

# Report

## 1.Perform General Data Analysis.

Performing general data analysis involves several steps. Those are,

### 1.Data Collection:

- Obtain the dataset from a reliable source. This could be in the form of a CSV file.

### 2.Data Cleaning:

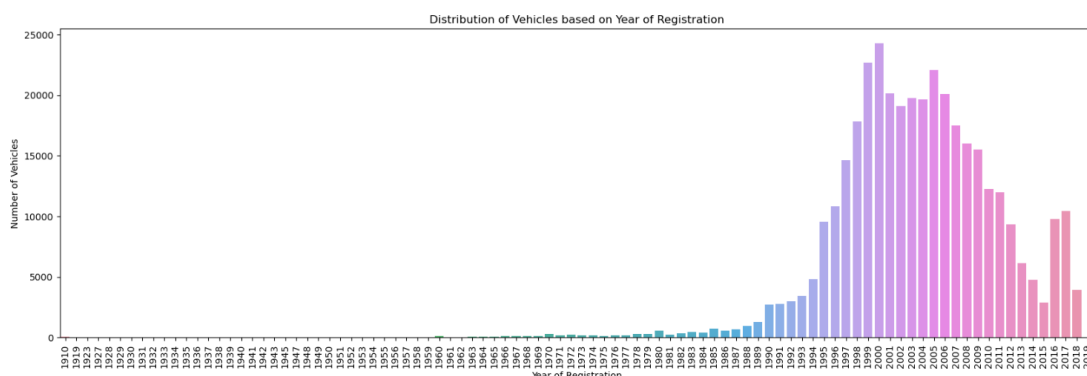
- Check for missing values, inconsistencies in the data.
- Handle missing data through imputation or removal.
- Standardize data formats.

### 3.Exploratory Data Analysis (EDA):

- Visualize the data using graphs, histograms, box plots, scatter plots etc.,
- to understand the distribution, relationships and patterns.
- Calculate summary statistics (mean, median, standard deviation, etc.) to describe the data.

## 2.Can you tell me the distribution of vehicles based on Year of Registration with the help of a plot.

```
In [6]: plt.figure(figsize=(20, 6))
sns.countplot(data=data, x='yearOfRegistration')
plt.title('Distribution of Vehicles based on Year of Registration')
plt.xlabel('Year of Registration')
plt.ylabel('Number of Vehicles')
plt.xticks(rotation=90) # Rotate x-labels for better readability
plt.show()
```

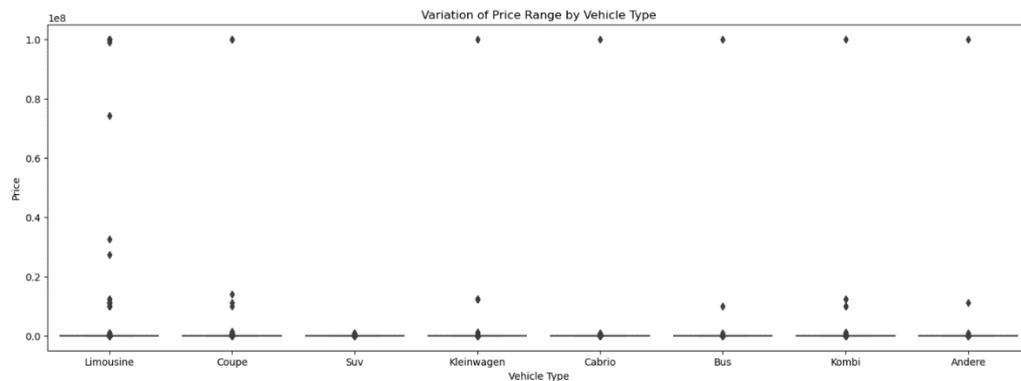


Here is the details of distribution of vehicles based on year of registration by using "Count Plot"

### 3.Create a plot based on the variation of the price range by the vehicle type

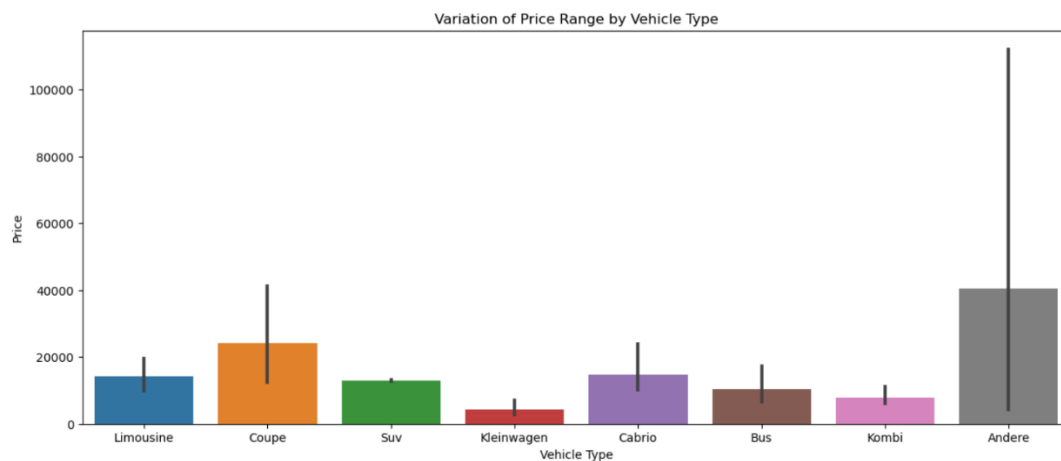
Here is the Box-plot based on the variation of the price range by vehicle type

```
plt.figure(figsize=(18, 6))
sns.boxplot(data=data, x='vehicleType', y='price')
plt.title('Variation of Price Range by Vehicle Type')
plt.xlabel('Vehicle Type')
plt.ylabel('Price')
plt.show()
```



Here is the bar-plot based on the variation of the price range by vehicle type

```
In [8]: plt.figure(figsize=(15, 6))
sns.barplot(data=data, x='vehicleType', y='price')
plt.title('Variation of Price Range by Vehicle Type')
plt.xlabel('Vehicle Type')
plt.ylabel('Price')
plt.show()
```



**4.Find out Total count of vehicles by type available on e-bay for sale. As well as create a visualization for the client.**

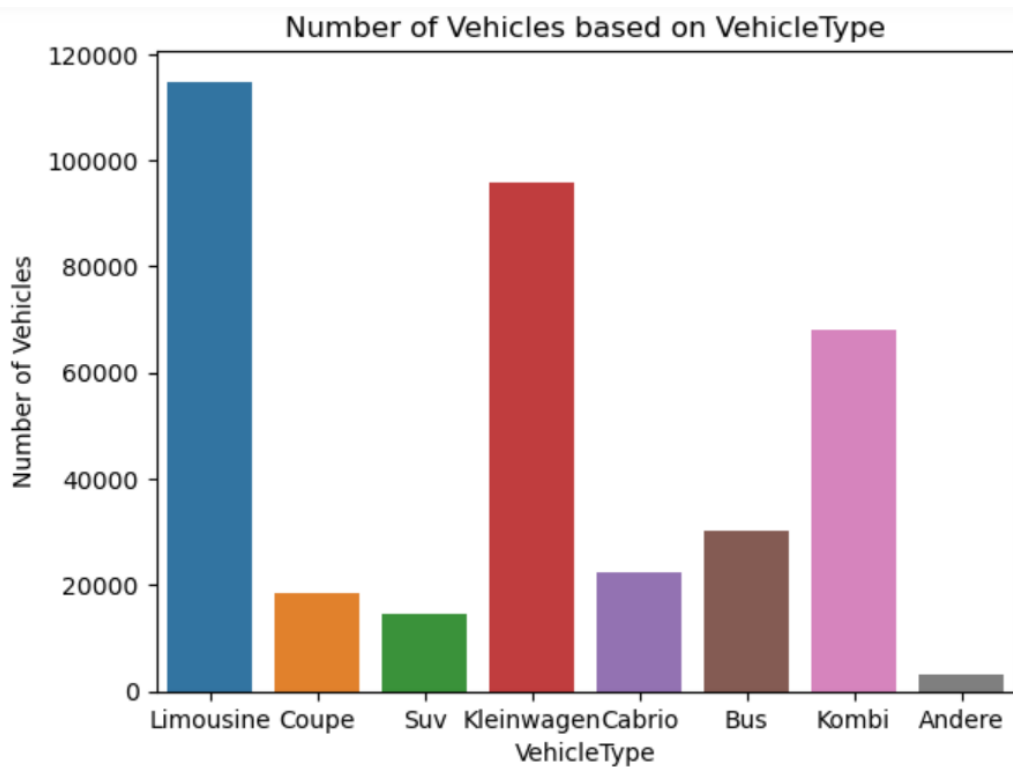
*Here is the visualization for the total count of vehicles by type available on e-bay for sale*

*By using count-plot.*

```
: data["vehicleType"].value_counts()
```

```
: Limousine      114818
   Kleinwagen    95859
   Kombi         68065
   Bus           30129
   Cabrio        22504
   Coupe         18567
   Suv           14510
   Andere        3090
   Name: vehicleType, dtype: int64
```

```
: sns.countplot(data=data,x="vehicleType")
   plt.title('Number of Vehicles based on VehicleType')
   plt.xlabel('VehicleType')
   plt.ylabel('Number of Vehicles')
   plt.show()
```

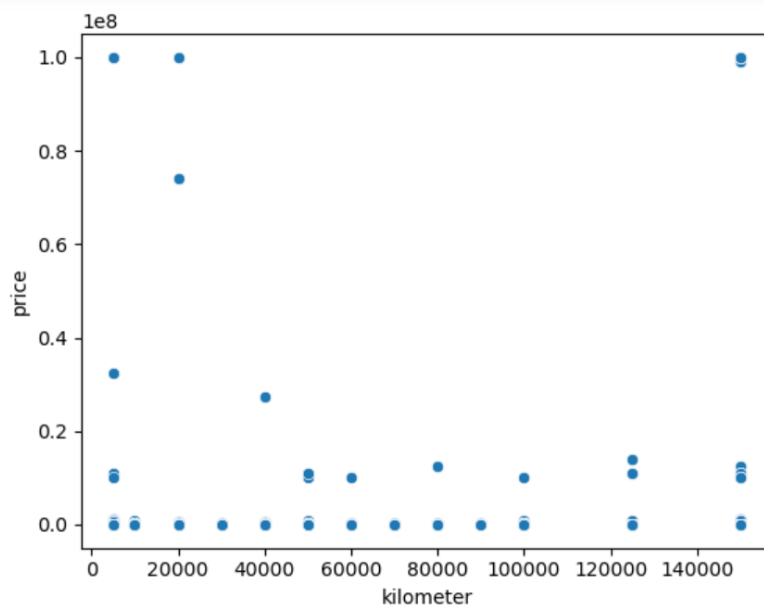


5. Is there any relationship between dollar price and kilometre? (Explain with appropriate analysis).

```
correlation=data["price"].corr(data["kilometer"])  
correlation
```

```
-0.007683223435760035
```

```
sns.scatterplot(x='kilometer', y='price', data=data)  
plt.show()
```



The correlation value ranges from -1 to 1, where:

- If the value is close to 1, it indicates a strong positive correlation.
- If the value is close to -1, it indicates a strong negative correlation.
- If the value is close to 0, it indicates no correlation.

The scatter plot visualizes the relationship between the two variables. If the points on the plot show a clear pattern, it suggests a relationship between the 'price' and 'kilometre'.

I have calculated correlation coefficient between price and kilometre is -0.0076 (approximately), then we can clarify that it indicates no correlation between "price" and "kilometre", because correlation coefficient is close to the 0.