REPORT ON SALES OF AUTOS

ANALYSIS-3

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

To reduce the data size from the data we have to use some techiniques in data using numpy and pandas

first of all we have to know the memory size of our data to know that we use the code as

datac.info()

or we can use the code as

memory=datac.memory\_usage().sum()

meminMB=memory/(1024\*\*2)

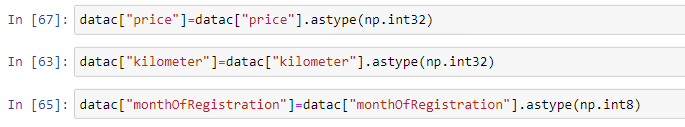
meminMB

will gives you the memory size of the data

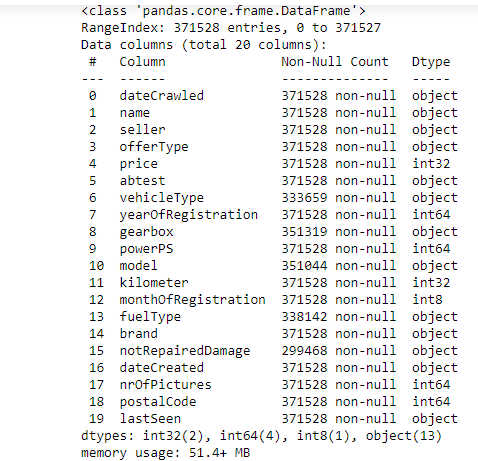
🡪here the size of the data is 56.690799713134766

i.e., approximately equal to 56.7 MB

🡪 To reduce the size of our data we can change the datatypes of the columns based on the precedence of the type of data and requirements as shown below



Now we see the memory size of the data

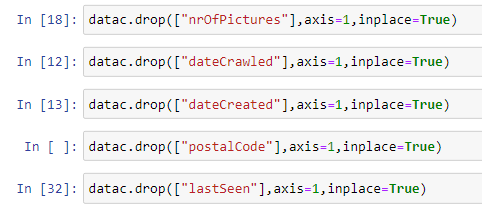


Here we can observe that the memory size is 51.4 MB but our data first was at 56.7MB so the data Is reduced to 5.3Mb

* To reduse this data still more we can remove the unnecessary columns
* present in our data we have some columns which can not be used at all

so it is necessary that we can remove the columns permanently from the data which are [dateceawled,nrofpictures,postalcode,lastseen]

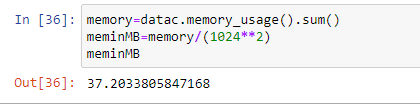
these columns are removed by using the code as



Here all the mentioned columns are temporarily removed from the data

Due to unnecessary data stored

Now we can look at the size of memory weather is redused or not by using the code as



This picture shows the memory size of our data after the modifications done

So we can see that the memory size of our data is 37.2033805847168 MB

which is far less than the data before modification i.e., ~56.7 MB

so the memory size of our data is reduced from 56.7 MB to 37.2 MB

we have saved the memory size of 19.5MB ~ equal to 20.0 MB

2) What is the Average price of vehicle by fuel type and gearbox type.Give a plot

To show the average price of vehicles based on the fuel type and gear type we should use bar plot for this

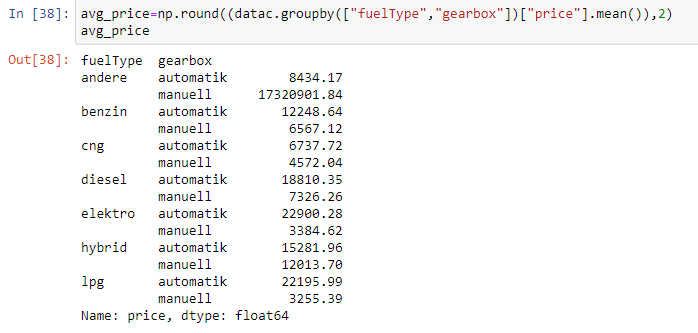
To collect the average prices of both the fuel type and gear box at a time

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

This is the code for the data to be collected ….

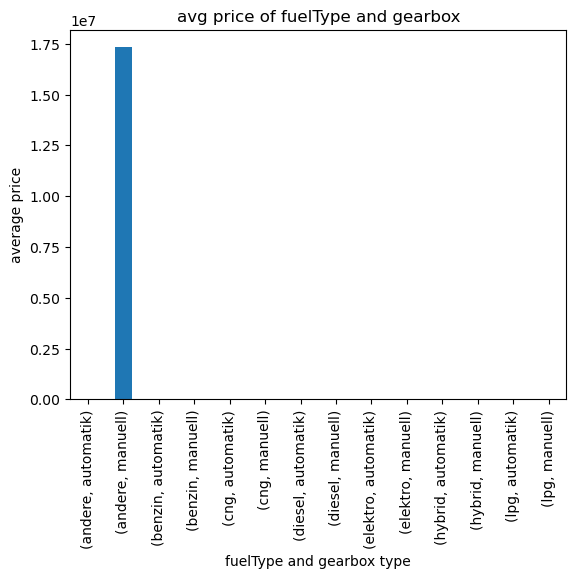
avg\_price=datac.groupby(["fuel Type","gearbox"])["price"].mean()

this will gives us the data as shown



This can also be shown by using the bar plot as the code avg\_price.plot(kind="bar")  
plt.xlabel("fuelType and gearbox type")  
plt.ylabel("average price")  
plt.title("avg price of fueltype and gearbox")  
plt.show()

this is the code for the data which gives you the plot as follows



From the figure above we can say that

The fuel name containing andere and gearbox as manuell has the highest average prize among all the vehicle combinations of vehicle type and gear box aswell

Due to the outlier in the fuel type the average price is effected so that the only one bar is at the highest and all other are lower

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

To show the average power of vehicles based on the vehicle type and gear type we should use bar plot for this

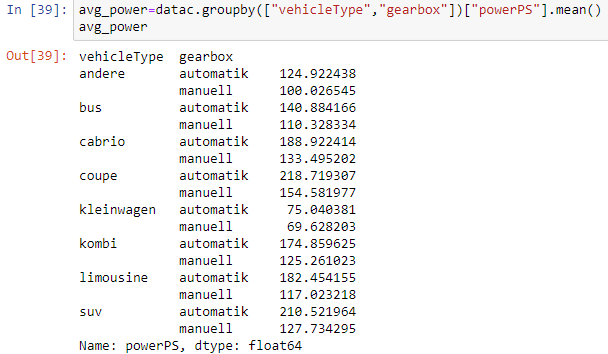
To collect the average power of both the vehicle type and gear box at a time

We should use the groupby method and in the groupby method we should use aggregate function as mean for the average for the power column

This is the code for the data to be collected ….

avg\_ power =datac.groupby(["vehicleType","gearbox"])["price"].mean()

this will gives us the data as shown

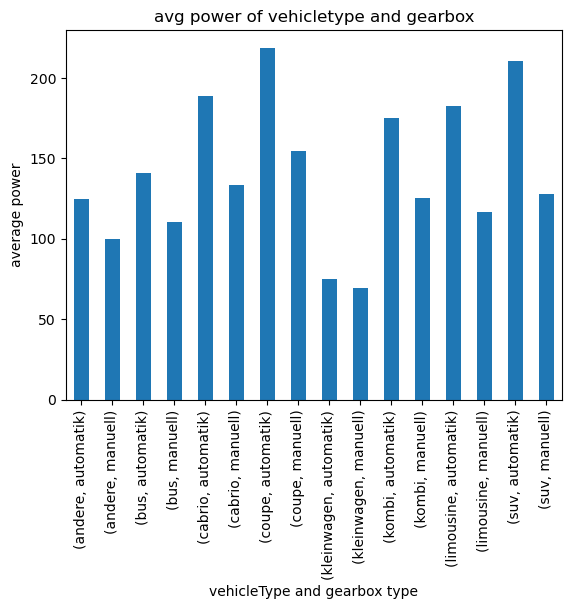


Here shows that the all average power about the combination of vehicletype and gearbox

This can also be shown by using the bar plot as the code

avg\_ power.plot(kind="bar")  
plt.xlabel("vehicleType and gearbox type")  
plt.ylabel("average power")  
plt.title("avg power of vehicletype and gearbox")  
plt.show()

this is the code for the data which gives you the plot as follows



From the figure above we can say that

The vehicle name containing coupe and gearbox as automatik has the highest average power of 230 among all the vehicle combinations of vehicle type and gear box aswell

These bar plot gives us all the combination of all the vehicle type and gear box type average powers

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

To show the average price of vehicles based on the vehicle type and brand by using heatmap plot for this

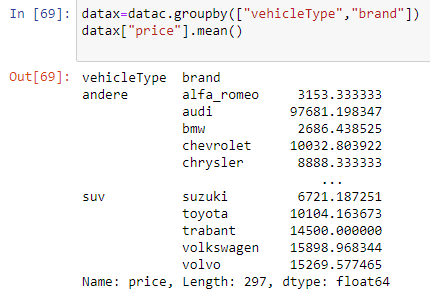
To collect the average prices of both the vehicle type and brand at a time

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

This is the code for the data to be collected ….

avg\_price=datac.groupby(["vehicleType","brand"])["price"].mean()

this will gives us the data as shown



This above picture shows you the average prices of all the brand and vehicle types as well at a time

This can also be shown by using the plot heatmap as the code

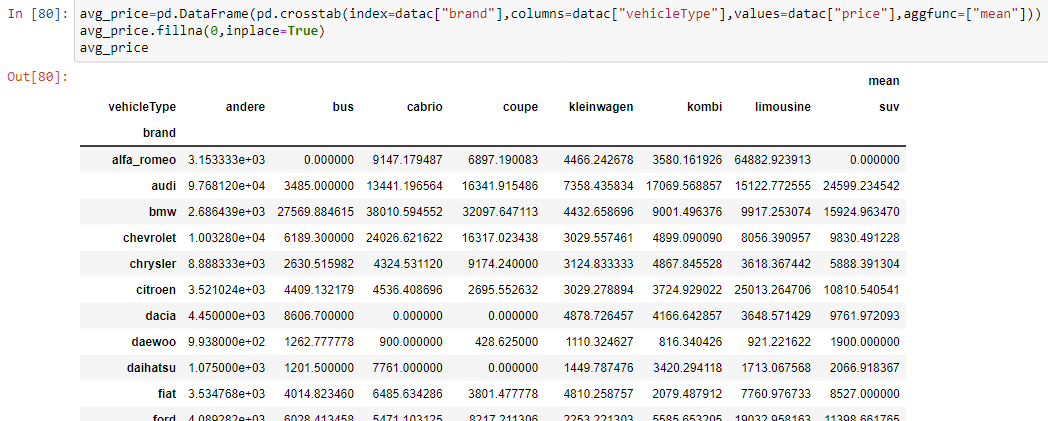
But before plotting the data we have some null values inside the data so we have to fill the null values with zero as follows

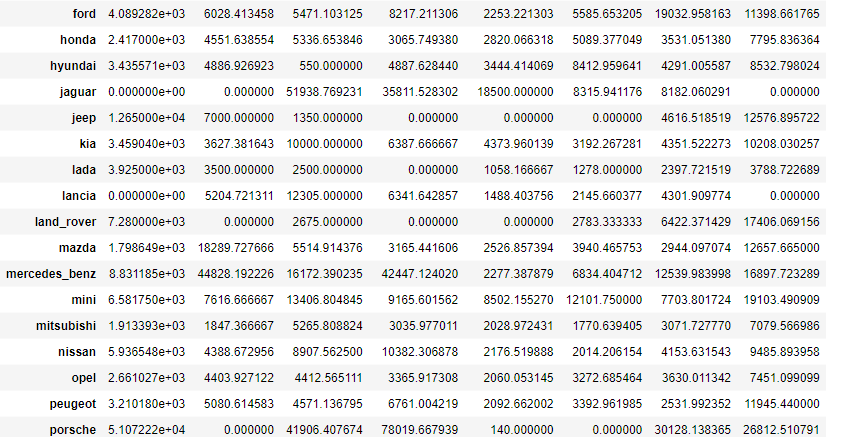
avg\_price=pd.DataFrame(pd.crosstab(index=datac["brand"],columns=datac["vehicleType"],values=datac["price"],aggfunc=["mean"]))

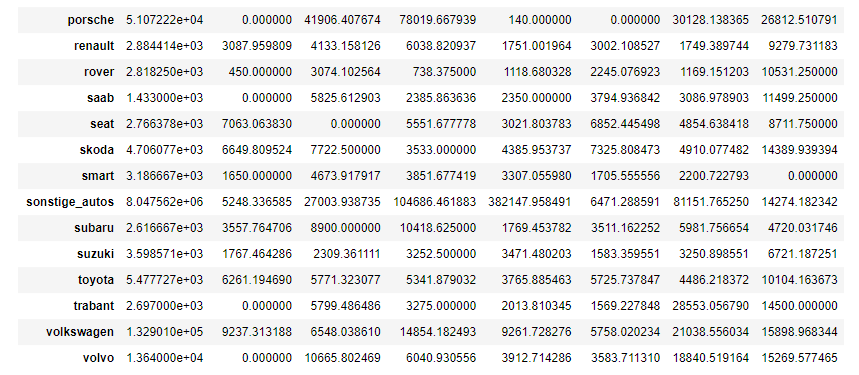
avg\_price.fillna(0,inplace=True)

avg\_price

this will gives the data frame result as







Hence all the data null values are filled with “0”

In the dataframe

To plot the data heat map for the data we we use the following code as

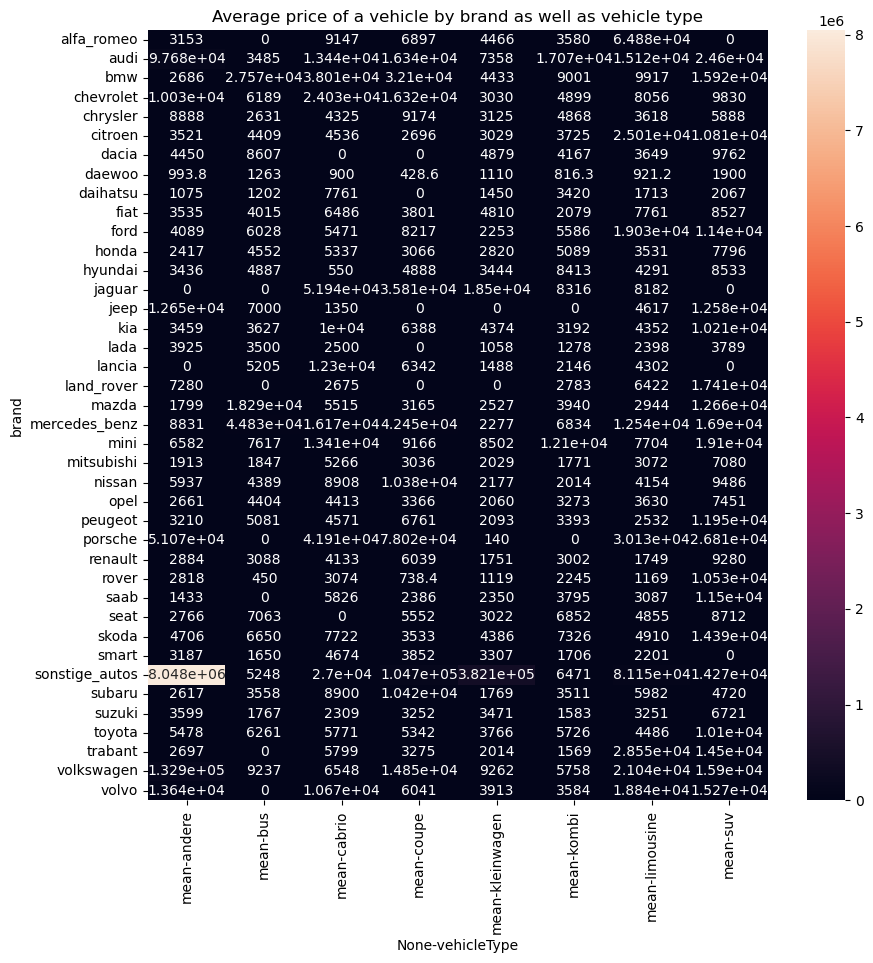
plt.figure(figsize=(10,10))

sns.heatmap(data=avg\_price,annot=True,fmt=".4g")

plt.title("Average price of a vehicle by brand as well as vehicle type")

plt.show()

where this above code gives us a heatmap plot for the average price for the vehicle type and brand as well



Here the image of heatmap shows that the all values of average prices for the vehicletype and brand at a time