Dataset Description

The dataset you provided contains various columns related to financial audit data. Here’s a breakdown of the columns and how you can start the project using this data:

**Columns Description:**

1. **Sector\_score**: A score representing the sector’s performance or risk.

2. **LOCATION\_ID**: An identifier for the location of the transaction or audit.

3. **PARA\_A, PARA\_B, PARA\_C, …**: Parameters A, B, C, etc., representing different metrics or features being audited.

4. **Score\_A, Score\_B, …**: Scores associated with the respective parameters.

5. **Risk\_A, Risk\_B, …**: Risk values associated with the respective parameters.

6. **TOTAL**: Total score or risk value.

7. **numbers**: Possibly the number of transactions or audits.

8. **RiSk\_E**: Another risk metric.

9. **History**: Historical data or flag.

10. **Prob**: Probability metric.

11. **Risk\_F**: Another risk factor.

12. **Score**: Overall score.

13. **Inherent\_Risk**: The risk that exists before any controls are applied.

14. **CONTROL\_RISK**: The risk that is managed by applying controls.

15. **Detection\_Risk**: The risk of not detecting a significant error.

16. **Audit\_Risk**: The overall risk from the audit.

17. **Risk**: A binary flag indicating high (1) or low (0) risk.

**Steps to Start the Project**

1. **Load and Understand the Data**:

Load the data into a pandas DataFrame and understand its structure. Check for missing values and decide on a strategy to handle them.

2. **Feature Engineering**:

Create additional features if needed. For example, you can combine some scores or risks to create composite metrics.

3. **Model Training**:

Use an anomaly detection algorithm like Isolation Forest to detect unusual patterns.

4. **Evaluation and Visualization**:

Evaluate the model’s performance and visualize the results to interpret the anomalies.