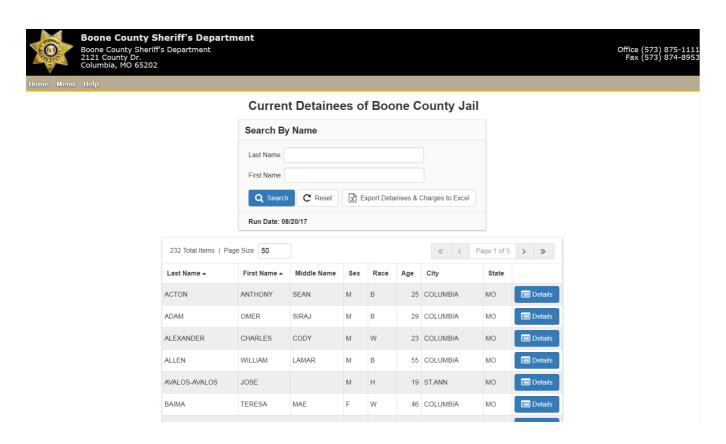
Project 4: Web scraping in Python + REGEX

In this project we use BS- BeautifulSoup and REGEX to find some whose last name starts with the letter 'A'

we'll cover the following ^
• Solution

Web scraping or web data extraction is data scraping used for extracting data from websites. In this project, we will extract tabular data from the Boone Country Sherrif's Dept website of criminal records (see the image below) and then find all the people whose name start with the word "A".



Solution

The problem solution uses BeautifulSoup. A detailed explanation of the code is out-of-scope for this course (hint: read the BS docs). In this code, we first extract HTML data and format/convert into BS's table using BS's

BeautifulSoup() function, then find and extract the table from the HTML

code.

```
import re
                                                                                        6
from pprint import pprint
import csv
import requests
from bs4 import BeautifulSoup
url = 'http://www.showmeboone.com/sheriff/JailResidents/JailResidents.asp'
response = requests.get(url)
html = response.content
soup = BeautifulSoup(html)
table = soup.find('tbody', attrs={'class': 'stripe'})
list of rows = []
for row in table.findAll('tr')[0:]:
   list_of_cells = []
    for cell in row.findAll('td'):
        text = cell.text.replace(' ', '')
        list_of_cells.append(text)
   list_of_rows.append(list_of_cells)
for line in list_of_rows:
    row = '\t'.join(str(i) for i in line) # python 2
    s=row[0:5] # Select only the Last names (1st column)
   m = re.search(r"^A",s,re.I)
    if m:
      print row
```

Expected output (Surnames started with the letter "A"):

```
ACTON
                                                   Details
       ANTHONY SEAN
                       Μ
                           В
                               25
                                   COLUMBIA
                                               MO
ADAM
       OMER
               SIRAJ
                           В
                               29 COLUMBIA
                                               MO
                                                   Details
ALEXANDER
           CHARLES CODY
                                                   MO Details
                             W
                                   23 COLUMBIA
ALLEN
       WILLIAM LAMAR
                               55 COLUMBIA
                                               MO Details
               JOSE
AVALOS-AVALOS
                                   19
                                       ST.ANN
                                               MO
                                                   Details
```

Easy!

This examples has been **adopted and extended** from the Python-BeautifulSoup's first web scraper, originally developed by Chase Davis, Jackie Kazil, Sisi Wei and Matt Wynn for bootcamps held by Investigative Reporters and Editors at the University of Missouri in Columbia, Missouri.