

# US Census URI Scraper Python Code

June 1, 2022

```
[1]: # import libraries
from bs4 import BeautifulSoup
import requests
import csv

[2]: # get the html code of the web page
htmlCode=requests.get('https://www.census.gov/programs-surveys/popest.html').
    ↪text
# assign html code to variable
soup=BeautifulSoup(htmlCode, 'html.parser')
# find all anchor elements
AllLinks=soup.find_all('a')

[3]: # extract links, convert from relative to absolute links, and add only good
    ↪links to the set
GoodLinks=set()
for i in AllLinks:
    URL=str(i.get('href'))
    if URL.startswith('/'):
        GoodLinks.add('https://www.census.gov'+URL)
    elif URL.startswith('#'):
        ''
    elif URL.startswith('None'):
        ''
    elif URL.startswith('#http'):
        GoodLinks.add(URL[1:])
    else:
        GoodLinks.add(URL)

[4]: # specify name and path to save the file to
f=open('C:/Users/todda/Desktop/Desktop/Career/Projects and Portfolio/Web
    ↪Scraper/Census_Links.csv','w', encoding='utf8')
# assign file to variable
writer=csv.writer(f, delimiter=' ', lineterminator='\r')
# create list to contain URI's
UniqueURIs=[]
```

```
[5]: # "append" adds absolute URI's to the list and "writerow" writes them to the ↵  
      ↪file
```

```
UniqueCount=0  
for q in GoodLinks:  
    if q not in UniqueURIs:  
        UniqueURIs.append(q)  
        writer.writerow(UniqueURIs)  
        UniqueCount=UniqueCount+1  
    del UniqueURIs[:]
```

```
[6]: #close file  
f.close()
```