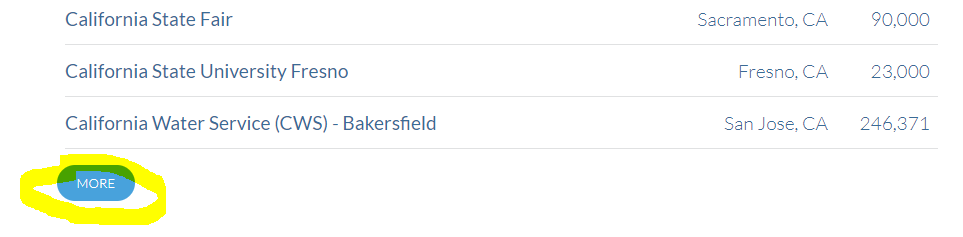
# Python Web Scraping practical exam

## Task 1: Python ‘Requests’ library and basic scraping knowledge test.

* Code the following requirement using Python Requests library and Beautiful Soup
* URL : <https://www.ewg.org/tapwater/state.php?stab=CA>
* This Url will show many water utilities in California



* Once you go down the list there is an option to view more (full list) called ‘More’, scraper should go to that page. ***(Hint : Simply check the url when you click ‘More’ button to find the URL)***

****

* For all the records in the list simply scrape Utility, Location and People served. Save the records in a CSV file.



## Task 2: Python ‘Selenium Web Driver (Chrome)’ and scraping data

* Use Selenium chrome driver and go to ‘<https://ikman.lk/en/ads>’
* Then click ‘Vehicles’ under All categories



* Type ‘lancer ex’ in search box and search.

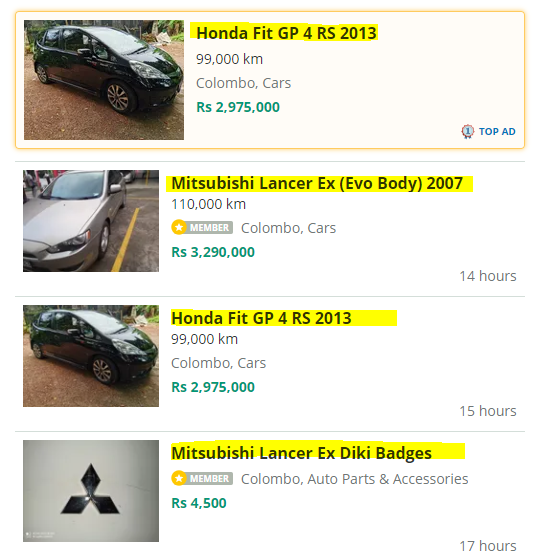


* Simply get the each advertisement title to an Array from the add list. And Print the array in console (Highlighted in Yellow)
* Then load the html source to beautifulSoup and for each record get **Vehicle title, and price** Script should go to next page to get the next page records also (Pagination). Save data to a CSV

Ex CSV:

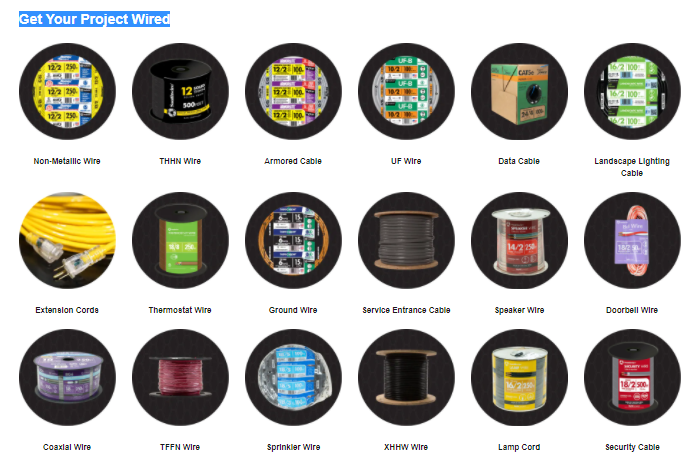
Honda Fit GP4 RS 2013,2975000

Mitsubishi Lancer Ex (Evo Body) 2007,3290000



## Task 3: Python requests library and beautifulsoup

* Write a python script to do the following, please use beautifulsoup to load the Html and parse the content
* Go to this URL : <https://www.lowes.com/b/southwire?searchTerm=southwire>
* This will list some categories under ‘Get Your Project Wired’ title, which is shown like bellow

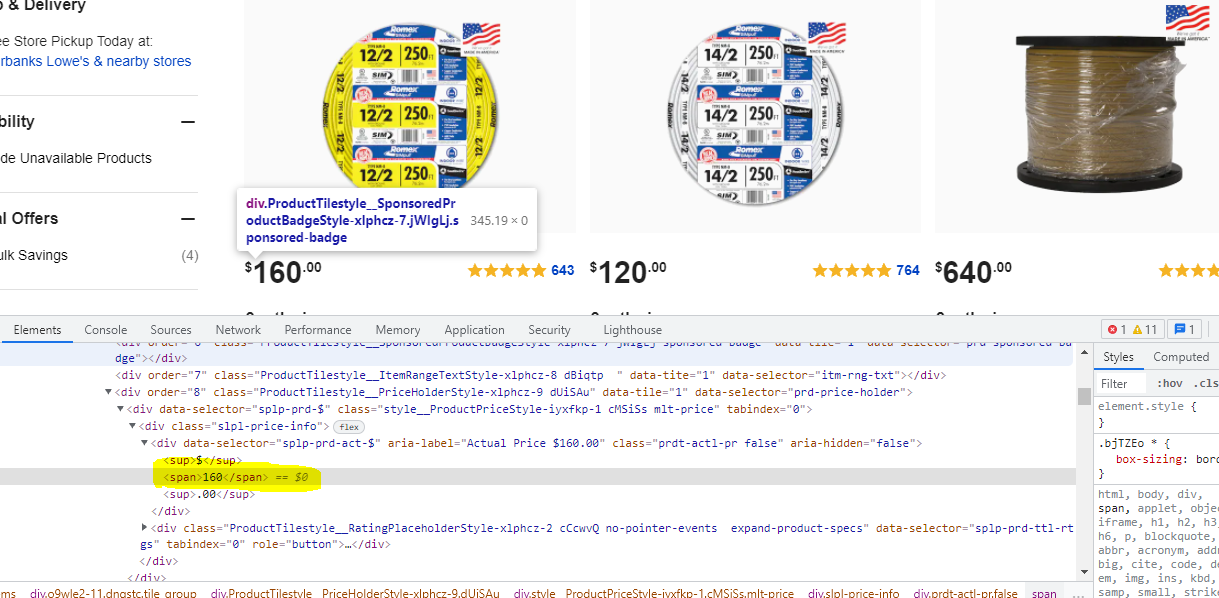


* Get the link and category name for each category (You have to use soup with css selectors)
* Save Category name and URL for each category as CSV called *categories.csv*

Ex : Non-Metallic Wire, <https://www.lowes.com/pl/Non-metallic-wire-Electrical-wire-cable-Electrical/4294546226?int_cmp=brand%3AA%3AElectrical%3AMerch%3ANon_Metallic_Wire>

THHN Wire, <https://www.lowes.com/pl/TFFN-thhn-wire-Electrical-wire-cable-Electrical/4294546249?int_cmp=brand%3AA%3AElectrical%3AMerch%3ATHHN_Wire>

* Go to each link in category, Ex for ‘Non-Metallic Wire’ go get the link and get the response using requests
* Load the response to soup and for each product listed there, Get the product name and price (see the next page), simply print product name and Price in console line by line.



## Task 4: Python POST() requests library and beautifulsoup

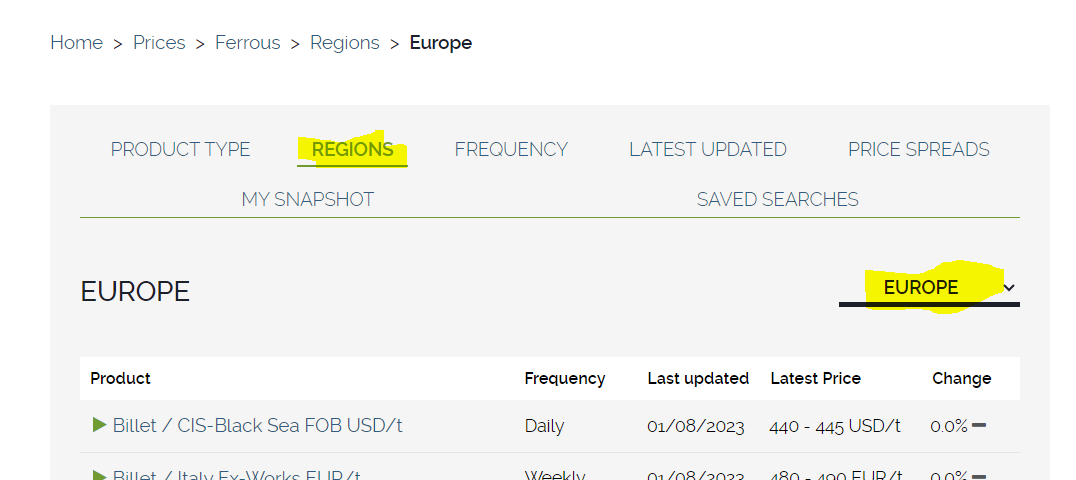
Navigate to URL : https://www.kallanish.com/en/prices/list/ferrous/

Login in the top right corner

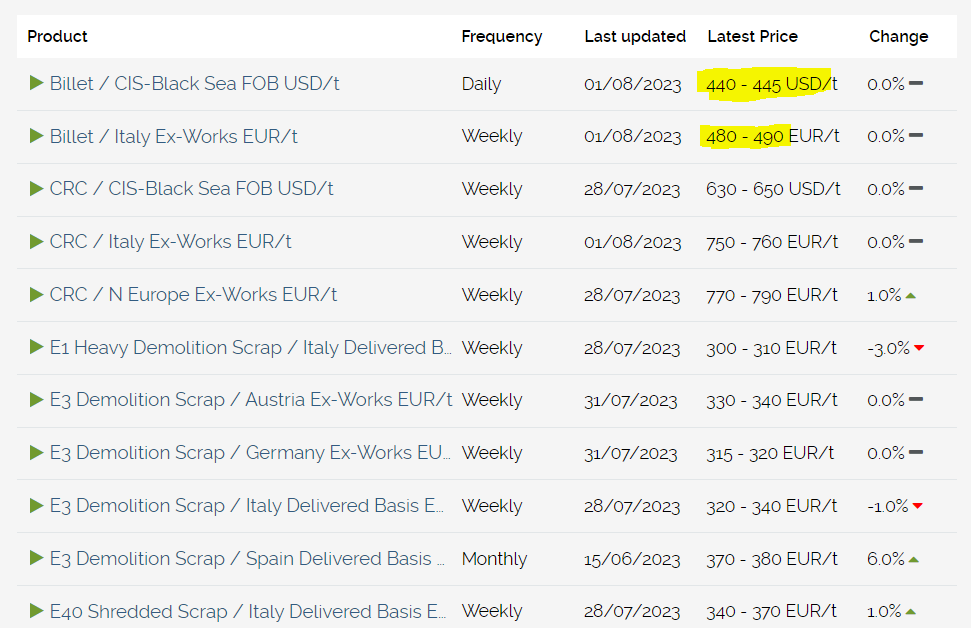
User name: [antony.fisher@mintecglobal.com](mailto:antony.fisher@mintecglobal.com)

Password: antony

Select the regions tab and select Europe from the drop down



Save all pages as html and add column for average of Latest Price column, if two numbers are given



## Task 5: Python PDF extraction using Camelot library

extract 1st page using Camelot and save as html table from **sample.pdf**