

Analysis of accessible food services to low income households

Introduction

In times of the 2020 global pandemic COVID-19, Singapore has implemented a lock down initiative, also known as the circuit breaker, to prevent outbreaks of the coronavirus. There are low income households who might find daily life difficult to manage due to the sudden circuit breaker implementations. These households struggle with high living expenses and insufficient financial support. In addition, if their neighbourhood are inaccessible to food services, it might be difficult for them to commute around as there might be additional transport cost incurred.

The objective of this project is to find neighbourhoods to prioritise according to their proportion of low-income residents and availability of food services in their vicinity. This analysis is to provide insight to interested charitable bodies or organisations, so that when there are willing community support initiatives that provides food charity handouts, they can be more informed over which neighbourhood to prioritise first in order to maximise the effectiveness of the compassionate handouts.

Data

The data set used in this analysis is from the Singapore Ministry of Trade and Industry - Department of Statistics, available at: <https://data.gov.sg/dataset/resident-working-persons-aged-15-years-over-by-planning-area-gross-monthly-income-from-work-2015>.

The data set consist of the 2015 income data of Singapore residents in each planning area. Planning Areas, also known as DGP areas or DGP zones, are the main urban planning and census divisions of Singapore delineated by the Urban Redevelopment Authority. They will be interpreted as 'Neighbourhood' for boundary division in this analysis.

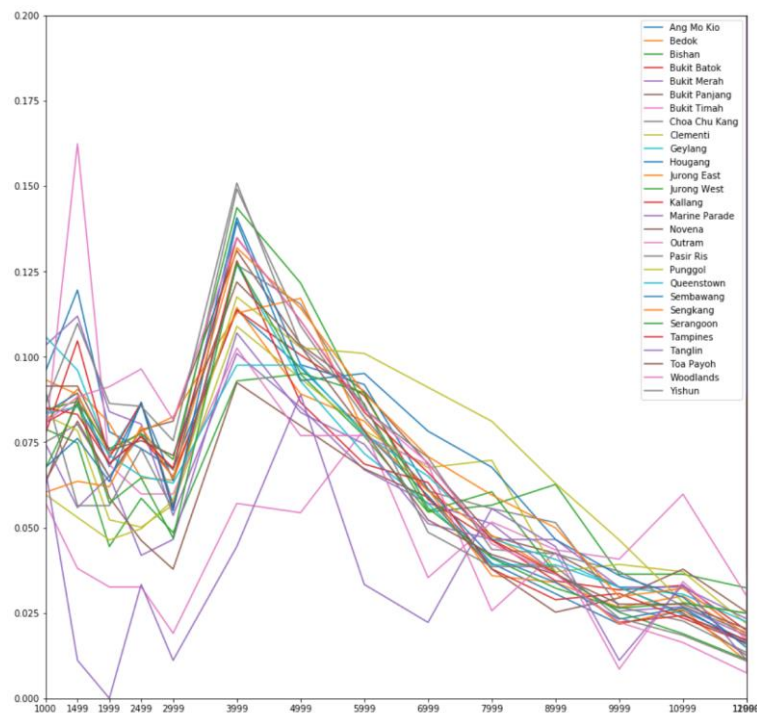
Since the data set is in population count, we have to convert the population data to proportion. Physical count of population in each income bracket is divided by total population in the planning area to obtain the proportion of the population in each income bracket. Since each planning area differs in area size and total population, it is more effective to target communities with larger proportion of in-need households.

For obtaining budget food services, Fourspace API is used to obtain them within the vicinity of the neighbourhood. In addition, geopy package is incorporated to obtain latitude and longitude coordinates for each planning area.

(Disclaimer: 2020 data set is not available at the point of this writing; this General Household Survey is a series of mid-decade national survey, so data are only available in 5-year intervals [2000, 2005, 2010, 2015].)

Methodology

For exploratory analysis, we plot the proportion per income bracket over each income bracket, to detect any pattern in the data that might help with our evaluation.



From the plot we obtained, we can see a general trend across all regions. This shows that middle to upper class households in Singapore are rather evenly distributed across each planning area the nation, following the similar distribution. However, for income range below 1999, there are a lot of variants in the proportion numbers for each planning area. There are some communities that are occupied with more low-income residents than others, which are the ones that will be the focus of our study.

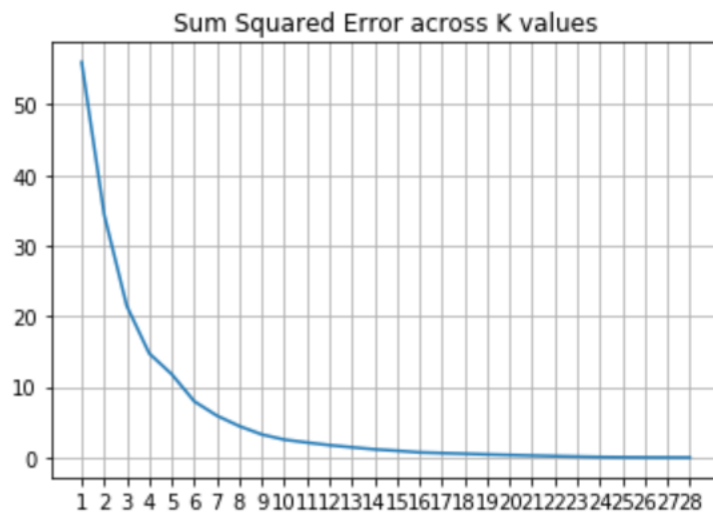
Since our goal is to assess the top priority of the in-need per planning area, we shall cumulate the population below 1999 salary and combine them to form our low-income category.

Next, we will query Fourspace API in the form of longitude and latitude to obtain food services in the vicinity. In order to get the geospatial coordinates, geopy package is used to convert the planning area to its respective longitude and latitude values. For which, the query will search the location point at a 2.5km radius, identifying nearby budget food venues.

As such, we will obtain our features for our analysis. Since the two features have different ranges, by normalising/standardizing the features around the center at 0 and with a standard deviation of 1, we can prevent the bias contributed by different scales. This is because the measurements of each feature is in different units and do not contribute equally to the analysis.

Then, K-means clustering is performed using the featurised data to obtain similar clusters. An optimal K is searched by fitting various values of K and finding the sum of squared differences.

The plot below shows the relationship between varying K and error values.



Cluster number K = 2 appears to receive the most drop in error. However, in this case, any number of clusters from 2 to 5 is feasible as we would like to narrow down the clusters. We'll select 5 to better segment and identify different kinds of in need households (low income, high number vs low number of available food services).

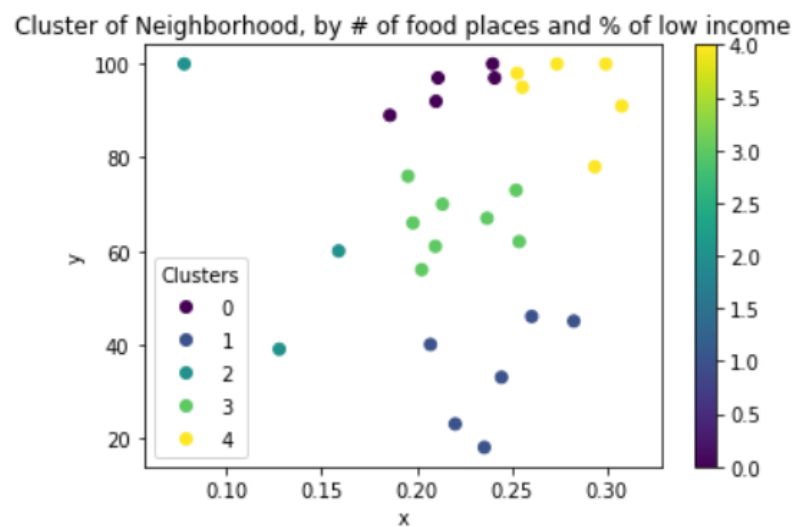
Results

The below two tables shows the sorted data according to percentage of low income first and number of budget food services nearby first respectively. This is to identify the pattern of the cluster IDs assigned to each planning area.

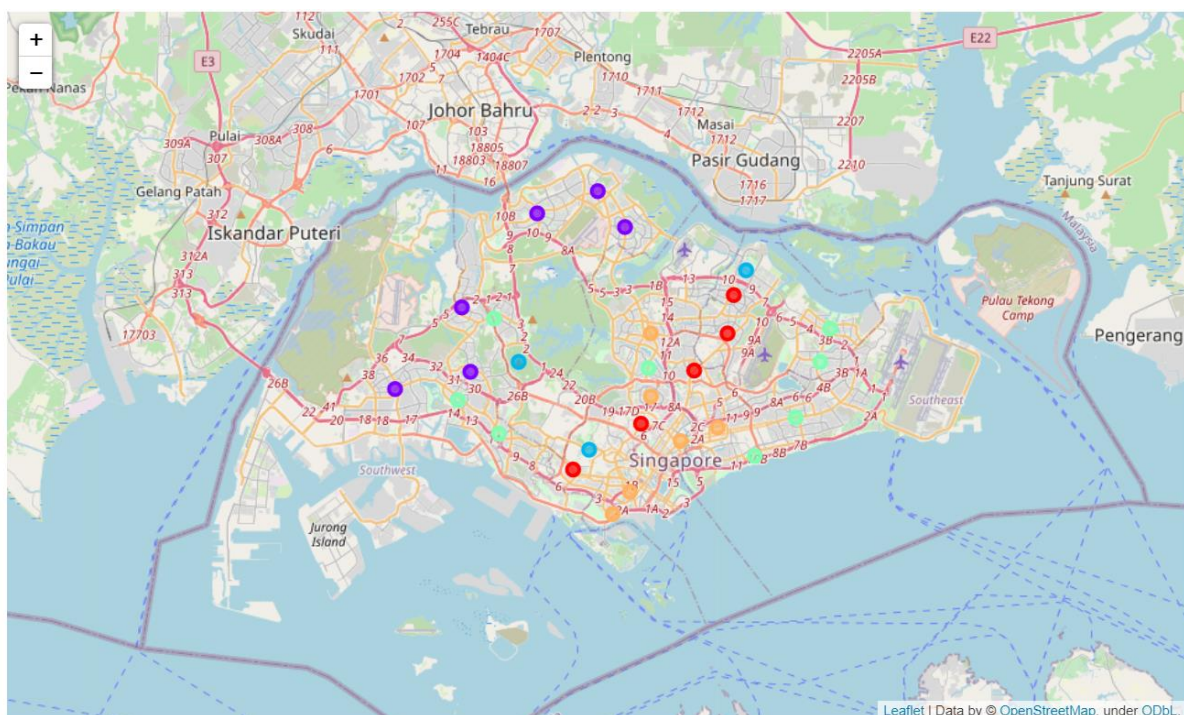
	Neighborhood	Latitude	Longitude	% of low income	# of budget food places	Cluster ID
16	Outram	1.282870	103.837860	0.307692	91	4
4	Bukit Merah	1.270439	103.828318	0.299270	100	4
0	Ang Mo Kio	1.370073	103.849516	0.293478	78	4
27	Yishun	1.429384	103.835028	0.282481	45	1
9	Geylang	1.318186	103.887056	0.273616	100	4
26	Woodlands	1.436897	103.786216	0.260386	46	1
25	Toa Payoh	1.335391	103.849741	0.255443	95	4
11	Jurong East	1.333115	103.742297	0.253796	62	3
13	Kallang	1.310759	103.866262	0.252708	98	4
1	Bedok	1.323976	103.930216	0.252162	73	3
12	Jurong West	1.339636	103.707339	0.244486	33	1
10	Hougang	1.370682	103.892545	0.240894	97	0
19	Queenstown	1.294623	103.806045	0.239837	100	0
23	Tampines	1.354653	103.943571	0.236911	67	3
3	Bukit Batok	1.349057	103.749591	0.235450	18	1
7	Choa Chu Kang	1.384749	103.744534	0.220165	23	1
8	Clementi	1.315100	103.765231	0.213508	70	3
22	Serangoon	1.349862	103.873729	0.211144	97	0
15	Novena	1.320526	103.843881	0.210084	92	0
5	Bukit Panjang	1.378629	103.762136	0.209738	61	3
20	Sembawang	1.449093	103.820055	0.207188	40	1
17	Pasir Ris	1.373031	103.949255	0.202564	56	3
2	Bishan	1.350986	103.848255	0.197980	66	3
14	Marine Parade	1.302689	103.907395	0.195349	76	3
21	Sengkang	1.391654	103.895364	0.185841	89	0
18	Punggol	1.405258	103.902330	0.158940	60	2
6	Bukit Timah	1.354690	103.776372	0.127717	39	2
24	Tanglin	1.306044	103.815280	0.077778	100	2

	Neighborhood	Latitude	Longitude	% of low income	# of budget food places	Cluster ID
3	Bukit Batok	1.349057	103.749591	0.235450	18	1
7	Choa Chu Kang	1.384749	103.744534	0.220165	23	1
12	Jurong West	1.339636	103.707339	0.244486	33	1
6	Bukit Timah	1.354690	103.776372	0.127717	39	2
20	Sembawang	1.449093	103.820055	0.207188	40	1
27	Yishun	1.429384	103.835028	0.282481	45	1
26	Woodlands	1.436897	103.786216	0.260386	46	1
17	Pasir Ris	1.373031	103.949255	0.202564	56	3
18	Punggol	1.405258	103.902330	0.158940	60	2
5	Bukit Panjang	1.378629	103.762136	0.209738	61	3
11	Jurong East	1.333115	103.742297	0.253796	62	3
2	Bishan	1.350986	103.848255	0.197980	66	3
23	Tampines	1.354653	103.943571	0.236911	67	3
8	Clementi	1.315100	103.765231	0.213508	70	3
1	Bedok	1.323976	103.930216	0.252162	73	3
14	Marine Parade	1.302689	103.907395	0.195349	76	3
0	Ang Mo Kio	1.370073	103.849516	0.293478	78	4
21	Sengkang	1.391654	103.895364	0.185841	89	0
16	Outram	1.282870	103.837860	0.307692	91	4
15	Novena	1.320526	103.843881	0.210084	92	0
25	Toa Payoh	1.335391	103.849741	0.255443	95	4
10	Hougang	1.370682	103.892545	0.240894	97	0
22	Serangoon	1.349862	103.873729	0.211144	97	0
13	Kallang	1.310759	103.866262	0.252708	98	4
4	Bukit Merah	1.270439	103.828318	0.299270	100	4
9	Geylang	1.318186	103.887056	0.273616	100	4
19	Queenstown	1.294623	103.806045	0.239837	100	0
24	Tanglin	1.306044	103.815280	0.077778	100	2

The scatter clutter plot is as shown below. We observe that cluster ID 1 (dark-blue) is our top priority, since the planning area have largest proportion of low income and very low number of food services nearby. The next on the list would be cluster ID 4 (yellow) since they have low income but rather high number of food services nearby.



By plotting the planning area labelled according to clusters on the satellite map of Singapore, we can identify the regions of similar planning areas. Our priority clusters are cluster 1 (purple) and 4 (orange).



Discussion

The analysis has shown that even though there is a clear distinction between planning areas with respect to the proportion of low-income residents, there is always a low-income population in each planning area. This might be contributed to non-working family households in middle and upper class families that are either retired or working a low salary job.

Therefore, the analysis has siphoned out the highest proportion of low-income residents in the segmented planning areas to support them since they might be a retiring community or an entirely low-income community. The generalisation of planning area can be wide and might contribute to any inaccuracy in the priority listing, therefore a better choice would to further divide the planning areas into sub regions to target more specific smaller communities.

However, due to the restraint of access to public government data, the analysis shall resolute around planning areas.

Conclusion

From this analysis, we have gotten insight over which planning areas are on top priority for food charity handouts, matching the criteria of high proportion of low-income residents and lack of food services in nearby vicinity. By targeting dense low-income communities without easy access to food, the similarity information can contribute to better decision making for any charitable entity to deliver their contributions in hopes to maximise the impact of their kind acts.

Not only is this analysis beneficial for providing information of top priority planning areas within need households, but also beneficial to identify the similar pattern of different classes within and between communities.