



BUSINESS INTELLIGENCE FOR USED CAR SALES ANALYSIS

CIS4008-N-BF1-2021

BIG DATA AND BUSINESS INTELLIGENCE

**SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL
TECHNOLOGIES**

Section 1



JANUARY 11, 2022

Prepared by

B1326237 GOWTHAMI NAGAPPAN

Contents

Acknowledgement.....	3
Dataset.....	4
Business Intelligence Questions.....	5
Loading Data.....	6
Data Pre-Processing and Data Cleansing.....	8
a. Rename the Table Name.....	8
b. Rename the Column Name.....	9
c. Removing the Column.....	10
d. Replacing the Values.....	10
e. Changing Data Types of the Column.....	14
f. Splitting Columns.....	16
g. Adding Custom Column.....	18
h. Creating New Index Column.....	22
Generating Fact and Dimension Tables.....	23
a. Creating Car Info Dimension Table.....	23
b. Creating Car Seller Dimension Table.....	25
c. Creating Car Type with Price Dimension Table.....	26
d. Creating Car Features Dimension Table.....	27
e. Creating Indian Used Car Sales 2021 Fact Table.....	28
Creating Primary and Secondary key in Fact and Dimension Table.....	29
Custom Column Calculation.....	19
M Query to Replace Column Values.....	12
DAX Expression.....	32
Creating Data Model and relationship.....	36
Section 2 Business Report.....	39
Introduction.....	40
Background.....	40
Key Findings.....	41
Recommendations.....	41
Dataset.....	43
Data Model.....	43
Findings based on analysis and evaluation.....	44
Conclusion.....	67
References.....	68

Acknowledgement

I am very thankfull to our module leader, Annalisa Occhipinti for her unwavering assistance and complete support through out the subject. She is a strong mentor and helped me in selecting the dataset for the ICA. Timely feedback on the clarification of queries coupled with highest degree of technical support, I was able to prepare the Business intelligence report in a professional manner. Without the support, I doubt on whether I would have been able to complete the report in a timely manner. Finally I want to express my sincere gratitude to everyone who has assisted and supported me along this journey.

Dataset

The dataset consists of the used car list, which was sold in 2021 through the online platform, Cardekho. The data list contains the details of Full name of the car model, Make, selling price of the car, Market price, Year of manufacture, Seller type, Kilometers driven, Owner type, Fuel type, Transmission type, Mileage of car, Engine CC, Maximum power, and Number of seats. The dataset was taken from a leading used car portal, www.cardekho.com which has operations in India. There were approximately data of 20,000 used cars in which there was 13 columns and 19,975 rows. The dataset was downloaded from www.kaggle.com. All the cars mentioned on the data set is manufactured (including assembled) and sold in India. Some of the car brands have discontinued their cars as the manufacturers have closed their operations.

Table mentioned below gives a brief overview of columns from ‘Used Car’ dataset.

Column Name	Description
Full Name	Name of the Car model with brand name
Selling Price	The price for which the car was sold. The sale price is based on multiple factors which include age of the car, kilometers driven and others
New Price	The current market price (showroom price) of a new car
Year	The year it was registered by the owner
Seller Type	There are three types of sellers who list the car on the portal. It is listed directly by the individual / dealer / trust mark dealer
KM Driven	Total kilometers the car has been driven. It is the main Influencer of the price
Owner Type	Number of owners for the car.
Fuel Type	Fuel used by the car. It is petrol, diesel, liquefied natural gas or electric.
Transmission Type	There are two different transmission types – Manual or Automatic. Manual variant involves the shifting of gears which is done by the driver whereas the gear shift is Automatic in Automatic cars.
Mileage	Mileage of Car is defined as number of Kilometers covered by Car for 1 liter of fuel. It changes according to the Driving Conditions and the model of the car.
Engine	Car's engine is measured in cubic centimeters (cc). It refers to the amount of air and fuel that can be pushed through the cylinders in the engine.
Max Power	The Max Power of Brake Horsepower (bhp) is the unit of power of an engine without any losses like heat and noise.
Seats	Total number of seats available in a car. It includes both the drivers and passenger's seat.

Business Intelligence Questions

The selected dataset has detailed information of Name of the car with model, Selling Price, New car Price, Manufacturing Year, Seller Type, Kilometers Driven, Owner Type, Fuel Type, Transmission Type, Mileage, Engine, Max Power and Seats.

The primary objective of the report is to analyze the best-selling car which is based on various factors like the kilometers driven, mileage, fuel type, age of the car and others. Various factors are taken into account for the comparison and the same has been given as an input in power BI tool to get valuable information.

The report provides answer to the following questions.

- Which is the highest selling car model?
- Which car model sale is high when compared with high / low mileage?
- How many cars are sold for each Fuel Type?
- Which car (based on manufactured year) sold the most?
- Which fuel type car sale is higher when compared with the transmission (Manual / Automatic) type?
- Which seat variant (number of seats) sale is higher?
- How many cars were sold for a particular year that has run for more than 100,000 Kilometers?
- How many numbers of car models were sold according to seller type?

Loading Data

Loading the raw data is the first stage in data analysis and business intelligence. Power BI can pull the data from a variety of places, including web APIs, databases, flat files in Excel, csv (comma-separated values) files and more.

Power BI displays a dialogue box on launch that allows you to load an existing project or create a new one. As illustrated in the diagram below, we can collect the data.

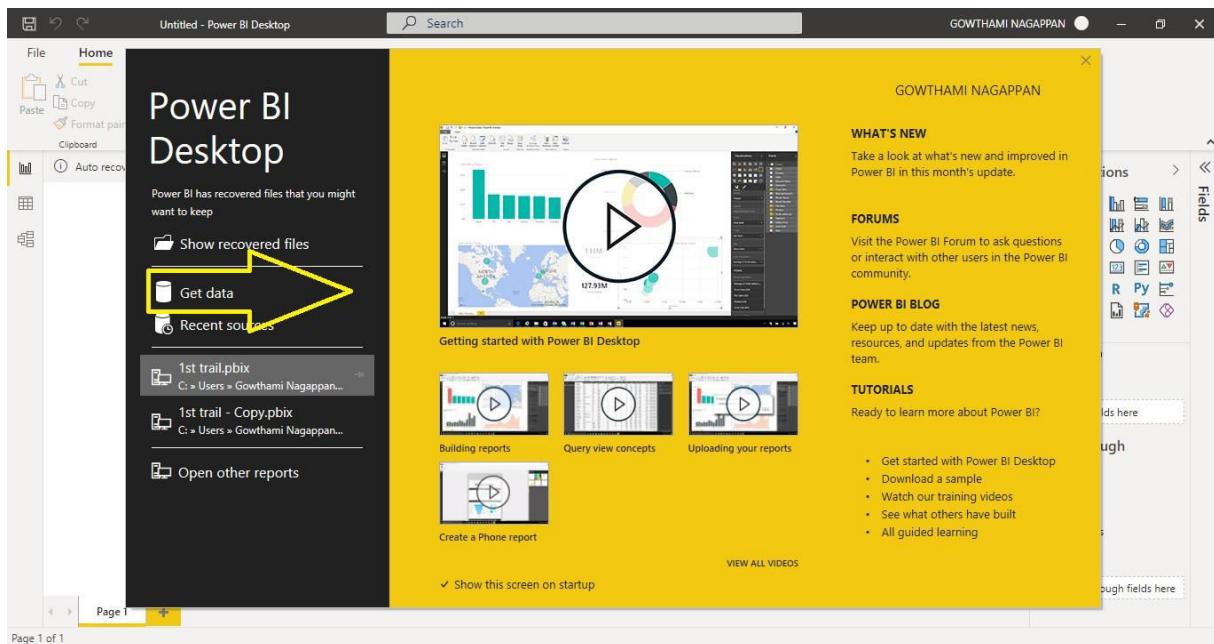


Fig: Power BI initialization

For this project, we had to load the dataset which was downloaded from Kaggle (Used Car Prices in India), and the format was in csv (cardekho updated.csv). The dataset was loaded by selecting "Text/CSV" from the "Get Data" menu.

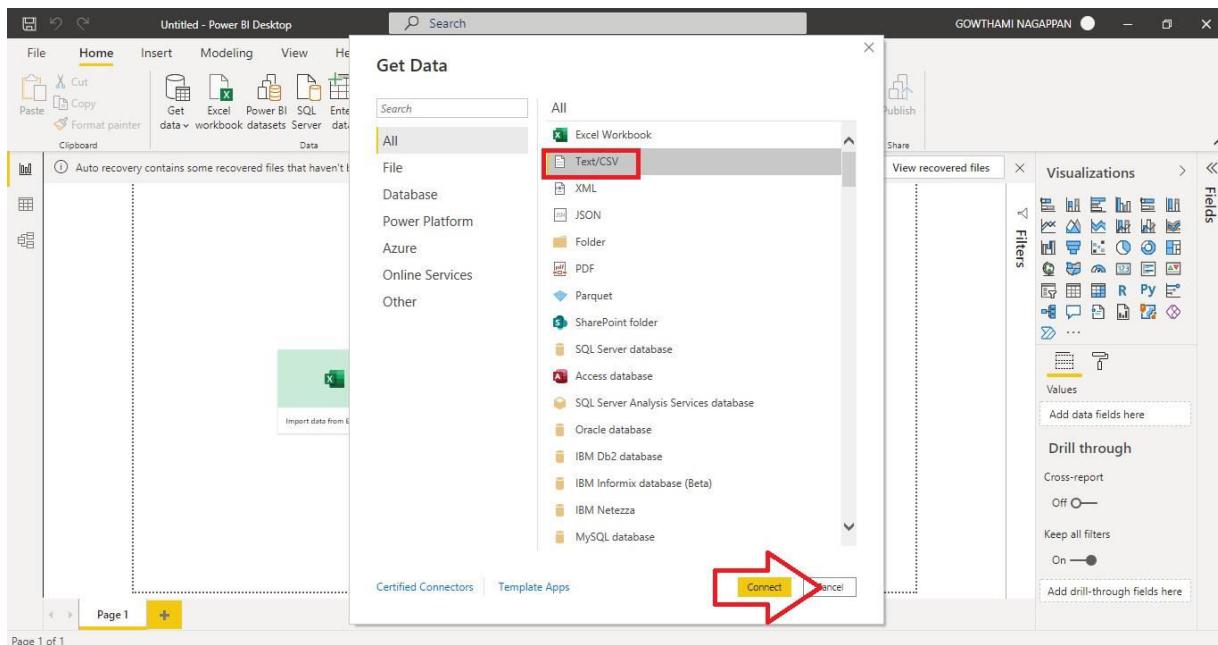


Fig: Dialog box for obtaining data

While picking a csv file, a dialogue box appeared with the options to either load or transform data as shown in the figure below.

By selecting the 'Load' option, the csv data loaded without being transformed at this time. Based on the values of the first 200 rows, this phase automatically identified the data type of the table's column.

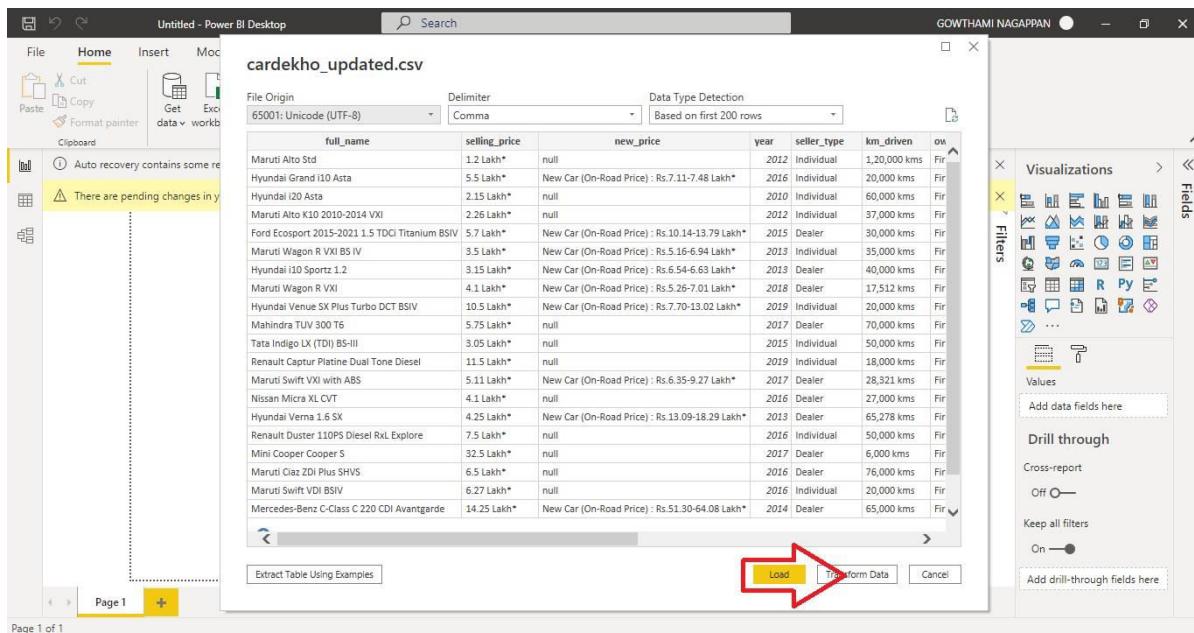


Fig: CSV file of Cardekho_updated is being loaded

Upon loading the relevant datasets, the data should be pre-processed, cleaned, and transformed to provide 'fact and dimension' tables for business intelligence.

Data Pre-Processing and Data Cleansing

After loading the CSV file into power-BI, data pre-processing steps is required to be done which includes adding new columns, renaming columns, replacing values for better understanding, changing data types for easier analysis, and removing unused / unwanted columns.

All these procedures are achievable with ‘Power Query Editor’. The 'Transform data' option in the 'Table tools' menu was selected to open the power query editor, as illustrated in the figure below.

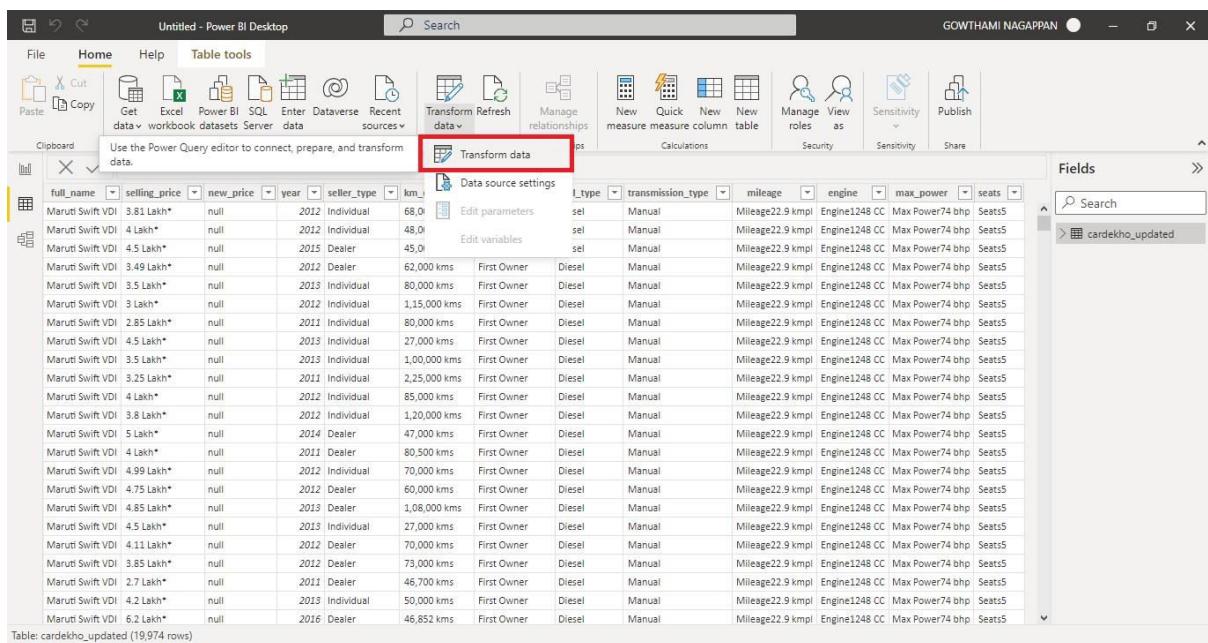


Fig: Selecting the Transform Data option

a. Rename the Table Name

The Table name was right-clicked and the ‘rename’ options was selected to change the Table name. The same step can be performed by double clicking the Table name cell.

The table name was changed from ‘Cardekho_updated’ into ‘Indian_Car_Sales_2021’

The screenshot shows the Power Query Editor interface. A context menu is open over the first row of the table, with the 'Rename' option highlighted. The table contains data about car models and their prices. The right pane shows 'APPLIED STEPS' with 'Changed Type' selected.

Fig: Renaming the Table Name

b. Rename the Column Name

In Indian_Car_Sales_2021, The column name ‘Full name’ was right clicked and ‘rename’ option was selected. The name was changed to ‘Car Brand with Model’.

In a similar way the ‘selling price’ was changed into ‘SellingPrice’. The same step was performed for other columns as well. The steps performed is shown in the image below.

The screenshot shows the Power Query Editor interface. A context menu is open over the 'full_name' column, with the 'Rename...' option highlighted. The table contains data about car models and their prices. The right pane shows 'APPLIED STEPS' with 'Changed Type' selected.

Fig: Renaming column name

c. Removing the Columns

We had new_price column (current price of new car) which had lots of null value which may affect our report. So, the new price column was deleted by right clicking on the column and the option ‘Remove column’ was selected as shown in the image below.

The new_price column had lots of unknown price values (approx. 19,000 missing values) which is not much use for our analysis. So, the columns were removed.

The screenshot shows the Power Query Editor interface. A table named 'Car Brand with Model' is displayed with three columns: 'full_name', 'Car Brand with Model', and 'new_price'. The 'new_price' column is highlighted with a yellow box. A context menu is open over this column, with the 'Remove' option highlighted with a red box. The menu also includes other options like 'Remove Other Columns', 'Duplicate Column', and 'Add Column From Examples...'. The 'APPLIED STEPS' pane on the right shows a step named 'Renamed Columns'.

Fig: Removing new_price column

d. Replacing the Values

For selecting the right data type of the column, the column should contain only the text or value.

Many of the used car data set had the text in conjunction with the values (example: KM Driven column had value with the text 1,20,000 kms). So, the text was replaced with empty space to keep the only values. Instead of 1,20,000 Kms it was replaced with 120000.

To replace columns data, the same option was followed by right-clicking the chosen column and the option ‘replace values’ was selected as shown below.

The screenshot shows the Power Query Editor interface with a table named 'Id with Model'. The 'KM_Driven' column is selected, and its context menu is open. The 'Replace Values...' option is highlighted with a red box.

Fig: Replacing column KM_Driven Column values

The KM_Driven Column had the word ‘kms’ and symbol (,) where we had removed all those words and symbol by empty space as shown below.

The screenshot shows the Power Query Editor with the 'Replace Values' dialog box open. The 'Value To Find' field contains 'kms' and the 'Replace With' field is empty. The 'OK' button is highlighted with a red box.

Fig: Replacing kms into empty space

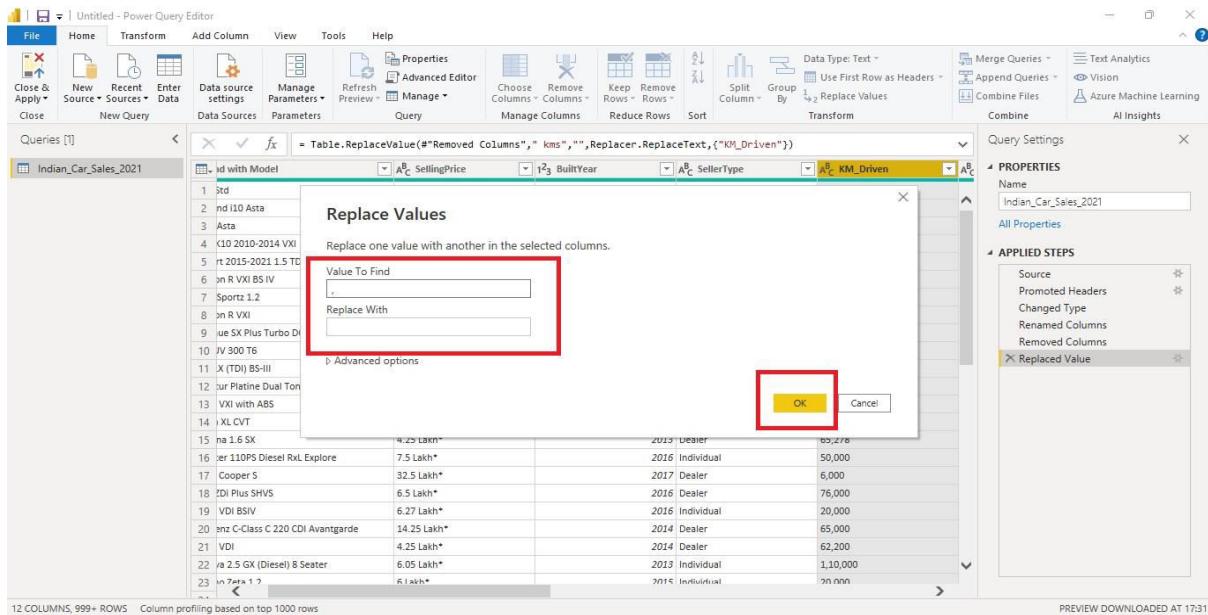


Fig: Replacing (,) into empty space

'M' Query to Replace Column Values

There are other columns which contain value with the text in similar to 'KM_Driven'. Instead of using 'replace value' option each time, M query can be used for all the columns.

As we did in the above step of removing the text from the values, the similar step was performed for all other column values which had the values with texts. For example, In Mileage_kmpl, 'Mileage 22.9 kmpl' was replaced with '22.9' only and in the case of Engine_cc (Engine 77 cc was replaced with 77).

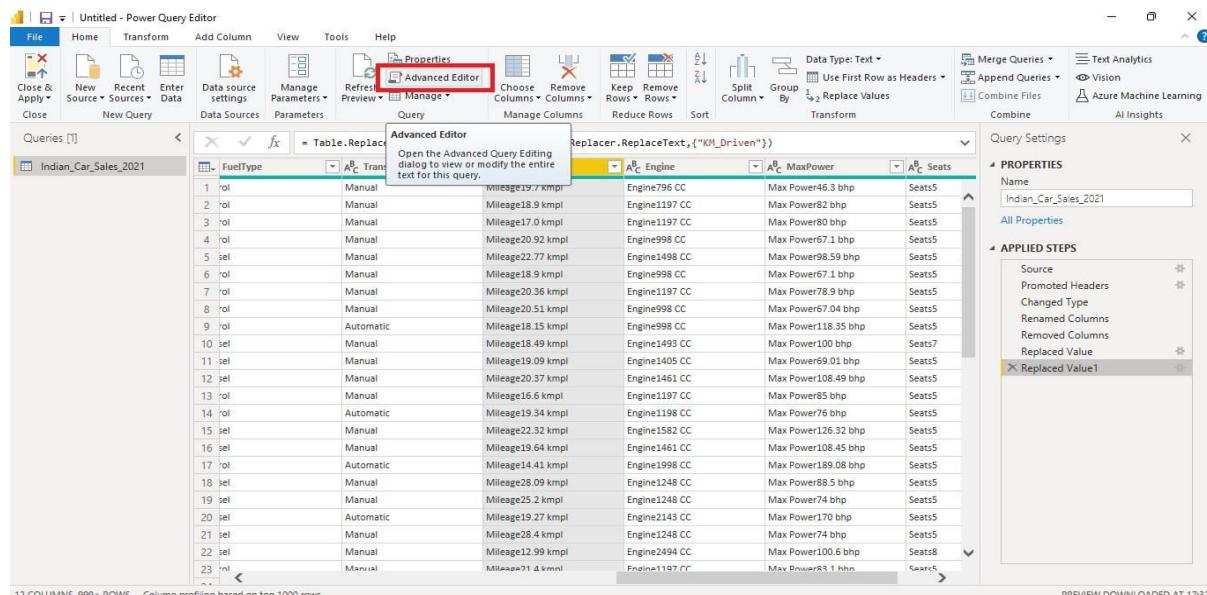


Fig: Advanced Editor to write M Query

The ‘Advanced Editor’ option in Toolbar was selected and the ‘M’ query was written to replace the values as shown below.

```

let
    Source = Csv.Document(File.Contents("C:\Users\Gowthami Nagappan\Desktop\archive\cardekho_updated.csv"),[Delimiter=",", Columns=13, Encoding=65001, QuoteStyle=QuoteStyle.None]),
    #Promoted Headers = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #Changed Type = Table.TransformColumnTypes(#"Promoted Headers",{{"full_name", type text}, {"selling_price", type text}, {"new_price", type text}, {"year", Int64.Type}, {"seller_type", type text}, {"kms", type number}, {"brand", type text}, {"model", type text}, {"seats", type number}, {"max_power", type number}, {"engine", type text}, {"transmission_type", type text}, {"fuel_type", type text}, {"milage", type number}),
    #Renamed Columns = Table.RenameColumns(#"Changed Type",{{"full_name", "Car Brand with Model"}, {"selling_price", "SellingPrice"}, {"new_price", "NewPrice"}, {"year", "BuiltYear"}, {"seller_type", "SellerType"}, {"kms", "KMs_Driven"}, {"brand", "Brand"}, {"model", "Model"}, {"seats", "Seats"}, {"max_power", "MaxPower"}, {"engine", "Engine"}, {"transmission_type", "TransmissionType"}, {"fuel_type", "FuelType"}, {"milage", "Mileage"}),
    #Removed Columns = Table.RemoveColumns(#"Renamed Columns",{"new_price"}),
    #Replaced Value = Table.ReplaceValue(#"Removed Columns", "kms", "", Replacer.ReplaceText,{"KMs_Driven"}),
    #Replaced Value2 = Table.ReplaceValue(#"Replaced Value", "Milage", "", Replacer.ReplaceText,{"Mileage"}),
    #Replaced Value3 = Table.ReplaceValue(#"Replaced Value2", "kmp1", "", Replacer.ReplaceText,{"Mileage"})
in
    #Replaced Value3

```

No syntax errors have been detected.

Done **Cancel**

Fig: ‘M’ Query for Replacing Values

‘M’ query was written to replace mileage column values as shown below.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. The 'Transform' tab is active. A red box highlights the 'Mileage' column in the table preview. The 'Applied Steps' pane on the right shows the steps taken: Promoted Headers, Changed Type, Renamed Columns, Removed Columns, Replaced Value, Replaced Value2, and Replaced Value3. The 'Replaced Value2' step is specifically highlighted with a red box.

FuelType	TransmissionType	Mileage	Engine	MaxPower	Seats
1 l	Manual	19.7	engine796 CC	Max Power63 bhp	Seats5
2 l	Manual	18.9	engine1197 CC	Max Power82 bhp	Seats5
3 l	Manual	17.0	engine1197 CC	Max Power80 bhp	Seats5
4 l	Manual	20.92	engine998 CC	Max Power67.1 bhp	Seats5
5 sel	Manual	22.77	engine1498 CC	Max Power90.59 bhp	Seats5
6 l	Manual	18.9	engine998 CC	Max Power71.1 bhp	Seats5
7 l	Manual	20.36	engine1197 CC	Max Power78.9 bhp	Seats5
8 sel	Manual	20.51	engine998 CC	Max Power67.04 bhp	Seats5
9 sel	Automatic	18.15	engine998 CC	Max Power18.35 bhp	Seats5
10 sel	Manual	18.49	engine1493 CC	Max Power100 bhp	Seats7
11 sel	Manual	19.09	engine1405 CC	Max Power69.01 bhp	Seats5
12 sel	Manual	20.37	engine1461 CC	Max Power108.49 bhp	Seats5
13 sel	Manual	16.6	engine1197 CC	Max Power85 bhp	Seats5
14 sel	Automatic	19.34	engine1198 CC	Max Power76 bhp	Seats5
15 sel	Manual	22.32	engine1582 CC	Max Power126.32 bhp	Seats5
16 sel	Manual	19.64	engine1461 CC	Max Power108.45 bhp	Seats5
17 sel	Automatic	14.41	engine1998 CC	Max Power189.08 bhp	Seats5
18 sel	Manual	28.09	engine1248 CC	Max Power85.5 bhp	Seats5
19 sel	Manual	25.2	engine1248 CC	Max Power74 bhp	Seats5
20 sel	Automatic	19.27	engine2143 CC	Max Power170 bhp	Seats5
21 sel	Manual	28.4	engine1248 CC	Max Power74 bhp	Seats5
22 sel	Manual	12.99	engine2494 CC	Max Power100.6 bhp	Seats8
23 sel	Manual	21.4	engine1197 CC	Max Power93.1 bhp	Seats5

Fig: The final Mileage Column after Replaced values

The same step was performed for other columns such as Engine_cc, MaxPower_bhp and Seats. The result will look as shown in the below image.

The screenshot shows the Power Query Editor interface with a table titled "Indian_Car_Sales_2021". The table has columns: TransmissionType, Mileage_kmpl, Engine_cc, MaxPower_bhp, and Seats. A red box highlights the "Replaced Value" section in the "APPLIED STEPS" pane on the right, which lists eight steps: Replaced Value, Replaced Value1, Replaced Value2, Replaced Value3, Replaced Value4, Replaced Value5, Replaced Value6, Replaced Value7, Renamed Columns1, and Replaced Value8. The "Name" field in the "PROPERTIES" pane is set to "Indian_Car_Sales_2021".

TransmissionType	Mileage_kmpl	Engine_cc	MaxPower_bhp	Seats
Manual	19.7	796	46.3	5
Manual	18.9	1197	82	5
Manual	17.0	1197	80	5
Manual	20.92	998	67.1	5
Manual	22.77	1498	98.59	5
Manual	18.9	998	67.1	5
Manual	20.36	1197	78.9	5
Manual	20.51	998	67.04	5
Automatic	18.15	998	118.35	5
Manual	18.49	1493	100	7
Manual	19.09	1405	69.01	5
Manual	20.37	1461	108.49	5
Manual	16.6	1197	85	5
Automatic	19.34	1198	76	5
Manual	22.32	1582	126.32	5
Manual	19.64	1461	108.45	5
Automatic	14.41	1998	189.08	5
Manual	28.09	1248	88.5	5
Manual	25.2	1248	74	5
Automatic	19.27	2143	170	5
Manual	28.4	1248	74	5
Manual	12.99	2494	100.6	8
Manual	21.4	1197	83.1	5

Fig: Replaced Values for all Columns

e. Changing Data Type of the Column

Data type for each column is an important aspect for future calculation and analysis as report need the column data type in a proper manner. For example, it could be a proper text or whole number.

Car Brand with Model as Text, SellingPrice as whole number in ‘currency’ format, SellerType as Text, KM_Driven as Number, OwnerType as Text, FuelType as Text, TransmissionType as Text, Mileage_kmpl as number, Engine_cc as number, MaxPower_bhp as number and Seats as number. Similar to the business terminology, the data type was changed.

To change data type in ‘Power Query Editor’, the left side icon on the column name was clicked as shown in below image and the corresponding data type was selected as mentioned in the above steps.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. In the main area, the 'Mileage_kmpl' column is highlighted and its data type is being changed. A context menu is open over the column header, with the 'Decimal Number' option highlighted by a red box. The 'Applied Steps' pane on the right shows the step 'Removed Columns1'.

Fig: Changing Text into Decimal Number Data type

Similar to the above, the same steps were repeated for all other column such as Mileage_kmpl as Decimal number, Engine_cc as number, Max_power_bhp as Decimal number and Seats as number data type.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. The 'Applied Steps' pane on the right shows the step 'Changed Type1'. Red boxes highlight the data types for 'Mileage_kmpl', 'Engine_cc', 'MaxPower_bhp', and 'Seats' in the column headers, which are now displayed as numbers. The 'Preview' pane at the bottom right shows the current state of the data.

Fig: After Changing Data type of all Columns

f. Splitting Columns

Indian_Car_Sales_2021 data contains SellingPrice values with Lakh, Cr and * with each value of columns. It means that the car soldout price was in Indian rupees. For better calculation, it was changed into whole value (Example: 1.2 Lakh into 120000 / 1 Cr into 10000000 / 90,000* into 90000).

To change the entire Column, the column was split into two by clicking the ‘Delimiter’ option which is shown below.

Fig: Splitting Column

Fig: Splitting by Space Right Most Delimiter

After using the above ‘Delimiter’ option, the space is considered for splitting the column.

After splitting the columns, the first column had the decimal number (1.2), and the other column had the text. Example of text in our case includes Lakh*, Cr* and * as null in second column.

The Second Column data was replaced into number (Lakh* into 100000 / Cr* into 1000000 / null into 1) by using the replaced values. It is shown in the below for reference.

The screenshot shows the Power Query Editor interface with the 'Replace Values' dialog open. The 'Value To Find' field contains 'Lakh*' and the 'Replace With' field contains '100000'. The 'OK' button is highlighted with a red box.

Fig: Replaced Lakh* into 100000

The screenshot shows the Power Query Editor interface with the 'Replace Values' dialog open. The 'Value To Find' field contains 'null' and the 'Replace With' field contains '1'. The 'OK' button is highlighted with a red box.

Fig: Replaced Null into 1

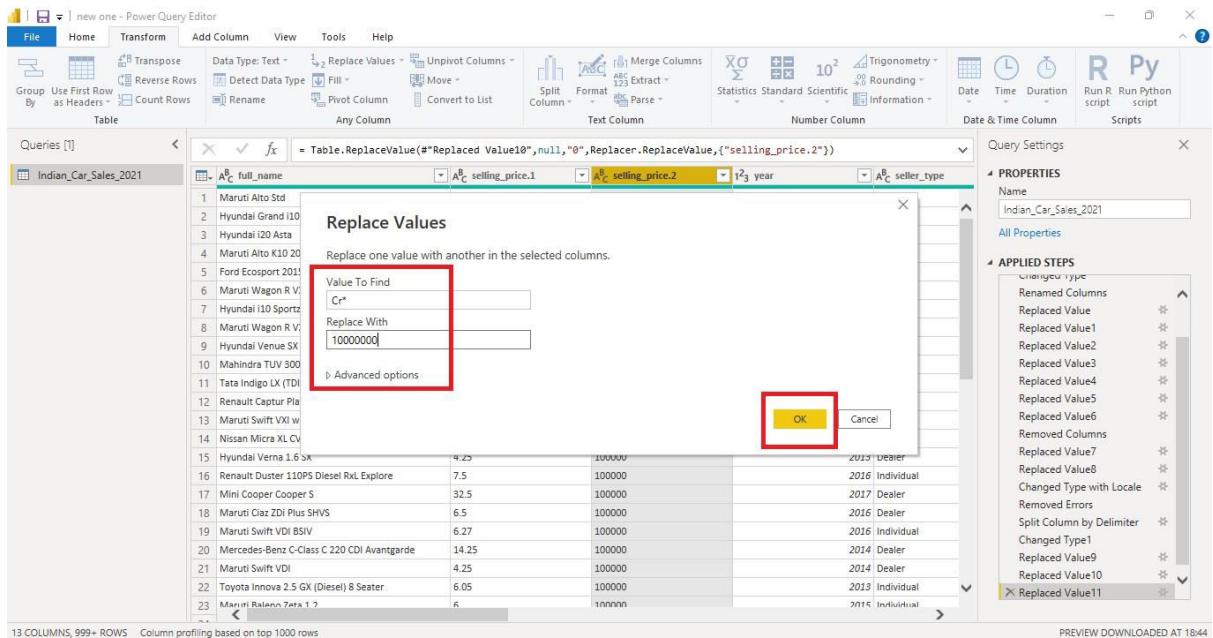


Fig: Replaced Cr* into 10000000

g. Adding Custom Column

Before adding custom column and to make the process simple the data type of column was changed into whole number to get the perfect value of selling price.

To add new column ‘add column’ option was clicked on the toolbar and the ‘custom column’ option was chosen as shown below.

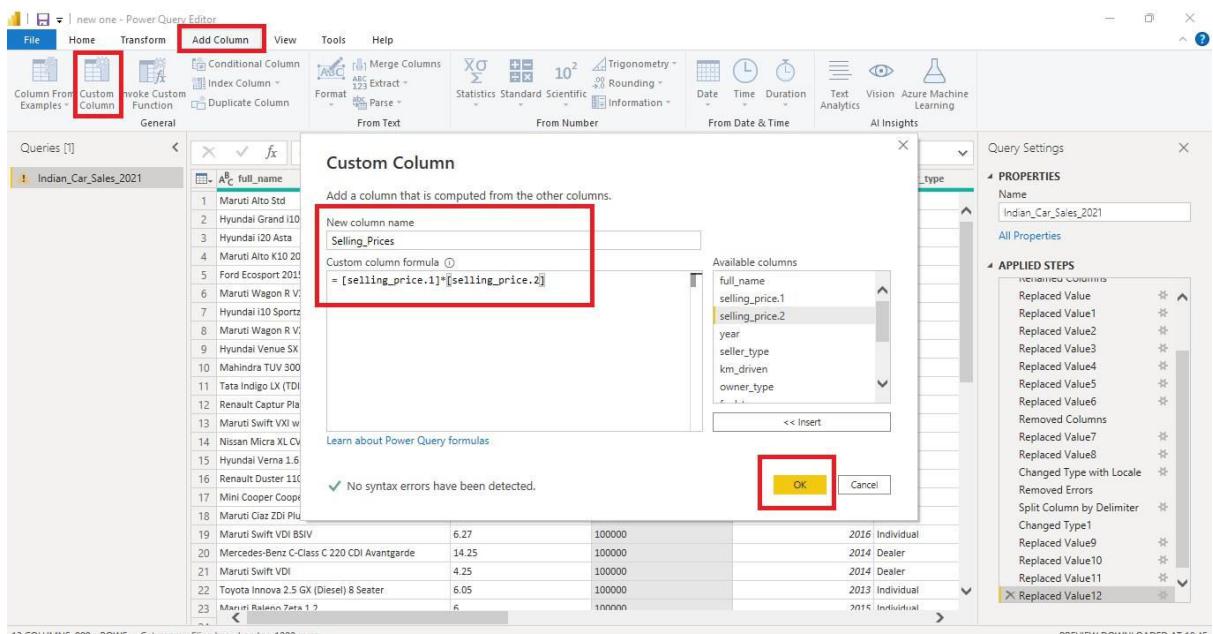


Fig: adding custom column

Custom Column Calculation

The new calculation for the column selling price was performed to convert lakhs / Cr into the exact value of the selling price.

Formula was used to convert the selling price $1.2 * \text{Lakhs}$ into 120000 in the new column by using the formula which is shown in the image below. The two splitted columns were multiplied to get the value. The new column name, SellingPrice and the values of the column will appear after executing the formula which is shown below.

$$\text{Selling_Prices} = [\text{Selling_price-1}] * [\text{Selling_price-2}]$$

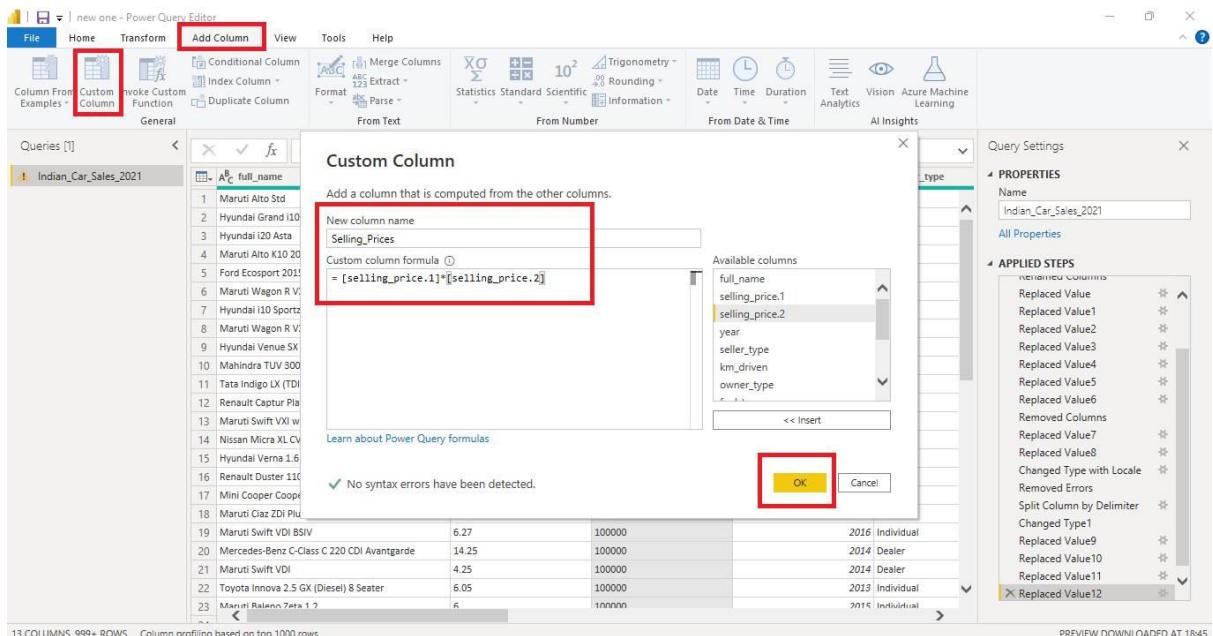


Fig: creation of new custom column Selling Price

After the execution of Formula, the ‘new custom as Selling_Prices’ column was created as shown in the image below.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. The 'Transform' tab is active. A red box highlights the 'SellingPrice' column, which has been moved to the rightmost position. The 'Applied Steps' pane on the right lists the 'Reordered Columns' step.

Fig: New SellingPrice column has added with values

Each line in the ‘let’ section was translated to an independent step in ‘advanced editor’ which is shown in the below image. The custom column is highlighted for better understanding.

After creating SellingPrice column with exact value the other column of selling prices was deleted as it is used only for analysis. The column was removed in similar to the one done before in section 2 under data preprocessing step.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. The 'Transform' tab is active. A red box highlights the 'Remove Column' option in the context menu for the 'selling_price' column.

Fig: Removing few column

The transformation done in ‘Power Query Editor’. The M language cannot be used in ‘Data View’ or ‘Report View’ in ‘Power BI’ instead we should use the DAX.

Table: Indian_Car_Sales_2021 (19,163 rows) Column: SellingPrice (1,184 distinct values)

Fig: Selecting the Currency to SellingPrice Column

The ‘Data View’ option was chosen to see the preprocessed Data. The Indian currency was added to the ‘SellingPrice’ column by clicking ‘column tools’. Then the currency icon was clicked, and the currency value was changed as per the data type.

Table: cardekho_updated (19,311 rows) Column: Soldout_Price (1,184 distinct values)

Fig: Selecting Indian Currency

As the dataset is related to India, the currency was selected as ‘Hindi (India) India Rupee’ which is shown below.

The screenshot shows the Power BI Desktop interface with the 'new one - Power BI Desktop' window open. The ribbon at the top has 'File', 'Home', 'Help', 'Table tools', and 'Column tools' selected. In the 'Column tools' tab, the 'SellingPrice' column is selected. The 'Format' dropdown is set to 'Currency'. The 'Value' section shows '\$' and '0' with a red box highlighting the '0'. The 'Fields' pane on the right lists various columns from the 'Indian_Car_Sales_2021' table, with 'SellingPrice' highlighted in yellow and a red box around its value range.

Fig: After Currency Change

h. Creating New Index Column

In Power Query Editor, adding the default index column was helpful where the ‘data modelling’ was used to create secondary keys for the relationship between tables.

The screenshot shows the Power Query Editor interface with the 'new one - Power BI Desktop' window open. The ribbon at the top has 'File', 'Home', 'Transform', and 'Add Column' selected. Under the 'Add Column' tab, the 'Index Column' option is highlighted with a red box. The main area shows a table with columns like 'Car ID', 'Car Brand with Model', 'BuiltYear', 'SellerType', and 'KM Driven'. The 'Query Settings' pane on the right shows the 'Name' as 'Indian_Car_Sales_2021' and a list of applied steps.

Fig: Creating Index Column

The ‘Transform data’ option was chosen from the options in the tool bar to add index column in power query editor with the option to start from 1 in Indian_Car_Sales_2021 as shown in the image above.

After creating index column, we have reordered the index by dragging the column and released it in the required location (Example: before Car Brand with Model). The index column was renamed with the CarID.

Generating Fact and Dimension Tables

A single flat table with 13 columns and 20,000 rows of ‘Used Car’ dataset from a firm having operations in India was selected. This dataset had Full name of the car with model, SellingPrice, Year, KM Driven, OwnerType, FuelType, Transmission Type and other details which were among the records in the database. Since all the data were stored in a single flat file or table, there were several difficulties in handling, managing, and visualizing / answering business related challenges. So, there was a duplication of data which resulted in a big table size. As a result, a star schema was created by normalizing this ‘flat table’ into several dimension tables and one fact table.

a. Creating Car Info Dimension Table

To make the ‘Car Info’ dimension table, a duplicate of Indian_Car_Sales_2021 table was created by right clicking on the table and clicking the ‘Duplicate’ option as shown in the below image.

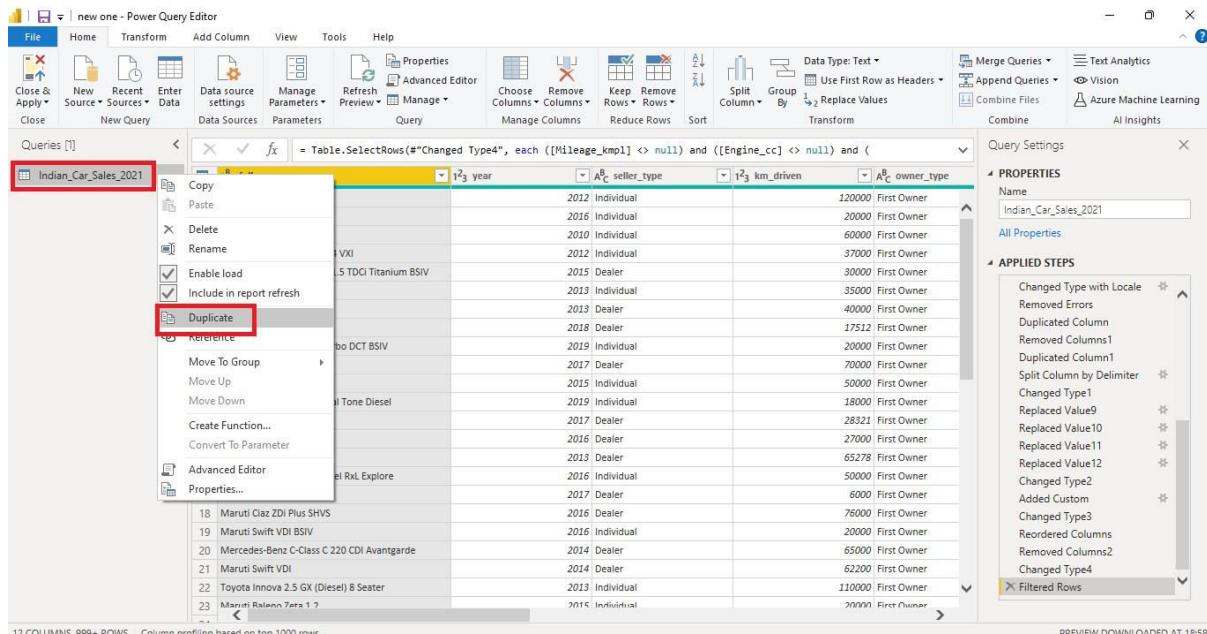


Fig: Duplicating Fact Table

Then the table name was renamed by right-clicking on the name. The same option can be performed by right clicking and selecting the ‘Rename’ option.

The table name was changed to ‘Car info’ and the following columns were selected: Car Brand with Model, Year and KM Driven by pressing the ctrl key.

The selected columns were chosen by right-clicking and the option ‘Remove other columns’ was clicked as shown below.

The screenshot shows the Power Query Editor interface. In the 'Queries' list, 'Car Info' is selected. In the main grid, four columns are highlighted with red boxes: 'Car Brand with Model', 'BuiltYear', 'SellerType', and 'KM Driven'. A context menu is open over these columns, with the 'Remove Other Columns' option highlighted. The menu also includes options like 'Remove Columns', 'Add Column From Examples...', 'Remove Duplicates', 'Replace Values...', 'Change Type', 'Merge Columns', 'Group By...', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', and 'Move'.

Fig: Removing other columns except selected columns

Dimension Table should contain unique rows to avoid ambiguity. All the columns were selected using ctrl key. Right-click and then ‘Remove Duplicates’ options was selected which eliminated the duplicate records from the table as shown in the image below.

The screenshot shows the Power Query Editor interface with the 'Car Info' dimension table selected. A context menu is open over the 'Car ID' column, with the 'Remove Duplicates' option highlighted. The table contains 23 rows of car information, including columns for Car ID, Car Brand with Model, BuiltYear, and KM Driven.

Fig: Removing duplicate records from Car Info dimension table

b. Creating Car Seller Dimension Table

Similar to Car Info dimension Table, new table was created by duplicating Indian_Car_Sales_2021 table and it was renamed to ‘Car Seller’ except certain columns which are Car Brand with Model, SellerType, OwnerType. All other columns were removed including the duplicates as shown in the image below.

The same procedure was followed to create 2 more-dimension tables.

The screenshot shows the Power Query Editor interface with the 'Car Seller' dimension table selected. A context menu is open over the 'SellerType' column, with the 'Remove Other Columns' option highlighted. The table contains 23 rows of car seller information, including columns for id with Model, BuiltYear, SellerType, and OwnerType.

Fig: Created Car Seller dimension table

After performing the above function, the duplicates were removed from the remaining columns as shown below.

The screenshot shows the Power Query Editor interface with the 'Car Seller' query selected. Two columns, 'SellerType' and 'OwnerType', are highlighted with yellow boxes. A context menu is open over these columns, with the 'Remove Duplicates' option highlighted and boxed in red. The 'Applied Steps' pane on the right shows various transformation steps taken on the query.

Fig: Removing Duplicates from Car Seller table

c. Creating Car Type with Price Dimension Table

The same procedure was followed to create 2 more-dimension tables where the Indian_Car_Sales_2021 was duplicated and renamed to Car Type with Price.

Leaving the following columns like Car Brand with Model, TransmissionType and FuelType, Seats and SellingPrice, all the other columns were deleted. Also, the duplicates were removed. The image is shown below for reference.

The screenshot shows the Power Query Editor interface with the 'Car Type with Price' query selected. Several columns are highlighted with yellow boxes, including 'Mileage_kmpl', 'Engine_cc', 'MaxPower_bhp', 'Seats', and 'SellingPrice'. A context menu is open over these columns, with the 'Remove Columns' option highlighted and boxed in red. The 'Applied Steps' pane on the right shows various transformation steps taken on the query.

Fig: Created Car Type with Price dimension table

The screenshot shows the Power Query Editor interface with the 'Car Type with Price' query selected. A context menu is open over three columns: 'FuelType', 'Transmission Type', and 'Seats'. The 'Remove Duplicates' option is highlighted in red. The 'APPLIED STEPS' pane on the right shows various transformations applied to the query.

Fig: Removing Duplicates in Car Type with Price dimension table

d. Creating Car Features Dimension Table

Like the above, the same steps were followed to create Car Features dimension table neglecting the other columns. The following columns were part of the Car Features: Car Brand with Model, Mileage_kmpl, MaxPower_bhp, Engine_cc.

The screenshot shows the Power Query Editor interface with the 'Car Feature' query selected. A context menu is open over three columns: 'Mileage_kmpl', 'Engine_cc', and 'MaxPower'. The 'Remove Other Columns' option is highlighted in red. The 'APPLIED STEPS' pane on the right shows various transformations applied to the query.

Fig: Created Car Features dimension table

The screenshot shows the Power Query Editor interface with the 'Car Feature' query selected. A context menu is open over the 'Car Feature' column, with the 'Remove Duplicates' option highlighted. The 'STEPS' pane on the right shows the history of operations, including 'Removed Other Columns'.

Fig: Removing Duplicates from Car Features table

e. Creating Indian Used Car Sales 2021 Fact Table

After creating dimension tables for Car Info, Car Type with Price, Car Features, Car Seller and Car Feature, the unused or unwanted fields and columns were removed from the parent table.

The screenshot shows the Power Query Editor interface with the 'Indian_Car_Sales_2021' query selected. A context menu is open over the 'Transmission Type' column, with the 'Remove Columns' option highlighted. The 'STEPS' pane on the right shows the history of operations, including 'Expanded Car Feature'.

Fig: Removing Columns from Fact Table

Creation of Primary and Secondary key in Fact with Dimension Table

Generally, the fact table does not store any information relating to any events, measure or perform any function. The table only acts as a mapping bridge between other dimension tables.

In this dataset, the unique values were not available in ‘Car Feature’ Table to create relationship between the parent to dimension table. This issue will create ‘many to many’ relationships in data modelling.

A ‘many-to-many’ relationship isn't the best form of connection to have. In respect to the column field, Car Feature relationship can only be formed as ‘many to many’ relationship. As the single product database does contain a list of unique product values but can't be used as a source of ‘one-to-many’ relationships.

In this scenario, the dimension table columns (in fact table) were merged to single index column as shown in the following steps.

Step 1:

The index column is created in ‘Car Features’ dimension table by Clicking Add Column with Index Column from 1 as shown below.

The screenshot shows the Power Query Editor interface. The ribbon at the top has 'Add Column' highlighted. A red box highlights the 'Index Column' button in the 'General' section of the ribbon. Another red box highlights the 'From 1' button in the dropdown menu for 'Index Column'. The main area shows a table with columns: id with Model, L2_Mileage_kmpl, Engine_cc, L2_Mileage_kmpl_bhp, and Index. The 'Index' column is the newly added column. The 'APPLIED STEPS' pane on the right shows the history of changes, with 'Added Index' being the last step.

Fig: Created Index Column

Step 2:

In the fact table (Indian_Car_Sales_2021) ‘merge queries’ option was chosen and the table Indian_Car_Sales_2021 and Car Feature was selected, and the columns need to be merged as shown in the image below

The screenshot shows the Power Query Editor interface with the 'Merge' dialog open. The 'Merge' dialog has two tables selected: 'Indian_Car_Sales_2021' and 'Car Feature'. In the 'Car Feature' table, the 'Index' column is highlighted with a red box. The 'Join Kind' dropdown is set to 'Left Outer (all from first, matching from second)'. A message at the bottom states: 'The selection matches 19581 of 19581 rows from the first table.' The 'OK' button is also highlighted with a red box.

Fig: Merging Columns

Step 3:

After the above step, the merged column ‘Car feature’ was added to the fact table.

The index was selected by unselecting all other column names in Car Feature column in the Fact Table as shown below.

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

- File, Home, Transform, Add Column, View, Tools, Help** menu bar.
- Data Sources**, **Parameters**, **Properties**, **Refresh**, **Preview**, **Manage** buttons.
- Queries [5]** list: Indian_Car_Sales_2021, Car Info, Car Seller, Car Type with Price, Car Feature.
- Table View**: Shows a list of rows with columns: 1, 1.2 Engine_cc, 1.2 MaxPower_bhp, 1.2 Seats, 1.2 SellingPrice, Car Feature.
- Transform ribbon**: Choose Columns, Remove Columns, Keep Rows, Remove Rows, Reduce Rows, Sort.
- Properties pane**: Name: Indian_Car_Sales_2021, All Properties.
- Applied Steps pane** (highlighted with a red box):
 - Removed Columns1
 - Duplicated Column1
 - Split Column by Delimiter
 - Changed Type1
 - Replaced Value9
 - Replaced Value10
 - Replaced Value11
 - Replaced Value12
 - Changed Type2
 - Added Custom
 - Changed Type3
 - Reordered Columns
 - Removed Columns2
 - Changed Type4
 - Renamed Columns1
 - Added Index
 - Renamed Columns2
 - Reordered Columns1
 - Merged Queries
- Query Settings pane**: Preview downloaded at 22:56.

Fig: Expanding the Car Feature

After successfully splitting the Fact and Dimension table, Close & Apply option was chosen to save the work.

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

- File, Home, Help** menu bar.
- Data** ribbon: Get data, Excel, Power BI, SQL, Enter data, Dataverse, Recent sources.
- Home ribbon**: Transform data, Manage relationships, Q&A, Language setup, Sensitivity, Publish.
- Load dialog box** (highlighted with a red box):
 - Car seller details: 19,311 rows loaded.
 - Car_info: 18,288 rows loaded.
 - Car_seller_details: 4,683 rows loaded.
 - Car_Type: 3,185 rows loaded.
 - Car_Features: 3,422 rows loaded.
- Fields pane** on the right: Search, Car_Features, Car_Info, Car_Price, Car_seller_details, Car_Type, Cardhekho_updated.
- Table view** at the bottom: All tables, +.

Fig: Save all the Tables created

DAX Expression

The full form of DAX is Data Analysis Expression Language. It is a language developed by Microsoft to interact with data in a variety of platforms like Power BI. DAX includes a library of functions and operators that can be used to build formulas, expression and create pivot tables.

DAX helps us to create new information from data which is already in our model. Using DAX, we can add three types of calculations to our data model:

1. Calculate tables
2. Calculate columns
3. Measures

The screenshot shows the Power BI Desktop interface with the title bar 'new one - Power BI Desktop'. The ribbon at the top has tabs 'File', 'Home', 'Help', and 'Table tools' (which is highlighted with a yellow background). Below the ribbon, there's a search bar and a 'Fields' pane on the right containing a tree view of data sources: 'Car Feature', 'Car Info', 'Car Seller', 'Car.Type with Price', and 'Indian_Car_Sales_2021'. The main area displays a table titled 'Car Brand with Model' with columns: FuelType, Transmission Type, Seats, and SellingPrice. A red box highlights the first row of the table. A second red box highlights the 'New column' icon in the 'Calculations' group of the 'Table tools' ribbon. The status bar at the bottom says 'Table: Car Type with Price (15,890 rows)'.

Fig: Types of calculation using DAX

a. New Column to Find Car Age using DAX

To find the age of the car, The DAX expression has been used. The manufacturing year of the car is subtracted from the current year. To perform the operation, Car Info table was clicked and under ‘Table tools’ option, ‘New column’ option is chosen as shown in the below.

New Column name = 2021 – Tablename (column name) – syntax

Car Age = 2021 – car Info [BuiltYear]

new one - Power BI Desktop

File Home Help Table tools Column tools

Summarization Sum Data category Uncategorized Sort by column Data groups Manage relationships New column

Struct Car ID BuiltYear KM Driven Car Brand with Model Car Age

16 2016 50000 Renault Duster 110PS Diesel RxL Explore 5
228 2016 50000 Hyundai i20 Active 1.4 SX 5
337 2016 50000 Maruti Alto 800 LXI 5
454 2016 50000 Hyundai Xcent 1.2 Kappa SX Option 5
569 2016 50000 Hyundai i20 2015-2017 Magna 1.2 5
596 2016 50000 Toyota Fortuner 2016-2021 2.7 2WD AT 5
659 2016 50000 Renault KWID RXT 5
685 2016 50000 Maruti Alto 800 LXI 5
724 2016 50000 Maruti Vitara Brezza ZDI Plus Dual Tone 5
735 2016 50000 Mahindra Scorpio 1.99 S5 Plus 5
1393 2016 50000 Maruti Alto 800 LXI 5
1466 2016 50000 Datsun GO T Option BSIV 5
1627 2016 50000 Tata Safari Storme VX 5
1696 2016 50000 Fiat Linea Classic 1.4 Petrol 5
1704 2016 50000 BMW X1 xDrive 20d M Sport 5
1738 2016 50000 Mercedes-Benz GLA Class 200 Sport 5
1796 2016 50000 Hyundai i20 Magna 1.4 CRDi 5
1817 2016 50000 Maruti Swift Dzire VDI 5
1993 2016 50000 Maruti Swift Dzire VXI 5
2215 2016 50000 Hyundai Xcent 1.2 Kappa S 5
2273 2016 50000 Hyundai i20 2015-2017 Asta Option 1.2 5
2616 2016 50000 Renault Duster 85PS Diesel RxZ 5
2865 2016 50000 Datsun GO Plus T BSIV 5

Table: Car Info (19,163 rows) Column: Car Age (26 distinct values)

Fig: Creating Calculated Column using DAX

After the calculation, the age of each car will be displayed in Car Age column as shown in the above image.

b. New Measure to Find Average Sold out price per Car:

For calculating the average sale of car price from ‘Car Type with Price’ Table, new measure options were used as shown in the below image.

new one - Power BI Desktop

File Home Help Table tools Measure tools

Format General Data category Uncategorized

Struct Car Brand with Model FuelType Transmission Type Seats SellingPrice

Maruti Wagon R VXI BS IV Petrol Manual 5 350000
Datsun RediGO Gold 1.0 Petrol Manual 5 350000
Maruti Wagon R VXI Petrol Manual 5 350000
Honda City 2017-2020 VTEC Petrol Manual 5 350000
Hyundai Xcent 1.2 Kappa SX Option Petrol Manual 5 350000
Renault Scala RxL Petrol Manual 5 350000
Hyundai i10 Sportz Petrol Manual 5 350000
Hyundai Grand i10 Magna Petrol Manual 5 350000
Maruti Swift VXI Petrol Manual 5 350000
Renault KWID 1.0 RXT Optional Petrol Manual 5 350000
Maruti Alto 800 LXI Petrol Manual 5 350000
Maruti Swift Dzire 1.2 Lxi BSIV Petrol Manual 5 350000
Skoda Superb 1.8 TFSI MT Petrol Manual 5 350000
Honda City 1.5 S MT Petrol Manual 5 350000
Maruti Wagon R Stingray LXI Petrol Manual 5 350000
Maruti Alto K10 VXI Petrol Manual 5 350000
Tata Zest Revotron 1.2T XE Petrol Manual 5 350000
Hyundai i20 Magna Petrol Manual 5 350000
Volkswagen Polo 1.2 MPI Trendline Petrol Manual 5 350000
Maruti Swift Dzire VXI Petrol Manual 5 350000
Maruti Wagon R VXI Optional Petrol Manual 5 350000
Maruti Celerio VXI Petrol Manual 5 350000
Hyundai Grand i10 Sportz Petrol Manual 5 350000

Table: Car Type with Price (15,890 rows) Column: Average Selling Price (0 distinct values)

Fig: Calculation of ‘Average price’ per car using ‘New measure’

To calculate average sale of a car price we used the ‘New Measure’ option and the following formula:

Average Selling Price = AVERAGE (Car Type with Price[Selling Price])

Using the formula, the average price per car were calculated for the year 2021 car sales (current year).

c. Quick Measure to Find total number of Manual / Automatic Car

Using this method, it was easy to calculate the total number of ‘Automatic transmission’ and ‘Manual transmission’ cars sold in 2021 from Car Type with Price dimension table.

In Quick measure option, the ‘filtered value’ was selected from the drop-down list of ‘calculation’ field. Then the column (Car Brand with Model) was dragged from the table on the right side to give base value and then Transmission Type Column is used for Filter Option. Finally, the ‘Automatic’ option was chosen. Upon which ‘ok’ options were selected as shown below.

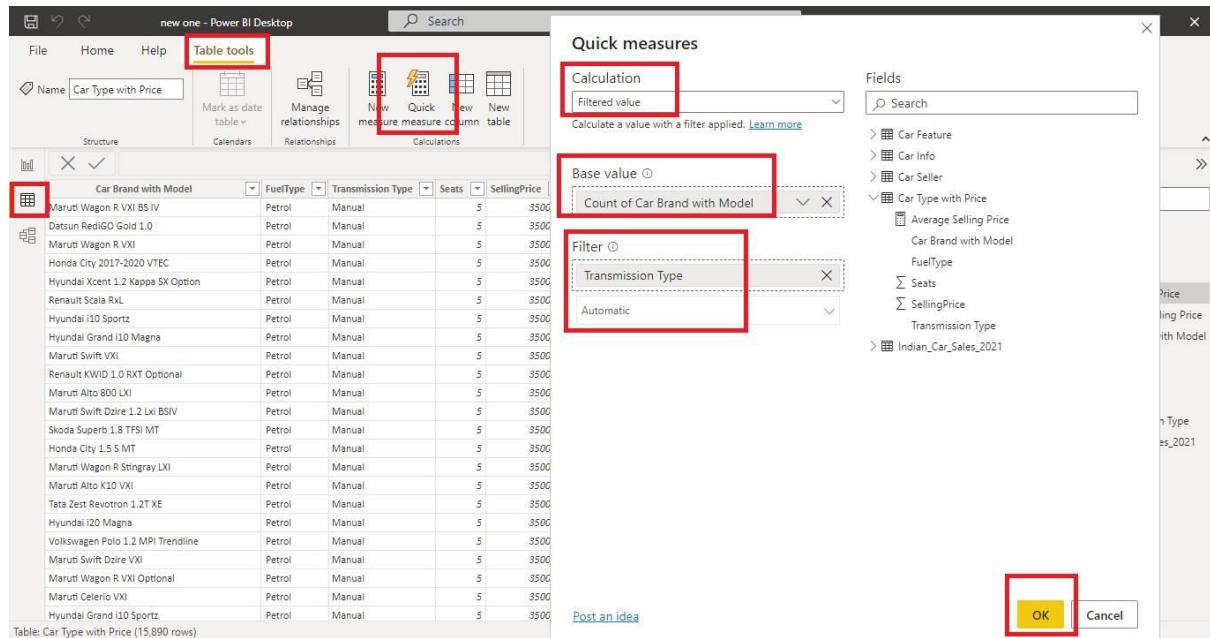


Fig: Quick Measure for Automatic Transmission Type

The same step was repeated by selecting ‘Manual transmission’ type in filtered values to find the total number of cars sold in manual transmission type.

After the above step is performed for calculating the count in both and ‘Manual’ and ‘Automatic’ transmission, it was added to the data set table as shown in the below image.

The screenshot shows the Power BI Desktop interface with the 'Table tools' tab selected. A 'Quick measures' dialog box is open, with the 'Calculation' field set to 'Filtered value'. Below it, the 'Base value' is 'Count of Car Brand with Model' and the 'Filter' is 'Transmission Type' set to 'Manual'. The Fields pane on the right lists various car-related dimensions and measures, including 'Car Type with Price', 'FuelType', 'SellingPrice', and 'Total No of Automatic Car SoldOut'. The 'OK' button at the bottom right of the dialog box is highlighted with a red box.

Fig: Quick Measure for Manual Transmission Type

After the quick measure calculation, the measure of automatic and manual type was added in Car Type with price table as shown below.

The screenshot shows the Power BI Desktop interface with the 'Measure tools' tab selected. In the 'Properties' section, a new measure named 'Total No of Manual Car SoldOut' has been created. The Fields pane on the right shows the measure has been added under the 'Car Type with Price' dimension. The table view shows the same data as before, with the new measure included in the columns.

Fig: Quick Measure Added in Car Type with Price Table

Creating Data Model and relationships

Following the segregation of flat table into fact and dimension tables, it is essential to link the tables together and establish relationships between them. To accomplish this, Power BI has a sophisticated capability called 'Power BI Data Modeling.'

The 'Model' option was chosen from the left side of the window to construct a Data Model, as illustrated in the figure below.

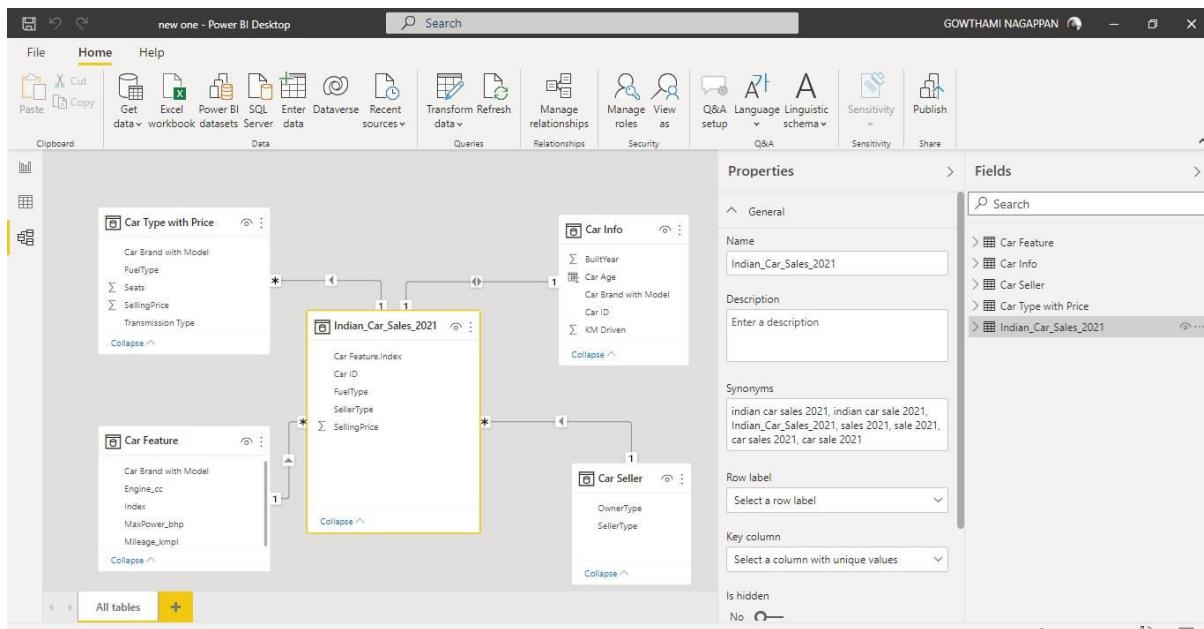


Fig: Data Model View - Fact and Dimension Tables

Power BI has the feature to automatically detect the relationships between the tables as shown above.

From the above image, it is visible that all the tables were arranged with the fact table (Indian_Car_Sales_2021) at the center and other dimension tables surrounding this fact table like '**Star Schema**'.

To create Relationship Manually, 'Manage relationships' option in tool bar was used as shown below.

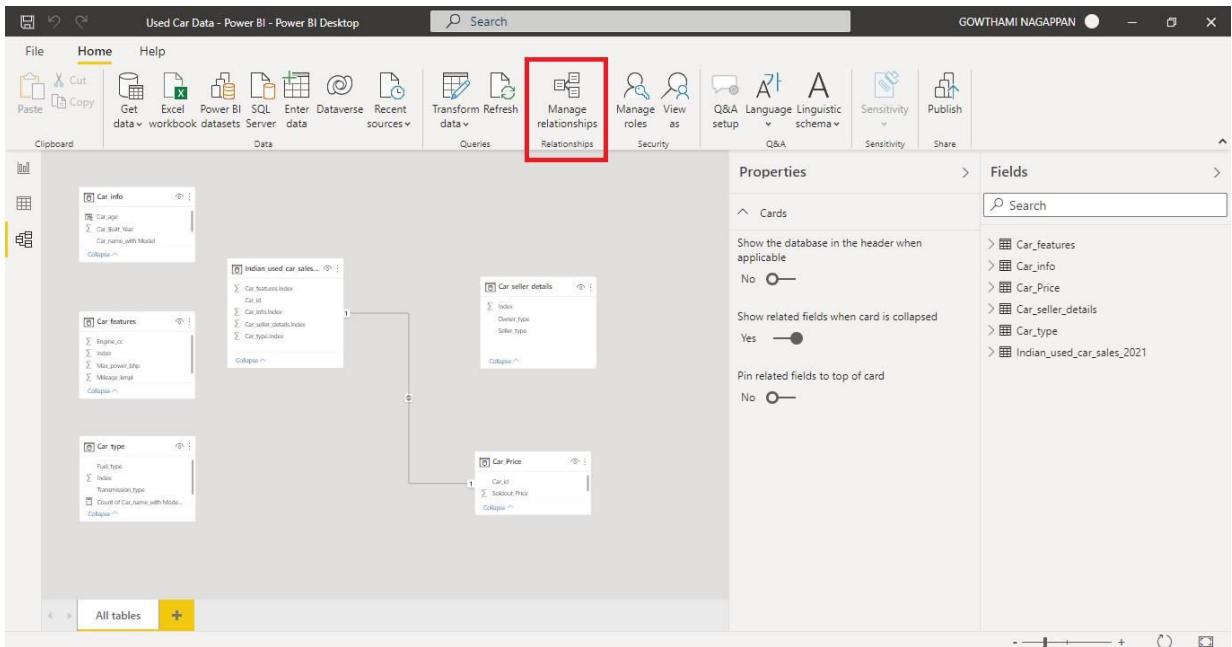


Fig: Managing Relationship

To create new relationship manually, new button option was chosen and the ‘Edit relationship’ option was visible where the tables are selected, and relationships were created as shown below.

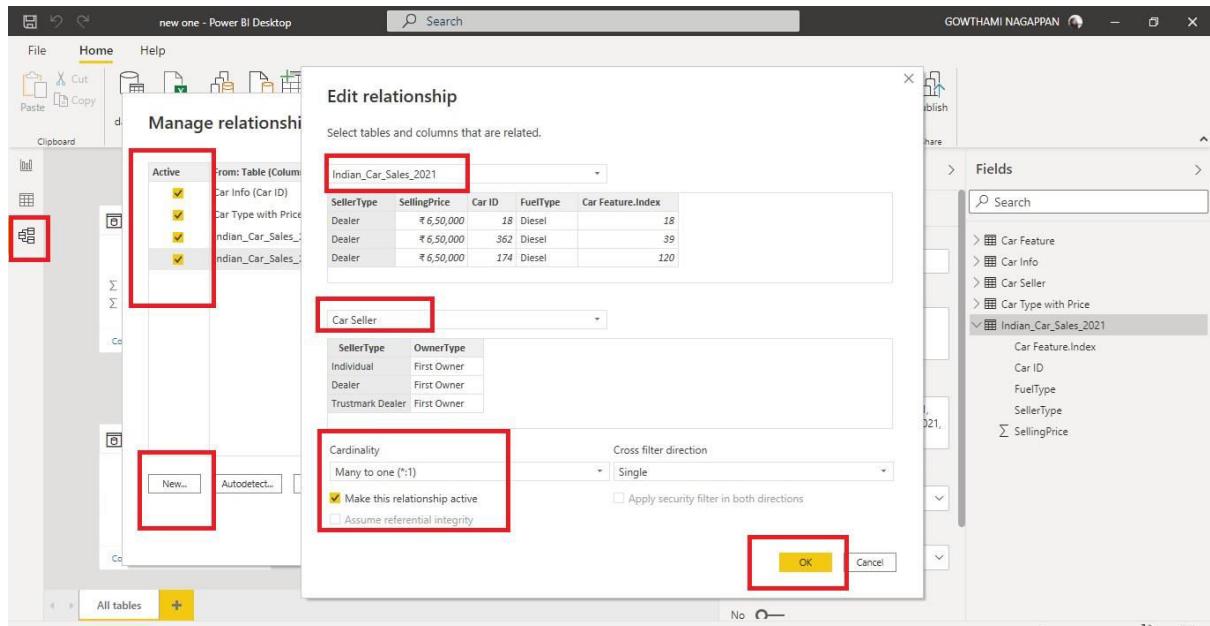


Fig: Creating New relationship

Then we can add tables which needs relationship to be created as above image.

After creation of all the relationship between tables, the ‘manage relationship’ dialogue box showed on how many relationships were active. It’s important that our manually / automatically created relationship should be active and click done as shown in the image below.

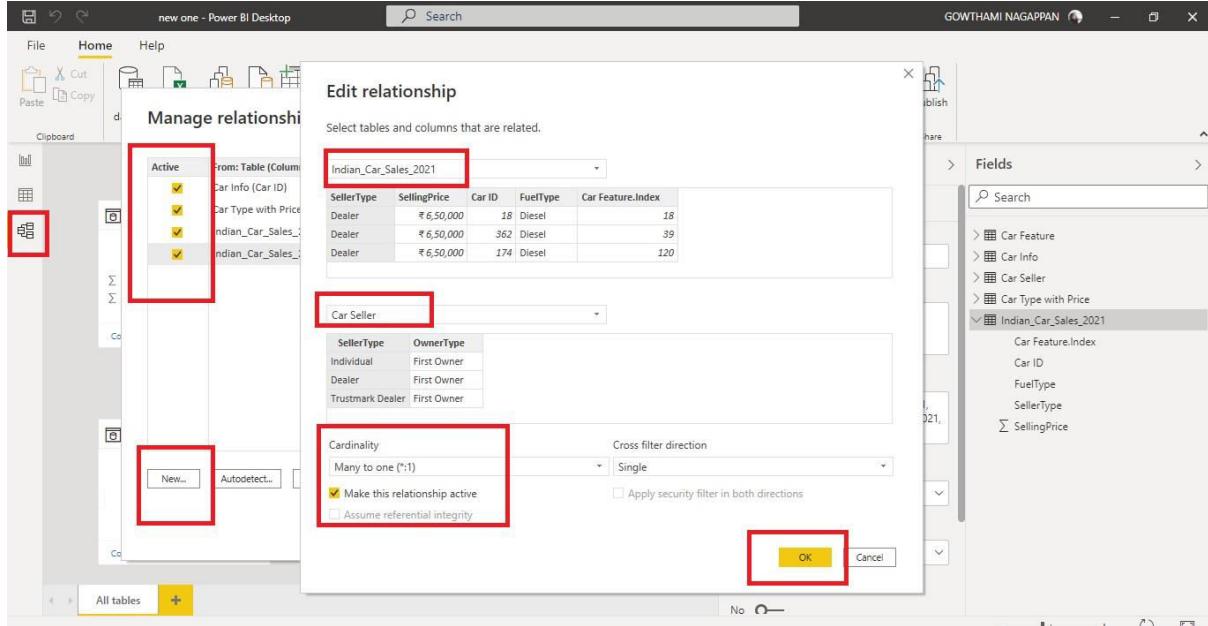


Fig: Active Relationship

Until we establish the connection, the data model replicates **Star Schema** with Dimension tables and Fact less fact table.

The final data model with relationships is shown in the figure below.

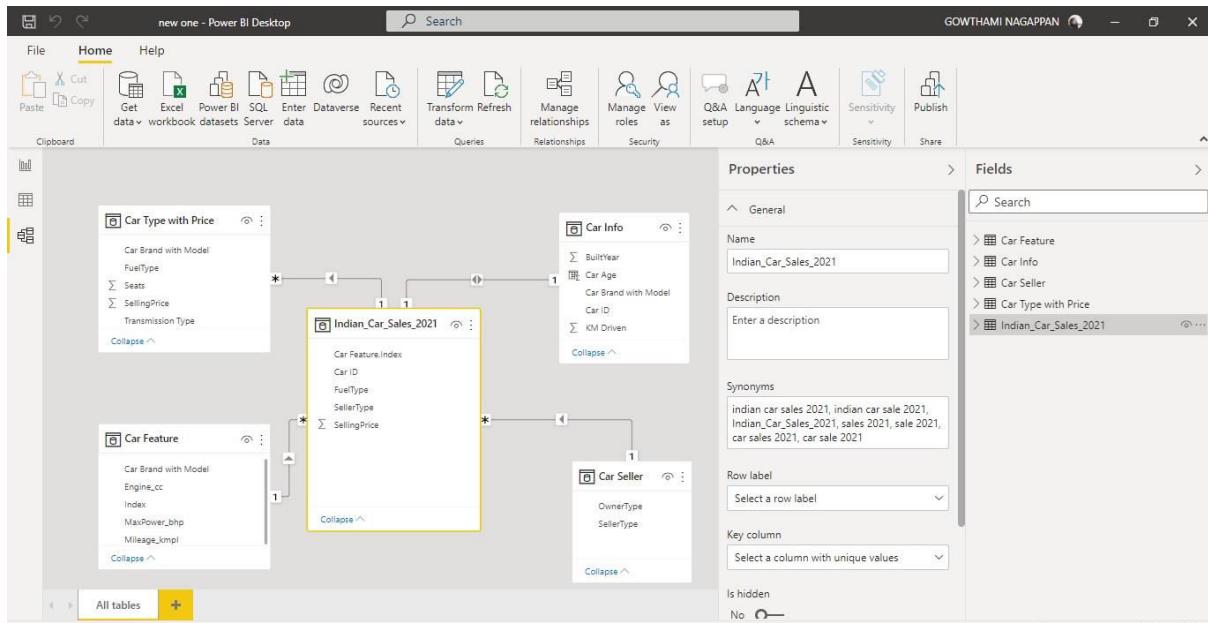


Fig: Fact and dimension tables with established relationships



BUSINESS INTELLIGENCE FOR USED CAR SALES ANALYSIS

CIS4008-N-BF1-2021

BIG DATA AND BUSINESS INTELLIGENCE

**SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL
TECHNOLOGIES**

Section 2



JANUARY 11, 2022

Prepared by

B1326237 GOWTHAMI NAGAPPAN

Executive summary

Introduction

Cars are the efficient ways of transportation where a person or family can travel from one place to the other at ease. It is also a sign of pride among the community if a person has a car. This has led to the increased sale of used cars in the market. Starting with the traditional players like the one situated at the corner of the road with cars stored at their site the customer can visit and view the cars, the segment has moved forward to the e-commerce space which is considered the easiest and cheapest way to buy and sell a car as there is no requirement for a warehouse which reduces the operation cost. There are many used car e-commerce companies operating on the segment and some of the companies which has grown from being small to get listed in the stock markets which shows a tremendous potential. In this report, we are analyzing and providing recommendation for the used car commerce giant ‘e-commerce’ on where the concentration can be done to increase the sale and profits for the firm. More companies are investing on technology which includes analytics ad data visualization, our report would highly benefit the company.

Background

There has been an increased demand of used car sales in India and the sector has seen a growth rate in the range of 12-15%. Due to higher disposable income among the lower communities including the middle-class population and farmer communities, the trade in the sector is rising at faster pace than before. Even the big multi-national car companies in India like Toyota, Hyundai, Volkswagen, Maruti Suzuki, Mahindra and Tata motors have entered in the used car segment which has been the long players in selling only the new cars. Also, there are several companies which operate on e-commerce space by allowing the users to buy and sell the used cars through their platform. The leading companies operating in this segment are CarTrade, Cardekho, Cars 24 and car wale are few of the top firms operating in this space. There is not a major difference between their business model and so there is a requirement for these companies to show a varied business model in comparison with their competitors. So, we are using our analysis to provide recommendations on how the business can increase their sales, drivers for listing the cars, segments to concentrate and luring the customers to post the cars for sale on the platform. Because of the rapid increase to data-driven insights which are used by most of the firms, they could provide with a valuable resource that will be a boon to the companies.

Key Findings

- Highest selling car brand is Maruti and the model ‘Swift Dzire VDI’ sells the most in the used car segment.
- Highest selling car model when compared with the mileage is ‘Maruti Swift AMT DDiS VDI’ and ‘Maruti Swift Dzire AGS ZDI plus’.
- Highest selling car model in the lower mileage segment is ‘BMW 6 series 650i Coupe’ and ‘Mercedes-Benz S-class S 500’.
- Number of petrol cars sold are more than the diesel, electric and LPG cars.
- Number of Cars sales based manufactured (built year) is high in recent built compared to old years built. specifically, the cars manufactured and registered in 2015, 2016 and 2017 sell the most in 2021.
- Manual transmission cars moving quickly in both fuel types, Petrol and Diesel.
- In relation to Seat variant, 5-seater cars have higher demand compared with another seat variant.
- Number of cars manufactured in the year 2017 & 2018 were sold more which had the run of more than 100,000 Kilometers.
- In relation to the seller type, number of cars sold by the Dealers are more than the Individual and Trust mark dealer.

Recommendation

- Firm should concentrate more on highest selling car brand and car model. These cars should be displayed on the website on priority basis when a customer is viewing the website.
- Firm must give preference for the manual transmission type cars instead of automatic version.
- Preference should be given for dealers when compared with individuals and trust mark dealers. Incentives can be given to the dealers in the form of listing prices reductions so more number of cars can be listed.
- Cars manufactured in the last 5 years should be on priority listing compared with the old model cars.
- Petrol car listing price can be reduced when compared to diesel and other fuel type which will lure more customers on the website.
- Out of the box thinking and researching onto various ecommerce websites, trust on the car is a main inducer for the customers to buy a particular car and the firm can partner with local quality check agencies to provide a report and so the same can be attached as a document while listing on the website.

The primary objective of this report is to analyze the number of cars sold through the platform, sale of cars according to fuel type, mileage, car model with high sale, car manufacturing year, transmission type, seat variants, number of owners and type of seller.

Final Dashboard Display after the analysis of sales and various factors in Used car sales.

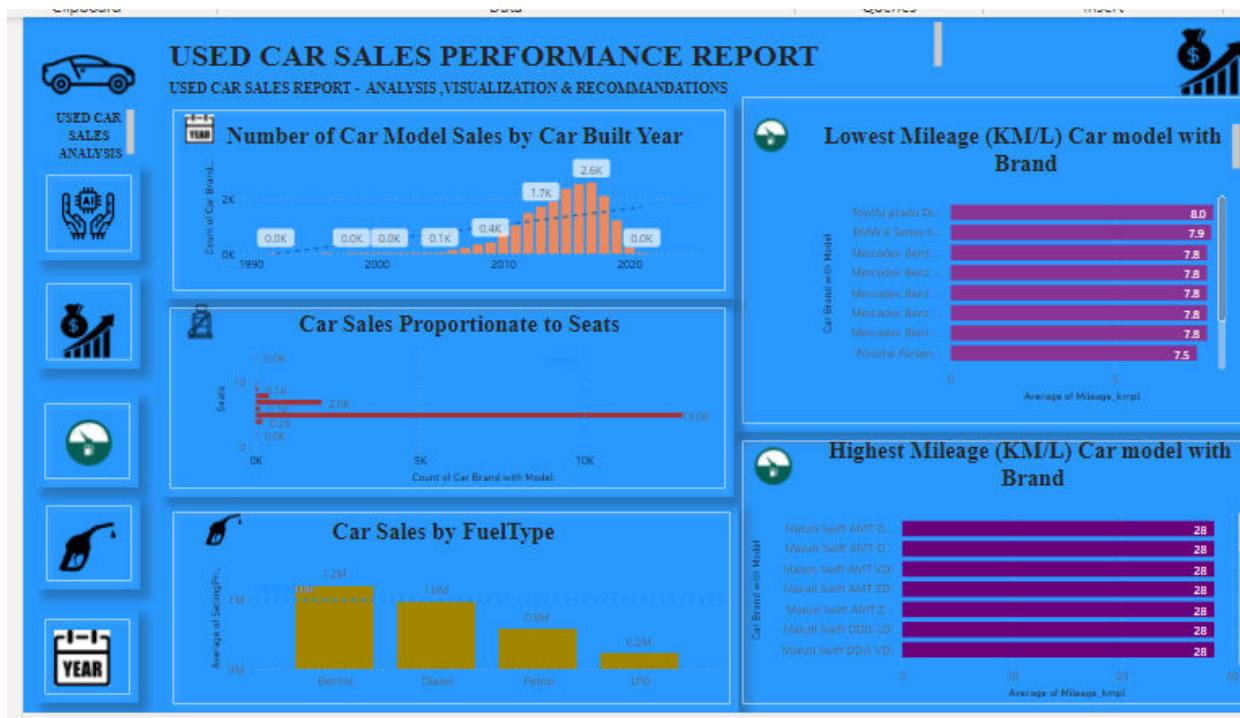


Fig: Sales Based on Features Dashboard 1



Fig: Sales Performance Dashboard 2

Analyzing the data against one with the other has given valuable results which is provided as an input for the companies in the way of key findings in above dashboards.



Fig: AI Feature Dashboard 3

Dataset

For this business intelligence report, the dataset from Kaggle.com (cardekho_updated.csv) was selected. Further details of these datasets are explained in section 1.

Data Model

As discussed in section 1, the flat file was divided into various dimension and a fact table. The data model and relationships between the tables were created using Power BI tool. The image below shows the final data model for selected datasets.

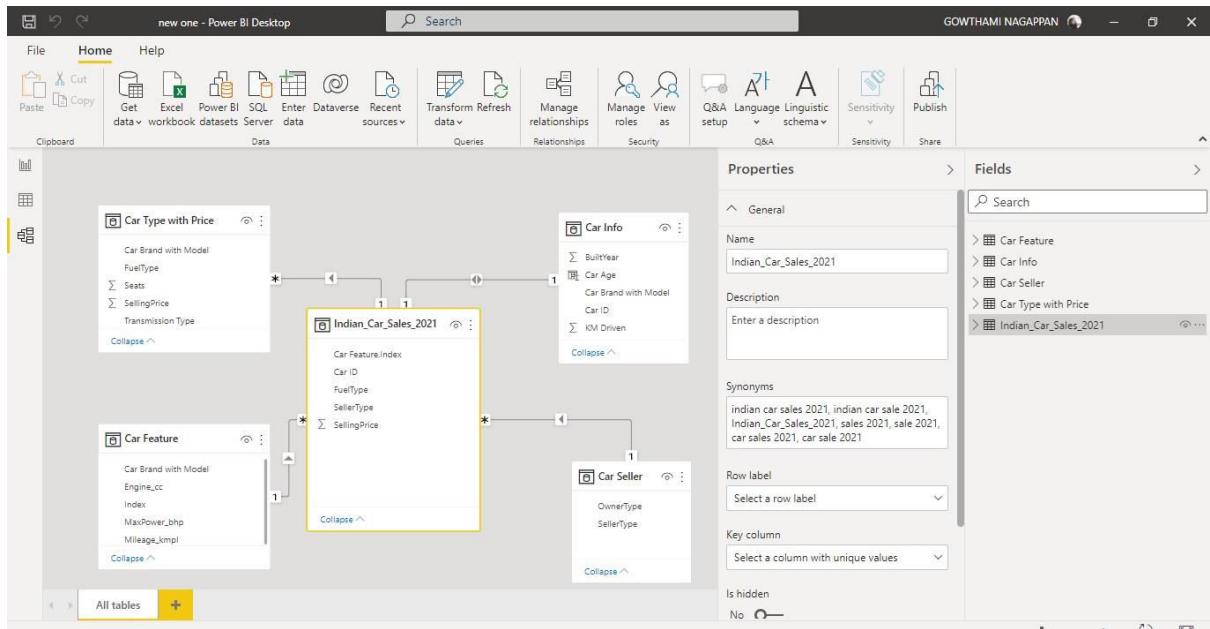


Fig: Fact and Dimension Table with Relationship

Findings based on analysis and evaluation

Business Report was created to visualize and analyze multiple Business Intelligence questions. Various graphs and tables were created using Power BI Visualization tools. The report provides answer to the complex questions which can provide various insights for the business improvement.

- Which is the highest selling car model?
- Which car model sale is high when compared with high / low mileage?
- How many cars are sold for each Fuel Type?
- Which car (manufacturing year) sold the most?
- Which fuel type car sale is higher when compared with the transmission (Manual / Automatic) type?
- Which seat variant (number of seats) sale is higher?
- How many cars were sold for a particular year that has run for more than 100,000 Kilometers?
- How many numbers of car models were sold according to seller type?

a. Highest selling car model when compared with different brands

To display the highest selling car model, ‘Clustered Column Chart’ option was chosen from the Visualization and the Columns were selected from the Car Type with Price table (Car Brand with Model in Axis and Selling Price in Values) field as shown below.

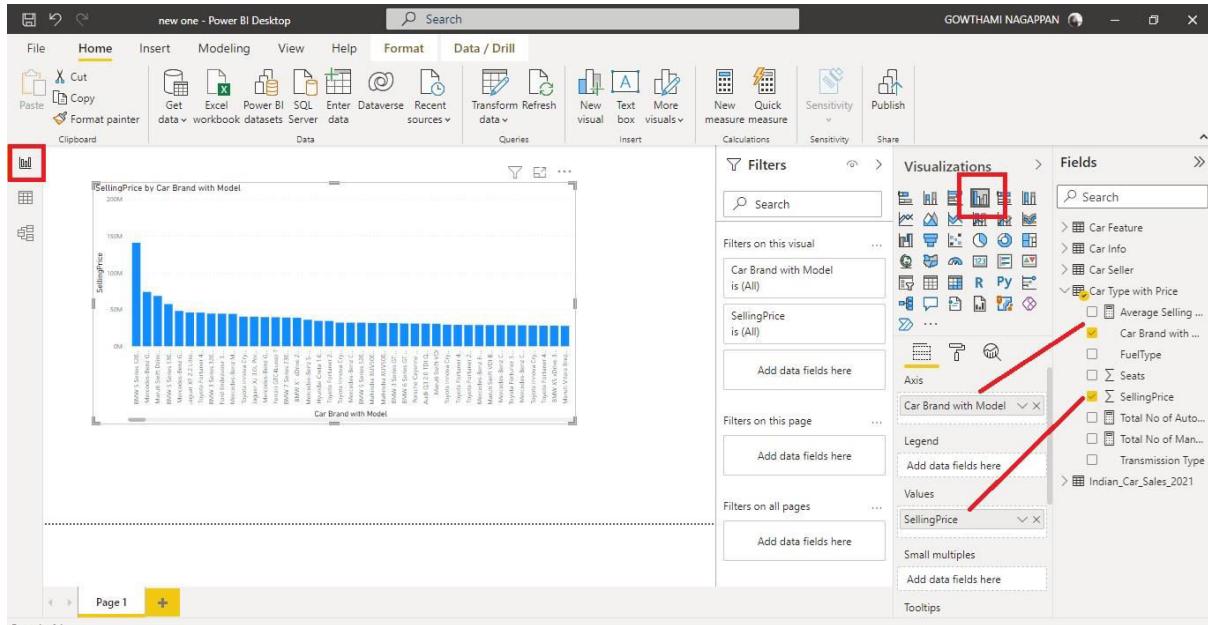


Fig: Clustered Column Chart

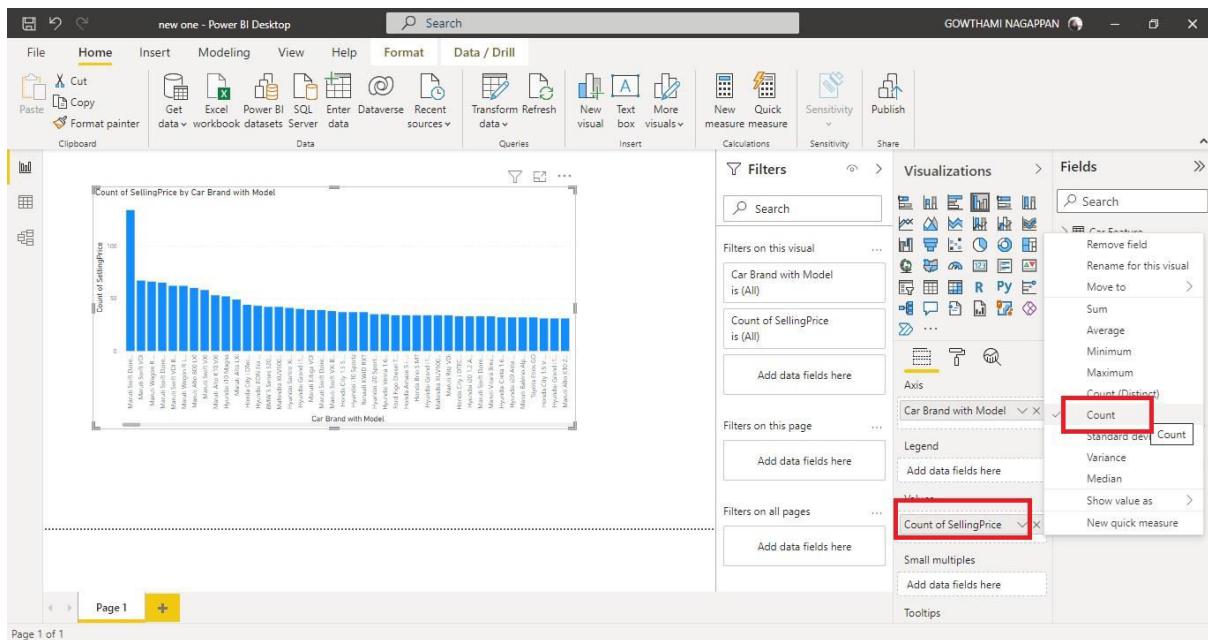


Fig: Changing Value into Count

Format tab was used for editing additional options in the chart which includes Color, Data label, Fonts, Font Size, Title etc. as shown below.

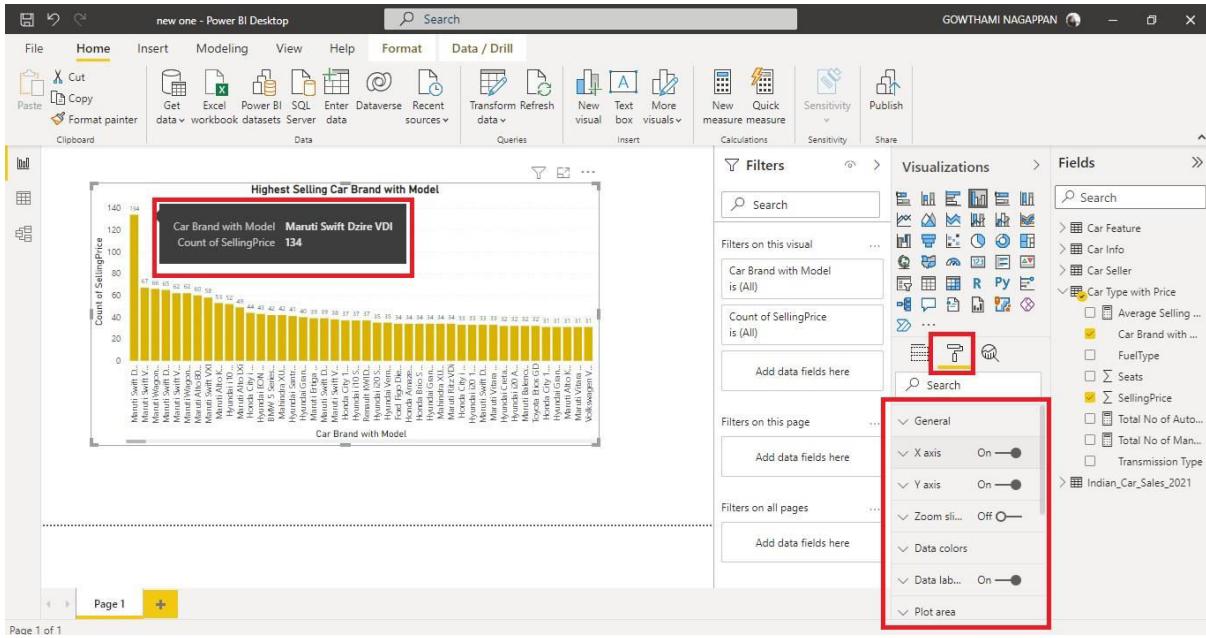


Fig: Format Tab for Editing Chart

Result

- Highest selling Car Model is Maruti Swift Dzire VDI
- Second highest selling Car Model is Maruti Swift VDI and Maruti Wagon R VXI

b. Cars sold with highest and lowest mileage

To display the car sales based on highest mileage (greater than or equal to 28.4 kmpl), ‘stacked bar chart’ was selected and the columns (Car brand with model, Mileage_cc) from Car feature Table were selected and the same was put on axis and values.

The below chart shows the number of car models which provides more than 28 Mileage (Kilometer/ liter of fuel) More than 15 Car brand has the Highest Mileage which can be seen from the below image.

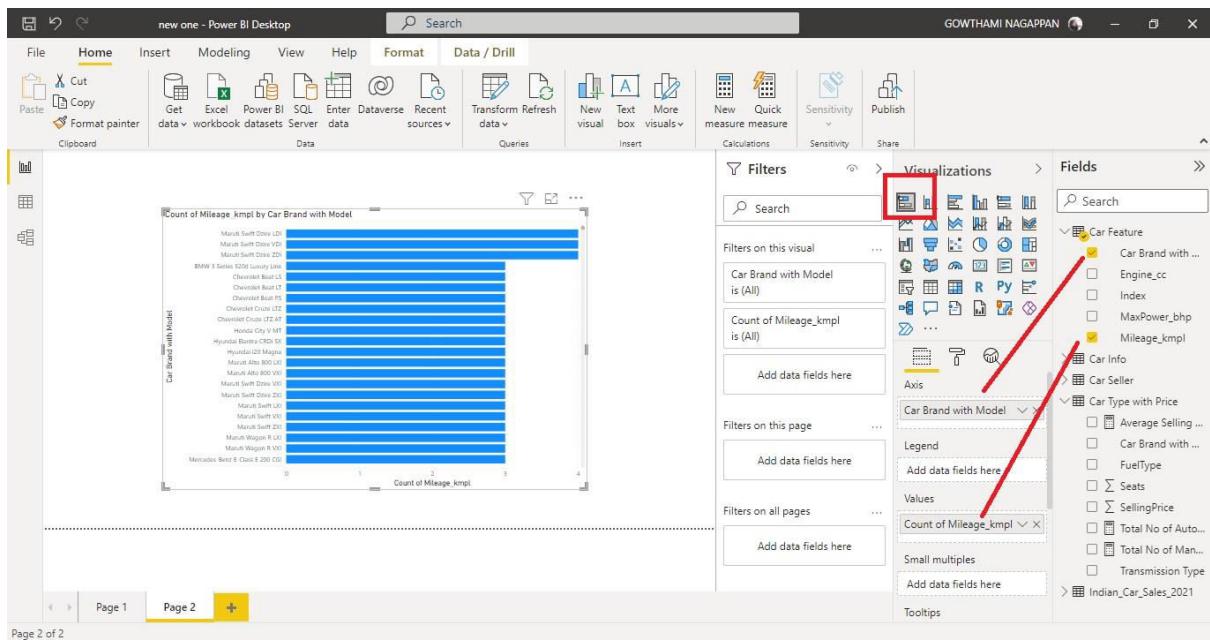


Fig: Stacked Bar Chart

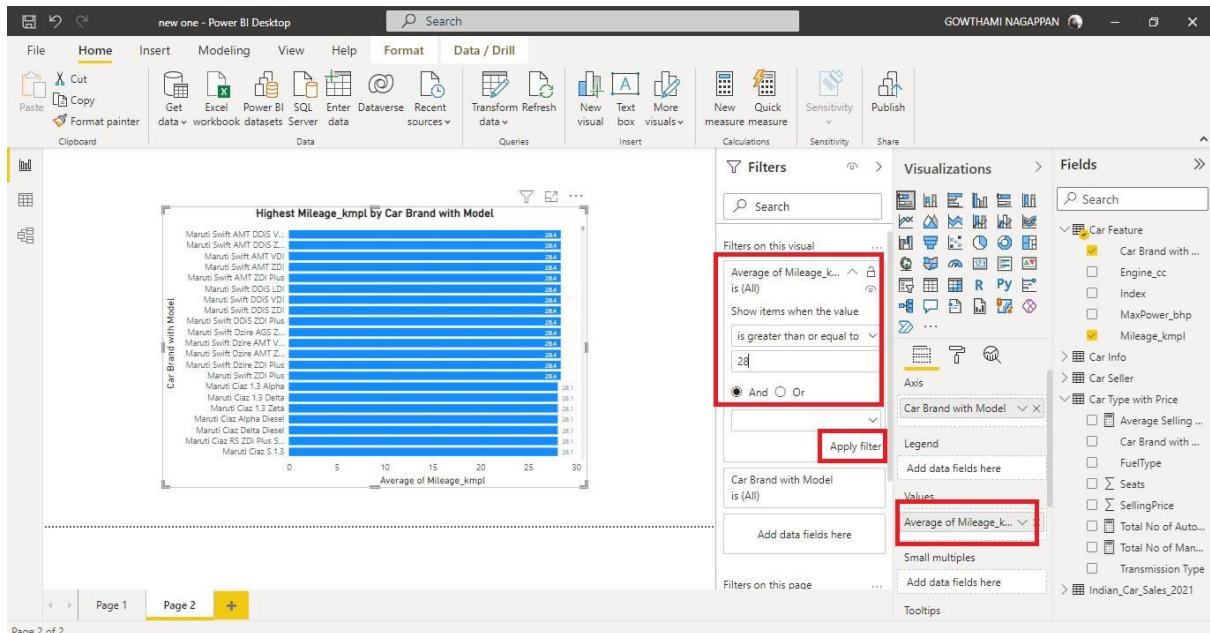


Fig: Apply Filter on Mileage and Format

Meanwhile we have checked the number of lowest mileage car brand with model sold in 2021 by following the above steps (visual, filter and format options). The filter used was less than or equal to 8 kilometer per liter of fuel.

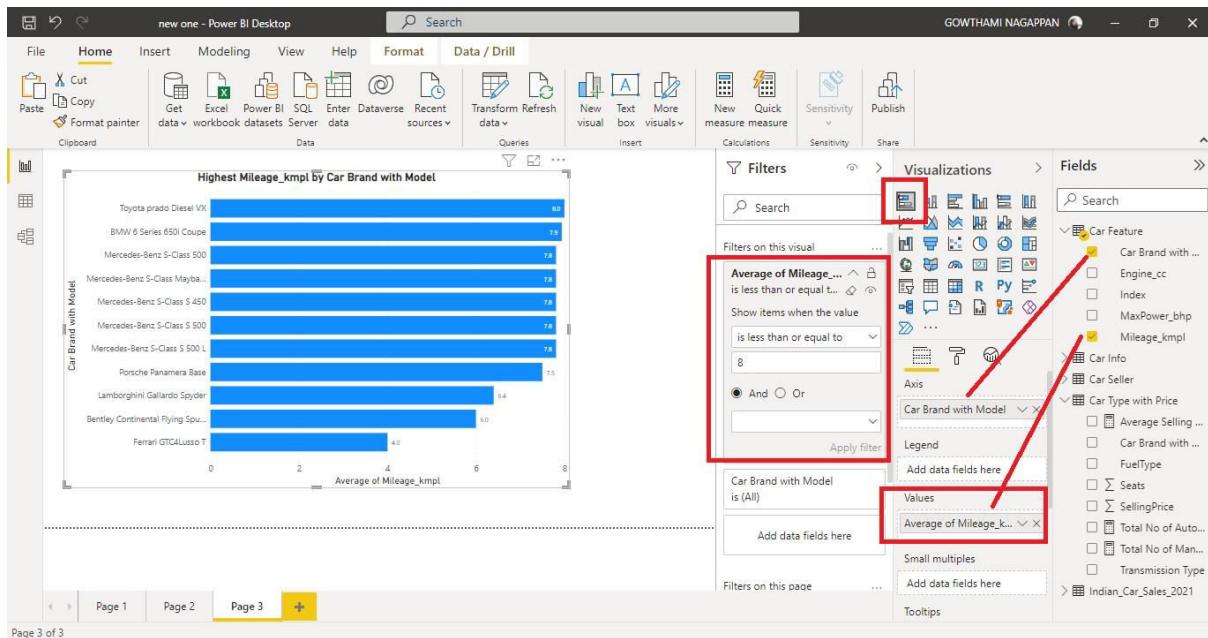


Fig: Stacked Bar Chart for Finding Lowest Mileage Cars

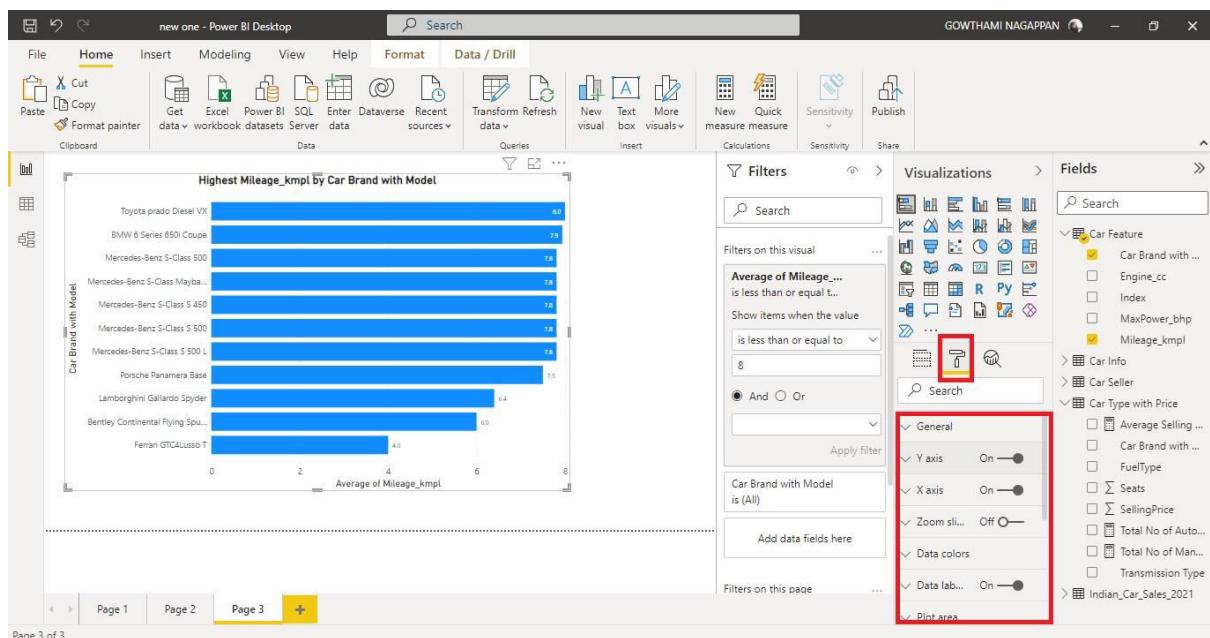


Fig: Stacked Bar Chart for Finding Lowest Mileage Cars

Result

More than 13 cars were sold with a mileage of more than 28 kmpl of fuel. The list of cars are shown below for reference:

Maruti Swift AMT DDiS VDI,
Maruti Swift AMT DDiS ZDI,
Maruti Swift AMT VDI,
Maruti Swift AMT ZDI,
Maruti Swift AMT ZDI plus,
Maruti Swift DDiS LDI,
Maruti Swift DDiS VDI,
Maruti Swift DDiS ZDI plus,
Maruti Swift Dzire AGS ZDI plus,
Maruti Swift Dzire AMT VDI,
Maruti Swift Dzire AMT ZDI plus,
Maruti Swift Dzire ZDI plus,
Maruti Swift ZDI plus

The 6 Models of cars which has the mileage less than 8 kmpl are shown below:

BMW 6 Series 650i Coupe,
Mercedes-Benz S-Class S 500,
Mercedes-Benz S-Class S 500 L,
Porsche Panamera Base,
Porsche Panamera Diesel,
Bentley Continental Flying Spur Speed BSIV

c. Average Selling price of the Car for each Fuel Type

To display the average selling price of car models sold for each Fuel type, ‘Stacked Column Chart’ was chosen by using average selling price of the car and Fuel Type from Car Type with Price Table.

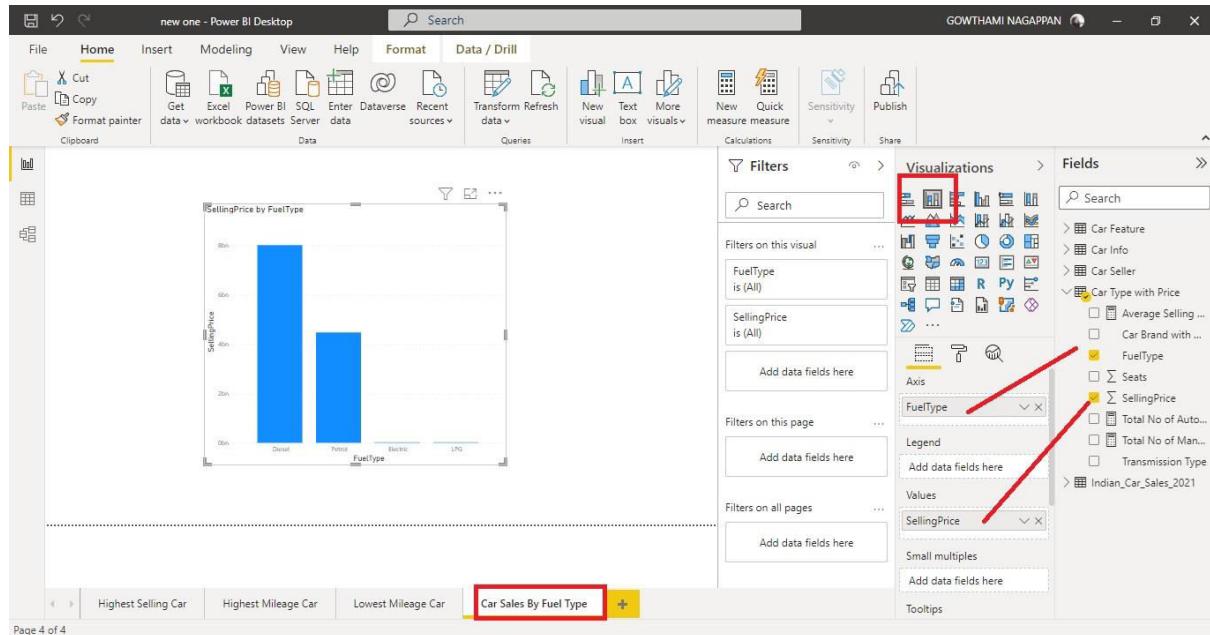


Fig: selling price for each Fuel type using stacked Column Chart

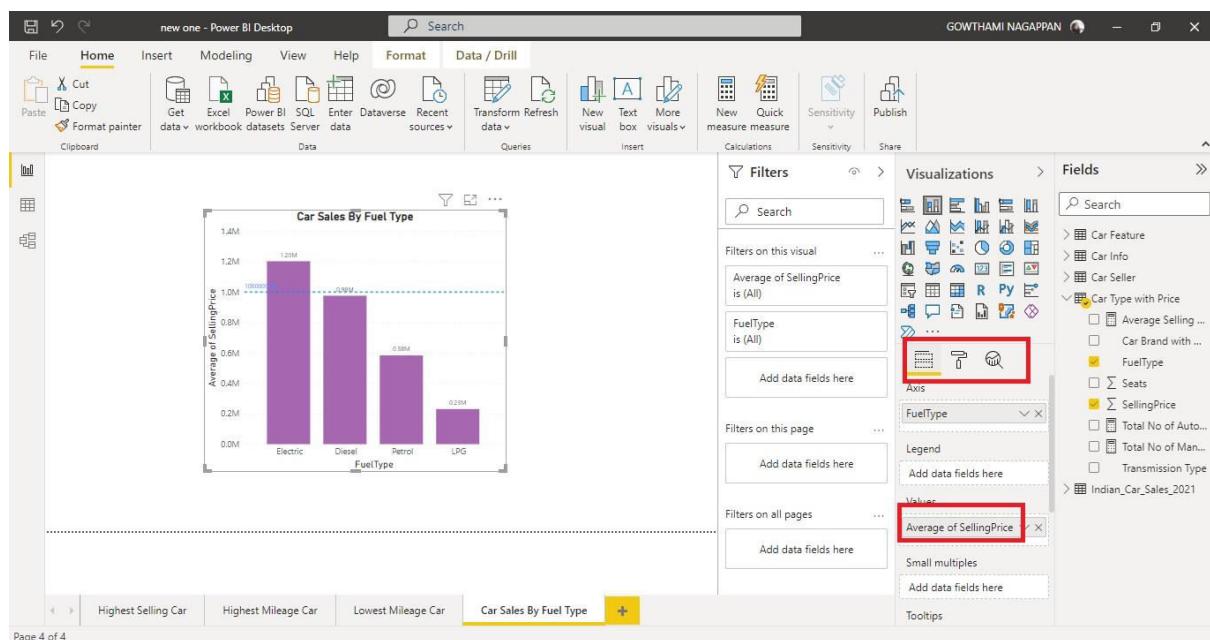


Fig: Average selling price for each Fuel type using stacked Column Chart

As shown in the earlier steps, the title, data label and font size were changed in Format Options.

The average Constantine Line was added to the chart to identify which fuel type car average selling price is higher.

Result

Average selling price of electric cars are higher compared to other fuel type like Diesel, Petrol and LPG.

d. Car sales based on the manufacturing year

The ‘line and Clustered Chart’ displayed the number of cars sold between 1990 to 2021.

The customers are showing interest to buy the cars which are manufactured between the years 2015 – 2020 which shows the customers preference to buy the cars built in the last 5 years.

The below image clearly shows that most of the sales happened during the past 5-year period.

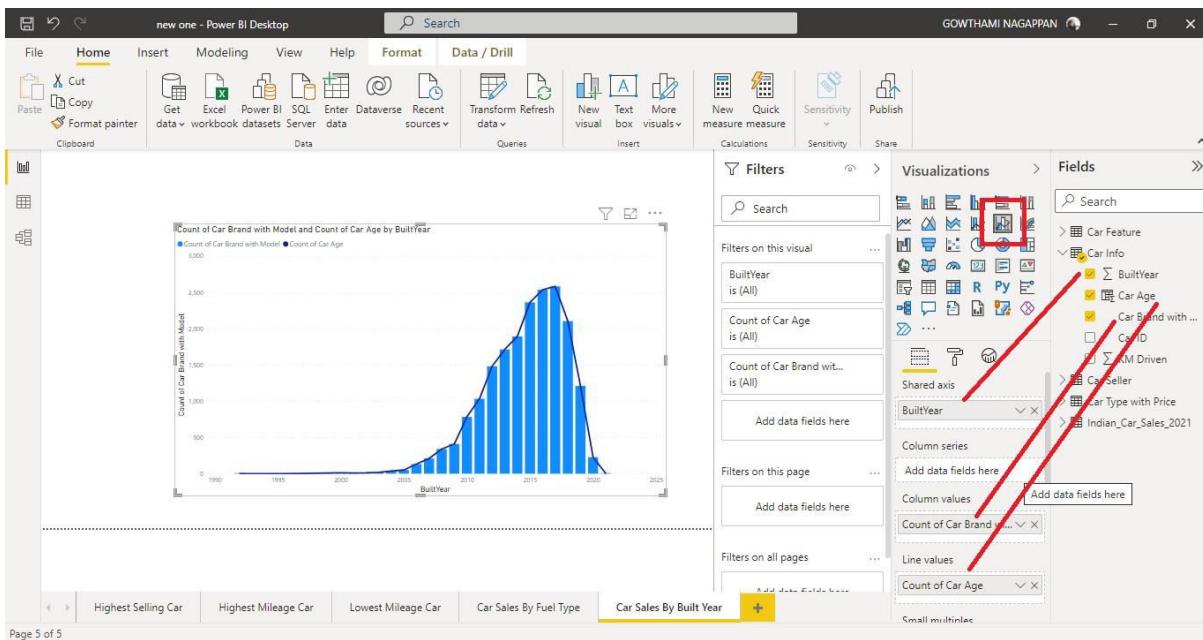


Fig: Line and Clustered Chart

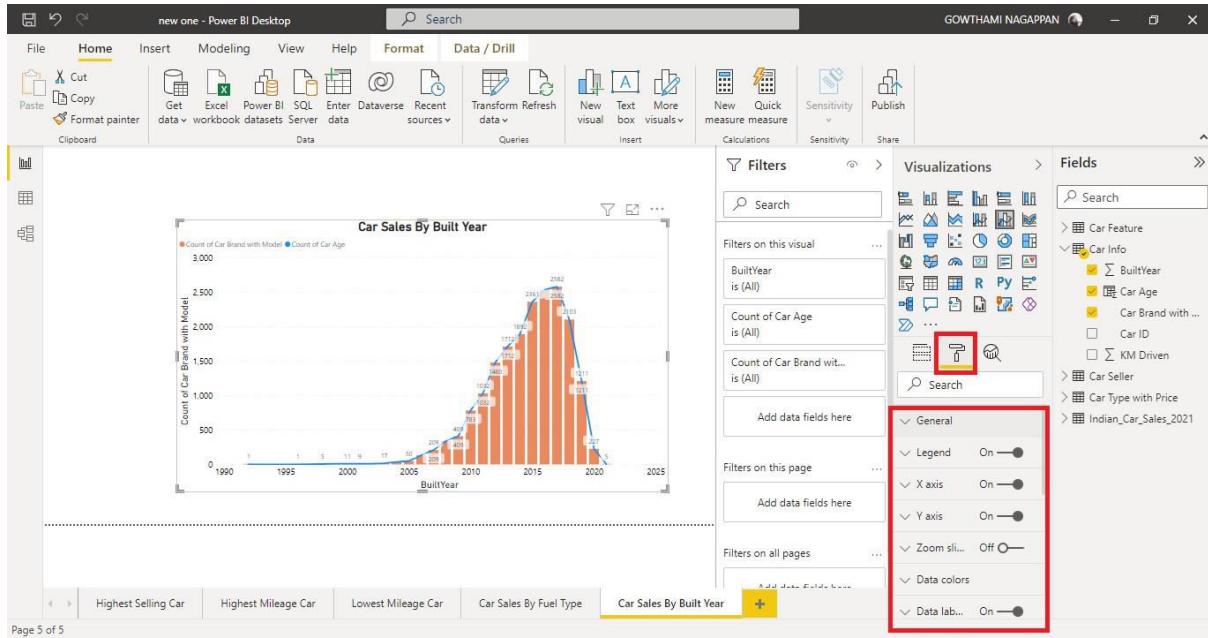


Fig: Line and Clustered Chart with Fields

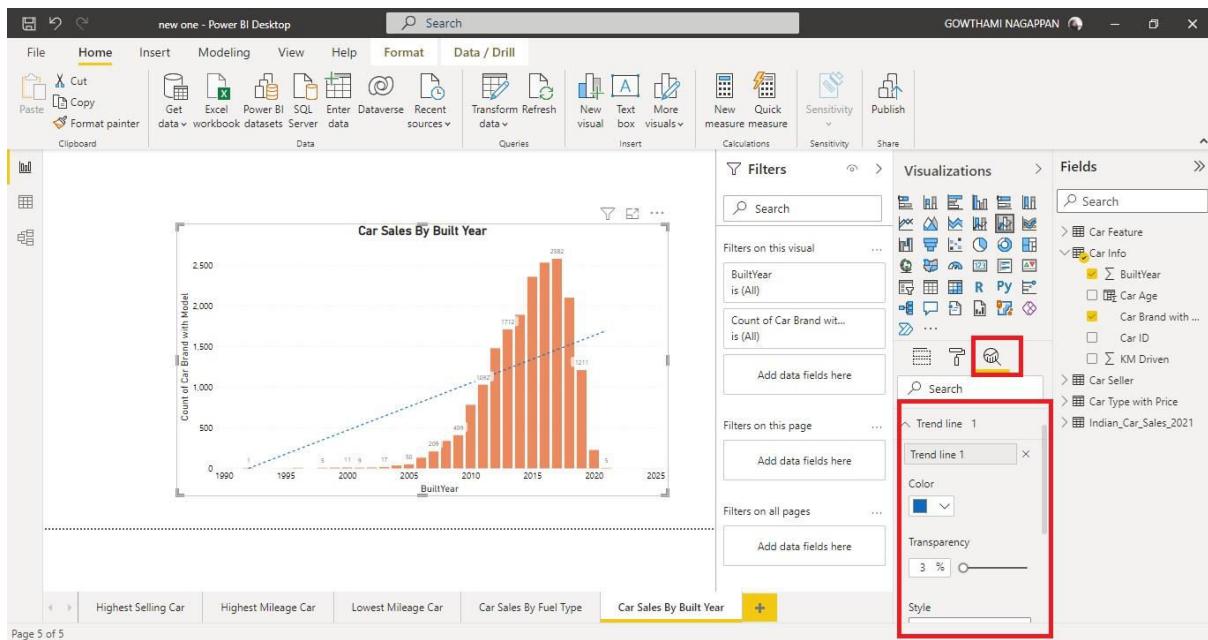


Fig: Line and Clustered Column Chart with Trend Line

Result

In terms of ranking cars manufactured in the year 2017, sales are higher and then it is 2016 and 2015.

e. Number of cars sold when compared with the Transmission type

Selected ‘Ribbon Chart’ to display the sales for each Fuel type. The below columns from the car type with prices were taken - Transmission Type (Axis), Fuel Type (Legends), Car Brand with Model (Right Click and select Count) in Values Field.

Editing was done in the format options for Title, Font size, Colors and Data label as shown in the image below.

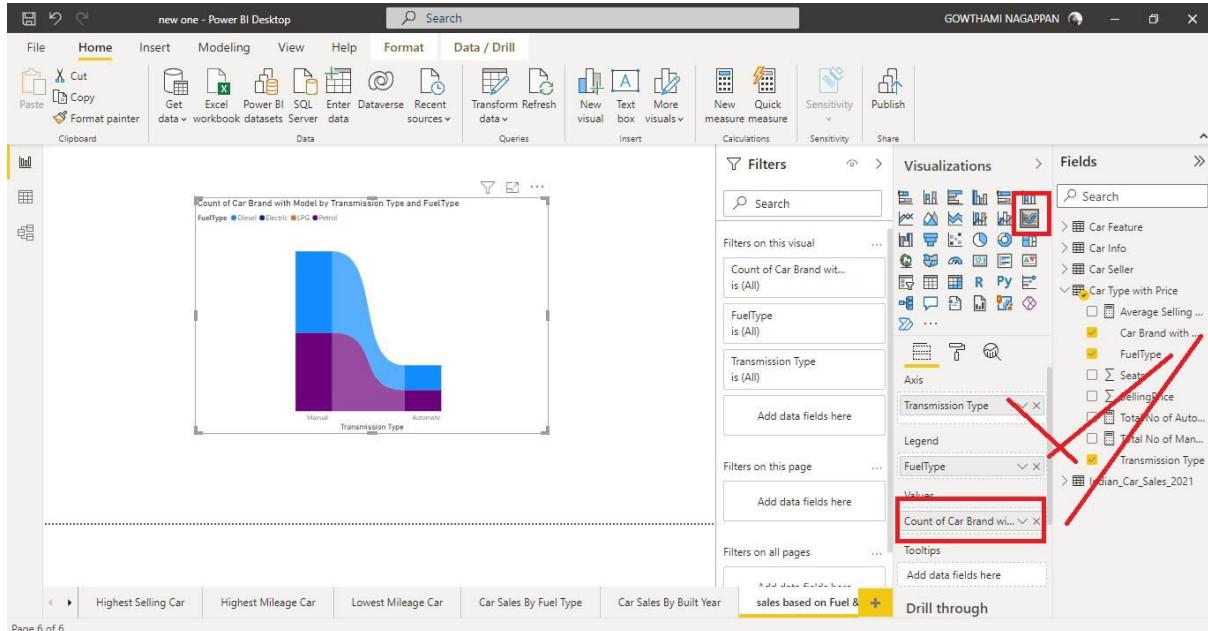


Fig: Ribbon Chart

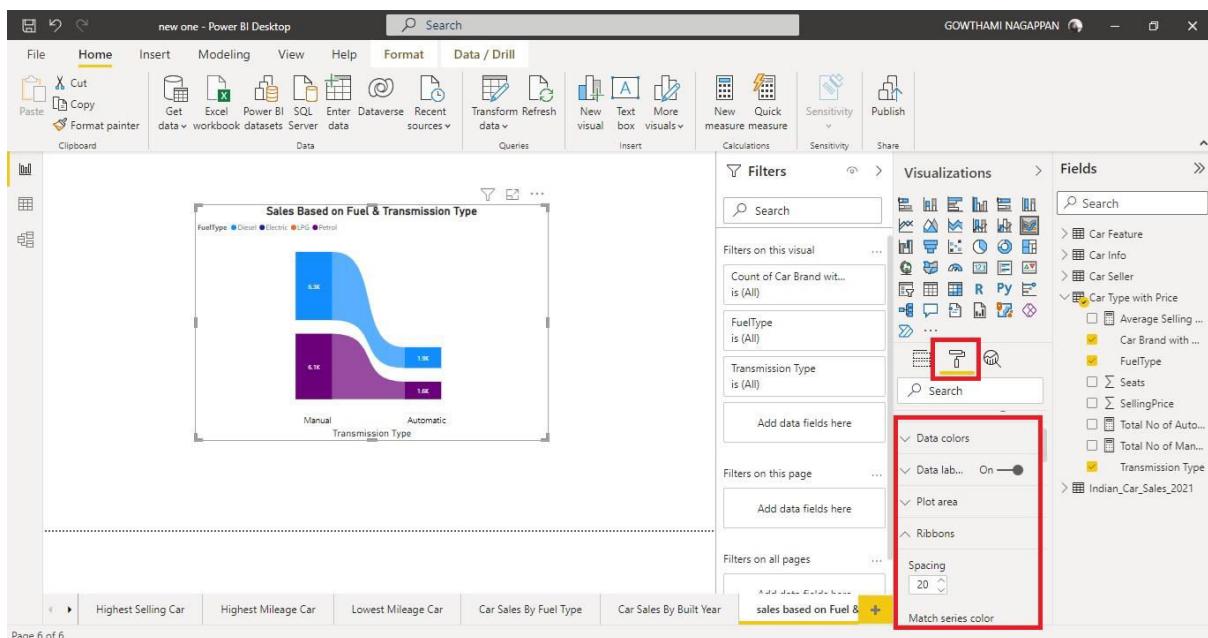


Fig: Fuel type Car Transmission Count using Ribbon Chart

It gives a better idea on which transmission type to focus for each fuel type. As per the analysis made, none of the LPG models were available for Automatic transmission and manual variant is not available for electric cars.

Result

- 6291 number of Diesel car models were sold in Manual Transmission type.
- 6050 number of petrol car models were sold in Manual Transmission type.
- 1921 number of Diesel car model were sold in Automatic Transmission type.
- 1620 number of petrol car model were sold in Automatic Transmission type.
- None of the LPG type cars were available in Automatic Transmission type.
- None of the Electric car type were available in Manual Type.

f. Number of cars sold according to the seat type

To show the number of sales based on the number of seats where ‘Clustered Bar Chart’ Visual with Car Brand with Model (Count) was used in values and the Seats in Axis field.

The format option was used to edit Data label, title and colour as shown below

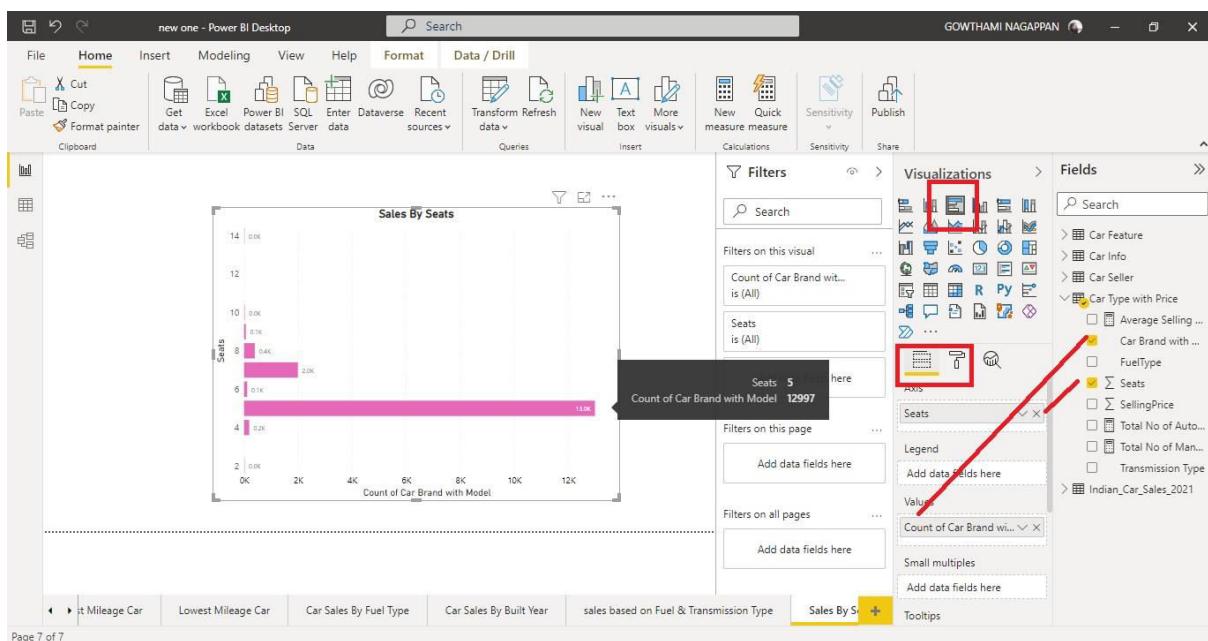


Fig: Clustered Bar Chart

The constant line was used to show the number of cars which were sold more than 10,000. The 5-seat car type has crossed that mark which is clearly shown in the image below.

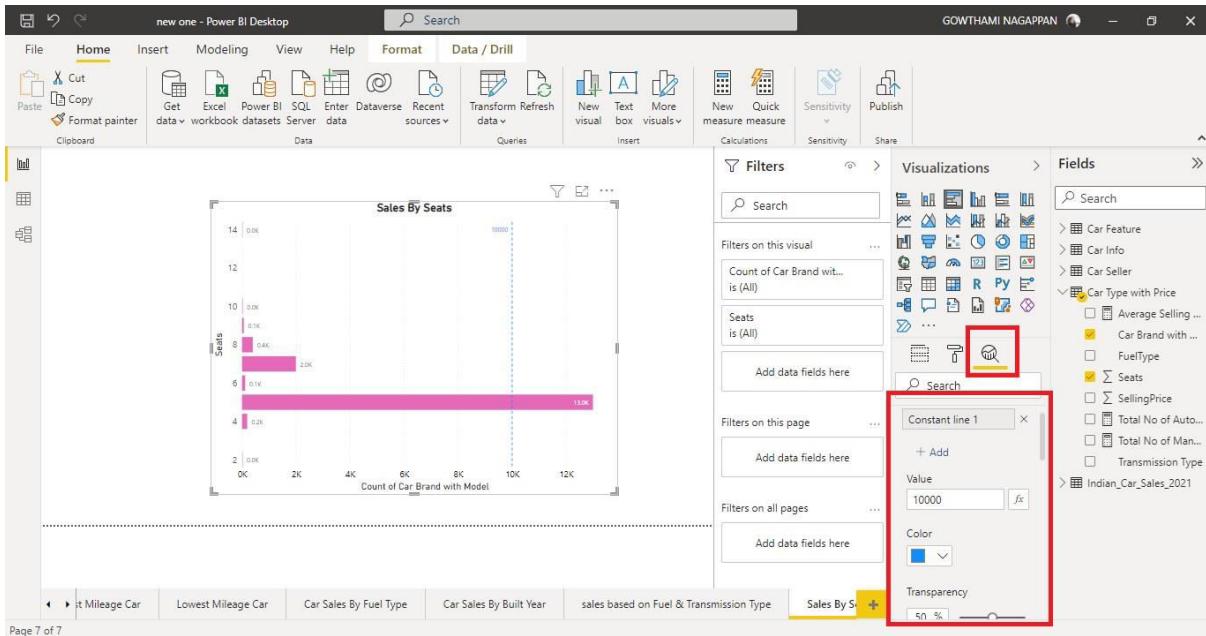


Fig: Clustered Bar Chart to find high sales based on Seats

The firm can concentrate on selling the 5-seater cars and that could be achieved by providing incentives to the dealers who list more number of cars with 5 seats.

Result

The number of cars sold under the '5 seat' type is higher when compared with the other categories.

g. Number of cars sold in each year which has clocked more than 100,000 Kilometers

To get new animation visual, the option 'Get more visual' option was chosen to add it in the Visualization Panel as shown below.

'Animated Bar Chart Race' was picked to show the number of cars sold for each year that has run for more than 100,000 KMs.

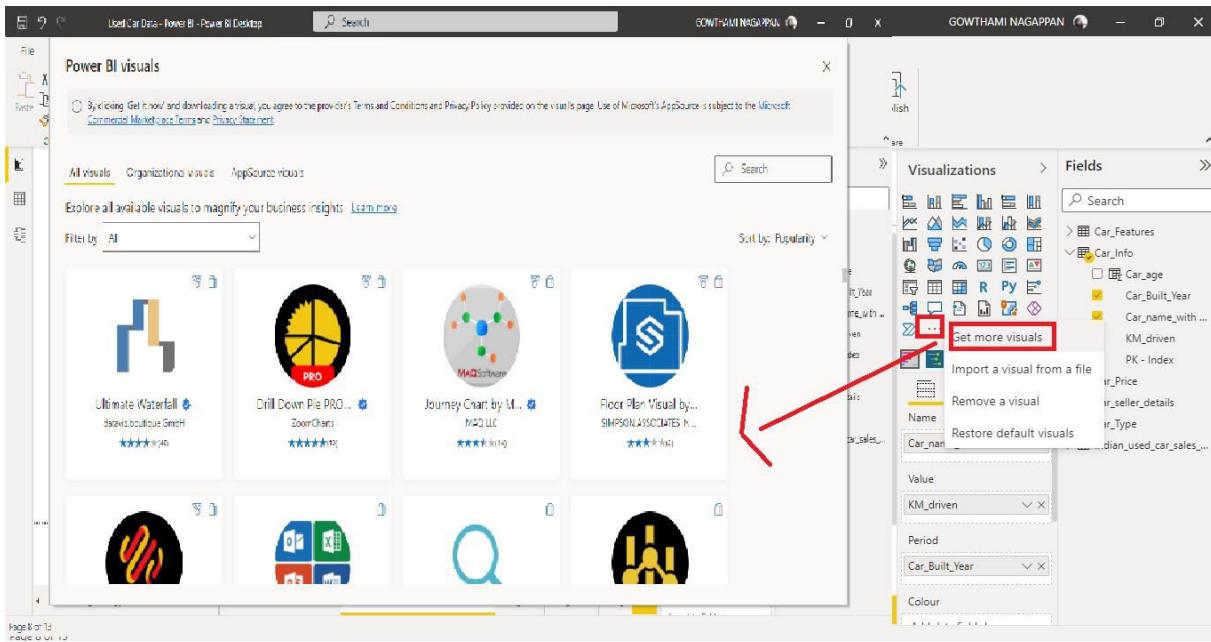


Fig: Getting More Visuals

The ‘Animated Chart Race’ Visual was used to display the details for each manufacturing year which includes minimum and maximum KM Driven, and the number of cars sold which had run above 100,000 Kms as shown below.

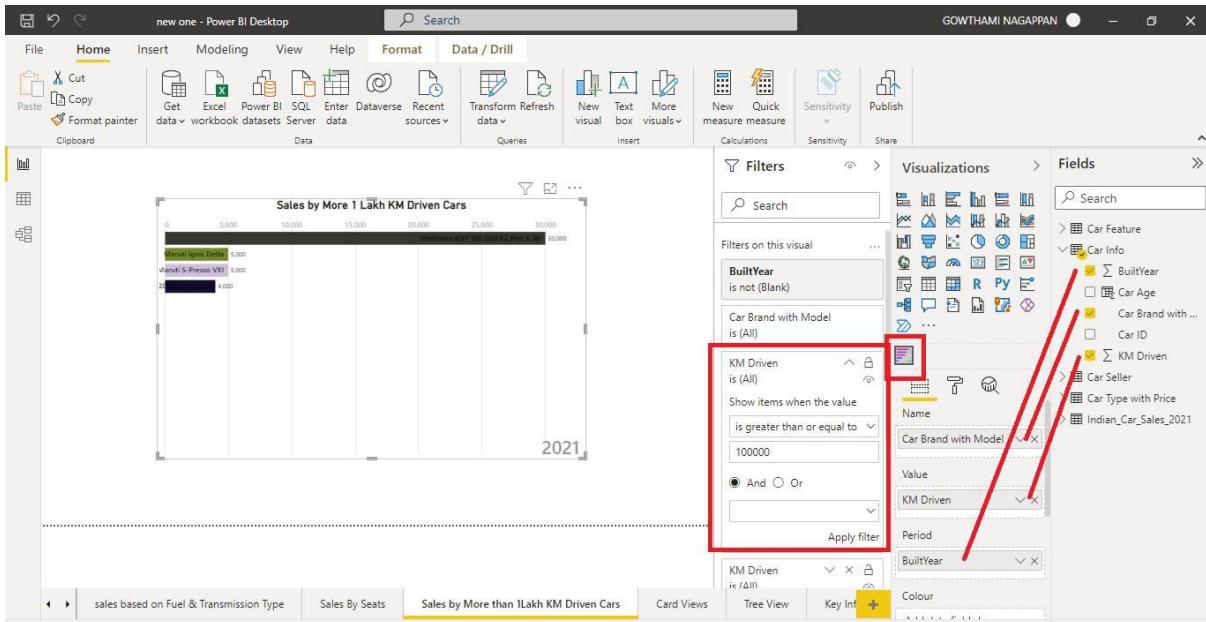


Fig: Applying Filter for More than 100,000 KM driven

In animated Chart Race, the cars sold for each year from 1990 to 2021 with a range of more than 100,000 Kms driven are displayed below.

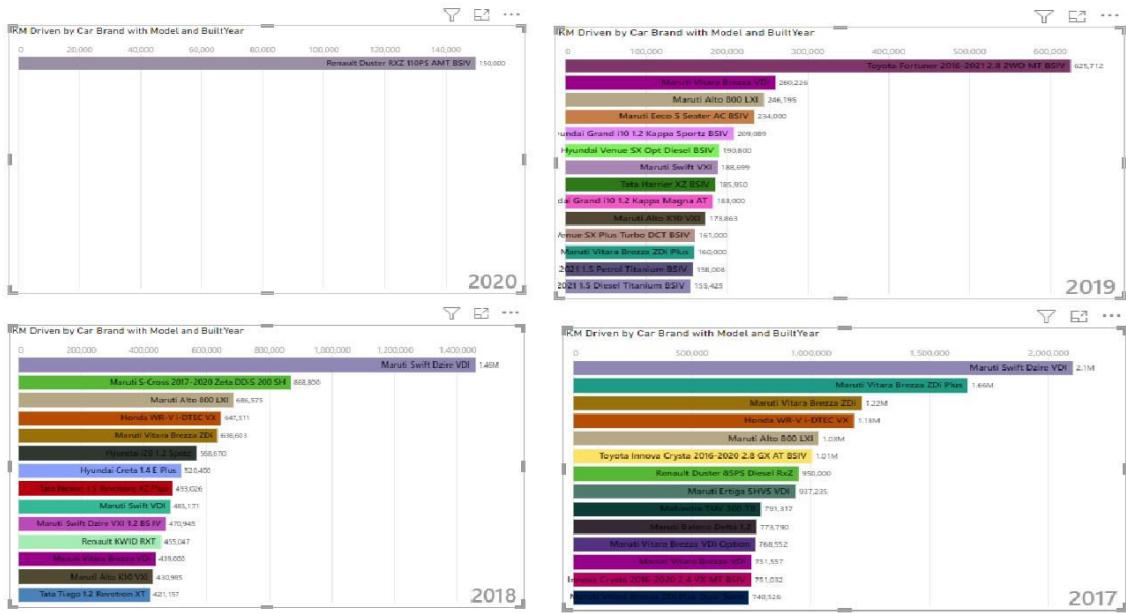


Fig: Animated Chart Race (Attached Screenshot for the past 4 years)

Result

In our sales data, the cars manufactured in 2021 year built were not driven more than 100k KMs, only one car manufactured in the year 2020 has been driven more than 100k KMs ,10 cars manufactured in the year 2017 & 2018 had been driven for more than 100k KMs

h. Number of cars sold according to seller type

The number of cars sold for each seller type is displayed using 'Multi row card view'.

Fig: Multi Row Car View for Variety of Sellers

Here Multi card view is used to display the sales of car for each sellers Type and Car Brand with Model (count) fields.

The format option was used for changing the title, data color and data label as shown above.

Result

The chart shows that the maximum sales happened from the dealers side when compared with the Individual and Trustmark Dealer. The cars sold by the Trustmark Dealer is very low.

i. Measures in Card View:

Few Basic Measures were done in section 1 before starting the Key Findings such as Average Sale price of the car and Total no of Manual & Automation Cars

Quick measure options were used to find the values which we explained already in section 1

The following Screenshot shows all the Card View with Values.

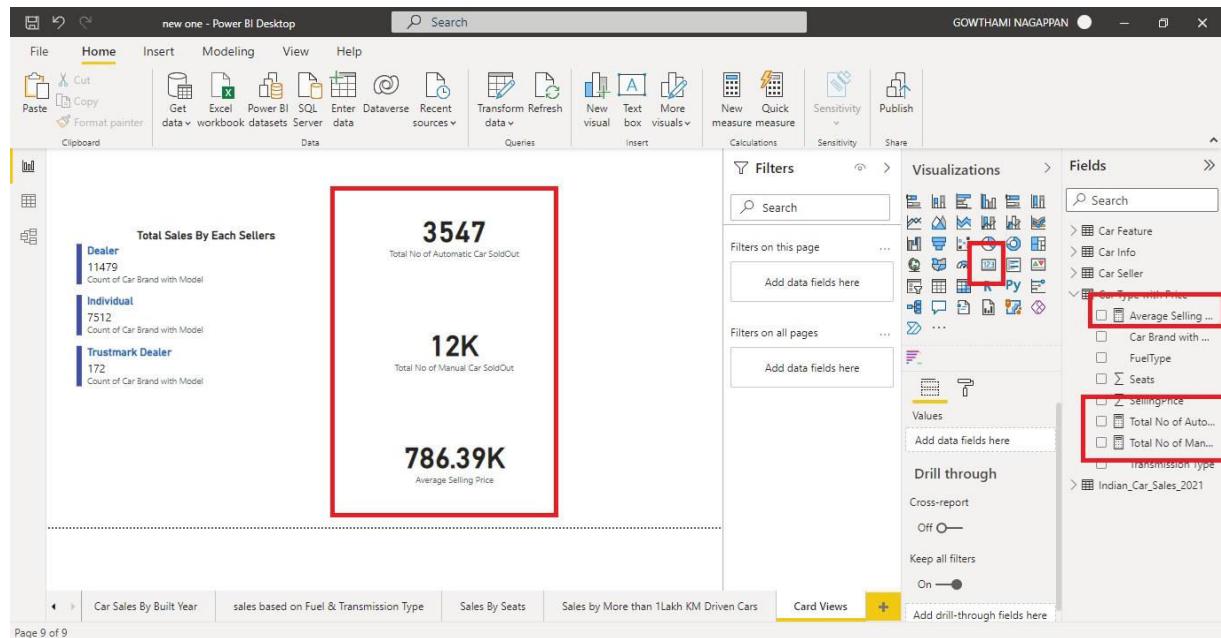


Fig: Card View

AI FEATURES IN POWER BI

Advanced analytics with the features of Artificial intelligence such as Decomposition Tree, Key Influencers and Q&A are available in Microsoft Power BI visualization.

a. Decomposition Tree

Decomposition Tree is an AI-powered tool which allows to view the data in multiple dimensions. It drills down and aggregates the data automatically. The request can be performed to find the next dimension to dig down based on certain parameters. The tree view is helpful to improve the investigation or root cause analysis.

Displaying no of cars varies by Engine_cc, Max_power and Mileage_kmpl as shown below.

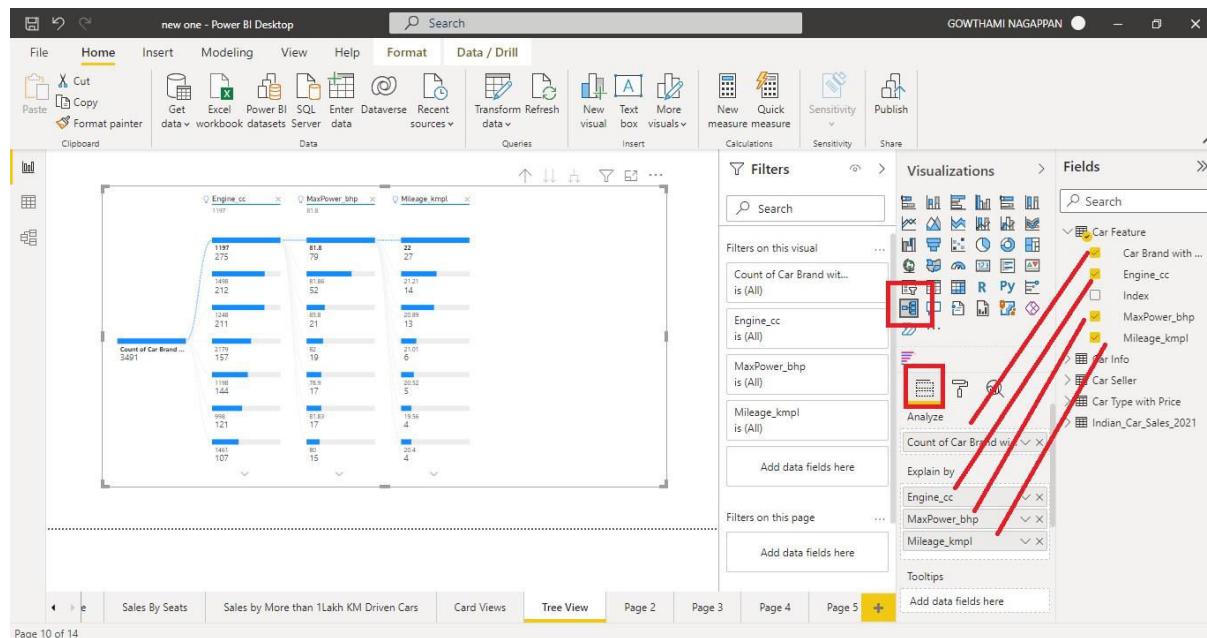


Fig: Decomposition Tree Expandby Engine, Max power and mileage



Fig: Another Screenshot from Different Expansion

b. Key Influencers

Visual key influencers aids in the understanding of the factors that determine critical business knowledge performance indicators. It analyses the information which will ranks the most important aspects and then it displays them as key influencers.

The top contributors to the given metric value are displayed in Key Influencers. Top segments display the top segments that contribute to the metric value that is chosen.

In this chart, the mileage is explained by MaxPower_bhp and Engine_cc as shown below

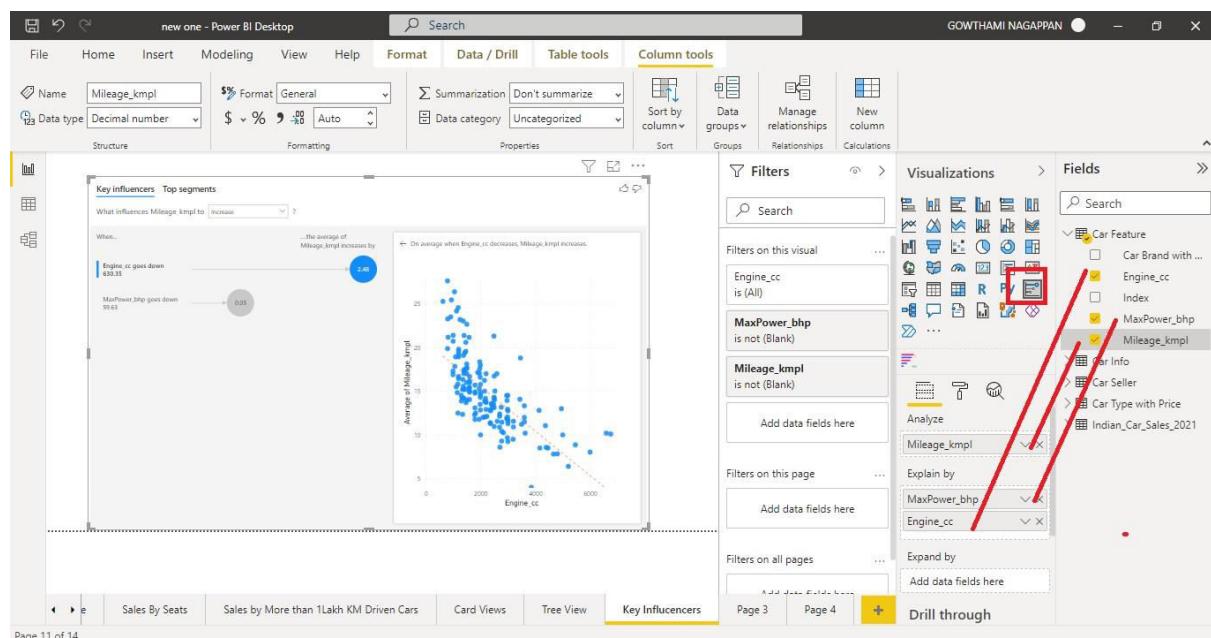


Fig: Key influencers Visual

The mileage will get increased when the power of the engine (MaxPower_bhp) and CC gets reduced. On the other way round, the mileage will decrease when the power of the engine (MaxPower_bhp) and CC increases.

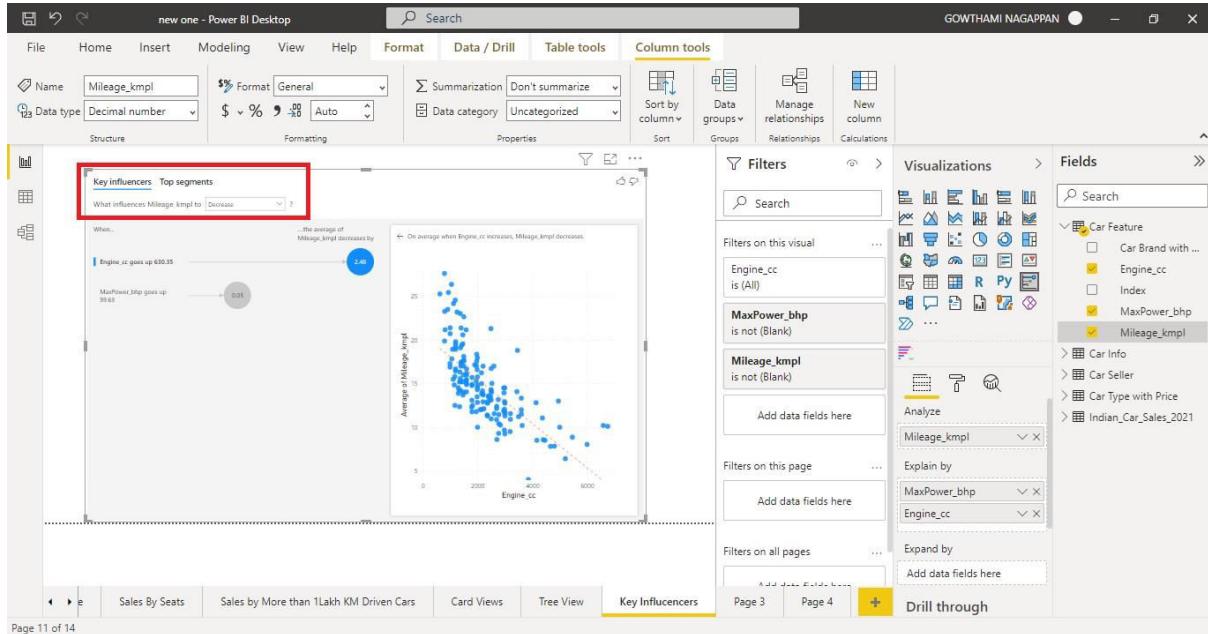


Fig: Increase and Decrease Factors influenced by mentioned Fields

Based on Engine_cc and MaxPower_bhp, the graph will show the average mileage corresponding to it which is shown in the image below.

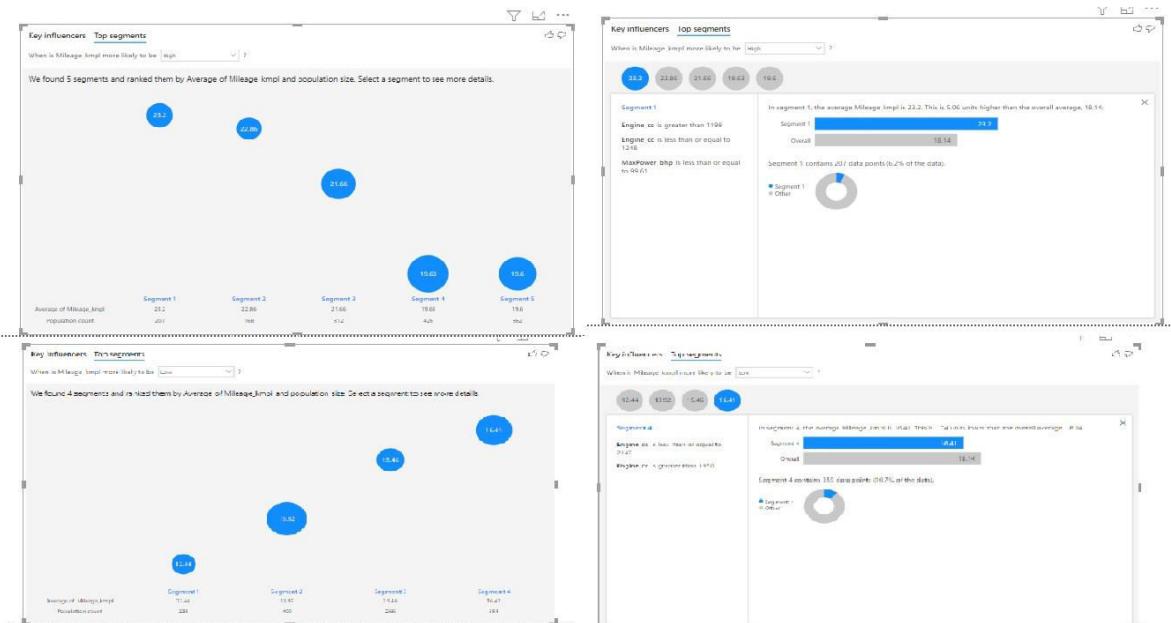


Fig: Top Segment Explains about Max_power and Engine_cc

c. Q & A Visual

The ‘Q&A’ graph is a useful tool that lets to ask questions in own terms and then it automatically answers them. In addition, it provides the examples on the visual type which could be used in the report.

Q&A visual icon option is selected from the Visualizations pane, and it will automatically display the question box with few suggested questions as shown below:

Fig: Q & A visual

Car ID	KM Driven	Car Brand with Model	Car Age	BuiltYear
G307	5000	Maruti Ignis Delta	0	2021
9810	5000	Mahindra KUV 100 G2 Plus 6 Str	0	2021
8242	2000	BMW 3 Series 320i Luxury Line	0	2021
14203	2000	BMW 3 Series 320d Luxury Line	0	2021
15127	5000	Maruti S-Presso VXi	0	2021

Fig: Autogenerated Q & A

Most of the questions and answers have already been explored and its possible to feed the query and make it understand. The automatically generated question charts are shown above.

After all the analysis, the dashboards are designed to show the sales which is based on certain factors and A.I features.

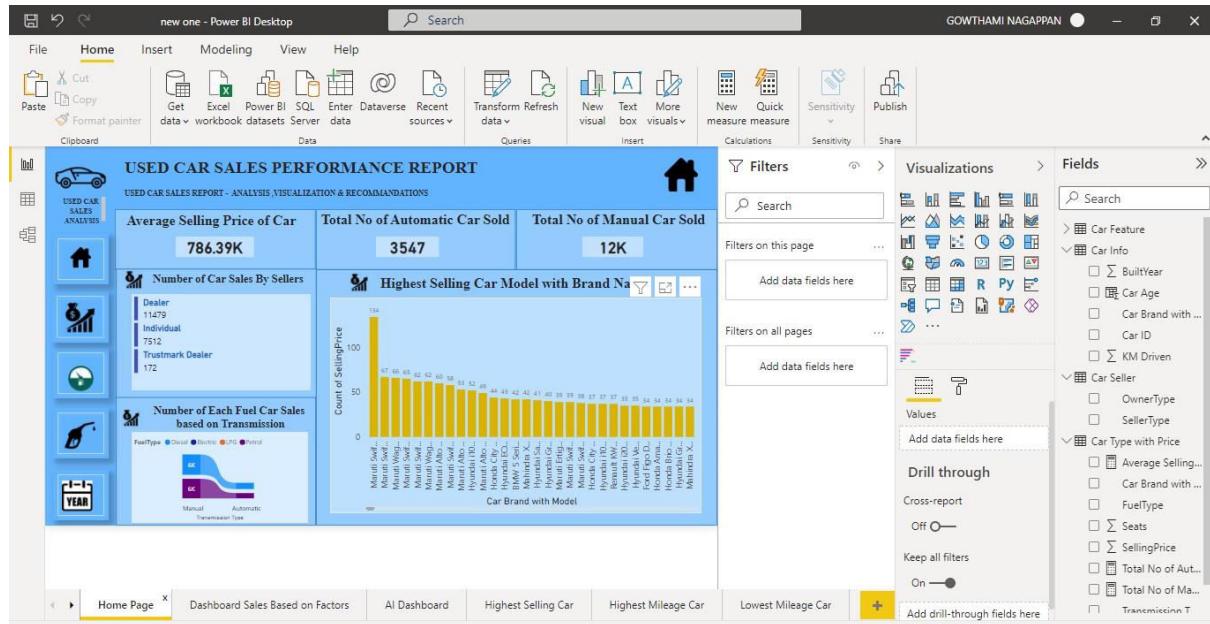


Fig: Dashboard for Sales

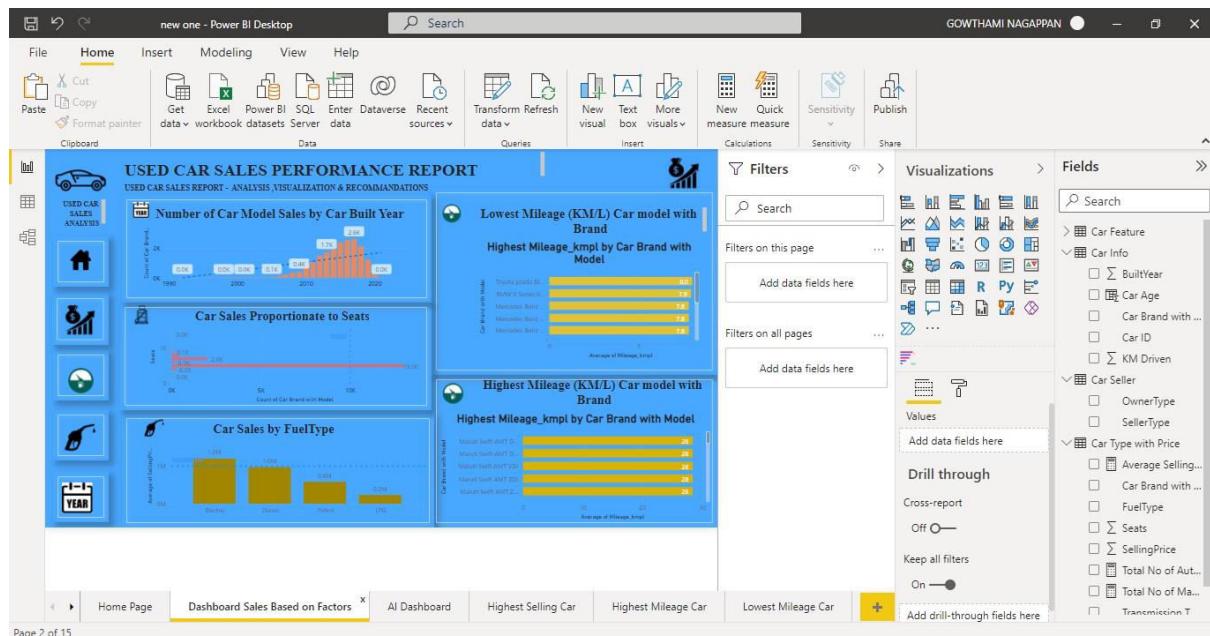


Fig: Sales based on certain factors

The screenshot shows the Power BI Desktop interface with the 'USED CAR SALES PERFORMANCE REPORT' dashboard open. The dashboard features a decomposition tree visual, an average price per car visual, and several filters and fields on the right side. The Power BI ribbon is visible at the top, and the Fields pane on the right lists various car-related fields like Car Feature, Car Info, and Car Seller.

Fig: Dashboard for AI

After the dashboards are created, publish options was chosen and all the work done until now will get saved in power BI online workspace.

The screenshot shows the Power BI Desktop interface with the same dashboard as before. The 'Publish' button in the ribbon is highlighted with a red box. A publishing dialog box is open, showing the progress of publishing the report to Power BI. The dialog box includes a progress bar and a message: 'Publishing to Power BI' and 'Did you know? You can create a portrait view of your report, tailored for mobile phones. On the View tab, select Mobile Layout. Learn more.'

Fig: Publishing Report

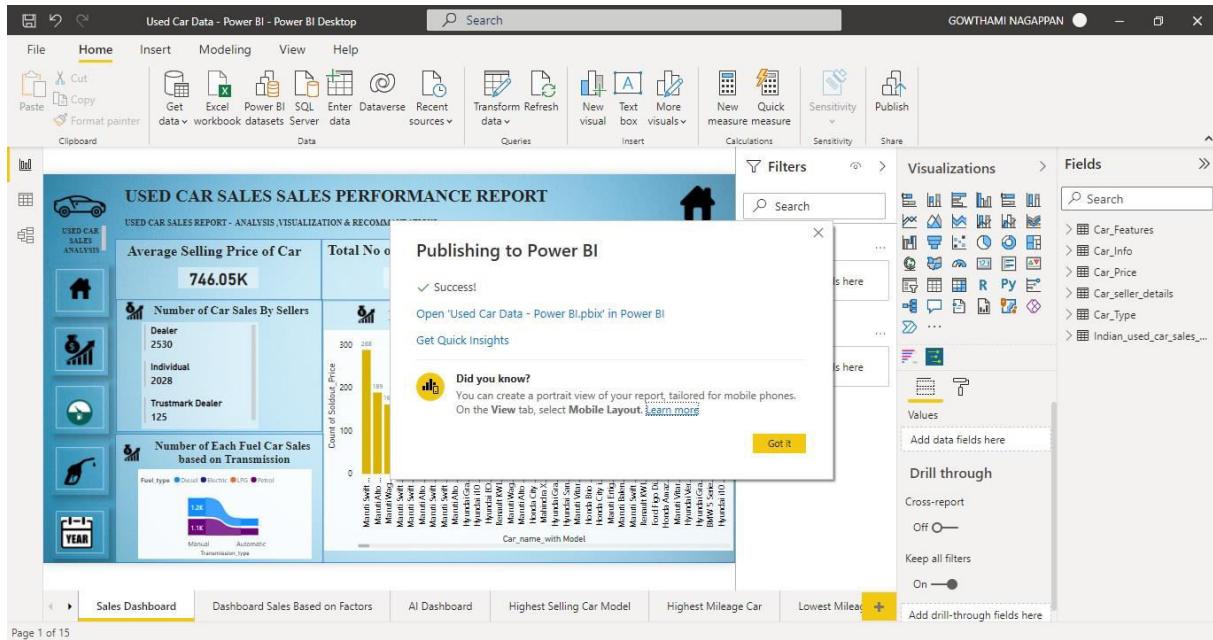


Fig: Report Published Successfully

Fig: Pages in Indian_Car_sales_2021

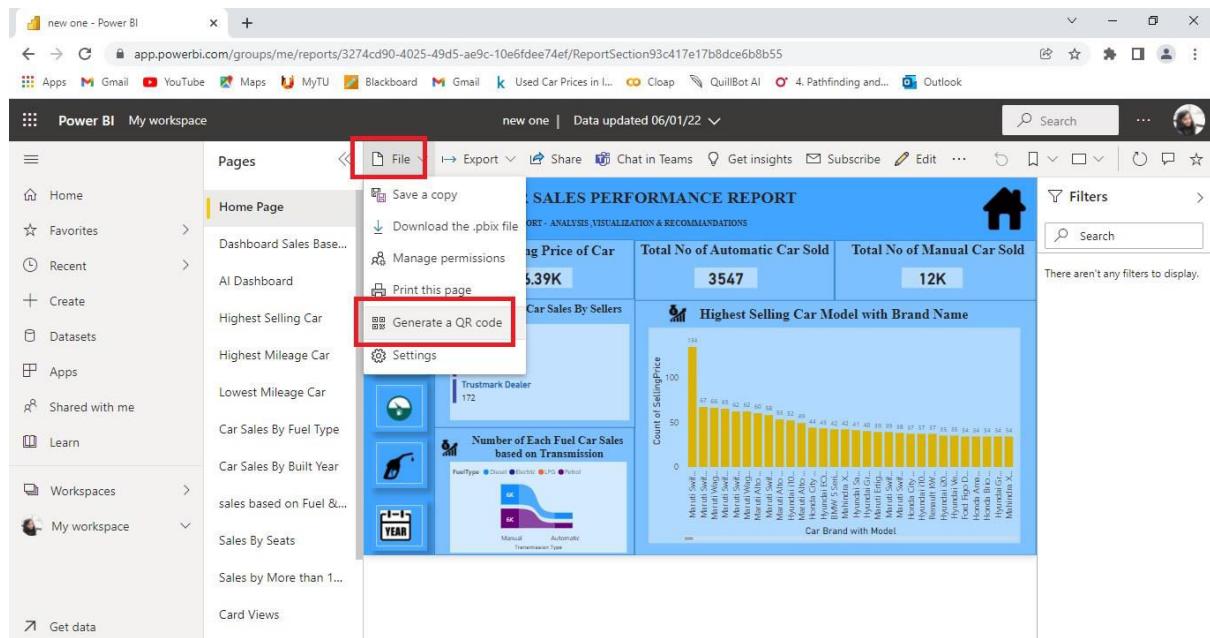


Fig: Generating QR Code

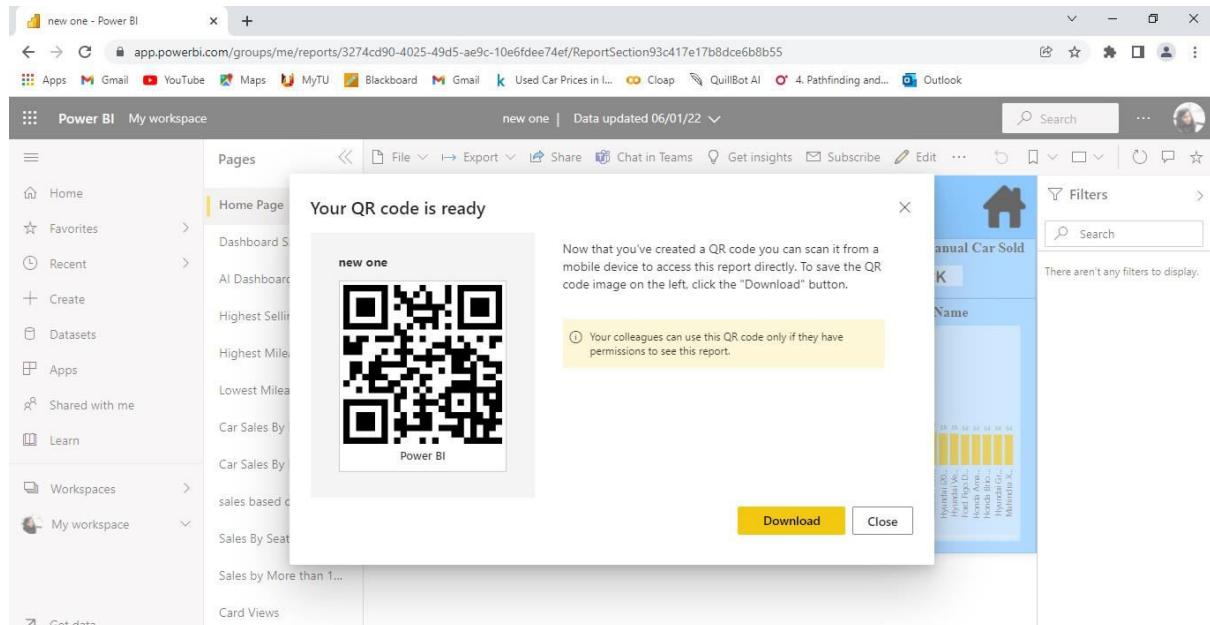


Fig: Sharing the report via QR code

Report sharing link: <https://app.powerbi.com/groups/me/reports/3274cd90-4025-49d5-ae9c-10e6fdee74ef/ReportSection93c417e17b8dce6b8b55>

Note: this link is generated using 60 days trial version and may get expired.

Conclusion

Business Intelligence report was prepared by using the cardekho used car sales data where we were able to find the key areas where there is a shortfall. The data was analyzed in detail and various relationship tables were created to identify each factor which had an impact on the sales and the same can be used as an input for making informed decisions. Through several analysis and metrics, it is found that the car sales can be improved by making the dealers to list more cars on the platform by providing incentives to them. Also, the key focus can be made on the cars which were manufactured within the five-year period as the customers preference for those cars are much higher than the older ones. Providing incentives to the customers who list the fast-moving cars and increasing the list price for the cars with less demand will induce the users to target only the fast-moving segment. Further scope of the research can be carried out to improve the customer experience who are using the ecommerce portal and diversifying into trusted service model.

References

Dataset from <https://www.kaggle.com/saisaathvik/used-cars-dataset-from-cardekho.com>

Tech startup transforms India's scattered used-car market:

<https://asia.nikkei.com/Business/Automobiles/Tech-startup-transforms-India-s-scattered-used-car-market>

Times of India about India's used-car market is booming:

<https://timesofindia.indiatimes.com/business/india-business/indias-used-car-market-is-booming-and-startups-are-capitalising-like-never-before/articleshow/88540155.cms>

Business-standard article about India saw three online used car platforms:

https://www.business-standard.com/article/economy-policy/india-online-used-car-sales-to-reach-8-3-mn-units-by-2026-report-121122700297_1.html

Economictimes - indiatimes about as startups drive up to the used-car market, a big lesson for Maruti and Mahindra: shift gears:

<https://economictimes.indiatimes.com/prime/transportation/as-startups-drive-up-to-the-used-car-market-a-big-lesson-for-maruti-and-mahindra-think-digital/primearticleshow/88151488.cms?from=mdr>