

```
In [0]: ▶ import csv
import requests
import bs4
```

```
In [0]: ▶ import pandas as pd
#df = pd.read_excel('HealthCanada.xlsx')
```

```
In [0]: ▶ #output = df.copy()
```

```
In [0]: ▶ res = requests.get('https://healthcanadians.gc.ca/recall-alert-rappel-avis/search-recherche/simple?s=&plain_text=&f_
```

```
In [0]: ▶ soup = bs4.BeautifulSoup(res.text, 'lxml')
tbody = soup('div', {"class": "margin-bottom-medium word_wrap"})
```

```
In [0]: ▶ sistertags = soup.find_all(class_='search_result_title')
title = []
link = []
for i in sistertags:
    t1 = i.text.strip()
    l1 = "https://healthcanadians.gc.ca" + i.get('href')
    title.append(t1)
    link.append(l1)
```

Work with each html here, try to do for first element first

```
In [0]: ▶ agency = []  
        date = []  
        category = []  
        issue = []  
        country = []  
        data = {'key': 'value'}
```

```
In [0]: ▶ manufacturerstring = " "  
        countrystring = " "
```

```
In [11]: #IN USE
for i in link:
    res = requests.get(i)
    soup = bs4.BeautifulSoup(res.text,'lxml')
    tbody = soup('dl',{'class':"dlSide"})
    containM = ""
    for row in tbody:
        cols = row.findChildren(recursive=False)
        cols = [ele.text.strip() for ele in cols]
        for i in range(len(cols)):
            if(cols[i]=='Manufacturer':
                containM="T"
#print(containM)
if(containM == "T"):
    print("//code for MORE appends goes here")
    for row in tbody:
        cols = row.findChildren(recursive=False)
        cols = [ele.text.strip() for ele in cols]
        for i in range(len(cols)-1):
            s = ''.join(cols[i])
            if(s == "Manufacturer"):
                s = ''.join(cols[i+1])
                c = s.split("\t")[-1]
                countrystring = countrystring + "," + c
                s = s.replace("\t","")
                s = s.replace("\n"," ")
                s = s.replace("\r","")
                manufacturerstring = manufacturerstring + "," + s
            else:
                data[cols[i]] = cols[i+1]
    country.append(countrystring)
    agency.append(manufacturerstring)
    date.append(data['Starting date:'])
    category.append(data['Subcategory:'])
    issue.append(data['Issue:'])
else:
    print("//code for LESS appends goes here")
    for row in tbody:
        cols = row.findChildren(recursive=False)
        cols = [ele.text.strip() for ele in cols]
        for i in range(len(cols)-1):
```

```
        data[cols[i]] = cols[i+1]
    date.append(data['Starting date:'])
    category.append(data['Subcategory:'])
    issue.append(data['Issue:'])
    agency.append("-")
    country.append("-")
```

```
//code for LESS appends goes here
//code for LESS appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for LESS appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for LESS appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for LESS appends goes here
//code for MORE appends goes here
//code for LESS appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
//code for MORE appends goes here
```

```
In [0]: ▶ #Take note of the order of the template
output = pd.DataFrame(
    {'Agency':agency,
     'Country':country,
     'Date':date,
     'Category':category,
     'Issue/Background':issue,
     'Title': title,
     'Url': link
    })
output.head() #make a new table like this

from pandas import ExcelWriter
import openpyxl

writer = ExcelWriter('HealthCanada.xlsx')
output.to_excel(writer,'Slide')
writer.save()
```

```
In [0]: ▶
```