GRT INSTITUTE OF
ENGINEERING AND
TECHNOLOGY, TIRUTTANI - 631209

Approved by AICTE, New Delhi Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PHASE 4

PROJECT TITLE

ELECTRICITY PRICE PREDICTION USING MACHINE LEARNING

COLLEGE CODE : 1103

LEELAVATHI N V
3rd yr, 5th sem
Reg no. : 110321104025
nvleela2004@gmail.com

FEATURE ENGINEERING:

Feature engineering in machine learning is the art of transforming raw data into a format that's more suitable for modeling. In simpler terms, you're jazzing up your data to make your machine learning model sing in harmony.

For an electricity price prediction project, feature engineering is like composing a symphony where each instrument contributes to the overall melody. Here are some ways you can engineer features for electricity price prediction:

- **Temporal Features:** Electricity prices often have a temporal pattern. Create features like time of day, day of the week, month, or even holidays. Maybe people use more electricity during certain seasons or at specific times of the day.
- **Weather-related Features**: Weather conditions can affect electricity usage. Include features like temperature, humidity, or even special events like storms or heatwaves.
- **Historical Prices:** Past prices can be strong indicators. Create features like rolling averages or percentage changes from previous periods.
- **Market Indicators:** Consider economic indicators or market trends that might influence electricity prices. GDP, inflation rates, or even trends in the energy market could be relevant.
- **Categorical Encoding:** If you have categorical variables, like types of energy sources or regions, encode them properly. One-hot encoding or label encoding can be used.
- **Interaction Features**: Sometimes, the combination of two or more features might have a more significant impact. Experiment with creating interaction terms.
- **Outlier Handling:** Identify and handle outliers in your data. Extreme values can distort predictions.
- **Scaling:** Ensure that your features are on similar scales. Scaling can be crucial, especially for algorithms sensitive to the magnitude of variables (e.g., distance-based algorithms).
- **Feature Selection:** Not every feature may contribute equally. Use techniques like recursive feature elimination or feature importance scores to select the most relevant features.

FEATURE SELECTION:

Selecting the right features for your electricity price prediction model is crucial for its performance. From the columns you've provided, it looks like you have a mix of temporal, categorical, and numerical features. Here's a breakdown of potential features you could consider:

1. Temporal Features:

- DateTime:Extract components like hour, minute, day of the week, month, etc.
- DayOfWeek: This can be redundant if you use DateTime, but sometimes it's useful on its own.

2. Categorical Features:

- Holiday: You can encode this as binary (1 for holiday, 0 for non-holiday).
- HolidayFlag: If it provides additional information beyond "is holiday," include it. Otherwise, it might be redundant.
- PeriodOfDay: Morning, afternoon, evening, night, etc.

3. Numerical Features:

- WeekOfYear, Day, Month, Year: These could be useful, especially if there are seasonal trends.
- ForecastWindProduction, ActualWindProduction: Wind production could influence electricity prices.
- SystemLoadEA, SystemLoadEP2: Past electricity consumption.
- SMPEA, SMPEP2: Social Market Price for Electricity, if relevant.
- ORKTemperature, ORKWindspeed: Weather-related features.
- CO2Intensity: Environmental impact could influence electricity prices.

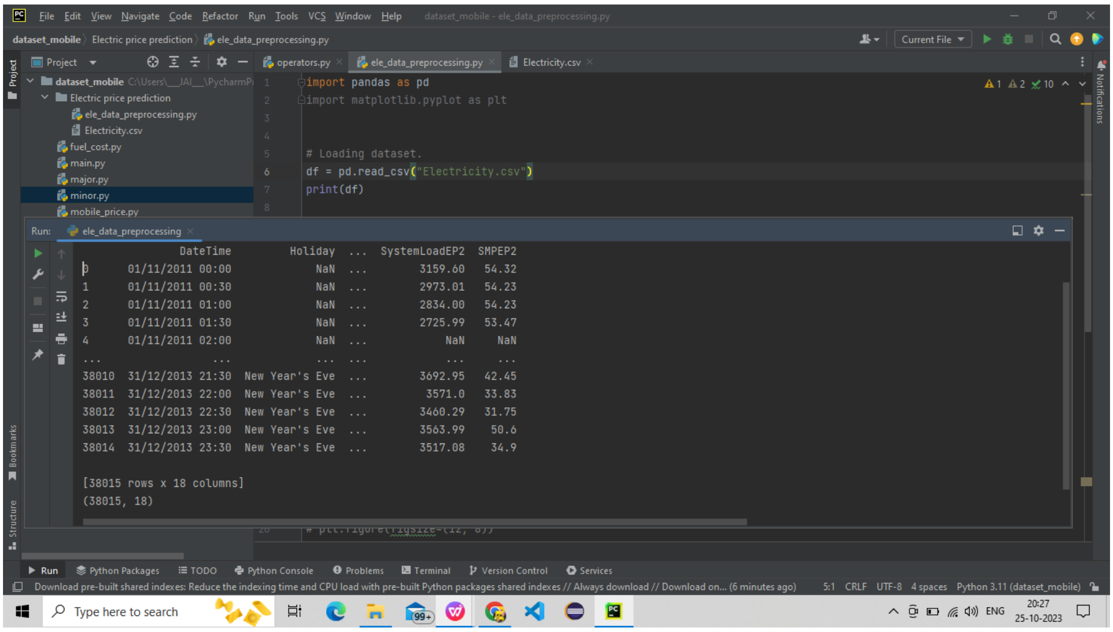
Correlation: Check for correlations between features. Highly correlated features might not add much new information.

Relevance: Ensure that each feature logically makes sense in the context of electricity price prediction. For example, if HolidayFlag is closely related to Holiday, including both might be unnecessary.

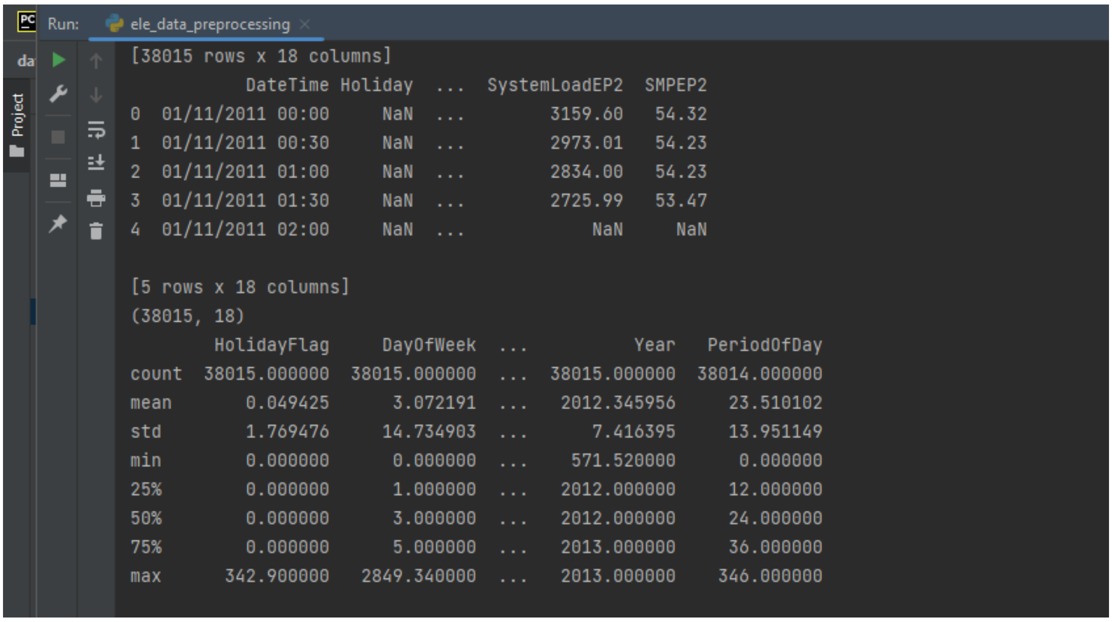
Dimensionality: Be mindful of the curse of dimensionality. Including too many irrelevant features could lead to overfitting.

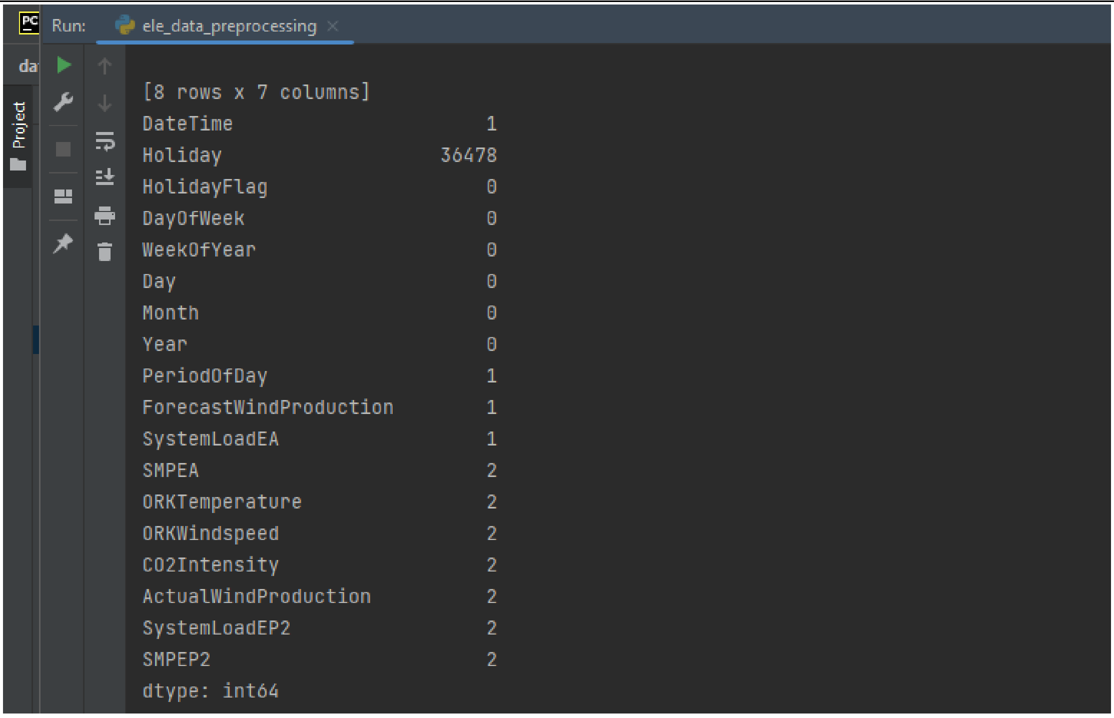
Feature Importance: If you're using tree-based models like Random Forest or XGBoost, you can check feature importance to see which features contribute the most to the model's predictions.

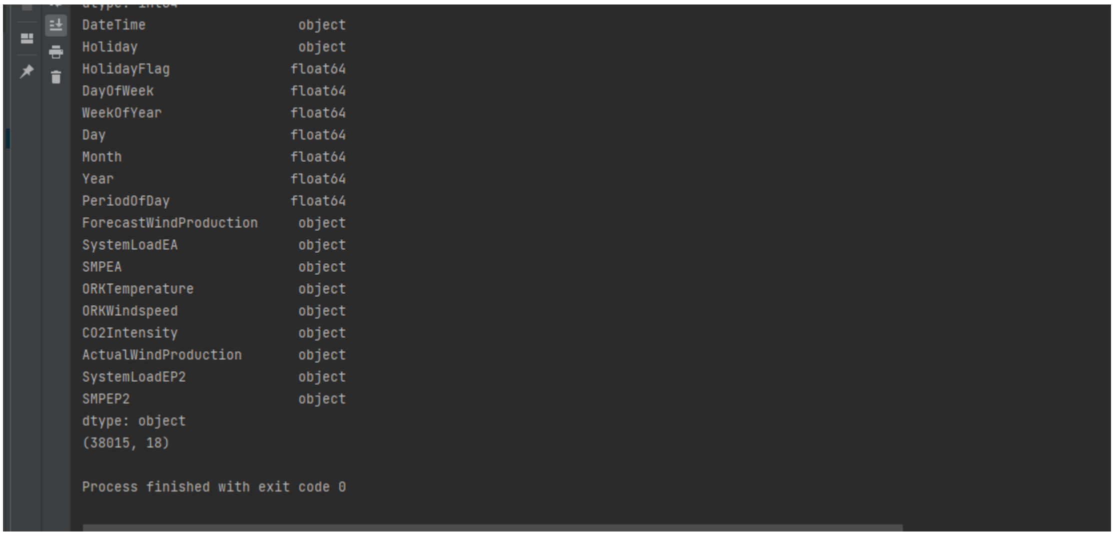
1. Load The Dataset:



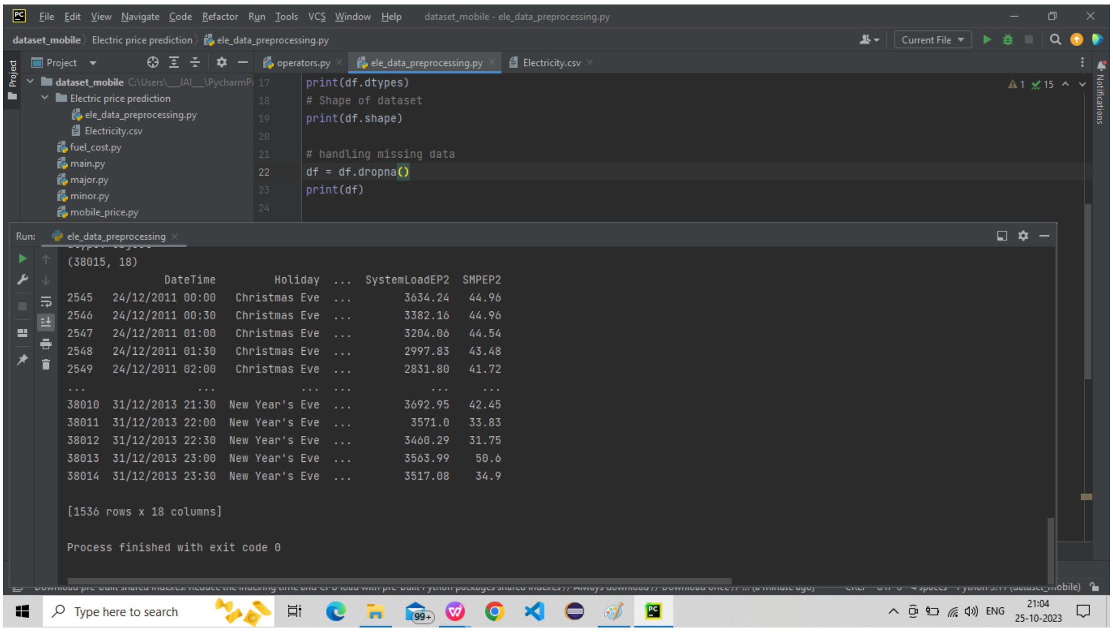
2.Explore The Dataset:



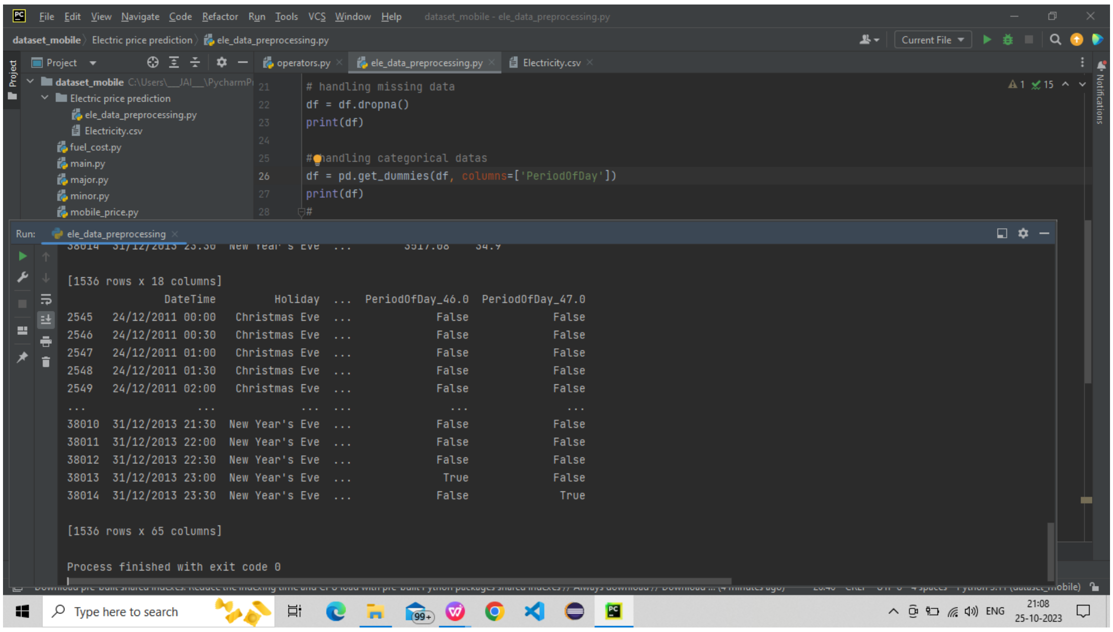




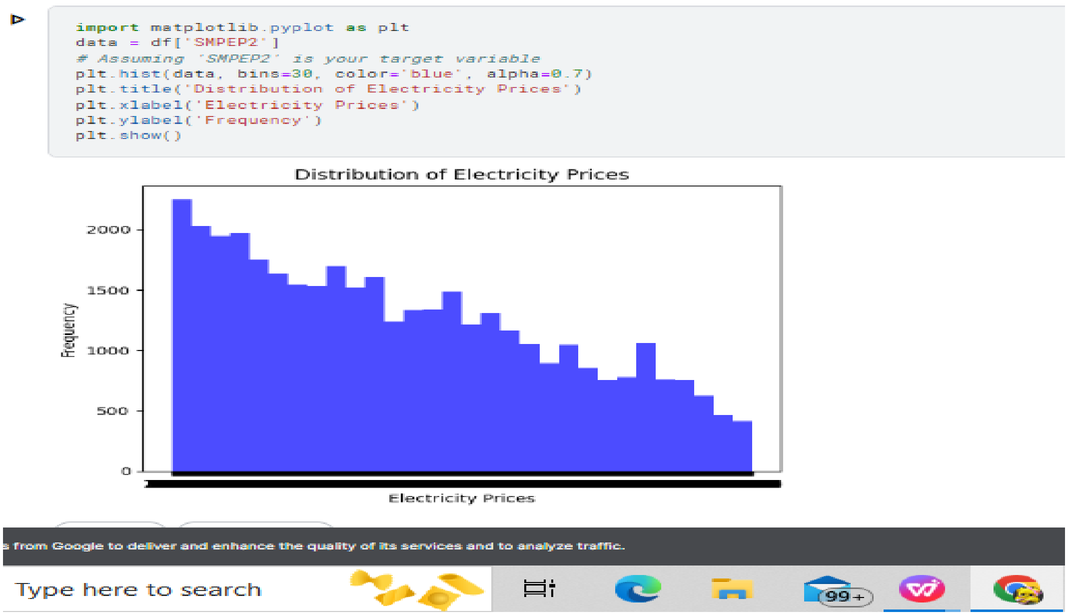
2. Handle Missing Values:



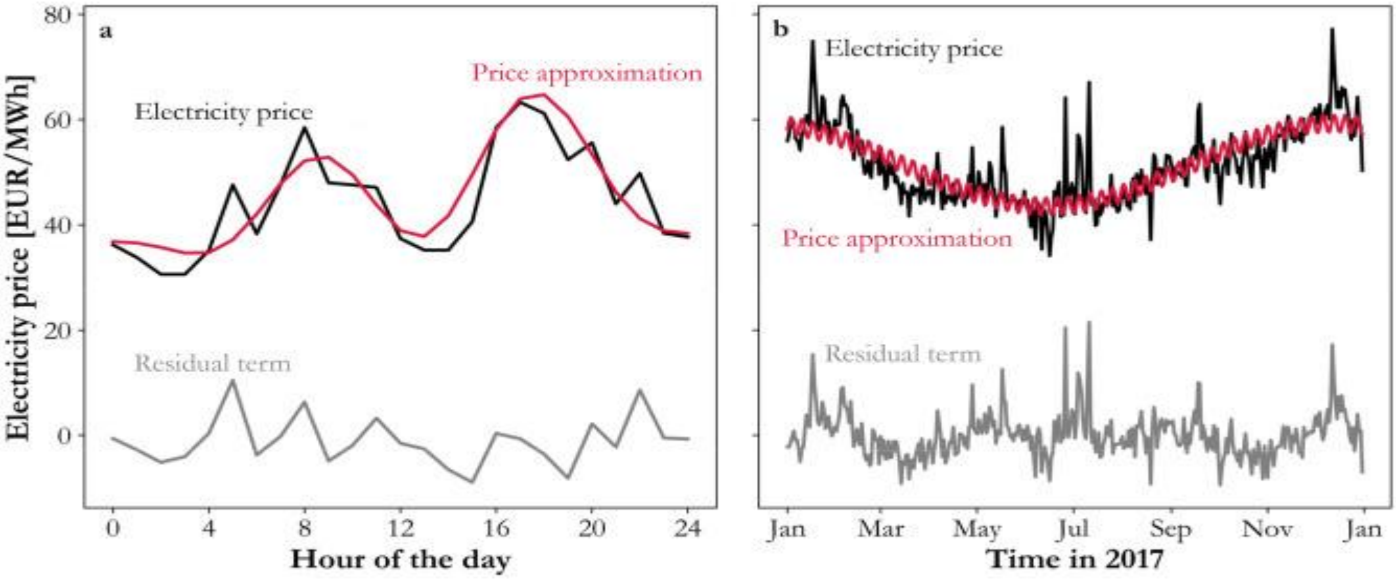
3. Handling categorical values:



4. Exploring Data:

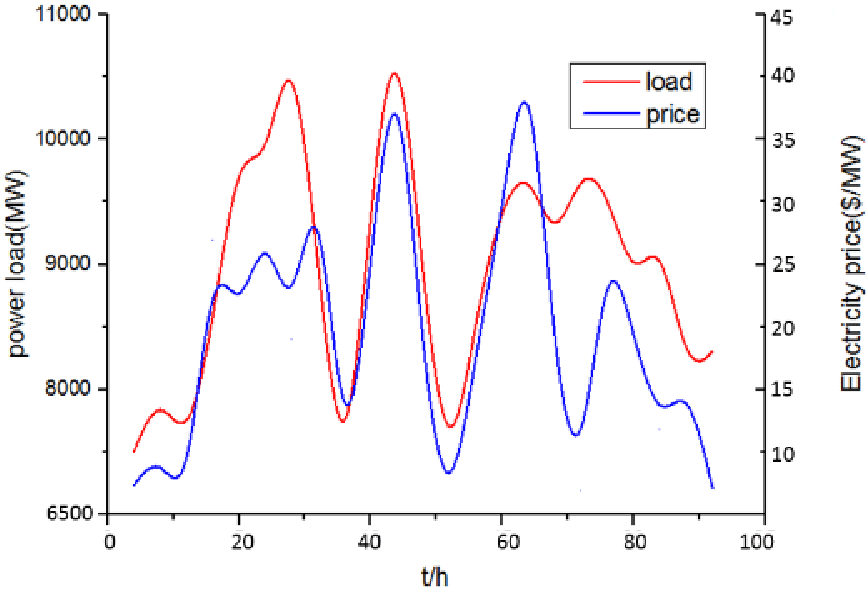


5.



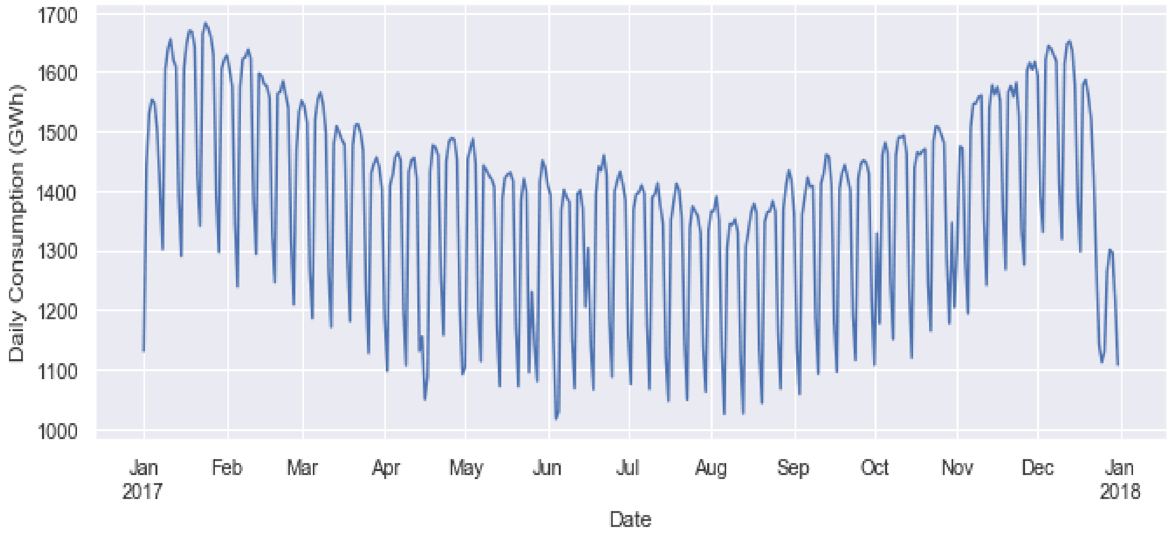
The above graph shows the relation between the Electricity price('SMPEP2') and the hours of the day, ('DateTime') and the price approximation on the year 2017 is predicted approximation electricity price prediction is visualized in the above picture.

6. Correlation Analysis

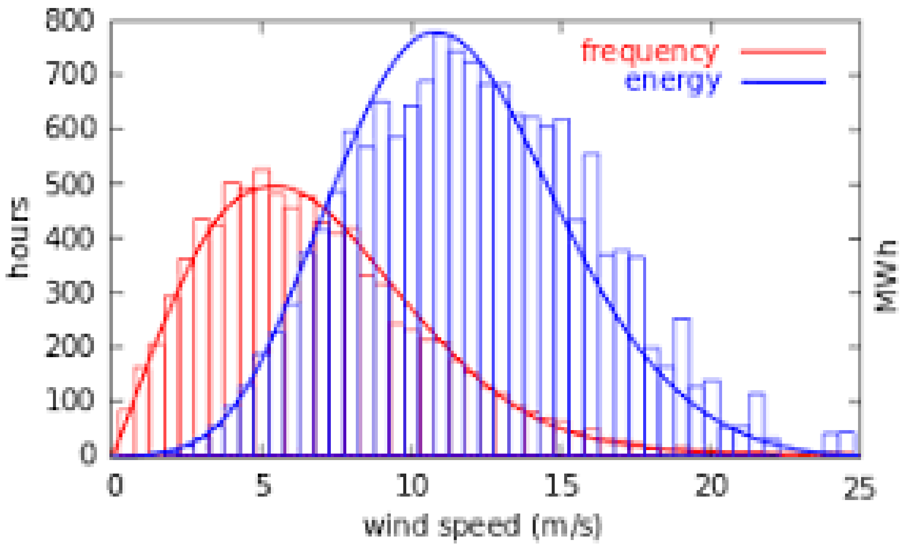


This gives an corelation between the electric prices and the power load in the given dataset.

7. **Temporal analysis :**



8. **Wind Production Analysis:**



These analysis provide a starting point of understanding your dataset and identifying potential patterns that can be useful for building an electricity price prediction model.

Summarization:

Feature Engineering:

1.**Time-related features:**Extracted additional features from the DateTime column, including day of the week, month, year, etc.

2. **Lag features:**Created lag features for the target variable ('SMPEP2') to capture temporal dependencies.

3. **Rolling statistics:** Calculated rolling mean, median, and standard deviation for relevant features to smooth out noise.

4.**Weather-related features:** Considered inclusion of weather-related features such as temperature, humidity, or wind speed if available.

5. **Special events:** Incorporated information about holidays or special events as binary features.

6. **Forecast features:** Utilized forecasted values for relevant features like 'ForecastWindProduction'.

7. **Interaction terms:** Considered creating interaction terms between features if they could have a significant impact.

8. **Cyclical features:** Encoded cyclical patterns, especially for time-related features, using sine and cosine transformations.

9. **Holiday indicators:** Created binary indicators for holidays or special events.

10.**Feature scaling:** Normalized or standardized numerical features to ensure they are on a similar scale.

Model Selection:

1.**Algorithm Selection:** Considered a variety of algorithms including time-series models (ARIMA, SARIMA) and machine learning algorithms (Random Forest, Gradient Boosting) for electricity price prediction.

2.**Data Splitting:**Split the data into training, validation, and testing sets to train and evaluate the models.

3.**Model Training:** Trained the selected models using historical data, tuning hyperparameters for optimal performance.

4. **Evaluation Metrics:** Evaluated models using appropriate metrics such as MAE, RMSE, or others.

5.**Comparison:** Compared the performance of different models to select the one that best suits the project's requirements.

6. **Deployment Considerations:** Prepared for the deployment of the chosen model for making future predictions.