To complete the project "Electricity Demand and Price Forecasting," here's a breakdown of the necessary steps for each module, from data ingestion to project presentation

1. Data Ingestion

- **Task**: Gather historical electricity demand and price data, along with relevant external factors like weather conditions, economic indicators, and calendar data (e.g., holidays, weekends).
- **Sources**:
  - Public datasets (such as the **Electric Reliability Council of Texas (ERCOT)**, **Open Power System Data**).
  - Weather data from APIs (like **OpenWeatherMap**, **NOAA**).
  - Energy market reports or datasets (e.g., **California ISO** or European electricity markets).
- **Output**: A comprehensive dataset containing:
  - Timestamp (date, time)
  - Electricity demand (MW)
  - Electricity price (in local currency or energy units)
  - Weather data (temperature, humidity, wind speed)
  - Other factors (holidays, time of day, day of the week)

2. Exploratory Data Analysis (EDA) and Data Preprocessing

- **EDA Tasks**:
  - Perform initial descriptive analysis (mean, median, standard deviation).
  - Visualize trends, seasonality, and patterns in electricity demand and price over time.
  - Check correlations between electricity demand/price and external factors (like temperature, day of the week).
  - Detect outliers and potential anomalies in demand and price data.
- **Preprocessing Tasks**:
  - Handle missing values (imputation or dropping).
  - Normalize or standardize data if necessary, especially for model compatibility.
  - Split the data into training and testing sets (time-based split to avoid data leakage).
- **Tools**: Use **Pandas**, **Matplotlib**, **Seaborn**, and **Scikit-learn** for EDA and preprocessing.
- **Output**: Cleaned dataset ready for model training, along with key insights from EDA.

3. Feature Engineering (if necessary)

- **Task**: Create additional features to improve model performance.
- **Feature Ideas**:
  - **Lag features**: Introduce past demand/price values as predictors (for example, demand or price 1 hour ago, 24 hours ago, etc.).
  - **Rolling averages**: Calculate moving averages of demand or price to capture trends and smooth fluctuations.
  - **Time-based features**: Extract features like hour of the day, day of the week, month, season, and holidays.

    - **Interaction terms**: Explore interactions between weather variables and demand/price to identify combined effects.
- **Output**: Enhanced dataset with engineered features that capture temporal and external factor relationships.

4. Model Selection and Training

- **Task**: Choose and train machine learning models to predict electricity demand and price.
- **Models to Consider**:
    - **Time Series Models**:
        - **ARIMA/SARIMA**: For demand/price forecasting based on historical patterns.
        - **Facebook Prophet**: Handles seasonality, holidays, and trend shifts in time series data.
    - **Machine Learning Models**:
        - **Random Forest Regression**: Captures complex, non-linear relationships.
        - **Gradient Boosting Machines (XGBoost/LightGBM)**: Powerful for multivariate prediction tasks.
        - **Long Short-Term Memory (LSTM)**: Deep learning model suitable for sequential data like electricity demand.
    - **Hybrid Models**: Combine time series techniques with machine learning to capture both temporal dependencies and external influences.
- **Tools**: Use **Scikit-learn**, **Statsmodels**, **Prophet**, and deep learning libraries like **TensorFlow** or **Keras**.
- **Output**: Trained models for predicting electricity demand and prices.

5. Model Evaluation, Selection, and Forecasting

- **Task**: Evaluate model performance and select the best model(s) for forecasting electricity demand and prices.
- **Evaluation Metrics**:
    - **RMSE (Root Mean Squared Error)**: Measures the average error magnitude.
    - **MAE (Mean Absolute Error)**: Provides error magnitude without exaggerating outliers.
    - **MAPE (Mean Absolute Percentage Error)**: Expresses error as a percentage of actual values, useful for comparing demand and price predictions.
- **Model Comparison**: Compare different models (time series vs machine learning) based on accuracy and performance metrics.
- **Forecasting**: Use the selected model(s) to forecast electricity demand and price for future periods (e.g., next day, week, or month).
- **Output**: Forecasts for future electricity demand and prices, along with a model evaluation report.

6. Project Presentation and Documentation

- **Project Report**:
    - **Introduction**: Define the problem, scope, and objectives of the project.

    - **Data Collection and EDA**: Discuss the data sources, structure, and key insights obtained from EDA.
    - **Feature Engineering**: Explain any new features created and why they were useful.
    - **Modeling**: Outline the models considered, training process, and hyperparameter tuning.
    - **Results**: Present model evaluation metrics, visualizations of actual vs. predicted demand/price, and insights.
    - **Conclusion**: Summarize the business impact (improved demand/price forecasting for smart grid operations).
- **Presentation**:
    - Create visual aids (charts, graphs) that showcase demand/price trends, model predictions, and the effect of external factors (e.g., weather).
    - Include comparisons between models and justify why the final model was chosen.
    - Provide actionable insights for grid operators or energy market participants.
- **Tools**: Use **PowerPoint**, **Jupyter Notebooks**, or **Google Slides** for the presentation.

Tools and Libraries:

- **Programming Languages**: Python
- **Libraries**:
    - **Pandas**, **NumPy**: Data manipulation and feature engineering
    - **Matplotlib**, **Seaborn**: Visualization
    - **Scikit-learn**: Model training, regression, and evaluation
    - **Statsmodels**, **Prophet**: Time series analysis
    - **TensorFlow**, **Keras**: Deep learning (LSTM)
- **APIs**: For data collection (e.g., weather APIs)
- **Jupyter Notebooks** for documentation and presentation.

Outcome:

- A machine learning model that accurately predicts electricity demand and price.
- Improved decision-making support for grid operators and energy market participants through reliable forecasts and insights.