Heart Disease project

🔍 Data Exploration & Visualization

1. How many patients in the dataset have heart disease (`target = 1`) vs those who don't? Plot a bar chart.
2. What percentage of males vs females had heart disease? Use a countplot or pie chart.
3. Compare heart disease rates across chest pain types (`cp`). Which type is most associated with heart issues?
4. Plot the distribution of age. Then, compare the average age of patients with and without heart disease.
5. Do patients with high fasting blood sugar (`fbs`) show higher heart disease risk?
6. What is the heart disease rate among patients with above-average cholesterol (`chol`) levels?
7. Create a new column categorizing patients into "young" (Age < 40), "middle-aged", and "elderly" (Age > 60). Which group is most affected?
8. Check for missing values in the dataset. How were they handled before modeling?
9. Create a heatmap showing correlation between numerical features. What medical or biological insights can you draw?
10. Compare heart disease rates by slope of the ST segment (`slope`). Which slope value shows the highest risk?

□ Modeling with Logistic Regression & Decision Tree

11. Train a Logistic Regression model and print the accuracy.
12. Train a Decision Tree Classifier. How does its accuracy compare to Logistic Regression?
13. Use `.predict()` to make predictions for the first 10 rows. How do the predicted values compare to actual results?
14. Plot and interpret the confusion matrix for both models. What does it tell you about false positives and negatives?
15. Use `classification_report()` to print precision, recall, and F1-score. Which model performs better on recall?
16. Use `cross_val_score` to evaluate each model. What's the average accuracy across all folds?
17. Try changing `max_depth` in the Decision Tree. How does it affect accuracy and overfitting?
18. Based on model scores, which features appear most important in predicting heart disease?

□ Critical Thinking & Interpretation

19. What 3 features seem to most influence heart disease prediction based on your analysis and model?
20. If you were to improve this model, what additional patient data (features) would you want to collect and why?
