

Homework

Problem Statement

Imagine a scenario where you are given the task to create a rerunnable **ETL** so that API data is consumable by business analysts. Fortunately, your co-workers have already done the Extract step and provided you with a zip file containing retail order data in raw Json format. Your project manager has put you on the task to support business analysts so that they can query that data using SQL from a PSQL database.

While you're at it, they would also want you to create a user table that would contain summary metrics that you think business analysts would find useful.

Note: Keep in mind that the newly created tables have to be sanely structured, and those steps should be reproducible with the exception that the ETL runs daily.

Provided information

The raw data resides at <https://s3.amazonaws.com/data-eng-homework/v1/data.zip>

Database Credentials

Type: Postgres

Version: 9.6.11

Hostname: datacandidatehomework.czwbfb7cwda.us-east-1.rds.amazonaws.com

Port: 5432

Username: jayg

Password: iheartmilkyj3lly

Database: data_balm_cleanser

Deliverables

- Provide us with a URL of your favorite version control system (for example: GitHub) containing your package
 - Please be conscientious of your commit messages as this will help us to understand your thought process
- Have at least one table for the orders data and one table for user data saved in the Postgres database provided above.
 - Do not restrain yourself to those two tables, you can get creative here!
- While there are endless ways to solve this problem, let's be pragmatic with the scope of the deliverable. We don't expect you to take more than 2 hours to work on this project.