1. Data Cleaning: Handling Missing Values, Outliers, and Multicollinearity

Done

2. Describe Your Fraud Detection Model:

- The fraud detection model used is **Logistic Regression** since the target variable (isFraud) is binary (fraud or not fraud).
- **Model Goal**: Predict whether a transaction is fraudulent based on features like transaction amount, balances, and transaction type.

Steps Taken:

1. Standardized the dataset using StandardScaler.
2. Applied logistic regression to classify transactions.
3. Evaluated the model using performance metrics like accuracy, precision, recall, and F1-score.

3. Variable Selection:

- Variables were selected based on domain knowledge and feature importance (coefficients).
- Highly correlated features were removed using VIF, and those with low relevance were excluded based on logistic regression coefficients.
- The final selected variables are:
    - newbalanceOrig, oldbalanceOrg, newbalanceDest, oldbalanceDest, amount, type_CASH_OUT, and type_TRANSFER.

4. Model Performance Demonstration:

Performance Metrics:

- Accuracy, Precision, Recall, F1-Score

5. Key Factors Predicting Fraudulent Customers:

- **Top predictive features (based on coefficients):**
    1. newbalanceOrig (-43.28)

2. oldbalanceOrg (38.65)
3. newbalanceDest (-14.55)
4. oldbalanceDest (12.45)
5. type_CASH_OUT (3.85)

These indicate that changes in balances before and after transactions are critical in predicting fraud.

6. Do These Factors Make Sense?

Yes, they make sense:

- Fraudulent transactions often involve large changes in account balances (newbalanceOrig, oldbalanceOrg).
- **type_CASH_OUT** and **type_TRANSFER** are common in fraud cases where funds are moved quickly to external accounts.
- newbalanceDest suggests that legitimate transactions usually update balances correctly, whereas fraud might leave them inconsistent.

7. Prevention Recommendations for Infrastructure Updates:

- Implement real-time **fraud detection systems** using machine learning models.
- Enhance **monitoring of large balance changes** or suspicious transaction types like CASH_OUT.
- Introduce **multi-factor authentication** for high-risk transactions.
- Use **behavioral analysis** to detect unusual transaction patterns.

8. Measuring Effectiveness of Implemented Actions:

- **Track fraud rate reduction** before and after implementation.
- **Monitor false positives and false negatives** to assess model accuracy.
- Evaluate **customer feedback** on security enhancements.