DATA ANALYTICS- UNIT 2

By

DEEPIKA KAMBOJ

Analysis of time series

- Time Series Data Analysis is a way of studying the characteristics of the response variable with respect to time as the independent variable.
- A Time-Series represents a series of time-based orders. It would be Years, Months, Weeks, Days, Horus, Minutes, and Seconds.
- It is an observation from the sequence of discrete time of successive intervals.
- Time Series Analysis (TSA) is used in different fields for time-based predictions – like Weather Forecasting models, Stock market predictions, Signal processing, Engineering domain – Control Systems, and Communications Systems.
- We could predict the future using AR, MA, ARMA, and ARIMA models.

Analysis of time series

“A time series is nothing but a sequence of various data points that occurred in a successive order for a given period of time.”

How to Analyze Time Series?

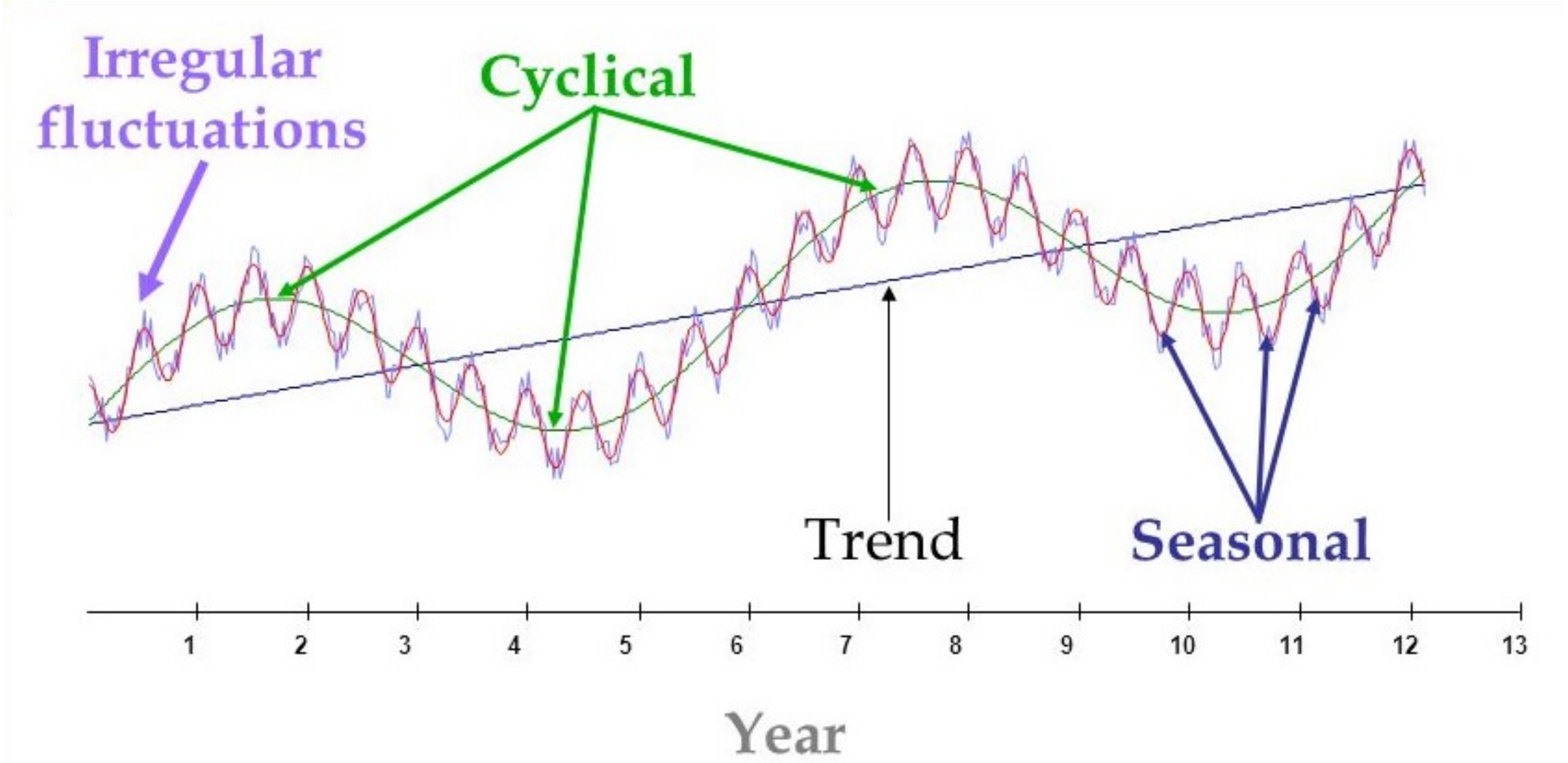
To perform the time series analysis, we have to follow the following steps:

- Collecting the data and cleaning it
- Preparing Visualization with respect to time vs key feature
- Observing the stationarity of the series
- Developing charts to understand its nature.
- Model building – AR, MA, ARMA and ARIMA
- Extracting insights from prediction

Components of Time Series Analysis

- **Trend**: In which there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend
- **Seasonality**: In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth
- **Cyclical**: In which there is no fixed interval, uncertainty in movement and its pattern
- **Irregularity**: Unexpected situations/events/scenarios and spikes in a short time span.

Components of Time Series Analysis



Data Types of Time Series

There are two major types –

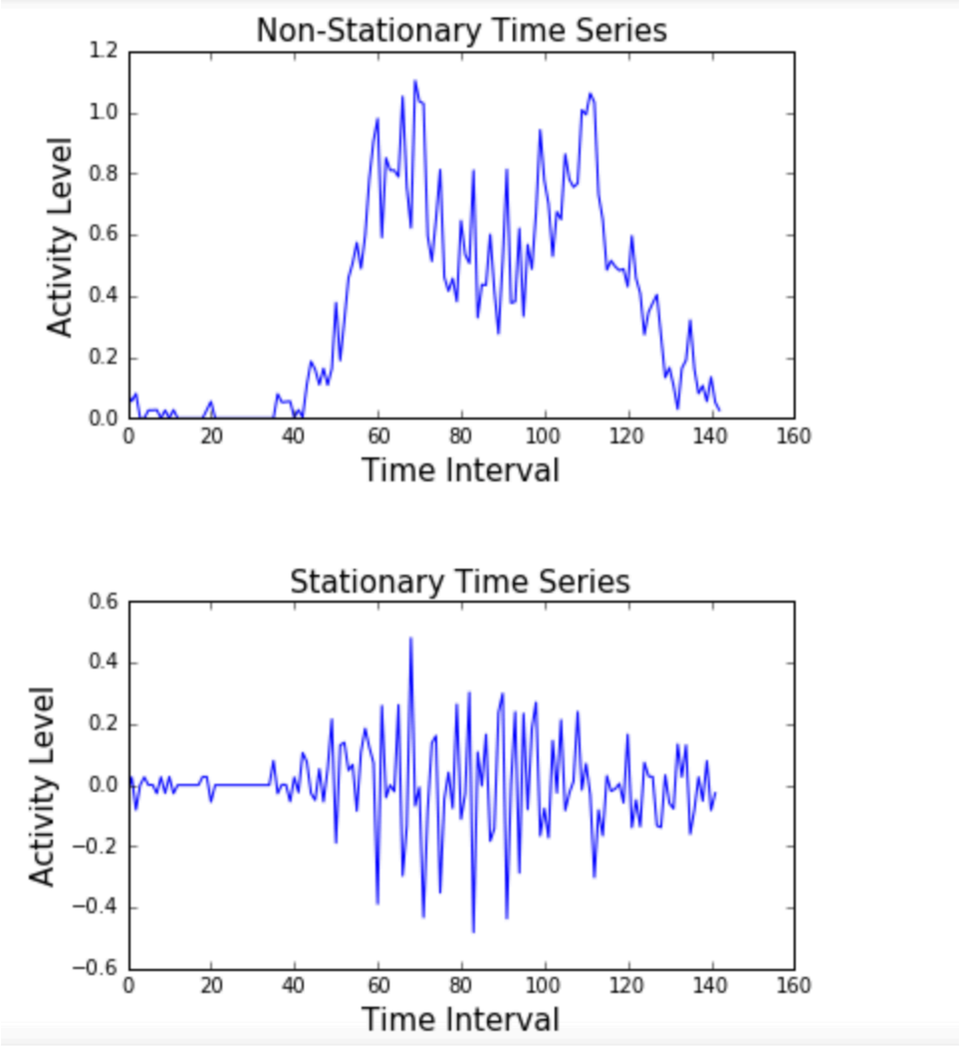
- Stationary
- Non-stationary

Stationary & Non- Stationary

Stationary: A time series is said to be stationary if its statistical properties, such as the mean and variance, remain constant over time.

Non- Stationary: If either the mean-variance or covariance is changing with respect to time, the dataset is called non-stationary.

Data Types of Time Series



Methods to Check Stationarity

During the TSA model preparation workflow, we must assess whether the dataset is stationary or not. This is done using **Statistical Tests.** There are two tests available to test if the dataset is stationary:

- Augmented Dickey-Fuller (ADF) Test
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Null and Alternate Hypothesis

- The null hypothesis is a statement that assumes there is no significant relationship between variables or no difference in the values being tested.
- The alternative hypothesis, on the other hand, is a statement that assumes there is a significant relationship between variables or a difference in the values being tested.

Augmented Dickey-Fuller (ADF) Test

The ADF test is the most popular statistical test. It is done with the following assumptions:

- Null Hypothesis (H0): Series is non-stationary
- Alternate Hypothesis (H1): Series is stationary
    - p-value >0.05 Fail to reject (H0)
    - p-value <= 0.05 Accept (H1)

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

- **Null hypothesis** (H0): The data is stationary.
- **Alternate hypothesis** (H1): The data is not stationary.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

- The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a statistical test used to determine whether a time series is stationary or non-stationary. Unlike the Augmented Dickey-Fuller (ADF) test, which tests for the presence of a unit root (i.e., non-stationarity), the KPSS test tests for the absence of a unit root (i.e., stationarity).

- The KPSS test involves computing a test statistic and a p-value based on the data. The null hypothesis of the KPSS test is that the time series is stationary, while the alternate hypothesis is that the time series is non-stationary. The test compares the test statistic to critical values based on the significance level and the number of observations in the time series.

Why we check for stationarity

1. Ease of modelling

2. Accurate predictions

Analysis of time series: linear systems analysis & nonlinear dynamics

Linear Systems Analysis

1. Deals with the study of linear relationships between variables in a time series and the behavior of the system over time.

2. Based on the assumption that changes in one variable can be expressed as a linear combination of previous values of the same variable and/or other related variables.

Linear Time Series Analysis Methods

- Autoregression (AR)
- Moving Average (MA)
- Autoregressive Integrated Moving Average (ARIMA)
- Vector Autoregression (VAR)

Analysis of time series: linear systems analysis & nonlinear dynamics

Nonlinear Dynamics:

1. Focuses on the study of nonlinear relationships between variables and the complex behavior that arises from these relationships.
2. Used to study systems that exhibit chaotic behavior, where small changes in initial conditions can result in large changes in the outcome over time.

Nonlinear Time Series Analysis Methods:

- Chaos Theory
- Fractal Analysis
- Neural Networks
- Support Vector Machines (SVM)
- Genetic Algorithms
- Hidden Markov Models (HMM)

Analysis of time series: Moving Averages

Simple Moving Average

SMA = (P1 + P2 + ... + Pn) / n

Weighted Moving Average (WMA):

WMA = (w1 * P1 + w2 * P2 + ... + wn * Pn) / (w1 + w2 + ... + wn)

Exponential Moving Average (EMA):

EMA = (P - EMA_previous) * SF + EMA_previous

Analysis of time series: Moving Averages

Moving Average Type	Formula	Calculation Method	Weighting	Sensitivity to Price Changes
Simple Moving Average (SMA)	(P1 + P2 + ... + Pn) / n	Takes the sum of a specified number of prices and divides by the number of prices	Equal weight to each price	Less sensitive to price changes due to equal weighting
Weighted Moving Average (WMA)	(w1 * P1 + w2 * P2 + ... + wn * Pn) / (w1 + w2 + ... + wn)	Gives different weights to different prices using a weighting function	Can use linear or exponential weighting functions	More sensitive to recent price changes due to higher weights given to recent prices
Exponential Moving Average (EMA)	(P - EMA_previous) * SF + EMA_previous	Gives more weight to recent prices using a smoothing factor	Uses an exponential weighting function	Highly sensitive to recent price changes due to the exponential weighting