REPORT ON SALES OF AUTOS

ANALYSIS-2

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

To say that number of brands available available on ebay by using visualization

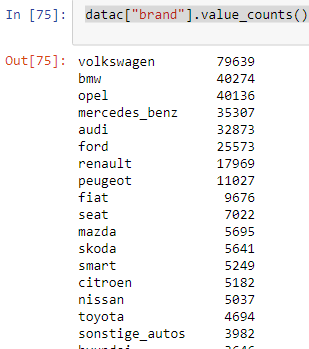
This can be done by using the bar plot or count plot

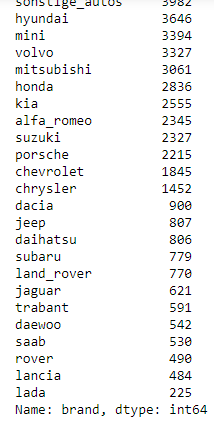
Both gives us the result of brand names on x-axis and count or frequency on y-axis

🡪before that we have to collect the required data column from data which is as follows

datac["brand"].value\_counts()

which will gives you the collected all brands and their count of number of times it sales as below





Now to plot the data we can use the count plot using the seaborn library

As the code shown below

sns.countplot(y=datac["brand"])

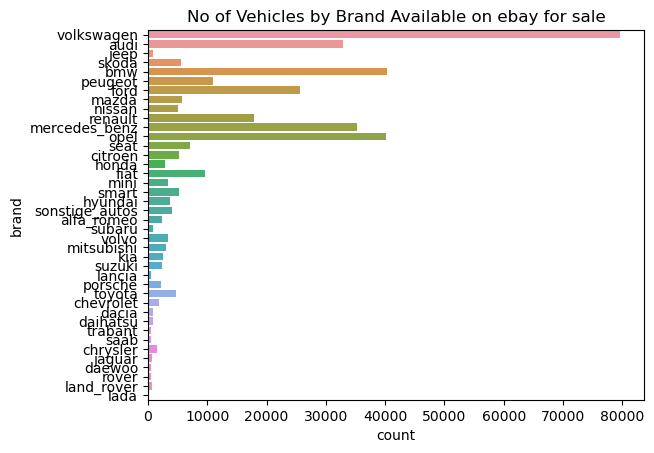
plt.xlabel("count")

plt.ylabel("brand")

plt.title("No of Vehicles by Brand Available on ebay for sale")

plt.show()

which gives you the plot as



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 e bay

And lada brand is the least count of saled

there are many brands which are saled between the counts of 0-80000

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

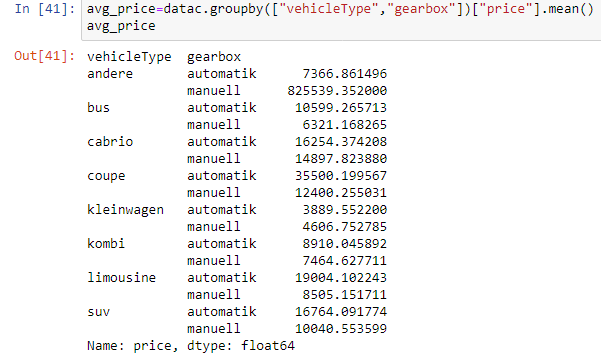
To collect the average prices of both the vehicle type and gear box at a time

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

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

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

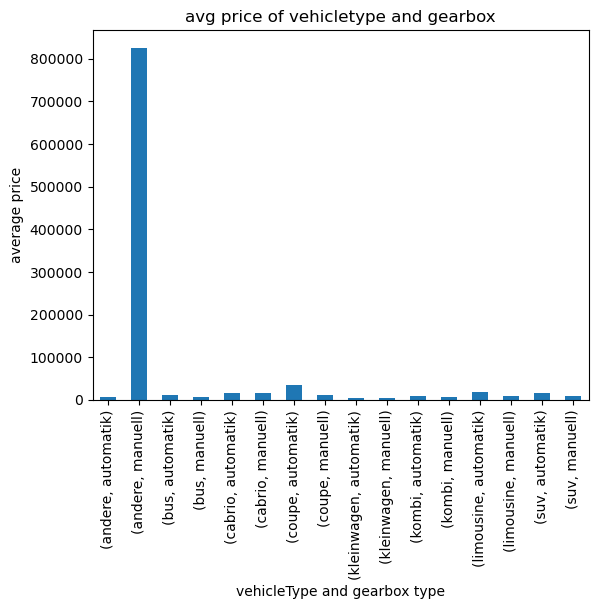
this will gives us the data as shown



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

This can also be shown by using the bar plot as the code avg\_price.plot(kind="bar")  
plt.xlabel("vehicleType and gearbox type")  
plt.ylabel("average price")  
plt.title("avg price 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 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 vehicle type the average price is effected so that the only one bar is at the highest and all other are lower

1. What is the marginal probability of private seller

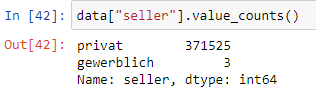
To find the marginal probability for the “private” in the seller column

First we should know about the all types of sales in the seller column

For that we need to collect the seller column data as

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

this will gives you the information as



Here we have 371525 count of private sellers in the total sellers

And only 3 gewerbich sellers in the data

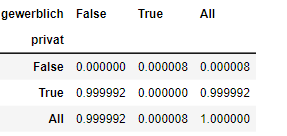
To get the marginal probability for the data we use the code as

pd.crosstab(index=datac["seller"]=="privat",columns=data["seller"]=="gewerblich",rownames=["privat"],colnames=["gewerblich"],normalize="all",margins=True)

here in this 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 below



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