|  |  |
| --- | --- |
| Dekho car market study | Data from Car Dekho : https://www.kaggle.com/datasets/akshaydattatraykhare/car-details-dataset/code?resource=download  Maclean Ebenezer |

Table of contents

[I. Dataset cleaning 2](#_Toc122389585)

[A. In Power Query 2](#_Toc122389586)

[B. In Power BI Desktop 3](#_Toc122389587)

[II. Exploratory work (not compatible with web publishing) 6](#_Toc122389588)

[A. Fuel type 6](#_Toc122389589)

[B. Transmission type 6](#_Toc122389590)

[C. Year group 7](#_Toc122389591)

[III. Analysis 8](#_Toc122389592)

[A. Market analysis 8](#_Toc122389593)

[1. Overview 8](#_Toc122389594)

[2. Market specificities 10](#_Toc122389595)

[B. New car selling strategy 11](#_Toc122389596)

[1. Depreciation rate 12](#_Toc122389597)

[2. Cars’ reliability 12](#_Toc122389598)

[3. Brand annual revenue 13](#_Toc122389599)

[4. Overall ranking 14](#_Toc122389600)

[C. Considerations 15](#_Toc122389601)

# Dataset cleaning

## In Power Query

1. Remove underscores from column names.
2. Replace “Opelcorsa” by “Opel” in Brand.
3. Group cars by brand:

* Create a new column named “Brand” by extracting brands from the column Name.
* For “Land”, use replace value to add “Land Rover”.

1. Group car models:

* Create a new column named “Car group” by extracting the text before the delimiter “ “ from model.
* Replace values for BMW.

1. Remove aggregation from “km\_driven”, ”selling\_price”, ”year” columns:

* Select the 3 columns in the data model, then go to Properties, Advanced, and select Summarise by “None”.

1. Create table manufacturers details:

* Import car manufacturers details from:

<https://www.car.info/en-se/brands?order=country>

* Create an XLS file and add manufacturers that have not been imported and import the file to Power Query.
* Append both manufacturers tables as “Manufacturer details”.
* Deactivate the loading of both previously created tables.

1. Create table Country list

* Import Country list from (Asia, Europe and North America only): <https://www.newworldencyclopedia.org/entry/list_of_countries_by_continent>
* Use first line as headers for all 3 tables.
* Create a personalised new column named “Continent” and using M language fill it down to the end depending on the table name.
* Extract text before “(” on the country column and delete all columns except from country and continent.
* Append all 3 tables as “Country list”.
* Replace “United States” by “USA”.
* Deactivate both the loading and the refresh of all 3 previously created tables.

## In Power BI Desktop

1. Go to model and create a 1-to-many relationship between the manufacturers details table and the car details table.
2. Create a 1-to-many relationship between the Country list table and the manufacturers details table.
3. Create a new group named “Year Group”:

This will allow analysis over aggregated time periods. As the analysis focuses on cars, and since the year distribution is over a 28 year period, it would be interesting to group cars over 4 years periods, resulting in 7 groups.

*Note: The groups won’t be equally populated.*

1. Create new columns:

Age = YEAR(TODAY())-'CAR DETAILS FROM CAR DEKHO'[year]

Car status = IF ('CAR DETAILS FROM CAR DEKHO'[owner] = "First Owner","New car","Resold

car")

Fuel type = if('CAR DETAILS FROM CAR DEKHO'[fuel]="Diesel","Diesel",

                IF('CAR DETAILS FROM CAR DEKHO'[fuel]="Petrol","Petrol","Other"))

Owner number = if ('CAR DETAILS FROM CAR DEKHO'[owner] = "Test Drive Car", 0 ,

                    if( 'CAR DETAILS FROM CAR DEKHO'[owner] = "First Owner", 1,

                        IF('CAR DETAILS FROM CAR DEKHO'[owner] = "Second Owner", 2,

                            IF('CAR DETAILS FROM CAR DEKHO'[owner]="Third Owner",3,

                                4

                        )

                    )

))

1. Create New Measures:

Average car price = AVERAGE('CAR DETAILS FROM CAR DEKHO'[selling price])

Average km per year = CALCULATE(DIVIDE([Total\_km\_driven],[Car age],0),'CAR DETAILS FROM

CAR DEKHO'[Car status]="New car")

Brand Depreciation = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[Depreciation

rate],,ASC,Dense)

Brand profitability = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[New car annual

revenue],,ASC,Dense)

Brand rank = [Brand profitability]+[Brand profitability]+[Brand reliability]+[Brand

Sales points]

Brand reliability = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[Average km per

year],,ASC,Dense)

Brand sales points = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[Sales

quantity],,ASC,Dense)

Car age = SUM('CAR DETAILS FROM CAR DEKHO'[Age])

Car Depreciation = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Depreciation

rate],,ASC,Dense)

Car profitability = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[New car annual revenue],,ASC,Dense)

Car quantity by brand = COUNTROWS(ALLSELECTED('CAR DETAILS FROM CAR DEKHO'))

Car quantity global = COUNTROWS(ALL('CAR DETAILS FROM CAR DEKHO'))

Car rank = [Car Depreciation]+[Car profitability]+[Car reliability]+[Car sales points]

Car reliability = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Average km per

year],,ASC,Dense)

Car revenue rank = if(ISINSCOPE('CAR DETAILS FROM CAR DEKHO'[Model

group]),RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Revenue],,DESC,Dense))

Car revenue ratio = DIVIDE([Revenue],[Car Total Revenue],0)

Car sales points = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Sales

quantity],,ASC,Dense)

Car sales rank = if(ISINSCOPE('CAR DETAILS FROM CAR DEKHO'[Model

group]),RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Sales quantity],,DESC,Dense))

Car sales Ratio = DIVIDE('CAR DETAILS FROM CAR DEKHO'[Sales quantity],'CAR DETAILS

FROM CAR DEKHO'[Car quantity by brand],0)+0

Car total revenue = CALCULATE(SUM('CAR DETAILS FROM CAR DEKHO'[selling

price]),ALLSELECTED('CAR DETAILS FROM CAR DEKHO'))

Continent rank = RANKX(ALL('Country list'[Continent]),[Sales quantity])

Depreciation rate =

VAR First\_known\_owner = MIN('CAR DETAILS FROM CAR DEKHO'[Owner number])

VAR Second\_known\_owner = CALCULATE(MIN('CAR DETAILS FROM CAR DEKHO'[Owner number]),'CAR DETAILS FROM CAR DEKHO'[Owner number]>First\_known\_owner)

Return DIVIDE([Average car price](FILTER('CAR DETAILS FROM CAR DEKHO',IF(Second\_known\_owner=0,'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner,'CAR DETAILS FROM CAR DEKHO'[Owner number]=Second\_known\_owner)))-CALCULATE([Average car price],'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner),CALCULATE([Average car price],'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner),0)

Fuel rank = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Fuel type]),[Sales

quantity],,DESC,Dense)

Model ranking = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Model group]),[Sales

quantity],,DESC,Dense)

New car annual revenue = DIVIDE([New car revenue],DISTINCTCOUNT('CAR DETAILS FROM CAR

DEKHO'[year]),0)

New car revenue = CALCULATE(SUM('CAR DETAILS FROM CAR DEKHO'[selling price]), 'CAR

DETAILS FROM CAR DEKHO'[Car status]="New car")

New car sales quantity = CALCULATE(COUNTROWS('CAR DETAILS FROM CAR DEKHO'), 'CAR

DETAILS FROM CAR DEKHO'[Car status]= "New car")

Revenue = SUM('CAR DETAILS FROM CAR DEKHO'[selling price])

Revenue rank = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[Revenue],,DESC,Dense)

Revenue ratio = DIVIDE([Revenue],[Total Revenue],0)

Sales quantity = COUNTROWS('CAR DETAILS FROM CAR DEKHO')

Sales rank = RANKX(ALL('CAR DETAILS FROM CAR DEKHO'[Brand]),[Sales

quantity],,DESC,Dense)

Sales Ratio = DIVIDE('CAR DETAILS FROM CAR DEKHO'[Sales quantity],'CAR DETAILS FROM

CAR DEKHO'[Car quantity global],0)

Standard deviation selling price = STDEV.P('CAR DETAILS FROM CAR DEKHO'[selling

price])

Total Revenue = CALCULATE(SUM('CAR DETAILS FROM CAR DEKHO'[selling price]),ALL('CAR

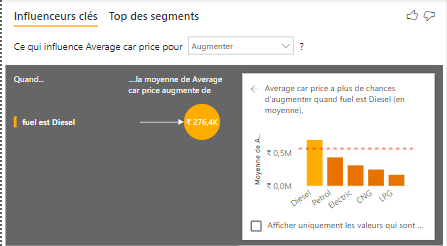
DETAILS FROM CAR DEKHO'))

# Exploratory work (not compatible with web publishing)

To get a better idea of where to start, Key influencers visuals can be interesting. Here they are being utilised to highlight factors influencing cars average selling price.

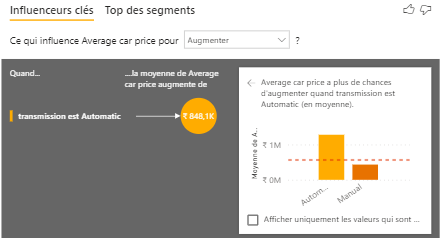
## Fuel type

Diesel cars tend to be more expensive than other cars.



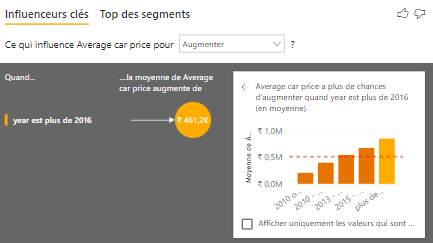
## Transmission type

Automatic cars tend to be more expensive than manual cars.



## Year group

Car sold after 2016 tend to be more expensive than other cars.



This exploratory work helps to understand what variable to use as filter when exploiting the data to get meaningful insights.

# Analysis

This analysis will be divided into 2 parts:

* An analysis of Dekho car market
* A New car selling strategy

In Power BI, corresponding visuals are divided in 2 groups:

* “(brand)” aiming to give insights at a brand level
* “(model)” aiming to give insights at a car model level. Visuals in this group are designed to be filtered by brand, allowing the comprehension of each model’s impact on its brand.

## Market analysis

This section aims to give an in-depth analysis of the market. Consuming trends will be explored, as well as factors influencing prices and sales.

### Overview

Une image contenant texte

Description générée automatiquementThe dataset focuses on Dekho car market accounting for over 2 billion Rupees worth of sales (24 million dollars) over 28 years *(See considerations).*

Most popular brands in terms of sales quantity are Maruti, Hyundai, Mahindra and Tata, dominating over 64% of the market and accounting for 55% of its revenue.

Une image contenant texte, moniteur, écran, noir

Description générée automatiquement

Brands with high selling prices attract less customers, therefore price has an impact on sales, however taking a closer look at top selling brands it appears that most popular brands are also Asian. The geographical location of the manufacturer also needs to be taken into consideration as people tend to buy more local cars.

Une image contenant carte

Description générée automatiquement

The 3 most selling car models are Marutis, followed by the Hyundai i20.

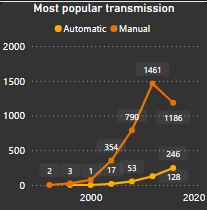
Une image contenant texte, moniteur, écran

Description générée automatiquement

Une image contenant texte, moniteur, écran

Description générée automatiquementFiltering the market analysis page on Maruti, allows to notice that Swifts alone account for 40% of the brand sales and more than half of its total revenue.

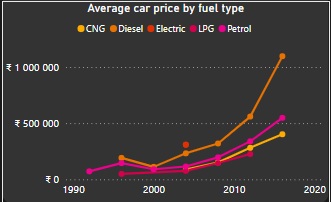
### Market specificities

People buy more cars as years advance, and they have very recognisable preferences, indeed the typical car is a cheap, diesel and manual car.

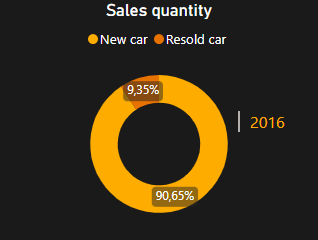
Manual cars have always dominated this market, however the switch from petrol cars to diesel cars was progressive diesel being the most popular fuel since 2012, slightly in front of petrol.

Une image contenant texte, équipement électronique, capture d’écran

Description générée automatiquement



Finally, customers also prefer new cars to resold cars with over 90% of cars being bought new since 2016.



Une image contenant texte

Description générée automatiquement

## New car selling strategy

This section aims to design a new car selling strategy based on available data. Brands depreciation rate, reliability, annual revenue, and new cars sales will be evaluated to build a model able to highlight top 3 brands and top 5 car models to sell to maximise revenue and client satisfaction.

This model relies on a point system attributed to brands and models separately, depending on their rank on each evaluated factor. The number of points attributed equals the inverse of the brand’s or car’s rank. For instance, Hyundai, Maruti and BMW are the 3 brands generating the most annual revenue on new cars out of 28 brands, therefore they respectively received 28, 27 and 26 points.

Une image contenant texte, écran, capture d’écran

Description générée automatiquement

### Depreciation rate

Here cars’ depreciation rate is being calculated as cars are often resold. Some customers might prioritise cars they can use and resell without losing too much money, therefore it is important to measure the average gap between the buying and the selling price of a car.

To do that, the average selling price of each car/brand for first known owners has been compared to the average selling price for second known owners.

Depreciation rate =

VAR First\_known\_owner = MIN('CAR DETAILS FROM CAR DEKHO'[Owner number])

VAR Second\_known\_owner = CALCULATE(MIN('CAR DETAILS FROM CAR DEKHO'[Owner number]),'CAR DETAILS FROM CAR DEKHO'[Owner number]>First\_known\_owner)

Return DIVIDE([Average car price](FILTER('CAR DETAILS FROM CAR DEKHO',IF(Second\_known\_owner=0,'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner,'CAR DETAILS FROM CAR DEKHO'[Owner number]=Second\_known\_owner)))-CALCULATE([Average car price],'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner),CALCULATE([Average car price],'CAR DETAILS FROM CAR DEKHO'[Owner number]=First\_known\_owner),0)

On average cars lose 37% of their value when they are being resold. This measure however needs to be considered with care as there is no data to estimate how long they are being used for before being resold.

### Cars’ reliability

Another interesting factor to consider when buying a car is its reliability, meaning how long is it going to last for.

Here, cars’ reliability is measured by calculating the average number of km driven per year. Because, once again there is no way to estimate how long a car has been used for before being resold, this measure focuses solely on new cars.

On average cars drive 7410 km/year. With respectively over 13k, 11,9k and 11,5k km driven per year on average, Volvo, Mitsubishi, and Toyota are the most reliable brands.

Une image contenant texte

Description générée automatiquement

Une image contenant texte, moniteur, capture d’écran, noir

Description générée automatiquement

Most reliable cars are Supro (Mahindra), Ingenio (Mahindra), CrossPolo (Volkswagen), NuvoSport (Mahindra).

Une image contenant texte, moniteur, noir, écran

Description générée automatiquement

### Brand annual revenue

Through the market analysis, we have evaluated the revenue generated by each brand/car over the 28 year period covered by the data however some brands have been on the market longer than others, it is therefore crucial to normalise the revenue variable to get meaningful insights. This can be achieved by measuring revenue annually.

New car annual revenue = DIVIDE([New car revenue],DISTINCTCOUNT('CAR DETAILS FROM CAR

DEKHO'[year]),0)+0

Top 3 brands in terms of annual revenue are Hyundai, Maruti and BMW. Overall, 2 strategies seem to thrive across the market:

* Sell a large amount of cheap cars
* Sell a lower amount of more expensive cars

The first approach seems to be the best working strategy, especially as customers tend to prioritise local cars. However, major international manufacturers such as BMW, Audi and Mercedes are also doing well as they are taking advantage of their brand names to distribute a decent amount of cars at a higher price. Location wise, German cars are the most profitable foreign cars.

Une image contenant texte, moniteur, intérieur, noir

Description générée automatiquement

This analysis is reinforced by the fact that the top 4 new cars in terms of annual revenue are all German cars.

Une image contenant texte, moniteur, écran, noir

Description générée automatiquement

### Overall ranking

This section was aiming to rank cars based on factors that customers might pay attention to when buying. Points were attributed to cars/brands depending on how well they performed in each area, consideration was also given to cars/brands’ popularity.

Une image contenant texte, moniteur, écran, noir

Description générée automatiquement

Overall, top 4 brands to sell on this market are Toyota, Mahindra equally ranked with Maruti and finally Hyundai. Top new cars to sell are Innova (Toyota), GLS (Mercedes), X5 Series (BMW), XC (Volvo) and finally Creta (Hyundai).

Une image contenant texte, périphérique, mètre

Description générée automatiquement



## Considerations

This study assumes that the year column describes the year when cars were sold however it is more likely that it was intended to describe cars’ manufacturing years, meaning that the dataset shows a picture of cars available on the market at a specific time rather than the historic of cars sold over the years. This choice highly affects interpretations, however the study can still be considered for its technical approach.