

The background of the slide is a black and white photograph of a modern building. On the right side, there is a large glass window or door with a grid pattern. On the left side, a staircase with a glass railing is visible, leading upwards. The overall aesthetic is clean and architectural.

Coursera Capstone

IBM Applied Data Science Capstone

Opening a coffee shop in Hong Kong, China, Asia

Business Problem

- The objective of this capstone project is to analyze the best location to start up a Coffee shop in Hong Kong, China.
- Using data science methodology and clustering by machine learning, this project would provide a suggestion to answer the business question: Where would be the recommended location to invest a coffee shop in Hong Kong, China.
- This project is particularly useful to anyone who would invest a coffee shop or plan to expend a coffee shop in city Hong Kong, where having a coffee craze.

Data

- List of neighborhoods in Hong Kong.

This defines the scope of this project which is confined to the city of Hong Kong, an international market of China in Asia.

- Latitude and longitude coordinates of those neighborhoods.

This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to coffee shop.

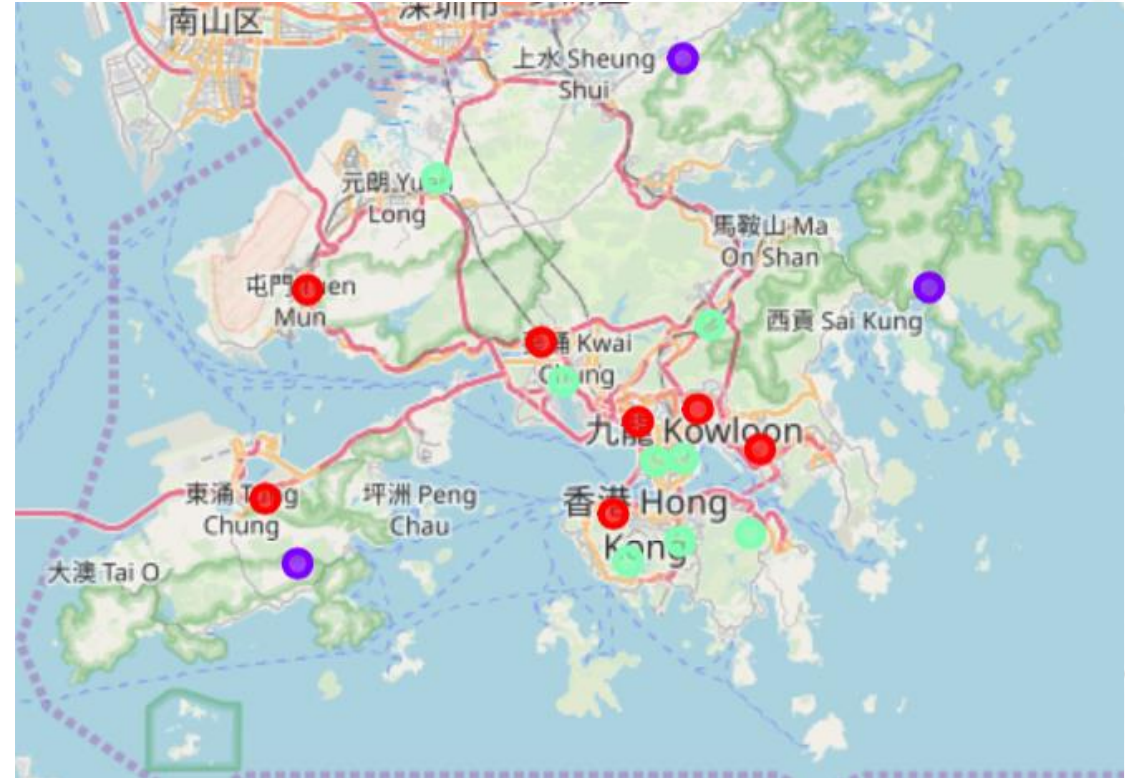
This is the data for clustering on the neighborhoods.

Sources of data and methods to collect

- This capstone make use of data science skills for analysis, including:
 - Web scraping from Wikipedia
(https://en.wikipedia.org/wiki/Districts_of_Hong_Kong)
 - API connection with Foursquare
 - Data cleaning
 - Data wrangling
 - Machine learning on K-means clustering
 - Map visualization with Folium

Results

- **Cluster 0** : Neighborhoods with high concentration of coffee shop
- **Cluster 1** : Neighborhoods with low number to no existence of coffee shop
- **Cluster 2** : Neighborhoods with moderate number of coffee shop



Discussion

- Cluster 1 represents a great opportunity and high potential areas to open new coffee shop as there is very little to no competition from existing.
- Most of the coffee shop are concentrated in the central area of Hong Kong, with the highest number in cluster 0 and moderate number in cluster 2.
- It is advised to avoid neighborhoods in cluster 0 which already have high concentration of coffee shop and suffering from intense competition.

Conclusion

- The neighborhoods in cluster 1 would be a preferred location to open a new coffee shop with the lowest number of competitors.
- This findings will help the relevant start up to capitalize on the opportunities on high potential locations while avoiding overcrowded area in the decisions top open a new coffee shop.
- In our observation, only consider one factor i.e. frequency of occurrence of coffee shop, but there should be others like population and rental payment etc. However, to the best knowledge of this analysis such data are not available to the neighborhood level required by this capstone.