00-Time-Series-Basics

October 19, 2022

Copyright Pierian Data

For more information, visit us at www.pieriandata.com

1 Time Series Basics

According to wikipedia, a time series is "a series of data points indexed in time order." The goal of time series analysis is to predict future values based on past behavior.

This differs from other analyses that compare x to y, since x is not really a variable. For example, if I measure the boiling point of water with 1%, 3%, 5% and 7% salt solutions, I should be able to predict the boiling point at 4%. But with time series, we can't go back and "plug in" a new value of x, since we can never go back in time. For this reason, a time series with one y variable is called univariate.

By the end of this course you will understand the following code, and be able to create your own time series plots.

For Further Reading:

Forecasting: Principles and Practice by Rob J Hyndman and George Athanasopoulos NIST/SEMATECH e-Handbook of Statistical Methods Introduction to Time Series Analysis

```
[1]: # RUN THIS CELL TO DISPLAY THE PLOT
     import pandas as pd
     %matplotlib inline

     # Source: https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html
     df = pd.read_csv('../Data/co2_mm_mlo.csv', encoding='utf8')
     df.head()
```

```
[1]:    year  month  decimal_date  average  interpolated
     0  1958      3      1958.208   315.71        315.71
     1  1958      4      1958.292   317.45        317.45
     2  1958      5      1958.375   317.50        317.50
     3  1958      6      1958.458      NaN        317.10
     4  1958      7      1958.542   315.86        315.86
```

```
[2]: title = 'Monthly Mean CO Levels (ppm) over Mauna Loa, Hawaii'
     df.plot('decimal_date','interpolated',figsize=(12,6),title=title).
      ↪autoscale(axis='both',tight=True);
```

Monthly Mean CO_2 Levels (ppm) over Mauna Loa, Hawaii
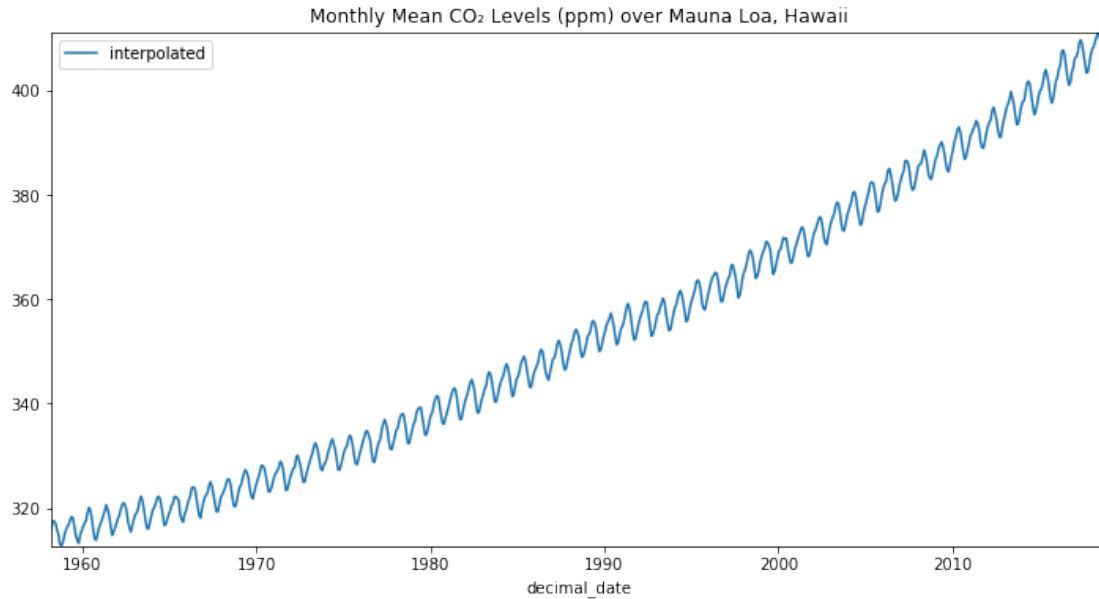
The above plot clearly illustrates some of the fundamentals of time series analysis: * **trend** - over the past 60 years there is a clear, non-linear upward trend in the levels of CO2 in the atmosphere at Mauna Loa, Hawaii * **seasonality** - within any given year there are cyclical patterns of rising and falling CO2 levels * **noise** - we also see random, non-systemic fluctuations in the data

1.1 What are "endog" and "exog"?

The data seen in a time series is described as either endogenous, that is, caused by factors within the system, or exogenous, caused by factors outside the system. We'll see that many statsmodels tools use endog to represent the incoming time series data in place of the constant y. For more information, visit http://www.statsmodels.org/stable/endog_exog.html

```
[ ]:
```