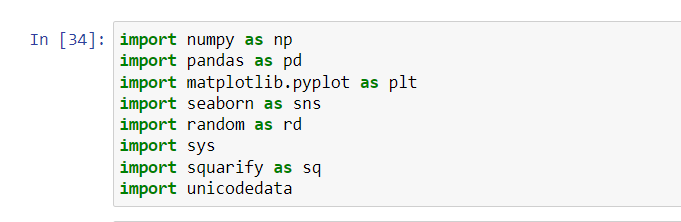
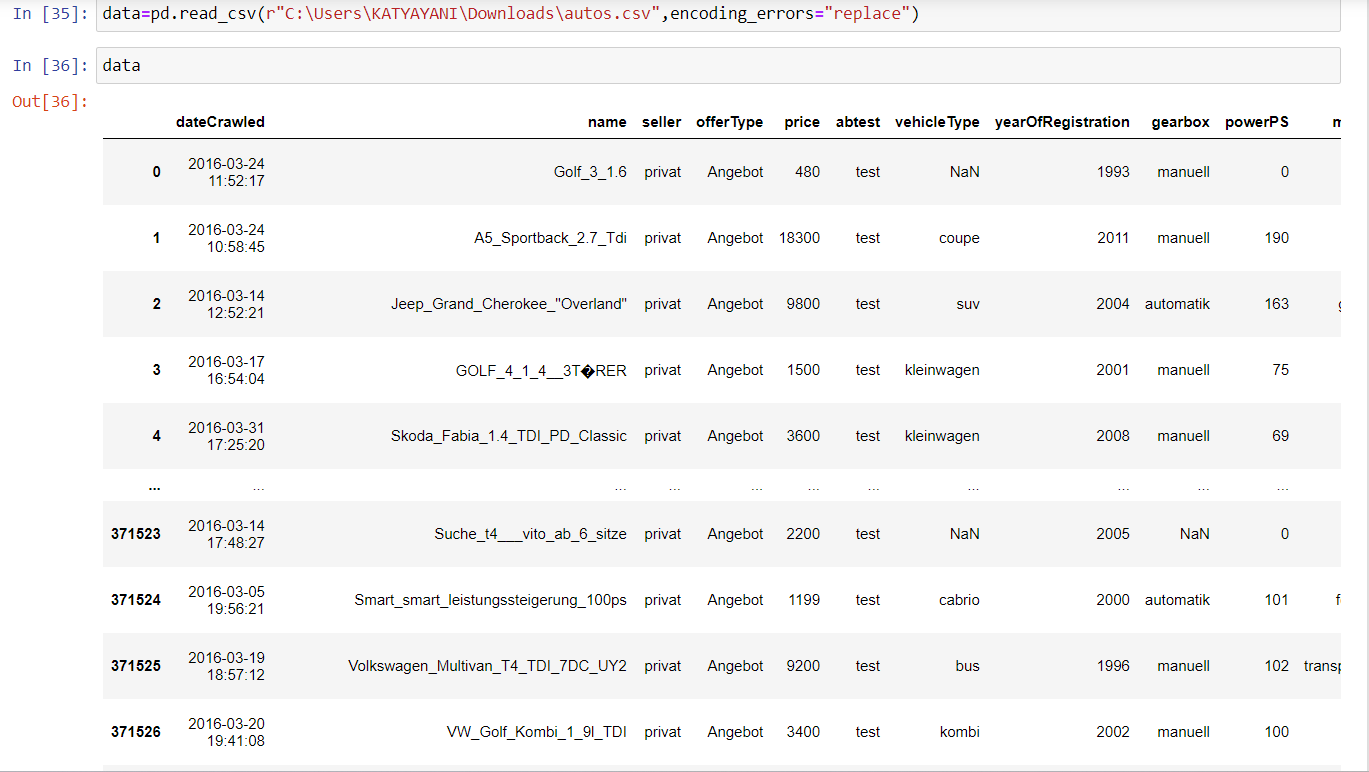
## Performing a general data analysis

**Importing libraries:**

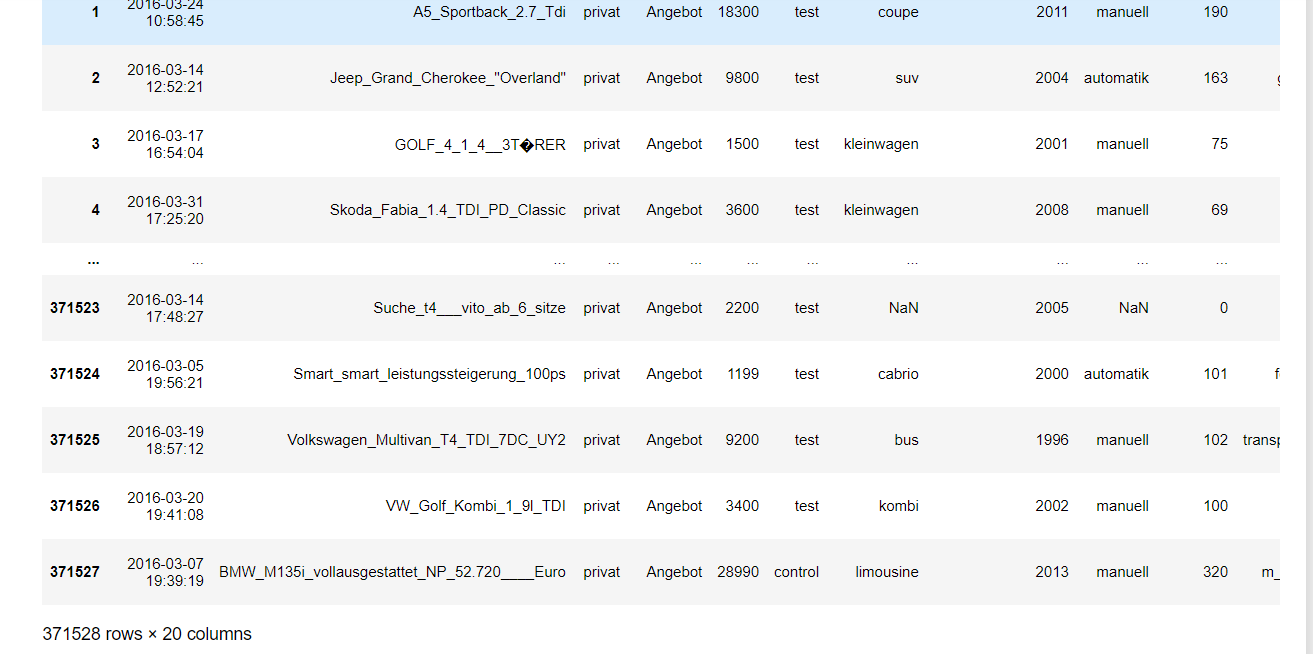
1. First of all I imported all the libraries into the jupyter notebook to perform analysis.



1. second step by using pandas library I read the csv into data frame.

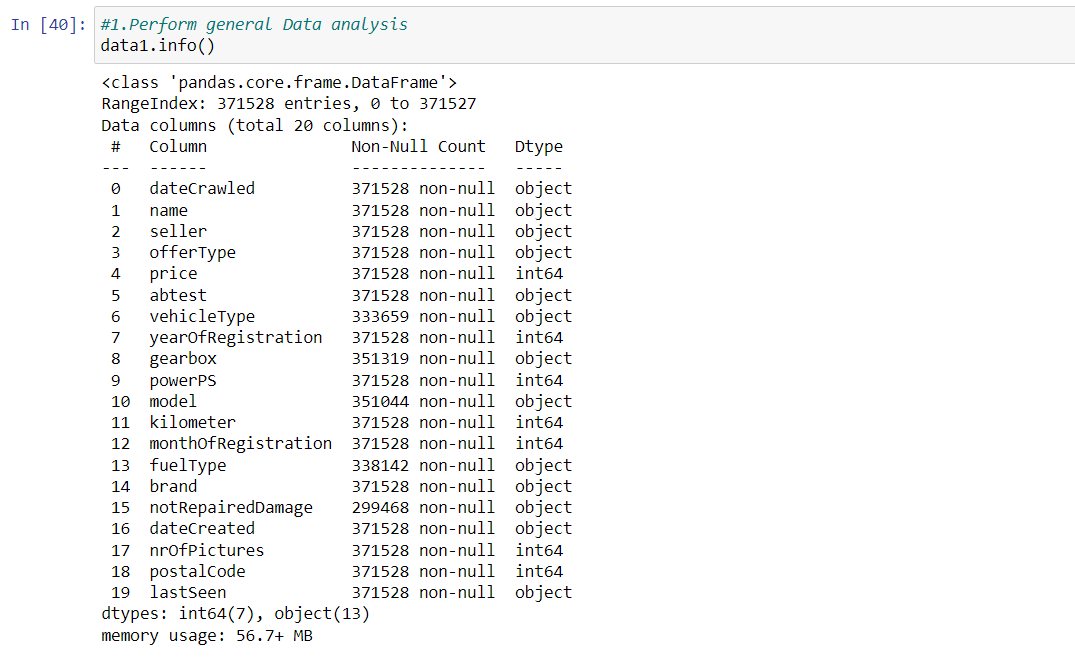


1. by converting csv file into data frame I can see perfectly how many columns is present in data set.
2. by knowing how many colums and rows presented in data we can perform analysis easily.
3. by seeing the below picture we can known that how many rows and columns present in data set.

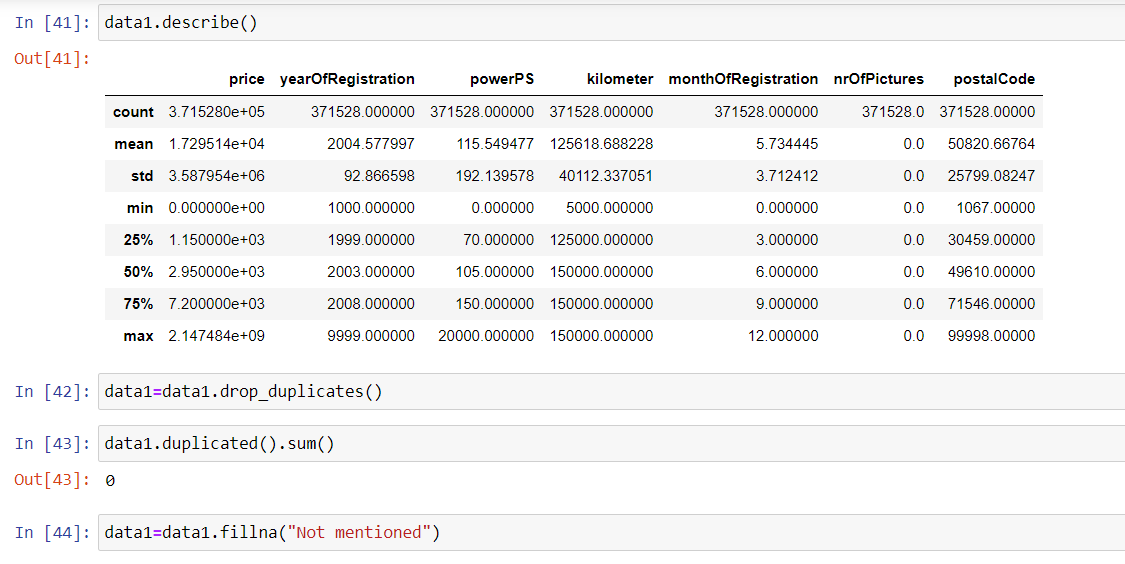


1. By seeing the data set I founded that The data set has 371528 data points. This data set is about the different vehicles available for sale.the data set has total 20 columns in that 7 are int64 data type and remaining 13 has object type.It takes total 56.7+ mb memory space.

#1.Perform general Data analysis



By using info function we can known that which column contain what type of data I found that there are 19 columns in the data set and in that 13 columns contain object data type .remaining int data type.

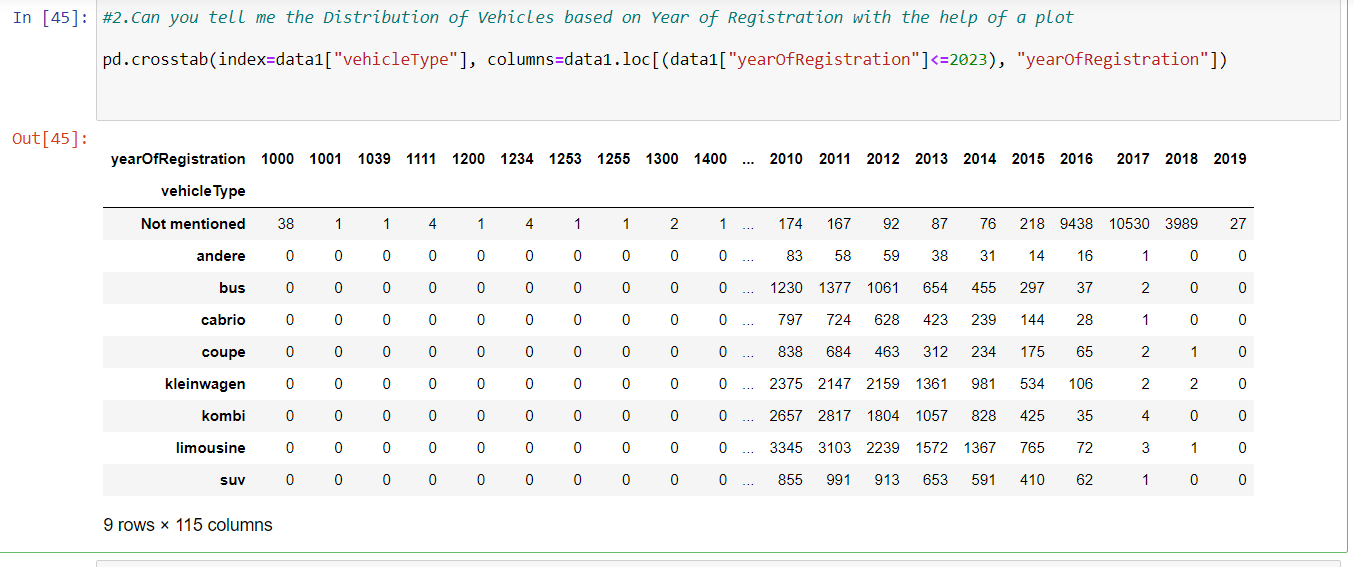


By using discribe function we can known that basic information about the data set.

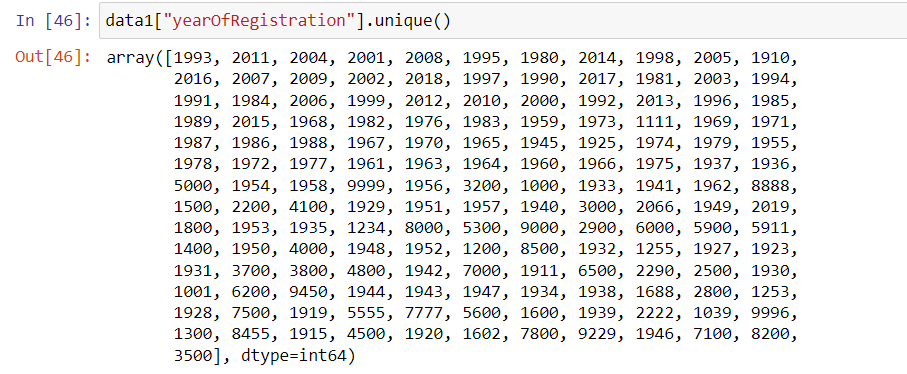
Next I copied the data into another variable called as data1 because while doing analysis if we did any mistake then copied data will be correpted original data will not distroy so that we have to copy the original data into another variable.

After I cleaned the data where na values is present on that place I replaced with “not mentioned” because if the data set contain nan values while doing analysis we will get error so we have drop nan values or we have to replace with another value.

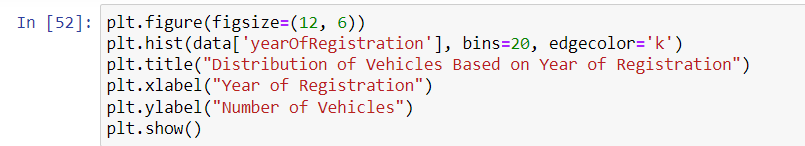
**#2.Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot**

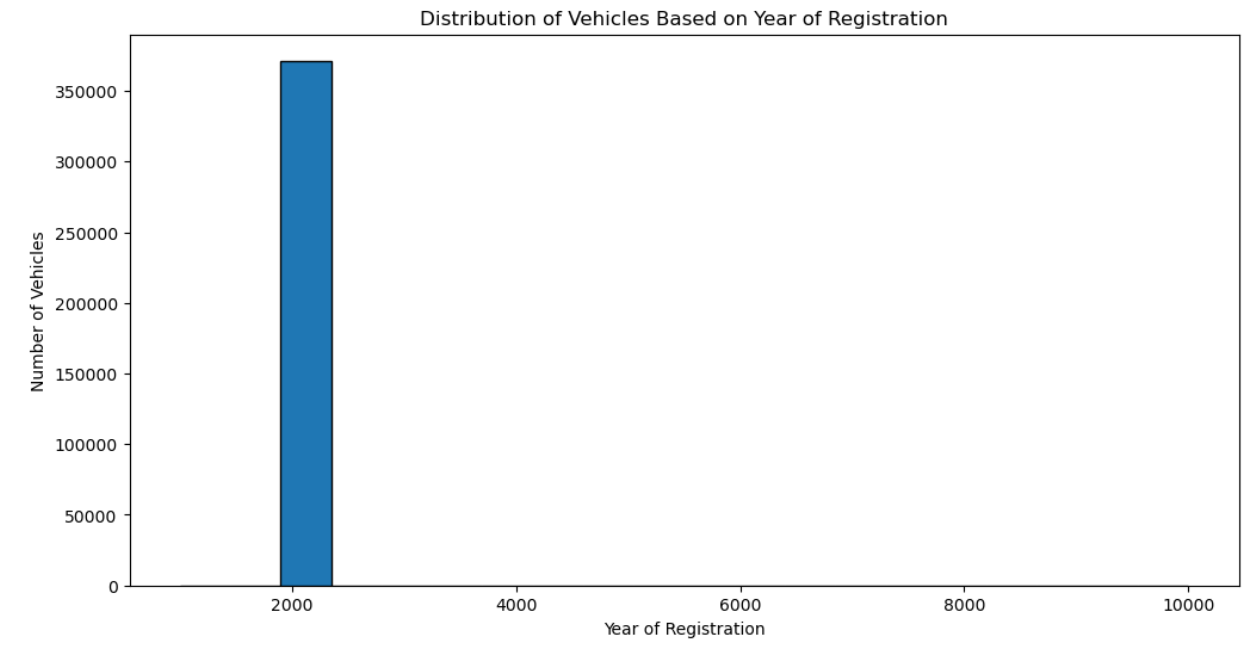
****

By using cross tab I found that Distribution of Vehicles based on Year of Registration . after that I represented that same distribution in plot shown in below picture.



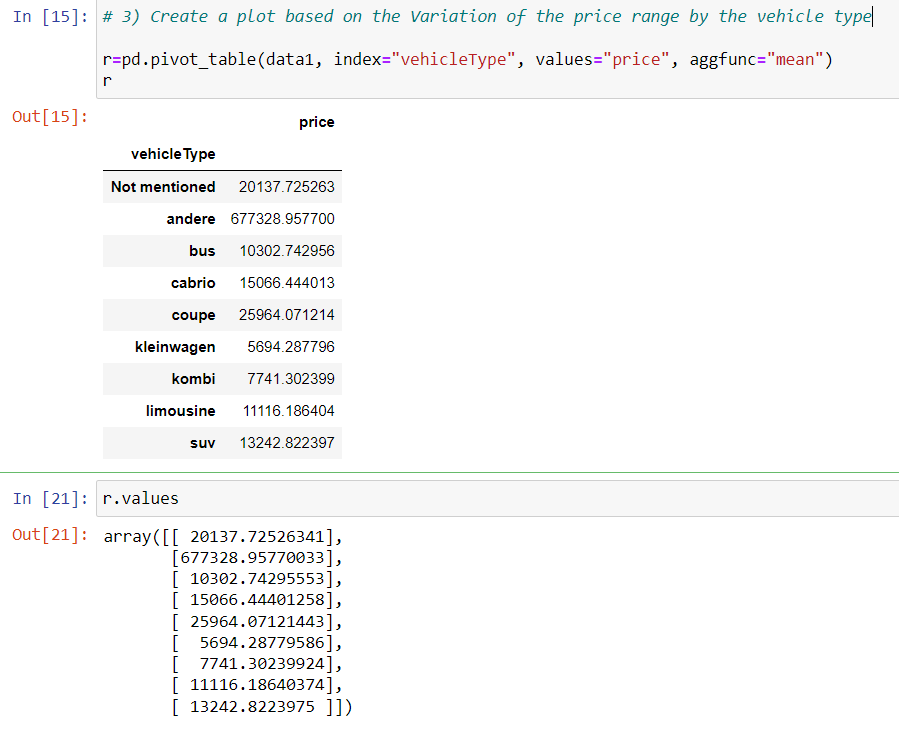
By using unique function I found that how many years present in the data set.

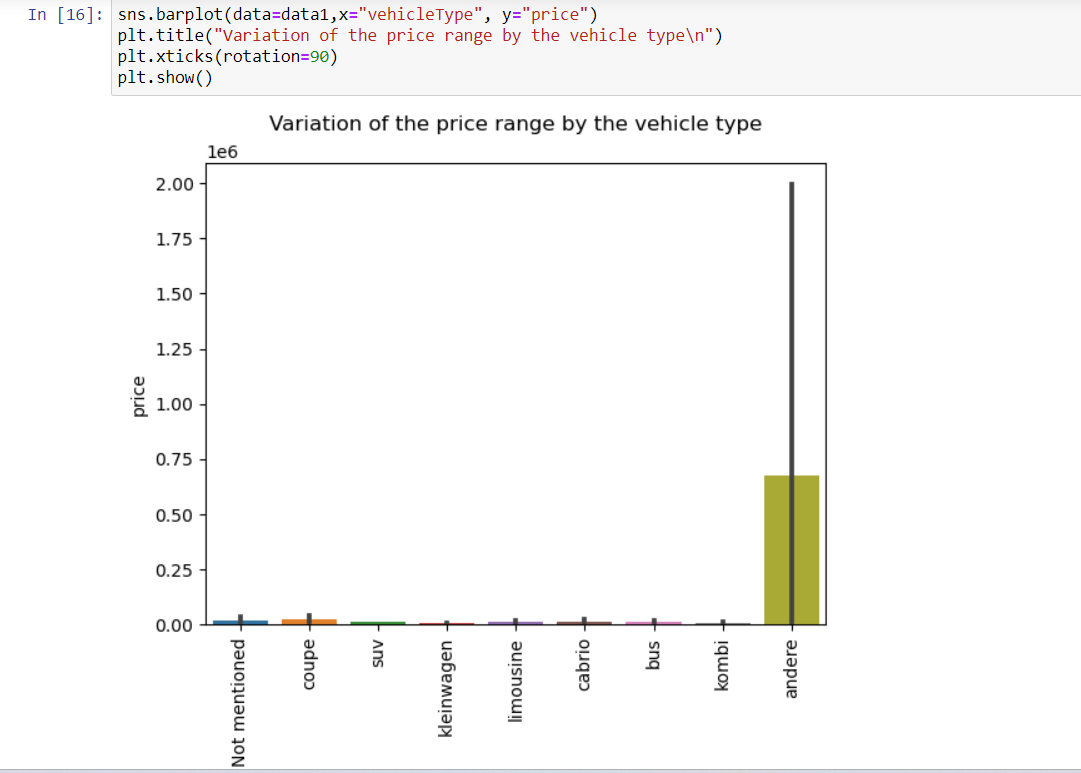




* From the above plot I analyse that in the year 2000 more vehicles are produced. Nearly 25k vehicles are produced in that year.
* And the less vehicles are produced in the years 1961-64.
* By using seaborn library and hist plot I represented that which year had highest production.

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

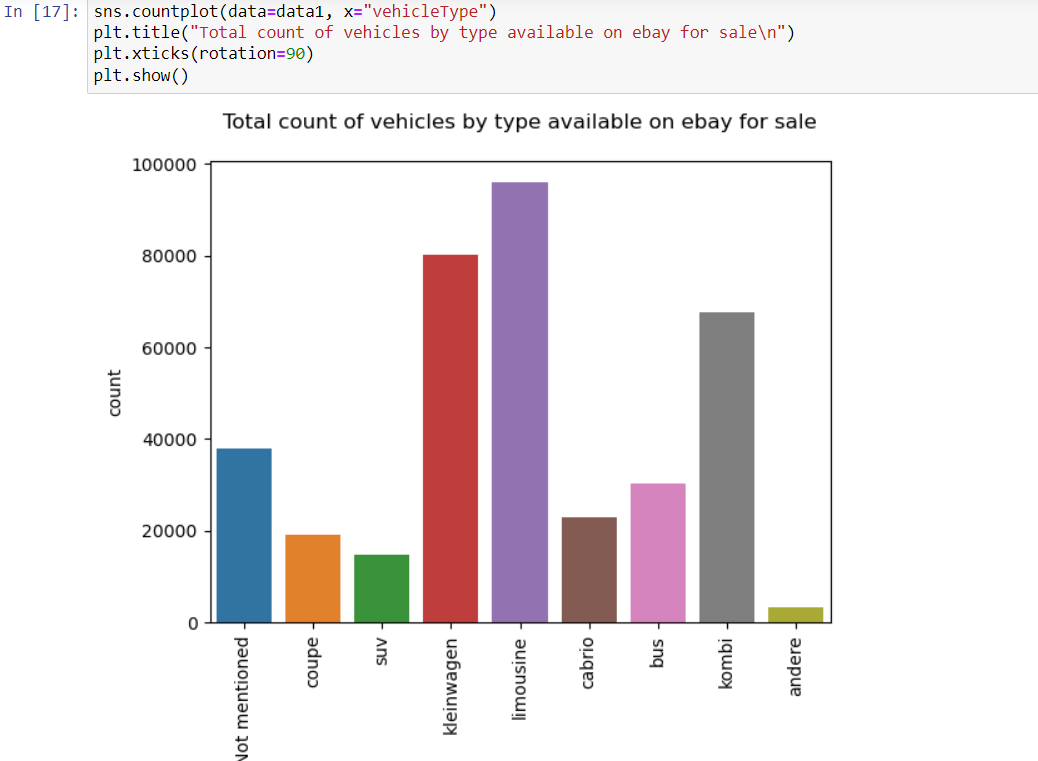




By using pivot table I founded that Variation of the price range by the vehicle type. By using bar plot I found that the vehicle andere type has more price and vehicle type SUV has less price.

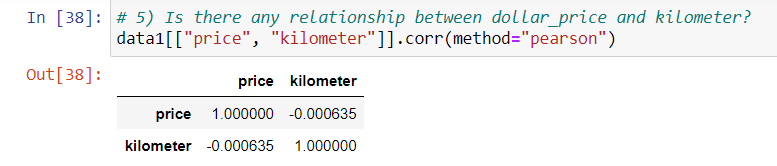
# 4) Find out Total count of vehicles by type available on ebay for sale.As well as create a visualization for the client

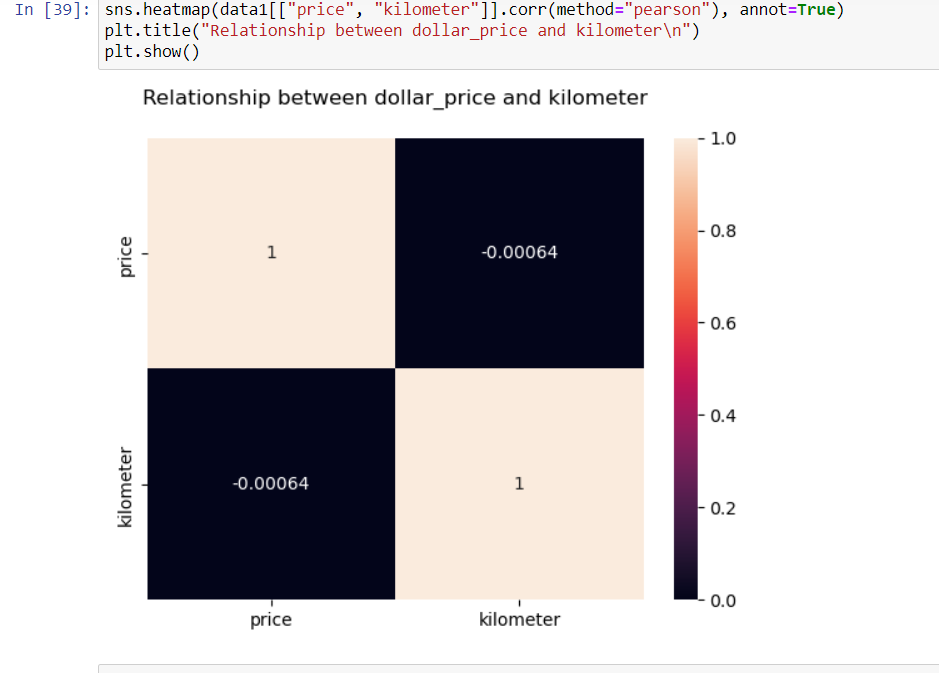




* From the above plot I found that limousine vehicle type have more number of vehicles with the count of 95894.
* And andere vehicle type has least number of vehicles with the count of 3357.

# 5) Is there any relationship between dollar\_price and kilometer?





* 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.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*END\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*