Phase 1: Problem Definition and Design Thinking

Problem Definition:

The main objective of this project is to develop a predictive model that leverages historical electricity prices and relevant factors to forecast future electricity prices. The goal is to create a valuable tool for both energy providers and consumers, enabling them to make well-informed decisions regarding electricity consumption and investments based on accurate price predictions. The project comprises several key stages, including data preprocessing, feature engineering, model selection, model training, and model evaluation.

Design Thinking:

Data Source:

Utilize the provided dataset, which includes the following columns:

* datetime
* holiday
* holiday flash
* dayofweek
* week of year
* day
* month
* year
* periods of day
* ForecastWindProduction
* SystemLoadEA
* SMPEA
* ORKTemperature
* ORKWindspeed
* ActualWindProduction
* SystemLoadEP2
* SMPEP2

Data Preprocessing:

2. Start by cleaning and preprocessing the dataset. This includes:

Handling missing values: Identify and fill in or remove missing data points appropriately.

Converting categorical features: Transform categorical data (e.g., "holiday") into numerical representations using methods like one-hot encoding or label encoding.

Feature Engineering:

3. Create additional features that could improve the model's predictive power. Consider generating:

Time-based features: Extract information such as day of the week, month, and year from the "datetime" column.

Lagged variables: Include historical electricity prices (lagged values) as features, as they often contain valuable information for forecasting.

Model Selection:

4. Choose suitable time series forecasting algorithms. Given your dataset, consider starting with the following models:

ARIMA (AutoRegressive Integrated Moving Average): A traditional time series forecasting model.

LSTM (Long Short-Term Memory): A deep learning model capable of capturing complex patterns.

Model Training:

5. Train the selected model using the preprocessed dataset. This involves feeding historical electricity prices and the engineered features into the chosen model for it to learn patterns and relationships.

Evaluation:

6. Assess the model's performance using appropriate time series forecasting metrics. Utilize metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or others relevant to your specific goals.

With these guidelines, you're ready to embark on your electricity price prediction project. Remember to iterate and fine-tune your approach as needed to achieve the most accurate forecasts for future electricity prices.