



Fairmoney Data Engineer Assignment - 2025

[DO NOT USE ChatGPT]

About FairMoney

FairMoney is a leading financial services provider in Nigeria. Over the past 7 years our android application has been downloaded by over 10 million people and we have been able to serve close to 4.5 million customers by providing them loans to fulfil their needs - two-thirds of whom have been small and micro businesses, using their loans to build and grow their businesses. We have disbursed over 16.8 million loans over this period. We have launched our lending product in other African countries like Uganda, Zambia last year.

FairMoney customer flow

1. User downloads the app & signs up.
2. User completes KYC steps & applies for a loan.
3. The user is either rejected or presented with a few loan offers.
4. The user either accepts one of the offers or declines them.
5. The user repays the loan.
6. If user defaults on the payment then collection executive is assigned to the user to ensure repayment of defaulted instalment.
7. The user re-applies for a new loan once previous loan is closed.
8. ... back to step 3

Problem Statement - Part A

You are working as a Data Engineer at FairMoney. Your task is to design a **relational data model** and an **Entity-Relationship Diagram (ERD)** for FairMoney's database. The design should be robust enough to address key business questions that stakeholders might want to answer from this data model.

What is the reason you chose normalized data schema or denormalized schema?

Problem Statement - Part B

Design the flow from raw data to reporting, including:

1. **Cleaning:** Handling nulls, duplicates, and standardising formats.
2. **Transforming:** Implementing business rules (e.g., calculating derived fields like days past due).
3. **Storing:** Organising the data into fact and dimension tables.

Use diagrams (e.g., flowcharts) or descriptions to illustrate the steps.

Problem Statement - Part C

1. Identify Key KPIs

Identify the top 5 Key Performance Indicators (KPIs) for the lending business that are critical for tracking its performance and health. For each KPI:

- Provide a clear definition and its business significance.
- Explain the formula or logic to calculate the KPI using the fact and dimension tables from the data model.

2. Write SQL Queries

Write SQL queries to calculate the identified KPIs using the analytical tables. Ensure the queries are efficient and optimized for scalability.

3. Bonus:

Creating and including sample data for your SQL queries will be considered a plus. Use either real or synthetically generated data that aligns with the data model structure.

Disclaimer:

Candidates are expected to complete this assignment independently. **Do not use ChatGPT or any other AI-based tools** to generate responses, SQL queries, or explanations. Submissions will be evaluated not only on correctness but also on originality, and any indications of AI-generated content may result in disqualification.