

education degree, mobility status, after tax income, scarp

November 22, 2024

1 Simple Webscraping Example with BeautifulSoup

```
[1]: # Import all neccessary libraries
```

```
from bs4 import BeautifulSoup
import urllib.request
import pandas as pd
```

```
[2]: url = "https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/
↳page.cfm?
↳Lang=E&SearchText=M1E&DGUIDlist=2021A0011M1E&GENDERlist=1,2,3&STATISTIClist=1,4&HEADERlist=
```

```
[3]: # 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
```

```
[3]: <http.client.HTTPResponse at 0x7f9f7184f9a0>
```

```
[4]: # Create the BeautifulSoup object
```

```
html_to_parse = BeautifulSoup(html, "html.parser")
```

```
[5]: # 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

```
[6]: # Create list of all the <th> tags in the table that has the title
      ↪ "2021A0011M1C - Population, 2021 - Counts - Total"

      td = tables[0].find(attrs={"title": "2021A0011M1C - Population, 2021 - Counts - Total"})
```

```
[7]: td
```

```
[7]: <td class="text-right text-nowrap" headers="rh1 r1 geo2021A0011M1C
      geo2021A0011M1Cstat1 geo2021A0011M1Cstat1gen1" title="2021A0011M1C - Population,
      2021 - Counts - Total"> 35,642</td>
```

```
[8]: # convert to float
      float(td.text.replace(",", ""))
```

```
[8]: 35642.0
```

1.1 Create a script that will look up from a list of Postal codes

```
[9]: import urllib.parse as urlparse
      from urllib.parse import urlencode
      import pandas as pd
```

```
[10]: df = pd.read_csv('Toronto_Neighborhoods_with_Coordinates.csv')
      df.head()
```

```
[10]:
```

| | Postalcode | Borough | Neighborhood \ |
|---|------------|-------------|--|
| 0 | M1B | Scarborough | Malvern / Rouge |
| 1 | M1C | Scarborough | Rouge Hill / Port Union / Highland Creek |
| 2 | M1E | Scarborough | Guildwood / Morningside / West Hill |
| 3 | M1G | Scarborough | Woburn |
| 4 | M1H | Scarborough | Cedarbrae |

| | Latitude | Longitude |
|---|-----------|------------|
| 0 | 43.806686 | -79.194353 |
| 1 | 43.784535 | -79.160497 |
| 2 | 43.763573 | -79.188711 |
| 3 | 43.770992 | -79.216917 |
| 4 | 43.773136 | -79.239476 |

```
[11]: # A list of postal code from the previous part

      postal = df['Postalcode']
      postal
```

```
[11]: 0      M1B
      1      M1C
      2      M1E
      3      M1G
      4      M1H
      ...
      98     M9N
      99     M9P
      100    M9R
      101    M9V
      102    M9W
      Name: Postalcode, Length: 103, dtype: object
```

```
[12]: # Creating Empty DataFrame and Storing it in variable df

df = pd.DataFrame(columns = ['postal_code', 'data', 'value'])
df
```

```
[12]: Empty DataFrame
      Columns: [postal_code, data, value]
      Index: []
```

```
[13]: df1 = pd.DataFrame(columns = ['postal_code', 'data', 'value'])
      df1
```

```
[13]: Empty DataFrame
      Columns: [postal_code, data, value]
      Index: []
```

```
[14]: df2 = pd.DataFrame(columns = ['postal_code', 'data', 'value'])
      df2
```

```
[14]: Empty DataFrame
      Columns: [postal_code, data, value]
      Index: []
```

```
[15]: # Loop through each postal code
      # median after tax income
      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_qs(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)}")
if len(tables) == 0:
    print(f"No tables starts with this postal code: {query}. Skipping...")#
↪no table under this postal code
else:
    title = (f"2021A0011{i} - Median after-tax income in 2020 among
↪recipients ($) - Counts - Total")
    td = tables[0].find(attrs={"title": title})
    print(td)
    if td and td.text.strip():
        print(f"Processing value('median after tax income'): {td.text}")
        try:
            value = float(td.text.replace(",", "").strip()) #convert string to
↪the float
        except ValueError:
            # cannot convert, assign it to zero
            print(f"Non-numeric value encountered: {td.text}. Assigning default
↪value 0.")
            value = 0.0
        else:
            # cannot find, assign it to zero
            print(f"Skipping: No valid data found for title '{title}'. Assigning
↪default value 0.")
            value = 0.0
    df.loc[len(df.index)] = [i, title, value]

```

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh7 r115 geo2021A0011M1B
geo2021A0011M1Bstat1 geo2021A0011M1Bstat1gen1" title="2021A0011M1B - Median
after-tax income in 2020 among recipients ($) - Counts - Total"> 31,600</td>
Processing value('median after tax income'): 31,600

```

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh7 r115 geo2021A0011M1C
geo2021A0011M1Cstat1 geo2021A0011M1Cstat1gen1" title="2021A0011M1C - Median
after-tax income in 2020 among recipients ($) - Counts - Total"> 39,600</td>
Processing value('median after tax income'): 39,600

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
Cell In[15], line 19
    16 query = urlparse.urlunparse(url_parts)
    18 # the following code is similar to the above
--> 19 html = urllib.request.urlopen(query)
    20 html_to_parse = BeautifulSoup(html, "html.parser")
    21 tables = html_to_parse.find_all("table")

File /opt/conda/lib/python3.11/urllib/request.py:216, in urlopen(url, data, timeout, cafile, capath, cadefault, context)
    214 else:
    215     opener = _opener
--> 216 return opener.open(url, data, timeout)

File /opt/conda/lib/python3.11/urllib/request.py:519, in OpenerDirector.open(self, fullurl, data, timeout)
    516     req = meth(req)
    518 sys.audit('urllib.Request', req.full_url, req.data, req.headers, req.get_method())
--> 519 response = self._open(req, data)
    521 # post-process response
    522 meth_name = protocol+"_response"

File /opt/conda/lib/python3.11/urllib/request.py:536, in OpenerDirector._open(self, req, data)
    533     return result
    535 protocol = req.type
--> 536 result = self._call_chain(self.handle_open, protocol, protocol +
    537                             '_open', req)
    538 if result:
    539     return result

File /opt/conda/lib/python3.11/urllib/request.py:496, in OpenerDirector._call_chain(self, chain, kind, meth_name, *args)
    494 for handler in handlers:
    495     func = getattr(handler, meth_name)
--> 496     result = func(*args)
    497     if result is not None:
    498         return result

File /opt/conda/lib/python3.11/urllib/request.py:1391, in HTTPSHandler._https_open(self, req)
    1390 def https_open(self, req):
-> 1391     return self.do_open(http.client.HTTPSConnection, req,
    1392                         context=self._context, check_hostname=self._check_hostname)

```

```

File /opt/conda/lib/python3.11/urllib/request.py:1352, in AbstractHTTPHandler.
-> do_open(self, http_class, req, **http_conn_args)
    1350     except OSError as err: # timeout error
    1351         raise URLError(err)
-> 1352     r = h.getresponse()
    1353 except:
    1354     h.close()

```

```

File /opt/conda/lib/python3.11/http/client.py:1378, in HTTPConnection.
-> getresponse(self)
    1376 try:
    1377     try:
-&> 1378         response.begin()
    1379     except ConnectionError:
    1380         self.close()

```

```

File /opt/conda/lib/python3.11/http/client.py:318, in HTTPResponse.begin(self)
    316 # read until we get a non-100 response
    317 while True:
-> 318     version, status, reason = self._read_status()
    319     if status != CONTINUE:
    320         break

```

```

File /opt/conda/lib/python3.11/http/client.py:279, in HTTPResponse.
-> _read_status(self)
    278 def _read_status(self):
-> 279     line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
    280     if len(line) > _MAXLINE:
    281         raise LineTooLong("status line")

```

```

File /opt/conda/lib/python3.11/socket.py:706, in SocketIO.readinto(self, b)
    704 while True:
    705     try:
-> 706         return self._sock.recv_into(b)
    707     except timeout:
    708         self._timeout_occurred = True

```

```

File /opt/conda/lib/python3.11/ssl.py:1311, in SSLSocket.recv_into(self, buffer,
-> nbytes, flags)
    1307     if flags != 0:
    1308         raise ValueError(
    1309             "non-zero flags not allowed in calls to recv_into() on %s" %
    1310             self.__class__)
-> 1311     return self.read(nbytes, buffer)
    1312 else:
    1313     return super().recv_into(buffer, nbytes, flags)

```

```

File /opt/conda/lib/python3.11/ssl.py:1167, in SSLSocket.read(self, len, buffer

```

```

1165 try:
1166     if buffer is not None:
-> 1167         return self._sslobj.read(len, buffer)
1168     else:
1169         return self._sslobj.read(len)

```

KeyboardInterrupt:

```

[ ]: #Bachelor's degree or higher
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)}")
    if len(tables) == 0:
        print(f"No tables starts with this postal code: {query}. Skipping...")#
    ↪no table under this postal code
    else:
        title = (f"2021A0011{i} - Bachelor's degree or higher - Counts - Total")
        td = tables[0].find(attrs={"title": title})
        print(td)
    if td and td.text.strip():
        print(f"Processing value('Bachelor's degree or higher): {td.text}")
        try:
            value = float(td.text.replace(",", "").strip()) #convert string to
    ↪the float
        except ValueError:
            # cannot convert, assign it to zero
            print(f"Non-numeric value encountered: {td.text}. Assigning default
    ↪value 0.")

```

```

        value = 0.0
    else:
        # cannot find, assign it to zero
        print(f"Skipping: No valid data found for title '{title}'. Assigning_
↳ default value 0.")
        value = 0.0
    df1.loc[len(df1.index)] = [i, title, value]

```

```

[16]: # for non-mover, we want to find the neighbour hood with the stable vibe
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_qs(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)}")
    if len(tables) == 0:
        print(f"No tables starts with this postal code: {query}. Skipping...")#_
↳ no table under this postal code
    else:
        title = (f"2021A0011{i} - Movers - Counts - Total")
        td = tables[0].find(attrs={"title": title})
        print(td)
        if td and td.text.strip():
            print(f"Processing value('Mover'): {td.text}")
            try:
                value = float(td.text.replace(",", "").strip()) #convert string to_
↳ the float
            except ValueError:
                # cannot convert, assign it to zero
                print(f"Non-numeric value encountered: {td.text}. Assigning default_
↳ value 0.")
                value = 0.0

```



```

else:
    # cannot find, assign it to zero
    print(f"Skipping: No valid data found for title '{title}'. Assigning_
↳default value 0.")
    value = 0.0
    df2.loc[len(df2.index)] = [i, title, value]

```

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1B
geo2021A0011M1Bstat1 geo2021A0011M1Bstat1gen1" title="2021A0011M1B - Movers -
Counts - Total"> 5,625</td>

```

Processing value('Mover'): 5,625

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1C
geo2021A0011M1Cstat1 geo2021A0011M1Cstat1gen1" title="2021A0011M1C - Movers -
Counts - Total"> 2,535</td>

```

Processing value('Mover'): 2,535

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1E
geo2021A0011M1Estat1 geo2021A0011M1Estat1gen1" title="2021A0011M1E - Movers -
Counts - Total"> 4,370</td>

```

Processing value('Mover'): 4,370

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1G
geo2021A0011M1Gstat1 geo2021A0011M1Gstat1gen1" title="2021A0011M1G - Movers -
Counts - Total"> 2,930</td>

```

Processing value('Mover'): 2,930

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1H
geo2021A0011M1Hstat1 geo2021A0011M1Hstat1gen1" title="2021A0011M1H - Movers -
Counts - Total"> 2,835</td>

```

Processing value('Mover'): 2,835

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1J
geo2021A0011M1Jstat1 geo2021A0011M1Jstat1gen1" title="2021A0011M1J - Movers -
Counts - Total"> 2,885</td>

```

Processing value('Mover'): 2,885

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1K
geo2021A0011M1Kstat1 geo2021A0011M1Kstat1gen1" title="2021A0011M1K - Movers -
Counts - Total"> 4,050</td>

```

Processing value('Mover'): 4,050

Number of tables found: 1

```

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1L
geo2021A0011M1Lstat1 geo2021A0011M1Lstat1gen1" title="2021A0011M1L - Movers -
Counts - Total"> 3,385</td>

```

Processing value('Mover'): 3,385

Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1M
geo2021A0011M1Mstat1 geo2021A0011M1Mstat1gen1" title="2021A0011M1M - Movers -
Counts - Total"> 2,210</td>
Processing value('Mover'): 2,210
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1N
geo2021A0011M1Nstat1 geo2021A0011M1Nstat1gen1" title="2021A0011M1N - Movers -
Counts - Total"> 2,190</td>
Processing value('Mover'): 2,190
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1P
geo2021A0011M1Pstat1 geo2021A0011M1Pstat1gen1" title="2021A0011M1P - Movers -
Counts - Total"> 4,425</td>
Processing value('Mover'): 4,425
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1R
geo2021A0011M1Rstat1 geo2021A0011M1Rstat1gen1" title="2021A0011M1R - Movers -
Counts - Total"> 2,585</td>
Processing value('Mover'): 2,585
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1S
geo2021A0011M1Sstat1 geo2021A0011M1Sstat1gen1" title="2021A0011M1S - Movers -
Counts - Total"> 4,035</td>
Processing value('Mover'): 4,035
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1T
geo2021A0011M1Tstat1 geo2021A0011M1Tstat1gen1" title="2021A0011M1T - Movers -
Counts - Total"> 3,330</td>
Processing value('Mover'): 3,330
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1V
geo2021A0011M1Vstat1 geo2021A0011M1Vstat1gen1" title="2021A0011M1V - Movers -
Counts - Total"> 4,325</td>
Processing value('Mover'): 4,325
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1W
geo2021A0011M1Wstat1 geo2021A0011M1Wstat1gen1" title="2021A0011M1W - Movers -
Counts - Total"> 4,425</td>
Processing value('Mover'): 4,425
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M1X
geo2021A0011M1Xstat1 geo2021A0011M1Xstat1gen1" title="2021A0011M1X - Movers -
Counts - Total"> 720</td>
Processing value('Mover'): 720
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2H
geo2021A0011M2Hstat1 geo2021A0011M2Hstat1gen1" title="2021A0011M2H - Movers -

Counts - Total"> 2,350</td>
Processing value('Mover'): 2,350
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2J
geo2021A0011M2Jstat1 geo2021A0011M2Jstat1gen1" title="2021A0011M2J - Movers -
Counts - Total"> 9,095</td>
Processing value('Mover'): 9,095
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2K
geo2021A0011M2Kstat1 geo2021A0011M2Kstat1gen1" title="2021A0011M2K - Movers -
Counts - Total"> 4,755</td>
Processing value('Mover'): 4,755
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2L
geo2021A0011M2Lstat1 geo2021A0011M2Lstat1gen1" title="2021A0011M2L - Movers -
Counts - Total"> 1,045</td>
Processing value('Mover'): 1,045
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2M
geo2021A0011M2Mstat1 geo2021A0011M2Mstat1gen1" title="2021A0011M2M - Movers -
Counts - Total"> 4,665</td>
Processing value('Mover'): 4,665
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2N
geo2021A0011M2Nstat1 geo2021A0011M2Nstat1gen1" title="2021A0011M2N - Movers -
Counts - Total"> 14,470</td>
Processing value('Mover'): 14,470
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2P
geo2021A0011M2Pstat1 geo2021A0011M2Pstat1gen1" title="2021A0011M2P - Movers -
Counts - Total"> 860</td>
Processing value('Mover'): 860
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M2R
geo2021A0011M2Rstat1 geo2021A0011M2Rstat1gen1" title="2021A0011M2R - Movers -
Counts - Total"> 4,365</td>
Processing value('Mover'): 4,365
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3A
geo2021A0011M3Astat1 geo2021A0011M3Astat1gen1" title="2021A0011M3A - Movers -
Counts - Total"> 3,155</td>
Processing value('Mover'): 3,155
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3B
geo2021A0011M3Bstat1 geo2021A0011M3Bstat1gen1" title="2021A0011M3B - Movers -
Counts - Total"> 1,210</td>
Processing value('Mover'): 1,210
Number of tables found: 1

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3C
 geo2021A0011M3Cstat1 geo2021A0011M3Cstat1gen1" title="2021A0011M3C - Movers -
 Counts - Total"> 4,440</td>
 Processing value('Mover'): 4,440
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3H
 geo2021A0011M3Hstat1 geo2021A0011M3Hstat1gen1" title="2021A0011M3H - Movers -
 Counts - Total"> 4,830</td>
 Processing value('Mover'): 4,830
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3J
 geo2021A0011M3Jstat1 geo2021A0011M3Jstat1gen1" title="2021A0011M3J - Movers -
 Counts - Total"> 3,075</td>
 Processing value('Mover'): 3,075
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3K
 geo2021A0011M3Kstat1 geo2021A0011M3Kstat1gen1" title="2021A0011M3K - Movers -
 Counts - Total"> 1,140</td>
 Processing value('Mover'): 1,140
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3L
 geo2021A0011M3Lstat1 geo2021A0011M3Lstat1gen1" title="2021A0011M3L - Movers -
 Counts - Total"> 1,425</td>
 Processing value('Mover'): 1,425
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3M
 geo2021A0011M3Mstat1 geo2021A0011M3Mstat1gen1" title="2021A0011M3M - Movers -
 Counts - Total"> 1,805</td>
 Processing value('Mover'): 1,805
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M3N
 geo2021A0011M3Nstat1 geo2021A0011M3Nstat1gen1" title="2021A0011M3N - Movers -
 Counts - Total"> 2,620</td>
 Processing value('Mover'): 2,620
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4A
 geo2021A0011M4Astat1 geo2021A0011M4Astat1gen1" title="2021A0011M4A - Movers -
 Counts - Total"> 1,305</td>
 Processing value('Mover'): 1,305
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4B
 geo2021A0011M4Bstat1 geo2021A0011M4Bstat1gen1" title="2021A0011M4B - Movers -
 Counts - Total"> 1,720</td>
 Processing value('Mover'): 1,720
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4C
 geo2021A0011M4Cstat1 geo2021A0011M4Cstat1gen1" title="2021A0011M4C - Movers -
 Counts - Total"> 5,220</td>

Processing value('Mover'): 5,220
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4E
geo2021A0011M4Estat1 geo2021A0011M4Estat1gen1" title="2021A0011M4E - Movers -
Counts - Total"> 2,330</td>
Processing value('Mover'): 2,330
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4G
geo2021A0011M4Gstat1 geo2021A0011M4Gstat1gen1" title="2021A0011M4G - Movers -
Counts - Total"> 1,940</td>
Processing value('Mover'): 1,940
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4H
geo2021A0011M4Hstat1 geo2021A0011M4Hstat1gen1" title="2021A0011M4H - Movers -
Counts - Total"> 1,400</td>
Processing value('Mover'): 1,400
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4J
geo2021A0011M4Jstat1 geo2021A0011M4Jstat1gen1" title="2021A0011M4J - Movers -
Counts - Total"> 2,895</td>
Processing value('Mover'): 2,895
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4K
geo2021A0011M4Kstat1 geo2021A0011M4Kstat1gen1" title="2021A0011M4K - Movers -
Counts - Total"> 2,990</td>
Processing value('Mover'): 2,990
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4L
geo2021A0011M4Lstat1 geo2021A0011M4Lstat1gen1" title="2021A0011M4L - Movers -
Counts - Total"> 3,870</td>
Processing value('Mover'): 3,870
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4M
geo2021A0011M4Mstat1 geo2021A0011M4Mstat1gen1" title="2021A0011M4M - Movers -
Counts - Total"> 3,770</td>
Processing value('Mover'): 3,770
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4N
geo2021A0011M4Nstat1 geo2021A0011M4Nstat1gen1" title="2021A0011M4N - Movers -
Counts - Total"> 1,715</td>
Processing value('Mover'): 1,715
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4P
geo2021A0011M4Pstat1 geo2021A0011M4Pstat1gen1" title="2021A0011M4P - Movers -
Counts - Total"> 6,030</td>
Processing value('Mover'): 6,030
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4R

geo2021A0011M4Rstat1 geo2021A0011M4Rstat1gen1" title="2021A0011M4R - Movers - Counts - Total"> 1,685</td>
 Processing value('Mover'): 1,685
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4S geo2021A0011M4Sstat1 geo2021A0011M4Sstat1gen1" title="2021A0011M4S - Movers - Counts - Total"> 6,900</td>
 Processing value('Mover'): 6,900
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4T geo2021A0011M4Tstat1 geo2021A0011M4Tstat1gen1" title="2021A0011M4T - Movers - Counts - Total"> 1,230</td>
 Processing value('Mover'): 1,230
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4V geo2021A0011M4Vstat1 geo2021A0011M4Vstat1gen1" title="2021A0011M4V - Movers - Counts - Total"> 2,825</td>
 Processing value('Mover'): 2,825
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4W geo2021A0011M4Wstat1 geo2021A0011M4Wstat1gen1" title="2021A0011M4W - Movers - Counts - Total"> 3,045</td>
 Processing value('Mover'): 3,045
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4X geo2021A0011M4Xstat1 geo2021A0011M4Xstat1gen1" title="2021A0011M4X - Movers - Counts - Total"> 2,735</td>
 Processing value('Mover'): 2,735
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M4Y geo2021A0011M4Ystat1 geo2021A0011M4Ystat1gen1" title="2021A0011M4Y - Movers - Counts - Total"> 10,910</td>
 Processing value('Mover'): 10,910
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5A geo2021A0011M5Astat1 geo2021A0011M5Astat1gen1" title="2021A0011M5A - Movers - Counts - Total"> 10,250</td>
 Processing value('Mover'): 10,250
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5B geo2021A0011M5Bstat1 geo2021A0011M5Bstat1gen1" title="2021A0011M5B - Movers - Counts - Total"> 5,775</td>
 Processing value('Mover'): 5,775
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5C geo2021A0011M5Cstat1 geo2021A0011M5Cstat1gen1" title="2021A0011M5C - Movers - Counts - Total"> 730</td>
 Processing value('Mover'): 730

Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5E
geo2021A0011M5Estat1 geo2021A0011M5Estat1gen1" title="2021A0011M5E - Movers -
Counts - Total"> 3,885</td>
Processing value('Mover'): 3,885
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5G
geo2021A0011M5Gstat1 geo2021A0011M5Gstat1gen1" title="2021A0011M5G - Movers -
Counts - Total"> 3,490</td>
Processing value('Mover'): 3,490
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5H
geo2021A0011M5Hstat1 geo2021A0011M5Hstat1gen1" title="2021A0011M5H - Movers -
Counts - Total"> 1,380</td>
Processing value('Mover'): 1,380
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5J
geo2021A0011M5Jstat1 geo2021A0011M5Jstat1gen1" title="2021A0011M5J - Movers -
Counts - Total"> 5,150</td>
Processing value('Mover'): 5,150
Number of tables found: 0
No tables starts with this postal code: <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M5K&DGUIDlist=2021A0011M5K>.
Skipping...
Processing value('Mover'): 5,150
Number of tables found: 0
No tables starts with this postal code: <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M5L&DGUIDlist=2021A0011M5L>.
Skipping...
Processing value('Mover'): 5,150
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5M
geo2021A0011M5Mstat1 geo2021A0011M5Mstat1gen1" title="2021A0011M5M - Movers -
Counts - Total"> 1,985</td>
Processing value('Mover'): 1,985
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5N
geo2021A0011M5Nstat1 geo2021A0011M5Nstat1gen1" title="2021A0011M5N - Movers -
Counts - Total"> 1,650</td>
Processing value('Mover'): 1,650
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5P
geo2021A0011M5Pstat1 geo2021A0011M5Pstat1gen1" title="2021A0011M5P - Movers -
Counts - Total"> 2,425</td>
Processing value('Mover'): 2,425
Number of tables found: 1

<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5R
 geo2021A0011M5Rstat1 geo2021A0011M5Rstat1gen1" title="2021A0011M5R - Movers -
 Counts - Total"> 4,780</td>
 Processing value('Mover'): 4,780
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5S
 geo2021A0011M5Sstat1 geo2021A0011M5Sstat1gen1" title="2021A0011M5S - Movers -
 Counts - Total"> 4,880</td>
 Processing value('Mover'): 4,880
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5T
 geo2021A0011M5Tstat1 geo2021A0011M5Tstat1gen1" title="2021A0011M5T - Movers -
 Counts - Total"> 3,700</td>
 Processing value('Mover'): 3,700
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M5V
 geo2021A0011M5Vstat1 geo2021A0011M5Vstat1gen1" title="2021A0011M5V - Movers -
 Counts - Total"> 21,160</td>
 Processing value('Mover'): 21,160
 Number of tables found: 0
 No tables starts with this postal code: [https://www12.statcan.gc.ca/census-
 recensement/2021/dp-
 pd/prof/details/page.cfm?Lang=E&SearchText=M5W&DGUIDlist=2021A0011M5W](https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M5W&DGUIDlist=2021A0011M5W).
 Skipping...
 Processing value('Mover'): 21,160
 Number of tables found: 0
 No tables starts with this postal code: [https://www12.statcan.gc.ca/census-
 recensement/2021/dp-
 pd/prof/details/page.cfm?Lang=E&SearchText=M5X&DGUIDlist=2021A0011M5X](https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M5X&DGUIDlist=2021A0011M5X).
 Skipping...
 Processing value('Mover'): 21,160
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6A
 geo2021A0011M6Astat1 geo2021A0011M6Astat1gen1" title="2021A0011M6A - Movers -
 Counts - Total"> 2,450</td>
 Processing value('Mover'): 2,450
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6B
 geo2021A0011M6Bstat1 geo2021A0011M6Bstat1gen1" title="2021A0011M6B - Movers -
 Counts - Total"> 2,400</td>
 Processing value('Mover'): 2,400
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6C
 geo2021A0011M6Cstat1 geo2021A0011M6Cstat1gen1" title="2021A0011M6C - Movers -
 Counts - Total"> 2,150</td>
 Processing value('Mover'): 2,150
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6E

geo2021A0011M6Estat1 geo2021A0011M6Estat1gen1" title="2021A0011M6E - Movers - Counts - Total"> 3,905</td>
 Processing value('Mover'): 3,905
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6G geo2021A0011M6Gstat1 geo2021A0011M6Gstat1gen1" title="2021A0011M6G - Movers - Counts - Total"> 4,330</td>
 Processing value('Mover'): 4,330
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6H geo2021A0011M6Hstat1 geo2021A0011M6Hstat1gen1" title="2021A0011M6H - Movers - Counts - Total"> 5,840</td>
 Processing value('Mover'): 5,840
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6J geo2021A0011M6Jstat1 geo2021A0011M6Jstat1gen1" title="2021A0011M6J - Movers - Counts - Total"> 5,950</td>
 Processing value('Mover'): 5,950
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6K geo2021A0011M6Kstat1 geo2021A0011M6Kstat1gen1" title="2021A0011M6K - Movers - Counts - Total"> 8,955</td>
 Processing value('Mover'): 8,955
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6L geo2021A0011M6Lstat1 geo2021A0011M6Lstat1gen1" title="2021A0011M6L - Movers - Counts - Total"> 1,220</td>
 Processing value('Mover'): 1,220
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6M geo2021A0011M6Mstat1 geo2021A0011M6Mstat1gen1" title="2021A0011M6M - Movers - Counts - Total"> 3,105</td>
 Processing value('Mover'): 3,105
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6N geo2021A0011M6Nstat1 geo2021A0011M6Nstat1gen1" title="2021A0011M6N - Movers - Counts - Total"> 3,080</td>
 Processing value('Mover'): 3,080
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6P geo2021A0011M6Pstat1 geo2021A0011M6Pstat1gen1" title="2021A0011M6P - Movers - Counts - Total"> 5,835</td>
 Processing value('Mover'): 5,835
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6R geo2021A0011M6Rstat1 geo2021A0011M6Rstat1gen1" title="2021A0011M6R - Movers - Counts - Total"> 2,610</td>
 Processing value('Mover'): 2,610

Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M6S
geo2021A0011M6Sstat1 geo2021A0011M6Sstat1gen1" title="2021A0011M6S - Movers -
Counts - Total"> 3,190</td>
Processing value('Mover'): 3,190
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M7A
geo2021A0011M7Astat1 geo2021A0011M7Astat1gen1" title="2021A0011M7A - Movers -
Counts - Total"><abbr title="suppressed to meet the confidentiality requirements
of the Statistics Act">x</abbr></td>
Processing value('Mover'): x
Non-numeric value encountered: x. Assigning default value 0.
Number of tables found: 0
No tables starts with this postal code: [https://www12.statcan.gc.ca/census-
recensement/2021/dp-
pd/prof/details/page.cfm?Lang=E&SearchText=M7R&DGUIDlist=2021A0011M7R](https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M7R&DGUIDlist=2021A0011M7R).
Skipping..
Processing value('Mover'): x
Non-numeric value encountered: x. Assigning default value 0.
Number of tables found: 0
No tables starts with this postal code: [https://www12.statcan.gc.ca/census-
recensement/2021/dp-
pd/prof/details/page.cfm?Lang=E&SearchText=M7Y&DGUIDlist=2021A0011M7Y](https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=M7Y&DGUIDlist=2021A0011M7Y).
Skipping..
Processing value('Mover'): x
Non-numeric value encountered: x. Assigning default value 0.
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M8V
geo2021A0011M8Vstat1 geo2021A0011M8Vstat1gen1" title="2021A0011M8V - Movers -
Counts - Total"> 6,885</td>
Processing value('Mover'): 6,885
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M8W
geo2021A0011M8Wstat1 geo2021A0011M8Wstat1gen1" title="2021A0011M8W - Movers -
Counts - Total"> 2,290</td>
Processing value('Mover'): 2,290
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M8X
geo2021A0011M8Xstat1 geo2021A0011M8Xstat1gen1" title="2021A0011M8X - Movers -
Counts - Total"> 1,030</td>
Processing value('Mover'): 1,030
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M8Y
geo2021A0011M8Ystat1 geo2021A0011M8Ystat1gen1" title="2021A0011M8Y - Movers -
Counts - Total"> 2,350</td>
Processing value('Mover'): 2,350
Number of tables found: 1
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M8Z

geo2021A0011M8Zstat1 geo2021A0011M8Zstat1gen1" title="2021A0011M8Z - Movers - Counts - Total"> 1,865</td>
 Processing value('Mover'): 1,865
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9A geo2021A0011M9Astat1 geo2021A0011M9Astat1gen1" title="2021A0011M9A - Movers - Counts - Total"> 4,525</td>
 Processing value('Mover'): 4,525
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9B geo2021A0011M9Bstat1 geo2021A0011M9Bstat1gen1" title="2021A0011M9B - Movers - Counts - Total"> 2,880</td>
 Processing value('Mover'): 2,880
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9C geo2021A0011M9Cstat1 geo2021A0011M9Cstat1gen1" title="2021A0011M9C - Movers - Counts - Total"> 3,215</td>
 Processing value('Mover'): 3,215
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9L geo2021A0011M9Lstat1 geo2021A0011M9Lstat1gen1" title="2021A0011M9L - Movers - Counts - Total"> 1,055</td>
 Processing value('Mover'): 1,055
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9M geo2021A0011M9Mstat1 geo2021A0011M9Mstat1gen1" title="2021A0011M9M - Movers - Counts - Total"> 1,895</td>
 Processing value('Mover'): 1,895
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9N geo2021A0011M9Nstat1 geo2021A0011M9Nstat1gen1" title="2021A0011M9N - Movers - Counts - Total"> 1,870</td>
 Processing value('Mover'): 1,870
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9P geo2021A0011M9Pstat1 geo2021A0011M9Pstat1gen1" title="2021A0011M9P - Movers - Counts - Total"> 1,495</td>
 Processing value('Mover'): 1,495
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9R geo2021A0011M9Rstat1 geo2021A0011M9Rstat1gen1" title="2021A0011M9R - Movers - Counts - Total"> 2,860</td>
 Processing value('Mover'): 2,860
 Number of tables found: 1
 <td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9V geo2021A0011M9Vstat1 geo2021A0011M9Vstat1gen1" title="2021A0011M9V - Movers - Counts - Total"> 4,480</td>
 Processing value('Mover'): 4,480

Number of tables found: 1

```
<td class="text-right text-nowrap" headers="rh33 r1962 geo2021A0011M9W
geo2021A0011M9Wstat1 geo2021A0011M9Wstat1gen1" title="2021A0011M9W - Movers -
Counts - Total"> 3,165</td>
```

Processing value('Mover'): 3,165

```
[ ]: df.head()
dfSort = df.sort_values(by = 'value', ascending = False)
dfSort
unique_df = dfSort.drop_duplicates(subset=["postal_code"], keep="first")# drop
↳the duplicate postal code
unique_df
unique_df['Score'] = unique_df['value'].rank(ascending=True, method='dense').
↳astype(int)
unique_df
```

```
[ ]: df1.head()
dfSort1 = df1.sort_values(by = 'value', ascending = False)
dfSort1
unique_df1 = dfSort1.drop_duplicates(subset=["postal_code"], keep="first")#
↳drop the duplicate postal code
unique_df1
unique_df1['Score'] = unique_df1['value'].rank(ascending=True, method='dense').
↳astype(int)
unique_df1
```

```
[18]: df2.head()
dfSort2 = df2.sort_values(by = 'value', ascending = False)
dfSort2
unique_df2 = dfSort2.drop_duplicates(subset=["postal_code"], keep="first")#
↳drop the duplicate postal code
unique_df2
unique_df2['Score'] = unique_df2['value'].rank(ascending=True, method='dense').
↳astype(int)
unique_df2
```

```
[18]:
```

| | postal_code | data | value | Score |
|----|-------------|--|---------|-------|
| 68 | M5V | 2021A0011M5V - Movers - Counts - Total | 21160.0 | 95 |
| 69 | M5W | 2021A0011M5V - Movers - Counts - Total | 21160.0 | 95 |
| 70 | M5X | 2021A0011M5V - Movers - Counts - Total | 21160.0 | 95 |
| 22 | M2N | 2021A0011M2N - Movers - Counts - Total | 14470.0 | 94 |
| 52 | M4Y | 2021A0011M4Y - Movers - Counts - Total | 10910.0 | 93 |
| .. | ... | ... | ... | ... |
| 55 | M5C | 2021A0011M5C - Movers - Counts - Total | 730.0 | 3 |
| 16 | M1X | 2021A0011M1X - Movers - Counts - Total | 720.0 | 2 |
| 85 | M7A | 2021A0011M7A - Movers - Counts - Total | 0.0 | 1 |
| 87 | M7Y | 2021A0011M7A - Movers - Counts - Total | 0.0 | 1 |

86 M7R 2021A0011M7A - Movers - Counts - Total 0.0 1

[103 rows x 4 columns]

```
[21]: # Now you can export this to a CSV file for further analysis or visulization
      #unique_df.to_csv('after tax income sorted value.csv')
      #unique_df1.to_csv('bachelors degress.csv')
      unique_df2.to_csv('movers.csv')
```

```
[ ]:
```

```
[ ]:
```