**REPORT ON SALES OF AUTO**

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, 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

* 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 'ISO-8859-1'. So data is loaded like below,

data=pd.read\_csv(r"C:\Users\DELL\Downloads\autos.csv",encoding='ISO-8859-1')

**ANALYSIS-1**

1. Perform general Data analysis

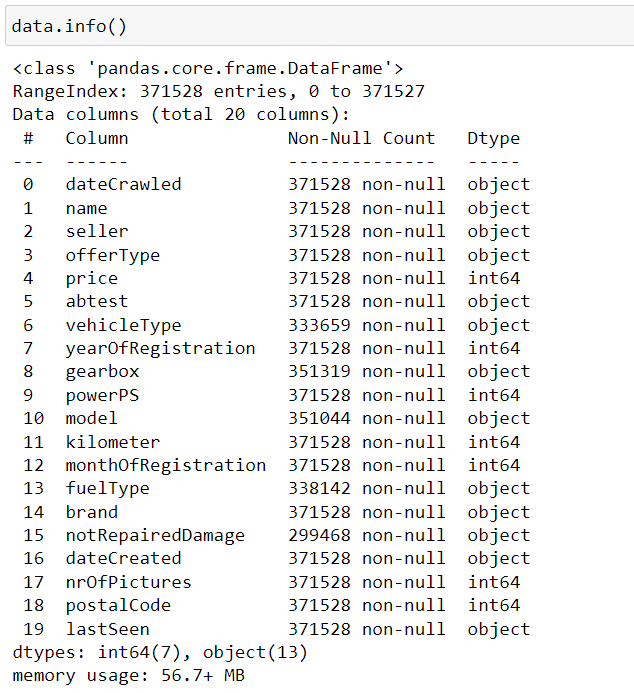
To perform general data we have to find the general information about the data like

1. Number of columns in the data
2. Number of rows in the data
3. Number of non null values in every column in the data
4. what are the data types of each column in the data

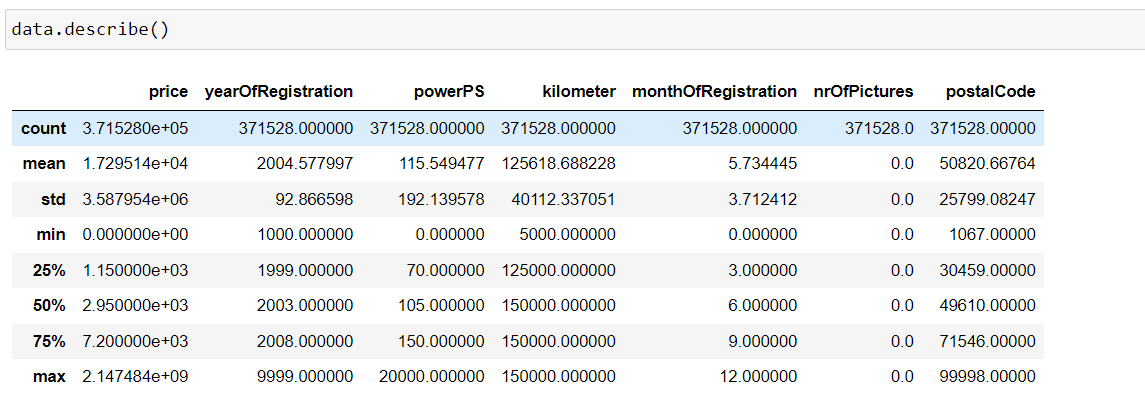
These are the basic data analysis should be done on the data as follows

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

Which will gives u the all values of data columns, rows, datatype of columns, non-null values in the column size of the data as follows

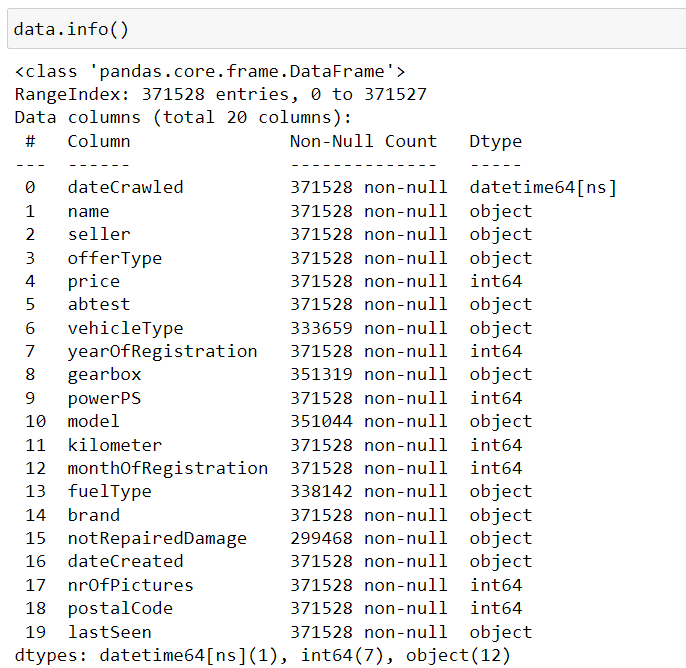


* Here we can confirm that there are total count of 18 columns in the autos sales dataset
* And there are 371528 rows in the dataset also there are some null values in some columns.



This is the general analysis of each numerical columns, of mean, median, mode, std and quantiles.

Here we chance the data type of column also. In the given data dateCrawled is in object data type but it should be in datetime type.



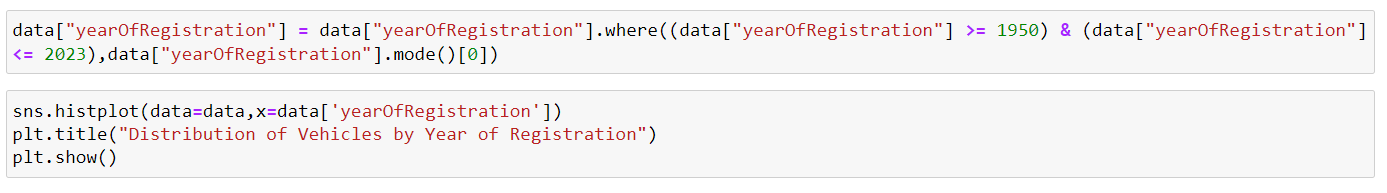
2) Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot

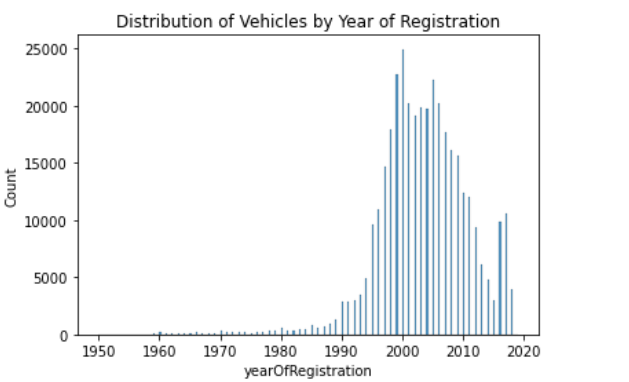
* To represent the distribution of plot we can use histogram plot.

Because the smoothened curve on the plots can represents the perfect distribution of that plots

1. We have to show how many no of vehicles have registered at that particular year so first we have to extract the data of registered year column and fit in a plot.
2. But the problem here is there are many years out of bound.
3. So we have to handle this by ranging the data to a certain value
   1. using the where function with applying a condition And replacing the other bounded data to the mode of its data values in the data by using the code
   2. data["yearOfRegistration"]=data["yearOfRegistration"].where((data["yearOfRegistration"] >= 1950) & (data["yearOfRegistration"] <= 2023),data["yearOfRegistration"].mode()[0])
   3. here we replaced the data of the column name year of registration which is not lying between 1950 and 2023 will be replaced with the mode of the data

1. Hence the data is converted to the particular minor ranged data.
2. To plot the data we use the code as





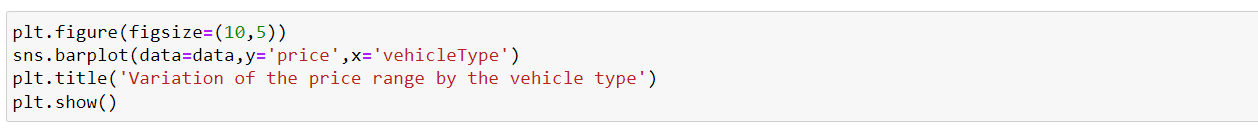
This image represents that the distribution of vehicles by the year of registration.

On x-axis we consider the year of registration and on y-axis we consider the count of vehicles at that particular year.

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

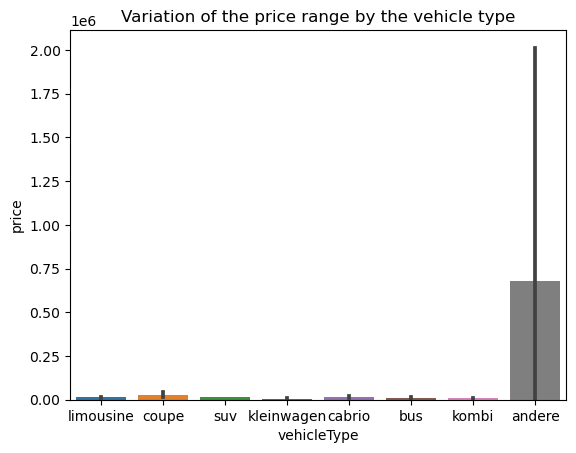
1. Create a plot based on the Variation of the price range by the vehicle type
2. To create a plot variation of price differ from one vehicle type to another vehicle we can select the bar plot.

* Because in bar plot it gives the count of vehicle type according to price.
* we use the code as



which represent the bar plot of vehicle type as x-axis and price on y-axis

as shown below.



From the figure we can say that the “andere” vehicle have the highest price of $0.74 and all other vehicle type very least price in the range of $0.12\*10^6-$0.0001.

And kleinwagen vehicle have the least cost price among all the vehicle type.

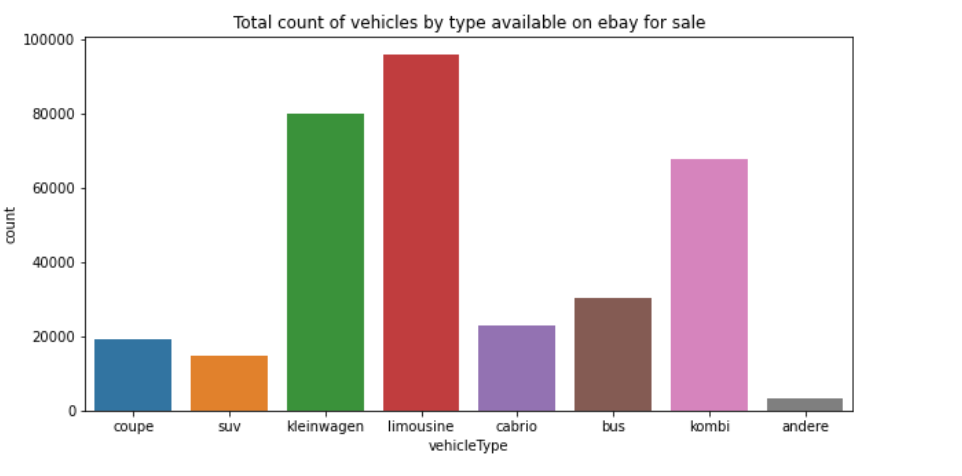
1. Find out Total count of vehicles by type available on ebay for sale. As well as create a visualization for the client

To create a plot on the count of vehicles by its type we can use the count plot for the data.

* Because in count plot we can get the count frequency on y-axis and vehicle type on x-axis.
* Before that we should take the consideration based on vehicle type by using the code (data[‘vehicleType’]).
* We used the code as below.



will gives the count plot about the count of each vehicle on showing the count of values in it



From the figure we can say that total count of vehicles by its type which are available on ebay is limousine have the highest count of 138900 cars soled. Also andere car has least count of soled which is around 5000.

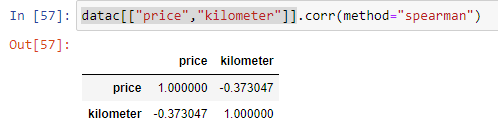
5)Is there any relationship between dollar\_price and kilometer? (Explain with appropriate analysis)

Yes, we can say the relation between the column price and kilometre which can be explained as below.

To know the relationship between the columns.

In the statistics there is a concept called correlation, here we can find the relation between two or more columns.





And the correlation between the price column and kilometre is -0.37.

Why because the correlation always between -1&1.

If the correlation lies between 0 to 0.5 we can say that moderately positively correlated.

If the correlation lies between 0 to -0.5 we can say that moderately negatively correlated.

If the correlation lies between 0.5-1 we can say that highly positively correlated.

If the correlation lies between -0.5 to -1 we can say that highly negatively correlated.

Finally we can say that relation between price and kilometre due to its negative correlation.