



Applied Data Science Capstone Project — Strategic Clustering for Opening a Cafe in Hong Kong

Cara Lo

1. Introduction

- ◇ Hong Kong is famous for being a 'busy city'
- ◇ Limitless choices on café
- ◇ Strategic location would be vital for a great start of a new café



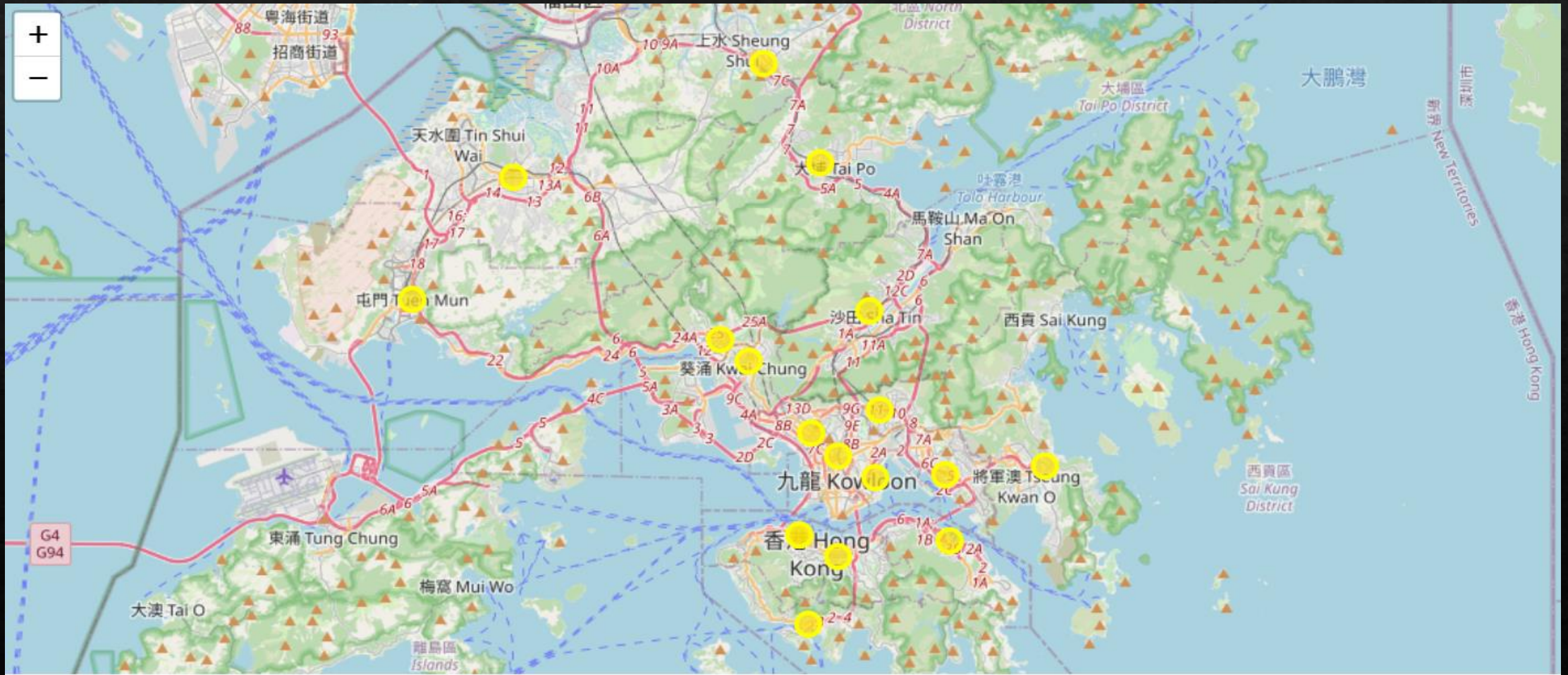
This Photo by Unknown Author is licensed under [CC BY-SA](#)

2.Data

◇ The following data is used to reach the solution:

1. List of districts in Hong Kong (Source: HK GeoSpatial Open Data)
2. Population of districts (Source: Census and Statistics Department)
3. Latitude and longitude of districts (Source: HK GeoSpatial Open Data)
4. Median monthly household income of districts (Source: Census and Statistics Department)
5. Venue data (Source: Foursquare API)

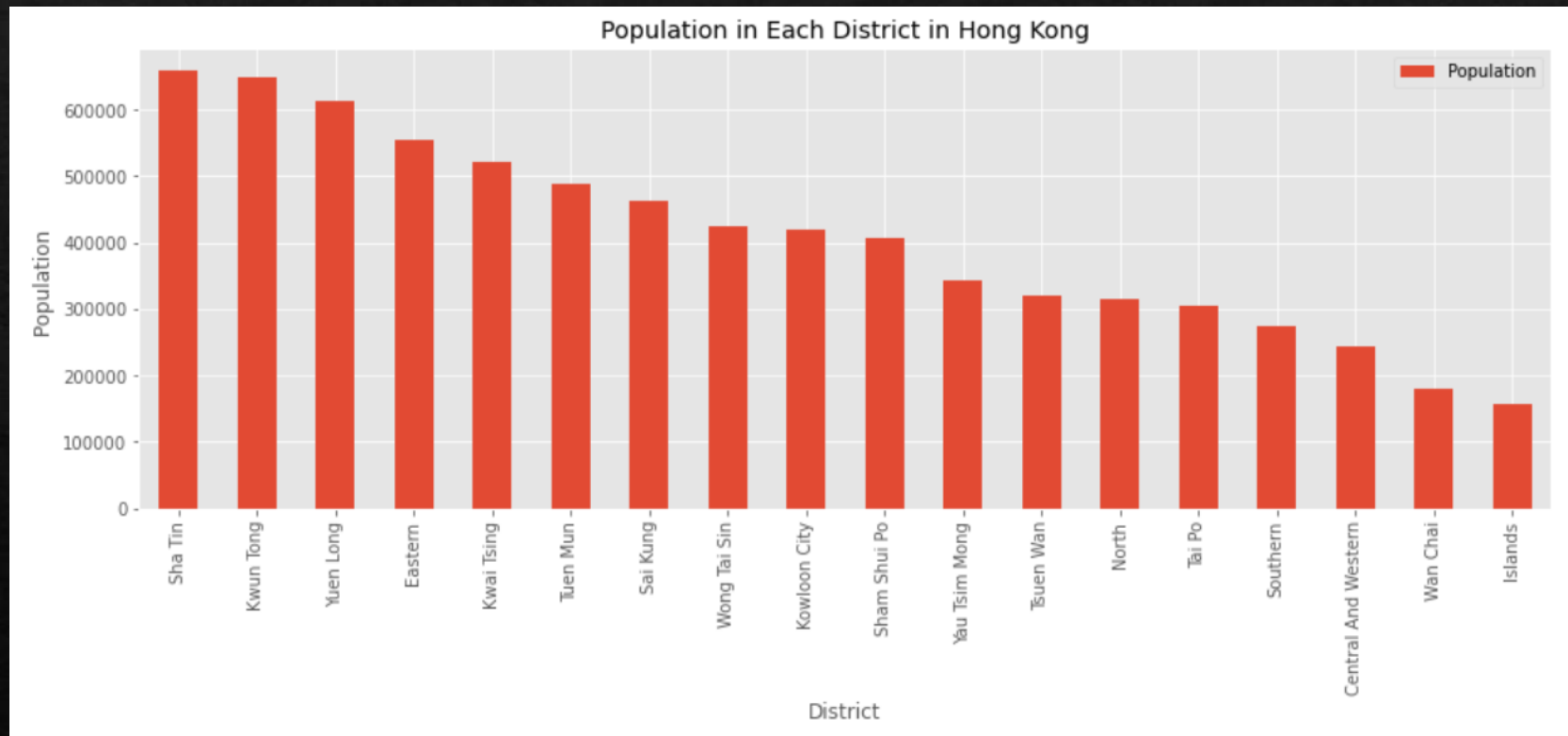
3. Methodology - 3.1 Folium Map



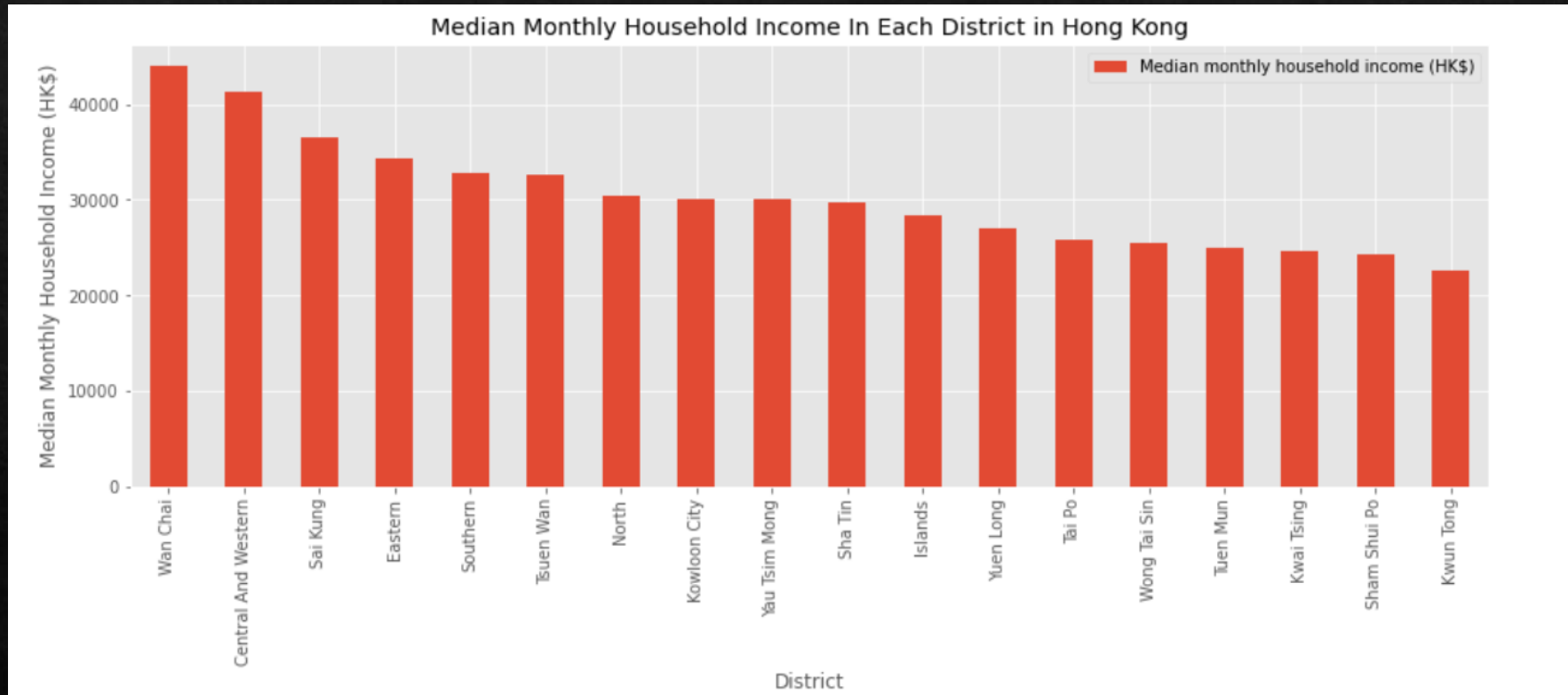
3.2 Foursquare API

- ◆ Foursquare API allows us to explore different venues, including restaurants, shopping malls and hotels etc. in each district. We are able to find out the number of cafés in each district by using Foursquare API.

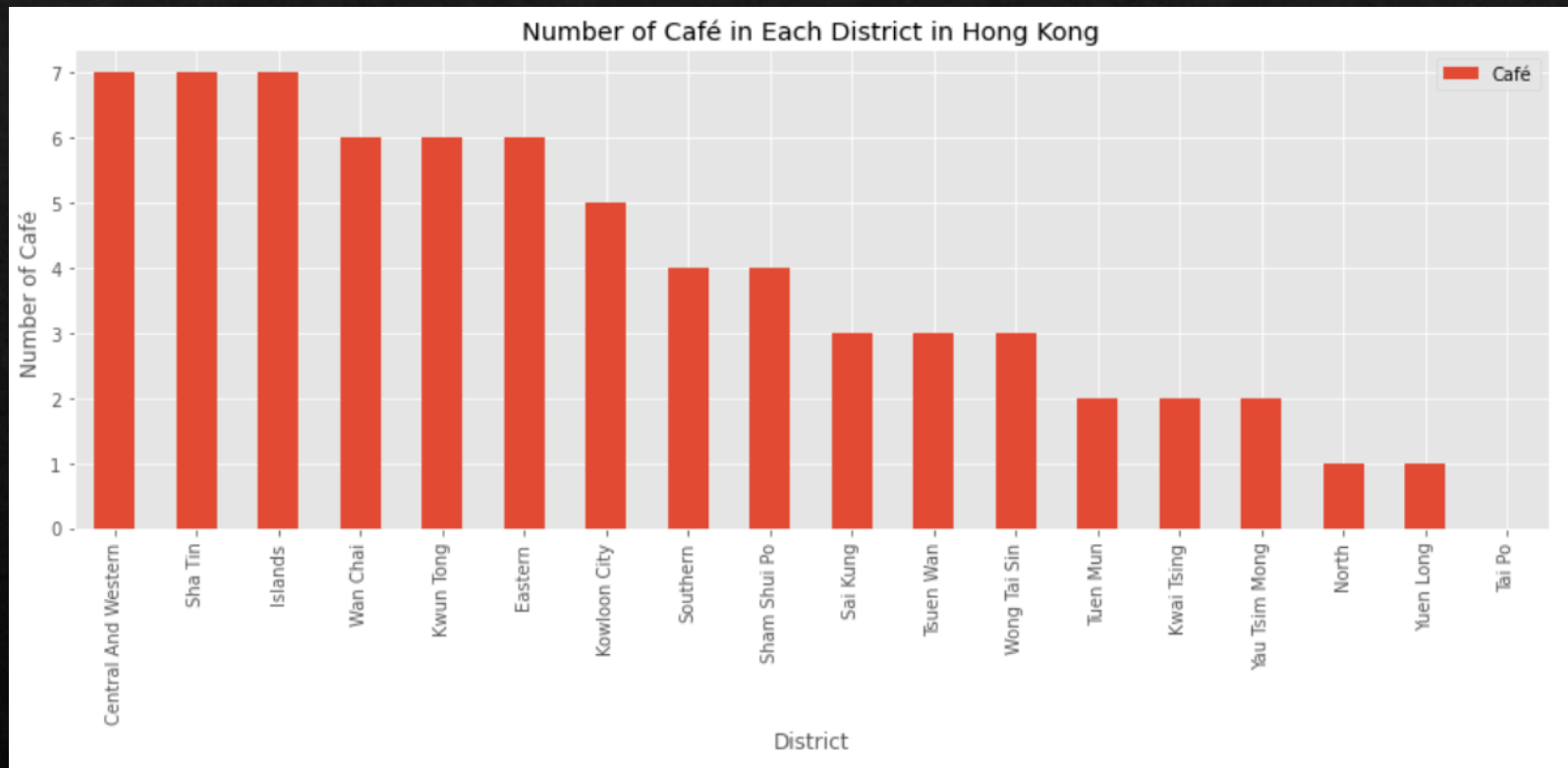
3.3 Bar Graph



3.3 Bar Graph



3.3 Bar Graph



3.4 K-Means Clustering

- ◆ K-means clustering, a machine learning algorithm used to create K clusters of data points based on feature similarity. In this project, the number of clusters is set to 3. Each district is then assigned with a cluster Label representing the cluster number (either 0 or 1 or 2).

ObjectID	District	Cluster Label	Population	Median monthly household income (HK\$)	Café
1	Central And Western	1	243266	41400	7
2	Eastern	2	555034	34300	6
3	Southern	1	274994	32800	4
4	Wan Chai	1	180123	44100	6
5	Sham Shui Po	0	405869	24300	4

4.Results

Cluster Label 0

- ◆ Med-Low Population
- ◆ Low Household Income
- ◆ Low Number Of Cafes

ObjectID	District	Cluster Label	Population	Median monthly household income (HK\$)	Café
5	Sham Shui Po	0	405869	24300	4
6	Kowloon City	0	418732	30000	5
8	Wong Tai Sin	0	425235	25500	3
11	Kwai Tsing	0	520572	24700	2
13	Sai Kung	0	461864	36500	3
17	Tuen Mun	0	489299	25000	2

4. Results

Cluster Label 1

- ◆ Med-Low Population
- ◆ High-Med Household Income
- ◆ High-Med Number Of Cafes

ObjectID	District	Cluster Label	Population	Median monthly household income (HK\$)	Café
1	Central And Western	1	243266	41400	7
3	Southern	1	274994	32800	4
4	Wan Chai	1	180123	44100	6
9	Yau Tsim Mong	1	342970	30000	2
10	Islands	1	156801	28400	7
12	North	1	315270	30400	1
15	Tai Po	1	303926	25800	0
16	Tsuen Wan	1	318916	32600	3

4. Results

Cluster Label 2

- ◆ High Population
- ◆ Med-Low Household Income
- ◆ Med Number Of Cafes

ObjectID	District	Cluster Label	Population	Median monthly household income (HK\$)	Café
2	Eastern	2	555034	34300	6
7	Kwun Tong	2	648541	22500	6
14	Sha Tin	2	659794	29700	7
18	Yuen Long	2	614178	27000	1

5. Recommendation & Conclusion

- ◆ Factors includes size of customer base, spending power and number of competitors
- ◆ Districts in cluster label 2 are recommended
- ◆ Eastern, Kwun Tong, Sha Tin and Yuen Long
- ◆ A large population, a moderate income and not many competitors
- ◆ Suitable for a new business to start