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:

- You will be required to use only Python and SQL for the below tasks
- After deriving each metric, print out the first 5 rows of the new dataframe
- Use comments to arrange and describe your work.
- For Task 1 and 3, you will be required to submit a jupyter notebook clearly demonstrating your working in completing the below tasks.
- For Task 2, submit the SQL Query either as a word document, .sql file or a markdown cell in jupyter notebook.
- Ensure your code is clean, well documented and modular(where applicable).

Task 1: Using Python read the provided dataset and derive the following metrics:

- 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: Let's assume that the provided dataset is in a relational database e.g Postgres or Microsoft sql, write an SQL query that assigns clients with PAR Statuses based on repayment, similar to the python task above. Note: This is just for the PAR Status and not all the above derived metrics.

Task 3: Using what you have derived above and the provided dataset, perform Exploratory Data Analysis in Python and share any key insights.

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