

Measure Energy Consumption

Development

Project Goal:

The goal of this project is to process the energy consumption data for time series forecasting, you'll typically need to perform several steps, including data loading, cleaning, and transformation.

Data Loading:

- Import the necessary Python libraries, such as Pandas and NumPy.
- Load the dataset from the provided Kaggle link. You can use Pandas to read CSV files or other formats.

Data Cleaning:

- Check for and handle any missing values.
- Convert the timestamp column to a datetime data type for time-based analysis.

Exploratory Data Analysis (EDA):

- Perform basic data exploration to understand the data's characteristics.
- Visualize the time series data to identify trends and patterns.

Feature Engineering:

- Create additional features, such as lag features (past values) or rolling statistics.
- These features can provide more information for forecasting.

Time Series Forecasting:

- Choose a time series forecasting model, such as ARIMA, Exponential Smoothing, or LSTM.
- Split the data into training and testing sets.
- Train the selected model on the training data and evaluate it on the test data.

Python Script:

Step 1: Setup

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import xgboost as xgb
from sklearn.metrics import mean_squared_error
```

customize the style

```
pd.options.display.float_format = '{:.5f}'.format
pd.options.display.max_rows = 12
```

#Load the data

```
filepath = '../input/hourly-energy-consumption/PJME_hourly.csv'
df = pd.read_csv(filepath)
```

```
print("Successfully Uploaded")
```

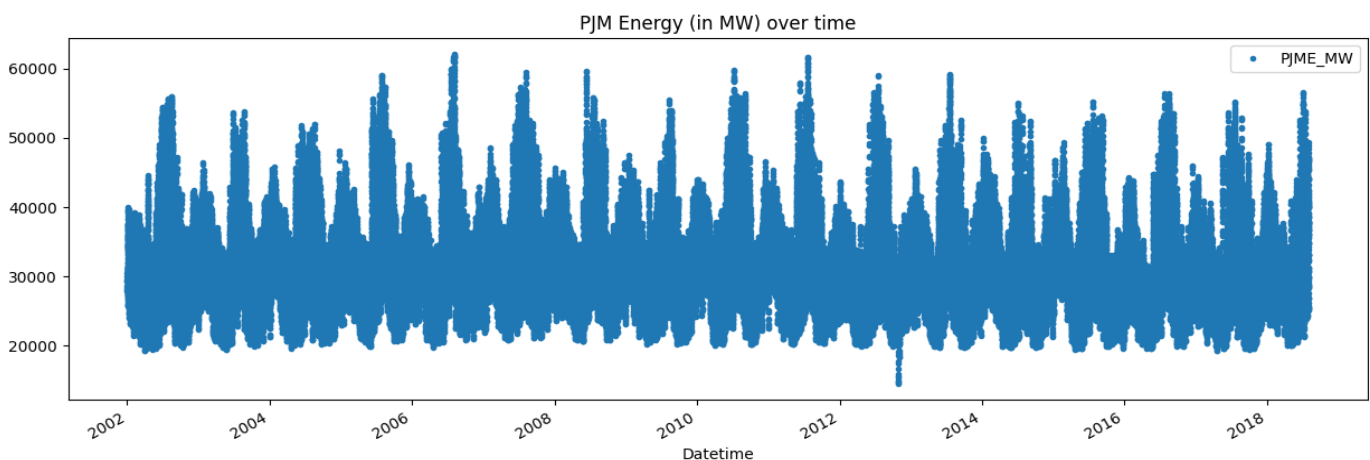
Step 2: Explore the data

turn data to datetime

```
df = df.set_index('Datetime')
df.index = pd.to_datetime(df.index)
```

create the plot

```
df.plot(style='.',
        figsize=(15, 5),
        title='PJM Energy (in MW) over time')
plt.show()
```



Step 2: Split the data

train / test split

```
train = df.loc[df.index < '01-01-2022']
```

```
test = df.loc[df.index >= '01-01-2022']
```

unfold_lessHide code

In [5]:

Linkcode

```
fig, ax = plt.subplots(figsize=(15, 5))
```

```
train.plot(ax=ax, label='Training Set', title='Train/Test Split')
```

```
test.plot(ax=ax, label='Test Set')
```

```
ax.axvline('01-01-2022', color='black', ls='--')
```

```
ax.legend(['Training Set', 'Test Set'])
```

```
plt.show()
```



After exploring the data, you need to prepare it for analysis, which involves

- ❖ Exploratory Data Analysis (EDA):
- ❖ Feature Engineering:
- ❖ Time Series Forecasting:

Setting up the data for analysis is crucial to ensure that it's in the right format and condition for modeling. It's the foundation for accurate predictions in the next steps of your project, which typically involve selecting a forecasting model and training it.