**Summary on sales of autos**

**Analysis 1:**

First we have to load the autos data for analysing and visualisation

For that we need to import python libraries like pandas , numpy, seaborn, mathplotlib, scipy as

Import numpy as np

Import pandas as pd

Import mathplotlib as plt

Import seaborn as sns

Import scipy as s

Import scipy.stat as ss

* After that load the data from the excel or csv file by copying the path as copypath
* But this data has a typical encoding error so the data should be in 'ISO-8859-1'. So data is loaded like below,

data=pd.read\_csv(r"C:\Users\DELL\Downloads\autos.csv",encoding='ISO-8859-1')

To perform general data we have to find the general information about the data like

1. Number of columns in the data
2. Number of rows in the data
3. Number of non null values in every column in the data
4. what are the data types of each column in the data

These are the basic data analysis should be done on the data as follows

To know the basic information about the data we use the code as data.info()

Which will gives u the all values of data columns, rows, datatype of columns, non-null values in the column size of the data

* Here we can confirm that there are total count of 18 columns in the autos sales dataset
* And there are 371528 rows in the dataset also there are some null values in some columns which are
* This is the general analysis of each numerical columns, of mean, median, mode, std and quantiles.
* Here we chance the data type of column also. In the given data dateCrawled is in object data type but it should be in datetime type.

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

1. We have to show how many no of vehicles have registered at that particular year so first we have to extract the data of registered year column and fit in a plot.
2. But the problem here is there are many years out of bound.
3. So we have to handle this by ranging the data to a certain value
   1. using the where function with applying a condition And replacing the other bounded data to the mode of its data values in the data by using the code
   2. data["yearOfRegistration"]=data["yearOfRegistration"].where((data["yearOfRegistration"] >= 1950) & (data["yearOfRegistration"] <= 2023),data["yearOfRegistration"].mode()[0])
   3. here we replaced the data of the column name year of registration which is not lying between 1950 and 2023 will be replaced with the mode of the data

1. Hence the data is converted to the particular minor ranged data.
2. To plot the data we use the code as
3. This image represents that the distribution of vehicles by the year of registration.
4. On x-axis we consider the year of registration and on y-axis we consider the count of vehicles at that particular year.
5. From this distribution we can say that a max no.of vehicles are registered between the years 1980-2018.

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

1. To create a plot variation of price differ from one vehicle type to another vehicle we can select the bar plot.

* Because in bar plot it gives the count of vehicle type according to price.
* we use the code as

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

From the figure we can say that the “andere” vehicle have the highest price of $0.74 and all other vehicle type very least price in the range of $0.12\*10^6-$0.0001.

And kleinwagen vehicle have the least cost price among all 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.

To create a plot on the count of vehicles by its type we can use the count plot for the data.

* Because in count plot we can get the count frequency on y-axis and vehicle type on x-axis.
* Before that we should take the consideration based on vehicle type by using the code (data[‘vehicleType’]).
* We used the code as below.

data[“vehicletypes”].value\_counts()

* 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, we can say the relation between the column price and kilometre which can be explained as below.
* To know the relationship between the columns.
* In the statistics there is a concept called correlation, here we can find the relation between two or more columns.
* And the correlation between the price column and kilometre is -0.37.
* Why because the correlation always between -1&1.
* If the correlation lies between 0 to 0.5 we can say that moderately positively correlated.
* If the correlation lies between 0 to -0.5 we can say that moderately negatively correlated.
* If the correlation lies between 0.5-1 we can say that highly positively correlated.
* If the correlation lies between -0.5 to -1 we can say that highly negatively correlated.
* Finally we can say that relation between price and kilometre due to its negative correlation.
* Finally we concluded that the price and kilometre are negatively corelated i.e, if kilometer increases price decreases both are inversely proportioned.

**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,we can say by using the count plot
* Both gives us the result of brand names on x-axis and count or frequency on y-axis. We can say each brand have how may vehicles. data[‘brand’].value\_counts()
* It will gives the count of each brand available on sale. To visualize the data I used the count plot as

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

* Will gives the count of each and every individual brand as a bar representation in the plot.

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 heat map.
* To collect the average prices of both the vehicle type and gear box,
* We can use pivort table by using it, we can take index as gear box, columns are vehicle type and values as price, pivort table default take the average.
* Code for pivort table:
* pd.pivot\_table(data=data,index='gearbox',columns='vehicleType',values='price')
* Here shows that the all average prices about the combination of vehicle type and gearbox.
* So now we can use bar plot and heatmap also.
* This is the code for bar plot ave\_price.plot(kind="bar")
* And we can aslo do by heatmap:

plt.figure(figsize=(10,5))

plt.title('Average price for vehicles based on the type of vehicle as well as on the type of gearbox')

sns.heatmap(ave\_price,annot=True)

plt.show()

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 collect the seller column data as
* data["seller"].value\_counts()
* 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
* In statstics we say that marginal probability that is normalize means marginal.
* 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

**Analysis 3:**

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 heat map.
* To collect the average prices of both the fuel type and gear box,
* We can use pivort table by using it, we can take index as gear box, columns are fuel type and values as price, pivort table default take the average of price.
  + Code for pivort table
  + pd.pivot\_table(data=data,index='gearbox',columns='fuelType',values='price')
* shows that the all average prices about the combination of fuel type and gearbox.
* Code for bar plot:
  + avg\_price1.plot(kind='bar')
  + plt.title('Average price of vehicle by fuel type and gearbox type')
  + plt.xlabel('fuelType and gearbox')
  + plt.ylabel('average power')
  + plt.show()
* By heat map also we can represent the average price of vehicles based on the fuel type and gear box.

plt.figure(figsize=(10,5))

plt.title('Average price of vehicle by fuel type and gearbox type')

sns.heatmap(avg\_price,annot=True)

plt.show()

* It will gives the combination of each individual fuel type and gearbox as a bar and mean price on another axis.

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 and heatmap.
* To collect the average power of both the vehicle type and gear box at a time
* We can use pivort table by using it, we can take index as gear box, columns are vehicle type and values as price, pivort table default take the average of power.
* Code for pivort table:

pd.pivot\_table(data=data,index='vehicleType',columns='gearbox',values='powerPS')

* shows that the all average power about the combination of fuel type and gearbox.
* Code for bar plot:

plt.figure(figsize=(15,10))

avg\_power.plot(kind='bar')

plt.title('Average power of a vehicle by vehicle type and gearbox type')

plt.xlabel('vehicleType and gearbox')

plt.ylabel('average power')

plt.show()

* By heat map also we can represent the average price of vehicles based on the fuel type and gear box.
* Code for heatmap:

plt.figure(figsize=(10,5))

plt.title('Average power of a vehicle by vehicle type and gearbox type')

sns.heatmap(avg\_power,annot=True)

plt.show()

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 can use pivort table by using it, we can take index as gear box, columns are vehicle type and values as price, pivort table default take the average of power.
* Code for pivort table:
* pd.pivot\_table(data=data,index='brand',columns='vehicleType',values='price')
* shows you the average prices of all the brand and vehicle types as well at a time.
* Code for heatmap:

plt.figure(figsize=(20,10))

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

sns.heatmap(avg\_price2,annot=True)

plt.show()

* Above code gives us a heatmap plot which gives the average prices for the vehicle type and brand at a time.