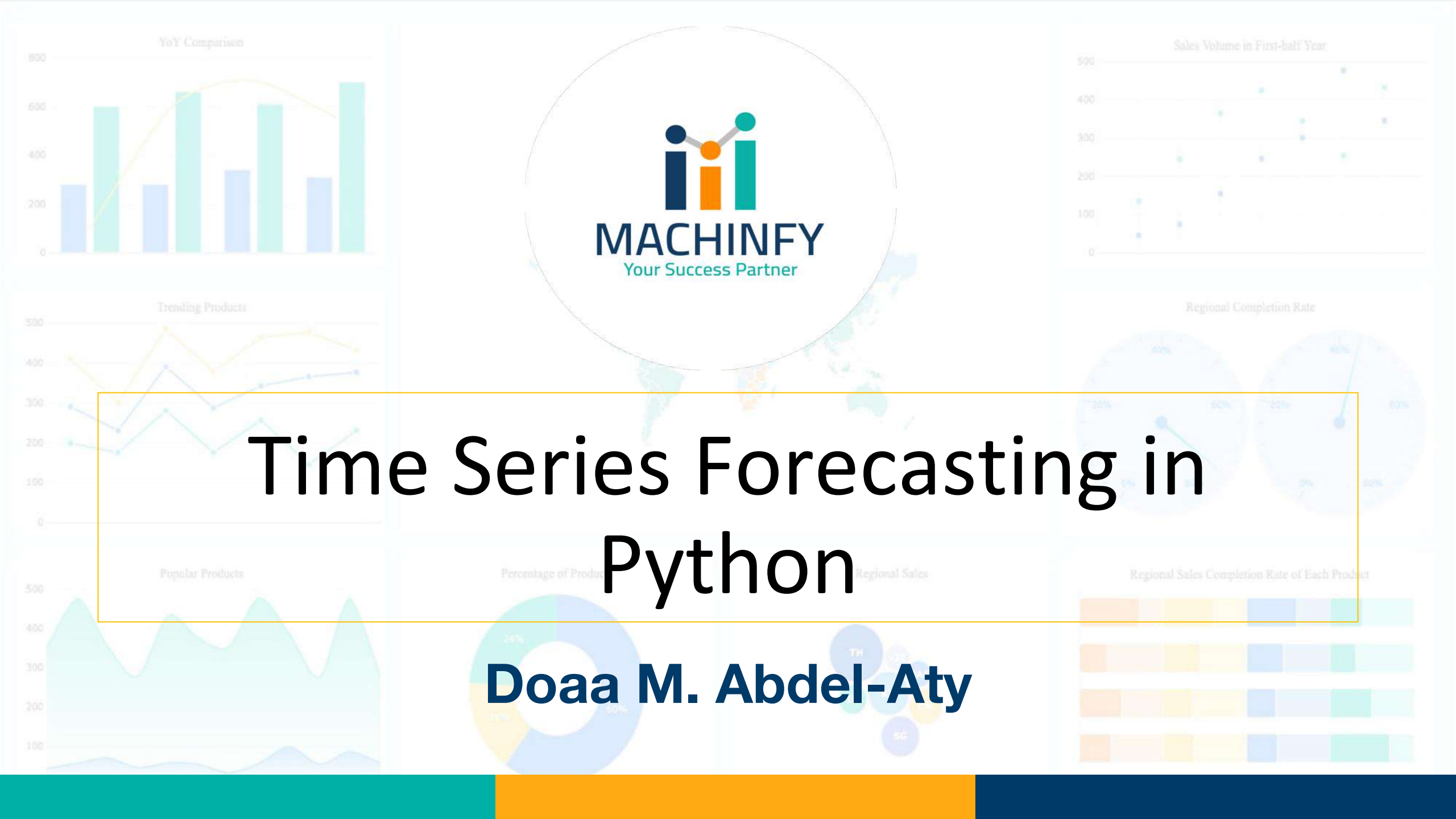


Time Series Forecasting in Python

Doaa M. Abdel-Aty

Doaa M. Abdel-Aty

- Data science researcher

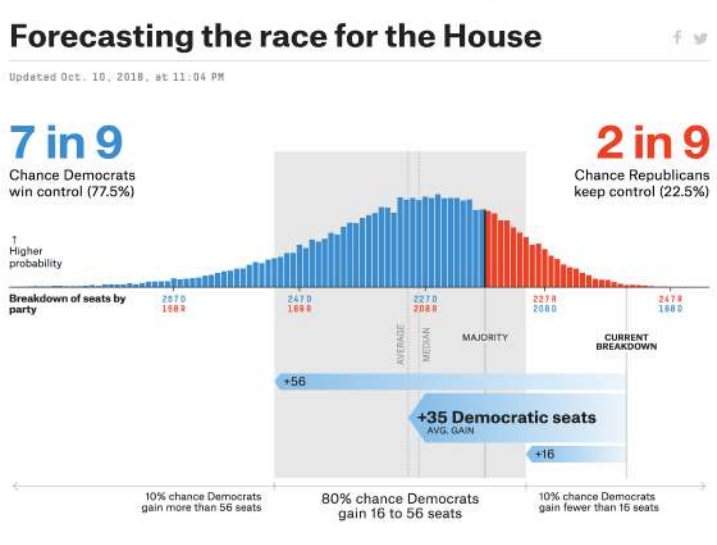
- Master of Data Science - Faculty of Statistical Studies and Research (FGSSR) - Cairo University
  (" Big Data analysis using Statistical intelligent Techniques")
  And precisely in " Energy Time series " (2020-2021).

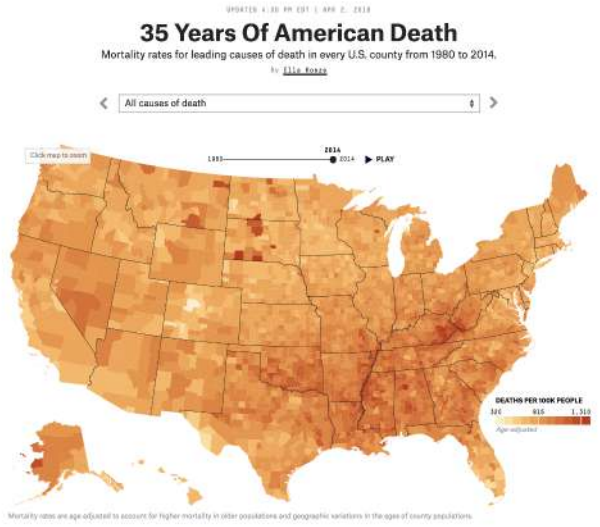
Some notes before we begin

- Please let me know if you want to go more or less in depth into a particular subject.
- Please feel free to stop me and ask any clarifying questions.

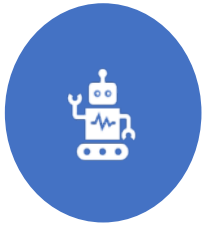
Why is forecasting important (or at least, interesting)?

Applications in many fields, including politics, finance, health, etc.





Time series Analysis for



Who is interested in learning to implement machine learning algorithms for time series data using Python.

Who needs to Enhance and learn practical data analytics

Who looking to work with real life time-series string data in Python.

people working in various fields like (and not limited to): academia, marketing, business, econometrics, finance, medicine, engineering and science

Time series models in several applications

Business	• Supply chain, booking, web traffic,..
Financial	• Stock option, Exchange, Econometrics, bitcoins …
Science	• Astronomy,Epidemiology,Weather, Earthquake prediction, Physics, …
Health	• Diagnosis, biomedical monitoring, …
Engineering	• Sensors & control Signal processing, …
Social sciences	• Population, birth rates, migration data, and political indicators.
Economics	• GDP, CPI, unemployment rates,…

Applications

Time Series

Web traffic forecasting

Control Systems

Malfunction Detection

Security

Threat Detection

Finance

Forecasting

Operations

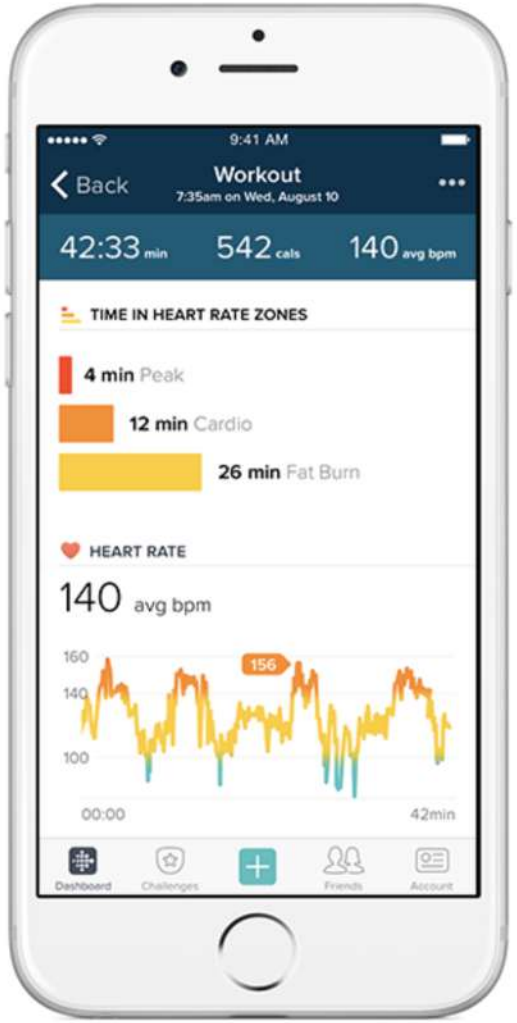
Availability
Performance

Media

Fake News

Retail sales forecasting

What is time-series data? Do I need time-series data to answer this question?

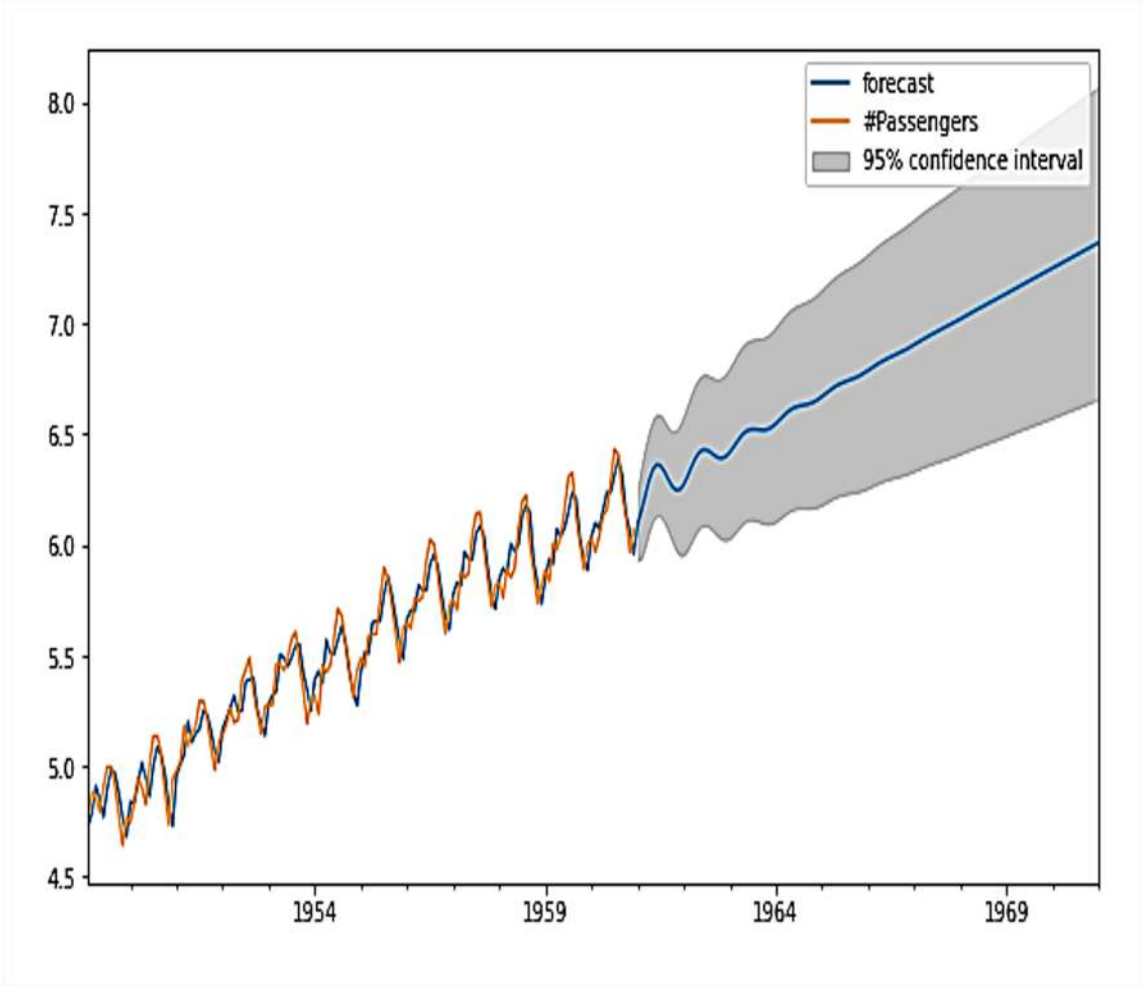


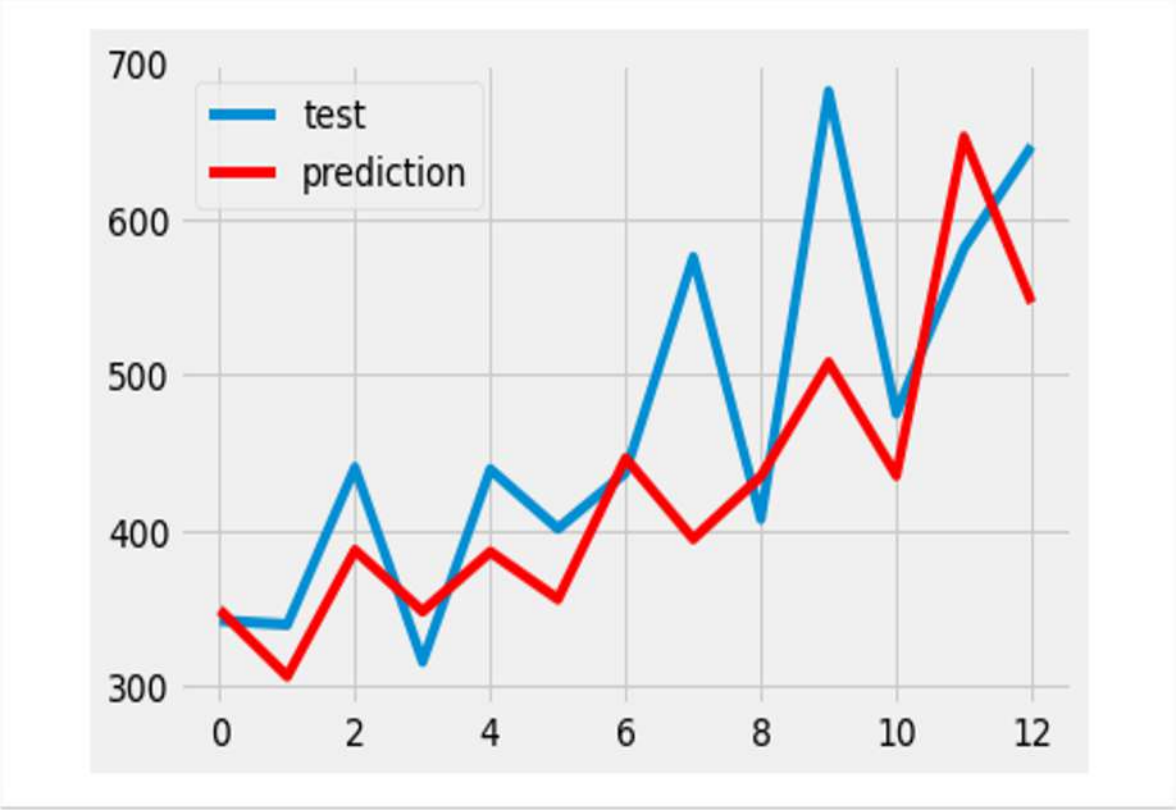
- Time-series data can be defined in many ways. I tend to describe it as “data collected on the same metric or same object at regular or irregular time intervals.”
- In terms of whether one needs time-series data depends on the question.
- I think about whether there is an inherent relationship or structure between data at various time points (e.g., is there a time-dependence), and whether we can leverage that time-ordered information.

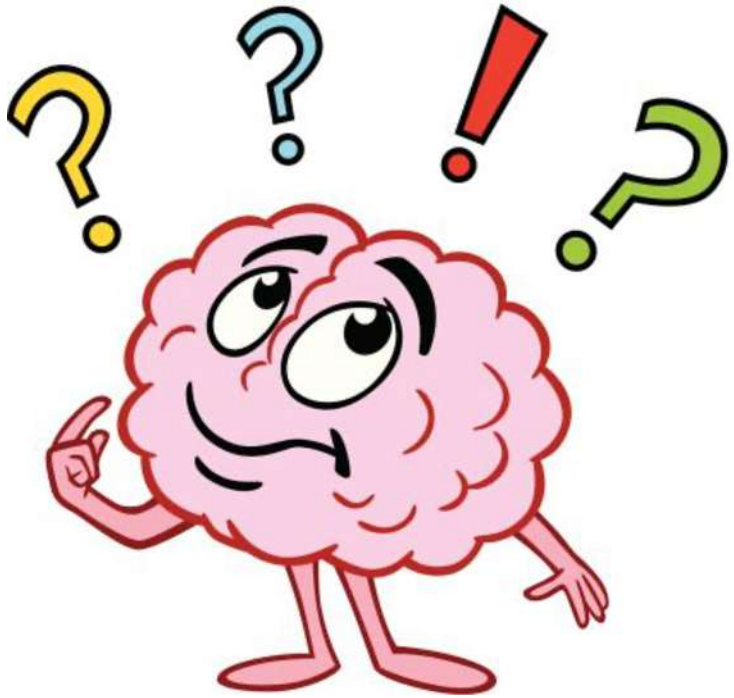
What are some things you may learn in this session?

- Why time-series data is different.
- How to process time-series data.
- How to better understand time-series data.
- How to apply statistical and machine learning methods to time-series problems.
- Understand some strengths and weaknesses of these models.
- How to evaluate, interpret, and convey output from forecasting models.

Session Objective

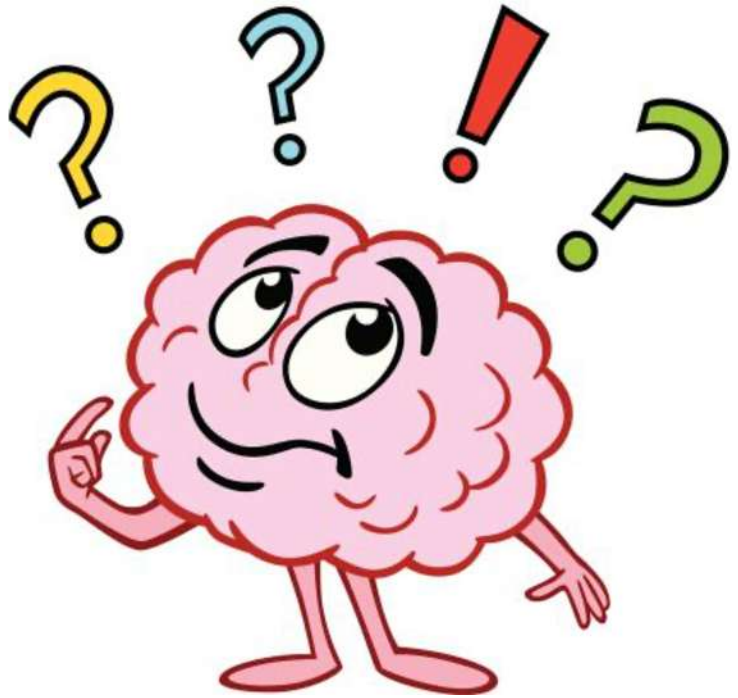






What question do I want to answer?

I think that it is important to start with a very concrete question that you think can be answered with data, specifically with time-series data.



- Some examples of questions that I might answer with forecasting:
  - What is the future expected price of Apple stock over the next year?
  - The number of cases of Covid 19 after two months?
  - What is the expected life expectancy of the average US female in 50 years?

What Is a time series ?

It is mathematically defined as a set of vectors

X(t),t = 0,1,2,...

where **t** represents the time points. The variable)
X (t is treated as a random variable (observation)

Why are the difference between time series and non-time series data?

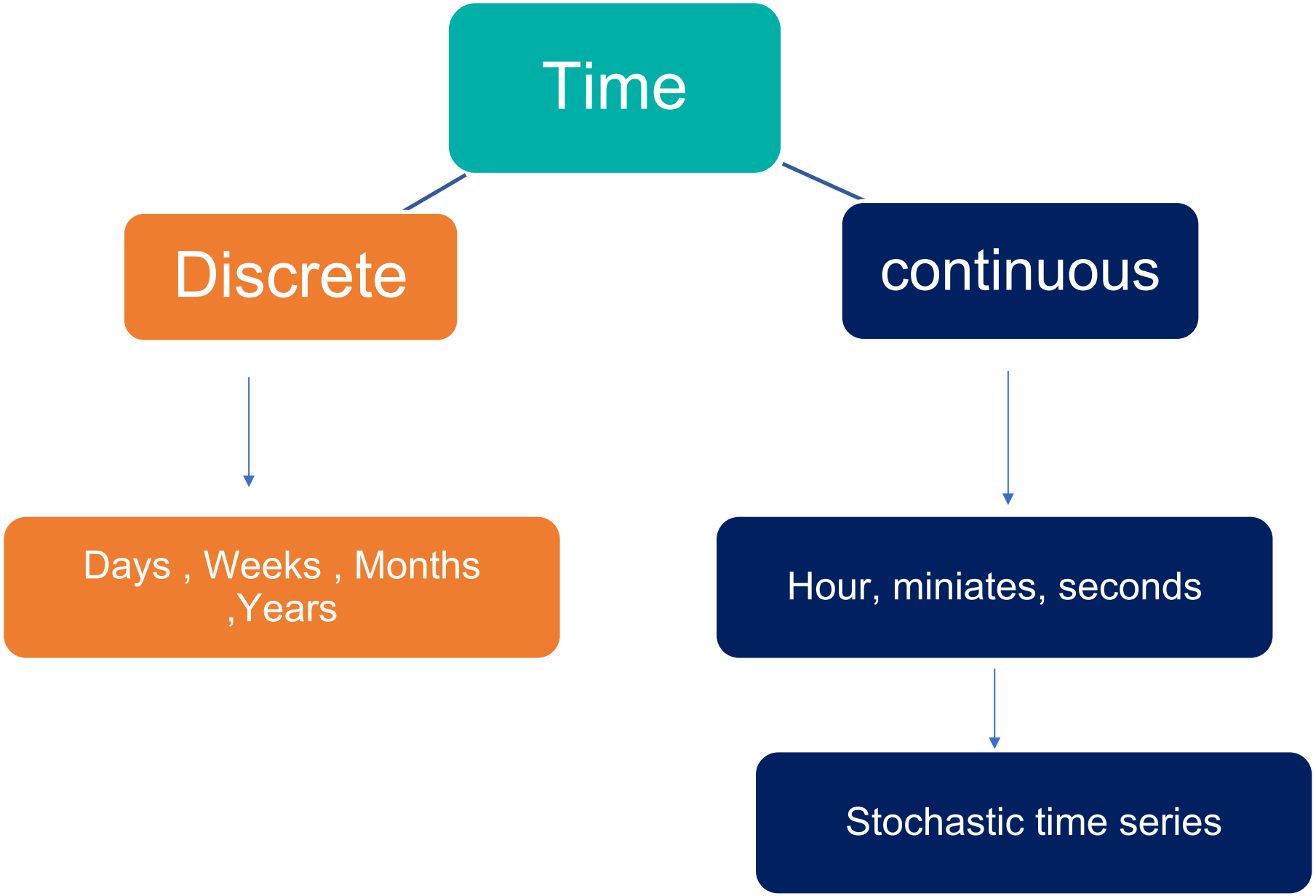
Importance of time series analysis:

- **Understand the past.**
  What happened over the last years, months?

- **Forecast the future.**

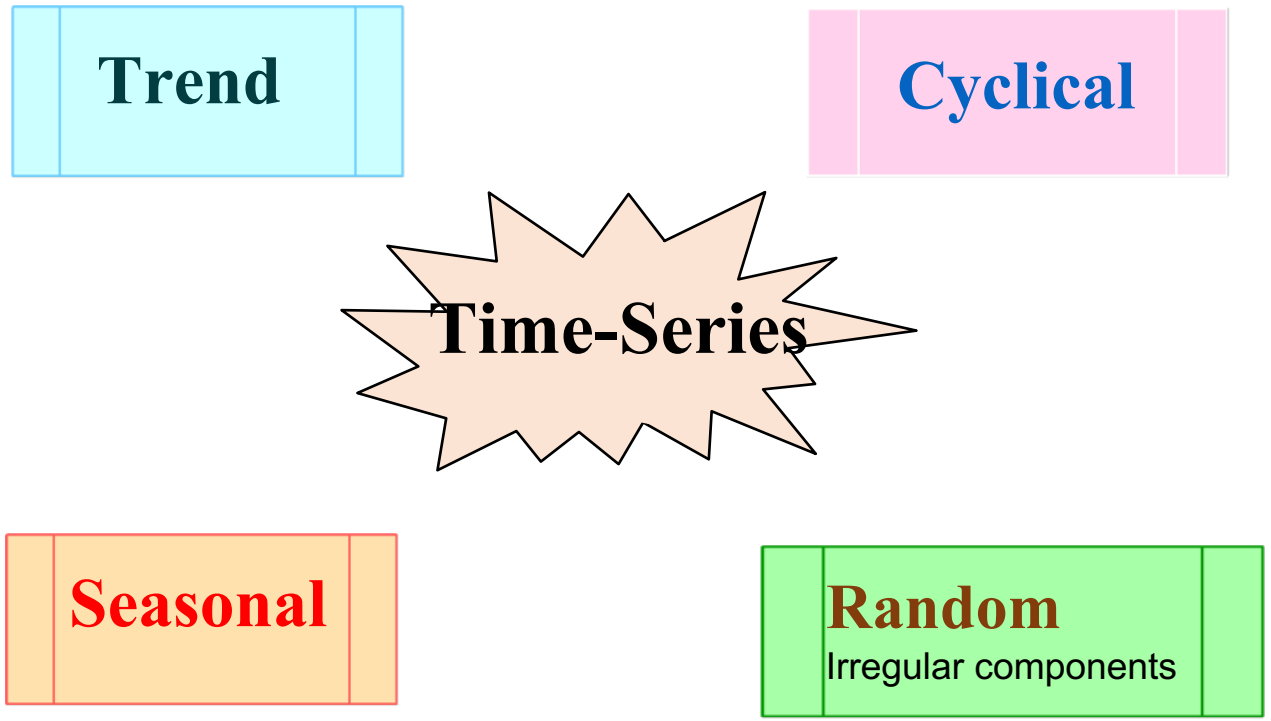
  Government wants to know future of unemployment rate, percentage increase in cost of living etc.

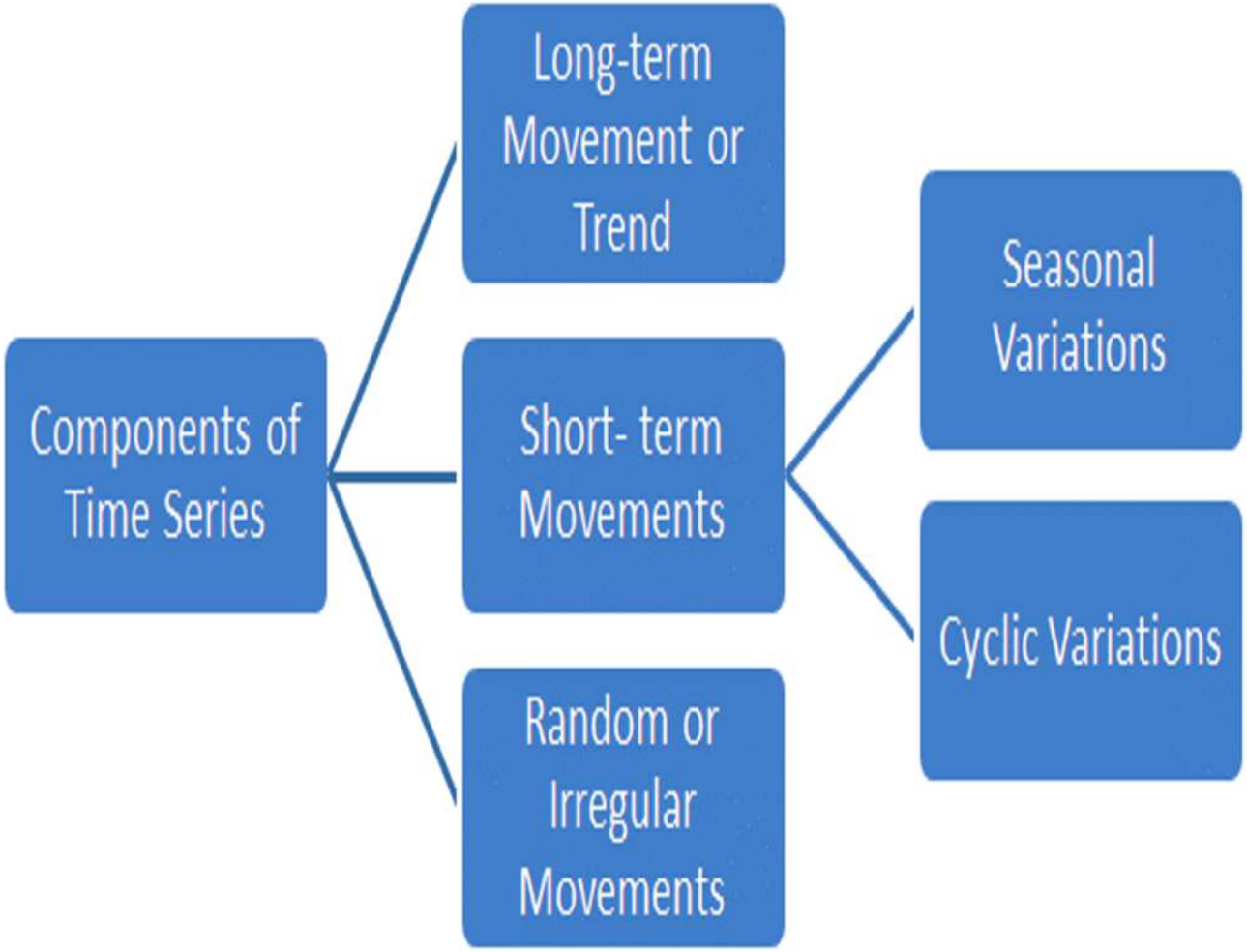
  For companies to predict the demand for their product etc.

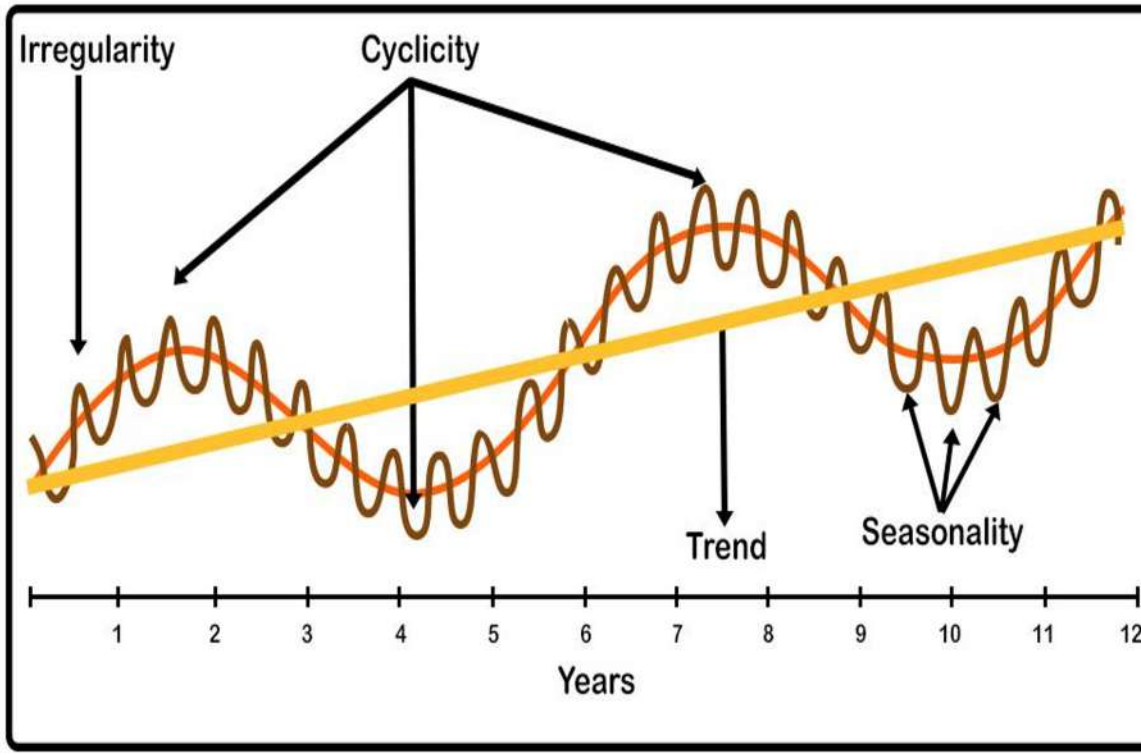

Time
Discrete
continuous
Days , Weeks , Months ,Years
Hour, miniates, seconds
Stochastic time series

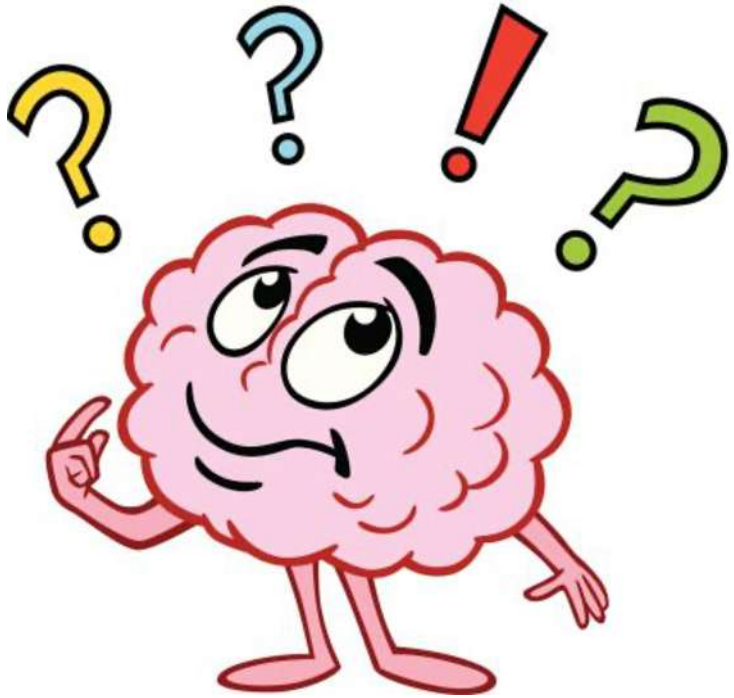
Components of a Time Series :

A time series in general is supposed to be affected by four main components, which can be separated from the observed data. These components are:




Components of Time Series
Long-term Movement or Trend
Short- term Movements
Random or Irregular Movements
Seasonal Variations
Cyclic Variations


Irregularity
Cyclicity
Trend
Seasonality
1
2
3
4
5
6
7
8
9
10
11
12
Years



Combination of Four Components

Xt = Original data at time t
Tt = Trend value at time t
St = Seasonal fluctuation at time t,
Ct = Cyclical fluctuation at time t,
It = Irregular variation at time t.

Considering the effects of these four components, two different types of models are generally used for a time series.

- **Additive Model**

$$Y(t) = T(t) + S(t) + C(t) + I(t)$$

Assumption: These four components are independent of each other.

- **Multiplicative Model**

$$Y(t) = T(t) \times S(t) \times C(t) \times I(t)$$

Assumption: These four components of a time series are not necessarily independent and they can affect one another.

The main objectives of time series analysis

The main objectives of time series analysis are:

Describe

- building of input-output models that represent the equivalent transfer functions of processes behind the time series

Control

- systems design, based on the result of analysis.

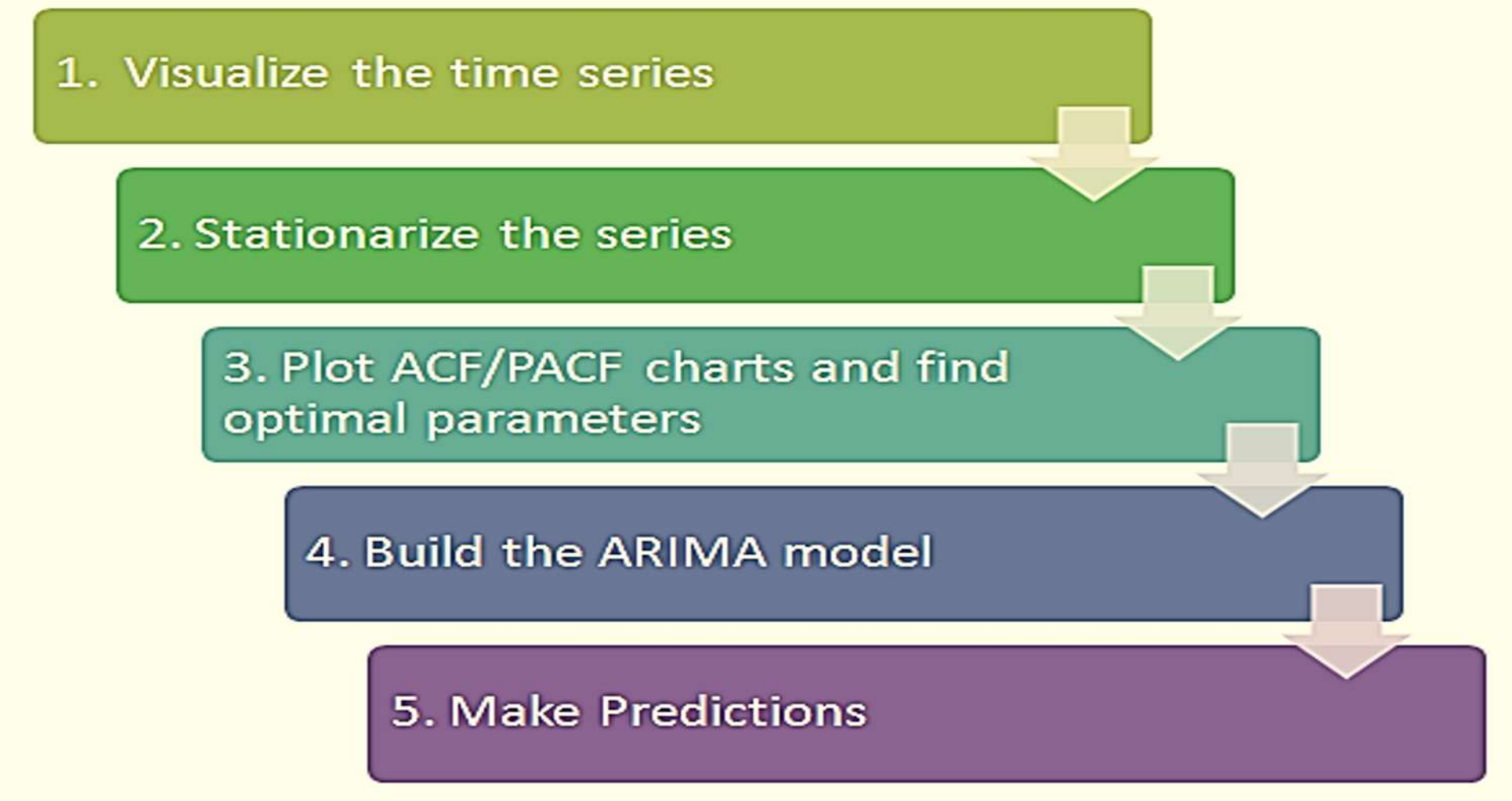
Modeling

- **Forecasting** the future time series values from the past values using the models developed

Time series analysis deals with:

- problems of identification of basic characteristic features of time series
- discovering - from the observation data on which the time series is built in- the internal time series structure.

steps of time series modeling:



Stationary and Non-Stationary Time Series

Stationary types

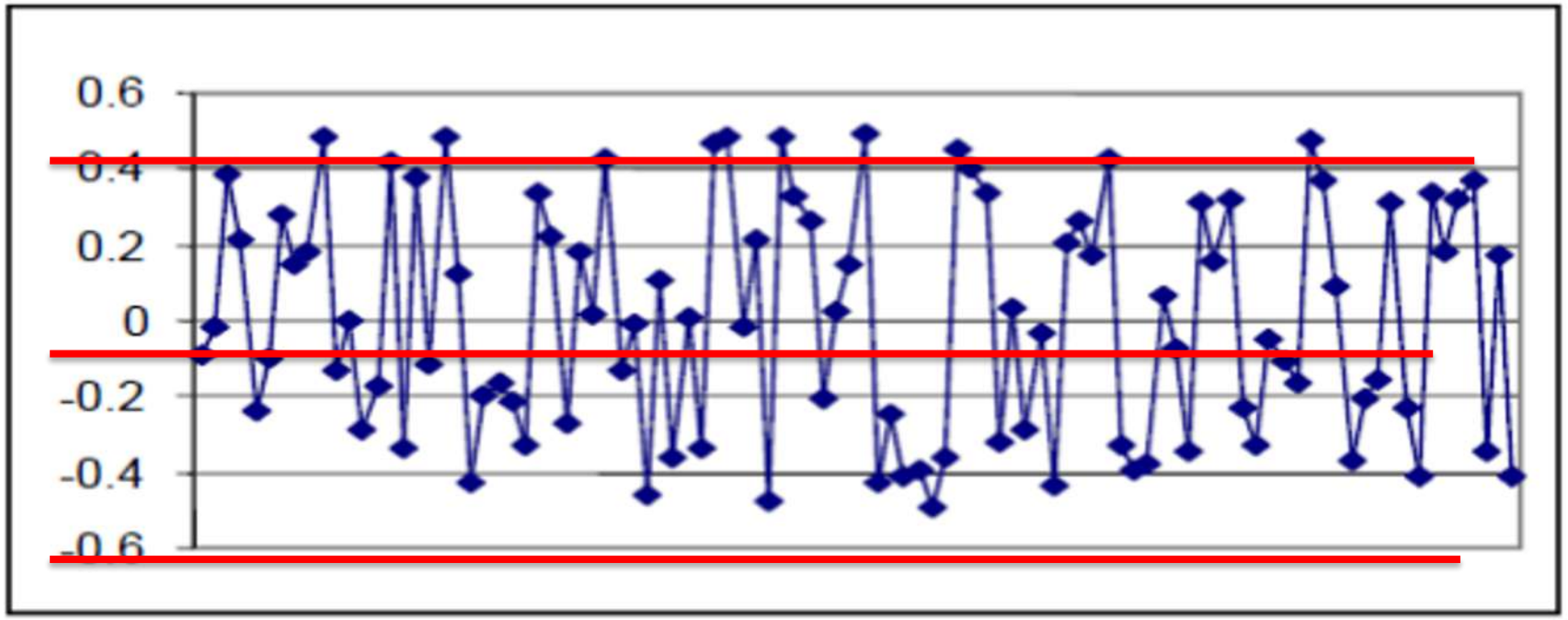
1- (Weak) Stationarity

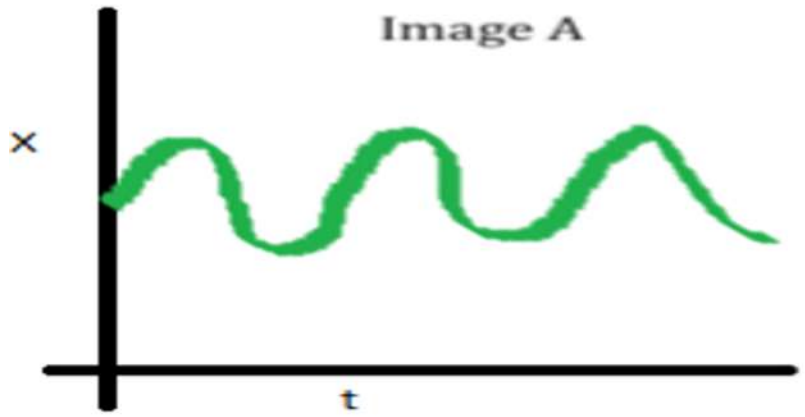
2- (strong)stationarity +Normality

$E(X_t)$ **= μ** (mean is constant in *t*)
$Var(X_t)$ **= σ2** (variance is constant in *t*)
$Cov(X_t, X_{t+k})$ **= χ(k)** (covariance is constant in *t*)

Mean

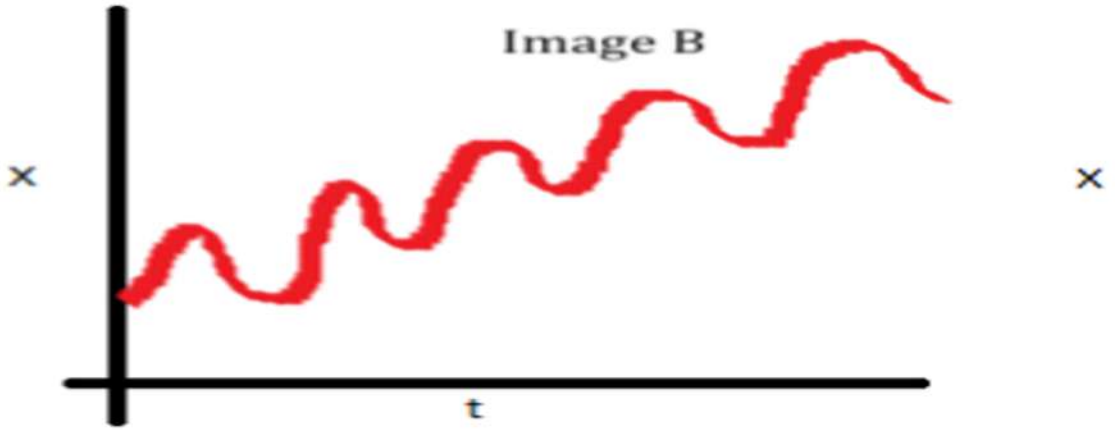
μ





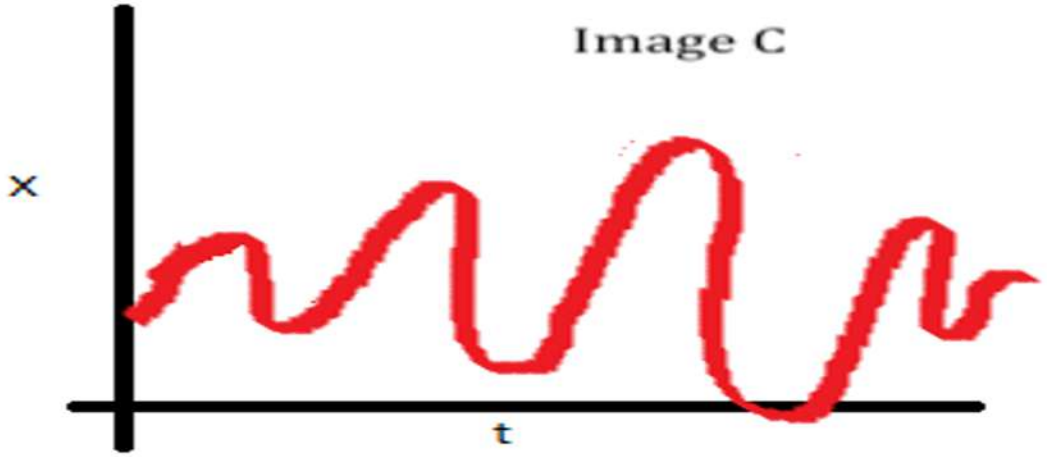
Stationary series

Time-Dependent Mean



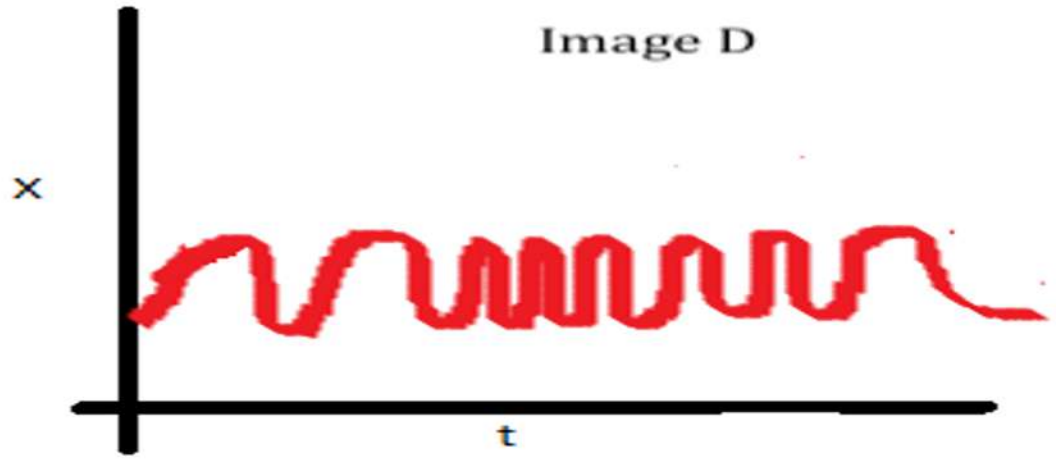
Non-Stationary series

Time-Dependent Variance

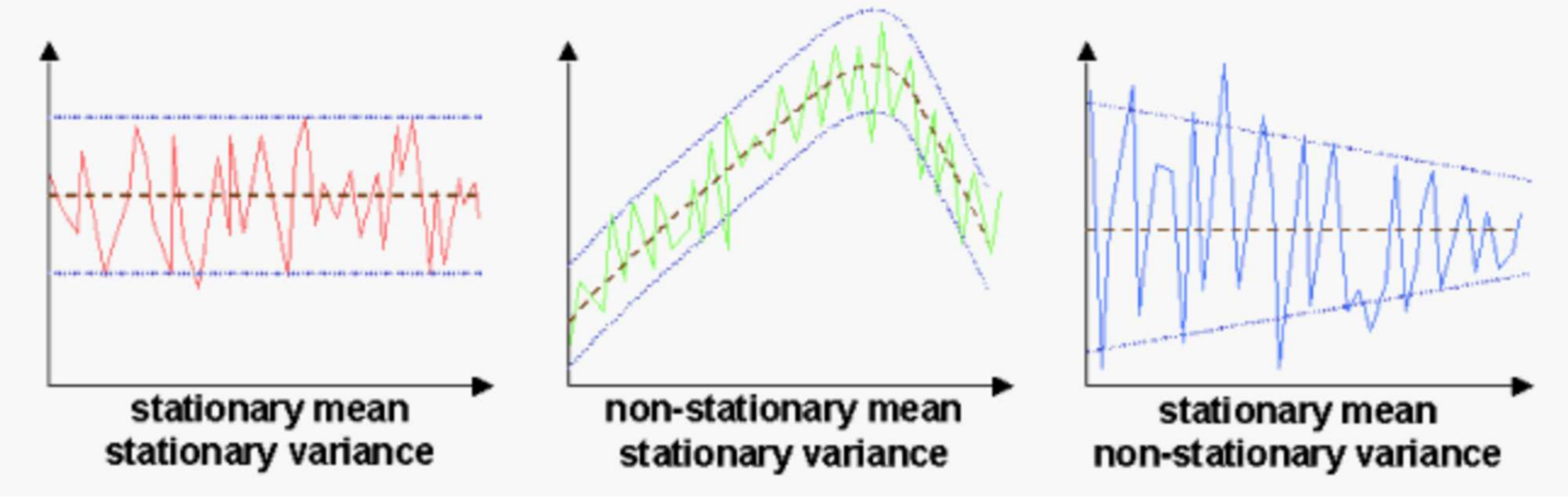


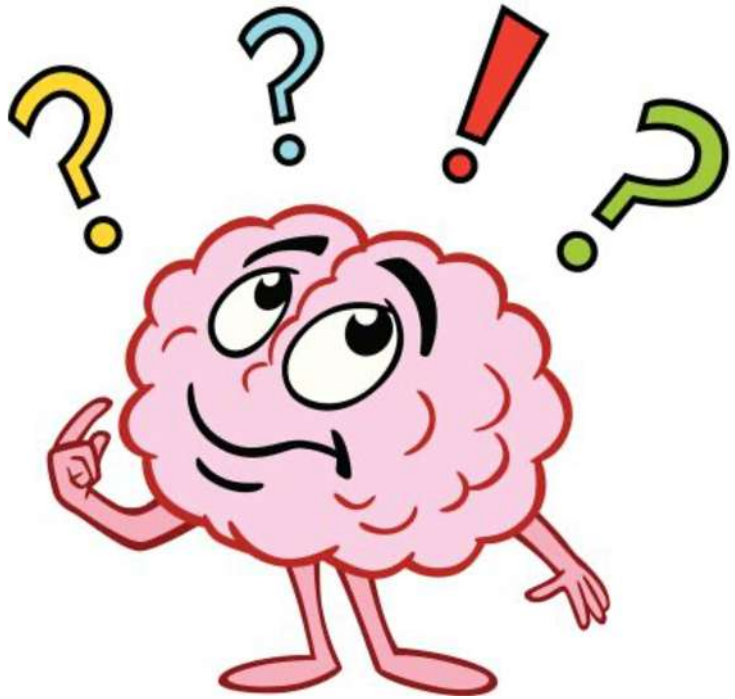
Non-Stationary series

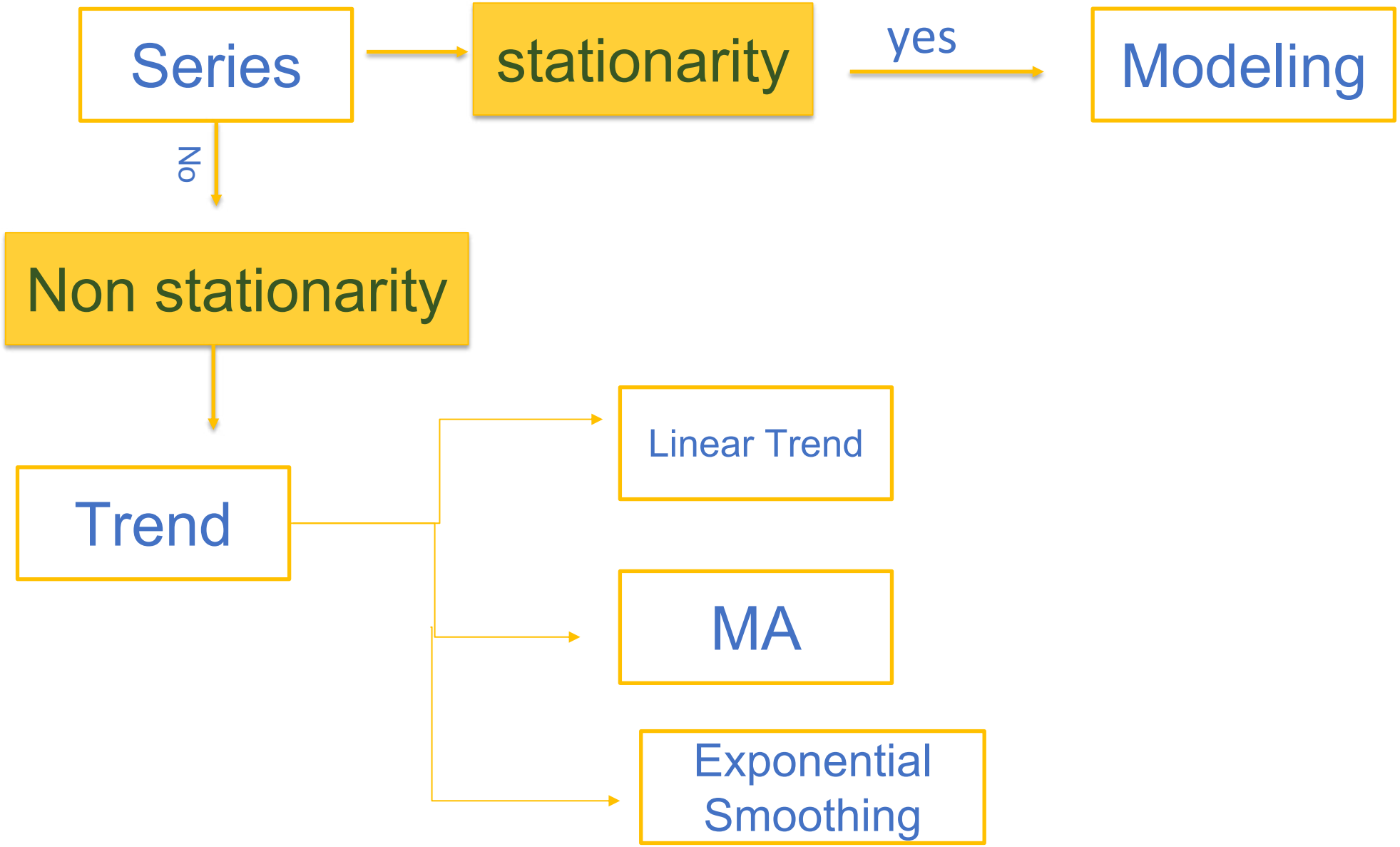
Time-Dependent Covariance

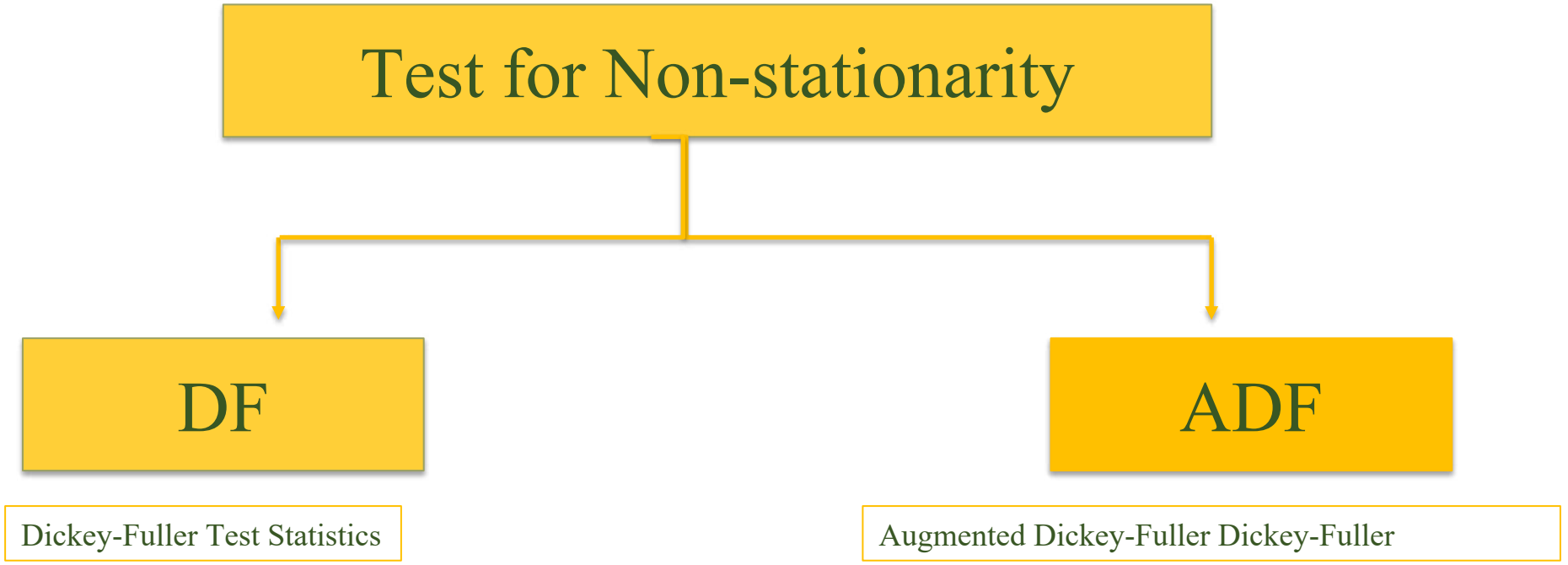


Non-Stationary series


stationary mean
stationary variance
non-stationary mean
stationary variance
stationary mean
non-stationary variance




Series
stationarity
yes
Modeling
no
Non stationarity
Trend
Linear Trend
MA
Exponential Smoothing



- H_o : Ts is Non – stationarity
  H_1 : Ts is Stationarity
- If P-Value > 5% ⟶ Non – stationarity

Non-stationarity(treatment)

Mean(μ)

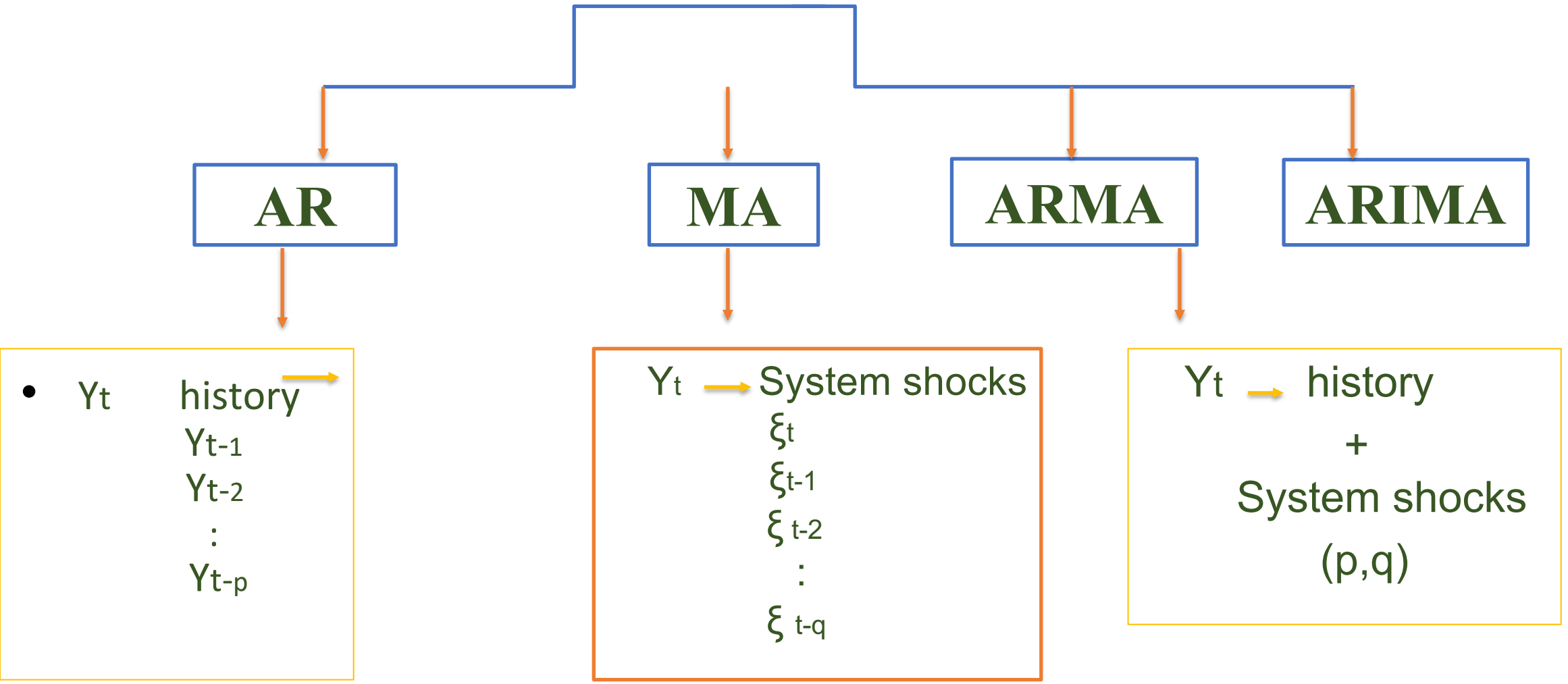
- The first diff. $(Y_t - Y_{t-1})$
- The second diff.

Variance (σ^2)

- Ln (Y_t)
- Root (Y_t)
- Y_t squared

- Ts ⟶ Stationarity , $Y_t \sim I_{(0)}$
- # diff Required to Achieve Stationarity. Now , $Ts = F(\xi_t)$

Time series Modeling



- **ARIMA** → **(p , d , q)**

I = # diff Required to Achieve Stationarity

- #diff already known
- Ts is Non – stationarity in mean

p: the number of lag observations in the model; also known as the lag order.

d: the number of times that the raw observations are differenced; also known as the degree of differencing.

q: the size of the moving average window; also known as the order of the moving average.

-

Determine the appropriate model for forecasting

Autocorrelations (AC) contain the same information as tha Autocovariances, with the advantage of not depending on the units of measurement.

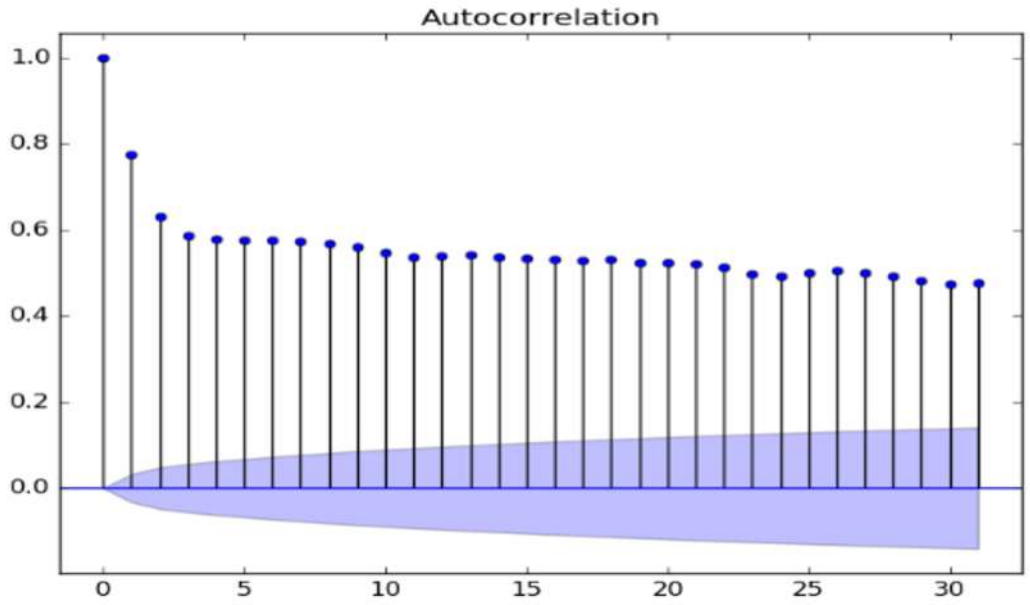
partial autocorrelation(PAC) is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed

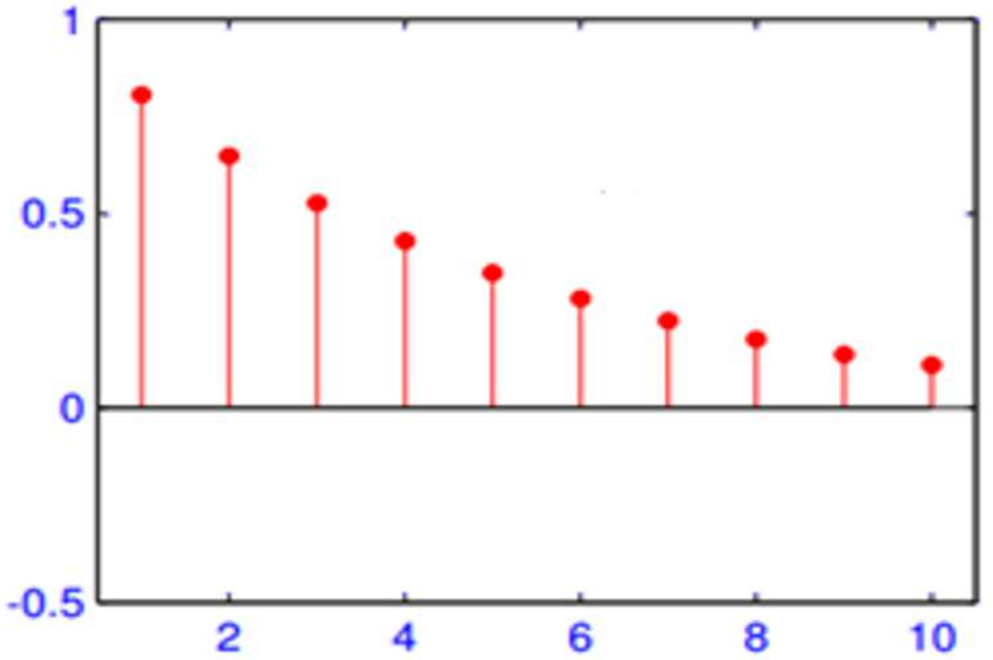
Determine the appropriate model for forecasting

- **Autoregressive (AR)**
- **[AC]** → Slowly

Autoregressive (AR)

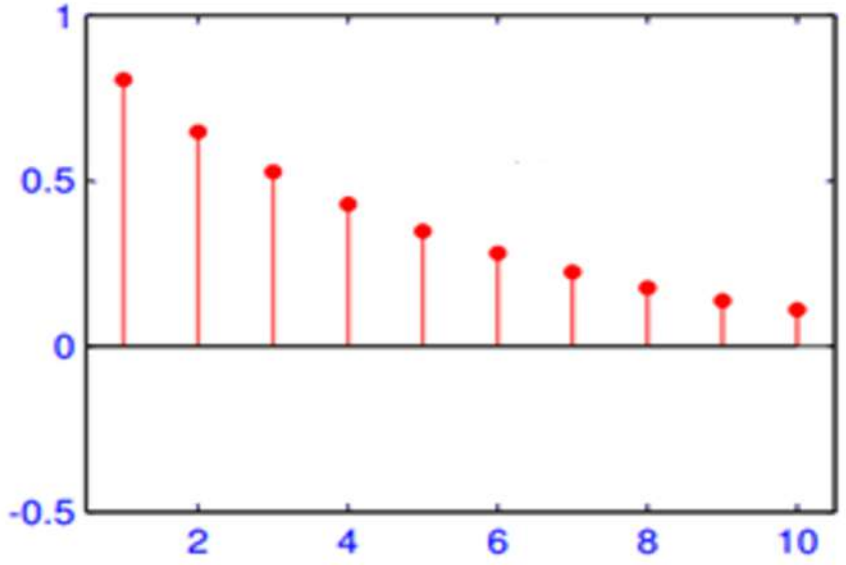
PAC → quickly





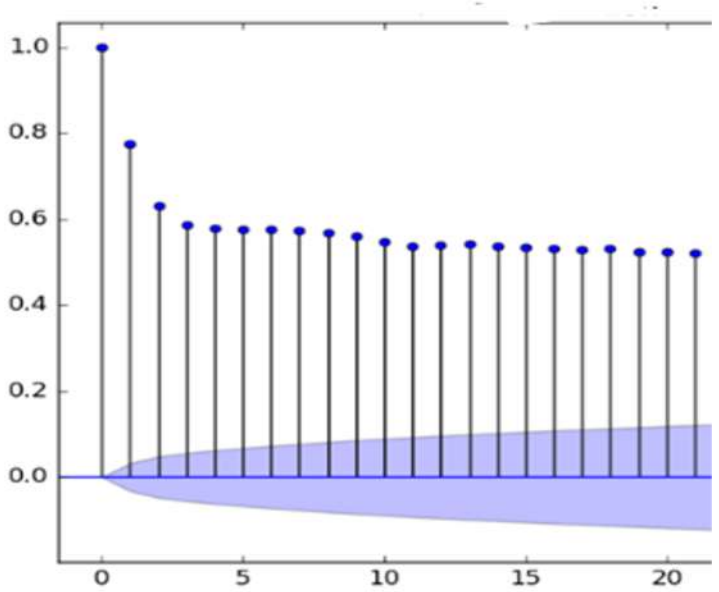
Determine the appropriate model for forecasting

- **Moving Averages (MA)**
- **[AC]** → quickly

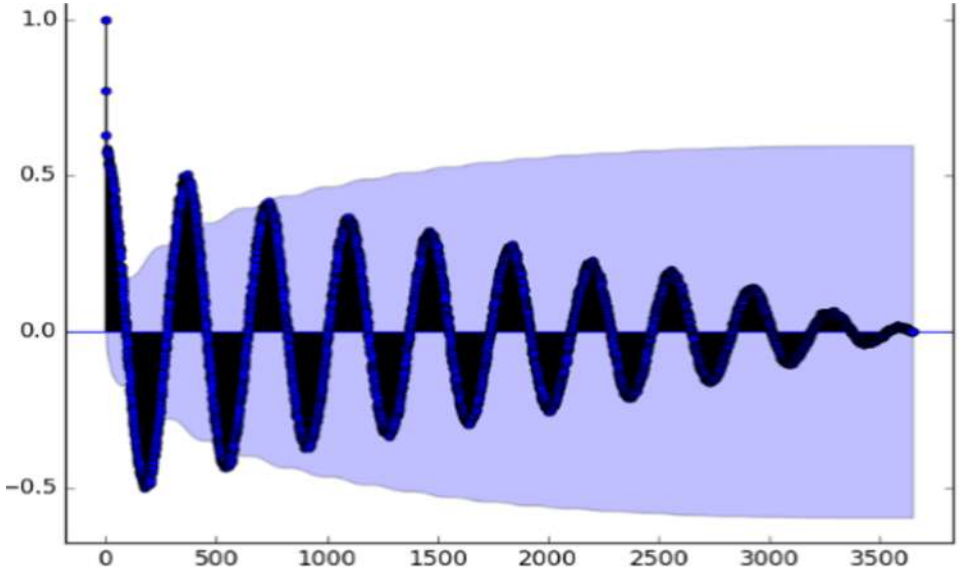


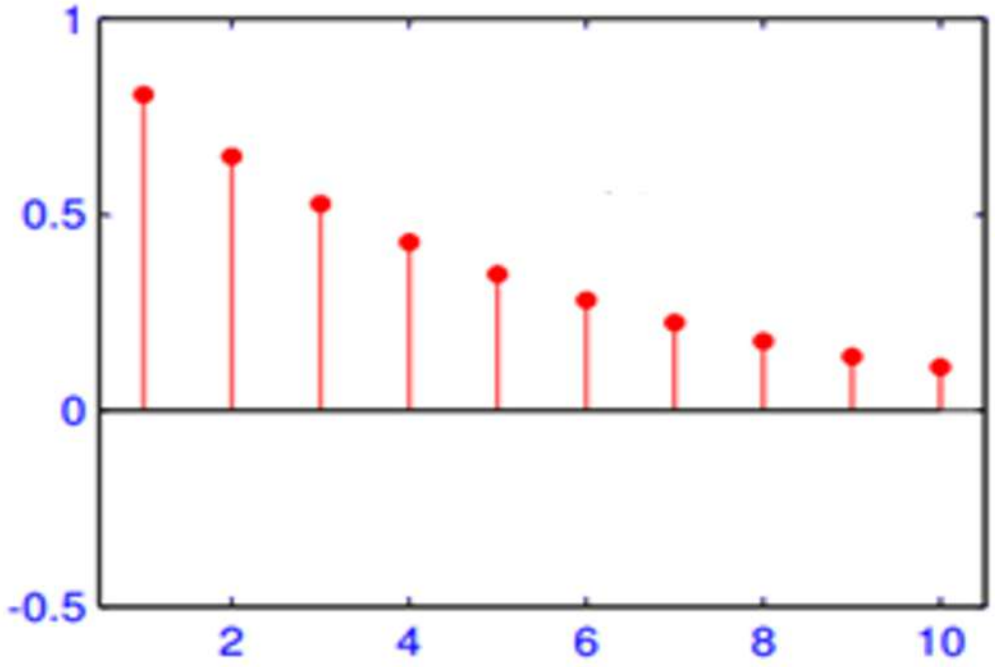
Moving Averages (MA)

[PAC] → Slowly



- **ARMA** → **AC** and **PAC** are quickly.





Diagnostic the Model

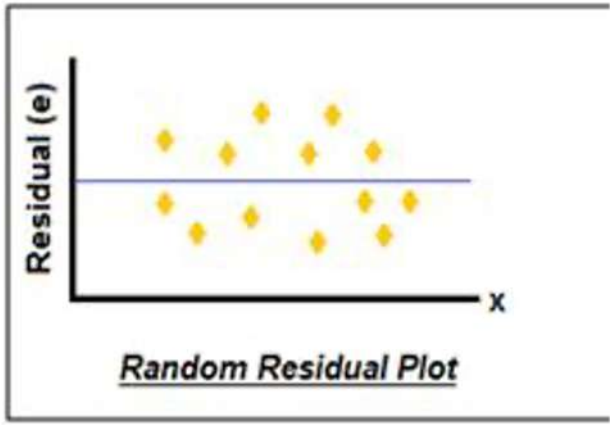
1- **ATC**

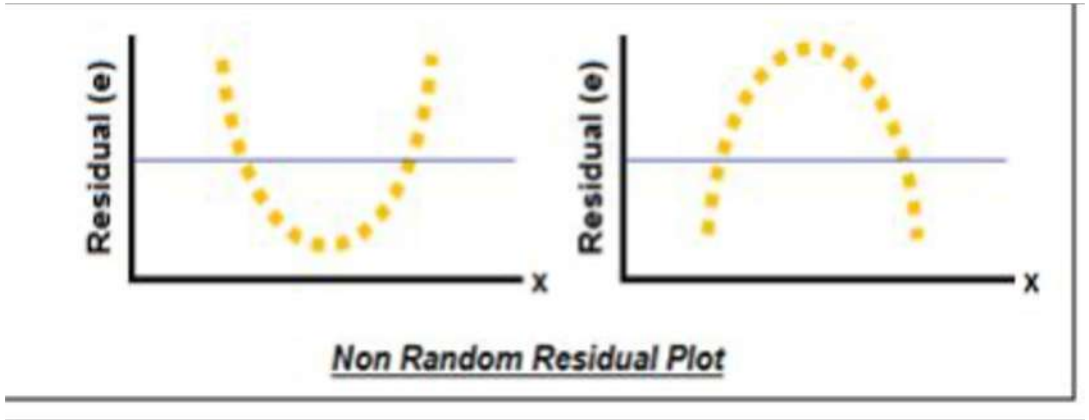
BIC

→ Minimum error

2- **MSE** → Minimum

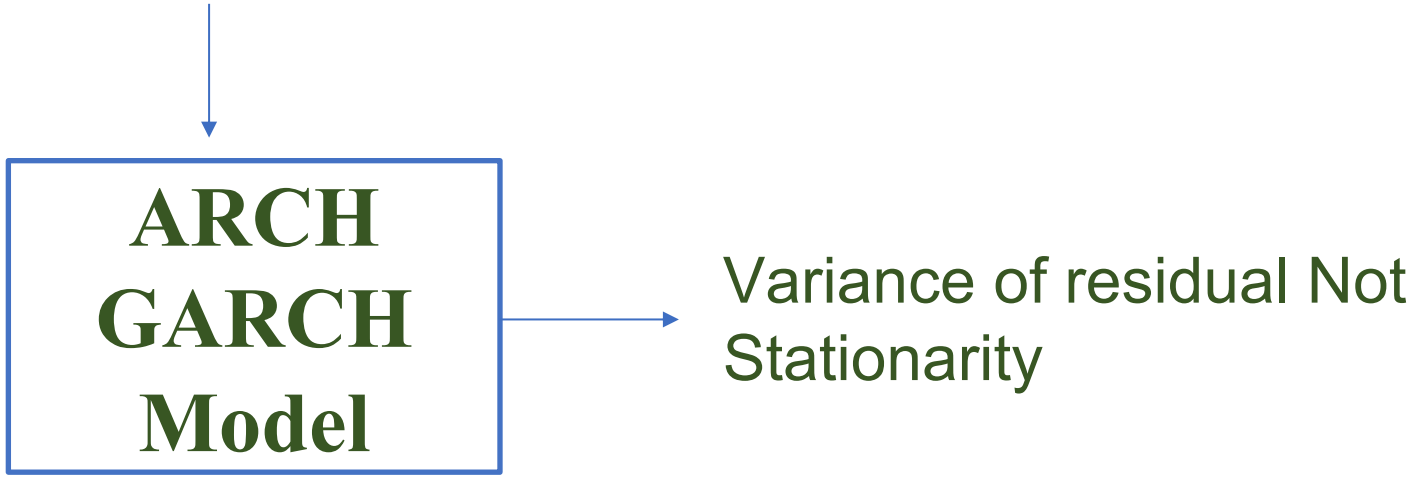
3- Random Residuals





4- Test For Stationarity For residual.

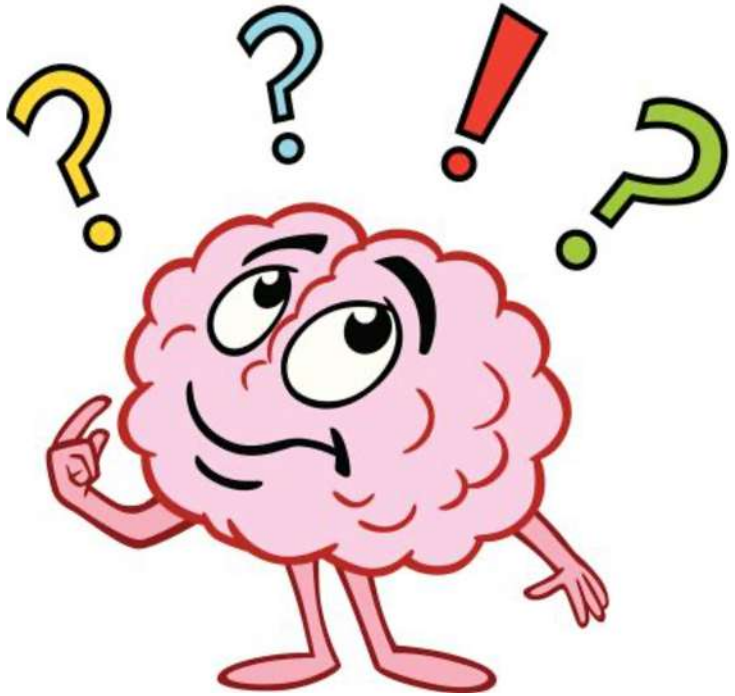
- If it’s Non – Stationarity



What do ARIMA models look like?

Machine/Deep learning Techniques and project

- Machine/Deep learning architectures and methods used for time series analysis
- Feature Engineering for Time Series
- Cross-Validation for Time Series
- Support Vector Regression (SVR)
- Prophet Model
- Long-Short Term Memory (LSTM)
- Recurrent neural networks (RNNs).



Thank you!

Keep in touch through

doaa.mahmoud2294@gmail.com

Contact Us

Website

Machinfy.com

email

sales@Machinfy.com

phones

+201154447771

WhatsApp

+201154447771

Facebook

FB.com/Machinfy

Twitter

Twitter.com/Machinfy

YouTube

Machinfy School

LinkedIn

https://www.linkedin.com/company/machinfy