Summary on sales of autos

Analysis 1:-

🡪In the analysis of data primarily we have to load the data in the jupyter notebook using pd.read\_csv(copypath) but there the data is at different format so I changed it using encoding=”utf-8” saved the data as data and loaded in pandas  
🡪later to stare the data safe I used datac=data.copy() for not effecting the original data

🡪finding duplicates as datac.duplicates(),

For unwanted columns I used the code for individual column as datac[column name].value\_counts()  
🡪and I have done the basic analysis like removing duplicates as

datac.drop\_duplicates(inplace=True)

and to remove unwanted columns and changing the data types of columns on requirement.

I have use the code as

datac.drop(["nrOfPictures"],axis=1,inplace=True)

datac.astype({"price":int64,"kilometer":float})

datac["dateCrawled"]=pd.to\_datetime(datac["dateCrawled"])

datac["yearOfRegistration"]=pd.to\_datetime(datac["yearOfRegistration"])

datac["dateCreated"]=pd.to\_datetime(datac["dateCreated"])

datac["lastSeen"]=pd.to\_datetime(datac["lastSeen"])

🡪finally I also have assigned the nan value with its object as datac.fillna({"notRepairedDamage":"no name","fuelType":"no fuelType","vehicleType":"no vehicleType","gearbox":"no gearbox","model":"no model"},inplace=True)

🡪Finally I have reviewed the data to check weather it is clear or not by usind the code as datac.info()

So the basic data analysis is done!

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

🡪 For this I have used the point plot to show how the data is distributed In a plot

As bi-variate analysis

The data of one variable on x axis and another variable as y-axis

🡪I plot the data using seaborn as sns.pointplot(datac,y=datac["yearOfRegistration"],x=datac["vehicleType"])

🡪It gave the final analysis report as plot for each and every individual type of vehicle to each and every individual year

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

🡪 For this I have collected the data of prices and the vehicletype in a data frame

* Now I plotted the data using bar plot as

sns.(datac,x=datac["vehicleType"],y=datac["price"])

* It gives the result as sales of the vehicles by the vatiation of prices as a dataframe
* Now with that data I used the bar plot which gave me the result as

A vehicle type on x-axis on the plot and count of vehicles prices on y-axis

There we can come to conclude that the different prices of vehicles by comparing one to another from the plot

Here the plotted data shows the difference of every vehicles prices with the vehicle type

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

🡪For this I have done the analysis of collecting all the vehicletypes in the data frame and used the code as

datac[“vehicletypes”].value\_counts()

🡪will gives the count of each type of vehicle

🡪It will be plotted in the bar plot as

sns.countplot(x=datac["vehicleType"])

🡪Finally it will shows the data of each individual type of vehicle with its count on another axis

5) Is there any relationship between dollar\_price and kilometer? (Explain with appropriate analysis)

* “YES” there is a relation ship between price and kilometre in the given data
* To show the relation between this I used the code as

datac[["price","kilometer"]].corr(method="spearman")

* It gave the result as matrix form of correlation where the relation between two variables is -0.37
* So finally we can conclude that the two variable are moderately negatively correlated
* To show this in a plotting method I used heatmap as

sns.heatmap(datac[["price","kilometer"]].corr(method="spearman"),annot=True)

🡪finally we concluded that the price and kilometre are negatively corelated

i.e, if kilometer increases price decreases

both are inversely proportioned.