

## **Introduction**

### **Background**

Mainland China remains gripped by a public health scandal in which a major drug maker was found to have supplied inferior vaccines for babies. According to a mainland drug watchdog, Jilin-based manufacturer Changsheng Bio-tech sold substandard DPT (diphtheria, pertussis and tetanus) vaccines to the Shandong Disease Prevention and Control Centre, the agency in charge of public health in a province of about 100 million people.

In Hong Kong, all vaccines for children are imported from foreign brands. Private clinics usually see about 20 per cent of their patients come from across the border to avoid jabs made on the mainland. The influx of Chinese consumers is generating much higher demand for the clinics' information in Hong Kong.

### **Business Problem**

The Mainland Chinese consumers usually are not familiar with the clinic locations in Hong Kong. Even though they might be able to search these locations on Internet, it is difficult for them to decide which clinic location is suitable for their stay. Like many other tourists, the stay in Hong Kong often needs information around the clinic, for instance, the number of hotels, restaurants, supermarkets, and even the shopping areas and entertainment to fulfil their necessities in Hong Kong.

The project is going to locate all the clinics in Hong Kong and explore the locations nearby, including the interests of these categories, to recommend some most suitable clinic location for these Mainland Chinese consumers to stay in Hong Kong for taking vaccines.

### **Interested Stakeholders**

The report is targeting for the parents from Mainland China or foreign countries that are willing to take their children in staying Hong Kong for vaccinations. The recommended clinic location would be beneficial for their consideration in staying which hotels nearby.

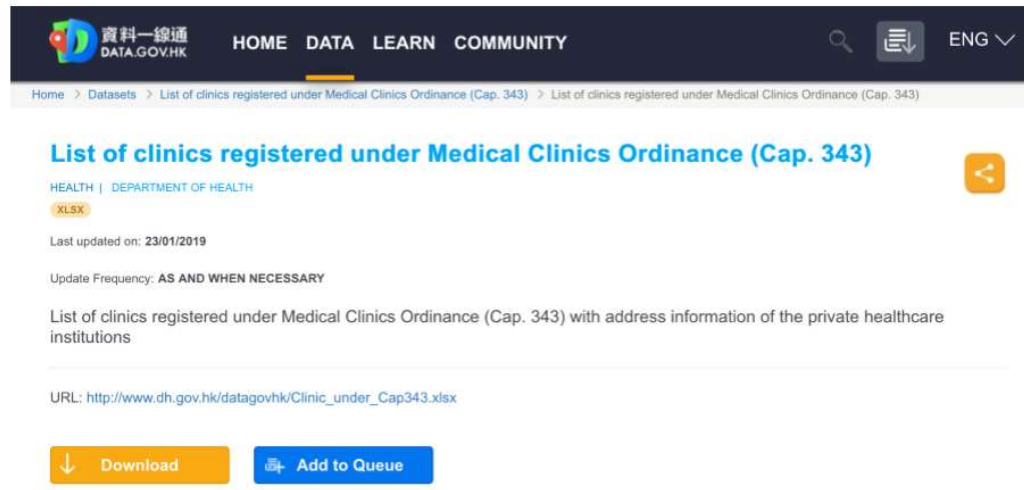
## **Data acquisition**

### **Data Source**

Hong Kong Government recently has an open data initiative (<https://data.gov.hk>) to publicly share government data and statistics for the public to use all datasets. The datasets can be grouped into different categories, including city management, climate, commerce,

development, education, employment, environment, finance, food, health, IT, law, population, recreation, social welfare and transport.

Regarding to the current business problem, the list of clinics registered dataset will be used for analysis. The dataset is useful as it consists of the geospatial data for each clinic location.



The screenshot shows the Data.gov.hk website interface. At the top, there is a navigation bar with the logo '資料一線通 DATA.GOV.HK' and links for HOME, DATA, LEARN, and COMMUNITY. Below the navigation bar, a breadcrumb trail indicates the path: Home > Datasets > List of clinics registered under Medical Clinics Ordinance (Cap. 343) > List of clinics registered under Medical Clinics Ordinance (Cap. 343). The main heading is 'List of clinics registered under Medical Clinics Ordinance (Cap. 343)' in blue. Below this, there is a sub-heading 'HEALTH | DEPARTMENT OF HEALTH' and a small 'XLSX' icon. The text 'Last updated on: 23/01/2019' and 'Update Frequency: AS AND WHEN NECESSARY' are displayed. A description states: 'List of clinics registered under Medical Clinics Ordinance (Cap. 343) with address information of the private healthcare institutions'. The URL 'http://www.dh.gov.hk/datagovhk/Clinic\_under\_Cap343.xlsx' is provided. At the bottom, there are two buttons: 'Download' (orange) and 'Add to Queue' (blue).

[https://data.gov.hk/en-data/dataset/hk-dh-dh\\_hq-dh-orphf-mc/resource/b7ef1714-1b03-4e6a-8b36-1d4be4bf6924](https://data.gov.hk/en-data/dataset/hk-dh-dh_hq-dh-orphf-mc/resource/b7ef1714-1b03-4e6a-8b36-1d4be4bf6924)

## Location Data

FourSquare location API is used to explore the nearby locations of the clinics. The result set consists of location information and predefined categories for analytics. However, the diversified categories from FourSquare become difficult to obtain the overall categories for decision making. Hence, the result set should be cleansed with appropriate grouping of categories for easier reporting and sorting.

## Methodology

### Understanding of clinic data

The dataset was imported into a Notebook (Watson Studio is used to run the notebook) to understand the data. Since the dataset consists of geospatial location of the clinics, the latitudes and longitudes of the clinics are associated with the clinic names and addresses.

The first 5 records of the clinic location are listed here for illustration.

Name of Clinic	Address	Latitude	Longitude
Assemblies of God Medical Clinic	No. 11-12, G/F, Suet Fung House, Fung Tak Esta...	22.344658	114.199971

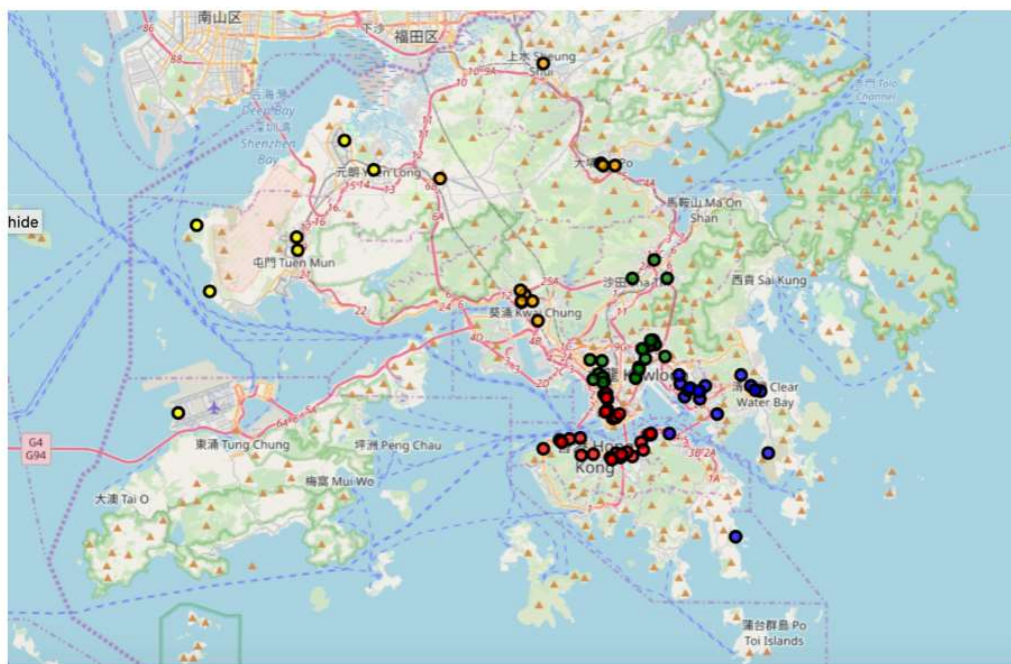
Name of Clinic	Address	Latitude	Longitude
BOCHK Health Care Division	Room 1460, Bank of China Tower, 1 Garden Road,...	22.279322	114.161557
Caritas Clinic – Caine Road	Medical Clinic - Room 032, G/F, Caritas House,...	22.278853	114.154302
Caritas Clinic – Ngau Tau Kok	Room 1, G/F, 1 On Tak Road, Ngau Tau Kok, Kowloon	22.320880	114.216369
Caritas Jockey Club Tsuen Wan Clinic	Room 3, G/F, Block B, 146 Chung On Street, Tsu...	22.368401	114.116045

The dataset looks good for analysis but it also requires data cleansing because there are duplicate records in the table and the headers are repeated in the middle of the dataset. In addition, from the FourSquare location data, there are clinics that return no result set. It is because FourSquare location data may not be prevalent data source in Hong Kong such as some remote clinic locations becomes inadequate for location-based analysis.

### Data Visualization

Dataset with latitudes and longitudes would be good for Folium Map for visualization. To cope with the distinct districts in Hong Kong, the K-Mean clustering of 5 clusters is used for analysis.

The clustering results is shown in the Folium Map below and it nearly matches with the distinct districts in Hong Kong (Hong Kong Island, Kowloon East, Kowloon West, East New Territories East and New Territories West). From the map, the red, green and blue dots are densely shown because these districts are the central districts in Hong Kong which is naturally more clinics to server the patients and consumers.



## Recommendation Approach

Each clinic data can be fitted into FourSquare location API to explore the nearby locations. To obtain a useful result set, a radius of 1000 meters is used. It is because consumers usually prefer to have a walkable distance from the hotels to the clinics for daily vaccination (if necessary). 1 kilometer is usually within 10-minute walk which is generally acceptable.

A total of 100 return items are used. There is a categories column in the result set but it is far too diversified, for example, Japanese restaurant and Dumpling restaurant as two different categories. It will not be effective for analysis and reporting. Hence, a re-grouping of the result set is necessary to perform better use. From the consumer's interest, 5 categories are summarized as entertainment, hotels, restaurants, shops and supermarket. These are usually the necessity for tourists' consideration to go for which clinic location in Hong Kong.

A weighting of the categories to calculate the overall score is necessary because the primary concerns of a clinic location should be the nearby hotels. And then the restaurants and supermarkets are the daily necessities while the entertainment and shops would be the least importance from the view of the tourists that visiting Hong Kong mainly for their children's vaccination in the clinics.

## Results

The top five clinics are listed in the following table and these are the recommended clinic location for the Mainland Chinese consumers to stay in Hong Kong.

Name	Entertainment	Hotels	Restaurants	Shops	Supermarket	Score
The Islamic Union Medical Clinic	4	5	73	8	2	262
Chung Sing Benevolent Society Clinic	4	3	72	14	2	255
Hong Kong Red Swastika Society Free Clinic	2	7	66	14	1	252
Hong Kong Breast Cancer Foundation Breast Heal...	2	8	65	14	0	251
HKFTU Medical & Healthcare Centre (Causeway Ba...	4	4	66	20	2	248

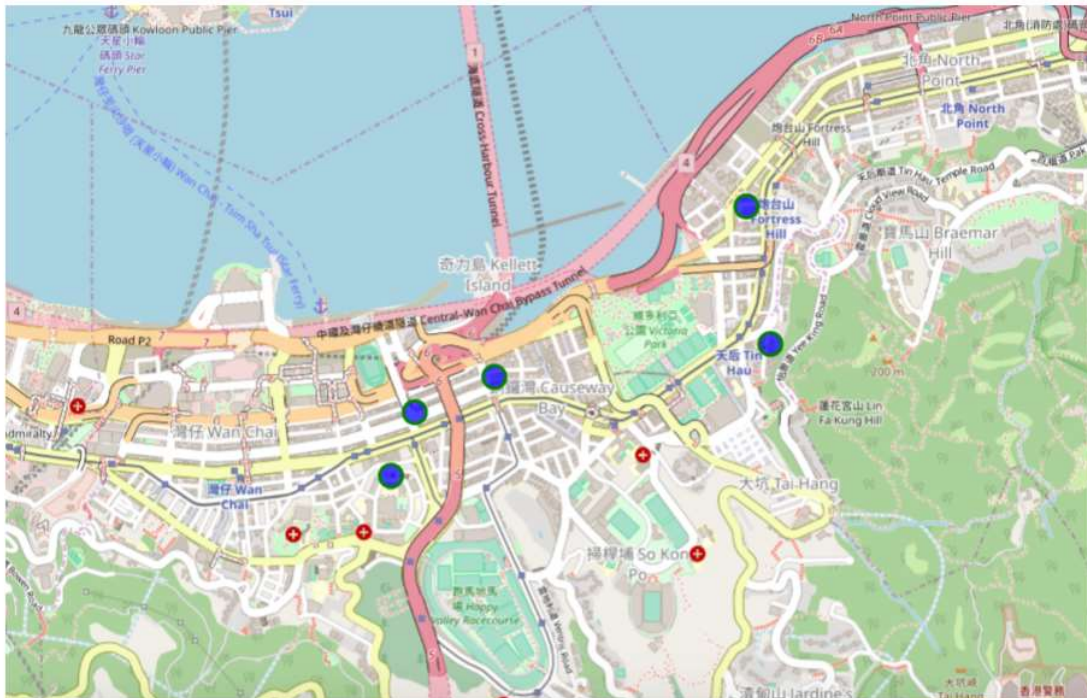
## Discussion

The top five clinics are shown in the Folium Map as shown below. Further looking into the locations, all these clinics are located at the center of the Hong Kong Island which are also the high dense population districts in Hong Kong (namely, Causeway Bay, Tin Hau and Fortress Hill). The analyzed result is matched with the Hong Kong situation that most of the tourists are preferred to stay in these districts for sight-seeing and leisure.



The recommendation is generally good to match with the clustering result in the first data visualization of the clinics. Going back to the first map that all red dots are located in these districts. Hence, in order to provide quick recommendation without further data analysis, all red dots in the first map can be used as general recommendation for the clinics to stay in Hong Kong for Mainland Chinese consumers.

Furthermore, as the result set is grouped into different categories, namely entertainment, hotels, restaurants, shops and supermarket. If a specific group of tourists prefer more entertainment nearby or shops nearby, the weighting parameter can be easily adjusted to reveal a more custom scoring for recommendation.



## Conclusion

In this analysis, the business problem of the influx of Mainland Chinese consumers into Hong Kong clinics will require a data analysis of the clinic locations with nearby interested necessities to provide recommendation. Top five clinic locations are suggested and all are matched with the Hong Kong situation that located in the center of Hong Kong Island.

A further analysis can be made to include the transportation data in FourSquare API or using Google API (which is more prevalent in Hong Kong). The data with metro stations and bus stops are also key factors for the tourists to stay in the appropriate location in Hong Kong.