**NAME**- AVROJIT DUTTA

**ROLL NO- 25/AIML-A6/NOV-8047**

**BATCH- AIML with Python Program | InternsElite**

**November/December 2025**

# Time Series Analysis and Sales Forecasting Using ARIMA

***Colab Link:***

https://colab.research.google.com/drive/1sELyqwGV3aBB8sk_Kf8j--_oApsm5LdI?usp=sharing

## Introduction

Time series analysis plays a crucial role in understanding patterns in data that are collected over regular time intervals. Many real-world datasets such as sales records, stock prices, weather data, and energy consumption exhibit time-dependent behavior. Accurate forecasting of such data helps organizations in planning, decision-making, and resource allocation.

Sales forecasting is one of the most common applications of time series analysis. By analyzing historical sales data, future trends can be estimated, allowing businesses to anticipate demand and plan accordingly. Unlike traditional machine learning models, time series models explicitly consider the temporal ordering of data.

In this project, time series analysis techniques are applied to forecast future sales values using historical monthly data. The ARIMA (AutoRegressive Integrated Moving Average) model is used due to its effectiveness in handling trend and seasonality in time-dependent data.

## Problem Statement

The objective of this project is to forecast future sales values based on historical time series data.

The problem can be defined as:

*Given historical monthly sales data, analyze its trend and seasonality, check for stationarity, and build a forecasting model that predicts future sales values accurately.*

The project involves visualizing the time series, decomposing it into components, testing for stationarity, building an ARIMA model, and evaluating forecasting accuracy using standard error metrics.

## Dataset Description

The dataset used in this project represents **monthly airline passenger counts**, which is a widely used benchmark dataset for time series forecasting tasks. In this project, the dataset is treated as **monthly sales data**.

The dataset contains:

- A time index representing months
- A numerical value representing sales (passenger count)

The data spans multiple years, making it suitable for identifying long-term trends and seasonal patterns. This dataset was chosen because it clearly demonstrates the key characteristics of time series data, such as trend and seasonality.

## Methodology

The following steps were followed to perform time series analysis and forecasting:

1. **Data Loading and Preprocessing**
   The dataset was loaded from an online source and converted into a time-indexed format. The time column was transformed into a datetime format and set as the index.
2. **Time Series Visualization**
   The sales data was plotted to visually inspect trends and seasonal patterns over time.
3. **Decomposition**
   The time series was decomposed into three components: trend, seasonality, and residuals. This helped in understanding the underlying structure of the data.
4. **Stationarity Check**
   The Augmented Dickey-Fuller (ADF) test was applied to check whether the time series is stationary. Since the data was non-stationary, differencing was applied.
5. **Model Building**
   An ARIMA model was built using appropriate parameters to capture the autoregressive, differencing, and moving average components of the time series.
6. **Forecasting and Evaluation**
   The trained model was used to forecast sales for the next 12 months. The predicted values were compared with actual values, and error metrics such as RMSE and MAE were calculated.

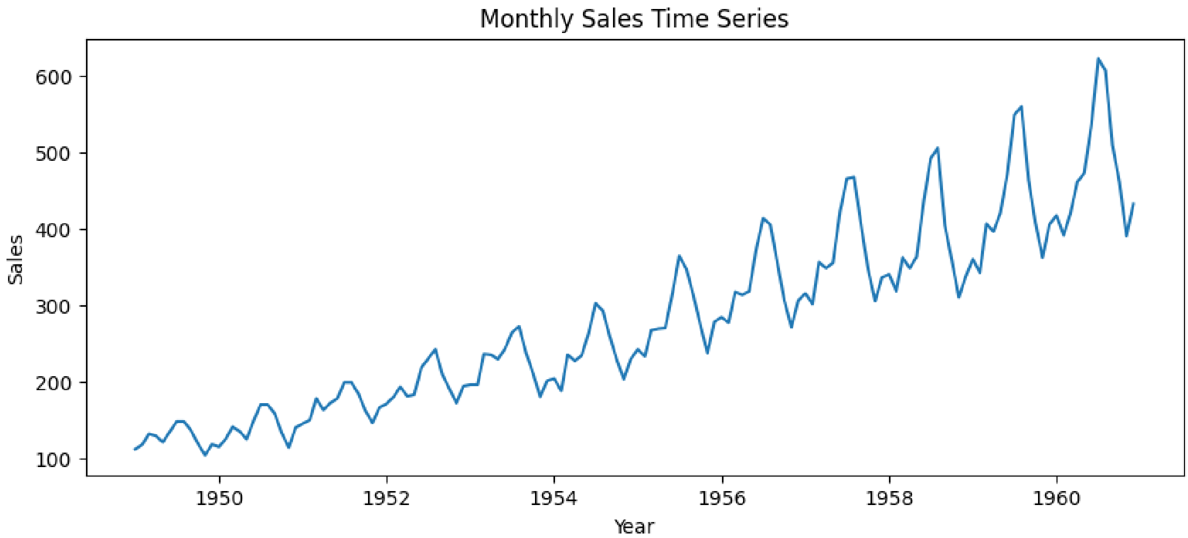
## Results and Discussion

The initial visualization of the time series showed a clear upward trend and recurring seasonal patterns, which justified the use of decomposition and ARIMA modeling. The decomposition results confirmed the presence of trend and seasonality in the data.

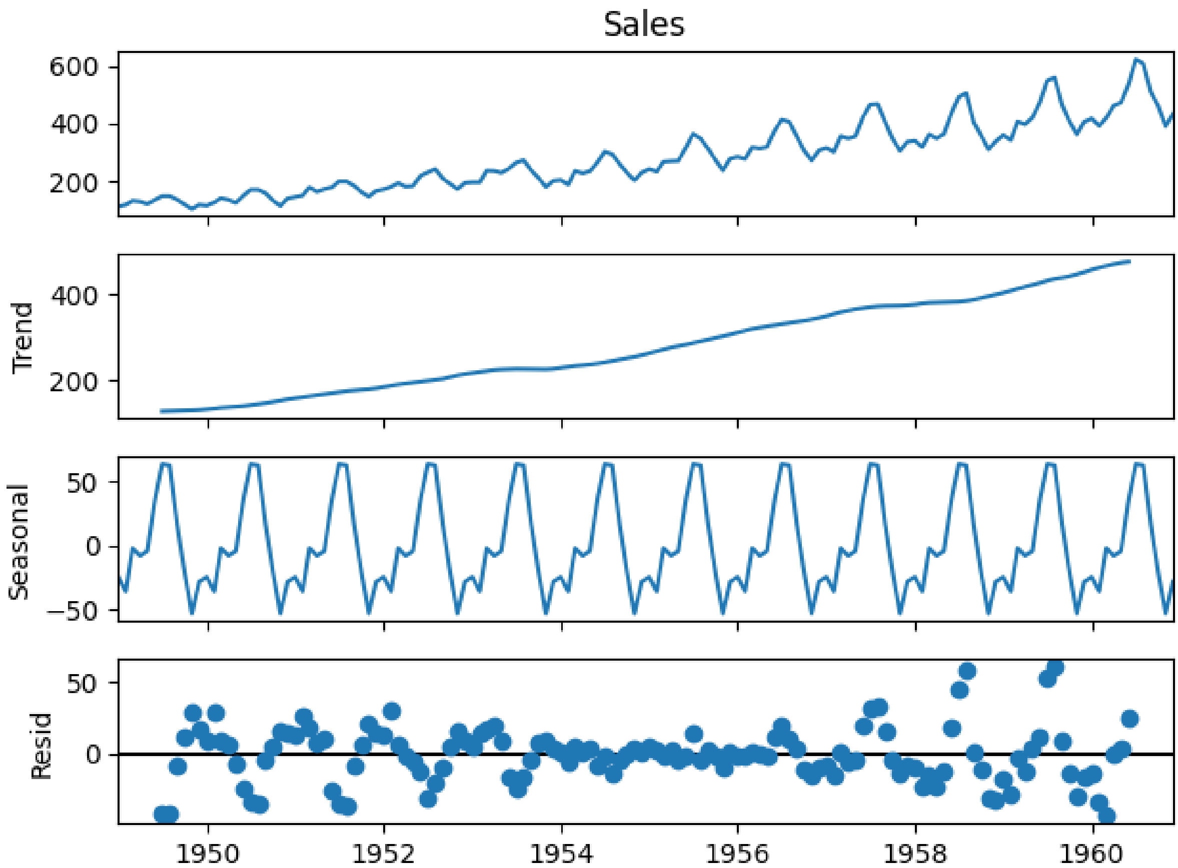
The ADF test indicated that the original time series was non-stationary. After applying differencing, the data became more suitable for ARIMA modeling.

The ARIMA model successfully captured the underlying patterns in the data and produced reasonable forecasts for future sales. The forecast plot showed a smooth continuation of the existing trend. Evaluation metrics such as RMSE and MAE indicated that the model performed adequately in predicting future values based on historical data.

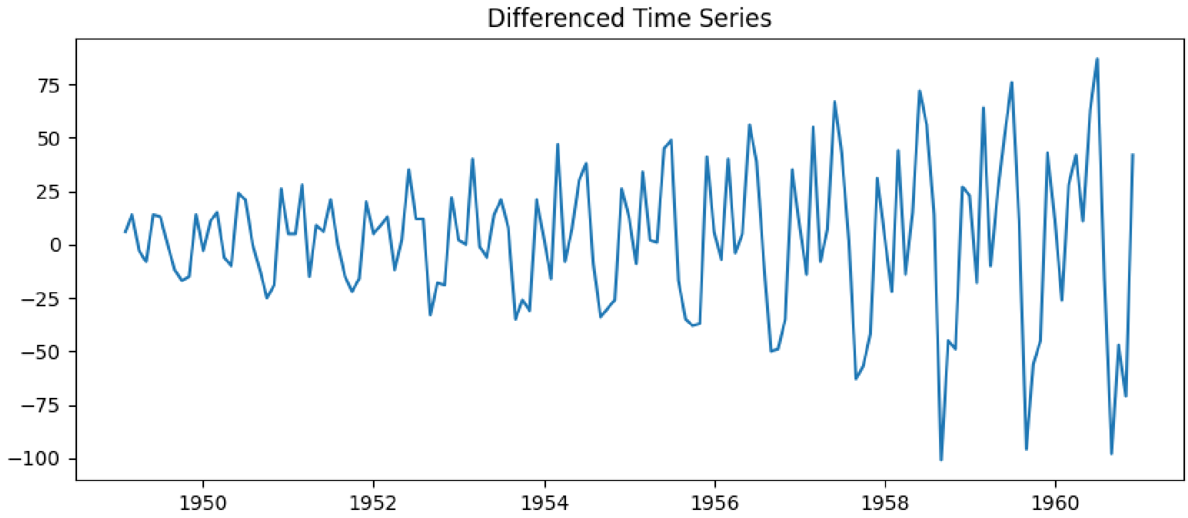
*Monthly Sales Time Series*



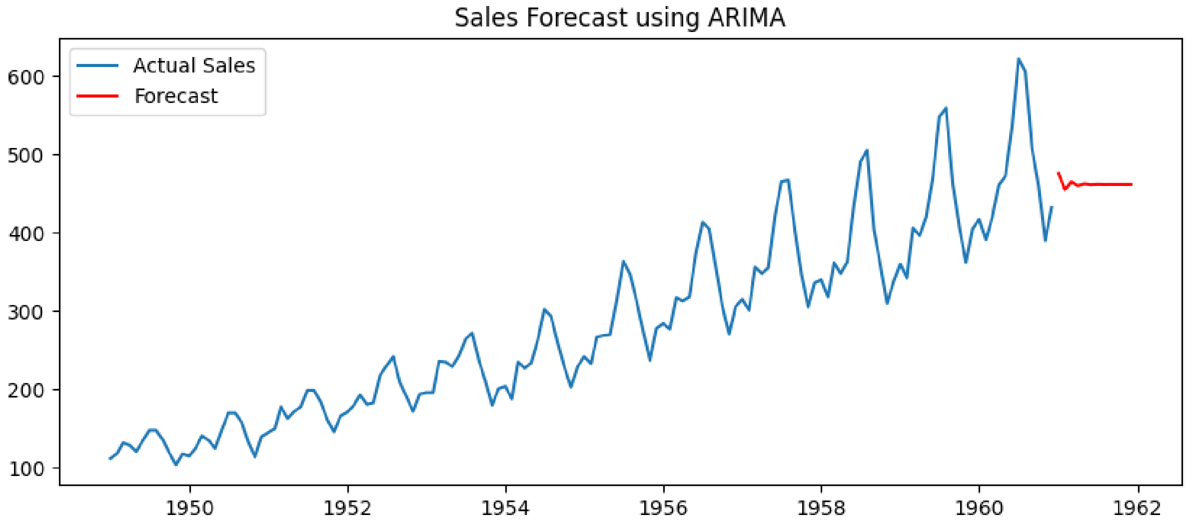

*Sales as per TREND/SALES/REIRD*




*Differenced Time Series*




*Sales Forecast using ARIMA*




## Conclusion

In this project, time series analysis techniques were applied to forecast monthly sales data using the ARIMA model. The study demonstrated the importance of understanding time dependency, stationarity, and seasonality before building forecasting models.

The ARIMA model proved to be effective in capturing temporal patterns and producing reliable forecasts. This approach can be extended to real-world applications such as retail sales forecasting, stock price prediction, and demand estimation.

Future work may include experimenting with advanced models such as SARIMA or LSTM-based neural networks and incorporating external factors to further improve forecasting accuracy.

## Reference

Brownlee, J. (2017). *Time Series Forecasting with the Airline Passengers Dataset*.