comoros
AF0041	Africa	Democratic Republic of the Congo
AF0011	Africa	Djibouti
AF00220	Africa	Egypt
AF0084	Africa	Equatorial Guinea
AF0086	Africa	Eritrea
AF00204	Africa	Ethiopia
AF00244	Africa	Gabon
AF0013	Africa	Gambia
AF00186	Africa	Ghana
AF0012	Africa	Guinea
AF00114	Africa	Guinea-Bissau
AF00243	Africa	Ivory Coast
AF00206	Africa	Kenya
AF00144	Africa	Lesotho
AF00170	Africa	Liberia
AF0052	Africa	Libya
AF0049	Africa	Madagascar

Fig.59. E-Commerce_Countries/Regions

The important steps and M-Language used in formatting the table:

Removed Other Columns = Table.SelectColumns(#"Removed Duplicates", {"CustomerCountry", "CustomerContinent"})

Duplicated Column = Table.DuplicateColumn(#"Removed Other Columns", "CustomerContinent", "CustomerContinent - Copy")

Split column by Position = Table.SplitColumn(#"Duplicated Column", "CustomerContinent - Copy", Splitter.SplitTextByPositions({0, 2}, false), {"CustomerContinent - Copy.1", "CustomerContinent - Copy.2"})

Sorted Rows = Table.Sort(#"Reordered Columns2", {"CustomerContinent", Order.Ascending}, {"CustomerCountry", Order.Ascending})

7.3.1.4 E-Commerce_Sales_Facts

Fig.60. E-Commerce_Sales_Facts

The important steps used in formatting and merging of columns and the M-Language:

```
Merged Queries = Table.NestedJoin(#"Renamed Columns2", {"CustomerName", "CustomerGender"}, #"E-Commerce_Customers", {"CustomerName", "CustomerGender"}, "Customer id", JoinKind.LeftOuter)
```

```
Expanded Customer id = Table.ExpandTableColumn(#"Merged Queries", "Customer id", {"Customer id"}, {"Customer id.1"})
```

```
Merged Queries1 = Table.NestedJoin(#"Reordered Columns1", {"ProductName", "ProductPrice", "CostOfGood", "ProductCategory"}, #"E-Commerce_Products", {"ProductName", "ProductPrice", "CostOfGood", "ProductCategory"}, "Product id", JoinKind.LeftOuter)
```

```
Expanded Product id = Table.ExpandTableColumn(#"Merged Queries1", "Product id", {"Product id"}, {"Product id.1"})
```

```
Merged Queries2 = Table.NestedJoin(#"Reordered Columns2", {"CustomerCountry", "CustomerContinent"}, #"E-Commerce_Countries/Regions", {"CustomerCountry", "CustomerContinent"}, "Country id", JoinKind.LeftOuter)
```

```
Expanded Country id = Table.ExpandTableColumn(#"Merged Queries2", "Country id", {"Country id"}, {"Country id.1"})
```

```
Removed Column2 = Table.RemoveColumns(#"Reordered Columns4", {"Country_Id", "Product_Id"})
```

```
Changed Type with Locale = Table.TransformColumnTypes(#"Changed Type3", {"salePrice", Currency.Type}, "en-US")
```

```
Removed Other Columns = Table.SelectColumns(#"Renamed Columns5", {"Sales id", "Customer id", "Product id", "Country id", "QuantitySold", "discountPerQuantity", "salePrice", "OrderDate"})
```

```
Merged Queries3 = Table.NestedJoin(#"Sorted Rows", {"Product id"}, #E-Commerce_Products, {"Product id"}, "E-Commerce_Products", JoinKind.LeftOuter)
```

```
Expanded E-Commerce_Products = Table.ExpandTableColumn(#"Merged Queries3", "E-Commerce_Products", {"ProductPrice", "CostOfGood"}, {"ProductPrice", "CostOfGood"})
```

```
Reordered Column5 = Table.ReorderColumns(#"Expanded E-Commerce_Products", {"Sales id", "Customer id", "Product id", "Country id", "ProductPrice", "CostOfGood", "QuantitySold", "discountPerQuantity", "SalesPrice", "OrderDate"})
```

7.3.1.5 Date Dim

This table was created in the Power Bi data view using DAX. It comprise the date, day, month, quarter and year.

The screenshot shows the Power BI Desktop interface with the 'Date Dim' table selected. The ribbon at the top has 'Table tools' selected. The 'Structure' tab is active, showing a preview of the table with columns: Date, Year, Month, Qtr, and Day. The preview data shows January 2016 with days numbered 1 through 31. To the right, the 'Data' pane shows the table structure with columns: Date, Day, Month, Qtr, and Year. Below the table preview, it says 'Table: Date Dim (1,827 rows)'.

Fig.61. Date Dim

DAX formulars:

Date = CALENDARAUTO()

Day = 'Date Dim'[Date].[Day]

Month = 'Date Dim'[Date].[Month]

Quarter = 'Date Dim'[Date].[Quater]

Year = 'Date Dim'[Date].[Year]

7.3.2 Star Schema

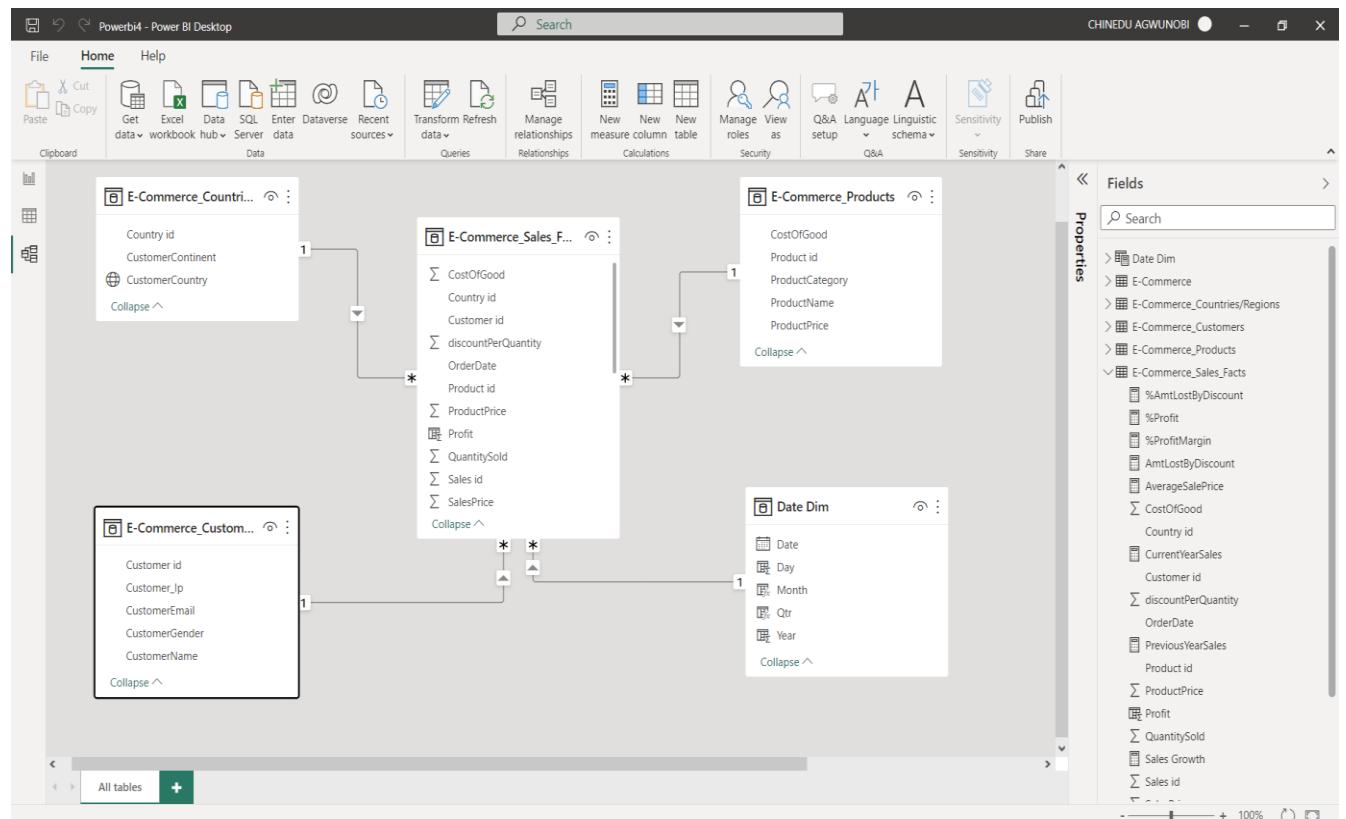


Fig.62. Star Schema Model

7.3.2.1 Relationships

The model has one to many relations and can also be many to many relations, this is because of the columns that were merged when the tables were created.

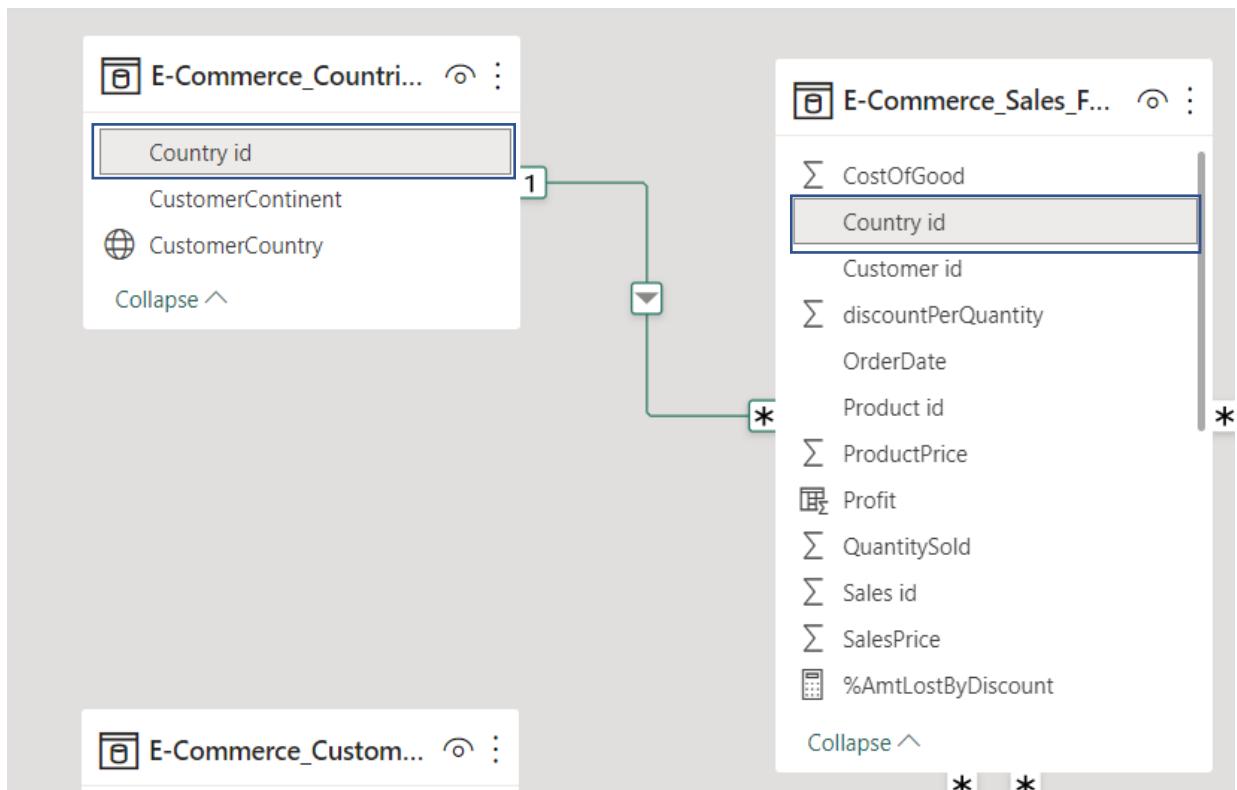


Fig.63. E-Commerce_Countries/Regions Relationship

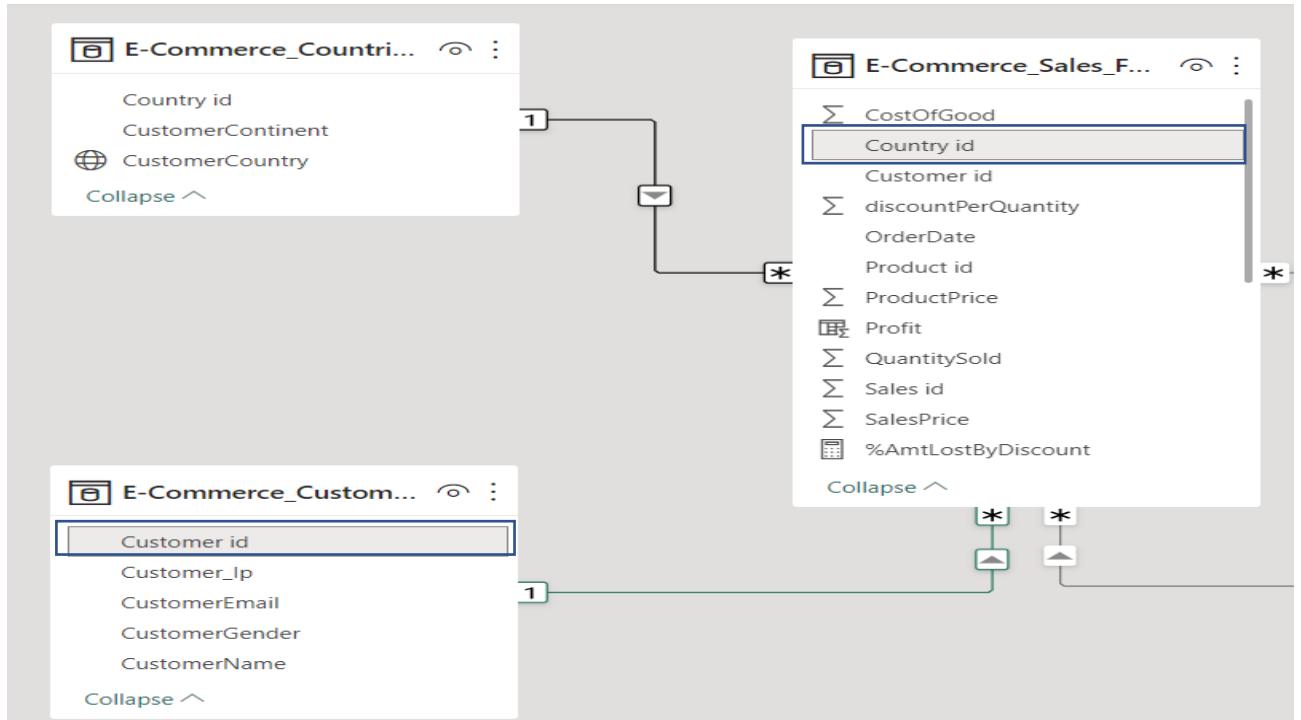


Fig.64. E-Commerce_Customer Relationship

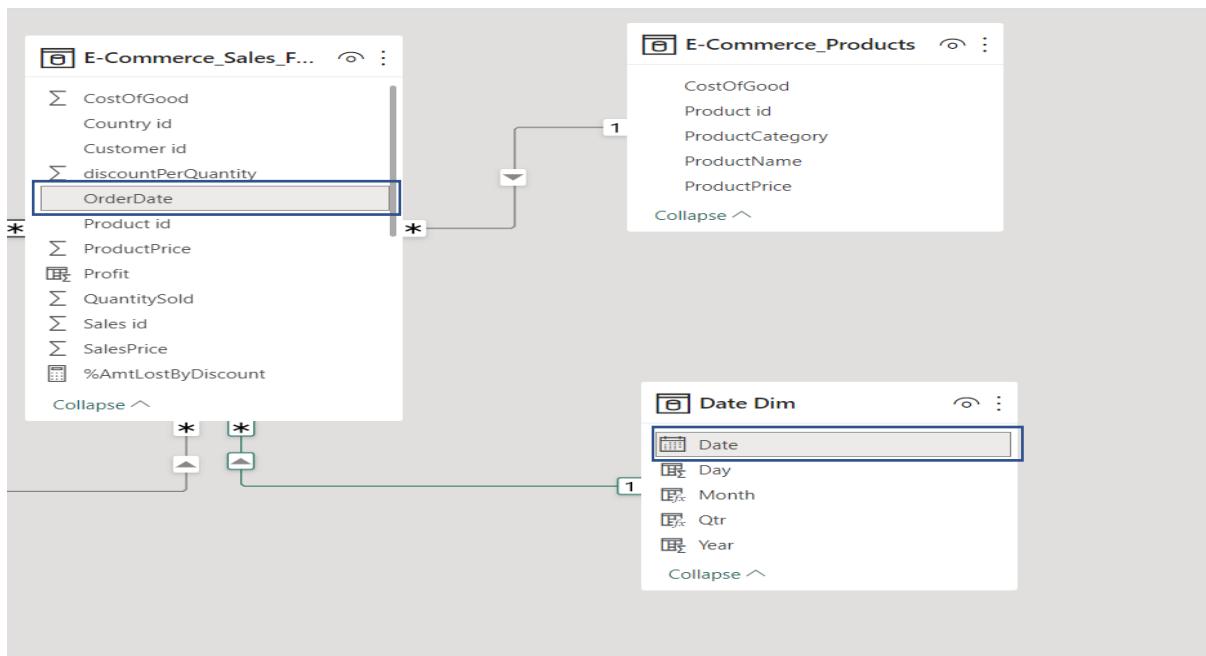


Fig.65. Date Dim Relationship

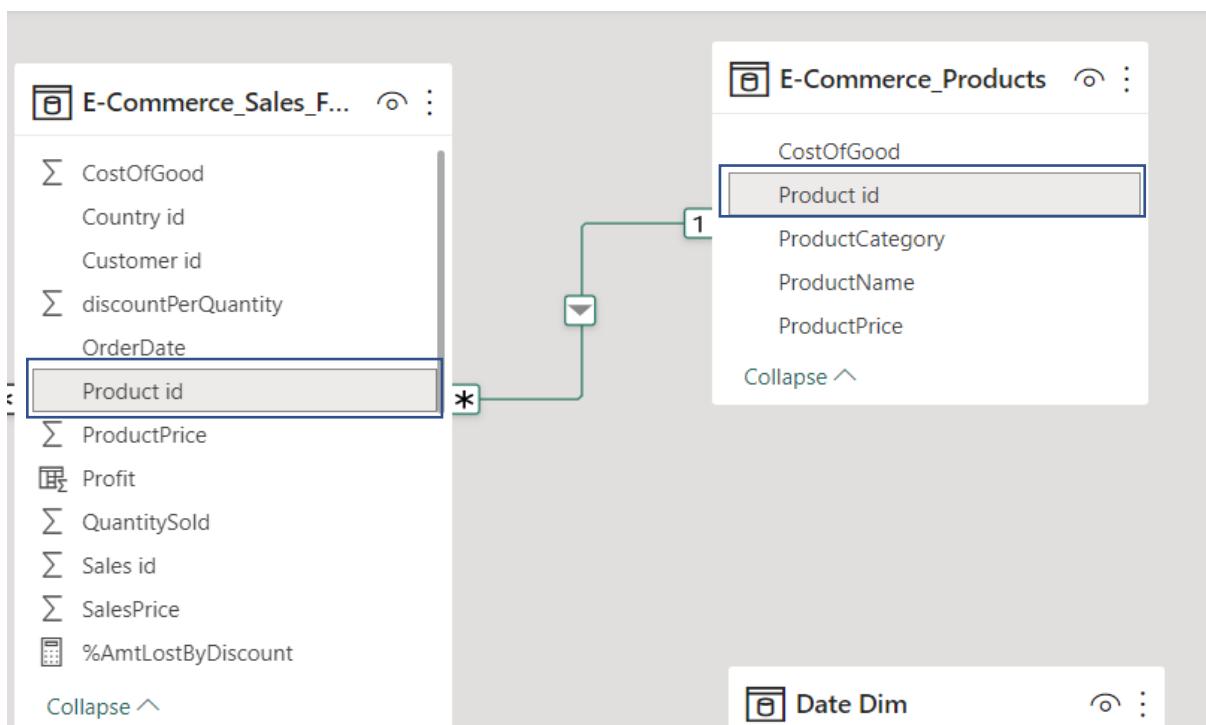


Fig.66. E-Commerce_Products Relationship

7.4 DAX and M-Language

Several DAX formulars were used in calculating measures, a column, and a table. The M-Language used were done during data pre-processing and cleaning.

On the E-Commerce_Sales_Facts table, 13 measure and a column were calculated using DAX.

7.4.1 DAX

TotalOrders = DISTINCTCOUNT('E-Commerce'[Sales id])

SalesRevenue = SUM('E-Commerce_Sales_Facts'[SalesPrice])

TotalCostofGoods = SUMX('E-Commerce_Sales_Facts', 'E-Commerce_Sales_Facts'[CostOfGood] * 'E-Commerce_Sales_Facts'[QuantitySold])

SalesRevenue = SUM('E-Commerce_Sales_Facts'[SalesPrice])

SalesPriceWithoutDiscount = CALCULATE(SUMX('E-Commerce_Sales_Facts', 'E-Commerce_Sales_Facts'[ProductPrice] * 'E-Commerce_Sales_Facts'[QuantitySold]))

Profit = 'E-Commerce_Sales_Facts'[SalesPrice] - 'E-Commerce_Sales_Facts'[TotalCostofGoods]

TotalProfit = 'E-Commerce_Sales_Facts'[SalesRevenue] - 'E-Commerce_Sales_Facts'[TotalCostofGoods]

%Profit = DIVIDE('E-Commerce_Sales_Facts'[TotalProfit], 'E-Commerce_Sales_Facts'[TotalCostofGoods], 0)

%ProfitMargin = DIVIDE('E-Commerce_Sales_Facts'[TotalProfit], 'E-Commerce_Sales_Facts'[SalesRevenue], 0)

AverageSalePrice = AVERAGE('E-Commerce_Sales_Facts'[SalesPrice])

AmtLostByDiscount = SUMX('E-Commerce_Sales_Facts', 'E-Commerce_Sales_Facts'[SalesPriceWithoutDiscount] - 'E-Commerce_Sales_Facts'[SalesPrice])

%AmtLostByDiscount = DIVIDE('E-Commerce_Sales_Facts'[AmtLostByDiscount], 'E-Commerce_Sales_Facts'[SalesRevenue], 0)

Date = CALENDARAUTO()

Day = 'Date Dim'[Date].[Day]

Month = 'Date Dim'[Date].[Month]

Quarter = 'Date Dim'[Date].[Quater]

Year = 'Date Dim'[Date].[Year]

PreviousYearSales =

```
VAR FirstYear = CALCULATE(MIN('E-Commerce'[OrderDate].[Year] ))  
RETURN  
CALCULATE ( 'E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = FirstYear )
```

CurrentYearSales =

```
VAR LatestYear = CALCULATE(MAX('E-Commerce'[OrderDate].[Year] ))  
RETURN  
CALCULATE ( 'E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = LatestYear )
```

Sales Growth =

```
var FirstYear = CALCULATE(MIN('Date Dim'[Year]))  
var LatestYear = CALCULATE(MAX('Date Dim'[Year]))  
var GDPFirstYear = CALCULATE('E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = FirstYear)  
var GDPLatestYear = CALCULATE('E-Commerce_Sales_Facts'[SalesRevenue], 'Date Dim'[Year] = LatestYear)  
var Diff = GDPLatestYear - GDPFirstYear  
return DIVIDE(Diff, GDPFirstYear, 0)
```

7.4.2 M-Language

The M-Language were used mostly during the data pre-processing and cleaning. The M-Language was majorly used for formatting and merging of columns. The Fig.67 below is the summary of the M-Language and can be seen in the power query advanced editor.



Fig.67. M-Language

7.5 Dashboard

The Power Bi dashboard was organized in an ordered form, it comprises of six pages starting with the Homepage which has buttons to navigate to other pages. A good business analyst first looks at the business summary or the sales overview to get the initial insight on the business before diving more to know the reasons why the sales went up or down. The dashboard is for retailers, eCommerce platforms, sales personnel and various businesses.

Sales Overview is the next page after the Homepage, to get a glimpse of business. We then look at the Sales Performance and see what product, continent, country, prices are really impacting the sales.

We went further to look at the Sales Growth over the years, are sales fluctuating or steady in growth. Knowing your Key Influencers are essential if you want your business not to fail.

We finally looked at the future sales, Sales Prediction, a wonderful technique in business analytics.

Below are the visuals of the six dashboards.

Big Data and Business Intelligence **Module Code: CIS4008-N**

E-Commerce Analysis on Sales Discounts

Homepage

Sales Overview

Sales Performance

Sales Growth

Key Influencers

Sales Prediction

Name: Chinedu Emmanuel Agwunobi

Student ID: B1555134

Fig.68. Homepage

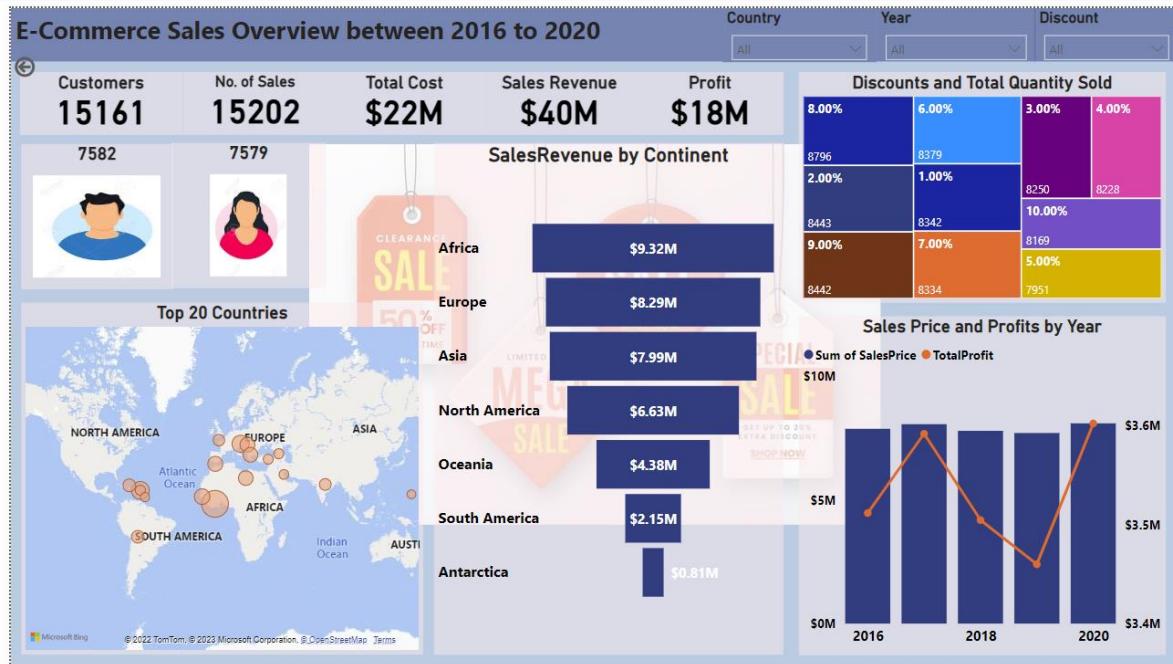


Fig.69. Sales Overview

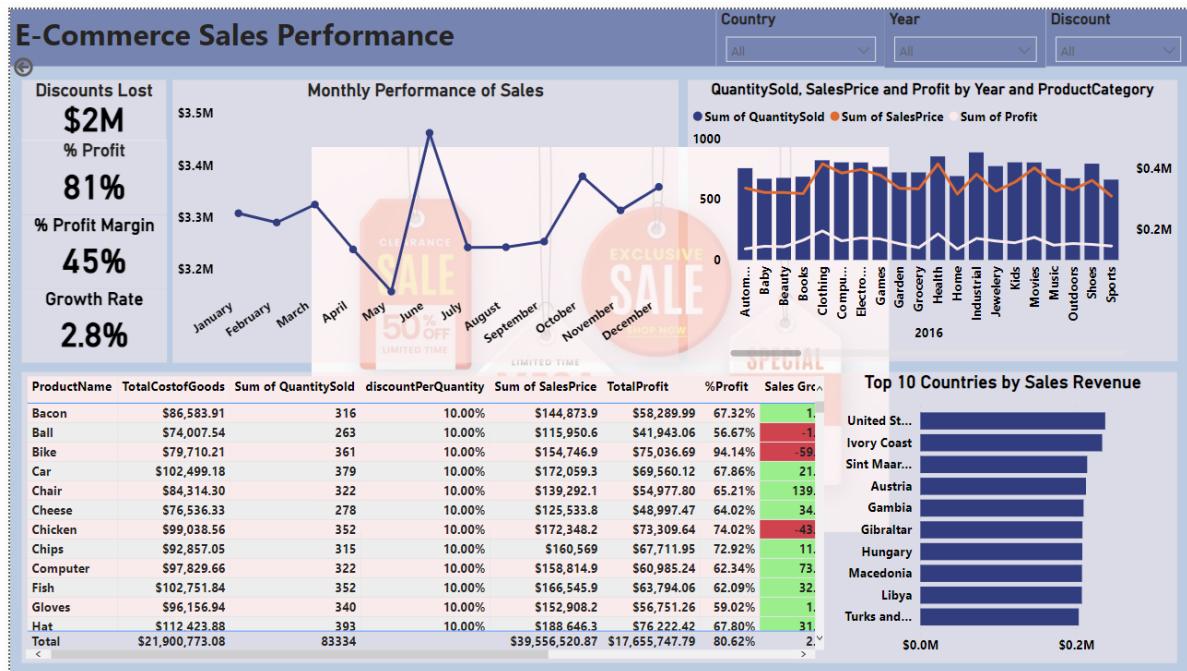


Fig.70. Sales Performance



Fig.71. Sales Growth

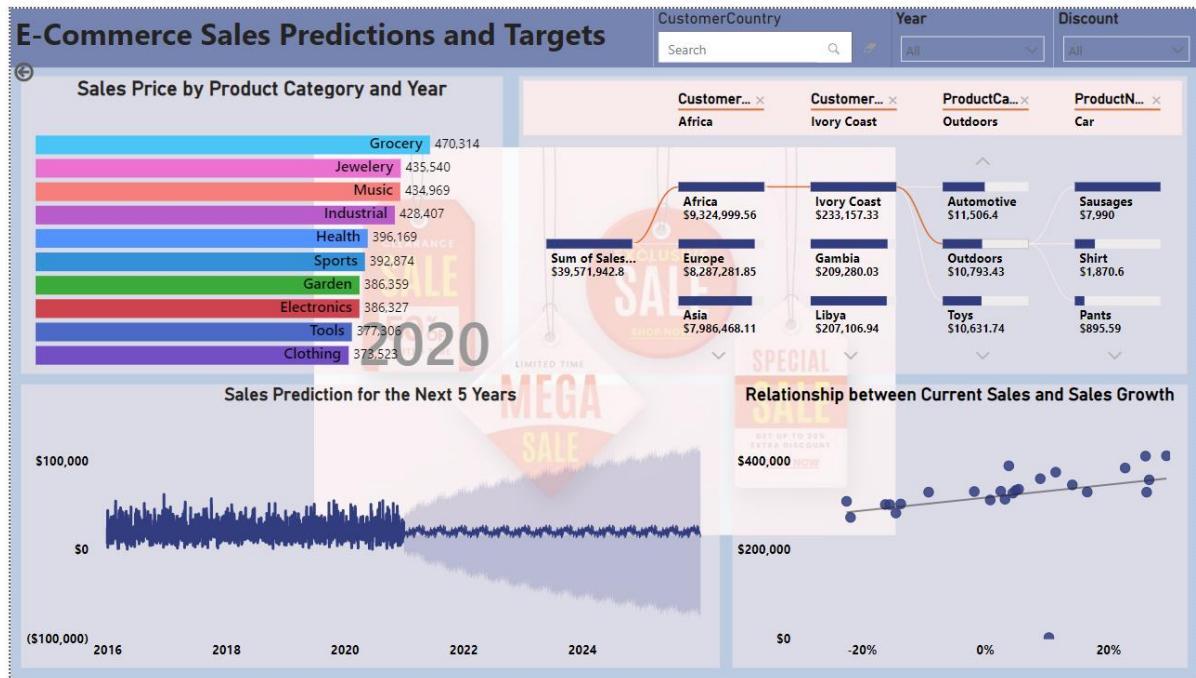


Fig.72. Sales Prediction

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	92
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	88
Dax and M language	Both DAX and M Language have been extensively used in the report	90
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	94
Average		Add below the average of the four cells above: 91