

# Correlating Weather Patterns with Residential Electricity Consumption



Here you will apply the skills you've learned in data science, statistical analysis, and data management to find the relationship between **residential electricity consumption** in New Jersey and **weather patterns**.

You will be using two datasets—one containing **monthly electricity sales data** and another with **corresponding weather conditions** (average temperature and precipitation). Your goal will be to explore the relationship between electricity usage and weather conditions by using **linear regression** to evaluate and predict consumption trends.

## Objectives

- *Data Cleaning and Preprocessing*
- *Merging the Datasets*
- *Exploratory Data Analysis (EDA)*
- *Building a Predictive Model*
- *Evaluating and Improving the Model*

## Part 1: Data Cleaning and Preprocessing

- Load both datasets (**residentialElectricity.csv** and **SM\_betweenmonths.csv**) into a Jupyter Notebook using Pandas.
- Identify missing values and decide how to handle them (e.g., dropping rows or imputing values).

- Convert dates into the correct datetime format and ensure consistency between the two datasets.

## Part 2: Merging the Datasets

- Align the weather data with the electricity data by matching each row with its corresponding period. You will need to use the **period** field for the electricity dataset and the **Date** field for the weather dataset.
- Ensure the merged dataset contains columns for all features: electricity consumption, average max temperature, average min temperature, mean temperature, and precipitation.

## Part 3: Exploratory Data Analysis (EDA)

- **Visualizations:** Plot monthly electricity usage against average temperatures and other weather features to observe trends.
- **Statistical Summaries:** Generate summary statistics of both datasets, such as average sales, temperatures, and precipitation levels.
- Identify any notable trends or seasonal effects in the data.

## Part 4: Building a Predictive Model

- **Feature Selection:** Decide which features from the merged dataset should be used to predict electricity usage (e.g., temperature and precipitation).
- **Model Training:** Split the data into training and test sets. Train a linear regression model to predict electricity consumption using average mean temperature, precipitation, and other selected features.
- **Manual R-squared Calculation:** After fitting your model, use the predicted and actual values to manually calculate the R-squared value, demonstrating your understanding of the statistical concepts involved.

## Part 5: Evaluating and Improving the Model

- Evaluate the model performance by comparing the **R-squared** value calculated manually and the one provided by Scikit-Learn.
- Discuss potential improvements to the model. What other features or adjustments could improve predictive power?