# Web Scraper for MA

Hi!

I have designed a web scraper to get the latest Covid-19 data regarding race and ethnicity from the Massachusetts Health website: https://www.mass.gov/info-details/covid-19-response-reporting

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 website.  If there is error in downloading the spreadsheet, the common solution is to re-paste the Full XPath (go to the Massachusetts 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 Massachusetts Health website title and date as data checks.

Then the scraper accesses the latest data then writes the new data to a csv.  The excel file "MA_covid_summary" will then contain the latest data.

Please note I built the scraper on the assumption that the excel file from the Massachusetts website does not change its name format.  This is not the case so now I need to adjust the scraper download the file with a general name (without assumption).

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.