You can map via NBViewer

Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Taku Sasaki

Table of contents

Introduction: Business Problem

Background

As COVID-19 has spread around the world for more than a year, vaccination is showing signs of convergence. However, there are still many unclear points about COVID-19 infection.

According to statistics, East Asians have a fairly low number of COVID-19 infections. It cannot be concluded that the cause has yet to be determined, whether it is due to genetic characteristics, many people already have immunity, or cultural differences. It is strange that the number of infected people is smaller than in the West in the big city of Tokyo, which has a murderous crowded train commuter.

People infected with COVID-19 vary from region to region. It cannot be said that there are many infected people because of the large population. The risk of COVID-19 infection varies depending on what kind of city you live in or stay in.

Note: For convenience, this document refers to the cities and wards of Tokyo as Borough.

Problem

The problem is to answer the question of what kind of city is the one with many COVID-19 infections.

There are various factors that cause a city with many infected people, such as many overseas travelers, many foreigners, and many bars where clusters are likely to occur. The purpose of this survey is to focus on the number of foreign residents and the number of bars to gain insight into the causal relationship with the number of people infected with COVID-19.

Sample use case:

- As a person planning to move to Tokyo, find out what kind of city has the same characteristics as a city with few infected people.
- As a visitor planning a trip to Tokyo, I would like to find a safe city with the same characteristics as a city with few infected people and enjoy eating and drinking.

It would be greatly appreciated if the causative factors could be clarified by solving this problem.

Target Audience

- Those who want to move to or stay in Tokyo
- Those who want to know which city is relatively safe against infection

---

Data

Data Requirements

Based on definition of our problem, factors that will influence our decision are:

- the number of foreign residents in Tokyo
- the number of COVID-19 test positives in Tokyo
- the number of existing bars in the neighborhood (any type of bar)

Data Collection

Following data sources will be needed to extract/generate the required information:

- The number of Foreign residents in Tokyo will be obtained from the following site:
    - Source: *"2-4 FOREIGN RESIDENTS BY DISTRICT AND NATIONALITY (2019)" in "TOKYO STATISTICAL YEARBOOK"*
        - by *Statistics Division, Bureau of General Affairs, Tokyo Metropolitan Government*
    - The data is as of 2019.
    - The data is aggregated by Borough, which means ward and city in Tokyo.

      Foreign residents here mean foreign nationals who are registered according to the Basic Resident Registration Act.
- The number of COVID-19 test positive in Tokyo will be obtained from the following site:
    - Source:
        - *COVID-19 The information website by Tokyo Metropolitan Government*
        - *Tokyo COVID-19 Task Force website (https://github.com/tokyo-metropolitan-gov)*
    - The data is as of the day before yesterday.

    - The data is aggregated by Borough, which means ward and city in Tokyo.
- The list of Special wards and districts in Tokyo
    - Source:
        - *Wikipedia: Special wards of Tokyo*
- The number of bars and their type and location in every neighborhood will be obtained using **Foursquare API**

Foreign residents in Tokyo

Raw data:

	11	年次\n(平成)	Year	階層\nコード	地域\nコード	地域	district	総数 Total	中国 China	韓国 Rep. of Korea	ベトナム Vietnam	フィリピン Philippines	ネパール Nepal	台湾 Taiwan	アメリカ U.S.A.	インド India	ミャンマー Myanmar	タイ Thailand	その他 Others	Unnamed: 19	Unnamed: 20
0	12	27	2015	0	13000	総数	Tokyo-to	417442	172769	94010	14645	28681	14355	...	16097	8730	5627	7154	55374	NaN	NaN
1	13	28	2016	0	13000	総数	Tokyo-to	449042	185982	93309	22131	29575	18412	...	16411	9475	7044	7370	59333	NaN	NaN
2	14	29	2017	0	13000	総数	Tokyo-to	486346	185883	88755	27762	30761	22660	17281	16939	10354	8249	7651	70051	NaN	NaN
3	15	30	2018	0	13000	総数	Tokyo-to	521500	199949	90438	32334	32089	26157	18568	17578	11153	9719	7958	75557	NaN	NaN
4	16	31	2019	0	13000	総数	Tokyo-to	551683	213767	92418	36227	33219	27290	19726	18508	12130	10395	8101	79902	NaN	NaN

In [15]

Prepared data:

	Municipal_code	Borough	Borough_suffix	FR_total	FR_China	FR_Rep_of_Korea	FR_Vietnam	FR_Philippines	FR_Nepal	FR_Taiwan	FR_USA	FR_India	FR_Myanmar	FR_Thailand	FR_Others
0	13101	Chiyoda	ku	2996	1250	455	76	64	24	194	210	79	16	46	582
1	13102	Chuo	ku	7651	3266	1401	192	153	95	385	391	275	38	89	1366
2	13103	Minato	ku	20057	3962	3461	144	1027	107	756	3257	649	56	191	6447
3	13104	Shinjuku	ku	43068	14153	10221	3484	747	3517	1884	1033	246	2218	735	4830
4	13105	Bunkyo	ku	10808	4646	1658	927	223	364	509	329	107	317	192	1536
5	13106	Taito	ku	15433	6489	3118	824	785	688	503	253	713	134	334	1592
6	13107	Sumida	ku	12645	5874	1966	679	1349	320	401	183	140	75	388	1270
7	13108	Koto	ku	29472	14783	4557	1030	1585	583	703	467	2065	436	366	2897
8	13109	Shinagawa	ku	13042	4317	2426	547	798	712	600	607	402	292	196	2145
9	13110	Meguro	ku	9102	1836	1498	202	537	252	469	979	198	54	165	2912
10	13111	Ota	ku	24199	8467	3562	1494	2497	2287	1033	621	291	275	473	3199
11	13112	Setagaya	ku	21379	5835	4367	864	983	512	999	1706	503	129	260	5221
12	13113	Shibuya	ku	10639	2013	1608	352	334	172	621	1383	177	57	168	3754
13	13114	Nakano	ku	19326	6786	3365	1889	531	1750	970	496	127	471	272	2669
14	13115	Suginami	ku	17722	5837	2854	1486	546	2226	1009	750	92	196	240	2486
15	13116	Toshima	ku	30223	12955	2545	3609	531	3439	1295	428	184	2232	286	2719
16	13117	Kita	ku	22621	10759	2432	2008	863	1342	565	247	194	1038	190	2983
17	13118	Arakawa	ku	19131	7284	5046	2075	528	1171	360	182	101	520	149	1715
18	13119	Itabashi	ku	26759	14177	3218	1657	1545	1148	975	345	125	335	316	2918
19	13120	Nerima	ku	19653	8177	4286	878	1099	690	827	527	120	164	290	2595
20	13121	Adachi	ku	31706	14001	7388	1554	3686	448	620	282	172	151	459	2945
21	13122	Katsushika	ku	21849	11322	3133	1076	1631	897	475	191	87	254	270	2513

	Municipal_code	Borough	Borough_suffix	FR_total	FR_China	FR_Rep_of_Korea	FR_Vietnam	FR_Philippines	FR_Nepal	FR_Taiwan	FR_USA	FR_India	FR_Myanmar	FR_Thailand	FR_Others
22	13123	Edogawa	ku	35710	15424	4390	2580	2812	1197	738	352	4148	391	505	3173
23	13201	Hachioji	shi	12936	4838	1815	1002	1360	592	311	302	153	94	193	2276
24	13202	Tachikawa	shi	4374	1928	754	209	396	179	118	130	47	9	51	553
25	13203	Musashino	shi	3240	1083	557	118	93	154	177	300	40	17	41	660
26	13204	Mitaka	shi	3813	1178	693	194	189	102	217	328	48	38	61	765
27	13205	Ome	shi	1877	363	211	283	416	33	81	63	8	3	55	361
28	13206	Fuchu	shi	5302	1809	814	269	536	97	218	264	54	43	103	1095
29	13207	Akishima	shi	2688	697	500	279	376	251	46	70	66	14	29	360
30	13208	Chofu	shi	4629	1662	1032	232	294	74	218	158	63	55	96	745
31	13209	Machida	shi	6228	2421	936	412	572	111	177	221	75	21	109	1173
32	13210	Koganei	shi	2792	1160	296	143	134	134	87	201	23	27	64	523
33	13211	Kodaira	shi	5204	1758	1065	285	265	116	191	107	36	20	91	1270
34	13212	Hino	shi	3139	1202	441	346	290	103	60	83	22	46	80	466
35	13213	Higashimurayama	shi	2826	1103	422	152	260	135	89	44	14	22	37	548
36	13214	Kokubunji	shi	2365	1030	365	109	111	215	84	74	13	9	18	337
37	13215	Kunitachi	shi	1706	634	337	99	71	85	71	65	30	8	26	280
38	13218	Fussa	shi	3816	740	234	944	408	510	86	108	70	7	110	599
39	13219	Komae	shi	1312	483	187	66	111	90	42	51	20	7	38	217
40	13220	Higashiyamato	shi	1157	404	193	38	245	23	26	23	5	3	25	172
41	13221	Kiyose	shi	1262	435	152	97	205	62	47	32	1	7	26	198
42	13222	Higashikurume	shi	2092	677	316	125	250	48	55	220	47	3	34	317
43	13223	Musashimurayama	shi	1640	591	155	252	351	23	19	33	3	4	27	182
44	13224	Tama	shi	2648	1133	479	147	233	74	60	62	26	38	36	360
45	13225	Inagi	shi	1321	383	267	95	173	19	35	33	13	2	23	278
46	13227	Hamura	shi	1392	209	91	99	292	25	20	51	13	1	21	570
47	13228	Akiruno	shi	839	130	107	181	127	7	22	59	2	1	19	184
48	13229	Nishitokyo	shi	4702	1961	899	260	298	82	263	142	40	39	68	650
49	13303	Mizuho	machi	782	113	50	140	213	4	10	27	3	7	26	189
50	13305	Hinode	machi	82	13	11	5	16	1	3	14	0	0	6	13
51	13307	Hinohara	mura	7	0	3	0	1	0	0	3	0	0	0	0
52	13308	Okutama	machi	44	3	2	4	18	0	0	2	0	0	1	14
53	13361	Oshima	shicho	103	8	13	8	7	0	0	5	0	1	5	56
54	13381	Miyake	shicho	35	0	16	3	13	0	0	0	0	0	0	3
55	13401	Hachijo	shicho	112	4	44	1	40	0	2	5	0	0	2	14
56	13421	Ogasawara	shicho	27	1	6	3	1	0	0	9	0	0	0	7

COVID19 Test positives in Tokyo

Date of covid19 patient.json : 2021/2/27

Raw data:

	code	area	label	ruby	count
0	131016.0	特別区	千代田区	ちよだく	601
1	131024.0	特別区	中央区	ちゅうおうく	1828
2	131032.0	特別区	港区	みなとく	3792
3	131041.0	特別区	新宿区	しんじゅくく	6551
4	131059.0	特別区	文京区	ぶんきょうく	1688

Prepared data:

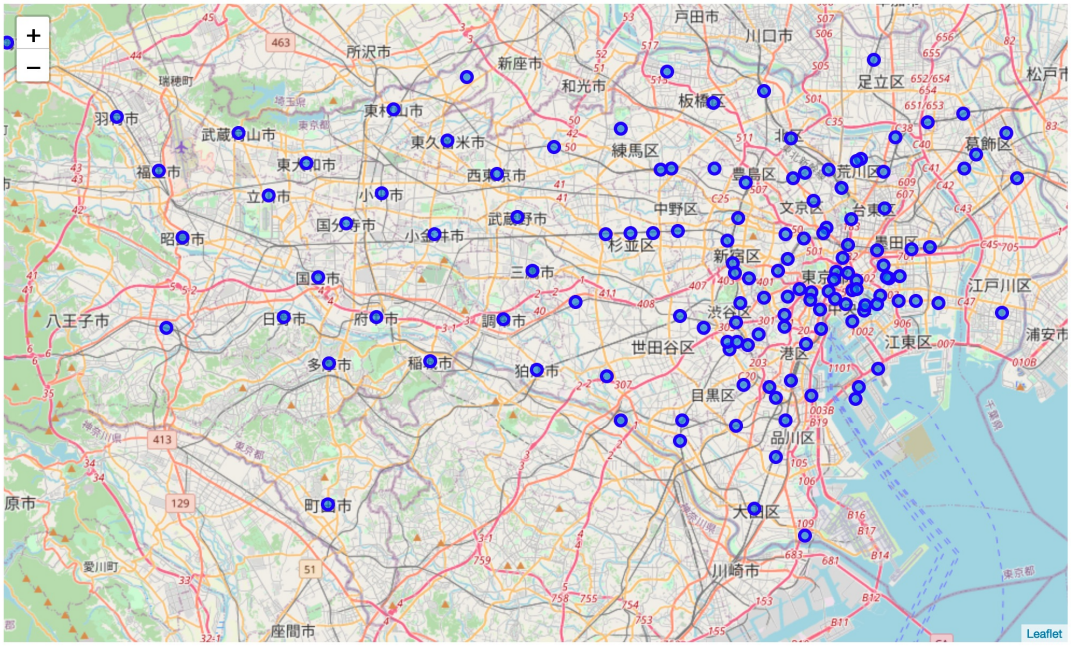
	Municipal_code	Borough_J	COVID_positive
0	13101	千代田区	601
1	13102	中央区	1828
2	13103	港区	3792
3	13104	新宿区	6551
4	13105	文京区	1688
5	13106	台東区	2069
6	13107	墨田区	2150
7	13108	江東区	3575
8	13109	品川区	3251
9	13110	目黒区	3053
10	13111	大田区	5860
11	13112	世田谷区	8188
12	13113	渋谷区	3315
13	13114	中野区	3710
14	13115	杉並区	4442
15	13116	豊島区	2940
16	13117	北区	2448
17	13118	荒川区	1756
18	13119	板橋区	4126
19	13120	練馬区	4828
20	13121	足立区	5225
21	13122	葛飾区	3968
22	13123	江戸川区	4529
23	13201	八王子市	2300
24	13202	立川市	856
25	13203	武蔵野市	883
26	13204	三鷹市	1103
27	13205	青梅市	559
28	13206	府中市	1197
29	13207	昭島市	578
30	13208	調布市	1357
31	13209	町田市	1942
32	13210	小金井市	598
33	13211	小平市	679
34	13212	日野市	731

	Municipal_code	Borough_J	COVID_positive
35	13213	東村山市	529
36	13214	国分寺市	504
37	13215	国立市	280
38	13218	福生市	374
39	13219	狛江市	395
40	13220	東大和市	323
41	13221	清瀬市	264
42	13222	東久留米市	440
43	13223	武蔵村山市	265
44	13224	多摩市	601
45	13225	稲城市	366
46	13227	羽村市	254
47	13228	あきる野市	380
48	13229	西東京市	1072
49	13303	瑞穂町	127
50	13305	日の出町	64
51	13307	檜原村	5
52	13308	奥多摩町	18
53	13361	大島町	21
54	13362	利島村	0
55	13363	新島村	0
56	13364	神津島村	1
57	13381	三宅村	4
58	13382	御蔵島村	1
59	13401	八丈町	7
60	13402	青ヶ島村	0
61	13421	小笠原村	3

Explore Dataset

View neighborhoods in Tokyo

[93]:

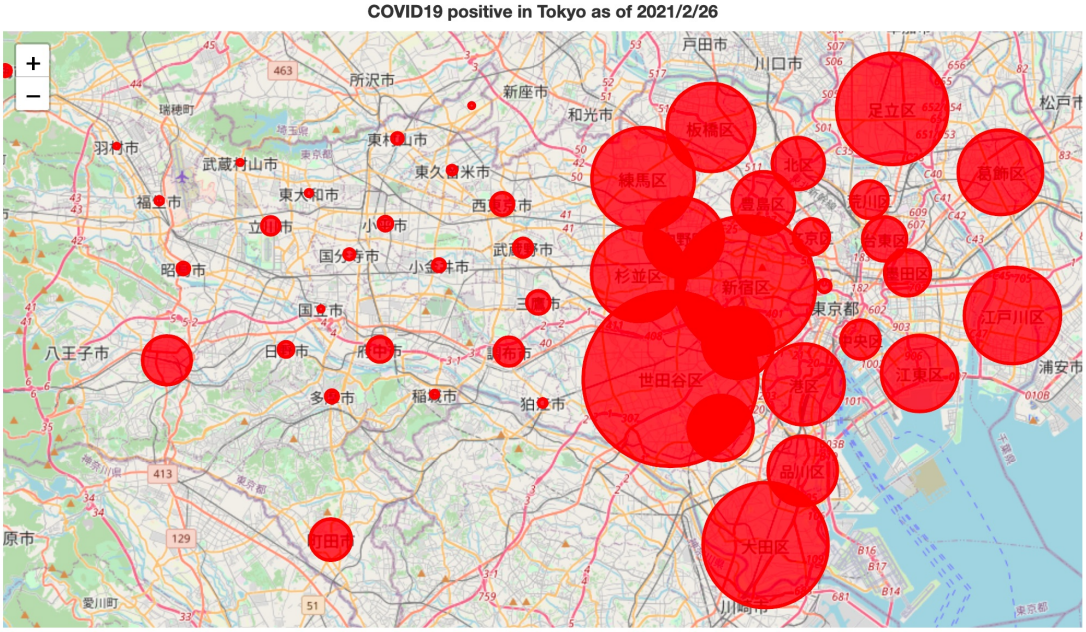


In [33]:

View COVID19 positive in Tokyo

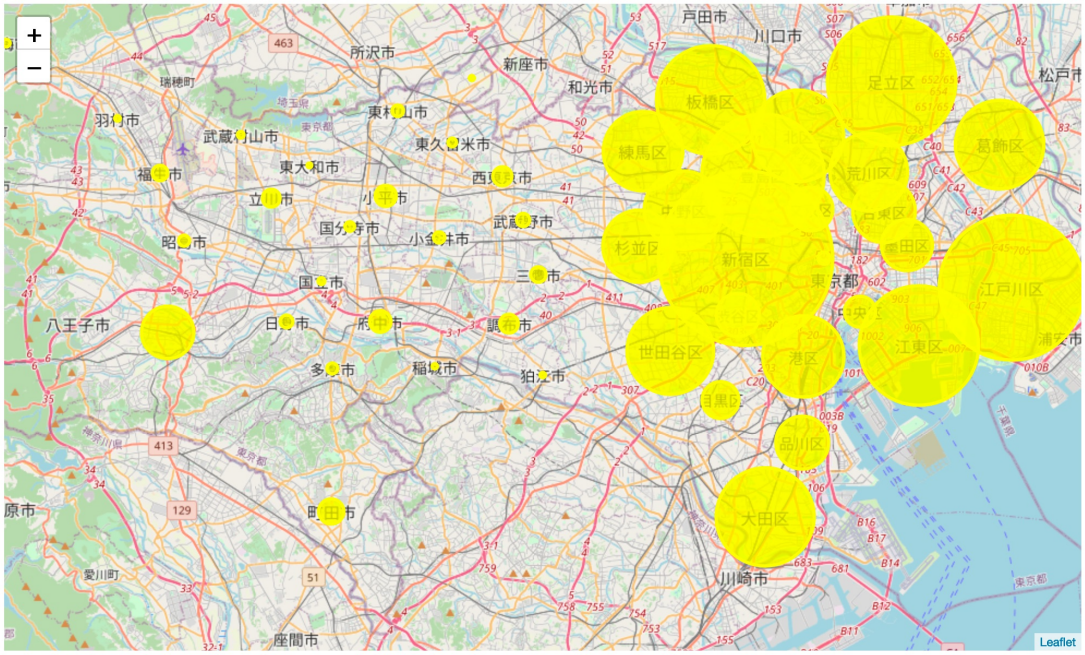
In [35]:

[36]:



View Foreign residents in Tokyo

[37]:



Out[50]:

Methodology

In the previous Data section, the following data frame were created to be investigated:

- `tokyo_venues_bar` : list of bar
    - Neighborhood (key)
    - Borough (key)
    - Neighborhood Latitude
    - Neighborhood Longitude
    - Venue
    - Venue Latitude
    - Venue Longitude
    - Venue Category
- `covid_foreign` : list of Borough with COVID positive and Foreign Residents
    - Municipal_code
    - Borough_J
    - COVID_positive
    - Borough (key)
    - Borough_suffix
    - FR_total
    - FR_China
    - FR_Rep_of_Korea

    - FR_Vietnam
    - FR_Philippines
    - FR_Nepal
    - FR_Taiwan
    - FR_USA
    - FR_India
    - FR_Myanmar
    - FR_Thailand
    - FR_Others
    - Latitude
    - Longitude

Now, let's proceed with the following two major analysis steps:

- analysis neighborhood bars
- analysis COVID positive, Foreign Residents and bars per Borough

1. Analysis Neighborhood bars

The purpose of this step is to investigate what kind of neighborhood belongs to the Borough.

- Use only `tokyo_venues_bar`
- Use **k-means** to cluster the Neighborhoods bar.
- Use KElbowVisualizer to determine the optimal k value
- Visualize the resulting Neighborhood bar clusters
- List each Neighborhood bar per cluster to investigate

2. Analysis COVID positive, Foreign Residents and bars per Borough

The purpose of this step is to investigate what Borough clusters are susceptible to COVID-19 infection.

- Aggregate each type of bar by Borough by using `tokyo_venues_bar`
- Merge it with `covid_foreign`
- Use sklearn.preprocessing.StandardScaler to normalize over the standard deviation
- Use **k-means** to cluster the Borough
- Use KElbowVisualizer to determine the optimal k value
- Visualize the resulting Borough clusters
- List each Borough per cluster to investigate

Analysis

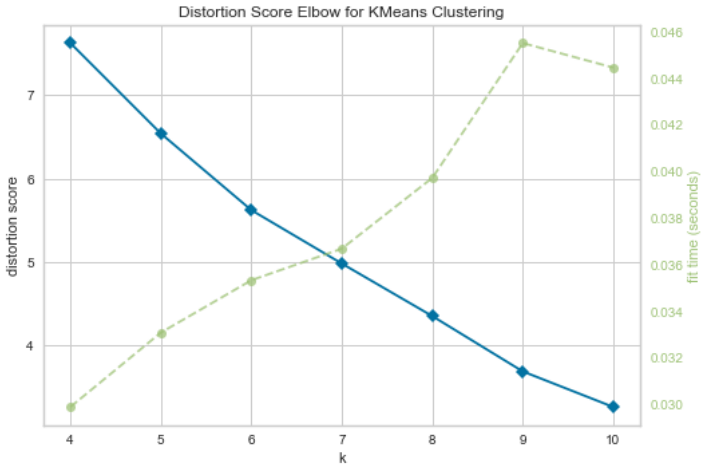
Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of bars in every area candidate**:

Cluster Neighborhoods

To analyze which neighborhood of Tokyo is good to open a new bar, I will use a K-means clustering: a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

So the first step is identify the best “K” using a famous analytical approach: the elbow method.

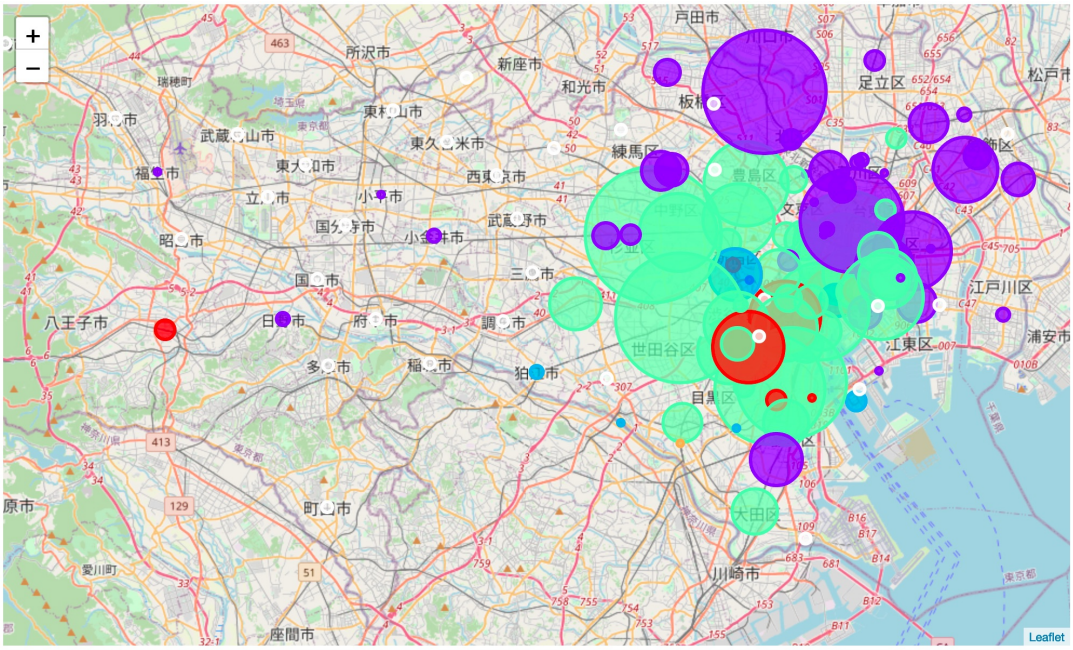
Let's see:



Out[68]:

View Neighborhood clusters

[69]:



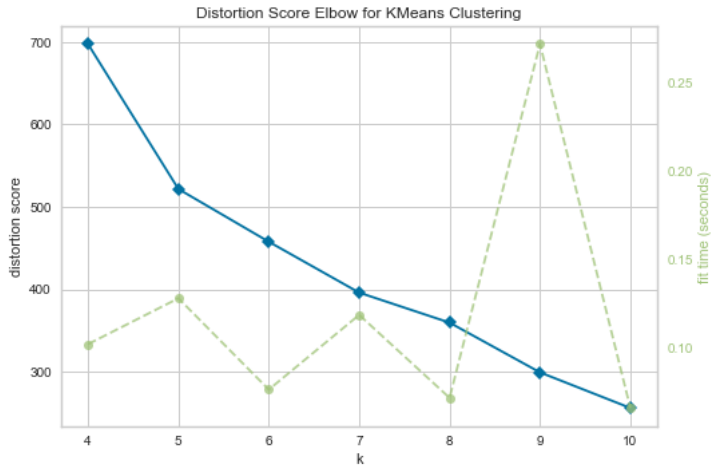
Out[69]:

Cluster COVID positive, foreign residents and bars

To analyze which neighborhood of Tokyo is good to open a new bar, I will use a K-means clustering: a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

So the first step is identify the best “K” using a famous analytical approach: the elbow method.

Let's see:



Out[83]:

From the plot up here, the best k value seems to be 6.

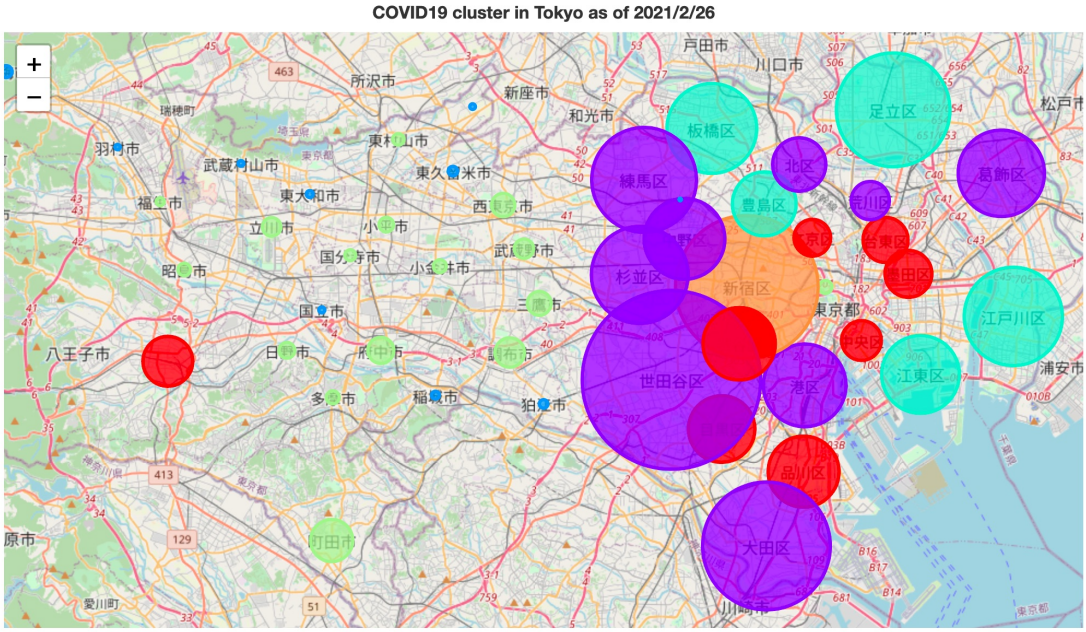
Run *k*-means to cluster the neighborhood into 6 clusters.

Out[85]:

View COVID Clusters

In [86]:

[86]:



Results and Discussion

1. Analysis Neighborhood bars

Neighborhood cluster view is a cluster created only by the neighborhood bar type. It is possible to grasp the atmosphere of the city by the type of bar.

- Cluster 1:
  - Neighborhood has many expensive bars, and the number of bars is 11 or less.
- Cluster 2:
  - There are many sake bars in the neighborhood, which is often found in downtown, and the number of bars is 19 or less.
- Cluster 3:
  - Neighborhood has many expensive bars, and the number of bars is 8 or less.
- Cluster 4:
  - There are few neighborhoods in downtown and many in the city center, and there are many neighborhoods and bars.
- Cluster 5:
  - Luxury residential area

2. Analysis COVID positive, Foreign Residents and bars per

Cluster	COVID_positive mean	FR_total mean	Bar_total mean
1	2451.25	11532.00	23.50
2	4321.88	20659.66	22.55
3	208.44	877.22	0.38
4	4065.00	30774.00	15.60
5	846.87	3797.62	3.81
6	6534.00	43068.00	40.00

- Cluster 1:
  - The number of infected people is relatively small in the city center or in some exceptional suburbs.
- Cluster 2:
  - It is a cluster with Setagaya Ward, which has the highest number of infected people, and there are many bars and foreign residents.
- Cluster 3:
  - Suburban clusters with the fewest bars and foreign residents and the least infected
- Cluster 4:
  - Boroughs with many foreign residents, bars and many infected people in downtown
- Cluster 5:
  - Suburban clusters with few bars and foreign residents and few infected
- Cluster 6:
  - Borough only in Shinjuku Ward, there are many bars, the number of foreign residents is the largest, and the second most infected.

Conclusion

The problem was to answer the question of what kind of city is a city with many COVID-19 infected people. It could be illustrated by clustering the features of Neigborhood according to the type of bar. Furthermore, Borough could be clustered and illustrated according to the number of foreign residents, the number of bars, and the number of COVID-19 infected persons. It can be speculated that Borough, which belongs to a cluster with a large number of infected people, may be susceptible to infection.

Future research

The following can be considered as the continuation of future research.

- The number of foreign residents is from 2019, so it cannot be said that it is valid data. If available, you should investigate with the latest data.
- It is natural that there are few infected people in areas with a small population and areas with a low population density. The analysis should also take into account the population and population density.
- In addition to foreign residents, the movement of immigrants and people from overseas should be considered as statistical values.