Exploring Hong Kong for Opening A New Restaurant

1. Introduction

"The City That Never Sleeps" - is a well known nickname for Hong Kong, my home town. It is an international city located in the East of Asia with a mixture of Asian and Western culture. It has a smaller area compared to other multicultural cities however it has a broad variety of food selection, from the roadside stalls to the most upscale restaurants. According to Google, it has one of the highest density of restaurants in the world containing over 15,000 restaurants which means almost 20.4 restaurants per 10,000 people. I can guarantee you can find almost any *cuisine* in this city within a 30 - 45 minutes ride, especially on Hong Kong Island, a very popular part of Hong Kong. So as you expected, the food/restaurant industry has a high contribution to the city's economy and tends to have a desirable return for owners or investors.

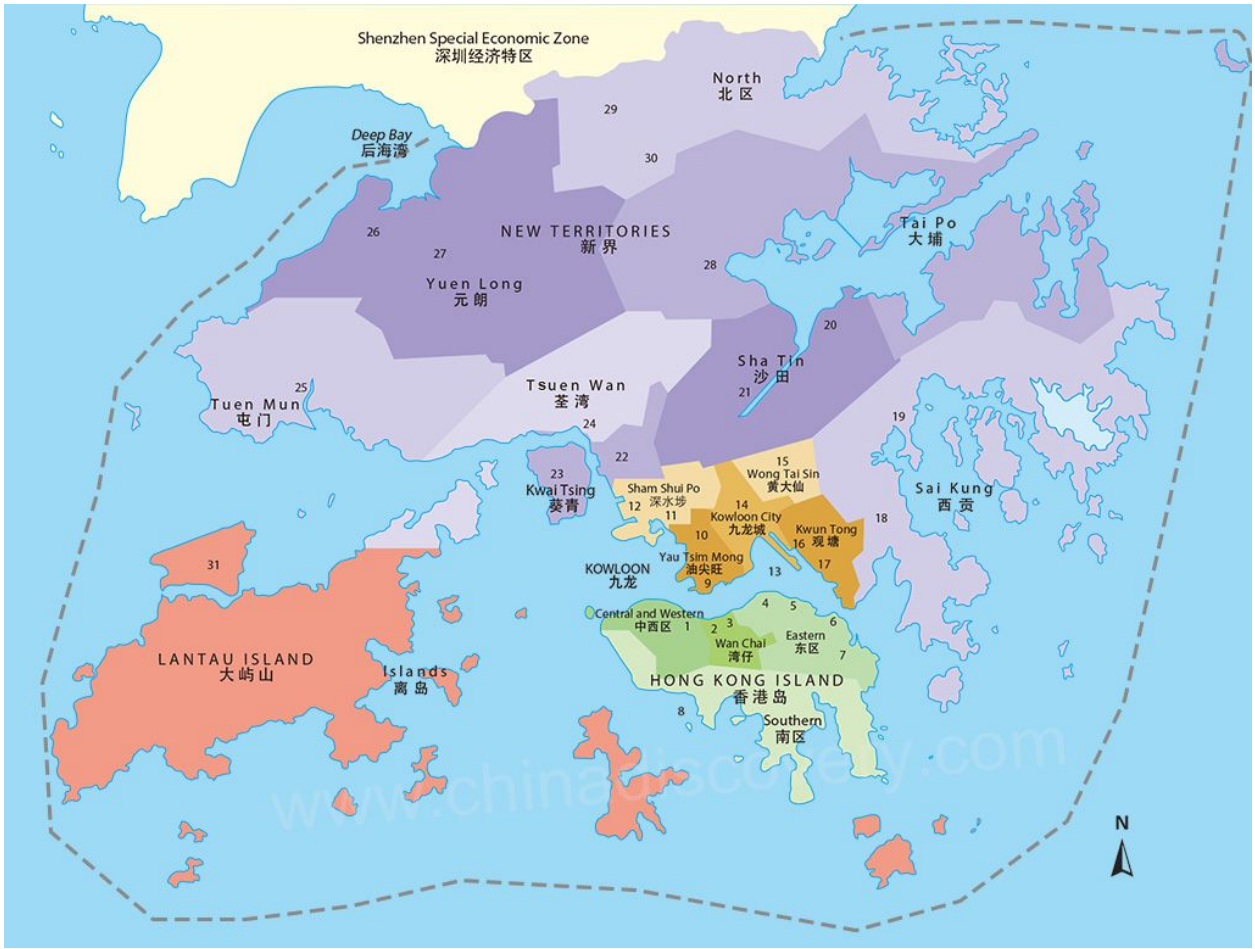


Figure 1.Hong Kong 18 District Map - [Photograph]. Retrieved from https://upload.wikimedia.org/wikipedia/commons/f/fe/Map_of_Hong_Kong_18_Districts_en.svg [Accessed 14 Dec.2020]

1.1. Objective and Interest

The objective of this project is to analyse and select the best locations within Hong Kong Island to open a new restaurant and decide on the cuisine with high potential. This can be useful to anyone who also has an interest in opening a restaurant in Hong Kong and in particular- me, as my family already own a few restaurants and thinking of expanding the business furthermore.

2. Data

2.1. Description of data

- Lists of different districts of the Hong Kong Island

  I used Wikipedia to research Hong Kong's geography (pretending I am not familiar with the city) and focused on the Hong Kong Island region - It has 4 districts that cover almost a dozen of smaller neighborhoods.

- Latitude and longitude coordinates of these districts for visualization and obtaining venue data via the foursquare API[2]

- Venue category that specific to "Restaurants"

2.2. How it will solve the problem

By identifying the restaurants' information, we can perform statistics to these data and have a more numeric idea on the popularity of restaurant or cuisine types in a particular area or cluster as shown later on in this report. Also, we can visually see the distributions on the map and better understand the density around the neighborhood. All these allow us to narrow down our options and give us clear instruction on the preferable locations.

2.3. *Data Source*

The main data source are as following:

- Google and Wikipedia - for obtaining the district information
- FourSquare API- all venue related information such as name, category, and address
- Geocoder ArcGIS[3] to obtain the latitude and longitude

3. Methodology

Several techniques are used in this project

- Wikipedia - web scraping techniques
  As mentioned above, there is a district table on the Hong Kong Wikipedia [1] page but in HTML format so I used the pandas *pd.read_html* function to load all the data frames (14 in

this case) within the page and focuses on the one in scope (the [6]th table).

Import a Hong Kong district dataframe obtained from wikipedia

```
hkdistrict18 = "https://en.wikipedia.org/wiki/Districts_of_Hong_Kong"

hkdistrict18_df=pd.read_html(hkdistrict18)
len(hkdistrict18_df)
```

```
14
```

```
hkdistrict18_df[6] #find out the table that show all the districts & region of Hong Kong
```

	District	Region	Chinese	Area(km2)	Comparable Territory
0	Central and Western	Hong Kong Island	中西區	12.44	Tokelau (New Zealand)
1	Eastern	Hong Kong Island	東區	18.56	Nauru

- Data Wrangling

  The district table we obtained from Wikipedia contains not applicable data such as Kowloon neighborhoods, area in km^2 ,etc so we have to clean up the data frame so it only shows the information of interest. The methods we used including ".loc == 'Hong Kong Island" and the Pandas *drop* function.

```
hkdistrict18_data=hkdistrict18_df[6] #assign a new table name

#since some of the district name is too general, an extra column is added to specify the address for fo
hkdistrict18_data['geo_address']=hkdistrict18_data['District'].str.cat(hkdistrict18_data['Region'],sep=

#drop unwanted columns
hkdistrict18_data1=hkdistrict18_data.drop(['Chinese', 'Area(km2)','Comparable Territory'], axis=1)

hkdistrict18_data1.head
```

```
#identify the rows that applicable to Hong Kong Island Region
hkisland_df=hkdistrict18_data1.loc[hkdistrict18_data1['Region']=='Hong Kong Island']
hkisland_df
```

	District	Region	geo_address
0	Central and Western	Hong Kong Island	Central and Western,Hong Kong Island
1	Eastern	Hong Kong Island	Eastern,Hong Kong Island
2	Southern	Hong Kong Island	Southern,Hong Kong Island
3	Wan Chai	Hong Kong Island	Wan Chai,Hong Kong Island

- Python Geocoder package

  Populating the addresses ("geo_address") as obtained from the cleaned data frame and passing it into the Geocoder code. We also need to specify the user agent so that the API services can retrieve the latitude and longitude of the said addresses and create a data frame ("hkisland_df_final") by appending it to the original table.

- Data Visualization API

  Folium & Word cloud
  Stipulate the variable name for the map we want to create and specifics within the codes all the details of desired features that are going to be shown after rendering such as the display location, colour, size, opacity, and label information of the markers and pop-ups. Similar

concepts with generating the Word cloud image.

Other matplotlib package related data visualization methods are also used e.g. bar chart.

- FourSquare API

  I have defined all the parameters (limit, radius, client ID, version, etc) that need to incorporate the foursquare API codes and generated an URL link. This URL link is then passed into the request. get() method to retrieve the JSON file information back.

  The limit we set for the venue quantity limit is 100 and a radius of 2000 meters.

```
    CentralNWestern_latitude,
    CentralNWestern_longitude,
    radius,
    LIMIT)
url # display URL
```

```
'https://api.foursquare.com/v2/venues/explore?&client_id=SIB1HHX3KY0MKB1MLAW0O1ATKYMWIEFRBNYABIC2WXJC
Y5RV&client_secret=EJZDHTTMN22S5EOSZYNXCCCETEEIXKKIKUVX3HZZR2QKAMJE&v=20180604&ll=22.2818286,114.1582
784&radius=2000&limit=100'
```

```
results = requests.get(url).json()
results
```

```
{'meta': {'code': 200, 'requestId': '5fd7eabce08ec61add58073e'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Central and Western District',
  'headerFullLocation': 'Central and Western District, Hong Kong',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 234,
```

- Define and regrouping of data

  For me this is a different approach with data cleaning as rather than removing unwanted information or filtering the table, additional data needs to be added for us to have a more inclusive and informative database and be able to perform a more confident K-means clustering method (see later points).

  I need to add in extra neighborhood information (the original data source from Wikipedia did not specify any neighborhood information and I realise the 4 districts hence cluster sample is not working well (4 single points instead of the “cluster" of data that we are looking for).

  Also, I need to align some of the restaurant categories so our statistics output will be most meaningful.
  (e.g. “Cantonese restaurant”, “Beijing Restaurant” and “Dim SUm restaurant” are the same as “Chinese restaurant)

	A	B	C	D	E	F	G
	District	Neighborhood	Category	Name	Latitude	Longitude	Address
	Central & Western	Central	Dim Sum Restaurant	Mott 32 (卅二公館)	22.28028	114.1590	B/F, Standard Chartered Bank Building, 4-4
	Central & Western	Central	Italian Restaurant	8½ Otto e Mezzo Bomba	22.28172	114.1587	Unit 202, 2/F, Alexandra House, 18 Chater
	Central & Western	Central	Pakistani Restaurant	New Punjab Club	22.28025	114.1554	G/F, World Wide Commercial Bldg, 34 Wyr
	Central & Western	Central	Italian Restaurant	Carbone	22.28069	114.1554	9/F, LKF Tower, 33 Wyndham St
	Central & Western	Central	Cantonese Restaurant	Man Wah (文華廳)	22.28199	114.1592	25/F, Mandarin Oriental Hong Kong, 5 Con
	Central & Western	Central	Japanese Restaurant	Sake Bar Ginn	22.28085	114.1553	Unit 4C, Ho Lee Commercial Bldg, 38-44 D
	Central & Western	Central	Chinese Restaurant	Howard's Gourmet (好酒	22.28146	114.1614	5/F, CCB Tower, 3 Connaught Road Central
	Central & Western	Central	Dumpling Restaurant	Cheung Hing Kee Shangh	22.28242	114.1535	52 Lyndhurst Terrace
	Central & Western	Central	Chinese Restaurant	China Tang (唐人館)	22.28053	114.1576	Shop 411-413, 4/F, Landmark Atrium, 15 C
	Central & Western	Central	Asian Restaurant	Brass Spoon	22.28332	114.1560	G/F, 10 Pottinger St
	Central & Western	Central	Japanese Restaurant	Toritama (酉玉)	22.27967	114.1552	2 Glenealy
	Central & Western	Central	Japanese Restaurant	Ronin	22.28452	114.1524	8 On Wo Ln
	Central & Western	Central	Thai Restaurant	Samsen (泰館)	22.28508	114.1522	G/F, 23 Jervois Street

For more reasons causing incompletion of data, please refer to the *Challenges* section.

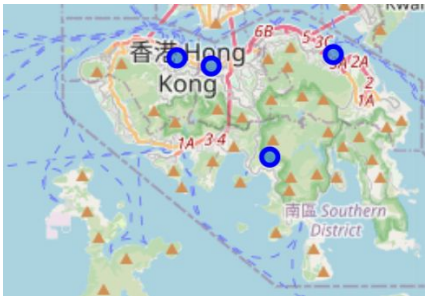
- <u>K-Means clustering (Unsupervised Machine Learning)</u>
  First, we use the elbow method to run the neighborhood data and find out a range of values for K and find out the optimal value for K for our data by determining the elbow point. After that, we can run SciKit-Learn's implementation of the K-means clustering method and divide our data into clusters with similar characteristics.

- <u>Others</u>
  Other minor techniques used in this project including value_counts(), sort values, and frequency to find out the occurrence of the restaurant category and using dummy encoding (or one-hot encoding) to append our data frame and prepare for necessary statistics work or data visualization.

4. Result

- Hong Kong has 4 regions (districts) within the Hong Kong Island and covers 16 neighborhoods
- The latitudes and longitudes for the 4 districts are as following:

	District	Region	geo_address	Latitude	Longitude
0	Central and Western	Hong Kong Island	Central and Western,Hong Kong Island	22.281829	114.158278
1	Eastern	Hong Kong Island	Eastern,Hong Kong Island	22.283121	114.224180
2	Southern	Hong Kong Island	Southern,Hong Kong Island	22.243200	114.197400
3	Wan Chai	Hong Kong Island	Wan Chai,Hong Kong Island	22.279015	114.172483



- Southern district has the busiest word cloud compare to the other 3 districts:

Top: Southern District venue category word cloud
Bottom: Central & Western venue category word cloud

- Before data modification, there are 114 different types of restaurants captured (different varieties of Chinese restaurants such as Beijing restaurant, Cantonese restaurant, Dim Sum restaurant); it has 28 types after redefining the venue category.
- The top 5 cuisines offered as per each of the 4 districts are:

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central & Western	Japanese Restaurant	Italian Restaurant	Cantonese Restaurant	Thai Restaurant	Chinese Restaurant
1	Eastern	Chinese Restaurant	Japanese Restaurant	Seafood Restaurant	French Restaurant	Cantonese Restaurant
2	Southern	Chinese Restaurant	Thai Restaurant	Restaurant	Italian Restaurant	Mediterranean Restaurant
3	Wan Chai	Japanese Restaurant	Chinese Restaurant	Italian Restaurant	Thai Restaurant	Middle Eastern Restaurant

Japanese & Chinese restaurants occupied the 1st most common venue position on a 1:1 ratio.

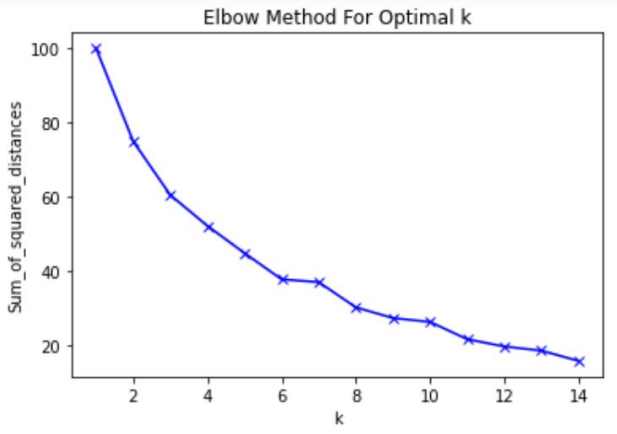
- The top 5 cuisines offered as per each neighborhood are:

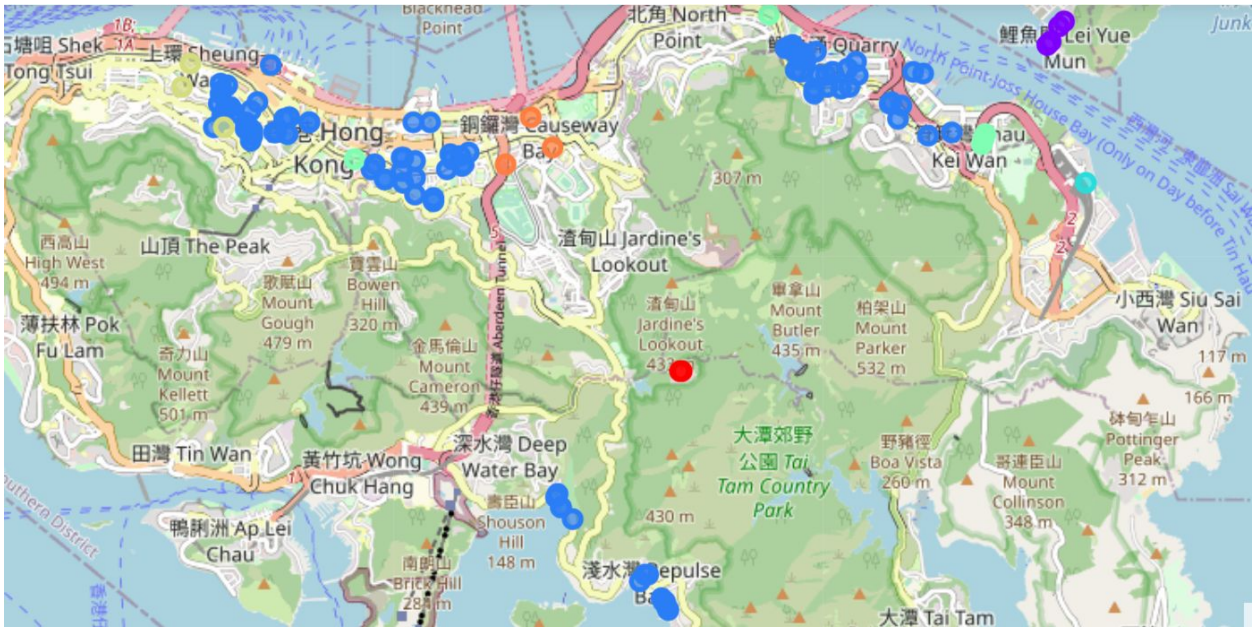
```
----Admiralty----
                          venue  freq
0             Chinese Restaurant   1.0
1            American Restaurant   0.0
2       Mediterranean Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4                Thai Restaurant   0.0


----Causeway Bay----
                          venue  freq
0          Vietnamese Restaurant  0.33
1           Taiwanese Restaurant  0.33
2            Japanese Restaurant  0.33
3       Mediterranean Restaurant  0.00
4  Vegetarian / Vegan Restaurant  0.00


----Central----
                  venue  freq
0   Chinese Restaurant  0.30
1   Italian Restaurant  0.17
2  Japanese Restaurant  0.13
3      Thai Restaurant  0.07
4    French Restaurant  0.07


----Chai Wan----
                          venue  freq
0           Fast Food Restaurant   1.0
1            American Restaurant   0.0
2       Mediterranean Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4                Thai Restaurant   0.0
----Deep Water Bay----
                        venue  freq
0            Thai Restaurant   0.4
1         Chinese Restaurant   0.2
2       Fast Food Restaurant   0.2
3  Modern European Restaurant   0.2
4        American Restaurant   0.0


----Lei Yue Mun----
                          venue  freq
0            Seafood Restaurant   1.0
1           American Restaurant   0.0
2       Mediterranean Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4                Thai Restaurant   0.0
```

```
----North Point----
                         venue  freq
0             Chinese Restaurant   1.0
1           American Restaurant   0.0
2       Mediterranean Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4               Thai Restaurant   0.0


----Quarry Bay----
                         venue  freq
0             Chinese Restaurant  0.29
1           Japanese Restaurant  0.21
2           American Restaurant  0.07
3  Vegetarian / Vegan Restaurant  0.07
4               Thai Restaurant  0.07


----Repulse Bay----
                         venue  freq
0          Chinese Restaurant  0.22
1  Modern European Restaurant  0.11
2       Australian Restaurant  0.11
3         Balinese Restaurant  0.11
4             Thai Restaurant  0.11


----Sai Wan Ho----
                         venue  freq
0          Chinese Restaurant  0.43
1  Modern European Restaurant  0.14
2           French Restaurant  0.14
3           Hainan Restaurant  0.14
4           Indian Restaurant  0.14


----Shau Kei Wan----
                          venue  freq
0             Chinese Restaurant  0.67
1              Hainan Restaurant  0.33
2           American Restaurant  0.00
3     Middle Eastern Restaurant  0.00
4  Vegetarian / Vegan Restaurant  0.00


----Sheung Wan----
                       venue  freq
0        Seafood Restaurant  0.33
1   Scandinavian Restaurant  0.33
2                Restaurant  0.33
3       American Restaurant  0.00
4  Mediterranean Restaurant  0.00
```

```
----Tai Koo----
                   venue  freq
0    Japanese Restaurant   0.5
1     Chinese Restaurant   0.2
2      Korean Restaurant   0.1
3      French Restaurant   0.1
4  Vietnamese Restaurant   0.1


----Tai Tam----
                           venue  freq
0       Mediterranean Restaurant   0.5
1             Chinese Restaurant   0.5
2            American Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4                Thai Restaurant   0.0


----Wan Chai----
                  venue  freq
0   Chinese Restaurant  0.29
1  Japanese Restaurant  0.19
2   Italian Restaurant  0.14
3      Thai Restaurant  0.10
4    Korean Restaurant  0.05


----Yau Tong----
                           venue  freq
0             Seafood Restaurant   1.0
1            American Restaurant   0.0
2       Mediterranean Restaurant   0.0
3  Vegetarian / Vegan Restaurant   0.0
4                Thai Restaurant   0.0
```

- The results from the K-means clustering show that we can categorize the data into 7 clusters based on the restaurant categories:





- Tai Tam, Chai Wan, Sheung Wan, and Causeway Bay are having a cluster in their neighborhood whereas other clusters are a combination of two or more neighborhoods:
    - Cluster 0 - Tai Tarm (dominated by Chinese restaurants)
    - Cluster 1 - Lei Yue Mun & Yau Tong (dominated by Seafood restaurants)
    - Cluster 2 - Central & Wan Chai (dominated by Chinese restaurant restaurants)
    - Cluster 3 - Chai Wan (dominated by Fast food restaurants)
    - Cluster 4 - North Point, Shau Kei Wan, and Admiralty (dominated by Chinese restaurants)
    - Cluster 5 - Sheung Wan (dominated by Seafood restaurants)
    - Cluster 6 - Causeway Bay (dominated by Vietnamese restaurants)

5. Discussion

5.1. *Observation*

According to the map, Southern districts are away from the other 3 districts which are connected so that already give a geographical decision to owner/investors as the traffic to and from the restaurant, transportation lead time for food resources or delivery may differ.

Also, the southern district is close to several famous beaches such as Deep Water Bay and Repulse Bay. Therefore one can assume the atmosphere around the area is more relaxed and consists of a variety of venue categories such as golf course and park etc. In comparison, the other 3 districts, especially the Central & western districts has more cuisines available but lack other venue types beyond restaurants.

The most popular restaurants or cuisine in Hong Kong Island are Chinese, Japanese and Vietnamese. The density is concentrated in the Admiralty, Central & Western, Wan Chai and North Point areas which are mostly within the Central & Western District. .

5.2. Recommendation

Based on our observations, it is certain that opening either a Chinese or Japanese restaurant will be well received in Hong Kong Island but to avoid heavy competition then maybe the owner or investors should consider somewhere in the Southern district. However considering the venue categories are mostly leisure-focused, there is a possibility that the business will do better in the holidays or weekend when people are travelling from urban areas than on normal weekdays.

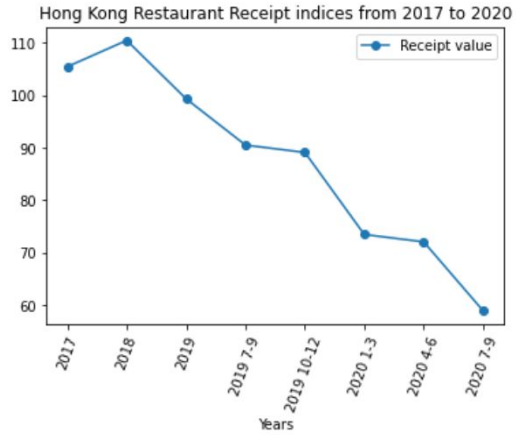
However, at the time of writing this report, we are in the Covid-19 pandemic which globally impacts the health and economies of people. Restaurants are severely affected as they are forced to be closed or not allow dining-in over a certain period or certain duration of the day. Based on the Hong Kong official report [4] from the Census and Statistics Department, the estimated business receipt indices and the restaurant receipts figure, both the value and volume are decreasing. For example for Chinese restaurants, the value in 2020 is less than half of the 2015 value.

表 6.6　食肆的收益
Table 6.6　Restaurant receipts

(甲)　按年及按季數字
(A)　Annual and quarterly figures

年 Year	月 Month	食肆總收益價值 (百萬元) Value of total restaurant receipts ($ million)	食肆收益指數 (2014年10月至2015年9月期內的平均指數=100) Index of restaurant receipts (Average index from October 2014 to September 2015=100)											
			所有食肆 All restaurants		中式餐館 Chinese restaurants		非中式餐館 Non-Chinese restaurants		快餐店 Fast food shops		酒吧 Bars		雜類飲食場所 Miscellaneous eating and drinking places	
			價值 Value	數量 Volume	價值 Value	數量 Volume	價值 Value	數量 Volume	價值 Value	數量 Volume	價值 Value	數量 Volume	價值 Value	數量 Volume
2017		112,737.4	109.0	102.0	105.4	97.9	111.0	105.4	114.5	106.2	106.3	105.8	111.8	103.3
2018		119,555.1	115.6	105.2	110.4	100.0	119.4	110.8	120.4	107.3	113.5	112.7	123.5	110.4
2019		112,447.2	108.7	96.8	99.3	87.9	111.8	101.5	122.6	107.4	103.3	100.6	125.5	108.5
2019	7 - 9	26,318.7	101.8	90.5	90.9	80.3	102.8	93.1	122.0	106.8	91.7	90.3	121.3	104.7
	10 - 12	26,001.3	100.6	89.1	91.6	80.6	101.7	92.1	116.6	101.9	103.2	97.7	114.1	98.2
2020	1 - 3	21,625.4	83.7	73.5	68.3	59.6	92.0	83.0	104.7	90.2	63.9	59.5	101.9	86.7
	4 - 6	21,192.4	82.0	72.1	69.4	60.5	87.8	79.6	98.8	85.1	62.7	59.5	103.7	87.6
	7 - 9 #	17,031.3	65.9	59.0	48.6	43.1	71.0	66.8	93.8	81.0	34.2	32.9	96.0	81.9

Additionally, I made a simple graph for the value of Hong Kong Chinese restaurant business receipt value from 2017 to 2020 below:



5.3. *Challenges*

There are several challenges when completing this project and they are mostly originated from the FourSquare venue API:

A. FourSquare somehow is unable to retrieve the correct coordinates for the Southern Districts. It shows (60.1802277, 24.910466) instead of (22.243200, 114.197400).

B. FourSquare seems to be unable to filter the restaurant category by passing in the "Category ID" for food/restaurant hence I need to use the “str: contains” method as shown in the coding.

C. Need to rework the FourSquare data externally on an excel because:

- Hong Kong does not have postal codes like other cities in the world
- There are smaller neighborhoods within the 4 districts of Hong Kong Island and we need to identify those to perform a better Clustering machine learning:
    - Translating any Chinese word into English
    - Find out the missing neighborhoods (via Google)
    - Specify existing neighborhoods as some are over general e.g. “Hong Kong (香港)” (the city) or “Central area(中區)”

Howard's Gourmet (好酒好蔡)	中環	22.28147	114.1614	Chinese Restaurant
Cheung Hing Kee Shanghai Pan-fried	中環	22.28242	114.1535	Dumpling Restaurant
China Tang (唐人館)	中区	22.28054	114.1577	Chinese Restaurant
Brass Spoon	中環	22.28333	114.1561	Asian Restaurant
Toritama (酉玉)	中環	22.27968	114.1553	Japanese Restaurant
Ronin	中環	22.28453	114.1524	Japanese Restaurant
Samsen (泰館)	香港	22.28508	114.1523	Thai Restaurant
Mak Mak	中区	22.28044	114.1579	Thai Restaurant
Yat Lok Restaurant (一樂燒鵝)	中環	22.28254	114.1554	Cantonese Restaurant
VEA Restaurant and Bar	中環	22.28489	114.153	French Restaurant
Posto Pubblico	中環	22.28174	114.1525	Italian Restaurant
Tate Dining Room & Bar	Sheung Wan	22.28095	114.1527	Restaurant
Chaiwala	香港	22.2809	114.1549	Indian Restaurant
Kinjo's Izakaya	香港	22.28204	114.1523	Japanese Restaurant
PiCi	中環	22.28325	114.1521	Italian Restaurant
Casa Lisboa	中環	22.28139	114.1549	Portuguese Restaurant
Uma Nota	香港	22.28267	114.1528	Brazilian Restaurant
Lung King Heen (龍景軒)	中環	22.28682	114.1574	Cantonese Restaurant
Pololi	中環	22.28284	114.1533	Hawaiian Restaurant
La Vache!	中環	22.28266	114.1527	French Restaurant
Bedu	香港	22.28422	114.1522	Middle Eastern Restaurant
Frantzén's Kitchen by Björn Frantzén	香港	22.28481	114.1482	Scandinavian Restaurant
Soho Banh Mi	中環	22.28095	114.1515	Vietnamese Restaurant

- Owner may fill out the restaurant information different from one another
- Some of the information is missing or only have an editorial difference:
    - Need to re-classify some of the restaurant categories with similar wording such as Dim Sum Restaurant, Cantonese Restaurant, Chinese Restaurant
- The completeness of Hong Kong data may not be sufficient to generate a confident output

6. Conclusion

According to the statistics and visualization of the data shown that opening either a Chinese or Japanese restaurant will be a popular choice. To avoid competition, the southern districts seem to be a good location especially when there are minimal (or none according to the FourSquare API) Japanese restaurants operating around the area. The relevant rental prices are generally lower in the southern district which further supports this conclusion when compared to the Central & Western or the Wan Chai district.

Nonetheless, this is only an assumption based on the available data of this project. Additional investigation of other information is required and consideration of other non-statistical factors will also need to be considered especially when the global economy is defeated by the 2020 Covid-19 pandemic.

7. Reference

[1] https://en.wikipedia.org/wiki/Hong_Kong

[2] https://developer.foursquare.com/developer/

[3]https://developer.here.com/products/geocoding-and-search?Freemium-Google-YT-0-Dev-Generic-E&utm_source=Google&utm_medium=ppc&utm_campaign=Dev_PaidSearch_DevPortal_AlwaysOn&utm_term=&gclid=CjwKCAiAoOz-BRBdEiwAyuvA69fR4gWD07SD3Iofz_QOsZow8QQ-fMIwqacZm1NrMTG0WoshBTdZvBoC1CoQAvD_BwE&gclsrc=aw.ds

[4] https://www.statistics.gov.hk/pub/B10100022020MM12B0100.pdf