**Time Series Analysis Assignment**

**Objective:** The objective of this assignment is to perform a time series analysis on the given data to understand and predict future trends. You will use Python and relevant libraries to preprocess the data, visualize trends, and apply time series forecasting models.

**Data Overview:** You have a dataset containing daily records, including the date, day of the week, holiday information, and various different restaurants (rest1, rest2, rest3, rest4, total). The aim is to predict the total metric.

**Steps:**

1. **Data Preprocessing:**
   * Load the dataset and convert the date column to a datetime format.
   * Set the date column as the index of the DataFrame.
   * Handle any missing values if present.
   * Encode the weekday and holiday columns to numerical values for analysis.
2. **Exploratory Data Analysis (EDA):**
   * Plot the total metric over time to visualize trends, seasonality, and any potential anomalies.
   * Use summary statistics to understand the distribution of total.
   * Check for any correlations between total and other metrics (rest1, rest2, rest3, rest4).
3. **Feature Engineering:**
   * Create additional time-based features such as month, day, year, day\_of\_week, is\_weekend, etc.
   * Use the holiday and holiday\_name columns to create features that might impact the total metric.
4. **Model Selection and Training:**
   * Split the data into training and testing sets.
   * Choose and implement time series forecasting models such as:
     + ARIMA (AutoRegressive Integrated Moving Average)
     + SARIMA (Seasonal ARIMA)
     + Prophet by Facebook
   * Train the models using the training data and evaluate their performance on the test data.
5. **Model Evaluation:**
   * Use metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to evaluate model performance.
   * Visualize the actual vs. predicted total values over time to assess the model's accuracy.
6. **Forecasting:**
   * Use the best-performing model to forecast future total values.
   * Plot the forecasted values alongside the actual values for comparison.
7. **Interpretation and Insights:**
   * Analyse the results and interpret the model's predictions.
   * Discuss any observed trends, seasonality, and the impact of holidays on the total metric.
   * Provide recommendations based on the forecasted data.

**Deliverables:**

1. **Python Notebook:**
   * A well-documented Jupyter notebook containing the code for data preprocessing, EDA, feature engineering, model training, evaluation, and forecasting.
2. **Report:**
   * A summary report explaining the approach, findings, and conclusions drawn from the analysis. Include visualizations and performance metrics.

Finally, select the algorithm that is most appropriate for this dataset. Provide a statistical justification for this choice. Consider how, in a real-time application, this solution will address the issue, outlining the potential benefits and improvements in planning that can be achieved.

Send the assignment to – divakarapm@gmail.com