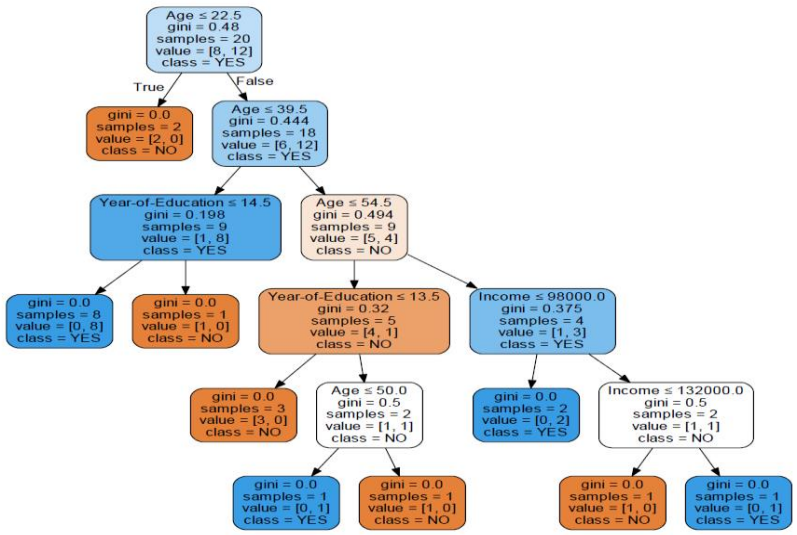
Before you run the python files, you need to run “Anaconda Prompt” in the same location as “Spyder”.

“Anaconda Prompt” is a command line window.

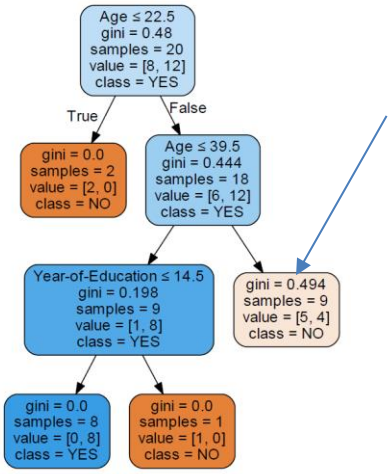
```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from sklearn import tree
4 import pandas as pd
5 my_data1=pd.read_csv('purchase2.csv')
6 clf = tree.DecisionTreeClassifier()
7 X=my_data1[['Age', 'Income', 'Year-of-Education']]
8 y=my_data1['Favorite']
9 clf.fit(X, y)
10 fig = plt.figure(figsize=(16,14))
11 tree.plot_tree(clf,feature_names=X.columns,fontsize=12,filled=True)
12
13 clf = tree.DecisionTreeClassifier(min_impurity_decrease =0.0672)
14 X=my_data1[['Age', 'Income', 'Year-of-Education']]
15 y=my_data1['Favorite']
16 clf.fit(X, y)
17 fig = plt.figure(figsize=(10,10))
18 tree.plot_tree(clf,feature_names=X.columns,fontsize=12,filled=True)
19
20
21 from sklearn.model_selection import train_test_split
22 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
23
24
25 for decrease in np.arange(0, 0.1,0.01):
26     clf = tree.DecisionTreeClassifier(min_impurity_decrease =decrease)
27     clf.fit(X_train, y_train)
28     print("min_impurity_decrease=%f score=%f" %(decrease,clf.score(X_test, y_test)))
29
30
```

This file contains 3 parts

The first part creates a tree



The second part creates another tree, which shows the effect of setting min_impurity_decrease =0.0674, since we know the Impurity Decrease for the third row right node is 0.0673, we know the tree will stop splitting at this node.



The third part shows the scores of tree by running min_impurity_decrease from 0 to 0.1

min_impurity_decrease=0.000000 score=0.500000

min_impurity_decrease=0.010000 score=0.500000

min_impurity_decrease=0.020000 score=0.500000

min_impurity_decrease=0.030000 score=0.500000

min_impurity_decrease=0.040000 score=0.500000

min_impurity_decrease=0.050000 score=0.500000

min_impurity_decrease=0.060000 score=0.500000

min_impurity_decrease=0.070000 score=0.250000

min_impurity_decrease=0.080000 score=0.250000

min_impurity_decrease=0.090000 score=0.250000

From the above results, we see min_impurity_decrease =0, 0.01,0.02,0.03,0.04,0.05,0.06 are better than min_impurity_decrease=0.07,0.08,0.09