Measure Energy Consumption

Problem Statement:

The measurement of energy consumption is critical in understanding and optimizing energy usage in various sectors, including manufacturing sites, homes, commercial buildings, and transportation. However, the manual collection and analysis of energy consumption data can be time-consuming and error-prone. Therefore, there is a need for an automated approach to collect, analyze and visualize energy consumption data for better decision-making.

Dataset

Link: https://www.kaggle.com/datasets/ro bikscube/hourly-energy-consumption

Problem Definition:

The problem at hand isto create an automated system that measures energy consumption, analyzes the data, and provides visualizations for informed decision-making. This solution aims to enhance efficiency, accuracy, and ease of understanding in managing energy consumption across various sectors.

Problem Definition: The problem at hand is to create an automated system that measures energy consumption, analyzes the data, and provides visualizations for informed decision-making. This solution aims to enhance efficiency, accuracy, and ease of understanding in managing energy consumption across various sectors.

LIBRARIES USED IN TIME SERIES ANALYSIS

When performing time series analysis for measuring energy consumption, you may need specialized libraries and tools that cater to the unique characteristics of energy consumption data. Here are some libraries and frameworks commonly used for this purpose:

1. **pandas**: As mentioned earlier, pandas is essential for data manipulation, and it can be used to handle and preprocess energy consumption data.

2. **numpy**: NumPy is valuable for numerical operations and calculations that may be required in energy consumption analysis.

3. **matplotlib** and **seaborn**: These libraries are helpful for visualizing energy consumption trends and patterns, making it easier to identify anomalies or seasonal variations.

4. **statsmodels**: It provides statistical tools for time series analysis, including regression models that can be used to study the relationship between energy consumption and other variables.

5. **PyStan**: If you need to perform Bayesian time series analysis for energy consumption, PyStan is a Python interface to Stan, a probabilistic programming language.

6. **Prophet**: Developed by Facebook, Prophet is suitable for forecasting energy consumption data, which often exhibits seasonality and holiday effects.

7. **tsfresh**: This library can be used to extract relevant features from energy consumption time series data, which can then be used for modeling and prediction.

8. **TensorFlow** and **PyTorch**: If you want to apply deep learning techniques to energy consumption analysis, these deep learning frameworks can be used to build complex neural network models for forecasting or anomaly detection.

9. **HMMlearn**: Hidden Markov Models (HMMs) can be useful for modeling energy consumption patterns, and HMMlearn is a library for working with HMMs in Python.

10. **OpenDSS (Distribution System Simulator)**: This is a specialized tool for simulating and analyzing electrical distribution systems. It can be used to model and analyze energy consumption in electrical networks.

11. **EnergyPlus**: EnergyPlus is an advanced building energy simulation program that can be used for detailed analysis of building energy consumption.

Time series forecasting

```python
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt
import xgboost as xgb
from xgboost import plot_importance, plot_tree
from sklearn.metrics import mean_squared_error, mean_absolute_error, mean_
absolute_percentage_error
plt.style.use('fivethirtyeight')
```

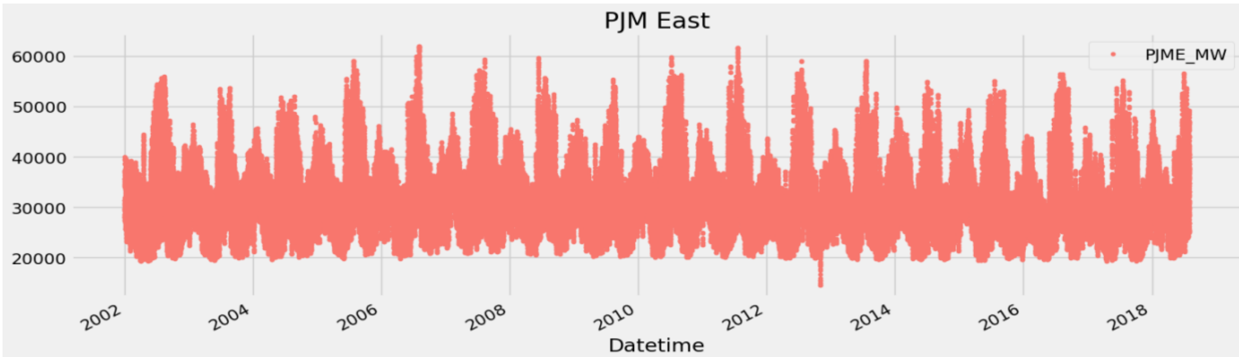
pjme = pd.read_csv('../input/hourly-energy-consumption/PJME_hourly.csv', index_col=[0], parse_dates=[0])

color_pal = ["#F8766D", "#D39200", "#93AA00", "#00BA38", "#00C19F", "#00B9E3", "#619CFF", "#DB72FB"]

_ = pjme.plot(style='.', figsize=(15,5), color=color_pal[0], title='PJM East')

Pjme.head(3)

Datetime	Pjme_MV
2002-12-31 01:00:00	26498.0
2002-12-31 02:00:00	25147.0
2002-12-31 03:00:00	24574.0



MACHINE LEARNING MODELS TO PREDICT ENERGY CONSUMPTION PATTERNS

Predicting future energy consumption patterns is a common task in the energy sector, and machine learning models can be highly effective for this purpose. Here are some machine learning models commonly used to predict future energy consumption patterns:

1. **Autoregressive Integrated Moving Average (ARIMA)**: ARIMA models are widely used for time series forecasting, including energy consumption. They capture trends, seasonality, and autoregressive components in the data.

2. **Seasonal Decomposition of Time Series (STL)**: STL decomposes time series data into seasonal, trend, and residual components, making it easier to model and forecast energy consumption patterns.

3. **Exponential Smoothing Methods**: Exponential smoothing methods like Holt-Winters are suitable for capturing seasonality and trends in energy consumption data.

4. **Prophet**: Developed by Facebook, Prophet is designed for forecasting time series data with daily observations and can handle holidays and special events, which are common in energy consumption.

5. **Long Short-Term Memory (LSTM) Networks**: LSTM is a type of recurrent neural network (RNN) that can capture long-term dependencies in time series data. It is often used for sequential data forecasting, including energy consumption.

6. **Gated Recurrent Unit (GRU) Networks**: Similar to LSTM, GRU networks are effective for capturing temporal dependencies in time series data and are computationally more efficient.

7. **Convolutional Neural Networks (CNNs)**: CNNs can be used for time series forecasting, especially when there are spatial or image-like aspects to the data, such as energy consumption heatmaps.

8. **XGBoost and LightGBM**: These gradient boosting algorithms are versatile and can be used for regression tasks to predict energy consumption patterns.

9. **Random Forests**: Random Forests are an ensemble learning method that can handle complex relationships in data and are suitable for energy consumption prediction when feature importance is essential.

10. **Support Vector Machines (SVMs)**: SVMs can be applied to time series forecasting tasks by transforming the time series data into appropriate feature representations.

11. **Gaussian Process Regression**: Gaussian Processes can be used for probabilistic time series forecasting, providing uncertainty estimates along with predictions.

12. **Neural Prophet**: An extension of the Prophet library that incorporates deep learning components, making it more flexible for complex energy consumption patterns.

13. **Hybrid Models**: Combining multiple models, such as ARIMA with neural networks or LSTM with traditional statistical methods, can often improve prediction accuracy.

The choice of machine learning model depends on the characteristics of your energy consumption data, such as seasonality, trend, noise, and the presence of external factors like weather. Experimentation and model selection based on performance metrics like Mean Absolute Error (MAE) or Root Mean Squared

Error (RMSE) on a validation dataset are crucial steps in finding the most suitable model for your specific energy consumption prediction task.

12. **GridLAB-D**: This is a power distribution system simulation and analysis tool that can be used for modeling and analyzing the consumption and distribution of electrical energy.

The choice of libraries will depend on the specific tasks you need to perform in your energy consumption analysis, such as forecasting, anomaly detection, feature extraction, or simulation. It's often a good practice to combine multiple libraries to cover various aspects of the analysis effectively.

Predicting Future Energy Consumption using python

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import warnings
warnings.filterwarnings("ignore") # hide warnings
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pprint
%matplotlib inline
```

```
df = pd.read_csv("/kaggle/input/hourly-energy-consumption/AEP_hourly.csv")
print("="*50)
print("First Five Rows ","\n")
print(df.head(2),"\n")

print("="*50)
print("Information About Dataset","\n")
print(df.info(),"\n")

print("="*50)
print("Describe the Dataset ","\n")
print(df.describe(),"\n")

print("="*50)
print("Null Values t ","\n")
print(df.isnull().sum(),"\n")
```

```
==================================================
First Five Rows

              Datetime   AEP_MW
0  2004-12-31 01:00:00  13478.0
1  2004-12-31 02:00:00  12865.0

==================================================
Information About Dataset

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121273 entries, 0 to 121272
Data columns (total 2 columns):
 #   Column    Non-Null Count   Dtype
---  ------    --------------   -----
 0   Datetime  121273 non-null  object
 1   AEP_MW    121273 non-null  float64
dtypes: float64(1), object(1)
memory usage: 1.9+ MB
None

==================================================
Describe the Dataset

              AEP_MW
count  121273.000000
mean    15499.513717
std      2591.399065
min      9581.000000
25%     13630.000000
50%     15310.000000
75%     17200.000000
max     25695.000000

==================================================
Null Values t

Datetime    0
AEP_MW      0
dtype: int64
```