Week3Toronto01

March 27, 2019

0.0.1 This notebook is to scrape the Wikipedia page, data wrangling and storing in the dataframe

0.0.2 Prior to start import the necessary libraries

```
In [30]: import pandas as pd
         import numpy as np
         import urllib
         import urllib.request
         import os
         import requests
         from bs4 import BeautifulSoup
```

0.0.3 Define a function to inovoke the Beautifulsoup prior to scraping the Website

```
In [31]: def make_soup(url):
             thepage = urllib.request.urlopen(url)
             soupdata = BeautifulSoup(thepage, "lxml")
             return soupdata
```

0.0.4 Call the function for the Postal codes website

```
In [32]: soup = make_soup("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
```

0.0.5 Scrap the website to retrieve the Post codes table

```
In [33]: Posttblbase = soup.find('table',class_="wikitable sortable")
```

0.0.6 Retrieve the headers of the table

```
In [34]: for record in Posttblbase.findAll('th'):
             print(record.text)
```

```
Postcode
Borough
Neighbourhood
```

0.0.7 Retrieve the rows and the data

```
In [35]: postcdfinl="\n"
         for record in Posttblbase.findAll('tr'):
             postcddata=""
             for data in record.findAll('td'):
                 postcddata=postcddata+","+data.text
             postcdfinl=postcdfinl+postcddata[1:]
```

0.0.8 Assign the Header

```
In [36]: header="Postcode,Borough,Neighbourhood"
```

```
In [37]: ### Check the raw data retrieved from the Website
         print(postcdfinl)
```

```
M1A,Not assigned,Not assigned
M2A,Not assigned,Not assigned
M3A,North York,Parkwoods
M4A,North York,Victoria Village
M5A,Downtown Toronto,Harbourfront
M5A,Downtown Toronto,Regent Park
M6A,North York,Lawrence Heights
M6A,North York,Lawrence Manor
M7A,Queen's Park,Not assigned
M8A,Not assigned,Not assigned
M9A,Etobicoke,Islington Avenue
M1B,Scarborough,Rouge
M1B,Scarborough,Malvern
M2B,Not assigned,Not assigned
M3B,North York,Don Mills North
M4B,East York,Woodbine Gardens
M4B,East York,Parkview Hill
M5B,Downtown Toronto,Ryerson
M5B,Downtown Toronto,Garden District
M6B,North York,Glencairn
M7B,Not assigned,Not assigned
M8B,Not assigned,Not assigned
M9B,Etobicoke,Cloverdale
M9B,Etobicoke,Islington
M9B,Etobicoke,Martin Grove
M9B,Etobicoke,Princess Gardens
M9B,Etobicoke,West Deane Park
M1C,Scarborough,Highland Creek
M1C,Scarborough,Rouge Hill
M1C,Scarborough,Port Union
M2C,Not assigned,Not assigned
M3C,North York,Flemingdon Park
```

```
M3C,North York,Don Mills South
M4C,East York,Woodbine Heights
M5C,Downtown Toronto,St. James Town
M6C,York,Humewood-Cedarvale
M7C,Not assigned,Not assigned
M8C,Not assigned,Not assigned
M9C,Etobicoke,Bloordale Gardens
M9C,Etobicoke,Eringate
M9C,Etobicoke,Markland Wood
M9C,Etobicoke,Old Burnhamthorpe
M1E,Scarborough,Guildwood
M1E,Scarborough,Morningside
M1E,Scarborough,West Hill
M2E,Not assigned,Not assigned
M3E,Not assigned,Not assigned
M4E,East Toronto,The Beaches
M5E,Downtown Toronto,Berczy Park
M6E,York,Caledonia-Fairbanks
M7E,Not assigned,Not assigned
M8E,Not assigned,Not assigned
M9E,Not assigned,Not assigned
M1G,Scarborough,Woburn
M2G,Not assigned,Not assigned
M3G,Not assigned,Not assigned
M4G,East York,Leaside
M5G,Downtown Toronto,Central Bay Street
M6G,Downtown Toronto,Christie
M7G,Not assigned,Not assigned
M8G,Not assigned,Not assigned
M9G,Not assigned,Not assigned
M1H,Scarborough,Cedarbrae
M2H,North York,Hillcrest Village
M3H,North York,Bathurst Manor
M3H,North York,Downsview North
M3H,North York,Wilson Heights
M4H,East York,Thorncliffe Park
M5H,Downtown Toronto,Adelaide
M5H,Downtown Toronto,King
M5H,Downtown Toronto,Richmond
M6H,West Toronto,Dovercourt Village
M6H,West Toronto,Dufferin
M7H,Not assigned,Not assigned
M8H,Not assigned,Not assigned
M9H,Not assigned,Not assigned
M1J,Scarborough,Scarborough Village
M2J,North York,Fairview
M2J,North York,Henry Farm
M2J,North York,Oriole
```

```
M3J,North York,Northwood Park
M3J,North York,York University
M4J,East York,East Toronto
M5J,Downtown Toronto,Harbourfront East
M5J,Downtown Toronto,Toronto Islands
M5J,Downtown Toronto,Union Station
M6J,West Toronto,Little Portugal
M6J,West Toronto,Trinity
M7J,Not assigned,Not assigned
M8J,Not assigned,Not assigned
M9J,Not assigned,Not assigned
M1K,Scarborough,East Birchmount Park
M1K,Scarborough,Ionview
M1K,Scarborough,Kennedy Park
M2K,North York,Bayview Village
M3K,North York,CFB Toronto
M3K,North York,Downsview East
M4K,East Toronto,The Danforth West
M4K,East Toronto,Riverdale
M5K,Downtown Toronto,Design Exchange
M5K,Downtown Toronto,Toronto Dominion Centre
M6K,West Toronto,Brockton
M6K,West Toronto,Exhibition Place
M6K,West Toronto,Parkdale Village
M7K,Not assigned,Not assigned
M8K,Not assigned,Not assigned
M9K,Not assigned,Not assigned
M1L,Scarborough,Clairlea
M1L,Scarborough,Golden Mile
M1L,Scarborough,Oakridge
M2L,North York,Silver Hills
M2L,North York,York Mills
M3L,North York,Downsview West
M4L,East Toronto,The Beaches West
M4L,East Toronto,India Bazaar
M5L,Downtown Toronto,Commerce Court
M5L,Downtown Toronto,Victoria Hotel
M6L,North York,Maple Leaf Park
M6L,North York,North Park
M6L,North York,Upwood Park
M7L,Not assigned,Not assigned
M8L,Not assigned,Not assigned
M9L,North York,Humber Summit
M1M,Scarborough,Cliffcrest
M1M,Scarborough,Cliffside
M1M,Scarborough,Scarborough Village West
M2M,North York,Newtonbrook
M2M,North York,Willowdale
```

```
M3M,North York,Downsview Central
M4M,East Toronto,Studio District
M5M,North York,Bedford Park
M5M,North York,Lawrence Manor East
M6M,York,Del Ray
M6M,York,Keelesdale
M6M,York,Mount Dennis
M6M,York,Silverthorn
M7M,Not assigned,Not assigned
M8M,Not assigned,Not assigned
M9M,North York,Emery
M9M,North York,Humberlea
M1N,Scarborough,Birch Cliff
M1N,Scarborough,Cliffside West
M2N,North York,Willowdale South
M3N,North York,Downsview Northwest
M4N,Central Toronto,Lawrence Park
M5N,Central Toronto,Roselawn
M6N,York,The Junction North
M6N,York,Runnymede
M7N,Not assigned,Not assigned
M8N,Not assigned,Not assigned
M9N,York,Weston
M1P,Scarborough,Dorset Park
M1P,Scarborough,Scarborough Town Centre
M1P,Scarborough,Wexford Heights
M2P,North York,York Mills West
M3P,Not assigned,Not assigned
M4P,Central Toronto,Davisville North
M5P,Central Toronto,Forest Hill North
M5P,Central Toronto,Forest Hill West
M6P,West Toronto,High Park
M6P,West Toronto,The Junction South
M7P,Not assigned,Not assigned
M8P,Not assigned,Not assigned
M9P,Etobicoke,Westmount
M1R,Scarborough,Maryvale
M1R,Scarborough,Wexford
M2R,North York,Willowdale West
M3R,Not assigned,Not assigned
M4R,Central Toronto,North Toronto West
M5R,Central Toronto,The Annex
M5R,Central Toronto,North Midtown
M5R,Central Toronto,Yorkville
M6R,West Toronto,Parkdale
M6R,West Toronto,Roncesvalles
M7R,Mississauga,Canada Post Gateway Processing Centre
M8R,Not assigned,Not assigned
```

```
M9R,Etobicoke,Kingsview Village
M9R,Etobicoke,Martin Grove Gardens
M9R,Etobicoke,Richview Gardens
M9R,Etobicoke,St. Phillips
M1S,Scarborough,Agincourt
M2S,Not assigned,Not assigned
M3S,Not assigned,Not assigned
M4S,Central Toronto,Davisville
M5S,Downtown Toronto,Harbord
M5S,Downtown Toronto,University of Toronto
M6S,West Toronto,Runnymede
M6S,West Toronto,Swansea
M7S,Not assigned,Not assigned
M8S,Not assigned,Not assigned
M9S,Not assigned,Not assigned
M1T,Scarborough,Clarks Corners
M1T,Scarborough,Sullivan
M1T,Scarborough,Tam O'Shanter
M2T,Not assigned,Not assigned
M3T,Not assigned,Not assigned
M4T,Central Toronto,Moore Park
M4T,Central Toronto,Summerhill East
M5T,Downtown Toronto,Chinatown
M5T,Downtown Toronto,Grange Park
M5T,Downtown Toronto,Kensington Market
M6T,Not assigned,Not assigned
M7T,Not assigned,Not assigned
M8T,Not assigned,Not assigned
M9T,Not assigned,Not assigned
M1V,Scarborough,Agincourt North
M1V,Scarborough,L'Amoreaux East
M1V,Scarborough,Milliken
M1V,Scarborough,Steeles East
M2V,Not assigned,Not assigned
M3V,Not assigned,Not assigned
M4V,Central Toronto,Deer Park
M4V,Central Toronto,Forest Hill SE
M4V,Central Toronto,Rathnelly
M4V,Central Toronto,South Hill
M4V,Central Toronto,Summerhill West
M5V,Downtown Toronto,CN Tower
M5V,Downtown Toronto,Bathurst Quay
M5V,Downtown Toronto,Island airport
M5V,Downtown Toronto,Harbourfront West
M5V,Downtown Toronto,King and Spadina
M5V,Downtown Toronto,Railway Lands
M5V,Downtown Toronto,South Niagara
M6V,Not assigned,Not assigned
```

```
M7V,Not assigned,Not assigned
M8V,Etobicoke,Humber Bay Shores
M8V,Etobicoke,Mimico South
M8V,Etobicoke,New Toronto
M9V,Etobicoke,Albion Gardens
M9V,Etobicoke,Beaumond Heights
M9V,Etobicoke,Humbergate
M9V,Etobicoke,Jamestown
M9V,Etobicoke,Mount Olive
M9V,Etobicoke,Silverstone
M9V,Etobicoke,South Steeles
M9V,Etobicoke,Thistletown
M1W,Scarborough,L'Amoreaux West
M1W,Scarborough,Steeles West
M2W,Not assigned,Not assigned
M3W,Not assigned,Not assigned
M4W,Downtown Toronto,Rosedale
M5W,Downtown Toronto,Stn A PO Boxes 25 The Esplanade
M6W,Not assigned,Not assigned
M7W,Not assigned,Not assigned
M8W,Etobicoke,Alderwood
M8W,Etobicoke,Long Branch
M9W,Etobicoke,Northwest
M1X,Scarborough,Upper Rouge
M2X,Not assigned,Not assigned
M3X,Not assigned,Not assigned
M4X,Downtown Toronto,Cabbagetown
M4X,Downtown Toronto,St. James Town
M5X,Downtown Toronto,First Canadian Place
M5X,Downtown Toronto,Underground city
M6X,Not assigned,Not assigned
M7X,Not assigned,Not assigned
M8X,Etobicoke,The Kingsway
M8X,Etobicoke,Montgomery Road
M8X,Etobicoke,Old Mill North
M9X,Not assigned,Not assigned
M1Y,Not assigned,Not assigned
M2Y,Not assigned,Not assigned
M3Y,Not assigned,Not assigned
M4Y,Downtown Toronto,Church and Wellesley
M5Y,Not assigned,Not assigned
M6Y,Not assigned,Not assigned
M7Y,East Toronto,Business Reply Mail Processing Centre 969 Eastern
M8Y,Etobicoke,Humber Bay
M8Y,Etobicoke,King's Mill Park
M8Y,Etobicoke,Kingsway Park South East
M8Y,Etobicoke,Mimico NE
M8Y,Etobicoke,Old Mill South
```

```
M8Y,Etobicoke,The Queensway East
M8Y,Etobicoke,Royal York South East
M8Y,Etobicoke,Sunnylea
M9Y,Not assigned,Not assigned
M1Z,Not assigned,Not assigned
M2Z,Not assigned,Not assigned
M3Z,Not assigned,Not assigned
M4Z,Not assigned,Not assigned
M5Z,Not assigned,Not assigned
M6Z,Not assigned,Not assigned
M7Z,Not assigned,Not assigned
M8Z,Etobicoke,Kingsway Park South West
M8Z,Etobicoke,Mimico NW
M8Z,Etobicoke,The Queensway West
M8Z,Etobicoke,Royal York South West
M8Z,Etobicoke,South of Bloor
M9Z,Not assigned,Not assigned
```

0.0.9 As the next steps, write the data onto a CSV file. For that purpose, open the file, write the header and write the data retrieved above

```
In [41]: file = open(os.path.expanduser("postcodes.csv"),"wb")

In [42]: file.write(bytes(header,encoding="ascii",errors='ignore'))

Out[42]: 30

In [43]: file.write(bytes(postcdfinl,encoding="ascii",errors='ignore'))

Out[43]: 8774
```

0.0.10 Convert the Postcodes data into a Pandas dataframe

```
In [44]: postcd_df = pd.read_csv('postcodes.csv')
```

0.0.11 Check the data in the dataframe

```
In [45]: postcd_df.head()

Out[45]:   Postcode           Borough     Neighbourhood
        0      M1A       Not assigned      Not assigned
        1      M2A       Not assigned      Not assigned
        2      M3A         North York          Parkwoods
        3      M4A         North York   Victoria Village
        4      M5A   Downtown Toronto       Harbourfront
```

0.0.12 Assign the values from column "Borough" to "Neighbourhood" where the value in "Neighbourhood" is equal to Not assigned

```
In [46]: postcd_df['Neighbourhood'] = np.where(postcd_df['Neighbourhood'] == 'Not assigned', pos
```

0.0.13 Remove Rows with Borough as "Not assigned"

```
In [47]: postcd_Onlyassigned = postcd_df[postcd_df.Borough != 'Not assigned']
```

0.0.14 Check if there are any rows with value "Not assigned" in column Burough...They should be none.

```
In [48]: postcd_Onlyassigned.loc[postcd_Onlyassigned['Borough'] == 'Not asigned']
```

```
Out[48]: Empty DataFrame
         Columns: [Postcode, Borough, Neighbourhood]
         Index: []
```

0.0.15 Check if there are any rows with value "Not assigned" in column Neighbourhood...They should be none.

```
In [49]: postcd_Onlyassigned.loc[postcd_Onlyassigned['Neighbourhood'] == 'Not asigned']
```

```
Out[49]: Empty DataFrame
         Columns: [Postcode, Borough, Neighbourhood]
         Index: []
```

0.0.16 Aggregate the rows based on the same value of Postcode

```
In [50]: Postcdgrouped = postcd_Onlyassigned.groupby(['Postcode','Borough'], as_index=False, sor
```

```
In [51]: Postcdgrouped.head()
```

```
Out[51]:    Postcode           Borough                     Neighbourhood
         0       M3A        North York                         Parkwoods
         1       M4A        North York                  Victoria Village
         2       M5A  Downtown Toronto        Harbourfront,Regent Park
         3       M6A        North York  Lawrence Heights,Lawrence Manor
         4       M7A       Queen's Park                      Queen's Park
```

0.0.17 Print the number of rows in the dataframe

```
In [52]: Postcdgrouped.shape
```

```
Out[52]: (103, 3)
```

```
In [ ]: ### Check the data in the dataframe
```

```
In [53]: Postcdgrouped
```

```
Out[53]:      Postcode              Borough  \
        0       M3A          North York
        1       M4A          North York
        2       M5A    Downtown Toronto
        3       M6A          North York
        4       M7A         Queen's Park
        5       M9A           Etobicoke
        6       M1B         Scarborough
        7       M3B          North York
        8       M4B           East York
        9       M5B    Downtown Toronto
        10      M6B          North York
        11      M9B           Etobicoke
        12      M1C         Scarborough
        13      M3C          North York
        14      M4C           East York
        15      M5C    Downtown Toronto
        16      M6C                York
        17      M9C           Etobicoke
        18      M1E         Scarborough
        19      M4E         East Toronto
        20      M5E    Downtown Toronto
        21      M6E                York
        22      M1G         Scarborough
        23      M4G           East York
        24      M5G    Downtown Toronto
        25      M6G    Downtown Toronto
        26      M1H         Scarborough
        27      M2H          North York
        28      M3H          North York
        29      M4H           East York
        ..      ...                 ...
        73      M4R     Central Toronto
        74      M5R     Central Toronto
        75      M6R        West Toronto
        76      M7R         Mississauga
        77      M9R           Etobicoke
        78      M1S         Scarborough
        79      M4S     Central Toronto
        80      M5S    Downtown Toronto
        81      M6S        West Toronto
        82      M1T         Scarborough
        83      M4T     Central Toronto
        84      M5T    Downtown Toronto
        85      M1V         Scarborough
        86      M4V     Central Toronto
        87      M5V    Downtown Toronto
        88      M8V           Etobicoke
```

```
89    M9V         Etobicoke
90    M1W        Scarborough
91    M4W   Downtown Toronto
92    M5W   Downtown Toronto
93    M8W         Etobicoke
94    M9W         Etobicoke
95    M1X        Scarborough
96    M4X   Downtown Toronto
97    M5X   Downtown Toronto
98    M8X         Etobicoke
99    M4Y   Downtown Toronto
100   M7Y        East Toronto
101   M8Y         Etobicoke
102   M8Z         Etobicoke


                                        Neighbourhood
0                                          Parkwoods
1                                   Victoria Village
2                          Harbourfront,Regent Park
3                   Lawrence Heights,Lawrence Manor
4                                       Queen's Park
5                                   Islington Avenue
6                                      Rouge,Malvern
7                                    Don Mills North
8                    Woodbine Gardens,Parkview Hill
9                           Ryerson,Garden District
10                                         Glencairn
11   Cloverdale,Islington,Martin Grove,Princess Gar...
12            Highland Creek,Rouge Hill,Port Union
13           Flemingdon Park,Don Mills South
14                                  Woodbine Heights
15                                   St. James Town
16                                Humewood-Cedarvale
17   Bloordale Gardens,Eringate,Markland Wood,Old B...
18               Guildwood,Morningside,West Hill
19                                        The Beaches
20                                        Berczy Park
21                               Caledonia-Fairbanks
22                                            Woburn
23                                           Leaside
24                                 Central Bay Street
25                                          Christie
26                                          Cedarbrae
27                                  Hillcrest Village
28    Bathurst Manor,Downsview North,Wilson Heights
29                                 Thorncliffe Park
..                                              ...
73                                 North Toronto West
```

```
74                  The Annex,North Midtown,Yorkville
75                         Parkdale,Roncesvalles
76              Canada Post Gateway Processing Centre
77   Kingsview Village,Martin Grove Gardens,Richvie...
78                                        Agincourt
79                                       Davisville
80                 Harbord,University of Toronto
81                               Runnymede,Swansea
82        Clarks Corners,Sullivan,Tam O'Shanter
83                    Moore Park,Summerhill East
84           Chinatown,Grange Park,Kensington Market
85   Agincourt North,L'Amoreaux East,Milliken,Steel...
86   Deer Park,Forest Hill SE,Rathnelly,South Hill,...
87   CN Tower,Bathurst Quay,Island airport,Harbourf...
88        Humber Bay Shores,Mimico South,New Toronto
89   Albion Gardens,Beaumond Heights,Humbergate,Jam...
90                 L'Amoreaux West,Steeles West
91                                         Rosedale
92              Stn A PO Boxes 25 The Esplanade
93                         Alderwood,Long Branch
94                                        Northwest
95                                      Upper Rouge
96                  Cabbagetown,St. James Town
97              First Canadian Place,Underground city
98        The Kingsway,Montgomery Road,Old Mill North
99                            Church and Wellesley
100  Business Reply Mail Processing Centre 969 Eastern
101  Humber Bay,King's Mill Park,Kingsway Park Sout...
102  Kingsway Park South West,Mimico NW,The Queensw...

[103 rows x 3 columns]
```

In [ ]:

In [ ]:

In [ ]: