**Growing New Dessert Business To One Or More Districts in Hong Kong**

Priscilla Yip

1 June 2019

1. **Introduction**
   1. Background

Hong Kong is named Gourmet Paradise not only because of the variety of cuisine, but also numerous restaurants and bistro in this little area, leading to intense competition among restaurants. To run a culinary business successfully in Hong Kong, location is of utmost importance. Among ground level location, shopping mall and upper level shop location, ground level location is usually preferred due to high footfall. However, this option is also the most expensive. That said, many business owners are still willing to invest the lion’s share on rents in exchange for higher business volume.

* 1. Problem Description

Mr A, a culinary business owner, opened a dessert shop in Mong Kok (in Yau Tsim Mong district) two years back. The dessert shop gradually develops a good reputation and starts making profits lately. Mr A would like to leverage on the reputation to open shops under the same brand in other areas. He believes that the location of his Mong Kok shop is the secret to his success, therefore, he wants to open shops in areas with similar surrounding environment. Mr A is puzzling how to select the best locations.

* 1. Surrounding Environment

Desserts are appealing to Hongkongers, as well as tourists. However, people normally savor desserts after a lunch or dinner. Hence, dessert shops usually begin their business after noon. So as to compensate for the short working period, it is very critical for these dessert shops to be opened in ground level areas, adjacent to other restaurants, shopping malls or other attractions with high footfall.

1. **Data Description**

Hong Kong is so tiny that there is no borough. The whole city can be divided into 18 cities, with area ranging from 9 km2 to 175 km2. Some data and statistics are available at this district level.

I tried running Geopy to get the coordinates of these 18 cities. However, based on my knowledge of the city, I found that the output is not very accurate. Thus, I searched their coordinates through the Internet. Then, I used Foursquare API to get the most common venues of each of the 18 districts.

I also need average footfall by district. Given that this data is unavailable, I took population density (population divided by area) as a proxy of footfall. 2017 population data is extracted from the Census and Statistics Department, whereas district areas are from Wikipedia. Average rental by district can be a good feature, but also not available. I did not attempt to take samples from the Internet to estimate average rentals because rentals fluctuate a lot even within the same district.

All the above data is quite clean, so there is no need for data wrangling.

1. **Methodology**

To identify districts having similar surrounding environment as Yau Tsim Mong, I use K-means to segment the 18 districts into clusters and then select districts that are in the same cluster as Yau Tsim Mong. The frequency of venue categories and population density are used as features to run the K-means model.

To determine the optimal number of clusters, I run the K-means model in a for loop from a count of 1 to 10. Within-cluster sum of squared errors, or SSE, is used to evaluate the performance of K-means. It appears that SSE declines as the count increases but stabilizes when number of clusters reaches 7. I decided to adopt 5 clusters. The distribution and clustering of the districts are visualized by using the python folium library. Each cluster is indicated by one color.

In the same cluster, districts should have similar number and categories of venues, as well as population density. That means these districts have similar surrounding environment and footfall. In the cluster of Yau Tsim Mong, if there are districts more than Mr A wants to select, preference will be given to districts having more restaurants, shopping malls or attractions. The higher the density the better.