```python
from datetime import datetime
import numpy as np
import pandas as pd
import matplotlib.pylab as plt
get_ipython().run_line_magic('matplotlib', 'inline')
from statsmodels.tsa.stattools import adfuller
from sklearn.metrics import mean_squared_error
from matplotlib.pylab import rcParams

rcParams['figure.figsize']=10,6


location=r"AirPassengers.csv"
df= pd.read_csv(location, encoding='gbk',parse_dates=['Month'],infer_datetime_format=True)
indf=df.set_index(['Month'])


indf.describe()
indf.rename(columns={'#Passengers':'Passengers'},inplace=True)
indf.head()
```
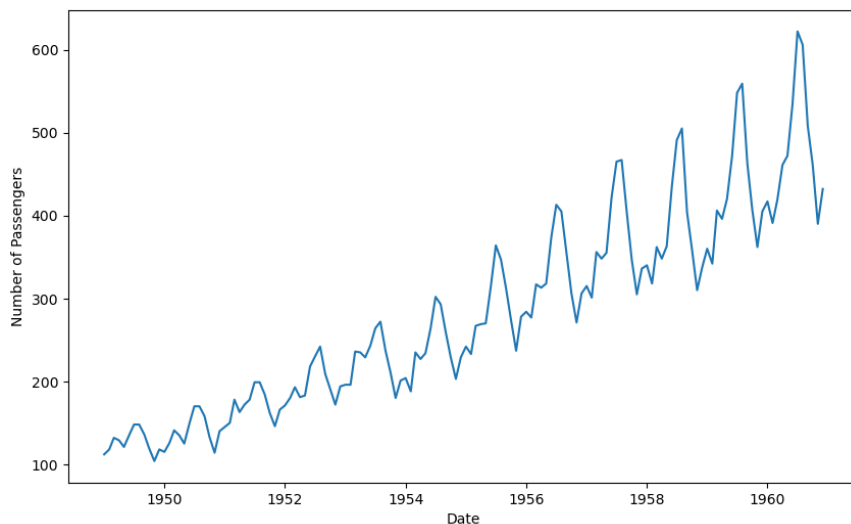
| | Passengers |
Month	
1949-01-01	112
1949-02-01	118
1949-03-01	132
1949-04-01	129
1949-05-01	121

```python
plt.xlabel('Date')
plt.ylabel('Number of Passengers')
plt.plot(indf)
```

[<matplotlib.lines.Line2D at 0x7fd0f009ae50>]


```python
indf['months'] = [x.month for x in indf.index]
indf['years'] = [x.year for x in indf.index]


indf.reset_index(drop=True, inplace=True)

X=indf.drop("Passengers",axis=1)
Y= indf["Passengers"]
X_train=X[:int (len(Y)*0.75)]
```

```
X_test=X[int(len(Y)*0.75):]
Y_train=Y[:int (len(Y)*0.75)]
Y_test=Y[int(len(Y)*0.75):]


from sklearn.ensemble import RandomForestRegressor

rf = RandomForestRegressor()
rf.fit(X_train, Y_train)


     RandomForestRegressor()


df1=df.set_index(['Month'])
df1.rename(columns={'#Passengers':'Passengers'},inplace=True)
train=df1.Passengers[:int (len(indf.Passengers)*0.75)]
test=df1.Passengers[int(len(indf.Passengers)*0.75):]
preds=rf.predict(X_test)
predictions=pd.DataFrame(preds,columns=['Passengers'])
predictions.index=test.index
plt.plot(train)
plt.plot(test, color='orange', label='actual')
plt.plot(predictions,color='green', label='Forecasts')
plt.xlabel('Date')
plt.ylabel('Number of Passengers')
plt.title("Forecast of AirPassengers")


     Text(0.5, 1.0, 'Forecast of AirPassengers')
```
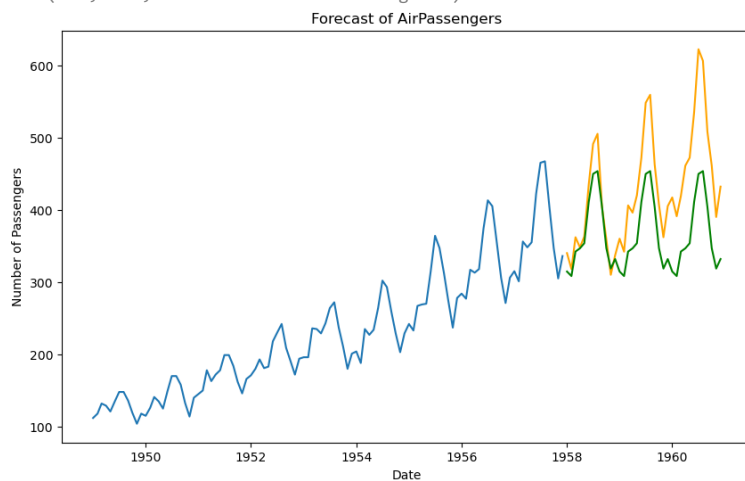


```
rf.score(X_train, Y_train)

     0.995041961894502


rf.score(X_test, Y_test)

     0.01408173241502042
```