General View of Covid-19 Cases and Covid-19 Clusters in Hong Kong

Jerry Lim

2020 Aug 31

Introduction:

Coronavirus disease 2019 (**COVID-19**) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. The outbreak was first detected in Wuhan, Hubei, China and it is widely spread to the world. The virus is infectious that the World Health Organization declared COVID-19 as a pandemic on 2020-03-11 and currently there are more than 21million confirmed cases worldwide. In Hong Kong, the third outbreak of Covid-19 started on early July and it has reached more than hundred cases daily during this wave.

For government bodies such as Department of Health Hong Kong, it is particular important for them to understand the latest situation of the cases and the outbreak situation for taking precautionary measures to prevent further outbreak. On the other hand, depends on the covid outbreak, Law enforcement agencies can have a better idea of whether the law needs to be tighten or whether it is good enough to reopen the services in Hong Kong.

The flow of this report analysis will separate into two parts. In part 1, a general analysis is performed to look at the big picture of the covid-19 cases in Hong Kong, including the number of new cases study and age distribution study. In part 2, different clustering methods are performed to find out the Covid-19 clusters in Hong Kong, and different results are shown and compared. Part 1 analysis will provide a general idea of the cases distribution in Hong Kong based on different categories, which is useful for determining high risk population. Part 2 analysis identifies the cluster outbreak for Department of health agencies to make immediate action to clean the infected buildings and provide testing for the residents involved.

Datasource:

The dataset are publicly available from government website. The first dataset is the details of propable/confirmed cases of COVID-19 in Hong Kong [2]. this dataset provides the general information of each confirmed or probable cases. The second dataset is the residential buildings in which probable/confirmed cases have resided in the past 14 days or non-residential building with 2 or more probable/confirmed cases in the past 14 days [3]. This dataset provides the complete building list or locations where the confirmed cases resided or gathered before. Both datasets are updated daily and the second dataset only remains the building where the confirmed Covid cases are in the past 14 days.

Data Cleaning:

The datasets are downloaded as a csv file. Since the dataset is updated daily, the first dataset data was gathered up to date 2020-09-04. For second dataset, the dataset was obtained back on 19/8/2020. The cases on the second dataset are cases between 2020-08-04 to 2020-08-19. The raw data need to be modified in order to make further analysis. For example, in the second dataset, the building name column has multiple data format, including address(12 Po Peng Street), building name(Chung Kin Building), shop name inside the building name(Fu Tai Shopping Centre Bao Dim Sin Seng) and building name with remarks(Wayson Commercial Building (non-residential)). In order to use foursquare api to find the corresponding building coordinates, it can only return coordinates with building name/address value that are not too specific and without remarks.

Hence, the second dataset is first modified by removing the remarks “(non-residential)”. Then, the building coordinate is searched using foursquare api. For address that are too specific, for example ‘Block 1, Lotus Tower, Kwun Tong Garden Estate’, foursquare api cannot locate this address. Instead the last string (Kwun Tong Garden Estate) was used as reference for the coordinate value. If the address/building still could not be found, a Null value is given for the building. A generalized dataset is shown in table 1.

Building name	Related probable/confirmed cases	count	lat_col	long_col
Chun Yu House, Sam Shing Estate	4455	1	22.381214	113.978212
Chung Kam House, Tin Chun Court	4171	1	37.300334	99.021896
Chung Kam House, Tin Chung Court	3744, 3905	2	22.460450	114.000154
Chung Kin Building	4159	1	22.322871	114.160516
Chung Lau House, Tin Chung Court	3793, 3934	2	22.460450	114.000154
...	...	...	...	...
Indi home Green River Restaurant	4413	1	NaN	NaN
Indi home Maxim's MX	4413	1	NaN	NaN
Island Lodge	4283	1	22.292321	114.201738
Java Road Municipal Services Building	2299, 2468	2	22.292333	114.199354
K-point	4412	1	50.462775	30.504672

Table 1: Dataset 2 including coordinates(lat_col,long_col) obtained by foursquare api

As seen in table 1, there are some buildings where foursquare api cannot identify or wrongly identify. For example, The building Indi home Maxim’s MX coordinate is not found, while K-point coordinate is located outside Hong Kong regime(Hong Kong area lies in longitude between 113 °E and 115 °E,

latitude between 22 °N and 23 °N). For simplicity, the analyzed dataset is filtered by removing coordinate value which is null and wrong.

Exploratory data analysis:

Part 1: General Understanding of Hong Kong Covid-19 Cases

Covid-19 Trend in Hong Kong: Daily New Cases Plot

To get a glimpse of the general confirmed cases in Hong Kong, fig.1 plot the daily new confirmed cases since 2020-01-23, when Hong Kong has the very first case(beginning of first wave) of Covid-19.

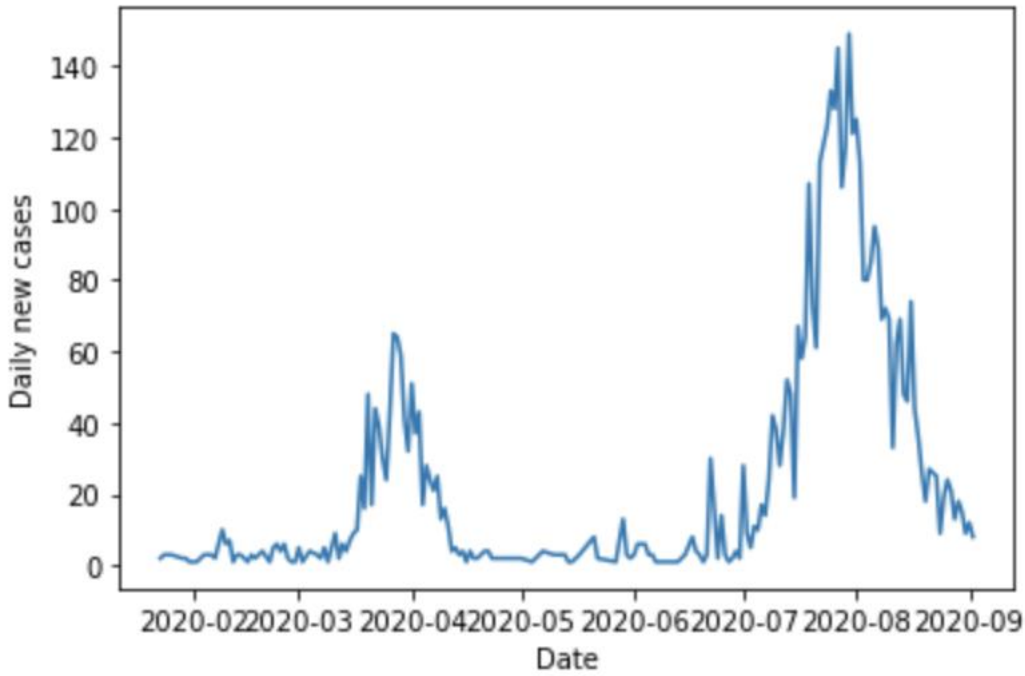


Fig.1: Daily new Covid-19 cases in Hong Kong

This figure showed that the second wave started around 2020-03-10, and the third wave started around 2020-07-04. The first wave did not show a bell curve so it can say that the first wave controlled well and did not cause an outbreak. As for second and third wave, a unique bell curve was formed and the peak reached up to 65 and 149 daily new cases respectively. The number of new cases then decreases to single digit around 2020-04-15 and 2020-09-03 respectively. Compared to the period of second and third waves, the second wave last around 35 days while third wave last around 50 days. Based on the data shown, it is possible that Covid-19 has become more infectious and spreading much quicker than before. This can be seen by the fact that Hong Kong government already had the experience dealing with

second wave, implemented much stricter measurement to contain and control the outbreak. Even so, the third outbreak remains the largest ever seen [4]. It is possible that the virus adapted well to human body and hence easier to get transmitted [5]. As the worldwide covid-19 cases still having an increasing trend, there may be a fourth outbreak in future, even though the current trend drops to single digit new cases daily. If that happens, law enforcement should consider act swiftly when starting to see an increasing trend in Hong Kong.

Gender plot of total Covid-19 cases in Hong Kong

Fig 2 shows the gender plot for covid-19 cases in Hong Kong. The result shows that the percentage between male and female are 50-50, meaning that covid-19 does not have gender bias towards infected person. Under this statistic, male or female have equal probability of getting infected.

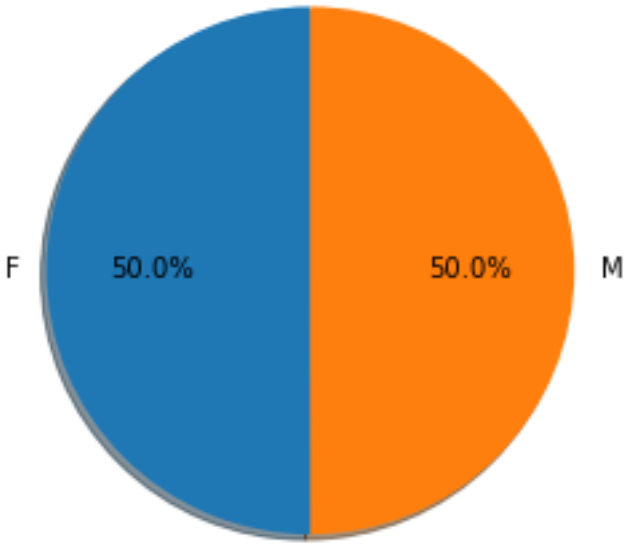


Fig 2: Gender Distribution of confirmed cases in Hong Kong.

Cases classification plot of Covid-19 cases in Hong Kong

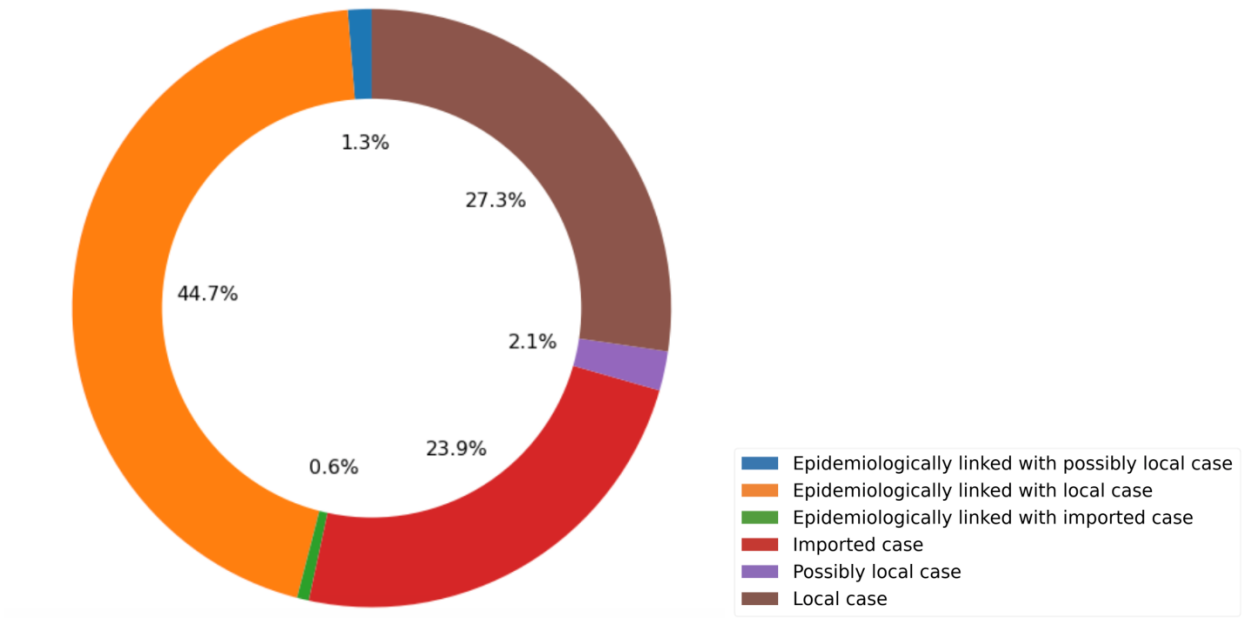


Fig 3: Case classification plot of covid-19 cases in Hong Kong

To get a closer look on infection details, we look at fig 3 which shows the confirmed cases classification. As seen in the figure, more than 70% are local/local transmitted case. By comparing local case to potential linked with local case, it is generally seen that for every local cases, there are 1.5 transmitted cases. That means in Hong Kong, for every 100 people infected, around 164 people get transmitted. On the contrary, for imported cases around 25% are imported/ imported transmitted case. For every 100 imported cases, only around 2.5 cases are transmitted in Hong Kong. Based on the statistics, it can see how easily the virus can spread locally, especially in Hong Kong where it is one of the densest populations in the world. Many policies need to be revised as the current law enforcement has created a lot of discomfort in the latest outbreak. On the other hand, the quarantine measurement applied by government to those travel to Hong Kong are preventing incoming transmission.

Age distribution plot of Covid-19 cases in Hong Kong

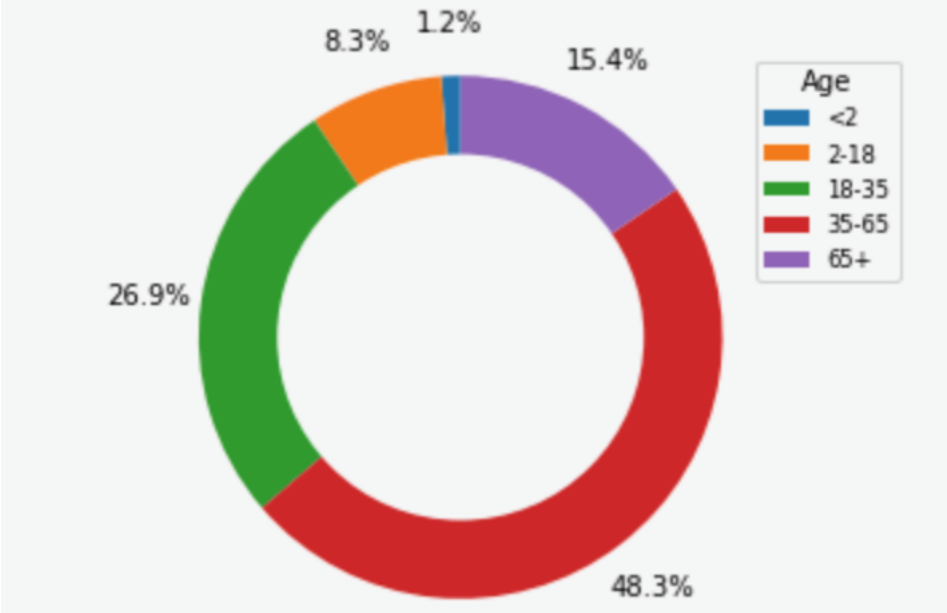


Fig.4 : Age distribution plot of Covid-19 cases in Hong Kong

Fig. 4 shows the pie chart of the percentage of confirmed cases based on different age category. We found that most cases are concentrated in middle age(35-65), which equal to 48.3% of the total cases. The following is the young adult(18-35) which equals to 26.9% of the total cases, whereas for old age(65+) only equals to 15.4% of total cases. However, when we look at the number of deceased cases, among the deceased record so far, 90.4% of them are in old age category, while the remaining deceased cases are in middle age group. This shows that the old age having the highest risk among all age group. Moreover, it is found that the confirmed cases which are Asymptomatic is huge, consist of more than 20% of total cases currently. This makes the detection of the virus more sophisticated as the individual does not show infected symptoms. Again, this increases the chance for everyone especially old age group of getting infected unknowingly. Although most cases can get recovered, we need to aware the fact that covid-19 can cause long lasting lung damage as well [6].

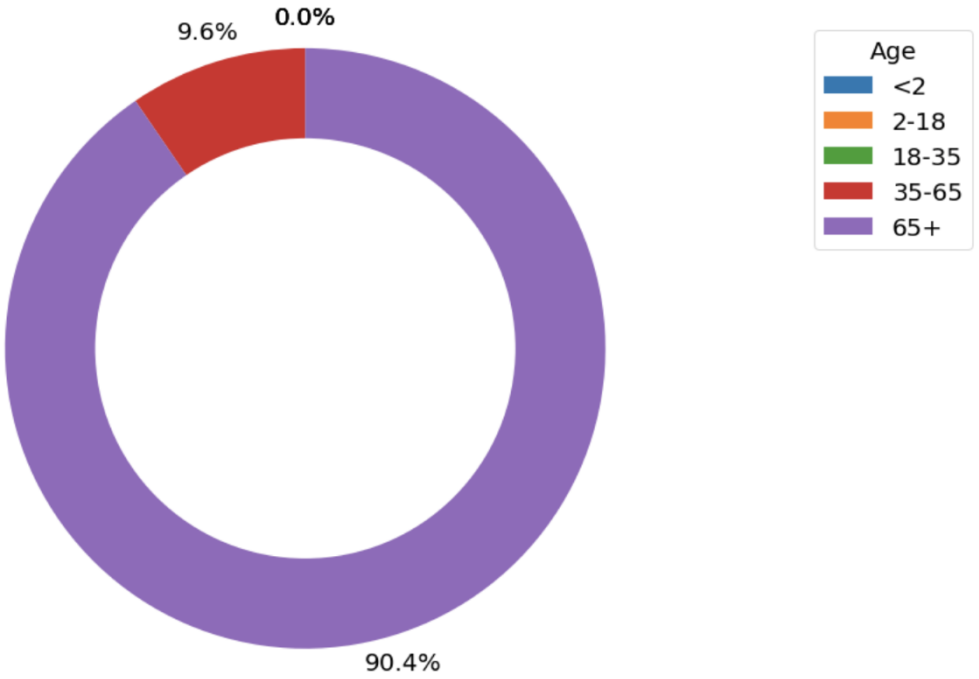


Fig.5 : Age distribution plot of Covid-19 deceased cases in Hong Kong

Part 2: Explore Covid-19 Third Wave Cluster Properties in Hong Kong.

Count Method

In the second dataset provided by government, they have included the Related probable/confirmed cases for each building name. Essentially we can directly count the number of Related probable/confirmed cases and find out the top 10 building name. The result is shown in Table 2.

Building name	Related probable/confirmed cases	count	lat_col	long_col
Tsz Wan Shan Shopping Centre	1311, 1312, 1404, 1444, 1477, 1491, 1507, 1508...	83	22.348022	114.200591
Yee On Court	2289, 2476, 2546, 2722, 3102, 3154, 3165, 3205...	35	22.320446	114.176428
Pei Ho Street Municipal Services Building	1557, 1608, 1611, 1808, 1852, 1905, 1926, 1927...	21	22.329372	114.160930
Shek Wu Hui Municipal Services Building	1947, 1989, 2053, 2072, 2179, 2284, 2393, 2576...	18	22.501996	114.130550
Ngau Chi Wan Municipal Services Building	1371, 1388, 1389, 1394, 1467, 1468, 1523, 1597...	17	22.334496	114.208999
Yeung Uk Road Municipal Services Building	1652, 1859, 2196, 2261, 2239, 3047, 3117, 3244...	17	22.368976	114.114582
Millennium City 3	2396, 2434, 2535, 2635, 2720, 2742, 2811, 2822...	16	22.314527	114.219643
San Hui Market	1471, 1561, 1907, 1931, 2154, 2381, 2458, 2493...	15	22.396796	113.976546
TKO Gateway	1590, 1983, 2123, 2330, 2479, 2572, 2670, 3245...	13	22.317220	114.266417
Kwai Chung Plaza	1820, 1996, 2332, 2455, 2459, 2465, 2602, 3018...	12	22.358099	114.127636

Table2: Top 10 building name having the most covid-19 cases

Fig 6: Folium map plot showing the top 10 building location. Radius of the circle indicates the number of cases in that building.

A general map plot is shown in fig 6, where the circle radius is drawn based on the number of cases. The biggest cluster observed is Tsz Wan Shan Shopping Centre, which have a total of 83 cases. The second biggest cluster is Yee On Court having 35 total cases and the third cluster is Pei Ho Street Municipal Services Building having 21 total cases. Under proper investigation, it was identified as Tze Wan Shan elderly home cluster, where spreading was centered at Tze Wan Shan elderly home [7]. The second cluster is related to direct-selling company Star Global at Yee On Court, in which they have reported having staff training without having mask on [8]. In Hong Kong, municipal buildings are places where markets and sport facilities are located at. In the top 10 buildings, we can see that all buildings are directly related to market or shopping centre, which are places that are less clean or places for gathering, which makes the perfect places for spreading the virus [9]. Moreover, since elderly having higher risk of getting infected, that may also act as a reason boost up the numbers of infections in particular for the first and third cluster, which has a high density of elderly people.

K-Means Method

a)

b)

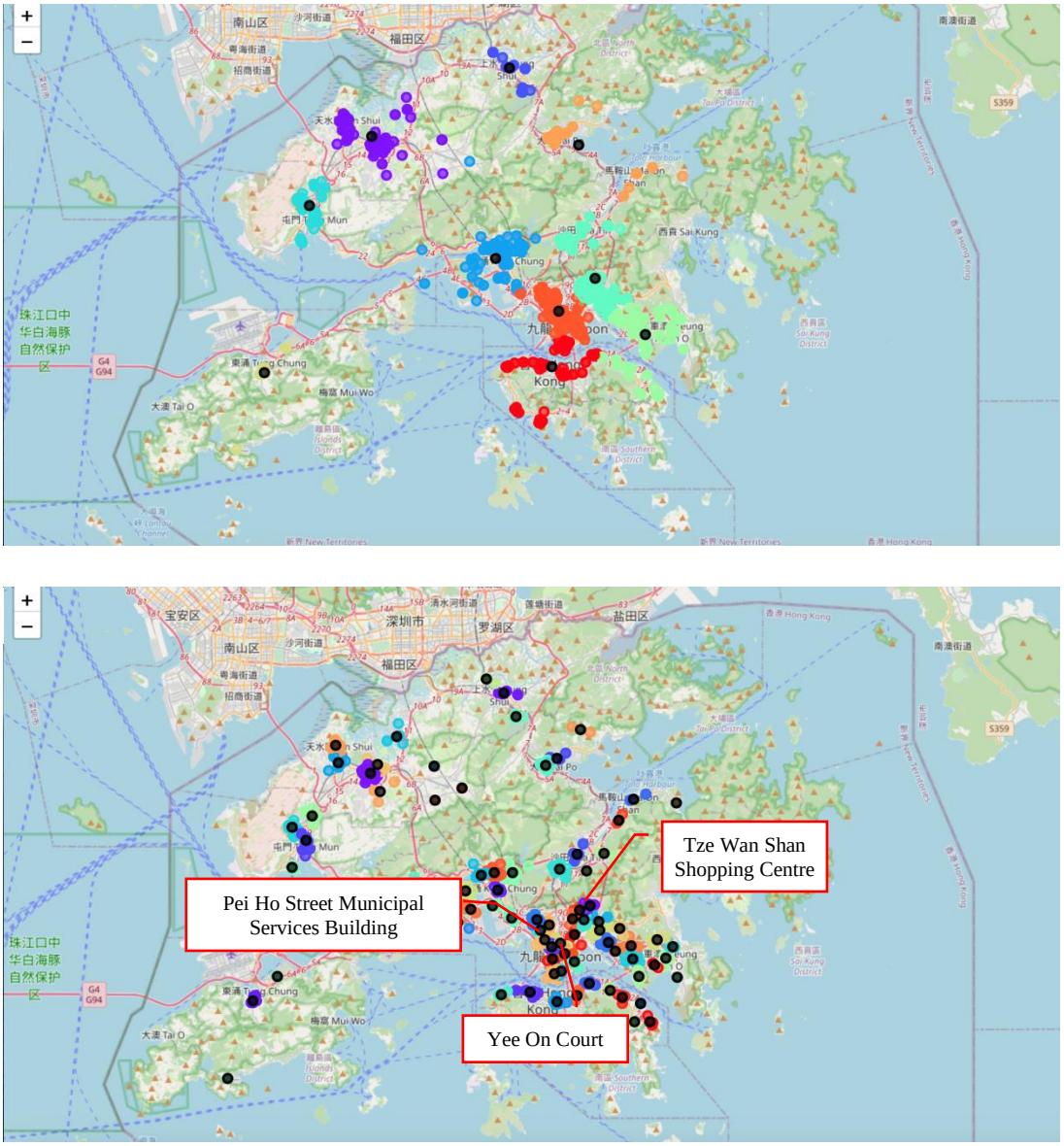


Fig 7. K-means clustering method. Different color denotes different cluster and black circle denotes each cluster center. (a) number of cluster =10. (b) number of cluster = 80.

In this section, the cases are grouped using k-means method by setting number of cluster as 10 and 80 respectively. The results are shown in fig 7. In fig 7(a), the numbers of cases are grouped into 10 different colors. Each cluster centers is denoted using black circle. In fig 7(a), it is seen that the cluster range is still too large. For example it groups cases at Tai Po and Ma On Shan as one cluster, though it is clear that they are separated quite far away. This set of cluster is unclear for government to perform any action. Setting number of cluster as 80 makes the cluster size much smaller, as shown in fig 7(b). If we compare fig 7(a) and 6(b), there are some similarities. For example we can see that both have cluster centered at Tze Wan Shan. This indicates that there is a high density of cases happening at that area

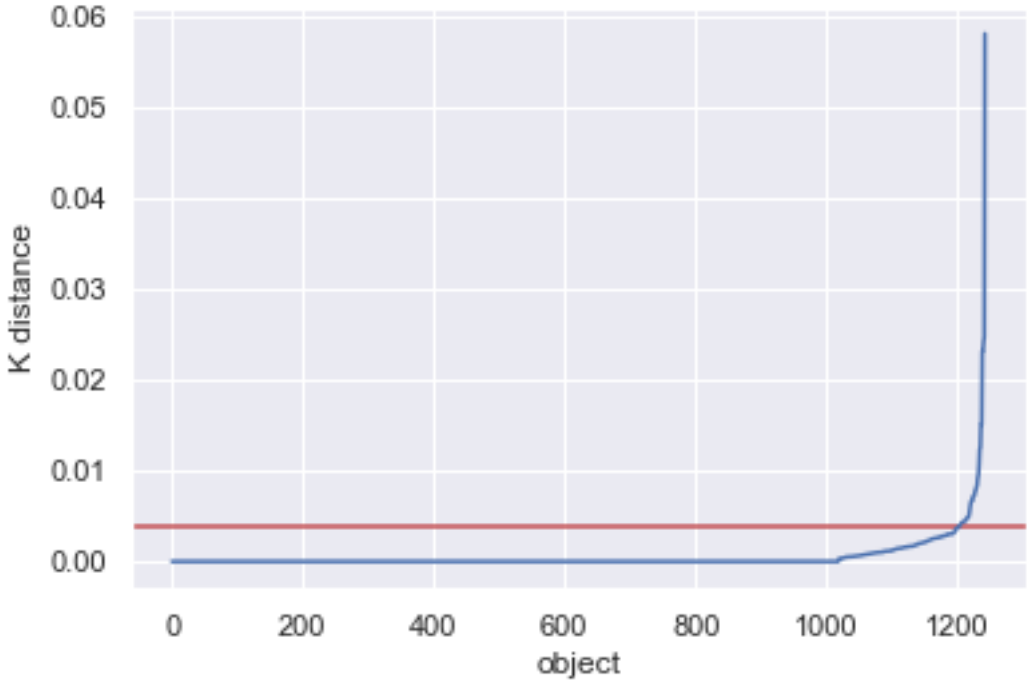
where it considers as a cluster point even though number of cluster breaks down to a larger size.

From this analysis we can see that k-means method can help us group the cases into different clusters based on our interest. Grouping cluster and find out the cluster center may help government to have a general idea on how the cases are distributed and clustered. This also helps government bodies to pinpoint certain areas to make better decision to control the outbreak more effectively.

DBSCAN

DBSCAN is another clustering method to identify the cluster based on density, which is effective to identify irregular cluster group. In DBSCAN method, two important parameters needed to define: epsilon(eps) and minimum points. Epsilon defines the radius of neighborhood around certain point, while minimum point defines the minimum number of points around that certain point within given eps value. We use the method introduced by Nadia Rahmah and Imas Sukaesih [10] to find the optimal eps value. Here the analysis follows the same procedure to generate k-dist plot as shown in fig 8(a).

a)



b)

c)

Fig 8: DBSCAN cluster plot. (a) The K-distance = 0.04 indicates optimal eps value eps = 0.04 using Nadia and Imas method. (b) DBSCAN result setting minimum points = 20. (c) DBSCAN result setting minimum points = 5.

In fig 8(a), the optimal eps value 0.04 was obtained for this dataset. As for minimum points, minimum points = 20 and 5 are chosen and the results are plotted in fig 8(b) and 8(c) respectively. In fig 8(b), there are at least 20 cases or above for each cluster, and it is observed that the top 3 building name where we obtained using count method are shown in this DBSCAN cluster. For moderate/smaller cluster, fig 8(c) is able to identify those, such as the cluster at Kwun Tong area, which considered as one of the largest local cluster back then [11].

Compared to k-means cluster, DBSCAN has the advantage of grouping the cluster in irregular shape, having the study to be more precise under certain area. The disadvantage of DBSCAN is that the results does not have a unique cluster point and more work need to be done in order to identify the exact spreading spot.

Conclusion:

This report provides a comprehensive study of the current covid-19 cases in Hong Kong. On the part 1 analysis results, it has shown that Hong Kong had experience an outbreak currently and it reached record of 145 new cases per day back in 2020-07-30. Looking at the latest new cases, it is seen that the outbreak is now contained and in control. Almost all deceased cases are at the old age which poses a high risk to Hong Kong elderly society since it is well known that Hong Kong has recorded in 2018 to have elderly consist around 1.27 million(17.9%) of the total Hong Kong population [12].

In part 2 analysis, three types of clustering method are used to identify the covid-19 clusters in Hong Kong, namely count method, k-means method and DBSCAN method. The count method count the total number of cases related to the building, which is useful to pinpoint the cluster outbreak point and government can perform immediate cleansing on the building. All cases can be grouped into clusters using k-means method. In this method it was found that the cluster center is close to the building name that having many cases which verifies the usefulness of k-means method. Finally, the grouping using DBSCAN method seems more convincing as it relates cases based on the density. The grouping results where min_samples = 20 are the cases of high density cases, while min_samples = 5 can show the cluster that are relatively low in density, but it also define the cluster as the virus spread quickly from one to another easily.

More analysis can be done in future. For example, if we have a time series of dataset 2, we can group and analyze the cluster behavior. From this analysis we are able to obtain the cluster movement whether the cluster spread even widely, or move towards other location, or even shrinking after quarantine has been made. All Health Department and Law enforcement agencies should be setting covid-19 control as priority as the outbreak already worsen the economics in Hong Kong and it will definitely getting much worse if Hong Kong having another outbreak in future.

Reference:

1. Severe acute respiratory syndrome coronavirus 2. https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2
2. Details of propable/confirmed cases of COVID-19 in Hong Kong. https://www.chp.gov.hk/files/misc/enhanced_sur_covid_19_eng.csv
3. Residential buildings in which probable/confirmed cases have resided in the past 14 days or non-residential building with 2 or more probable/confirmed cases in the past 14 days http://www.chp.gov.hk/files/misc/building_list_eng.csv
4. Hong Kong third wave: record 145 Covid-19 cases trigger toughest preventive measures yet. https://www.scmp.com/news/hong-kong/health-environment/article/3094787/hong-kong-third-wave-social-distancing-measures
5. Coronavirus: Hong Kong on verge of 'large-scale' outbreak, says Carrie Lam. https://www.bbc.com/news/world-asia-china-53575875
6. Coronavirus can cause lasting lung damage – but the effects may ease over time. https://theconversation.com/coronavirus-can-cause-lasting-lung-damage-but-the-effects-may-ease-over-time-140398
7. Coronavirus: Hong Kong sees daily record high of 67 infections, death toll rises to 10. https://hongkongfp.com/2020/07/16/coronavirus-hong-kong-sees-daily-record-high-of-67-infections-death-toll-rises-to-10/
8. Hong Kong reports 125 new cases and four deaths within a day. https://www.thestandard.com.hk/breaking-news/section/4/152194/Hong-Kong-reports-125-new-cases-and-four-deaths-within-a-day
9. 42 個高危街市一貼睇晒 最多確診者到訪北河街街市. https://hk.news.yahoo.com/42-%E5%80%8B%E9%AB%98%E5%8D%B1%E8%A1%97%E5%B8%82%E4%B8%80%E8%B2%BC%E7%9D%87%E6%99%92-%E6%9C%80%E5%A4%9A%E7%A2%BA%E8%A8%BA%E8%80%85%E5%88%B0%E8%A8%AA%E5%8C%97%E6%B2%B3%E8%A1%97%E8%A1%97%E5%B8%82-132221700.html
10. Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. Nadia Rahmah1, Imas Sukaesih Sitanggang. https://iopscience.iop.org/article/10.1088/1755-1315/31/1/012012/pdf
11. Covid-19 surge: Hong Kong admits quarantine exemptions may be to blame, as city sees 106 new infections. https://hongkongfp.com/2020/07/28/covid-19-surge-hong-kong-admits-quarantine-exemptions-may-be-to-blame-as-city-sees-106-new-infections/
12. Population ageing trend of Hong Kong. Dr. Kelvin Wong, Matthew Yeung. https://www.hkeconomy.gov.hk/en/pdf/el/el-2019-02.pdf