

PROJECT 01

Acquiring and Processing information on world's largest banks Data

Project Description:

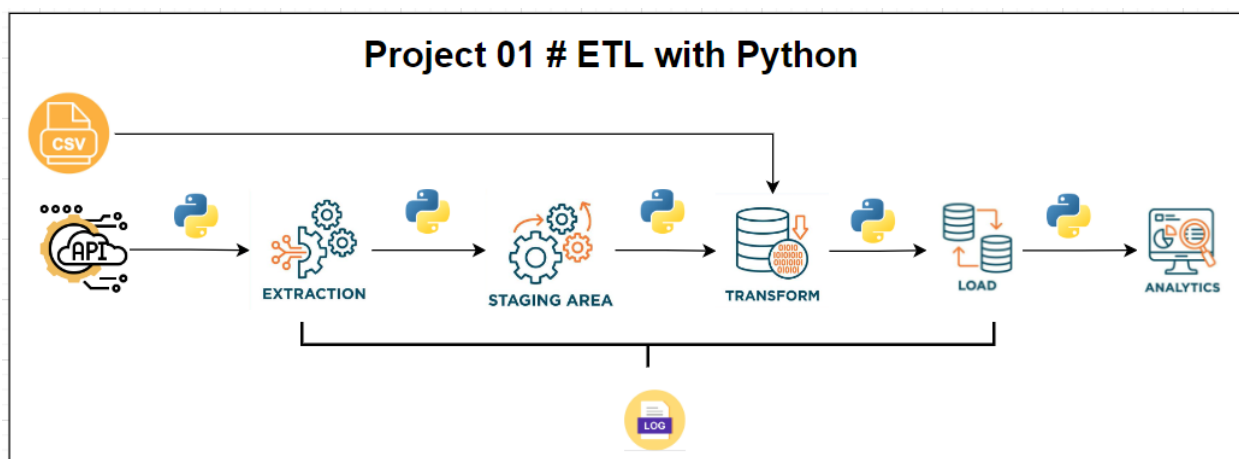
Data Engineering is the field where individuals do Extract, Transform & load (ETL) if I say the responsibilities of Data in a very summarized way. This project is regarding the acquiring and processing information on world's largest banks with having Python as skill in hand we majorly used Python and its libraries like Pandas & matplotlib for transformation & visualization, Beautiful soap for data scraping & SQLite for creating database.

Workflow:

The data should be scraped from Wikipedia using beautiful soap python library. After scraping the data, it should be saved into some staging area like Data Frame from where data is been taken for processing where we have to leverage the NumPy and Pandas capabilities to process data while using some additional csv files. After preprocessing the data should be loaded into SQLite database from where we'll query the data & use matplotlib to create some visualizations.

Architecture Diagram:

Architecture diagram of the whole project is given below



Tech Stack:

1. Python
2. Beautiful Soap
3. Pandas
4. SQLite
5. Matplotlib

Dataset:

URL for data scraping is given below:

https://web.archive.org/web/20230908091635/https://en.wikipedia.org/wiki/List_of_largest_banks

CSV data (**exchange_rate.csv**)

Currency	Rate
EUR	0.93
GBP	0.8
INR	82.95

Steps of Working:

- Task 1: Logging function
- Task 2: Extraction of data
- Task 3: Transformation of data
- Task 4: Loading to CSV
- Task 5: Loading to Database
- Task 6: Function to Run queries on Database
- Task 7: Verify log entries

Submission Guideline:

1. Deadline for the project completion is 9th July 2024
2. Project Should be uploaded on GitHub.
3. Following Google form should be filled.
Link: <https://forms.gle/xegLNv7NczQrQpFV8>