

HK District Cluster Research for the Relocating Individual

Introduction

Hong Kong is an international financial centre and one of the most densely populated cities in the world. The living standard in Hong Kong is also comparable to the most developed cities in the world. The direct gateway of investing to China from an international standpoint means that Hong Kong is a very attractive place for expats and international talents to pursue their career.

As such, there is a huge market for people who would be researching the type of area they wish to relocate to. This report is meant to serve audience who wish to relocate to Hong Kong and looking for where they would like to pursue their house rental or house purchase or for businessman who wish to understand the neighbourhood in different districts of Hong Kong to understand the market landscape.

Data Collection

Firstly, I will extract the data from Wikipedia to define the districts of Hong Kong: https://en.wikipedia.org/wiki/Districts_of_Hong_Kong. The longitudes and latitudes of different districts are then obtained by geocoder or any other data source if geocoder cannot locate the district properly.

Foursquare API is then utilized to query for popular venues in the neighborhood, forming one hot encoding and applying k-means clustering machine learning technique to explore the data.

Methodology

The district data from Wikipedia are scraped using the `read_html` function in the pandas package. A first and preliminary data cleaning technique would be applied to ensure the data is consistent.

The Foursquare query will then be applied to obtain the nearby popular venues to check the conditions of various districts. The data would then be turned into one-hot encoding and standardized to apply k-means clustering. The districts would then be clustered into different types for presentation.

Results

The 18 districts among Hong Kong were further filtered through to only include districts not in the New Territories given we are targeting audience that are probably interested to have a flexible commute within the CBD area of Hong Kong. As such, New Territories would be too far out for individuals to consider.

The 18 districts were then further distilled to only 9 districts that are most likely interested by expats.

These 9 districts were then clustered to 3 different clusters using K-means technique.

The Image below illustrates the clustering done using the data from FourSquare.



Discussion

The three clusters generated involve 7 of them being in cluster 1 and 1 district in each of cluster 2 and 3. By in large, the results make sense given the data obtained from FourSquare indicated that among cluster 1, there is a vibrant mix of shopping malls, restaurants, hotels and coffee shops. In Cluster 2, the location is further away from the main area of Hong Kong island with parks being the most common venue, which is a rare phenomenon given the scarcity of land in Hong Kong. Lastly, in cluster 3, the district includes only restaurants of different types. It does make sense from a local standpoint given that area is known for eating and industrial work (which wouldn't be captured by FourSquare).

Conclusion

The clustering makes sense in how it is divided. Cluster 1 would probably be the most sought after by expats or individuals who wish to pursue a good vibrant city life in general. Cluster 2 would be particularly attractive for individuals who have a strong interest in outing sports or hiking. Lastly, cluster 3 would seem viable for only the biggest food appreciating individual when relocating.