Teesside University MIDDLESBROUGH TS1 3BA

School of Computing – MSc Data Science

BIKE SALES ANALYSIS

Big Data and Business Intelligence, CIS 4008-N

Name: ATHIRA REGHUNATH

Student ID: W9561613

Submitted On: 11-01-2023

Table of Contents

SECTION 1: ICA – BUSINESS REPORT

SECTION 2- ICA – APPENDIX: BI DESIGN

BIKE SALES ANALYSIS

NAME: ATHIRA REGHUNATH

STUDENT ID: W9561613

SECTION 1: ICA – BUSINESS REPORT

1. Executive Summary

This report is based on the analysis of the BikeSales data and created a dashboard for finding solutions for the 10 business questions. The data can be mainly divided into four types Year wise, Product wise, Customer wise and Country Wise. It will help to create data visualisation. First, the data cleansing and preprocessing section after data modelling have done. Then the relationships are created. If we want to connect more than one column and making data visualisation is only possible after relationship creation. After that dashboard is designed. The dashboard contains five pages. The first one is the home page, Year wise bike sales, Customer wise bike sales, Product wise bike sales and Country Wise bike sales are the other pages. Each page has different graphs related to the data and finding solutions for the business questions from these graphs.

1.1. Introduction

The major problem faced in any sales sector is the lack of analyses of the previous sales data. The data can tell more than we expect. If we analyse sales data we can know about which country, which kind of people or which product is mostly sold or which gave most profit etc and it will help to the business. Here analyse the bike sales data and seek answers to some questions.

1.1.1. Data Source

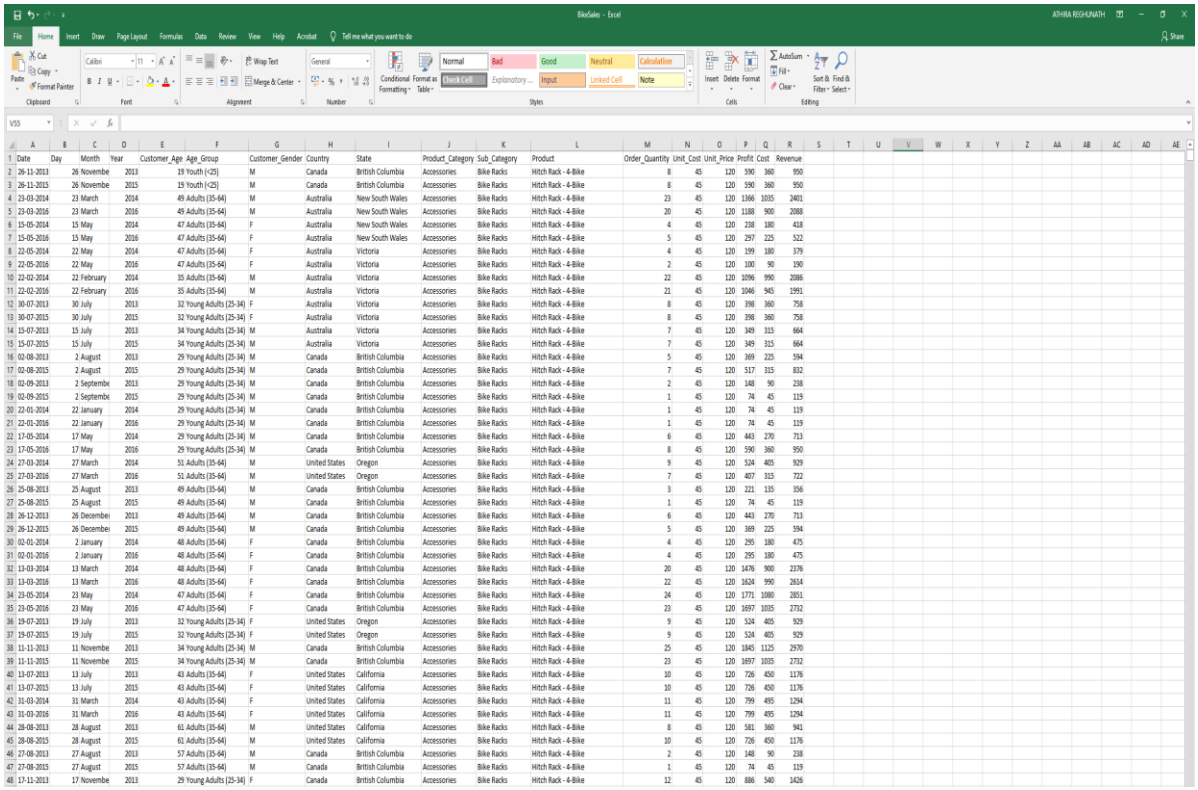
Here I am going to visualise the Bike Sales dataset, this data contains how many bike sales are in Europe and America and how profitable for different age groups, different products and different countries. We can get an idea about how prices are likely to change in the future based on this analysis. The Bike Sales dataset is taken from Kaggle https://www.kaggle.com/sadiqshah/bike-sales-in-europe.This dataset contains 18 columns and 113036 rows.

1.1.2. Description

The dataset consists of the following columns:

1. Date = Date of the order
2. Day = Day of the order
3. Month = Month of the order
4. Year = Year of the order
5. Customer_Age = The Age of the customer
6. Age_Group = Age range for the customer
7. Customer_Gender = The customer gender
8. Country = The country where the order was made
9. State = The state where the order was made
10. Product_Category = Product category for each order
11. Sub_Category = Product sub-category for each order
12. Product = The product for each order

13. Order_Quantity = Quantity of the product
14. Unit_Cost = Cost of inventory holding cost
15. Unit_Price = The price of the product
16. Profit = Net profit of each order
17. Cost = The sum cost of all products for each order
18. Revenue = Total revenue of each order



Date	Day	Month	Year	Customer_Age	Age_Group	Customer_Gender	Country	State	Product_Category	Sub_Category	Product	Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue
26-11-2013	26	Novembe	2013	19	Youth (<25)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
26-11-2015	26	Novembe	2015	19	Youth (<25)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
23-03-2014	23	March	2014	49	Adults (35-64)	M	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	23	45	120	1366	1035	2401
23-03-2016	23	March	2016	49	Adults (35-64)	M	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	20	45	120	1188	900	2088
15-05-2014	15	May	2014	47	Adults (35-64)	F	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	238	180	418
15-05-2016	15	May	2016	47	Adults (35-64)	F	Australia	New South Wales	Accessories	Bike Racks	Hitch Rack - 4-Bike	5	45	120	297	225	522
22-05-2014	22	May	2014	47	Adults (35-64)	F	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	199	180	379
22-05-2016	22	May	2016	47	Adults (35-64)	F	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	2	45	120	100	90	190
22-02-2014	22	February	2014	35	Adults (35-64)	M	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	22	45	120	1096	990	2086
22-02-2016	22	February	2016	35	Adults (35-64)	M	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	21	45	120	1046	945	1991
30-07-2013	30	July	2013	32	Young Adults (25-34)	F	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	398	360	758
30-07-2015	30	July	2015	32	Young Adults (25-34)	F	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	398	360	758
15-07-2013	15	July	2013	34	Young Adults (25-34)	M	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	7	45	120	349	315	664
15-07-2015	15	July	2015	34	Young Adults (25-34)	M	Australia	Victoria	Accessories	Bike Racks	Hitch Rack - 4-Bike	7	45	120	349	315	664
02-08-2013	2	August	2013	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	5	45	120	369	225	594
02-08-2015	2	August	2015	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	7	45	120	517	315	832
02-09-2013	2	Septembe	2013	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	2	45	120	140	90	230
02-09-2015	2	Septembe	2015	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	1	45	120	74	45	119
22-01-2014	22	January	2014	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	1	45	120	74	45	119
22-01-2016	22	January	2016	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	1	45	120	74	45	119
17-05-2014	17	May	2014	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	6	45	120	443	270	713
17-05-2016	17	May	2016	29	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
27-03-2014	27	March	2014	51	Adults (35-64)	M	United States	Oregon	Accessories	Bike Racks	Hitch Rack - 4-Bike	9	45	120	524	405	929
27-03-2016	27	March	2016	51	Adults (35-64)	M	United States	Oregon	Accessories	Bike Racks	Hitch Rack - 4-Bike	7	45	120	407	315	722
25-08-2013	25	August	2013	49	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	3	45	120	221	135	356
25-08-2015	25	August	2015	49	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	1	45	120	74	45	119
26-12-2013	26	Decembei	2013	49	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	6	45	120	443	270	713
26-12-2015	26	Decembei	2015	49	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	5	45	120	369	225	594
02-01-2014	2	January	2014	48	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	295	180	475
02-01-2016	2	January	2016	48	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	295	180	475
13-03-2014	13	March	2014	48	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	20	45	120	1476	900	2376
13-03-2016	13	March	2016	48	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	22	45	120	1624	990	2614
23-05-2014	23	May	2014	47	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	24	45	120	1771	1080	2851
23-05-2016	23	May	2016	47	Adults (35-64)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	23	45	120	1697	1035	2732
19-07-2013	19	July	2013	32	Young Adults (25-34)	F	United States	Oregon	Accessories	Bike Racks	Hitch Rack - 4-Bike	9	45	120	524	405	929
19-07-2015	19	July	2015	32	Young Adults (25-34)	F	United States	Oregon	Accessories	Bike Racks	Hitch Rack - 4-Bike	9	45	120	524	405	929
11-11-2013	11	Novembe	2013	34	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	25	45	120	1345	1125	2970
11-11-2015	11	Novembe	2015	34	Young Adults (25-34)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	23	45	120	1697	1035	2732
13-07-2013	13	July	2013	43	Adults (35-64)	F	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	10	45	120	726	450	1176
13-07-2015	13	July	2015	43	Adults (35-64)	F	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	10	45	120	726	450	1176
31-03-2014	31	March	2014	43	Adults (35-64)	F	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	11	45	120	799	495	1294
31-03-2016	31	March	2016	43	Adults (35-64)	F	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	11	45	120	799	495	1294
28-08-2013	28	August	2013	61	Adults (35-64)	M	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	581	360	941
28-08-2015	28	August	2015	61	Adults (35-64)	M	United States	California	Accessories	Bike Racks	Hitch Rack - 4-Bike	10	45	120	726	450	1176
27-08-2013	27	August	2013	57	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	2	45	120	148	90	238
27-08-2015	27	August	2015	57	Adults (35-64)	M	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	1	45	120	74	45	119
17-11-2013	17	Novembe	2013	29	Young Adults (25-34)	F	Canada	British Columbia	Accessories	Bike Racks	Hitch Rack - 4-Bike	12	45	120	886	540	1426

Figure 1: Bikesales Dataset

1.1.3. Reasons for Choosing Dataset

I have so many reasons for choosing this dataset,

- This dataset has enough values for analysis and making a good dashboard
- The size of the dataset is also useful for making powerful business questions, data preprocessing, data modelling, making relationships, and using Dax and M language
- The dataset having values from 2011 to 2016, will help to make a good years analysis
- This data set is almost clean and the null values are very less compared to another dataset

2. Findings Based On Analysis and Evaluation

2.1. Business Intelligence Requirements/Questions

This project has mainly focused on bike sale data in Europe and America. Following are the business questions that needed answers through this project and data visualisations.

1) Which year had the greatest overall revenue increase?
2) In which month of the year make the highest revenue?
3) In which year the profit margin is the largest?
4) Which state has the highest average profit per capita?
5) What are the top countries by total revenue?
6) Which state have the highest average revenue in 2016?
7) What are the top age groups by total sales?
8) Which customer gender makes the highest revenue in total?
9) Which product subcategory influences profit to increase?
10) Which year does the most bike sales occur?

2.1. Power BI Visuals with Description

In this Power BI dashboard, there are 5 pages. The first page is the Home page and it contains buttons to navigate to the other four pages. The other four pages are Year Wise Sales, Country Wise Sales, Customer Wise Sales and Product Wise Sales.

Home Page

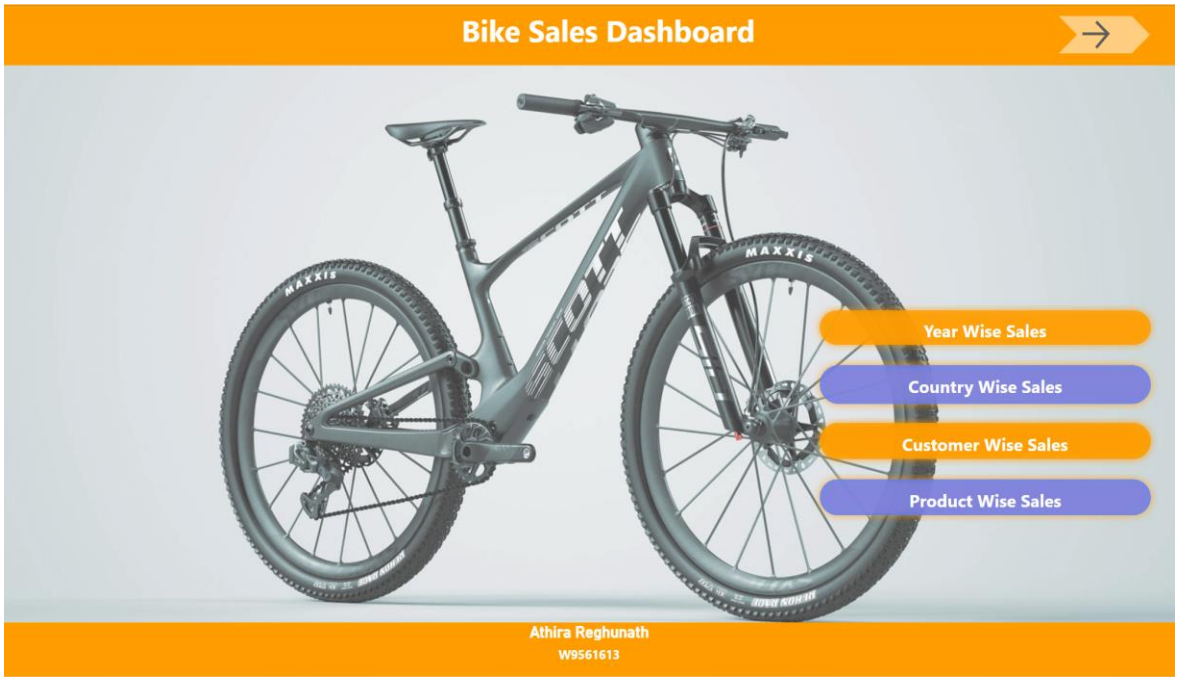


Figure 29: Home Page

The above image is the home page of the dashboard. This has no visualisations. Only included some buttons to navigate to the inner pages and an arrow button to go to the next page.

Year Wise Bike Sales Page

This page included all year-wise visualisations. Different visualisations were created to answer the business questions stated above.

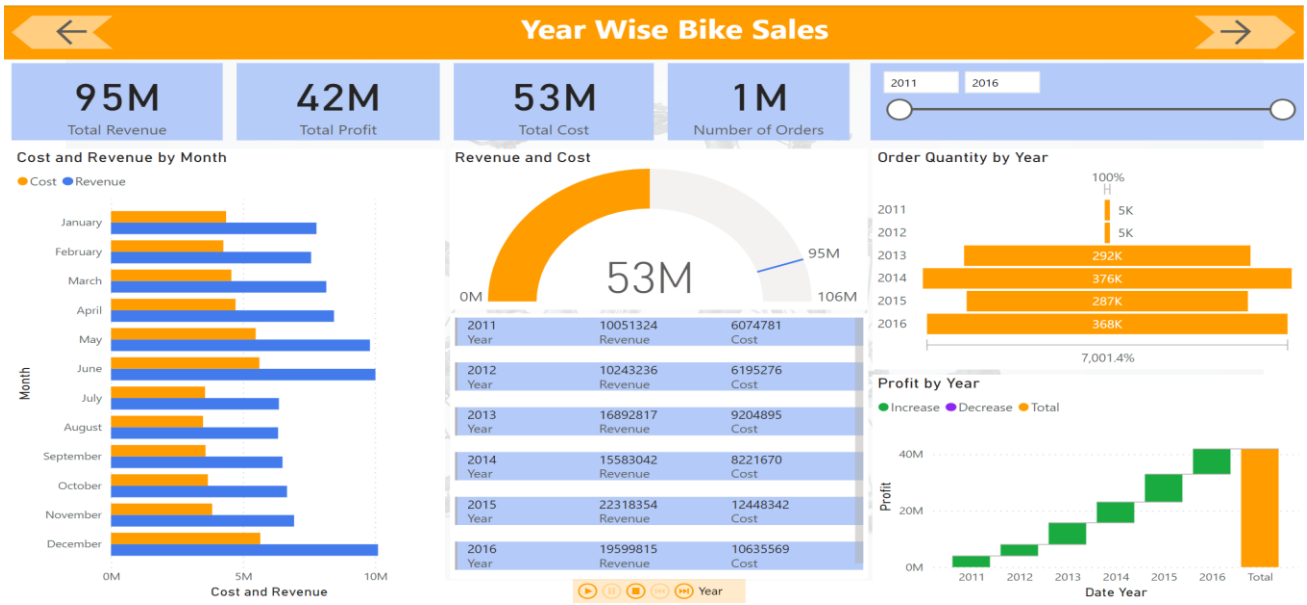


Figure 30: Year-Wise Sales Page

Here I added some measures for visualisation using DAX formulas

Total Revenue and Total Profit I Millions are found out for better understanding and visualisation nof the data

The DAX formula used here for calculating the Total Revenue is;

Total Revenue = SUM(RevenueData[Revenue])

The DAX formula used here for calculating the Total Revenue is;

Total Profit = SUM(RevenueData[Profit])

The DAX formula used here for calculating the Total Cost is;

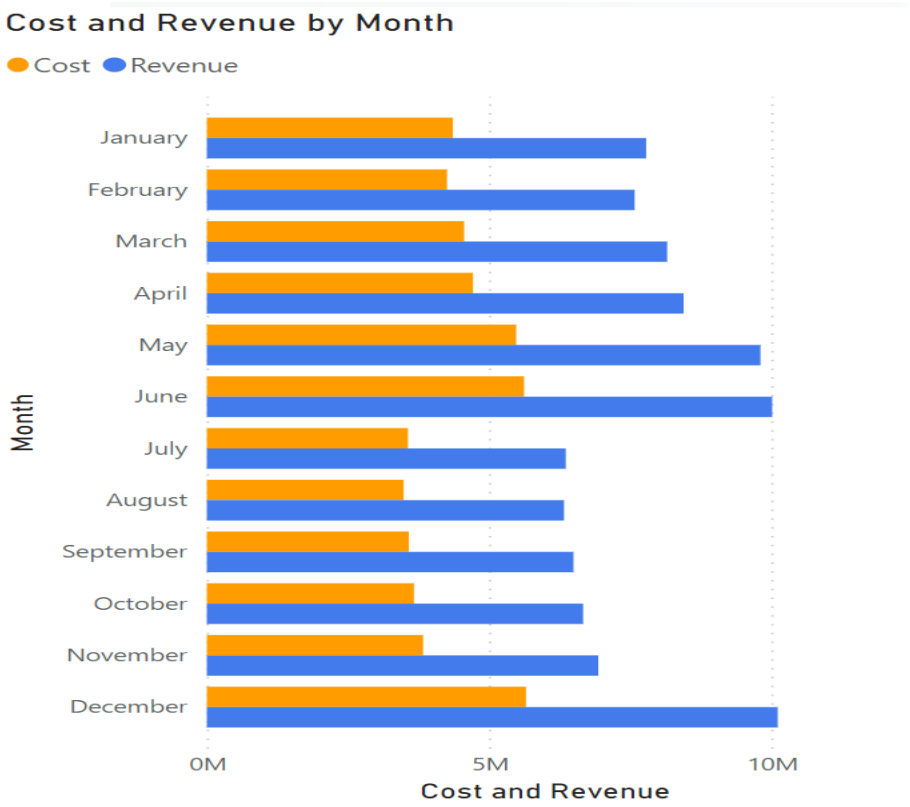
Total Cost = SUM(RevenueData[Cost])

The DAX formula used here for calculating the Total Number of Orders is;

Number of Orders = SUM(Order_Quantity)

1) In which month of the year make the highest revenue?

To answer this question here added a Clustered Bar Chart the x-axis of this chart has Cost and Revenue and the y-axis is Month. From this chart, we can clearly understand December is the month which makes the highest revenue and June is in the second position.

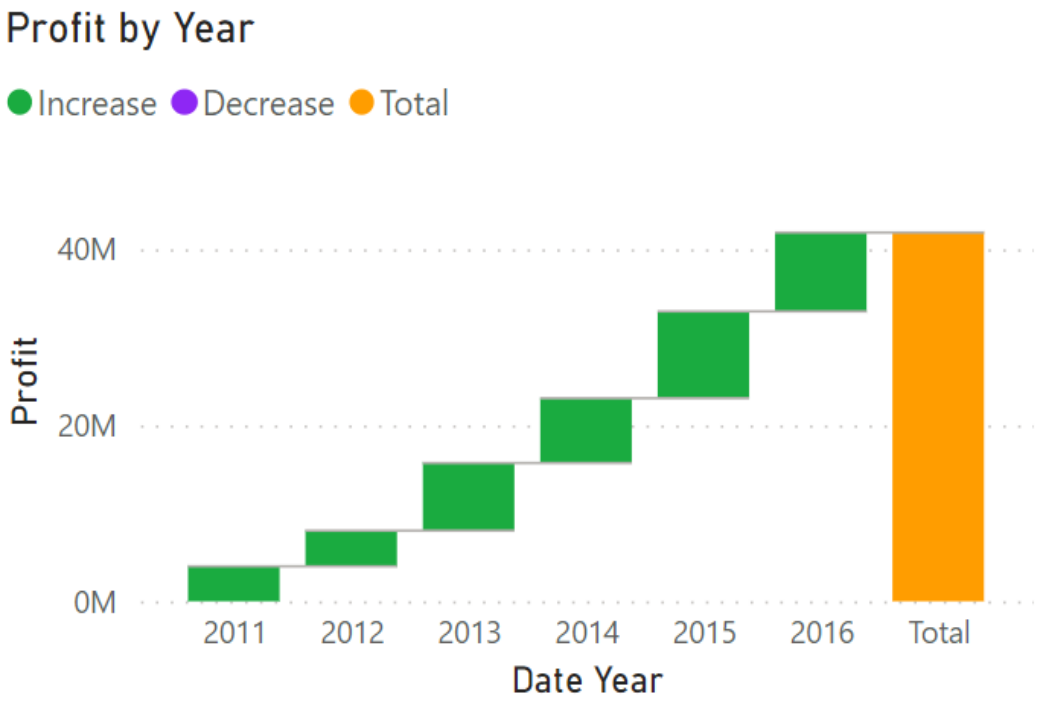


The second business question getting answered in this dashboard is,

1) **In which year the profit margin is the largest?**

A Waterfall Chart is used to answer this question. Here the x–axis is the years from 2011 to 2016 and the y–axis shows the Profit Data. From this graph, we can easily understand the largest profit margin is in the year 2015 and the very next place is for 2016. Here the profit is only increasing year by year so the profit is in the upward direction.

The first business question which getting answered on this page is,



The third business question is,

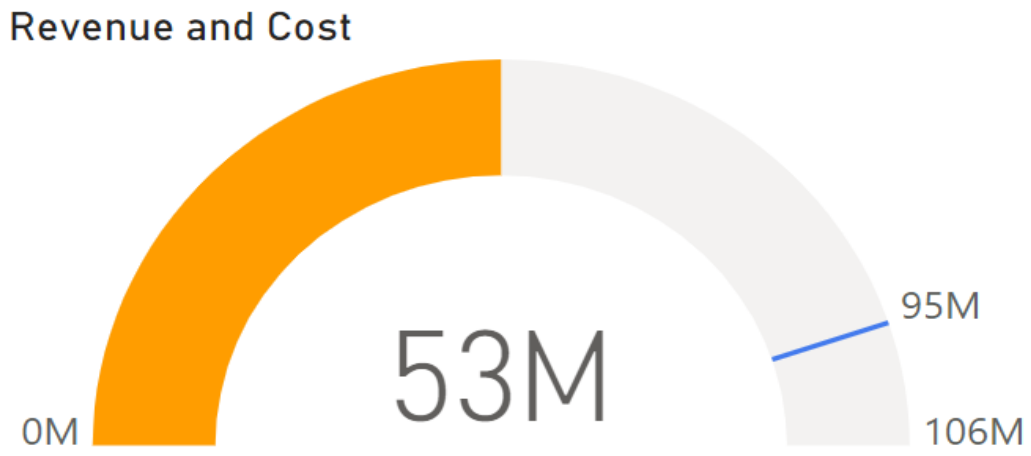
2) **Which year had the greatest overall revenue increase?**

A Multi Row Card is used to visualise the answer to this question. The data in the multi raw card are Year, Revenue and Cost. From this card, we can understand that the year 2015 have the highest revenue increase

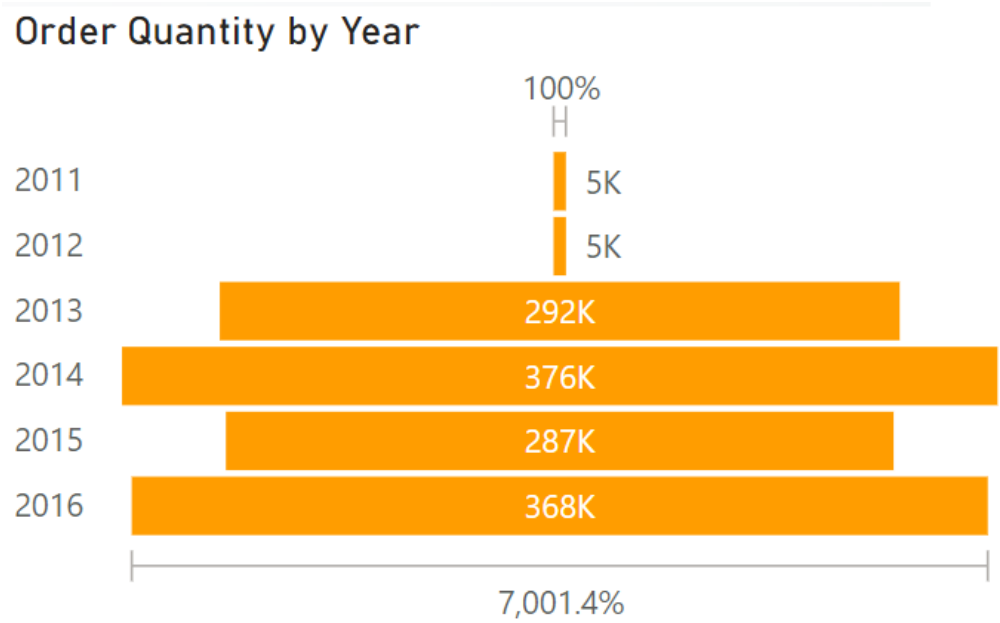
2011 Year	10051324 Revenue	6074781 Cost
2012 Year	10243236 Revenue	6195276 Cost
2013 Year	16892817 Revenue	9204895 Cost
2014 Year	15583042 Revenue	8221670 Cost
2015 Year	22318354 Revenue	12448342 Cost
2016 Year	19599815 Revenue	10635569 Cost

Other than these graphs some other supporting graphs are also included,

The Gauge graph is used to show the progress towards a particular goal. Here the graph is showing the progress of cost towards the revenue



The other graph used here is a Funnel Graph, It shows the Order Quantity or sales by Year. From this graph, we can understand that the year 2014 have a greater number of Orders placed.

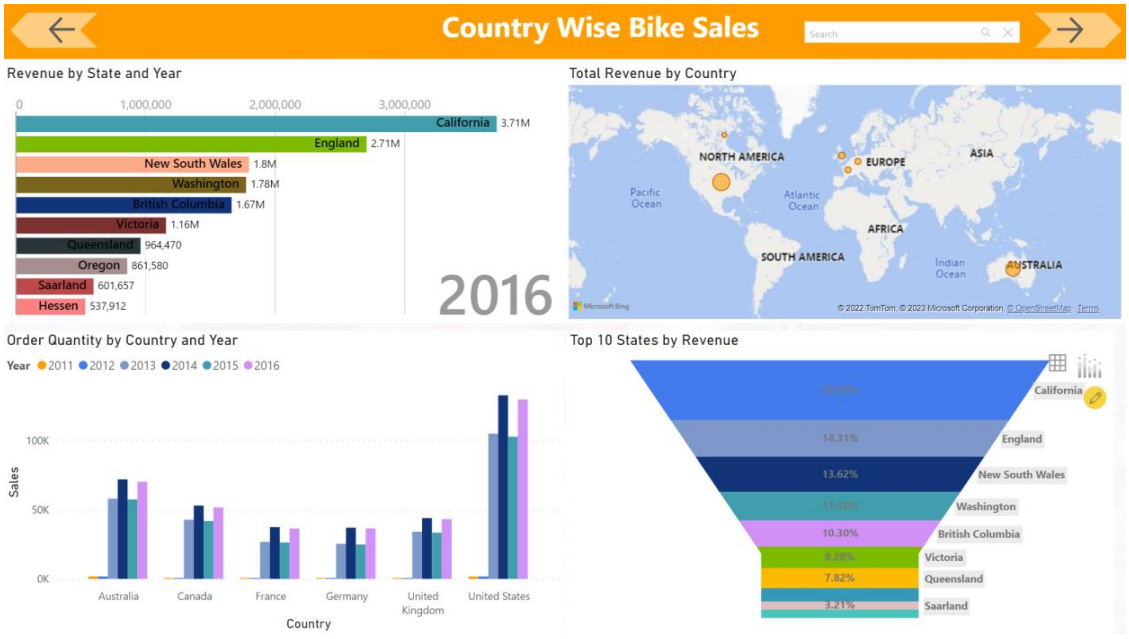


Other than graphs included an Year filter added for filtering the whole data by year

A play button is also added for getting an animation effect to the whole charts

Country Wise Bike Sales Page

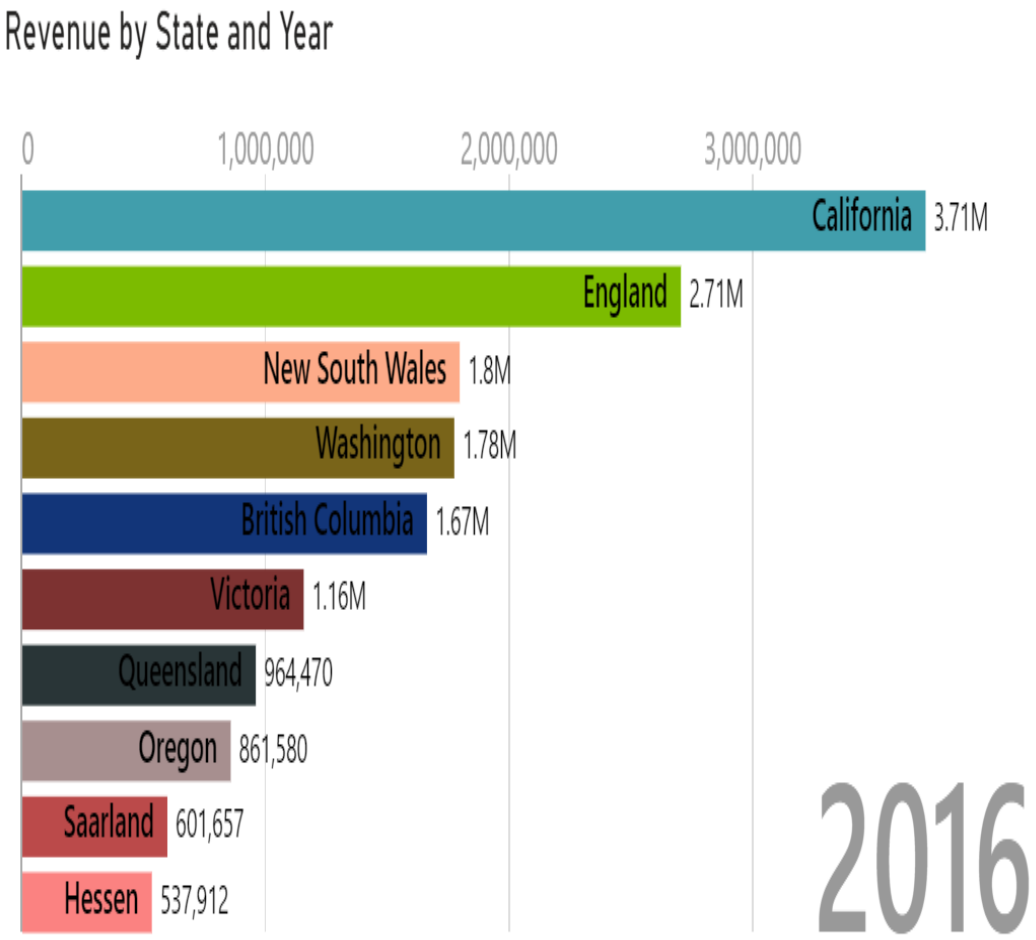
This page includes all country and state wise visualisations,



The first question seeks answers related to the country from this page is,

3) Which state have the highest average revenue in 2016?

Here used an Animated Bar Chart to visualise the answer to this question. The chart represents Revenue by State and Year. The x-axis shows the Revenue and the y-axis shows the states and the chart is changed according to the year change. Different states are denoted using different colours.



The next visualised business question is,

4) What are the top countries by total revenue?

For visualising this used a map and the bubble size is set according to the Revenue of that country.

Total Revenue by Country

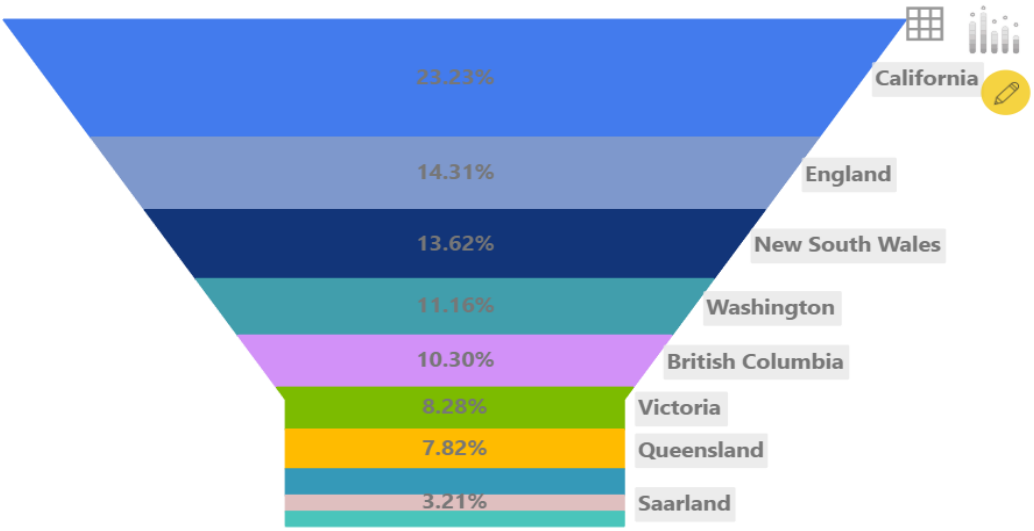
From this graph, we can easily understand that America is the Top country in the case of revenue and Australia is in the second position.

The visualised third business question is,

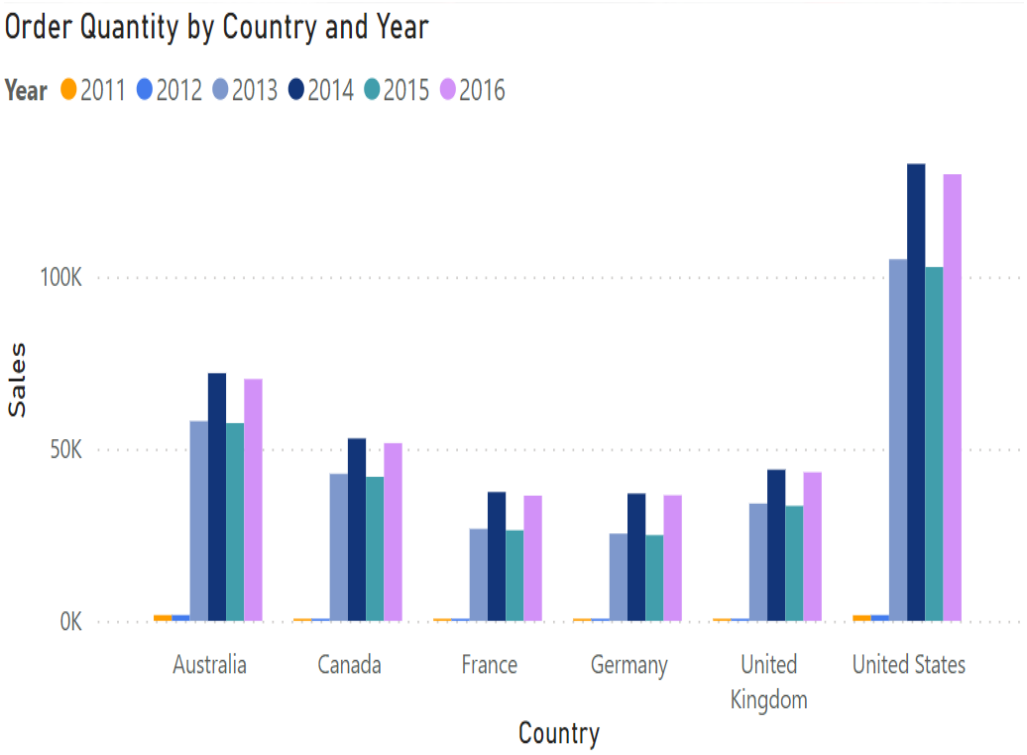
5) **Which state has the highest average profit per capita?**

This is visualised using a Funnel Chart. This chart shows the top 10 states by Revenue. The highest revenue-making state is shown at the top of the funnel. So we can clearly understand that California has the highest average revenue-making state.

Top 10 States by Revenue



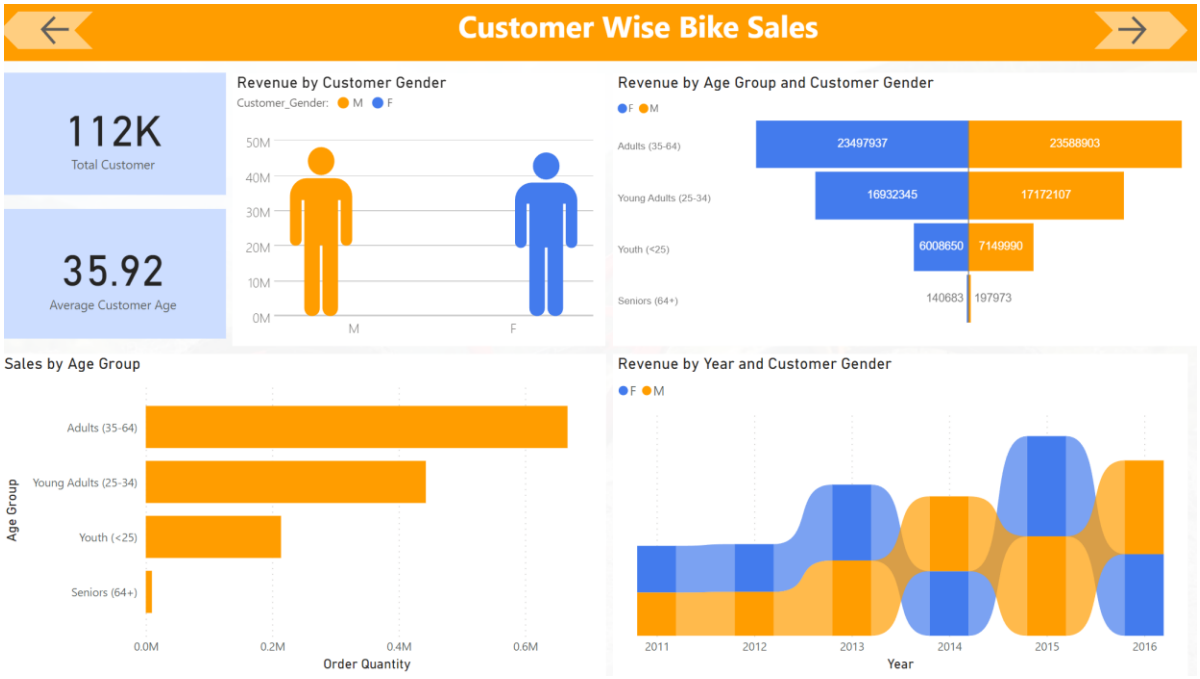
Along with these graphs, a Clustered Column Chart chart is included to visualise order quantity by country and Year. 6 years are showing in different colours. The x-axis showing country names and Y axis showing the Sales data. Wecan clearly say that, from 2013 to 2016 United Staes is in the first position in the case of Sales



A search filter is added to search for the Country and State

Customer Wise Bike Sales Page

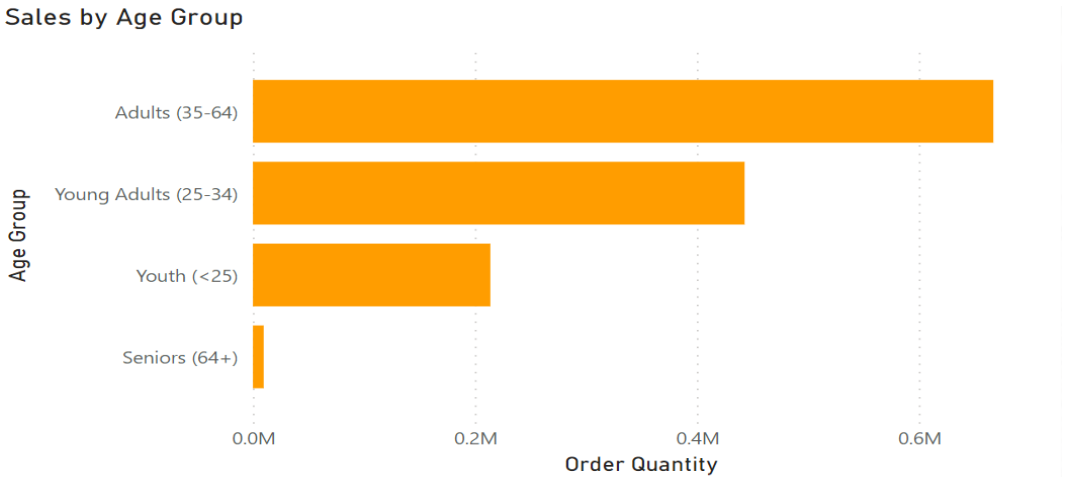
This page contains bike sales visualisation related to customer only



The first customer-wise sales business question is,

6) **What are the top age groups by total sales?**

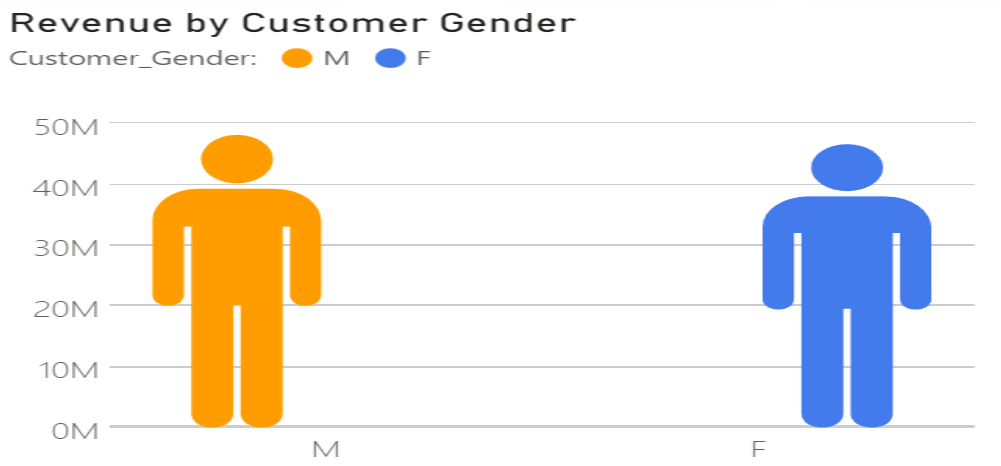
This question is visualised using a Stacked Bar Chart. The graph is showing Sales by age group data. The x-axis shows the order quantity and the y-axis shows the age group. We can clearly understand that the highest sales are done by the age group Adults(35-64).



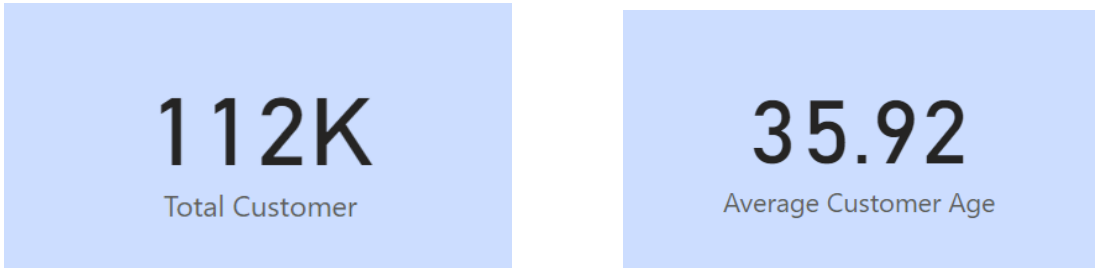
The second business question seeking an answer is,

7) **Which customer gender makes the highest revenue in total?**

For this visualisation, an Infographic Designer is used. This visualisation gives the Gender wise Revenue. The x-axis shows the customer gender and the y-axis shows the revenue dat in millions. From this graph we can easily understand the highest revenue making customer gender is Males.



Two measures are added using the DAX formula, for the data visualisation,



Total Customer and Average Customer Age are calculated using DAX

The DAX formula for calculating Total Customers is,

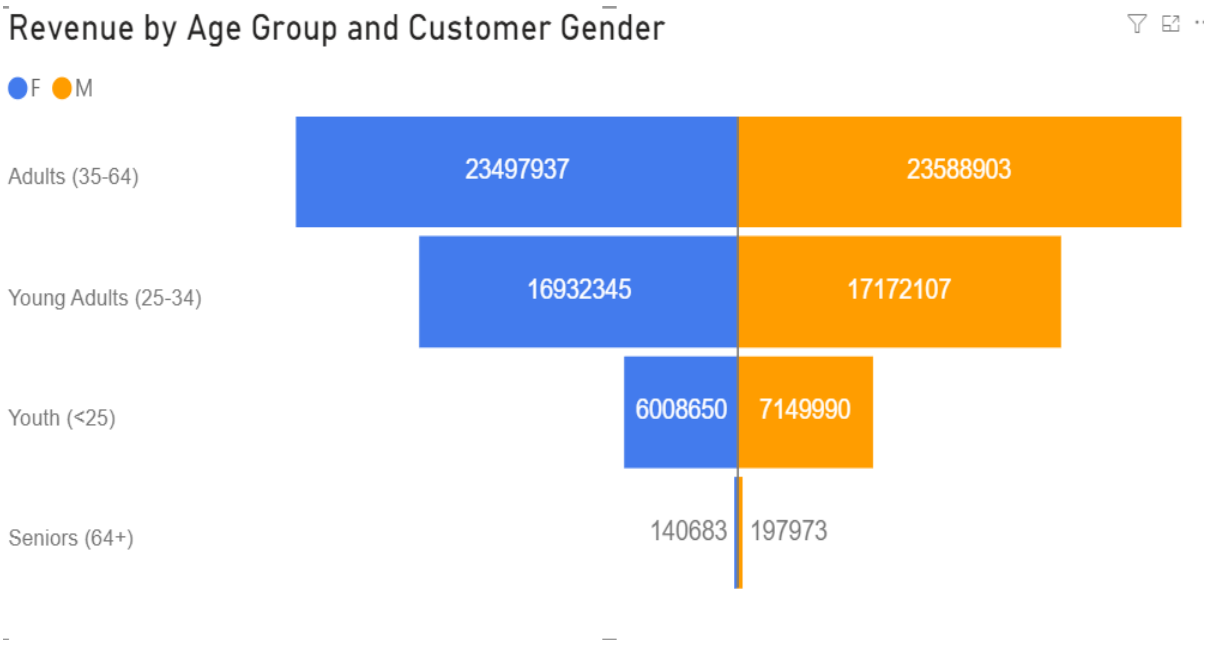
```
Total Customer = DISTINCTCOUNT(CustomerData[CustomerIndex])
```

The DAX formula for calculating Average Customer Age is,

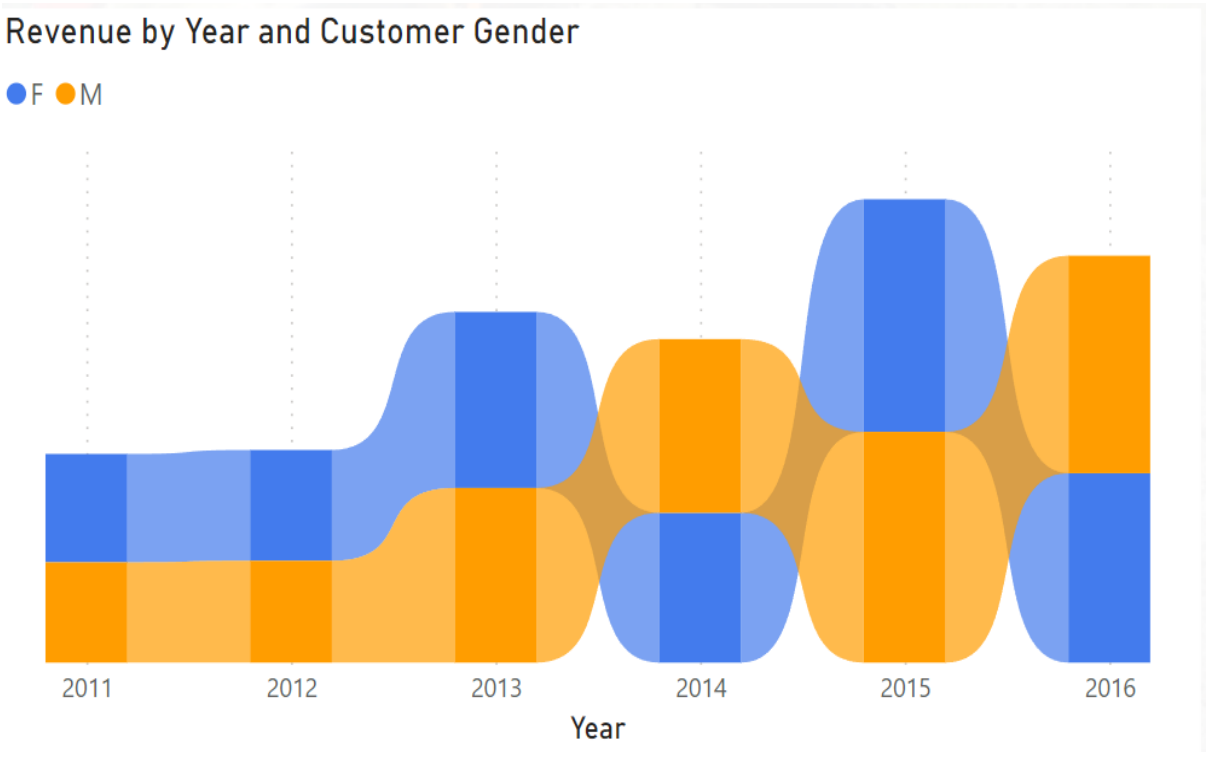
```
Average Customer Age = AVERAGE(CustomerData[Customer_Age])
```

The other two graphs are used in this visualisation,

One is a Tornado chart. This chart is showing revenue by age group by customer gender. Its simple and easily undestandable.

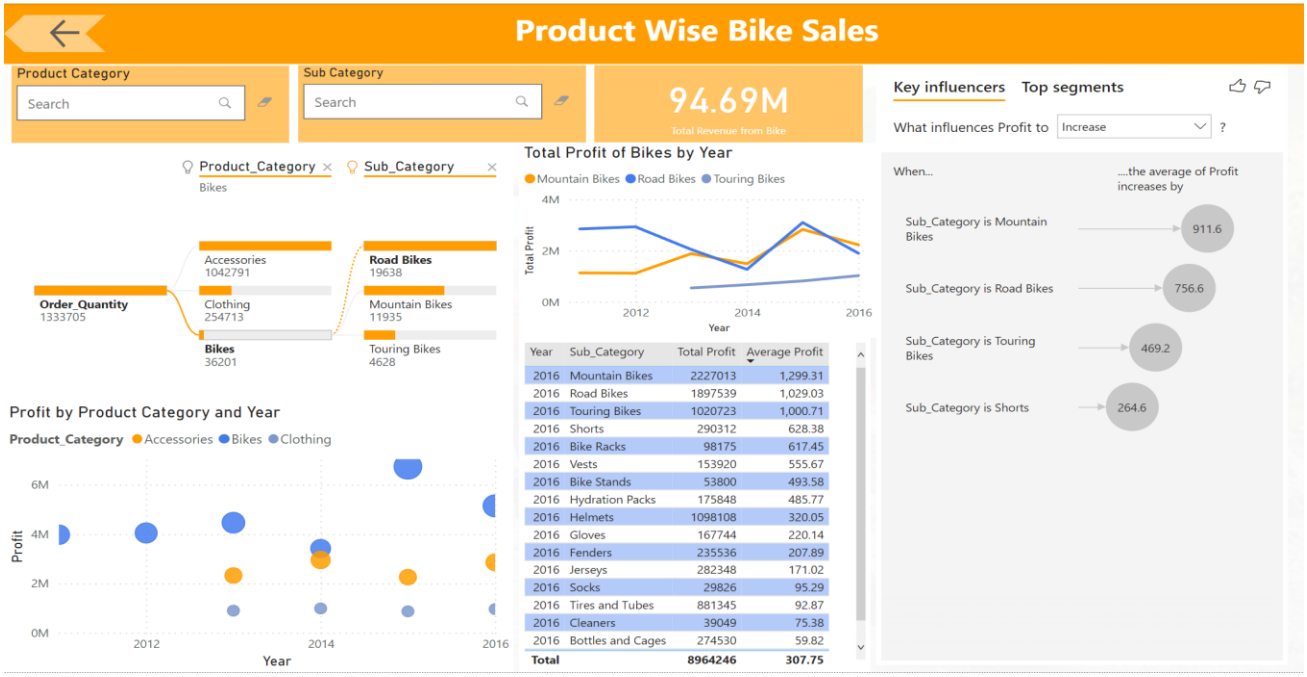


The other chart used here is the ribbon chart, This chart is showing the Revenue by Year and Customer Gender. Year is showing in the x-axis and y axis showing customer gender. From this graph we can easily understand each alternative year each gender showing the first position. In 2016 Males are in the first position.



Product Wise Bike Sales Page

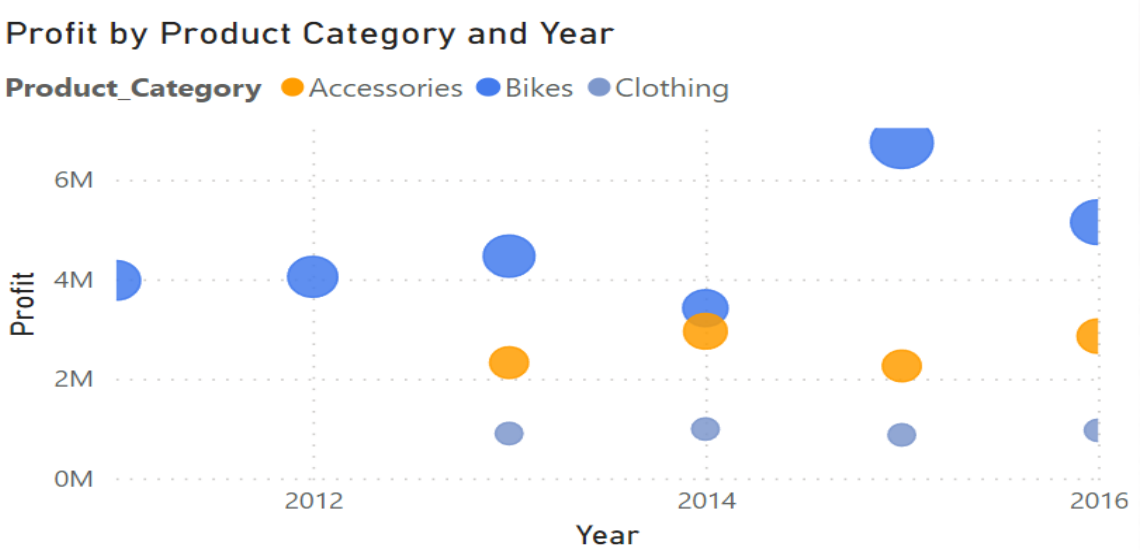
This page contains the product and product category-wise sales visualisation



The first business question related to the product is,

8) Which year does the most bike sales occur?

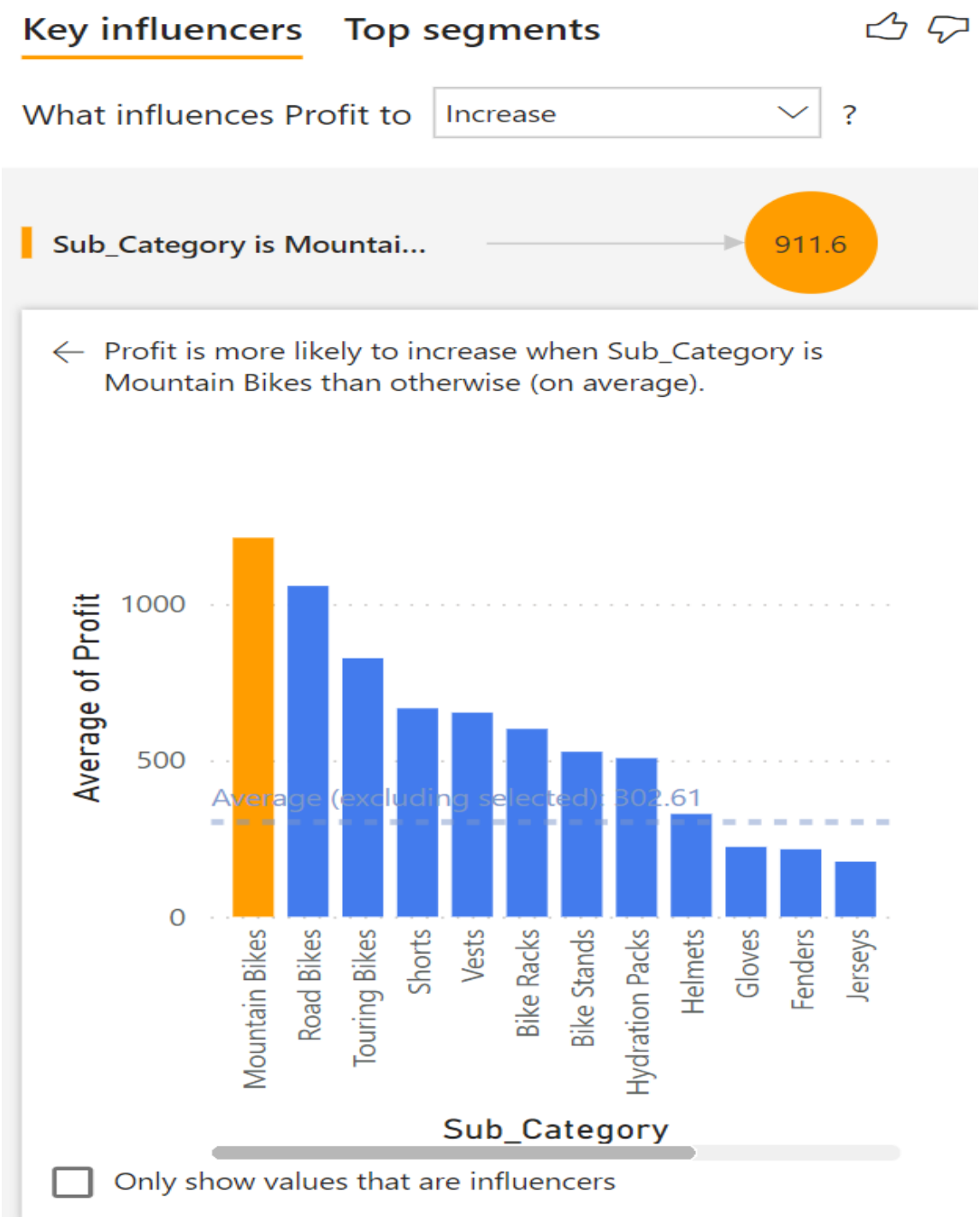
This is visualised by a scatter plot, Scatter plots display profit by category and year. Here the x-axis is a year and the y-axis is the profit in millions. Categories are given as legends. Bubble size is increased by the increase in the profit. Here we can understand that the most bike sales happened in 2015.



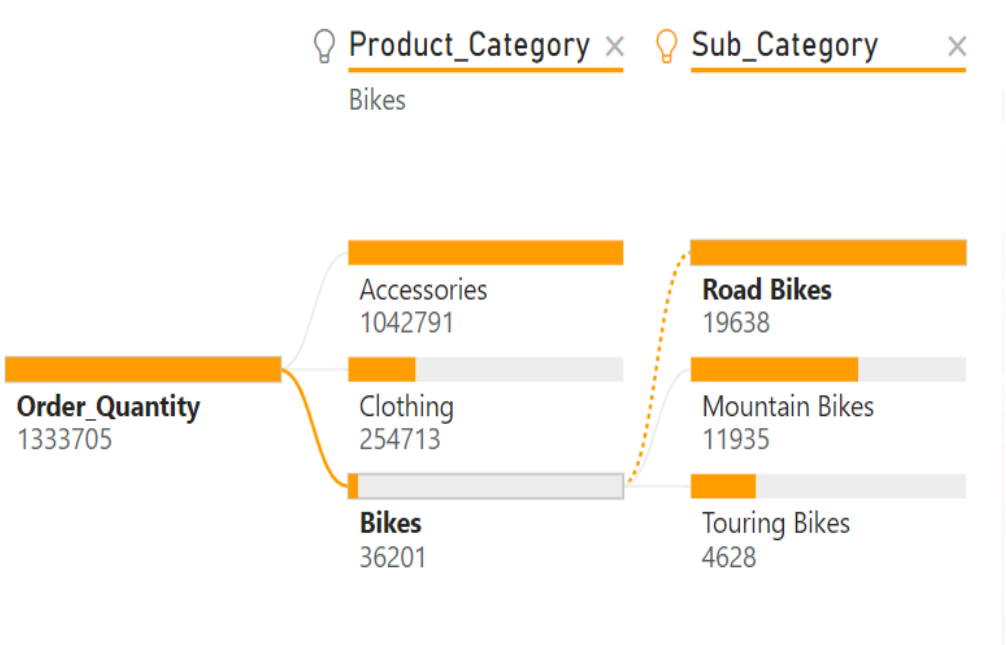
The second business question related to the product is,

9) Which product subcategory influences profit to increase?

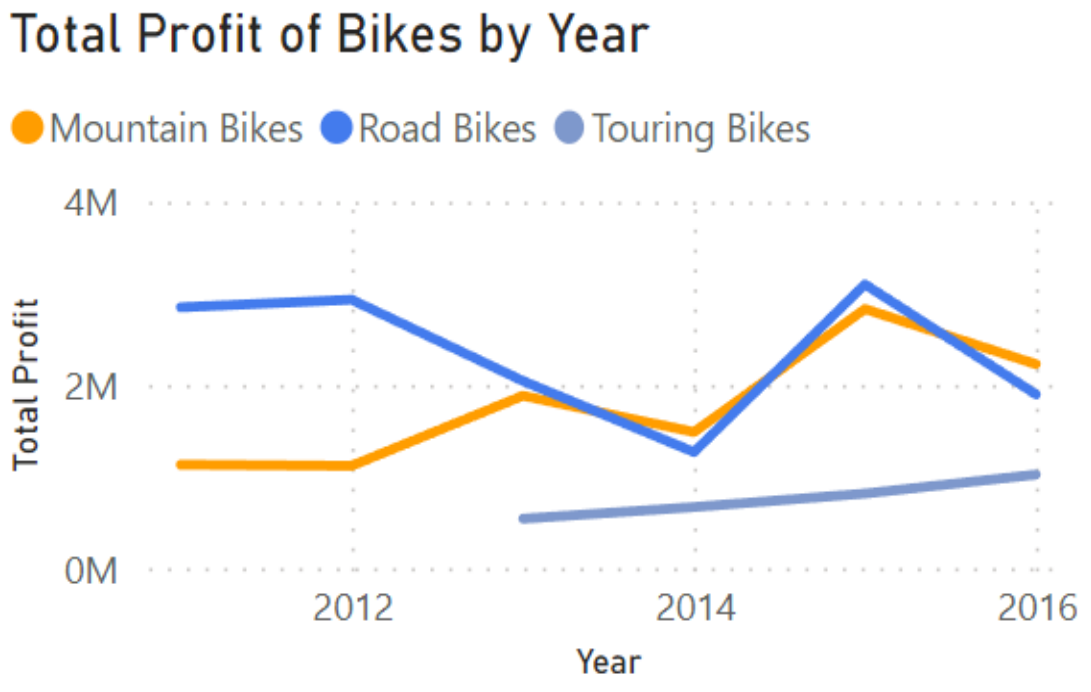
For visualising this key influencers are used. It gives lots of data about a dataset. Here wecan understand that profit is more likely to increase when sub category is mountain bike than otherwise.



This is visualised using a tree map. While clicking the order quantity product categories are expanding. Clicking on a product expands the sub category and all are shown with the sales. Revenue and profit are showing in the tooltip.



A line graph is used to display the total profit of bikes by year. The x-axis shows the year and the y-axis shows the total profit in million. Here we can see that mountain bikes got more profit in 2016.



Calculated total revenue frome bikes using the measure Total Revenue.

Here is a table added with Year, Subcategory, Total Profit and average profit. Total profit is already calculated. We can calculate the average profit measure using DAX formula.

Year	Sub_Category	Total Profit	Average Profit
2016	Mountain Bikes	2227013	1,299.31
2016	Road Bikes	1897539	1,029.03
2016	Touring Bikes	1020723	1,000.71
2016	Shorts	290312	628.38
2016	Bike Racks	98175	617.45
2016	Vests	153920	555.67
2016	Bike Stands	53800	493.58
2016	Hydration Packs	175848	485.77
2016	Helmets	1098108	320.05
2016	Gloves	167744	220.14
2016	Fenders	235536	207.89
2016	Jerseys	282348	171.02
2016	Socks	29826	95.29
2016	Tires and Tubes	881345	92.87
2016	Cleaners	39049	75.38
2016	Bottles and Cages	274530	59.82
Total		**8964246**	**307.75**

DAX formula to calculate the Average Profit is,

```
Average Profit = AVERAGE(RevenueData[Profit])
```

Two Text Filters are added to search by category or sub-category

Arrow buttons are used in all pages to navigate to net or previous pages

Previous Button

Next Button

2.2. Key Findings from the Chart

December had the highest Revenue and was 62.14% higher than August

Cost and Total Revenue are positively correlated with each other.

Revenue and Cost diverged the most when the Month was December

The United States had the highest Total Revenue and was 283.23% higher than Canada

United States accounted for 32.32% of Total Revenue.

Adults (35-64) accounted for 50.02% of Order_Quantity.

Between 2011 and 2016, Mountain Bikes had the largest increase in Total Profit (97.24%) while Road Bikes had the largest decrease (33.36%).

3. Conclusion

Here from this data visualisation, we can conclude that 2015 is the year which has the highest revenue and profit. December is the month having the highest revenue and profit. The United States is the country and California is the state which has the highest revenue. From the customer's analysis Adults are the major source of profit

BIKE SALES ANALYSIS

NAME: ATHIRA REGHUNATH

STUDENT ID: W9561613

SECTION 2- ICA – APPENDIX: BI DESIGN

1. Dataset Description and BI requirements

1.1. Introduction

The major problem faced in any sales sector is the lack of analyses of the previous sales styles. That means which country, which kind of people or which product is mostly sold or which gave most profit etc. Here in the Bike sales industry also need to know the same details from the previous records of data. If we visualise the whole data according to our needs it will be easy to understand and easy to explain to others also. Here I am going to make a dashboard related to Bike Sales dataset and this needs to go through these steps Data Preprocessing, Data Modelling, Creating Relationships, Data Visualisation etc.

1.2. Data Source and Description

1.2.1. Source

Here I am going to visualise the Bike Sales dataset, this data contains how many bike sales are in Europe and America and how profitable for different age groups, different products and for different countries. We can get an idea about how prices are likely to change in the future based on this analysis. The Bike Sales dataset is taken from Kaggle https://www.kaggle.com/sadiqshah/bike-sales-in-europe.This dataset contains 18 columns and 113036 rows.

2. Data Pre-Processing and Data Cleansing

2.1. Data Loading

The data pre-processing start only after loading the BikeSales.csv dataset to powerbi. For this first, we open the powerbi and it will open like the below figure.

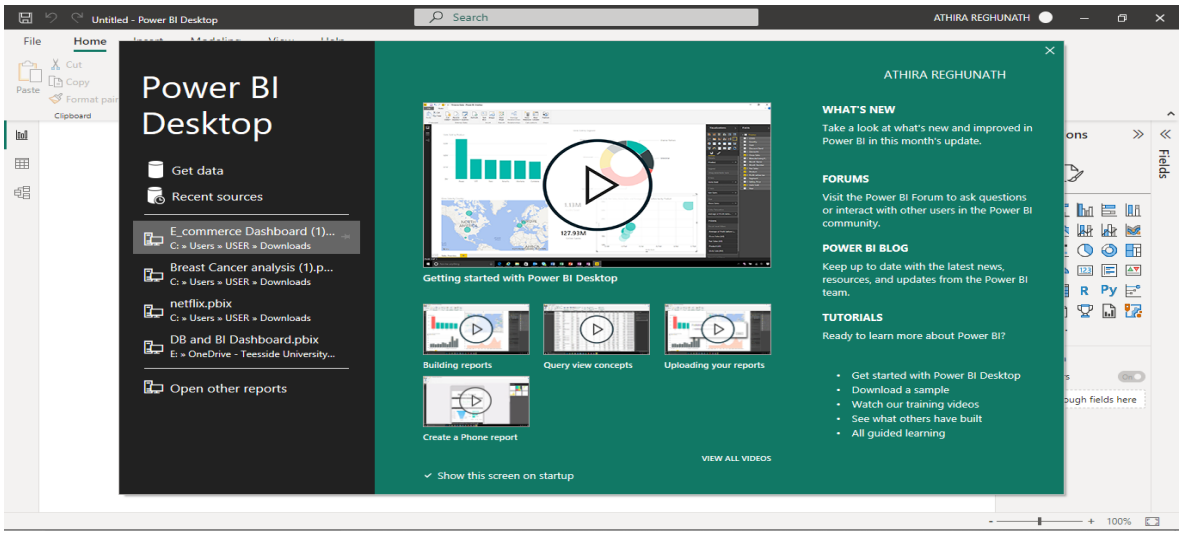


Figure 2: Power BI startup dialogue box

We can load the BikeSales dataset by clicking the Get data menu and then importing it. Otherwise, we can import it using an M language query.

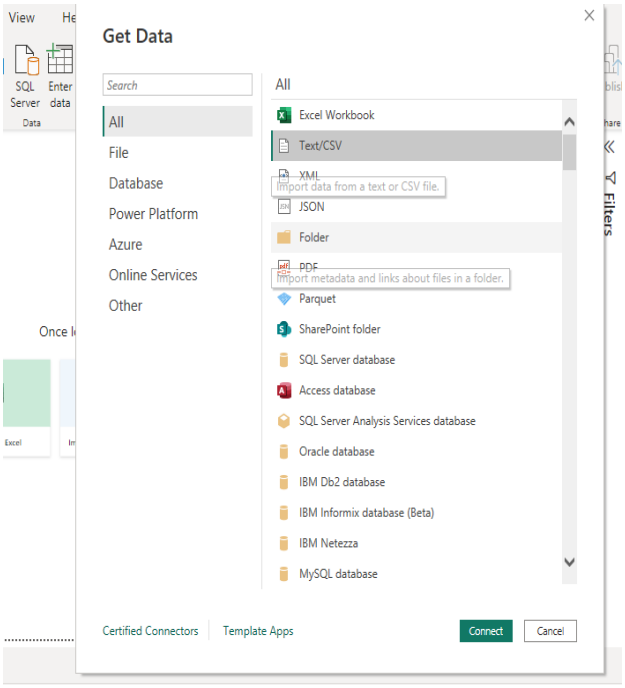


Figure 3: Select Test/CSV and press Correct button

Selecting Test/CSV and pressing the correct button. Then select the file from the system and open, below page is opened,

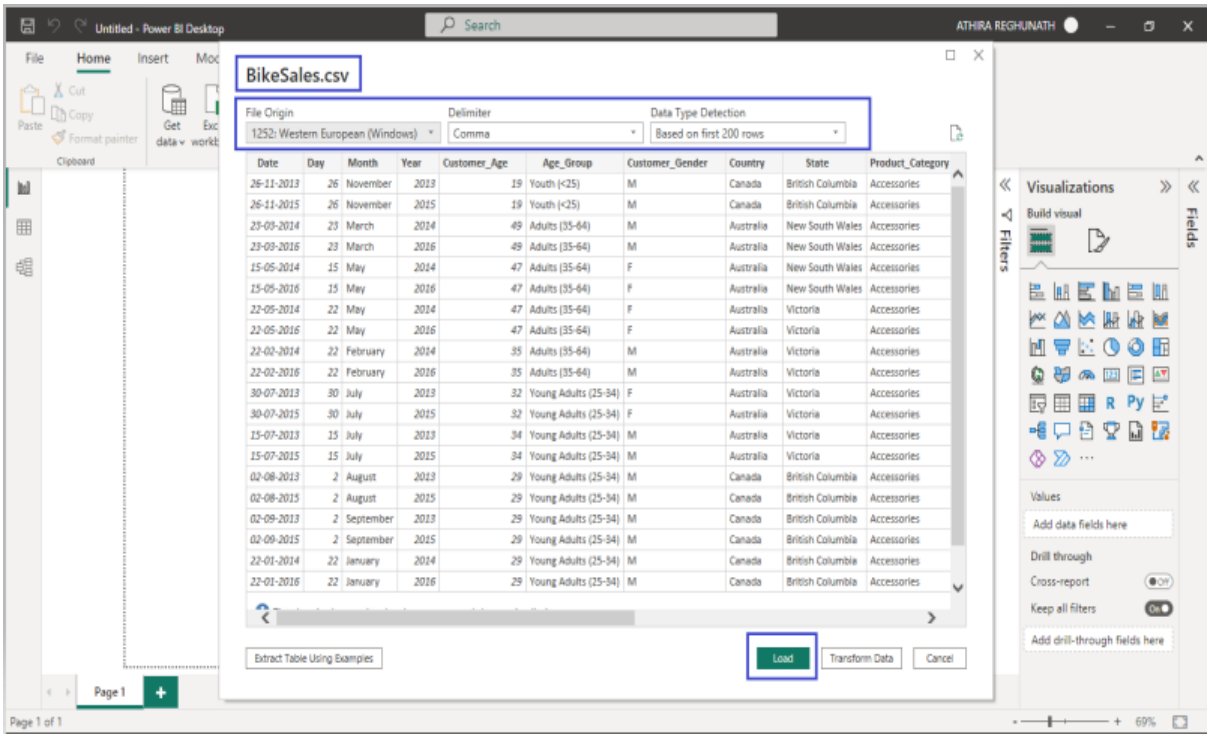


Figure 4: Click Load for loading the dataset to Power BI

Here 200 rows of BikeSales dataset are showing and we can press the load button on the bottom of the page and load the data.

When loading the data the power bi detected some errors. There was no error in the dataset but I added some null and different data type values in two rows for just showing how to manage errors.

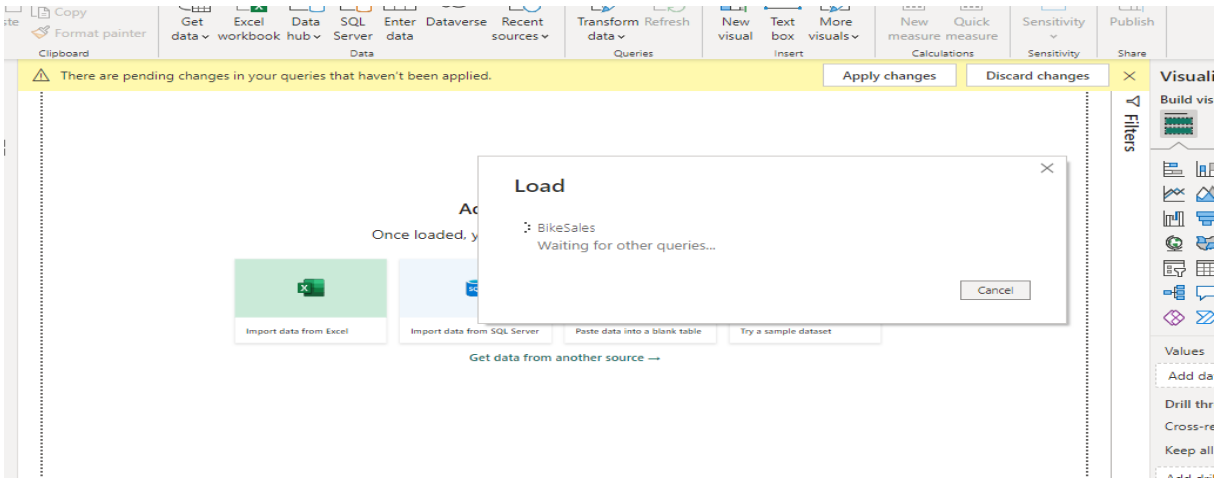


Figure 5: Error showing while loading

2.2 Data Cleansing

After loading the data we need to clean the data. So click on the Transform Data

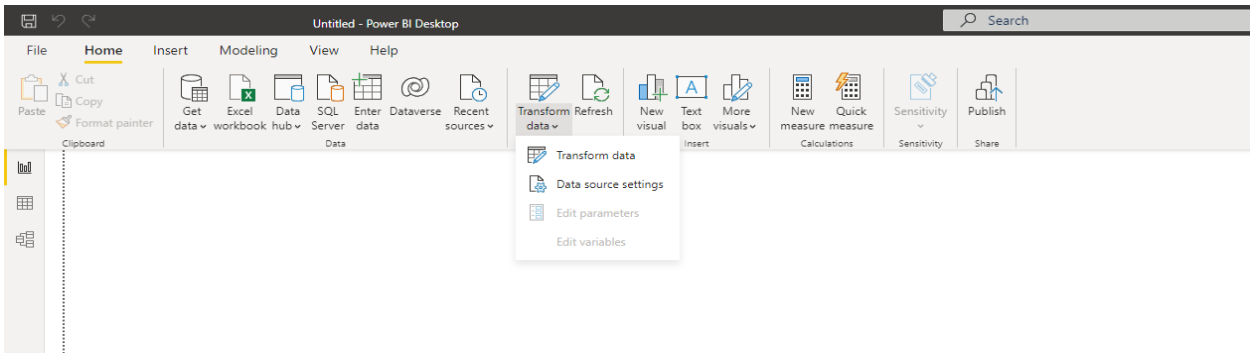


Figure 6: Transform Data

Removing Errors

Go to transform data and click on the errors it will show the errors

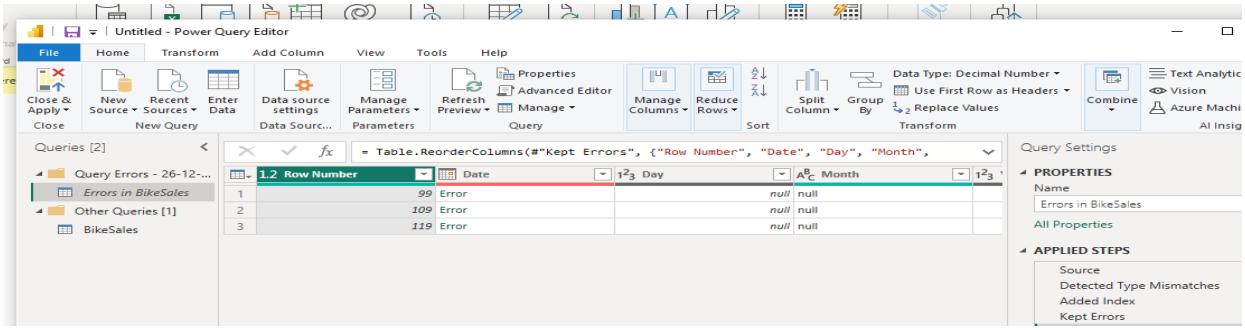


Figure 7: Error showing in transform data page

We can remove these errors by right-clicking and select Remove Errors menu

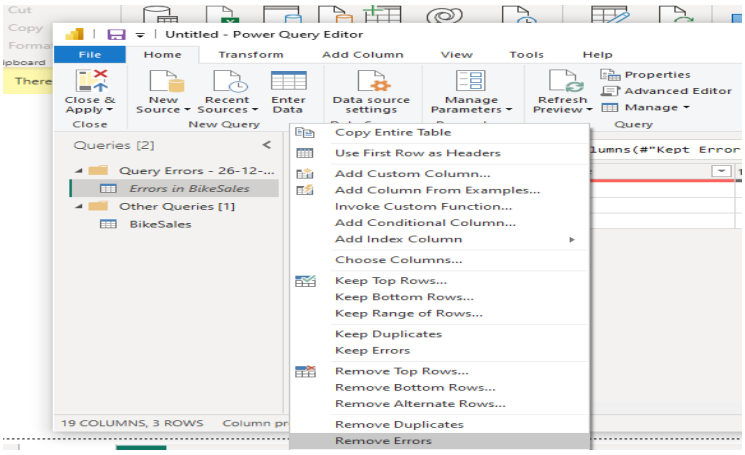


Figure 8: Remove errors menu

The BikeSales dataset in the transform data page is shown below

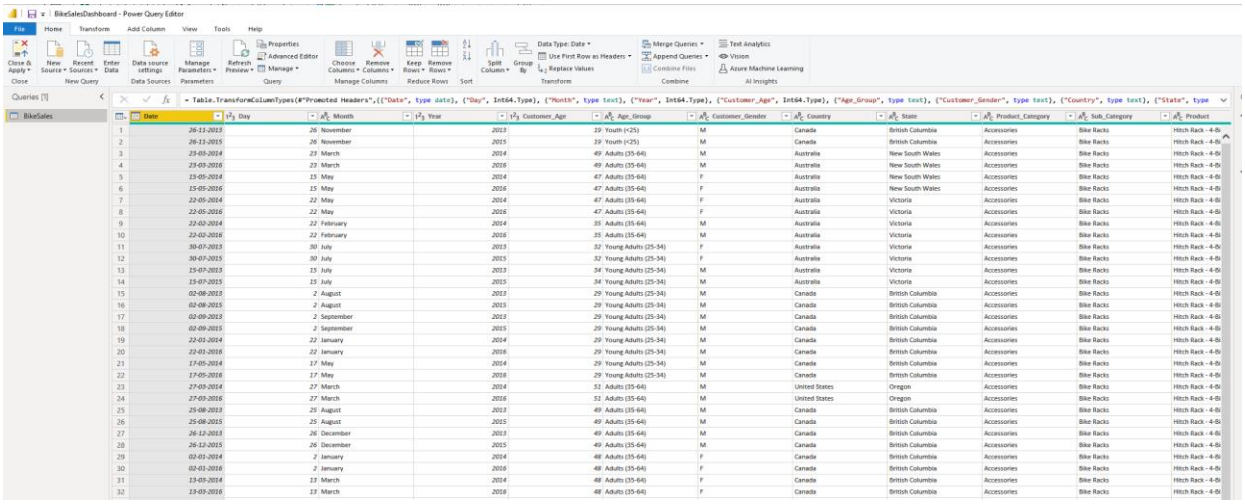


Figure 9: BikeSales dataset in the transform data page

Changing Data Type

Changed the data type of Date as Date and order_Quantity, Unit_Cost, Unit_Price, Cost, Profit and Revenue as whole numbers

M language for changing data type is,

= Table.TransformColumnTypes(#"Promoted Headers",{{"Date", type date}, {"Order_Quantity", Int64.Type}, {"Unit_Cost", Int64.Type}, {"Unit_Price", Int64.Type}, {"Profit", Int64.Type}, {"Cost", Int64.Type}, {"Revenue", Int64.Type}})

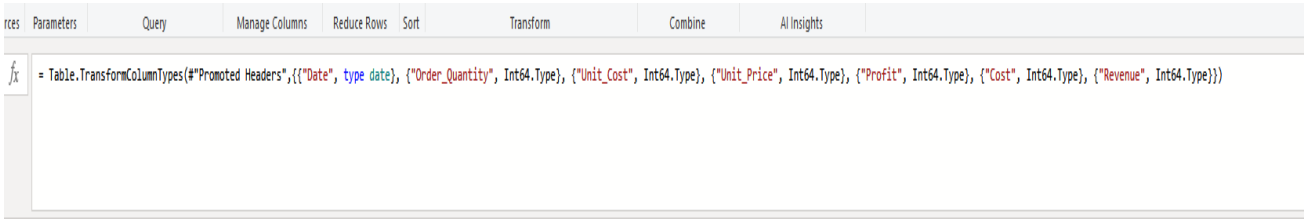


Figure 10: M language for changing the data type

Removing Columns

Here Day, Month and Year columns are given separately along with the Date column. We can derive day, month and year details from the date column. So need to delete columns Day, Month and Year. For this select the column we want to remove and select Remove Column

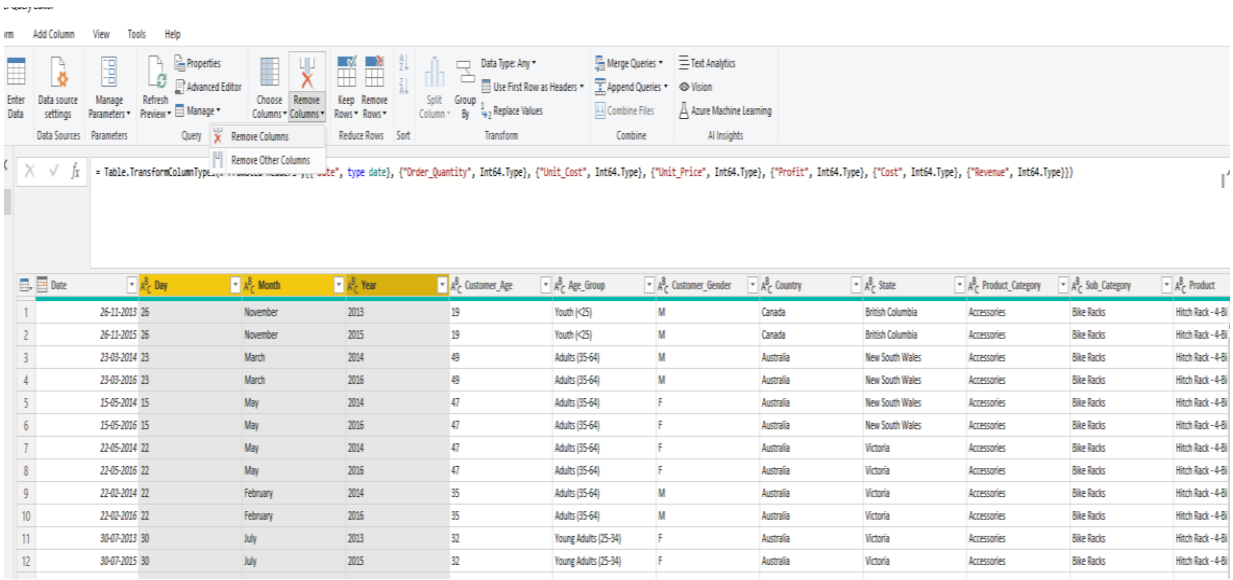


Figure 11: Remove Columns

M language for removing Day, Month and Year columns are

= Table.RemoveColumns(#"Changed Type",{"Day", "Month", "Year"})

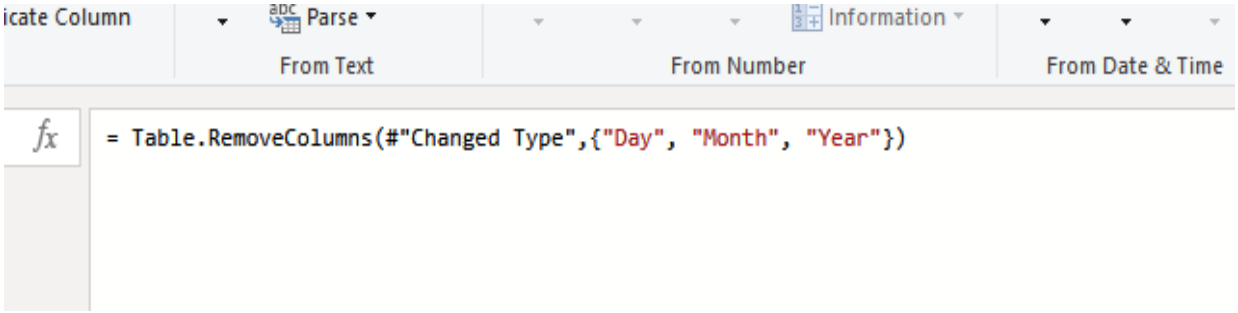


Figure 10: M language for removing columns

Removing Duplicates

For removing duplicate values, Select all columns then right-click on one of the columns and select Remove Duplicates

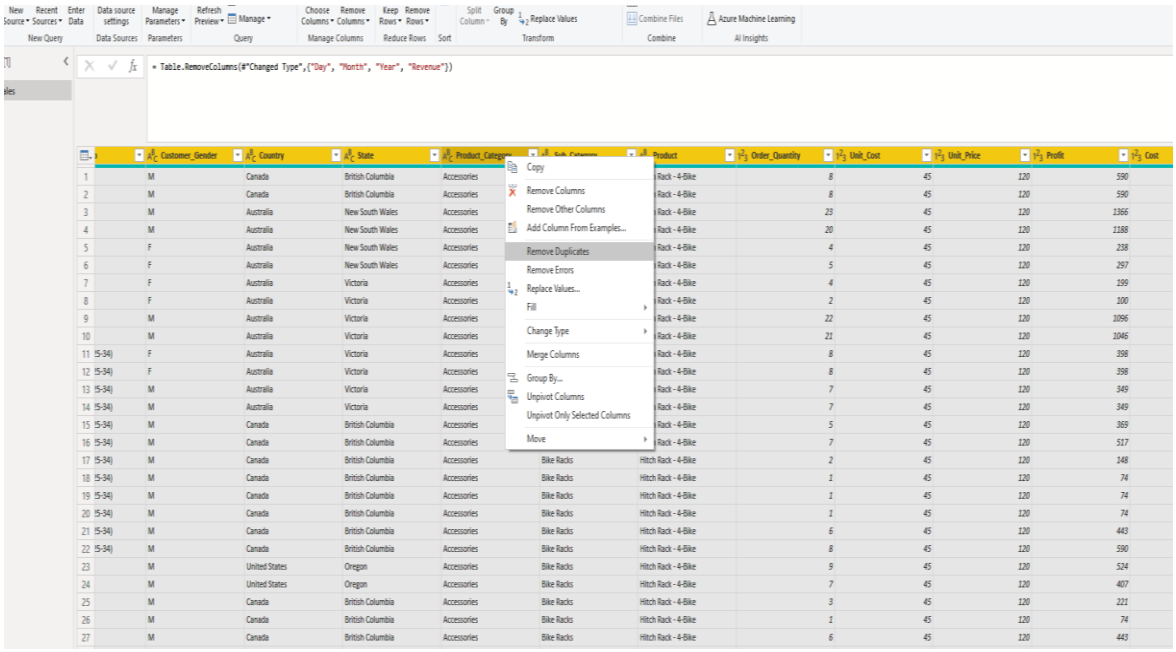


Figure 11: Selecting all columns for removing duplicates

Custom Column with Power Query Formula

We can see that the Profit and Revenue columns are showing some false data.

Here,

Cost = Order_Quantity * Unit_Cost

Revenue = Order_Quantity * Unit_Price

Profit = Revenue – Cost

We can check with the first raw data values

Cost = 8*45 =360

Revenue = 8*120 = 960 (In the table revenue is calculated as 950)

Profit = 960-360 = 600 (In the table Profit is calculated as 590)

So, need to create two new tables for Revenue and Profit

Product_Category	Sub_Category	Product	Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue
Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	590	360	950
Accessories	Bike Racks	Hitch Rack - 4-Bike	23	45	120	1366	1035	2401
Accessories	Bike Racks	Hitch Rack - 4-Bike	20	45	120	1188	900	2088
Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	238	180	418
Accessories	Bike Racks	Hitch Rack - 4-Bike	5	45	120	297	225	522
Accessories	Bike Racks	Hitch Rack - 4-Bike	4	45	120	199	180	379
Accessories	Bike Racks	Hitch Rack - 4-Bike	2	45	120	100	90	190
Accessories	Bike Racks	Hitch Rack - 4-Bike	22	45	120	1096	990	2086
Accessories	Bike Racks	Hitch Rack - 4-Bike	21	45	120	1046	945	1991
Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	398	360	758
Accessories	Bike Racks	Hitch Rack - 4-Bike	8	45	120	398	360	758
Accessories	Bike Racks	Hitch Rack - 4-Bike	7	45	120	349	315	664

Figure 12: Incorrect values for Revenue and Profit

Select Custom Column from Add Column

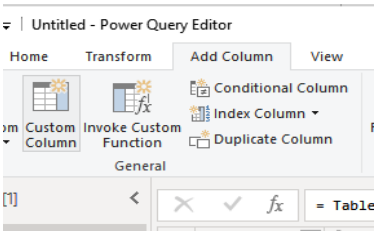


Figure 12: Selecting Custom Column

In the custom column dialogue box give the new column name as Revenue1 and the column formula as [Order_Quantity] * [Unit_Price] and click the OK button

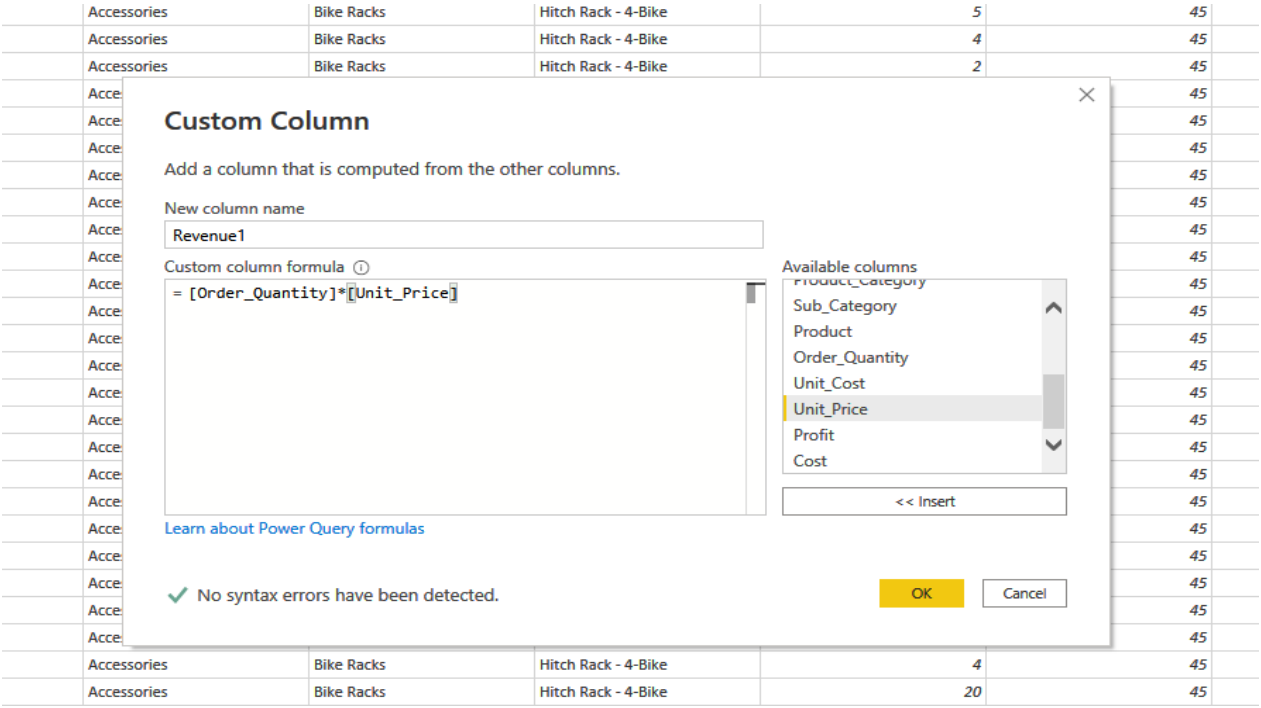


Figure 13: Figure 15: Power query formula to create column Profit1

M language for adding the Revenue1 column with the new formula is,

= Table.AddColumn(#"Removed Duplicates", "Revenue1", each [Order_Quantity][Unit_Price])*

Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue	Revenue1
8	45	120	590	360	950	960
8	45	120	590	360	950	960
23	45	120	1366	1035	2401	2760
20	45	120	1188	900	2088	2400
4	45	120	238	180	418	480
5	45	120	297	225	522	600
4	45	120	199	180	379	480
2	45	120	100	90	190	240
22	45	120	1096	990	2086	2640
21	45	120	1046	945	1991	2520
8	45	120	398	360	758	960

Figure 14: New column Revenue1 is added to the table

The same steps are repeated with the Profit column also. We need to take Custom Column from Add column. In the custom column dialogue box give the new column name as Profit1 and the column formula as [Revenue1] - [Cost] and click the OK button

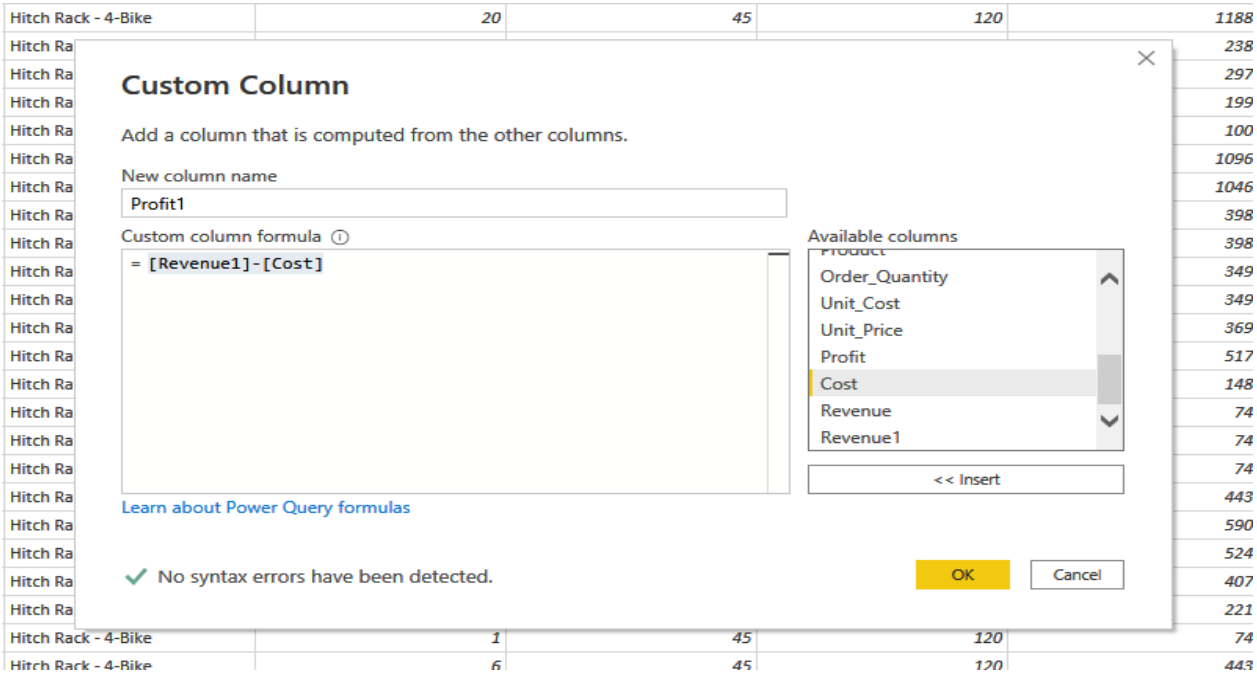


Figure 15: Power query formula to create column Profit1

M language for adding the Profit1 column with the new formula is,

= Table.AddColumn(#"Added Custom", "Profit1", each [Revenue1]-[Cost])

	Unit_Price	Profit	Cost	Revenue	Revenue1	Profit1
45	120	590	360	950	960	600
45	120	590	360	950	960	600
45	120	1366	1035	2401	2760	1725
45	120	1188	900	2088	2400	1500
45	120	238	180	418	480	300
45	120	297	225	522	600	375
45	120	199	180	379	480	300
45	120	100	90	190	240	150
45	120	1096	990	2086	2640	1650
45	120	1046	945	1991	2520	1575
45	120	398	360	758	960	600
45	120	398	360	758	960	600
45	120	349	315	664	840	525
45	120	349	315	664	840	525
45	120	369	225	594	600	375

Figure 16: New column Profit1 is added to the table

We added Revenue1 and Profit1 columns instead of Revenue and profit columns. Now going to delete the Revenue and Profit columns to avoid data redundancy. For this, select the Revenue and Profit columns together and remove columns by selecting the Remove Column from the menu that appeared after the right click.

M language for removing Revenue and Profit columns is,

= Table.RemoveColumns(#"Added Custom1",{"Profit", "Revenue"})

Now change the name of Revenue1 and Profit1 to Revenue and Profit. For this right click on the column heading and select Rename and change the names.

M language for renaming Revenue1 and Profit1 columns to Revenue and Profit is,

= Table.RenameColumns(#"Removed Columns1",{{"Revenue1", "Revenue"}, {"Profit1", "Profit"}})

There are no values for setting a primary key so adding an Index Column from Add Column and set value from 1.

M language for adding an index column is,

= Table.AddIndexColumn(#"Renamed Columns", "Index", 1, 1, Int64.Type)

Then drag the column to the beginning of the table and the M language for this is,

= Table.ReorderColumns(#"Added Index",{"Index", "Date", "Customer_Age", "Age_Group", "Customer_Gender", "Country", "State", "Product_Category", "Sub_Category", "Product", "Order_Quantity", "Unit_Cost", "Unit_Price", "Cost", "Revenue", "Profit"})

Then rename it as SalesIndex, M language for renaming is,

= Table.RenameColumns(#"Reordered Columns",{{"Index", "SalesIndex"}})

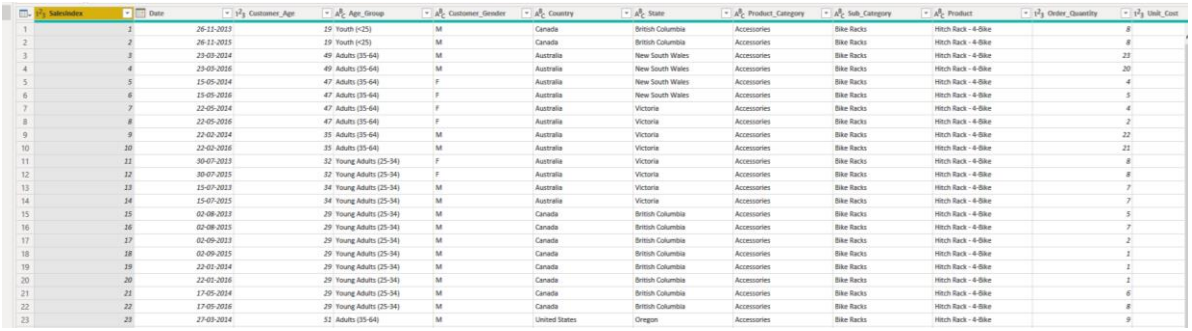


Figure 17: BikeSales table after adding the SalesIndex column

3. Data Modelling

After the data cleaning the current model is shown below,

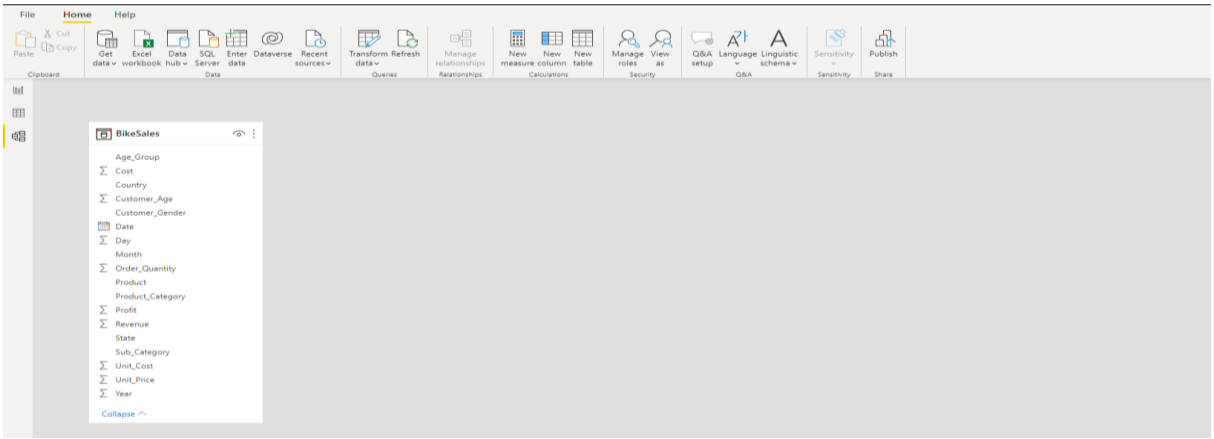


Figure 18: Data model after cleaning

Now we have only a big table, this table is called Fact Table. We need to create small tables from this big one to reduce complexity, avoid data duplication and for ease of

analysing and creating visuals. Those newly created small tables are called Dimension Tables.

3.1. Creating Dimension Tables

First going to create the dimension table RevenueData. For this, take the power query editor and by right click select duplicate. Now the BikeSales table is duplicated and gets a BikeSales2 Table.

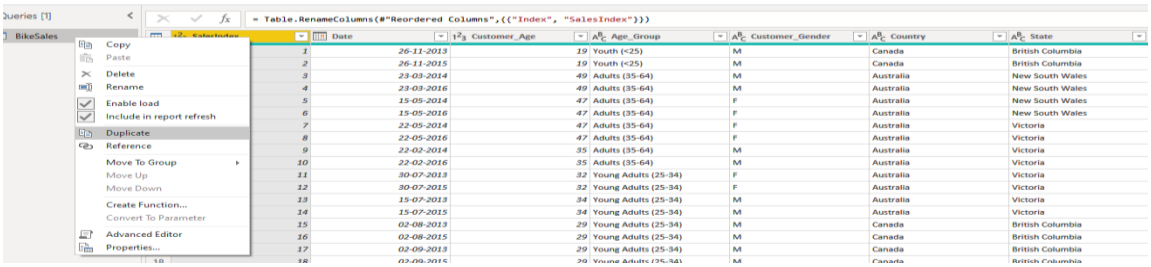


Figure 19: Duplicating BikeSales Table

For the RevenueData dimension table select all columns related to the Revenue ie, Revenue, Profit, Cost, Order_Quantity and Date. Right click on one of the selected table and select Remove Other Columns from the menu.

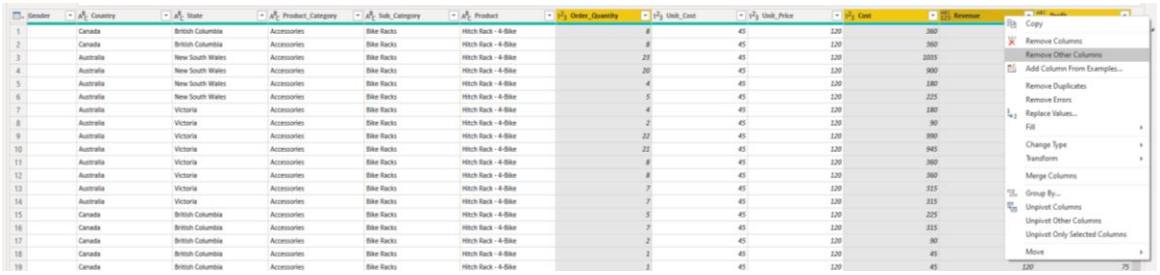
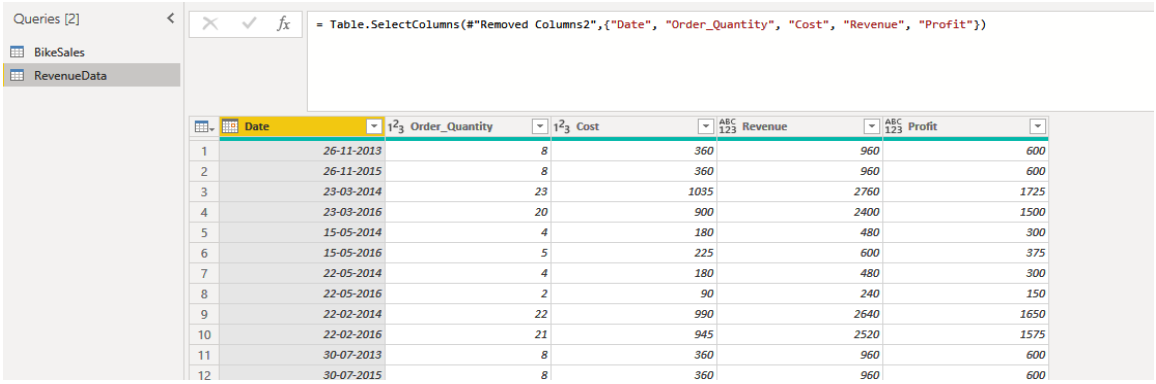


Figure 20: Remove unwanted columns from the dimension table

Now the dimension table has only five columns and renames as RevenueData



	Date	Order_Quantity	Cost	Revenue	Profit
1	26-11-2013	8	360	960	600
2	26-11-2015	8	360	960	600
3	23-03-2014	23	1035	2760	1725
4	23-03-2016	20	900	2400	1500
5	15-05-2014	4	180	480	300
6	15-05-2016	5	225	600	375
7	22-05-2014	4	180	480	300
8	22-05-2016	2	90	240	150
9	22-02-2014	22	990	2640	1650
10	22-02-2016	21	945	2520	1575
11	30-07-2013	8	360	960	600
12	30-07-2015	8	360	960	600

Figure 21: RevenueData Dimension Table

At last, we want to add an ID column to the dimension table. For this select IndexColumn from add a column and select From 1. Now the Index Column added and rename as Revenue Index.

BikeSales
RevenueData

	RevenueIndex	Date	Order_Quantity	Cost	Revenue	Profit
1	1	26-11-2013	8	360	960	600
2	2	26-11-2015	8	360	960	600
3	3	23-03-2014	23	1035	2760	1725
4	4	23-03-2016	20	900	2400	1500
5	5	15-05-2014	4	180	480	300
6	6	15-05-2016	5	225	600	375
7	7	22-05-2014	4	180	480	300
8	8	22-05-2016	2	90	240	150
9	9	22-02-2014	22	990	2640	1650
10	10	22-02-2016	21	945	2520	1575
11	11	30-07-2013	8	360	960	600
12	12	30-07-2015	8	360	960	600
13	13	15-07-2013	7	315	840	525
14	14	15-07-2015	7	315	840	525

Figure 22: RevenueData Dimension Table with RevenueIndex column

In the same way, we can create other three dimension tables

1. CustomerData table with columns CustomerIndex, Customer_Age, Customer_Gender and Age_Group
2. CountryData table with columns CountryIndex, Country and State
3. ProductData table with ProductIndex, Product, Product Category, Sub_Category, Unit_Cost and Unit_Price

Adding Id/Key Column to Fact Table

The ID columns from all the dimension tables have to be added as foreign keys to the fact table to create relationships. For this, the dimension table is merged with the fact table. This is done by clicking on Merge Queries on the home tab and connecting them to their shared columns. For the RevenueData table, the shared columns are Date, Order_Quantity, Profit, Cost and Revenue. These columns are selected on both tables and merged using the Left Outer join.

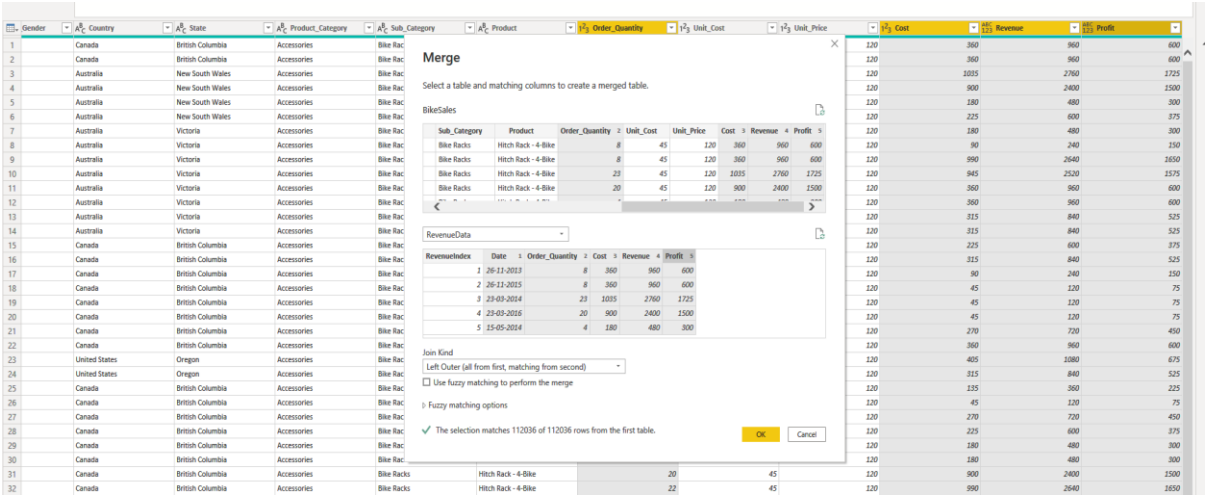


Figure 23: Merge Query for adding ID column of RevenueData table to Fact Table

Order_Quantity	Unit_Cost	Unit_Price	Cost	Revenue	Profit	RevenueData
8	45	120	360	960	600	Table
8	45	120	360	960	600	Table
23	45	120	1035	2760	1725	Table
20	45	120	900	2400	1500	Table
4	45	120	180	480	300	Table
5	45	120	225	600	375	Table
4	45	120	180	480	300	Table
2	45	120	90	240	150	Table
22	45	120	990	2640	1650	Table
21	45	120	945	2520	1575	Table
8	45	120	360	960	600	Table

Figure 24: Fact Table after Merged to RevenueData table

Expand the RevenueIndex column and select only the RevenuIndex and click Ok

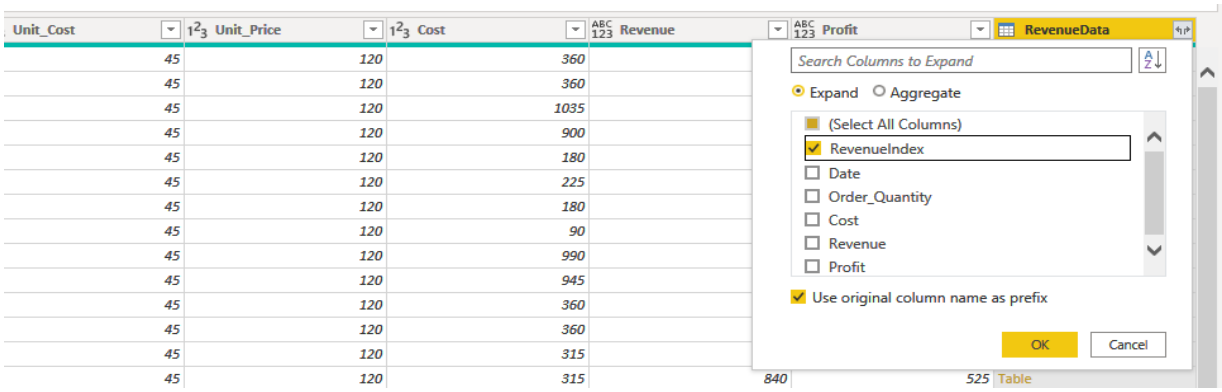


Figure 25: Expanding RevenueData Table and selecting RevenueIndex column

Order_Quantity	Unit_Cost	Unit_Price	Cost	Revenue	Profit	RevenueIndex
8	45	120	360	960	600	1
8	45	120	360	960	600	2
8	45	120	360	960	600	66
8	45	120	360	960	600	2
8	45	120	360	960	600	66
23	45	120	1035	2760	1725	3
20	45	120	900	2400	1500	4
4	45	120	180	480	300	5
4	45	120	180	480	300	415
4	45	120	180	480	300	5
4	45	120	180	480	300	415
5	45	120	225	600	375	6
4	45	120	180	480	300	7
4	45	120	180	480	300	445

Figure 26: RevenueIndex Added to the BikeSales Table

These steps are repeated for the other 3 tables

Then we need to remove all other columns except the Date, CustomerIndex, CountryIndex, RevenueIndex, SalesIndex and ProductIndex. At last, rename the BikeSales table to SalesData Table.

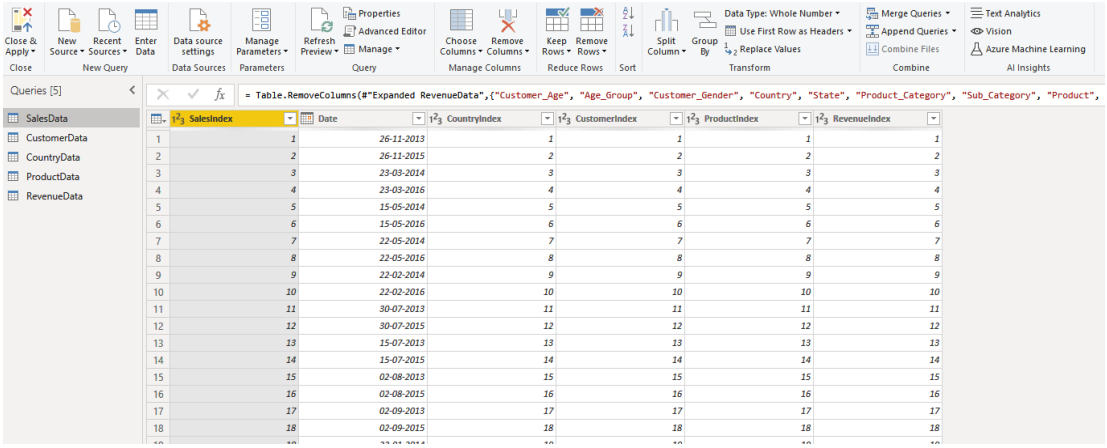


Figure 27: Power Query editor after all data pre-processing and modelling

Exit the power query editor by clicking Close& Apply.

3.2. Creating Relationships

If we want to make visualisations there is a relationship between the fact table and the dimension table. For creating a relationship choose "Manage relationships" from the data model sections home tab to carry out this action. When a dialogue box appears, select "New". The first relationship is between the SalesData fact table and the CustomerData dimension table based on the connecting column CustomerIndex and we get a relationship with cardinality Many to One. Repeat the same for the other three dimension tables.

At last, we have a Star schema model relationship where one fact table is connected to many dimension tables. Here, the fact table is the SalesData table and the dimension tables are CustomerData, RevenueData, CountryData and ProductData.

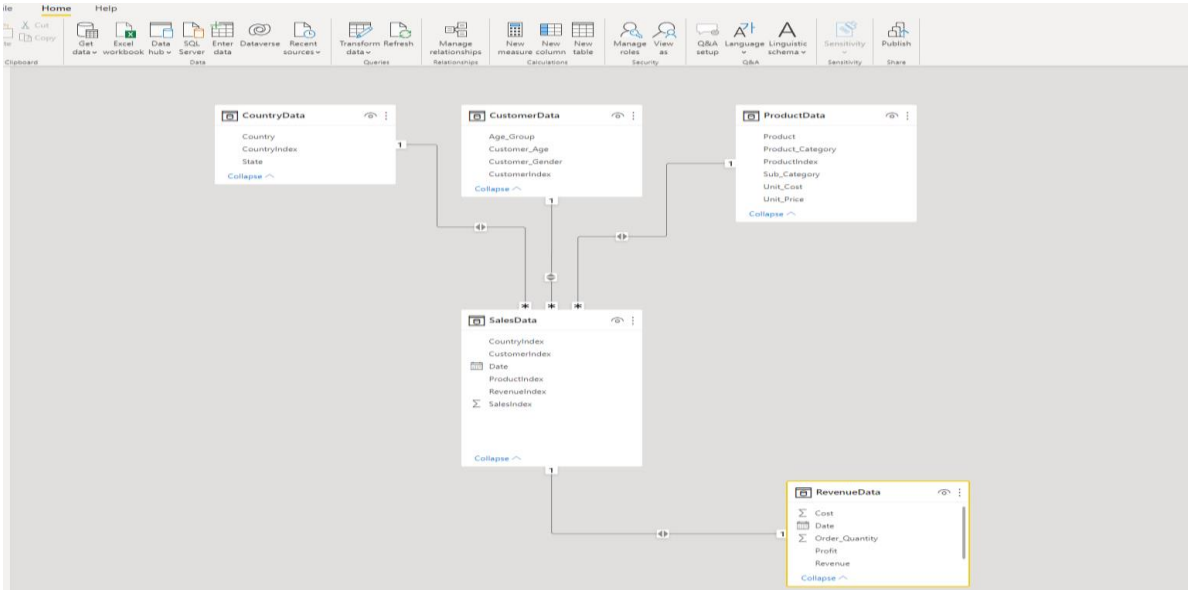


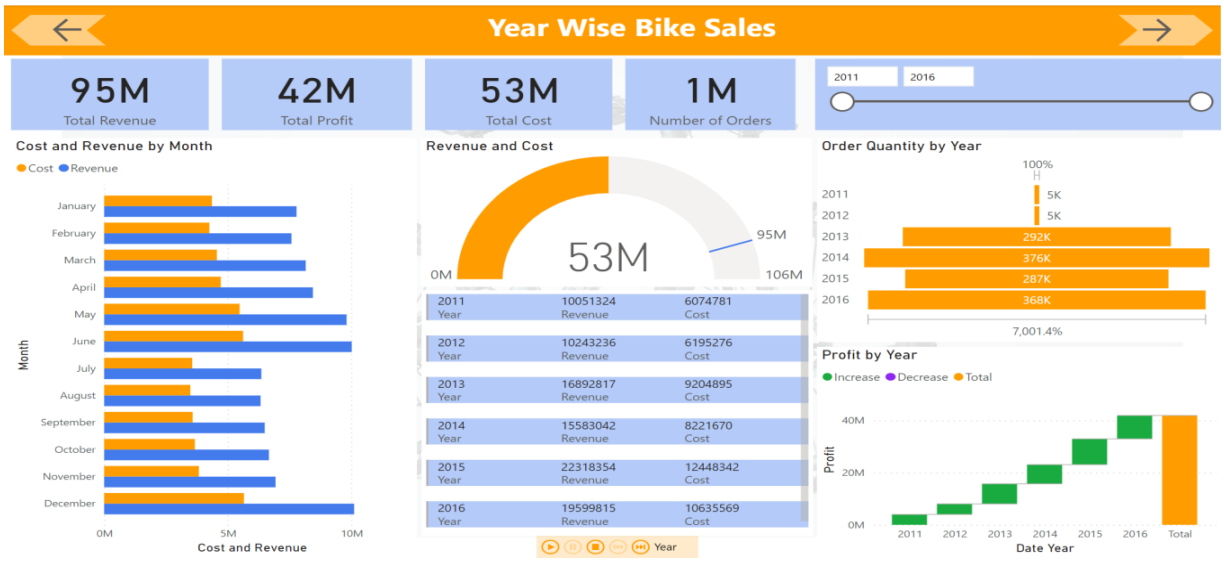
Figure 28: Data Model after relationships are created

4. Dashboards

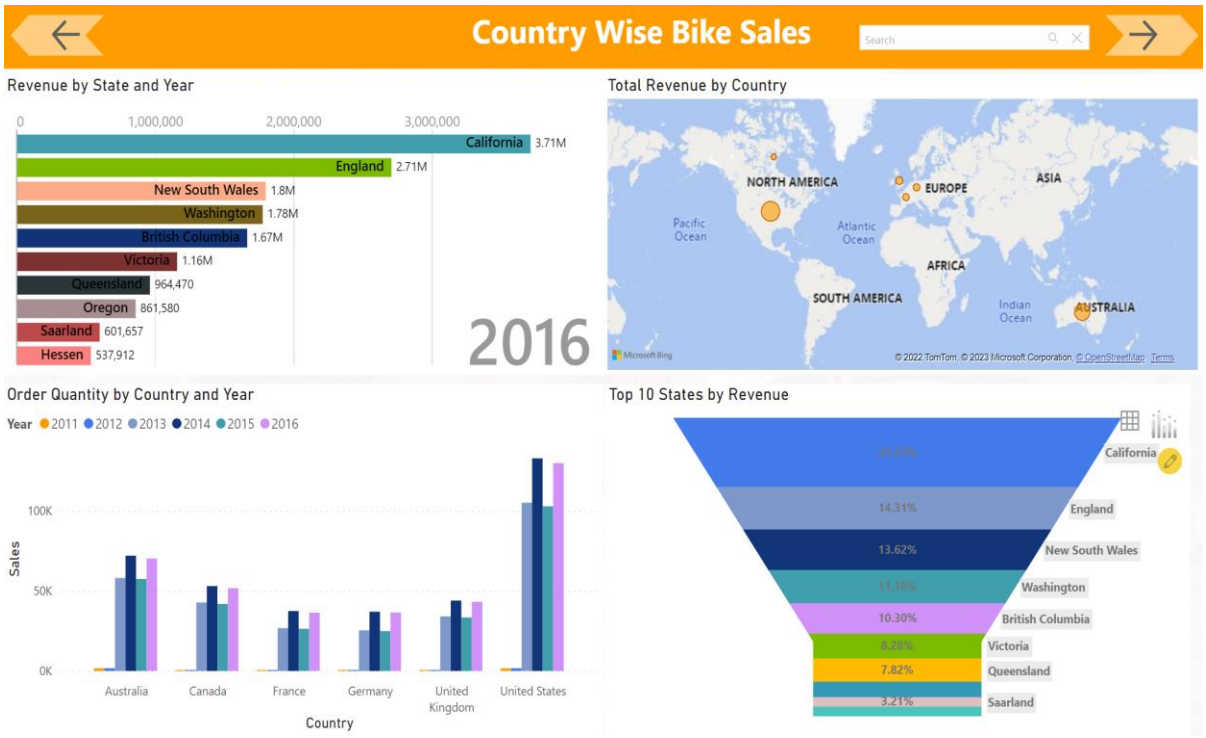
In this Power BI dashboard, there are 5 pages. The first page is the Home page and it contains buttons to navigate to the other four pages. The other four pages are Year Wise Sales, Country Wise Sales, Customer Wise Sales and Product Wise Sales. More Dashboard explanations are in first section.

Home Page

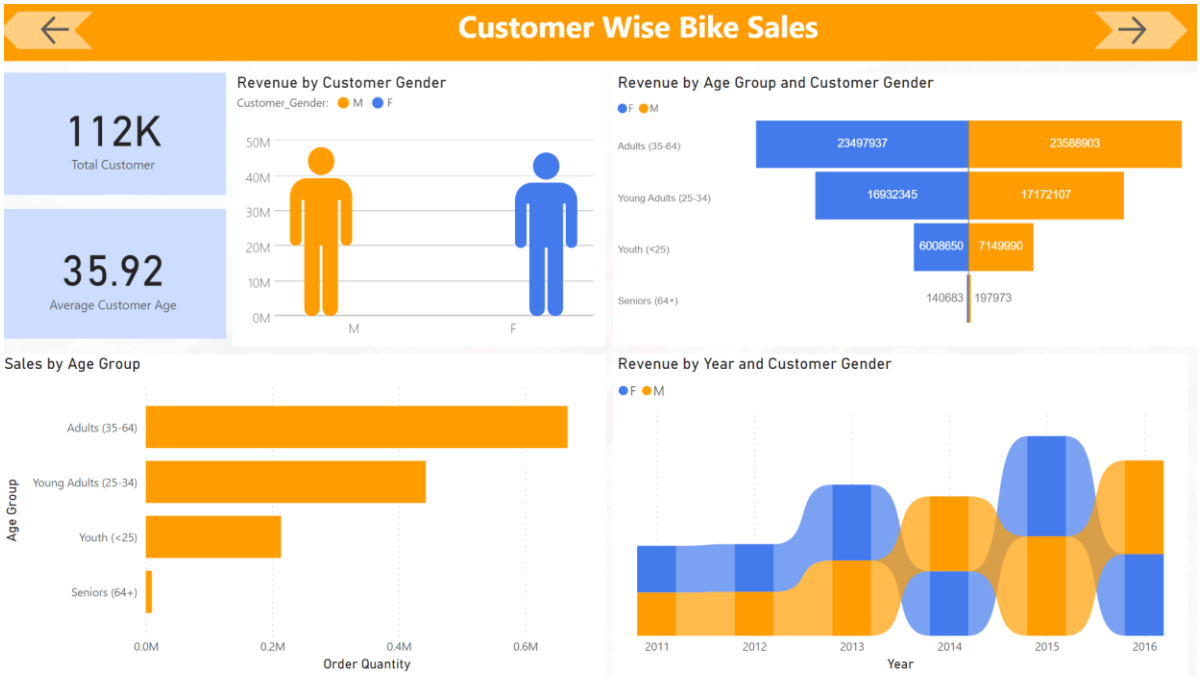
Year Wise Bike Sales Page



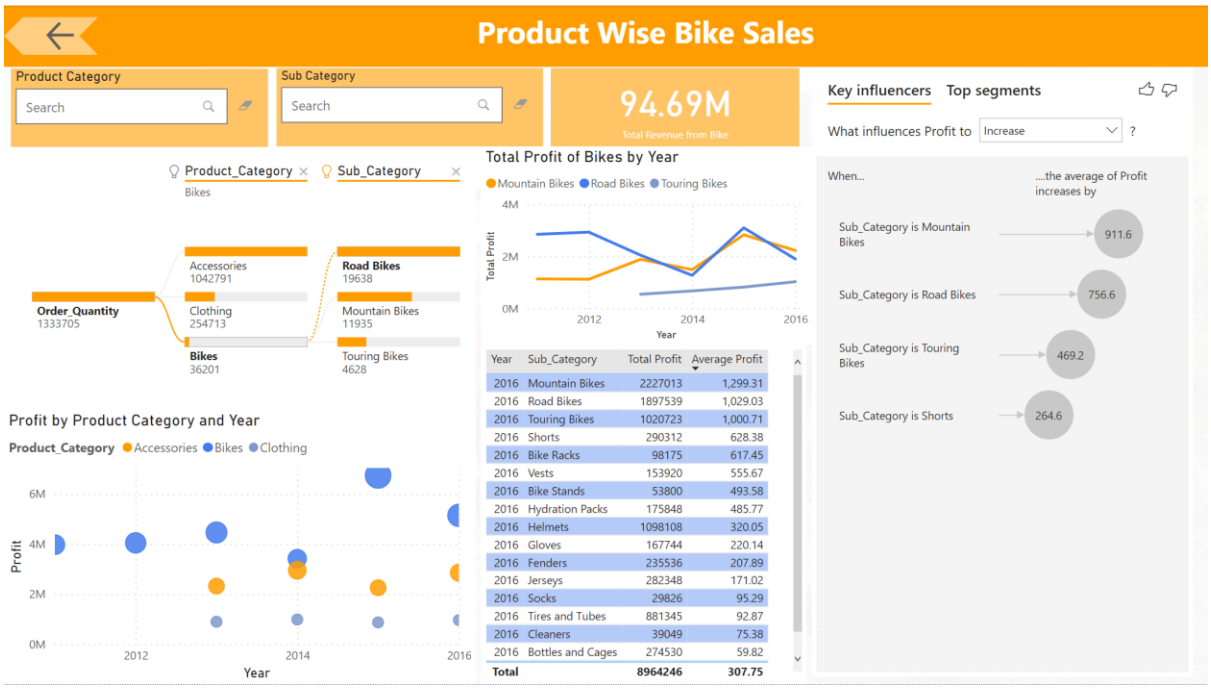
Country Wise Bike Sales Page



Customer Wise Bike Sales Page



Product Wise Bike Sales Page



5. Reference

https://learn.microsoft.com/en-us/power-bi/

https://www.pbiusergroup.com/home

https://learn.microsoft.com/en-us/training/powerplatform/power-bi?WT.mc_id=powerbi_landingpage-docs-link

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	93
Dax and M language	Both DAX and M Language have been extensively used in the report	90
Dashboard Design	The dashboard contains a variety of charts, including advanced ones.	95
Average		**Add below the average of the four cells above:** **93.25**