HCAI5DS02 – Data Analytics and Visualization.
Lecture – 10
Introduction to Time Series Analysis.

Siman Giri

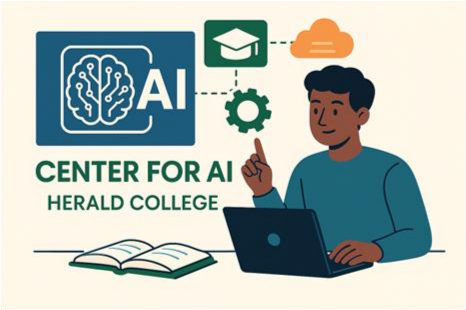


image generated via copilot.

1.1 What is Time Series Data?

- **Time series Data** is a **sequence of data points** collected or recorded at **specific, consistent time intervals**.
- Think of **it as any dataset** where you can ask;
  - *"What was the value at this specific time ?"*
- **Key characteristics:**
  - **Time series data** has **three fundamental** components:
    1. **Time stamps:** The specific points in time when each measurement is taken
       - **e.g. 2023 – 1 – 27 , 14:30:00, Jan 2023, Q2 2022**
    2. **Metric/Value:** The actual measurement or observation recorded at that time
       - **e.g. 25^{o}C, $145.67, 10,000 website visits**
    3. **Regular Intervals (usually):** The data is most valuable when collected at consistent intervals.
       - **e.g. every second, every hour, every day**
       - However, **irregular intervals like emergency room visits can also form a time series.**

1.2 Time Series Data: Example.

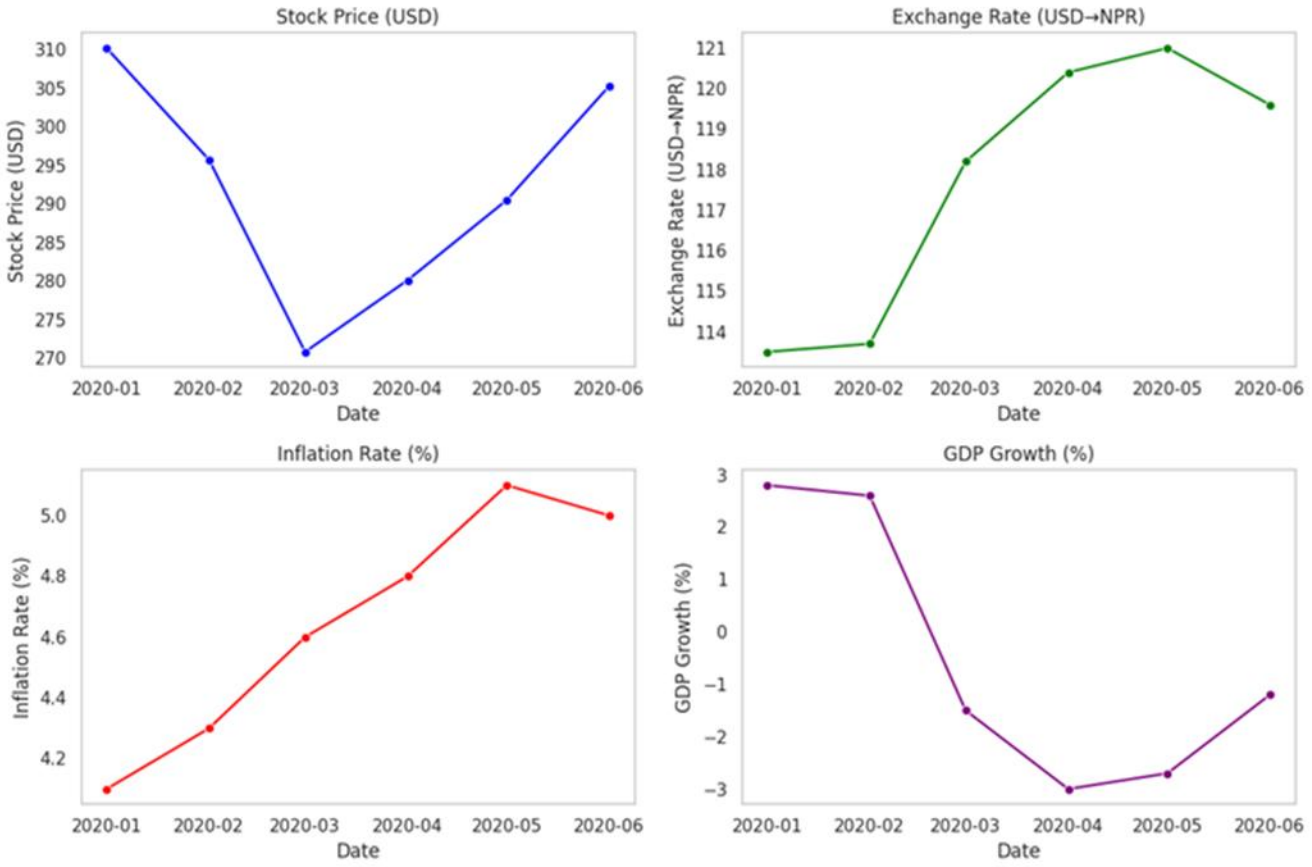
- **Data gathered sequentially in time are called a time series.**

1. **Economics and Finance:**
   - Daily closing of Apple stock (**X_t = price on day t**).
   - Monthly inflation rate or unemployment rate.
2. **Environmental Modelling:**
   - Hourly **CO_2 concentration levels** in **the atmosphere**.
   - Daily river discharge measurements.
   - Ocean temperature variations across.
3. **Medicine:**
   - Electrocardiogram (ECG) signal of a patient (continuous time series).
   - Daily number of COVID – 19 cases.
   - Hourly blood sugar levels for a diabetic patient.

Date	Stock Price (USD)	Exchange Rate (USD→NPR)	Inflation Rate (%)	GDP Growth (%)
Jan 2020	310.2	113.5	4.1	2.8
Feb 2020	295.7	113.7	4.3	2.6
March 2020	270.8	118.2	4.6	-1.5
Apr 2020	280.1	120.4	4.8	-3.0

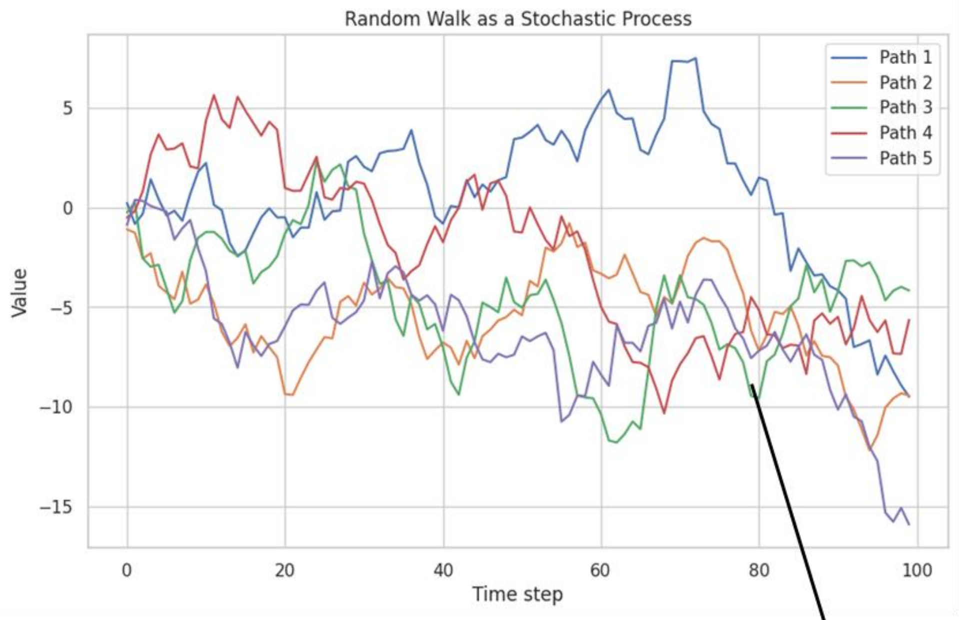
1.2.1 Visualizing Time Series Data.

Date	Stock Price (USD)	Exchange Rate (USD→NPR)	Inflation Rate (%)	GDP Growth (%)
Jan 2020	310.2	113.5	4.1	2.8
Feb 2020	295.7	113.7	4.3	2.6
March 2020	270.8	118.2	4.6	-1.5
Apr 2020	280.1	120.4	4.8	-3.0



1.3 Mathematical Definition of a Time Series.

- A **time series** can be **formally defined as**: $\{X_t : t \in T\}$ **where:**
  - X_t = observed value of the varaible at time t
  - t = time index (discrete: daily, monthly, or continuous: every second
  - T = ordered set of time points
- So, **a time series** is essentially a **stochastic process indexed by time**.
- A **stochastic process** is a family of random variables:
  - $\{Y_t : t \in T\}$ defined on a **common probability space** $\{\Omega, \mathcal{F}, \mathbb{P}\}$,
    - where **t belongs to an index set**
    - (e.g. **time:** $T = \mathbb{N}$ **for discrete or** $T = \mathbb{R}^+$ **for continuous**).
    - Y_t = **the random variable at time t.**
  - **Together, they describe the evolution of randomness across time.**
- **Relation to Time Series:**
  - **Time series = one realization (sample path) of a stochastic process.**
  - For example:
    - **Stochastic process: all possible ways a stock price could evolve.**
    - Time Series: the actual observed stock price sequence in reality.

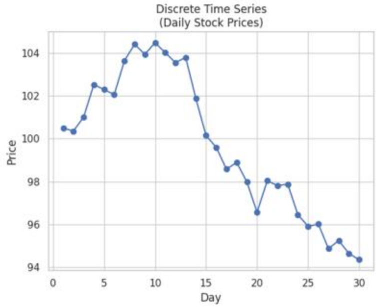


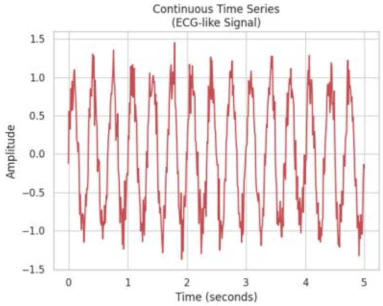
A Time Series Data.

1.3.1 Discrete Vs. Continuous Time Series Data.

- **Discrete Time Series:**
  - A **discrete – time series** is a **sequence** of
    - **random variables** observed at
      - **equally spaced** discrete time points.
  - **Formally:**
    - $\{\mathbf{X_t}: \mathbf{t} \in \mathbb{Z}^+\}, \ \mathbf{t = 1, 2, 3, \ldots}$
    - where **each** $\mathbf{X_t}$ is a random variable
    - observed at **time step t.**
  - **Example:**
    - Daily closing price of a stock,
    - monthly unemployment rates,
    - annual rainfall.
  - observations at ***fixed intervals***

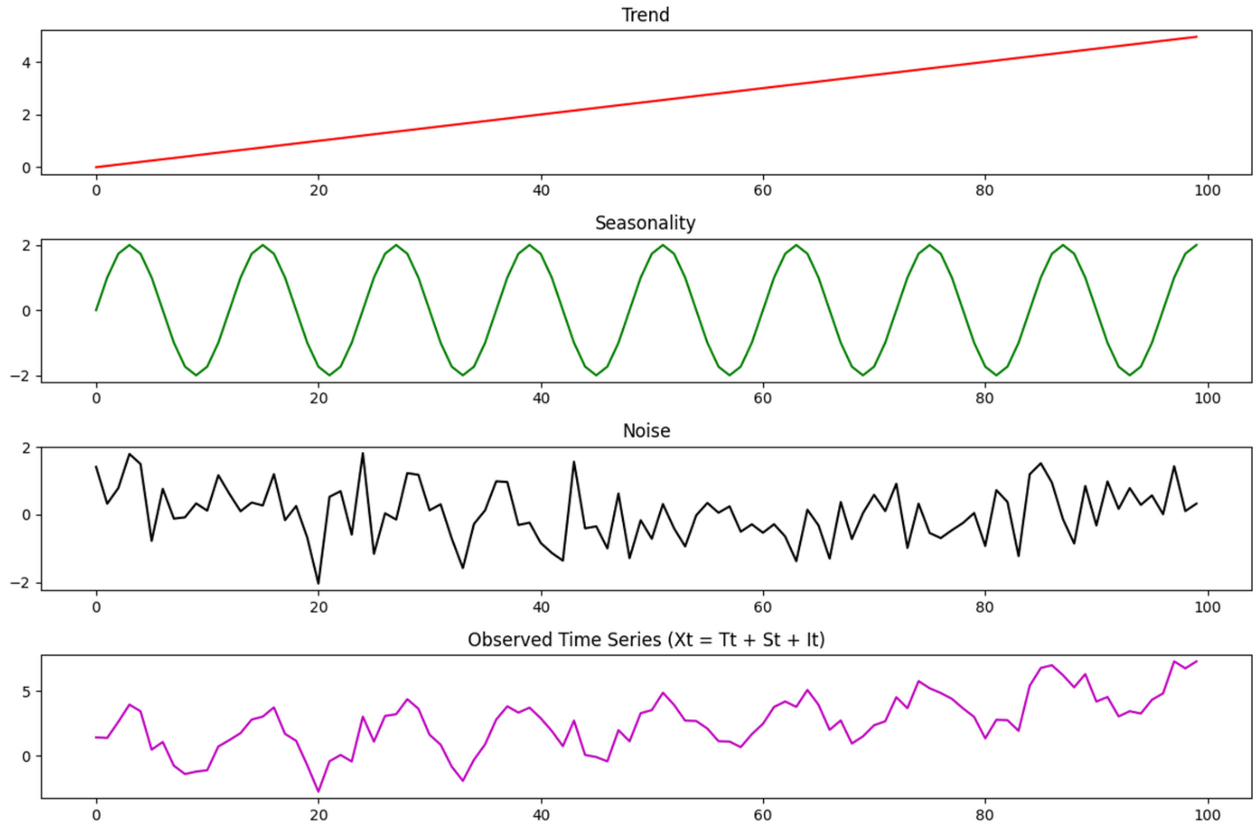
- **Continuous Time Series:**
  - A **continuous – time series** (or continuous – time stochastic process)
  - is a family of **random variables** indexed
    - by **continuous time**.
  - Formally:
    - $\{\mathbf{X(t)}: \mathbf{t} \in \mathbb{R}, \mathbf{t} \geq \mathbf{0}\}$
    - where $\mathbf{X(t)}$ gives the value of the process
    - at any **real – valued time t.**
  - **Example:**
    - ECG signal, temperature recorded continuously.
  - observations can be taken ***at any point in time*** (theoretically infinite resolution).





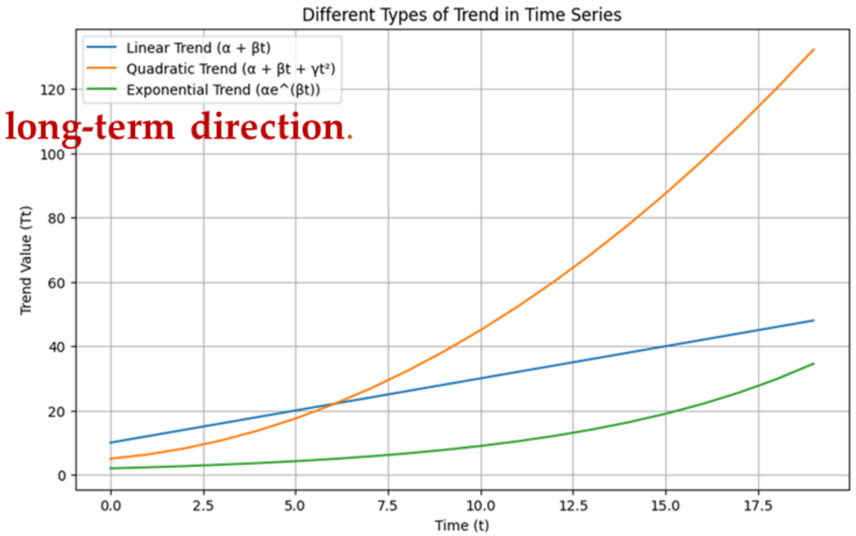
1.4 Basic Components of Time Series Data.

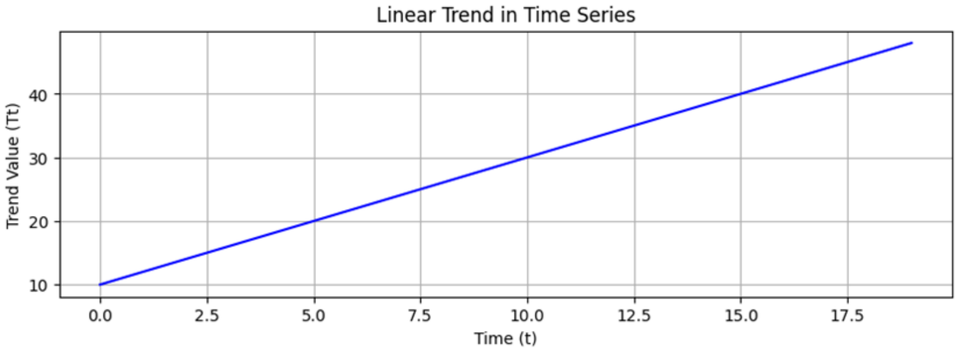
- Any **time series** can be broadly viewed as a combination of:
  - **Trend (T_t)** → Long-term movement or direction of the series (upward, downward, flat).
  - **Seasonality (S_t)** → Regular, repeating short-term patterns with fixed frequency.
  - **Noise/Irregularity (I_t)** → Random, unpredictable fluctuations.



1.4.1 Basic Component – Trend (T_t).

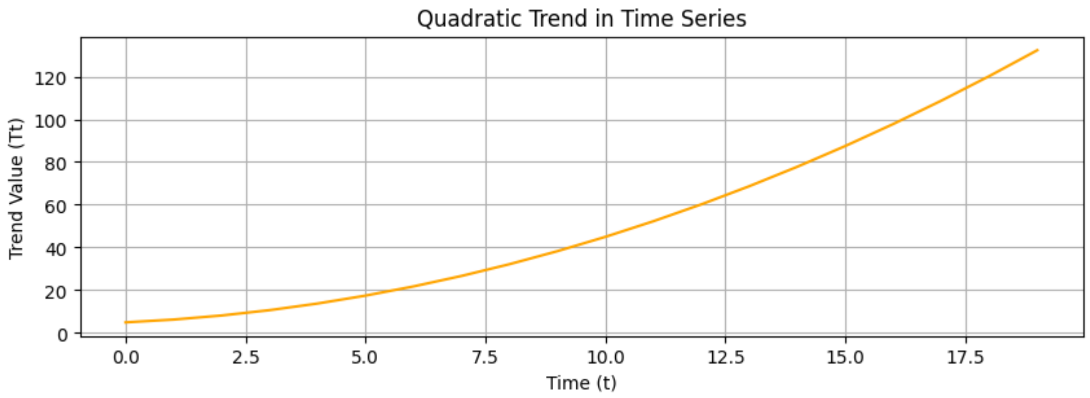
- **Definition:** The smooth, slowly varying part of the series that shows **the long-term direction**.
- **Mathematical Form:** Often modeled as a deterministic function of time.
    1. Linear Trend:
        - $\mathbf{T_t = \alpha + \beta \cdot t}$
        - Here:
            - **α: intercept, baseline value of the series at time $t = 0$.**
            - **β: slope, rate of change per time unit.**
        - This **reperesents a straight-line trend**,
    - **Example:**
        - **Sales increasing by 100 units per month.**





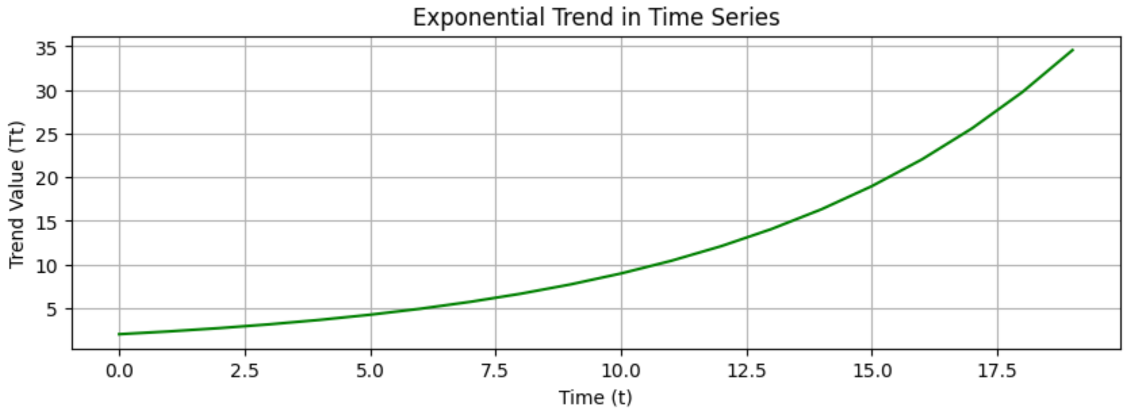
1.4.1 Basic Component – Trend (T_t).

- **Definition:** The smooth, slowly varying part of the series that shows **the long-term direction**.
- **Mathematical Form:** Often modeled as a deterministic function of time.
  2. **Quadratic or Polynomial Trend:**
     - $\mathbf{T_t = \alpha + \beta \cdot t + \gamma \cdot t^2}$
     - Adds curvature to the trend.
     - **γ: controls acceleration or deceleration of the trend.**
       - **Positive γ → trend bends upwards (accelerating growth).**
       - **Negative γ → trend bends downwards (slowing growth).**
     - Can **be extended to higher – order polynomials** (e.g., cubic, quartic).



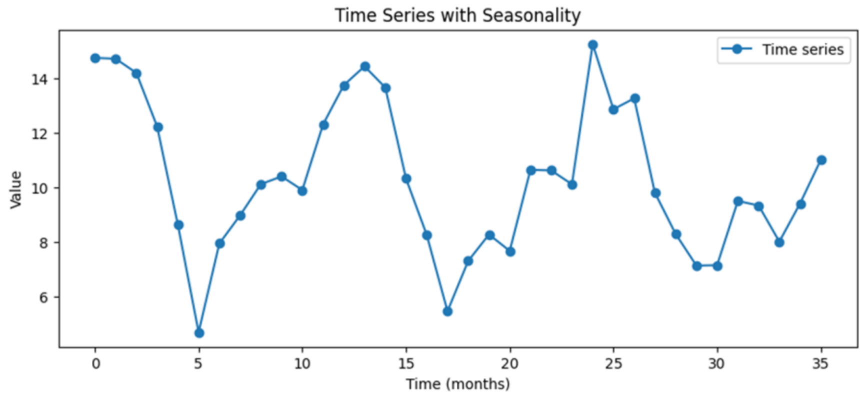
1.4.1 Basic Component – Trend (T_t).

- **Definition:** The smooth, slowly varying part of the series that shows **the long-term direction**.
- **Mathematical Form:** Often modeled as a deterministic function of time.
  3. **Exponential Trend:**
     - $\mathbf{T_t = \alpha e^{\beta t}}$
       - Growth or decay is multiplicative rather than additive.
         - If $\beta > 0 \rightarrow$ the trend grows exponentially (**compounding growth**)
         - If $\beta < 0 \rightarrow$ the trend decays exponentially (**compounding decline**)
       - Often used in finance, population growth, epidemiology etc.



1.4.2 Basic Component – Seasonality (S_t).

- **Definition:** Regular, **periodic fluctuations** with fixed and known period **p**.
  - **Mathematical:** $S_{t+p} = S_t \ \forall t$
  - This is the formal definition of a **perfectly periodic (strict) seasonal component** in a time series.
  - Here:
    - **S_t**: The value of **the seasonal component** at **time t.**
    - **S_{t+p}**: The value of **the seasonal component** at exactly one **period p later**.
    - **$\forall t$**: For all time **points t.**
  - The equation means:
    - The **seasonal pattern repeats every p time steps**.
      - **For example:**
        - If **sales peak every December**,
          - then **the seasonal effect for December 2024** is **the same as December 2025**.



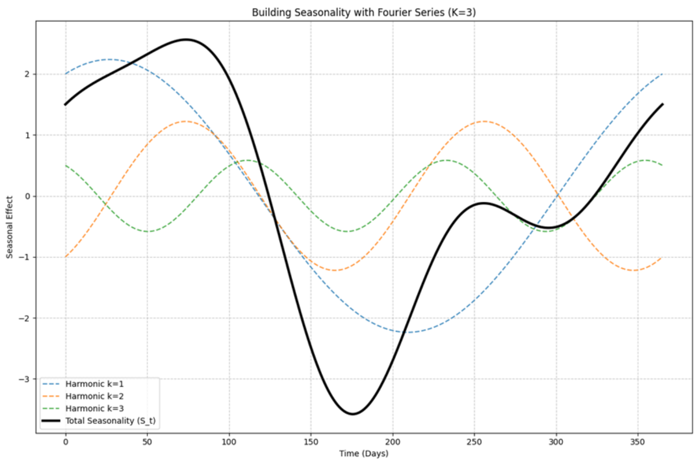
1.4.2 Basic Component – Seasonality (S_t).

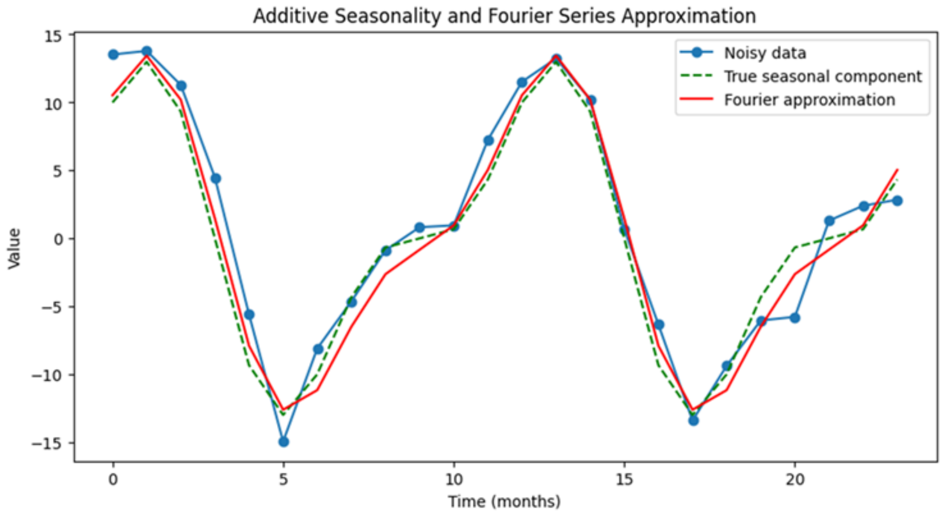
- **Representing Seasonality with Fourier Series :**
  - The Modeling Tool → **Fourier Series:**
    - Sometimes the seasonal component is **complex** and not just a simple sine wave.
    - In such cases, we can **decompose it into a sum of sines and cosines**:
      - $\mathbf{S_t = \sum_{k=1}^{K} \left[a_k \cos\left(\frac{2\pi kt}{p}\right) + b_k \sin\left(\frac{2\pi kt}{p}\right)\right]}$
      - This is the how do we actually create a function S_t that obeys the rule $\mathbf{S_{t+p} = S_t \; \forall t}$.
      - Here we use **sum of sine and cosine waves** , which are **naturally periodic**.
        - This sum is called a **Fourier Series**.
- **Let's dissect the components:**

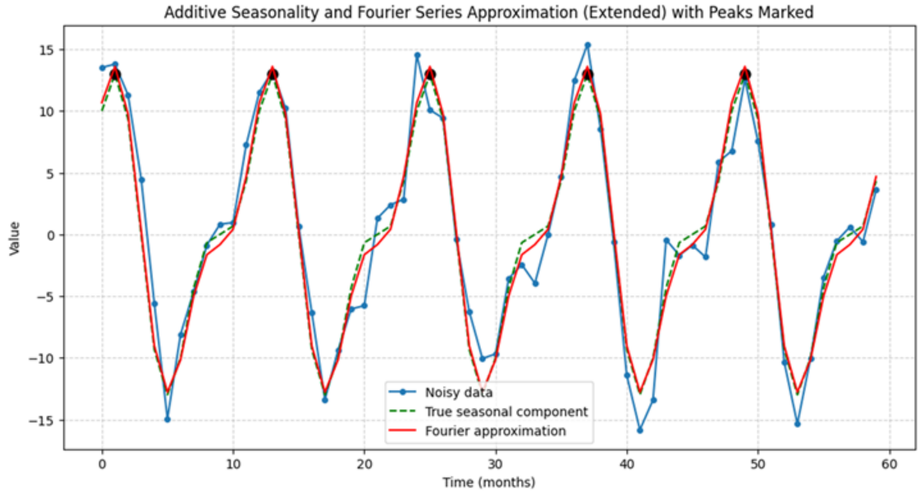
Symbol	Meaning
$\mathbf{k}$	Harmonic number. • k = 1 is the fundamental frequency; • k > 1 are higher harmonics capturing more complex seasonal patterns.
$\mathbf{a_k, b_k}$	• Coefficients that determine the amplitude of each cosine and sine wave. • They are estimated from the data.
$\mathbf{\frac{2\pi kt}{p}}$	• Frequency of wave. • Ensures the wave completes k cycles over the period p.
$\mathbf{K}$	• Number of harmonics included. • More harmonics → more detailed/complex seasonal pattern.

1.4.2 Basic Component – Seasonality (S_t).

- **Representing Seasonality with Fourier Series :**
  - **Example:**
    - Suppose monthly sales have yearly seasonality **(p = 12):**
      - $S_t = a_1 \cos\left(\frac{2\pi t}{12}\right) + b_1 \sin\left(\frac{2\pi t}{12}\right) + a_2 \cos\left(\frac{4\pi t}{12}\right) + b_2 \sin\left(\frac{4\pi t}{12}\right)$
        - **k = 1 captures the main yearly peak.**
        - **k = 2 captures secondary peaks e.g. mid year sales.**
    - Adding more **harmonics k > 2 captures** finer details.
  - In short, **additive seasonality** assumes a **repeating pattern**, and
    - Fourier series provides a flexible mathematical way to represent it using sine and cosine waves.

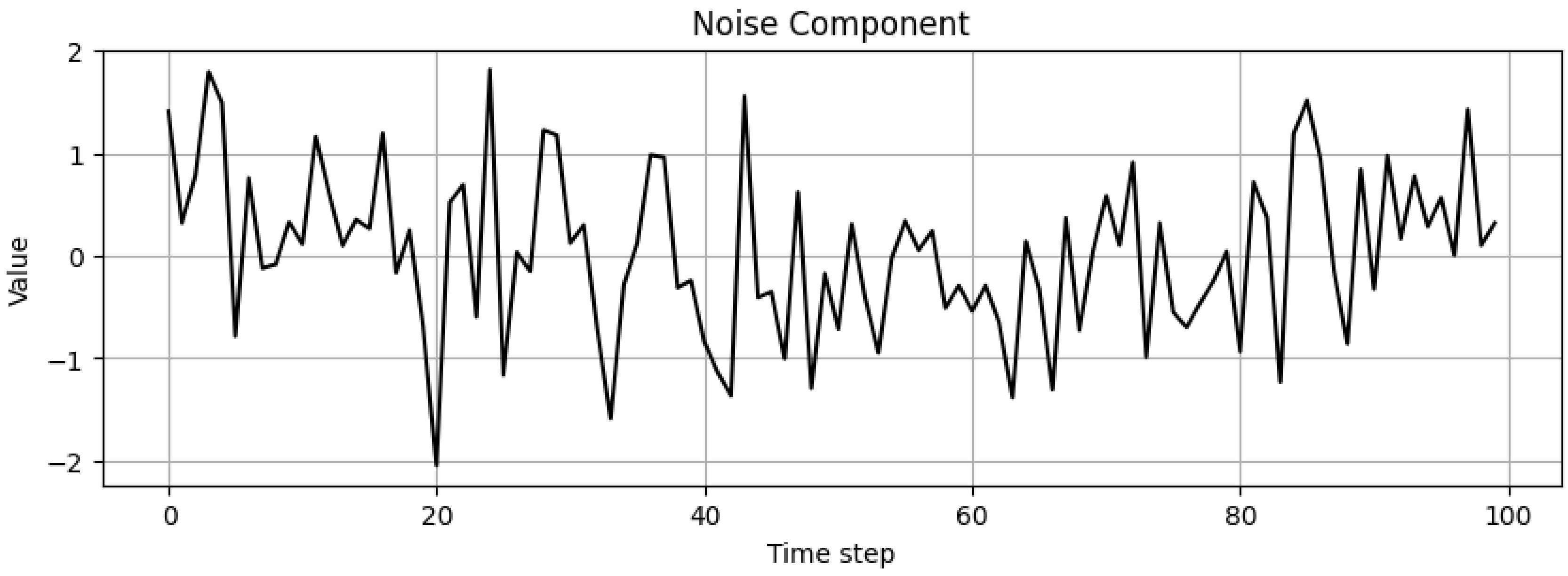






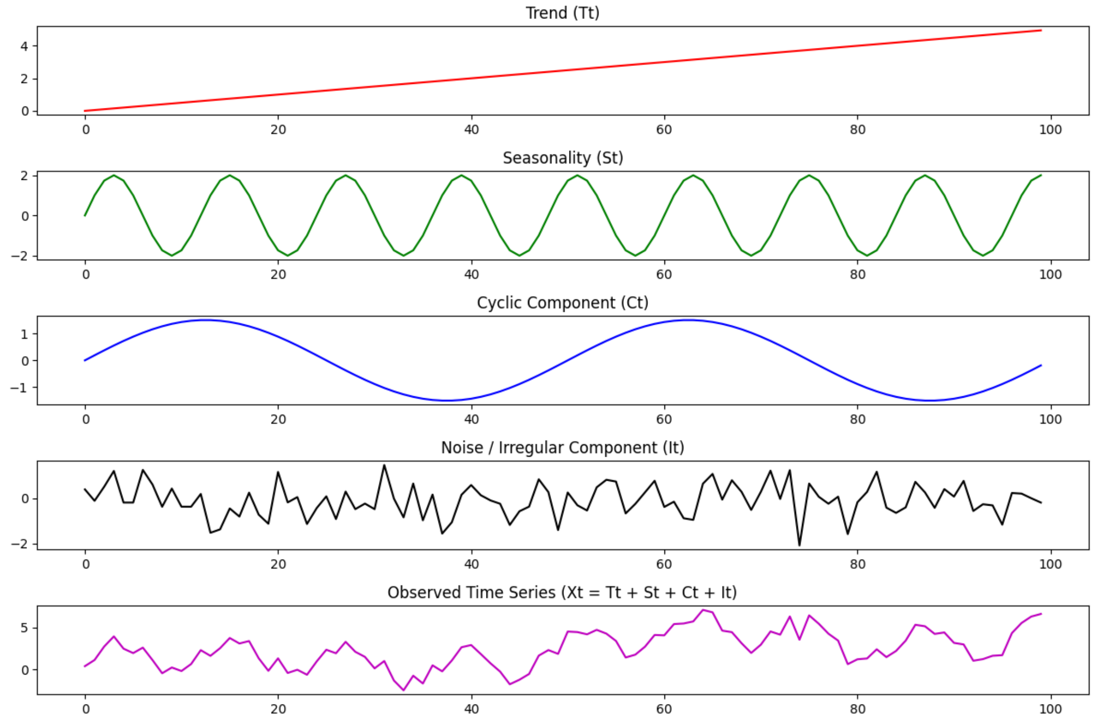
1.4.3 Basic Component – Irregular Component (I_t).

- **Definition:** Residual part of **the time series** after **removing trend, seasonality, and cycles**.
- **Mathematical Form:**
  - $\mathbf{I_t \sim \epsilon_t}$, **here** $\mathbb{E}[\epsilon_t] = 0$, & $\mathrm{Var}(\epsilon_t) = \sigma^2$.
    - where ϵ_t is random error term also known as white noise $\left(\epsilon_t \sim \text{i.i.d } N(0, \sigma^2)\right)$



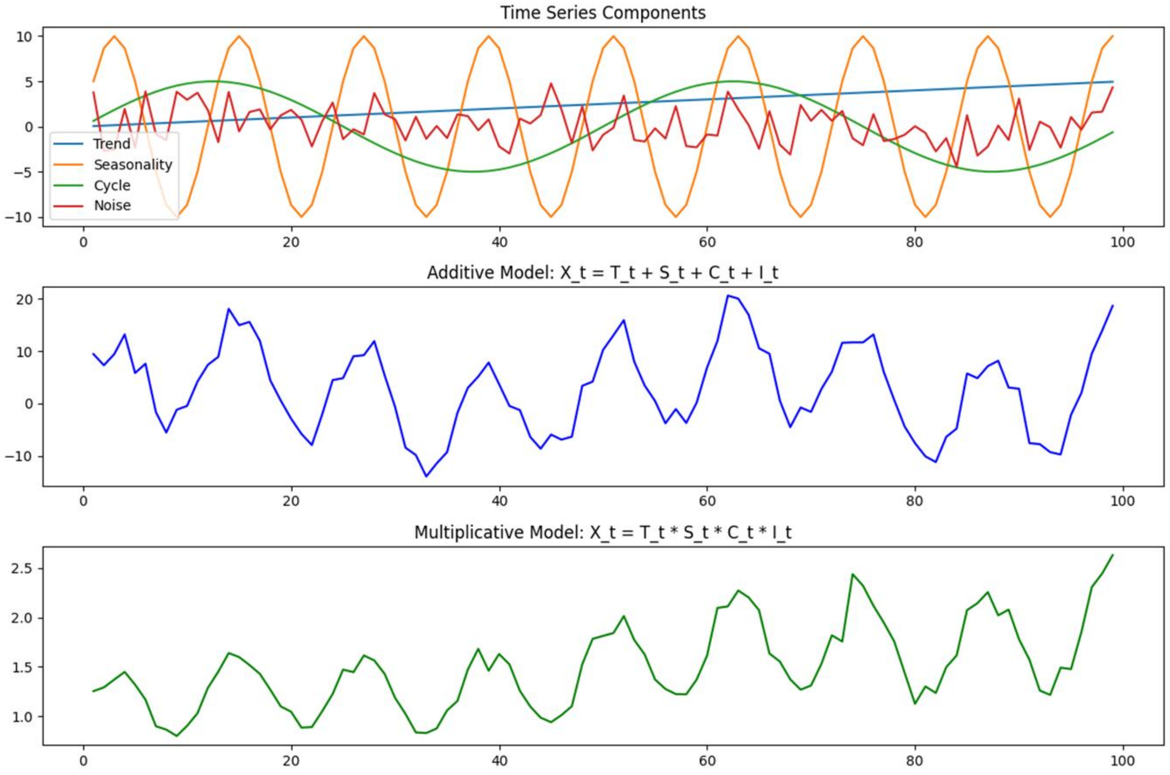
1.5 Extended Component – Time Series Data.

- **Cyclic Component (C_t):**
  - **Definition:** Long – term oscillations without a fixed period, typically influenced by external factors such as economy, politics.
  - **Mathematical Form:**
    - Modeled as smooth fluctuations, often approximated by autoregressive processes or low – frequency sine/cosine terms:
      - $C_t \sim f(t)$ **with period not constant.**
    - Example: boom – recession cycles.



1.6 Decomposition of Time Series.

- A **time series** can generally be expressed as a sum or product of four distinct components:
  - **Additive Form (aka Additive Model):**
    - $\mathbf{X_t = T_t + S_t + C_t + I_t}$
  - **Or, Multiplicative forms (aka Multiplicative Model):**
    - $\mathbf{X_t = T_t \cdot S_t \cdot C_t \cdot I_t}$



2. General Approach to Time Series Modeling.

2.1 Time Series Forecasting: Workflow in Practice.

1. **Visual Inspection:**
   - Plot the series and examine the main features of the graph, checking whether there is
     - a trend,
     - a seasonal or another periodic component,
     - any apparent sharp changes in behavior,
     - any outlying observations.
2. **Remove Trend and Periodic Components:**
   - Remove the trend and periodic components to get stationary residuals (this step is called *detrending and deseasonalizing*).
   - Broadly speaking, a time series is said to be ***stationary*** if there is
     - no systematic change in the mean (no trend);
     - no systematic change in the variance, and
     - no strictly periodic variations.
3. **Model Stationary Residuals:**
   - Choose a model to fit the residuals, making use of various sample statistics,
     - including the ***sample autocorrelation function* (ACF).**
   - Forecast the residuals and then invert the transformations to arrive at forecasts of the original series.

2.2 Classes of Forecasting Models.

- Time series forecasting models can be broadly divided into two general classes:
  - **Univariate Time Series Models:**
    - **Definition:**
      - These models use **only the historical values** of a **single variable** to **make predictions.**
        - No other **external factors (covariates)** are considered.
      - The **goal** is to **model patterns** in the series it self: trend, seasonality and **auto – correlation**.
    - Key Type is **Autoregressive (AR) Models:**
      - "Auto" means self – the series **predicts itself using past values**.
      - General form:
        - $\mathbf{Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}) + \epsilon_t}$
          - where p is the number of lags (how many past values we use).
        - Linear AR Model example:
          - $\mathbf{Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t}$
    - **Example:**
      - Predicting monthly airline passengers using only past passenger numbers.
    - Key feature: only the target series is needed, no external variables.

2.2.1 How Autoregressive Works?

- An **autoregressive (AR) model predicts** the current value of a time series using its own past values.
  - It is like saying:
    - "What happens today depends on what happened on the past?"
  - General Representations for $AR(p)$ is:
    - $\mathbf{Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t}$
    - Where:
      - Y_t = value of the series at time t
      - c = constant (intecept)
      - ϕ_i = coefficinets for past values
      - p = order (number of past lags used)
      - ϵ_t = random error at time t
  - Special Cases:
    - AR(1): First order autoregression $\rightarrow Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$
      - Today depends on yesterday.
    - AR(2): Second order autoregression $\rightarrow Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$
      - Today depends on yesterday and the day before yesterday.
    - In General AR(p) Today looks like the last p days combined.
- **Q: How do you choose the order p?**

2.2 Classes of Forecasting Models.

- **Mulitvariate Time Series Models:**
  - **Definition:**
    - These models use **additional variables (covariates, regressors, predictors)** along with the target series to improve forecasts.
    - They can **capture how external factors influence the target variable**.
  - **Simple Example – Multiple Regression:**
    - $\mathbf{Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \epsilon_t}$
    - here:
      - $\mathbf{Y_t}$ = **target variable at time t**
      - $\mathbf{X_{i,t}}$ = **external predictors (covariates)**
      - $\boldsymbol{\beta_i}$ = **coefficients**

2.2 Classes of Forecasting Models.

- **Mulitvariate Time Series Models:**
  - Advanced Example – Multivariate AR Models (VAR or ARIMAX):
    - Here, past values of both the target and predictors are used:
      - $\mathbf{Y_t = f\left(Y_{t-1}, \dots, Y_{t-p}, X_{t-1}, \dots, X_{t-q}\right) + \epsilon_t}$
        - p = lags of the target variable
        - q = lags of the predictors
  - Example:
    - Forecasting **electricity demand** using:
      - **Past electricity demand (autoregressive part)**
      - **Temperature, day of week, holidays (external covariates)**
  - Key feature: **Combines self – history with influencing factors.**

2.3 Stationarity in Time Series Model.

- **What is Stationarity?**
  - In simple terms, a **stationary time series is one whose statistical properties do not change over time**.
    - Think of it like this:
      - A stationary process is in a state of **"statistical equilibrium".**
      - Its behavior is consistent, **making it predictable.**
  - **Example:**
    - If you were to take a "snapshot" of any chunk of the series,
      - it would look roughly like any other chunk.
    - A time series is said to be **strictly stationary if the entire distribution of its values**
      - is **constant over time**.
        - This is very strong condition and often impractical to check.
    - Therefore,
      - we usually work with a **weaker but more practical form called weak stationarity** (or covariance stationarity).

2.3.1 Weak – Stationarity in Time Series Model.

- A **weakly stationary** series must satisfy **three conditions:**
  1. **Constant Mean**: The average value of the series remains constant over time.
     - Non – stationary example: A series with an upward trend has a mean that increases over time.
     - Stationary example: A series that fluctuates around a fixed horizontal line.
  2. **Constant Variance:** The volatility or spread of the data points around the mean does not change over time. This is also known as homoscedasticity.
     - Non stationary example: A series where the fluctuations become wider and more volatile as time goes on.
     - Stationary example: The "wobbliness" of the series remains consistent throughout.
  3. **Constant Autocovariance:** The relationship (covariance) between values at two time points depends only on the *distance* or *lag* between those two points and not on the actual time at which they are observed.
     - *Simplified:* The relationship between today's value and yesterday's value is the same as the relationship between the value on June 10th and June 9th, or any other two consecutive days.
- **Q: How to make Time – Series Data Stationary or weak stationary?**

2.3.1 Weak – Stationarity in Time Series Model.

- A **weakly stationary** series must satisfy **three conditions:**
  1. **Constant Mean**: The average value of the series remains constant over time.
     - Non – stationary example: A series with an upward trend has a mean that increases over time.
     - Stationary example: A series that fluctuates around a fixed horizontal line.
  2. **Constant Variance:** The volatility or spread of the data points around the mean does not change over time. This is also known as homoscedasticity.
     - Non stationary example: A series where the fluctuations become wider and more volatile as time goes on.
     - Stationary example: The "wobbliness" of the series remains consistent throughout.
  3. **Constant Autocovariance:** The relationship (covariance) between values at two time points depends only on the *distance* or *lag* between those two points and not on the actual time at which they are observed.
     - *Simplified:* The relationship between today's value and yesterday's value is the same as the relationship between the value on June 10th and June 9th, or any other two consecutive days.
- **Q: How to make Time – Series Data Stationary or weak stationary?**
  - **Smoothing , Detrending & Deseasonalization**

2.4 Smoothing.

- **Smoothing** is the process of reducing short-term fluctuations (noise) in a time series to reveal the underlying long-term pattern (trend or seasonality).
- **Finite Moving Average (MA) Smoothing:**
  - Finite Moving Average (MA) smoothing is a technique to **remove random noise** in a time series by replacing each observation with the **average of its neighbors** within a fixed window.
  - Mathematically, for a time series $\{y_t\}$, the centered moving average with window size m is:
    - $\tilde{y}_t = \frac{1}{m}\sum_{i=-k}^{k} y_{t+i}$
    - where:
      - $m = 2k + 1$ (odd window size, centered at t)
      - $k =$ number of points before or after t included.
  - If we don't have symmetry we use trailing moving average:
      - $\tilde{y}_t = \frac{1}{m}\sum_{i=0}^{m-1} y_{t-i}$

For Further please follow the

{Notebook.}