

Project GitHub - https://github.com/patricklukowicz/507_repo

Packages Used: Beautiful Soup, requests, collections, plotly

Interaction Instructions: There are no specific instructions to interact with this program beyond running the python code in an IDE. Please see below for further detail on this portion of the project.

Data Sources

For each:

- Origin, including URLs for data and documentation
 - o MC5 MusicBrainz Page - <https://musicbrainz.org/artist/08b736bb-1c82-40b4-8b0b-49e2182a067a>
 - o Detroit MusicWiki Bands Page - <http://musicwikidetroit.org/index.php/Category:Bands>
- Format(s) (e.g., JSON, CSV, HTML):
 - o Data sources are web pages formatted as HTML
 - o Cache is formatted as a JSON file
- How you accessed the data, and whether caching was used:
 - o Using web scraping methods, I was able to identify the embedded URLs that house the data I'm looking for and cached that initial web page scrape into a JSON file. I attempted to cache the URL links in both scrape_cache files but in the second cache file I was only able to access the URLs by inserting them all into an empty list. However, I was unable to successfully crawl the Detroit MusicWiki site as I received permission denied errors indicating to me that I was not allowed to access their site data by crawling.
 - o I attempted to check to see if there was any robots.txt information on the kinds of permission I could glean from this page, but it doesn't appear that the Detroit MusicWiki site had anything like this despite the permission denials I was receiving
- Summary of data
 - o The numbers of records available are web pages that I've scraped from the MusicBrainz database for one Detroit area rock band (MC5) as well as a list of 200+ URLs of present day bands from the Detroit MusicWiki site.
- Description of records, including important fields/attributes of each record for the purpose of and what they represent:
 - o I have pulled about a dozen URLs pertaining to genre, music release and member relationship data pertaining to the MC5 and roughly 200+ URLs of bands from Detroit off the MusicWiki site.
 - o My intention was to pull the band information above to draw ties between Detroit bands base on their genre styles, labels, and members. However, I ran into multiple data sourcing issues which forced me to restart my project several times over. The Detroit MusicWiki is my fourth data source that I've attempted to access this information from.

Interaction and Presentation Design

- Please see the video I included as part of my submission for this final project. In it I explain why I was unable to complete this portion of the assignment effectively.