📈 Time Series Forecasting Assignment

Problem Statement:

This assignment aims to analyze and forecast stock price trends using **time series forecasting techniques**. By working with real stock data (Tesla and Apple) from Kaggle, students will perform **trend analysis**, **seasonality identification**, and build models such as **ARIMA and SARIMA** to predict future prices. The objective is to gain practical exposure to time series modeling and understand the statistical patterns underlying financial time series.

Guidelines:

1. Foundational Knowledge:

- Understand time series components: **trend**, **seasonality**, **residuals**.
- Learn about **additive vs multiplicative decomposition models**.
- Familiarize yourself with **Autocorrelation Function (ACF)** and **Partial Autocorrelation Function (PACF)**.
- Understand the modeling techniques: **ARIMA** and **SARIMA**, including parameter roles (`p`, `d`, `q`, `P`, `D`, `Q`, `s`).

2. Data Exploration:

- Select either **Tesla (TSLA)** and **Apple (AAPL)** stock dataset from Kaggle.
- Perform exploratory data analysis:
    - Check for missing data, outliers, and datatypes.
    - Plot closing prices over time.
    - Analyze trends using rolling mean and standard deviation.

3. Preprocessing and Decomposition:

- Convert date column to datetime format and set it as the index.
- Handle any null values or irregularities in the dataset.
- Use `seasonal_decompose` to decompose the time series.
- Analyze both **additive** and **multiplicative** models to determine which fits better.

4. ACF and PACF Analysis:

- Generate and interpret ACF and PACF plots using `plot_acf()` and `plot_pacf()` from `statsmodels`.
- Use the plots to identify suitable values for `p`, `d`, and `q` for the ARIMA model.

5. Forecasting Model Construction:

- Build the following models:
    - **ARIMA** using `statsmodels.tsa.arima.model.ARIMA`
    - **SARIMA** using `SARIMAX` for seasonality
- Train and validate the models using a train-test split.
- Forecast future prices and visualize them.

6. Model Evaluation:

- Evaluate predictions using:
    - **Root Mean Squared Error (RMSE)**
    - **Mean Absolute Error (MAE)**
    - **Mean Absolute Percentage Error (MAPE)**

- Plot actual vs predicted values to visually assess the model.
- Analyze residuals to validate model assumptions.

Step-by-Step Approach to Time Series Forecasting:

1. Setup and Data Preparation:

- Import libraries: `pandas`, `matplotlib`, `statsmodels`, `seaborn`, etc.
- Load the Kaggle dataset and parse the `Date` column.
- Set the `Date` column as index, handle missing values if any.

2. Exploratory Time Series Analysis:

- Visualize the closing price over time.
- Apply rolling statistics (mean and std) to observe trend stability.
- Perform seasonal decomposition and visualize components.

3. ACF and PACF Analysis:

- Use `plot_acf()` and `plot_pacf()` to identify significant lags.
- Choose appropriate values for `p`, `d`, `q` based on visual inspection and differencing.

4. Model Building:

- Train ARIMA and SARIMA models on historical data.
- Forecast next 30 or 60 days of closing prices.
- Visualize predicted prices alongside actual data.

5. Evaluation and Visualization:

- Use RMSE, MAE and MAPE to evaluate performance.
- Compare ARIMA vs SARIMA predictions.
- Plot residuals to confirm randomness (white noise assumption).

📁 Links to Kaggle Datasets for the Assignment:

- Tesla Stock Price History:
  🔗https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021/
- Apple Stock Price History:
  🔗https://www.kaggle.com/datasets/varpit94/apple-stock-data-updated-till-22jun2021/