REPORT ON SALES OF AUTO  
 ----------------------------------------

Import data:-  
🡪First we have to load the autos data for analysing and visualisation

🡪For that we need to import python libraries like PANDAS, NUMPY, SEABORN, MATHPLOTLIB, and SCIPY as

Import numpy as np

Import pandas as pd

Import mathplotlib as plt

Import seaborn as sns

Import scipy as s

Import scipy.stat as ss

LOAD DATA:-

* After that load the data from the excel or csv file by copying the path as copypath
* But this data has a typical encoding error so the data should be in UTF-8 format and saved and while loading the data we have to load the data as

Data=pd.read\_csv(r”copypath”,encoding=utf-8)

* Now to make the data not to be effected by the analysis changes so we have to take a copy of the data as

Datac=data.copy()

**Report of ANALYSIS-1**

1) Perform general Data analysis

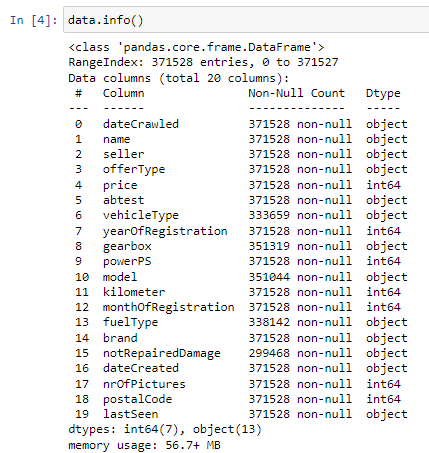
To perform general data we have to find the general information about the data like

Number of columns in the data   
Number of rows in the data  
Number of non null values in every column in the data  
Number of duplicate rows in the data  
what are the data types of each column in the data

These are the basic data analysis should be done on the data as follows

To know the basic information about the data we use the code as datac.info()

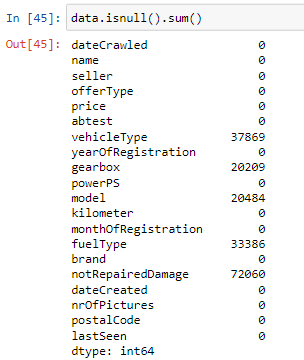
Which will gives u the all values of data columns, rows, data type of columns, on-null values in the column size of the data as follow



🡪Here we can confirm that there is total count of 18 columns in the autos sales dataset

🡪And there are 371528 rows in the dataset also there are some null values in some columns which are

**FINDING NULL VALUES**:-



So for we have 37869 null values in vehicle type column

And 20209 null values in gearbox column, 20484 null values in model column

33386 null values in fuel type column and 72060 null values in not repaired damage column

🡪Here all the null valued columns have same object-data type

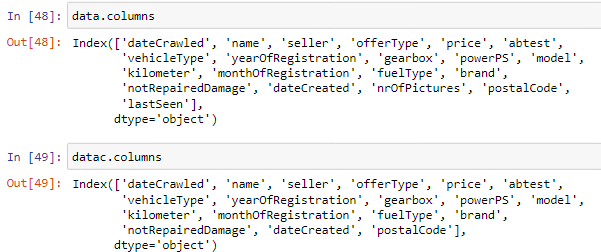
**REMOVING UNNECESSARY COLUMNS**:-

For our data we have unnecessary columns which we can remove by using drop method as

datac.drop(["nrOfPictures"],axis=1,inplace=True)

which can remove the column in the data permanently in the dataset.

🡪The result is “nrofpictures” column has removed permanently from datac. If we observe both the cells of our outputs from the below picture

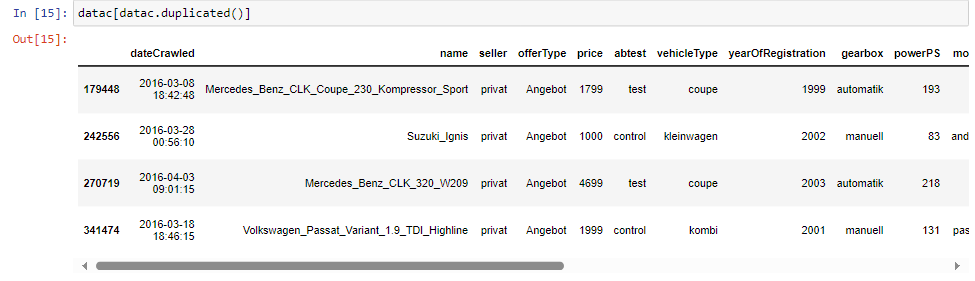


**REMOVING DUPLICATED ROWS**:-

* If we have duplicated rows we can find them by using the code as

datac.duplicated()

it will gives you the result as



i.e., we have four(4) no of duplicated rows present In the datac set

🡪 to remove these duplicated rows in the data use the code as

datac.drop\_duplicates(inplace=True)

* If we check the data these duplicated 4 rows will be removed permanently from the datac

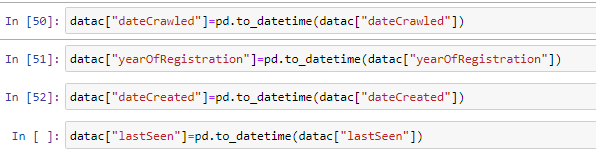
**REMOVING OR REPLACING NULL VALUES**:-  
From the above data all the null values are from object type so all are categorical columns for the categorical columns if the data contain less than 50% of its datapoint we can replace the null values with the mode of the column here all the columns are having null values very less than the 50% of the data of that particular column .So we can replace the data with its mode using the code as

datac.fillna({"notRepairedDamage":datac["notRepairedDamage"].mode()[0],"fuelType":datac["fuelType"].mode()[0],"vehicleType":datac["vehicleType"].mode()[0],"gearbox":datac["gearbox"].mode()[0],"model":datac["model"].mode()[0]},inplace=True)

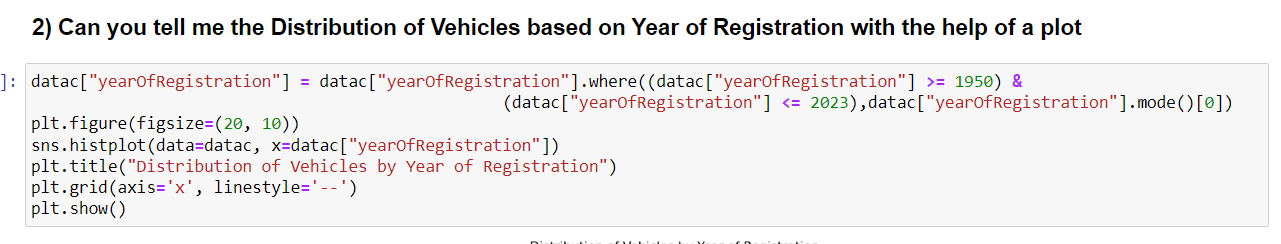
it will replace the null values with their respective column modes.

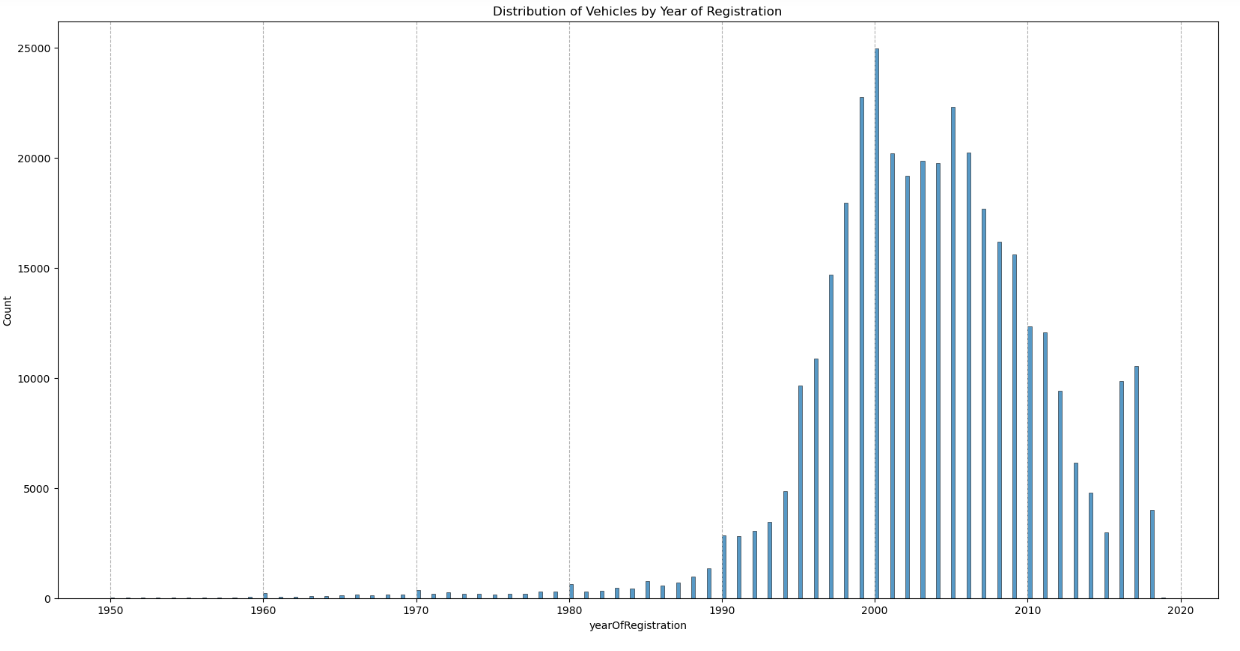
**CHANGING DATA TYPE**:-

🡪Here we obtain some of the columns are in object type instead of being its original type.so we can change their data type with their corresponding type as



Here all the columns are changed to their corresponding data type





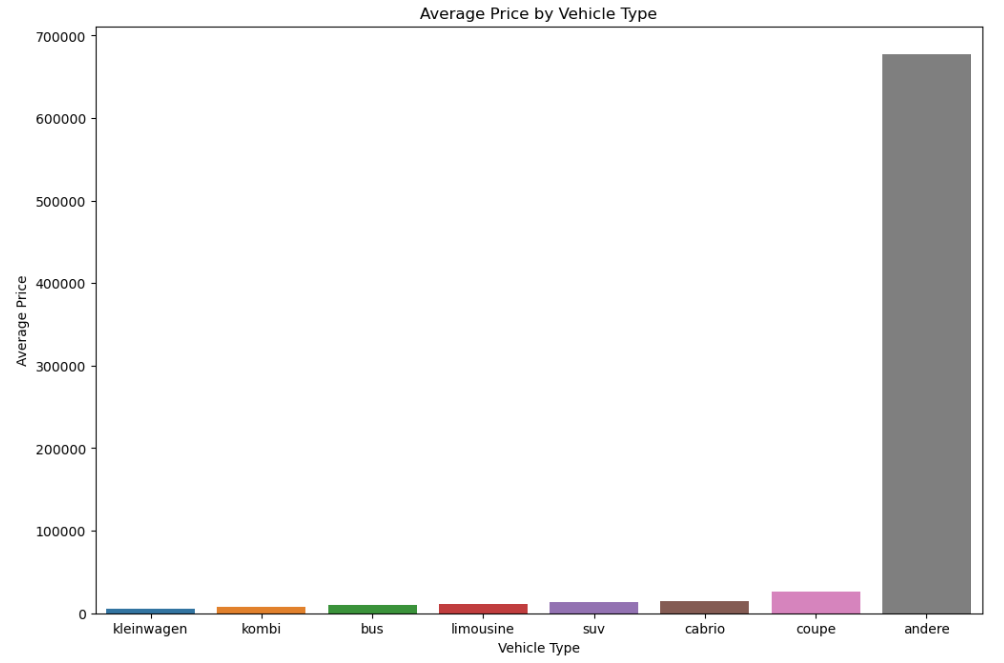
This image represents that the distribution of vehicles by the year of registration

At x-axis we consider the year of registration and on y-axis we consider the count of vehicles at that particular year on bar graph

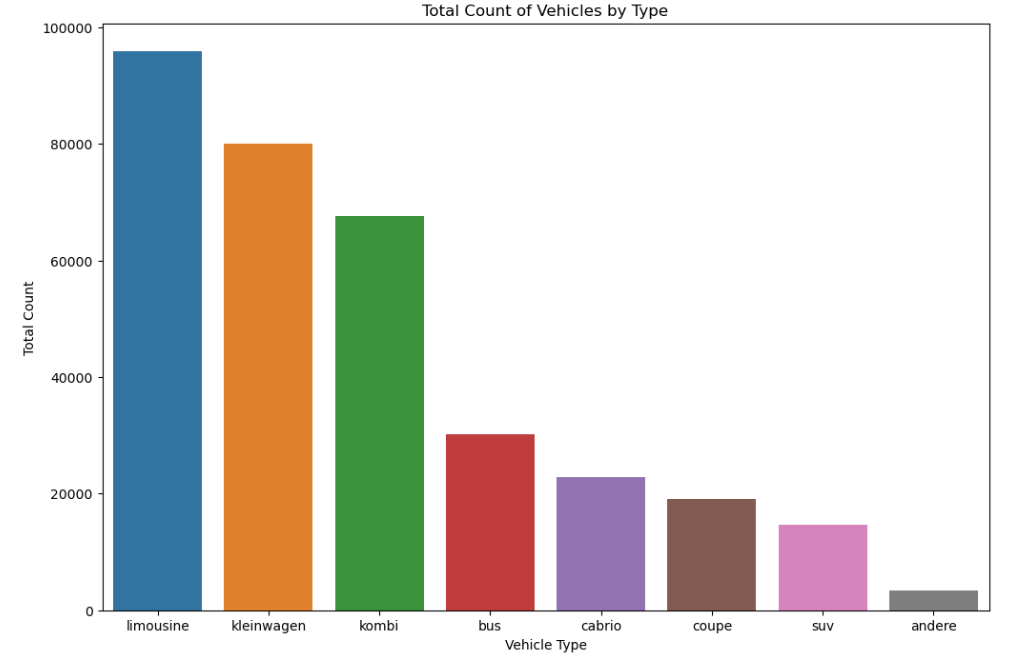
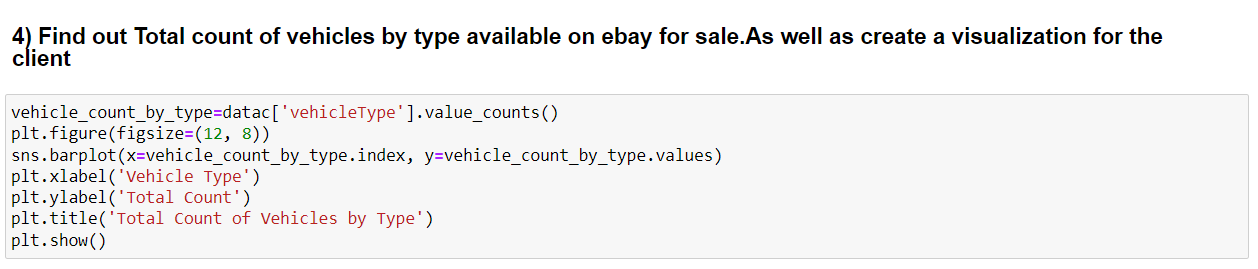
From this distribution we can conclude that a max no of vehicles are registered between the years 1980-2018

And there are very less count of vehicles which are registered on that particular years

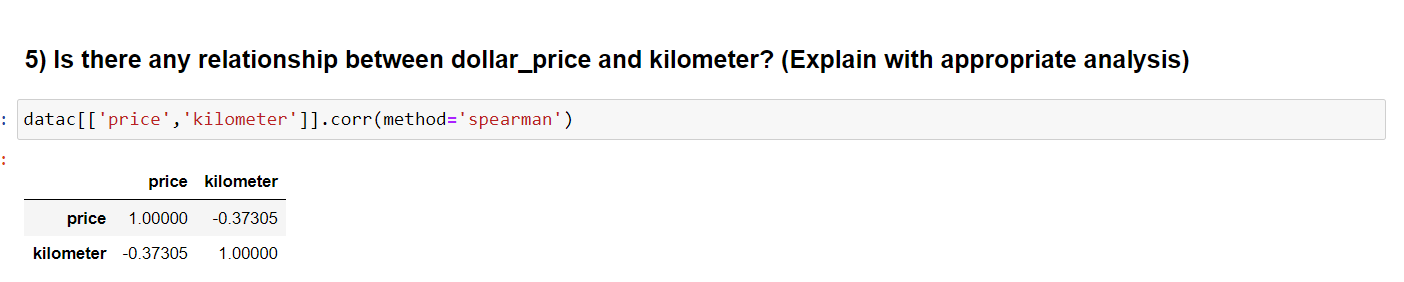
### 3) Create a plot based on the Variation of the price range by the vehicle type



The generated bar plot offers a clear visual representation of how average prices vary among different vehicle types. This information can be valuable for understanding the price distribution within the dataset and making informed decisions related to vehicle type pricing.



The generated bar plot provides a clear overview of the distribution of vehicle types in the dataset. It helps in understanding which types are more prevalent and which are less common. This information can be valuable for various analytical and decision-making purposes, such as inventory management or market analysis.



The Spearman correlation analysis between 'price' and 'kilometre' is a valuable step in understanding the relationship between these two variables in the dataset. The obtained correlation coefficient can help in making data-driven decisions, such as pricing strategies or assessing the impact of kilometres on vehicle pricing