In [ ]:
```
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import export_graphviz
from IPython.display import Image
```

In [ ]:
```
data = pd.read_csv("data.csv")
data
```

Out[ ]:

	Age	Income	Gender	MaritialStatus	Buys
0	<21	High	Male	Single	No
1	<21	High	Male	Married	No
2	21-35	High	Male	Single	Yes
3	>35	Medium	Male	Single	Yes
4	>35	Low	Female	Single	Yes
5	>35	Low	Female	Married	No
6	21-35	Low	Female	Married	Yes
7	<21	Medium	Male	Single	No
8	<21	Low	Female	Married	Yes
9	>35	Medium	Female	Single	Yes
10	<21	Medium	Female	Married	Yes
11	21-35	Medium	Male	Married	Yes
12	21-35	High	Female	Single	Yes
13	>35	Medium	Male	Married	No

In [ ]:
```
le=LabelEncoder();
x=data.iloc[:,:-1]
x=x.apply(le.fit_transform)
print("Age:",list( zip(data.iloc[:,0], x.iloc[:,0])))
print("\nIncome:",list( zip(data.iloc[:,1], x.iloc[:,1])))
print("\nGender:",list( zip(data.iloc[:,2], x.iloc[:,2])))
print("\nmaritialStatus:",list( zip(data.iloc[:,3], x.iloc[:,3])))
```

```
Age: [('<21', 1), ('<21', 1), ('21-35', 0), ('>35', 2), ('>35', 2), ('>35',
2), ('21-35', 0), ('<21', 1), ('<21', 1), ('>35', 2), ('<21', 1), ('21-35',
0), ('21-35', 0), ('>35', 2)]

Income: [('High', 0), ('High', 0), ('High', 0), ('Medium', 2), ('Low', 1),
('Low', 1), ('Low', 1), ('Medium', 2), ('Low', 1), ('Medium', 2), ('Medium'
, 2), ('Medium', 2), ('High', 0), ('Medium', 2)]

Gender: [('Male', 1), ('Male', 1), ('Male', 1), ('Male', 1), ('Female', 0),
('Female', 0), ('Female', 0), ('Male', 1), ('Female', 0), ('Female', 0), ('
Female', 0), ('Male', 1), ('Female', 0), ('Male', 1)]

maritialStatus: [('Single', 1), ('Married', 0), ('Single', 1), ('Single', 1
), ('Single', 1), ('Married', 0), ('Married', 0), ('Single', 1), ('Married'
, 0), ('Single', 1), ('Married', 0), ('Married', 0), ('Single', 1), ('Marri
ed', 0)]
```

In [ ]:
```
x
```

Out[ ]:

	Age	Income	Gender	MaritialStatus
0	1	0	1	1
1	1	0	1	0
2	0	0	1	1
3	2	2	1	1
4	2	1	0	1
5	2	1	0	0
6	0	1	0	0
7	1	2	1	1
8	1	1	0	0
9	2	2	0	1
10	1	2	0	0
11	0	2	1	0
12	0	0	0	1
13	2	2	1	0

In [ ]:
```
y=data.iloc[:,-1]
```

In [ ]:
```
y
```

Out[ ]:
```
0      No
1      No
2     Yes
3     Yes
4     Yes
5      No
6     Yes
7      No
8     Yes
9     Yes
10    Yes
11    Yes
12    Yes
13     No
Name: Buys, dtype: object
```

In [ ]:
```
dt=DecisionTreeClassifier()
dt.fit(x,y)
```

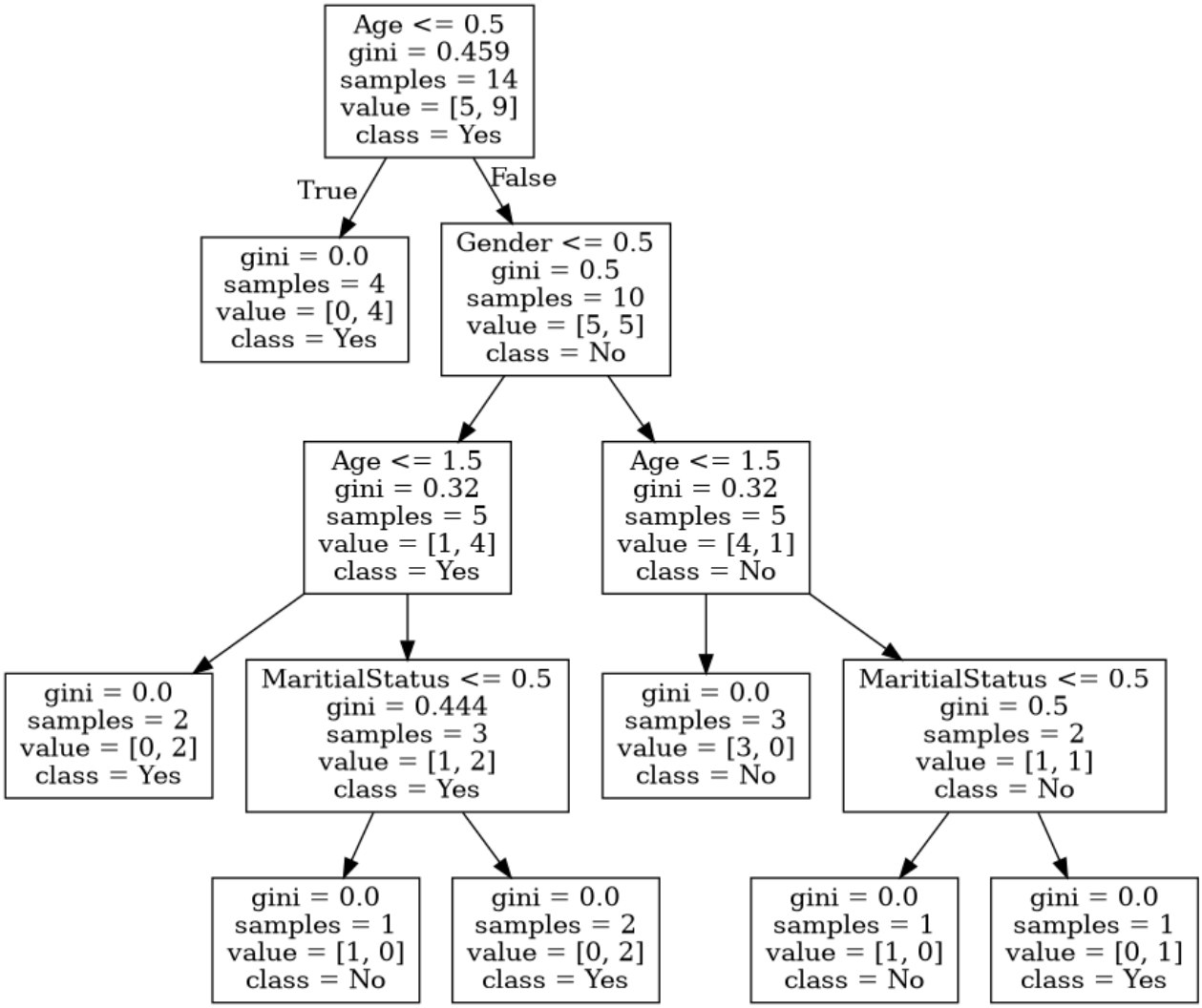
Out[ ]:
```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',
                       max_depth=None, max_features=None, max_leaf_nodes=No
ne,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort='deprecated',
                       random_state=None, splitter='best')
```

In [ ]:
```
#[Age < 21, Income = Low,Gender = Female, Marital Status = Married]
query=np.array([1,1,0,0])
pred=dt.predict([query])
pred[0]
```

Out[ ]:
```
'Yes'
```

In [ ]:
```
export_graphviz(dt,out_file="data.dot",feature_names=x.columns,class_names=
!dot -Tpng data.dot -o tree.png
Image("tree.png")
```

Out[ ]:


Age <= 0.5
gini = 0.459
samples = 14
value = [5, 9]
class = Yes
True
False
gini = 0.0
samples = 4
value = [0, 4]
class = Yes
Gender <= 0.5
gini = 0.5
samples = 10
value = [5, 5]
class = No
Age <= 1.5
gini = 0.32
samples = 5
value = [1, 4]
class = Yes
Age <= 1.5
gini = 0.32
samples = 5
value = [4, 1]
class = No
gini = 0.0
samples = 2
value = [0, 2]
class = Yes
MaritialStatus <= 0.5
gini = 0.444
samples = 3
value = [1, 2]
class = Yes
gini = 0.0
samples = 3
value = [3, 0]
class = No
MaritialStatus <= 0.5
gini = 0.5
samples = 2
value = [1, 1]
class = No
gini = 0.0
samples = 1
value = [1, 0]
class = No
gini = 0.0
samples = 2
value = [0, 2]
class = Yes
gini = 0.0
samples = 1
value = [1, 0]
class = No
gini = 0.0
samples = 1
value = [0, 1]
class = Yes