Time Series Data

Time series data is a collection of observations obtained through repeated measurements over time. Plot the points on a graph, and one the axes would always be time.

Time series lends itself naturally to visualization. Line plots of observations over time are popular, but there is a suite of other plots that you can use to learn more about your problem. The more you learn about your data, the more likely you are to develop a better forecasting model. There are six different types of plots that can be used to visualize time series data with Python. They are:

* Line Plots
* Histograms and Density Plots
* Box and Whisker Plots
* Heat Maps
* Lag Plots or Scatter Plots
* Autocorrection Plots

A stationary time series is one whose properties do not depend on the time at which the series is observed. Thus, time series with trends or seasonal effects are not stationary, since the trend and seasonality will affect the value of the time series at different times. The statistical properties (mean, variance) do not change over time for a stationary time series.  
To test for stationarity, there are various unit-root tests: ADF (Augmented Dickey-Fuller Test) test, KPSS test, Philip-Perron test.

Multivariate Linear Regression

Similar to simple linear regression, however there are multiple independent variables contributing to the dependent variable, hence multiple coefficients to determine, and complex computation due to the added variables.

The basis of Vector Autoregression is that each of the time series in the system influences each other. This means you can predict the series with past values of itself along with other series in the system. Using Granger’s Causality Test, it is possible to test this relationship before even building the model. Granger’s causality tests the null hypothesis, that the coefficients of past values in the regression equation is zero. In simpler terms, the past values of time series (X) do not cause the other series (Y). So, if the p-value obtained from the test is lesser than the significance level of 0.05, then, you can safely reject the null hypothesis.

Cointegration occurs when two or more nonstationary time series: have a long-run equilibrium, move together in such a way that their linear combination results in a stationary time series, they share an underlying common stochastic trend.