Problem Statement:

This internship project is focused on leveraging Python for time series prediction of household electricity consumption. The dataset includes essential features such as date, time, global active power, global reactive power, voltage, global intensity, and sub-metering values. The objective is to build robust time series forecasting models that can accurately predict future electricity consumption trends based on historical data. The insights derived from this analysis can empower households to optimize energy usage, plan efficiently, and contribute to sustainable energy practices.

Dataset Description:

1. Date: Date of the electricity consumption recording.

2. Time: Time of the electricity consumption recording.

3. Global_active_power: Total active power consumed by the household.

4. Global_reactive_power: Total reactive power consumed by the household.

5. Voltage: Voltage level during the electricity consumption period.

6. Global_intensity: Total current intensity consumed by the household.

7. Sub_metering_1: Electricity consumption in sub-metering 1 (e.g., kitchen).

8. Sub_metering_2: Electricity consumption in sub-metering 2 (e.g., laundry).

9. Sub_metering_3: Electricity consumption in sub-metering 3 (e.g., water heater).

Project Objectives:

1. Data Preprocessing:

   - Clean and preprocess the dataset, handling any missing values or outliers.

   - Combine the date and time columns into a datetime format for effective time series analysis.

2. Exploratory Data Analysis (EDA):

   - Conduct EDA to uncover patterns, trends, and seasonality in electricity consumption.

   - Visualize the relationships between different features to gain insights.

3. Time Series Forecasting Models:

   - Implement time series forecasting models such as ARIMA, SARIMA, or LSTM.

   - Evaluate the performance of the models using appropriate metrics.

4. Feature Engineering:

- Investigate the impact of various features on electricity consumption.
- Explore the creation of new features that might enhance prediction accuracy.

5. Model Evaluation and Tuning:

- Fine-tune model hyperparameters for optimal performance.
- Validate and optimize the model using a separate test dataset.

6. Future Consumption Prediction:

- Generate forecasts for future electricity consumption based on the trained models.
- Visualize and interpret the predictions to identify potential consumption patterns.

Deliverables:

- Python scripts for data preprocessing, EDA, and time series prediction models.
- Visualizations illustrating consumption patterns, model evaluation metrics, and predicted future trends.
- A comprehensive report summarizing the findings, challenges encountered, and recommendations for optimizing household electricity consumption.

This project equips interns with hands-on experience in time series analysis, forecasting, and feature engineering, contributing to the broader goal of promoting energy-efficient practices in households.