

WK3 assignment

June 17, 2020

```
[51]: import numpy as np #handle data in a vectorized manner
import pandas as pd #data analysis
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)
import matplotlib.pyplot as plt # generate figure
import requests # library to handle requests
import json #library to handle JSON files
import folium #map rendering library
!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you
    ↪ haven't completed the Foursquare API lab
import xml
import matplotlib.cm as cm # Matplotlib and associated plotting modules
import matplotlib.colors as colors # Matplotlib and associated plotting modules
!conda install -c conda-forge geopy --yes # uncomment this line if you haven't
    ↪ completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and
    ↪ longitude values
from bs4 import BeautifulSoup #for webscraping import BeautifulSoup
from pandas.io.json import json_normalize # tranform Json file into a pandas
    ↪ dataframe
from sklearn.cluster import KMeans # import k-means from clustering stage
```

Collecting package metadata (current_repodata.json): done

Solving environment: done

All requested packages already installed.

Collecting package metadata (current_repodata.json): done

Solving environment: done

All requested packages already installed.

Get the data from Wikipedia

```
[52]: url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
result=requests.get(url) #request data
```

```
soup=BeautifulSoup(result.content, 'html.parser') # parse data from the url
↳ into a beautifulsoup object
print(url)
print(result.status_code)
print(result.headers)
```

```
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
200
{'Date': 'Wed, 17 Jun 2020 16:48:47 GMT', 'Vary': 'Accept-
Encoding, Cookie, Authorization', 'Server': 'ATS/8.0.7', 'X-Content-Type-Options':
'nosniff', 'P3p': 'CP="See
https://en.wikipedia.org/wiki/Special:CentralAutoLogin/P3P for more info."',
'Content-Language': 'en', 'Last-Modified': 'Sun, 14 Jun 2020 17:04:41 GMT',
'Content-Type': 'text/html; charset=UTF-8', 'Content-Encoding': 'gzip', 'Age':
'5033', 'X-Cache': 'cp3064 hit, cp3058 hit/19', 'X-Cache-Status': 'hit-front',
'Server-Timing': 'cache;desc="hit-front"', 'Strict-Transport-Security': 'max-
age=106384710; includeSubDomains; preload', 'Set-Cookie': 'WMF-Last-
Access=17-Jun-2020; Path=/; HttpOnly; secure; Expires=Sun, 19 Jul 2020 12:00:00 GMT,
WMF-Last-Access-
Global=17-Jun-2020; Path=/; Domain=.wikipedia.org; HttpOnly; secure; Expires=Sun, 19
Jul 2020 12:00:00 GMT, GeoIP=GB:ENG:Walthamstow:51.59:-0.02:v4; Path=/; secure;
Domain=.wikipedia.org', 'X-Client-IP': '82.15.178.91', 'Cache-Control':
'private, s-maxage=0, max-age=0, must-revalidate', 'Accept-Ranges': 'bytes',
'Content-Length': '13149', 'Connection': 'keep-alive'}
```

```
[53]: post_data=soup.find('table')
fields=post_data.find_all('td')
# create three lists to store table data
postcode = []
borough = []
neighbourhood = []
```

```
[65]: # append the data into the respective lists
for i in range(0, len(fields), 3):
    postcode.append(fields[i].text.strip())
    borough.append(fields[i+1].text.strip())
    neighbourhood.append(fields[i+2].text.strip())

df_post=pd.DataFrame(data=[postcode,borough,neighbourhood]).transpose()
df_post.columns=['Postcode','Borough','Neighborhood']
df_post.head()
```

```
[65]:
```

	Postcode	Borough	Neighborhood
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 Regent Park, Harbourfront

```
[66]: df_post.shape
```

```
[66]: (720, 3)
```

delete 'Not assigned' value in Borough

```
[71]: df_postcode=df_post[df_post.Borough !='Not assigned'].reset_index(drop=True)
df_postcode.head()
```

```
[71]:
```

	Postcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Group neighborhood in the same borough

```
[69]: df_post_group=df_postcode.groupby(['Postcode','Borough'],as_index=False).
      ↪agg(lambda x:','.join(x))
df_post_group
```

```
[69]:
```

	Postcode	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	Downtown Toronto
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

Neighborhood

0	Malvern, Rouge,Malvern, Rouge,Malvern, Rouge,M...
1	Rouge Hill, Port Union, Highland Creek,Rouge H...
2	Guildwood, Morningside, West Hill,Guildwood, M...
3	Woburn,Woburn,Woburn,Woburn
4	Cedarbrae,Cedarbrae,Cedarbrae,Cedarbrae
5	Scarborough Village,Scarborough Village,Scarbo...
6	Kennedy Park, Ionview, East Birchmount Park,Ke...
7	Golden Mile, Clairlea, Oakridge,Golden Mile, C...
8	Cliffside, Cliffcrest, Scarborough Village Wes...
9	Birch Cliff, Cliffside West,Birch Cliff, Cliff...
10	Dorset Park, Wexford Heights, Scarborough Town...
11	Wexford, Maryvale,Wexford, Maryvale,Wexford, M...

12 Agincourt,Agincourt,Agincourt,Agincourt
 13 Clarks Corners, Tam O'Shanter, Sullivan,Clarks...
 14 Milliken, Agincourt North, Steeles East, L'Amo...
 15 Steeles West, L'Amoreaux West,Steeles West, L'...
 16 Upper Rouge,Upper Rouge,Upper Rouge,Upper Rouge
 17 Hillcrest Village,Hillcrest Village,Hillcrest ...
 18 Fairview, Henry Farm, Oriole,Fairview, Henry F...
 19 Bayview Village,Bayview Village,Bayview Villag...
 20 York Mills, Silver Hills,York Mills, Silver Hi...
 21 Willowdale, Newtonbrook,Willowdale, Newtonbroo...
 22 Willowdale, Willowdale East,Willowdale, Willow...
 23 York Mills West,York Mills West,York Mills Wes...
 24 Willowdale, Willowdale West,Willowdale, Willow...
 25 Parkwoods,Parkwoods,Parkwoods,Parkwoods
 26 Don Mills,Don Mills,Don Mills,Don Mills
 27 Don Mills,Don Mills,Don Mills,Don Mills
 28 Bathurst Manor, Wilson Heights, Downsview Nort...
 29 Northwood Park, York University,Northwood Park...
 30 Downsview,Downsview,Downsview,Downsview
 31 Downsview,Downsview,Downsview,Downsview
 32 Downsview,Downsview,Downsview,Downsview
 33 Downsview,Downsview,Downsview,Downsview
 34 Victoria Village,Victoria Village,Victoria Vil...
 35 Parkview Hill, Woodbine Gardens,Parkview Hill,...
 36 Woodbine Heights,Woodbine Heights,Woodbine Hei...
 37 The Beaches,The Beaches,The Beaches,The Beaches
 38 Leaside,Leaside,Leaside,Leaside
 39 Thorncliffe Park,Thorncliffe Park,Thorncliffe ...
 40 East Toronto, Broadview North (Old East York),...
 41 The Danforth West, Riverdale,The Danforth West...
 42 India Bazaar, The Beaches West,India Bazaar, T...
 43 Studio District,Studio District,Studio Distric...
 44 Lawrence Park,Lawrence Park,Lawrence Park,Lawr...
 45 Davisville North,Davisville North,Davisville N...
 46 North Toronto West, Lawrence Park,North Toron...
 47 Davisville,Davisville,Davisville,Davisville
 48 Moore Park, Summerhill East,Moore Park, Summer...
 49 Summerhill West, Rathnelly, South Hill, Forest...
 50 Rosedale,Rosedale,Rosedale,Rosedale
 51 St. James Town, Cabbagetown,St. James Town, Ca...
 52 Church and Wellesley,Church and Wellesley,Chur...
 53 Regent Park, Harbourfront,Regent Park, Harbour...
 54 Garden District, Ryerson,Garden District, Ryer...
 55 St. James Town,St. James Town,St. James Town,S...
 56 Berczy Park,Berczy Park,Berczy Park,Berczy Park
 57 Central Bay Street,Central Bay Street,Central ...
 58 Richmond, Adelaide, King,Richmond, Adelaide, K...

59 Harbourfront East, Union Station, Toronto Isla...
 60 Toronto Dominion Centre, Design Exchange,Toron...
 61 Commerce Court, Victoria Hotel,Commerce Court,...
 62 Bedford Park, Lawrence Manor East,Bedford Park...
 63 Roselawn,Roselawn,Roselawn,Roselawn
 64 Forest Hill North & West, Forest Hill Road Par...
 65 The Annex, North Midtown, Yorkville,The Annex,...
 66 University of Toronto, Harbord,University of T...
 67 Kensington Market, Chinatown, Grange Park,Kens...
 68 CN Tower, King and Spadina, Railway Lands, Har...
 69 Stn A PO Boxes,Stn A PO Boxes,Stn A PO Boxes,S...
 70 First Canadian Place, Underground city,First C...
 71 Lawrence Manor, Lawrence Heights,Lawrence Mano...
 72 Glencairn,Glencairn,Glencairn,Glencairn
 73 Humewood-Cedarvale,Humewood-Cedarvale,Humewood...
 74 Caledonia-Fairbanks,Caledonia-Fairbanks,Caledo...
 75 Christie,Christie,Christie,Christie
 76 Dufferin, Dovercourt Village,Dufferin, Doverco...
 77 Little Portugal, Trinity,Little Portugal, Trin...
 78 Brockton, Parkdale Village, Exhibition Place,B...
 79 North Park, Maple Leaf Park, Upwood Park,North...
 80 Del Ray, Mount Dennis, Keelsdale and Silvertho...
 81 Runnymede, The Junction North,Runnymede, The J...
 82 High Park, The Junction South,High Park, The J...
 83 Parkdale, Roncesvalles,Parkdale, Roncesvalles,...
 84 Runnymede, Swansea,Runnymede, Swansea,Runnymed...
 85 Queen's Park, Ontario Provincial Government,Qu...
 86 Canada Post Gateway Processing Centre,Canada P...
 87 Business reply mail Processing Centre, South C...
 88 New Toronto, Mimico South, Humber Bay Shores,N...
 89 Alderwood, Long Branch,Alderwood, Long Branch,...
 90 The Kingsway, Montgomery Road, Old Mill North,...
 91 Old Mill South, King's Mill Park, Sunnylea, Hu...
 92 Mimico NW, The Queensway West, South of Bloor,...
 93 Islington Avenue, Humber Valley Village,Isling...
 94 West Deane Park, Princess Gardens, Martin Grov...
 95 Eringate, Bloordale Gardens, Old Burnhamthorpe...
 96 Humber Summit,Humber Summit,Humber Summit,Humb...
 97 Humberlea, Emery,Humberlea, Emery,Humberlea, E...
 98 Weston,Weston,Weston,Weston
 99 Westmount,Westmount,Westmount,Westmount
 100 Kingsview Village, St. Phillips, Martin Grove ...
 101 South Steeles, Silverstone, Humbergate, Jamest...
 102 Northwest, West Humber - Clairville,Northwest,...

For Neighborhood="Not assigned", make the value the same as Borough

```
[73]: for inex,row in df_post_group.iterrows():
      if row['Neighborhood']=='Not assigned':
          row['Neighborhood']=row['Borough']
```

```
df_post_group.head()
```

```
[73]:   Postcode      Borough      Neighborhood
0      M1B  Scarborough  Malvern, Rouge,Malvern, Rouge,Malvern, Rouge,M...
1      M1C  Scarborough  Rouge Hill, Port Union, Highland Creek,Rouge H...
2      M1E  Scarborough  Guildwood, Morningside, West Hill,Guildwood, M...
3      M1G  Scarborough                Woburn,Woburn,Woburn,Woburn
4      M1H  Scarborough  Cedarbrae,Cedarbrae,Cedarbrae,Cedarbrae
```

Check if data is match the requiry of question

```
[75]: #create new dataframe for test
columns_name=['Postcode','Borough','Neighborhood']
df_test=pd.DataFrame(columns=columns_name)

list_test=['M5G','M2H','M1B','M1J','M4M','M1C','M1H']

for postcode in list_test:
    df_test=df_test.
    →append(df_post_group[df_post_group['Postcode']==postcode],ignore_index=True)

df_test
```

```
[75]:   Postcode      Borough \
0      M5G  Downtown Toronto
1      M2H      North York
2      M1B      Scarborough
3      M1J      Scarborough
4      M4M  East Toronto
5      M1C      Scarborough
6      M1H      Scarborough

      Neighborhood
0  Central Bay Street,Central Bay Street,Central ...
1  Hillcrest Village,Hillcrest Village,Hillcrest ...
2  Malvern, Rouge,Malvern, Rouge,Malvern, Rouge,M...
3  Scarborough Village,Scarborough Village,Scarbo...
4  Studio District,Studio District,Studio Distric...
5  Rouge Hill, Port Union, Highland Creek,Rouge H...
6      Cedarbrae,Cedarbrae,Cedarbrae,Cedarbrae
```

```
[76]: df_post_group.shape
```


[76]: (103, 3)

Load the coordinates

```
[78]: df_geo = pd.read_csv('http://cocl.us/Geospatial_data')
df_geo.columns = ['Postcode', 'Latitude', 'Longitude']

df_po=pd.merge(df_post_group,df_geo,on=['Postcode'],how='inner')
df_tor=df_po[['Borough','Neighborhood','Postcode','Latitude','Longitude']].
    ↪copy()

df_tor.head()
```

```
[78]:
```

	Borough	Neighborhood	Postcode	\
0	Scarborough	Malvern, Rouge,Malvern, Rouge,Malvern, Rouge,M...	M1B	
1	Scarborough	Rouge Hill, Port Union, Highland Creek,Rouge H...	M1C	
2	Scarborough	Guildwood, Morningside, West Hill,Guildwood, M...	M1E	
3	Scarborough	Woburn,Woburn,Woburn,Woburn	M1G	
4	Scarborough	Cedarbrae,Cedarbrae,Cedarbrae,Cedarbrae	M1H	

	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

```
[79]: address='Toronto, Canada'
geolocator=Nominatim()
location=geolocator.geocode(address)
latitude=location.latitude
longitude=location.longitude
print('The coordinate of the city of Totonto are{} ,{}.'.
    ↪format(latitude,longitude))
```

The coordinate of the city of Totonto are43.6534817,-79.3839347.

/Users/arielchien/opt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2: DeprecationWarning: Using Nominatim with the default "geopy/1.22.0" `user_agent` is strongly discouraged, as it violates Nominatim's ToS <https://operations.osmfoundation.org/policies/nominatim/> and may possibly cause 403 and 429 HTTP errors. Please specify a custom `user_agent` with `Nominatim(user_agent="my-application")` or by overriding the default `user_agent`: `geopy.geocoders.options.default_user_agent = "my-application"`. In geopy 2.0 this will become an exception.

```
[82]: #create map of Toronto using latitude and longitude values
map_toronto=folium.Map(location=[latitude,longitude],zoom_start=14)

#adding markers to map

for lat,lng,borough,neighborhood in zip(df_tor['Latitude'],df_tor['Longitude'],df_tor['Borough'],df_tor['Neighborhood']):
    label='{},{}'.format(neighborhood,borough)
    label=folium.Popup(label,parse_html=True)
    folium.CircleMarker([lat,lng],radius=5,popup=label,color='blue',fill=True,fill_color='#3186cc',fill_opacity=0.6,parse_html=False).add_to(map_toronto)

map_toronto
```

```
[82]: <folium.folium.Map at 0x1a20e3c8d0>
```