Project Objective

This project aims to reproduce the outcomes and figures from the paper titled "Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries"[1]. Students will engage in data cleaning, data imputation, conducting statistical and machine learning analyses, and critically evaluating the outcomes in comparison to the original study.

You work should cover(but not be limited to) the following.

Data Preparation

- Data Cleaning: Ensure the dataset is free of errors or irrelevant data points. This includes handling outliers, missing values, and incorrect entries.
- Data Imputation: Apply appropriate techniques(Try MissForest) to handle missing data ensuring the integrity and consistency of the dataset.

Analysis Requirements

1. Spearman Heatmap Analysis: Conduct Spearman correlation analysis to explore relationships between variables and visualize these relationships in the form of a heatmap.
2. Principal Component Analysis (PCA): Perform PCA to identify the principal components and explain the variance they capture in the data.
3. *Machine Learning Models:*
   - Implement four different machine learning models to analyze the data. Suggested models include:
     - Support Vector Machine (SVM)
     - Random Forest Regressor
     - Gradient Boosting Regressor
     - Another model of your choice
   - Evaluate the models based on their performance metrics (e.g., R^2, RMSE) and select the best performing model.
4. *SHAP Analysis: Use SHAP (SHapley Additive exPlanations) to interpret the model selected as performing the best, focusing on the contribution of each feature to the predictive outcomes.*

[1] Yang, Y., Shahbeik, H., Shafizadeh, A., Masoudnia, N., Rafiee, S., Zhang, Y., Pan, J., Tabatabaei, M. and Aghbashlo, M., 2022. Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries. *Renewable Energy*, *201*, pp.70-86.