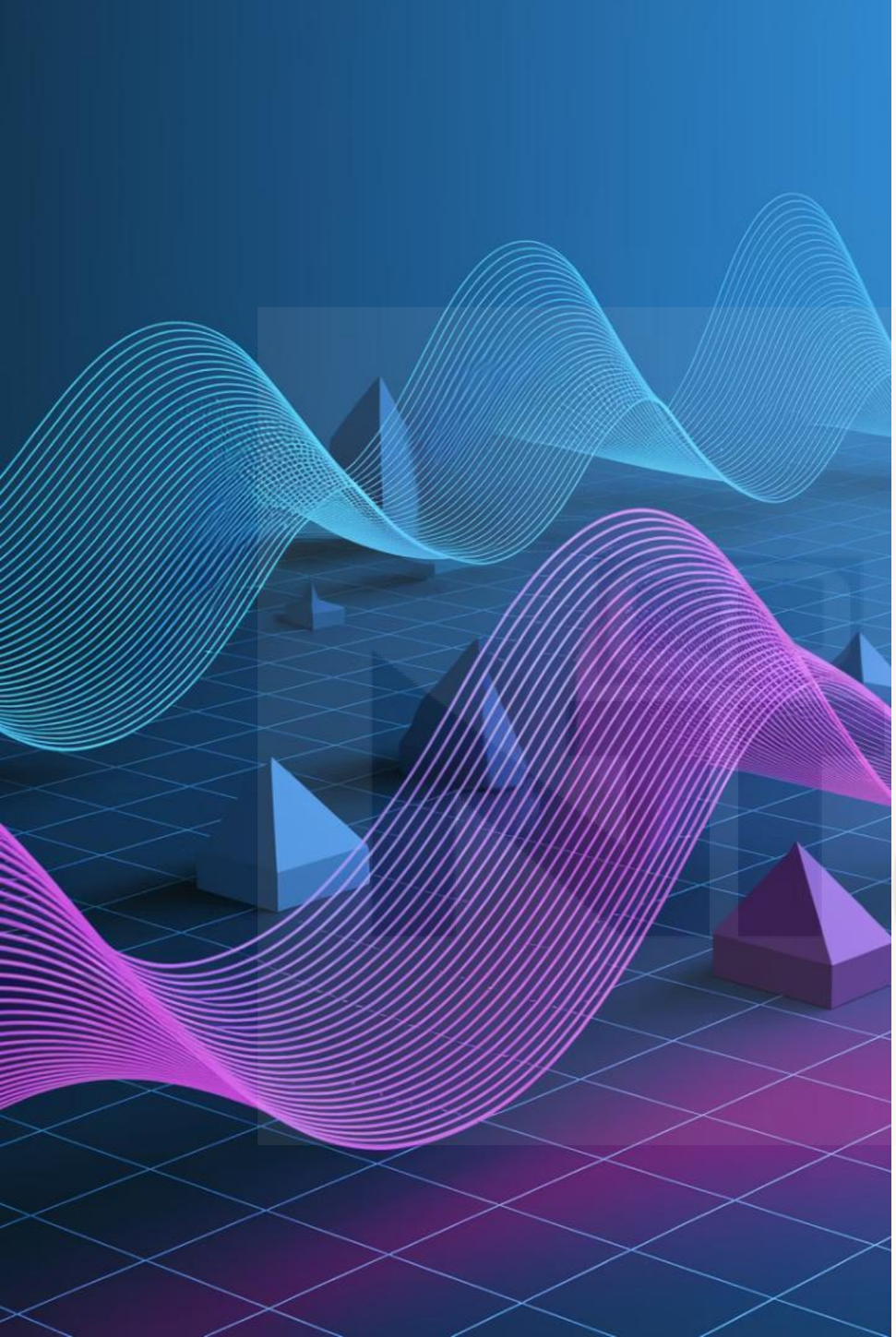


ARIMA: Mastering Time Series Analysis

MUKESH KUMAR

What is ARIMA?

Autoregressive

Predicts future values based on past values in the time series.

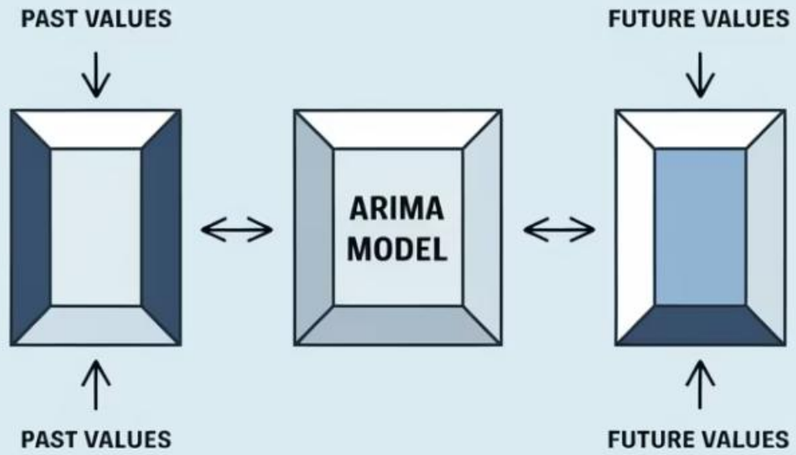
Integrated

Makes the time series stationary through differencing.

Moving Average

Uses past forecast errors to improve future predictions.

ARIMA = AR + I + MA



Autoregression Explained

What It Is

Autoregression predicts future values based on past values in the time series.

It assumes recent events have a stronger impact on the future than older events.

Example

To predict tomorrow's temperature, we look at today's temperature and previous days.

Recent temperatures have more influence on the prediction than temperatures from weeks ago.

Making Data Stationary

What is Stationarity?

A stationary time series has statistical properties that don't change over time.

The mean and variance remain constant throughout the series.

Why It Matters

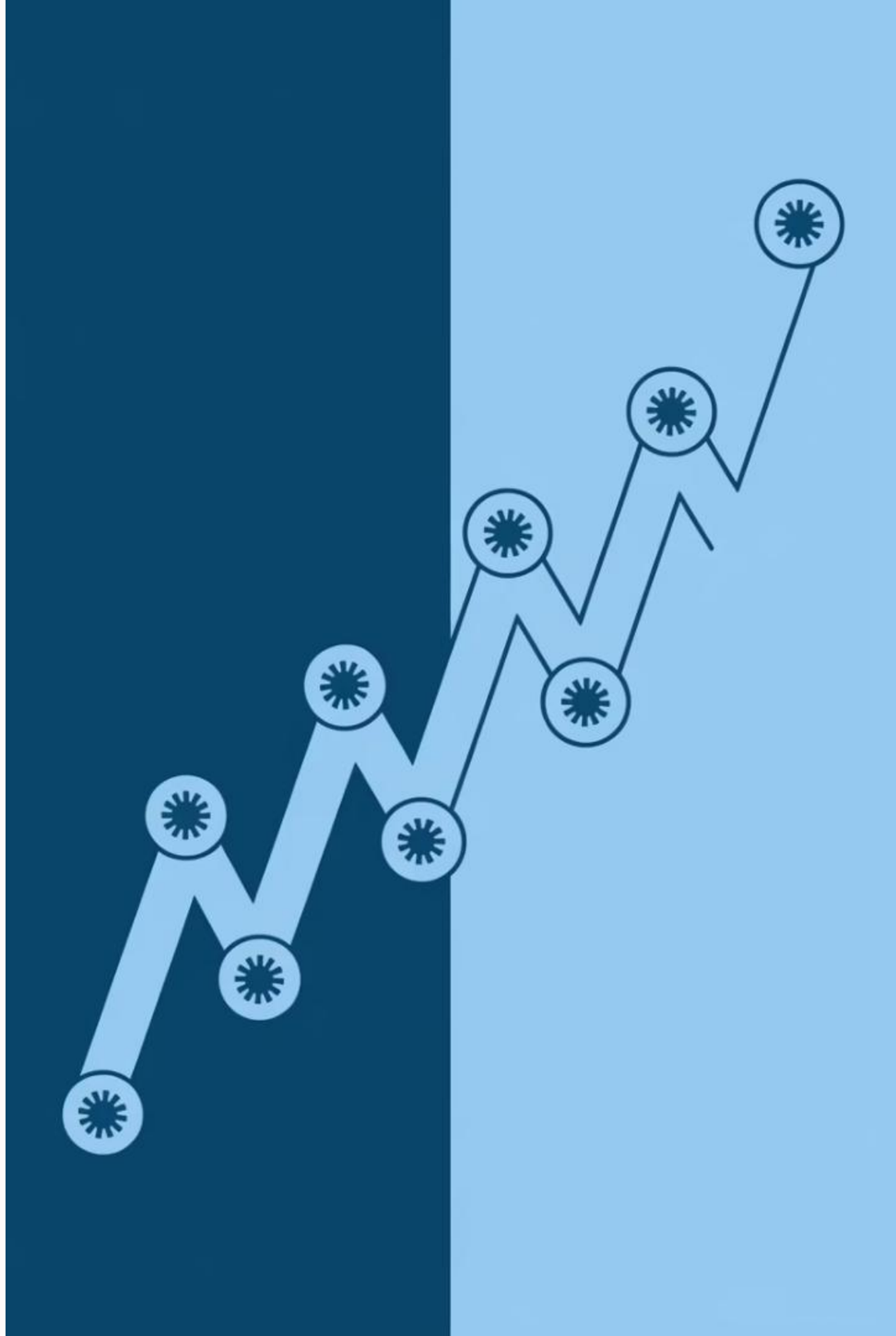
ARIMA assumes patterns in the past will continue into the future.

Non-stationary data makes it hard to find stable patterns for accurate predictions.

The "I" in ARIMA

The Integrated component makes data stationary through differencing.

It transforms the data to focus on changes rather than absolute values.



Differencing Technique

Original Data

Start with your raw time series data, which may have trends or seasonality.

First Differencing

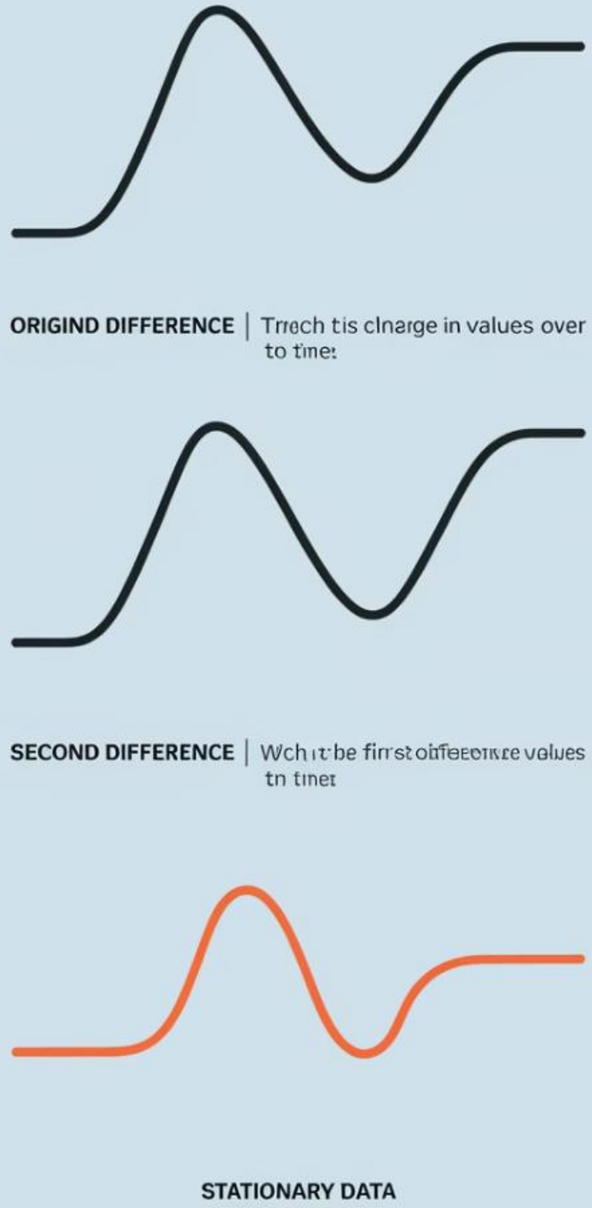
Calculate the change between consecutive observations.

Second Differencing

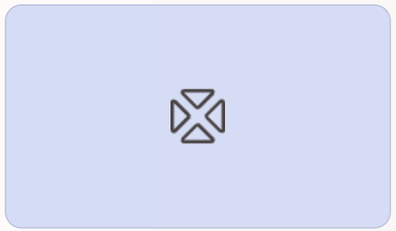
If needed, take the difference of the differences.

Stationary Result

The differenced series should now have constant statistical properties.

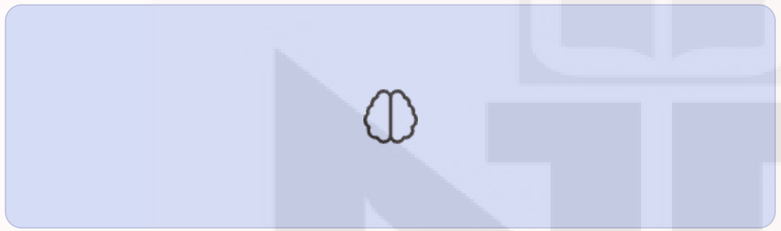


Moving Average Component



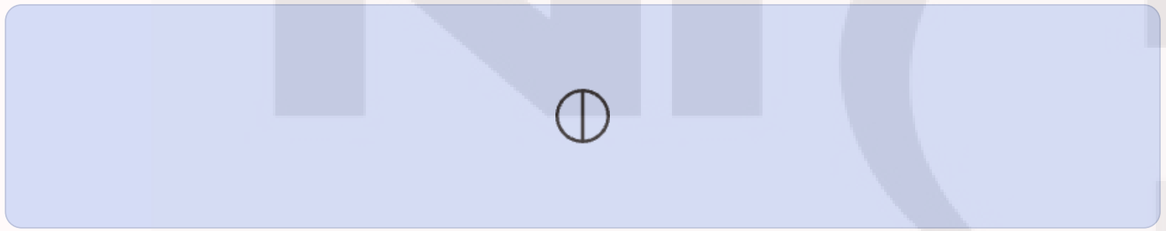
Identify Forecast Errors

Track where predictions missed actual values.



Learn From Mistakes

Recognize patterns in these errors.



Adjust Future Predictions

Correct forecasts based on past error patterns.

The MA component helps the model improve over time by learning from its mistakes.

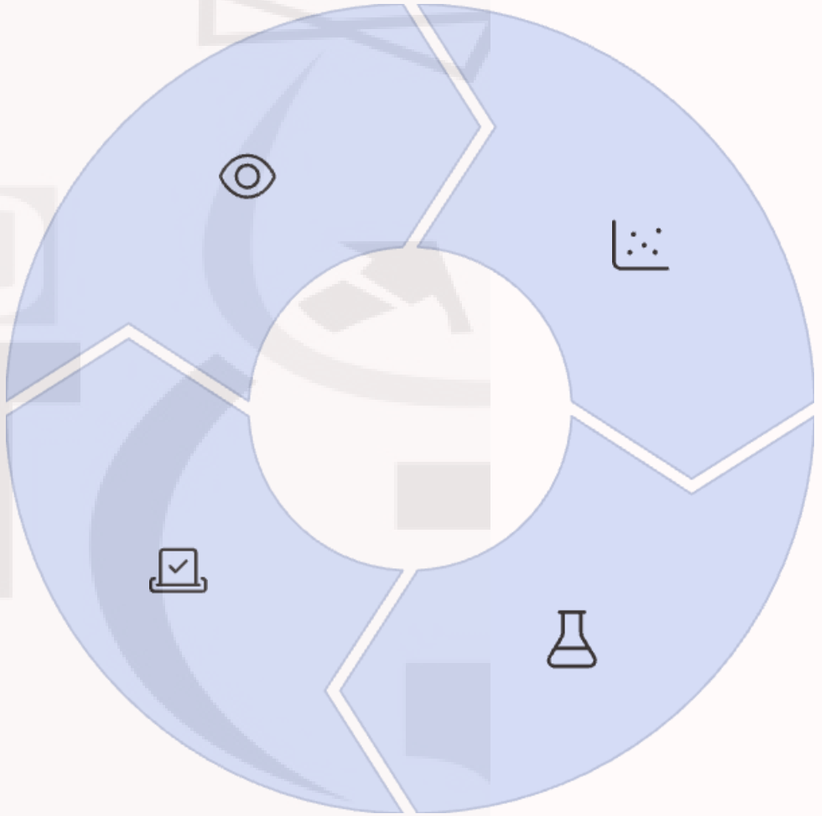
Checking for Stationarity

Visual Inspection

Plot the data and look for obvious trends or changing variance.

Rolling Statistics

Calculate mean and standard deviation in sliding windows over time.



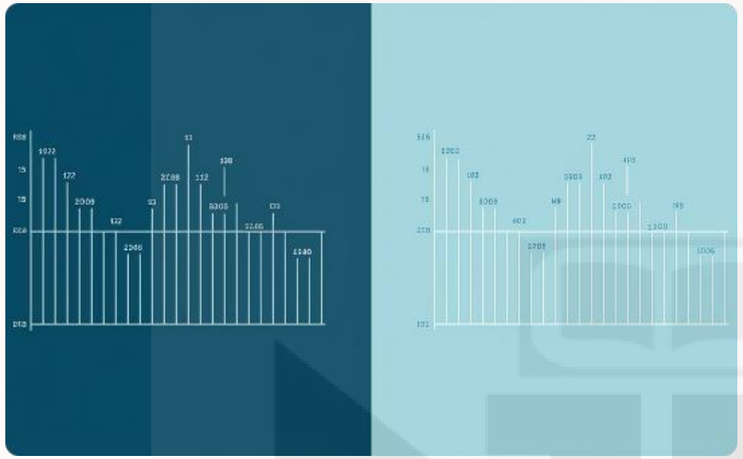
Interpretation

A p-value below 0.05 suggests the series is stationary.

Statistical Tests

Use the Augmented Dickey-Fuller test to check for stationarity.

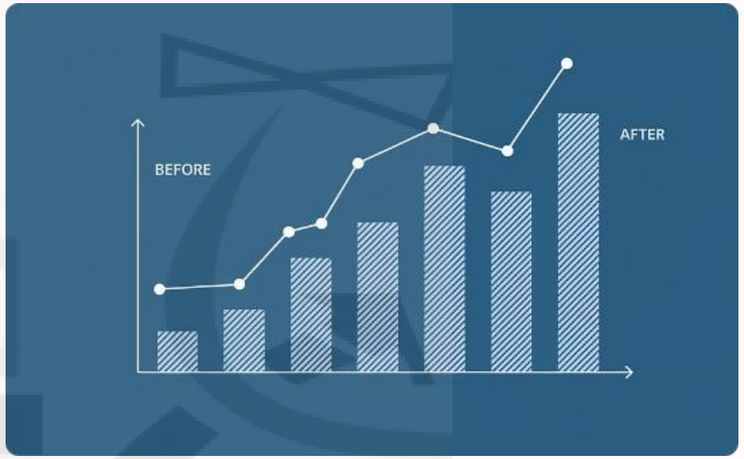
Making Non-Stationary Data Stationary



Logarithmic Transformation

Helps stabilize variance when variability increases with the level of the series.

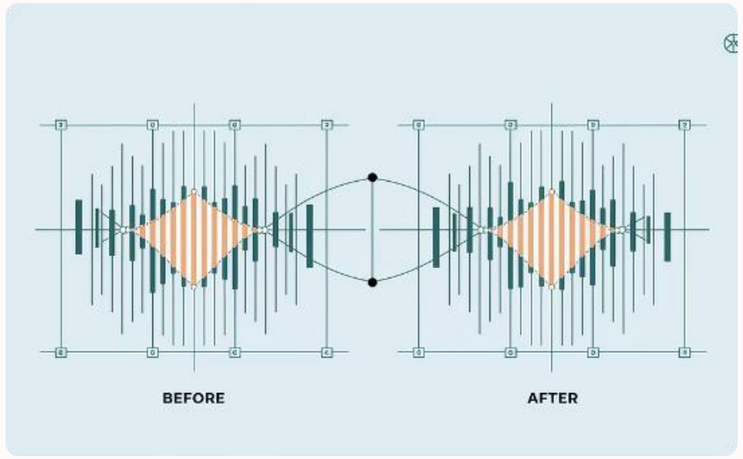
Useful for exponentially growing data like user counts.



Differencing

Focuses on changes rather than absolute values.

Effective at removing trends from the data.



Combined Approach

Sometimes both transformations are needed for complex data.

Remember to reverse transformations when interpreting predictions.

Determining ARIMA Parameters

ACF Plot

Shows correlation between time series and its past values at different lags.

Helps determine the Moving Average (q) parameter.

PACF Plot

Shows direct correlation after removing intermediate lag influences.

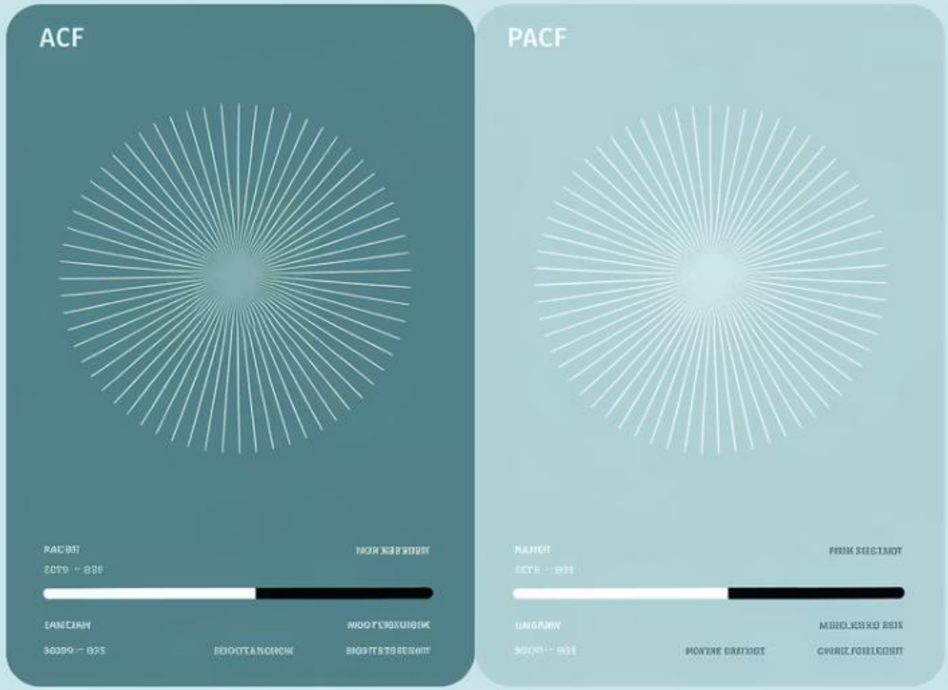
Helps determine the Autoregressive (p) parameter.

Differencing Order

The number of times you differenced the data equals d.

Auto ARIMA

Automates parameter selection by testing multiple combinations.



Implementing ARIMA in Python

Import Libraries

Use StatsModels library which includes ARIMA functionality.

Import pandas for data handling and matplotlib for visualization.

Fit the Model

Pass your time series data and p, d, q values to the ARIMA function.

The function estimates model parameters automatically.

Make Predictions

Use the fitted model to forecast future values.

Remember to reverse any transformations applied to the data.

Evaluating ARIMA Performance

Visual Comparison

Plot predictions against actual values.

Check how closely they align on the graph.

Look for patterns in where the model performs well or poorly.

Root Mean Squared Error (RMSE)

Measures the average difference between predictions and actual values.

Lower RMSE indicates more accurate predictions.

Compare RMSE across different model configurations.

Key Takeaways

Stationarity is Crucial

Always check if your data is stationary before applying ARIMA.

Parameter Selection Matters

Use ACF/PACF plots or auto_arima to find optimal p, d, q values.

No One-Size-Fits-All

Experiment with different methods to find what works best for your data.

Remember Transformations

Always reverse transformations to interpret predictions correctly.