Advance Data Analysis
On
Cars Dataset

Name :- **Adnan Baig**

Batch :- **251**

- **Analysis 1:--**

1) Perform general Data analysis:-

* Importing Libraries

* Importing Dataset

* Data Cleaning and Manipulation

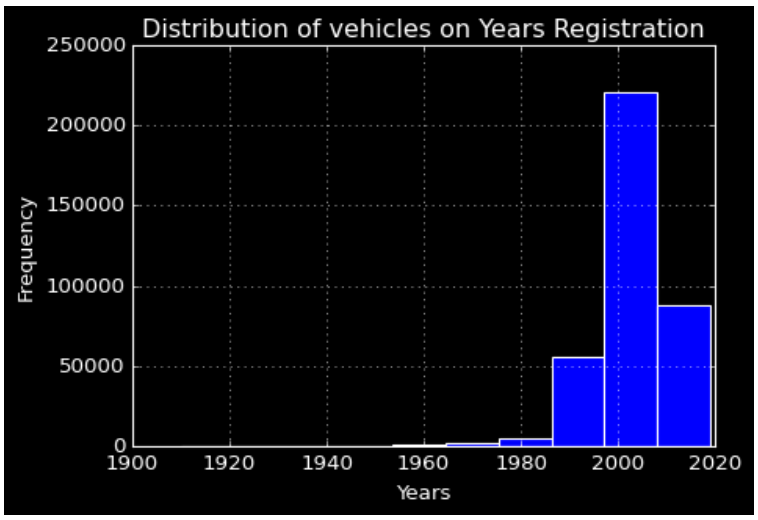
    * Filling missing values

    * Handling with structural error

    * Handling with Typecasting

    * Columns proper sequence

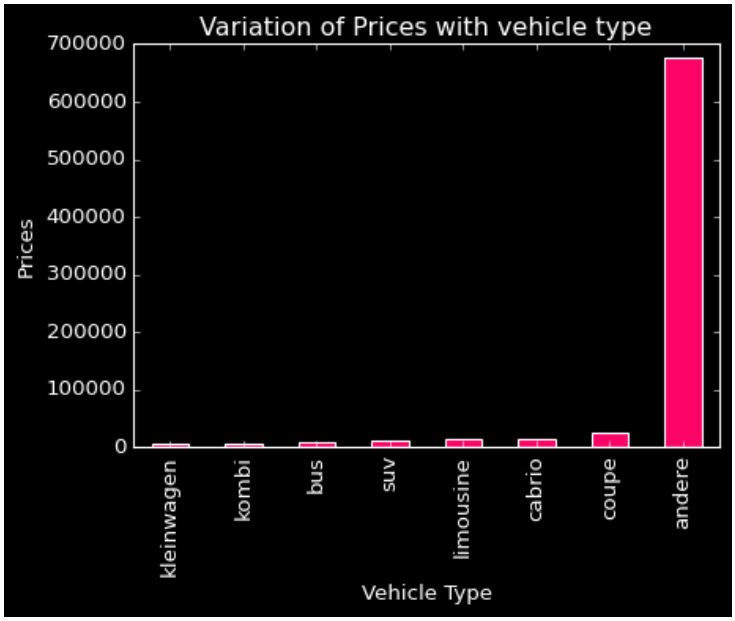
2) Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot :-



* Above bar plot is showing the distribution of Vehicle based on Registration Year , the maximum vehicles are having registration year 2000.

* There less vehicles with old registration years.

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



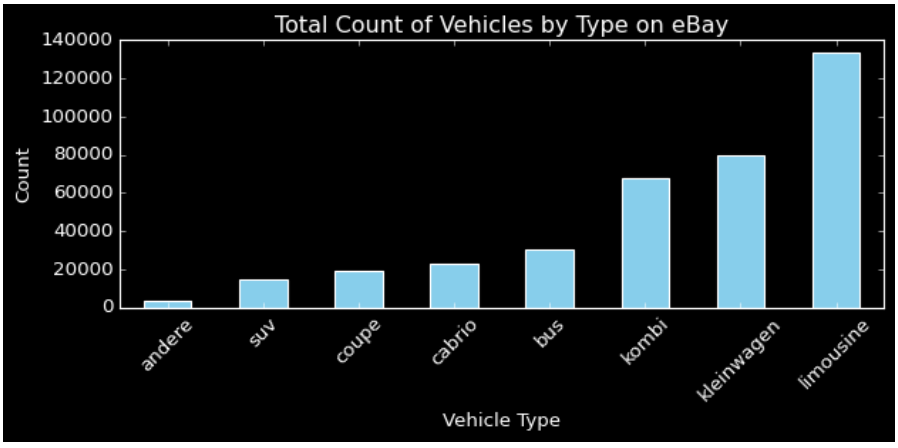
* Above plot is showing there is high average price in andere type of vehicle.

* Rest of vehicle types are less prices.

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

```
vehicle_count = df['Vehicle Type'].value_counts().sort_values()
vehicle_count
```

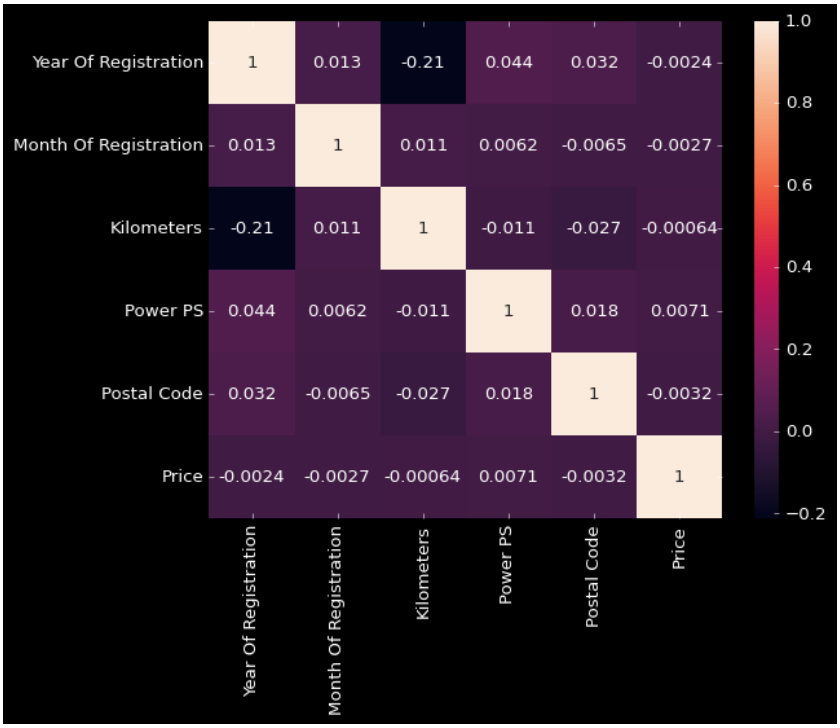
```
Vehicle Type
andere          3357
suv            14707
coupe          19015
cabrio         22898
bus            30201
kombi          67564
kleinwagen     80023
limousine     133581
Name: count, dtype: int64
```



* Above plot is showing there is maximum and near about 130000 vehicles are limousine type vehicle

* Less vehicle are from andere type of vehicle

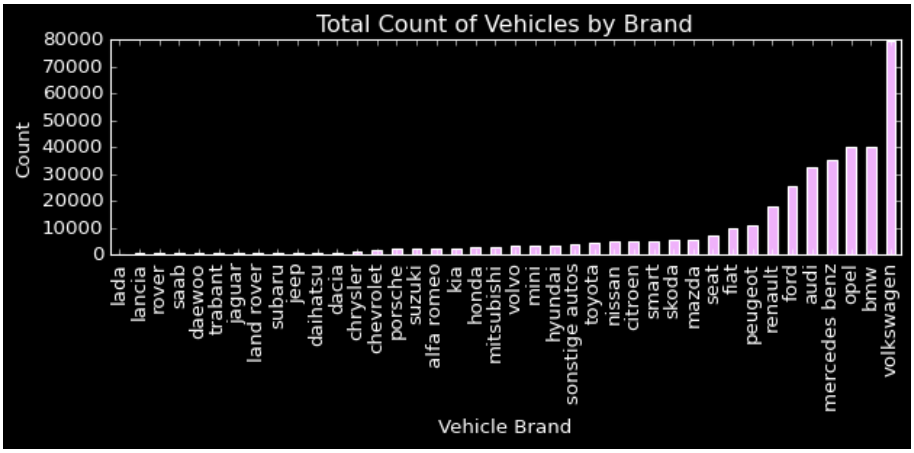
5) Is there any relationship between dollar price and kilometer? (Explain with appropriate analysis



* From above scatter plot and heatmap we can see there is -0.2 negative week relationship between registration years and prices, prices is not influencing by years.

- **Analysis 2:--**

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

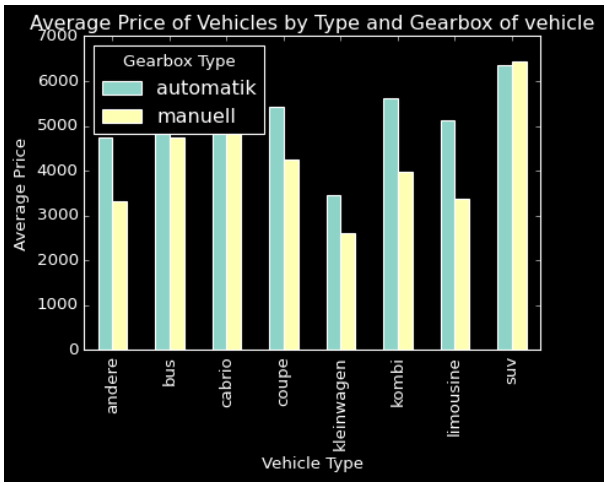


***** Above plot is showing no of vehicle based on vehicles brand , there is maximum vehicle from Volkswagen near about 80000

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

```
avg_prices = df.groupby(['Vehicle Type', 'Gearbox Type'])['Price'].mean().unstack()
avg_prices
```

Gearbox Type	automatik	manuell
Vehicle Type		
andere	4748.554017	3314.511682
bus	5632.606970	4743.225859
cabrio	5981.279909	5297.472116
coupe	5434.073321	4249.401552
kleinwagen	3445.203710	2613.711917
kombi	5623.585119	3987.081556
limousine	5121.658741	3379.571481
suv	6346.135982	6430.416722



***** Above plot is showing manual gearbox from andere vehicle type having maximum average price

3) What is the marginal probability of private seller

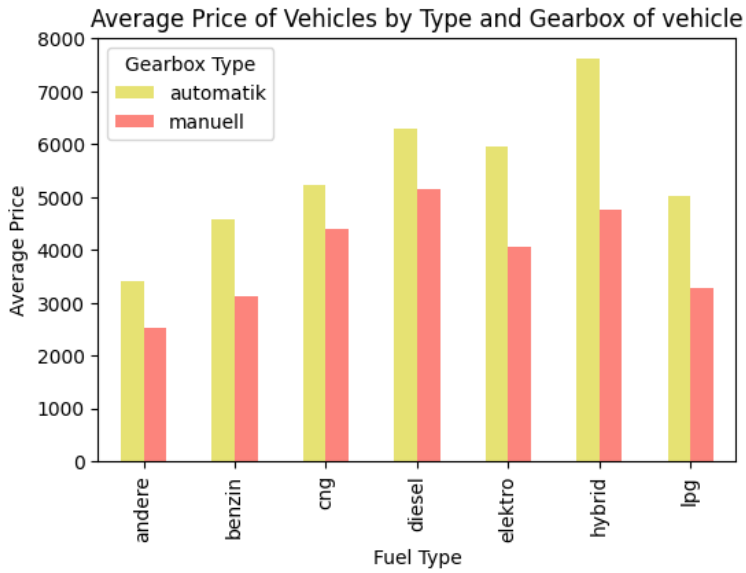
* Marginal probability of 'private' seller in the 'seller' column is ---> 0.9999919212809617

- **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?

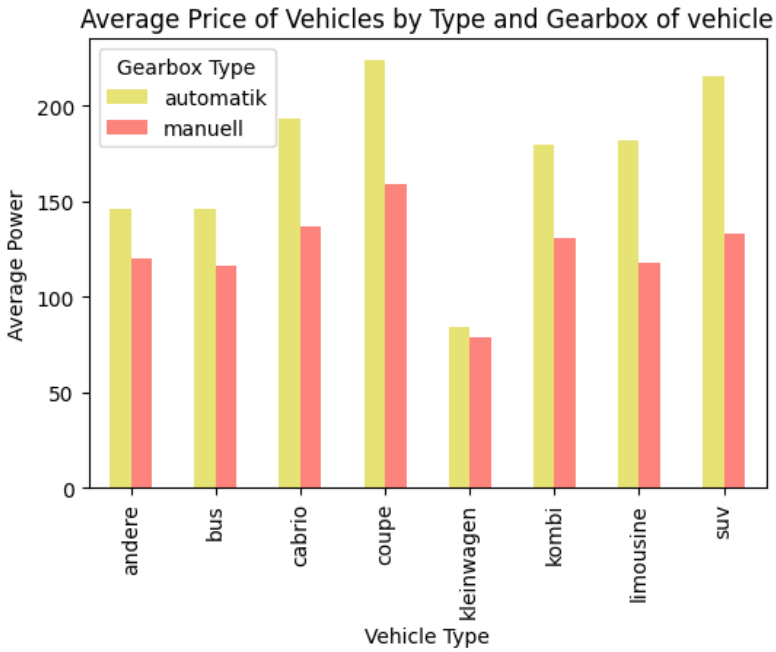
* We can reduce the usage of the data by using appropriate and proper datatypes because its contains different bytes for different datatypes.

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



* Above plot is showing manual gearbox and andere fuel type having maximum avg price

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



* Above plot is showing automatic gearbox from coupe vehicle type having maximum avg power more than 200 power ps

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