Financial Transactions Anomaly Detection(TASK-3)

Objective: Develop a system to identify and report anomalies in a dataset of financial transactions.

Task Description:

You are tasked with creating a Python script that processes a dataset of financial transactions to detect potential anomalies. This involves preprocessing data, calculating statistical thresholds, and flagging transactions that significantly deviate from typical patterns.

Requirements:

1. **Data Preprocessing:**
   - **Loading Data:** Load the dataset into a Pandas DataFrame.
   - **Cleaning Data:** Handle missing values appropriately. Convert necessary columns to their respective data types (e.g., dates to datetime, amounts to numeric).
2. **Statistical Analysis:**
   - **Calculate Category Statistics:** Compute statistical metrics such as mean and standard deviation for transaction amounts grouped by categories.
   - **Establish Thresholds:** Use statistical methods (e.g., Z-score) to establish thresholds for anomaly detection.
3. **Anomaly Detection:**
   - **Implement Logic:** Develop logic to detect anomalies based on the thresholds established in the statistical analysis.
   - **Types of Anomalies:** Detect unusually high transaction amounts, unexpected frequency patterns, or other deviations from expected norms.
4. **Reporting:**
   - **Generate Anomaly Report:** Create a detailed report listing all detected anomalies.
   - **Report Details:** Include transaction ID, date, category, amount, and a clear reason for flagging each transaction as an anomaly.

Implementation Overview:

Step-by-Step Approach:

1. **Data Loading and Preprocessing:**
   - Load the dataset into a DataFrame.
   - Check for missing values and handle them (e.g., dropping or imputing missing data).
   - Convert data types (e.g., dates to datetime objects, amounts to numeric values).
2. **Statistical Analysis:**
   - Calculate statistical metrics (mean, standard deviation) for transaction amounts grouped by categories using Pandas functions.
   - Use these metrics to define thresholds for anomaly detection. Common methods include Z-score, IQR (Interquartile Range), or domain-specific rules.
3. **Anomaly Detection:**
   - Implement logic to compare each transaction against the established thresholds.
   - Flag transactions that exceed the defined thresholds as anomalies.
   - Store information about detected anomalies (transaction ID, date, category, amount, reason for anomaly).

4. **Reporting:**
   ○ Generate a comprehensive report that summarizes all detected anomalies.
   ○ Display relevant details for each anomaly, making it clear why each transaction was flagged.
   ○ Use visualizations (e.g., boxplots, histograms) to enhance understanding and visualization of anomalies.