In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
from sklearn.linear_model import LogisticRegression
```

In [3]:
```
df=pd.read_csv("c7 used csv").dropna()

df
```

Out[3]:

	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 [4]:
```
df.dropna(inplace=True)
```

In [32]:
```
df['Make'].value_counts()
```

Out[32]:
```
ford        17965
VW          15157
vauxhall    13632
merc        13119
BMW         10781
Audi        10668
toyota       6738
skoda        6267
hyundi       4860
Name: Make, dtype: int64
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 99187 entries, 0 to 99186
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Unnamed: 0    99187 non-null  int64  
 1   model         99187 non-null  object 
 2   year          99187 non-null  int64  
 3   price         99187 non-null  int64  
 4   transmission  99187 non-null  object 
 5   mileage       99187 non-null  int64  
 6   fuelType      99187 non-null  object 
 7   tax           99187 non-null  int64  
 8   mpg           99187 non-null  float64
 9   engineSize    99187 non-null  float64
 10  Make          99187 non-null  object 
dtypes: float64(2), int64(5), object(4)
memory usage: 9.1+ MB
```

In [7]:
```
feature_matrix = df[['Unnamed: 0','year','price','mileage','tax','mpg','engine
target_vector = df['transmission']
```

In [8]:
```
feature_matrix.shape
```

Out[8]:
```
(99187, 7)
```

In [9]:
```
target_vector.shape
```

Out[9]:
```
(99187,)
```

In [10]:
```
from sklearn.preprocessing import StandardScaler
```

In [11]:
```
fs = StandardScaler().fit_transform(feature_matrix)
```

In [12]:
```
logr = LogisticRegression()
logr.fit(fs,target_vector)
```

Out[12]:
```
LogisticRegression()
```

In [13]:
```
feature_matrix.shape
```

Out[13]:
```
(99187, 7)
```

In [14]:
```
target_vector.shape
```

Out[14]:
```
(99187,)
```

In [15]:
```
from sklearn.preprocessing import StandardScaler
```

In [16]:
```
fs = StandardScaler().fit_transform(feature_matrix)
```

In [17]:
```
logr = LogisticRegression()
logr.fit(fs,target_vector)
```

Out[17]:
```
LogisticRegression()
```

In [18]:
```
observation=df[['Unnamed: 0','year','price','mileage','tax','mpg','engineSize'
```

In [19]:
```
prediction = logr.predict(observation)
prediction
```

Out[19]:
```
array(['Semi-Auto', 'Semi-Auto', 'Semi-Auto', ..., 'Automatic',
       'Automatic', 'Automatic'], dtype=object)
```

In [20]:
```
logr.classes_
```

Out[20]:
```
array(['Automatic', 'Manual', 'Other', 'Semi-Auto'], dtype=object)
```

In [21]:
```
logr.predict_proba(observation)[0][1]
```

Out[21]:
```
0.0
```

Random Forest

In [22]:
```
df['Make'].value_counts()
```

Out[22]:
```
Manual       56445
Semi-Auto    22677
Automatic    20056
Other            9
Name: transmission, dtype: int64
```

In [23]:
```
x=df[['Unnamed: 0','year','price','mileage','tax','mpg','engineSize']]
y=df['Make']
```

In [33]:
```
g1={'Make':{'ford':1,'VW':2, 'vauxhall':3, 'merc':4, 'BMW':5, 'Audi':6, 'toyot
df=df.replace(g1)
df
```

Out[33]:

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

99187 rows × 11 columns

In [34]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

In [35]:
```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[35]:
```
RandomForestClassifier()
```

In [36]:
```
parameters = {'max_depth':[1,2,3,4,5],'min_samples_leaf':[5,10,15,20,25],
              'n_estimators': [10,20,30,40,50]
              }
```

In [37]:
```
from sklearn.model_selection import GridSearchCV
grid_search = GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="a
grid_search.fit(x_train,y_train)
```

Out[37]:
```
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')
```

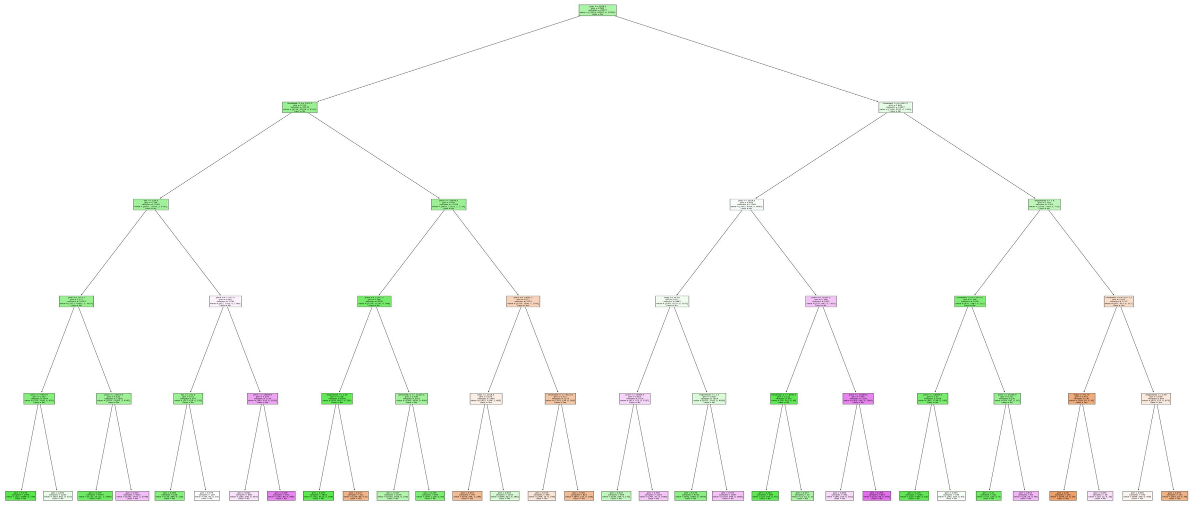
In [38]:
```
grid_search.best_score_
```

Out[38]:
```
0.731715396802535
```

```
In [39]: rfc_best = grid_search.best_estimator_
```

In [41]:

```
from sklearn.tree import plot_tree
plt.figure(figsize = (80,40,))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','
Out[41]: [Text(2232.0, 1993.2, 'year <= 2018.5\ngini = 0.582\nsamples = 43977\nvalue =
         [13929, 39563, 4, 15934]\nclass = No'),
          Text(1116.0, 1630.8000000000002, 'Unnamed: 0 <= 7307.5\ngini = 0.529\nsample
         s = 30150\nvalue = [8739, 30208, 4, 8520]\nclass = No'),
          Text(558.0, 1268.4, 'tax <= 155.0\ngini = 0.536\nsamples = 17961\nvalue = [3
         887, 17655, 3, 6791]\nclass = No'),
          Text(279.0, 906.0, 'year <= 2015.5\ngini = 0.511\nsamples = 16165\nvalue =
         [3275, 16602, 3, 5623]\nclass = No'),
          Text(139.5, 543.5999999999999, 'price <= 9745.5\ngini = 0.442\nsamples = 324
         9\nvalue = [570, 3700, 0, 876]\nclass = No'),
          Text(69.75, 181.19999999999982, 'gini = 0.179\nsamples = 1996\nvalue = [151,
         2848, 0, 154]\nclass = No'),
          Text(209.25, 181.19999999999982, 'gini = 0.642\nsamples = 1253\nvalue = [41
         9, 852, 0, 722]\nclass = No'),
          Text(418.5, 543.5999999999999, 'price <= 15937.5\ngini = 0.526\nsamples = 12
         916\nvalue = [2705, 12902, 3, 4747]\nclass = No'),
          Text(348.75, 181.19999999999982, 'gini = 0.315\nsamples = 8579\nvalue = [111
         2, 11111, 1, 1369]\nclass = No'),
          Text(488.25, 181.19999999999982, 'gini = 0.625\nsamples = 4337\nvalue = [159
         3, 1791, 2, 3378]\nclass = No'),
          Text(837.0, 906.0, 'price <= 12701.0\ngini = 0.645\nsamples = 1796\nvalue =
         [612, 1053, 0, 1168]\nclass = No'),
          Text(697.5, 543.5999999999999, 'tax <= 232.5\ngini = 0.51\nsamples = 657\nva
         lue = [227, 679, 0, 135]\nclass = No'),
          Text(627.75, 181.19999999999982, 'gini = 0.464\nsamples = 528\nvalue = [133,
         584, 0, 116]\nclass = No'),
          Text(767.25, 181.19999999999982, 'gini = 0.579\nsamples = 129\nvalue = [94,
         95, 0, 19]\nclass = No'),
          Text(976.5, 543.5999999999999, 'price <= 19699.0\ngini = 0.578\nsamples = 11
         39\nvalue = [385, 374, 0, 1033]\nclass = No'),
          Text(906.75, 181.19999999999982, 'gini = 0.642\nsamples = 548\nvalue = [182,
         290, 0, 367]\nclass = No'),
          Text(1046.25, 181.19999999999982, 'gini = 0.458\nsamples = 591\nvalue = [20
         3, 84, 0, 666]\nclass = No'),
          Text(1674.0, 1268.4, 'price <= 15979.5\ngini = 0.497\nsamples = 12189\nvalue
         = [4852, 12553, 1, 1729]\nclass = No'),
          Text(1395.0, 906.0, 'price <= 11922.5\ngini = 0.346\nsamples = 9056\nvalue =
         [2276, 11247, 0, 698]\nclass = No'),
          Text(1255.5, 543.5999999999999, 'engineSize <= 2.05\ngini = 0.238\nsamples =
         6034\nvalue = [976, 8211, 0, 290]\nclass = No'),
          Text(1185.75, 181.19999999999982, 'gini = 0.206\nsamples = 5833\nvalue = [75
         2, 8123, 0, 284]\nclass = No'),
          Text(1325.25, 181.19999999999982, 'gini = 0.427\nsamples = 201\nvalue = [22
         4, 88, 0, 6]\nclass = Yes'),
          Text(1534.5, 543.5999999999999, 'Unnamed: 0 <= 11824.5\ngini = 0.508\nsample
         s = 3022\nvalue = [1300, 3036, 0, 408]\nclass = No'),
          Text(1464.75, 181.19999999999982, 'gini = 0.547\nsamples = 2076\nvalue = [10
         10, 1910, 0, 325]\nclass = No'),
          Text(1604.25, 181.19999999999982, 'gini = 0.395\nsamples = 946\nvalue = [29
         0, 1126, 0, 83]\nclass = No'),
          Text(1953.0, 906.0, 'price <= 20085.5\ngini = 0.611\nsamples = 3133\nvalue =
         [2576, 1306, 1, 1031]\nclass = Yes'),
          Text(1813.5, 543.5999999999999, 'year <= 2016.5\ngini = 0.614\nsamples = 196
         0\nvalue = [1423, 1168, 1, 462]\nclass = Yes'),
          Text(1743.75, 181.19999999999982, 'gini = 0.533\nsamples = 767\nvalue = [75
         9, 246, 1, 196]\nclass = Yes'),
          Text(1883.25, 181.19999999999982, 'gini = 0.603\nsamples = 1193\nvalue = [66
4, 922, 0, 266]\nclass = No'),
 Text(2092.5, 543.5999999999999, 'Unnamed: 0 <= 9121.5\ngini = 0.517\nsamples
= 1173\nvalue = [1153, 138, 0, 569]\nclass = Yes'),
 Text(2022.75, 181.19999999999982, 'gini = 0.519\nsamples = 346\nvalue = [30
6, 15, 0, 243]\nclass = Yes'),
 Text(2162.25, 181.19999999999982, 'gini = 0.501\nsamples = 827\nvalue = [84
7, 123, 0, 326]\nclass = Yes'),
 Text(3348.0, 1630.8000000000002, 'Unnamed: 0 <= 9281.5\ngini = 0.649\nsample
s = 13827\nvalue = [5190, 9355, 0, 7414]\nclass = No'),
 Text(2790.0, 1268.4, 'year <= 2019.5\ngini = 0.648\nsamples = 11044\nvalue =
[3902, 6956, 0, 6685]\nclass = No'),
 Text(2511.0, 906.0, 'mpg <= 38.95\ngini = 0.652\nsamples = 9443\nvalue = [35
65, 6114, 0, 5352]\nclass = No'),
 Text(2371.5, 543.5999999999999, 'price <= 25965.0\ngini = 0.605\nsamples = 1
614\nvalue = [913, 371, 0, 1257]\nclass = No'),
 Text(2301.75, 181.19999999999982, 'gini = 0.618\nsamples = 409\nvalue = [14
6, 334, 0, 172]\nclass = No'),
 Text(2441.25, 181.19999999999982, 'gini = 0.505\nsamples = 1205\nvalue = [76
7, 37, 0, 1085]\nclass = No'),
 Text(2650.5, 543.5999999999999, 'engineSize <= 1.7\ngini = 0.636\nsamples =
7829\nvalue = [2652, 5743, 0, 4095]\nclass = No'),
 Text(2580.75, 181.19999999999982, 'gini = 0.475\nsamples = 4732\nvalue = [94
2, 5194, 0, 1404]\nclass = No'),
 Text(2720.25, 181.19999999999982, 'gini = 0.573\nsamples = 3097\nvalue = [17
10, 549, 0, 2691]\nclass = No'),
 Text(3069.0, 906.0, 'price <= 20550.0\ngini = 0.588\nsamples = 1601\nvalue =
[337, 842, 0, 1333]\nclass = No'),
 Text(2929.5, 543.5999999999999, 'Unnamed: 0 <= 8582.0\ngini = 0.222\nsamples
= 405\nvalue = [28, 552, 0, 49]\nclass = No'),
 Text(2859.75, 181.19999999999982, 'gini = 0.203\nsamples = 386\nvalue = [25,
537, 0, 42]\nclass = No'),
 Text(2999.25, 181.19999999999982, 'gini = 0.547\nsamples = 19\nvalue = [3, 1
5, 0, 7]\nclass = No'),
 Text(3208.5, 543.5999999999999, 'price <= 26369.5\ngini = 0.484\nsamples = 1
196\nvalue = [309, 290, 0, 1284]\nclass = No'),
 Text(3138.75, 181.19999999999982, 'gini = 0.608\nsamples = 392\nvalue = [88,
224, 0, 291]\nclass = No'),
 Text(3278.25, 181.19999999999982, 'gini = 0.366\nsamples = 804\nvalue = [22
1, 66, 0, 993]\nclass = No'),
 Text(3906.0, 1268.4, 'engineSize <= 1.8\ngini = 0.593\nsamples = 2783\nvalue
= [1288, 2399, 0, 729]\nclass = No'),
 Text(3627.0, 906.0, 'Unnamed: 0 <= 13871.0\ngini = 0.396\nsamples = 1658\nva
lue = [431, 1985, 0, 212]\nclass = No'),
 Text(3487.5, 543.5999999999999, 'price <= 18996.0\ngini = 0.37\nsamples = 14
08\nvalue = [362, 1738, 0, 145]\nclass = No'),
 Text(3417.75, 181.19999999999982, 'gini = 0.218\nsamples = 1063\nvalue = [12
0, 1483, 0, 82]\nclass = No'),
 Text(3557.25, 181.19999999999982, 'gini = 0.593\nsamples = 345\nvalue = [24
2, 255, 0, 63]\nclass = No'),
 Text(3766.5, 543.5999999999999, 'price <= 20296.0\ngini = 0.521\nsamples = 2
50\nvalue = [69, 247, 0, 67]\nclass = No'),
 Text(3696.75, 181.19999999999982, 'gini = 0.31\nsamples = 182\nvalue = [41,
222, 0, 9]\nclass = No'),
 Text(3836.25, 181.19999999999982, 'gini = 0.613\nsamples = 68\nvalue = [28,
25, 0, 58]\nclass = No'),
 Text(4185.0, 906.0, 'Unnamed: 0 <= 10117.5\ngini = 0.633\nsamples = 1125\nva
lue = [857, 414, 0, 517]\nclass = Yes'),
 Text(4045.5, 543.5999999999999, 'mpg <= 60.75\ngini = 0.447\nsamples = 181\n
value = [204, 39, 0, 42]\nclass = Yes'),
 Text(3975.75, 181.19999999999982, 'gini = 0.36\nsamples = 155\nvalue = [194,
27, 0, 26]\nclass = Yes'),
 Text(4115.25, 181.19999999999982, 'gini = 0.654\nsamples = 26\nvalue = [10,
12, 0, 16]\nclass = No'),
 Text(4324.5, 543.5999999999999, 'engineSize <= 2.05\ngini = 0.649\nsamples =
944\nvalue = [653, 375, 0, 475]\nclass = Yes'),
 Text(4254.75, 181.19999999999982, 'gini = 0.662\nsamples = 793\nvalue = [48
6, 359, 0, 419]\nclass = Yes'),
 Text(4394.25, 181.19999999999982, 'gini = 0.452\nsamples = 151\nvalue = [16
7, 16, 0, 56]\nclass = Yes')]
```



```
In [ ]:
```