# Importing the Libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#%matplotlib inline#for encoding
from sklearn.preprocessing import LabelEncoder#for train test splitting
from sklearn.model_selection import train_test_split#for decision tree objec
from sklearn.tree import DecisionTreeClassifier#for checking testing results
from sklearn.metrics import classification_report, confusion_matrix#for visu
from sklearn.tree import plot_tree
```

Read the files

In [2]:
```
from google.colab import files
uploaded = files.upload()
```

Choose Files no files selected Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving Weather.csv to Weather.csv
```

In [3]:
```
df = pd.read_csv("Weather.csv")
print(df)
```

```
     Outlook  Temp Humidity  Windy Play Golf
0      Rainy   Hot     High  False        No
1      Rainy   Hot     High   True        No
2      Rainy  Mild     High  False        No
3      Rainy  Cool   Normal  False       Yes
4      Rainy  Mild   Normal   True       Yes
5   Overoact   Hot     High  False       Yes
6   Overoact  Cool   Normal   True       Yes
7   Overoact  Mild     High   True       Yes
8   Overoact   Hot   Normal  False       Yes
9      Sunny  Mild     High  False       Yes
10     Sunny  Cool   Normal  False       Yes
11     Sunny  Cool   Normal   True        No
12     Sunny  Mild   Normal  False       Yes
13     Sunny  Mild     High   True        No
```

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   Outlook    14 non-null     object
 1   Temp       14 non-null     object
 2   Humidity   14 non-null     object
 3   Windy      14 non-null     bool  
 4   Play Golf  14 non-null     object
dtypes: bool(1), object(4)
memory usage: 590.0+ bytes
```

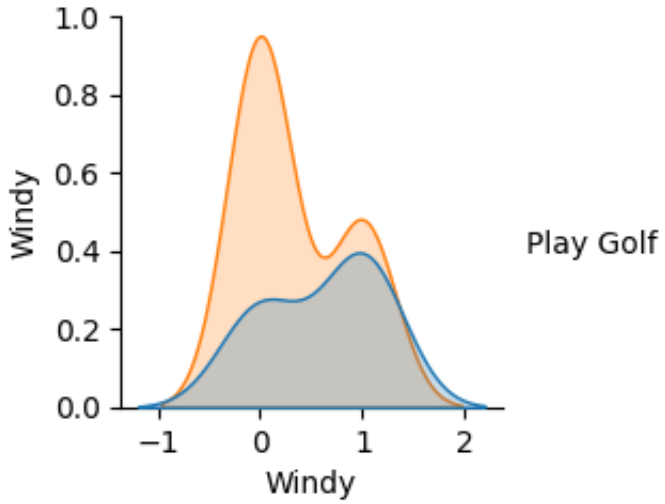
Check if there are missing values

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

Out[5]:
```
Outlook      False
Temp         False
Humidity     False
Windy        False
Play Golf    False
dtype: bool
```

In [6]:
```
# let's plot pair plot to visualise the attributes all at once
sns.pairplot(data=df, hue = 'Play Golf')
```

Out[6]:
```
<seaborn.axisgrid.PairGrid at 0x7fe9ad8ba850>
```




In [7]:
```
# correlation matrix
sns.heatmap(df.corr())
```

Out[7]:
```
<Axes: >
```


# Data Preprocessing

Identify the target columns

In [8]:
```
#separate the target variable(y) and features(X) as follows
target = df['Play Golf']
df1 = df.copy()
df1 = df1.drop('Play Golf', axis =1)
```

Convert the variables to numeric values

In [9]:
```
# Data has categorical variables stored in it we will encode it in numeric v
X = df1.apply(LabelEncoder().fit_transform)
print(X)
```

| | Outlook | Temp | Humidity | Windy |
|---|---|---|---|---|
| 0 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 | 1 |
| 2 | 1 | 2 | 0 | 0 |
| 3 | 1 | 0 | 1 | 0 |
| 4 | 1 | 2 | 1 | 1 |
| 5 | 0 | 1 | 0 | 0 |
| 6 | 0 | 0 | 1 | 1 |
| 7 | 0 | 2 | 0 | 1 |
| 8 | 0 | 1 | 1 | 0 |
| 9 | 2 | 2 | 0 | 0 |
| 10 | 2 | 0 | 1 | 0 |
| 11 | 2 | 0 | 1 | 1 |
| 12 | 2 | 2 | 1 | 0 |
| 13 | 2 | 2 | 0 | 1 |

In [10]:
```
target
```

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

In [11]:
```
#label encoding
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
target = le.fit_transform(target)
target
```

Out[11]:
```
array([0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0])
```

In [12]:
```
y = target
print(y)
```

```
[0 0 0 1 1 1 1 1 1 1 1 0 1 0]
```

In [13]:
```
from sklearn.model_selection import KFold, cross_val_score, train_test_split
# Splitting the data - 80:20 ratio
X_train, X_test, y_train, y_test = train_test_split(X , y, test_size = 0.2,
print("Testing split input- ", X_test.shape)
```

```
Testing split input-  (3, 4)
```

# Modeling Tree and testing it

In [14]:
```
from sklearn import tree
# Defining the decision tree algorithmdtree=DecisionTreeClassifier()
dtree = tree.DecisionTreeClassifier()
dtree.fit(X_train,y_train)
print('Decision Tree Classifier Created')
```

```
Decision Tree Classifier Created
```

In [15]:
```
# Predicting the values of test data
y_pred = dtree.predict(X_test)
print("Classification report - \n", classification_report(y_test,y_pred))
```

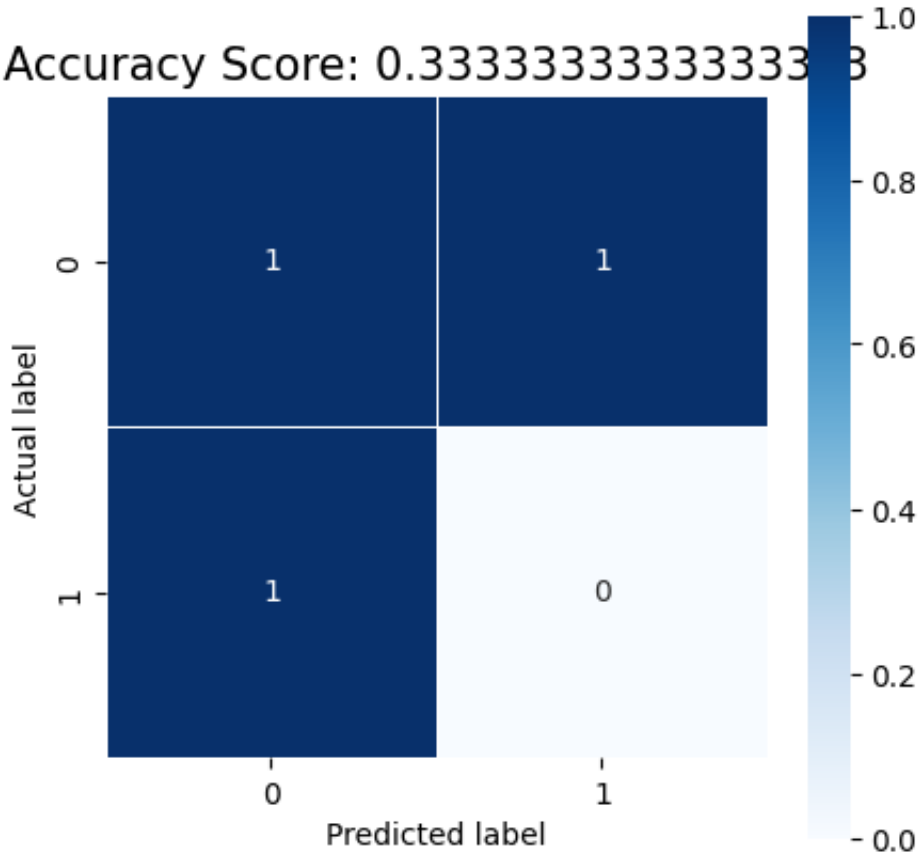
```
Classification report -
               precision    recall  f1-score   support

           0       0.50      0.50      0.50         2
           1       0.00      0.00      0.00         1

    accuracy                           0.33         3
   macro avg       0.25      0.25      0.25         3
weighted avg       0.33      0.33      0.33         3
```

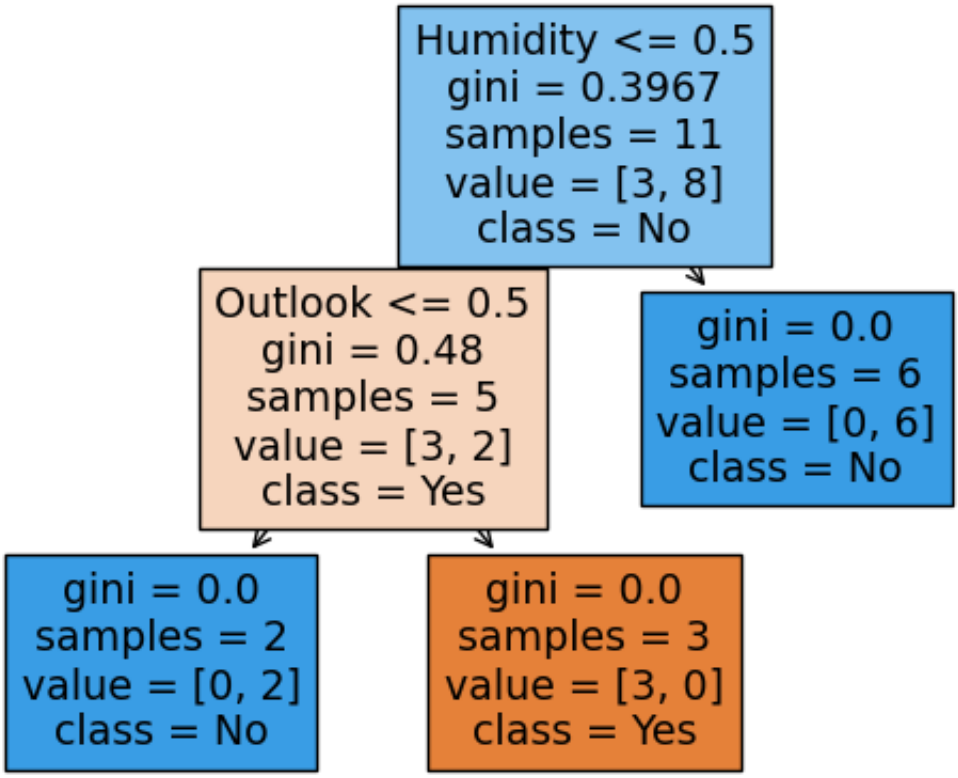
In [16]:
```
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(5,5))
sns.heatmap(data=cm,linewidths=.5, annot=True,square = True,  cmap = 'Blues'
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
all_sample_title = 'Accuracy Score: {0}'.format(dtree.score(X_test, y_test))
plt.title(all_sample_title, size = 15)
```

Out[16]:
```
Text(0.5, 1.0, 'Accuracy Score: 0.3333333333333333')
```




# Visualizing the decision tree

In [17]:
```
# Visualising the graph without the use of graphvizplt.figure(figsize = (20,
dec_tree = plot_tree(decision_tree=dtree, feature_names = df1.columns, class
```


Humidity <= 0.5
gini = 0.3967
samples = 11
value = [3, 8]
class = No
Outlook <= 0.5
gini = 0.48
samples = 5
value = [3, 2]
class = Yes
gini = 0.0
samples = 6
value = [0, 6]
class = No
gini = 0.0
samples = 2
value = [0, 2]
class = No
gini = 0.0
samples = 3
value = [3, 0]
class = Yes