**How to Progress with These Data Sources (Bank, Mpesa, Call Logs)**

We now have **rich feature sets** from each service. The next step is to integrate them into a **rule-based scoring engine** (and later a hybrid ML model). Here’s a clear roadmap:

**1) Bank Account Features → Rules**

**Indicators of financial health & risks**

* **Income & Liquidity**
  + average\_balance ≥ 1500 → **+3 points**
  + closing\_to\_opening\_ratio ≥ 1.05 → **+1 point** (growth in savings)
  + average\_to\_closing\_balance\_ratio ≥ 0.8 → **+1 point**
* **Cashflow & Activity**
  + net\_cash\_flow > 0 → **+2 points**
  + percentage\_active\_days ≥ 0.5 → **+1 point**
  + average\_transactions\_per\_active\_day ≥ 2 → **+1 point**
* **Risk Signals**
  + has\_bounced\_cheques = 1 → **–2 points**
  + total\_charges\_amount > 3% of turnover → **–1 point**
  + balance\_volatility\_std\_dev > 2× average\_balance → **–1 point**
* **Profile Flags**
  + has\_salary\_inflow = 1 → **+1 point**
  + has\_loan\_repayment = 1 **and** net\_cash\_flow > 0 → **+1 point**
  + has\_betting\_transactions = 1 → **–2 points**
  + has\_single\_dominant\_beneficiary = 1 **and** withdrawals ratio > 60% → **–1 point**

**2) Mpesa Features → Rules**

**Indicators of mobile money behavior**

* **Volume & Frequency**
  + total\_transactions > 500 → **+2 points**
  + avg\_transaction\_size > 200 → **+1 point**
  + total\_inflow > total\_outflow → **+2 points**
* **Spending Patterns**
  + merchant\_ratio > 0.25 → **+1 point** (productive use)
  + paybill\_ratio > 0.4 → **+1 point**
  + airtime\_ratio > 0.3 → **–1 point** (too much consumption on airtime)
* **Risk & Stability**
  + balance\_volatility too high (e.g., > 3× avg\_balance) → **–1 point**
  + Very low end\_balance (< 100) despite high inflows → **–1 point**
  + recurring\_payments > 10 → **+1 point** (stability)

**3) Call Logs Features → Rules**

**Indicators of social stability & reliability**

* **Frequency & Engagement**
  + call\_frequency > 15/month → **+1 point**
  + call\_duration > 60 minutes total → **+1 point**
  + active\_behavior ≥ 1.2 (consistent engagement) → **+1 point**
* **Network Stability**
  + stable\_contacts\_ratio > 0.7 → **+1 point**
  + unique\_contacts high & not just missed\_only → **+1 point**
* **Risk Patterns**
  + missed\_only > 0.5 → **–1 point**
  + night\_vs\_day > 0.3 (too many night calls) → **–1 point**
  + Very high regular\_patterns\_std (unstable calling habits) → **–1 point**

**4) Integration into Scoring**

* Each domain (Bank, Mpesa, Call Logs) can contribute **up to 10–15 points**.
* Total score can be mapped into **credit decisions**:
  + **≥ 20 points** → Approved: **$500**
  + **15–19 points** → Approved: **$400**
  + **10–14 points** → Approved: **$150**
  + **< 10 points** → **Disapproved**

**5) Next Steps for the Project**

1. **Feature Mapping** → Build a unified table per customer (bank + mpesa + call logs).
2. **Rule Implementation** → Add these rules into your scoring engine (Python functions).
3. **Visualization** → Show how different features influence scores (bar charts, pie charts).
4. **Demo Preparation** → Present:
   * Example customer dataset
   * Step-by-step scoring explanation (why approved or disapproved)
   * Visual results