Background

One Acre Fund is a non-profit social enterprise that supplies financing and training to help smallholders grow their way out of hunger and build lasting pathways to prosperity. The organization consists of many departments, one of which is the Data Analytics Team. One of their main duties is to support Data Analysis needs of strategic program initiatives. The team handles customer data which is stored in a centralized database and later accessed through various exports that are used to perform various analyses.

Scenario: Financial Advisory Services is one of the departments supported by the Kenya Analytics Team. They have reached out to have a Loan Book designed that highlights key client metrics.

Task Notes:

This is a 4-hour test designed to assess your technical skills in Python, SQL, and data visualisation, as well as your ability to think critically and derive actionable insights from data.

You are required to:

1. Save and export the Jupyter notebook as a PDF, ensuring the Python code is well-documented and easy to follow before submitting.
2. Include SQL queries where appropriate.
3. Provide visualisations to support your analysis.
4. Make business recommendations based on your finding

If One Acre Fund wants to rebalance the portfolios so that no single agent has disproportionate influence, what strategy would you recommend? How would you redistribute the loans?You are provided with a dataset containing loan repayment information. Your first task is to derive key risk-related metrics to assess loan performance. Using Python, perform the following calculations and assign values to new columns in the dataset:

- PAR (Portfolio at Risk) Status. Assuming the Finance request came in today (the day you receive the exercise), assign each client a PAR Status based on their repayment progression. This should be as a new column labeled ‘PAR status’ in the dataset.

  Status to be assigned:

  - On Time: Contains clients; who as of today, are on track with their payments i.e the next contract payment due date is some time in the future.
  - PAR0-7: Contains clients; who as of today, are between 0 to 7 days past the next contract payment due date
  - PAR8-30: Contains clients; who as of today, are between 8 to 30 days past the next contract payment due date
  - PAR31-90: Contains clients; who as of today, are between 31 to 90 days past the next contract payment due date
  - PAR90+: Contains clients; who as of today, are more than 90 days past the next contract payment due date.
- Current Collection Rate which is derived by taking the Cumulative Amount Paid divided by (Expected Cumulative Amount Paid - Deposit)
- Derive each client’s total amount in arrears which is the expected amount to have been paid at this time minus what has been paid.
- Payment Progression for each client. This is cumulative amount paid divided by the nominal contract value
- Expected Payment Progression for each client. This is expected cumulative amount paid divided by the nominal contract value
- Derive loan type from ‘name’ column: using the ‘name’ column, create a new column called Loan Type. Any entry in the name column that contains ‘Individual’ is an Individual Loan, any entry that contains ‘Group’ is a Group Loan, any entry that contains ‘Paygo’ is a Paygo Loan and any entry that contains ‘Cash’ is a Cash Sale.

Task 2: Hidden Risk Analysis

One Acre Fund is concerned about hidden risks in the portfolio. The obvious risks are already managed, but your goal is to uncover clients or loans that appear low-risk but could become high-risk in the near future.

Deliverable:

1. Create an extra column in the dataset with the indicator(s) defined, and provide a summary of the clients flagged by each indicator(s)
2. Based on your analysis, propose a proactive strategy to address these hidden risks. What actions should One Acre Fund take to prevent these clients from falling into arrears? (Please insert a markdown and answer this in less than 200 words)

Task 3: Identifying High Impact Agents

Not all agents are created equal. Some agents manage loans that are more difficult to collect, while others manage relatively easy portfolios. Your goal is to identify high-impact agents—agents whose portfolios have the biggest effect on the organisation's overall financial health.

Deliverable:

1. Provide a list of the top 5 high-impact agents, explain your calculation of the Agent Impact Score, and visualize the distribution of the scores across all agents.
2. If One Acre Fund wants to rebalance the portfolios so that no single agent has disproportionate influence, what strategy would you recommend? How would you redistribute the loans? (Insert a markdown in your notebook, and discuss in less than 200 words)

Definitions of Columns in the dataset provided:

Column/Metric	Definition
Contract Reference	Uniquely identifies a contract between One Acre Fund and the client. A client can have multiple contracts.

Status	Loan repayment status. Active means the loan is on going, complete means the loan is fully paid and defaulted means the loan is defaulted
Start Date	Date when contract starts
End date	Date when a client completes repaying their loans
next_contract_payment_due_date	Date a client is expected to make a payment towards their loan
cumulative_amount_paid	Total amount repaid to date.
expected_cumulative_amount_paid	Total amount that should have been paid to date
nominal_contract_value	Total loan amount a client has signed up for
deposit_amount	Total amount a client paid to qualify for the loan
birthdate	A client's date of birth
gender	A client's gender
l3_entity_id	Id that denotes the region a client is enrolled from
Name	Loan type offering. Types of loan offerings: Individual Loan, Group Loan, PAYGO(Pay As You Go) Loan. Name is appended with these loan types]
Region	The geographical area where the client is based (e.g., Mt Kenya, Rift Valley, etc.).
Loan product Type	The type of loan provided (e.g., Agriculture Loan, Individual Loan, Paygo, etc.).
Agent ID	A unique identifier assigned to the agent responsible for managing the client's loan

Loan Interest Rates	The percentage rate applied to the loan, representing the cost of borrowing for the client.
Repayment Frequency	The frequency of loan repayments (Weekly, Monthly, etc.).
Credit Score	The total amount the client has repaid.