

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

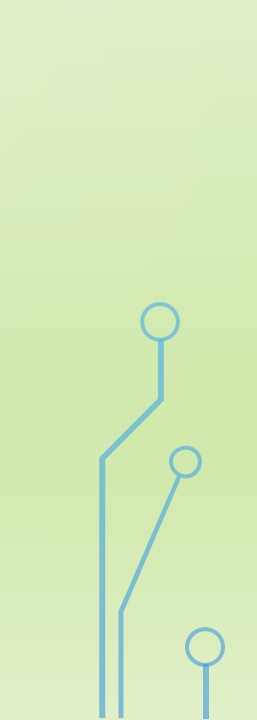
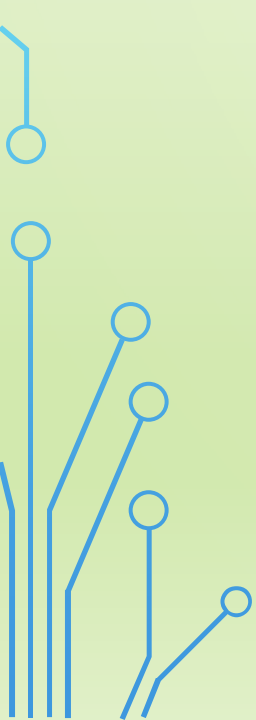
Presentation - The Battle of Neighborhoods (2)

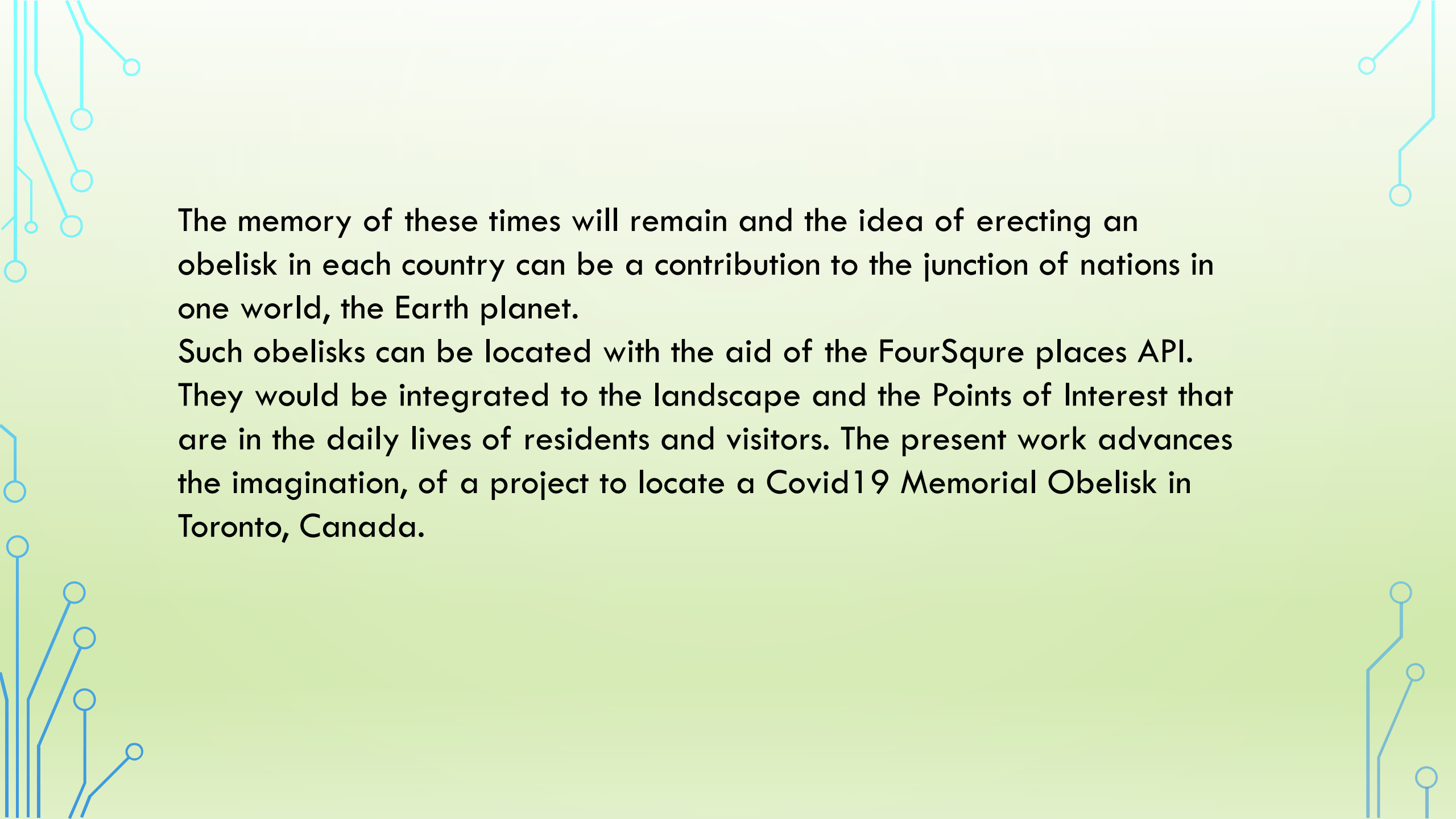


INTRODUCTION

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 FourSquare 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 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).

METHODOLOGY

Primary data and quantitative information on neighborhoods

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).

```
[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

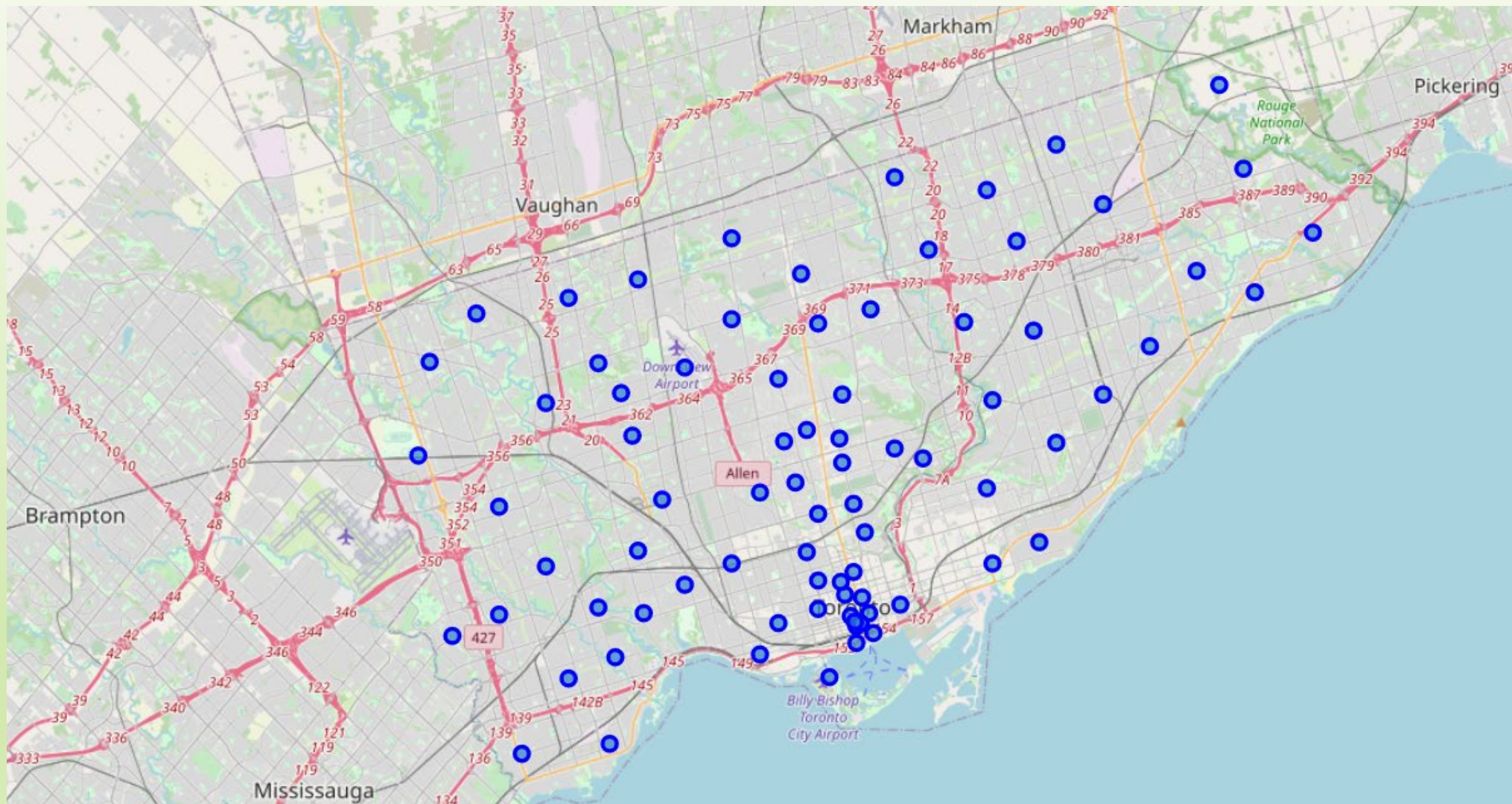
```
[16]: toronto_venues.groupby('Neighbourhood').count()
```

[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

Geospatial representation of neighborhoods

A map of the Greater Toronto Area (GTA) illustrating a geospatial representation of neighborhoods. The map shows a network of red lines, likely representing transit routes, with blue dots indicating specific neighborhood locations. Key areas labeled include Markham, Vaughan, Brampton, Mississauga, and Pickering. The map also shows geographical features like Lake Ontario and Rouge National Park. The blue dots are distributed across the urban areas, with a higher concentration in the central part of the map, particularly around the downtown core and the airport area. The red lines form a complex web connecting various parts of the region, with some lines labeled with numbers like 37, 35, 33, 32, 69, 66, 65, 63, 58, 59, 54, 53, 48, 44, 42, 39, 333, 336, 340, 342, 346, 122, 121, 119, 134, 136, 139, 1428, 145, 149, 157, 154, 15, 13, 12, 10, 7, 5, 3, 2, 350, 351, 352, 354, 356, 201, 21, 23, 25, 26, 27, 29, 31, 33, 35, 37, 63, 65, 69, 73, 75, 77, 79, 83, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000.

Types of venues retrieved 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	Berczy 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)

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 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

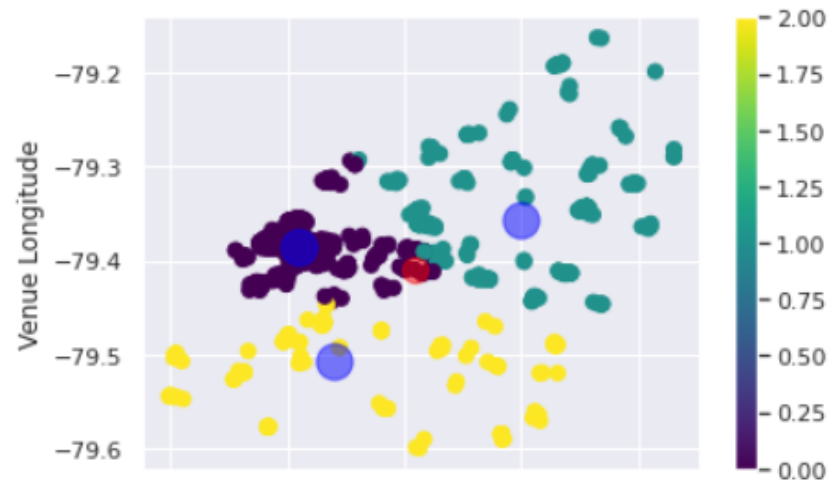
Elbow curve determining the number of optimal centroids equals 3.

```
[38]: K_clusters = range(1,10)
kmeans = [KMeans(n_clusters=i) for i in K_clusters]
Y_axis = df_c[['Venue Latitude']]
X_axis = df_c[['Venue Longitude']]
score = [kmeans[i].fit(Y_axis).score(Y_axis) for i in range(len(kmeans))]
# Visualize
plt.plot(K_clusters, score)
plt.xlabel('Number of Clusters')
plt.ylabel('Score')
plt.title('Elbow Curve')
plt.show()
```



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

```
[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 ]])
```

The Covid19 Monumental Obelisk is highlighted.

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

	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]: label11.head()
```

	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]: label12.head()
```

	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
```

CONCLUSION

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?lti=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.