# MEASURE ENERGY CONSUMPTION

## *TEAM MEMBER*

**510521104039:SASIKUMAR R**

***PHASE 1: DOCUMENT SUBMISSION***

***PROJECT TITLE:*** **Measure Energy Consumption**

***OBJECTIVE:***

The objective of this project is to develop an automated system that measures energy consumption, analyze 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.

***OUTLINES:***

1,Data source

2.Data Preprocessing

3.Feature Extraction

4.Model Development

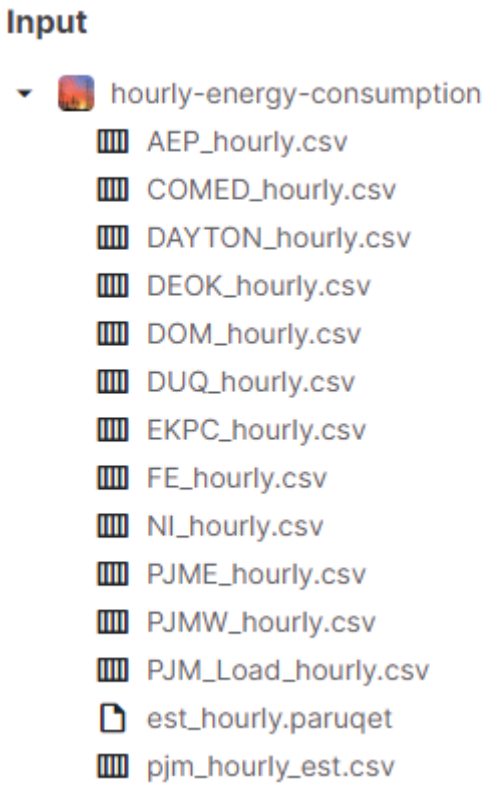
5.Visualization

6.Automation

# 1.Data Source

An data source is the one of the crucial part in measuring energy consumption it can consists of an various datasets to analyze the energy consumption. we can use the datasets with the help of programming we can make much more understandable.

Dataset Link: https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption




# 2.Data Preprocessing

The Data preprocessing involves the Clean, transform, and prepare the dataset for analysis. From the above Data source we have data set files such as .csv files with different type of names **NOTE: All the .csv files are linked in one single file : est_hourly.paruqet**

## Viewing the clean and transform dataset for analysis

***CODE:***

```
import pandas as pd

import os

df=pd.read_parquet('../input/hourly-energy-consumption/est_hourly.paruqet')

df.head()
```

***OUTPUT:***

| Datetime | AEP | COMED | DAYTON | DEOK | DOM | DUQ | EKPC | FE | NI | PJME | PJMW | PJM_Load |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1998-12-31 01:00:00 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 29309.0 |
| 1998-12-31 02:00:00 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 28236.0 |
| 1998-12-31 03:00:00 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 27692.0 |
| 1998-12-31 04:00:00 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 27596.0 |
| 1998-12-31 05:00:00 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 27888.0 |

# 3.Feature Extraction

Feature extraction is the process of selecting and transforming raw data into a set of relevant features or metrics that can be used for analysis or as inputs to machine learning models. In the context of energy consumption data, feature extraction aims to capture meaningful patterns, trends, and characteristics of the data that can help in understanding or predicting energy consumption behaviour.

## *CODE:*

```
import pandas as pd

import os

df=pd.read_parquet('../input/hourly-energy-consumption/est_hourly.paruqet')

df.describe().T
```

## *OUTPUT:*

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| AEP | 121273.0 | 15499.513717 | 2591.399065 | 9581.0 | 13630.0 | 15310.0 | 17200.00 | 25695.0 |
| COMED | 66497.0 | 11420.152112 | 2304.139517 | 7237.0 | 9780.0 | 11152.0 | 12510.00 | 23753.0 |
| DAYTON | 121275.0 | 2037.851140 | 393.403153 | 982.0 | 1749.0 | 2009.0 | 2279.00 | 3746.0 |
| DEOK | 57739.0 | 3105.096486 | 599.859026 | 907.0 | 2687.0 | 3013.0 | 3449.00 | 5445.0 |
| DOM | 116189.0 | 10949.203625 | 2413.946569 | 1253.0 | 9322.0 | 10501.0 | 12378.00 | 21651.0 |
| DUQ | 119068.0 | 1658.820296 | 301.740640 | 1014.0 | 1444.0 | 1630.0 | 1819.00 | 3054.0 |
| EKPC | 45334.0 | 1464.218423 | 378.868404 | 514.0 | 1185.0 | 1386.0 | 1699.00 | 3490.0 |
| FE | 62874.0 | 7792.159064 | 1331.268006 | 0.0 | 6807.0 | 7700.0 | 8556.00 | 14032.0 |
| NI | 58450.0 | 11701.682943 | 2371.498701 | 7003.0 | 9954.0 | 11521.0 | 12896.75 | 23631.0 |
| PJME | 145366.0 | 32080.222831 | 6464.012166 | 14544.0 | 27573.0 | 31421.0 | 35650.00 | 62009.0 |
| PJMW | 143206.0 | 5602.375089 | 979.142872 | 487.0 | 4907.0 | 5530.0 | 6252.00 | 9594.0 |
| PJM_Load | 32896.0 | 29766.427408 | 5849.769954 | 17461.0 | 25473.0 | 29655.0 | 33073.25 | 54030.0 |

# 4.Model Development

The model development involves Utilization of statistical analysis to uncover trends ,patterns and anomalies in the data With the help of matplotlib we can uncover the dataset for better understanding. As an example I'm gonna to take the

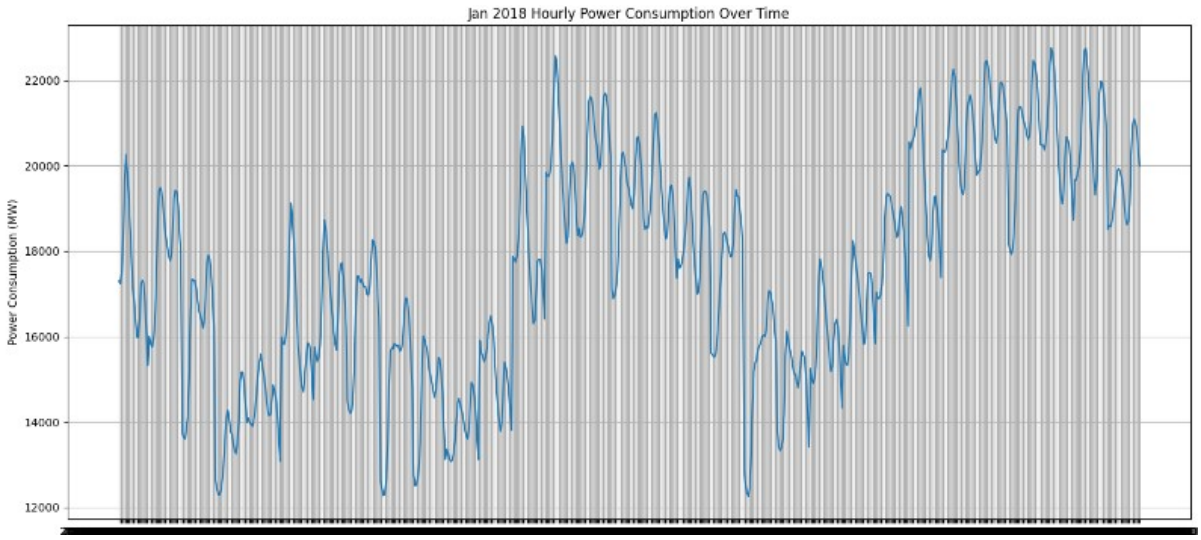
**DATASET: (hourly-energy-consumption/AEP_hourly.csv)**

From the dataset I'm gonna to extract the JAN 2018 hourly power consumption

## ***CODE:***

```
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('../input/hourly-energy-consumption/AEP_hourly.csv')
plt.figure(figsize=(15, 7))
plt.plot(df['Datetime'][120531:121275], df['AEP_MW'][120531:121275])
plt.title('Jan 2018 Hourly Power Consumption Over Time')
plt.xlabel('Datetime')
plt.ylabel('Power Consumption (MW)')
plt.grid(True)
plt.tight_layout()
plt.savefig('2018_Jan.png')
plt.show()
```

## ***OUTPUT:***



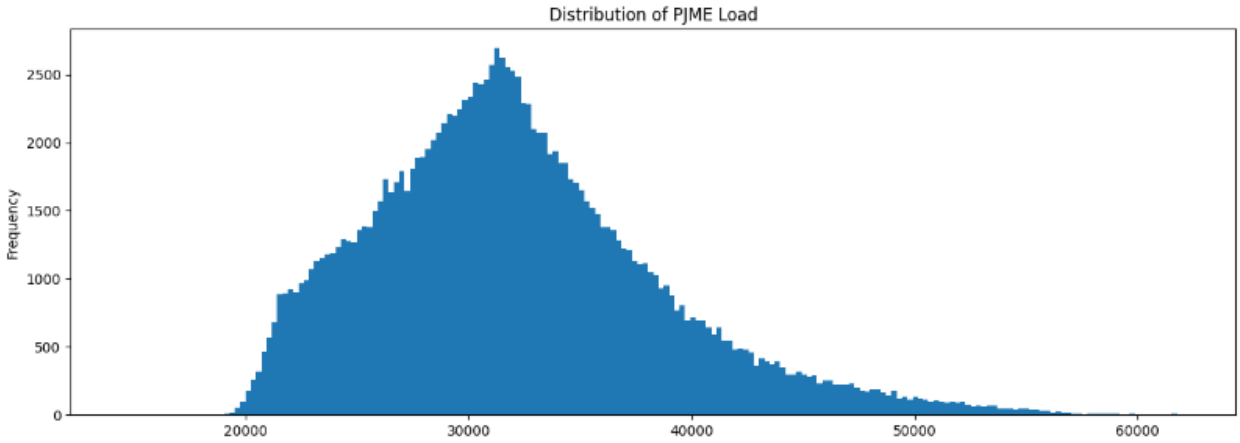

# 5.Visualization

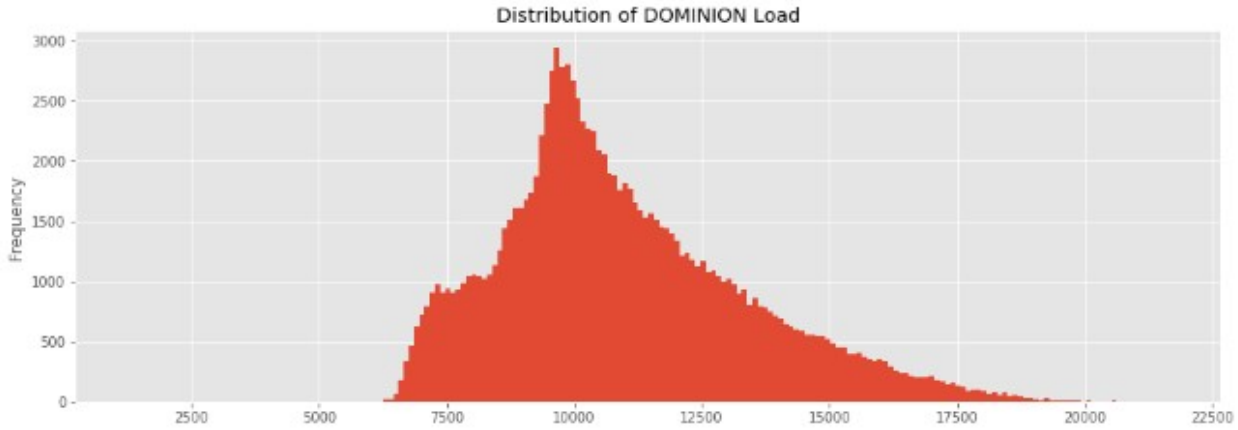
Creating visualizations to present energy consumption trends and insights is a powerful way to convey complex information in a clear and understandable manner. Below, I'll provide examples of different types of graphs and charts that can be used to visualize energy consumption trends and insights. These visualizations can be created using tools like Microsoft Excel, Google Sheets, Tableau, or Python libraries such as Matplotlib and Seaborn.

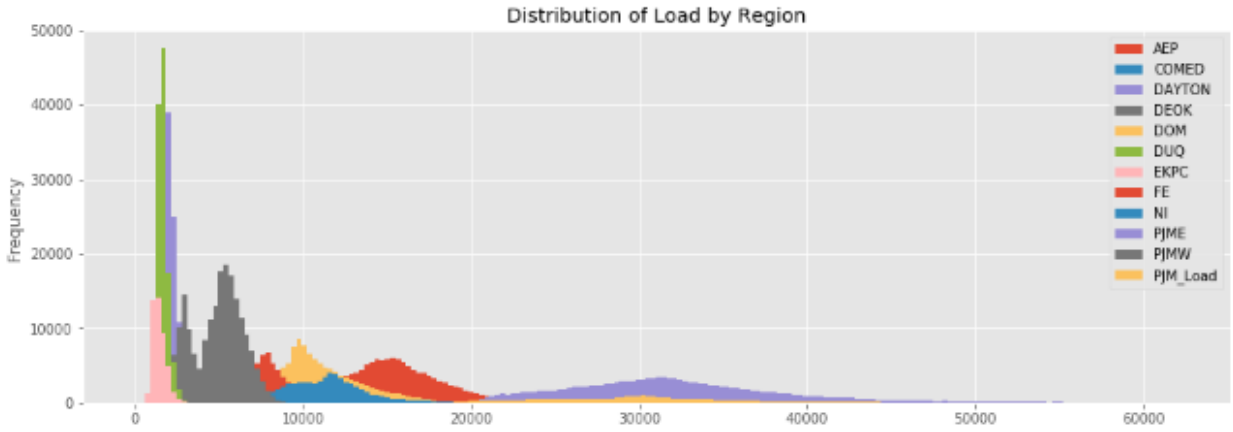
## ***CODE:***

```
import pandas as pd

import os

df=pd.read_parquet('../input/hourly-energy-consumption/est_hourly.paruqet')

D1= df['PJME'].plot.hist(figsize=(15, 5), bins=200, title='Distribution of PJME Load')

D2= df['DOM'].plot.hist(figsize=(15, 5), bins=200, title='Distribution of DOMINION Load')

D3= df.plot.hist(figsize=(15, 5), bins=200, title='Distribution of Load by Region')
```

## ***OUTPUT:***




# 6. Automation

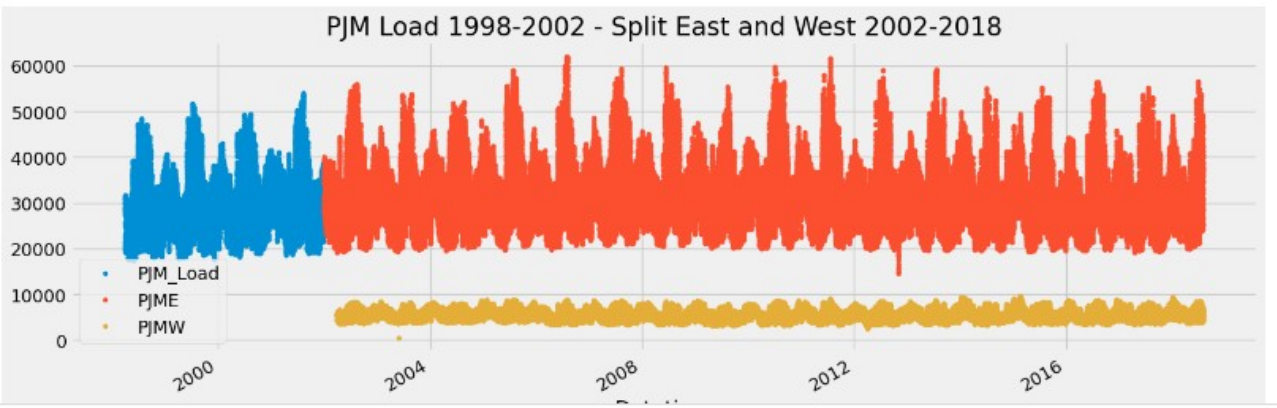
Building a script to automate data collection, analysis, and visualization processes involves several steps. I'll outline a high-level process using Python, a popular

programming language for such tasks. You can adapt this process to your specific needs and datasets.

## ***CODE:***

```
import matplotlib.pyplot as plt

import numpy as np

import os

df = pd.read_parquet('../input/hourly-energy-consumption/est_hourly.paruqet')

D1= df[['PJM_Load','PJME','PJMW']] \
    .plot(style='.', figsize=(15, 5), title='PJM Load 1998-2002 - Split East and West 2002-2018')
```

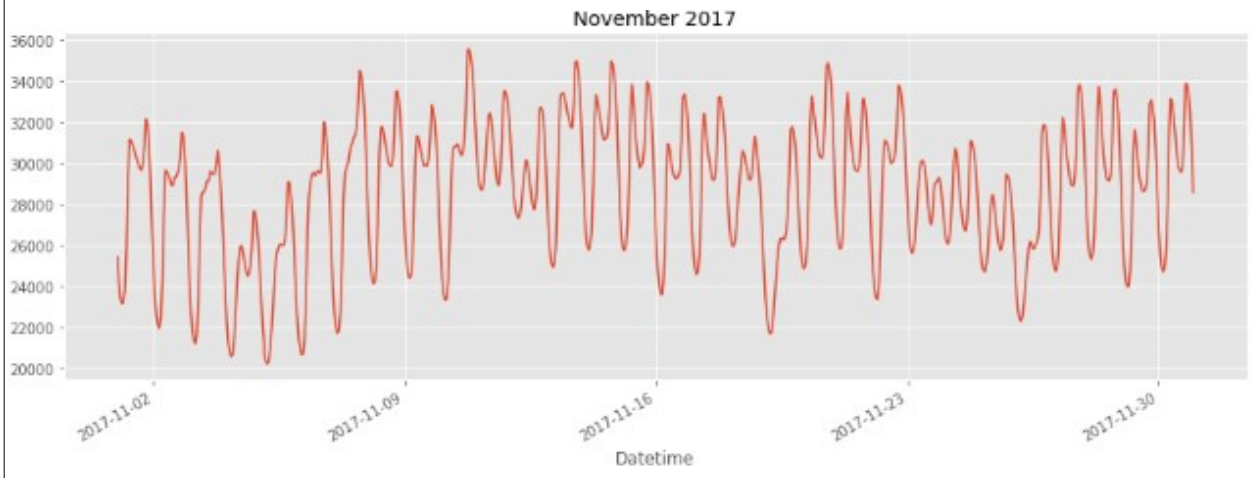
## ***OUTPUT:***




## ***CODE:***

```
import matplotlib.pyplot as plt

import numpy as np

import os

df = pd.read_parquet('../input/hourly-energy-consumption/est_hourly.paruqet')

D2= df['PJME'].loc[(df['PJME'].index >= '2017-11-01') &
                   (df['PJME'].index < '2017-12-01')] \
    .plot(figsize=(15, 5), title = 'November 2017')
```

## ***OUTPUT:***


November 2017
36000
34000
32000
30000
28000
26000
24000
22000
20000
2017-11-02
2017-11-09
2017-11-16
2017-11-23
2017-11-30
Datetime