



Chapter 19 - exercise 1: Monthly milk production

- Cho dữ liệu monthly-milk-production-pounds.csv. Áp dụng mô hình HoltWinters để dự báo lượng sản phẩm sữa cho 6 tháng đầu năm 1976.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.holtwinters import ExponentialSmoothing
```

Đọc dữ liệu, kiểm tra/định dạng thời gian

```
In [2]: df = pd.read_csv('monthly-milk-production-pounds.csv',
                        parse_dates=['Month'],
                        index_col='Month')
```

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 168 entries, 1962-01-01 to 1975-12-01
Data columns (total 1 columns):
milk_production_pounds_per_cow    168 non-null int64
dtypes: int64(1)
memory usage: 2.6 KB
```

```
In [4]: # https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html
# freq='H', 'D', 'W', 'M', 'MS': Hour, Day, Week, Month, Calendar month begin
df.index.freq = 'MS'
```

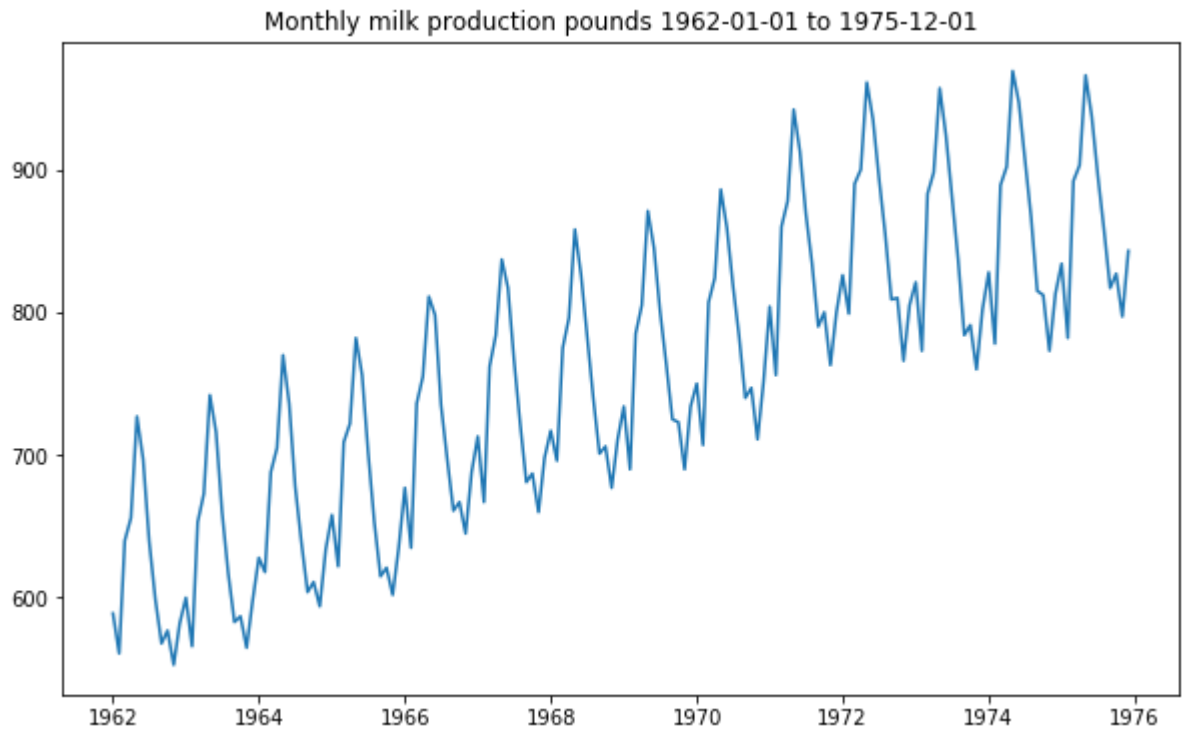
```
In [5]: df.head()
```

```
Out[5]:
```

milk_production_pounds_per_cow	
Month	
1962-01-01	589
1962-02-01	561
1962-03-01	640
1962-04-01	656
1962-05-01	727



```
In [6]: plt.figure(figsize=(10,6))  
plt.plot(df)  
plt.title("Monthly milk production pounds 1962-01-01 to 1975-12-01")  
plt.show()
```



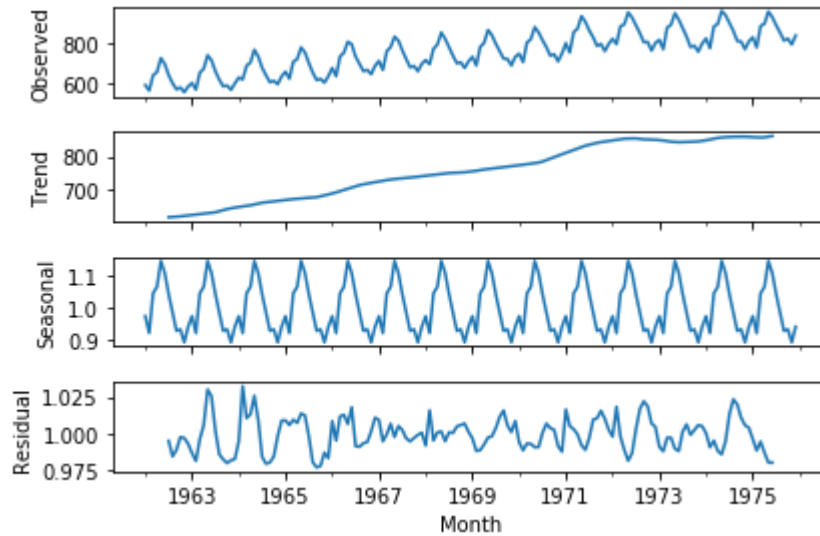
Decomposition

```
In [7]: from statsmodels.tsa.seasonal import seasonal_decompose  
result = seasonal_decompose(df, model='multiplicative')  
result
```

```
Out[7]: <statsmodels.tsa.seasonal.DecomposeResult at 0x212d8c4c358>
```



```
In [8]: result.plot()
plt.show()
```



Chia dữ liệu train/test => Áp dụng mô hình

```
In [9]: train, test = df.iloc[:130, 0], df.iloc[130:, 0]
```

```
In [10]: train[0:5]
```

```
Out[10]: Month
1962-01-01    589
1962-02-01    561
1962-03-01    640
1962-04-01    656
1962-05-01    727
Freq: MS, Name: milk_production_pounds_per_cow, dtype: int64
```

```
In [11]: test[0:5]
```

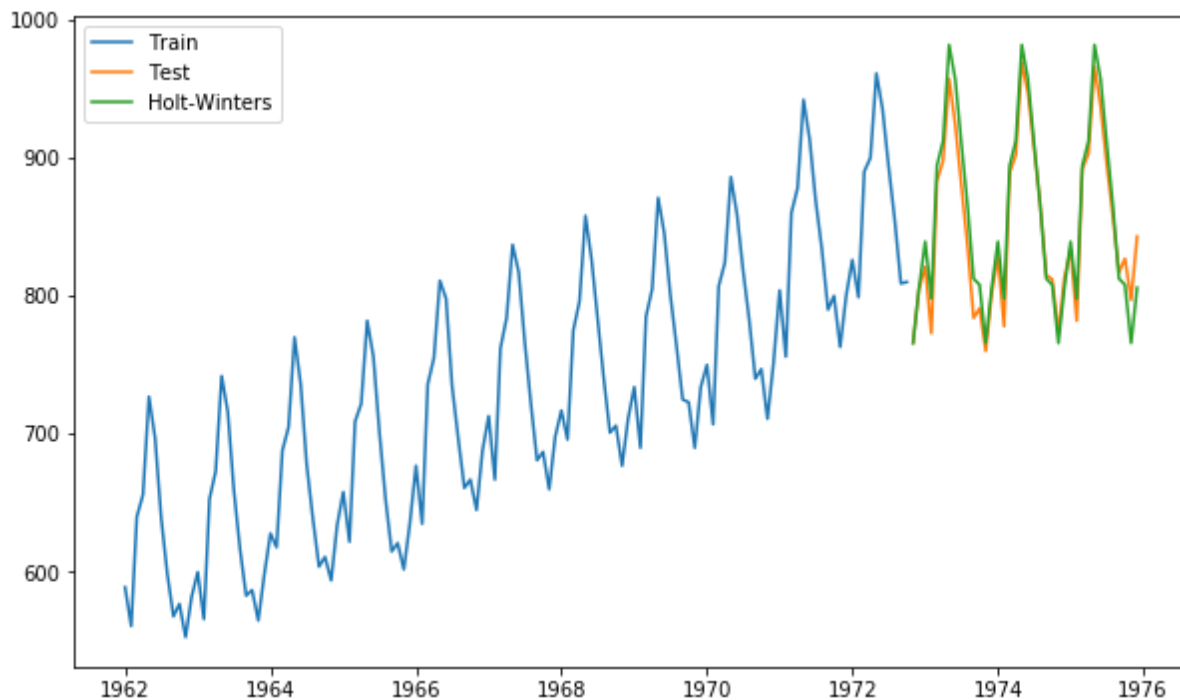
```
Out[11]: Month
1972-11-01    766
1972-12-01    805
1973-01-01    821
1973-02-01    773
1973-03-01    883
Freq: MS, Name: milk_production_pounds_per_cow, dtype: int64
```

```
In [12]: # https://www.statsmodels.org/stable/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html
model = ExponentialSmoothing(train, seasonal='mul', seasonal_periods=12).fit()
# https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html
pred = model.predict(start=test.index[0], end=test.index[-1])
```



```
In [13]: plt.figure(figsize=(10,6))
plt.plot(train.index, train, label='Train')
plt.plot(test.index, test, label='Test')
plt.plot(pred.index, pred, label='Holt-Winters')
plt.legend(loc='best')
```

Out[13]: <matplotlib.legend.Legend at 0x212dc0c46d8>



Dự đoán

```
In [14]: df.tail()
```

Out[14]:

milk_production_pounds_per_cow	
Month	
1975-08-01	858
1975-09-01	817
1975-10-01	827
1975-11-01	797
1975-12-01	843



```
In [15]: import datetime
s = datetime.datetime(1976, 1, 1)
e = datetime.datetime(1976, 6, 1)
pred_next_6_month = model.predict(start=s, end=e)
pred_next_6_month
```

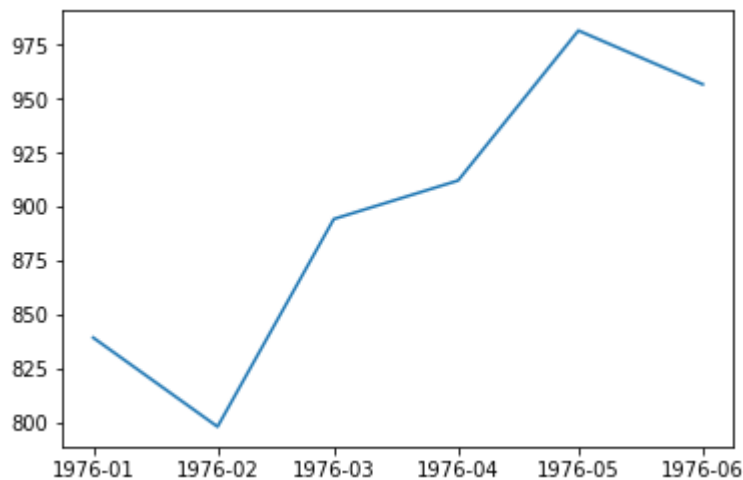
```
Out[15]: 1976-01-01    839.255997
1976-02-01    797.959544
1976-03-01    894.299974
1976-04-01    912.154098
1976-05-01    981.713246
1976-06-01    956.770929
Freq: MS, dtype: float64
```

```
In [16]: x = pd.Series(pred_next_6_month)
type(x)
```

```
Out[16]: pandas.core.series.Series
```

```
In [17]: plt.plot(x.index, x.values)
```

```
Out[17]: [<matplotlib.lines.Line2D at 0x212dc11cef0>]
```

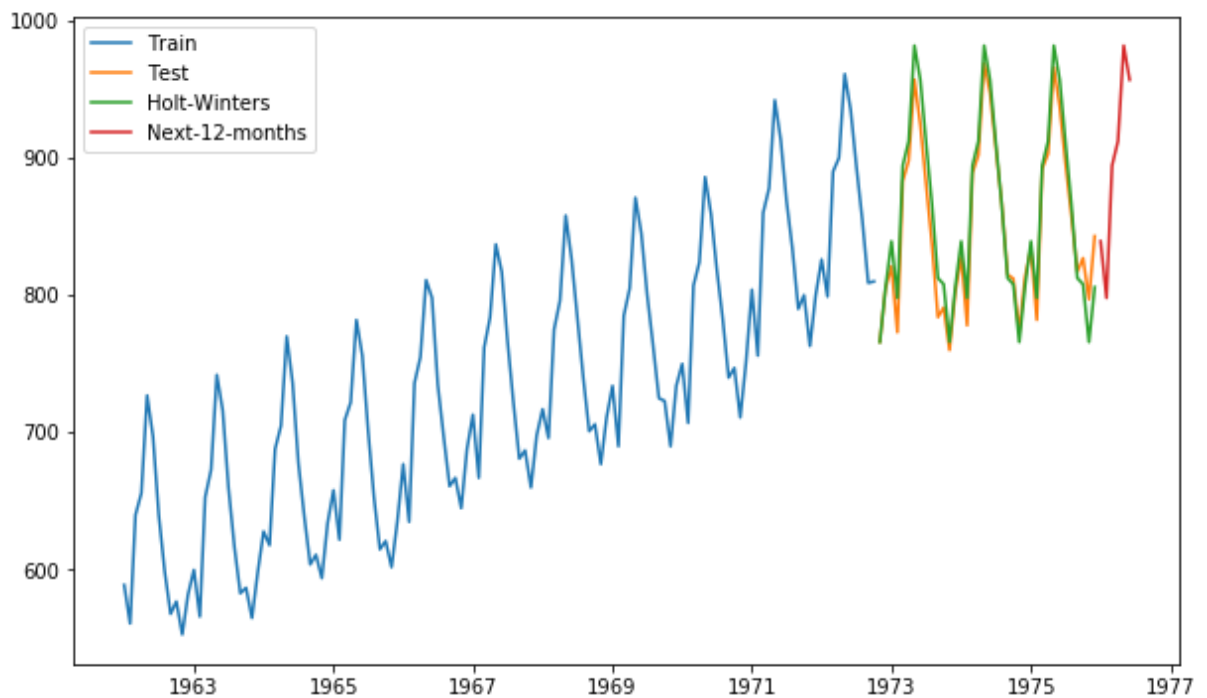


Trực quan hóa dữ liệu



```
In [18]: plt.figure(figsize=(10,6))  
plt.plot(train.index, train, label='Train')  
plt.plot(test.index, test, label='Test')  
plt.plot(pred.index, pred, label='Holt-Winters')  
plt.plot(x.index, x.values, label='Next-12-months')  
plt.legend(loc='best')
```

Out[18]: <matplotlib.legend.Legend at 0x212dc175898>



In []: