



# Bike Sales

## BUSINESS INTELLIGENCE AND BIG DATA



JESUBUKADE AJAKAYE

E4216209

Bike Sales

## Table of Contents

List of Table .....	3
List of Figures .....	4
Executive Summary .....	6
Introduction .....	8
Data Source .....	8
Methodology/ Modelling .....	9
BI Requirements/ Questions .....	11
Analysis and Discussion .....	12
Product and Category Insights .....	12
Key Findings .....	17
Geographic/ Location Analysis .....	18
Key Findings .....	21
Demographic Analysis .....	23
Key Findings .....	27
Marketing Analysis: Examine the new marketing strategies based on existing data to be developed to improve sales, revenue and profitting.....	28
Key Findings .....	34
Conclusion and Recommendation .....	34
ICA – Appendix: BI Design .....	36
Appendix A: Data Pre-Processing and Data Cleaning.....	36
Dealing with errors .....	37
Manipulating Columns .....	38
Appendix B: BI Data Modelling via Star Schema – Fact and Dimension Tables.....	41
Relationships .....	42

Keys.....	42
Appendix C: DAX and M Language .....	43
Measures and Calculated Columns using DAX .....	43
Application of M Language.....	46
Appendix D: Dashboards .....	48
Navigator Page.....	48
Home Page .....	49
Product Page .....	50
Location Page .....	51
Demography Page .....	52
Marketing Page.....	53
Question Tab .....	54
Appendix E: Self-Assessment.....	54

## List of Table

Table 1: General Description of Dataset .....	8
Table 2: Data Description Based on Table.....	9
Table 3: Analysis of Chart showing Total revenue by product category .....	12
Table 4: Analysis of Chart showing Order Quantity by Product Category .....	13
Table 5: Analysis of Chart showing Total Revenue by Sub-Category and Year .....	14
Table 6: Analysis of Chart showing Order Quantity and Total Revenue by Year and Product Category.....	15
Table 7: Analysis of Chart showing Total Profit and Total Cost by Product .....	16
Table 8: Analysis of Chart showing Matrix of Location by Date .....	18
Table 9: Analysis of Chart showing Map of Total Revenue by Country.....	19
Table 10: Analysis of Chart showing Total Profit Margin by Year and Country .....	20
Table 11: Analysis of Chart showing Total Revenue by State .....	21
Table 12: Analysis of Chart showing Total Revenue by Age Group .....	23
Table 13: Analysis of Chart showing Total Revenue and Order Quantity by Gender .....	24
Table 14: Analysis of Chart showing Total Profit by Age Group and Gender .....	25
Table 15: Analysis of Chart showing Matrix of Age Group by Gender.....	26
Table 16: Analysis of Chart showing Total Revenue by Month and Year .....	28
Table 17: Analysis of Chart showing Total Revenue by State .....	29
Table 18: Analysis of Chart showing Total Revenue and Total Profit Margin (%) by Quarter .....	30
Table 19: Analysis of Chart showing Total Revenue by Date.....	31
Table 20: Analysis of Chart showing Product by Order Profitability.....	32
Table 21: Analysis of Chart showing Matrix of Product Category by Order Profitability..	33
Table 22: Other Measure Created .....	45
Table 23: Self-Assessment Score .....	54

## List of Figures

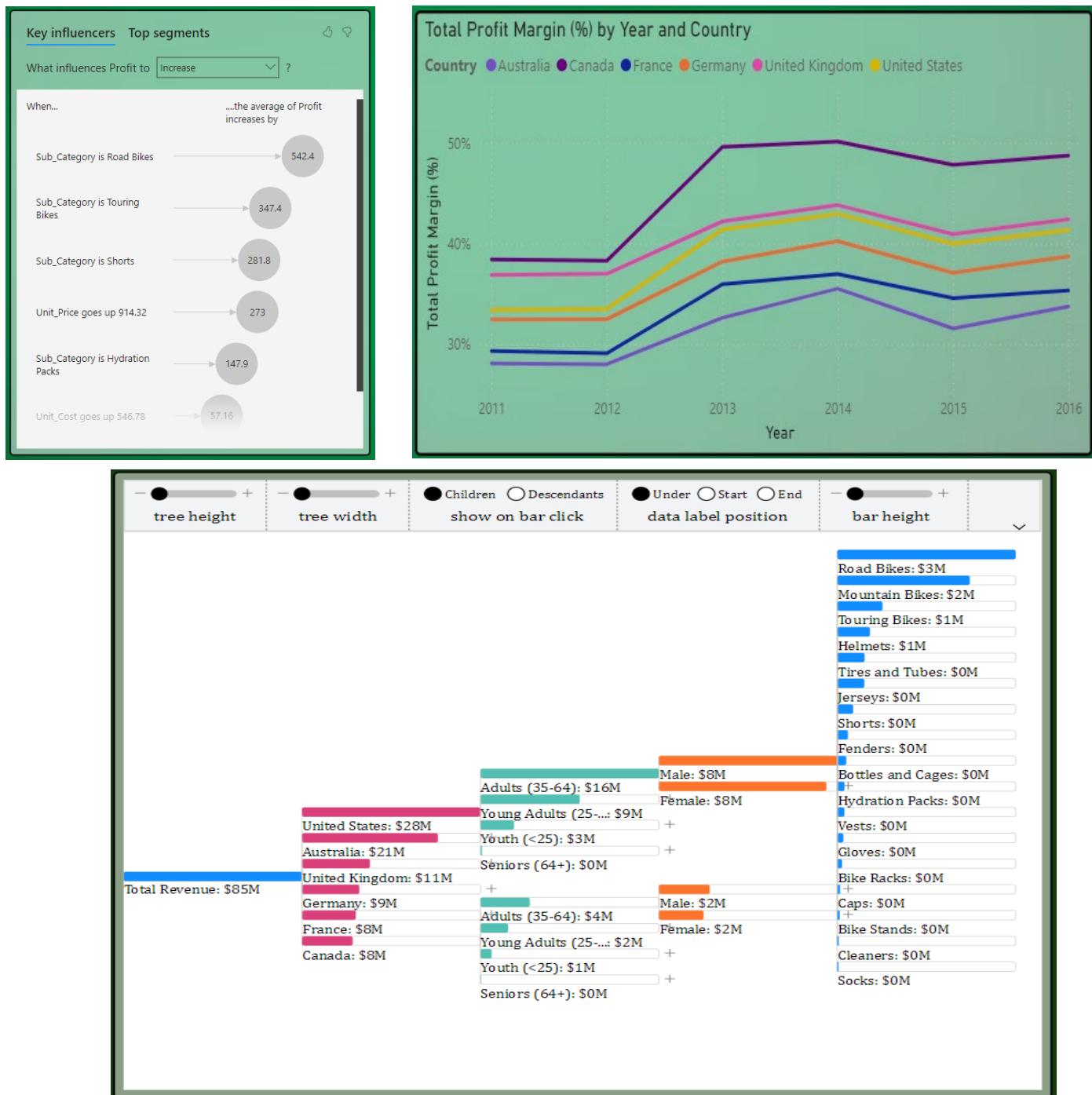
Figure 1: Chart showing total revenue by product category.....	12
Figure 2: Chart showing Order Quantity by Product Category .....	13
Figure 3: Chart showing Total Revenue by Sub-Category and Year .....	14
Figure 4: Chart showing Order Quantity and Total Revenue by Year and Product Category .....	15
Figure 5: Chart showing Total Profit and Total Cost by Product .....	16
Figure 6: Chart showing Matrix of Location by Date .....	18
Figure 7: Chart showing Map of Total Revenue by Country.....	19
Figure 8: Chart showing Total Profit Margin by Year and Country.....	20
Figure 9: Chart showing Total Revenue by State .....	21
Figure 10: Chart showing Total Revenue by Age Group .....	23
Figure 11: Chart showing Total Revenue and Order Quantity by Gender .....	24
Figure 12: Chart showing Total Profit by Age Group and Gender.....	25
Figure 13: Chart showing Matrix of Age Group by Gender.....	26
Figure 14: Chart showing Total Revenue by Month and Year .....	28
Figure 15: Chart showing Total Revenue by State .....	29
Figure 16: Chart showing Total Revenue and Total Profit Margin (%) by Quarter .....	30
Figure 17: Chart showing Total Revenue by Date.....	31
Figure 18: Chart showing Product by Order Profitability.....	32
Figure 19: Chart showing Matrix of Product Category by Order Profitability .....	33
Figure 20: Preview of dataset after introducing faulty rows on excel.....	36
Figure 21: Preview of data when imported to Power BI.....	36
Figure 22: Interface showing error loaded .....	37
Figure 23: Query interface showing errors.....	37
Figure 24: Showing dataset processes after cleaning .....	38
Figure 25: Remove other columns when creating the date table .....	39
Figure 26: Before reordering columns.....	39
Figure 27: After reordering columns .....	40
Figure 28: Before renaming columns .....	40
Figure 29: After renaming columns.....	41

Figure 30: Table created from the flat table.....	41
Figure 31: Model view before relationships .....	42
Figure 32: Model view after relationships using Star Schema.....	43
Figure 33: DAX for profit margin (%) .....	43
Figure 34: DAX for order profitability .....	44
Figure 35: DAX for total revenue .....	44
Figure 36: DAX for total profit margin (%) .....	44
Figure 37: DAX for order count .....	45
Figure 38: Interface before creating measures and calculated columns.....	45
Figure 39: Interface after creating measures and calculated columns .....	46
Figure 40: Error view when trying to transform on new PC.....	46
Figure 41: Source before manipulating the M Language.....	47
Figure 42: Source after manipulating the M Language .....	47
Figure 43: View of the Navigator Dashboard.....	48
Figure 44: View of Home Dashboard .....	49
Figure 45: View of Product Dashboard.....	50
Figure 46: View of Location Dashboard .....	51
Figure 47: View of Demography Dashboard .....	52
Figure 48: View of Marketing Dashboard.....	53
Figure 49: view of Questioning Dashboard.....	54

# Executive Summary

This report is on the appraisal of the GoGREEN Bike Stores from January 2011 to July 2016 using data analytics and business intelligence (BI) to generate insight to the dataset generated by the company over this period. Four (4) business requirements were developed for analysis including product and category insight with five (5) BI questions, geographic insight with four (4) BI questions, demographic insight with four (4) BI questions, and marketing insight with six (6) BI questions.

## Key Findings



- ⌚ The Bike product category returns the most revenue, but the Accessories category returns in the biggest profit margin.
- ⌚ United States and Australia are the countries returning the most revenue as California, England and New South Wales are the states returning the most revenue.
- ⌚ There is a steady increase in profit margin yearly except for the general depression in 2015 and Canada returns the highest profit margin.
- ⌚ Most of the customers are the working population as the average age was 36 years. There is no significant difference between the transactions by male and female customers.
- ⌚ Almost 500 transactions from 9 products in the clothing category were unprofitable in all the years considered.

## **Recommendations**

- ⌚ For sales to keep moving in an upward trajectory, the company needs to sell more quantity as it was deduced that the quantity sold increased yearly, which increased total revenue and total profit.
- ⌚ The company needs to holistically review the performance of states not making up to the average total revenue and improve marketing strategies to improve sales.
- ⌚ Since sales are doing well with the working population, the company needs to develop marketing strategies to improve sales for the age group under 25 years.
- ⌚ The slight differences between the male and female purchasing history should be considered and marketing and sales offers should be designed as appropriate.
- ⌚ A market assessment of products that are not profitable should be carried out and some of these products can be dropped as contribute little to the total revenue of the company.
- ⌚ The company must be more consistent in sales and record keeping, avoiding gaps that occurred in some parts of 2014 and 2015.

## Introduction

The turn of the 21st century has seen more advocacy for an improved and sustainable environment. Countries look towards cutting down on carbon emissions, especially for vehicles which created bigger markets for electric vehicles and bicycles. Some countries like the United Kingdom now implement cycle to work policy to encourage its citizens to engage in sustainable practices. The GoGREEN Bike Stores on this premise in 2011 entered the Bike Sales Market and has seen generated records of sales over the year. Understanding the importance of Data Analytics and Business Intelligence to stay competitive in this market, necessitated the development of this report to show key insights of how the company has performed over the years and other areas to improve upon in order to achieve the company's business goals. The key users of the report include the CEO and management for informed decision making, marketing team for development of strategies and expansion activities, and finally for all employees to understand how far they have gone in achieving set goals.

## Data Source

The dataset from Kaggle titled "[Bike Sales in Europe](#)" was updated by Sadiq Shah in 2020. This data is about sales of bikes, clothing and accessories. The data extends more than Europe (United Kingdom, France, Germany) as it contains countries including United States of America, Canada and Australia. The dataset gives information about sales progression from Jan 2011 to July 2016 and for this analysis and report a dummy company named GoGREEN Bike Stores is used.

**Table 1: General Description of Dataset**

Item	Description
File Name	Sales.csv
File Link	<a href="https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe/data">https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe/data</a>
File Size	14,884kB
Total Row	113,036
Total Columns	18
Numerical Columns	7
Categorical Columns	7
Date Columns	4

## Methodology/ Modelling

After much pre-processing and data modelling, the Flat Table was split into a Fact Table, 5 Dimension Tables and a Bridge Table in a star data modelling style. “*Sales Inventory Table*” is the fact table while “*Age Table*”, “*Date Table*”, “*Gender Table*”, “*Location Table*”, and “*Product Table*” are the dimension tables. The modelling was carried out to avoid any many to many relationship between the Dimension tables and the Fact Table. Table 2 describes each table's columns and includes calculated columns created using DAX language.

**Table 2: Data Description Based on Table**

Table	Column Name	Description	Data Type
Sales Inventory Table (Fact Table)	Sales_ID	The number of each transaction	Numeric
	Date (Foreign Key)	Information about the day sales were made	DateTime
	Customer_Age (Foreign Key)	The exact age of the customer	Numeric
	Customer_Gender (Foreign Key)	The symbol for gender as applicable	Categorical
	State (Foreign Key)	The geographical location of the customer in a country	Nominal
	Product (Foreign Key)	The name of the product sold	Nominal
	Order_Quantity	The total number of products sold in each transaction	Numeric
	Profit	The amount made as profit on each transaction (revenue – cost)	Numeric
	Cost	The total cost in each transaction (unit_cost * order_quantity)	Numeric
	Revenue	The total amount realised in each transaction (unit_price * order_quantity)	Numeric
	Profit_Margin (%) (Calculated Column)	Derived by dividing profit by revenue in each transaction. (profit/revenue)	Numeric
	Order profitability	An if statement checked on	Nominal

	(Calculated Column)	the Profit_Margin (%) to see if a transaction is profitable or not	
Age Table (1:*	Customer_Age	Used as Table Key to connect the Age Table to the Fact Table	Numerical
	Age_Group	Places the Age in various corresponding groups	Categorical
Date Table (1:*	Date	Used as Table Key to connect the Date Table to the Fact Table	DateTime
	Day	Day number between 1 and 31 depending on the month	Numerical
	Day of the Week	Name of the day in the week	Nominal
	Month	Name of the month in a year	Nominal
	Month Number	Corresponding number with the month name for arrangement	Numerical
	Quarter	Number of the quarter in the year to group the months	Numerical
	Year	Year number	Numerical
	Year_Month (Calculated Column)	A combination of the month and the year.	Mixed
Gender Table (1:*	Customer_Gender	Used as Table Key to connect the Gender Table to the Fact Table	Categorical
	Gender	The name of the gender either Male or Female	Categorical
Location Table (1:*	Country	Name of the country where the transaction took place.	Nominal
	State	Used as Table Key to connect the Location Table to the Fact Table	Nominal
Product Table (1:*	Product	Used as Table Key to connect the Product Table to the Fact Table	Nominal
	Product_Category	Category of the product which could be Bikes, Accessories or Clothing	Nominal
	Sub_Category	Name of different subcategories under each product category	Nominal
	Unit_Cost	The cost of each product	Numerical
	Unit_Price	The price the product was sold for.	Numerical

## **BI Requirements/ Questions**

1. Product and Category Insights
  - a. What product category brings in the most revenue?
  - b. What product category has the most quantity sold?
  - c. What is the yearly total revenue by product subcategory?
  - d. What is the spread of order quantity and revenue by product category in the year across months?
  - e. How do each product's total cost and revenue compare?
2. Geographic/ Location Analysis
  - a. What is the total revenue of each location (Country/State) by date (Year/Month) and where do we have performance below average?
  - b. What is the total revenue for each country during the period?
  - c. What is the total profit margin yearly for each country?
  - d. What are the top states bringing in revenue?
3. Demographic Analysis
  - a. What is the revenue spread by age group?
  - b. What is the total revenue and order quantity between genders?
  - c. What is the spread of total profit by gender and age group?
  - d. Where does the most revenue come in within the gender and age group matrix?
4. Marketing Analysis: Examine the new marketing strategies based on existing data to be developed to improve sales, revenue and profiting
  - a. What are the sales spread monthly in each year?
  - b. What are the lowest revenue states?
  - c. What is the total revenue and total profit in each quarter?
  - d. What is the revenue forecast over the next 12 months?
  - e. How many products are not profitable?
  - f. How many orders are not profitable and in which sub-category and product category are they in?

## Analysis and Discussion

### Product and Category Insights

- i. What product category brings in the most revenue?

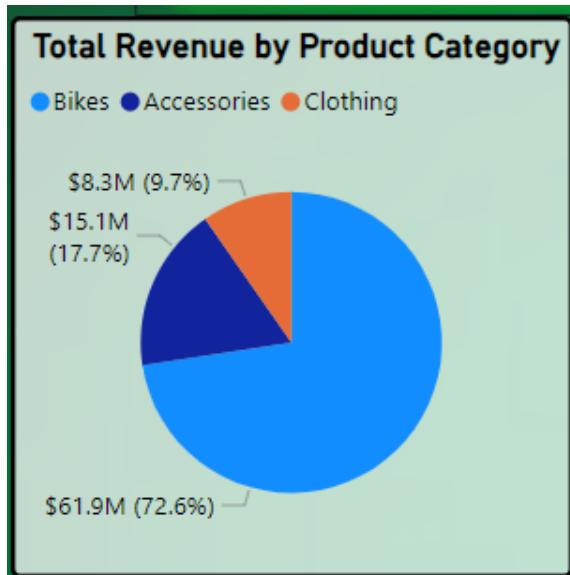
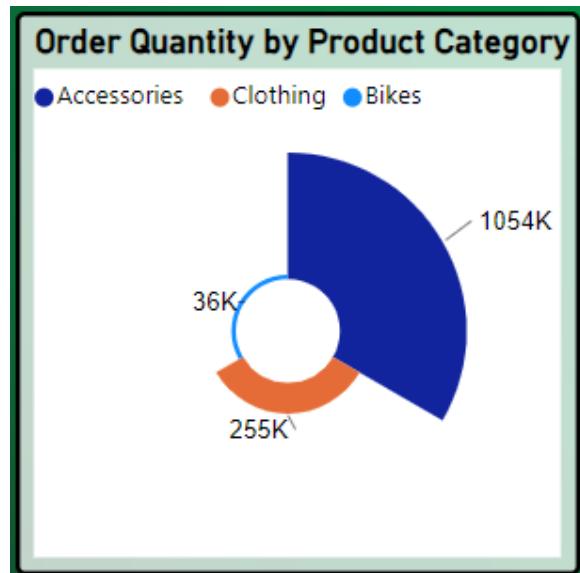


Figure 1: Chart showing total revenue by product category

Table 3: Analysis of Chart showing Total revenue by product category

Title	Total Revenue by Product Category
Chart Type	Pie Chart
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To show the distribution of product category (categorical) according to total revenue (numerical) at a glance in pie size.
Analysis	The visual revealed that the bike category brought in the most revenue as it is more expensive than any other product category.

ii. What product category has the most quantity sold?



**Figure 2: Chart showing Order Quantity by Product Category**

**Table 4: Analysis of Chart showing Order Quantity by Product Category**

<b>Title</b>	Order Quantity by Product Category
<b>Chart Type</b>	Aster Plot 1.4.0
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show the distribution of product category (categorical) according to the sum of order quantity (numerical). The aster plot is like a doughnut chart but shows variation in the diameter.
<b>Analysis</b>	Findings from this visual revealed that the accessories category sold more in quantity, this could be due to the fact that one bike might need more than one item to accessorize it. Also, despite the bike bringing in the most revenue, the quantity percentage is very little when compared supporting the findings from the previous analysis.

iii. What is the yearly total revenue by product subcategory?

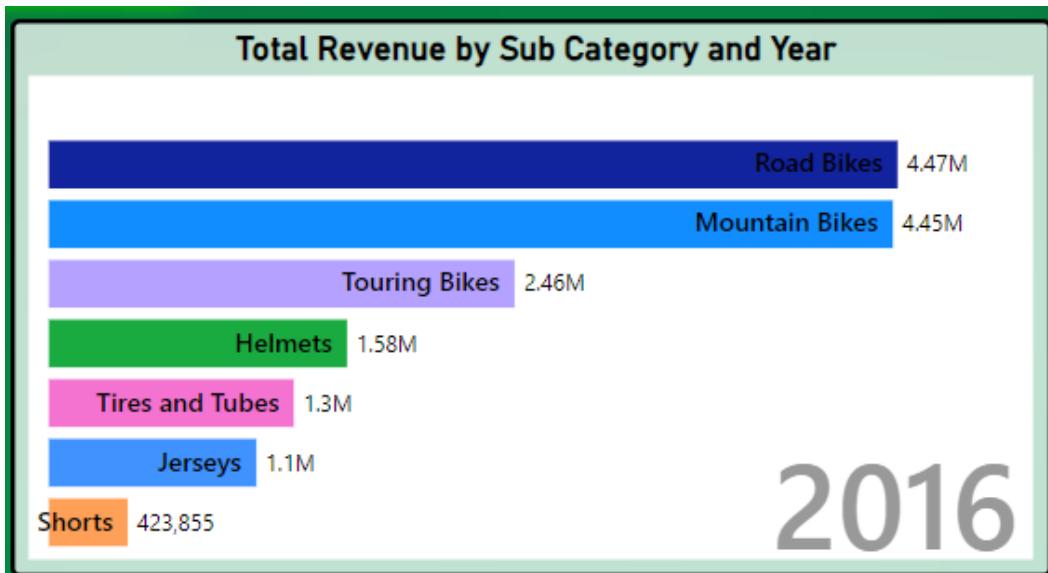


Figure 3: Chart showing Total Revenue by Sub-Category and Year

Table 5: Analysis of Chart showing Total Revenue by Sub-Category and Year

Title	Total Revenue by Sub-Category and Year
Chart Type	Animated Bar Chart Race
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To show the total revenue (numerical) of subcategories (categorical) after each year and how the rank.
Analysis	The visual revealed that while the bikes subcategories brought in the most revenue, at around 2014, Helmets ranked 3 <sup>rd</sup> in total revenue and Touring bikes gained popularity again in 2015. This underscores the importance of Helmets for safety and an important item to recommend with bikes during sales.

- iv. What is the spread of order quantity and revenue by product category in the year across months?

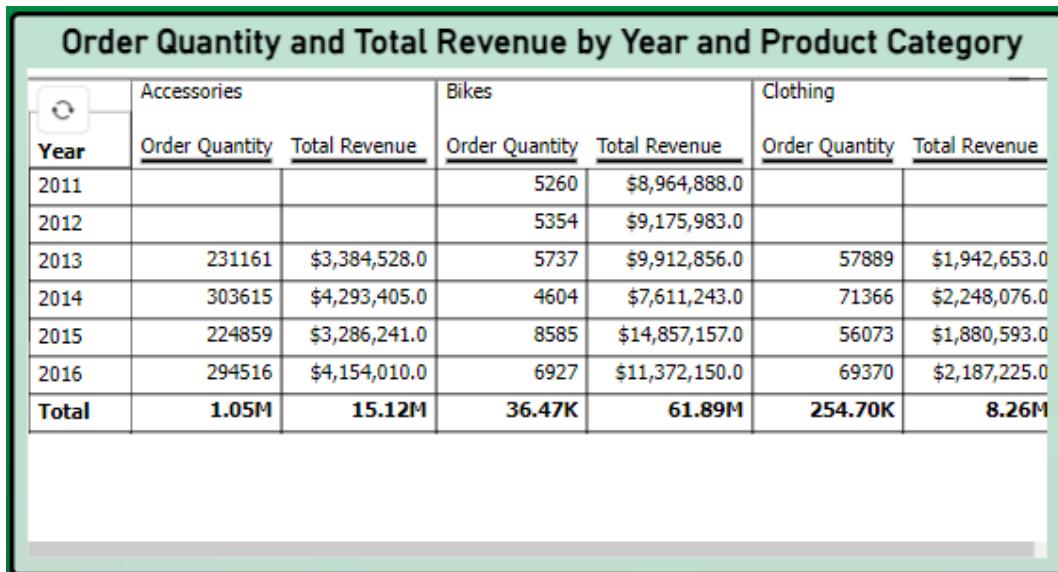


Figure 4: Chart showing Order Quantity and Total Revenue by Year and Product Category

Table 6: Analysis of Chart showing Order Quantity and Total Revenue by Year and Product Category

<b>Title</b>	Order Quantity and Total Revenue by Year and Product Category
<b>Chart Type</b>	Business Performance Planning – Reporting v1.6.02728
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show how the product categories (categorical) are performing over the years and months based on order quantity (numerical) and total revenue (numerical).
<b>Analysis</b>	The visual revealed that the company started with sales of bikes in 2011 and 2012 which grew over the years. The figures for order quantity in 2014 in the bike category were low because it's only for half a year. 2016 also showed the order quantity for half a year. The sales in the accessories and clothing categories started in 2013 and also grew over the years both in total revenue and order quantity. This showed an improvement in sales and revenue generation over the years

v. How do each product's total cost and revenue compare?



**Figure 5: Chart showing Total Profit and Total Cost by Product**

**Table 7: Analysis of Chart showing Total Profit and Total Cost by Product**

<b>Title</b>	Total Profit and Total Cost by Product
<b>Chart Type</b>	Clustered Bar Chart
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To compare Total Cost (numerical) and Total Profit (numerical) for each Product to see which product is more profitable for the company. Helps the company to decide which product to finance more to get more profits.
<b>Analysis</b>	The visual is filtered to Top 15 with the Total Profit. Findings revealed that Mountain-200 Black, 38 (Bike) is the most

	profitable Product and its Total cost is less than Road-150 Red, 62 when compared. It is worth of note that products in the accessories category such as Sport-100 Helmet, Red has more Total profit than Total Cost.
--	---

## Key Findings

Here is the summary of Key Findings from the Product and Category Insight (see Appendix for full Dashboard)

- ✓ The bikes category brought in the most revenue while the accessories category sold more quantity.
- ✓ The company only sold bikes in 2011 and 2012 but expanded to accessories and clothing in 2013. No sales were recorded from July to December 2014 in the bike category while no sales were recorded from August 2014 to June 2015 in the Accessories and Clothing categories.
- ✓ The Helmets sub-category appeared to be important items for sales and turnover.
- ✓ There is a slight change when the visuals are filtered to show dynamism in the analysis for example when filtered to Canada, the bike categories bring in about 50% of the total revenue as against higher figures from other countries.
- ✓ The findings showed an improvement in sales and revenue generation over the years studied.

## Geographic/ Location Analysis

- i. What is the total revenue of each location (Country/State) by date (Year/Month) and where do we have performance below average?

Country	2011	2012	2013	2014	2015	2016	Total
■ Australia	\$2,529,914.0	\$2,557,511.0	\$4,107,470.0	\$2,802,842.0	\$5,712,339.0	\$3,591,983.0	\$21,302,059.0
New South Wales	\$1,175,012.0	\$1,171,056.0	\$1,697,649.0	\$1,229,432.0	\$2,354,993.0	\$1,575,353.0	\$9,203,495.0
Queensland	\$622,846.0	\$642,944.0	\$975,721.0	\$618,695.0	\$1,392,224.0	\$813,837.0	\$5,066,267.0
South Australia	\$138,199.0	\$141,107.0	\$307,194.0	\$179,033.0	\$413,574.0	\$218,167.0	\$1,397,274.0
Tasmania	\$72,073.0	\$61,490.0	\$134,687.0	\$58,222.0	\$187,793.0	\$65,919.0	\$580,184.0
Victoria	\$521,784.0	\$540,914.0	\$992,219.0	\$717,460.0	\$1,363,755.0	\$918,707.0	\$5,054,839.0
■ Canada	\$789,798.0	\$835,527.0	\$1,443,537.0	\$1,471,522.0	\$1,731,954.0	\$1,663,400.0	\$7,935,738.0
Alberta	\$16,553.0	\$11,638.0	\$7,348.0	\$6,161.0	\$7,468.0	\$7,635.0	\$56,803.0
British Columbia	\$773,245.0	\$823,889.0	\$1,435,633.0	\$1,465,361.0	\$1,723,997.0	\$1,655,765.0	\$7,877,890.0
Ontario			\$556.0		\$489.0		\$1,045.0
■ France	\$946,624.0	\$962,153.0	\$1,431,870.0	\$1,411,100.0	\$1,877,928.0	\$1,803,197.0	\$8,432,872.0
■ Germany	\$833,603.0	\$835,188.0	\$1,452,704.0	\$1,702,508.0	\$1,931,091.0	\$2,223,502.0	\$8,978,596.0
Bayern	\$117,195.0	\$120,829.0	\$193,034.0	\$267,226.0	\$278,047.0	\$365,654.0	\$1,341,985.0
Brandenburg	\$7,268.0	\$11,249.0	\$34,355.0	\$33,486.0	\$48,520.0	\$34,892.0	\$169,770.0
Hamburg	\$124,618.0	\$137,850.0	\$277,548.0	\$281,576.0	\$368,779.0	\$368,634.0	\$1,559,005.0
Hessen	\$230,224.0	\$230,652.0	\$316,229.0	\$386,423.0	\$409,265.0	\$500,598.0	\$2,073,391.0
Nordrhein-Westfalen	\$155,142.0	\$141,564.0	\$288,261.0	\$365,008.0	\$395,595.0	\$459,330.0	\$1,804,900.0
Saarland	\$199,156.0	\$193,044.0	\$343,277.0	\$368,789.0	\$430,885.0	\$494,394.0	\$2,029,545.0
■ United Kingdom	\$823,481.0	\$834,885.0	\$1,868,796.0	\$2,014,044.0	\$2,484,181.0	\$2,620,809.0	\$10,646,196.0
England	\$823,481.0	\$834,885.0	\$1,868,796.0	\$2,014,044.0	\$2,484,181.0	\$2,620,809.0	\$10,646,196.0
■ United States	\$3,041,468.0	\$3,150,719.0	\$4,935,660.0	\$4,750,708.0	\$6,286,498.0	\$5,810,494.0	\$27,975,547.0
Total	\$8,964,888.0	\$9,175,983.0	\$15,240,037.0	\$14,152,724.0	\$20,023,991.0	\$17,713,385.0	\$85,271,008.0

Figure 6: Chart showing Matrix of Location by Date

Table 8: Analysis of Chart showing Matrix of Location by Date

Title	Location by Date
Chart Type	Matrix
Data Type(s) in the Visual	Numerical
Reason for choosing Visual	To show the Total Revenue in each Location. The Country column is drilled down to States while the Date row can be drilled down to Months. The color was added to show locations that did not make up to average total revenue per state.
Analysis	The visual revealed that majority of the State and even most countries (except for Australia and United States) were not making up to the average total revenue. Which improved in 2015 and 2016 for all countries.

ii. What is the total revenue for each country during the period?

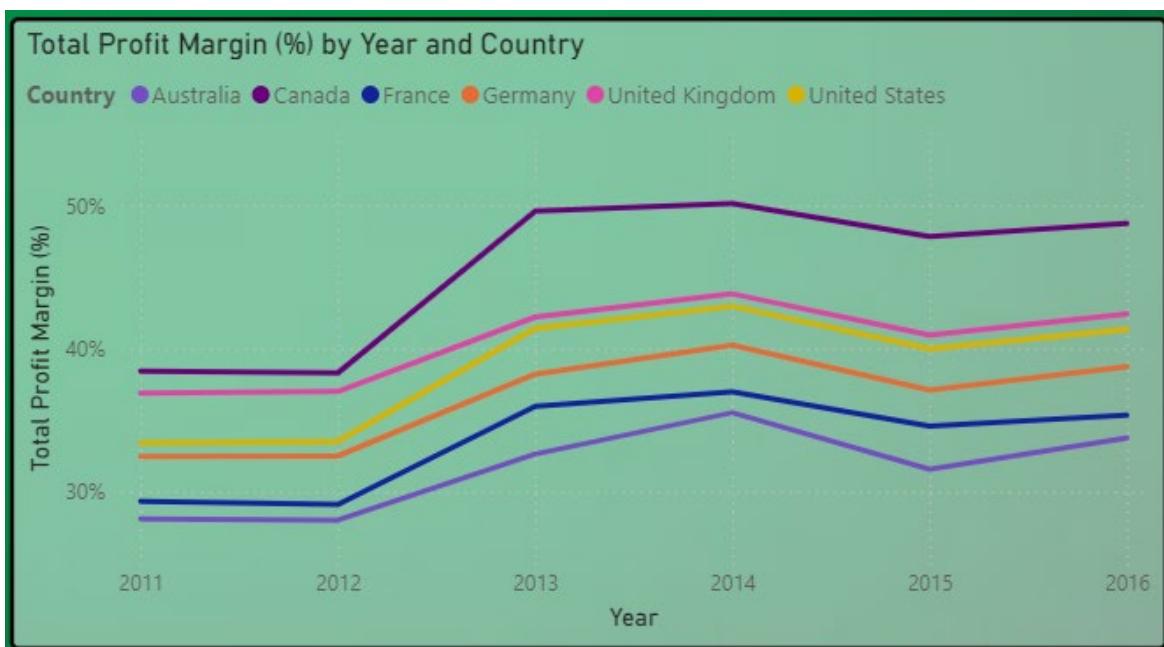


**Figure 7: Chart showing Map of Total Revenue by Country**

**Table 9: Analysis of Chart showing Map of Total Revenue by Country**

<b>Title</b>	Total Revenue by Country
<b>Chart Type</b>	Map
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	The bubble size was adopted to visualise the Total Revenue by each Country on the Map.
<b>Analysis</b>	The visual revealed that the United States has the highest Total Revenue followed by Australia.

iii. What is the total profit margin yearly for each country?



**Figure 8: Chart showing Total Profit Margin by Year and Country**

**Table 10: Analysis of Chart showing Total Profit Margin by Year and Country**

<b>Title</b>	Total Profit Margin (%) by Year and Country
<b>Chart Type</b>	Line Chart
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show trends in the Total profit margin over the years for each country
<b>Analysis</b>	The visual revealed that there was a steady increase in the profit margin over the years with a slight drop in 2015. The drop was a global phenomenon during this period which shows international economic interactions.

iv. What are the top states bringing in revenue?

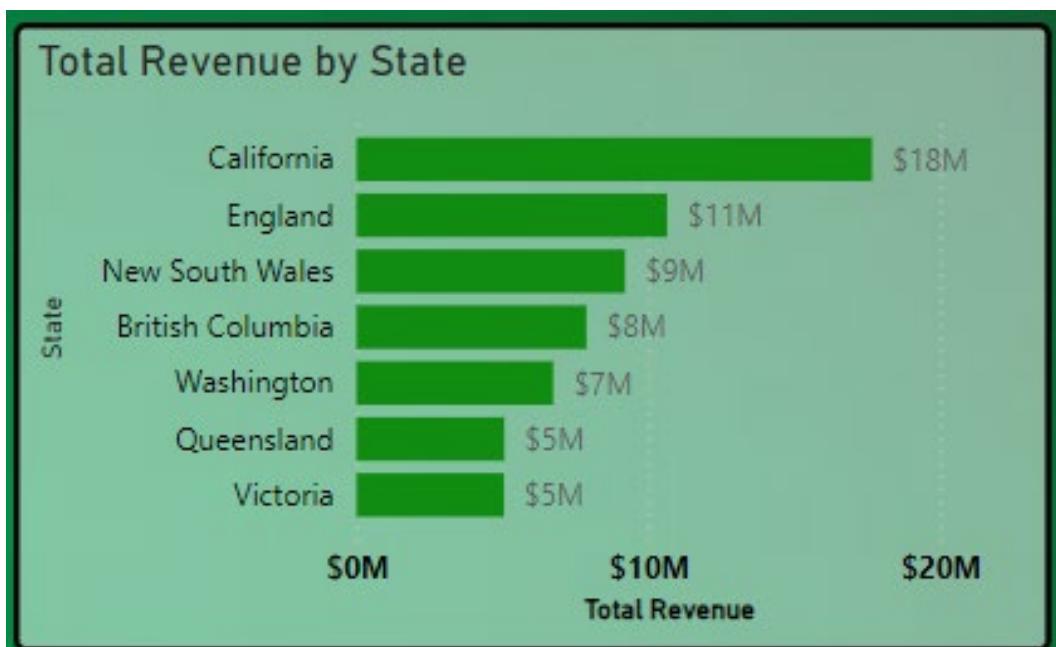


Figure 9: Chart showing Total Revenue by State

Table 11: Analysis of Chart showing Total Revenue by State

Title	Total Revenue by State
Chart Type	Clustered Bar Chart
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	This chart is filtered by the Top States with the highest total profit. Such that we can understand where the highest total revenue is coming in based on profit.
Analysis	The findings revealed that California (One of the largest economies in the World) in the United States generates the most revenue which is more than 20% of the Total revenue. England and New South Wales also generated more than 10% of the Total revenue each.

## Key Findings

Here is the summary of Key Findings from the Geographic/ Location Analysis (see Appendix for full Dashboard)

- ✓ The Total Revenue of the majority of the States is not up to the Average total revenue by states which showed we had very few states which are very

economically viable for the business such as California in United States, England (which is a country) in the United Kingdom and New South Wales in Australia. (Seems Marketing was concentrated in certain regions and should be expanded. This is explained by 1 state in Canada having over 90% of the Total revenue generated)

- ✓ United States and Australia are the biggest markets for the company generating about 50% of the Total Revenue for the Company.
- ✓ There is steady growth in profit margin yearly in all the countries except for 2015 depression recorded in the visual
- ✓ States like California, England and New South Wales has the highest Total Revenue ranked by the Total Profit made.

## Demographic Analysis

- i. What is the revenue spread by age group?

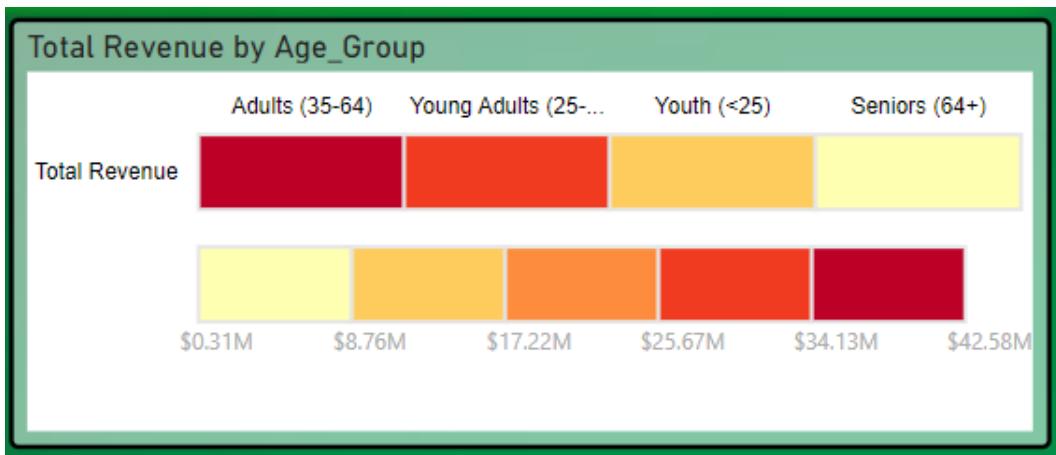


Figure 10: Chart showing Total Revenue by Age Group

Table 12: Analysis of Chart showing Total Revenue by Age Group

<b>Title</b>	Total Revenue by Age Group
<b>Chart Type</b>	Table Heatmap 3.5.0
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show the Total Revenue according to age group in a heat map.
<b>Analysis</b>	The visual revealed Adults of ages between 35-64 brings in the highest total revenue followed by Young Adults of ages between 25-34. The campaign of Cycle to work has started gaining ground at this period and active workers are between ages 25 – 64, which could have contributed to the big revenue chunk. Also, the majority of those older than 64 now prefer riding on a scooter to a bicycle while for teenagers, their parents might have completed the record of purchase of bicycles.

ii. What is the total revenue and order quantity between genders?

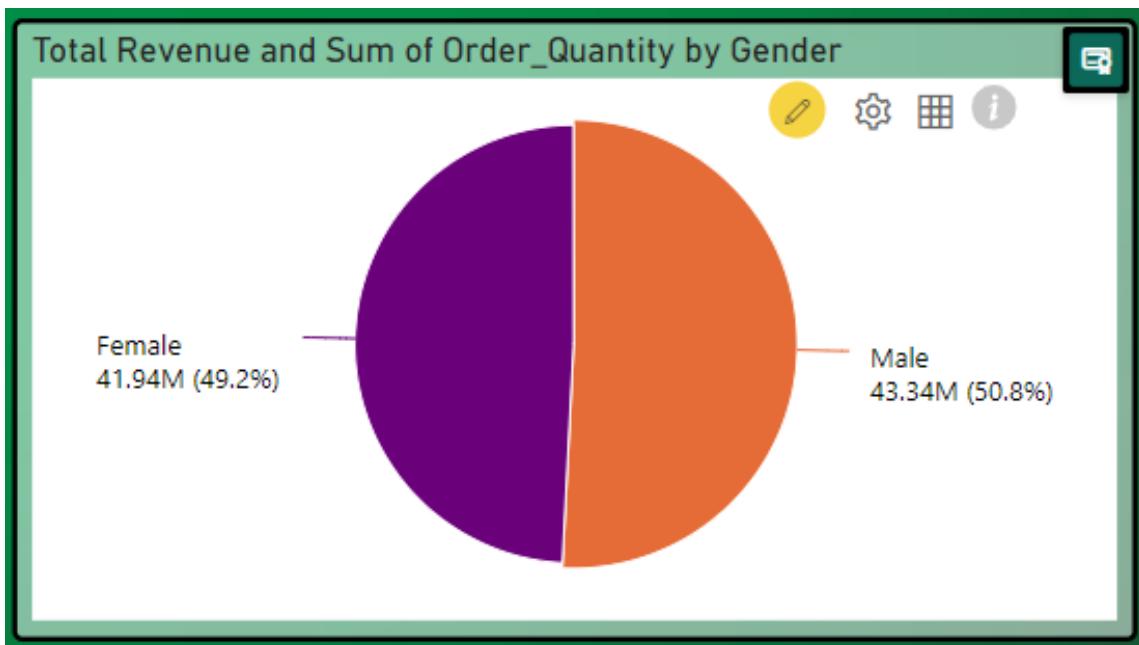


Figure 11: Chart showing Total Revenue and Order Quantity by Gender

Table 13: Analysis of Chart showing Total Revenue and Order Quantity by Gender

<b>Title</b>	Total Revenue and Order Quantity by Gender
<b>Chart Type</b>	Advanced Pie and Donut – xViz v3.0.1
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	This chart helps to show variations between the total revenue and order quantity. The pie details the total revenue while the diameter details the order quantity
<b>Analysis</b>	Findings revealed that Male customers purchase more quantities and therefore generates more revenue through their purchase than their female counterparts.

iii. What is the spread of total profit by gender and age group?

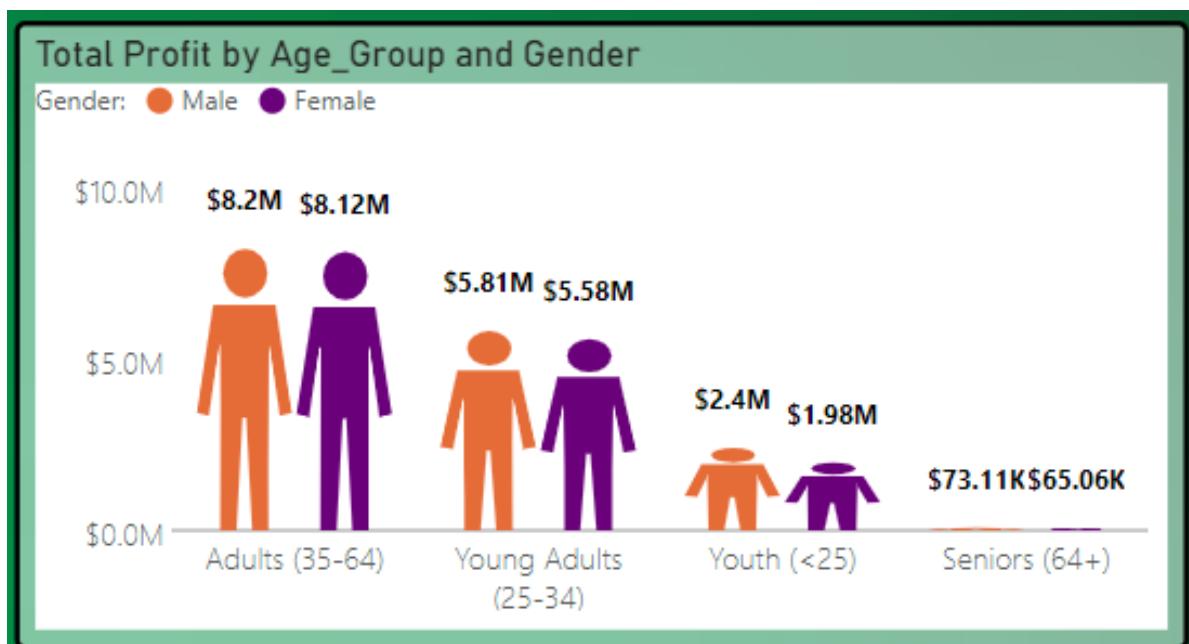


Figure 12: Chart showing Total Profit by Age Group and Gender

Table 14: Analysis of Chart showing Total Profit by Age Group and Gender

<b>Title</b>	Total Profit by Age Group and Gender
<b>Chart Type</b>	Infographic Designer 1.9.7
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show the total profit between genders within different age groups based on Product Categories.
<b>Analysis</b>	The visual revealed that the company slightly profit more from male population than female in all age groups.

- iv. Where does the most revenue come in within the gender and age group matrix?

Gender	Female		Male		Total	
Age_Group	Order Qty	Total Revenue	Order Qty	Total Revenue	Order Qty	Total Revenue
Adults (35-64)	328479	\$21,269,149.0	343938	\$21,315,004.0	672417	\$42,584,153.0
Accessories	257982	\$3,779,749.0	267745	\$4,066,982.0	525727	\$7,846,731.0
Bikes	8978	\$15,481,547.0	8652	\$14,966,763.0	17630	\$30,448,310.0
Clothing	61519	\$2,007,853.0	67541	\$2,281,259.0	129060	\$4,289,112.0
Seniors (64+)	4506	\$130,050.0	5371	\$177,992.0	9877	\$308,042.0
Accessories	3858	\$64,693.0	4469	\$60,061.0	8327	\$124,754.0
Bikes	27	\$44,895.0	53	\$90,164.0	80	\$135,059.0
Clothing	621	\$20,462.0	849	\$27,767.0	1470	\$48,229.0
Young Adults (25-34)	214441	\$15,194,242.0	232434	\$15,461,372.0	446875	\$30,655,614.0
Accessories	166059	\$2,309,015.0	180366	\$2,690,577.0	346425	\$4,999,592.0
Bikes	6844	\$11,585,517.0	6901	\$11,330,039.0	13745	\$22,915,556.0
Clothing	41538	\$1,299,710.0	45167	\$1,440,756.0	86705	\$2,740,466.0
Youth (<25)	98307	\$5,342,158.0	117840	\$6,381,041.0	216147	\$11,723,199.0
Accessories	79225	\$939,825.0	94447	\$1,207,282.0	173672	\$2,147,107.0
Bikes	2200	\$3,870,559.0	2812	\$4,524,793.0	5012	\$8,395,352.0
Clothing	16882	\$531,774.0	20581	\$648,966.0	37463	\$1,180,740.0
<b>Total</b>	<b>645733</b>	<b>\$41,935,599.0</b>	<b>699583</b>	<b>\$43,335,409.0</b>	<b>1345316</b>	<b>\$85,271,008.0</b>

Figure 13: Chart showing Matrix of Age Group by Gender

Table 15: Analysis of Chart showing Matrix of Age Group by Gender

Title	
Chart Type	Matrix
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To show the revenue and Order quantity between genders within different age groups based on Product Categories.
Analysis	Findings from this visual revealed that Females contributed more Total revenue than their Male counterpart in the Adult and Young Adult in the Bikes category while the male contributed more Total Revenue at the Clothing and Accessories Categories. This could be due to a number of reasons, Adult female may order more quantity or female Young Adults at this stage prefers more expensive and fancy bikes.

## **Key Findings**

Here is the summary of Key Findings from the Demographic Analysis (see Appendix for full Dashboard)

- ✓ Total revenue emanated from the working population (as the average customer age is 36 years) which may be traced to cycle-to-work campaigns for improved sustainable environments.
- ✓ Male customers purchase more quantities and therefore generates more revenue through their purchase than their female counterparts.
- ✓ The Females between 25 – 64 generated more total revenue than their male counterpart in the Bike Categories which creates a catchment opportunity for marketing and sales.
- ✓ Sales to male buyers is slightly profitting than to their female counterparts across all age groups

## Marketing Analysis: Examine the new marketing strategies based on existing data to be developed to improve sales, revenue and profiting

- i. What are the sales spread monthly in each year?

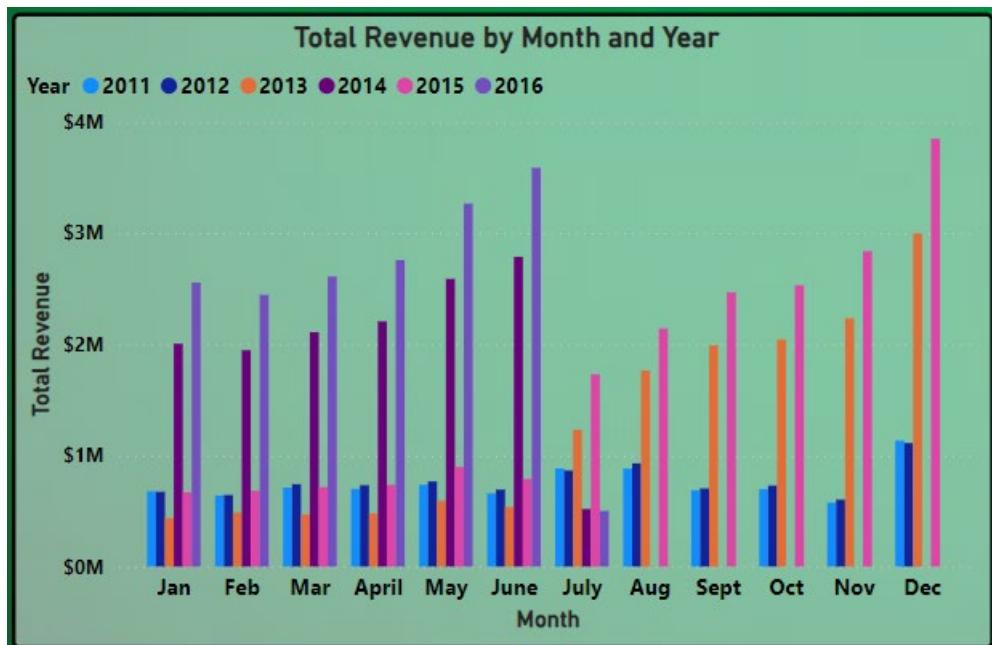


Figure 14: Chart showing Total Revenue by Month and Year

Table 16: Analysis of Chart showing Total Revenue by Month and Year

Title	Total Revenue by Month and Year
Chart Type	Clustered Column Chart
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To show the distribution of revenue yearly by month.
Analysis	The visual revealed that total revenue increases towards the ending of each year. This could be due to more favourable weather for biking between the end of 2 <sup>nd</sup> quarter and the beginning of 4 <sup>th</sup> quarter of the year.

ii. What are the lowest revenue states?

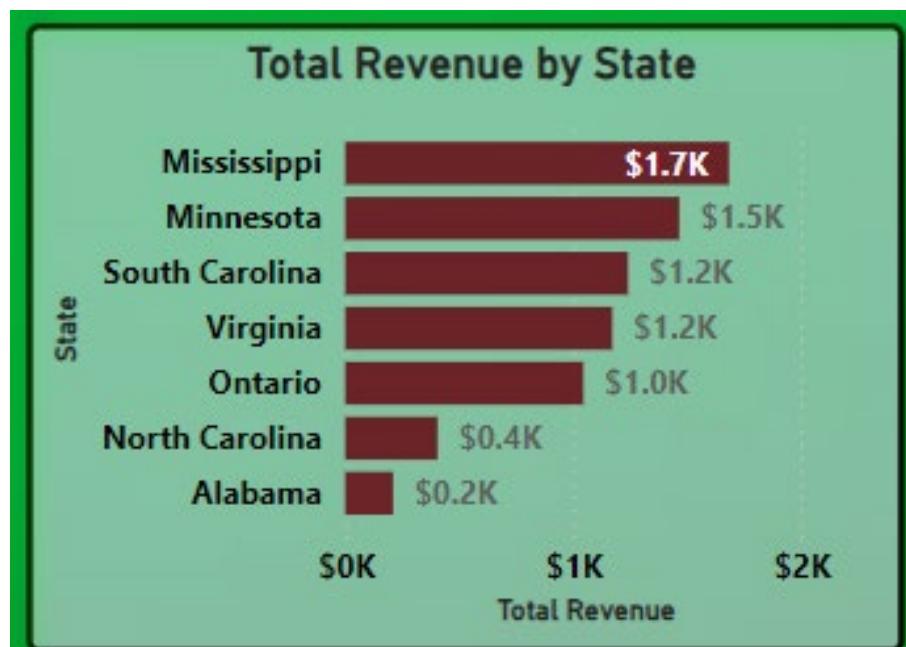
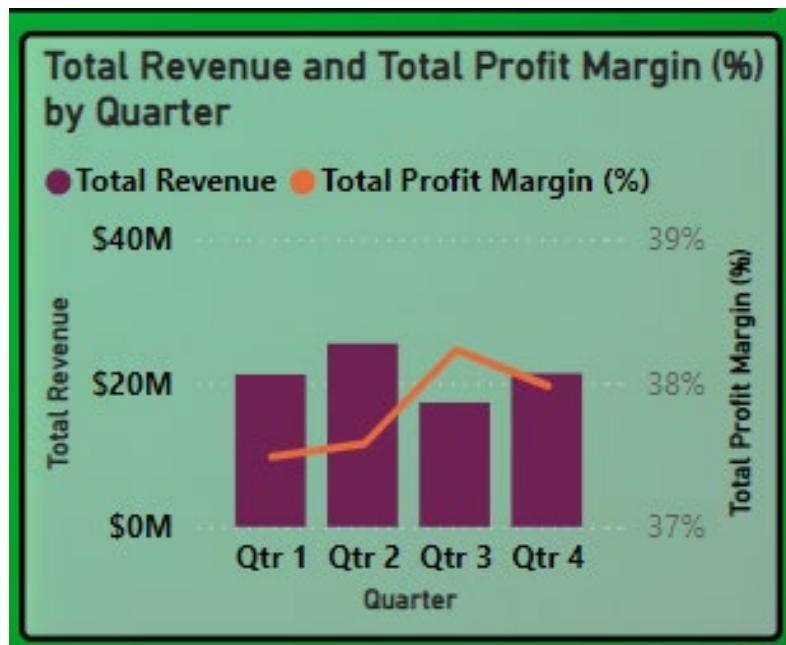


Figure 15: Chart showing Total Revenue by State

Table 17: Analysis of Chart showing Total Revenue by State

Title	Total Revenue by State
Chart Type	Clustered Bar Chart
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To show the ranking of lowest performing states.
Analysis	Findings revealed that despite United States contributing most to the total revenue, the lowest ranking states are majorly from the Country.

- iii. What is the total revenue and total profit in each quarter?

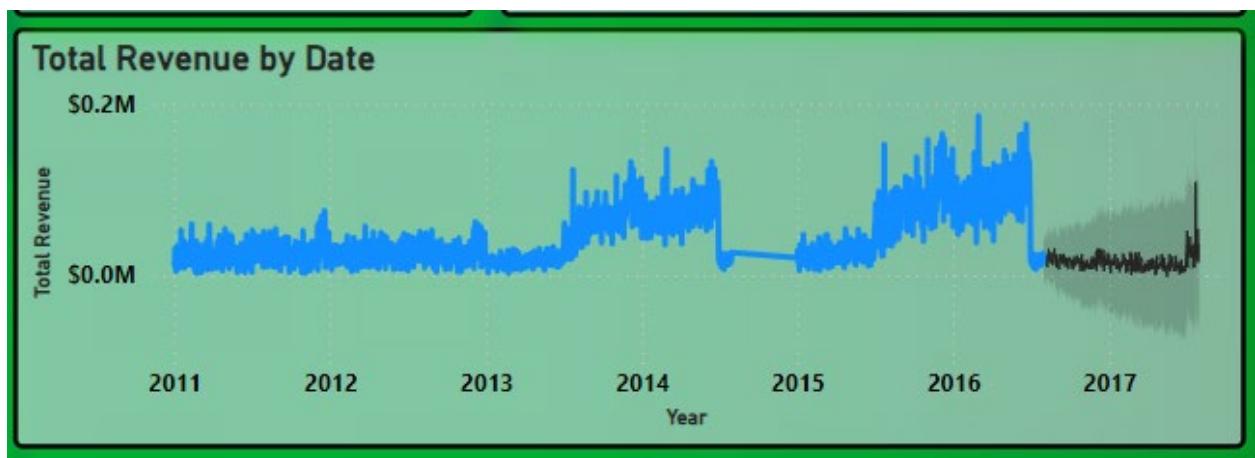


**Figure 16: Chart showing Total Revenue and Total Profit Margin (%) by Quarter**

**Table 18: Analysis of Chart showing Total Revenue and Total Profit Margin (%) by Quarter**

<b>Title</b>	Total Revenue and Total Profit Margin (%) by Quarter
<b>Chart Type</b>	Line and Clustered Column Chart
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To compare the Total Revenue in each quarter and total profit margin (%) generated during each quarter.
<b>Analysis</b>	The visual revealed that the 2nd quarter generated the most revenue but has the 2 <sup>nd</sup> lowest profit margin while 3 <sup>rd</sup> quarter with the least total revenue has the biggest profit margin (%). This could be because most accessories and clothing are bought around this period which are less expensive than bikes but are more profitable.

iv. What is the revenue forecast for the next 12 months?



**Figure 17: Chart showing Total Revenue by Date**

**Table 19: Analysis of Chart showing Total Revenue by Date**

<b>Title</b>	Total Revenue by Date
<b>Chart Type</b>	Line Chart
<b>Data Type(s) in the Visual</b>	Categorical & Numerical
<b>Reason for choosing Visual</b>	To show the sales prediction over the next 12 months.
<b>Analysis</b>	The visual revealed that total revenue will be stable till around July 2017 where a boom is expected.

v. How many products are not profitable?

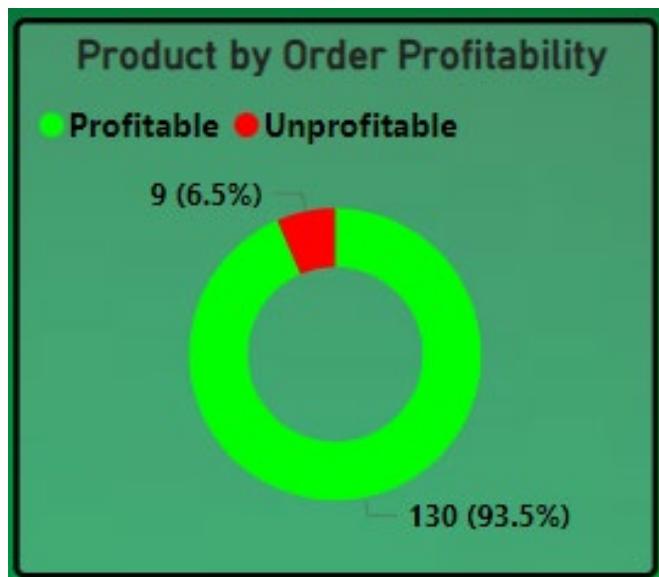


Figure 18: Chart showing Product by Order Profitability

Table 20: Analysis of Chart showing Product by Order Profitability

Title	Product by Order Profitability
Chart Type	Donut Chart
Data Type(s) in the Visual	Categorical
Reason for choosing Visual	To show the distribution of product (categorical) according to order profitability (categorical).
Analysis	Findings from the visual revealed that some transactions under 9 Products were not profitable which is about 6.5% of the total product count.

- vi. How many orders are not profitable and in which subcategory and product category are they in?

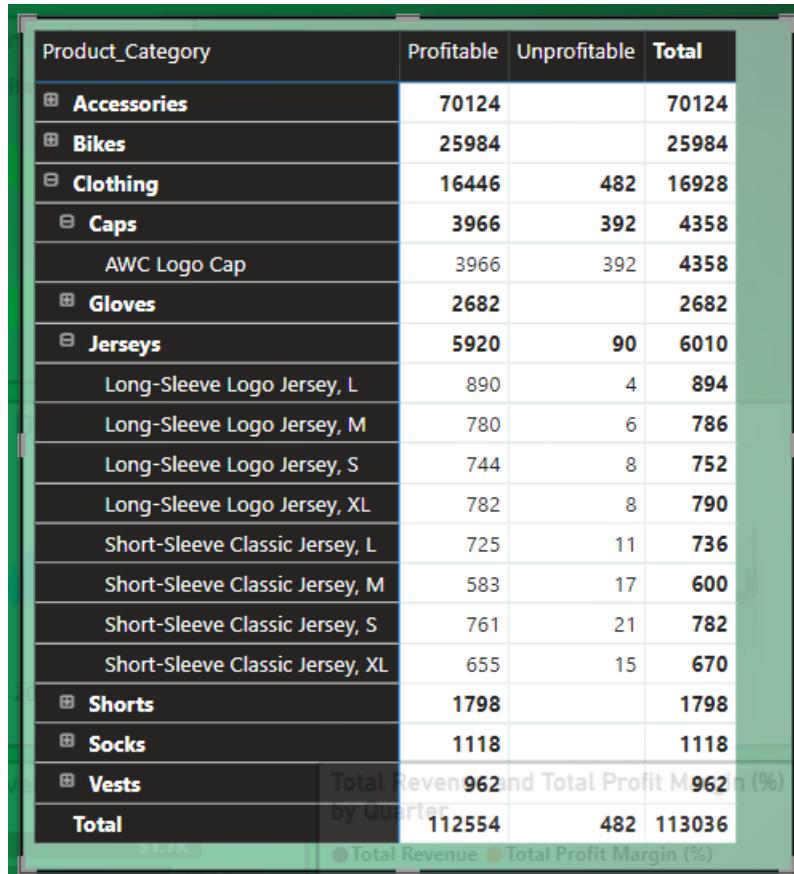


Figure 19: Chart showing Matrix of Product Category by Order Profitability

Table 21: Analysis of Chart showing Matrix of Product Category by Order Profitability

Title	
Chart Type	Matrix
Data Type(s) in the Visual	Categorical & Numerical
Reason for choosing Visual	To investigate further the categories and how many transactions that are not profitable.
Analysis	The visual revealed that a total of four hundred and eighty-two (482) transactions were not profitable. There was no transaction that was not profitable under the Bike and Accessories Categories. 392 and 90 transactions in the Caps and Jerseys subcategories respectively were not profitable under the Clothing Category.

## **Key Findings**

Here is the summary of Key Findings from the Marketing Analysis (see Appendix for full Dashboard)

- ✓ Total revenue increases towards the end of each year.
- ✓ Despite United States contributing most to the total revenue, the lowest ranking states are majorly from the Country.
- ✓ The company makes more profit in the 3<sup>rd</sup> quarter of the year when compared to other quarters.
- ✓ A total of four hundred and eighty-two (482) transactions were not profitable under 9 products. Further investigation showed that they were under the Caps and Jersey Subcategories.

## **Conclusion and Recommendation**

Cycling not only helps our environment but also helps individuals to keep fit and bikes companies like GoGREEN have taken advantage to flourish in this market of bike sales over the year. The bikes category generates the most revenue for the company as it is capital intensive, but the biggest profit margin is from the accessories category. The United States and Australia are the biggest market for bike sales. While some successes have been recorded, it is important for the company to improve its total revenue and profits through the following recommendations:

- 🚲 The company should work more on marketing key influencers as they improve profits for the company.
- 🚲 The company must be more consistent in sales and record keeping, avoiding gaps that occurred in some parts of 2014 and 2015.
- 🚲 Accessories such as helmets should be offered with bike packages as seen that helmets are important accessories sold based on analysis.
- 🚲 For sales to keep moving in an upward trajectory, the company needs to sell more quantity as it was deduced that the quantity sold increased yearly, which

increased total revenue and total profit.

- 🚲 The company needs to holistically review the performance of states not making up to the average total revenue and improve marketing strategies to improve sales.
- 🚲 Since sales are doing well with the working population, the company needs to develop marketing strategies to improve sales for the age group under 25 years.
- 🚲 The slight differences between the male and female purchasing history should be considered and marketing and sales offers should be designed as appropriate.
- 🚲 More focus should be placed on low performing states as they looked to be an untapped market available for generation of new streams of income.
- 🚲 A market assessment of products that are not profitable should be carried out and some of these products can be dropped as contribute little to the total revenue of the company.

## ICA – Appendix: BI Design

### Appendix A: Data Pre-Processing and Data Cleaning

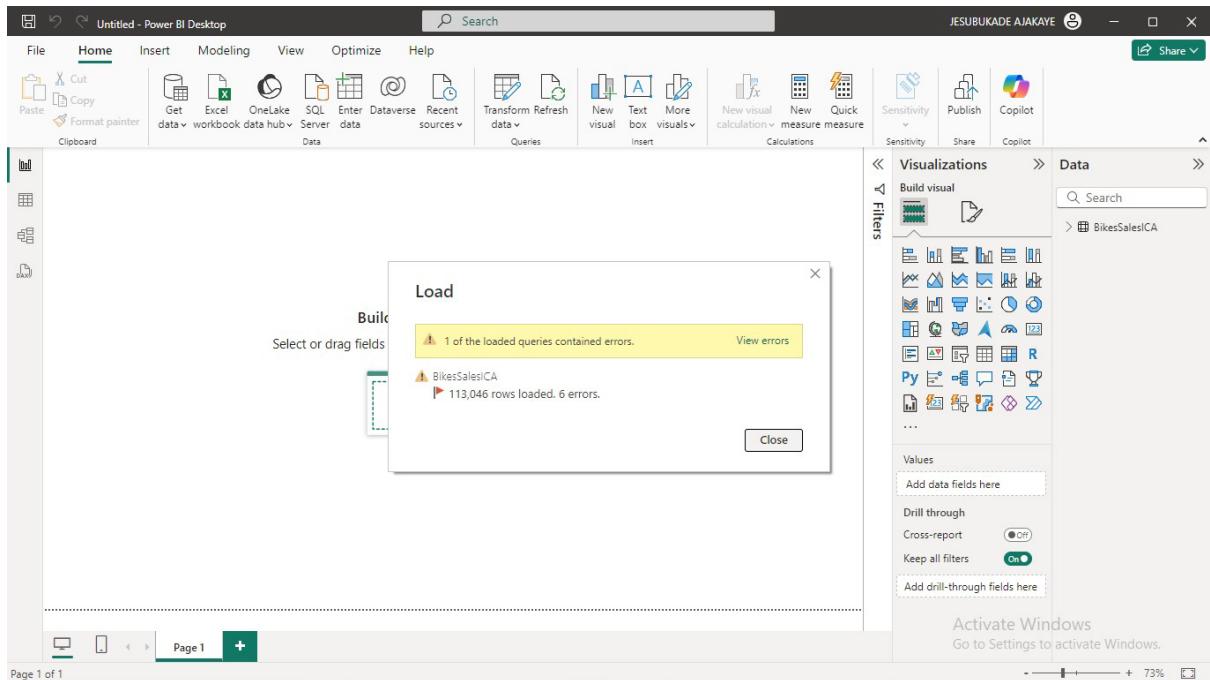
The dataset I worked on was a clean dataset, so I added 10 rows with some missing values to be able to demonstrate my pre-processing and data cleaning skills.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	Date	Day	Month	Year	Customer_Age_Group	Customer_Country	State	Product_C	Sub_Catag	Product	Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue			
2	07/08/2013	8	July	2013	29	Young	Adu M	Germany	Hessen	Clothing	Vests	NA	20	24	64	710	480	1190	
3	07/08/2015	8	July	2015	29	Young	Adu M	Germany	Hessen	Clothing	NA	Classic Ve	21	24	64	746	504	1250	
4	12/28/2013	28	December	2013	41	Adults	(35- M	United King	England	Clothing	NA	Classic Ve	2	24	64	75	48	123	
5	12/28/2015	28	December	2015	41	Adults	(35- M	United King	England	Clothing	Vests	NA	NA	2	24	64	75	48	123
6	04/12/2014	12	April	2014	41	Adults	(35- M	United King	England	Clothing	Vests	Classic Ve	NA	24	64	225	144	369	
7	04/12/2016	12	April	2016	41	Adults	(35- M	United King	England	Clothing	Vests	Classic Ve	3	NA	64	112	72	184	
8	04/02/2014	2	April	2014	18	Youth	(<25 M	Australia	Queensl	Clothing	Vests	Classic Ve	22	24	NA	655	528	1183	
9	04/02/2016	2	April	2016	18	Youth	(<25 M	Australia	Queensl	Clothing	Vests	Classic Ve	22	24	64	NA	528	1183	
10	03/04/2014	4	March	2014	37	Adults	(35- F	France	Seine (Pari	Clothing	Vests	Classic Ve	24	24	64	684	NA	1260	
11	03/04/2016	4	March	2016	37	Adults	(35- F	France	Seine (Pari	Clothing	Vests	Classic Ve	23	24	64	655	552	NA	
12	11/26/2013	26	November	2013	19	Youth	(<25 M	Canada	British Col	Accessory	Bike Racks	Hitch Rack	8	45	120	590	360	950	
13	11/26/2015	26	November	2015	19	Youth	(<25 M	Canada	British Col	Accessory	Bike Racks	Hitch Rack	8	45	120	590	360	950	
14	3/23/2014	23	March	2014	49	Adults	(35- M	Australia	New South	Accessory	Bike Racks	Hitch Rack	23	45	120	1366	1035	2401	
15	3/23/2016	23	March	2016	49	Adults	(35- M	Australia	New South	Accessory	Bike Racks	Hitch Rack	20	45	120	1188	900	2088	
16	5/15/2014	15	May	2014	47	Adults	(35- F	Australia	New South	Accessory	Bike Racks	Hitch Rack	4	45	120	238	180	418	
17	5/15/2016	15	May	2016	47	Adults	(35- F	Australia	New South	Accessory	Bike Racks	Hitch Rack	5	45	120	297	225	522	
18	5/22/2014	22	May	2014	47	Adults	(35- F	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	4	45	120	199	180	379	
19	5/22/2016	22	May	2016	47	Adults	(35- F	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	2	45	120	100	90	190	
20	2/22/2014	22	February	2014	35	Adults	(35- M	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	22	45	120	1096	990	2086	
21	2/22/2016	22	February	2016	35	Adults	(35- M	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	21	45	120	1046	945	1991	
22	7/30/2013	30	July	2013	32	Young	Adu F	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	8	45	120	398	360	758	
23	7/30/2015	30	July	2015	32	Young	Adu F	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	8	45	120	398	360	758	
24	7/15/2013	15	July	2013	34	Young	Adu M	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	7	45	120	349	315	664	
25	7/15/2015	15	July	2015	34	Young	Adu M	Australia	Victoria	Accessory	Bike Racks	Hitch Rack	7	45	120	349	315	664	
26	8/02/2013	2	August	2013	29	Young	Adu M	Canada	British Col	Accessory	Bike Racks	Hitch Rack	5	45	120	369	225	594	
27	8/02/2015	2	August	2015	29	Young	Adu M	Canada	British Col	Accessory	Bike Racks	Hitch Rack	7	45	120	517	315	832	
28	09/02/2013	2	September	2013	29	Young	Adu M	Canada	British Col	Accessory	Bike Racks	Hitch Rack	2	45	120	148	90	238	

Figure 20: Preview of dataset after introducing faulty rows on excel

The screenshot shows the Microsoft Power Query Editor interface. The top menu bar includes File, Home, Transform, Add Column, View, Tools, and Help. The main area displays a preview of a query named 'BikesSalesICA'. The preview shows 26 rows of data with columns: Sub\_Catag, Product, Order\_Quantity, Unit\_Cost, Unit\_Price, and Profit. Some cells contain errors or invalid data, such as 'Error' in the Sub\_Catag column for rows 5, 6, 8, 9, 11, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, and 26. The right side of the screen features the 'Properties' pane with settings for the query, and the 'Applied Steps' pane, which currently lists a single step: 'Changed Type'.

Figure 21: Preview of data when imported to Power BI



**Figure 22: Interface showing error loaded**

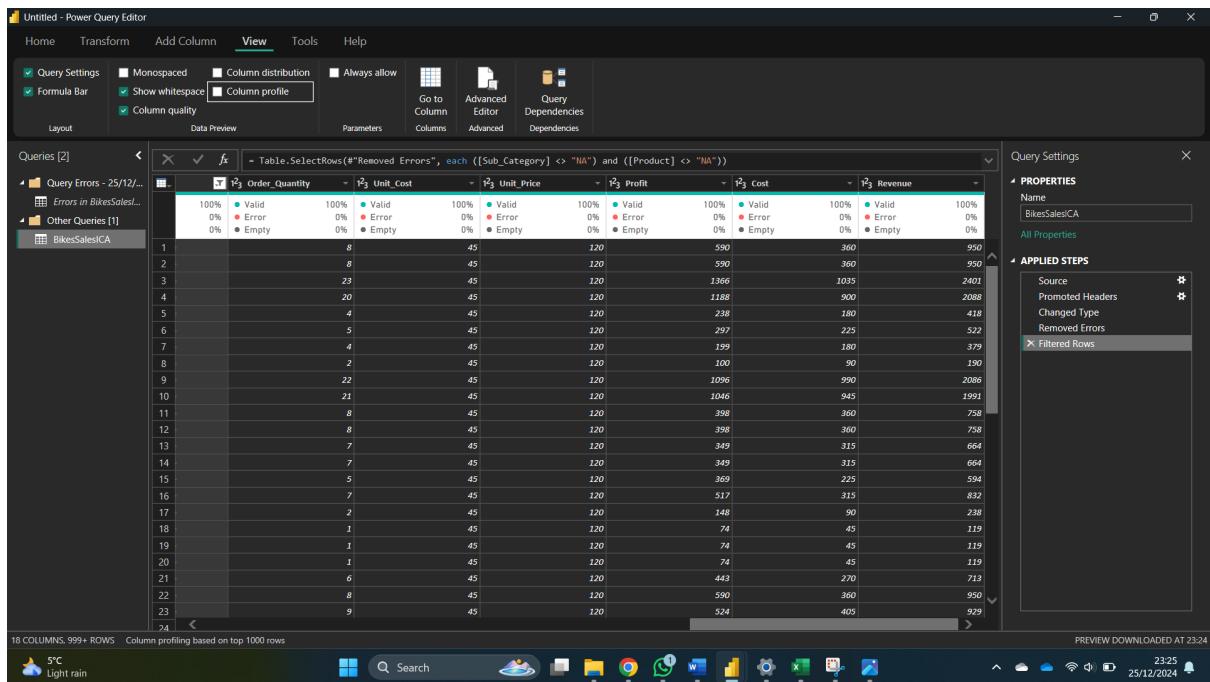
## Dealing with errors

Sub_Category	Product	Order_Quantity	Unit_Cost	Unit_Price	Profit
Vests	NA	20	24	64	
NA	Classic Vest, L	21	24	64	
	Classic Vest, S	2	24	64	
Vests		2	24	64	
Vests	Classic Vest, S	Error	24	64	
Vests	Classic Vest, S	3	Error	64	
Vests	Classic Vest, M	22	24	Error	
Vest	Classic Vest, M	22	24	64	Error
Vests	Classic Vest, L	24	24	64	
Vests	Classic Vest, L	23	24	64	
Bike Racks	Hitch Rack - 4-Bike	8	45	120	
Bike Racks	Hitch Rack - 4-Bike	8	45	120	
Bike Racks	Hitch Rack - 4-Bike	23	45	120	
Bike Racks	Hitch Rack - 4-Bike	20	45	120	
Bike Racks	Hitch Rack - 4-Bike	15	45	120	
Bike Racks	Hitch Rack - 4-Bike	16	45	120	
Bike Racks	Hitch Rack - 4-Bike	17	45	120	
Bike Racks	Hitch Rack - 4-Bike	18	45	120	
Bike Racks	Hitch Rack - 4-Bike	19	45	120	
Bike Racks	Hitch Rack - 4-Bike	20	45	120	
Bike Racks	Hitch Rack - 4-Bike	21	45	120	
Bike Racks	Hitch Rack - 4-Bike	22	45	120	
Bike Racks	Hitch Rack - 4-Bike	23	45	120	
Bike Racks	Hitch Rack - 4-Bike	24	45	120	
Bike Racks	Hitch Rack - 4-Bike	25	45	120	

**Figure 23: Query interface showing errors**

- Clicked on the columns to remove errors
- Filtered to remove NAs and blanks
- Checked data types to ensure that they are correct

Note: All data cleaning steps carried out are shown in the “APPLIED STEP” section (see *Figure 24*)



**Figure 24: Showing dataset processes after cleaning**

## Manipulating Columns

The following steps were carried out when manipulating columns to show my pre-processing

- Removing Columns – majorly carried out when duplicating tables from the flat table
- Renaming Columns – carried out when duplicating columns
- Reordering Columns – to ensure the columns are logically ordered
- Removing Duplicates – duplicate values were removed from the dimension tables (see APPLIED STEPS in *Figure 25-29*)

Queries [2]

BikesSalesICA

Date

Date

Day

Month

Year

11/26/2013

26 November

2013

11/26/2015

26 November

2015

3/23/2014

23 March

2014

3/23/2016

23 March

2016

5/15/2014

15 May

2014

5/15/2016

15 May

2016

5/22/2014

22 May

2014

5/22/2016

22 May

2016

2/22/2014

22 February

2014

2/22/2016

22 February

2016

7/30/2013

30 July

2013

7/30/2015

30 July

2015

7/15/2013

15 July

2013

7/15/2015

15 July

2015

8/2/2013

2 August

2013

8/2/2015

2 August

2015

9/2/2013

2 September

2013

9/2/2015

2 September

2015

1/22/2014

22 January

2014

1/22/2016

22 January

2016

5/17/2014

17 May

2014

5/17/2016

17 May

2016

3/27/2014

27 March

2014

3/27/2016

27 March

2016

8/25/2013

25 August

2013

8/25/2015

25 August

2015

8/25/2016

25 August

2016

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

PREVIEW DOWNLOADED AT 11:53 PM

**Figure 25: Remove other columns when creating the date table**

Queries [6]

BikesSalesICA

DateTable

AgeTable

GenderTable

LocationTable

ProductTable

Product

Order\_Quantity

Profit

Cost

Revenue

Hitch Rack - 4-Bike

8

590

360

950

Hitch Rack - 4-Bike

8

590

360

950

Hitch Rack - 4-Bike

23

1366

1035

2401

Hitch Rack - 4-Bike

20

1188

900

2088

Hitch Rack - 4-Bike

4

238

180

418

Hitch Rack - 4-Bike

5

297

225

522

Hitch Rack - 4-Bike

4

199

180

379

Hitch Rack - 4-Bike

2

100

90

190

Hitch Rack - 4-Bike

22

1096

990

2086

Hitch Rack - 4-Bike

21

1046

945

1991

Hitch Rack - 4-Bike

8

398

360

758

Hitch Rack - 4-Bike

8

398

360

758

Hitch Rack - 4-Bike

7

349

315

664

Hitch Rack - 4-Bike

7

349

315

664

Hitch Rack - 4-Bike

5

369

225

594

Hitch Rack - 4-Bike

7

517

315

832

Hitch Rack - 4-Bike

2

148

90

238

Hitch Rack - 4-Bike

1

74

45

119

Hitch Rack - 4-Bike

1

74

45

119

Hitch Rack - 4-Bike

1

74

45

119

Hitch Rack - 4-Bike

6

443

270

713

Hitch Rack - 4-Bike

8

590

360

950

Hitch Rack - 4-Bike

9

524

405

929

Hitch Rack - 4-Bike

7

407

315

722

Hitch Rack - 4-Bike

3

221

135

356

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

PREVIEW DOWNLOADED AT 12:22 AM

**Figure 26: Before reordering columns**

SalesInventoryTable

	Product	Order_Quantity	Cost	Revenue	Profit
1	Hitch Rack - 4-Bike	8	360	950	590
2	Hitch Rack - 4-Bike	8	360	950	590
3	Hitch Rack - 4-Bike	23	1035	2401	1366
4	Hitch Rack - 4-Bike	20	900	2088	1188
5	Hitch Rack - 4-Bike	4	180	418	238
6	Hitch Rack - 4-Bike	5	225	522	297
7	Hitch Rack - 4-Bike	4	180	379	199
8	Hitch Rack - 4-Bike	2	90	190	100
9	Hitch Rack - 4-Bike	22	990	2086	1096
10	Hitch Rack - 4-Bike	21	945	1991	1046
11	Hitch Rack - 4-Bike	8	360	758	398
12	Hitch Rack - 4-Bike	8	360	758	398
13	Hitch Rack - 4-Bike	7	315	664	349
14	Hitch Rack - 4-Bike	7	315	664	349
15	Hitch Rack - 4-Bike	5	225	594	369
16	Hitch Rack - 4-Bike	7	315	832	517
17	Hitch Rack - 4-Bike	2	90	238	148
18	Hitch Rack - 4-Bike	1	45	119	74
19	Hitch Rack - 4-Bike	1	45	119	74
20	Hitch Rack - 4-Bike	1	45	119	74
21	Hitch Rack - 4-Bike	6	270	713	443
22	Hitch Rack - 4-Bike	8	360	950	590
23	Hitch Rack - 4-Bike	9	405	929	524
24	Hitch Rack - 4-Bike	7	315	722	407
25	Hitch Rack - 4-Bike	3	135	356	221
26					

9 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 12:25 AM

Figure 27: After reordering columns

SalesInventoryTable

	Month	Year	Day of the Week	Date - Copy	Quarter
1	26 November	2013	Tuesday	11	4
2	26 November	2015	Thursday	11	4
3	23 March	2014	Sunday	3	1
4	23 March	2016	Wednesday	3	1
5	15 May	2014	Thursday	5	2
6	15 May	2016	Sunday	5	2
7	22 May	2014	Thursday	5	2
8	22 May	2016	Sunday	5	2
9	22 February	2014	Saturday	2	1
10	22 February	2016	Monday	2	1
11	30 July	2013	Tuesday	7	3
12	30 July	2015	Thursday	7	3
13	15 July	2013	Monday	7	3
14	15 July	2015	Wednesday	7	3
15	2 August	2013	Friday	8	3
16	2 August	2015	Sunday	8	3
17	2 September	2013	Monday	9	3
18	2 September	2015	Wednesday	9	3
19	22 January	2014	Wednesday	1	1
20	22 January	2016	Friday	1	1
21	17 May	2014	Saturday	5	2
22	17 May	2016	Tuesday	5	2
23	27 March	2014	Thursday	3	1
24	27 March	2016	Sunday	3	1
25	25 August	2013	Sunday	8	3
26					

7 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 1:58 PM

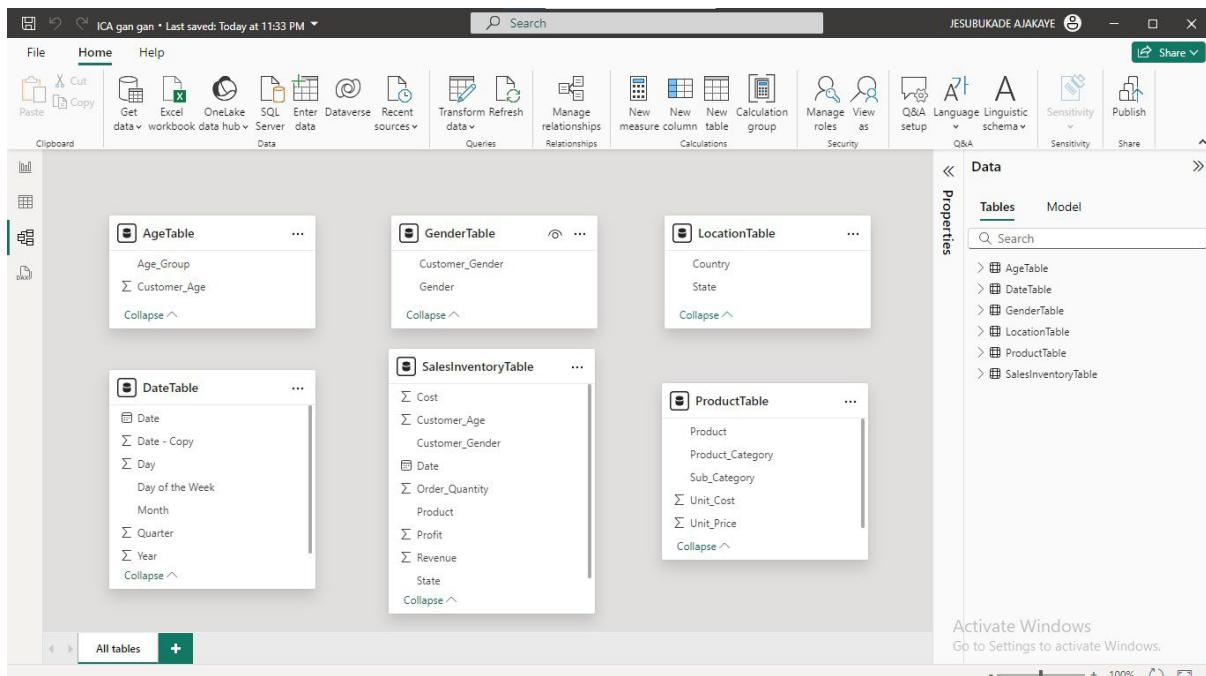
Figure 28: Before renaming columns

**Figure 29: After renaming columns**

## Appendix B: BI Data Modelling via Star Schema – Fact and Dimension Tables

Based on the BI questions, five-dimension tables and a flat table were created during data pre-processing. This is carried out by right clicking on the flat table and selecting “duplicate table”. Columns not needed were removed as shown in Appendix A and duplicates were removed especially to ensure there were no “many to many” relationships when connecting tables. “SalesInventoryTable” is the Fact Table, while “DataTable”, “AgeTable”, “GenderTable”, “LocationTable” and “ProductTable” are the Dimension Tables.

**Figure 30: Table created from the flat table**



**Figure 31: Model view before relationships**

## Relationships

All the dimension tables were connected to the fact table and previous preprocessing steps ensured that One to Many (1:\*) relationships was achieved between all the dimension tables and the fact table. Also, the cross-filter direction was set to both ways such that any dimension table can be able to access information on another dimension table and still get accurate results.

## Keys

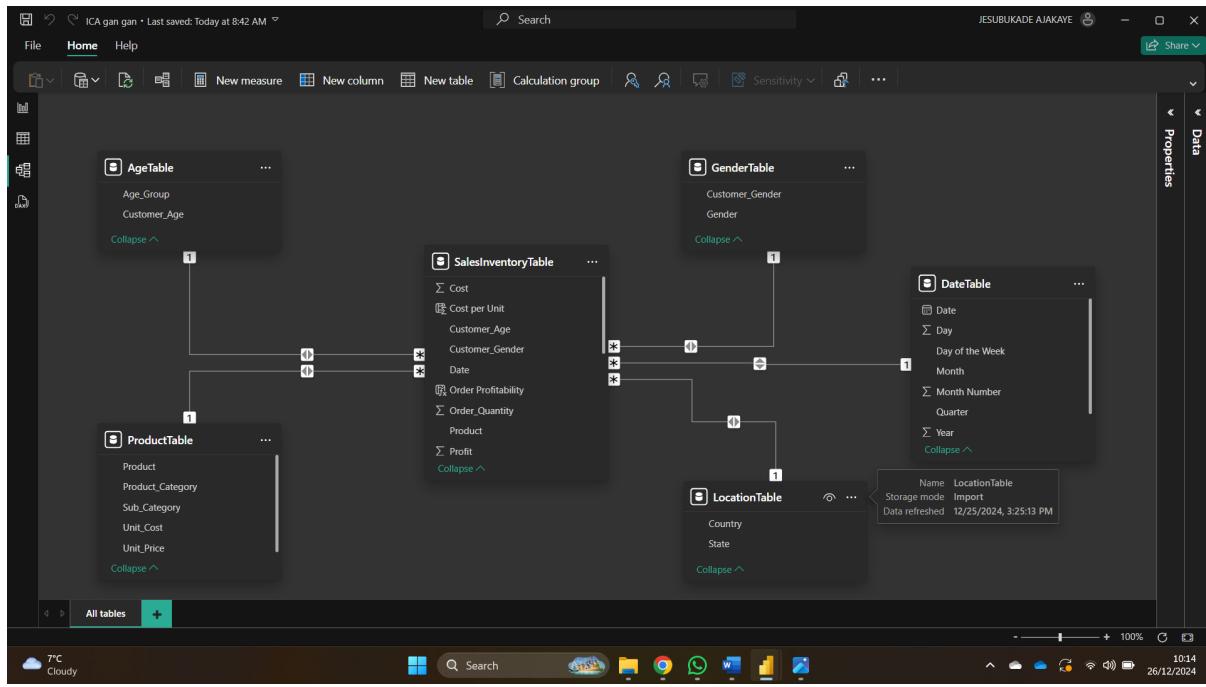
**AgeTable** – “Customer\_Age” was the key used to connect this table to the fact table. Duplicates had already been removed hence they could perform effectively when used.

**DateTable** – The “Date” column was used as the connecting key. The unique days has been filtered and serves as a feature that can be used to connect this table to the fact table.

**GenderTable** - “Customer\_Gender” which is the initials of the available genders in the dataset was used to connect this table. M – Male and F – Female and served as unique identifiers on the table.

**LocationTable** – “State” it was better to use the state as a key for this table as the state links to their country name. Duplicate values were already removed hence this performed effectively when used as key in this dataset.

**ProductTable** – “Product” column was used as the connecting key as it’s a set of unique values linking to another column on this table. Performed effectively in the model.



**Figure 32: Model view after relationships using Star Schema**

## Appendix C: DAX and M Language

### Measures and Calculated Columns using DAX

Some measures and calculated columns were created using DAX to enable smooth analysis and for display on cards and are listed as follows:

#### Profit Margin (%) (Calculated Column)

This was calculated dividing Profit by Revenue (see *Figure 33*) It helps us to understand how much profit the company was making from each transaction.

The screenshot shows the Power BI Column Tools view. A calculated column named "Profit Margin (%)" is being defined. The formula is `Profit Margin (%) = DIVIDE(SalesInventoryTable[Profit], SalesInventoryTable[Revenue])`. The column is set to have a decimal number data type and a percentage format. The table below shows a single row of data.

Date	Customer_Age	Customer_Gender	State	Product	Order_Quantity	Profit	Cost	Revenue	Sales_ID	Profit Margin (%)	Order Profitability
09/08/2011	24	M	California	Mountain-200 Silver, 42	1	1008	1266	2274	49946	44.33%	Profitable

**Figure 33: DAX for profit margin (%)**

#### Order Profitability (Calculated Column)

The column was created by checking a condition on the profit margin (%) column (see *Figure 34*). If the margin was greater than 0, the transaction was recorded as profitable. This helps the company to understand where they are making a profit or not.

The screenshot shows the Power BI DAX Editor interface. The top navigation bar includes File, Home, Help, Table tools, and Column tools. The Column tools tab is selected. The 'Name' field is set to 'Order Profitability'. The 'Format' dropdown is set to 'Text'. The 'Summarization' dropdown is set to 'Don't summarize'. The 'Data category' dropdown is set to 'Uncategorized'. Below these settings, there are sections for 'Structure', 'Formatting', 'Properties', 'Sort', 'Groups', 'Relationships', and 'Calculations'. A preview pane at the bottom shows a single row of data with the calculated value 'Profitable'.

**Figure 34: DAX for order profitability**

### Total Revenue (Measure)

This measure was created using the SUM function in DAX (see *Figure 35*). It adds up all the revenue column and provides output for display or analysis.

The screenshot shows the Power BI DAX Editor interface. The top navigation bar includes File, Home, Help, Table tools, and Measure tools. The Measure tools tab is selected. The 'Name' field is set to 'Total Revenue'. The 'Format' dropdown is set to 'Currency'. The 'Home table' dropdown is set to 'SalesInventoryTable'. Below these settings, there are sections for 'Structure', 'Formatting', 'Properties', and 'Calculations'. A preview pane at the bottom shows the formula 'Total Revenue = SUM(SalesInventoryTable[Revenue])'.

**Figure 35: DAX for total revenue**

### Total Profit Margin (%) (Measure)

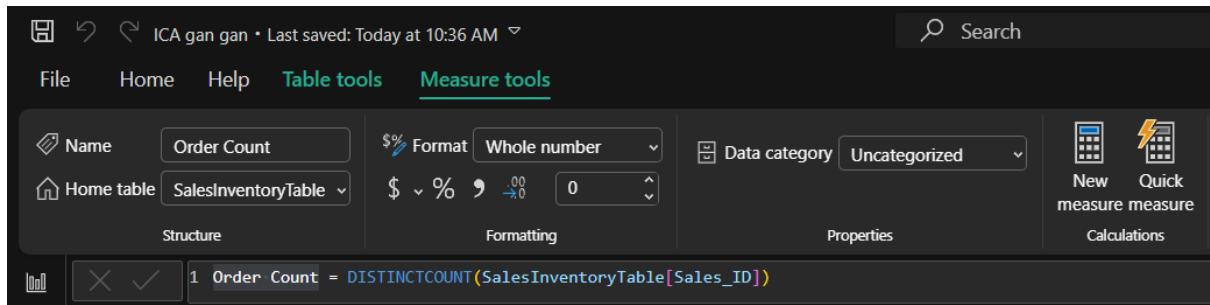
This measure was created using the DIVIDE function in DAX (see *Figure 36*). It divides the Total Profit measure by the Total revenue measure and provides output for display or for analysis.

The screenshot shows the Power BI DAX Editor interface. The top navigation bar includes File, Home, Help, Table tools, and Measure tools. The Measure tools tab is selected. The 'Name' field is set to 'Total Profit Margin...'. The 'Format' dropdown is set to 'Percentage'. The 'Home table' dropdown is set to 'SalesInventoryTable'. Below these settings, there are sections for 'Structure', 'Formatting', 'Properties', and 'Calculations'. A preview pane at the bottom shows the formula 'Total Profit Margin (%) = DIVIDE([Total Profit], [Total Revenue])'.

**Figure 36: DAX for total profit margin (%)**

### Order Count (Measure)

This measure was created using the DISTINCTCOUNT function in DAX (see *Figure 37*). It performs an addition of all the entries (transactions) in the SaleInventoryTable and provides output for display or for analysis.



**Figure 37: DAX for order count**

**Table 22: Other Measure Created**

Name	DAX function used	Description
Average Order Quantity	DIVIDE, SUM DISTINCTCOUNT	Returns the average order quantity per State
Average Total Revenue	DIVIDE DISTINCTCOUNT	Returns the average total revenue per State
Total Cost	SUM	Returns the sum of all entries in the cost column.
Total Profit	SUM	Returns the sum of all entries in the profit column.

The screenshot shows the Power BI interface with the 'Table tools' ribbon selected. The main area displays a table with columns: Date, Customer\_Age, Customer\_Gender, State, Product, Order\_Quantity, Profit, Cost, Revenue, and Sales\_ID. The data consists of 113,036 rows. To the right, the 'Data' pane is open, showing the structure of the 'SalesInventoryTable' with columns: Cost, Customer\_Age, Customer\_Gender, Date, Order\_Quantity, Product, Profit, Revenue, Sales\_ID, and State. The 'Activate Windows' watermark is visible at the bottom right.

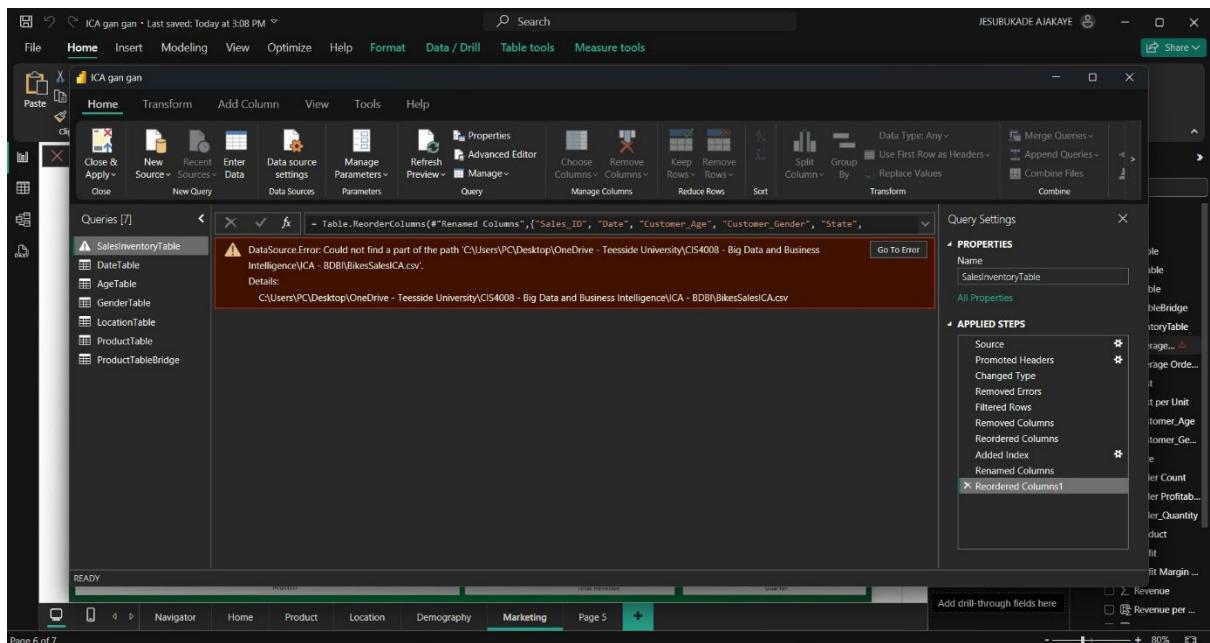
**Figure 38: Interface before creating measures and calculated columns**

The screenshot shows the Power BI desktop application interface. The main area displays a table named "SalesInventoryTable" with 113,036 rows. The columns include Date, Customer\_Age, Customer\_Gender, State, Product, Order\_Quantity, Profit, Cost, Revenue, Sales\_ID, Profit Margin (%), and Order Profitability. A context menu is open over the table, and the "Data" pane on the right lists various measures and calculated columns such as Average Order Quantity, Average Total Revenue, Cost, Order Count, Order Profitability, Order\_Quantity, Profit, Profit Margin (%), Revenue, Sales\_ID, State, Total Cost, Total Profit, Total Profit Margin (%), and Total Revenue.

**Figure 39: Interface after creating measures and calculated columns**

## Application of M Language

During the completion of this report, I had to switch the personal computer (PC) I was working on hence an error occurred while I tried to do some transformation because data source cannot be found (see error in figure below). I then imported the dataset into my Power BI work view, opened the advanced editor to see the M Language statements, then, copied the correct data source to all my tables therefore I did not need to start my data preprocessing and cleaning all over again (see figure below for before and after manipulating the M Statement)



**Figure 40: Error view when trying to transform on new PC**

```

let
    Source = Csv.Document(File.Contents("C:\Users\lucas\Desktop\OneDrive - Teesside University\CIS4008 - Big Data and Business Intelligence\ICA - BOBI\BikesSalesICA.csv"),[Delimiter=",", Columns=18, Encoding=1252, QuoteStyle=QuoteStyleNone]),
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Date", type date}, {"Day", Int64.Type}, {"Month", type text}, {"Year", Int64.Type}, {"Customer_Age", Int64.Type}, {"Age_Group", type text}, {"Customer_Gender", Int64.Type}, {"State", type text}, {"Product", type text}, {"Product_Category", type text}, {"Sub_Category", type text}, {"Unit_Cost", "Unit_Price"}),
    #"Removed Errors" = Table.RemoveRowsWithErrors(#"Changed Type"),
    #"Filtered Rows" = Table.SelectRows(#"Removed Errors", each ((Product) <> "N/A" and ({Sub_Category} <> "N/A")),
    #"Reordered Columns" = Table.ReorderColumns(#"Filtered Rows",{"Day", "Month", "Year", "Age_Group", "Country", "Product_Category", "Sub_Category", "Unit_Cost", "Unit_Price"}, {"Customer_Age", "Customer_Gender", "State", "Product", "Order_Quantity", "Cost", "Revenue", "Profit"}),
    #"Added Index" = Table.AddIndexColumn(#"Reordered Columns", "Index", 1, 1, Int64.Type),
    #"Renamed Columns" = Table.RenameColumns(#"Added Index",{{"Index", "Sales_ID"}}),
    #"Reordered Columns1" = Table.ReorderColumns(#"Renamed Columns",{"Sales_ID", "Date", "Customer_Age", "Customer_Gender", "State", "Product", "Order_Quantity", "Cost", "Revenue", "Profit"})
in
    #"Reordered Columns1"

```

✓ No syntax errors have been detected.

Done Cancel

**Figure 41: Source before manipulating the M Language**

```

let
    Source = Csv.Document(File.Contents("C:\Users\lucas\Desktop\OneDrive - Teesside University\CIS4008 - Big Data and Business Intelligence\ICA - BOBI\BikesSalesICA.csv"),[Delimiter=",", Columns=18, Encoding=1252, QuoteStyle=QuoteStyleNone]),
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Date", type date}, {"Day", Int64.Type}, {"Month", type text}, {"Year", Int64.Type}, {"Customer_Age", Int64.Type}, {"Age_Group", type text}, {"Customer_Gender", Int64.Type}, {"State", type text}, {"Product", type text}, {"Product_Category", type text}, {"Sub_Category", type text}, {"Unit_Cost", "Unit_Price"}),
    #"Removed Errors" = Table.RemoveRowsWithErrors(#"Changed Type"),
    #"Filtered Rows" = Table.SelectRows(#"Removed Errors", each ((Product) <> "N/A" and ({Sub_Category} <> "N/A")),
    #"Reordered Columns" = Table.ReorderColumns(#"Filtered Rows",{"Day", "Month", "Year", "Age_Group", "Country", "Product_Category", "Sub_Category", "Unit_Cost", "Unit_Price"}, {"Customer_Age", "Customer_Gender", "State", "Product", "Order_Quantity", "Cost", "Revenue", "Profit"}),
    #"Added Index" = Table.AddIndexColumn(#"Reordered Columns", "Index", 1, 1, Int64.Type),
    #"Renamed Columns" = Table.RenameColumns(#"Added Index",{{"Index", "Sales_ID"}}),
    #"Reordered Columns1" = Table.ReorderColumns(#"Renamed Columns",{"Sales_ID", "Date", "Customer_Age", "Customer_Gender", "State", "Product", "Order_Quantity", "Cost", "Revenue", "Profit"})
in
    #"Reordered Columns1"

```

✓ No syntax errors have been detected.

Done Cancel

**Figure 42: Source after manipulating the M Language**

## Appendix D: Dashboards

Seven dashboards were created for the whole project.

### Navigator Page

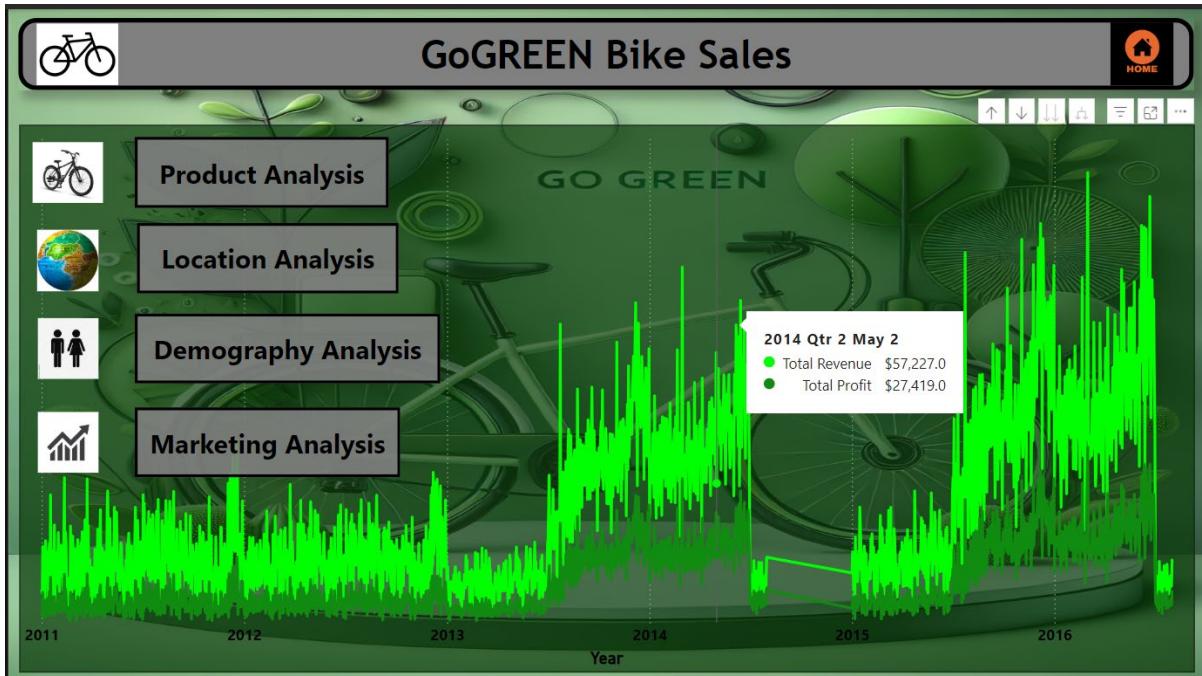


Figure 43: View of the Navigator Dashboard

- The navigation page is designed to project the sustainable vision of the company and contains buttons that leads to the four BI Question dashboards.
- A timeseries visual of the total revenue and total profit was added to this page to complement the design and hovering over this area will give some insight into the company performance over the period studied.

## Home Page

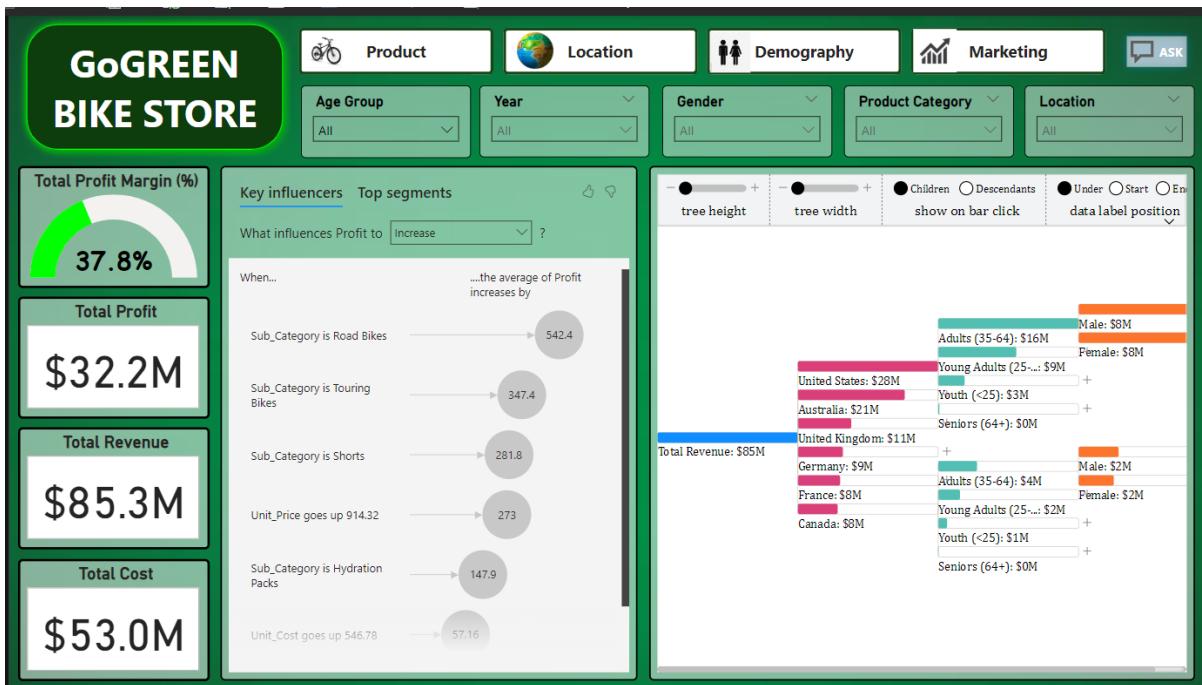


Figure 44: View of Home Dashboard

- The idea of Home page was to capture key insights and summary that on-the-spot decision can be made from, and AI insight was one of the most productive ways of achieving this.
- Buttons to BI Dashboards are arranged for seamless navigation and the Q&A button to lead to the Questioning Dashboard.
- Under which are the slicers/filters to get different insights from the Home Dashboard.
- To the left side of the page are cards showing the output of the key measures created using DAX
- Key influencers of Profit are then showed then a decomposition tree shows the spread of total revenue from Countries to Age Group to Gender to Products.

## Product Page

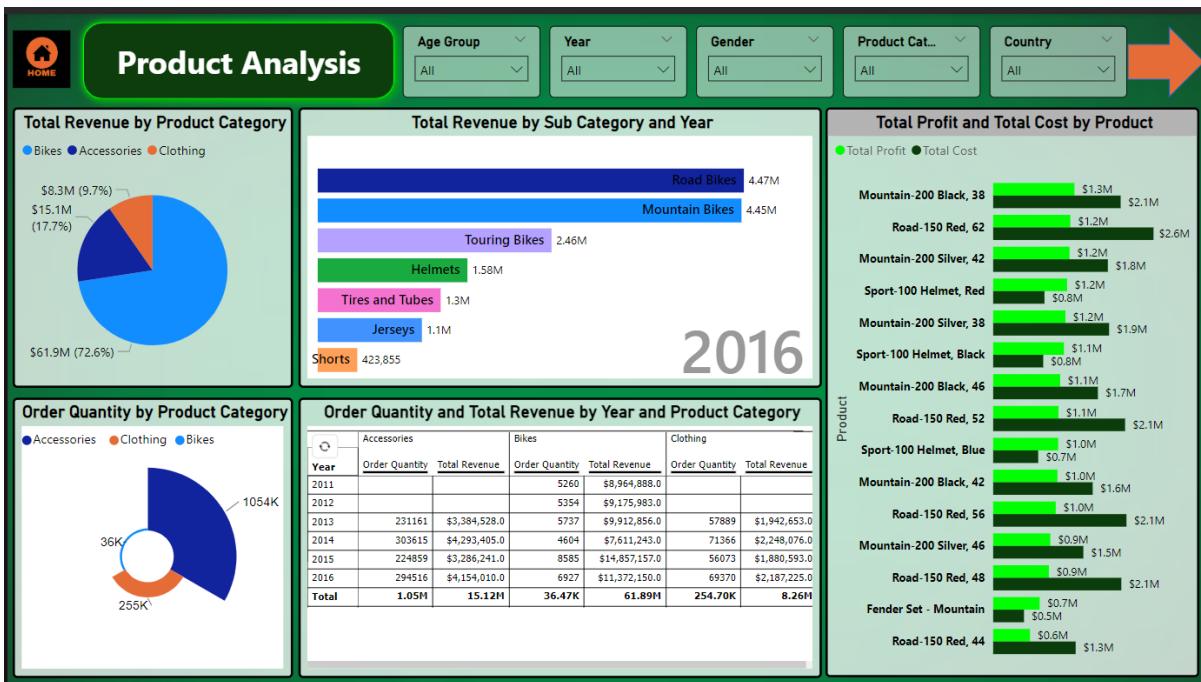


Figure 45: View of Product Dashboard

- This page provides the first BI insight – which is the product analysis
- Slicers/filters to provide insights at various levels were added at the top and the Home and Next buttons were added.
- The pie chart to the left of the dashboard shows the distribution of total revenue according to product categories and complemented by the aster plot under it which shows the distribution of order quality in these categories.
- In the middle is an animated bar chart which shows the total revenue in each subcategory from the start year to the ending year of the report.
- Below this is a business performance table which shows the spread of order quantity and total revenue in the product category in each year which can be drilled down to months (this chart revealed that the company started only with bikes in 2011 and 2012)
- At the right of the dashboard is a clustered bar chart which compares total profits of the top 15 products to their total cost.

## Location Page

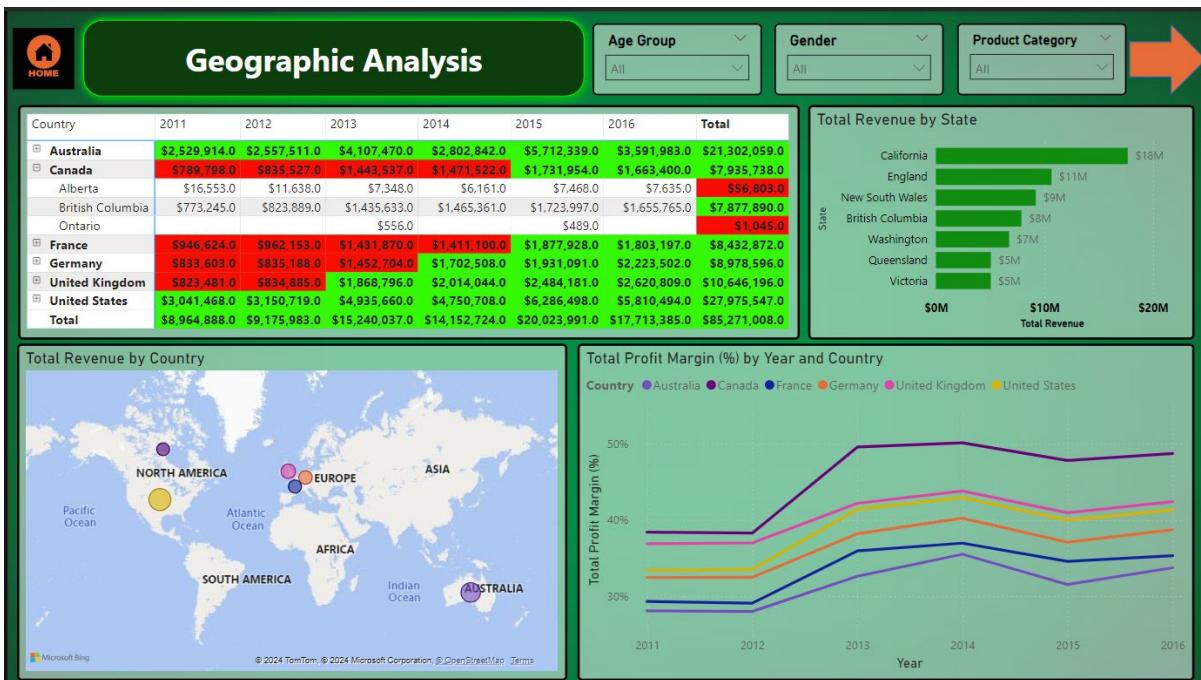


Figure 46: View of Location Dashboard

- ✚ The second BI questions were analysed on this Dashboard which gives insight on the geographic locations represented in the dataset.
- ✚ Slicers/filters were added with Home and Next Buttons
- ✚ Top visuals showed a Matrix Table for Country by Year with background colours added to show total values lower than the average total revenue per state and a cluster column bar chat showing the top performing states in total revenue filter to the Top 7 on Total profit.
- ✚ The Lower visuals is of a Map which shows each country total revenue with the size of a bubble and a Line chart of profit margin of these countries over the years. These two charts applied the same colour code for each country.

## Demography Page

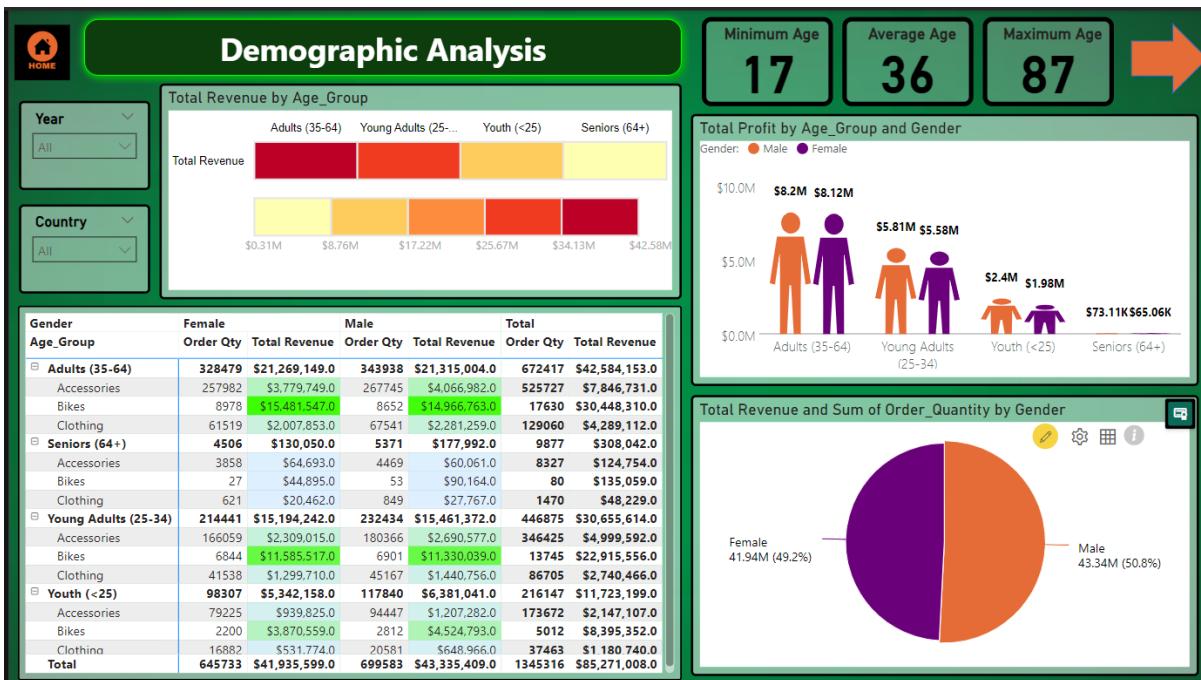


Figure 47: View of Demography Dashboard

- ⊕ This Dashboard provides insight to the third categories of the BI questions on the demography of the customer and how the company performs withing groups.
- ⊕ Filters added on the top left side with Home and Next buttons added on the page.
- ⊕ Three cards at the top to show the minimum, average and maximum age of customers.
- ⊕ Beside the filters is a heatmap of total revenue distribution among the various age groups and below this is a Matrix of Age Group by Gender explaining the total revenue made from the product categories within these groups.
- ⊕ To the right of Dashboard is an infographic visual showing the total profit made from each gender within their age group.
- ⊕ Below is a Pie-Donut chart showing the total revenue and order quantity between genders. The diameter of each pie (total revenue) is based on the sum of order quantity by each group.

## Marketing Page

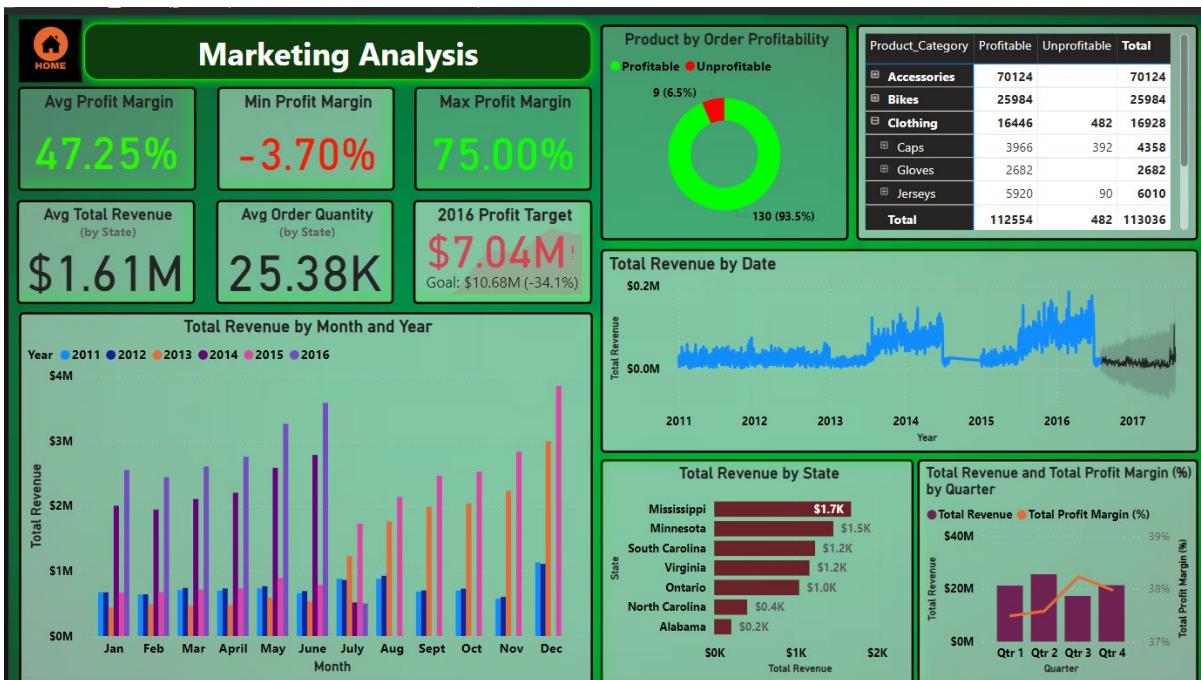
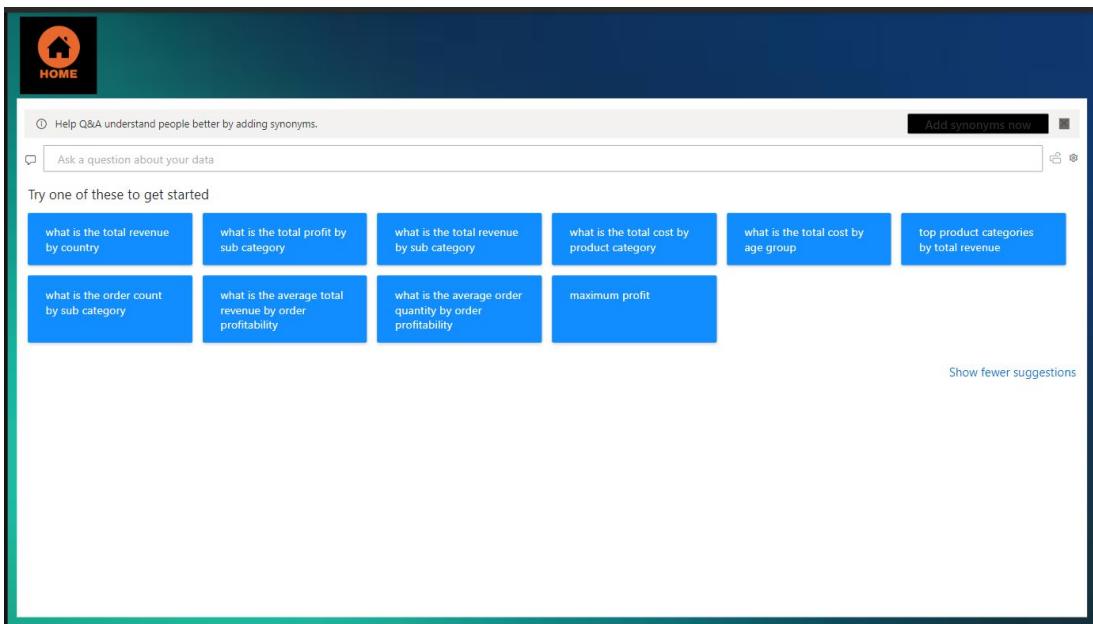


Figure 48: View of Marketing Dashboard

- This Dashboard addresses the last BI questions on areas to improve and help develop marketing strategies to improve total revenue as well as total profit.
- Cards were used to display the average, minimum and maximum profit margin recorded, the average total revenue by state, the average order quantity by state and a KPI card for the Total Profit Goal.
- Under the cards is a visual of a cluster column chart showing the spread of total revenue in each year in each month.
- At the top right of the Dashboard are charts on product profitability explained with a donut chart and further investigated with a matrix
- Below them is a forecast of sales using the time series of total revenue for this forecast to predict the total revenue over the next 12 months.
- At the bottom right is visuals on the lowest performing state on total revenue and a line and column chart for the comparison of revenue and total profit margin quarterly.

## Question Tab



**Figure 49: view of Questioning Dashboard**

- ➡ The Q&A on the Home Dashboard leads here and the user can ask questions in natural language and would get visuals that answer those questions or select from the suggestions provided.

## Appendix E: Self-Assessment

**Table 23: Self-Assessment Score**

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	95
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	90
Dax and M language	Both DAX and M Language have been <b>extensively</b> used in the report	99
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	80
Average		<b>Add below the average of the four cells above:</b> <b>91</b>