The Battle of Neighborhoods - Exploring the possibility of start Oneline Food Delivery or Virtual Restaurant at Hong Kong

1. Introduction and Background of the Business problem:

Hong Kong is one of the most populous metropolitan area in the world, and a lot of top tier multinational companies are choice Hong Kong as APAC Headquarter. As result of average office rents continue to rise in core business areas (e.g. Central district), those companies no long to consider to start their business at core business areas and relocated their office to areas with lower rents.

Darning working day, many office workers in Hong Kong start discussing where to have their lunch and that provide huge opportunities for restaurants in office area. However, those restaurants are always full during the lunch hours (12:00 to 14:00) and office workers usually spend 15 to 30 mins line up for table in restaurant or fast food shop. Therefore, the need for food delivery industry (or Virtual Restaurant) raise.

The food delivery industry in Asia market worth 53 billion US-dollars and which representing more than 50% of global demand. Let's focus in Hong Kong, the consumer are well-known for demanding speed at lunch and which supported the online food delivery segment to reaching 615 million US-dollars for 2019, which grow 9.9% from 2018.

We will concentrate on one most condencsd business district; and we will use data science technology to generate the best location of each areas for stackholders.

Target Audience

1. Business personnel who have interest to invest on Virtual Restaurant. This analysis will focuse on large pools of office workers in Hong Kong during lunch hours.
2. Office worker who want to judge reasonable lunch/breakfast place for physical restaurant or switch to Virtual Restaurant

2. Data Preparation

Based on the defination of business problem, the following factores will help to address problem:

- The number of existing restaurants in the neighborhood
- The distance of restaurant in the neightborhood
- Distance of neighborhood from selected business district

2.1. Data Source

We will use regularly spaced circular grids of selected business district at Hong Kong to define neighborhoods.

- Getting coordinates of selected business district by **Geopy Client**
- List out the number of restaurants in ervery neighborhood by **Foursquare API**
- Collect the Hong Kong district name, pupulation, area and density from **Wikipedia Districts of Hong Kong**

Result set of Districts of Hong Kong

	District	Chinese	Population	Area	Density	Region	latitude	longitude
0	Central and Western	中西區	244600	12.44	19983.92	Hong Kong Island	22.281322	114.160258
1	Eastern	東區	574500	18.56	31217.67	Hong Kong Island	1.178718	38.602580
2	Southern	南區	269200	38.85	6962.68	Hong Kong Island	31.990978	-102.071297
3	Wan Chai	灣仔區	150900	9.83	15300.10	Hong Kong Island	22.279015	114.172483
4	Sham Shui Po	深水埗區	390600	9.35	41529.41	Kowloon	22.328190	114.160854
5	Kowloon City	九龍城區	405400	10.02	40194.70	Kowloon	22.330160	114.189937
6	Kwun Tong	觀塘區	641100	11.27	56779.05	Kowloon	22.312937	114.225610
7	Wong Tai Sin	黃大仙區	426200	9.30	45645.16	Kowloon	22.341654	114.193859
8	Yau Tsim Mong	油尖旺區	318100	6.99	44864.09	Kowloon	22.307404	114.165526
9	Islands	離島區	146900	175.12	825.14	New Territories	33.838992	-96.759999
10	Kwai Tsing	葵青區	507100	23.34	21503.86	New Territories	22.341003	114.104264
11	North	北區	310800	136.61	2220.19	New Territories	64.573154	11.528036
12	Sai Kung	西貢區	448600	129.65	3460.08	New Territories	22.382249	114.272828
13	Sha Tin	沙田區	648200	68.71	9433.85	New Territories	22.381056	114.188879
14	Tai Po	大埔區	307100	136.15	2220.35	New Territories	22.449402	114.171133
15	Tsuen Wan	荃灣區	303600	61.71	4887.38	New Territories	22.371661	114.113470
16	Tuen Mun	屯門區	495900	82.89	5889.38	New Territories	22.390827	113.973169
17	Yuen Long	元朗區	607200	138.46	4297.99	New Territories	22.442646	114.030434

2.2. Data Cleaning

After combine result sets of office building in Hong Kong District and Central Business Districts (CBD) in Hong Kong, we start gathering process of food and restaurant by Foursqure API. However, as of result, we figure out there are 4965 duplicated addresses in result data frame and also there are same city with multiple name; for example: "`Tsim Sha Tsui`", "尖沙咀", "`Jordan`", "佐敦", etc. Therefore I build function which base on the Geopy library to locate the collect address, city and district information by latitude and longitude of the restaurant.

```
Wan Chai            83
Kwun Tong           61
Yau Ma Tei          57
Central District    31
West Kowloon        22
Sheung Wan          20
Tsim Sha Tsui        7
Hong Kong            5
Jordan               4
Ngau Tau Kok         3
Cha Kwo Ling         1
Central              1
Kwung Tong           1
tsim sha tsui        1
Soho                 1
Name: city, dtype: int64
```

Coffee Shop	31
Café	27
Chinese Restaurant	23
Fast Food Restaurant	21
Cha Chaan Teng	17
Cantonese Restaurant	13
Noodle House	11
Japanese Restaurant	10
Hong Kong Restaurant	10
Sandwich Place	8
Italian Restaurant	8
Dim Sum Restaurant	7
Bakery	7
Burger Joint	6
Vietnamese Restaurant	6
Thai Restaurant	6
Seafood Restaurant	6
Dessert Shop	5
Sushi Restaurant	4

3. Methodology

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

As the data collected as above, we can make use of contextualize data/cluster to help us to resolve the problem.

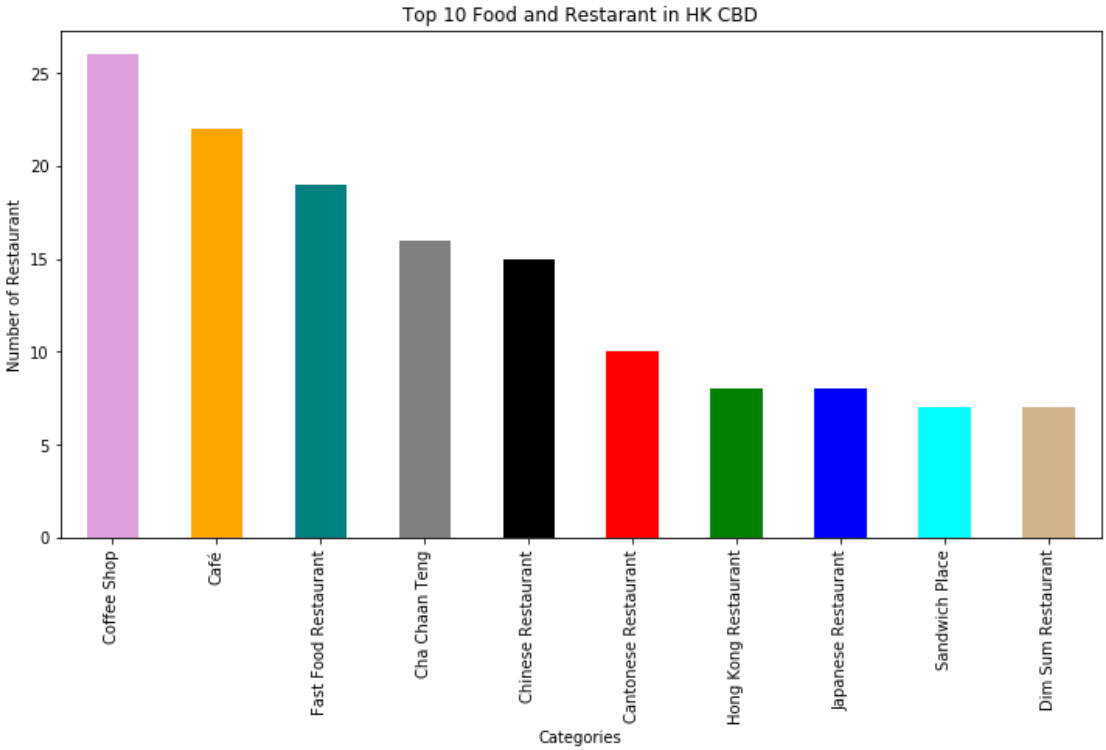
Combined the Foursquare venues data of different category of Hong Kong district, a matrix which captured restaurants information inside each district/city. The weighted matrix able to help to shortlist 5 target locations with venues information to generate the resul that help to select best location to start new virtual restaurant business.

Before building the matrix, I need to prepare the required data and apply some data analysis. Here is the top 10 food and restaurant can find in Hong Kong CBD:

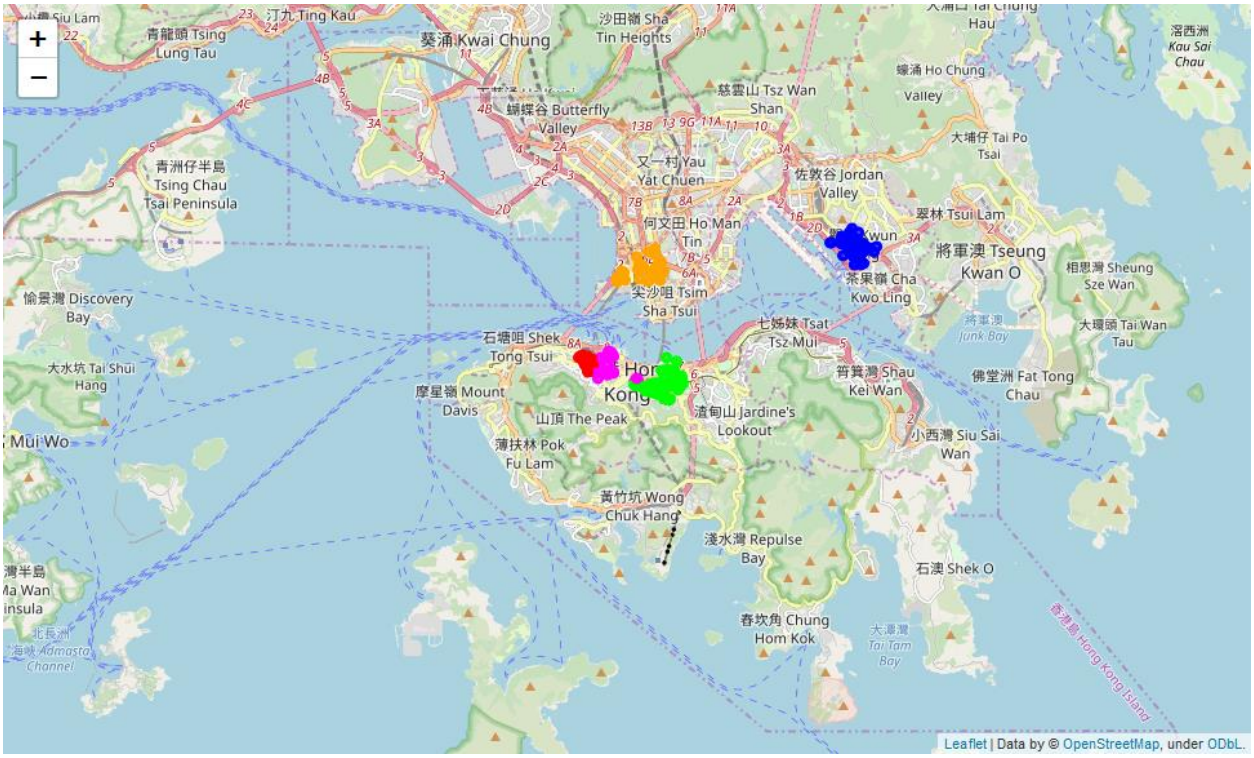
```
Wan Chai            83
Kwun Tong           61
Yau Ma Tei          57
Central District    31
West Kowloon        22
Sheung Wan          20
Tsim Sha Tsui        7
Hong Kong            5
Jordan               4
```

```
Ngau Tau Kok          3
Name: city, dtype: int64
```

Furthermore stockholder would like to focus on this 5 citys (Wan Chai, Central, Kwun Tong, Yau Ma Tei, Sheung Wan) out of top 10 as above, because there are more business offices and factories building. Here is data virtualized chat:



Visualize the data & Leaftlet Map for Top 5 Hong Kong CBD



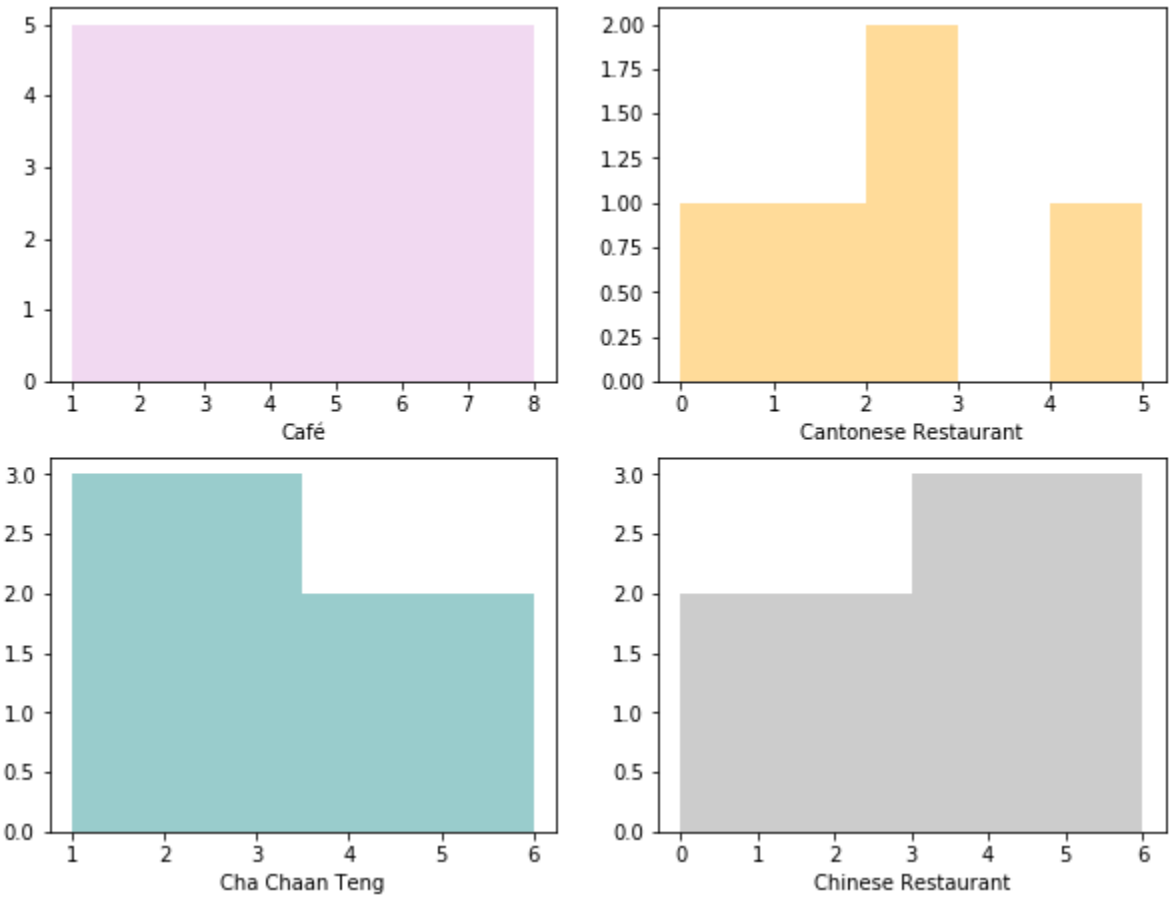
Here is another result set for each category of dataframe

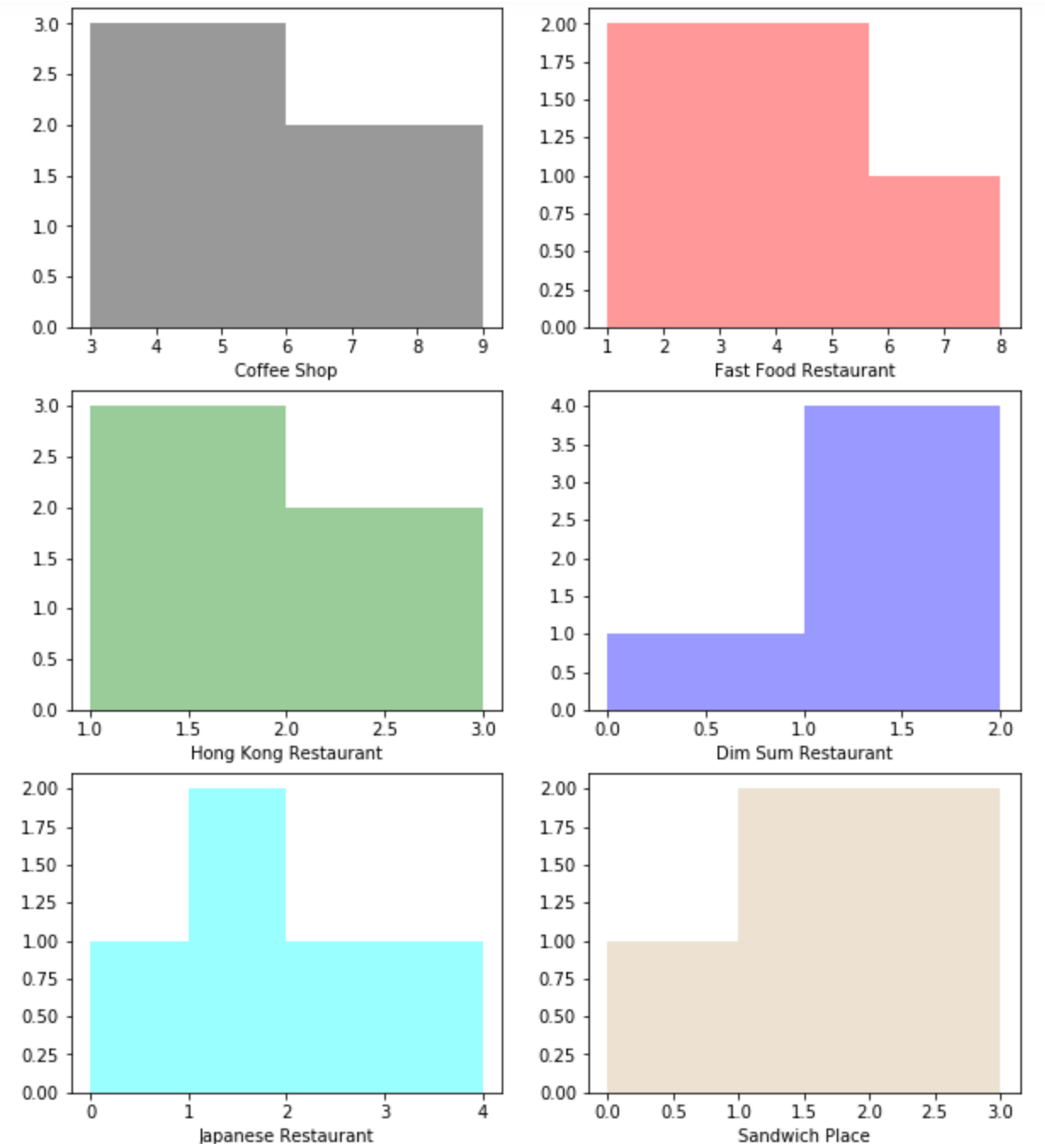
categories city	Café	Cantonese Restaurant	Cha Chaan Teng	Chinese Restaurant	Coffee Shop	Dim Sum Restaurant	Fast Food Restaurant	Hong Kong Restaurant	Japanese Restaurant	Sandwich Place
Central District	1.0	2.0	2.0	1.0	5.0	2.0	2.0	1.0	2.0	2.0
Kwun Tong	7.0	1.0	6.0	3.0	6.0	0.0	8.0	1.0	4.0	1.0
Sheung Wan	1.0	0.0	1.0	0.0	3.0	2.0	1.0	1.0	0.0	0.0
Wan Chai	8.0	5.0	5.0	6.0	9.0	1.0	4.0	3.0	1.0	3.0
Yau Ma Tei	5.0	2.0	2.0	5.0	3.0	2.0	4.0	2.0	1.0	1.0

Here is statistics result for top 10 restaurants

categories	Café	Cantonese Restaurant	Cha Chaan Teng	Chinese Restaurant	Coffee Shop	Dim Sum Restaurant	Fast Food Restaurant	Hong Kong Restaurant	Japanese Restaurant	Sandwich Place
count	5.000000	5.000000	5.000000	5.00000	5.00000	5.000000	5.000000	5.000000	5.000000	5.000000
mean	4.400000	2.000000	3.200000	3.00000	5.20000	1.400000	3.800000	1.600000	1.600000	1.400000
std	3.286335	1.870829	2.167948	2.54951	2.48998	0.894427	2.683282	0.894427	1.516575	1.140175
min	1.000000	0.000000	1.000000	0.00000	3.00000	0.000000	1.000000	1.000000	0.000000	0.000000
25%	1.000000	1.000000	2.000000	1.00000	3.00000	1.000000	2.000000	1.000000	1.000000	1.000000
50%	5.000000	2.000000	2.000000	3.00000	5.00000	2.000000	4.000000	1.000000	1.000000	1.000000
75%	7.000000	2.000000	5.000000	5.00000	6.00000	2.000000	4.000000	2.000000	2.000000	2.000000
max	8.000000	5.000000	6.000000	6.00000	9.00000	2.000000	8.000000	3.000000	4.000000	3.000000

Here is distplot for each restaurant categorifies:

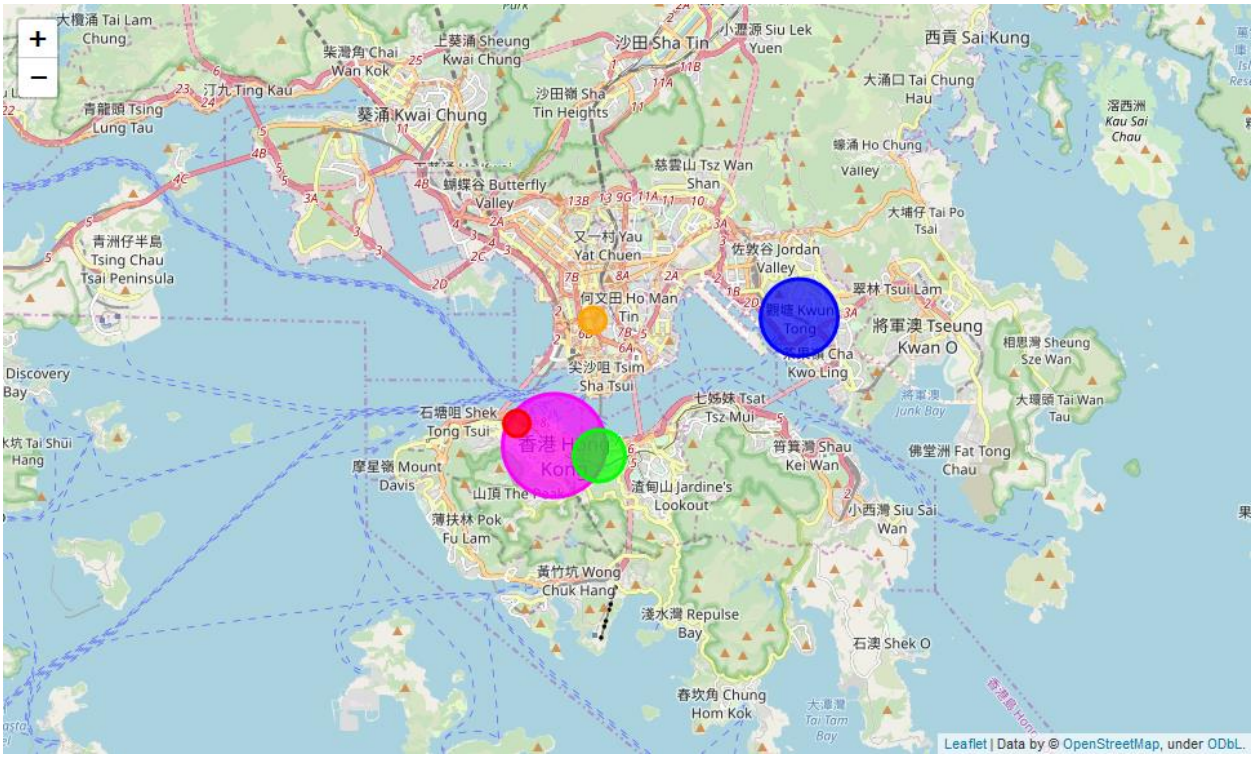




4. Clustering the Central Business District

Afterall, we are ready to cluster 5 selected Central Business District of Hong Kong. And we based on the venue categories and use K-Means clustering. Here is clustering result:

	city	cluster_label	latitude	longitude
0	Central District	4	22.281322	114.160258
1	Kwun Tong	3	22.312937	114.225610
2	Sheung Wan	1	22.286870	114.150267
3	Wan Chai	2	22.279015	114.172483
4	Yau Ma Tei	0	22.312365	114.170779



5 stackholder selected CBD of Hong Kong segmented into 4 clusters based on the most common venues. The size of circles represents relation between food/restaurants and office building for each city. Stockholder can consider locate more resource on ***Central District*** *and* ***Kwun Tong***

5. Results and Discussion

With stackholder interested 5 cities and Food & Restaurants list which nearby the Hong Kong central business district, we successfully clustered it into 4 clusters based on the most common venues. As the above leaflet map showing, we can finalize our recomentation to stakcholder that **Central District** and **Kwun Tong** are the areas they can start their Virtual Restaurants/Food Delivery business. The followings are the finding in this analysis:

- The data corrected from Foursquire API are outdated and untrustworthy, and which spend too many time to perform clean data
- It was very difficult to find data source for this analysis

6. Conclusion

Finally in conclusion state, I am happy to said that the stockholder problem is resolved in this Data Science simulation project. And also we experenced the real-live project of data science, which highly enhanced my python programming skil like web scraping and also experenced on call Foursqure API, geopy.geocoder for geting location information and generate leaflet map to help on data analysis process.

And of course, in real life, final decission will be made by stakeholders which based on more precise data source, and consideraton additional factors by profesional agents.