Web Scraper for ME

Hi!

I have designed a web scraper to get the latest Covid-19 data regarding race and ethnicity from the Maine Health website: https://www.maine.gov/dhhs/mecdc/infectious-disease/epi/airborne/coronavirus/data.shtml

My scraper uses selenium to download the latest excel file with race and ethnicity information. Selenium navigates to the spreadsheet using the full XPath. The XPath is very sensitive to changes on the wesbite. If there is error in downloading the spreadsheet, the common solution is to re-paste the Full XPath (go to the Maine Health website on a Chrome browser) in order to fix the navigation path. This spreadsheet goes to the downloads folder on Mac (not tested on Window devices).

The scraper prints the Maine Health website title and date as data checks.

Then the scraper accesses the latest data and concatenates the old race and ethnicity data to the new data from the new downloaded csv. After this procedure, the scraper writes the new total data to a csv. The excel file "maine_covid_summary" will then contain the latest data.

Note the last step in the scraper is to delete the current csv in the downloads folder. If the current spreadsheet is not deleted the scraper will look at the old data always and not be able to access the new data.