**Forest Cover Type Assignment**

**Background Story:** You are working for a national forest service tasked with automatically classifying forest cover types based on cartographic variables. The service wants to use ML to predict forest cover types for conservation and land-use planning.

**Dataset:** Forest Cover Type Dataset

* ~5,000 sample subset (downsample the full dataset for simplicity)
* 54 features
* Target: 7 cover types (Spruce/Fir, Lodgepole Pine, etc.)

**Instructions:**

* Conduct EDA including class distribution, correlation analysis, and visualizations
* Train and evaluate:
  + Logistic Regression
  + Decision Tree
  + Random Forest
  + Gradient Boosting
  + XGBoost
* Use Cross-Validation throughout
* Apply RFE to reduce feature space
* Apply PCA for dimensionality reduction and create a 2D scatter plot
* Use hyperparameter tuning (GridSearchCV or RandomizedSearchCV) for Gradient Boosting and XGBoost
* Visualize confusion matrices and feature importances

**Objectives:**

* Recommend a classification model for use in national forest monitoring
* Suggest whether fewer features can maintain predictive accuracy
* Provide intuitive visualization of cover type separation

**Deliverables:**

* Jupyter Notebook (well-structured with commentary and visual outputs)