# The Battle of Neighborhood

-- Coursera IBM Capstone Project

# Introduction

## Problem Background:

Hong Kong is one of the most populous cities in Asia. It is diverse and is the financial capital of China. It is multicultural with business opportunities provided and business friendly environment. It has attracted many different players into the market as a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance in Asia.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give a good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

## Problem Description:

A restaurant is a business which prepares and serves food and beverages to customers in return for money, either paid before the meal, after the meal, or with an open account. Hong Kong is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's colonial and immigrant history.

1. Chinese immigrants - different types of authentic Chinese cuisines can be found in Hong Kong
2. British immigrants - as a former colony there are a lot of western cuisine
3. South East Asian - relocated in Hong Kong for a few generations and there are a lot of local community here
4. Expats community who works and relocate to Hong Kong - some open restaurants here to provide authentic food choices
5. Hotel or fine dining Michelin starred restaurants
6. Famous local street food

So it is evident that to survive in such a competitive market it is very important to have a strategic plan. Various factors need to be studied in order to decide on the Location such as :

1. Hong Kong Population
2. Hong Kong Demographics
3. The average income of the areas
4. The rental cost of shop in different district
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the district
8. Untapped markets
9. Saturated markets etc

   The list can go on...

ABC Company Ltd., as a traditional Jewish chained restaurant management group with mid to high price range, need to choose the correct location to start its first venture.If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

## Target Audience:

To recommend the correct location, ABC Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood of Hong Kong will be the best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

They are interested in building in an area that meets the following criteria:

- A neighborhood with an average to above average income to the total population
- Above average populations of 25–40-year-old male and female professionals
- Average to above average median net household incomes

With these criteria given by the investing group, based on previous success in other markets, the objective is to locate and recommend to the investors, the target audience, which neighborhood(s) of Hong Kong will be the best choice to start their international growth plan. The information gained will assist in choosing the right location by providing data about the population of each neighborhood, in addition to other established venues present in these areas.

Additionally, this information could be of interest to other potential investors looking to open a new restaurant or entertainment venue in Toronto.

### Success Criteria:

The success criteria of the project will be a good recommendation of Neighborhood choice to ABC Company Ltd based on the lack of such restaurants in that location and the suitable target customer basis who has the purchasing power.

---

# Data

One city will be analysed in this project : ***Hong Kong***.

We will be using the below datasets for analysing Hong Kong

***Data 1 :*** There are 18 districts in Hong Kong. In order to segment the districts and explore them, we will essentially need a dataset that contains the 18 districts that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is :
http://opendata.esrichina.hk/datasets/geodetic-survey-control-stations-trigonometric-stations-in-hong-kong

***Data 2 :*** Working Population Distribution by Income Distribution
http://opendata.esrichina.hk/datasets/working-population-distribution-by-income-distribution-in-2016

***Data 3***: Population Distribution by Quinquennial Age Group
http://opendata.esrichina.hk/datasets/population-distribution-by-quinquennial-age-group-2016-1

***Data 4***: Statistics on Valuation List and Government Rent Roll of Hong
http://opendata.esrichina.hk/datasets/summary-statistics-on-valuation-list-and-government-rent-roll-of-hong-kong/data

***Data 4 :*** For the below analysis we will get data from wikipedia as given below :

1. Population of different district in Hong Kong
   https://en.wikipedia.org/wiki/Districts_of_Hong_Kong
2. Demographics of Hong Kong
   https://en.wikipedia.org/wiki/Hong_Kong#Demographics

***Data 5 :*** Foursquare API to collect information on other venues/competitors in the neighborhoods of Hong Kong

---

# Methodology

In order to establish the targeted districts, we will explore the demographics of the districts in Hong Kong by segmenting the data and conducting descriptive analysis using Panda. Additional data will be gleaned by web scraping and API will be used to generate data.

Data group 1: Demographics data from Wikipedia

1. We scraped the data by using *Beautiful Soup* from Wikipedia https://en.wikipedia.org/wiki/Districts_of_Hong_Kong in order to start understanding how are the districts divided in Hong Kong, and how many of them total.
2. After putting it into *Pandas* dataframe, we can see that Hong Kong is officially divided into 18 districts. From here we can start using district as our unit and find more data

| | District | Chinese | Population | Area | Density | Region |
|---|---|---|---|---|---|---|
| 0 | None | None | None | None | None | None |
| 1 | Central and Western | 中西區 | 244,600 | 12.44 | 19,983.92 | Hong Kong Island\n |
| 2 | Eastern | 東區 | 574,500 | 18.56 | 31,217.67 | Hong Kong Island\n |
| 3 | Southern | 南區 | 269,200 | 38.85 | 6,962.68 | Hong Kong Island\n |
| 4 | Wan Chai | 灣仔區 | 150,900 | 9.83 | 15,300.10 | Hong Kong Island\n |
| 5 | Sham Shui Po | 深水埗區 | 390,600 | 9.35 | 41,529.41 | Kowloon\n |
| 6 | Kowloon City | 九龍城區 | 405,400 | 10.02 | 40,194.70 | Kowloon\n |
| 7 | Kwun Tong | 觀塘區 | 641,100 | 11.27 | 56,779.05 | Kowloon\n |
| 8 | Wong Tai Sin | 黃大仙區 | 426,200 | 9.30 | 45,645.16 | Kowloon\n |
| 9 | Yau Tsim Mong | 油尖旺區 | 318,100 | 6.99 | 44,864.09 | Kowloon\n |
| 10 | Islands | 離島區 | 146,900 | 175.12 | 825.14 | New Territories\n |
| 11 | Kwai Tsing | 葵青區 | 507,100 | 23.34 | 21,503.86 | New Territories\n |
| 12 | North | 北區 | 310,800 | 136.61 | 2,220.19 | New Territories\n |
| 13 | Sai Kung | 西貢區 | 448,600 | 129.65 | 3,460.08 | New Territories\n |
| 14 | Sha Tin | 沙田區 | 648,200 | 68.71 | 9,433.85 | New Territories\n |
| 15 | Tai Po | 大埔區 | 307,100 | 136.15 | 2,220.35 | New Territories\n |
| 16 | Tsuen Wan | 荃灣區 | 303,600 | 61.71 | 4,887.38 | New Territories\n |
| 17 | Tuen Mun | 屯門區 | 495,900 | 82.89 | 5,889.38 | New Territories\n |
| 18 | Yuen Long | 元朗區 | 607,200 | 138.46 | 4,297.99 | New Territories\n |

Table 1. Demographics of Hong Kong

Data group 2: Scoring system - income, age group and shop rent

1. We found the Working Population Distribution by Income Distribution from: http://opendata.esrichina.hk/datasets/working-population-distribution-by-income-distribution-in-2016/data?geometry=113.150%2C22.135%2C115.127%2C22.580
   Then we put it into *Pandas* dataframe for later use

| | English_Name | income_less_than_6000 | F6000_to_less_than_10000 | F10000_to_less_than_20000 | F20000_to_less_than_30000 | equal_or_more_than30000 | Total |
|---|---|---|---|---|---|---|---|
| 0 | WONG TAI SIN | 23511 | 27046 | 99256 | 32044 | 27420 | 209277 |
| 1 | KWAI TSING | 30404 | 36719 | 122595 | 37173 | 32536 | 259427 |
| 2 | SHAM SHUI PO | 28509 | 29595 | 83472 | 27203 | 36786 | 205565 |
| 3 | YAU TSIM MONG | 27612 | 23779 | 63191 | 25263 | 43239 | 183084 |
| 4 | KOWLOON CITY | 38036 | 21940 | 74167 | 31089 | 51992 | 217224 |

Table 2. Working Population Distribution by Income Distribution

2. We extract the spatial distribution of age group data from http://opendata.esrichina.hk/datasets/population-distribution-by-quinquennial-age-group-2016-1
   Then we put it into *Pandas* dataframe for later use.

| | English_Name | Age_0_to_4 | Age_5_to_9 | Age_10_to_14 | Age_15_to_19 | Age_20_to_24 | Age_25_to_29 | Age_30_to_34 | Age_35_to_39 | Age_40_to_44 | Age_45_to_49 | Age_50_to_54 | Age_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | WONG TAI SIN | 14529 | 16767 | 14841 | 20313 | 27154 | 28101 | 30654 | 31169 | 31429 | 32337 | 37869 | |
| 1 | KWAI TSING | 19220 | 19693 | 19388 | 25446 | 31313 | 36107 | 40875 | 40584 | 39885 | 38781 | 44190 | |
| 2 | SHAM SHUI PO | 17104 | 16703 | 13808 | 18432 | 24403 | 28828 | 33246 | 31951 | 31888 | 31597 | 34472 | |
| 3 | YAU TSIM MONG | 13063 | 12873 | 11153 | 15500 | 22642 | 26533 | 27191 | 25855 | 25949 | 25909 | 31943 | |
| 4 | KOWLOON CITY | 17189 | 17609 | 13802 | 18367 | 26325 | 31147 | 34215 | 31973 | 30705 | 30208 | 37240 | |

Table 3. Spatial distribution of age

3. We scrap the shop rent of different district from http://opendata.esrichina.hk/datasets/summary-statistics-on-valuation-list-and-government-rent-roll-of-hong-kong/data
   Then we put it into *Pandas* dataframe for later use.

| | DISTRICT | Shop_Number | 鋪位數量 | Shop_Rateable_Value_000 |
|---|---|---|---|---|
| 0 | WONG TAI SIN | 3427 | 3427 | 2092898 |
| 1 | KWAI TSING | 3888 | 3888 | 2496396 |
| 2 | SHAM SHUI PO | 9430 | 9430 | 4403466 |
| 3 | YAU TSIM MONG | 21099 | 21099 | 21964082 |
| 4 | KOWLOON CITY | 7728 | 7728 | 3252365 |

Table 4. Shop rent of different district

4. Based on the income, age, rent we find above, we can build a scoring system to get a standardized score based the category being divided by its median score and then multiplied by a factor of given importance.

First, we will find the median of different items, including: income, age(3 groups), and rent. According to our client's will, we will prioritize income and age, then rent will be the less important factor. The weight of different factors will be:

- Income = 0.4
- Age = 0.5 (Age 25-29 = 0.15; Age 30-34 = 0.2 ; Age 35-39 = 0.15)
- Rent = 0.1

| | English_Name | income_score | total_score | rent_score |
|---|---|---|---|---|
| 0 | CENTRAL & WESTERN | 1.073166 | 0.301956 | 0.281265 |
| 1 | EASTERN | 1.668297 | 0.615261 | 0.107065 |
| 2 | ISLANDS | 0.423931 | 0.191529 | 0.160306 |
| 3 | KOWLOON CITY | 1.188389 | 0.507146 | 0.088646 |
| 4 | KWAI TSING | 0.743680 | 0.587907 | 0.068042 |

Table 5. Score of different categories

5. Finally, we will add the above 3 scores together to get the final score of the district. The districts which have the greatest total score means they are the most suitable ones based on the criteria from the clients: with people's earning above average, most population between age 25-39, and the relatively cheaper shop rent.

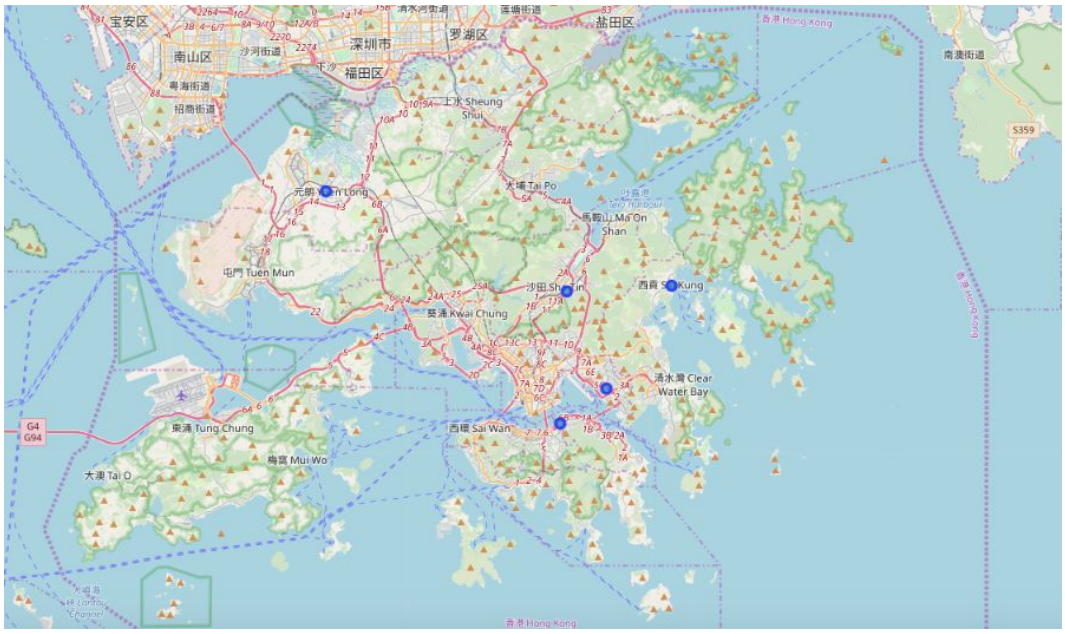
| | English_Name | final_score |
|---|---|---|
| 8 | SHA TIN | 2.435851 |
| 1 | EASTERN | 2.283558 |
| 7 | SAI KUNG | 1.897748 |
| 17 | YUEN LONG | 1.748027 |
| 5 | KWUN TONG | 1.739904 |
| 3 | KOWLOON CITY | 1.695535 |
| 16 | YAU TSIM MONG | 1.420340 |
| 13 | TUEN MUN | 1.377624 |
| 0 | CENTRAL & WESTERN | 1.375122 |
| 4 | KWAI TSING | 1.331587 |
| 9 | SHAM SHUI PO | 1.310211 |
| 12 | TSUEN WAN | 1.291935 |
| 15 | WONG TAI SIN | 1.084293 |
| 11 | TAI PO | 1.055925 |
| 10 | SOUTHERN | 1.050335 |
| 14 | WAN CHAI | 1.017270 |
| 6 | NORTH | 0.854667 |
| 2 | ISLANDS | 0.615461 |

Table 6. Final score of each district

6. After getting the final score, we narrow down our scope by picking the top 5 districts for further investigation, which are:
    a. Sha Tin
    b. Eastern
    c. Sai Kung
    d. Yuen Long
    e. Kwun Tong

<u>Data group 3: Foursquare API and clustering</u>

1. We manually explore the latitude and longitude of the 5 districts that we picked.
2. We put it into a csv file for further use.
3. The geographical data in the csv file is utilized as the input into the Foursquare API, that returned venues from the districts that we selected based on the highest score.




Map 1. Visualization of the district on Hong Kong map

4. As we can see from the table below, here we are able to find the number of restaurants(venue) and the number of categories(type of cuisine) in each district. Later, let's find out more in detail.

| District | District Latitude | District Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| **Eastern** | 90 | 90 | 90 | 90 | 90 | 90 |
| **Kwun Tong** | 87 | 87 | 87 | 87 | 87 | 87 |
| **Sai Kung** | 62 | 62 | 62 | 62 | 62 | 62 |
| **Sha Tin** | 45 | 45 | 45 | 45 | 45 | 45 |
| **Yuen Long** | 51 | 51 | 51 | 51 | 51 | 51 |

Table 7. Number of venue and number of categories in the top 5 districts

5. After that, we can look into district and see what is the top 10 venues in each district and arrange them in order.

| | District | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Eastern | Noodle House | Hotel | Chinese Restaurant | Cantonese Restaurant | Coffee Shop | Sushi Restaurant | Dessert Shop | Italian Restaurant | Café | Korean Restaurant |
| 1 | Kwun Tong | Chinese Restaurant | Fast Food Restaurant | Coffee Shop | Café | Cha Chaan Teng | Noodle House | Hong Kong Restaurant | Japanese Restaurant | Sushi Restaurant | Restaurant |
| 2 | Sai Kung | Café | Seafood Restaurant | Thai Restaurant | Chinese Restaurant | Pub | Dessert Shop | Coffee Shop | Pizza Place | Burger Joint | Park |
| 3 | Sha Tin | Chinese Restaurant | Fast Food Restaurant | Cantonese Restaurant | Train Station | Hotel | Café | Convenience Store | Pizza Place | Playground | Pool |
| 4 | Yuen Long | Chinese Restaurant | Noodle House | Japanese Restaurant | Fast Food Restaurant | Shopping Mall | Dessert Shop | Café | Coffee Shop | Pool | Indian Restaurant |

Table 8. Top 10 venues in each district

---

# Results

Our research shows that the districts have different demographics based on various characteristics. This time, our clients have specify on the class of income and the age range of their target customers, as well as taking shop rent of the area into consideration to find the most suitable location for their first flagship Jewish restaurant in Hong Kong.

First of all, we have scraped the data from the 18 districts in Hong Kong. However, it is not enough and it is hard to compare. Therefore, we have built a scoring system based on the importance of the factors. We have weighting for each factor and multiply the score to the weighting. Finally, we came up with the final score which shows the top 5 most suitable districts to open the restaurant in this order: Sha Tin, Eastern, Sai Kung, Yuen Long, and Kwun Tong.

Based on our analysis using scoring system, we can narrow down our research into 5 districts. The next step is to identify the types of restaurants in the same district to check whether there is potential competitor, and clustering, we found that Sha Tin has the overall best conditions for our potential new flagship restaurant in Hong Kong. Given that among the 18 districts, people who live in Sha tin have above-average monthly income, good amount of young to middle-aged group, also with reasonable shop rent. Moreover, within the district, there is no Jewish restaurant that makes no direct competition to our client. With the suitable target customer and also convenient transportation, we can suggest that Sha Tin could be a good place for the new restaurant.

Last but not least, the analysis does not imply that those zones are actually optimal locations for a new restaurant. The purpose of this analysis was to only provide recommendation based on the specific criteria of the client. Recommended districts should therefore be considered only as a starting point for a more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met. For example, the ratio of local/expat community, the education level of the district, and

---

# Discussion

From the results discovered and presented, the following observations and recommendations can be made:

1. When opening a foreign cuisine restaurant, we may want to take more factors into consideration, such as the ratio of local/expat community, the level of education, the demographics of population(people are single or have a family?). They are factors that reflect people's open mindedness to new type of cuisine.
2. The second recommendation is that we can also think of the distance from the central business district(CBD). Since the working class in CBD is normally those with higher purchasing power. That will be a good indication as well.

---

# Conclusion

The purpose of this project was to identify which area in Hong Kong will be the best choice for our client to open their first Jewish restaurant. Since they already set the criteria of target customers based on income and age range, and taking the shop rent into consideration, we have narrowed down the district based on the scoring system as the first step.

In order to aid our client to look for the optimal location for a new Jewish restaurant, we categorized the restaurant type from Foursquare data so we can identify whether there is any direct competitor in the same district. Clustering helped us to see the density of the type of the restaurant to come up with the final conclusion.

Final decision on optimal location of the Jewish restaurant location will be made by stakeholders based on specific characteristics of district and locations in every recommended zone, taking into consideration additional factors like convenience of each location (proximity to station), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every district etc.