Financial Inclusion Marketing Campaign

# 🎯 Objective

Design an ETL (Extract, Transform, Load) pipeline using Python and SQL for a financial institution's marketing campaign data. The goal is to process raw customer, campaign, and transaction data to create a clean, integrated dataset for analytics.

# 🔄 ETL Pipeline Overview

The ETL pipeline will be implemented using Python and SQL as follows:

## 1. Extract

* Tools: Python (Pandas)
* Tasks:
* • Read raw CSV files: customers.csv, campaign.csv, and transactions.csv
* • Validate schema consistency and record counts

## 2. Transform

* Tools: Python (Pandas), SQL
* Tasks:
* • Clean missing and inconsistent data (e.g., nulls in response column)
* • Create derived metrics like total transactions per customer
* • Join data from multiple tables to create a single consolidated view
* • Apply business rules for campaign analysis (e.g., only deposits count towards campaign success)

## 3. Load

* Tools: SQL (PostgreSQL / SQLite)
* Tasks:
* • Create destination tables in a relational database
* • Insert transformed data into target tables
* • Enable indexes for performance optimization

# 📐 Data Model

* Target Tables:
* • customers (customer\_id, name, gender, age, region, income\_level, occupation)
* • campaign (campaign\_id, customer\_id, campaign\_type, response, channel, date)
* • transactions (transaction\_id, customer\_id, transaction\_type, amount, transaction\_date)
* • customer\_campaign\_summary (customer\_id, total\_amount, response\_flag, region, income\_level)

**🧮 Phase 2: SQL (Data Modeling & Querying)**

* Example Tasks:
* Find regions with the highest response rates.
* Determine transaction activity among campaign responders.
* Group campaign effectiveness by channel, income level, or occupation.

**📊 Phase 3: Visualization**

* Suggestions:
* Bar plots of response rates by region or income level.
* Line chart of transactions over time.
* Boxplot comparing transaction amounts for responders vs non-responders.