Data Science Certification

Module 9

Time Series Analysis

Agenda

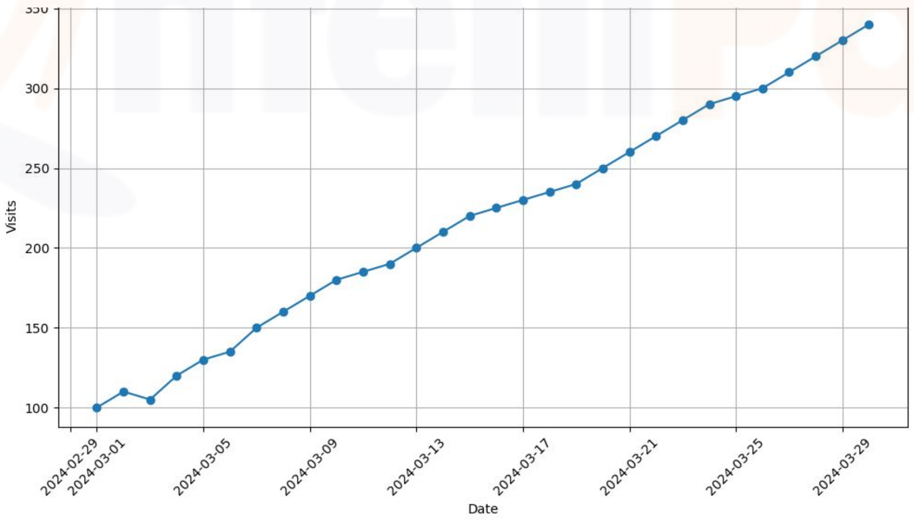
1. Introduction to Time Series
2. Importance of Time Series
3. Components of Time Series
4. Understanding Stationarity
5. ML Models for Time Series Analysis
6. Optimizing Models

Introduction to Time Series

Time Series Forecasting

Time series forecasting is the process of using historical data to make **predictions about future values of a time-dependent variable**. It involves analyzing the **patterns and trends** within the time series data to make informed predictions for the future.

For example, let's say you have data on how many people visit a website each day. Time series forecasting would involve studying the patterns in that data - like more visits during weekends or holidays - to make educated guesses about how many people might visit the website in the coming days or weeks.



Time Series Data

Time series data is a collection of observations obtained or recorded over a period of time.

It represents the evolution of a particular variable over a period of time.

The time intervals in this data are equally spaced.

Importance of Time Series

Importance of Time Series

Financial Planning and Budgeting

To forecast financial metrics, plan budgets, and make strategic financial decisions. Using time series analysis, we forecasted next year's sales revenue to optimize budget allocation and strategic financial decision-making.

Marketing and Sales Planning

To plan marketing strategies, optimize inventory, and forecast sales. By leveraging time series analysis, we optimized inventory levels and forecasted sales, enabling strategic marketing decisions for increased market share.

Importance of Time Series

Healthcare Planning

Through time series analysis, we predicted disease outbreaks, forecasted patient admission rates, and optimized resource planning for hospitals, ensuring efficient healthcare delivery.

Environmental Monitoring

Utilizing time series analysis, we monitored environmental variables like temperature, precipitation, and pollution levels to inform sustainable resource management and mitigate environmental impacts.

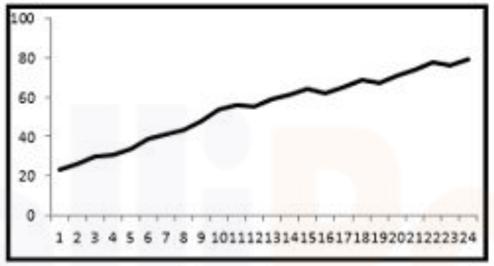
Components of Time Series

Components of Time Series

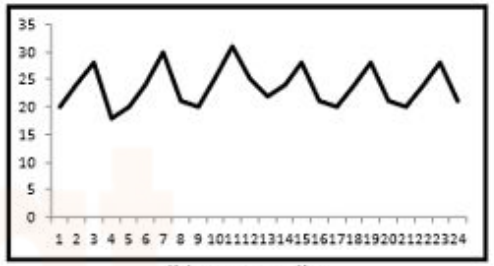
The components of a Time Series **affect the overall pattern and behavior of the data.**

Components of a Time Series consists of:

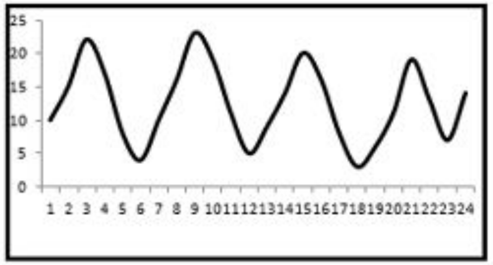
1. **Trend**
2. **Seasonality**
3. **Cyclic pattern**
4. **Noise**



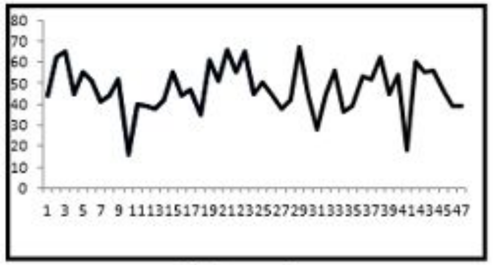
(a) Trend



(b) Seasonality



(c) Cyclicality

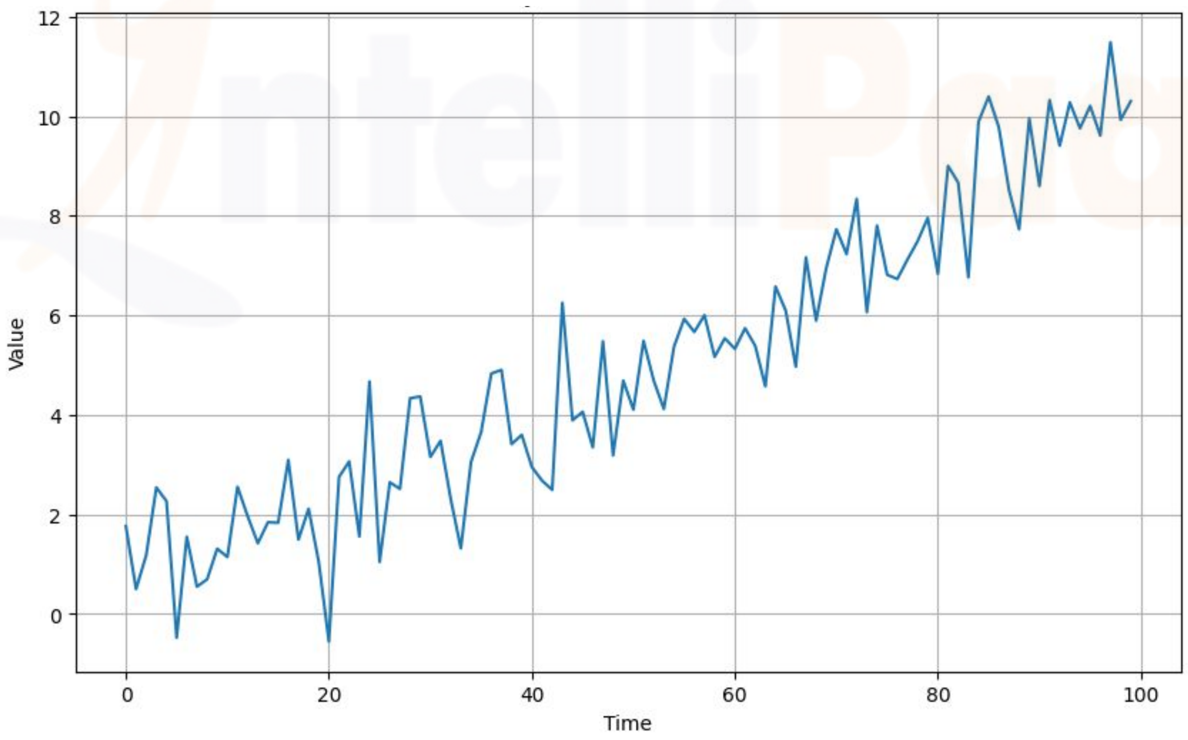


(d) Irregular

Trend

- Trend is the long-term **movement in the data** that shows whether it's increasing, decreasing, or remaining constant over an extended period of time.
- The data having trend will have **varying standard mean.**

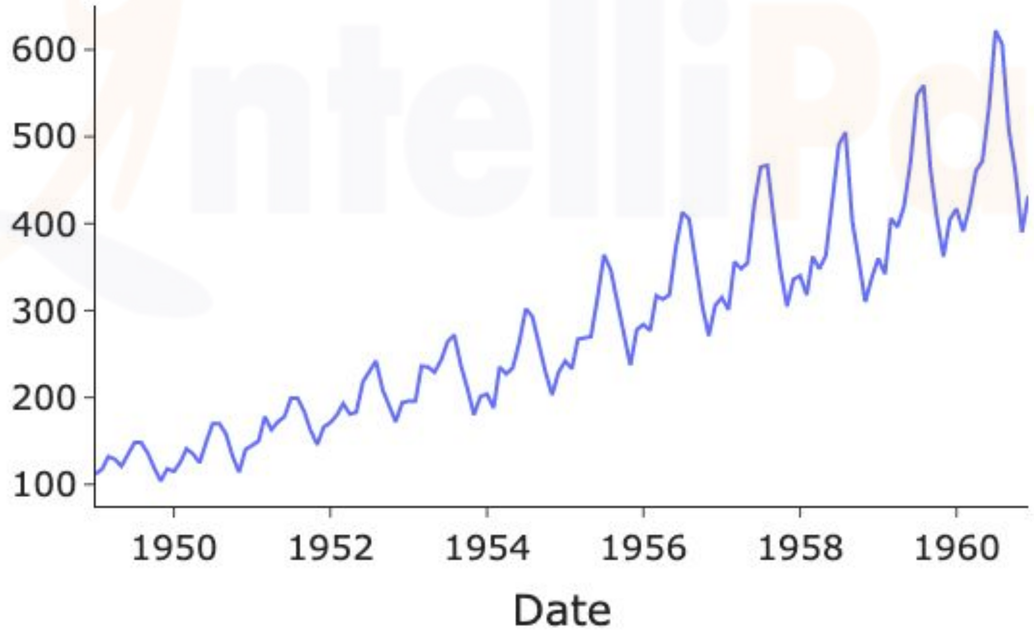
  For example, a business experiencing a steady increase in sales over several years.



Seasonality

- Seasonality is the **predictable patterns within the data** that occur at regular intervals, such as daily, weekly, monthly, or annually.

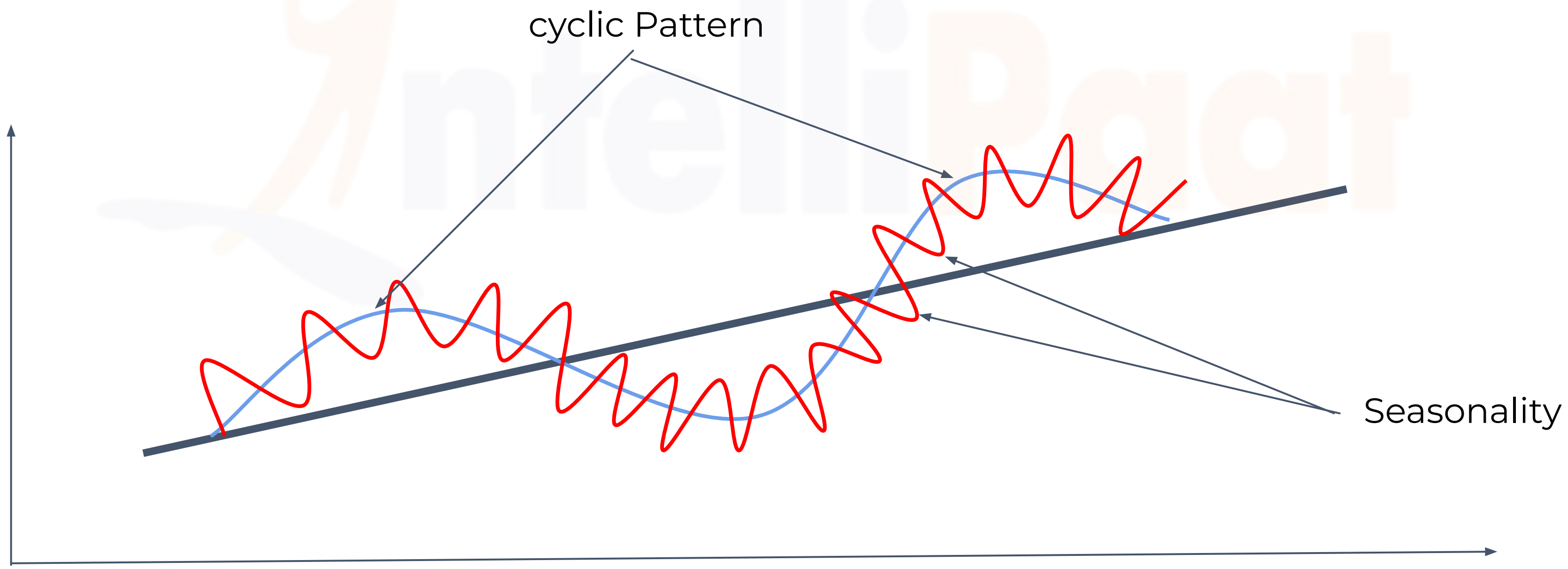
  For example, the sale of umbrellas increases during the rainy season, and the sale of firecrackers increase during Diwali.



Cyclic Pattern

- A cyclic pattern is a **fluctuation in a time series that does not have a fixed or predictable time interval.**
- Seasonality on the other hand is a fluctuation that repeats regularly such as monthly or quarterly. But cycles often **lack regularity in terms of frequency, duration and amplitude.**

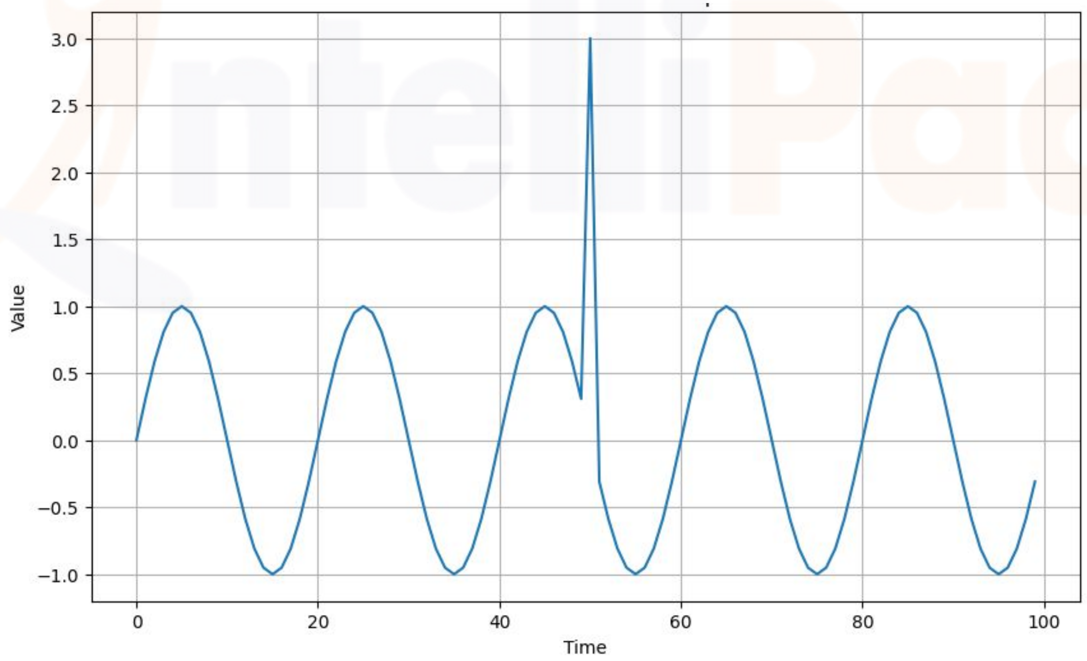
For example, the real estate market may experience sudden cycles of boom and bust over a span of several years.



Noise

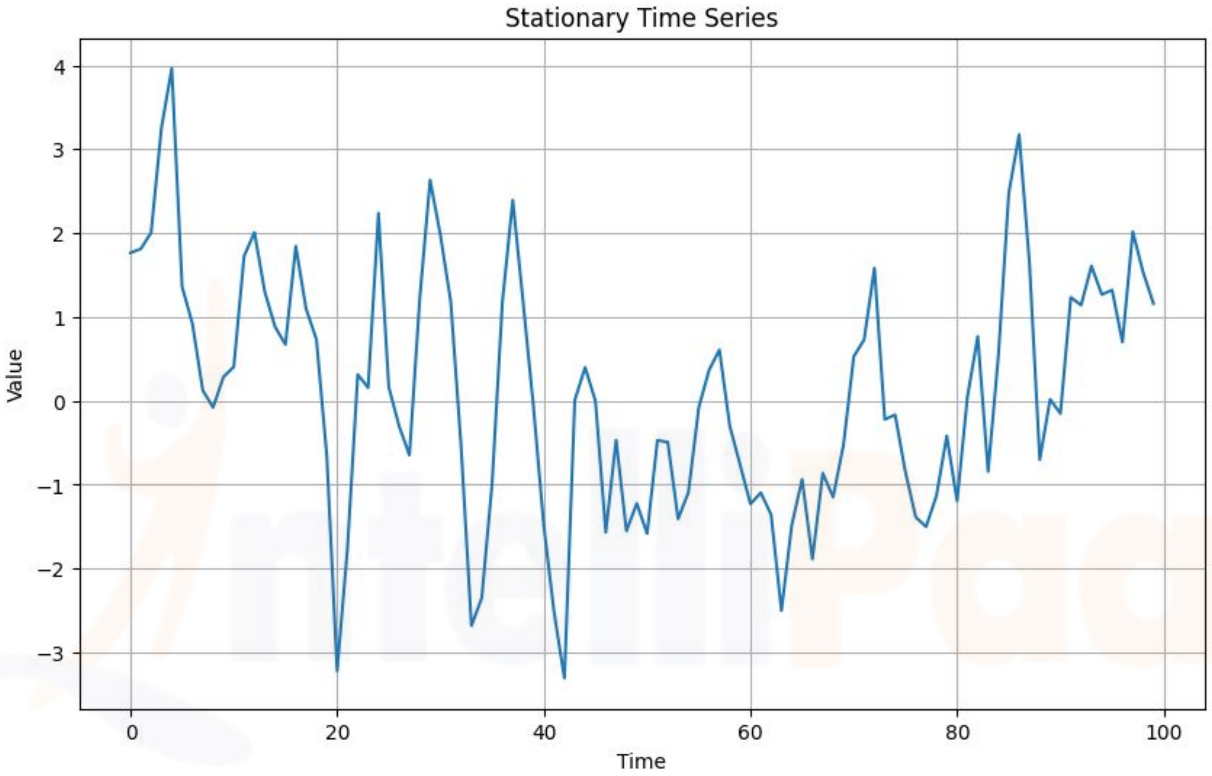
- Noise refers to the **random and unpredictable irregularities** present in the data.
- Noise is characterized by its **lack of pattern.** It appears as random fluctuations around the components of the time series data.

  For example, sudden spikes or drop in sales due to unforeseen events like natural disasters.



Understanding Stationarity

Introduction to Stationarity



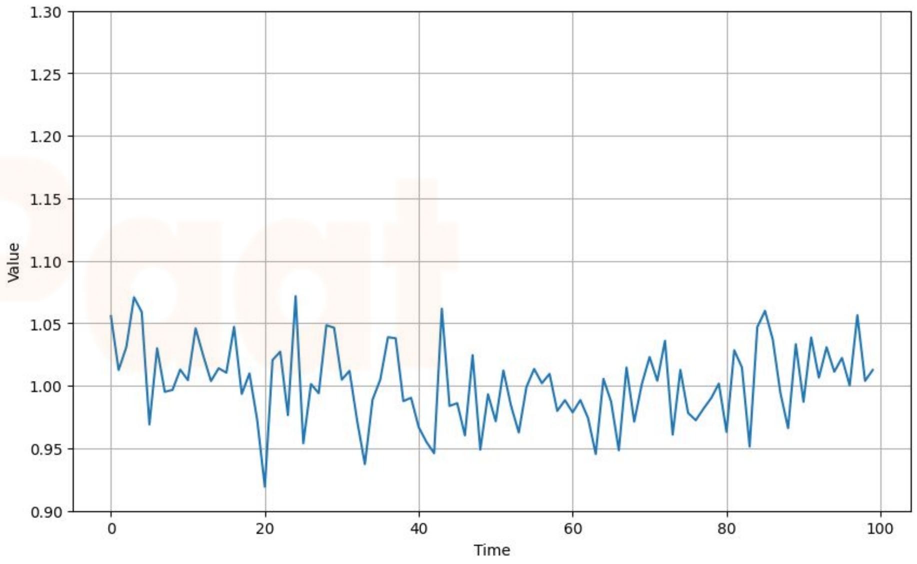
- Stationarity is a property of Time Series data.
- A stationary time series is one whose **mean and variance do not change with time.**
- Understanding stationarity is **crucial for accurate modeling and forecasting in time series analysis.**

Introduction to Stationarity

Constant Mean

- In a stationary time series, the mean value of the data points remain constant across different time intervals, which also implies that the trend is removed. This implies that the series does not exhibit a upward or downward trend movement over time.

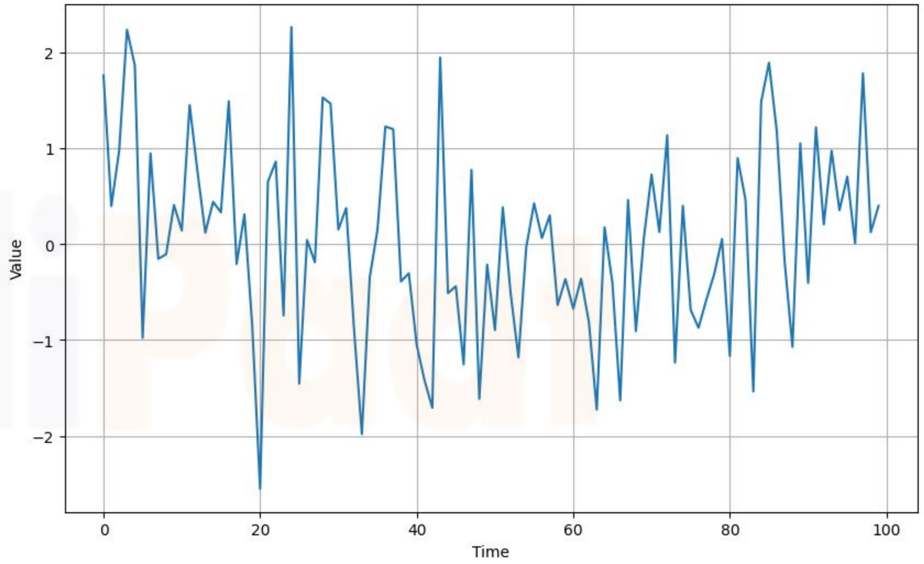
- When we talk about a "constant mean" in a stationary time series, it's like saying the average value of our data doesn't change as we move through different time periods. Imagine you're tracking something over time, and on average, it stays pretty much the same without showing a clear trend going up or down.



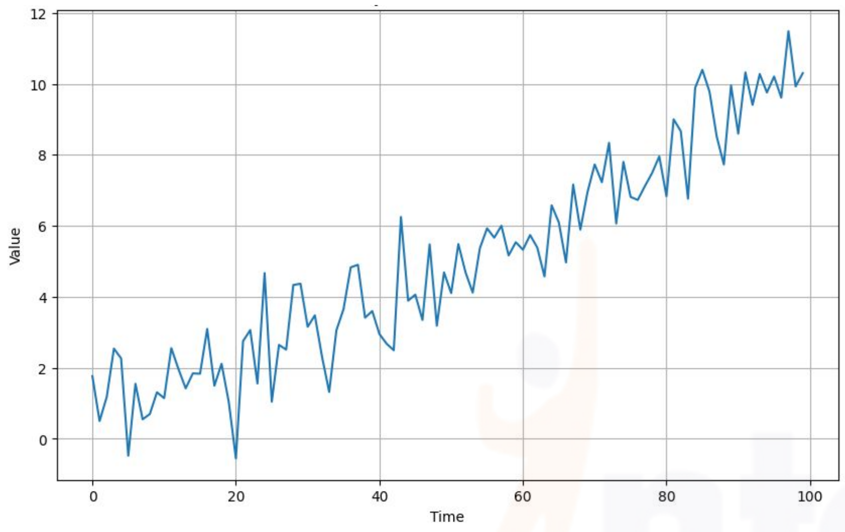
Introduction to Stationarity

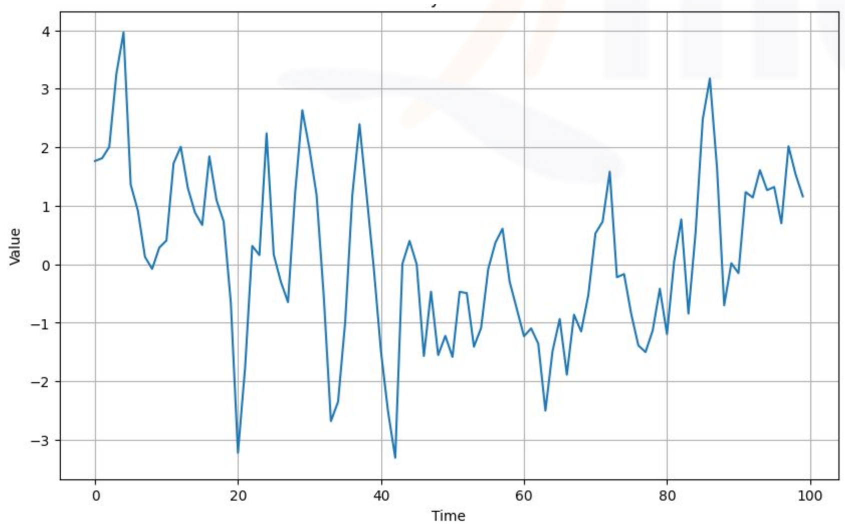
Constant Variance

- ❑ In a stationary time series, the variance of the data points remain constant across time. This implies that the spread of the values does not change when we mention "constant variance" in a stationary time series, which also implies that the seasonality is removed.
- ❑ It means that the range or how spread out our data is stays consistent over time. Picture the data points kind of staying within a certain range without the spread getting wider or narrower.



Stationary vs. Non-stationary Time Series





A non-stationary time series does not have a constant mean
or a constant variance.

A stationary time series has a constant mean
and a constant variance.

Achieving Stationarity

Differencing

Subtract each observation from its preceding one. This helps remove trends and seasonality.

Transformations

Apply mathematical transformations, such as log transformations. This stabilizes variance.

Check and Repeat

Visualize the data, examine rolling statistics, and conduct the Augmented Dickey-Fuller (ADF) test to check stationarity.

ML Models
for Time Series Analysis

ML Models for Time Series Analysis

ARIMA

ARIMA is a widely used time series forecasting model that combines autoregression (AR), differencing (I), and moving averages (MA). It is effective for capturing linear trends and seasonality.

Auto ARIMA

Auto ARIMA is an automated version of the ARIMA model. It determines the optimal values for the ARIMA parameters (p, d, q) by searching through various combinations.

ML Models for Time Series Analysis

SARIMA

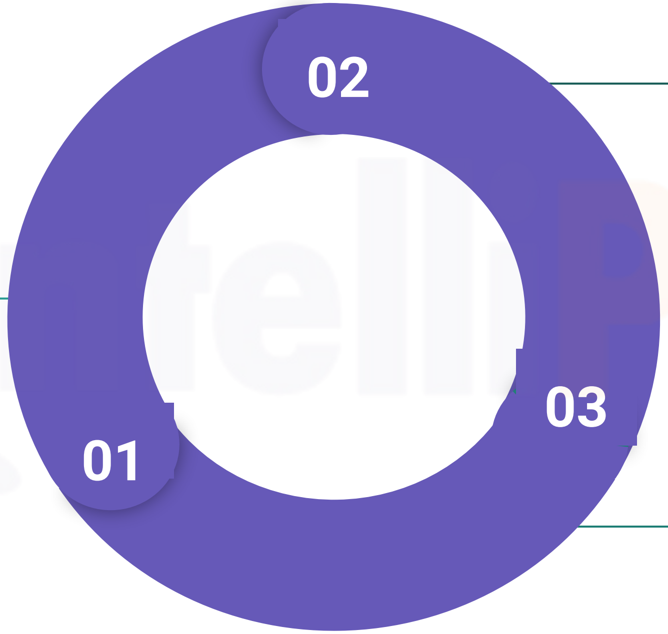
SARIMA extends the ARIMA model to handle seasonality in time series data. It includes additional parameter (s) to account for seasonal components. Here 's' represents the length of the season. SARIMA is effective for forecasting data with recurring patterns at regular intervals.

Optimizing Models

Optimizing Time Series Models: A Closer Look at p, d, and q

p (Autoregressive Order)

It represents the number of lag observations included in the model. It indicates the number of previous time steps to consider when predicting the current time step. A higher p value means the model considers more historical data.



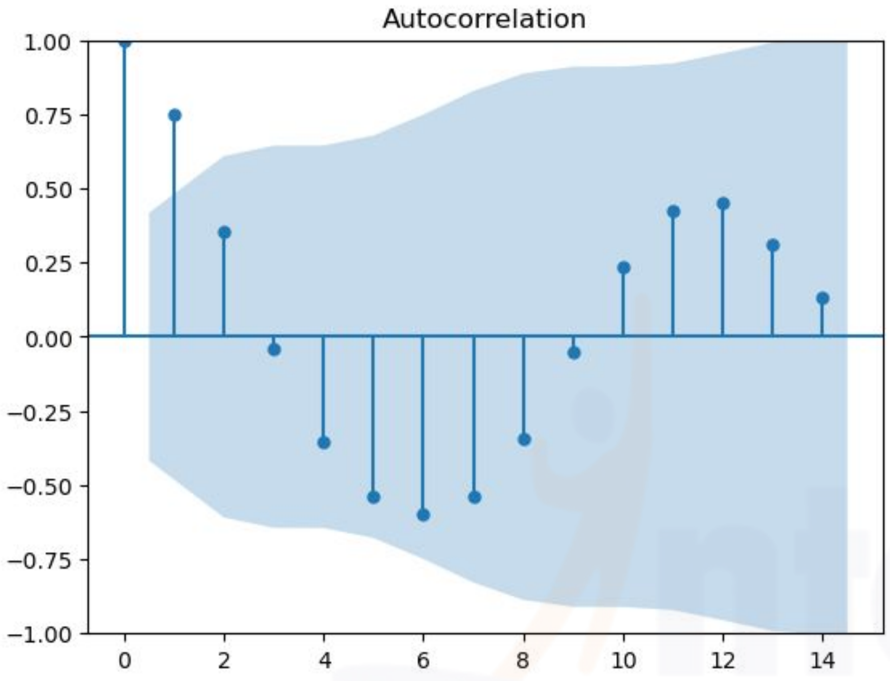
d (Integrated Order)

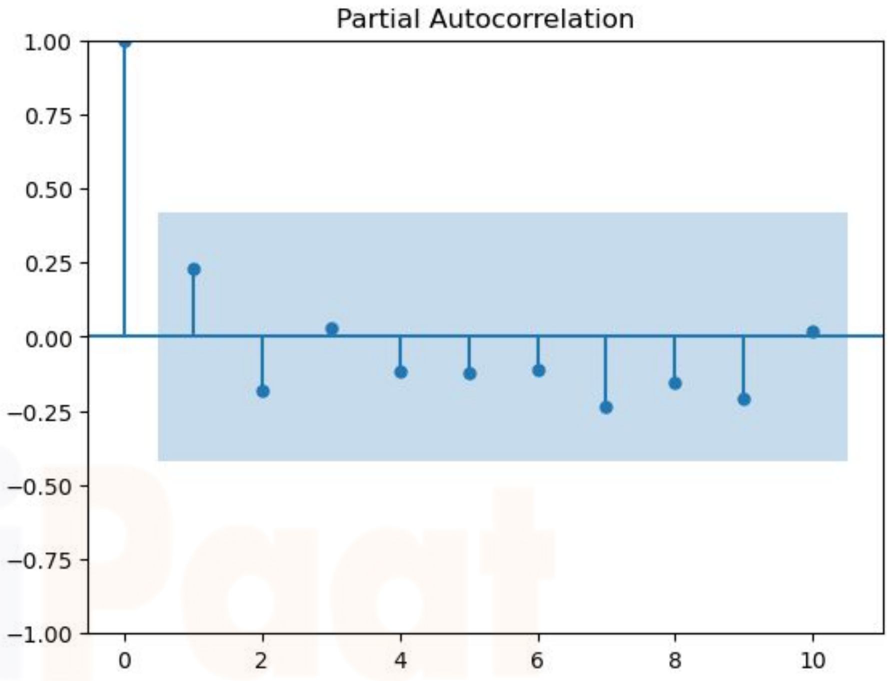
It represents the number of differences needed to make the time series data stationary. Stationarity is a property of time series data where statistical properties, such as mean and variance, do not change over time.

q (Moving Average Order)

The moving average order (q) represents the number of lagged forecast errors included in the model. It indicates the influence of past forecast errors on the current time step. A higher q value means the model considers more historical forecast errors.

Visual Inspection Using ACF and PACF





1. Identify the q where the ACF plot shows a significant spike. This is the moving average order.
2. Identify the p where the PACF plot shows a significant spike. This is the autoregressive order.
3. It should be taken into account that ACF and PACF only provides us an estimate and not a exact value.

Hands-On:

Airline Passenger Forecasting Hands-On

Contact Us

080-4524-9465

support@intellipaat.com