Find Optimal Locations for Opening Up a Bubble Tea Shop in Toronto

Yuxin "Melanie" Cui

1. Introduction

1.1 Background

Bubble tea, a sweet drink originally from Taiwan, impresses people with its foamy layer and chewy "pearls". And it has become exceedingly popular in North America with a growing market. Given the fact that bubble tea has a stable customer source especially among Chinese, Toronto, the most populous city in Canada and a large Asian population of around 20 to 30 percent, is considered a potentially huge market for Chinese cuisine.

1.2 Business Problem

The goal of this project is to find the optimal locations for a new bubble tea shop in the city of Toronto, Canada using the Foursquare location data. This report is targeted to potential investors who are interested in opening up a bubble tea shop in Toronto. Only the choice of location in the business plan of opening a bubble tea shop will be explored. And areas where Asian restaurants are located will be targeted in consideration of a shared customer base for bubble tea.

1.3 Interest

Entrepreneurs or small business investors who are seeking opportunities in bubble tea business might be interested.

2. Data Acquisition

2.1 Data Source

The list of neighborhoods of the city of Toronto, Canada is scraped from webpage https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M using pandas html table scraping method into dataframe. The coordinates of boroughs are provided by http://cocl.us/Geospatial_data and geographical coordinate of Toronto is obtained by Geopy library. The venue data related to the interested regions was retrieved by using Foursquare API.

2.2 Data Manipulation

The list of neighborhood names and coordinates are merged into one table. Rows with missing values of "Not assigned" in borough list are dropped. And the neighborhood list is them grouped in the same borough.

East York, North York, Scarborough and Toronto are selected into further considerations because of their demographic components of Asian population. In conclusion, 7 unique boroughs are chosen in the project: Central Toronto, Downtown Toronto, East Toronto, West Toronto, East York, North York, and Scarborough.

Foursquare API was used to pull a list of top 100 venues within the radius of 500 meters including their names, categories, latitude and longitude. From which, it is shown that there are 258 unique categories of venues. To analyze each neighborhood, rows are grouped by neighborhood, and the mean on the frequency of occurrence of each venue category is calculated.

## 3. Clustering Analysis

3.1 K-means Clustering

The analysis starts off by setting the number of clusters to be 3. And venue category of "Chinese Restaurant" from the frequency data table is listed out to assign clustering labels. K-means clustering is performed on the interested dataset of the top 100 venues within 500 meters of radius. This clustering method identifies 3 (or k number of) centroids of the data and allocates every data point to the nearest cluster with the nearest mean (centroid).

In this project, the neighborhoods of Toronto are clustered into 3 clusters according to their frequency of occurrence of "Chinese Restaurant" in the venue category. And

3.2 Clustering Results

Cluster 0	Neighborhoods with least Chinese restaurants
Cluster 1	Neighborhoods with most Chinese restaurants
Cluster 2	Neighborhoods with some Chinese restaurants

The results of clustering in the map of Toronto are visualized using Folium package are shown in the following figure. The neighborhoods are classified into 3 clusters with 3 colors. Cluster 0 is shown with red marker, cluster 1 with purple marker and cluster 2 with green marker. Observations can be made that Chinese restaurants are mostly found in cluster 1 and least found in cluster 0.
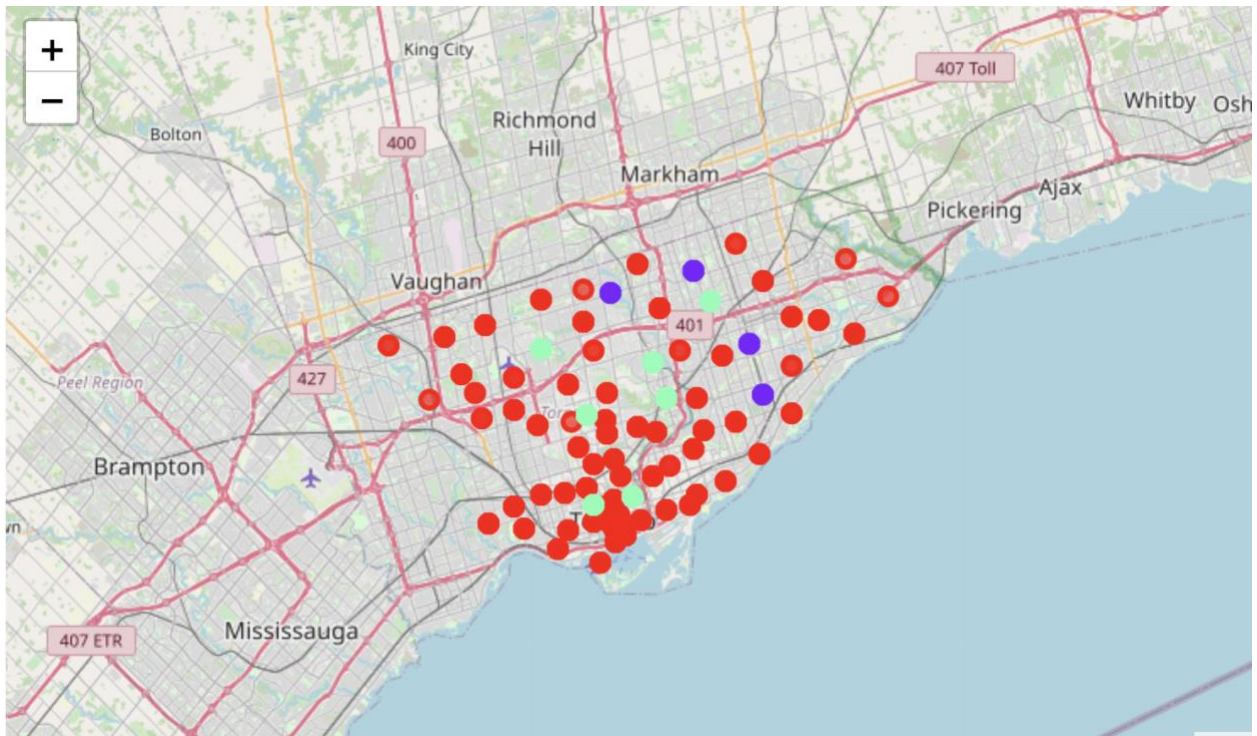
4. Recommendations

While all bubble tea shops are found in cluster 1, areas such as Church and Wellesley, Central Bay Street, Garden District and Ryerson in Downtown Toronto. These neighborhoods are outside of our recommendation in consideration of potential competitions. However, bubble tea shops and Chinese restaurants usually share customer base, and that the relationship between Chinese food and bubble tea is more similar to complements. This project recommends locations where Chinese restaurant are nearby but competitions with other bubble tea shops are low. So, neighborhoods with Chinese restaurants like Steeles West and L'Amoreaux West in the borough of Scarborough are highly favored.

5. Limitations and Future Research

In this project, the only factor that is examined is the frequency of Chinese Restaurant. However, many other elements might affect this recommendation such as location of other Asian restaurants, location of coffee shops, location of CBD (Central Business District), local population density and average income etc.

6. Conclusion

Clustering is unsupervised learning that targeted at finding patters and grouping in unlabeled data. K-means clustering is a partitioning clustering that divides data points into non-overlapping subsets without internal structure. It is easy to implement and usually efficient for medium or large size databases. But the number of clusters (k) needs to be specified first.

This project identifies the business problem of finding the ideal locations for opening up a bubble tea shop. Venue and geographical data required is extracted and prepared. K-means clustering method is used for analysis in order to provide proper recommendations to potential investors.

7. References

1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. https://en.wikipedia.org/wiki/Demographics_of_Toronto
3. http://cocl.us/Geospatial_data