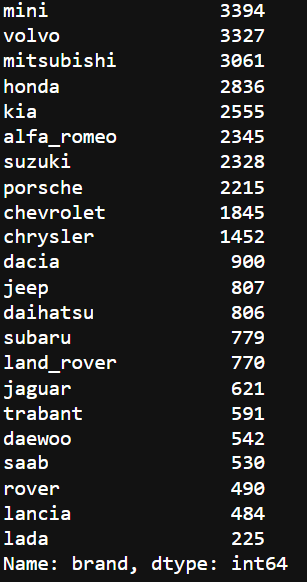
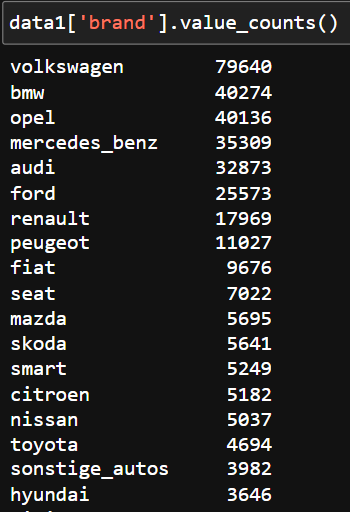
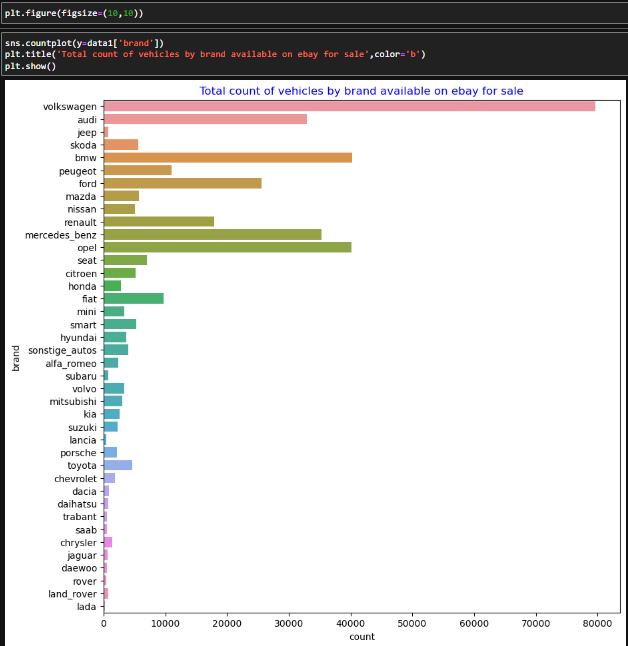
REPORT ON ANALYSIS - 2

1.Can you tell me No of Vehicles by Brand Available on ebay for sale with the help of visualization?

* To know the number of brands available on ebay can be done by using bar plot or count plot.
* Both gives us the result of brand names on x-axis and count or frequency on y-axis for that we have to collect the required data column from data which is as follows,
* data1["brand"].value\_counts()
* The above code will give us the count of all brands and their count of number of times it’s sales as shown below.



* We are going to see the visualization of the above table with the help of plot.



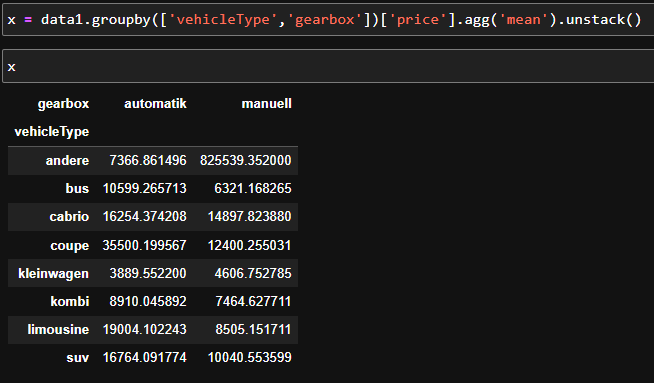
* From the figure we can say that this is a plot of brand and its count, Where “Volkswagen” has the highest count of sales on ebay and “lada” brand is the brand with least count of sales there are many brands which are sold between the count of 0-80000.
* The above visualization is done using count plot by applying it on the “brand” column.

2. What is the Average price for vehicles based on the type of vehicle as well as on the type of gearbox? Explain me with both numerical and visualization analysis.

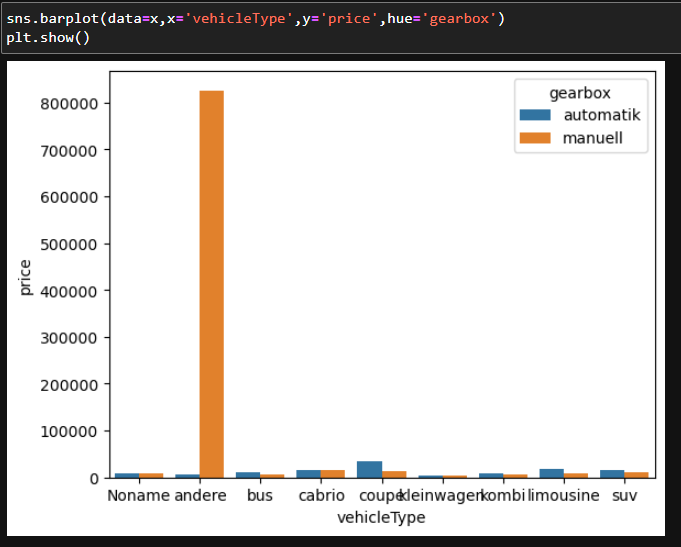
* To show the average price of vehicles based on the vehicle type and gear type we should use bar plot for this
* To collect the average prices of both the vehicle type and gear box at a time
* We should use the groupby method and in the groupby method we should use aggregate function as mean for the average for the price column
* By using the code given below we can extract the data.

“x = data1.groupby(["vehicleType","gearbox"])["price"].aggmean().unstack()”

* This will give the data as shown below.



* The above image shows the vehicle type and gearbox type and mean of price accordingly.
* We can visualize the same by using plot. So, we can understand clearly.



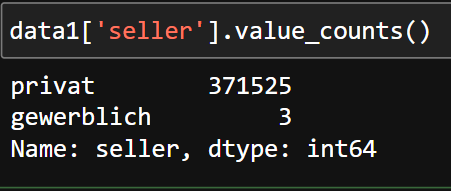
* From the visualization from the above we can say that the vehicle name containing “andere” and gearbox as manuel has the highest average prize among all the vehicle combinations of vehicle type and gear box.

3.What is the marginal probability of private seller?

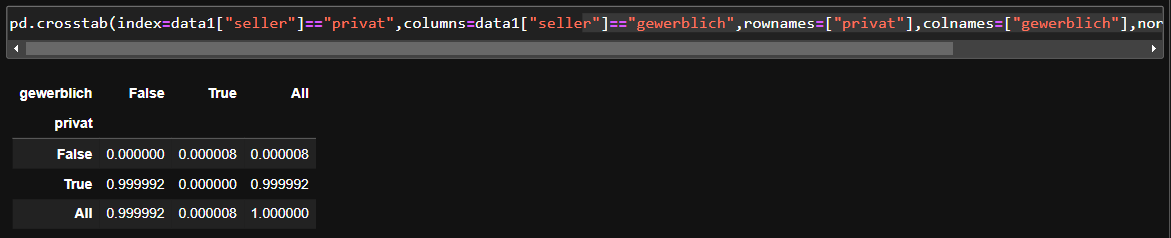
* To find the marginal probability for the “private” in the seller column first we should know about the all types of sales in the seller column.
* For that we need to access the seller column using “value\_counts”.

data1["seller"].value\_counts()

* After executing the above code the below data is given.



* From the above we can see that “privat” seller has more sales than “gewerblich”.
* “Privat” seller has 371525 sales and “gewerblich” has only 3 sales.
* By using the crosstab we can see the marginal probability of the columns we need. The code given below is used to find the marginal probability.
* “pd.crosstab(index=data1["seller"]=="privat",columns=data1["seller"]=="gewerblich",rownames=["privat"],colnames=["gewerblich"],normalize="all",margins=True)”



* In the above code we use crosstab for the seller column to combine all the private seller to the gewerblich, to get the marginal probability we use normalize = ”all” will gives the joint probability of all the type.
* When we use the parameter in the code “margin = True” the margin option in the data will be activated and gives the marginal probability as shown above.
* This will represent that the gewerblitch as a column and private as a row.
* All the data point are cross tab.
* And the to know the marginal probability we can confirm that the all.
* Values true column gives the marginal probability for the private seller.
* Here the marginal probability for private seller is 0.999992.
* Here the marginal probability for not a private seller is 0.000008.
* Here the marginal probability for gewerblitch seller is 0.000008
* Here the marginal probability for not a gewerblitch seller is 0.999992.