**Content**

[1. Matplotlib 2](#_Toc48128681)

[2. Concepts 3](#_Toc48128682)

[3. Statistical concepts 5](#_Toc48128683)

[1. Trend 6](#_Toc48128684)

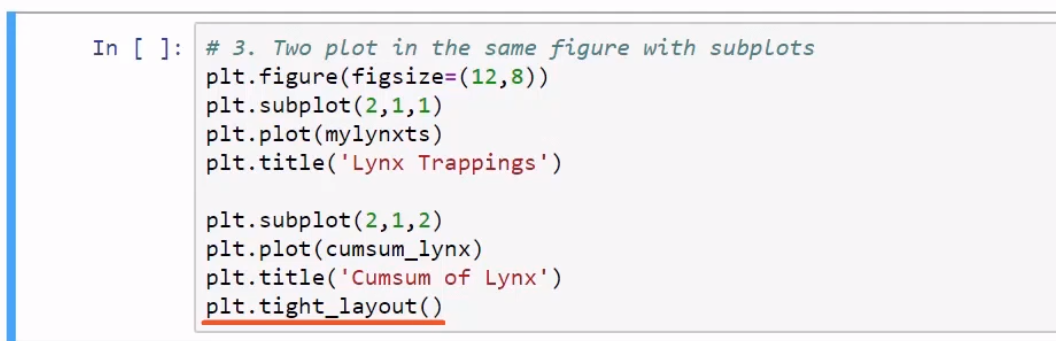
[2. Indications of non-stationarity 6](#_Toc48128685)

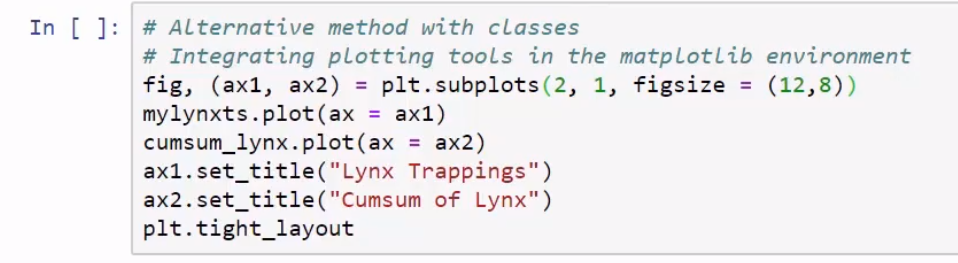
[3. ACF and PACF plots 7](#_Toc48128686)

[4. Smoothers 7](#_Toc48128687)

[4. ARIMA 9](#_Toc48128688)

1. Matplotlib





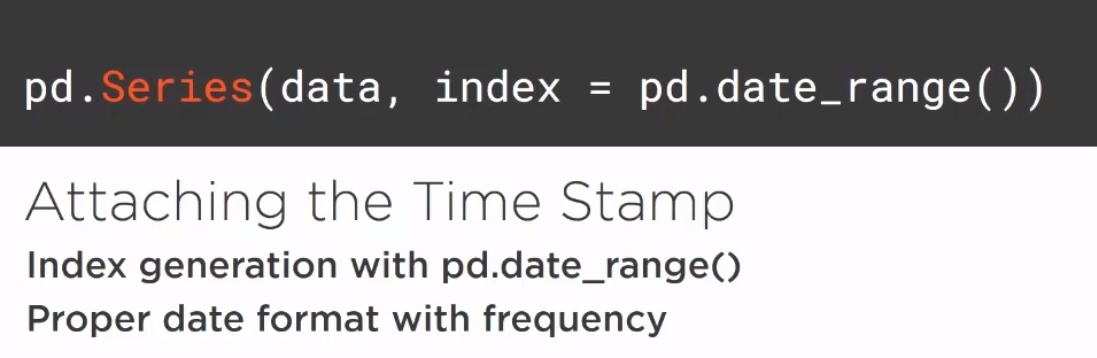
1. Concepts

* Exponential smooting – models should put more weight in most recent events.
* Python: StatsModels library, a statistical toolbox
  + TSA module for time series analysis
  + Third party modules: STLdecompose and pmdarima
* R has better documentation and more tools for time series analysis. Beginner and intermediate topics however are covered in Python.
* Bonus: Integration of R in Python is possible (rpy2 module)

A dataset is not fully seasonal if the periods between peeks are not the same.

Constant variance: It means that when you plot the individual error against the predicted value, the variance of the error predicted value should be constant.

TimeSeries !!!

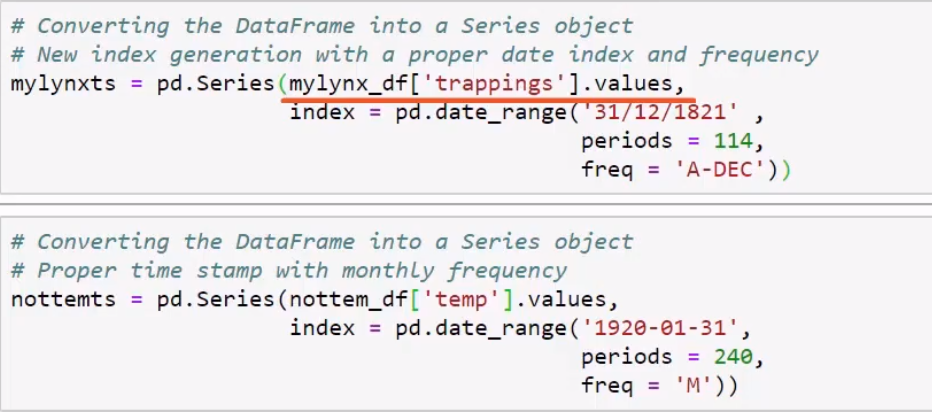


Stationarity – it tells if the data has the same statistical traits throughout the series. Augmented Dickey-Fuller test.

Autocorrelation: correlation between the observations of a time series variable. ACF and PACF plots.

Moving averages and smoothers.

To solve the problem of scaling – logarithmic scaling or same padding into the minus.



1. Statistical concepts

* Trend
* Stationary

In mathematics and statistics, a stationary process (or a strict/strictly stationary process or strong/strongly stationary process) is a stochastic process whose unconditional joint probability distribution does not change when shifted in time.

* Stochastic

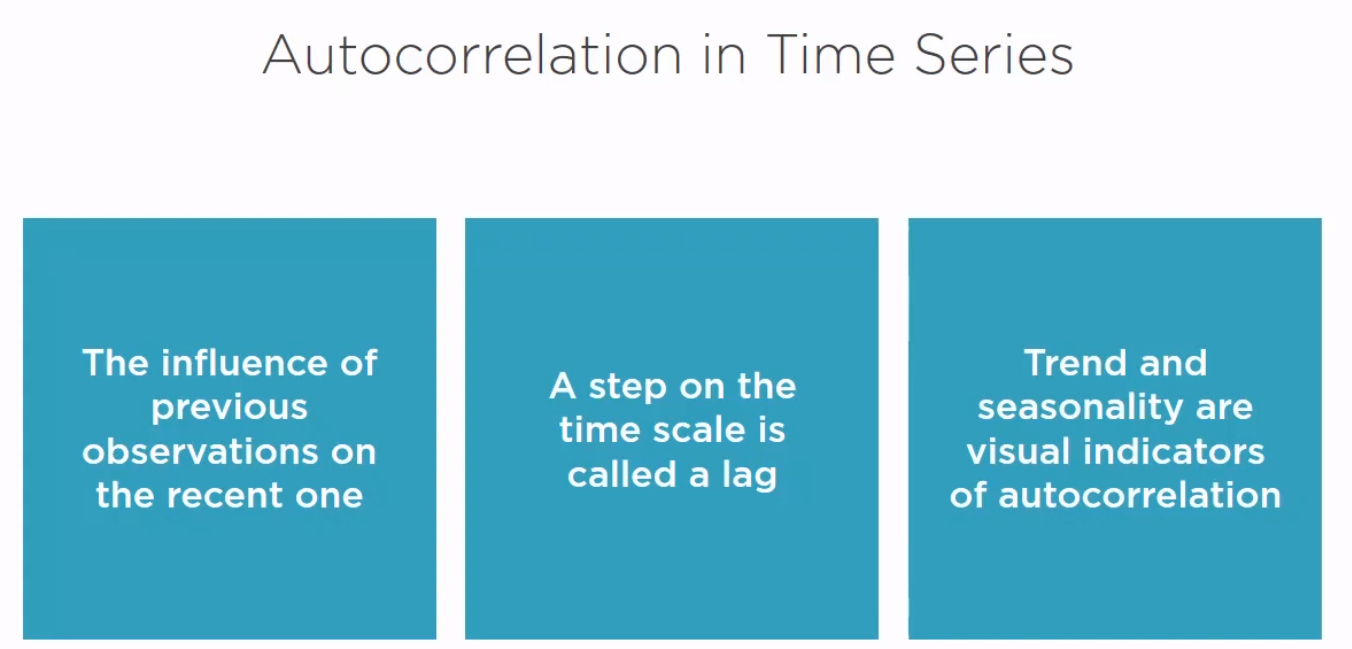
Stochastic models possess some inherent randomness. The same set of parameter values and initial conditions will lead to an ensemble of different outputs. Obviously, the natural world is buffeted by stochasticity stochasticity. But, stochastic stochastic models are considerably more complicated than the deterministic ones.

* Deterministic

In deterministic models, the output of the model is fully determined by the parameter values and the initial conditions.

* Autocorrelation

also known as serial correlation, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. Describes the correlation between the values of an ordered series at different time points.



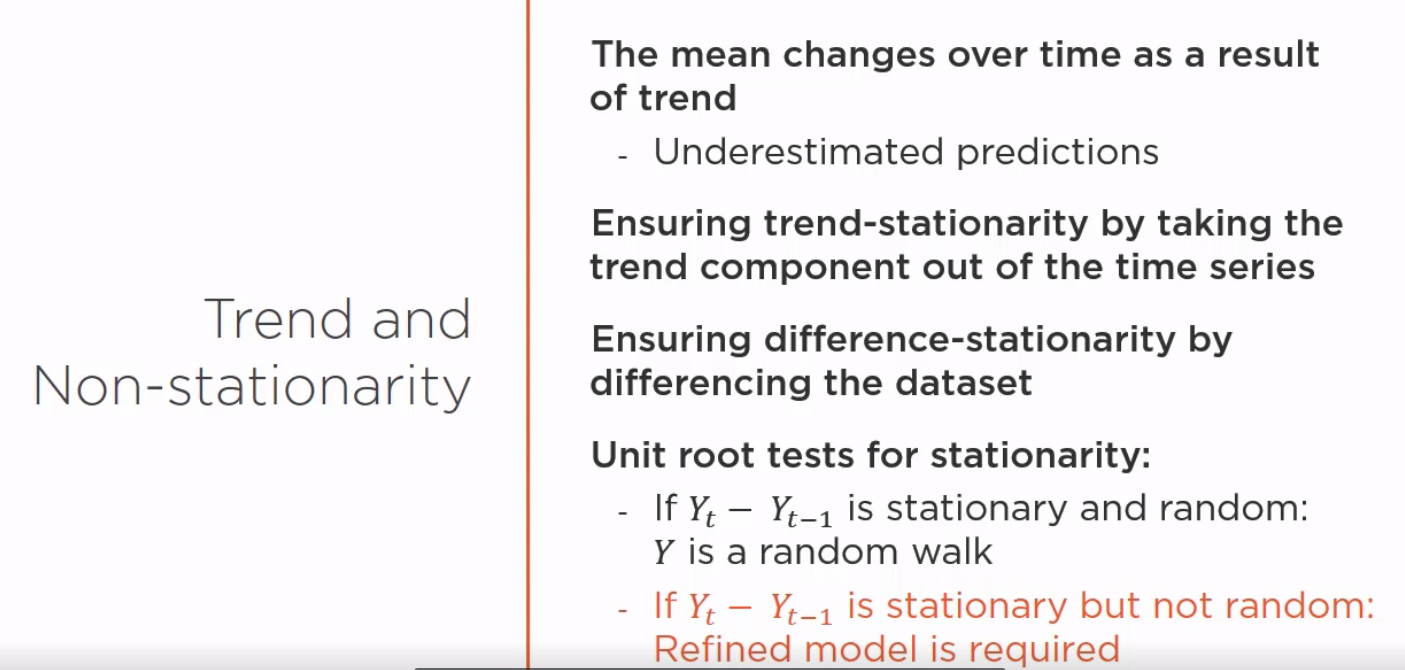
* 1. Trend
* Trending mean

A common violation of stationarity. There are two popular models for nonstationary series with a trending mean.

* Trend stationary: The mean trend is deterministic. Once the trend is estimated and removed from the data, the residual series is a stationary stochastic process.
* Difference stationary: The mean trend is stochastic. Differencing the series D times yields a stationary stochastic process.

The distinction between a deterministic and stochastic trend has important implications for the long-term behavior of a process:

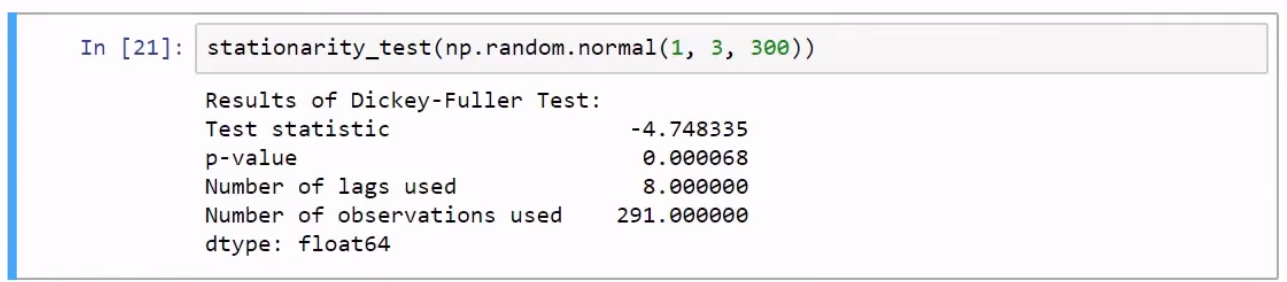
* Time series with a deterministic trend always revert to the trend in the long run (the effects of shocks are eventually eliminated). Forecast intervals have constant width.
* Time series with a stochastic trend never recover from shocks to the system (the effects of shocks are permanent). Forecast intervals grow over time.

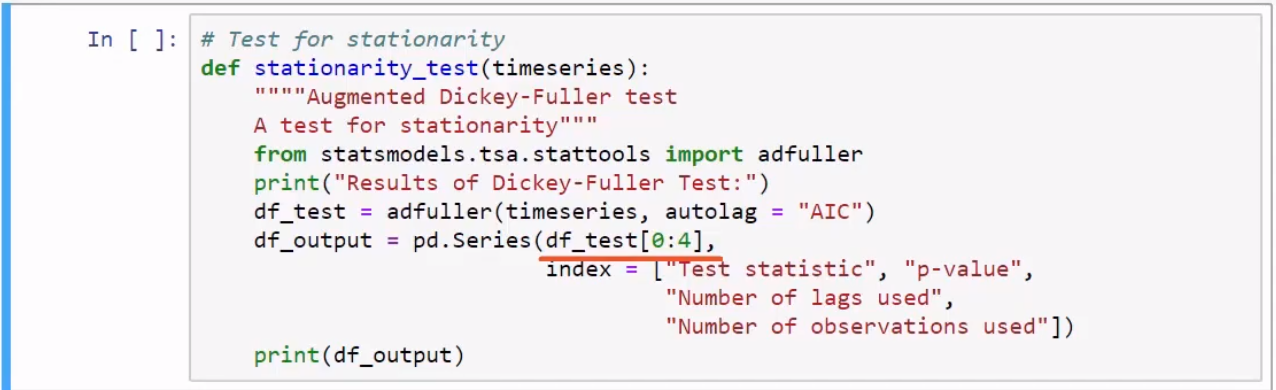


* One lag differencing
  1. Indications of non-stationarity
* The data has a clear trend and/or seasonality
* Changes in variance and/or mean
* There are statistical tests in python which tells if the data is stationary or not – Unit Root Tests
* Augmented Dickey-Fuller test

Removes the autocorrelation and tests for non-stationarity. Equal mean and variance throughout the time series. Null hypothesis: Non-stationarity. The more negative the test statistic gets, the stronger the rejection og h-null. Only if the P-value < 0.05 you can assume stationarity.

Random numbers - Test statistic





* 1. ACF and PACF plots
* ACF – Autocorrelation Function

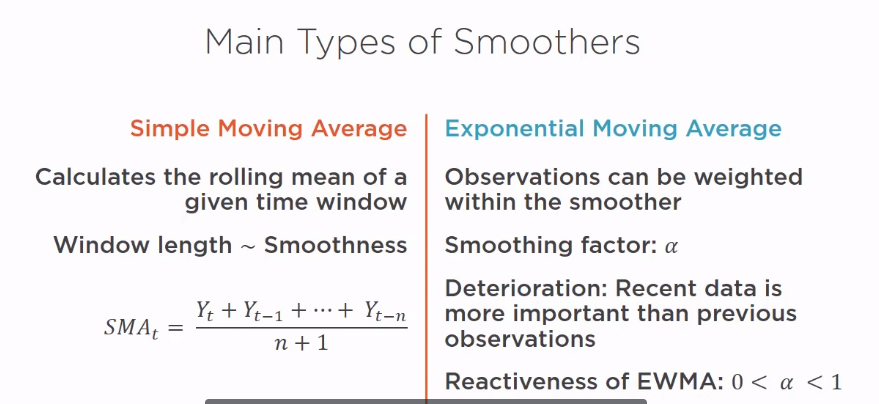
The correlation between the observation at the current time spot and the observations at previous time spots

* PACF – Partial Autocorrelation Function

The correlation between observations at two time spots given that we consider both observations are correlated to observations at other time spots.

* 1. Smoothers

A way how to deal with distractions in the pattern (outliers, extreme values). Smoothers show the data via decimating the hights and lows.



1. ARIMA

In statistics, econometrics and signal processing, an autoregressive (AR) model is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature, economics, etc. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation (or recurrence relation which should not be confused with differential equation). Together with the moving-average (MA) model, it is a special case and key component of the more general autoregressive–moving-average (ARMA) and autoregressive integrated moving average (ARIMA) models of time series, which have a more complicated stochastic structure; it is also a special case of the vector autoregressive model (VAR), which consists of a system of more than one interlocking stochastic difference equation in more than one evolving random variable.

Contrary to the moving-average (MA) model, the autoregressive model is not always stationary as it may contain a unit root.

