

ENR3312: Assignment 4 on Data Visualization

Instructor: Honggeun Jo, Teaching Assistant: Eunsil Park, Junghwan Seol

Please load the solar farm energy data from the file `processed_solar_farm.csv`. After analyzing the data, follow the instructions and prepare a detailed report (3-5 pages). Make sure to include your Python script in a Jupyter notebook (*.ipynb) format. Submit both the report and the notebook by April 4th, 9 PM.

Assignment Instructions:

1. Add a New Feature to the DataFrame Based on Temperature:

- Create a new feature called **"weather"** in the DataFrame based on the temperature column (T).
- Define the "weather" feature as:
 - **'warm'** if temperature (T) is greater than 15.
 - **'cold'** if temperature (T) is less than or equal to 15.

2. Compute Descriptive Statistics of the "Weather" Feature:

- Manually compute the following statistics for the "weather" feature:
 - Mean
 - Variance
 - Standard Deviation
 - Skewness
 - Kurtosis

3. Compare Manual Computations with Results from Standard Libraries:

- Using a standard module (e.g., `scipy` or `statistics`), compute the same statistics listed in Step 2.
- Compare the results of your manual calculations with those produced by the external library/module.

4. Visualization: Histogram & Boxplot of generated_power_kw by "Weather":

- Plot a **histogram** and a **boxplot** of the generated_power_kw variable, with separate labels for each "weather" category (i.e., 'warm' and 'cold').
- Ensure that the plots are clearly labeled with proper titles, axis labels, and legends if necessary.

5. Correlation Analysis between generated_power_kw and Temperature:

- Compute the following correlation measures between generated_power_kw and temperature (T):
 - **Covariance**
 - **Pearson Correlation Coefficient**
 - **Spearman Rank Correlation Coefficient**
- Compute these metrics manually and also using an external module (e.g., scipy.stats).
- Provide a brief explanation of **Spearman's rank correlation coefficient**—its meaning and how to compute it.

6. Heatmap of Correlation Coefficients:

- Generate a **heatmap** to visualize the correlation coefficients:
 - Include both **Pearson** and **Spearman** correlation matrices.
 - Ensure the heatmap is well-labeled with clear axis titles and a color scale.

7. Scatter Plots:

- Plot **scatter plots** between generated_power_kw and each of the variables in the DataFrame.
- Each scatter plot should have appropriate labels for the axes and a title.

Report Guidelines:

- **Length:** The report should be between 3-5 pages in length.
- **Content:**

- Include your interpretation of each figure (histogram, boxplot, correlation heatmap, scatter plots).
 - Discuss the significance of your findings, such as the relationship between generated_power_kw and temperature, and any other interesting insights you observe.
 - Ensure that all statistical results and figures are clearly explained.
- **Figures:** Insert the generated plots directly into the report with proper captions and references in the text.