**1. Data Loading and Exploration**

The project starts by loading a dataset of financial transactions. Initial exploration of the data helps identify its structure, including column names, data types, and statistical summaries. This step ensures a clear understanding of the dataset before processing.

**2. Handling Missing Values**

Missing values are addressed to ensure data completeness:

* **Categorical Columns**: Missing values are replaced with the most common value (mode), ensuring consistency in categorical data.
* **Numerical Columns**: Missing values are replaced with the median, preserving the central tendency of the data without introducing extreme values.

**3. Data Transformation**

* **Binary Columns**: Columns representing binary outcomes, such as isFraud and isFlaggedFraud, are converted to integer data types for compatibility with machine learning algorithms.
* **Categorical Encoding**: Categorical variables, such as transaction types, are encoded into numerical values for machine learning models that require numerical input.

**4. Duplicate Detection**

The dataset is checked for duplicate rows, which are identified and reported. This step ensures data quality and avoids bias during model training.

**5. Feature Selection**

Relevant features for prediction are identified. Columns that do not contribute meaningfully to the classification task, such as account names, are excluded from the analysis. This reduces noise and focuses the model on relevant data.

**6. Train-Test Split**

The dataset is split into training and testing sets. This separation allows the model to learn patterns from the training data and be evaluated on unseen test data, ensuring it generalizes well.

**7. Model Training**

A Random Forest Classifier is used as the machine learning model. It is well-suited for classification tasks due to its ability to handle both numerical and categorical data and its resistance to overfitting.

**8. Model Evaluation**

The trained model is evaluated using several metrics:

* **Accuracy Score**: Measures the overall correctness of the model.
* **Classification Report**: Provides precision, recall, and F1-score for each class, offering insights into how well the model distinguishes between fraudulent and non-fraudulent transactions.
* **Confusion Matrix**: Summarizes prediction results, showing the number of correct and incorrect classifications for each class.

**9. Visualization**

A confusion matrix is visualized using a heatmap. This graphical representation makes it easier to interpret the model's performance by highlighting true positives, false positives, true negatives, and false negatives.