Final project overview

Estimated effort: 5 mins

In this project, you will put all the skills acquired throughout the course and your knowledge of basic Python to test. You will work on real-world data and perform the operations of Extraction, Transformation, and Loading as required. Throughout the project, you will note some outputs you need to answer questions on the graded quiz. You will also take snapshots, which you will upload in the peer-graded assignment.

Project Scenario

A multi-national firm has hired you as a data engineer. Your job is to access and process data as per requirements.

Your boss asked you to compile the list of the top 10 largest banks in the world ranked by market capitalization in billion USD. Further, you need to transform the data and store it in USD, GBP, EUR, and INR per the exchange rate information made available to you as a CSV file. You should save the processed information table locally in a CSV format and as a database table. Managers from different countries will query the database table to extract the list and note the market capitalization value in their own currency.

Directions

1. Write a function to extract the tabular information from the given URL under the heading *By Market Capitalization*, and save it to a data frame.
2. Write a function to transform the data frame by adding columns for Market Capitalization in *GBP*, *EUR*, and *INR*, rounded to 2 decimal places, based on the exchange rate information shared as a CSV file.
3. Write a function to load the transformed data frame to an output CSV file.
4. Write a function to load the transformed data frame to an SQL database server as a table.
5. Write a function to run queries on the database table.
6. Run the following queries on the database table:
   a. Extract the information for the London office, that is *Name* and *MC_GBP_Billion*
   b. Extract the information for the Berlin office, that is *Name* and *MC_EUR_Billion*
   c. Extract the information for New Delhi office, that is *Name* and *MC_INR_Billion*
7. Write a function to log the progress of the code.
8. While executing the data initialization commands and function calls, maintain appropriate log entries.

Author(s)

Abhishek Gagneja