## 

### **📦 Project Title**

**"Detect Anomalous Transactions in Unstructured Financial Records"**

### **📜 Dataset Description**

The dataset consists of **raw transaction logs** with minimal structure. Each row is a log entry from a financial system or a user-facing event log. Example formats:

{

"log": "2023-05-14 14:05:31 | user: 1023 | txn: withdrawal of £500 from ATM near Liverpool | device: Samsung Galaxy S10 | location: 53.4084,-2.9916"

}



### **🎯 Objectives**

#### **1. Data Understanding & Parsing**

#### **2. Feature Engineering**

#### **3. Anomaly Detection**

Choose one or more:

* **Rule-based** (e.g., amount > X + new location → suspicious)
* **Statistical / clustering** (e.g., DBSCAN, Isolation Forest)
* **Sequence modeling** (if temporal sequences are valuable)
* **Embedding + autoencoders** (textual descriptions vectorized)

#### **4. Evaluation**

* If labeled anomalies are available: precision, recall, etc.
* If not:  
  + Flag top-N anomalies with explanations
  + Visualize clusters or outliers
  + Provide a manual validation plan

#### **5. Interpretability**

* Explain what makes an event anomalous (e.g., rare device+location pair)
* Include confidence scores if possible

### **💾 Deliverables**

* Code (in Jupyter notebook or script)
* Parser module for extracting structured features
* Notebook visualizations of detected anomalies
* README with:  
  + Parsing logic explanation
  + Modeling approach
  + Findings & business impact
* Optional: web demo (Streamlit, Flask) or CLI to parse and flag logs

### **✅ Evaluation Rubric**

| **Area** | **Criteria** |
| --- | --- |
| Parsing & Cleaning | Robustness, generalization, edge-case handling |
| Feature Engineering | Creativity, domain relevance, statistical richness |
| Modeling | Sound method, parameter tuning, justification |
| Evaluation | Clear metrics, qualitative and quantitative results |
| Explainability | Interpretability of flagged anomalies |
| Code Quality | Modularity, clarity, reproducibility |
| Business Thinking | Actionable insights, real-world impact articulation |