**MEASURING ENERGY CONSUMPTION**

**Sanjay Kumar V** (2021503046**) Achyut Prasad DC** (2021503002)

**Ragul N** (2021503542) **Santhosh Kumar** **S** (2021503302)

**Phase 3: Development Part 1: Dataset Loading and Pre-processing**

**Introduction**

**Introduction**

In this phase of the project, we will focus on loading and pre-processing the PJM Hourly Energy Consumption dataset. The PJM Interconnection LLC (PJM) is a regional transmission organization in the United States, operating an electric transmission system serving several states. The dataset provides hourly power consumption data in megawatts (MW) and is a valuable resource for understanding energy consumption patterns.

**PJM Hourly Energy Consumption Data:**

**Source:** PJM's website

**Data Format:** Hourly time series data

**Unit of Measurement:** Megawatts (MW)

**Geographical Coverage:** Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia.

The dataset captures energy consumption trends over time, and it is particularly useful for various analyses and modelling tasks. Here are some ideas of what can be done with this dataset:

**Energy Consumption Prediction:** Splitting the last year into a test set, we can build a model to predict energy consumption. This can be valuable for optimizing energy resources and planning.

**Trend Analysis:** We can explore trends in energy consumption related to different factors, such as hours of the day, holidays, or long-term trends. Understanding these trends can help in resource allocation and grid management.

**Seasonal Analysis:** Investigate how daily trends change depending on the time of year. The difference between summer and winter trends can be substantial and understanding these variations is essential for energy management.

In "Development Part 1" of the "Measure Energy Consumption" project, our primary objective is to initiate the project by loading and preprocessing the PJM Hourly Energy Consumption dataset. We aim to ensure that the dataset is clean, accurate, and well-structured for subsequent analyses. This phase also involves gaining a comprehensive understanding of the dataset's characteristics, exploring potential use cases, and laying the groundwork for data cleaning, transformation, and documentation. By achieving these objectives, we establish a solid foundation for deriving valuable insights and building predictive models in the following project phases.

**Data Cleaning**

**1. DESCRIPTION OF DATA**

* Rename Columns
* Data Dimensions
* Data Types
* Check NA
* Fillout NA
* Change Types
* Descriptive Statistical
* Numerical Attributes
* Categorical Attributes

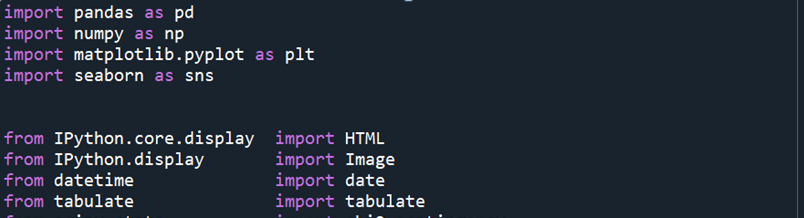
**2. FEATURE ENGINEERING**

* Creation of Hyphoteses
  + Demographic Hyphoteses
  + Geographic Hyphoteses
  + Sociocultural Hyphoteses
* Final list of Hypotheses
* Feature Engineering

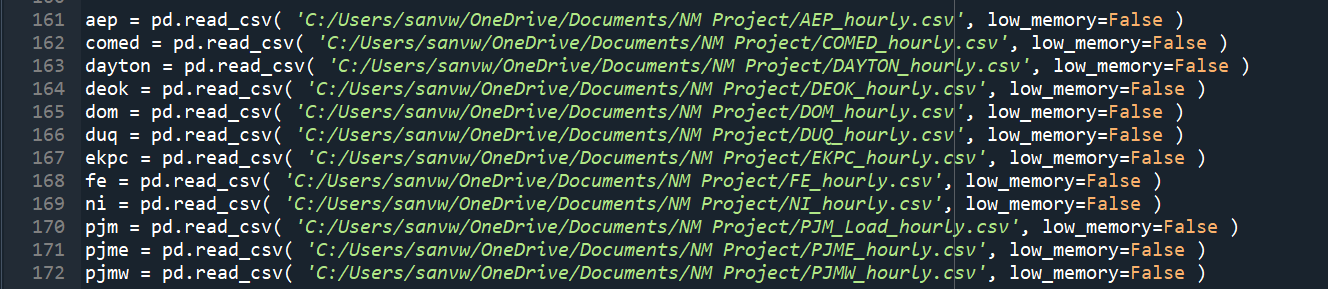
**3. VARIABLE FILTERING**

* Line filtering
* Column Selection

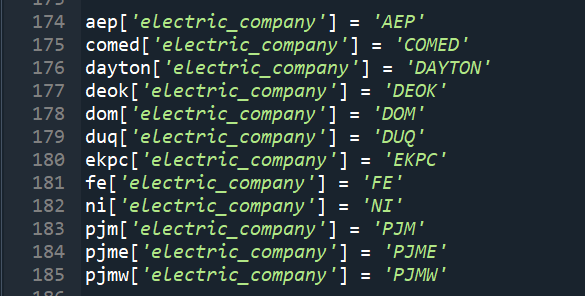
**IMPORTS:**

****

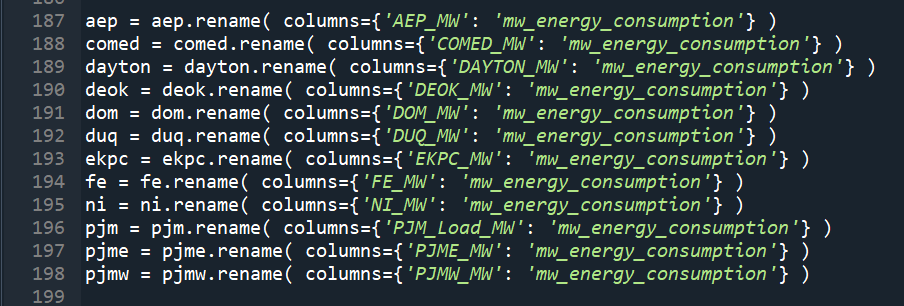
**LOADING DATA:**



# combining the datasets into one.



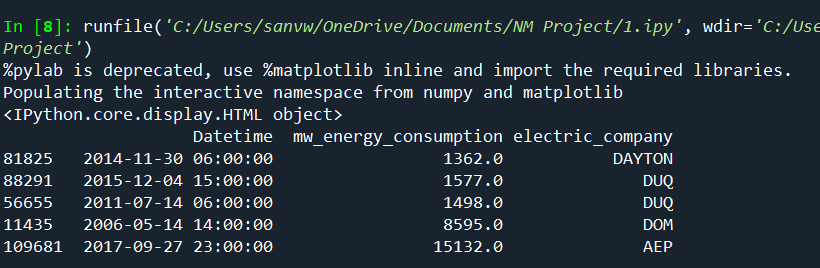
# The column in each dataset that shows the Megawatt Energy Consumption.



# Concatenate all the datasets



**Output:**

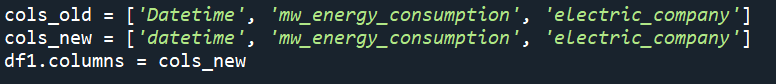


df\_raw.shape

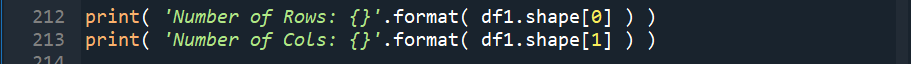


DESCRIPTION OF DATA

1. Rename coulumns

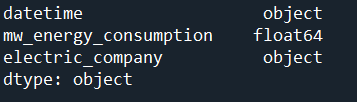


2. Data Dimensions



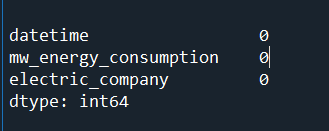


3. Datatypes



4. Check NA





**Conclusion:**

In "Development Part 1," we have taken the initial steps in our journey to measure energy consumption effectively and make informed decisions based on data-driven insights. This phase focused on loading and preprocessing the PJM Hourly Energy Consumption dataset, setting the stage for more advanced analyses and modeling.We began by retrieving the dataset from PJM's website and gained a thorough understanding of its structure, format, unit of measurement, and geographical coverage. This foundational knowledge is pivotal in interpreting the data accurately and aligning our objectives with the dataset's characteristics.

Our primary objectives in this phase were to prepare the dataset for subsequent phases of the project and to ensure its quality and reliability. We explored potential use cases, including energy consumption prediction, trend analysis, and seasonal variations, clarifying our project goals.Data preprocessing preparation included addressing issues related to missing values, outliers, data inconsistencies, and anomalies. These issues were identified, and we laid the groundwork for data cleaning, transformation, and documentation.

This phase sets the stage for the deeper analysis and modeling work that follows in the project. The dataset is now in a state that allows us to derive meaningful insights, make informed decisions, and build predictive models to optimize energy consumption.As we move forward, we will build upon this foundation to explore the dataset's trends, patterns, and seasonal variations, with the ultimate goal of enhancing energy efficiency and accuracy across various sectors. This phase's accomplishments have positioned us for success in the subsequent stages of our "Measure Energy Consumption" project.