Documentation

For Links with no Catcha pages :

<https://www.tradinghours.com/exchanges>

<https://www.globalbrandsmagazine.com/list-of-banks-by-country>

<http://geni.org/globalenergy/library/organizations/#global>

<https://www.federalregister.gov/documents/2019/02/01/2019-00897/indian-entities-recognized-by-and-eligible-to-receive-services-from-the-united-states-bureau-of>

<https://www.relbanks.com/stock-exchanges>

<https://fiscal.treasury.gov/surety-bonds/list-admitted-reinsurers.html>

The external data pull task required to get information from various websites.

For this a python script has been used. The script visits very website using requests library and scape the content from the website using Beautiful Soup library. Both the libraries are open source python libraries.

Most of the websites have data in the tabular format. The script tries to retain the data in the same tabular format.

However, the data is not in tabular format like for few links. In that case the script is modified to take relevant data and converts it into a tabular data with fileds like name, address, state, country, zip code.

For link https://www.tradinghours.com/exchanges: some stock exchanges doesn’t have address in a proper format and some have missing information about zip code, state.

The data pulled is temporarily stored using pandas DataFrame on the memory. Finally the data is converted into a readable csv file. The script is fully automated and the user just needs to run the script file and the end result will be a csv file downloaded on the system.