

# Masterclass

Time series forecasting



- Introduction to Time series forecasting
- Key concepts of time series forecasting
  - Stationarity, decomposing series
- Build Time series forecasting models using ARIMA, SARIMA, ARIMAX and SARIMAX



Introduction to time series forecasting

#### Introduction to Time series

#### Time series is anything observed sequentially over time

- Most commonly, time series data is observed at equally spaced successive intervals of time
- The measurements taken during an event in a time series are arranged in a proper chronological order

#### Time series use cases

- Predicting daily stock prices
- Sales forecasting or demand forecasting of products for inventory planning
- Predicting energy demand for future
- Predicting footfall (visitors) for retail stores at different times of day
- Traffic forecasting transport and route optimization, road facility design
- Weather Forecasting across seasons

#### Introduction to Time series

How is time series forecasting different from machine learning predictions?

#### Time is an important factor

A machine learning dataset is a collection of observations. Time does not play the primary role in machine learning datasets. Time Series Models try to understand changes/patterns in data (variable of interest) in response to time

#### Feature creation

Most of the time series forecasting techniques are univariate, i.e., there is only one feature. So, feature creation does not play a major role. However, certain specific techniques like ARIMAX do use additional features.



Quiz time!

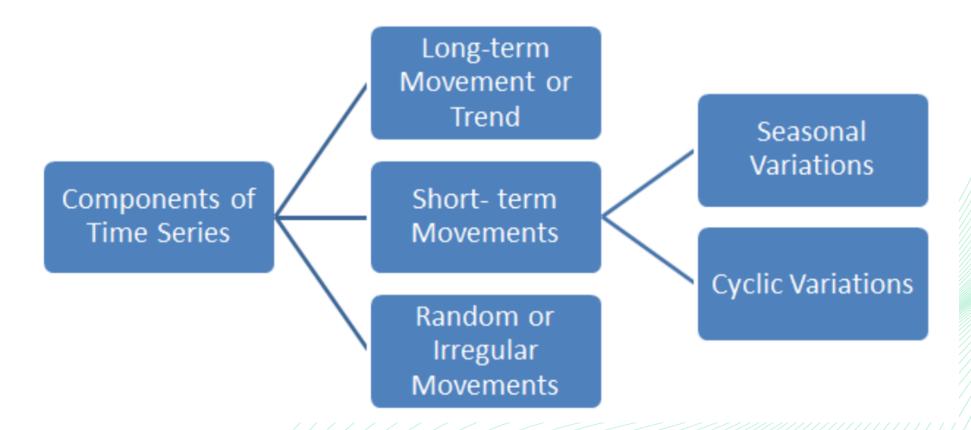
# Quiz time!

### Predict sales of Oct 2021 based on previous records

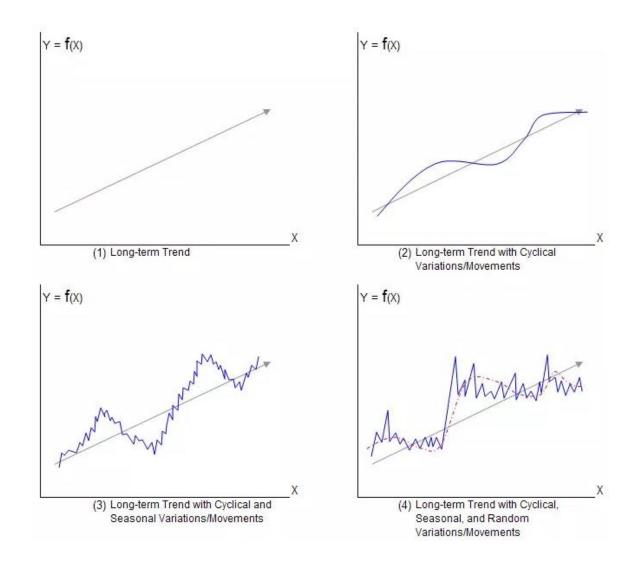
Q1	Time period	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21
QΙ	Sales(\$M)	1	1	1	1	1	1	1	1	1	?
0.0	Time period	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21
Q2	Sales(\$M)	1	2	3	4	5	6	7	8	9	?
Q3	Time period	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21
QJ	Sales(\$M)	1	5	1	5	1	5	1	5	1	?
Q4	Time period	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21
QТ	Sales(\$M)	1	1	3	1	1	1	3	1	1	?
	Holidays	0	0	1	0	0	0	1	0	0	1

# Components of Time series model

The various reasons or the forces which affect the values of an observation in a time series are the components of a time series. The four categories of the components of time series are

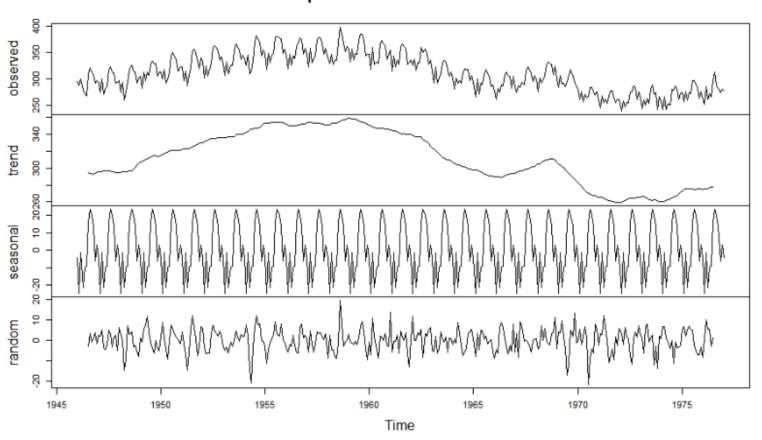


# Components of Time series model



# Decomposition of a time series

#### Decomposition of additive time series



## A lot of forecasting techniques to choose from

#### **Time Series Forecasting Techniques**

Smoothing Based Techniques

- Simple Moving Average
- Simple Exponential Smoothing
- Holt's Linear Trend
- Holt Winter's Exponential Smoothing

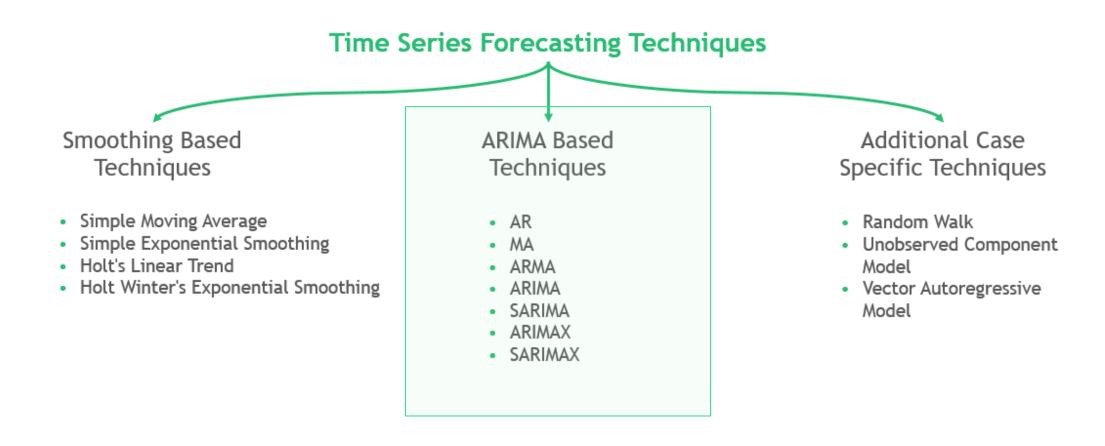
ARIMA Based Techniques

- AR
- MA
- ARMA
- ARIMA
- SARIMA
- ARIMAX
- SARIMAX

Additional Case Specific Techniques

- Random Walk
- Unobserved Component Model
- Vector Autoregressive Model

### A lot of forecasting techniques to choose from



#### Introduction to ARIMA Models

An ARIMA model is a class of statistical models for analyzing and forecasting time series data

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

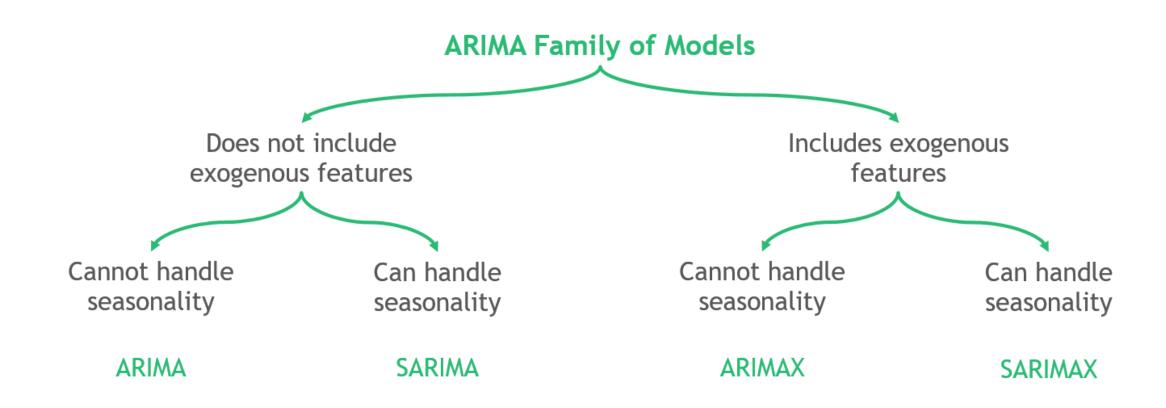
• AR: Autoregression

• I: Integrated

• MA: Moving Average

A key requirement for ARIMA models is stationarity of time series\*

# ARIMA Family of models



### Forecasting quality scoring metrics

- R squared: coefficient of determination. This can be interpreted as variance explained by the model. Ranges from (-inf, 1]
- Mean Absolute Error (MAE): It is mean of mod(actual predicted value). This metric is easy to interpret as it has the same unit of measurement as the initial series. It can range from [0, +inf)
- Mean Squared Error(MSE): This is mean of (actual-predicted)^2. The squared error gives higher penalty to large deviations. This ranges from [0,+inf)
- Mean Squared Logarithmic Error: This is practically same as MSE. However, we compute logarithm of the series while calculating this. It is used when data has exponential trends. It ranges from, [0, +inf)
- Root mean squared error (RMSE): This is standard deviation of the residuals. Range: [0, +inf)
- Mean Absolute Percentage Error(MAPE): This is MAE as a percentage of the actual. Range: [0,+inf)

