


Download this dataset in csv

 farmer\_loans

Upload to Claude, ChatGPT or Gemini

Copy this Prompt

I'm analyzing a dataset of loans from our agriculture program.  
Take a look at the attached CSV and tell me how the program is performing. What are the one biggest insight I should bring to leadership?  
No followup questions, give it your best shot.

## Prompt 2

Here some additional context to refine your analysis

### ## Program context

We're a nonprofit that loans seeds and fertilizer to smallholder farmers in Kenya. Loans are disbursed at the start of the planting season and repaid after harvest, with interest. Each loan is identified by loan\_id — a single farmer may take multiple loans across seasons or as top-ups, so farmer\_id can repeat across rows.

Program versions:

- v1: traditional loan only — 15% interest per season
- v2: loan + training — 18% interest per season
- v3: loan + training + crop insurance — 20% interest per season

### ## Data dictionary

- farmer\_id: unique farmer identifier (CAN REPEAT across rows — a farmer may have several loans)
- loan\_id: unique identifier for each loan. THIS is the unit of analysis for any "per-loan" metric.
- season: loan disbursement year

- region: Kenya region (North / South / East)
- gender: legacy entries use 1=Female, 2=Male; newer entries use F/M; blanks mean unknown
- plot\_acres: farm size in acres (NOT hectares)
- crop: crop planted with the loan
- program: program version — see context above
- loan\_amount\_kes: PRINCIPAL disbursed, in Kenyan Shillings.  
This is the only "denominator" you should use for repayment rates.
- expected\_total\_due\_kes: principal + expected interest at season end. Do NOT use this as a denominator for repayment rate — it's only for accounting / what we hope to collect.
- disbursement\_date: date loan was given. WARNING: column has mixed formats — some YYYY-MM-DD, some DD/MM/YYYY.
- status: loan outcome. Codes:
  - R = Repaid in full (principal AND interest)
  - PR = Partial Repayment (some principal, no interest)
  - OD = Overdue (still tracking, not yet defaulted)
  - D = Defaulted (will not be recovered)
  - W = Written off (drought-forgiven — EXCLUDE from rates)
  - P = Pending (not yet due — EXCLUDE from rates until resolved)
- principal\_repaid\_kes: principal repaid in KES. 0 is real data.
- interest\_repaid\_kes: interest repaid in KES. 0 is real data.
- loan\_purpose: free-text, NOISY (inconsistent capitalization, abbreviations). Do not segment by this without normalizing first.

Repayment rate is defined as:

- Unit of analysis: count of unique loan\_ids (one row = one loan)
- Numerator: count of loan\_ids where status = 'R'
- Denominator: count of loan\_ids EXCLUDING:
  - \* status = 'W' (drought-forgiven, not failures)
  - \* status = 'P' (pending, not yet due)
- Trust the status field; do not re-derive status from the repayment amounts (some 'R' loans may show 0 interest, etc.)

## My specific question

Leadership is interested into how loans are being repaid, to inform our 2025 fundraising pitch.

- 1) Calculate the **\*\*on-time full repayment rate by for 2024 season\*\***
- 2) Tell me in 3 short bullet points what was wrong in your past analysis