In [18]:
```
import pandas as pd
import numpy as np
import sklearn
import matplotlib.pyplot as plt
from sklearn.model_selection \
import train_test_split
from sklearn.naive_bayes import GaussianNB
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler , LabelEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
```

In [19]:
```
df=pd.read_csv("NVDA_weekly_return_volatility.csv")
year=df['Year'].unique()
Q1_label=[]
yearly_mean=df.groupby('Year')['mean_return'].mean().values
for i in range(len(year)):
    for j in range(len(df)):
        if df['Year'][j]==year[i] and df["mean_return"][j]>yearly_mean[i]:
            Q1_label.append('green')
        elif df['Year'][j]==year[i]:
            Q1_label.append('red')
df['label']=Q1_label
Q1_X=df[df["Year"]==2017][["mean_return","volatility"]]
Q1_y=df[df["Year"]==2017]["label"]
Q2_X=df.loc[df["Year"]==2018][["mean_return","volatility"]]
Q2_y=df.loc[df["Year"]==2018]["label"]
NB_classifier = GaussianNB().fit(Q1_X, Q1_y)
prediction = NB_classifier.predict(Q2_X)
error_rate = np.mean(prediction != Q2_y)
print(error_rate)
```

```
0.22641509433962265
```

1. implement a Gaussian naive bayesian classifier and compute its accuracy for year 2

In [20]:
```
print("the year2 accuracy is",accuracy_score(Q2_y, NB_classifier.predict(Q2_X)))
```

```
the year2 accuracy is 0.7735849056603774
```

1. compute the confusion matrix for year 2

In [21]:
```
a= confusion_matrix(Q2_y, NB_classifier.predict(Q2_X))
print('the confusion matrix is\n',a)
```

```
the confusion matrix is
 [[23  6]
 [ 6 18]]
```

1. what is true positive rate and true negative rate for year 2

In [22]:
```
Q4_TN, Q4_FP, Q4_FN, Q4_TP = confusion_matrix(Q2_y, NB_classifier.predict(Q2_X))
Q4_TPR=Q4_TP/(Q4_TP+Q4_FN)
```

```
Q4_TNR=Q4_TN/(Q4_TN+Q4_FP)
print('true positive rate',Q4_TPR)
print('true negative rate',Q4_TNR)
```

```
true positive rate 0.75
true negative rate 0.7931034482758621
```

1. implement a trading strategy based on your labels for year 2 and compare the performance with the "buy-and-hold" strategy. Which strategy results in a larger amount at the end of the year?