

ETL Banks Project – Project Summary

This project implements an automated ETL (Extract, Transform, Load) pipeline to compile a quarterly report of the top 10 largest banks in the world ranked by market capitalization.

Objectives

- Extract banking data from Wikipedia (archived source)
- Transform market capitalization values from USD to GBP, EUR, and INR
- Store processed data locally as CSV and in a SQLite database
- Enable repeatable, automated execution every financial quarter

ETL Workflow

1. Extraction: Scrape the 'By market capitalization' table using BeautifulSoup.
2. Transformation: Apply exchange rates and round results appropriately.
3. Load: Save results to CSV and SQLite database.
4. Query: Execute SQL queries to validate transformations.
5. Logging: Log each pipeline stage for traceability.

Technologies Used

Python, Pandas, BeautifulSoup, SQLite, NumPy, Requests, ReportLab

This project demonstrates core data engineering skills including web scraping, data transformation, database integration, logging, and automation-ready design.