**1.Perform general data analysis**

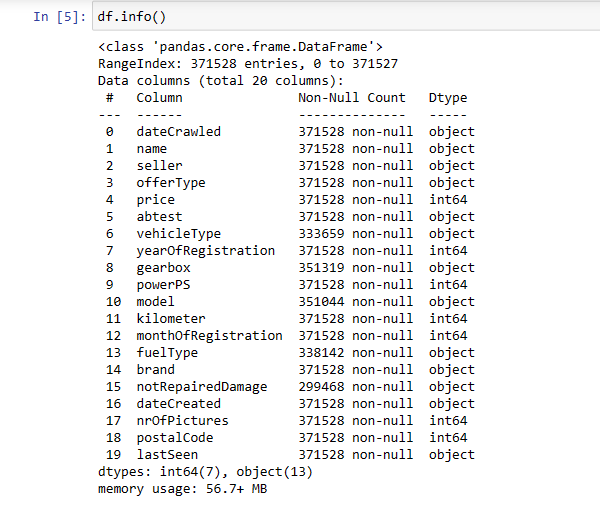
**Data set description:**

The data set has 371528 data points. This data set is about the different vehicles available for sale.

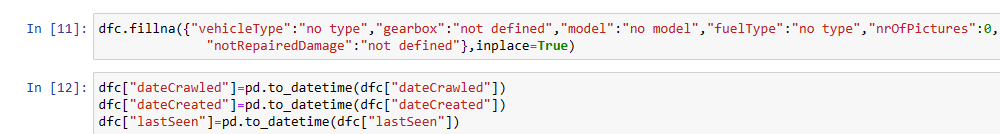
The data set has total 20 columns in that 7 are int64 data type and remaining 13 has object type.

It takes total 56.7+ mb memory space.

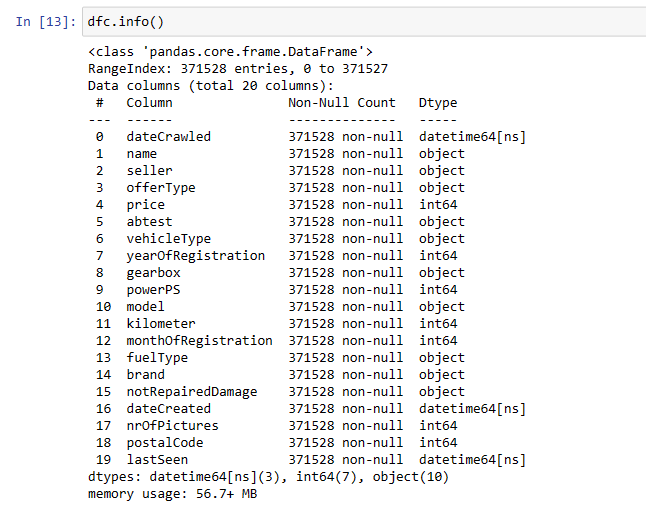
**Data Cleaning:**



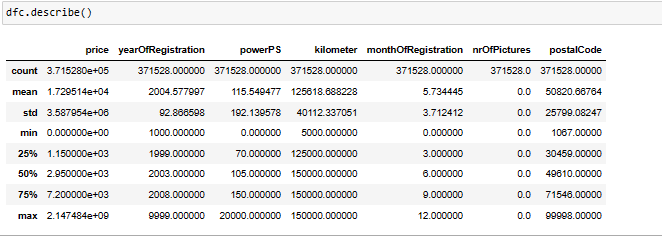
First, we fill the all-nan values and we’ll change the data type of date crawled, date created and last seen columns to datetime data type using this code.



Then we got the info of data set like this.



By using Descriptive statistics, we got a summary of our entire data.

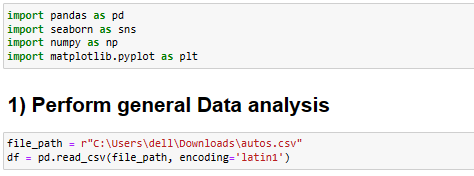


From that in the column price we can tell the max price of vehicle which is for sale is 2.14e+09, minimum is 0.0e+00 and standard deviation is 3.5879e+06.

From that in the column powerPS max power is 20K, minimum power is 70 and standard deviation is 192.1395.

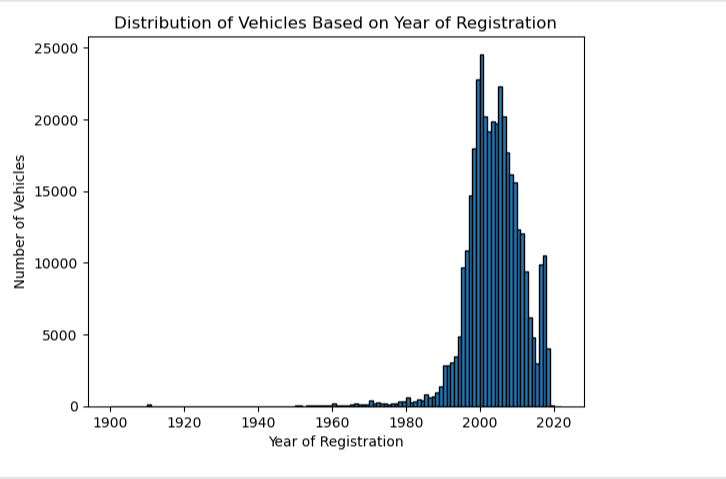
**2.** **)Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot.**

I have the cleaned data set with me and I import the modules which I may use for plots.



The plot looks like this for distribution of vehicles based on year of registration with the help of the given code.



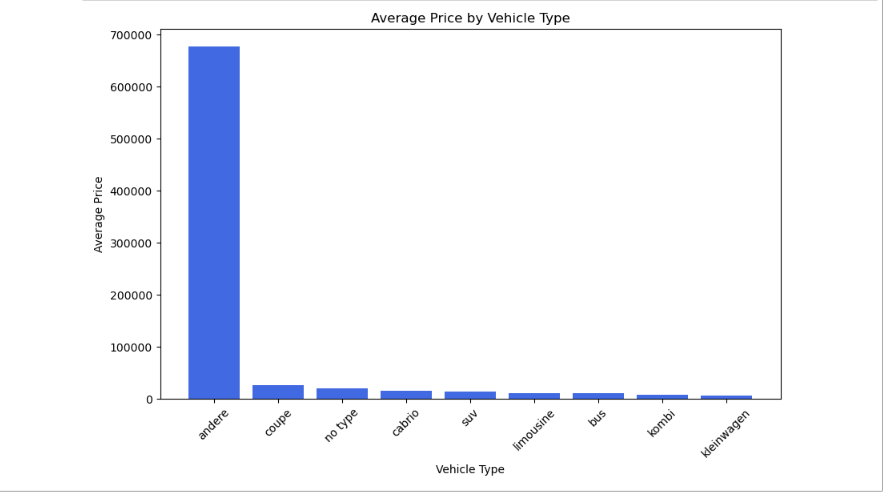


From the above plot I analyse that in the year 2000 more vehicles are produced. Nearly 25k vehicles are produced in that year.

And the less vehicles are produced in the years 1961-64.

**3.** **Create a plot based on the Variation of the price range by the vehicle type.**

From the below code we can get the plot.



From the above plot I can tell that the vehicle andere type has more price and vehicle type SUV has less price.

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

To find the total count of vehicles by type I used groupby() function.

A screenshot of a computer

Description automatically generated

From that group by data I have created a new data frame named da.

And then I plot a bar plot on that.

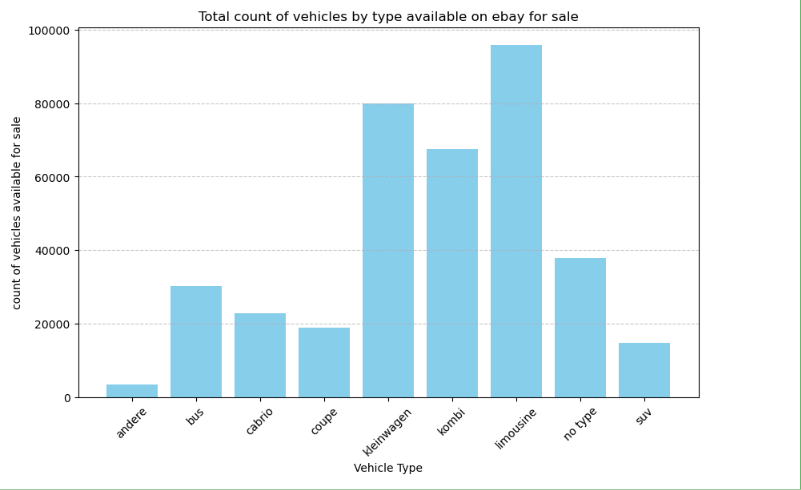
A screenshot of a computer

Description automatically generated

The code I used to plot a bar plot.



And the plot I got.

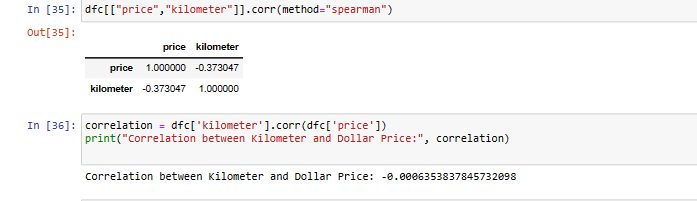


From the above plot I can tell that limousine vehicle type have a greater number of vehicles with the count of 95894.

And andere vehicle type has least number of vehicles with the count of 3357.

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

Has the relationship between price and the kilo meter is as the price increases the kilo meter decreases.to find that I used Pearson correlation co-efficient.



Because of the Pearson correlation co-efficient is in negative that means the price increases the kilo meter decreases.