Project Development Phase

SPRINT 1

Date	11 November 2022
Team ID	PNT2022TMID34636
Project Name	Car Resale Value Prediction
Maximum Marks	4 Marks

Car Resale Value Prediction

Import Libraries

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib as plt
from sklearn.preprocessing import LabelEncoder
import pickle
```

Read the Dataset

In [3]:
```
df = pd.read_csv("autos.csv", header=0, sep=',', encoding='Latin1',)
df
```

Out[3]:

	dateCrawled	name	seller	offerType	price	abtest	vehicleType	yearOfRegistration	gearbox	powerPS	n
0	2016-03-24 11:52:17	Golf_3_1.6	privat	Angebot	480	test	NaN	1993	manuell	0	
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	privat	Angebot	18300	test	coupe	2011	manuell	190	

pre-process the Data

In [4]:
```
df=df.drop('offerType',axis=1)
```

In [6]:
```
df=df[(df.powerPS > 50) & (df.powerPS < 900)]
```

In [7]:
```
df = df[(df.yearOfRegistration >= 1950) & (df.yearOfRegistration < 2017)]
```

In [11]:
```
df.drop(['name', 'abtest', 'dateCrawled', 'nrOfPictures', 'lastSeen', 'postalCode','dateCreated'],
axis='columns',inplace=True)
```

In [12]:
```
new_df = df.copy()
```

In [13]:
```
new_df = new_df.drop_duplicates ([ 'price', 'vehicleType', 'yearOfRegistration' ,'gearbox',
'powerPS', 'model', 'kilometer', 'monthOfRegistration', 'fuelType' ,'notRepairedDamage'])
```

In [14]:
```
new_df.gearbox.replace(('manuell', 'automatik'), ('manual', 'automatic'), inplace=True)
```

In [15]:
```
new_df.gearbox.replace(('manuell', 'automatik'), ('manual', 'automatic'), inplace=True)
```

In [20]:
```
new_df.vehicleType.replace(('kleinwagen', 'cabrio', 'komb', 'andere'), ('small car', 'convertible',
'combination', 'others'), inplace=True)
```

In [22]:
```
 new_df.notRepairedDamage.replace(('ja', 'nein'), ('Yes',
'No'),inplace=True)
```

In [22]:
```
 new_df.notRepairedDamage.replace(('ja', 'nein'), ('Yes',
'No'),inplace=True)
```

In [23]:
```
new_df = new_df[(new_df.price >= 100) & (new_df.price <= 150000)]
```

In [24]:
```
new_df['notRepairedDamage'].fillna(value='not-declared', inplace=True)
```

In [28]:
```
new_df[ 'fuelType'].fillna(value='not-declared', inplace=True)
```

In [29]:
```
new_df[ 'gearbox'].fillna(value='not-declared', inplace=True)
```

In [30]:
```
new_df[ 'vehicleType'].fillna (value='not-declared', inplace=True)
```

In [32]:
```
new_df['model'].fillna(value='not-declared',inplace=True)
```

In [33]:
```
new_df['model'].fillna(value='not-declared',inplace=True)
```