A) To extract the links from the “Current Estimates” website, I used the BeautifulSoup package to parse the HTML and search for links with an ‘a’ tag.
Code segment:

```
In [2]: #defining URL
        url= "https://www.census.gov/programs-surveys/popest.html"
```

```
In [3]: #sending/retrieving requests for response
        r= requests.get(url)
```

```
In [4]: html= r.text
        #parsing html using BeautifulSoup
        census= BeautifulSoup(html, 'html.parser')
```

B) To determine if a link points to another HTML page, I searched the ‘a’ tags for the ‘href’ attribute. If it has that then it is a link to another page.
Code segment:

```
In [5]: find= census.find_all("a")
```

```
In [6]: len(find)
Out[6]: 252
```

C) The program ensures that relative links are saved as absolute URIs in the first for loop segment. I clean up the links as I add them to the list by adding/removing characters so that they end up in the format https://www.link.gov/page. This ensures that they are saved as absolute URIs and not relative links.

```
In [8]: for link in find:
            hrefs= str(link.get("href"))
            if hrefs.startswith('#http'):
                links.add((hrefs[1:]))
            elif hrefs.startswith('None'):
                ''
            elif hrefs.startswith('#'):
                ''
            elif hrefs.startswith('/'):
                links.add('https://www.census.gov' +hrefs)
            elif hrefs.endswith('.gov'):
                links.add(hrefs + '/')
            else:
                links.add(hrefs)
```

D) The program ensures that there are no duplicated links by creating a set and adding the links to it. Duplicated links will be skipped over, leaving a set of 118 unique links. I also filter out any unwanted duplicates in the second for loop, by only adding items not already in the list.

```
In [7]: links= set()
```

E) The full code I wrote to extract all the unique web links that point to other pages is below.

```
From bs4 import BeautifulSoup
Import csv
Import requests
#defining URL
url= "https://www.census.gov/programs-surveys/popest.html"
#retrieving requests
r= requests.get(url)
html= r.text
#parsing the html with BeautifulSoup
census= BeautifulSoup(html, 'html.parser')
find= census.find_all("a")
len(find)
links= set()
For link in find:
        hrefs= str(link.get("href"))
        If hrefs.startswith('#http):
                links.add((hrefs[1:]))
        Elif hrefs.startswith('None'):
                ''
        Elif hrefs.startswith('#'):
                ''
        Elif hrefs.startswith('/'):
                links.add('https://www.census.gov' + hrefs)
        Elif hrefs.endswith('.gov'):
                links.add(hrefs + '/')
        Else: links.add(hrefs)
len(links)
file=open('project.csv', 'w+')
write= csv.writer(file, delimiter= ' ', lineterminator='\r')
links_list= []
For x in links:
        links_list.append(x)
        If not link_list:
                write.writerow(link_list)
Else:
        del
```

F) The html code of the website at the time it was scraped is saved under census.html.

G) The csv file my script created is saved under my_export.csv.
H) Below is a screenshot of my results, also included in the folder for clarity.

References

Population and Housing Unit Estimates. (2020, March 09). Retrieved July 07, 2020, from https://www.census.gov/programs-surveys/popest.html