Certainly, when working on a project to predict future electricity prices, loading and preprocessing the dataset is a crucial first step. Here’s a step-by-step guide on how to load and preprocess the dataset, along with some initial analysis:

1. \*\*Loading the Dataset\*\*:

- Import the necessary libraries in your chosen programming environment (e.g., Python with pandas).

- Load the historical electricity price dataset, which may be available in various formats such as CSV, Excel, or database queries.

```python

Import pandas as pd

# Load the dataset

Data = pd.read\_csv(‘electricity\_prices.csv’)

```

2. \*\*Data Exploration\*\*:

- Explore the dataset to gain an initial understanding of its structure and content.

```python

# Display the first few rows of the dataset

Print(data.head())

# Check data types, missing values, and summary statistics

Print(data.info())

Print(data.describe())

```

3. \*\*Data Preprocessing\*\*:

- Data preprocessing is essential to ensure data quality and prepare it for analysis and modeling.

- Handle missing values, outliers, and duplicates.

- Convert date/time columns to datetime objects.

- Ensure a consistent time interval for time series analysis.

```python

# Handle missing values

Data.dropna(inplace=True)

# Convert date/time column to datetime object

Data[‘timestamp’] = pd.to\_datetime(data[‘timestamp’])

# Set the timestamp column as the index

Data.set\_index(‘timestamp’, inplace=True)

# Ensure a consistent time interval (e.g., hourly data)

Data = data.resample(‘H’).mean()

```

4. \*\*Data Visualization\*\*:

- Create visualizations to understand the dataset’s patterns and seasonality.

- Plot time series, histograms, and box plots to identify trends and outliers.

```python

Import matplotlib.pyplot as plt

# Visualize time series data

Plt.figure(figsize=(12, 6))

Plt.plot(data.index, data[‘electricity\_price’])

Plt.title(‘Electricity Price Time Series’)

Plt.xlabel(‘Date’)

Plt.ylabel(‘Price’)

Plt.show()

# Create additional visualizations as needed

```

5. \*\*Data Analysis\*\*:

- Calculate descriptive statistics to understand the dataset’s central tendencies and variability.

- Identify any seasonality or trends in the data using time series decomposition techniques.

- Test for stationarity, as it’s a prerequisite for many time series forecasting models.

6. \*\*Feature Engineering\*\*:

- Create additional features that may help in forecasting, such as lag variables, rolling statistics, or external factors like weather data.

7. \*\*Data Splitting\*\*:

- Split the dataset into training and testing sets for model evaluation.

- Ensure that you reserve a portion of the data for future predictions.

With these steps completed, you will have a well-preprocessed dataset, and you can proceed to select and implement forecasting models like Prophet or deep learning techniques as mentioned earlier. This dataset preprocessing and initial analysis are essential for building accurate predictive models for future electricity prices.

**Done by**

**K Gunendra**

**Au720921244021**