Census Data Analysis

November 26, 2024

1 Simple Webscraping Example with Beautiful Soup

```
[1]: # Import all neccessary libraries
     from bs4 import BeautifulSoup
     import urllib.request
     import pandas as pd
[2]: # Assign the URL to a variable
     url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/
      →page.cfm?
      ⇒Lang=E&SearchText=M1C&DGUIDlist=2021A0011M1C&GENDERlist=1,2,3&STATISTIClist=1,4&HEADERlist=
     # use the urlopen function to open the webpage
     html = urllib.request.urlopen(url)
     # show object html
     html
[2]: <a href="http.client.HTTPResponse">http.client.HTTPResponse</a> at 0x7f248c60ead0>
[3]: # Create the BeautifulSoup object
     html_to_parse = BeautifulSoup(html, "html.parser")
[4]: # create a list of tables. There is only 1 table in this webpage
     tables = html_to_parse.find_all("table")
     print(f"Number of tables found: {len(tables)}")
    Number of tables found: 1
[5]: # Create list of all the > tags in the table that has the title_
      →"2021A0011M1C - Population, 2021 - Counts - Total"
```

```
td = tables[0].find(attrs={"title":"2021A0011M1C - Population, 2021 - Counts -
                    Gradure of the state of t
  [6]: td
  [6]: <td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M1C
                geo2021A0011M1Cstat1 geo2021A0011M1Cstat1gen1" title="2021A0011M1C - Population,
                2021 - Counts - Total"> 35,642
  [7]: # convert to float
                float(td.text.replace(",", ""))
  [7]: 35642.0
              1.1 Create a script that will look up from a list of Postal codes
  [8]: import urllib.parse as urlparse
                from urllib.parse import urlencode
  [9]: # A list of postal code from the previous part
                postal = ['M4N','M4P','M4R','M4S','M4T','M4V','M5N','M5P','M5R']
[10]: # Creating Empty DataFrame and Storing it in variable df
                df = pd.DataFrame(columns = ['postal_code', 'data', 'value'])
[11]: # Loop through each postal code
                for i in postal:
                          url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

                          params = {
                                     'SearchText': i,
                                      'DGUIDlist': '2021A0011'+i
                          }
                           # this part switches up the postal code parameter in the url
                          url_parts = list(urlparse.urlparse(url))
                          query = dict(urlparse.parse_qsl(url_parts[4]))
                          query.update(params)
                          url_parts[4] = urlencode(query)
                          query = urlparse.urlunparse(url_parts)
                           # the following code is similar to the above
```

```
html = urllib.request.urlopen(query)
    html_to_parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    # change the title to find the data you want
    try:
       title = (f"2021A0011{i} - Population, 2021 - Counts - Total")
       td = tables[0].find(attrs={"title":title})
       print(td)
       df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
       print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Population,
2021 - Counts - Total"> 16,058
Number of tables found: 1
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Population,
2021 - Counts - Total"> 25,057
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Population,
2021 - Counts - Total"> 11,909
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Population,
2021 - Counts - Total"> 30,754
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M4T
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Population,
2021 - Counts - Total"> 10,332
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M4V
geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Population,
2021 - Counts - Total"> 19,273
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M5N
geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Population,
2021 - Counts - Total"> 16,154
Number of tables found: 1
geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Population,
```

```
Number of tables found: 1
     geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Population,
     2021 - Counts - Total"> 26,197
[12]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
         # the following code is similar to the above
         html = urllib.request.urlopen(query)
         html_to_parse = BeautifulSoup(html, "html.parser")
         tables = html_to_parse.find_all("table")
         print(f"Number of tables found: {len(tables)}")
         # change the title to find the data you want
         try:
             title = (f"2021A0011{i} - 15 to 19 years - Counts - Total")
             td = tables[0].find(attrs={"title":title})
             print(td)
             df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
         except IndexError:
             print(f"No matching tables found for title: {title}")
         continue
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4N
     geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - 15 to 19
     years - Counts - Total"> 1,125
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4P
     geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - 15 to 19
     years - Counts - Total">
                               855
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4R
```

2021 - Counts - Total"> 19,791

```
years - Counts - Total">
                             795
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4S
    geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - 15 to 19
    years - Counts - Total"> 1,115
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4T
    geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - 15 to 19
    years - Counts - Total">
                             455
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M4V
    geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - 15 to 19
    years - Counts - Total">
                             695
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r14 geo2021A0011M5N
    geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - 15 to 19
    years - Counts - Total"> 1,140
    Number of tables found: 1
    geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - 15 to 19
    years - Counts - Total"> 1,025
    Number of tables found: 1
    geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - 15 to 19
    years - Counts - Total">
                             745
[13]:
       for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
         # the following code is similar to the above
         html = urllib.request.urlopen(query)
         html_to_parse = BeautifulSoup(html, "html.parser")
         tables = html_to_parse.find_all("table")
```

geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - 15 to 19

```
print(f"Number of tables found: {len(tables)}")
    # change the title to find the data you want
   try:
       title = (f"2021A0011{i} - 20 to 24 years - Counts - Total")
       td = tables[0].find(attrs={"title":title})
       print(td)
       df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
       print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh2 r15 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - 20 to 24
years - Counts - Total">
                       890
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh2 r15 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - 20 to 24
years - Counts - Total"> 1,500
Number of tables found: 1
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - 20 to 24
vears - Counts - Total">
                       665
Number of tables found: 1
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - 20 to 24
years - Counts - Total"> 1,420
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh2 r15 geo2021A0011M4T
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - 20 to 24
years - Counts - Total">
                       480
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh2 r15 geo2021A0011M4V
geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - 20 to 24
years - Counts - Total">
                       855
Number of tables found: 1
geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - 20 to 24
years - Counts - Total">
                       980
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh2 r15 geo2021A0011M5P
geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - 20 to 24
years - Counts - Total"> 1,060
Number of tables found: 1
```

geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - 20 to 24

```
years - Counts - Total"> 1,760
```

```
for i in postal:
Γ14] : |
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/
       →details/page.cfm?Lang=E"
         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
         # the following code is similar to the above
         html = urllib.request.urlopen(query)
         html_to_parse = BeautifulSoup(html, "html.parser")
         tables = html_to_parse.find_all("table")
         print(f"Number of tables found: {len(tables)}")
         try:
             title = (f"2021A0011{i} - 25 to 29 years - Counts - Total")
             td = tables[0].find(attrs={"title":title})
             print(td)
             df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
         except IndexError:
             print(f"No matching tables found for title: {title}")
         continue
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M4N
     geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - 25 to 29
     years - Counts - Total">
                               715
     Number of tables found: 1
     geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - 25 to 29
     years - Counts - Total"> 3,110
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M4R
     geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - 25 to 29
     years - Counts - Total">
                               825
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M4S
     geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - 25 to 29
```

```
Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M4T
    geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - 25 to 29
    years - Counts - Total">
                             610
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M4V
    geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - 25 to 29
    years - Counts - Total"> 1,580
    Number of tables found: 1
    geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - 25 to 29
    years - Counts - Total">
                             870
    Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh2 r16 geo2021A0011M5P
    geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - 25 to 29
    years - Counts - Total"> 1,585
    Number of tables found: 1
    geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - 25 to 29
    years - Counts - Total"> 2,940
[15]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

¬details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
         # the following code is similar to the above
         html = urllib.request.urlopen(query)
         html_to_parse = BeautifulSoup(html, "html.parser")
         tables = html_to_parse.find_all("table")
         print(f"Number of tables found: {len(tables)}")
         try:
            title = (f"2021A0011{i} - Public transit - Counts - Total")
            td = tables[0].find(attrs={"title":title})
```

years - Counts - Total"> 3,165

```
print(td)
            df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
         except IndexError:
            print(f"No matching tables found for title: {title}")
         continue
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4N
    geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Public
                               705
    transit - Counts - Total">
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4P
    geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Public
    transit - Counts - Total"> 2,560
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4R
    geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Public
    transit - Counts - Total">
                               825
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4S
    geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Public
    transit - Counts - Total"> 2,860
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4T
    geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Public
    transit - Counts - Total">
                               560
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M4V
    geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Public
    transit - Counts - Total"> 1,190
    Number of tables found: 1
    geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Public
    transit - Counts - Total"> 1,025
    Number of tables found: 1
    geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Public
    transit - Counts - Total"> 1,280
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2607 geo2021A0011M5R
    geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Public
    transit - Counts - Total"> 1,540
[16]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
```

```
'SearchText': i,
        'DGUIDlist': '2021A0011'+i
    }
    # this part switches up the postal code parameter in the url
    url_parts = list(urlparse.urlparse(url))
    query = dict(urlparse.parse_qsl(url_parts[4]))
    query.update(params)
    url_parts[4] = urlencode(query)
    query = urlparse.urlunparse(url parts)
    # the following code is similar to the above
    html = urllib.request.urlopen(query)
    html_to_parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    try:
        title = (f"2021A0011{i} - Walked - Counts - Total")
        td = tables[0].find(attrs={"title":title})
        print(td)
        df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
        print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Walked -
Counts - Total">
                  295
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Walked -
Counts - Total">
                  825
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Walked -
Counts - Total">
                   265
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Walked -
Counts - Total">
                  820
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4T
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Walked -
```

Counts - Total">

250

```
Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M4V
     geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Walked -
     Counts - Total">
                       790
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M5N
     geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Walked -
     Counts - Total">
                       300
     Number of tables found: 1
     geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Walked -
     Counts - Total">
                       455
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2608 geo2021A0011M5R
     geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Walked -
     Counts - Total"> 1,440
[17]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
         # the following code is similar to the above
         html = urllib.request.urlopen(query)
         html to parse = BeautifulSoup(html, "html.parser")
         tables = html_to_parse.find_all("table")
         print(f"Number of tables found: {len(tables)}")
         try:
             title = (f"2021A0011{i} - Bicycle - Counts - Total")
             td = tables[0].find(attrs={"title":title})
             print(td)
             df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
         except IndexError:
             print(f"No matching tables found for title: {title}")
         continue
```

```
<td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M4N
    geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Bicycle -
    Counts - Total">
                       40
    Number of tables found: 1
    geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Bicycle -
    Counts - Total">
                      175
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M4R
    geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Bicycle -
    Counts - Total">
                       55
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M4S
    geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Bicycle -
    Counts - Total">
                      160
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M4T
    geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Bicycle -
    Counts - Total">
                       60
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M4V
    geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Bicycle -
    Counts - Total">
                      140
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M5N
    geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Bicycle -
    Counts - Total">
                       70
    Number of tables found: 1
    geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Bicycle -
    Counts - Total">
                      145
    Number of tables found: 1
    <td class="text-right text-nowrap" headers="rh50 r2609 geo2021A0011M5R
    geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Bicycle -
    Counts - Total">
                      455
[18]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
```

Number of tables found: 1

```
query = dict(urlparse.parse_qsl(url_parts[4]))
    query.update(params)
    url_parts[4] = urlencode(query)
    query = urlparse.urlunparse(url_parts)
    # the following code is similar to the above
    html = urllib.request.urlopen(query)
    html to parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    try:
        title = (f"2021A0011{i} - Total - Main mode of commuting for the
  ⇔employed labour force aged 15 years and over with a usual place of work or ⊔
  →no fixed workplace address - 25% sample data - Counts - Total")
        td = tables[0].find(attrs={"title":title})
        print(td)
        df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
        print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Total -
Main mode of commuting for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 3,065
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Total -
Main mode of commuting for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 6,735
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Total -
Main mode of commuting for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 2.645
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Total -
Main mode of commuting for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 7,385
```

```
<td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4T
     geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Total -
     Main mode of commuting for the employed labour force aged 15 years and over with
     a usual place of work or no fixed workplace address - 25% sample data - Counts -
     Total"> 1,920
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M4V
     geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Total -
     Main mode of commuting for the employed labour force aged 15 years and over with
     a usual place of work or no fixed workplace address - 25% sample data - Counts -
     Total"> 4,575
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M5N
     geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Total -
     Main mode of commuting for the employed labour force aged 15 years and over with
     a usual place of work or no fixed workplace address - 25% sample data - Counts -
     Total"> 4,050
     Number of tables found: 1
     geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Total -
     Main mode of commuting for the employed labour force aged 15 years and over with
     a usual place of work or no fixed workplace address - 25% sample data - Counts -
     Total"> 4,865
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh50 r2603 geo2021A0011M5R
     geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Total -
     Main mode of commuting for the employed labour force aged 15 years and over with
     a usual place of work or no fixed workplace address - 25% sample data - Counts -
     Total"> 6,095
[19]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/
       →details/page.cfm?Lang=E"
         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
```

Number of tables found: 1

```
# the following code is similar to the above
    html = urllib.request.urlopen(query)
    html_to_parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    try:
        title = (f"2021A0011{i} - Worked at home - Counts - Total")
        td = tables[0].find(attrs={"title":title})
        print(td)
        df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
        print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Worked at
home - Counts - Total"> 3,695
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Worked at
home - Counts - Total"> 7,335
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Worked at
home - Counts - Total"> 3,505
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Worked at
home - Counts - Total"> 9,655
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4T
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Worked at
home - Counts - Total"> 2,985
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M4V
geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Worked at
home - Counts - Total"> 5,670
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh48 r2594 geo2021A0011M5N
geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Worked at
home - Counts - Total"> 3.610
Number of tables found: 1
geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Worked at
home - Counts - Total"> 5,380
```

```
Number of tables found: 1  7,630
```

```
[20]: for i in postal:
          url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

          params = {
              'SearchText': i,
              'DGUIDlist': '2021A0011'+i
          }
          # this part switches up the postal code parameter in the url
          url_parts = list(urlparse.urlparse(url))
          query = dict(urlparse.parse_qsl(url_parts[4]))
          query.update(params)
          url parts[4] = urlencode(query)
          query = urlparse.urlunparse(url_parts)
          # the following code is similar to the above
          html = urllib.request.urlopen(query)
          html_to_parse = BeautifulSoup(html, "html.parser")
          tables = html_to_parse.find_all("table")
          print(f"Number of tables found: {len(tables)}")
          try:
              title = (f"2021A0011{i} - Total - Place of work status for the employed ∪
       ⇔labour force aged 15 years and over - 25% sample data - Counts - Total")
              td = tables[0].find(attrs={"title":title})
              print(td)
              df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
          except IndexError:
              print(f"No matching tables found for title: {title}")
          continue
```

```
Number of tables found: 1
 6,820

Number of tables found: 1
 14,115

Number of tables found: 1
```

```
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 6,190
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M4S
     geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 17,145
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M4T
     geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 4,935
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M4V
     geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 10,320
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M5N
     geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 7,710
     Number of tables found: 1
     geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 10,275
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M5R
     geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Total -
     Place of work status for the employed labour force aged 15 years and over - 25%
     sample data - Counts - Total"> 13,835
[21]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
```

<td class="text-right text-nowrap" headers="rh48 r2593 geo2021A0011M4R

query.update(params)

```
url_parts[4] = urlencode(query)
    query = urlparse.urlunparse(url_parts)
    # the following code is similar to the above
    html = urllib.request.urlopen(query)
    html_to_parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    try:
       title = (f"2021A0011{i} - Between 7 a.m. and 7:59 a.m. - Counts -<math>\Box

¬Total")
       td = tables[0].find(attrs={"title":title})
        df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
       print(f"No matching tables found for title: {title}")
    continue
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2620 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Between 7
a.m. and 7:59 a.m. - Counts - Total">
                                     695
Number of tables found: 1
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Between 7
a.m. and 7:59 a.m. - Counts - Total"> 1,610</td>
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2620 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Between 7
a.m. and 7:59 a.m. - Counts - Total">
                                     655
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2620 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Between 7
a.m. and 7:59 a.m. - Counts - Total"> 1,755
Number of tables found: 1
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Between 7
a.m. and 7:59 a.m. - Counts - Total">
                                     425
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2620 geo2021A0011M4V
geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Between 7
a.m. and 7:59 a.m. - Counts - Total"> 1,095 
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2620 geo2021A0011M5N
geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Between 7
```

```
Number of tables found: 1
    geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Between 7
    a.m. and 7:59 a.m. - Counts - Total"> 1,050
    Number of tables found: 1
    geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Between 7
    a.m. and 7:59 a.m. - Counts - Total"> 1,145 
[22]: for i in postal:
        url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/
      →details/page.cfm?Lang=E"
        params = {
            'SearchText': i,
            'DGUIDlist': '2021A0011'+i
        }
         # this part switches up the postal code parameter in the url
        url_parts = list(urlparse.urlparse(url))
        query = dict(urlparse.parse_qsl(url_parts[4]))
        query.update(params)
        url_parts[4] = urlencode(query)
        query = urlparse.urlunparse(url_parts)
        # the following code is similar to the above
        html = urllib.request.urlopen(query)
        html_to_parse = BeautifulSoup(html, "html.parser")
        tables = html to parse.find all("table")
        print(f"Number of tables found: {len(tables)}")
        try:
            title = (f"2021A0011{i} - Between 8 a.m. and 8:59 a.m. - Counts -11
      Gradul
            td = tables[0].find(attrs={"title":title})
            print(td)
            df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
         except IndexError:
            print(f"No matching tables found for title: {title}")
         continue
    Number of tables found: 1
```

980

a.m. and 7:59 a.m. - Counts - Total">

 $\tt geo2021A0011M4Nstat1\ geo2021A0011M4Nstat1gen1"\ title="2021A0011M4N\ -\ Between\ 8"$

<td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M4N

a.m. and 8:59 a.m. - Counts - Total"> 1,115

Number of tables found: 1

```
a.m. and 8:59 a.m. - Counts - Total"> 1,925 
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M4R
     geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Between 8
     a.m. and 8:59 a.m. - Counts - Total">
                                           785
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M4S
     geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Between 8
     a.m. and 8:59 a.m. - Counts - Total"> 2,260
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M4T
     geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Between 8
     a.m. and 8:59 a.m. - Counts - Total">
                                           600
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M4V
     geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Between 8
     a.m. and 8:59 a.m. - Counts - Total"> 1,380
     Number of tables found: 1
     geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Between 8
     a.m. and 8:59 a.m. - Counts - Total"> 1,165
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M5P
     geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Between 8
     a.m. and 8:59 a.m. - Counts - Total"> 1,365 
     Number of tables found: 1
     <td class="text-right text-nowrap" headers="rh52 r2621 geo2021A0011M5R
     geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Between 8
     a.m. and 8:59 a.m. - Counts - Total"> 1,980 
[23]: for i in postal:
         url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/

¬details/page.cfm?Lang=E"

         params = {
             'SearchText': i,
             'DGUIDlist': '2021A0011'+i
         }
         # this part switches up the postal code parameter in the url
         url_parts = list(urlparse.urlparse(url))
         query = dict(urlparse.parse_qsl(url_parts[4]))
         query.update(params)
         url_parts[4] = urlencode(query)
         query = urlparse.urlunparse(url_parts)
```

geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Between 8

```
# the following code is similar to the above
    html = urllib.request.urlopen(query)
    html_to_parse = BeautifulSoup(html, "html.parser")
    tables = html_to_parse.find_all("table")
    print(f"Number of tables found: {len(tables)}")
    try:
        title = (f"2021A0011{i} - Total - Time leaving for work for the
  employed labour force aged 15 years and over with a usual place of work or,
  ono fixed workplace address - 25% sample data - Counts - Total")
        td = tables[0].find(attrs={"title":title})
        print(td)
        df.loc[len(df.index)] = [i, title, float(td.text.replace(",", ""))]
    except IndexError:
        print(f"No matching tables found for title: {title}")
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2617 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Total -
Time leaving for work for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 3,065
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2617 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Total -
Time leaving for work for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 6,735
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2617 geo2021A0011M4R
geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Total -
Time leaving for work for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 2,645
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2617 geo2021A0011M4S
geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Total -
Time leaving for work for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 7,385
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh52 r2617 geo2021A0011M4T
geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Total -
Time leaving for work for the employed labour force aged 15 years and over with
a usual place of work or no fixed workplace address - 25% sample data - Counts -
Total"> 1,920
```

Number of tables found: 1

4,575

Number of tables found: 1

4,050

Number of tables found: 1

4,865

Number of tables found: 1

6,095

[24]: # Now you can export this to a CSV file for further analysis or visulization df.to_csv("Central_Toronto_Census_Data.csv", index=False)