FINDING OPTIMUM LOCATIONS FOR OPENING A RESTAURANT IN HONG KONG

AMANPREET SINGH

2 June, 2019

1. Introduction

1.1 Background

Hong Kong has emerged as one of the topmost alpha + world cities. Being emerged as one of the crucial center for economic activities, it has secured the seventh position among the largest trading centers of the world. Interestingly, it is not only recognized as a mere financial port but also as a multicultural hub. It is also viewed as a bridge between Asia and America. Apart from these, it is also considered as one of the prime tourist zones. Its colorful traditions, natural scenes and glamorous views magnetize hundreds of thousands of tourists every year.

1.2 Problem

The purpose of this project is to locate some best spots for opening up a Vegetarian Restaurant in Hong Kong. The primary requirement is that the location should not be already busted with other restaurants. It must be noted that Hong Kong is already crowded with hundreds of restaurants which vary over sixty categories. Moreover, location should be near some scenic view point as well as farther from dense residential areas. The availability of nearby raw stuff (i.e. vegetables etc) market in a seafood optimum area would be another issue as well.

And meeting up these requirements in world's fourth densely populated, and multicultural and international business hub, is not that easy. So, we will be using Data Science to hit the target.

1.3 Problem Statement

Finally, in a nutshell, the business problem can be summed up as: Locate the optimum areas for setting up a vegetarian restaurant in the stiff competitive world of Hong Kong.

1.4 Target Audience

This project would interest any individual or group who wants to set up a restaurant or more specifically, a vegetarian restaurant in Hong Kong.

2. Data Requirements

2.1 Data Acquisition

In order to solve our problem, we will require a large amount of data regarding the various neighborhoods of Hong Kong. The details of the neighborhoods of along with their geographical latitudes and longitudes shall be received via following link: https://www.geodatos.net/en/coordinates/hong-kong

A screenshot regarding whereabouts of geographical coordinates of various neighborhoods can be seen Fig 2.1.

We would also require to analyze these neighborhoods for getting better insight. This requirement shall be fulfilled using Foursquare location data which provides smart service for exploring the venues of a particular area.

This tool will not only aid in locating the restaurants in a particular area but also various other scenic or professional venues in that area. Apart from delivering their names, it will also hand over their respective geographical latitudes and longitudes.

An example of the usage of Foursquare location service can be seen in Fig. 2.2.

These geographical coordinates are the most important features which will play the key role in our project.

City	Coordinates
Hong Kong	22.2783203, 114.1746902
Kowloon	22.3166695, 114.1833267
Tsuen Wan	22.3706608, 114.1047897
Yuen Long Kau Hui	22.4500008, 114.0333328
Tung Chung	22.2878304, 113.9424286

Fig. 2.1 Screenshot of Neighborhoods and respective Coordinates of Hong Kong

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hong Kong	22.27832	114.17469	Zahrabel	22.278194	114.175912	Middle Eastern Restaurant
1	Hong Kong	22.27832	114.17469	The Optimist	22.278049	114.175854	Spanish Restaurant
2	Hong Kong	22.27832	114.17469	Kam's Roast Goose (甘牌燒鵝)	22.277647	114.175361	Cantonese Restaurant
3	Hong Kong	22.27832	114.17469	Seorae (喜來稀肉)	22.278280	114.174143	Korean Restaurant
4	Hong Kong	22.27832	114.17469	Sang Kee Seafood Restaurant (生記海鮮飯店)	22.277755	114.172093	Seafood Restaurant

Fig. 2.2 Screenshot of Dataframe using Foursquare location service

	City	Latitude	Longitude
0	Hong Kong	22.2783203	114.1746902
1	Kowloon	22.3166695	114.1833267
2	Tsuen Wan	22.3706608	114.1047897
3	Yuen Long Kau Hui	22.4500008	114.0333328
4	Tung Chung	22.2878304	113.9424286

Fig. 2.3 Screenshot of Dataframe showing Latitudes and Longitudes of various neighborhoods

2.2 Data Cleaning and Data Preparation

Initially, we just had a list of neighborhoods of Hong Kong along with their coordinates. We divided the coordinates column into two columns on basis of latitudes and longitudes for ease of calculations. It can be seen in the Fig 2.3.

Then, using foursquare, we acquired information regarding various venues in these neighborhoods. As explained later, we kept only those records which belonged to top five competitive neighborhoods only.

We also added distance between the venue from the respective central point of that neighborhood and area in which that venue lies with respect to the central point of that neighborhood. These columns were added for analysis purposes and their explanation will be given in the coming sections.

3. Methodology

3.1 General Overview of Data

Now, we have gathered the information regarding the restaurants in Hong Kong. To begin with our analysis, the first thing we need to know is the distribution of these restaurants which has been depicted in the Table 3.1.

It is clear that Hong Kong and Central share a major portion of these restaurants while the remaining neighborhoods barely tip to half of their share. Moreover, about 8 neighborhoods do not contribute even 10 restaurants.

Before proceeding further, it would be wise to have overview of various food venue categories from Fig 3.2. We can see there are over forty types of different restaurants. It must be noted that there are only 7 vegetarian restaurants. So, we have barely any competition with vegetarian restaurants. But as far as other restaurants are concerned, we have a very tough competition. There are several other types of food venues too, but we are not considering them in our analysis.

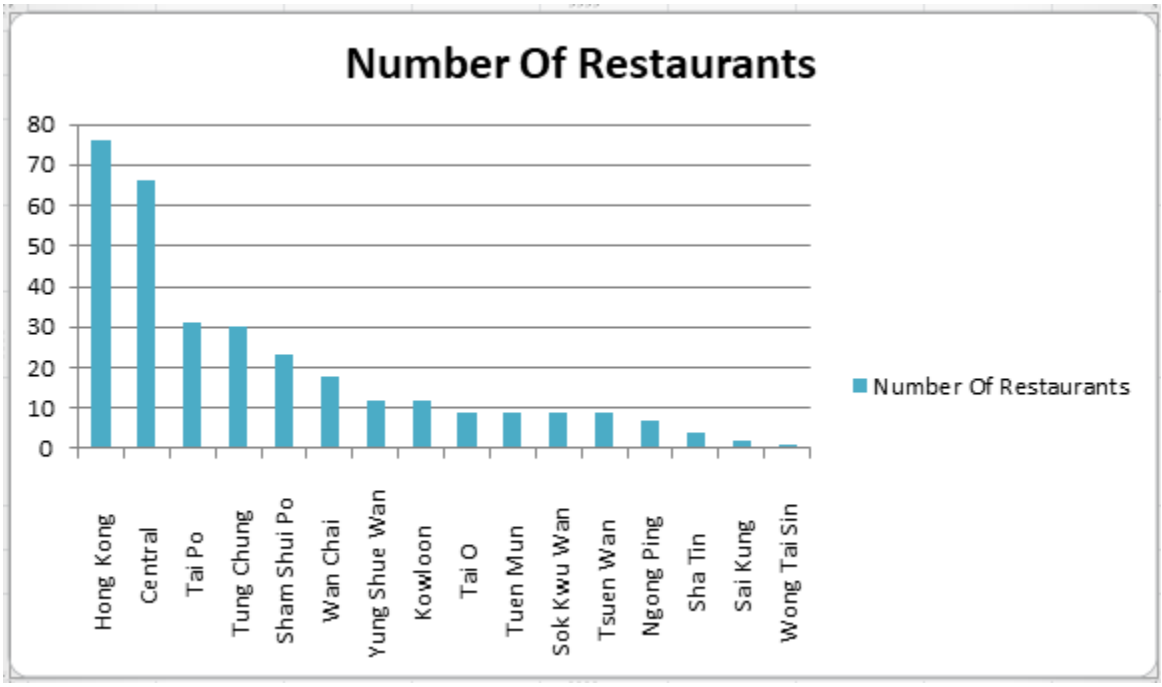

Fig. 3.1 Bar Chart showing number of restaurants in various neighborhoods

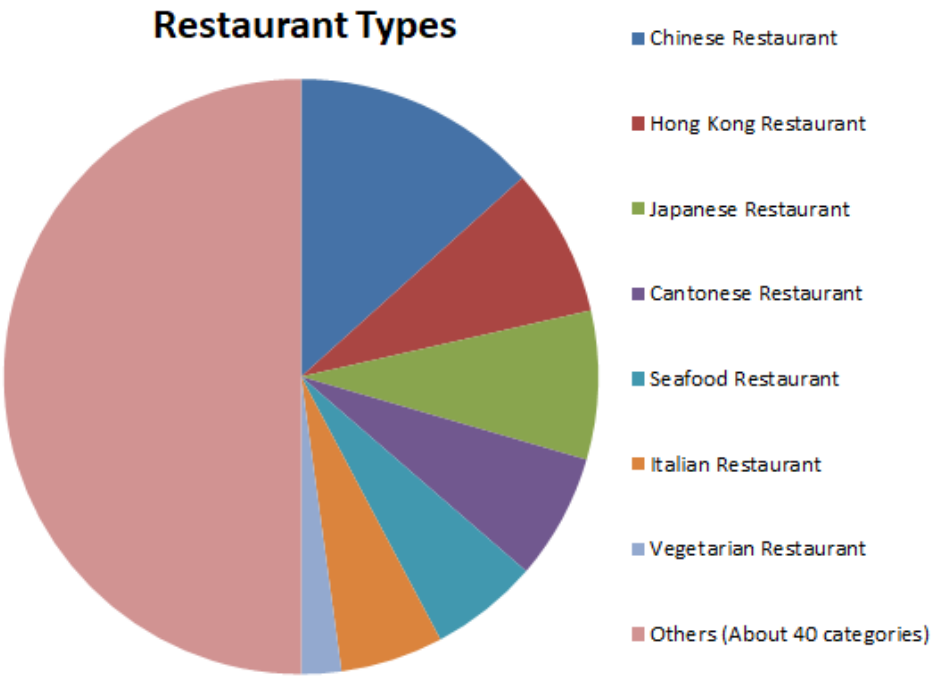

Fig. 3.2 Pie Chart showing various types of restaurants and their contribution

3.2 Identifying Focusing Neighborhoods

We have seen that most of the neighborhoods indeed have very less number of restaurants. To look out for a location to open up a restaurant in these low density neighborhoods, doesn't seems to be a big deal. So, we restricted our analysis to only those neighborhoods where number of restaurants is atleast 20. In this way, we are left with only following five neighborhoods.

Table 3.1 Distribution of Restaurants in Top Five Neighborhoods

Neighborhood	Number Of Restaurants
Hong Kong	76
Central	66
Tai Po	31
Tung Chung	30
Sham Shui Po	23

It is worth mentioning that these neighborhoods contribute 71% of the total restaurants. And we are to find optimum locations in these top competitive neighborhoods.

3.3 Determining the distribution pattern of these neighborhoods

We need to look out for low density areas in five topmost stiff competitive neighborhoods. In order to do that, we need to might need to divide them into smaller sub areas. So, we calculate the distance between our restaurants and the center of their respective neighborhoods. We have used haversine distance formula which calculates distance between two geographical coordinates as follows:

$$d = 2r\, arcsin\left(\sqrt{sin^2\left(\frac{\varphi_2-\varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \quad (3.1)$$

where:

- d is distance between two points
- r is radius of Earth
- φ_1, φ_2 are latitudes of those two points
- λ_1, λ_2 are longitudes of those two points

But unfortunately, the farthest distance between center of the neighborhood and a venue, comes out to be just 497 m. It means that our neighborhoods are indeed very small. So, it would not be a good choice to divide these neighborhoods on the basis of length or area.

Moreover, the following map also depicts that our venues are closely located to each other. While large portions surrounding these central locations seem untouched. It's most probably because those regions might not be that suitable.

Fig. 3.3 Screenshot of Map showing location of restaurants of various neighborhoods

So, we need to locate those regions which are not farther from central point as well as close to some scenic spot. In a nutshell, we are not in a position to divide these neighborhoods further on the basis of area. The areas with which we are left are, already very small. atan2

That's why, we moved to another approach. We determined the direction of our venues with respect to the center of the neighborhood by calculating angle between them. For this we used the following formula:

$$\theta = \text{degree}\left(\text{atan2}\left((y_1 - y_2), (x_1 - x_2)\right)\right) \tag{3.2}$$

where:

- θ is angle between two coordinates
- x_1 is destination latitude
- x_2 is source latitude
- y_1 is destination longitude
- y_2 is source longitude

In this way, we got the relative directions of venues with respect to their respective central points. We can see this distribution in Hong Kong neighborhood from the Table 3.2.

It can be seen that nearly 40% of the restaurants lie in South Western region while North Eastern and North Western regions barely have 10%.

Table 3.2 Distribution of Restaurants with respect to central point of Hong Kong neighborhood

Direction With Respect To Central Point	Number Of Restaurants
SW	31
E	13
W	10
S	9
SE	6
NW	3
NE	2
N	2

The scatter plot in Fig 3.4 also depicts higher density in a particular region. It can be seen in bottom left part of this plot.

Similarly, the distribution patterns for other four neighborhoods can be seen in the form of pie charts from Fig 3.5 to Fig 3.8.

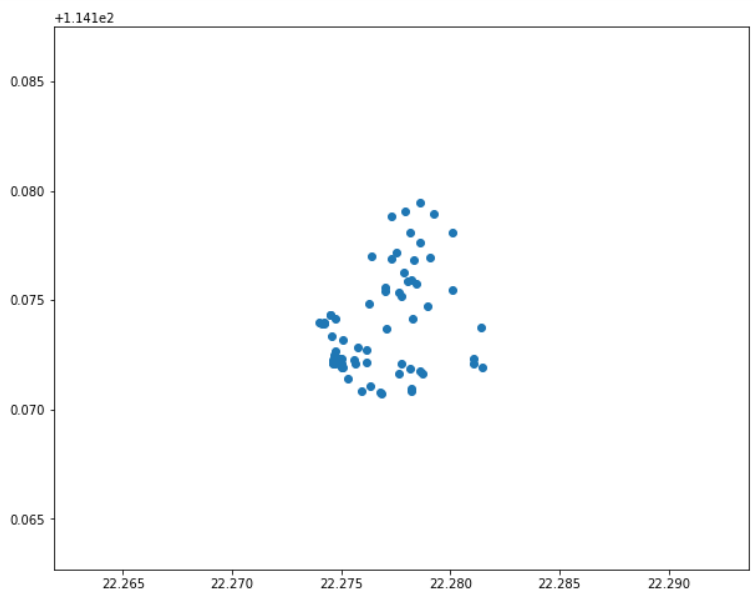

Fig. 3.4 Scatter Plot showing the distribution of restaurants in Hong Kong neighborhood

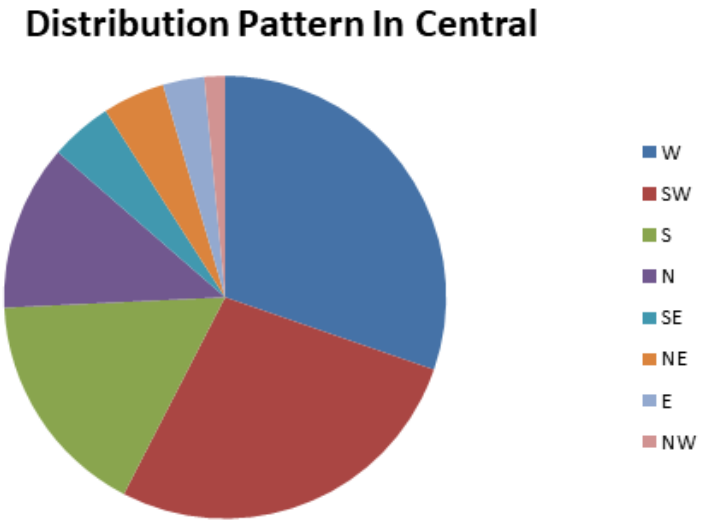

Fig. 3.5 Pie Chart showing Distribution Pattern of Restaurants in Central

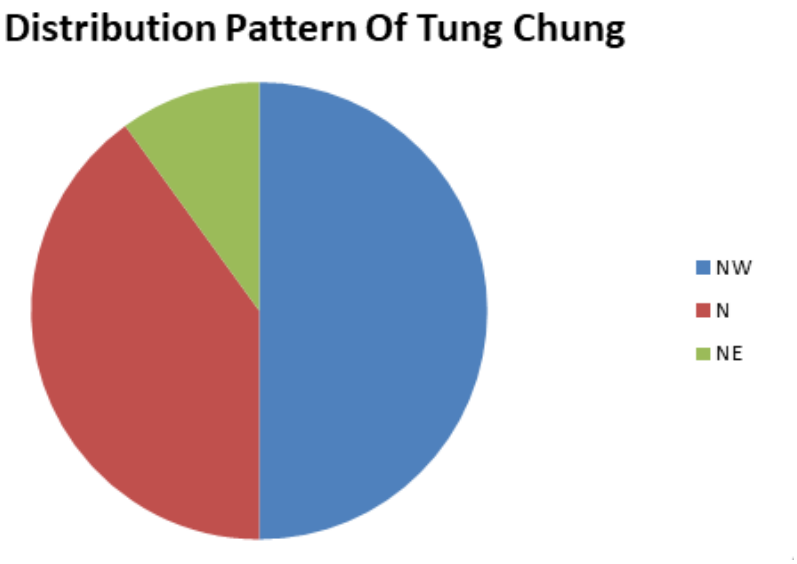

Fig. 3.6 Pie Chart showing Distribution Pattern of Restaurants in Tung Chung

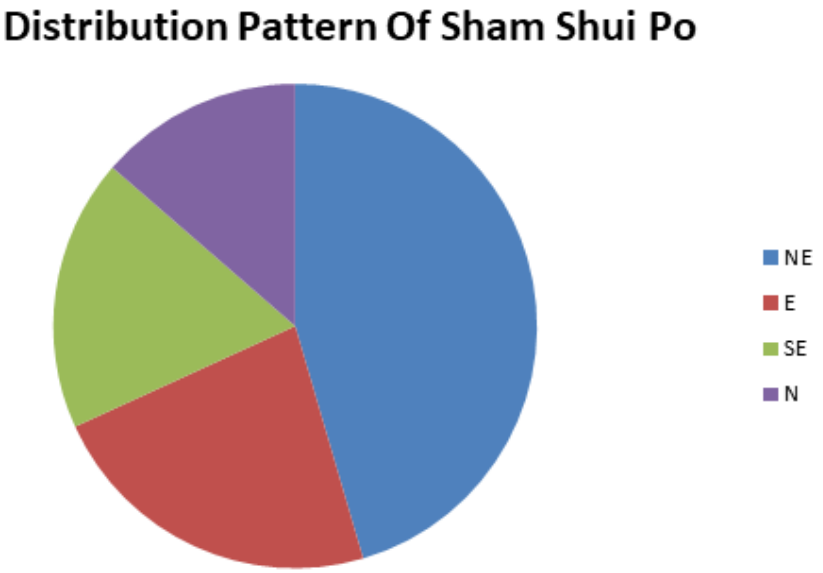

Fig. 3.7 Pie Chart showing Distribution Pattern of Restaurants in Sham Shui Po

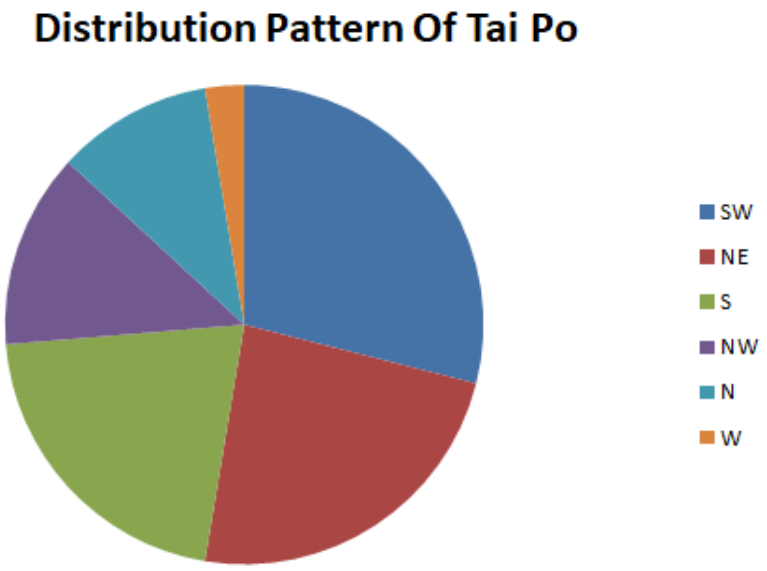

Fig. 3.8 Pie Chart showing Distribution Pattern of Restaurants in Tai Po

3.4 Defining Low Restaurant Density Areas

By studying the direction of locations of the venues with respect to their central points, we can define low density areas for all of our selected neighborhoods. The criteria we have set is, for a particular neighborhood, only those areas would be considered as low density area where number of restaurants is less than or equal to the average number of restaurants in that area.

The table below represents the areas of their respective neighborhoods with low restaurant density.

Table 3.3 Low Restaurant Density Regions

Neighborhood	**Low Restaurant Density**
Hong Kong	W, S, SE, NW, NE, N
Tung Chung	NE
Tai Po	NE, NW, N
Sham Shui Po	E, SE, N
Central	N, SE, NE, E, NW

3.5 Scenes and Parks

The best location for opening up a restaurant is usually near some scenic spot. So, we fetched out all the scenes and parks of these five neighborhoods. In addition, we also defined their location areas in the same manner as we did with the restaurants.

The table below represents the areas of their respective neighborhoods in which scenes and parks lie.

Table 3.4 Regions of Scenes and Parks

Neighborhood	**Scenes/Parks**
Hong Kong	NE, NW, SW, SE
Tung Chung	N
Tai Po	NW
Sham Shui Po	W, S, SE
Central	N, SE, NE, E

3.6 Low Residential Areas

Another aspect to be kept in mind is restaurant should not be in densely populated areas. Though, Hong Kong is pretty crowded, still we need to look out for comparatively scarcely populated areas. The scatter plot in Fig 3.9 shows denser and scarcer regions of Hong Kong neighborhood.

We defined such regions in the same way as we defined low density restaurant areas i.e. the regions where the number of residential venues was less than or equal to the average number of residential units in that area, were considered low density regions.

The table below represents the areas of their respective neighborhoods with low residential density.

Table 3.5 Low Residential Density Regions

Neighborhood	**Low Residence**
Hong Kong	NW, SE, E, NE, W, S
Tung Chung	NE, W, S, SE
Tai Po	NE, N, NW
Sham Shui Po	NW, W, NE, SE
Central	N, SW

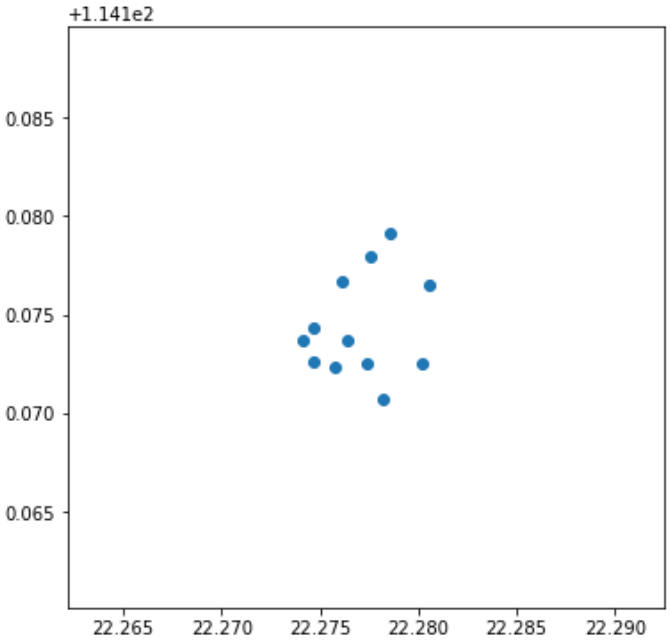

Fig. 3.9 Scatter Plot showing Distribution Pattern of Residential Units in Hong Kong neighborhood

3.7 Optimum Zones

The best location is near some scenic spot, far from crowded residential areas and that too where there are a very few restaurants. We will try to define all those areas where at least two of these three conditions are met.

The table given below illustrates what we have been going through.

Table 3.6 Optimum Regions

Neighborhood	**Low Restaurant Density**	**Low Residence**	**Scenes/Parks**	**Optimum Regions**
Hong Kong	W, S, SE, NW, NE, N	NW, SE, E, NE, W, S	NE, NW, SW, SE	SE, NW, NE, W, S
Tung Chung	NE	NE, W, S, SE	N	NE
Tai Po	NE, NW, N	NE, N, NW	NW	NE, N, NW
Sham Shui Po	E, SE, N	NW, W, NE, SE	W, S, SE	W, SE
Central	N, SE, NE, E, NW	N, SW	N, SE, NE, E	N, E, NE, SE

In the above table, we can see Low Restaurant Density regions, Low Residence regions and Scenes/ Parks regions. We gathered all those regions in which values from any two of these columns were common. In this way, we identified our optimum regions.

So now, we get all the venues of these five neighborhoods using Foursquare. Again, we calculated their respective location areas. But we kept only those venues which lie in Optimum Regions. We simply deleted all those venues which belonged to any non-optimum zone.

In this way, we got the coordinates of all those venues which come under our optimum zones. We assume that these venue coordinates represent the important locations of those zones. We will be using these coordinates for K-means clustering.

We have over 100 venues from optimum zones of these five neighborhoods. It is worth mentioning that among these venues, there is only one vegetarian restaurant as it can be seen in Table 3.7.

Table 3.7 Restaurants in Optimum Regions

Neighborhood	Optimum Regions	Restaurants	Vegetarian
Hong Kong	SE, NW, NE, W, S	25	1
Tung Chung	NE	3	0
Tai Po	NE, N, NW	7	0
Sham Shui Po	W, SE	4	0
Central	N, E, NE, SE	12	0

3.8 K-means clustering

k- means clustering groups out data based on similarity. It can be categorized into unsupervised learning. It divides data into k non-overlapping subsets without any cluster internal structure. It is worth mentioning that points within cluster are similar to each other while the points across different clusters are different.

We will be creating five clusters, one for each neighborhood. These clusters will be created using the geographical coordinates of venues from the optimum zones of these neighborhoods. Since, the area is not much large, we don’t prefer large number of clusters. Large number of clusters may cause overlapping. It must be mentioned that average distance between venue and their respective central point is just 300 meters.

Fig 3.10 illustrates the map showing clusters created using k-means clustering.

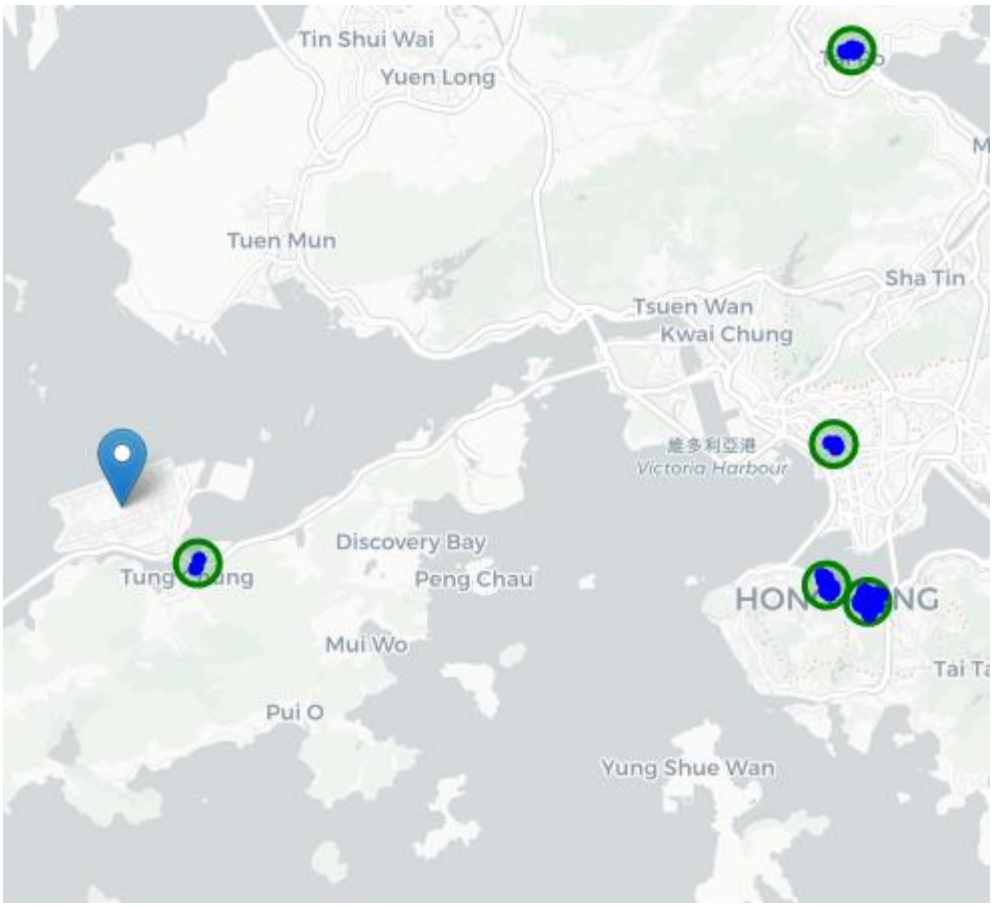

Fig. 3.10 Screenshot of Map showing Clusters created using k-means clustering

4. Result Section

Finally, we are able to define optimum locations for opening up a vegetarian restaurant in five most competitive neighborhoods of Hong Kong.

The table below enlists the cluster centroid coordinates of respective neighborhoods along with their addresses. These addresses were obtained by reverse geocoding tool via following link: https://www.latlong.net/Show-Latitude-Longitude.html

Table 4.1 Cluster Centroids and their respective address

Neighborhood	Center Latitude	Center Longitude	Address
Tung Chung	22.28976767	113.94332793	North Lantau Highway, Hong Kong Island Tung Chung
Central	22.28332206	114.15918593	1 Connaught Place, Central & Western District, Hong Kong Island
Tai Po	22.45266454	114.16960888	On Po Road, Hong Kong Island Tai Po District
Hong Kong	22.27802033	114.1732967	Hong Kong Island Wan Chai District, Wan Chai District, Wan Chai, Hong Kong Island Hong Kong Island Hong Kong
Sham Shui Po	22.32812909	114.16167531	259 Tai Nan Street, Sham Shui Po District

5. Discussion and Future Directions

5.1 Discussion

Hong Kong is one of the most important global centers of the world. Though, it is full of opportunities but one must not ignore the fact that it is stiff competitive place. We emphasized only over restaurants and it seems that Hong Kong is really a very tough market for restaurants.

The results also depict that Chinese or Japanese food is most preferred. There is not much competition for vegetarian restaurant but challenges may arise due to preference of non-vegetarian or sea food or other restaurants.

Hong Kong is a really very crowded area. The facts that almost all the venues are tightly packed and small market areas, created a big challenge. More specifically, Hong Kong and Central neighborhoods were the toughest ones.

But still we located low density areas by distributing the restaurants based on their directions with respect to central point. We also looked out for scenic views and low density residential areas. Our approach was that if we cannot meet all the three conditions, then atleast we should meet any of the two.

With this approach, we succeeded in identifying optimum zones. The small distance between the venues posed a challenge in deciding the optimum number of clusters too. So as to prevent overlapping, we selected five clusters, one for each neighborhood. Their centroids of these clusters were reversely geocoded for getting addresses. These addresses should be the beginning points for exploration.

5.2 Further Directions

This analysis should not be considered as final work, rather, it is more like a beginning work. There can be many ways by which this work can be further tuned. For example, if we are able to collect data regarding price of land or rent prices, we can further improve our results. Similarly, we know that most of the population in Hong Kong belongs to China, or Korean countries, they prefer seafood or non-veg. If we are able to locate any patch where vegetarian population or immigrants reside, it would become a golden option for us. We have considered only scenic spots in our analysis, we can get more results if we included cultural, historical or other landmarks as well.

6. Conclusion

In a nutshell, we have analyzed the various neighborhoods of Hong Kong, especially, in terms of restaurants. There are over 300 restaurants of over 40 different types. Hong Kong and Central neighborhoods alone constitute about 45% of these restaurants. We defined the low density restaurant regions of top five competitive neighborhoods of Hong Kong. Then we looked for some scenic spots as well. We also explored low density residential areas. Finally, we defined optimum regions where atleast two of the above conditions were met and created clusters from the locations of those regions. Not only we created clusters, rather, we generated addresses of the points from where exploration work for restaurant should begin.