# DATA ANALYSIS PROJECT-1

ANALYSIS-1

SUMMARY

In Analysis-1

- Firstly, I have performed general data analysis like data collection, data cleaning, Exploratory data analysis (EDA) and Reporting.
- Filling Null values.
- Secondly, I have observed that the distribution of vehicles based on the year of registration provides valuable insights into the age composition of the autos.
- For instance, a higher concentration of vehicles might be observed in certain years, indicating popular models.
- The variation of the price range by vehicle type provides valuable insights into the pricing dynamics across different auto categories.
- For instance, luxurious types like "Coupe" and "SUV" tend to have higher average prices, reflecting their premium status.
- In contrast, more practical types like "Kleinwagen" (Compact Cars) have considerably lower average prices, suggesting is a key factor for buyers in these categories.
- The total count of vehicles by type available on e-bay for sale provides a comprehensive overview of the platform's inventory.
- "Kleinwagen" and sedans might dominate indicating their high market demand and availability. "Carbio" and "SUV" might have lesser listings reflecting their limited market presence.
- Understanding these are crucial for market analysis.
- The correlation coefficient ranges from -1 to 1.
    1. If the correlation coefficient is close to 1, wit indicates a strong positive correlation.
    2. If the correlation coefficient is close to -1, it indicates a strong negative correlation.
    3. If the correlation coefficient is close to 0, it indicates a weak or no correlation.
- Here I have observed correlation is close to 0, between price and kilometer.

1.

```
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()
```


Distribution of Vehicles based on Year of Registration
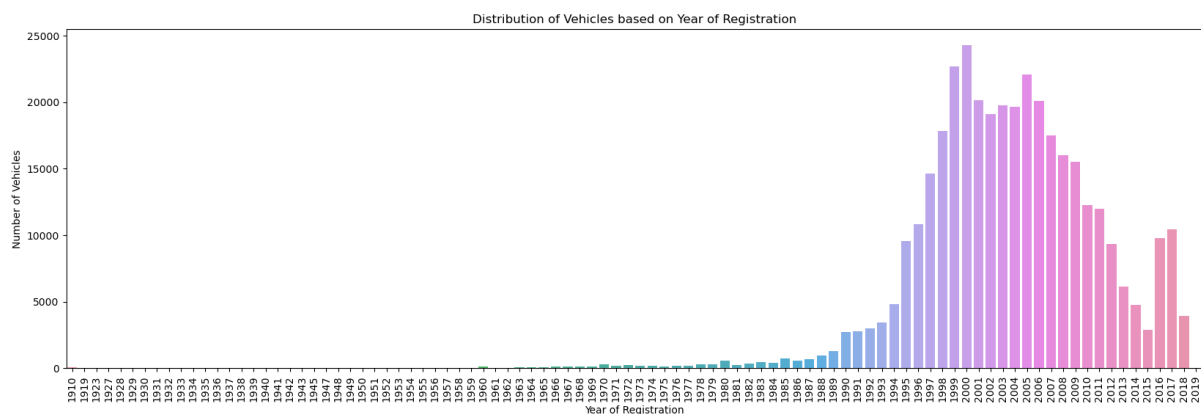
Here, count plot talks

- I have observed that the distribution of vehicles based on the year of registration provides valuable insights into the age composition of the autos. Additionally, the distribution can reveal trends in the market, such as the prevalence of vintage cars or influx of newer models. For instance, a higher concentration of vehicles might be observed in certain years, indicating popular models. Overall, understanding this distribution is crucial for both buyers and sellers, helping them make informed decisions about pricing demand, and market trends.

2.

```python
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()
```
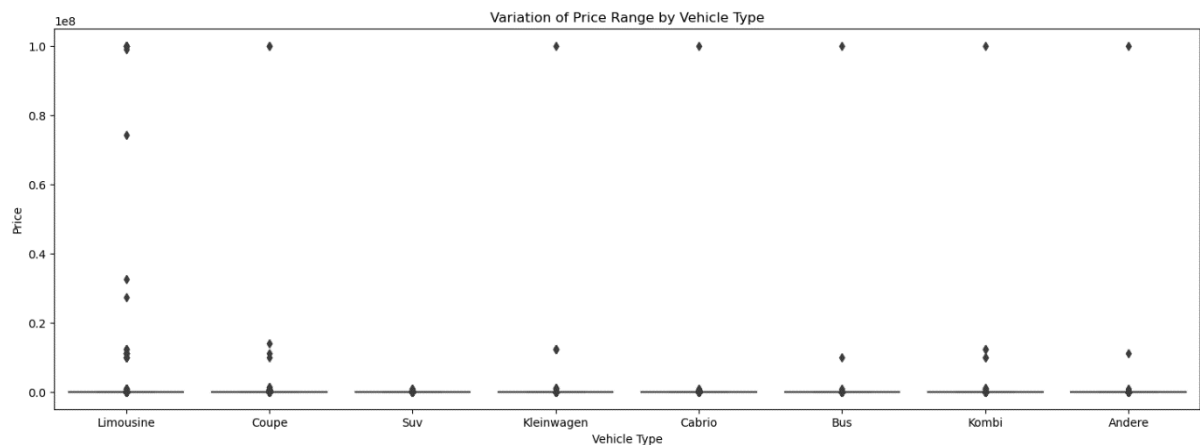


```python
plt.figure(figsize=(15, 6))
sns.barplot(data=data, x='vehicleType', y='price')
plt.title('Variation of Price Range by Vehicle Type')
plt.xlabel('Vehicle Type')
plt.ylabel('Price')
plt.show()
```

Variation of Price Range by Vehicle Type
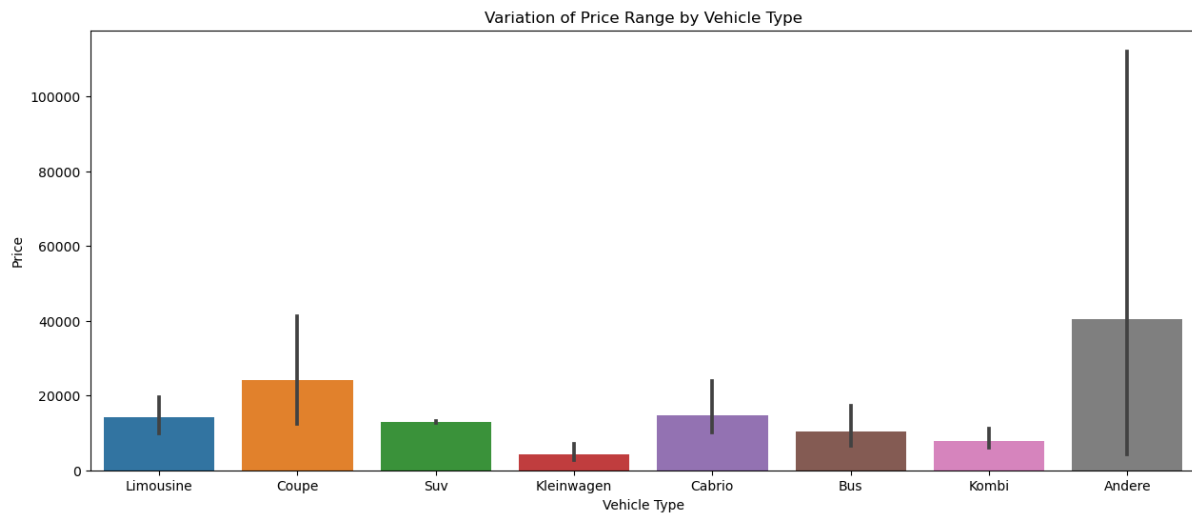
Here bar plot and boxplot tell

- The variation of the price range by vehicle type provides valuable insights into the pricing dynamics across different auto categories. For instance, luxurious types like "Coupe" and "SUV" tend to have higher average prices, reflecting their premium status. In contrast, more practical types like "Kleinwagen" (Compact Cars) have considerably lower average prices, suggesting is a key factor for buyers in these categories. This analysis aids both sellers and buyers in understanding the market trends. and making well-informed purchasing decisions.

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


Number of Vehicles based on VehicleType
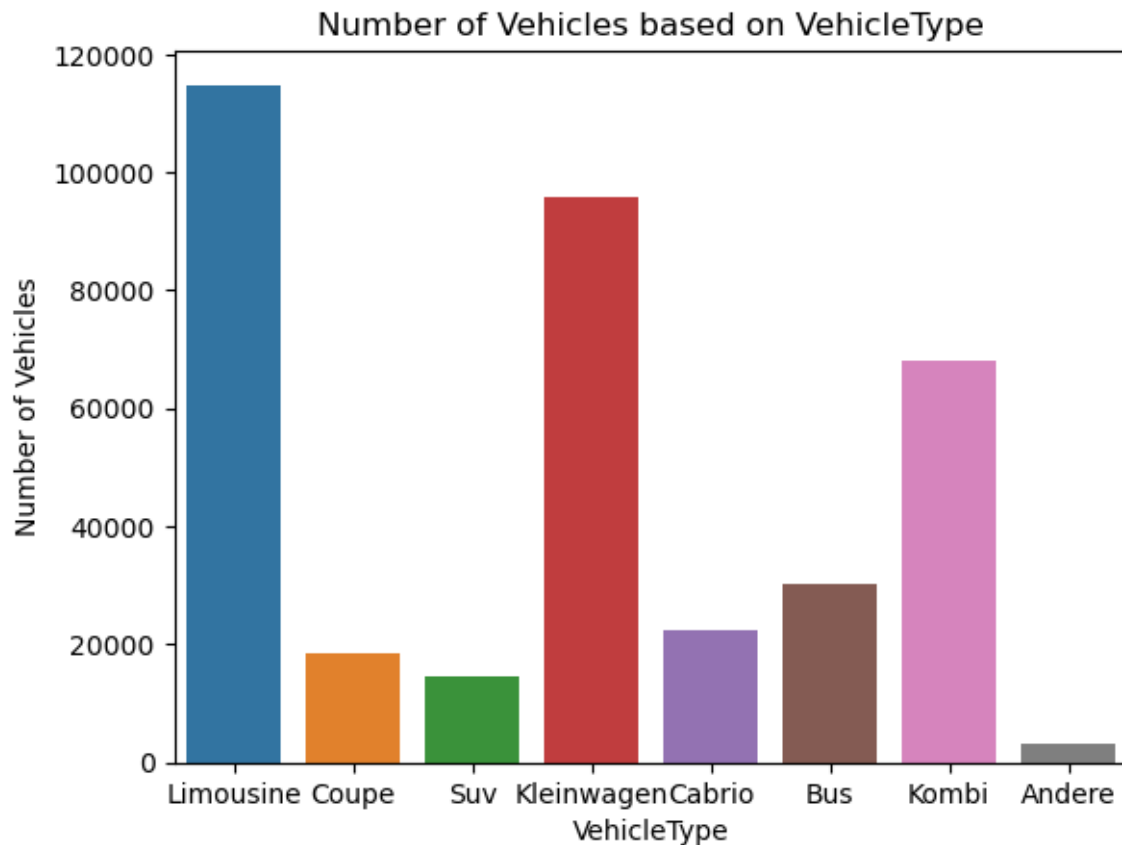
Here, count plot tells about

- The total count of vehicles by type available on e-bay for sale provides a comprehensive overview of the platform's inventory. "Kleinwagen" and sedans might dominate indicating their high market demand and availability. "Carbio" and "SUV" might have lesser listings reflecting their limited market presence. Understanding these are crucial for market analysis.

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

```
-0.007683223435760035
```

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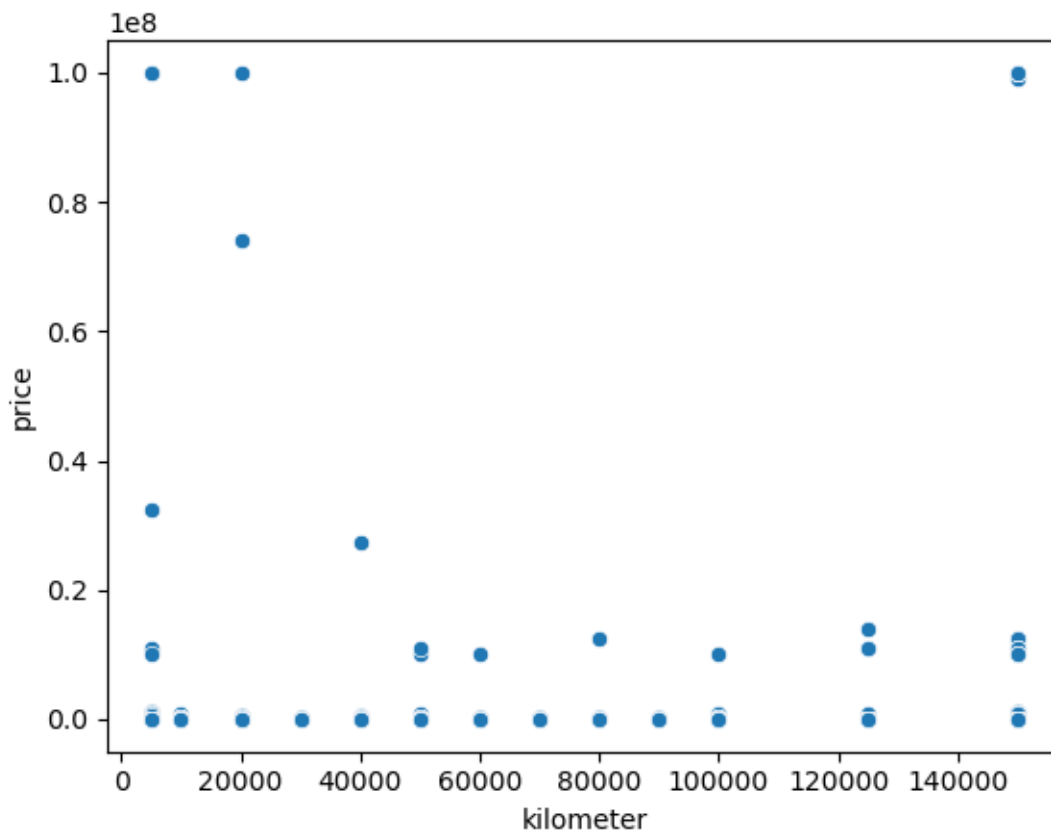

- The correlation coefficient ranges from -1 to 1.
    4. If the correlation coefficient is close to 1, with
    5. indicates a strong positive correlation.
    6. If the correlation coefficient is close to -1, it indicates a strong negative correlation.
    7. If the correlation coefficient is close to 0, it indicates a weak or no correlation.
- Here I have observed correlation   is close to 0, between price and kilometer. If the points on the scatter plot are widely dispersed and do not follow a clear pattern, it suggests a weak or no correlation between the two variables.