

BIKE AND ACCESSORY SALES ANALYSIS



**BIG DATA AND BUSINESS INTELLIGENCE MODULE
CIS 4008-N**

SHIRLEY AKUA KONNOR
B1889318
SUBMISSION DATE:11/01/2023

Table of Contents

1. EXECUTIVE SUMMARY.....	5
1.1 Introduction.....	5
1.2 Key Findings.....	5
2. DATA SET DESCRIPTION	6
2.1 Introduction.....	6
2.2 Data Source.....	6
2.2.1 Description	6
2.2.2 Rationale for choosing the Dataset	7
2.3 BI Requirements and Questions.....	8
3. FINDINGS BASED ON ANALYSIS AND EVALUATION	9
3.1 Analysis and Evaluation	9
3.2 Key Findings.....	16
4. CONCLUSIONS AND RECOMMENDATIONS.....	18
4.1 Recommendations Based on Findings	18
4.2 Conclusion	18
1. DATA PRE-PROCESSING AND CLEANING	20
1.1 Importing the Dataset.....	20
1.2 Data Cleaning.....	21
1.2.1 Removing Errors and Columns.....	21
1.2.2 Removing Blanks and Null	22
1.2.3 Renaming Column	23
1.2.4 Replacing Values	25
1.2.5 Changing data type.....	27
1.2.6 Adding a column (Index column) and Reordering the column.....	28
2. DATA MODELLING	30
2.1 Creating Dimension Tables.....	30
2.2 Creating Fact Table.....	32
2.3 Creating a Bridge Table	33
2.4 Creating Relationships	34
3. USING DAX AND M LANGUAGE	36
3.1 M language used in data pre-processing	36
3.2 Adding columns to the Calender dimension table using DAX	37
3.3 Adding columns to the Fact table using DAX	38
3.2.1 Changing data types in Sales Fact Table.....	40
3.4 Creating a new measure using DAX.....	41

4. DASHBOARD	45
4.1 HOMEPAGE	45
4.2 GEOGRAPHICAL ANALYSIS PAGE	46
4.3 CUSTOMER ANALYSIS PAGE.....	46
4.4 PRODUCT ANALYSIS PAGE.....	47
4.5 ANNUAL SALES ANALYSIS PAGE	48
4.6 ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS PAGE	49
5. REFERENCES.....	51

SECTION 1:

BUSINESS INTELLIGENCE REPORT

1. EXECUTIVE SUMMARY

1.1 Introduction

Bikes are increasingly becoming a preferred mode of transportation in recent times due to the increased cost of fuel and gas across the globe, most especially after the Covid-19 pandemic. With the rise in demand for the use of bikes, the cost of bikes and their accessories have also increased. Therefore analysing the sales of bikes and their accessories will be beneficial to companies in the bike market and also to manufacturers of these products. This will help determine the factors that will help boost the revenue and profit of their business and also help them make informed decisions in the future to ensure the growth of the business across the United States and Europe.

The aim of the project was to perform various analyses on the bikes and accessory sales in United States and Europe and determine the factors that boosted the market. The analysis was based on various categories in relation to the data set such as geography, customer, product and annual sales. It also helped in the estimating of future revenue based on past performance and trends from the dataset for businesses so as to help them make decisions effectively .

1.2 Key Findings

- The United States and Australia had the highest orders for bike and accessory products which generated high revenue and profit for the business.
- Canada was the country with the least generated profit and revenue.
- Bikes and accessories were in high demand as compared to the demand for clothing.
- The accessories were products were the highest ordered product but the bikes generated more profit than the accessories.
- The adult age group (35-64) patronized most of the bikes and accessories and this intended generated a lot of revenue and profit for the business
- The senior age group (64+) were the least in generating revenue and profit due to low patronage.
- More males purchased these products as compared to females.
- 2015 was the year with highest generated revenue
- The business's profit margin was fairly good which made the bike and accessory sales a profitable business.

1.3. Recommendations

- More marketing should be done in Canada and European countries like Germany, the United Kingdom and France so as to increase the order for the products which will also increase revenue and profit for the business
- The clothing products should be improved in both marketing and quality in order to increase their demand on the market.
- Promotion sales too can be done to increase the sales of the products
- Different types of bikes should be manufactured to suit the senior age group to also increase their demand for bikes.
- The adult age group should be focused on improving the quality of the products they purchase since they are the highest revenue and profit age group.

2. DATA SET DESCRIPTION

2.1 Introduction

Bikes are being preferred for short-distance commutes in recent times as they save considerable time as compared to other modes of transport. They not only serve as a mode of transport but are used for various activities such as commuting, exercise, sport, hiking and adventure purposes among others. The demand for bike usage has increased over the past years due to the global Covid-19 pandemic as compared to pre-pandemic periods. The bikes industry is dominated by all age groups across different countries and therefore manufacturers offer various types of bikes such as sports bikes, touring bikes, road bikes, mountain bikes and others. The rising demand for bikes is driving the global market growth for them and their accessories.

This Business Intelligence project aims to analyse the data on bikes and accessory sales in Europe and the USA in order to gain insights into the bike and accessory market in various countries in Europe and USA. This will help the bike company make informed decisions about the business.

2.2 Data Source

The dataset used is Bike sales which were obtained from Kaggle.

Link for dataset: <https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe>

The dataset contains information on sales of bikes and accessories in Europe and the USA from 2011-2016. It includes the profits and revenue, age groups of people and gender of people that purchased them. The data is characterized by a lot of diversity in bikes, their accessory and clothes.

2.2.1 Description

Column Name	Description
Date	Date sale occurred
Day	Day the sale occurred
Month	The month the sale occurred
Year	Year the sale occurred
Customer_Age	Age of the customer
Age_group	The age ranges of the customer grouped into four
Customer_Gender	Gender of the customer
Country	Country the sales took place
State	State the sales took place
Product_category	Product category for each order (bikes, accessories and clothing)
Sub_category	A more detailed description of the products(for example road bikes, tires, shorts)
Product	Specific Product of each order
Order_quantity	Quantity of the product ordered
Unit_cost	The individual cost of each product
Unit_price	Individual price of each product

Profit	Total profit made on the count of products sold
Cost	The total cost of the number of products in order
Revenue	Total revenue for each individual order

The figure below shows a screenshot of the dataset:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Date	Day	Month	Year	Customer_Age_Group	Customer_Country	State	Product_Cat	Product_CSub_Cat	Product	Order_Qu	Unit_Cost	Unit_Price	Profit	Cost	Revenue										
2	26/11/2013	26	November	2013	19 Youth (<21M)	Canada	British Columbia	Bike Rack	Hitch Rack	8	45	120	590	360	950											
3	26/11/2015	26	November	2015	19 Youth (<21M)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	8	45	120	590	360	950											
4	23/03/2014	23	March	2014	49 Adults (35M)	Australia	New South Wales	Accessory Bike Rack	Hitch Rack	23	45	120	1366	1035	2401											
5	23/03/2016	23	March	2016	49 Adults (35M)	Australia	New South Wales	Accessory Bike Rack	Hitch Rack	20	45	120	1188	900	2088											
6	15/05/2014	15	May	2014	47 Adults (35F)	Australia	New South Wales	Accessory Bike Rack	Hitch Rack	4	45	120	238	180	418											
7	15/05/2016	15	May	2016	47 Adults (35F)	Australia	New South Wales	Accessory Bike Rack	Hitch Rack	5	45	120	297	225	522											
8	22/05/2014	22	May	2014	47 Adults (35F)	Australia	Victoria	Accessory Bike Rack	Hitch Rack	4	45	120	199	180	379											
9	22/05/2016	22	May	2016	47 Adults (35F)	Australia	Victoria	Accessory Bike Rack	Hitch Rack	2	45	120	100	90	190											
10	22/02/2014	22	February	2014	35 Adults (35M)	Australia	Victoria	Accessory Bike Rack	Hitch Rack	22	45	120	1096	990	2086											
11	22/02/2016	22	February	2016	35 Adults (35M)	Australia	Victoria	Accessory Bike Rack	Hitch Rack	21	45	120	1046	945	1991											
12	30/07/2013	30	July	2013	32 Young Adt F	Australia	Victoria	Accessory Bike Rack	Hitch Rack	8	45	120	398	360	758											
13	30/07/2015	30	July	2015	32 Young Adt F	Australia	Victoria	Accessory Bike Rack	Hitch Rack	8	45	120	398	360	758											
14	15/07/2013	15	July	2013	34 Young Adt M	Australia	Victoria	Accessory Bike Rack	Hitch Rack	7	45	120	349	315	664											
15	15/07/2015	15	July	2015	34 Young Adt M	Australia	Victoria	Accessory Bike Rack	Hitch Rack	7	45	120	349	315	664											
16	02/08/2013	2	August	2013	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	5	45	120	369	225	594											
17	02/08/2015	2	August	2015	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	7	45	120	517	315	832											
18	02/09/2013	2	September	2013	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	2	45	120	148	90	238											
19	02/09/2015	2	September	2015	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	1	45	120	74	45	119											
20	22/01/2014	22	January	2014	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	1	45	120	74	45	119											
21	22/01/2016	22	January	2016	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	1	45	120	74	45	119											
22	17/05/2014	17	May	2014	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	6	45	120	443	270	713											
23	17/05/2016	17	May	2016	29 Young Adt M	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	8	45	120	590	360	950											
24	27/03/2014	27	March	2014	51 Adults (35M)	United States	Oregon	Accessory Bike Rack	Hitch Rack	9	45	120	524	405	929											
25	27/03/2016	27	March	2016	51 Adults (35M)	United States	Oregon	Accessory Bike Rack	Hitch Rack	7	45	120	407	315	722											
26	25/08/2013	25	August	2013	49 Adults (35M)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	3	45	120	221	135	356											
27	25/08/2015	25	August	2015	49 Adults (35M)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	1	45	120	74	45	119											
28	26/12/2013	26	December	2013	49 Adults (35M)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	6	45	120	443	270	713											
29	26/12/2015	26	December	2015	49 Adults (35M)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	5	45	120	369	225	594											
30	02/01/2014	2	January	2014	48 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	4	45	120	295	180	475											
31	02/01/2016	2	January	2016	48 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	4	45	120	295	180	475											
32	13/03/2014	13	March	2014	48 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	20	45	120	1476	900	2376											
33	13/03/2016	13	March	2016	48 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	22	45	120	1624	990	2614											
34	23/05/2014	23	May	2014	47 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	24	45	120	1771	1080	2851											
35	23/05/2016	23	May	2016	47 Adults (35F)	Canada	British Columbia	Accessory Bike Rack	Hitch Rack	23	45	120	1697	1035	2732											
36	19/07/2013	19	July	2013	32 Young Adt F	United States	Oregon	Accessory Bike Rack	Hitch Rack	9	45	120	524	405	929											
37	19/07/2015	19	July	2015	32 Young Adt F	United States	Oregon	Accessory Bike Rack	Hitch Rack	9	45	120	524	405	929											

FIG 1: THE BIKE SALES DATASET

2.2.2 Rationale for choosing the Dataset

This data was chosen for the following reasons:

- To gain insights on the bike and accessory market by detecting trends and patterns to effectively analyse and help formulate informed business decisions, determine the products in demand for the bike and accessory business, the products that boost revenue and show profitability for the business, determine the type of customers that patronize the products and compare sales by country and know the geographical demand for the bikes and accessory.
- To develop and demonstrate business skills like data cleaning and pre-processing, data modelling, the use of M language in Power BI, creating columns and measures using DAX and analysing the dataset through visualisations and presenting it in dashboards.
- To address Big Data problems such as removing errors and dealing with missing values.

2.3 BI Requirements and Questions

This Business intelligence project seeks to analyse data on bikes and accessories sales for the business based on key metrics. It seeks to gain insights into sales in each country, according to gender and age group of customers or product types and also makes future predictions to help measure the business performance.

The questions that this project seeks to answer are as follows:

- Which country had the highest order quantity and generated the most profit and revenue?
- Which product category and subcategory had the highest orders and generated the most profit?
- Which customer age group and gender had the most orders?
- Which year was the most profitable?
- What was the profit margin for the business and what it means for the business
- What was the average profit and revenue generated and is there a correlation between them?
- Would there be an increase or decrease in the revenue in the next year?

3. FINDINGS BASED ON ANALYSIS AND EVALUATION

3.1 Analysis and Evaluation

Various charts and graphs were created to analyse the data set and to answer the business questions stated above:

Question 1

Which country had the highest order quantity and generated the most profit and revenue?

Answer:

The different charts used to answer this question were created in the geographical analysis dashboard as shown below:



FIG 2: THE MAP SHOWING THE COUNTRIES BY ORDER QUANTITY

The chart in the figure above was created using the map visual in Power BI. The countries were used for the location and the bubble size represented the order quantity for the products for each country.

The chart shows that the United States had the highest order quantity according to the bubble size.

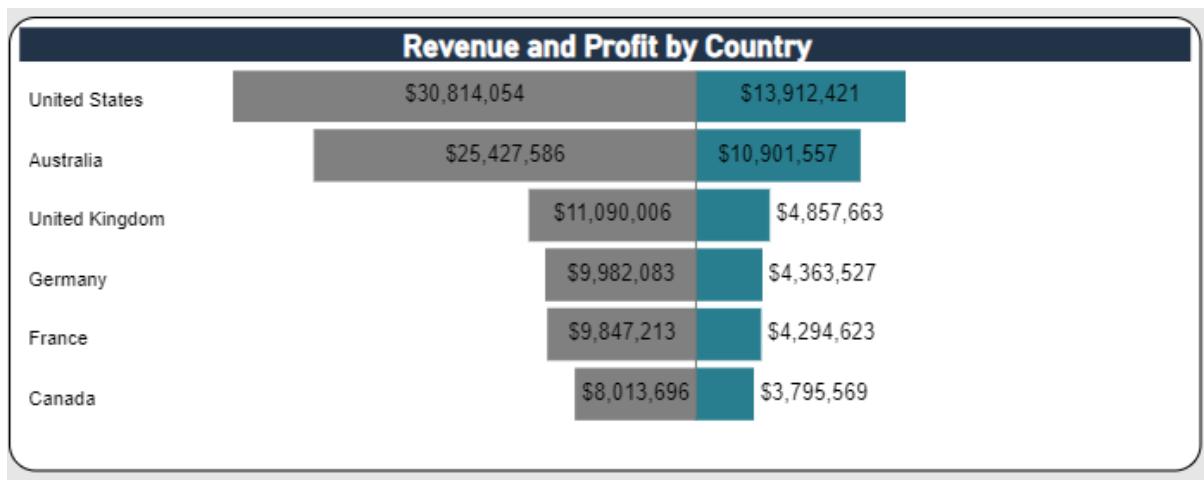


FIG 3: THE TORNADO CHART REPRESENTING THE REVENUE AND PROFIT GENERATED BY EACH COUNTRY

The chart in the figure above was created using the Tornado chart visual in Power BI. The countries were used for the group and the revenue and profit for the values. This chart shows the amount of revenue and profit each country generated.

The chart shows that the United States had the highest revenue and profit as compared to the other countries.

Question 2

Which product category and subcategory had the highest orders and generated the most profit?

Answer:

The different charts used to answer this question were created in the product analysis dashboard as shown below:

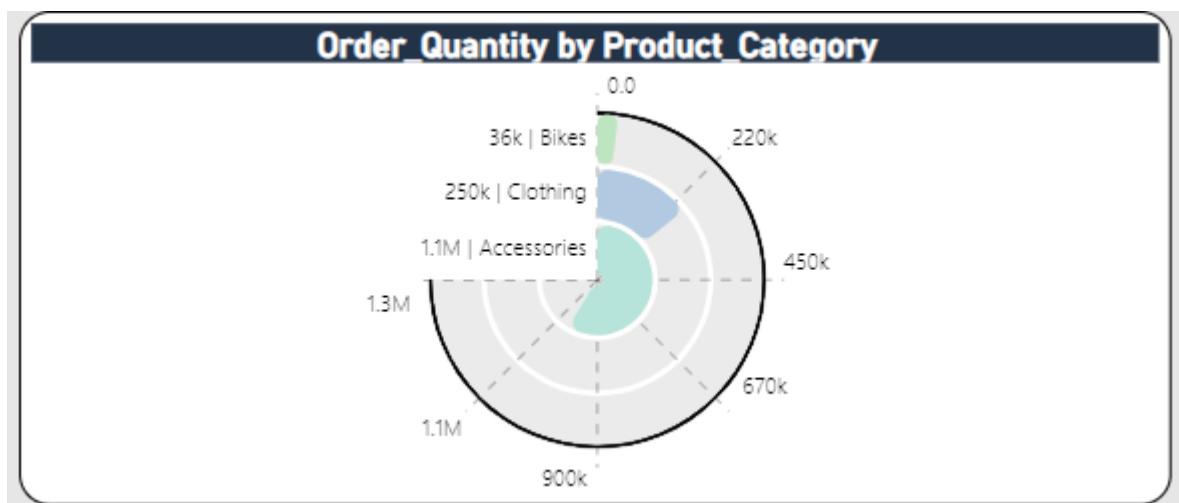


FIG 4:THE DRILL-DOWN BAR CHART REPRESENTING THE ORDER QUANTITY BY PRODUCT CATEGORY

The chart in the figure above was created using the Drill-down bar chart in Power BI. The product categories were used for the categories and the order quantity for the measures. This chart shows the order quantity for each product category. The chart shows that accessories had the highest order quantity.



FIG 5:THE CLUSTERED COLUMN BAR CHART REPRESENTING THE ORDER QUANTITY BY PRODUCT SUB-CATEGORY

The chart in the figure above was created using Power BI's clustered column bar chart. The product categories were used for the x-axis and the order quantity for the y-axis. This chart shows the order quantity for the top 10 product subcategory. The chart shows the top 10 sub-categories with the highest order quantity with Tires and tubes as the subcategory with the highest order.



FIG 6:THE SIMPLE WATERFALL CHART REPRESENTING THE PROFIT GENERATED BY EACH PRODUCT CATEGORY

The chart in the figure above was created using Power BI's simple waterfall chart. The product categories were used for the categories and the profit for the values. This chart shows the profit generated by each product category. The chart shows that the bikes category generated the highest profit.

Question 3

Which customer age group and gender had the most orders?

Answer:

The different charts used to answer this question were created in the customer analysis dashboard as shown below:

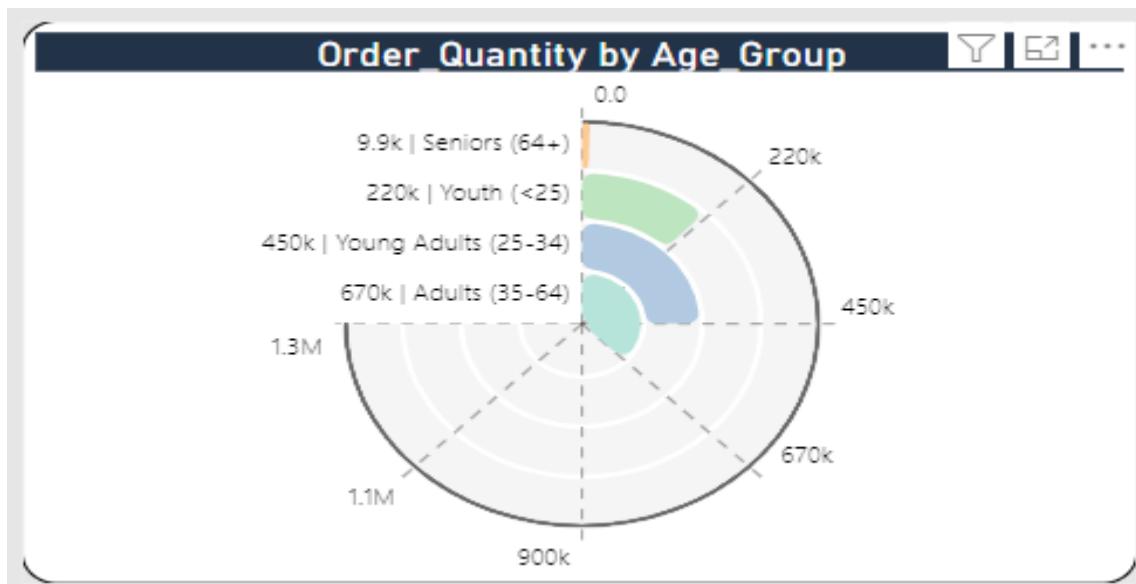


FIG 7:THE DRILL-DOWN BAR CHART REPRESENTING THE ORDER OF QUANTITY OF PRODUCTS BY EACH AGE GROUP

The chart in the figure above was created using the Drill-down bar chart in Power BI. The age groups of the customers were used for the categories and the order quantity for the measures. This chart shows the order quantity for each age group. The chart shows that the age group for adults that is (35-64) had the highest order quantity.

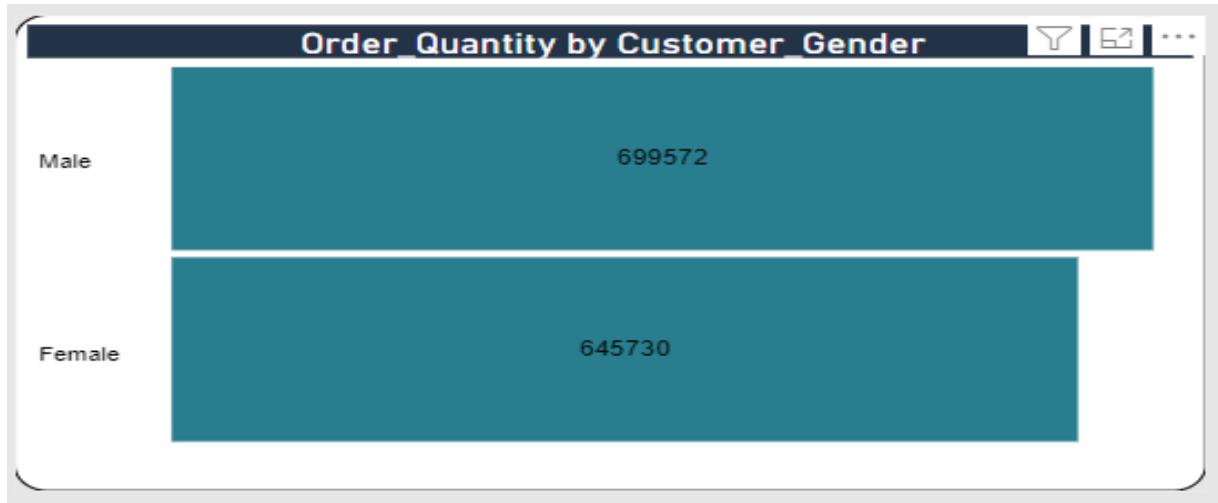


Fig 8:THE TORNADO CHART REPRESENTING THE ORDER OF QUANTITY OF PRODUCTS BY CUSTOMER GENDER

The chart in the figure above was created using the tornado visual in Power BI. The customer gender of the customers was used for the categories and the order quantity for the measures. This chart shows the order quantity for gender. The chart shows that the Males had the highest order quantity.

Question 4

Which year was the most profitable?

Answer:

The chart used to answer this question were created in the annual sales analysis dashboard as shown below:

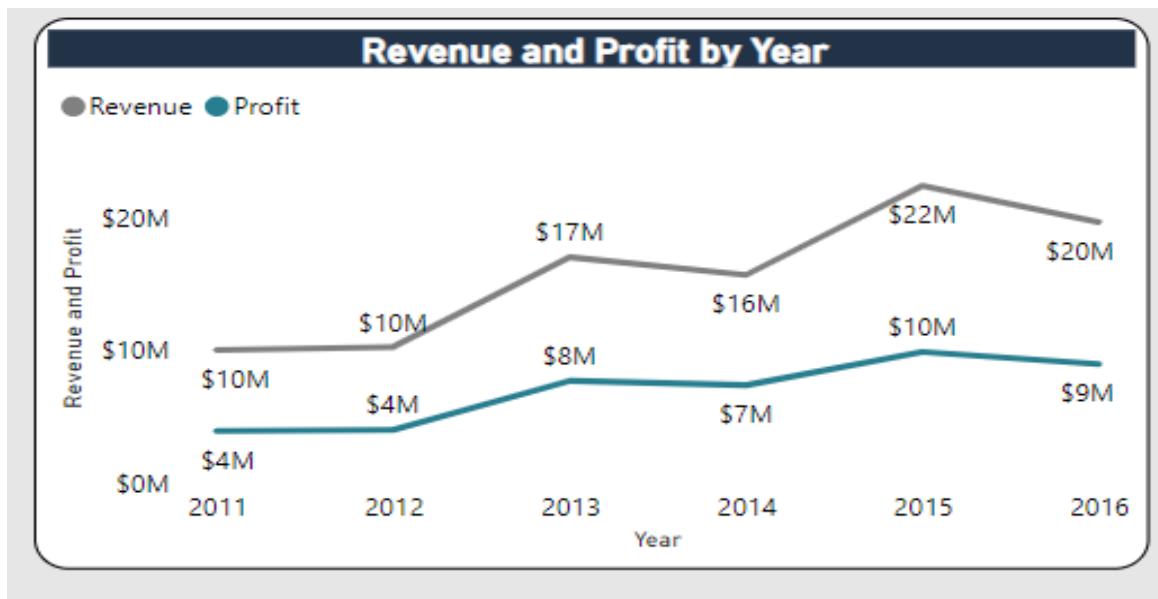


FIG 9:THE LINE CHART REPRESENTING THE REVENUE AND PROFIT GENERATED BY YEAR

The chart in the figure above was created using the line chart in Power BI. The x-axis represented the year and revenue and profit was represented on the y-axis. This chart shows the revenue and profit generated each year. The chart shows that 2015 was the most profitable year.

Question 5

What was the profit margin for the business and what it means for the business?

Answer:

This question was analysed by calculating the formula using DAX and a card was used to represent it on the annual sales analysis dashboard as shown below:

Formula using DAX:

$$\text{Profit_Margin} = (\text{SUM}(\text{Sales}[\text{Revenue}]) - \text{SUM}(\text{Sales}[\text{Total_cost}])) / \text{SUM}(\text{Sales}[\text{Revenue}])$$



FIG 10:THE PROFIT MARGIN CALCULATED

The figure above was created using the card visual in Power BI. The calculated value using the above formula is used as the field for this represented. This card shows the profit margin for the business. The profit margin measures the business profitability in the industry. Therefore a 44% profit margin means that an income of \$ 0.44 is generated for each dollar of revenue.

Question 6

What was the average profit and revenue generated and is there a correlation between them?

Answer:

This question was analysed by calculating the formula using DAX and a card was used to represent the values and a chart was used to represent their correlation on the annual sales analysis dashboard as shown below:

Formula using DAX:

Average_Revenue = SUM(Sales[Revenue])/SUM(Sales[Order_Quantity])

Average_Profit= Sales[Average_Revenue]- Sales[Average_Cost]

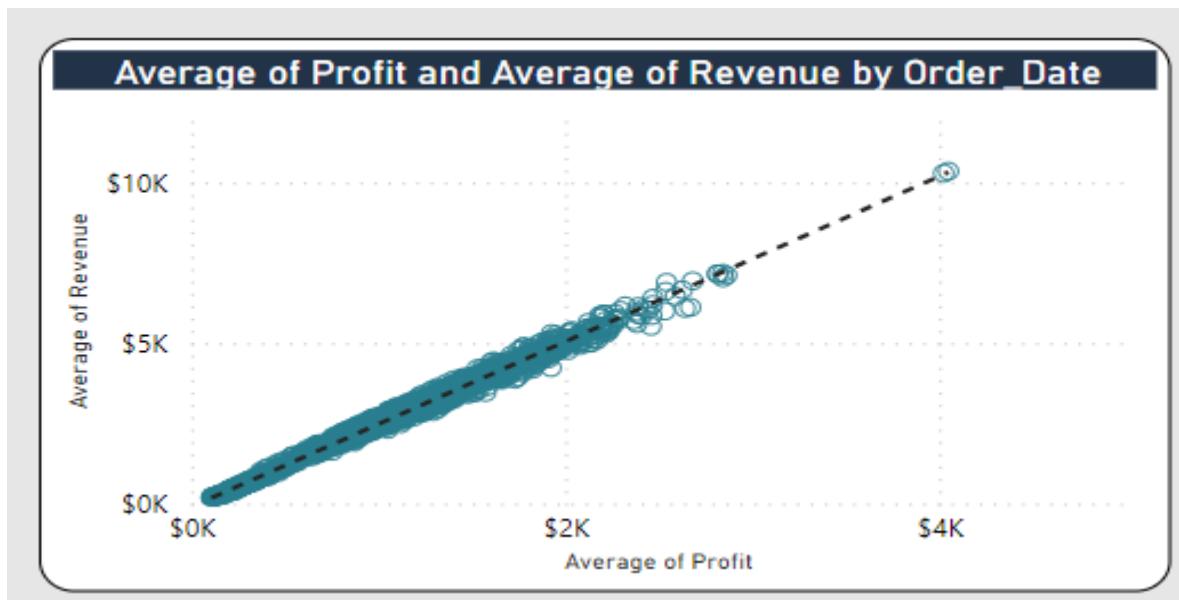


FIG 11:THE SCATTER CHART SHOWING THE CORRELATION BETWEEN AVERAGE REVENUE AND PROFIT

The chart in the figure above was created using the Scatter chart in Power BI. The x-axis represented the average profit and the average revenue was represented by the y-axis and the order date was used as the values. This chart shows the average revenue and average profit generated each year. The chart shows that there was a positive correlation between the average profit and average revenue which means that as the revenue increases the profit also increases.

Question 7

Would there be an increase or decrease in the revenue in the next year?

Answer:

The different charts used to answer this question were created in the artificial intelligence and data analytics analysis dashboard as shown below:

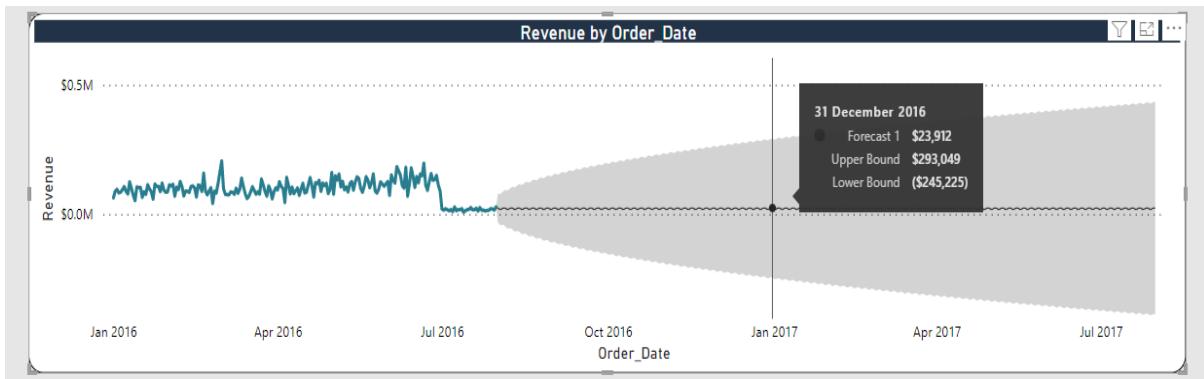


FIG 12:THE LINE CHART SHOWING THE FORECAST FOR DECEMBER 2016

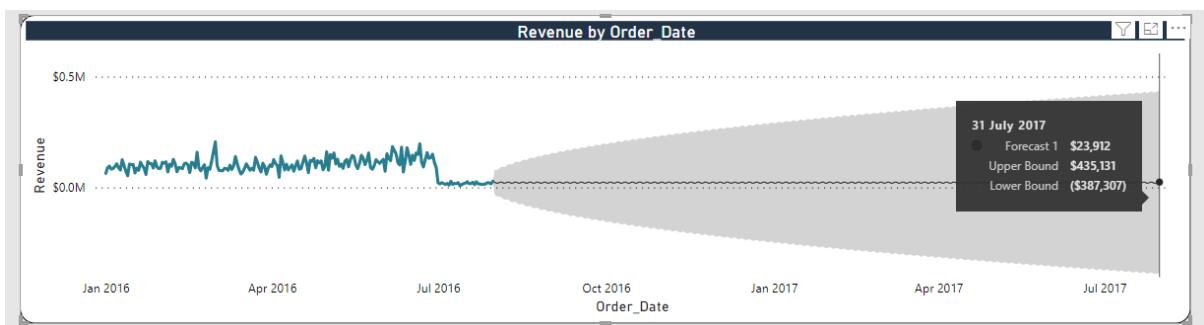


FIG 13:THE LINE CHART SHOWING THE FORECAST FOR JULY 2017

The chart in the figure above was created using the line chart in Power BI. The x-axis represented the order date and revenue was represented by the y-axis. This chart was used to predict the business revenue over the next year with the upper and lower bounds indicating the range within which this forecast will be certain. The chart shows that the forecasted revenue for the business will be the same within different ranges of certainty.

3.2 Key Findings

The following key findings from the dashboard are summarised below:

- The United States and Australia were the countries with the highest orders which generated more revenue and profit for the business.
- Canada was the country with the least generated profit and revenue
- The adult age group (35-64) ordered more of the products followed by the young adults(25-64)
- The age group with the least order quantities were the seniors(64+) which made them the age group to generate the least profit.
- Males ordered the most products as compared to females
- The product category with the highest orders was accessories followed by bikes but the bikes generated more profit than the accessories
- The clothing category generated the least profit and was the least ordered product category
- The year with the highest revenue and profit was 2015

- The years 2011 and 2015 had the same revenue and profit which was the least as compared to the other years.
- There was a decrease in revenue and profit for 2016 comparing it to 2015.
- There was a positive correlation between the average revenue and average profit

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 Recommendations Based on Findings

- More marketing should be done in Canada and European countries like Germany, the United Kingdom and France so as to increase the order for the products which will also increase revenue and profit for the business
- The clothing products should be improved in both marketing and quality in order to increase their demand on the market.
- Promotion sales too can be done to increase the sales of the products
- Different types of bikes should be manufactured to suit the senior age group to also increase their demand for bikes.
- The adult age group should be focused on improving the quality of the products they purchase since they are the highest revenue and profit age group.

4.2 Conclusion

This project exposed me to various ways Power Bi can be used to analyse a data set since this is my first time using it for a project. It helped me improve my skills such as data cleaning and preprocessing, choosing the appropriate charts to analyse the dataset and creating a business report. In general, I learnt how to create dashboards that look professional and are able to help a business make informed decisions based on the analysis of the data.

I look forward to applying the Power BI skills I learnt throughout the course and this project to real-world challenges throughout my career.

SECTION 2:

BUSINESS INTELLIGENCE DESIGN

1. DATA PRE-PROCESSING AND CLEANING

1.1 Importing the Dataset

To begin this analysis, the first step was to import the dataset in Power BI. This was done by clicking on the “Get Data” button on the home tab and this gives a drop-down that shows different ways of importing the data set. This dataset is in a CSV format so the “Text/CSV” was chosen and imported through the dialogue box that opened by clicking on “Load”. This imported the data successfully in Power BI. These steps are shown in the figures below:

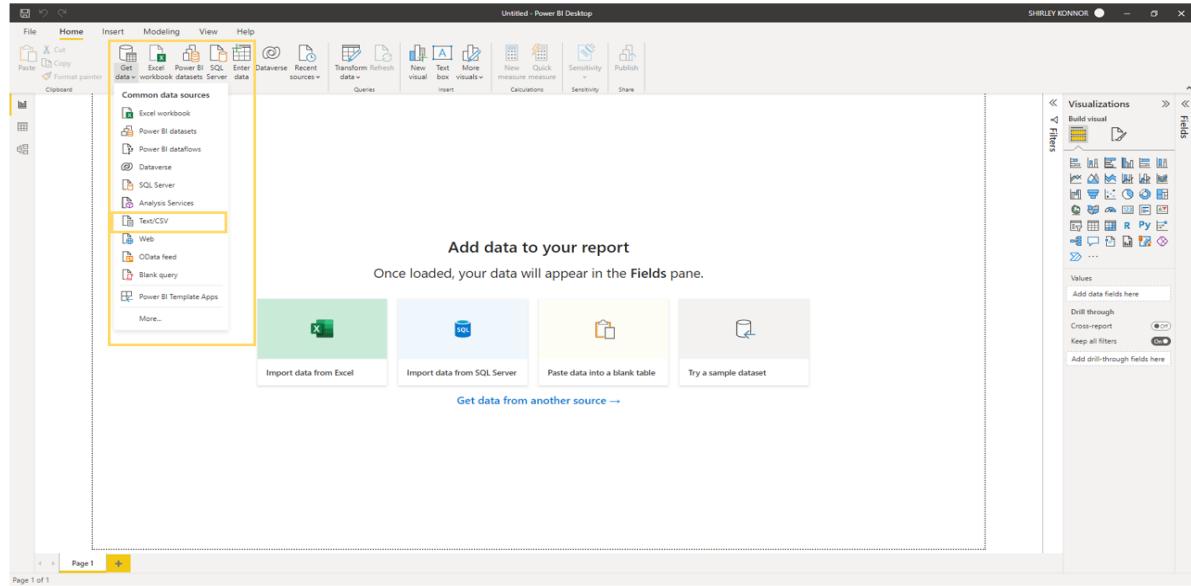


FIG 14:THE “GET DATA” DROP-DOWN SHOWING THE “TEXT/CSV” OPTION

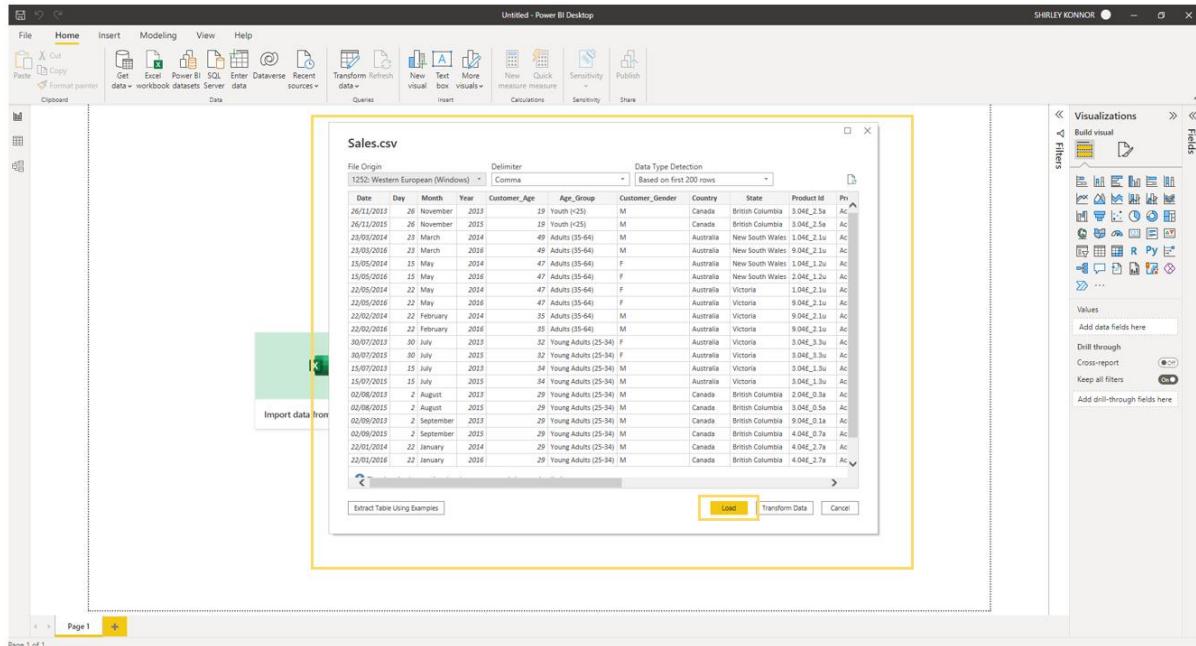


FIG 15: THE DIALOGUE SHOWING THE “LOAD” OPTION

1.2 Data Cleaning

The dataset is successfully imported into Power BI so the next step is to perform some data cleaning and preprocessing steps in the Power Query editor. To get to the Power query editor, the “Transform data” was chosen from the home tab. The figure below shows this step:

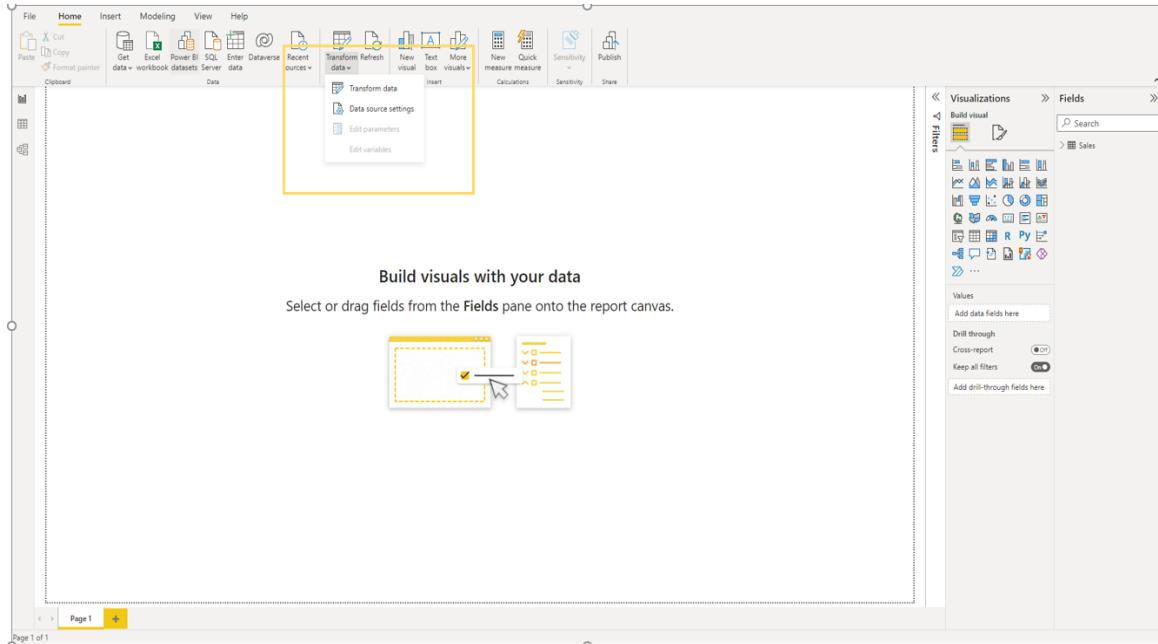


FIG 16: THE “TRANSFORM DATA” ON THE HOME TAB TO GET TO POWER QUERY EDITOR

1.2.1 Removing Errors and Columns

There were no errors in the data set some errors were created by adding a Column named “Product Id”. This column contained errors because they contained both numbers and letters that Power Bi could not convert. The column was not needed hence it was removed by right-clicking on the column and selecting “remove column”. This step is shown in the figures below:

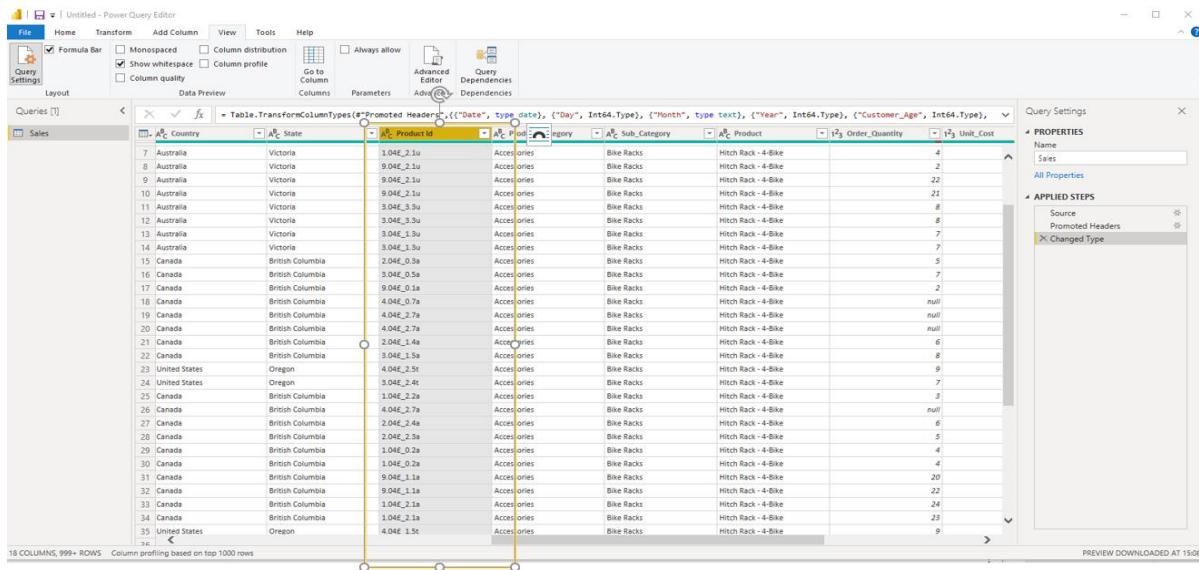


FIG 17:: THE “TRANSFORM DATA” ON THE HOME TAB TO GET TO POWER QUERY EDITOR

The screenshot shows the Power Query Editor interface with a table named 'Sales'. The 'Product Id' column is highlighted in yellow, indicating errors. A tooltip at the bottom left says 'DataFormat.Error: We couldn't convert to Number. Details: 3.04E-25a'. The 'APPLIED STEPS' pane on the right shows a step named 'Changed Type1'.

FIG 18: THE “PRODUCT ID “ COLUMN CONTAINING ERRORS.

Also, the columns for “Day”, “Month” and “Year” were removed from the data set since a column for “Date” was provided. This was done by selecting all three columns and right-clicking them to select “remove columns” from the dropdown. This step can be seen in the figure below:

The screenshot shows the Power Query Editor interface with a table named 'Sales'. The 'Day', 'Month', and 'Year' columns are highlighted in yellow, indicating they are selected for removal. The 'APPLIED STEPS' pane on the right shows a step named 'Removed Columns'.

FIG 19: THE “DAY “, “MONTH” AND “YEAR” COLUMNS SELECTED TO BE REMOVED.

1.2.2 Removing Blanks and Null

The dataset did not contain some blanks and nulls so these were created in the CSV file to show knowledge of dealing with removing blanks and nulls. These blanks were removed from the “order_quantity” column by clicking on the arrow by the column header and unselecting the tick box by the “null” option in the dialogue box. This step is shown in the figure below:

FIG 20: THE “DAY”, “MONTH” AND “YEAR” COLUMNS SELECTED TO BE REMOVED.

FIG 21: THE NULLS HAVE BEEN REMOVED FROM THE “ORDER_QUANTITY” COLUMN

1.2.3 Renaming Column

The column “Date” was renamed to “order_date”, the column “sub_category” was renamed to “product_sub_category” and the column “cost” was renamed to “Total_cost”. These were done by clicking on the column header and selecting “rename” from the dropdown. The names were then typed to replace the previous ones. This step is shown in the figure below:

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. The 'Date' column is highlighted in yellow, and the 'Sub_Category' column is also highlighted in yellow. The 'APPLIED STEPS' pane on the right shows a single step named 'Filtered Rows'. The data table contains 999+ rows and 13 columns.

FIG 22: THE COLUMNS “DATE” AND “SUB_CATEGORY” TO BE RENAMED

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. The 'Order_Date' column is highlighted in yellow, and the 'Product_Sub_Category' column is highlighted in yellow. The 'APPLIED STEPS' pane on the right shows a step named 'Renamed Columns'. The data table contains 999+ rows and 13 columns.

FIG 23: THE COLUMNS SHOWING THE RENAMED COLUMNS “ORDER_DATE” AND “PRODUCE_SUB_CATEGORY”.

The screenshot shows the Power Query Editor interface with a table named 'Table.SelectRows(*#Removed Columns*, each [Order_Quantity] > null)'. The columns are: #_c0 State, #_c1 Country, #_c2 Product_Catagory, #_c3 Sub_Catagory, #_c4 Product, #_c5 Order_Quantity, #_c6 Unit_Cost, and #_c7 Cost. The 'Cost' column is highlighted in yellow. The 'Applied Steps' pane on the right shows the step 'Renamed Column'.

FIG 24: THE COLUMNS “COST” TO BE RENAMED

The screenshot shows the Power Query Editor interface with the same table structure as Figure 24. The 'Cost' column has been renamed to 'Total_Cost'. The 'Applied Steps' pane on the right shows the step 'Renamed Column'.

FIG 25: THE COLUMN SHOWING THE RENAMED COLUMN “TOTAL_COST”.

1.2.4 Replacing Values

The values “M” and “F” in the “customer_gender” column were replaced with “Male” and “Female” respectively. This was done by highlighting the column and selecting “transform” on the home tab in Power query. Then “replace values” was selected and a dialogue box popped up. In the “values to find”, M was typed and in the “replace with”, “Male” was typed in. This step was repeated for the F too. This replaced M with Male and F with Female. The figure below shows this step:

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. The 'Customer_Gender' column is highlighted, showing values 'M' and 'F'. The 'Applied Steps' pane on the right shows the step 'Renamed Columns'.

FIG 26: THE COLUMN “CUSTOMER_GENDER” SHOWS THE “M” AND “F” VALUES.

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. A 'Replace Values' dialog box is open over the 'Customer_Gender' column, which contains 'F' values. The dialog box shows 'Value To Find' as 'F' and 'Replace With' as 'Female'. The 'OK' button is highlighted.

FIG 27: THE “REPLACE VALUE” BOX SHOWS THE “F” VALUES BEING REPLACED.

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. In the 'Applied Steps' pane, the last step is highlighted with an orange border, labeled 'Replaced Value1'. This step is associated with the formula `Table.ReplaceValue(#"Replaced Value", "F", "Female", Replacer.ReplaceText, {"Customer_Gender"})`. The main table view shows rows where the 'Customer_Gender' column has been updated from 'F' to 'Female'.

FIG 28: THE “REPLACE VALUE” BOX SHOWS THE “F” VALUES BEING REPLACED.

1.2.5 Changing data type

The columns “Unit_cost”, “Unit_price” and “Total_cost” were formatted as a whole number. this was converted to fixed decimal type. This was done by right-clicking on the top right corner of the column headers and selecting “fixed decimal number” in the drop-down box. This can be seen in the figure below:

The screenshot shows the Power Query Editor with the 'Sales 2' query selected. The 'Applied Steps' pane indicates that the 'Replaced Value1' step has been applied. The main table view shows the 'Unit_Cost', 'Unit_Price', and 'Total_Cost' columns, which have been converted to fixed decimal numbers. The 'Properties' pane shows the 'Name' is set to 'Sales 2'.

FIG 29: THE COLUMN “UNIT_PRICE”, “UNIT_COST” AND “TOTAL_COST” TO BE CONVERTED TO A FIXED DECIMAL

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. The ribbon at the top includes 'File', 'Home', 'Transform', 'Add Column', 'View', 'Tools', and 'Help'. The 'Transform' tab is active. The main area displays a table with columns: 'Index', 'State', 'Product_Category', 'Product_Sub_Category', 'Product', 'Order_Quantity', 'Unit_Cost', 'Unit_Price', and 'Total_Cost'. The 'Unit_Cost' and 'Unit_Price' columns are highlighted in yellow, indicating they have been changed to fixed decimal numbers. The 'Total_Cost' column is also highlighted. The 'Applied Steps' pane on the right lists steps like 'Promoted Headers', 'Changed Type', and 'Replaced Value1'. A specific step, 'Changed Type1', is highlighted with a yellow box and has the description 'Replaced Value1'.

FIG 30: THE COLUMN “UNIT_PRICE”, “UNIT_COST” AND “TOTAL_COST” CHANGED TO FIXED DECIMALS.

1.2.6 Adding a column (Index column) and Reordering the column

An “index” column was added to the dataset and this was done by clicking on the “Add column” tab and selecting the “index column” button. Then select the start number for the index column and then to reorder the column by dragging it to the left of the table. The following figures show these steps:

The screenshot shows the Power Query Editor interface with the 'Sales' query selected. The ribbon at the top includes 'File', 'Home', 'Transform', 'Add Column', 'View', 'Tools', and 'Help'. The 'Add Column' tab is active. The main area displays a table with columns: 'Order_Date', 'Customer_Age', 'Age_Group', 'Customer_Gender', 'Country', 'State', 'Product_Sub_Category', and 'Product'. The 'Order_Date' column is highlighted in yellow, indicating it is the current active column. The 'Applied Steps' pane on the right lists steps like 'Promoted Headers', 'Changed Type', and 'Replaced Value1'. A specific step, 'Changed Type1', is highlighted with a yellow box and has the description 'Replaced Value1'.

FIG 31: SELECTING THE “ADD COLUMN” TAB TO ADD AN INDEX COLUMN

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Column From Examples Custom Column Insert Column Function General

Merge Columns Index Column Duplicate Column Format Parse From Text From Number From Date & Time Text Analytics Vision AI Machine Learning

Trigonometry Statistics Standard Scientific Information Date Time Duration

10²

From Text

From Number

From Date & Time

Text Analytics

Vision

AI Insights

Query Settings

Name: Sales

All Properties

APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Removed Columns
- Filtered Rows
- Renamed Columns
- Replaced Value
- Replaced Value 1
- Change Type 1
- Added Index**
- Reordered Columns**

PREVIEW DOWNLOADED AT 1052

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

Index	Date	Customer_Age	Age_Group	Customer_Gender	Country	State	Product_Catagory	Product_Sub_Catagory	Product	Order_Quantity
1	26/11/2015	29 Youth (<25)	Male	Canada	British Columbia	Accessories	Bike Racks			
2	26/11/2015	29 Youth (<25)	Male	Canada	British Columbia	Accessories	Bike Racks			
3	23/03/2014	49 Adults (35-64)	Male	Australia	New South Wales	Accessories	Bike Racks			
4	23/03/2016	49 Adults (35-64)	Male	Australia	New South Wales	Accessories	Bike Racks			
5	25/05/2014	47 Adults (35-64)	Female	Australia	New South Wales	Accessories	Bike Racks			
6	25/05/2016	47 Adults (35-64)	Female	Australia	New South Wales	Accessories	Bike Racks			
7	22/06/2014	47 Adults (35-64)	Female	Australia	Victoria	Accessories	Bike Racks			
8	22/06/2016	47 Adults (35-64)	Female	Australia	Victoria	Accessories	Bike Racks			
9	22/03/2014	85 Adults (35-64)	Male	Australia	Victoria	Accessories	Bike Racks			
10	22/03/2016	85 Adults (35-64)	Male	Australia	Victoria	Accessories	Bike Racks			
11	30/07/2013	82 Young Adults (25-34)	Female	Australia	Victoria	Accessories	Bike Racks			
12	30/07/2015	82 Young Adults (25-34)	Female	Australia	Victoria	Accessories	Bike Racks			
13	15/07/2013	34 Young Adults (25-34)	Male	Australia	Victoria	Accessories	Bike Racks			
14	15/07/2015	34 Young Adults (25-34)	Male	Australia	Victoria	Accessories	Bike Racks			
15	02/08/2013	29 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
16	02/08/2015	29 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
17	02/08/2013	29 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
18	17/05/2014	29 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
19	17/05/2016	29 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
20	27/03/2014	51 Adults (35-64)	Male	United States	Oregon	Accessories	Bike Racks			
21	27/03/2016	51 Adults (35-64)	Male	United States	Oregon	Accessories	Bike Racks			
22	29/05/2013	49 Adults (35-64)	Male	Canada	British Columbia	Accessories	Bike Racks			
23	29/05/2015	49 Adults (35-64)	Male	Canada	British Columbia	Accessories	Bike Racks			
24	26/12/2013	49 Adults (35-64)	Male	Canada	British Columbia	Accessories	Bike Racks			
25	03/01/2014	48 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
26	03/01/2016	48 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
27	13/03/2014	48 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
28	13/03/2016	48 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
29	23/05/2014	47 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
30	23/05/2016	47 Adults (35-64)	Female	Canada	British Columbia	Accessories	Bike Racks			
31	19/07/2013	32 Young Adults (25-34)	Female	United States	Oregon	Accessories	Bike Racks			
32	19/07/2015	32 Young Adults (25-34)	Female	United States	Oregon	Accessories	Bike Racks			
33	11/11/2013	34 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
34	11/11/2015	34 Young Adults (25-34)	Male	Canada	British Columbia	Accessories	Bike Racks			
35	13/07/2013	43 Adults (35-64)	Female	United States	California	Accessories	Bike Racks			
36	13/07/2015	43 Adults (35-64)	Female	United States	California	Accessories	Bike Racks			
37	31/03/2014	43 Adults (35-64)	Female	United States	California	Accessories	Bike Racks			
38	31/03/2016	43 Adults (35-64)	Female	United States	California	Accessories	Bike Racks			

FIG 32: THE “INDEX” COLUMN ADDED AND REORDERED

2. DATA MODELLING

This is where our Fact table was normalised into Dimension tables to avoid data duplication and to increase the data quality, data consistency and proper structure for easy analysis and visual creation.

2.1 Creating Dimension Tables

The first dimension created was the Customer table and this was also done in the Power Query editor. The Customer table was created by duplicating the Sales table and renaming it. This was done by right-clicking on the Sales table and selecting “duplicate” in the drop-down box and then the duplicated table is renamed to “Customer” by right-clicking it and selecting “rename”. After renaming the duplicated tablet to “Customer”, every column was deleted excluding the “Index”, “Customer_age”, “Age_group” and “Customer_gender”. This was done by selecting these columns and then right-clicking on them to select “remove other columns”. The figure below shows the Customer table created:

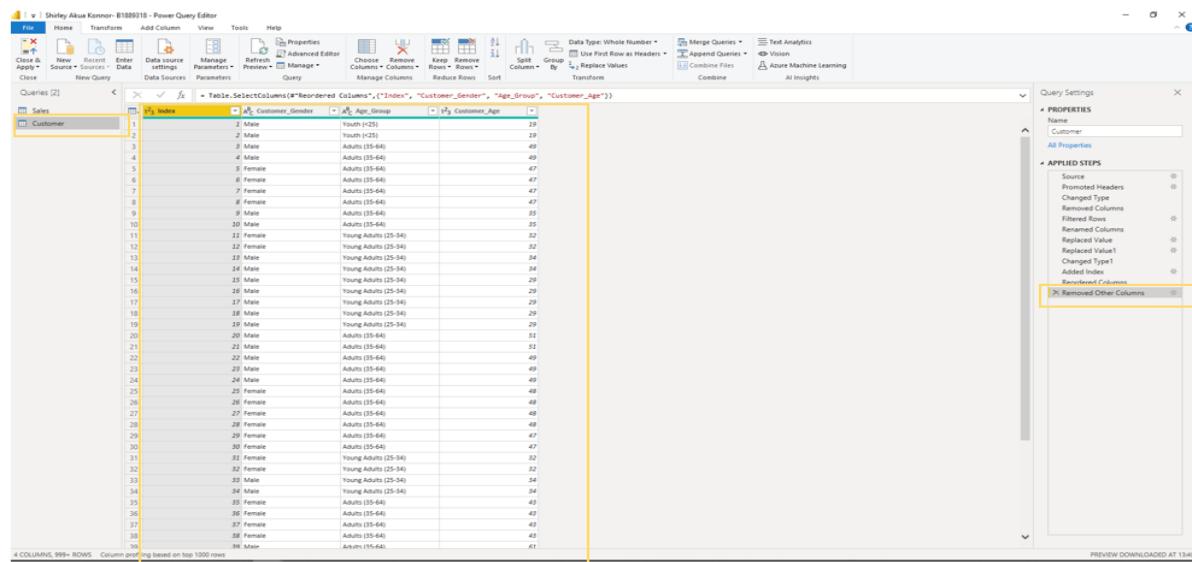


FIG 33: THE “CUSTOMER” DIMENSION TABLE WAS CREATED.

The next dimension created was the Product Dimension table. This was also done by right-clicking on the Sales table and selecting “duplicate” in the drop-down box and then the duplicated table was renamed to “Product” by right-clicking it and selecting “rename”. After renaming the duplicated tablet to “Product”, every column was deleted excluding the “Index”, “Product_category”, “Product_sub_category”, “Product” and “Order_quantity” columns. This was done by selecting these columns and then right-clicking on them to select “remove other columns”. In the Product dimension table, duplicates were removed by selecting all the columns and right-clicking to select “remove duplicates” in the drop-down box. The figure below shows the Product table created:

The screenshot shows the Power Query Editor interface with the 'Customer' dimension table loaded. The table contains approximately 300 rows and 5 columns. The 'Product' column is highlighted with a yellow box. The 'APPLIED STEPS' pane on the right shows a step named 'Removed Duplicates'.

FIG 34: THE “CUSTOMER” DIMENSION TABLE WAS CREATED.

The next dimension table created was the Location table. This was also done by right-clicking on the Sales table and selecting “duplicate” in the drop-down and then the duplicated table was renamed to “Location” by right-clicking it and selecting “rename”. After renaming the duplicated table to “Location”, every column was deleted excluding the “Country” and “State” columns. This was done by selecting these columns and then right-clicking on them to select “remove other columns”. In the Location dimension table, duplicates were removed by selecting all the columns and right-clicking to select “remove duplicates” in the drop-down box. The figure below shows the Location table created:

The screenshot shows the Power Query Editor interface with the 'Location' dimension table loaded. The table contains 53 rows and 2 columns. The 'Country' column is highlighted with a yellow box. The 'APPLIED STEPS' pane on the right shows a step named 'Removed Duplicates'.

FIG 35: THE “LOCATION” DIMENSION TABLE WAS CREATED.

The final dimension table created was the Calendar table. This was done by right-clicking on the Sales table and selecting “duplicate” in the drop box and then the duplicated table was renamed to “Calender” by right-clicking it and selecting “rename”. After renaming the duplicated table to “Calendar”, every column was deleted excluding the “Index” and “Order_date” columns. This was done by selecting these columns and then right-clicking on them to select “remove other columns”.

The screenshot shows the Microsoft Power Query Editor interface. In the left pane, under 'Queries [5]', the 'Calendar' query is selected. The main area displays a table with two columns: 'Index' and 'Order_Date'. The 'Index' column contains integers from 1 to 999, and the 'Order_Date' column contains dates starting from 26/12/2003. The 'Applied Steps' pane on the right lists several steps, with the 'Removed Other Columns' step highlighted by a yellow box. The 'Properties' pane on the far right shows the query is named 'Calender'.

FIG 36: THE “CALENDAR” DIMENSION TABLE WAS CREATED.

2.2 Creating Fact Table

The Sales table was modified to create the Fact table. Columns were removed from the sales table except for the “Index”, “country”, “Order_quantity”, “unit_cost”, “unit_price” and “Total_cost” columns. This was done by selecting these columns and then right-clicking on them to select “remove other columns”. The figure below shows the Sales table created:

The screenshot shows the Power Query Editor interface. In the center, there's a table titled "Sales" with the following data:

Index	Country	Order_Quantity	Unit_Cost	Unit_Price	Total_cost
1	Canada	8	45.00	120.00	360.00
2	Canada	8	45.00	120.00	360.00
3	Australia	23	45.00	120.00	1,035.00
4	Australia	20	45.00	120.00	900.00
5	Australia	4	45.00	120.00	180.00
6	Australia	5	45.00	120.00	225.00
7	Australia	4	45.00	120.00	180.00
8	Australia	2	45.00	120.00	90.00
9	Australia	22	45.00	120.00	990.00
10	Australia	21	45.00	120.00	945.00
11	Australia	8	45.00	120.00	360.00
12	Australia	8	45.00	120.00	360.00
13	Australia	7	45.00	120.00	315.00
14	Australia	7	45.00	120.00	315.00
15	Canada	5	45.00	120.00	225.00
16	Canada	7	45.00	120.00	315.00
17	Canada	2	45.00	120.00	90.00
18	Canada	6	45.00	120.00	270.00
19	Canada	8	45.00	120.00	360.00
20	United States	9	45.00	120.00	405.00
21	United States	7	45.00	120.00	315.00
22	Canada	3	45.00	120.00	135.00
23	Canada	6	45.00	120.00	270.00
24	Canada	5	45.00	120.00	225.00
25	Canada	4	45.00	120.00	180.00
26	Canada	4	45.00	120.00	180.00
27	Canada	20	45.00	120.00	900.00
28	Canada	22	45.00	120.00	990.00
29	Canada	24	45.00	120.00	1,080.00
30	Canada	23	45.00	120.00	1,035.00
31	United States	9	45.00	120.00	405.00
32	United States	9	45.00	120.00	405.00
33	Canada	25	45.00	120.00	1,125.00
34	Canada	23	45.00	120.00	1,035.00
35	United States	19	45.00	120.00	855.00
36	United States	10	45.00	120.00	450.00
37	United States	11	45.00	120.00	495.00
38	United States	11	45.00	120.00	495.00
39	United States	8	45.00	120.00	360.00

Below the table, it says "COLUMNS: 999 ROWS: 3999" and "Column profiling based on top 1000 rows".

FIG 37: THE “SALES” FACT TABLE BEING MODIFIED

After all the dimension and fact tables were created, the figure below shows what the data model looks like before relationships were created:

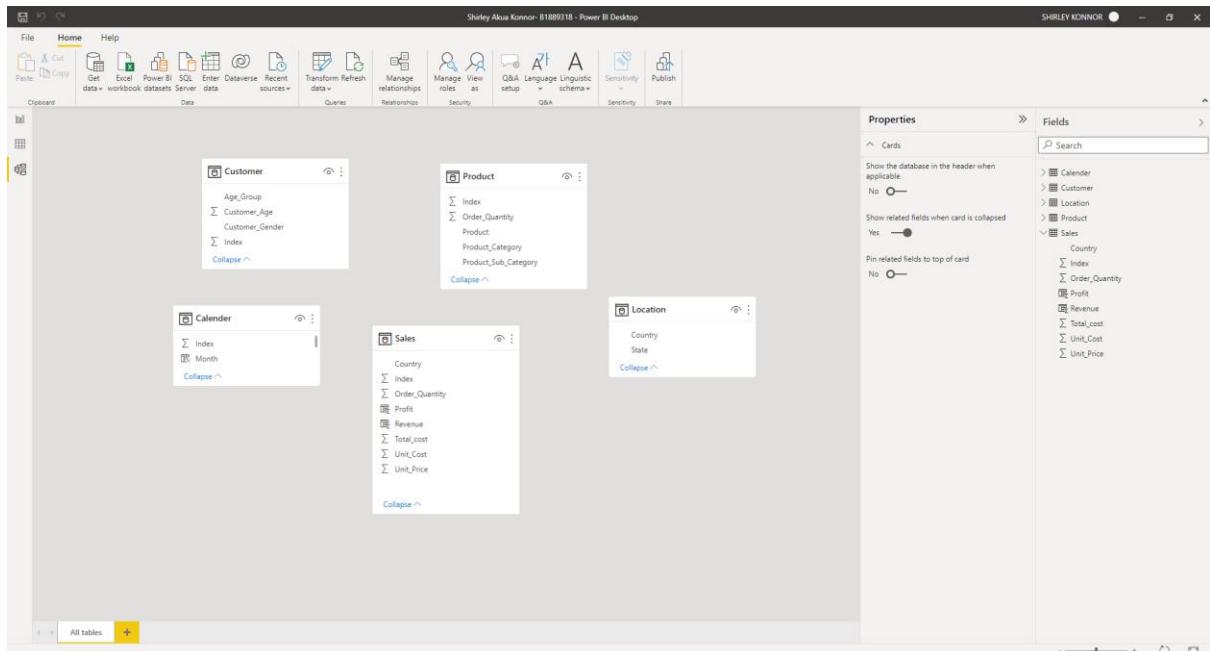


FIG 38: THE DATA MODEL BEFORE CREATING RELATIONSHIPS

2.3 Creating a Bridge Table

In order to avoid a many-to-many relationship in the model, a bridge table was created between the Fact table(Sales table) and the dimension table (Location). This bridge table was named “Country”. This was done by duplicating the “Location Table” and all columns except “country” were removed. Duplicates were also removed as shown in the figure below:

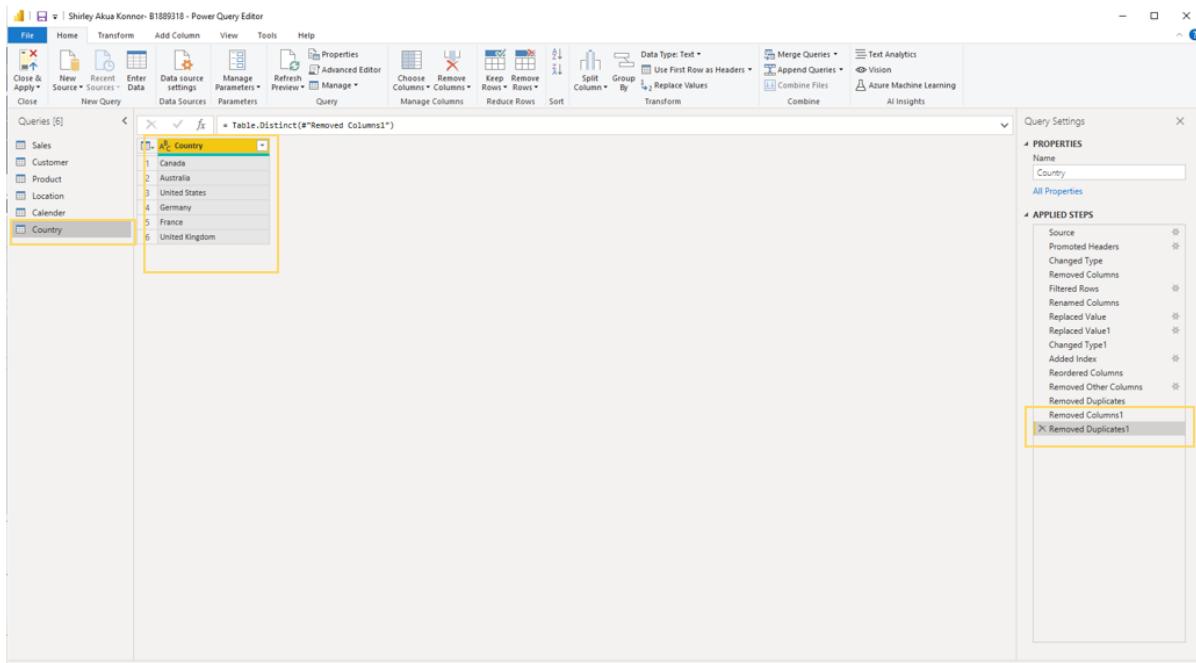


FIG 39: THE “COUNTRY” BRIDGE TABLE CREATED

2.4 Creating Relationships

In order to perform analysis for this project, relationships were created between the Fact table and Dimension tables in the model. This was done by clicking on the data model section in Power BI, and then selecting “Manage Relationships” from the home tab. A dialogue box popped up and “New” was selected and each relationship was created based on the connecting column in each dimension table in the model. Relationships with Cardinality one-to-one and many-to-one were created. A Snowflakes schema model was created after all the relationships were created. This means that one dimension table was connected through a bridge table to the Fact table. The other three dimension tables were connected to the fact table as well. The Fact table was the Sales table and the Dimensions tables were Customer, Product, Location, Calender and the bridge table was Country. These steps are shown in the figure below:

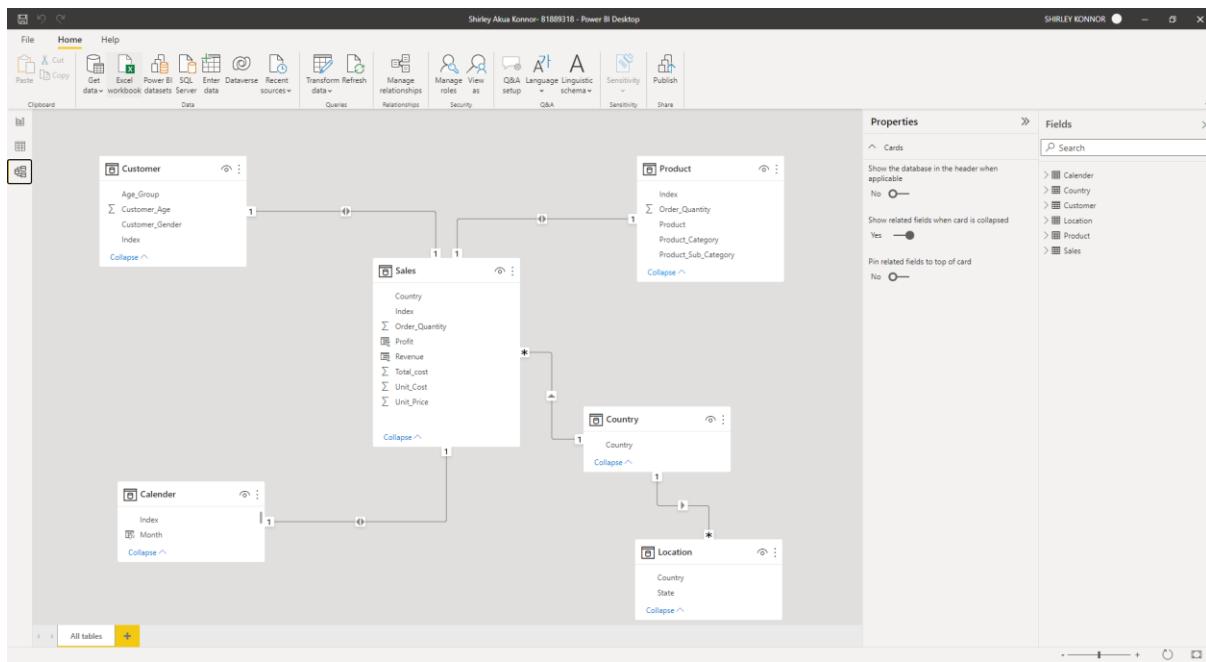


FIG 40: THE DATA MODEL AFTER CREATING RELATIONSHIPS

3. USING DAX AND M LANGUAGE

3.1 M language used in data pre-processing

During data pre-processing, M language was used to enhance the model and to help repeat the same queries more quickly. This was done in the Power Query editor, to get to the Power query editor, the “Transform data” was chosen from the home tab. Then “advanced editor” was selected in the home tab of the power query editor as shown in the figures below:

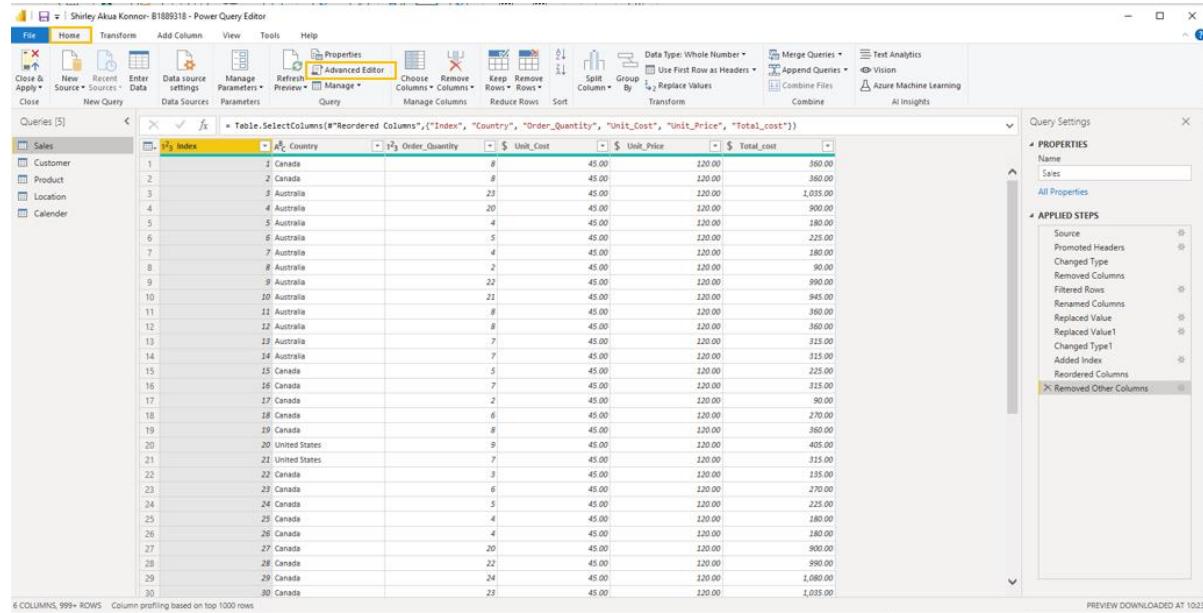


FIG 41: THE POWER QUERY EDITOR SHOWING THE ADVANCED EDITOR

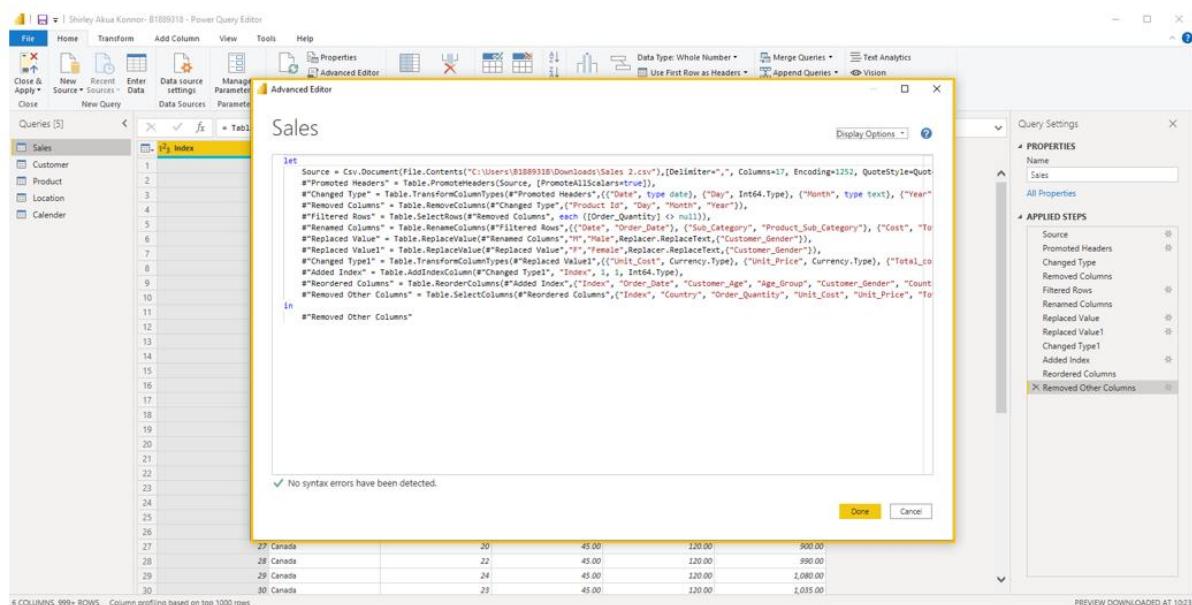


FIG 42: THE ADVANCED EDITOR SHOWING THE MLANGUAGE USED FOR THE SALES FACT TABLE.

3.2 Adding columns to the Calender dimension table using DAX

In the Calendar dimension table, a few columns were added using Dax formulas. The columns created were “Year”, “Month” and “Quarter”. This was done by going to the “Data” view in Power BI, then clicking on the calender table, and then the “New column” tab selected in the “Table Tools” ribbon. The following DAX formulas were used:

Year = 'Calender'[Order_Date].[Year]

Month = 'Calender'[Order_Date].[Month]

Qtr = 'Calender'[Order_Date].[Quarter]

These steps are shown in the figures below:

FIG 43: THE DATA VIEW WHERE THE YEAR, MONTH AND QUARTER COLUMNS WERE CREATED WITH DAX

FIG 44: THE “YEAR” COLUMN BEING CREATED WITH DAX

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. In the 'Structure' section, a new column named 'Month' is being defined with the formula `Month = 'Calendar'[Order_Date], [Month]`. The 'Data type' is set to 'Text'. The 'Fields' pane on the right shows the hierarchy of the 'Calendar' table, with 'Month' highlighted.

FIG 45: THE “MONTH” COLUMN BEING CREATED WITH DAX

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. In the 'Structure' section, a new column named 'Qtr' is being defined with the formula `Qtr = 'Calendar'[Order_Date], [Quarter]`. The 'Data type' is set to 'Text'. The 'Fields' pane on the right shows the hierarchy of the 'Calendar' table, with 'Qtr' highlighted.

FIG 46: THE “QTR” COLUMN BEING CREATED WITH DAX

3.3 Adding columns to the Fact table using DAX

After a thorough examination of the dataset, it was realized that the columns “Revenue” and “Profit” were not accurately calculated. This can be seen in the figure below:

A screenshot of Microsoft Excel showing a table of Bike Sales data. The table includes columns for Date, Day, Month, Year, Customer_Age_Group, Customer_Country, State, Product_CSub, Sales_Person, Order_Quantity, Unit_Cost, Unit_Price, Profit, Cost, Revenue, and Profitability. The Profit and Revenue columns are highlighted with a yellow border.

FIG 47: THE “REVENUE” AND “PROFIT” COLUMN WERE WRONGLY CALCULATED

Therefore these columns were calculated using DAX formulas. This was done by going to the “Data” view in Power BI, then clicking on the Sales table, and then the “New column” tab selected in the “Table Tools” ribbon. The following DAX formulas were used:

$$\text{Revenue} = \text{'Sales'}[\text{Order_quantity}] * \text{'Sales'}[\text{Unit_price}]$$

$$\text{Profit} = \text{'Sales'}[\text{Revenue}] - \text{'Sales'}[\text{Total_cost}]$$

These steps are shown in the figures below:

A screenshot of the Power BI Data View showing the Sales table. The Fields pane on the right shows the Sales table selected. The table itself displays data with columns: Country, Order_Quantity, Unit_Cost, Unit_Price, Total_cost, and Index. The Revenue and Profit columns are not yet visible in this view.

FIG 48: THE DATA VIEW WHERE THE REVENUE AND PROFIT COLUMNS WERE CREATED WITH DAX

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. A new column named 'Revenue' is being created, defined by the DAX formula: `Revenue = Sales[Revenue] * Sales[Unit_Price]`. The 'Format' dropdown is set to 'Currency'. The 'Summarization' dropdown is set to 'Sum'. The 'Data type' dropdown is set to 'Fixed decimal number'. The 'Name' dropdown is set to 'Revenue'. The 'Fields' pane on the right shows the Sales table with various columns like Order_Quantity, Unit_Cost, and Unit_Price.

FIG 49: THE REVENUE COLUMN CREATED WITH DAX

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. A new column named 'Profit' is being created, defined by the DAX formula: `Profit = 'Sales'[Revenue] - 'Sales'[Total_cost]`. The 'Format' dropdown is set to 'Currency'. The 'Summarization' dropdown is set to 'Sum'. The 'Data type' dropdown is set to 'Fixed decimal number'. The 'Name' dropdown is set to 'Profit'. The 'Fields' pane on the right shows the Sales table with various columns like Order_Quantity, Unit_Cost, and Unit_Price.

FIG 50: THE PROFIT COLUMN CREATED WITH DAX

3.2.1 Changing data types in Sales Fact Table

The columns “unit_cost”, “unit_price”, “Total_cost”, “Revenue” and “Profit” were changed to English(United States currency). This was done by selecting the currency on the “columns tools” tab for each column. The following figures show this steps:

Shirley Alusa Konnor - B1889318 - Power BI Desktop

SIRLEY KONNOR

File Home Help Table tools Column tools

Name Revenue Data type Fixed decimal numbers... Format Currency Summarization Sum Data category Uncategorized

Structure Sort by column Sort Groups Manage relationships New column Calculations

Properties

1 Revenue = 'Sales'[Order_Quantity]*Sales[Unit_Price]

Country Order_Quantity Unit_Cost Unit_Price Total_cost Index Revenue Profit

United States 1 2 \$5 2 1146 \$5.00 \$3
United States 1 2 \$5 2 1632 \$5.00 \$3
United States 1 2 \$5 2 1634 \$5.00 \$3
United States 1 2 \$5 2 1647 \$5.00 \$3
United States 1 2 \$5 2 1665 \$5.00 \$3
United States 1 2 \$5 2 1670 \$5.00 \$3
United States 1 2 \$5 2 1674 \$5.00 \$3
United States 1 2 \$5 2 1829 \$5.00 \$3
United States 1 2 \$5 2 1830 \$5.00 \$3
United States 1 2 \$5 2 1844 \$5.00 \$3
United States 1 2 \$5 2 2668 \$5.00 \$3
United States 1 2 \$5 2 3134 \$5.00 \$3
United States 1 2 \$5 2 3602 \$5.00 \$3
United States 1 2 \$5 2 3605 \$5.00 \$3
United States 1 2 \$5 2 3630 \$5.00 \$3
United States 1 2 \$5 2 4094 \$5.00 \$3
United States 1 2 \$5 2 4779 \$5.00 \$3
United States 1 2 \$5 2 4787 \$5.00 \$3
United States 1 2 \$5 2 4792 \$5.00 \$3
United States 1 2 \$5 2 4801 \$5.00 \$3
United States 1 2 \$5 2 4851 \$5.00 \$3
United States 1 2 \$5 2 4853 \$5.00 \$3
United States 1 2 \$5 2 4855 \$5.00 \$3
United States 1 2 \$5 2 4861 \$5.00 \$3
United States 1 2 \$5 2 4863 \$5.00 \$3
United States 1 2 \$5 2 4866 \$5.00 \$3
United States 1 2 \$5 2 4867 \$5.00 \$3
United States 1 2 \$5 2 4868 \$5.00 \$3
United States 1 2 \$5 2 4957 \$5.00 \$3
United States 1 2 \$5 2 5002 \$5.00 \$3
United States 1 2 \$5 2 5021 \$5.00 \$3
United States 1 2 \$5 2 5022 \$5.00 \$3
United States 1 2 \$5 2 5332 \$5.00 \$3
United States 1 2 \$5 2 5334 \$5.00 \$3
United States 1 2 \$5 2 6155 \$5.00 \$3
United States 1 2 \$5 2 6177 \$5.00 \$3
United States 1 2 \$5 2 6178 \$5.00 \$3

Table: Sales (113,022 rows) Column: Revenue (436 distinct values)

FIG 51: THE DATA TYPES BEEN CHANGED FOR THE COLUMNS

Shirley Alusa Konnor - B1889318 - Power BI Desktop

SIRLEY KONNOR

File Home Help Table tools Column tools

Name Profit Data type Fixed decimal numbers... Format Currency Summarization Sum Data category Uncategorized

Structure Sort by column Sort Groups Manage relationships New column Calculations

Properties

1 Profit = 'Sales'[Revenue]-Sales[Total_cost]

Country Order_Quantity Unit_Cost Unit_Price Total_cost Index Revenue Profit

United States 1 \$2 \$5 \$2 1146 \$5.00 \$3
United States 1 \$2 \$5 \$2 1632 \$5.00 \$3
United States 1 \$2 \$5 \$2 1634 \$5.00 \$3
United States 1 \$2 \$5 \$2 1647 \$5.00 \$3
United States 1 \$2 \$5 \$2 1665 \$5.00 \$3
United States 1 \$2 \$5 \$2 1670 \$5.00 \$3
United States 1 \$2 \$5 \$2 1674 \$5.00 \$3
United States 1 \$2 \$5 \$2 1829 \$5.00 \$3
United States 1 \$2 \$5 \$2 1830 \$5.00 \$3
United States 1 \$2 \$5 \$2 1844 \$5.00 \$3
United States 1 \$2 \$5 \$2 2668 \$5.00 \$3
United States 1 \$2 \$5 \$2 3134 \$5.00 \$3
United States 1 \$2 \$5 \$2 3602 \$5.00 \$3
United States 1 \$2 \$5 \$2 3605 \$5.00 \$3
United States 1 \$2 \$5 \$2 3630 \$5.00 \$3
United States 1 \$2 \$5 \$2 4094 \$5.00 \$3
United States 1 \$2 \$5 \$2 4779 \$5.00 \$3
United States 1 \$2 \$5 \$2 4787 \$5.00 \$3
United States 1 \$2 \$5 \$2 4792 \$5.00 \$3
United States 1 \$2 \$5 \$2 4801 \$5.00 \$3
United States 1 \$2 \$5 \$2 4851 \$5.00 \$3
United States 1 \$2 \$5 \$2 4853 \$5.00 \$3
United States 1 \$2 \$5 \$2 4855 \$5.00 \$3
United States 1 \$2 \$5 \$2 4861 \$5.00 \$3
United States 1 \$2 \$5 \$2 4863 \$5.00 \$3
United States 1 \$2 \$5 \$2 4866 \$5.00 \$3
United States 1 \$2 \$5 \$2 4867 \$5.00 \$3
United States 1 \$2 \$5 \$2 4868 \$5.00 \$3
United States 1 \$2 \$5 \$2 4957 \$5.00 \$3
United States 1 \$2 \$5 \$2 5002 \$5.00 \$3
United States 1 \$2 \$5 \$2 5021 \$5.00 \$3
United States 1 \$2 \$5 \$2 5022 \$5.00 \$3
United States 1 \$2 \$5 \$2 5332 \$5.00 \$3
United States 1 \$2 \$5 \$2 5334 \$5.00 \$3
United States 1 \$2 \$5 \$2 6155 \$5.00 \$3
United States 1 \$2 \$5 \$2 6177 \$5.00 \$3
United States 1 \$2 \$5 \$2 6178 \$5.00 \$3

Table: Sales (113,022 rows) Column: Profit (309 distinct values)

FIG 52: THE DATA TYPES CHANGED FOR THE COLUMNS

3.4 Creating a new measure using DAX

A new measure was added to the data for analysis. The measure created were “Average_Revenue”, “Average_Cost ” and “Profit_Margin”. This was done by going to the “Data” view in Power BI, then clicking on the sales table, and then the “New measure” tab selected in the “Table Tools” ribbon. The following DAX formulas were used:

$$\text{Average_Revenue} = \text{SUM}(\text{Sales}[Revenue])/\text{SUM}(\text{Sales}[Order_Quantity])$$

Average_Cost = SUM(Sales[Total_cost])/SUM(Sales[Order_Quantity])

Average_Profit= Sales[Average_Revenue]- Sales[Average_Cost]

Profit_Margin = (SUM(Sales[Revenue])-SUM(Sales[Total_cost]))/SUM(Sales[Revenue])

These steps are shown in the figures below:

FIG 53: THE DATA VIEW WHERE THE NEW MEASURES WERE CREATED WITH DAX

FIG 54: THE “AVERAGE_REVENUE” MEASURE CREATED WITH DAX

Shirley Akua Konnor: B1B89318 - Power BI Desktop

Average_Cost = SUM(Sales[Total_Cost])/SUM(Sales[Order_Quantity])

Table: Sales (113,022 rows) Column: Average_Cost (0 distinct values)

FIG 55: THE “AVERAGE_COST” MEASURE CREATED WITH DAX

Shirley Akua Konnor: B1B89318 - Power BI Desktop

Average_Profit = Sales[Average_Revenue]-Sales[Average_Cost]

Table: Sales (113,022 rows) Column: Average_Profit (0 distinct values)

FIG 56: THE “AVERAGE_PROFIT” MEASURE CREATED WITH DAX

Sales

Profit Margin = $\frac{\text{SUM}(\text{Sales}[\text{Revenue}]) - \text{SUM}(\text{Sales}[\text{Total_Cost}])}{\text{SUM}(\text{Sales}[\text{Revenue}])}$

Fields

- Calender
- Country
- Customer
- Location
- Product
- Sales
 - Average_Cost
 - Average_Profit
 - Average_Revenue
 - Country
 - Index
 - Order_Quantity
 - Profit
 - Revenue
 - Total_Cost
 - Unit_Cost
 - Unit_Price
 - Profit Margin

FIG 57: THE “PROFIT_MARGIN” MEASURE CREATED WITH DAX

4. DASHBOARD

The Power BI dashboard is made of six pages. Each page contains different contents and charts which were used to analyse the data based on specific criteria from the data set. The dashboard was organized by the following:

- HOMEPAGE
- GEOGRAPHICAL ANALYSIS
- CUSTOMER ANALYSIS
- PRODUCT ANALYSIS
- ANNUAL SALES ANALYSIS
- ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS

4.1 HOMEPAGE

This is the first page of the dashboard. It shows a brief description of what the entire dashboard contains and details of the rest of the dashboard. It also shows details of the person who created the dashboard. The homepage contains the following:

- Title of the project
- The course name
- The course code
- The logo of Teesside University
- My name
- Student id
- A navigator button showing what all the pages represent.

The dashboard is shown in the figure below:



FIG 58: THE HOMEPAGE OF THE DASHBOARD

4.2 GEOGRAPHICAL ANALYSIS PAGE

This is the second page of the dashboard. This page contains different charts used to analyse the data based on geography. Each chart on this page analysed the data set in relation to the location that is the country and states where sales occurred. The page consists of the following:

- An icon representing geography
- The page title
- A back button used to navigate to the previous page
- A home button used to navigate to the home page
- A slicer which contains all the countries in the dataset.
- A card showing the number of states
- A tornado chart representing the amount of revenue and profit based on each country
- A map showing the order quantity of the products based on the continents where the countries where the products were purchased
- A simple waterfall chart used to represent the total cost of the products based on the countries they were purchased
- A next button used to navigate to the next page

The dashboard is shown in the figure below:

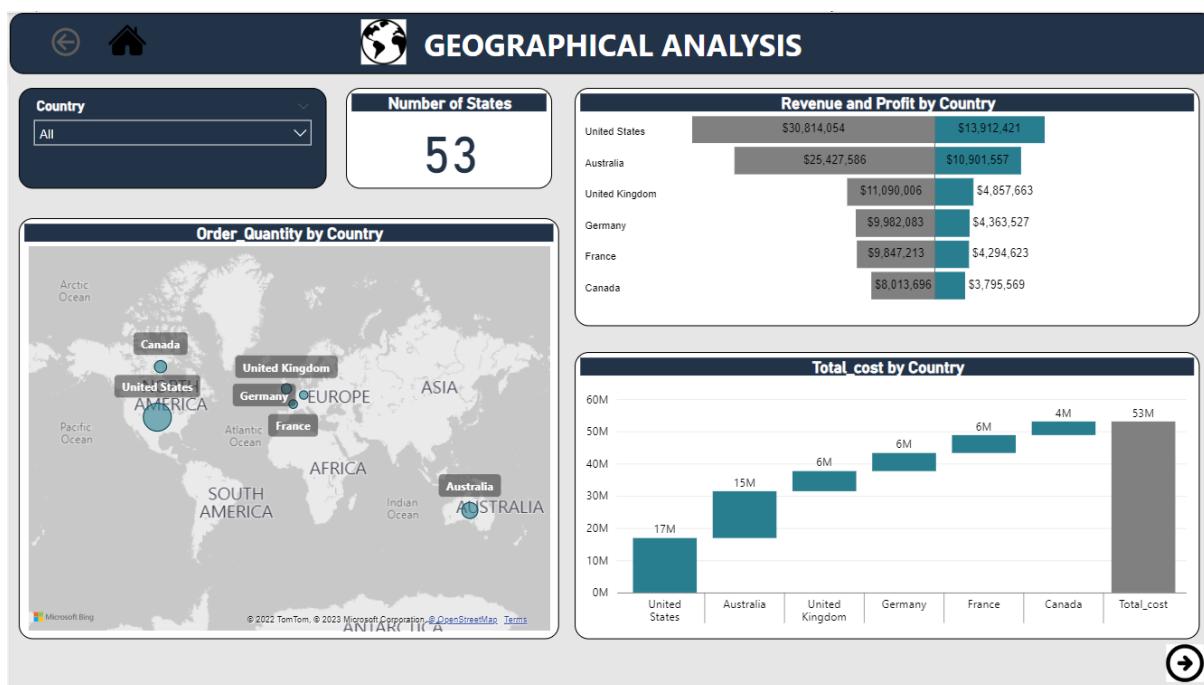


FIG 59: THE GEOGRAPHICAL ANALYSIS DASHBOARD

4.3 CUSTOMER ANALYSIS PAGE

This is the third page of the dashboard. This page contains different charts used to analyse the data based on Customer information. The charts on this page analysed the data in relation to details of the customers in the data set. The page is made of:

- An icon representing customers

- The page title
- A slicer which contains all the age groups of the customers in the dataset.
- A back button used to navigate to the previous page
- A home button used to navigate to the home page
- Four cards representing the number of age groups, the total number of orders, total cost, total revenue and the total profit of the dataset.
- A drill-down radial bar chart representing the amount of profit generated by the age group and gender of customers.
- A tornado chart representing the order quantity based on the gender of the customers.
- A stacked bar chart representing the amount of profit generated based on the age group of the customers.
- A simple waterfall chart representing the amount of revenue based on the age group of the customers.
- A next button used to navigate to the next page

The dashboard is shown in the figure below:

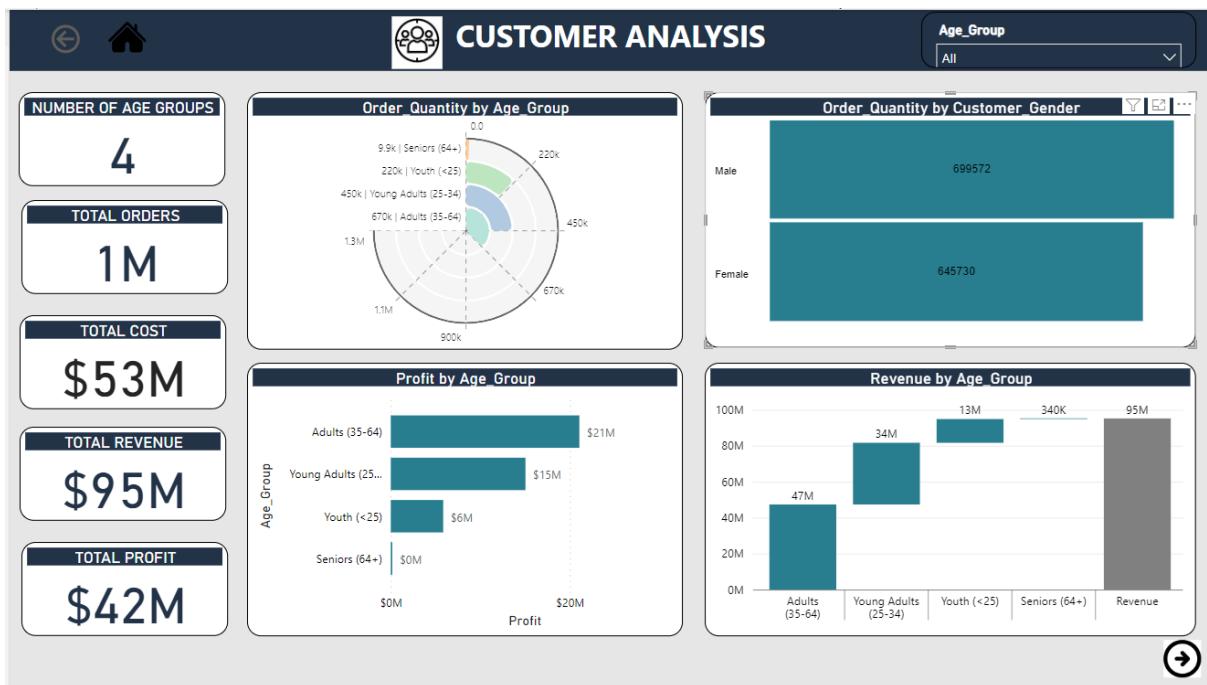


FIG 60: THE CUSTOMER ANALYSIS DASHBOARD

4.4 PRODUCT ANALYSIS PAGE

This is the fourth page of the dashboard. This page contains different charts used to analyse the data based on product information. The various charts on this page represent the data in relation to details of the products in the data set. The page is made of:

- An icon representing products
- The page title
- A back button used to navigate to the previous page
- A home button used to navigate to the home page
- An enlightened slicer displaying the three product categories

- A drill-down radial bar chart used to represent the order quantity for the product category
- Three cards are used to display the number of product categories, the number of products and the total number of orders.
- A simple waterfall chart is used to represent the profit of the product categories
- A clustered column chart that displays the order quantity of the top 10 product sub-categories
- A next button used to navigate to the next page

The dashboard is shown in the figure below:

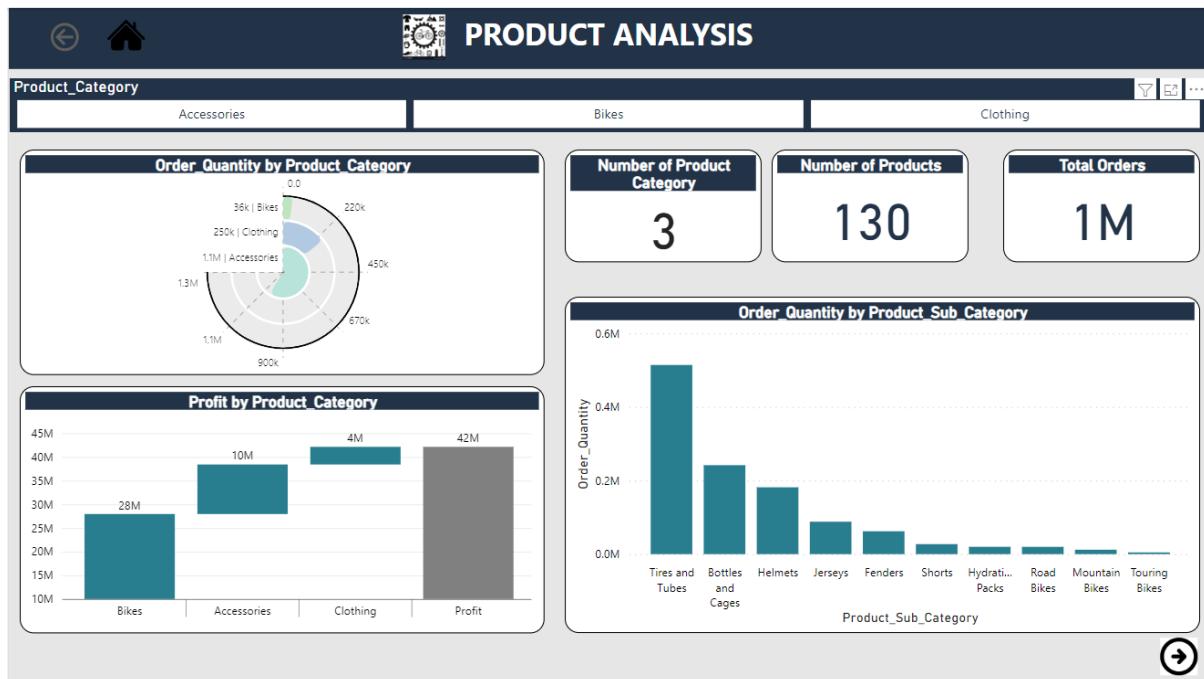


FIG 61: THE PRODUCT ANALYSIS DASHBOARD

4.5 ANNUAL SALES ANALYSIS PAGE

This is the fifth page of the dashboard which has various charts used to represent and display the data to analyse the annual sales. This page is made of:

- An icon representing sales
- The page title
- A back button used to navigate to the previous page
- A home button used to navigate to the home page
- A card used to display the average revenue, average cost, average profit and profit margin
- A line chart representing the revenue and profit generated yearly
- A simple waterfall chart used to represent the order quantity of products by year.
- A scatter chart used to show the correlation between the average profit and average revenue
- An animated bar chart used to represent the profits of the products generated yearly.
- A next button used to navigate to the next page

The dashboard is shown in the figure below:

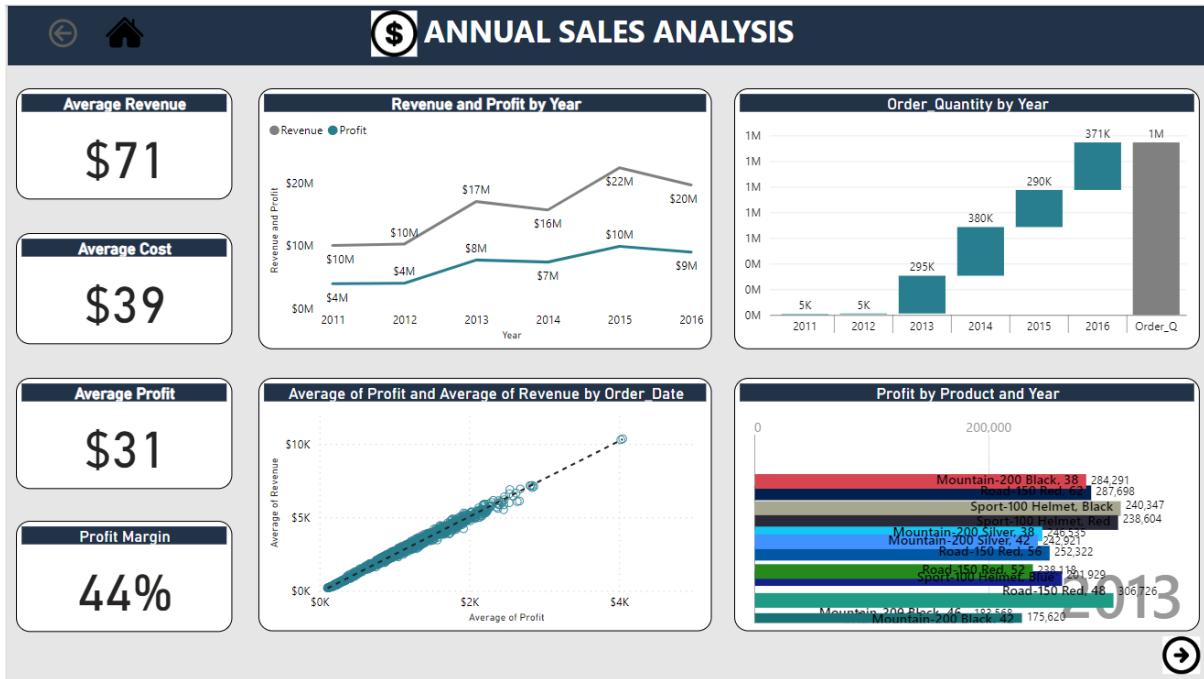


FIG 62: THE ANNUAL SALES ANALYSIS DASHBOARD

4.6 ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS PAGE

This is the last page of the dashboard which contains charts used to represent insights about artificial intelligence. This page is made of:

- An icon representing products
- The page title
- A back button used to navigate to the previous page
- A home button used to navigate to the home page
- A slicer showing the order date containing years, months and quarter
- A decomposition tree representing the revenue generated by country, age group, Product category, product subcategory and Products.
- A line chart used for forecasting revenue by the order date.

The dashboard is shown in the figure below:

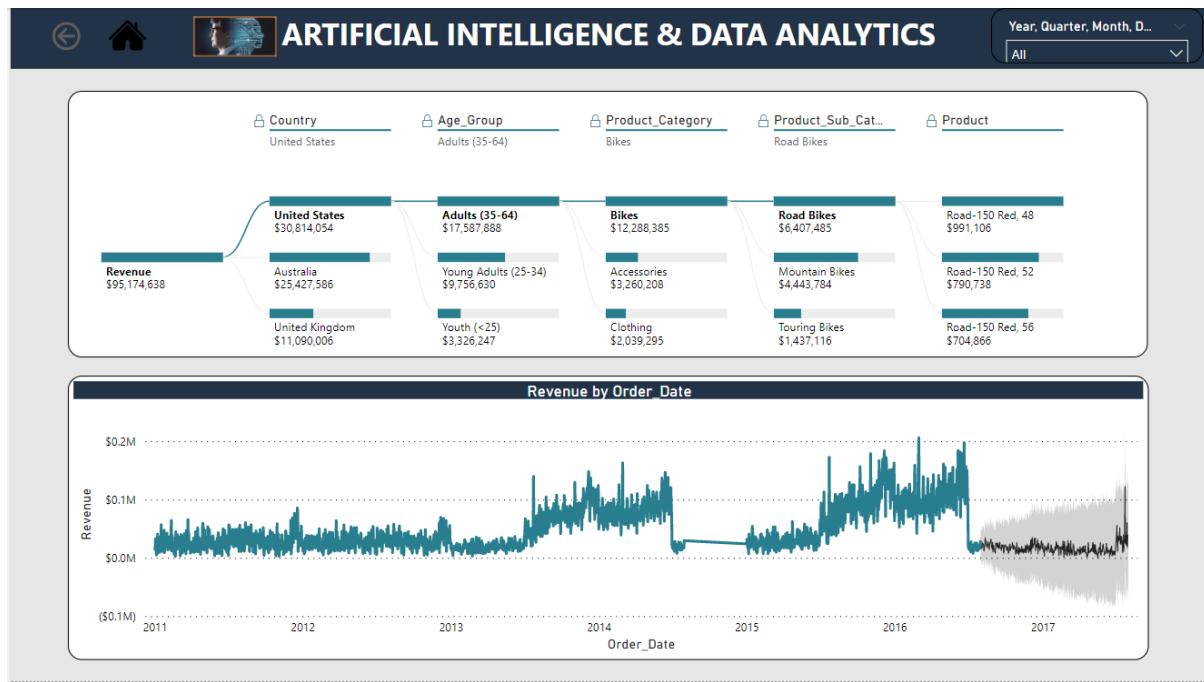


FIG 63: THE ARTIFICIAL INTELLIGENCE AND DATA ANALYTICS DASHBOARD

5. REFERENCES

Bicycle market size, share & covid-19 impact analysis, by type (road, Mountain, hybrid, and others), by design (folding and regular), by end-user (men, women, and kids), by technology (electric and conventional), and regional forecasts, 2022-2029 (2022) *Bicycle Market Size, Share & Value / Trends Analysis [2029]*. Available at: <https://www.fortunebusinessinsights.com/bicycle-market-104524> (Accessed: January 10, 2023).

Puneeth, R.B. and Nethravathi, S.P. (2021) *Bicycle Industry in India and its Challenges – A Case Study. International Journal of Case Studies in Business, IT, and Education* . Available at: <http://doi.org/10.5281/zenodo.5205299>.

A.Use the table below to **self-assess** your work. This will help reflect on your work.
You must keep this table in your report.

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	80
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	70
Dax and M language	Both DAX and M Language have been extensively used in the report	75
Dashboard Design	The dashboard contains a variety of charts, including advanced ones.	75
Average		Add below the average of the four cells above: 75