**Instructions:**

1. Load this dataset into a Pandas DataFrame and perform data exploration using Pandas and NumPy.
2. Utilize Matplotlib and Seaborn to create visualizations for better data understanding.
3. Apply the concepts of K-Nearest Neighbors (KNN), Decision Trees (DT), and Logistic Regression to build predictive models.
4. Perform Exploratory Data Analysis (EDA) on the dataset to gain insights.
5. Answer the following questions:

**Pandas and NumPy Questions (15):**

1. How many rows and columns are there in the dataset?
2. What is the average age of the patients?
3. How many patients are male and how many are female?
4. What is the maximum value of cholesterol (chol) in the dataset?
5. How many patients have a resting blood pressure (trestbps) greater than 140?
6. Calculate the correlation between age and maximum heart rate achieved (thalach).
7. Create a new DataFrame with columns 'age', 'sex', 'chol', 'thalach', and 'target'.
8. Add a new column 'chol\_category' to the DataFrame based on the cholesterol values: 'normal' if chol < 200, 'borderline' if 200 <= chol < 240, and 'high' if chol >= 240.
9. Replace the target values: 0 as 'no disease' and 1 as 'disease' in the 'target' column.
10. Create a crosstab to show the relationship between 'sex' and 'target'.
11. Calculate the mean and standard deviation of 'oldpeak' for each target class.
12. Replace the 'thal' values with corresponding strings: 1 as 'fixed defect', 2 as 'normal', 3 as 'reversible defect'.
13. Drop the 'slope' column from the DataFrame.
14. Split the DataFrame into input features (X) and the target variable (y).
15. Normalize the input features using Min-Max scaling.

**Machine Learning Questions (20):**

1. Split the data into training and testing sets (80-20 ratio) for all features (X) and the target variable (y).
2. Build a K-Nearest Neighbors (KNN) classifier with k=5 and train it on the training data.
3. Evaluate the KNN classifier on the testing data and calculate its accuracy.
4. Build a Decision Tree classifier and train it on the training data.
5. Evaluate the Decision Tree classifier on the testing data and calculate its accuracy.
6. Build a Logistic Regression classifier and train it on the training data.
7. Evaluate the Logistic Regression classifier on the testing data and calculate its accuracy.
8. Compare the accuracies of all three classifiers (KNN, Decision Tree, Logistic Regression).
9. Plot a bar chart to show the feature importances in the Decision Tree classifier.
10. Visualize the confusion matrix for the Logistic Regression classifier.
11. Calculate precision, recall, and F1-score for the Logistic Regression classifier.
12. Plot an ROC curve for the Logistic Regression classifier and calculate the AUC score.
13. Use cross-validation (k=5) to evaluate the Decision Tree classifier and report the mean accuracy.
14. Perform feature scaling using StandardScaler and train the KNN classifier again. Compare its accuracy with the previous result.
15. Implement GridSearchCV to find the best hyperparameters for the Logistic Regression classifier (C and penalty) using a 5-fold cross-validation.
16. Apply the trained Logistic Regression model on the following new data:

bashCopy code

age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal 63 1 3 145 233 1 0 150 0 2.3 0 0 1

What is the predicted target for this patient?

1. Apply the Decision Tree classifier on the same data and compare the predicted target.
2. Calculate the Pearson correlation coefficient between 'thalach' and 'oldpeak'.
3. Visualize the distribution of 'thalach' using a histogram.
4. Create a box plot to show the distribution of 'chol\_category' across the target classes.

Please work on the assignment, and if you have any questions or need further assistance, feel free to ask. Good luck!