Wk3TorontoGraded

January 8, 2019

1 GRADED ASSIGNMENT FOR WEEK3 Segmenting and Clustering Neighborhoods in Toronto

Import some code from "Segmenting and Clustering Neighborhoods in New York City" lab

In [ ]:
```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed
from geopy.geocoders import Nominatim # convert an address into latitude and longitude v

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't co
import folium # map rendering library

print('Libraries imported.')
```

Import BeautfulSoup to scrape Wikipedia page that lists Toronto neighborhoods and postal codes

In [2]:
```
from bs4 import BeautifulSoup as BS
import urllib.request
```

Now turn Wikipedia table into a dataframe

```
In [3]: res = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
        soup = BS(res.content,'lxml')
        table = soup.find_all('table')[0]
        df = pd.read_html(str(table))[0]
        postcode = df[0].tolist()
        borough = df[1].tolist()
        neigh = df[2].tolist()

        df.head(10)
```

Out[3]:

	0	1	2
0	Postcode	Borough	Neighbourhood
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village
5	M5A	Downtown Toronto	Harbourfront
6	M5A	Downtown Toronto	Regent Park
7	M6A	North York	Lawrence Heights
8	M6A	North York	Lawrence Manor
9	M7A	Queen's Park	Not assigned

Delete false header row 0 - Can rename headers later

```
In [4]: df = df.drop(0)
```

Now eliminate rows with unassigned postcode

```
In [5]: df = df.drop(df[df[1] == "Not assigned"].index)
```

Now replace "Not assigned" Neighbourhood values with Borough value

```
In [6]: df[2] = df[2].replace("Not assigned", df[1])
```

Now combine neighbourhoods in the same postcode/borough into one record with Neighbourhood values comma delimited

```
In [7]: df = pd.DataFrame({0:df[0],1:df[1],2:df[2]})

        df = df.groupby([0,1])[2].apply(lambda tags : ','.join(tags)).reset_index()
```

Now rename columns as "PostalCode" "Borough" and "Neighbourhood"

```
In [8]: df = df.rename(columns = {0:"PostalCode", 1:"Borough", 2:"Neighbourhood"})

        df
```

Out[8]:

	PostalCode	Borough	\
0	M1B	Scarborough	
1	M1C	Scarborough	

2	M1E	Scarborough
3	M1G	Scarborough
4	M1H	Scarborough
5	M1J	Scarborough
6	M1K	Scarborough
7	M1L	Scarborough
8	M1M	Scarborough
9	M1N	Scarborough
10	M1P	Scarborough
11	M1R	Scarborough
12	M1S	Scarborough
13	M1T	Scarborough
14	M1V	Scarborough
15	M1W	Scarborough
16	M1X	Scarborough
17	M2H	North York
18	M2J	North York
19	M2K	North York
20	M2L	North York
21	M2M	North York
22	M2N	North York
23	M2P	North York
24	M2R	North York
25	M3A	North York
26	M3B	North York
27	M3C	North York
28	M3H	North York
29	M3J	North York
30	M3K	North York
31	M3L	North York
32	M3M	North York
33	M3N	North York
34	M4A	North York
35	M4B	East York
36	M4C	East York
37	M4E	East Toronto
38	M4G	East York
39	M4H	East York
40	M4J	East York
41	M4K	East Toronto
42	M4L	East Toronto
43	M4M	East Toronto
44	M4N	Central Toronto
45	M4P	Central Toronto
46	M4R	Central Toronto
47	M4S	Central Toronto
48	M4T	Central Toronto
49	M4V	Central Toronto

50	M4W	Downtown Toronto
51	M4X	Downtown Toronto
52	M4Y	Downtown Toronto
53	M5A	Downtown Toronto
54	M5B	Downtown Toronto
55	M5C	Downtown Toronto
56	M5E	Downtown Toronto
57	M5G	Downtown Toronto
58	M5H	Downtown Toronto
59	M5J	Downtown Toronto
60	M5K	Downtown Toronto
61	M5L	Downtown Toronto
62	M5M	North York
63	M5N	Central Toronto
64	M5P	Central Toronto
65	M5R	Central Toronto
66	M5S	Downtown Toronto
67	M5T	Downtown Toronto
68	M5V	Downtown Toronto
69	M5W	Downtown Toronto
70	M5X	Downtown Toronto
71	M6A	North York
72	M6B	North York
73	M6C	York
74	M6E	York
75	M6G	Downtown Toronto
76	M6H	West Toronto
77	M6J	West Toronto
78	M6K	West Toronto
79	M6L	North York
80	M6M	York
81	M6N	York
82	M6P	West Toronto
83	M6R	West Toronto
84	M6S	West Toronto
85	M7A	Queen's Park
86	M7R	Mississauga
87	M7Y	East Toronto
88	M8V	Etobicoke
89	M8W	Etobicoke
90	M8X	Etobicoke
91	M8Y	Etobicoke
92	M8Z	Etobicoke
93	M9A	Etobicoke
94	M9B	Etobicoke
95	M9C	Etobicoke
96	M9L	North York
97	M9M	North York

98	M9N	York	
99	M9P	Etobicoke	
100	M9R	Etobicoke	
101	M9V	Etobicoke	
102	M9W	Etobicoke	

	Neighbourhood
0	Rouge,Malvern
1	Highland Creek,Rouge Hill,Port Union
2	Guildwood,Morningside,West Hill
3	Woburn
4	Cedarbrae
5	Scarborough Village
6	East Birchmount Park,Ionview,Kennedy Park
7	Clairlea,Golden Mile,Oakridge
8	Cliffcrest,Cliffside,Scarborough Village West
9	Birch Cliff,Cliffside West
10	Dorset Park,Scarborough Town Centre,Wexford He...
11	Maryvale,Wexford
12	Agincourt
13	Clarks Corners,Sullivan,Tam O'Shanter
14	Agincourt North,L'Amoreaux East,Milliken,Steel...
15	L'Amoreaux West,Steeles West
16	Upper Rouge
17	Hillcrest Village
18	Fairview,Henry Farm,Oriole
19	Bayview Village
20	Silver Hills,York Mills
21	Newtonbrook,Willowdale
22	Willowdale South
23	York Mills West
24	Willowdale West
25	Parkwoods
26	Don Mills North
27	Flemingdon Park,Don Mills South
28	Bathurst Manor,Downsview North,Wilson Heights
29	Northwood Park,York University
30	CFB Toronto,Downsview East
31	Downsview West
32	Downsview Central
33	Downsview Northwest
34	Victoria Village
35	Woodbine Gardens,Parkview Hill
36	Woodbine Heights
37	The Beaches
38	Leaside
39	Thorncliffe Park
40	East Toronto

```
41                         The Danforth West,Riverdale
42                       The Beaches West,India Bazaar
43                                    Studio District
44                                      Lawrence Park
45                                   Davisville North
46                                 North Toronto West
47                                          Davisville
48                           Moore Park,Summerhill East
49  Deer Park,Forest Hill SE,Rathnelly,South Hill,...
50                                           Rosedale
51                           Cabbagetown,St. James Town
52                                 Church and Wellesley
53                             Harbourfront,Regent Park
54                              Ryerson,Garden District
55                                       St. James Town
56                                          Berczy Park
57                                   Central Bay Street
58                                Adelaide,King,Richmond
59    Harbourfront East,Toronto Islands,Union Station
60            Design Exchange,Toronto Dominion Centre
61                       Commerce Court,Victoria Hotel
62                    Bedford Park,Lawrence Manor East
63                                           Roselawn
64                  Forest Hill North,Forest Hill West
65                   The Annex,North Midtown,Yorkville
66                       Harbord,University of Toronto
67             Chinatown,Grange Park,Kensington Market
68  CN Tower,Bathurst Quay,Island airport,Harbourf...
69                      Stn A PO Boxes 25 The Esplanade
70                First Canadian Place,Underground city
71                      Lawrence Heights,Lawrence Manor
72                                           Glencairn
73                                 Humewood-Cedarvale
74                               Caledonia-Fairbanks
75                                           Christie
76                          Dovercourt Village,Dufferin
77                              Little Portugal,Trinity
78          Brockton,Exhibition Place,Parkdale Village
79              Maple Leaf Park,North Park,Upwood Park
80         Del Ray,Keelesdale,Mount Dennis,Silverthorn
81                        The Junction North,Runnymede
82                        High Park,The Junction South
83                               Parkdale,Roncesvalles
84                                   Runnymede,Swansea
85                                        Queen's Park
86                Canada Post Gateway Processing Centre
87  Business Reply Mail Processing Centre 969 Eastern
88          Humber Bay Shores,Mimico South,New Toronto
```

```
89                                Alderwood,Long Branch
90         The Kingsway,Montgomery Road,Old Mill North
91   Humber Bay,King's Mill Park,Kingsway Park Sout...
92   Kingsway Park South West,Mimico NW,The Queensw...
93                                    Islington Avenue
94   Cloverdale,Islington,Martin Grove,Princess Gar...
95   Bloordale Gardens,Eringate,Markland Wood,Old B...
96                                      Humber Summit
97                                    Emery,Humberlea
98                                             Weston
99                                          Westmount
100  Kingsview Village,Martin Grove Gardens,Richvie...
101  Albion Gardens,Beaumond Heights,Humbergate,Jam...
102                                         Northwest
```

Use .shape method to display number of rows in dataframe

```
In [9]: df.shape
```

```
Out[9]: (103, 3)
```

Dataframe has 103 rows and 3 columns