IBM DATA SCIENCE PROFESIONAL CAPSTONE PROJECT

WEN YIDA
2019-12-23

IBM

Background: Immigration to Canada

70% from Europe before 1960s

Universal point system in 1967

64% from non-European regions in the 1970s

Business Immigration Program in 1978

60% from Asia since 1980s
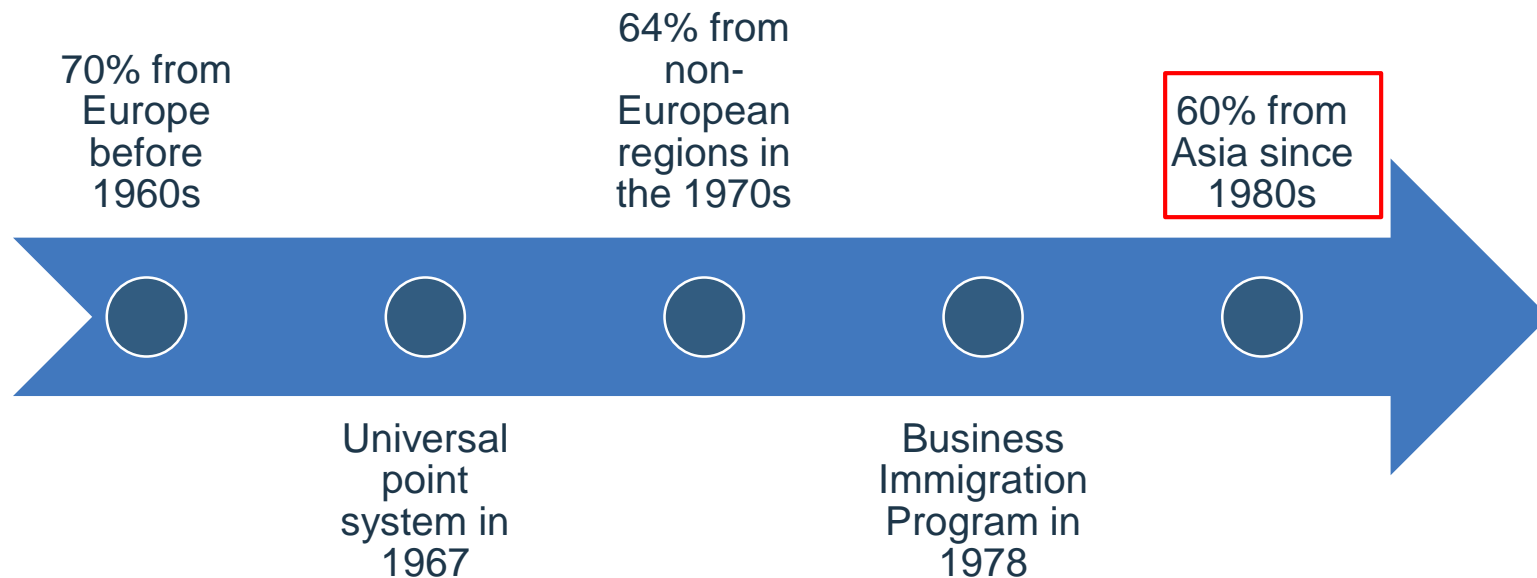
Quick Fact: Toronto

- Toronto – Pollution 6.1 million by 2019

- 10.8% of the city's population, or 658 K(*) are Chinese, based on 2016 Census -> **Big Market Share**

- Location is key to run a successful Chinese Cuisine Restaurant.

IBM

About this project

- This is the capstone project for IBM Data Science Professional Certificate.

- Due to very high market share, it is a very good opportunity to open an Chinese Cuisine restaurant in Toronto.

  – This project is trying to find out the ideal location for open Chinese Cuisine restaurant in Toronto:
    - Neighborhoods list in Canada Toronto (week 3, from Wikipedia);
    - Use Foursquare API to get list of venues and coordinates;
    - Coordinates of Toronto Neighborhoods;
    - Use K-means Clustering to group by Toronto Neighborhoods by Chinese Cuisine restaurants.
    - Visualize the map and data by using Folium taught during visualization class.

  – Target Audience:
    - Entrepreneurs
    - Joint Venture or Private Equity who has interests on acquiring Chinese restaurants.

IBM

Data

- To solve this problem, we will need below data:

  List of neighborhoods in Toronto, Canada;

  Latitude and Longitude of these neighborhoods;

  Venue data related to Chinese restaurants;

  Neighborhoods that are good to open an Chinese Restaurants.

EXTRACTING THE DATA

- The list of Toronto neighborhoods in Wikipedia;

- Latitude and Longitude data of neighborhoods via CSV provided during Week 3;

- Foursquare API to get venue information related to Chinses restaurants.

K-MEANS Clustering

K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible.

For this project clustered the neighborhoods in Toronto into 3 clusters (**Cluster 0, Cluster 1, Cluster 2**)

Clustering based on their frequency of occurrence for "Chinese food". Based on the results (the concentration of clusters).

Cluster 0: Neighborhoods with the less number of Chinese Restaurants
Cluster 1: Neighborhoods with the middle size of Chinese restaurants.
Cluster 2: Neighborhoods with a more number of Chinese restaurants

Result and Visualization

In [65]: `#Cluster 0`
`to_merged.loc[(to_merged['Cluster Labels'] ==0) & (to_merged['Venue Category'] == 'Chinese Restaurant') ]`

Out[65]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
20	Harbourfront East, Toronto Islands, Union Station	0.010000	0	43.640816	-79.381752	Pearl Harbourfront	43.638157	-79.380688	Chinese Restaurant
30	Ryerson, Garden District	0.010000	0	43.657162	-79.378937	GB Hand-Pulled Noodles	43.656434	-79.383783	Chinese Restaurant
9	Church and Wellesley	0.011765	0	43.665860	-79.383160	Crown Princess Fine Dining 伯爵名宴	43.666455	-79.387698	Chinese Restaurant

In [66]: `#Cluster 1`
`to_merged.loc[(to_merged['Cluster Labels'] ==1) & (to_merged['Venue Category'] == 'Chinese Restaurant') ]`

Out[66]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
18	Harbord, University of Toronto	0.028571	1	43.662696	-79.400049	River Tai Restaurant	43.662902	-79.403167	Chinese Restaurant
6	Central Bay Street	0.034091	1	43.657952	-79.387383	GB Hand-Pulled Noodles	43.656434	-79.383783	Chinese Restaurant
6	Central Bay Street	0.034091	1	43.657952	-79.387383	Kowloon Seafood Dim Sum Restaurant	43.656217	-79.392418	Chinese Restaurant
6	Central Bay Street	0.034091	1	43.657952	-79.387383	Yueh Tung Chinese Restaurant	43.655281	-79.385337	Chinese Restaurant
5	Cabbagetown, St. James Town	0.023810	1	43.667967	-79.367675	China Gourmet	43.664180	-79.368359	Chinese Restaurant

In [67]: `#Cluster 2`
`to_merged.loc[(to_merged['Cluster Labels'] ==2) & (to_merged['Venue Category'] == 'Chinese Restaurant') ]`

Out[67]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
25	North Toronto West	0.052632	2	43.715383	-79.405678	C'est Bon	43.716785	-79.400406	Chinese Restaurant
7	Chinatown, Grange Park, Kensington Market	0.054945	2	43.653206	-79.400049	Swatow Restaurant 汕頭小食家	43.653866	-79.398334	Chinese Restaurant
7	Chinatown, Grange Park, Kensington Market	0.054945	2	43.653206	-79.400049	New Sky Restaurant 小沙田食家	43.655337	-79.398897	Chinese Restaurant
7	Chinatown, Grange Park, Kensington Market	0.054945	2	43.653206	-79.400049	Asian Legend 味香村	43.653603	-79.395047	Chinese Restaurant
7	Chinatown, Grange Park, Kensington Market	0.054945	2	43.653206	-79.400049	Rosewood Chinese Cuisine	43.653171	-79.396710	Chinese Restaurant
7	Chinatown, Grange Park, Kensington Market	0.054945	2	43.653206	-79.400049	House of Gourmet 滿庭芳	43.653273	-79.397230	Chinese Restaurant

Result and Visualization
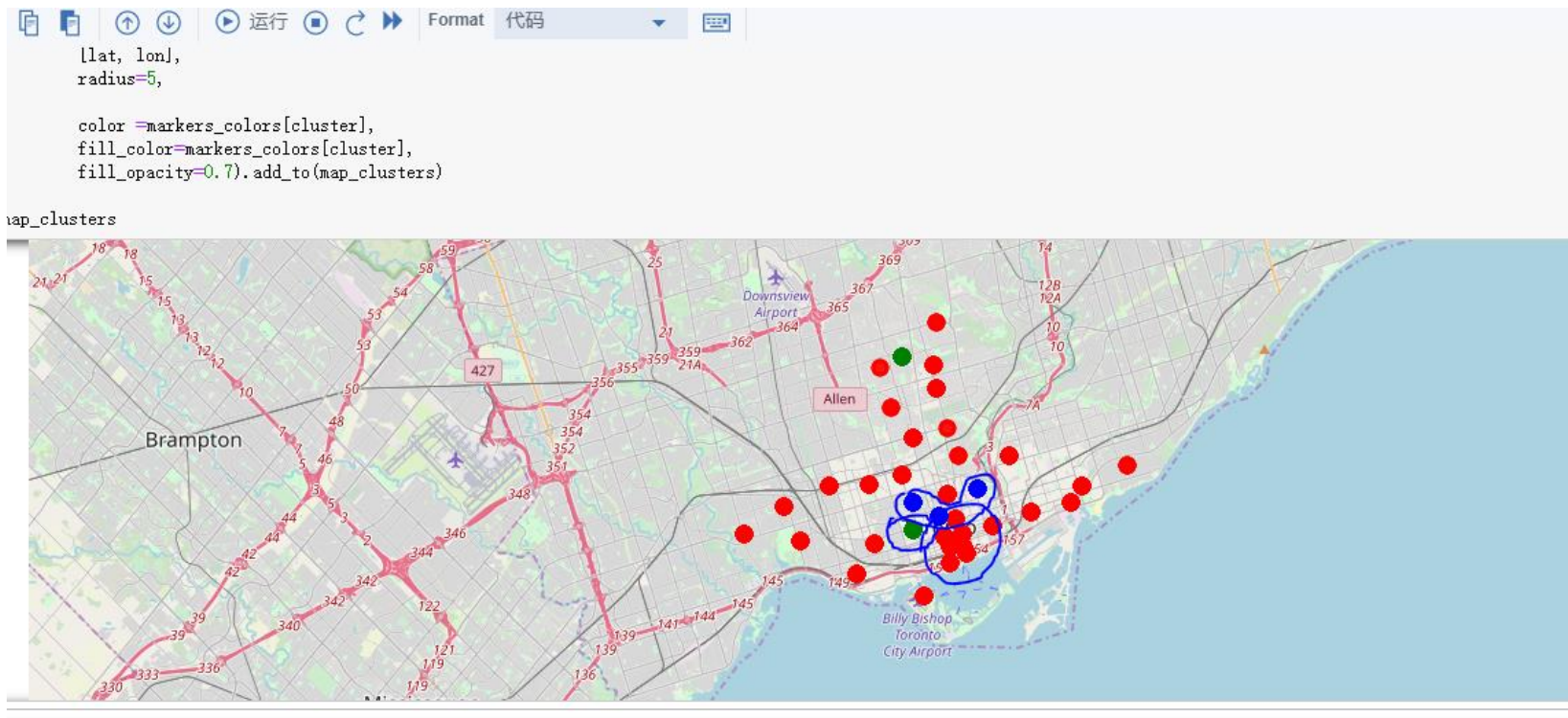
Conclusion

- There are a lot of Chinese Restaurants in Toronto already;

- Most Chinese restaurants are in Chinatown, Grange Park, Kensington Market (Cluster 2).

   And in Cluster 1 areas which are more concentrated on Central Bay Street.

- In cluster 0 area, there are less Chinese restaurants. For restaurant like "Crown Princess Fine Dining 伯爵名宴", is near University of Toronto.

Therefore, if you want to open a Chinese restaurants, maybe the Church and Wellesley is a good place, since is near University of Toronto got a lots of Chinese students as customer, but less competition.

Since main customer group is university students, food should be cheaper but tasty, decoration and healthy diet(like organic) for restaurants will not be so important.

THANK YOU

IBM®