## Assignment: Advanced Model Training and Evaluation

For this assignment, you'll be working with the [Wine Quality dataset](https://www.kaggle.com/datasets/rajyellow46/wine-quality), which contains physicochemical tests and quality ratings for both red and white wines. Your task is to predict the quality of wine (a regression task) based on its physicochemical properties.

**Data Preparation:**

* Load the Wine Quality dataset.
* Handle any missing values if they exist.
* Split the dataset into training and testing sets.

**Modeling with Hyperparameter Tuning:**

* Train and tune an SVM, KNN, and Gradient Boosting model using GridSearchCV.
* For each model:
* Define a set of hyperparameters to search.
* Use GridSearchCV to find the best hyperparameters.
* Evaluate the best model on the test set.
* Print the performance metric (e.g., Mean Squared Error) for the test set.

**Model Comparison**:

* Compare the performance of the three models.
* Which one performed best?
* Why do you think it performed best?

**Bonus Challenges (Advanced):**

**Feature Engineering**:

* Create new features that might improve the performance of your models.
* For example, can you create interaction terms or polynomial features?

**Ensemble Methods**:

Combine the predictions of your three models to create an ensemble. Does this improve performance?

*Hint: Consider techniques like weighted average, stacking, etc.*

**Interpretability**:

* Can you explain which features are most important in predicting wine quality?

*Hint: Consider techniques like SHAP values, feature importance from tree-based models, etc.*

*Data Visualization:*

* Create meaningful visualizations that can help in understanding the dataset better. For example, which features correlate most with wine quality?

**Deliverables**:

1. Jupyter Notebook:
   1. Your notebook should contain all the code, visualizations, and explanations for each step.
2. Report:
   1. A brief report (1-2 pages) summarizing your findings, insights, and the decisions you made throughout the assignment.
   2. Discuss your results from the bonus challenges and any other interesting observations.

**Evaluation Criteria:**

* Code Quality and Clarity: Your code should be clean, well-organized, and commented.
* Data Exploration and Cleaning: Handling of missing values, anomalies, and data splits.
* Model Training and Evaluation: Proper implementation and use of GridSearchCV, and evaluation metrics.
* Interpretation: Insights from your results, understanding of model performance, and the impact of features.
* Bonus Challenges: Implementation and results from the advanced tasks.