**Dry Bean Classification Assignment**

**Background Story:** You are hired by an agricultural research center to build an automated system that classifies dry beans into seven different species based on their morphological features. Manual sorting is slow and prone to errors, so the center is moving towards ML-driven solutions.

**Dataset:** Dry Bean Dataset

* 13,611 samples
* 16 features
* Target: 7 bean types (e.g., Seker, Barbunya, Bombay, etc.)

**Instructions:**

* Perform thorough EDA (distribution plots, boxplots, pairplots)
* Train and evaluate:
  + Logistic Regression
  + Decision Tree
  + Random Forest
  + Gradient Boosting
  + XGBoost
* Use Cross-Validation for evaluation
* Use RFE to find the minimal feature subset
* Apply PCA for visualization
* Apply hyperparameter tuning for Gradient Boosting and XGBoost using GridSearchCV or RandomizedSearchCV
* Visualize confusion matrix and class-wise accuracy

**Objectives:**

* Identify the most important features for bean classification
* Evaluate whether high accuracy can be achieved with fewer features
* Recommend a model for deployment in automatic bean sorting lines
* Present a 2D PCA plot showing bean class separation

**Deliverables:**

* Jupyter Notebook with clean, well-documented code