Alexander Paredes

aparedes0212

Crack Those Macros

What you planned to do

- Web scrape at least 10 stores
    - To obtain nutrition facts and pricing information
    - Includes restaurants and grocery stores
- Create a database to store results from web scraping
    - Create database tables, views, and indexes.
- Provide instructions and methods for teammates to access database
    - Creates separate read-only and admin database users
    - Ensure teammates can test functionality

What you did not do

- Did not get pricing information for restaurant items.
- Did not create indexes for database tables.

What problems you encountered

- Encountered Cloudflare CAPTCHA when web scraping.
- Nutrition facts came in a variety of formats.

Issues completed

- Web Scrape 1
- Database Creation
- Database Access Seteup

Files you worked on

- backend/webscraping/connectTest.py
- backend/webscraping/Instacart_CSV_to_DB.py
- backend/webscraping/Instacart_Get_Food_Urls.py
- backend/webscraping/Instacart_Store_Availability_to DB.py
- backend/webscraping/Instacart_Urls_To_CSV.py
- backend/webscraping/Nutrinix_Get_Items.py

- backend/webscraping/Nutrinix_Items_To_PKL.py
- backend/webscraping/Nutrinix_PKL_To_DB.py
- backend/webscraping/schema_generate.py

Open Source Code Committed Into Repository

- pyodbc
- pandas
- re
- numpy
- selenium
- time
- os
- bs4
- concurrent.futures
- threading

What you accomplished

I implemented a web scraping solution that gathered nutrition facts from 22 locations, including both restaurants and grocery stores. Despite encountering challenges such as Cloudflare CAPTCHA and varied data formats, I resolved these issues. I slowed down the scraping speed and used random web scraping agents. Additionally, I created Python scripts specifically designed for data cleansing and standardization. These scripts addressed the inconsistencies in the nutrition facts data by converting various formats into a uniform structure. This ensured that the data was consistent and ready for further analysis and storage in the database.

The nutrition facts and grocery store pricing data were obtained and stored in an Azure database, designed with tables and views to facilitate data retrieval and analysis. I provided instructions for database access, set up separate read-only and admin users, and ensured functionality testing for collaboration.