

```
In [81]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [241]: df=pd.read_csv(r"C:\USERS\user\Downloads\c7_used_cars - c7_used_cars.csv")
```

Out[241]:

	Unnamed: 0	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
0	0	T-Roc	2019	25000	Automatic	13904	Diesel	145	49.6	2.0
1	1	T-Roc	2019	26883	Automatic	4562	Diesel	145	49.6	2.0
2	2	T-Roc	2019	20000	Manual	7414	Diesel	145	50.4	2.0
3	3	T-Roc	2019	33492	Automatic	4825	Petrol	145	32.5	2.0
4	4	T-Roc	2019	22900	Semi-Auto	6500	Petrol	150	39.8	1.5
...
99182	10663	A3	2020	16999	Manual	4018	Petrol	145	49.6	1.0
99183	10664	A3	2020	16999	Manual	1978	Petrol	150	49.6	1.0
99184	10665	A3	2020	17199	Manual	609	Petrol	150	49.6	1.0
99185	10666	Q3	2017	19499	Automatic	8646	Petrol	150	47.9	1.4
99186	10667	Q3	2016	15999	Manual	11855	Petrol	150	47.9	1.4

99187 rows × 11 columns

```
In [242]:
```

```
Out[242]: Index(['Unnamed: 0', 'model', 'year', 'price', 'transmission', 'mileage',
                'fuelType', 'tax', 'mpg', 'engineSize', 'Make'],
                dtype='object')
```

```
In [243]: df=df.head(15)
```

Out[243]:

	Unnamed: 0	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize	Mal
0	0	T-Roc	2019	25000	Automatic	13904	Diesel	145	49.6	2.0	V
1	1	T-Roc	2019	26883	Automatic	4562	Diesel	145	49.6	2.0	V
2	2	T-Roc	2019	20000	Manual	7414	Diesel	145	50.4	2.0	V
3	3	T-Roc	2019	33492	Automatic	4825	Petrol	145	32.5	2.0	V
4	4	T-Roc	2019	22900	Semi-Auto	6500	Petrol	150	39.8	1.5	V
5	5	T-Roc	2020	31895	Manual	10	Petrol	145	42.2	1.5	V
6	6	T-Roc	2020	27895	Manual	10	Petrol	145	42.2	1.5	V
7	7	T-Roc	2020	39495	Semi-Auto	10	Petrol	145	32.5	2.0	V
8	8	T-Roc	2019	21995	Manual	10	Petrol	145	44.1	1.0	V
9	9	T-Roc	2019	23285	Manual	10	Petrol	145	42.2	1.5	V
10	10	T-Roc	2019	23985	Semi-Auto	10	Petrol	145	39.8	1.5	V
11	11	T-Roc	2019	23585	Manual	10	Petrol	145	42.2	1.5	V
12	12	T-Roc	2020	25785	Semi-Auto	10	Petrol	145	39.8	1.5	V
13	13	T-Roc	2019	23995	Semi-Auto	1069	Petrol	145	39.8	1.5	V
14	14	T-Roc	2018	17495	Manual	21645	Petrol	145	53.3	1.5	V

In [244]: `a=df[['Unnamed: 0', 'year', 'price', 'mileage', 'tax', 'mpg', 'engineSize', 'Make']]`

Out[244]:

	Unnamed: 0	year	price	mileage	tax	mpg	engineSize	Make
0	0	2019	25000	13904	145	49.6	2.0	VW
1	1	2019	26883	4562	145	49.6	2.0	VW
2	2	2019	20000	7414	145	50.4	2.0	VW
3	3	2019	33492	4825	145	32.5	2.0	VW
4	4	2019	22900	6500	150	39.8	1.5	VW
5	5	2020	31895	10	145	42.2	1.5	VW
6	6	2020	27895	10	145	42.2	1.5	VW
7	7	2020	39495	10	145	32.5	2.0	VW
8	8	2019	21995	10	145	44.1	1.0	VW
9	9	2019	23285	10	145	42.2	1.5	VW
10	10	2019	23985	10	145	39.8	1.5	VW
11	11	2019	23585	10	145	42.2	1.5	VW
12	12	2020	25785	10	145	39.8	1.5	VW
13	13	2019	23995	1069	145	39.8	1.5	VW
14	14	2018	17495	21645	145	53.3	1.5	VW

In [245]:

Out[245]: VW 15
Name: Make, dtype: int64

In [246]: `x=a.drop('Make',axis=1)`

In [247]: `g1={"Make":{"VW":1}}
a=a.replace(g1)`

	Unnamed: 0	year	price	mileage	tax	mpg	engineSize	Make
0	0	2019	25000	13904	145	49.6	2.0	1
1	1	2019	26883	4562	145	49.6	2.0	1
2	2	2019	20000	7414	145	50.4	2.0	1
3	3	2019	33492	4825	145	32.5	2.0	1
4	4	2019	22900	6500	150	39.8	1.5	1
5	5	2020	31895	10	145	42.2	1.5	1
6	6	2020	27895	10	145	42.2	1.5	1
7	7	2020	39495	10	145	32.5	2.0	1
8	8	2019	21995	10	145	44.1	1.0	1
9	9	2019	23285	10	145	42.2	1.5	1
10	10	2019	23985	10	145	39.8	1.5	1
11	11	2019	23585	10	145	42.2	1.5	1
12	12	2020	25785	10	145	39.8	1.5	1
13	13	2019	23995	1069	145	39.8	1.5	1
14	14	2018	17495	21645	145	53.3	1.5	1

```
In [248]: from sklearn.model_selection import train_test_split
```

```
In [249]: from sklearn.ensemble import RandomForestClassifier  
  
rfc=RandomForestClassifier()
```

```
Out[249]: RandomForestClassifier()
```

```
In [250]: parameters={'max_depth':[1,2,3,4,5],  
                    'min_samples_leaf':[5,10,15,20,25],  
                    'n_estimators':[10,20,30,40,50]}
```

```
In [251]: from sklearn.model_selection import GridSearchCV  
  
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
```

```
Out[251]: GridSearchCV(cv=2, estimator=RandomForestClassifier(),  
                    param_grid={'max_depth': [1, 2, 3, 4, 5],  
                                'min_samples_leaf': [5, 10, 15, 20, 25],  
                                'n_estimators': [10, 20, 30, 40, 50]},  
                    scoring='accuracy')
```

```
In [252]:
```

```
Out[252]: 1.0
```

```
In [253]:
```

```
In [254]: from sklearn.tree import plot_tree  
  
plt.figure(figsize=(80,40))
```

```
Out[254]: [Text(2232.0, 1087.2, 'gini = 0.0\nsamples = 6\nvalue = 10.0')]
```

gini = 0.0
samples = 6
value = 10.0

```
In [ ]:
```

