**perform general data analysis:**

1. **A1.First of all I imported all the libraries into the jupyter notebook to perform analysis.**

* First point data set has 371528 data points this data set is about the different vehicles available for sale on eBay.
* Data set has total 20 columns in that 7 are in 64 data type and remaining 13 has object data type.
* it takes total 56.7 + MB memory space.
* After cleaning the entire data set by using fillna() and changing some columns to date time data type Now it has 3 columns which have datetime data type and 7 columns which have int 64 data type and 10 columns which have object data type.
* By using descriptive statistics I can tell that maximum of prize is 2.147e+09 ,minimum price is 0 and standard deviation is 3.5 8E + 06.
* In the column power Ps the maximum power is 20,000 and the minimum power is 0 and the standard deviation is 192.139.
* In the column kilometre maximum kilometres is 15 K and minimum kilometre is zero understand deviation is 3.712.

1. **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.

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

* By using the below code I’ll show the visualization of me no.of vehicles by brand available on ebay for sale.
* From the above plot I can tell that Volkswagen brand has more number of vehicle available on ebay for sale and the count is 79640.
* And the brand lada have less number of vehicles available on ebay for sale with the count of 225.

***#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***

* From the above plot the average price for vehicles based on vehicle type is high for andere with the 677329.
* And kleinwagen ha less price with 5695.
* For average price for vehicles based on gearbox type.
* To find the average price for vehicles by gearbox type I used pivot table() function.
* from the above plot I can tell that the vehicles which don’t have any type has more average price when compared to others with 44019.
* And automatic has less average price with 15146.

**#3.What is the marginal probability of private seller**

* From that the marginal probability of private seller is 0.99.
* For above analysis I used heat map.

**C1**.**The memory usage of the data is around 6.1 mb. How can we reduce the memory usage of the data set?**

* The memory usage of the data set is as you can see 59.5+ MB.
* I dropped some columns which are not use full and I changed the data type of some columns.
* After dropping those columns and changing the data types of some columns the memory usage of the data set is decreased to 41.8+MB.

**2.what is the average price of vehicle by fuel type and gearbox type And give a plot.**

* From the above plot I can tell that andere fuel type in manuell gearbox type has more average price when compared to all others.
* To represent this I used heat map.

**3.What is the Average power of a vehicle by vehicle type and gearbox type Give a plot.**

From the above plot we can see that coupe vehicle type has more average price when compared to other vehicle types. And in that coupe vehicle type automatic gear box type has more average price than other two gear box types.

**4.What is the Average price of a vehicle by brand as well as vehicle type.Use heatmap to explain this.**

* to plot a heat map visualization first I have created a cross table for mean of prices based on brand and vehicle type.
* With the help of that cross table I have created a heat map where on X axis we take vehicle type and on y axis we take brand.

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