Anomaly Detection Report

1. Methodology

This report presents an analysis of anomaly detection techniques applied to financial transaction data. The two methods used are:
- Z-Score Method
- Isolation Forests Algorithm

1.1 Z-Score Method

The Z-Score method calculates the standard deviation of each data point from the mean. Data points that exceed a threshold (e.g., ±3 standard deviations) are considered anomalies. This method works well when the data follows a normal distribution but may struggle with skewed data.

1.2 IsolationForest Algorithm

IsolationForest is an unsupervised machine learning algorithm that isolates anomalies by randomly partitioning data points. Since anomalies are rare and different, they are isolated faster than normal instances. This method is effective for high-dimensional and skewed datasets.

2. Comparison Analysis

Both Z-Score and IsolationForest have their strengths and weaknesses:

Method	Advantages	Disadvantages
Z-Score	Simple, fast, and interpretable. Works well on normally distributed data.	Struggles with skewed or non-Gaussian distributions.
IsolationForest	Handles high-dimensional and non-Gaussian data well. Robust to outliers.	Requires tuning of contamination parameter. Can misclassify rare normal points as anomalies.

3. Insights Derived from the Analysis

1. **Anomaly Distribution**: The dataset is highly imbalanced, making supervised approaches challenging. Both methods identified outliers, but IsolationForest performed better in handling data imbalance.
2. **Use-Case Suitability**:
- Z-Score is useful when the data follows a normal distribution, and we want an explainable metric.
- IsolationForest is better for large, complex, and high-dimensional datasets.
3. **Real-World Impact**: Anomaly detection can prevent fraudulent transactions, reduce

financial losses, and improve security in financial institutions. Combining both methods can enhance accuracy.