Concise Explanation of Code

1. **Data Loading**: load the dataset from a CSV file using pandas.
2. **Data Preprocessing**:
    - Basic dataset information is displayed, including data types.
    - Missing values are checked and counted for each column.
    - Unnecessary columns ("id" and "last_major_derog_none") are dropped.
    - Missing values in the "dti" column are filled with the mean.
    - Categorical columns ("home_ownership", "purpose", "grade") are one-hot encoded.
    - The "term" column is cleaned, removing "months" and converting to integers.
3. **Data Balancing**:
    - The dataset is split into normal and fraud classes.
    - Additional samples are generated from the fraud class to balance the data.
    - The original normal and additional fraud samples are combined and shuffled.
4. **Data Splitting**: The data is split into training and testing sets, preserving class distribution.
5. **Model Training**: A Random Forest Classifier model is initialized and trained on the training data.
6. **Model Evaluation**:
    - Numerical features are standardized.
    - After Oversampling , Random Forest Classifier model is trained on the standardized data.
    - Predictions are made on the test set.
7. **Metrics Calculation**:
    - Classification metrics (accuracy, precision, recall, F1-score) are calculated for model evaluation.
8. **Model Performance Reporting**:

- Evaluation metrics are printed to assess the model's performance.
- A classification report is displayed with precision, recall, F1-score, and support.
- The confusion matrix is printed, showing true positives, true negatives, false positives, and false negatives.

This workflow covers data preparation, model training, evaluation, and reporting for credit risk modeling.