Dataset used:
https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention
Prompts Used to Develop the Code:
1. "Can you help me load a CSV dataset into a pandas dataframe and inspect the first few rows?"
2. "Please show me how to clean this dataset by removing duplicates and checking for missing values."
3. "The dataset contains multiple student outcomes. How can I convert the target variable into a binary classification label for graduate vs dropout?"
4. "Can you generate some exploratory data analysis plots to understand the distribution and correlations within the dataset?"
5. "Show me how to split the dataset into training and testing sets using stratification."
6. "Can you apply feature scaling so models that depend on distance or gradients perform correctly?"
7. "Please help me train multiple machine-learning models and compare their accuracy, precision, recall, and F1-score."
8. "Can you make a clear benchmark table and plot to compare the models visually?"
9. "Help me identify the best-performing model and generate a confusion matrix."
10. "How can I compute and visualize feature importance for this dataset?"
11. "I want to create a function that takes a student profile as input and predicts whether they will graduate or drop out."
12. "Can you write a function that generates a random student from the dataset so I can test the model on new profiles?"
13. "Please create a simulation tool that predicts outcomes for multiple random students and summarizes the results."
14. "How can I write a set of unit tests to automatically validate the model’s behavior?"
15. "Please help me format the outputs so they look clean and readable for presentation."