

Report

K-MEANS SEGMENTING AND CLUSTERING VENUES IN TORONTO TO INSERT A COVID19 MEMORIAL OBELISK AS A POINT OF INTEREST

Alejandro Duarte

fd.alejandro@uol.com.br

Applied Data Science Capstone by IBM/Coursera
Capstone Project - The Battle of Neighborhoods (2)

INTRODUCTION

The years 2019, 2020 and 2021 are marked in history by the experiences of loss of human lives for the SARS-CoV-2, by the impossibility of maintaining the routine of social contacts, by the dedication of medical personnel to care and save lives, by the interdisciplinary union of science and technology in the development of vaccines, for the sufferings and the hope in all parts of the world.

At the end of the present work, It's the situation: **Globally**, as of **9:38am CET, 6 January 2021**, there have been **84.780.171 confirmed cases** of COVID-19, including **1.853.525 deaths** (WHO, 2021).

The memory of these times will remain and the idea of erecting an obelisk in each country can be a contribution to the junction of nations in one world, the Earth planet.

Such obelisks can be located with the aid of the FourSquire places API. They would be integrated to the landscape and the Points of Interest that are in the daily lives of residents and visitors. The present work advances the imagination, of a project to locate a Covid19 Memorial Obelisk in Toronto, Canada.

Objectives

The present task has two main purposes:

1. A description of the problem and a discussion of the background

We crossed the year 2020 and arrived in 2021 six days ago. These times has been marked by the covid19 pandemic. Possibly every country in the world will erect a monument to the victims of the disease. In this work, an obelisk memorial will be "located" in Roselawn neighborhood, Toronto, Canada.

The location was carried out a priori, the maps and places of daily interest were structured using the FourSquare platform. However, this platform does not offer suggestions for visiting points related to the historical memory of places, but for utilitarian purposes of services. Anyway, the locations offered by FOURSQUARE (2020) are extremely useful because they can be associated with any other purpose related to points of interest. Thus, segmentation and clustering of venues in Toronto were used to highlight the insertion of the memorial obelisk.

2. A description of the data and how it will be used to solve the problem

The data for carrying out the task were taken from the World Health Organization (WHO) website and from the Segmenting and Clustering Neighborhoods in Toronto notebook (COURSERA, 2020a; b), and WIKIPEDIA (2020).

The data consists of names, postal codes and geographical coordinates related to boroughs, neighborhoods, venues in Toronto and, to allow the location of a monument in honor to the victims of the covid19. Dataframes were organized and, using two clustering schemes, a place

in Roselawn (Forest Hill North neighborhood) was positioned on a map to raise the obelisk. Postal code information for Toronto were transformed to geographical coordinates. Data archives were saved to local disk, in csv or xlsx format, to guarantee immediate access.

Notebook documentation

Each step is documented on the notebook under the following headings:

- Installing libraries for data analysis
- Data collection and dataframes
- Neighborhood Geolocation and mapping
- The 100 important venues in Toronto
- Including Covid19 Memorial Obelisk geolocation, in Roselawn, Toronto – CA
- Venues for each neighborhood
- Most common venues grouped by neighborhood
- Clustering in map (5 clusters)
- Localization of the Toronto Covid19 Memorial Obelisk
- Determination of optimal number of clusters on venues (Elbow curve)
- Optimal number of clusters = 3
- Graphic of venues segmentation around the Covid19 Memorial Obelisk in Toronto
- Centroids in blue color, venues grouped in yellow, green, and violet color, Covid19 Monumental Obelisk in red color

Various models for elaborating projects and report to fulfill the objectives outlined has been published (KOROS, 2020; MAGIYA, 2019).

METHODOLOGY

```
[4]: gc = pd.read_excel('Geospatial_Coordinates_1.xlsx')
gc
```

```
[4]:
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.77314	-79.239476
...
98	M9N	York	Weston	43.70688	-79.518188
99	M9P	Etobicoke	Westmount	43.69632	-79.532242
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054

103 rows × 5 columns

Venues for each neighborhood

```
[16]: toronto_venues.groupby('Neighbourhood').count()
```

```
[16]:
```

	Unnamed: 0	Neighbourhood	Latitude	Neighbourhood	Longitude	Venue	Venue	Latitude	Venue	Longitude	Venue	Category
Neighbourhood												
Agincourt	4		4			4	4		4		4	4
Alderwood, Long Branch	6		6			6	6		6		6	6
Bathurst Manor, Wilson Heights, Downsview North	22		22			22	22		22		22	22
Bedford Park, Lawrence Manor East	26		26			26	26		26		26	26
Berczy Park	58		58			58	58		58		58	58
...
Willowdale, Willowdale East	35		35			35	35		35		35	35
Willowdale, Willowdale West	5		5			5	5		5		5	5
Woburn	3		3			3	3		3		3	3
Woodbine Heights	8		8			8	8		8		8	8
York Mills West	2		2			2	2		2		2	2

73 rows × 7 columns

Figure 1. Primary data and quantitative information on neighborhoods.

Figure 1 is composed by two tables that present the primary data and quantitative information and identifications of postal code, neighborhoods and geographic coordinates of venues in Toronto. This information is the basis for its geospatial representation on the map in Figure 2.

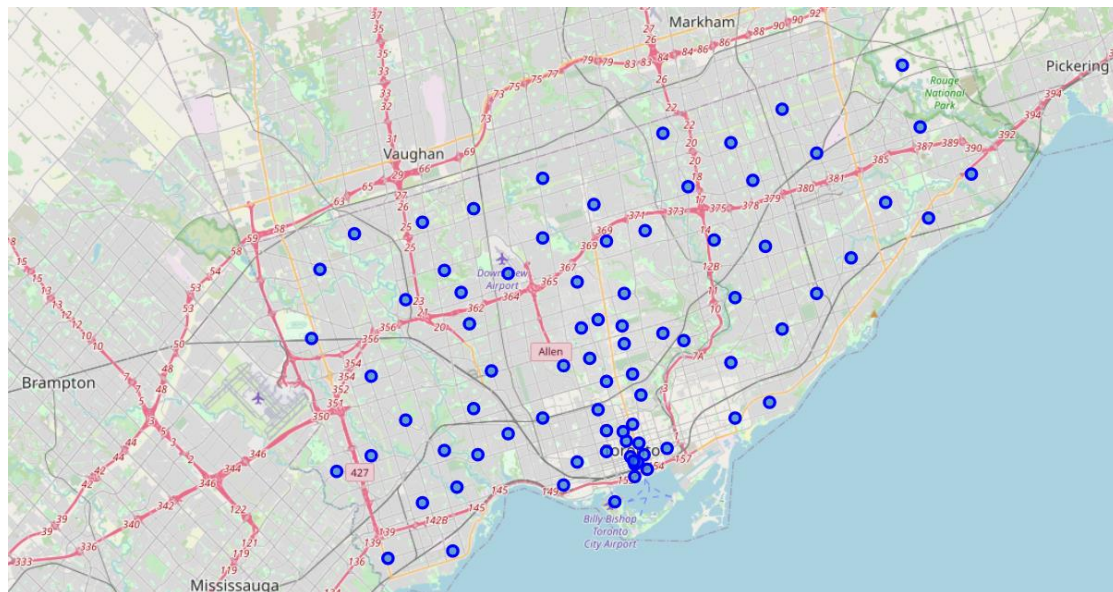


Figure 2. Geospatial representation of neighborhoods.

Among the most common venues are restaurants, pizzerias, coffee shops, and other places of consumption of food and drinks; also, shopping malls, yoga and fitness, as shown in Figure 3. Thus, in general, the types of venues depends upon the information service provided by FourSquare.

[23]:

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Latin American Restaurant	Skating Rink	Breakfast Spot	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
1	Alderwood, Long Branch	Pizza Place	Gym	Sandwich Place	Coffee Shop	Pub	Yoga Studio	Doner Restaurant	Discount Store	Distribution Center	Dog Run
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Gift Shop	Fried Chicken Joint	Shopping Mall	Bridal Shop	Sandwich Place	Diner	Restaurant	Deli / Bodega
3	Bedford Park, Lawrence Manor East	Pizza Place	Sandwich Place	Italian Restaurant	Coffee Shop	Thai Restaurant	Cupcake Shop	Indian Restaurant	Café	Pub	Sushi Restaurant
4	Berzcy Park	Coffee Shop	Cocktail Bar	Farmers Market	Beer Bar	Seafood Restaurant	Bakery	Cheese Shop	Restaurant	Juice Bar	Hotel

[24]: `neighbourhoods_venues_sorted.shape`

[24]: (73, 11)

Figure 3. Types of venues retrieved by FourSquare.

Although, there exist information about parks, monuments, and other points of interest not especially correlated to commercial sites. To this type of venue, the Covid19 Memorial Obelisk is added, with the social and historical significance proceeding from a life-changing event occurring at the end of the first twenty years of the 21st century.

Points of interest that shape the urban landscape in terms of feelings, experiences, and other inclinations, in addition to the daily needs for food, transportation and others.

The integration of a Covid19 Monumental Obelisk into the landscape of a city in every country of the world could become a symbol for the arrival of new times of cooperation in science and technology for life, to reduce inequalities and to preserve the environment.

RESULTS

Initially, 5 centroids were established for clustering. The method of K-means clustering was used (GARBADE, 2018; IBM, 2020). A partial representation of the content of the fifth clustering is shown in Figure 4.

[34]: `toronto_merged.loc[toronto_merged['Cluster_Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]`

[34]:

	Borough	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Scarborough	4	Construction & Landscaping	Bar	Yoga Studio	Discount Store	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
2	Scarborough	4	Intersection	Restaurant	Breakfast Spot	Rental Car Location	Electronics Store	Medical Center	Bank	Doner Restaurant	Distribution Center	Dog Run
3	Scarborough	4	Coffee Shop	Korean BBQ Restaurant	Yoga Studio	Discount Store	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
5	Scarborough	4	Playground	Convenience Store	Yoga Studio	Diner	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant
6	Scarborough	4	Discount Store	Bus Station	Convenience Store	Department Store	Coffee Shop	Hobby Shop	Chinese Restaurant	Drugstore	Dog Run	Doner Restaurant
11	Scarborough	4	Middle Eastern Restaurant	Auto Garage	Sandwich Place	Bakery	Shopping Mall	Vietnamese Restaurant	Distribution Center	Dog Run	Doner Restaurant	Drugstore
12	Scarborough	4	Lounge	Latin American Restaurant	Skating Rink	Breakfast Spot	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
13	Scarborough	4	Pizza Place	Fast Food Restaurant	Italian Restaurant	Fried Chicken Joint	Noodle House	Convenience Store	Bank	Gas Station	Chinese Restaurant	Thai Restaurant
15	Scarborough	4	Fast Food Restaurant	Pharmacy	Coffee Shop	Sandwich Place	Bank	Furniture / Home Store	Chinese Restaurant	Indian Restaurant	Breakfast Spot	Pizza Place
17	North York	4	Fast Food Restaurant	Golf Course	Dog Run	Pool	Mediterranean Restaurant	Escape Room	Electronics Store	Ethiopian Restaurant	Dessert Shop	Eastern European Restaurant
18	North York	4	Clothing Store	Coffee Shop	Fast Food Restaurant	Restaurant	Shoe Store	Juice Bar	Bank	Japanese Restaurant	Cosmetics Shop	Convenience Store
22	North York	4	Ramen Restaurant	Sandwich Place	Restaurant	Café	Pizza Place	Coffee Shop	Sushi Restaurant	Shopping Mall	Japanese Restaurant	Lounge
24	North York	4	Pharmacy	Coffee Shop	Discount Store	Pizza Place	Grocery Store	Airport Lounge	Colombian Restaurant	Farmers Market	Falafel Restaurant	Event Space
28	North York	4	Coffee Shop	Bank	Gift Shop	Fried Chicken Joint	Shopping Mall	Bridal Shop	Sandwich Place	Diner	Restaurant	Deli / Bodega
29	North York	4	Furniture / Home Store	Bar	Massage Studio	Caribbean Restaurant	Coffee Shop	Metro Station	Drugstore	Dog Run	Doner Restaurant	Donut Shop
30	North York	4	Grocery Store	Park	Athletics & Sports	Shopping Mall	Food Truck	Baseball Field	Bank	Discount Store	Liquor Store	Airport
31	North York	4	Grocery Store	Park	Athletics & Sports	Shopping Mall	Food Truck	Baseball Field	Bank	Discount Store	Liquor Store	Airport
32	North York	4	Grocery Store	Park	Athletics & Sports	Shopping Mall	Food Truck	Baseball Field	Bank	Discount Store	Liquor Store	Airport
33	North York	4	Grocery Store	Park	Athletics & Sports	Shopping Mall	Food Truck	Baseball Field	Bank	Discount Store	Liquor Store	Airport
34	North York	4	Hockey Arena	Coffee Shop	Portuguese Restaurant	Intersection	French Restaurant	Pizza Place	Electronics Store	Escape Room	Eastern European Restaurant	Drugstore
36	East York	4	Intersection	Park	Skating Rink	Spa	Beer Store	Athletics & Sports	Bus Stop	Curling Ice	Donut Shop	Distribution Center
37	East Toronto	4	Neighborhood	Trail	Health Food Store	Coffee Shop	Asian Restaurant	Pub	Discount Store	Distribution Center	Dog Run	Doner Restaurant
38	East York	4	Sporting Goods Shop	Coffee Shop	Bank	Burger Joint	Furniture / Home Store	Liquor Store	Sandwich Place	Supermarket	Sports Bar	Fish & Chips Shop

Figure 4. Partial content of the 5th cluster.

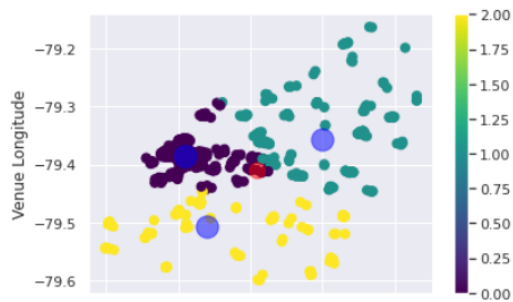
It was also considered the choose of the number of clustering based on the Elbow curve, resulting in the optimal number of centroids equal to 3, as observed in Figure 5.

The clustering process was performed with 3 centroids and the result is shown in Figure 6, where the obtained clusters, coordinates of the centroids and geographic distribution of the related venues appear.



Figure 5. Elbow curve determining the number of optimal centroids equals 3.

```
[40]: #plt.figure(figsize=(10,8)) viridis
X.plot.scatter(x = 'Venue Latitude', y = 'Venue Longitude', c=labels, s=50, cmap='viridis')
plt.scatter(centers[:, 0], centers[:, 1], c='blue', s=300, alpha=0.5)
plt.scatter(x = 43.7043, y = -79.4094, c='red', s=150, alpha=0.5)
plt.show()
```



Centroids in blue color, venues grouped in yellow, green, and violet color, Covid19 Monumental Obelisk in red color

```
[41]: centers = kmeans.cluster_centers_
centers

[41]: array([[ 43.65510346, -79.38492301],
 [ 43.74987769, -79.35691288],
 [ 43.66969076, -79.5068592 ]])
```

Figure 6. Segmentation and clustering with 3 centroids. Covid19 Monumental Obelisk highlighted in red color.

Some of the venues in each cluster appear in Figure 7, where the presence of the Covid19 Memorial Obelisk, in Toronto, is highlighted.

```
[47]: label0.tail()
```

```
[47]:
```

	Venue	Venue Latitude	Venue Longitude	cluster_label
1675	SUDS	43.659880	-79.394712	0
1676	Convocation Hall	43.660828	-79.395245	0
1677	Tim Hortons	43.658906	-79.388696	0
1678	Understudy Café at Gerstein	43.662308	-79.394098	0
1747	Covid19 Memorial Obelisk	43.704312	-79.409354	0

```
[48]: label1.head()
```

```
[48]:
```

	Venue	Venue Latitude	Venue Longitude	cluster_label
0	Wendy's	43.807448	-79.199056	1
1	Chris Effects Painting	43.784343	-79.163742	1
2	Royal Canadian Legion	43.782533	-79.163085	1
3	RBC Royal Bank	43.766790	-79.191151	1
4	G & G Electronics	43.765309	-79.191537	1

```
[49]: label2.head()
```

```
[49]:
```

	Venue	Venue Latitude	Venue Longitude	cluster_label
203	MUSE Massage Spa	43.765686	-79.489318	2
204	Carribean Heat	43.764155	-79.490227	2
205	Tim Hortons	43.764289	-79.488790	2
206	Fox & Fiddle	43.763795	-79.488497	2
207	Bad Boy Furniture - North York	43.764314	-79.486588	2

Only an idea!

Localization of the Toronto Covid19 Memorial Obelisk

```
[35]: display(toronto_venues.loc[1747])
```

Unnamed: 0	1747
Neighbourhood	Roselawn
Neighbourhood Latitude	43.7043
Neighbourhood Longitude	-79.4094
Venue	Covid19 Memorial Obelisk
Venue Latitude	43.7043
Venue Longitude	-79.4094
Venue Category	Monument
Name: 1747, dtype: object	

Figure 7. The Covid19 Monumental Obelisk is highlighted.

CONCLUSION

In this work It was possible to use data and information on spatial distribution of venues, as well as to use Python language tools and libraries, and to apply the k-means segmentation and clustering procedures to group venues by neighborhoods in Toronto and idealize the construction of a Covid19 Memorial Obelisk in Roselawn (Forest Hill North).

Although the results reflect the fulfillment of the objectives of the work through an idealization, this does not mean that the idea cannot be turned into reality, given the social and historical relevance of the Covid19 pandemic started in 2019 and still ongoing in the present year of 2021. The outline left by this terrible circumstance, leads to a reflection for the union of efforts for the fulfillment of the goals of sustainable development of the United Nations. For that, there exist the necessity of continuation of the international cooperation in science and technology that has been a hallmark of the race to save human lives and produce a vaccine against the disease during the pandemic. The Covid19 Memorial Obelisk can bring together governments, architects, people in general, to build ideas around a better future.

BIBLIOGRAPHY

COURSERA. **Segmenting and Clustering Neighborhoods in New York City**. 2020a. Available at: <https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/DS0701EN/DS0701EN-3-3-2-Neighborhoods-New-York-py-v1.0.ipynb?iti=true>. Accessed: Dec.15,2020.

COURSERA. **Segmenting and Clustering Neighborhoods in Toronto**. 2020b. Available at: <https://www.coursera.org/learn/applied-data-science-capstone/peer/I1bDq/segmenting-and-clustering-neighborhoods-in-toronto>. Accessed: Dec.2,2020.

FOURSQUARE. **FourSquare City Guide**. 2020. Available at: <https://foursquare.com/>. Accessed: Dec.10,2020.

GARBADE, M. J. **Understanding K-means Clustering in Machine Learning**. 2018. Available at: <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>. Accessed: Dic.3,2020.

IBM. **Unsupervised Learning**. 2020. Available at: <https://www.ibm.com/cloud/learn/unsupervised-learning>. Accessed: Dec.1,2020.

KOROS, G. **IBM Data Science Course Capstone Project**. 2020. Available at: <https://medium.com/@gaborkoros/ibm-data-science-course-capstone-project-ae681ac9b4bd>. Accessed: Dec.26,2020.

MAGIYA, J. **Clustering GPS Coordinates and Forming Regions with Python**. 2019. Available at: <https://levelup.gitconnected.com/clustering-gps-co-ordinates-forming-regions-4f50caa7e4a1>. Accessed: Dec.4,2020.

WHO. **WHO Coronavirus Disease (COVID-19) Dashboard**. Geneva, 2021. Available at: <https://covid19.who.int/>. Accessed: Jan.6,2021.

WIKIPEDIA. **List of postal codes of Canada: M**. 2020. Available at: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Accessed: Dec.10,2020.