



Introduction and business problem



- Hong Kong is world's global city for various market players
 - Outdoor events are popular in Hong Kong
 - The increase needs of outdoor supply related products took place from 2020
- ☐ ABC Outdoor Supply Store plan to open 1st shop in Hong Kong
 - O How to choose the best district to open the 1st shop?
 - The district should have high demand and low competition
 - This study is to find out the district to fulfill above two factors



Data Description

- Data Requirement:
- List of major Hong Kong district (contains district name)
- Latitude and Longitude coordinates of these districts.
- Venue data for each district, (contains the venue name, venue latitude, venue longitude, venue category)
- Population of district

- Data Source:
- Internet (Wikimedia) for Hong Kong district information
- Foursquare API for venue information
- Geocoder information for latitude and longitude coordinates
- Data Processing
- Data Cleansing is needed
- Data must put into structured format to analysis

District	Chinese	Population of 2016	Population Growth from 2006	Density	Latitude	Longitude
Central and Western	中西區	243,266	-2.7%	19,391	22.282190	114.144860
Eastern	東區	555,034	-2.8%	30,861	22.272090	114.221396
Southern	南區	274,994	-0.6%	7,080	22.258010	114.153080
Wan Chai	灣仔區	180,123	-0.1%	17,137	22.277101	114.173837

District	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Central and Western	22.282190	114.144860	Morning Trail, The Peak (山頂異種症)	22.278008	114.144432	Trail
Central and Western	22.282190	114.144860	Frantzén's Kitchen by Björn Frantzén	22.284808	114.148220	Scandinavian Restaurant
Central and Western	22.282190	114.144860	Okra Hong Kong	22.286108	114.146104	Japanese Restaurant
Central and Western	22.282190	114.144860	Craftissimo	22.284589	114.148293	Beer Store
Central and Western	22.282190	114.144860	Bruti	22.286156	114.143600	Tapas Restaurant
Eastern	22.272090	114.221396	Master Low-key Food Shop (您讀宴手大街小食)	22.279212	114.229606	Snack Place
Eastern	22.272090	114.221396	Quarry Gap (大風坳)	22.266827	114.213317	Other Great Outdoors

Qua	rry Gap (大風和) 2.
District	Outdoor Supply Store
Central and Western	0.00
Eastern	0.00
Islands	0.00
Kowloon City	0.00
Kwai Tsing	0.00
Kwun Tong	0.00
North	0.00
Sai Kung	0.00
Sha Tin	0.00
Sham Shui Po	0.01



Methodology

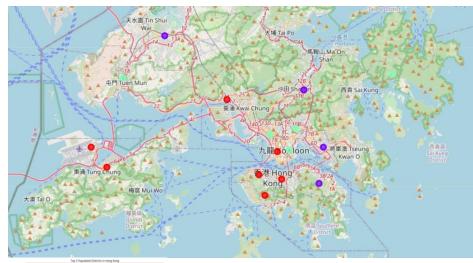
- Data Scraping from internet (Wikimedia) to get list of districts
- Get geographical coordinates (each district' latitude, longitude) by Geocoder
- Get the venue information through Foursquare API
- Group the data by district and taking the mean of frequency occurrence of venue category
- Filter the district by venue category "Outdoor Supply Store", merge with district population data to prepare clustering data
- Process K-Means machine learning technique to segment and cluster similar districts
- Visualize the clustering results into map by Folium

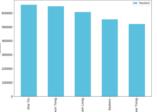


Result and Discussion

- K-Means clustering categorized the districts into 3 x clusters
 - base on the frequency of the occurrence of venues category as "Outdoor Supply Store" and the population of the district

List top districts with highest population







Result and Discussion

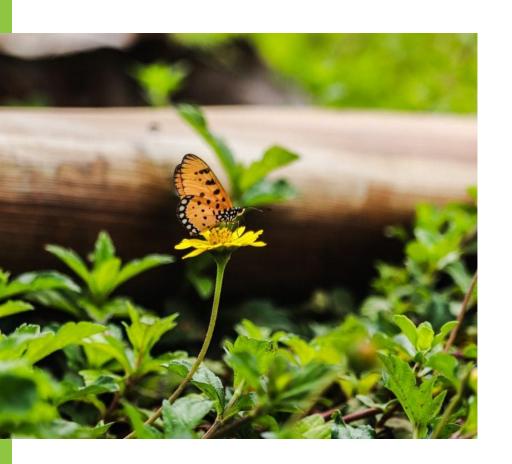
- Base on the K-Means clustering, we found Cluster 1 has the highest population districts and also has low supply of "Outdoor Supply Store"
- In Cluster 1, we checked out the district "Sha Tin", has the highest population in the cluster, and very few "Outdoor Supply Store"
- Further, the Bart Chart of top districts with highest population also show that the district "Sha Tin" is the district with highest popular among all the other districts.
- Both the factor of "high demand" and "less competition" are fulfilled.

	District	Outdoor Supply Store	Chinese	Population	Population Growth	Density	Latitude	Longitude	Cluster Labels
5	Sha Tin	0.0	沙田區	659794.0	+8.6%	9,602	22.382036	114.202102	1
	5 Kwun Tong	0.0	觀塘區	648541.0	+10.4%	57,530	22.314236	114.226625	1
	17 Yuen Long	0.0	元朗區	607200.0	+15.0%	4,435	22.445131	114.025732	1



Conclusion

- The recommendation of the district to open 1st outdoor supply store:
 - Cluster selection: Cluster 1
 - District : Sha Tin
- Because the factor that High demand and Less competition are considered
- Further suggestion
 - Run the program again to have updated results
 - consider the result for selecting 2nd shop for growth of business
 - The result could be used to re-validate the findings
 - For better decision making for the project team



Thank You!

And welcome to Hong Kong!