# REPORT

## INTRODUCTION

This data set contains the information of cars includes name of the car,year,selling price,Kilometer driven,fuel,seller type,transmission,owner,mileage,engine,max power,torque,seat capacity etc.

The data set now available were collected to analyse the key characteristics of the cars in the data set and also to enhance the customer experiences in the automobile industry.

## AIM

The main goal of this data analysis is to facilitate informed decision making. Business and organizations are inundated with data from different sources, including customer interactions, market trends and internal operations.

OBJECTIVE

1.Data Collection

By acquiring a well-structured data set. This data set included details of the cars such as name, fuel, selling price, transmission, owner and more. The data was stored in a database named car_dekho and the table name is cartable.

2.Data Cleaning

By data cleaning we can remove or correct the data set according to the needs by doing data cleaning we can improve the data accuracy.

3.SQL Queries

Designed and executed SQL Queries to extract relevant information from the database. This involved SQL operations such as SELECT, FROM, WHERE, GROUP BY,ORDER BY,LIMIT and aggregate functions.

Using SQL Queries, we have extracted and analysed key data points from the cartable to gain a deep understanding of the data set.

ANALYSIS

-->The transmission in the given data set is 4 types “Manual” , “ Manual” , “Automatic” , “ Automatic” . This is changed to 2 types “Manual” , “Automatic”.

-->Trimmed the Name column to remove the unwanted spaces and ordered it.

-->Updated the datatype of mileage, torque, max power, engine to int.

-->Add a new column named “Company” by cutting the name into substring and add the sub string to the newly created column.

-->Remove the all the duplicates existed in the data set and store it into a new table named “New1”.

-->Categorize mileage into low, avg, high.

-->Categorize the cars based on the kilometre driven by them.

# CONCLUSION

By analysing the data set cartable with in the database car_dekho, we find the valuable insights through SQL Queries. We also find  details of the most expensive car,average km driven of the data set.

1.Total number of vehicles:

Total number of vehicles in the database is 7927.By removing the duplicates occurred in the data set we get the total number of vehicles is 6738.from this we get that more than 1000 vehicles in the data set are duplicates.

2.Total number of vehicles based on the owner type:

First owner=419

Second owner=1892

Third owner= 493

Test Drive owner=5

Fourth and above owner= 155

From the above analysis we get that almost above half portion of the cars in the data set is first owner.

3.Total number of vehicles that based on the fuel type used:

Petrol=3663

Diesel=2987

CNG=52

LPG=35

Electric=1

Here we can clearly get that the usage of petrol vehicles is very high in the data set. And the other one of the striking point that only one person is use electric vehicle.

4.Most expensive car details:

The Most expensive car in the data set is Volvo on the basis of the selling price.

5.Average KM driven by owner:

First Owner=61089.3868

Second Owner=90305.6934

Third Owner=101238.6897

Fourth & Above Owner=105492.2258

Test Drive Owner=14631.4000

6.Highest max power:

The highest max power is 400 for Volvo car in the data set.