

The battle of Toronto neighborhoods

Green edition





What will be discussed? And why?

Environmental related topics are getting popular everyday as mankind know more about the opportunity cost of economic (especially manufacturing) activities. People may want to know how polluted his/her neighborhoods is as pollution may affect,

- Health of the residents
- Values of properties



Tools utilized

- IBM Cloud Object Storage (<https://www.ibm.com/cloud/object-storage>)
- IBM Watson Studio (<https://www.ibm.com/watson/studio>)
- Pandas (<https://pandas.pydata.org/pandas-docs/stable/index.html#>)
- Scikit-learn (<https://scikit-learn.org/stable/>)
- Folium (<https://python-visualization.github.io/folium/>)



Data acquisition and cleaning

1. Postal codes of neighborhoods in Toronto provided by Wikipedia.
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
 - a. 103 rows in dataframe
2. Provided csv which contains the latitude and longitude of Toronto's neighborhoods mapped by their postal codes.
3. Merge result from 2 into dataframe obtained from 1

	Postal code		Borough	Neighborhood	Latitude	Longitude
0	M3A		North York	Parkwoods	43.753259	-79.329656
1	M4A		North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto		Regent Park / Harbourfront	43.654260	-79.360636
3	M6A		North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government		43.662301	-79.389494



Data acquisition and cleaning (cont.)

1. Call Foursquare's Venue Search API (<https://api.foursquare.com/v2/venues/search>) for each pair of latitude and longitude from step 3 of previous page
 - a. For each response, group and count for each desired categories and put into another dataframe

	Garden	National Park	Other Great Outdoors	Park	Total Number of Green Zones	Industrial Estate	Power Plant	Waste Facility	Total Number of W.I.P. Zones	Postal code
0	0	0	8	14	22	0	0	0	0	M3A
0	3	0	9	15	27	0	0	0	0	M4A
0	8	0	16	23	47	1	0	0	1	M5A
0	1	0	3	12	16	1	0	4	5	M6A
0	8	0	12	28	48	1	0	0	1	M7A



What is next? What more need to be done?

- KMeans clustering from scikit-learn (<https://scikit-learn.org/>)
 - $k = 5$
 - Hope to be able to get 5 clusters (greenest, above avg, avg, below avg, most polluted)



Result summary

```
In [4]: # Numbers of neighborhoods in each cluster:  
result_df['Cluster label'].value_counts()
```

```
Out[4]:  0      38  
        1      23  
        4      21  
        2      11  
        3      10
```



Result summary (Cont.)

```
In [10]: # Mean of total number of green zones for each cluster:  
result_df.groupby('Cluster label')['Total Number of Green Zones'].agg(numpy.mean)
```

```
Out[10]: Cluster label  
0      45.763158  
1      23.086957  
2       7.818182  
3      32.300000  
4      15.476190
```

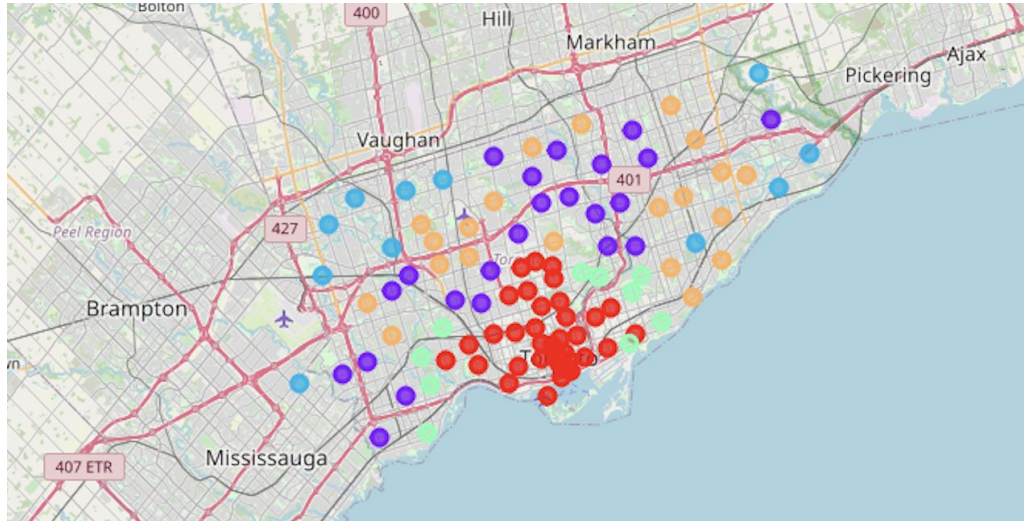



Result summary (Cont.)

```
In [11]: # Mean of total number of W.I.P. (Waste Facility, Industrial Estate, Power Station) for each cluster:  
result_df.groupby('Cluster label')['Total Number of W.I.P. Zones'].agg(numpy.mean)
```

```
Out[11]: Cluster label  
0      0.736842  
1      0.217391  
2      1.000000  
3      0.500000  
4      1.238095
```

The big picture



- Red: Cluster 0
- Purple: Cluster 1
- Cyan: Cluster 2
- Light blue: Cluster 3
- Orange: Cluster 4



Result analysis

Cluster	
0:	Relatively high # of green zones, relatively moderate # of W.I.P. zones
1:	Relatively moderate # of green zones, relatively low # of W.I.P. zones
2:	Relatively low # of green zones, relatively high # of W.I.P. zones
3:	Relatively high # of green zones, relatively low # of W.I.P. zones
4:	Relatively moderate # of green zones, relatively high # of W.I.P. zones

- Do not be tricked by the description.
 - I.e. Cluster 2:
 - Means of total number W.I.P. zones for the cluster is only 1
 - Means of total number of green zones for the cluster is ~8
- Within Toronto, neighborhoods do seem green.
 - May mean that industrial areas, power stations, waste facilities were moved away from Toronto



Result analysis (Cont.)

- Inconclusive
 - Need to take a look at cities/town around Toronto also
 - More data (i.e. Air pollution index of each neighborhood) is needed
 - May need to involve weather forecast/pattern