

# Capstone Finance Project – Full Checklist

**Project Goal:** Build a full-stack data pipeline for synthetic banking transactions, microloan analysis, and behavioral scoring.

## Step 0 - Environment Setup (Completed)

- Create project folder `capstone_finance/`
- Setup virtual environment `.venv` and install dependencies (`pandas`, `pyyaml`)
- Initialize Git repository (`git init`)
- Create `.gitignore` to exclude `__pycache__/, data/raw/, .venv/`
- Create `README.md`
- Create `requirements.txt`

## Step 1 - Data Ingestion (Completed)

- Define folder structure: `src/ingest/, data/raw/, config/`
- Create `config/config.yml` with `raw_dir`, `n_customers`, `start_date`, `end_date`, `seed`
- Write `generate_transactions.py` with:
  - Customer metadata (salary, EMI, rent, utilities)
  - Daily transactions (salary, rent, EMI, utilities, random spending)
  - Unique transaction IDs
  - Running balance computation
  - Partitioned CSV output `data/raw/YYYY/MM/DD/transactions.csv`
- Generate `data/customers.csv`
- Git: commit and push Step 1 branch

## Step 2 - Data Lake Structure (Next)

- Define Bronze / Silver / Gold folder structure
- Write loader script: raw CSV → bronze parquet
- Verify parquet partitioning

- Git: branch step2\_data\_lake, commit, push

## **Step 3 - Data Cleaning & Transformation**

- Handle missing or inconsistent values
- Standardize transaction categories
- Detect salary, rent, EMI events programmatically
- Compute features: avg daily balance, total debit/credit per month

## **Step 4 - Feature Engineering for Microloans**

- Identify cash crunch periods
- Compute repayment score, fixed obligations, spending habits
- Generate monthly features per customer
- Save features in gold zone for modeling

## **Step 5 - Modeling & Scoring**

- Build predictive model for short-term liquidity gaps
- Use features from gold zone
- Generate Foneloan candidate score
- Evaluate and validate model

## **Step 6 - Reporting & Dashboard**

- Aggregate metrics for each customer
- Generate dashboards / charts for decision makers
- Include alerts for Foneloan candidates

## **Step 7 - Deployment & Version Control**

- Git workflow: branch per step, PR/merge to main after verification

- CI/CD for ETL scripts (optional)
- Backup raw and processed data