

Covid-19 cases & Venues Data Analysis of Stockholm

Sabrina El Mouhib

August 15, 2020

Introduction

Covid-19 outbreak sparked a global emergency. This deadly epidemic that originated in Wohan, China find its way to reach every country in the world. After WHO declared the outbreak a global health emergency, most countries rushed in imposing quarantines and lockdown in order to control the spread the virus. Sweden, was one of the few countries that chose to follow another approach. In fact, Sweden's main goal was to lower the rate of infection so that the need of hospital beds does not overwhelm. The only difference is that Sweden relied on voluntary compliance of its citizens instead of enforcing strict laws that will limit the movement of essential workers. As a result, café and restaurants in Stockholm stayed open.

Business Problem

The goal of this report is to analyze the total number of Covid-19 cases in the 14 districts of Stockholm and study the correlation between the number of café in each district and the number of covid-19 cases. Was the Swedish government right by allowing Café in Stockholm to remain open? This question will be answered at the end of this report

Data Description

Data used in this project came from different sources:

- Coordinate of Stockholm districts were found on StockholmCity website that were later used to create a choropleth map of Stockholm.[1]
- Foursquare API was used to get Café in Stockholm districts.
- Google Maps was used to get the center coordinates of each districts in Stockholm.
- Covid-19 cases per district were downloaded from the official Stockholm City website responsible for publishing covid-19 news on a weekly basis. [2]. It is important to note that these data were collected up to the 31 week of 2020. Any cases after this date were not included in this report.

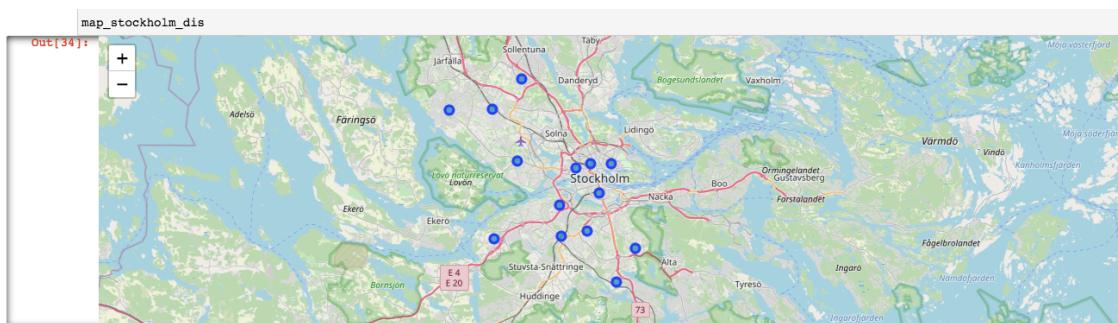
Methodology

First, the data were collected and transformed into a dataframe as shown below. The dataframe included name of the Stockholm district, number of covid-19 cases, number of covid-19 cases by 10000 people, latitude and longitude of the centers of the districts.

Out[28]:

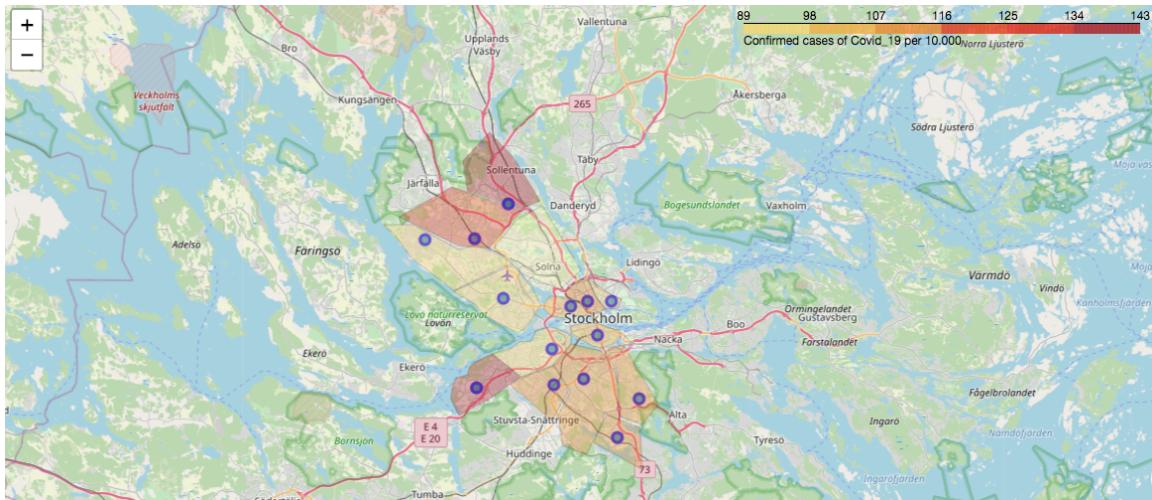
	Districts	Confirmed Cases Per 10.000	Latitude	Longitude	Total Confirmed Cases
0	Alvsjo	110	59.277860	18.008880	347
1	Bromma	93	59.338889	17.938889	753
2	Enskede-Arsta-Vantor	115	59.281755	18.051066	1183
3	Farsta	115	59.240278	18.097222	687
4	Hagersten-Liljeholmen	93	59.303268	18.005921	857
5	Hasselby-Vallingby	97	59.379777	17.830623	739
6	Kungsholmen	103	59.333016	18.033094	734
7	Norrmal	110	59.336997	18.056697	806
8	Ostermal	90	59.336701	18.088621	708
9	Rinkeby-Kista	142	59.405409	17.945848	726
10	Skarholmen	140	59.275407	17.902173	530
11	Skarpnack	113	59.268037	18.128015	525
12	Sodermalm	101	59.312879	18.069671	1318
13	Spanga-Tensta	121	59.381009	17.899786	472

Python folium library was used to visualize geographic details of Stockholm and its districts and I created a map of Stockholm with districts centers superimposed on top. I used latitude and longitude values to get the visual as below:

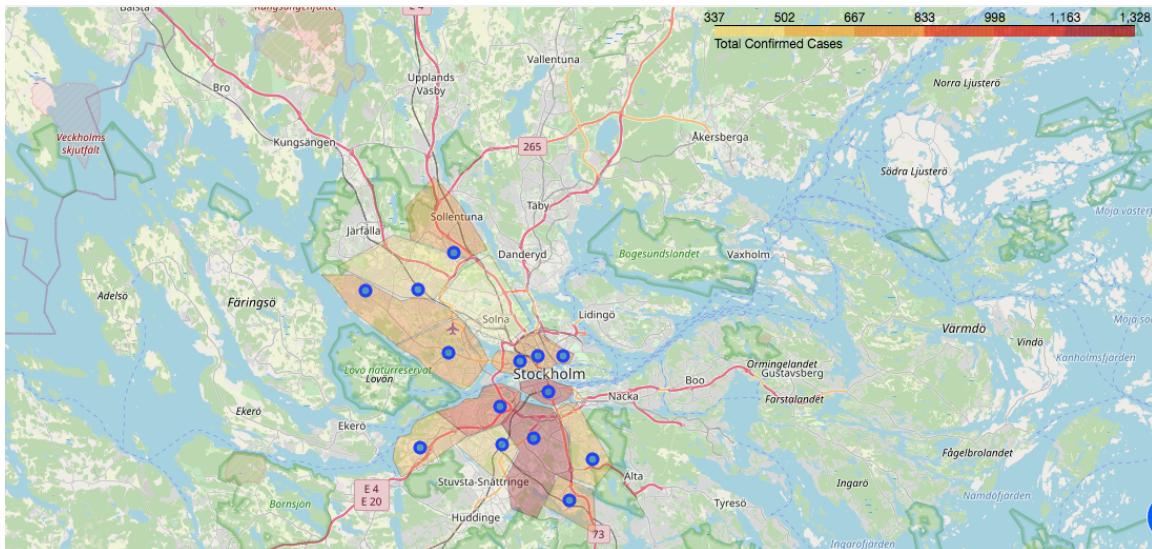


Results:

Then, choropleth map was created to visualize regions with the highest, medium and lowest confirmed cases of Covid 19 per 10.000:



Similarly, choropleth map was created to visualize regions with the highest, medium and lowest total confirmed cases of Covid 19:



Using Foursquare, the number of Café within each district was retrieved using a radius of 1000m and limit of 50 Café. The results were as follow:

===== Alvsjo =====

1 cafe were returned by Foursquare in the district Alvsjo

===== Bromma =====

7 cafe were returned by Foursquare in the district Bromma

===== Enskede-Arsta-Vantor =====

3 cafe were returned by Foursquare in the district Enskede-Arsta-Vantor

===== Farsta =====

9 cafe were returned by Foursquare in the district Farsta

===== Hagersten-Liljeholmen =====

17 cafe were returned by Foursquare in the district Hagersten-Liljeholmen

===== Hasselby-Vallingby =====

2 cafe were returned by Foursquare in the district Hasselby-Vallingby

===== Kungsholmen =====

50 cafe were returned by Foursquare in the district Kungsholmen

===== Norrmalm =====

50 cafe were returned by Foursquare in the district Norrmalm

===== Ostermalm =====

50 cafe were returned by Foursquare in the district Ostermalm

===== Rinkeby-Kista =====

28 cafe were returned by Foursquare in the district Rinkeby-Kista

===== Skarholmen =====

5 cafe were returned by Foursquare in the district Skarholmen

===== Skarpnack =====

8 cafe were returned by Foursquare in the district Skarpnack

===== Sodermalm =====

50 cafe were returned by Foursquare in the district Sodermalm

===== Spanga-Tensta =====

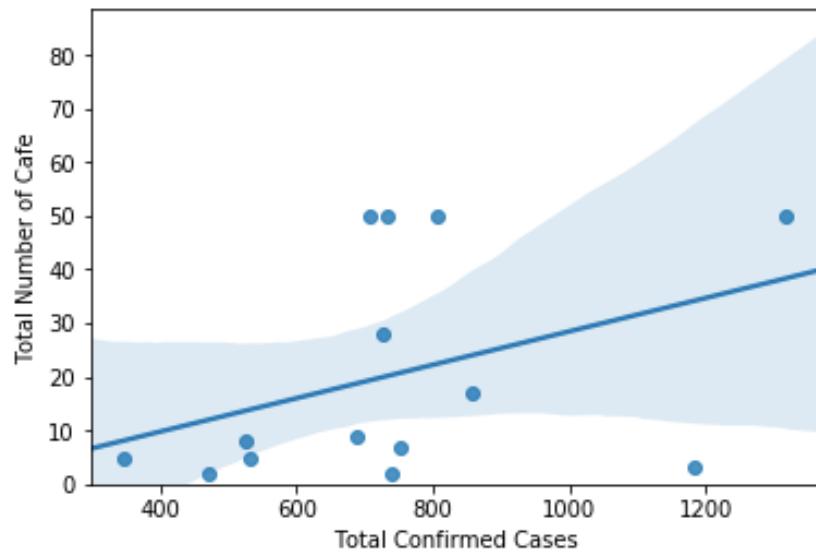
2 cafe were returned by Foursquare in the district Spanga-Tensta

The number of cafe was transformed into a data frame and merged with the initial dataframe. The result is as follow:

Out[125]:

	Districts	Total Number of Cafe	Confirmed Cases Per 10.000	Latitude	Longitude	Total Confirmed Cases
0	Alvsjo	5	110	59.277860	18.008880	347
1	Bromma	7	93	59.338889	17.938889	753
2	Enskede-Arsta-Vantör	3	115	59.281755	18.051066	1183
3	Farsta	9	115	59.240278	18.097222	687
4	Hägersten-Liljeholmen	17	93	59.303268	18.005921	857
5	Hässelby-Vällingby	2	97	59.379777	17.830623	739
6	Kungsholmen	50	103	59.333016	18.033094	734
7	Norrmalm	50	110	59.336997	18.056697	806
8	Ostermalm	50	90	59.336701	18.088621	708
9	Rinkeby-Kista	28	142	59.405409	17.945848	726
10	Skarholmen	5	140	59.275407	17.902173	530
11	Skarpnack	8	113	59.268037	18.128015	525
12	Södermalm	50	101	59.312879	18.069671	1318
13	Spånga-Tensta	2	121	59.381009	17.899786	472

To answer the problematic of the project, it was necessary to conduct a correlation study between the number of cafes and the total number of covid-19 cases. To do that, a scatter plot was performed. In addition, P-value and The Pearson Correlation Coefficient were calculated to study the correlation. The results were as follow:

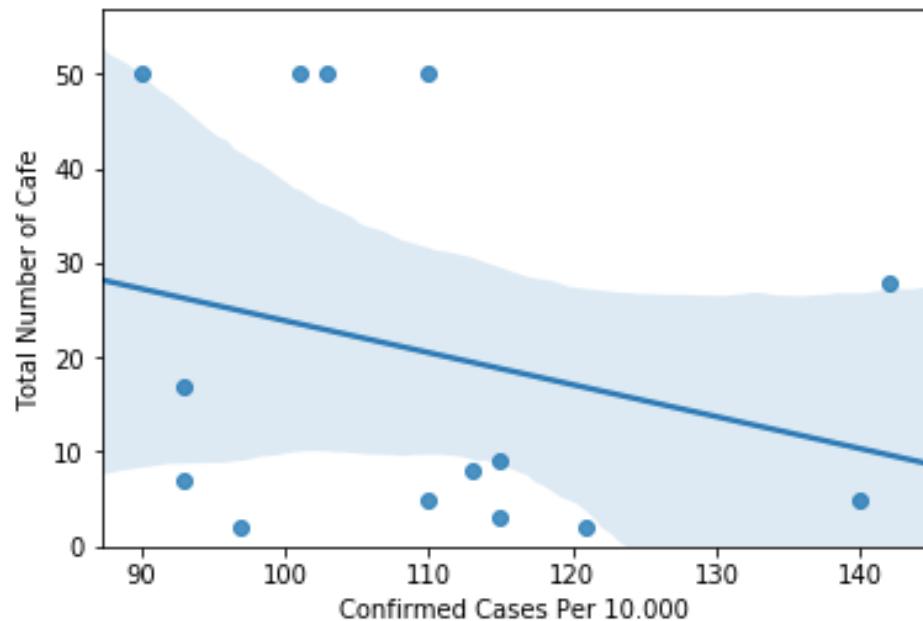


Out[138]:

	Total Number of Cafe	Total Confirmed Cases
Total Number of Cafe	1.000000	0.391628
Total Confirmed Cases	0.391628	1.000000

The Pearson Correlation Coefficient is 0.39162828705926034 with a P-value of P = 0.16611689213047273. It was concluded that number of café does not seem as good predictor of the total number of Covid-19 cases. Also, the data points are very scattered and far from the fitted line, showing lots of variability. With a high P-value and low Pearson Correlation Coefficient, number of café is not good determinant of the total number of Covid-19 cases.

Similarly, the same process was repeated to compare the number of café and Confirmed-cases of Covid-19 per 10000.

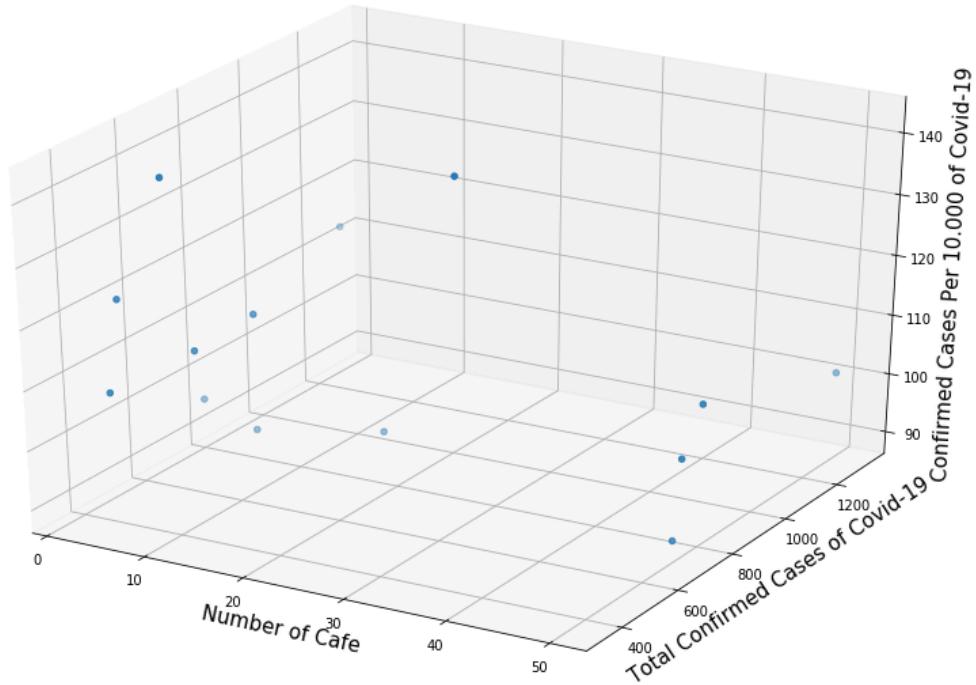


Out[128]:

	Total Number of Cafe	Confirmed Cases Per 10.000
Total Number of Cafe	1.000000	-0.265357
Confirmed Cases Per 10.000	-0.265357	1.000000

The Pearson Correlation Coefficient is -0.26535716209863547 with a P-value of 0.35919654319129873. Consequently, the number of café is not a good predictor of the confirmed cases of Covid-19 per 10.000. To conclude, the Swedish government was right by letting the cafes open, as there is no prove of a strong correlation between the number of cases and the number of café in a district. Then, the high number of cases is the cause of other aspects, which is interesting to explore in future work.

In last part of this report, Kmeans algorithm was used to group the latest generated dataframe into 3 clusters. The analyzed data were: Number of Café, Total number of Cases, Total Number of cases per 10.000 shown in the following graph:

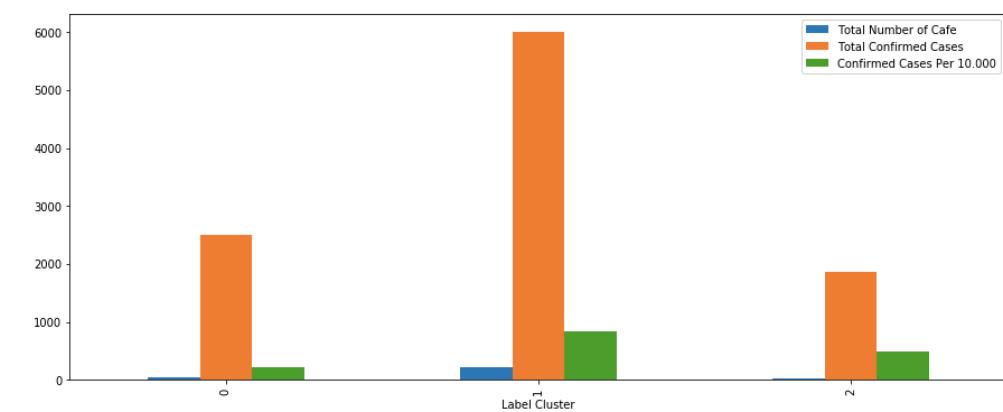


Using the Kmeans algorithm, a new dataframe was generated included the cluster label shown as follow:

Out[148]:

Label Cluster	Districts	Total Number of Cafe	Total Confirmed Cases	Confirmed Cases Per 10.000	Latitude	Longitude
0	2	Alvsjo	5	347	110	59.277860 18.008880
1	1	Bromma	7	753	93	59.338889 17.938889
2	0	Enskede-Arsta-Vantör	3	1183	115	59.281755 18.051066
3	1	Farsta	9	687	115	59.240278 18.097222
4	1	Hägersten-Liljeholmen	17	857	93	59.303268 18.005921
5	1	Hässelby-Vällingby	2	739	97	59.379777 17.830623
6	1	Kungsholmen	50	734	103	59.333016 18.033094
7	1	Norrmalms distriktsförening	50	806	110	59.336997 18.056697
8	1	Ostermalm	50	708	90	59.336701 18.088621
9	1	Rinkeby-Kista	28	726	142	59.405409 17.945848
10	2	Skarholmen	5	530	140	59.275407 17.902173
11	2	Skarpnäck	8	525	113	59.268037 18.128015
12	0	Södermalm	50	1318	101	59.312879 18.069671
13	2	Spånga-Tensta	2	472	121	59.381009 17.899786

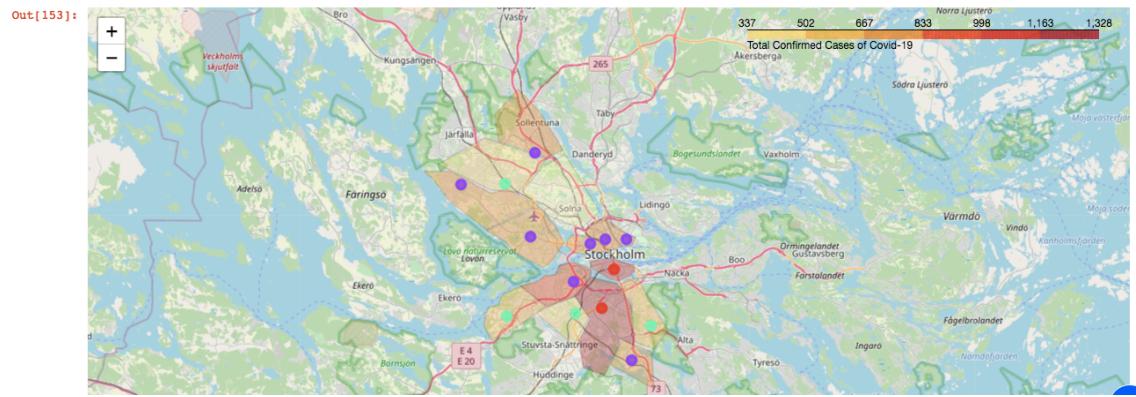
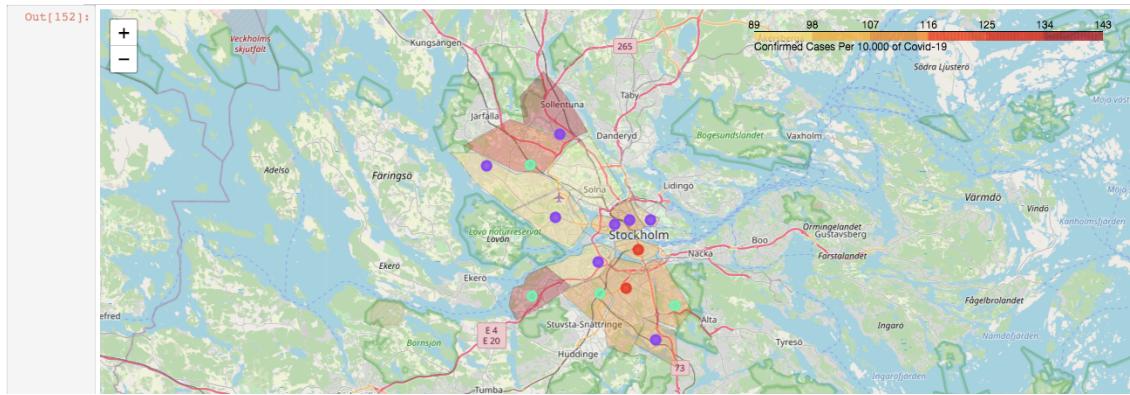
To understand the clusters, a graph was generated:



Therefore, we can name the clusters as follow:

1. Cluster 0 = Medium number of Café shops
2. Cluster 1 = High number of Café Shops
3. Cluster 2 = Low number of Café Shops

Now that the districts were grouped into 3 groups, maps visualizing the clusters and the Covid-19 cases were generated:



Discussion:

From the maps above, we can see that Enskede-Årsta-Vantör and Södermalm are the two districts with the highest number of Covid-19 cases and also high number of Café shops available in the district. However, this correlation is not true in all cases. A counter example is Hasselby-Vallingby, in north of Stockholm, in which the number of café shops is high but low confirmed cases of Covid-19. It is also important to mention that Total number of covid-19 differ from covid-19 cases per 10000. For example, Södermalm has high total number of Covid-19 cases but medium number of cases per 10000. Despite the high number of café shops in that district, the cases per 10,000 are still low compared to other districts such as Rinkeby-Kista. Spanga-Tensta district has one of the highest cases per 10000, but also has the lowest number of café shops. Therefore, we can conclude that there is no correlation between the number of café shops open and the number of positive covid-19 cases.

Conclusion:

The Swedish government made a good choice by remaining café shops open as there is no proof of correlation between number of café shops open and the number of covid-19 cases.

Reference:

[1]: Start. (1970, August 14). Retrieved August 15, 2020, from <https://www.sll.se/>

[2]: (n.d.). Retrieved August 15, 2020, from
<https://foursquare.com/developers/apps/>

[3]: Ny rapportering av antal smittade med covid-19 per kommun och stadsdel.
(n.d.). Retrieved August 15, 2020, from <https://www.sll.se/verksamhet/halsa-och-vard/nyheter-halsa-och-vard/2020/06/ny-rapportering-av-antal-smittade-med-covid-19-per-kommun-och-stadsdel/>