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

In [2]:
```
df=pd.read_csv(r"C9_Data.csv")
df
```

Out[2]:

	row_id	user_id	timestamp	gate_id
0	0	18	2022-07-29 09:08:54	7
1	1	18	2022-07-29 09:09:54	9
2	2	18	2022-07-29 09:09:54	9
3	3	18	2022-07-29 09:10:06	5
4	4	18	2022-07-29 09:10:08	5
...	...	...	...	...
37513	37513	6	2022-12-31 20:38:56	11
37514	37514	6	2022-12-31 20:39:22	6
37515	37515	6	2022-12-31 20:39:23	6
37516	37516	6	2022-12-31 20:39:31	9
37517	37517	6	2022-12-31 20:39:31	9

37518 rows × 4 columns

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37518 entries, 0 to 37517
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   row_id     37518 non-null  int64 
 1   user_id    37518 non-null  int64 
 2   timestamp  37518 non-null  object
 3   gate_id    37518 non-null  int64 
dtypes: int64(3), object(1)
memory usage: 1.1+ MB
```

In [4]:
```
df=df.dropna()
```

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
row_id       0
user_id      0
timestamp    0
gate_id      0
dtype: int64
```

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

Out[6]:

	row_id	user_id	gate_id
count	37518.000000	37518.000000	37518.000000
mean	18758.500000	28.219015	6.819607
std	10830.658036	17.854464	3.197746
min	0.000000	0.000000	-1.000000
25%	9379.250000	12.000000	4.000000
50%	18758.500000	29.000000	6.000000
75%	28137.750000	47.000000	10.000000
max	37517.000000	57.000000	16.000000

In [7]:
```
df.columns
```

Out[7]:
```
Index(['row_id', 'user_id', 'timestamp', 'gate_id'], dtype='object')
```

In [8]:
```
df['Education'].value_counts()
```

Out[8]:
```
Graduate        383
Not Graduate     97
Name: Education, dtype: int64
```

In [9]:
```
g1={"Education":{'Graduate':1,'Not Graduate':2}}
df=df.replace(g1)
print(df)
```

```
      Loan_ID  Gender Married Dependents  Education Self_Employed  \
1    LP001003    Male     Yes          1          1            No
2    LP001005    Male     Yes          0          1           Yes
3    LP001006    Male     Yes          0          2            No
4    LP001008    Male      No          0          1            No
5    LP001011    Male     Yes          2          1           Yes
..        ...     ...     ...        ...        ...           ...
609  LP002978  Female      No          0          1            No
610  LP002979    Male     Yes         3+          1            No
611  LP002983    Male     Yes          1          1            No
612  LP002984    Male     Yes          2          1            No
613  LP002990  Female      No          0          1           Yes

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
1               4583             1508.0       128.0             360.0
2               3000                0.0        66.0             360.0
3               2583             2358.0       120.0             360.0
4               6000                0.0       141.0             360.0
5               5417             4196.0       267.0             360.0
..               ...                ...         ...               ...
609             2900                0.0        71.0             360.0
610             4106                0.0        40.0             180.0
611             8072              240.0       253.0             360.0
612             7583                0.0       187.0             360.0
613             4583                0.0       133.0             360.0

     Credit_History Property_Area Loan_Status
1               1.0         Rural           N
2               1.0         Urban           Y
3               1.0         Urban           Y
4               1.0         Urban           Y
5               1.0         Urban           Y
..              ...           ...         ...
609             1.0         Rural           Y
610             1.0         Rural           Y
611             1.0         Urban           Y
612             1.0         Urban           Y
613             0.0     Semiurban           N

[480 rows x 13 columns]
```

In [10]:
```
x=df[['ApplicantIncome','CoapplicantIncome','LoanAmount','Loan_Amount_Term','C
y=df["Education"]
```

In [11]:
```
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 [12]:
```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

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

In [13]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]}
```

In [14]:
```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
grid_search.fit(x_train,y_train)
```

Out[14]:
```
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 [15]:
```
grid_search.best_score_
```

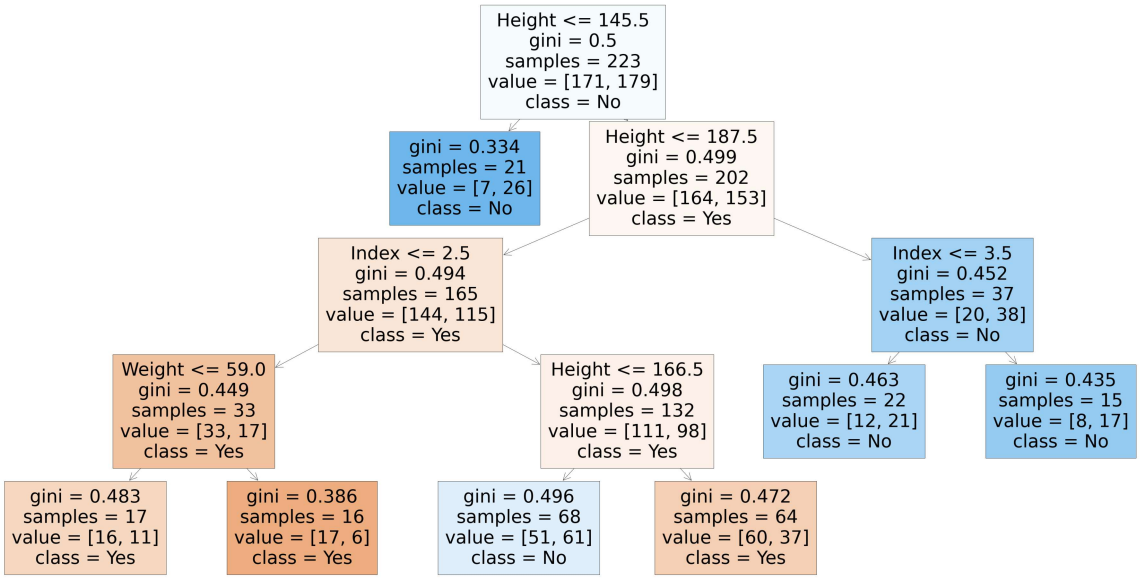
Out[15]:
```
0.5571428571428572
```

In [16]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]}
```

In [17]:
```
rfc_best=grid_search.best_estimator_
```

In [18]:
```
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[18]:
```
[Text(2232.0, 1956.96, 'Height <= 145.5\ngini = 0.5\nsamples = 223\nvalue =
[171, 179]\nclass = No'),
 Text(1826.1818181818182, 1522.0800000000002, 'gini = 0.334\nsamples = 21\nva
lue = [7, 26]\nclass = No'),
 Text(2637.818181818182, 1522.0800000000002, 'Height <= 187.5\ngini = 0.499\n
samples = 202\nvalue = [164, 153]\nclass = Yes'),
 Text(1623.2727272727273, 1087.2, 'Index <= 2.5\ngini = 0.494\nsamples = 165
\nvalue = [144, 115]\nclass = Yes'),
 Text(811.6363636363636, 652.3200000000002, 'Weight <= 59.0\ngini = 0.449\nsa
mples = 33\nvalue = [33, 17]\nclass = Yes'),
 Text(405.8181818181818, 217.44000000000005, 'gini = 0.483\nsamples = 17\nval
ue = [16, 11]\nclass = Yes'),
 Text(1217.4545454545455, 217.44000000000005, 'gini = 0.386\nsamples = 16\nva
lue = [17, 6]\nclass = Yes'),
 Text(2434.909090909091, 652.3200000000002, 'Height <= 166.5\ngini = 0.498\ns
amples = 132\nvalue = [111, 98]\nclass = Yes'),
 Text(2029.090909090909, 217.44000000000005, 'gini = 0.496\nsamples = 68\nval
ue = [51, 61]\nclass = No'),
 Text(2840.7272727272725, 217.44000000000005, 'gini = 0.472\nsamples = 64\nva
lue = [60, 37]\nclass = Yes'),
 Text(3652.3636363636365, 1087.2, 'Index <= 3.5\ngini = 0.452\nsamples = 37\n
value = [20, 38]\nclass = No'),
 Text(3246.5454545454545, 652.3200000000002, 'gini = 0.463\nsamples = 22\nval
ue = [12, 21]\nclass = No'),
 Text(4058.181818181818, 652.3200000000002, 'gini = 0.435\nsamples = 15\nvalu
e = [8, 17]\nclass = No')]
```



In [ ]:
```
```