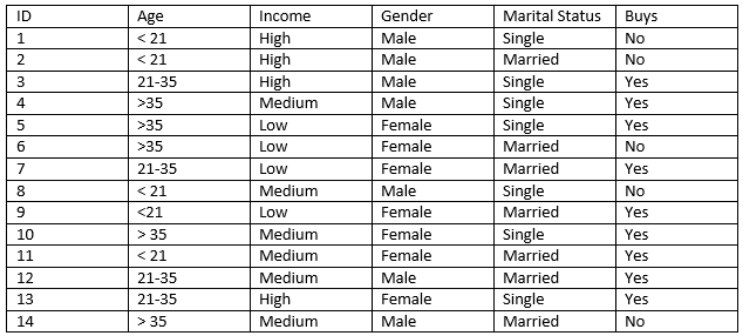
NAME: Kshitija Nalawade

ROLL NO: 4202041

BATCH: BB2

**Problem Statement:** A dataset collected in a cosmetics shop showing details of customers and whether or not they responded to a special offer to buy a new lip-stick is shown in table below. Use this dataset to build a decision tree, with Buys as the target variable, to help in buying lip-sticks in the future. Find the root node of decision tree. According to the decision tree you have made from previous training data set, what is the decision for the test data: [Age < 21, Income = Low, Gender = Female, Marital Status = Married]?



**CODE:**

In [1]:

**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 [2]:

data = pd.read\_csv("data.csv")

data

Out[2]:

|  | **ID** | **Age** | **Income** | **Gender** | **Marital Status** | **Buys** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 1 | <21 | High | Male | Single | No |
| **1** | 2 | <21 | High | Male | Married | No |
| **2** | 3 | 21-35 | High | Male | Single | Yes |
| **3** | 4 | >35 | Medium | Male | Single | Yes |
| **4** | 5 | >35 | Low | Female | Single | Yes |
| **5** | 6 | >35 | Low | Female | Married | No |
| **6** | 7 | 21-35 | Low | Female | Married | Yes |
| **7** | 8 | <21 | Medium | Male | Single | No |
| **8** | 9 | <21 | Low | Female | Married | Yes |
| **9** | 10 | >35 | Medium | Female | Single | Yes |
| **10** | 11 | <21 | Medium | Female | Married | Yes |
| **11** | 12 | 21-35 | Medium | Male | Married | Yes |
| **12** | 13 | 21-35 | High | Female | Single | Yes |
| **13** | 14 | >35 | Medium | Male | Married | No |

In [3]:

le=LabelEncoder();

x=data.iloc[:,:-1]

x=x.apply(le.fit\_transform)

print("Age:",list( zip(data.iloc[:,0], x.iloc[:,0])))

print("**\n**Income:",list( zip(data.iloc[:,1], x.iloc[:,1])))

print("**\n**Gender:",list( zip(data.iloc[:,2], x.iloc[:,2])))

print("**\n**maritialStatus:",list( zip(data.iloc[:,3], x.iloc[:,3])))

Age: [(1, 0), (2, 1), (3, 2), (4, 3), (5, 4), (6, 5), (7, 6), (8, 7), (9, 8), (10, 9), (11, 10), (12, 11), (13, 12), (14, 13)]

Income: [('<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)]

Gender: [('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)]

maritialStatus: [('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)]

In [4]:

x

Out[4]:

|  | **ID** | **Age** | **Income** | **Gender** | **Marital Status** |
| --- | --- | --- | --- | --- | --- |
| **0** | 0 | 1 | 0 | 1 | 1 |
| **1** | 1 | 1 | 0 | 1 | 0 |
| **2** | 2 | 0 | 0 | 1 | 1 |
| **3** | 3 | 2 | 2 | 1 | 1 |
| **4** | 4 | 2 | 1 | 0 | 1 |
| **5** | 5 | 2 | 1 | 0 | 0 |
| **6** | 6 | 0 | 1 | 0 | 0 |
| **7** | 7 | 1 | 2 | 1 | 1 |
| **8** | 8 | 1 | 1 | 0 | 0 |
| **9** | 9 | 2 | 2 | 0 | 1 |
| **10** | 10 | 1 | 2 | 0 | 0 |
| **11** | 11 | 0 | 2 | 1 | 0 |
| **12** | 12 | 0 | 0 | 0 | 1 |
| **13** | 13 | 2 | 2 | 1 | 0 |

In [5]:

y=data.iloc[:,-1]

In [6]:

y

Out[6]:

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 [9]:

**from** **sklearn.tree** **import** DecisionTreeClassifier

classifier=DecisionTreeClassifier(criterion='entropy')

classifier.fit(x,y)

Out[9]:

DecisionTreeClassifier(criterion='entropy')

In [10]:

*#predict value for the given Expression*

test\_x=np.array([1,0,1,1,0])

pred\_y=classifier.predict([test\_x])

print("Predicted class for input [Age < 21, Income = High,Gender = Male, Marital Status = Single]**\n**", test\_x," is ",pred\_y[0])

Predicted class for input [Age < 21, Income = High,Gender = Male, Marital Status = Single]

[1 0 1 1 0] is No

In [10]:

export\_graphviz(dt,out\_file="data.dot",feature\_names=x.columns,class\_names=["No","Yes"])

!dot -Tpng data.dot -o tree.png

Image("tree.png")

Out[23]:

