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('loan_train.csv')
df
```

d	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_
o	0	Graduate	No	5849	0.0	NaN	
s	1	Graduate	No	4583	1508.0	128.0	
s	0	Graduate	Yes	3000	0.0	66.0	
s	0	Not Graduate	No	2583	2358.0	120.0	
o	0	Graduate	No	6000	0.0	141.0	
..	...	...	...	...	...	...	
o	0	Graduate	No	2900	0.0	71.0	
s	3+	Graduate	No	4106	0.0	40.0	
s	1	Graduate	No	8072	240.0	253.0	

In [15]:

```
df.columns
```

Out[15]:

```
Index(['Loan_ID', 'Gender', 'Married', 'Dependents', 'Education',
       'Self_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmoun
t',
       'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Statu
s'],
      dtype='object')
```

In [16]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 614 entries, 0 to 613
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Loan_ID            614 non-null    object 
 1   Gender             601 non-null    object 
 2   Married            611 non-null    object 
 3   Dependents         599 non-null    object 
 4   Education          614 non-null    object 
 5   Self_Employed      582 non-null    object 
 6   ApplicantIncome    614 non-null    int64  
 7   CoapplicantIncome  614 non-null    float64
 8   LoanAmount         502 non-null    float64
 9   Loan_Amount_Term   598 non-null    float64
 10  Credit_History     564 non-null    float64
 11  Property_Area      614 non-null    object 
 12  Loan_Status        614 non-null    object 
dtypes: float64(4), int64(1), object(8)
memory usage: 62.5+ KB
```

In [17]:

```
df['Loan_Status'].value_counts()
```

Out[17]:

```
Y    422
N    192
Name: Loan_Status, dtype: int64
```

In [18]:

```
x=df[['ApplicantIncome']]
y=df['Loan_Status']
```

In [19]:

```
d={"Loan_Status":{'Y':1,'N':2}}
df=df.replace(df)
print(df)
```

```
      Loan_ID  Gender Married Dependents     Education Self_Employed  \
0    LP001002    Male      No          0      Graduate            No   
1    LP001003    Male     Yes          1      Graduate            No   
2    LP001005    Male     Yes          0      Graduate           Yes   
3    LP001006    Male     Yes          0  Not Graduate            No   
4    LP001008    Male      No          0      Graduate            No   
..        ...     ...     ...        ...           ...           ...   
609  LP002978  Female      No          0      Graduate            No   
610  LP002979    Male     Yes         3+      Graduate            No   
611  LP002983    Male     Yes          1      Graduate            No   
612  LP002984    Male     Yes          2      Graduate            No   
613  LP002990  Female      No          0      Graduate           Yes   

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
0               5849                0.0         NaN             360.0   
1               4583             1508.0       180.0             360.0   
2               3000                0.0       133.0             360.0   
3               2583             2358.0       180.0             360.0   
4               6000                0.0         NaN             360.0   
..               ...                ...         ...               ...   
609             2900                0.0       164.0             360.0   
610             4106                0.0        95.0             360.0   
611             8072             5000.0       180.0             360.0   
612             7583                0.0       180.0             360.0   
613             4583                0.0         NaN             360.0   

     Credit_History Property_Area Loan_Status  
0               1.0         Urban           Y  
1               1.0         Rural           N  
2               1.0         Urban           Y  
3               1.0         Urban           Y  
4               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             1.0     Semiurban           N  

[614 rows x 13 columns]
```

In [20]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.70)
```

In [21]:

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[21]:

```
RandomForestClassifier()
```

Depth of Tree

In [22]:

```
parameters={"max_depth":[1,2,3,4,5],"min_samples_leaf":[5,23,45,76,78],'n_estimators':[1
```

Cross Validate

In [23]:

```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

Out[23]:

```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 23, 45, 76, 78],
                         'n_estimators': [10, 23, 45, 65, 7]},
             scoring='accuracy')
```

Score

In [24]:

```
grid_search.best_score_
```

Out[24]:

```
0.7228260869565217
```

In [25]:

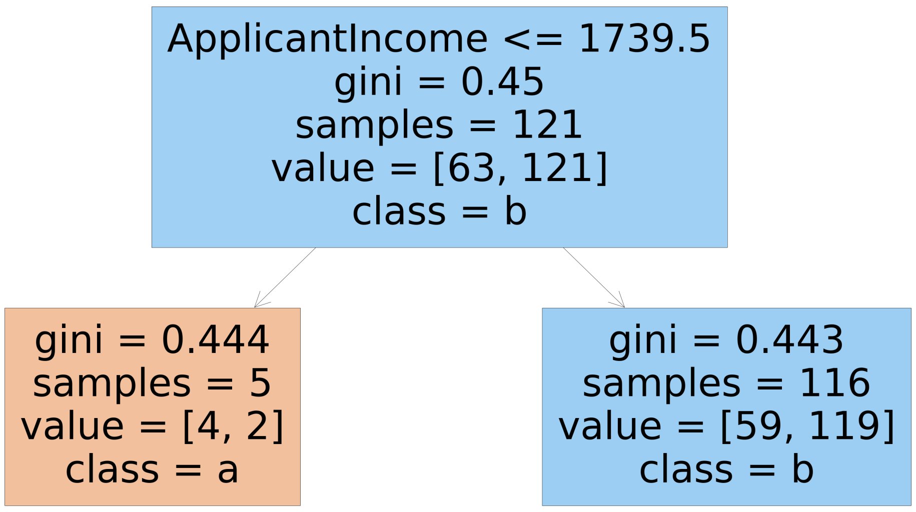
```
rfc_best=grid_search.best_estimator_
```

In [26]:

```
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['a','b'],filled=Tr
```

Out[26]:

```
[Text(2232.0, 1630.8000000000002, 'ApplicantIncome <= 1739.5\ngini = 0.45
\nsamples = 121\nvalue = [63, 121]\nclass = b'),
 Text(1116.0, 543.5999999999999, 'gini = 0.444\nsamples = 5\nvalue = [4,
2]\nclass = a'),
 Text(3348.0, 543.5999999999999, 'gini = 0.443\nsamples = 116\nvalue = [5
9, 119]\nclass = b')]
```



In [ ]: