
Project Migration & Data Loading Summary

Objective:
Migrate a microfinance dataset project to Replit and Supabase environments, ensuring database schema integrity and successful CSV data loading.

---

1. Project Setup

Verified project structure and logs for Replit compatibility.

Ensured Postgres database creation on Replit.

Resolved port conflicts and workflow issues.

Confirmed proper drizzle configuration for schema management.

---

2. Dataset Generation

Ran generate_kechita_dataset.py to generate CSV datasets for:

100 branches

777 officers

120,000 customers

150,000 loans

1,883,077 repayments

36,500 daily branch performance records

1,200 monthly branch summaries

9,324 officer performance records

6,000 fraud signals

72,000 AI customer features

---

3. Database Schema Management

Generated SQL schema for Supabase: supabase-schema-generated.sql.

Applied schema via npx drizzle-kit generate and npm run db:push.

Required user to create tables in Supabase before actual data load.

---

4. Loader Script Preparation

Dry-run mode added to scripts/load-data.ts for safe validation.

Detected camelCase vs snake_case column mismatch (avgTargetTier → avg_target_tier).

Added a mapping function to convert camelCase to snake_case for Supabase compatibility.

---

5. Data Loading

Ran loader in batches due to memory issues (especially large repayments table: 1.8M records).

Fixed duplicate record handling in Supabase, including tables without id columns.

Dry-run validation completed successfully; all CSVs parsed correctly.

Actual data load attempted but encountered memory limits; batch size adjustments needed.

Next Steps for Agent

1. Ensure Supabase tables are created using supabase-schema-generated.sql.

2. Validate loader mapping and batch processing for large tables.

3. Execute the data loader with batch adjustments to prevent memory issues.

4. Verify data integrity (row counts, duplicates, field formats) in both Replit Postgres and Supabase.

5. Update progress tracker and checkpoints.

How do we proceed here? Here are my supabase keys for you to see what's already been populated etc. And give me the next steps forward. Let's cut the data generation by half across all boards so that we get 6 month records but we need to fill in the Postgress and supabase so that it will always be the source of truth moving forward, and ensure the UI is properly tied to the db:

URl - https://xjbjwwzxhigynoltrthd.supabase.co
Anon -
eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJpc3MiOiJzdXBhYmFzZSIsInJlZiI6InhqYmp3d3p4aGlneW5vbHRydGhkIiwicm9sZSI6ImFub24iLCJpYXQiOjE3NjMzNDM2NDAsImV4cCI6MjA3ODkxOTY0MH0.J6YiCHq5HXWdIvv19D90LzRnafltfKmgpRTAqKRFTWI

Service Role -
eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJpc3MiOiJzdXBhYmFzZSIsInJlZiI6InhqYmp3d3p4aGlneW5vbHRydGhkIiwicm9sZSI6InNlcnZpY2Vfcm9sZSIsImlhdCI6MTc2MzM0MzY0MCwiZXhwIjoyMDc4OTE5NjQwfQ.NEGE9ZZ_-zw8AaVWppgpR8pZXnpE07HRFbYZB-Y1bC0