IBM Data Science Capstone
Final Project

Week 5 - Project Report

Opening of a new Chinese Restaurant Business and Finding best place to live for a Chinese Migrant Chef in Toronto Canada

By
Salman Azam

INTRODUCTION

This project is for a Chinese migrant to Canada who is a chef and wants to migrate to Toronto and setup his own Chinese restaurant business. He prefers to both live in a Chinese neighborhood and open restaurant in a Chinese neighborhood too obviously. As we know there are not too many restaurants with Chinese cuisine in Toronto Canada. So, this might be an interesting idea for a new startup.

As such cuisines are very popular in Chinese community, and also among some locals, it might be a good idea to start a setup but it might work even better in areas of Chinese communities so the first objective before starting this idea will be to identify such localities and to figure out if we have enough of such localities in Toronto to open up such a business and if yes, then where to start such a business.

Business Objectives

Our business idea has following key problems to identify and resolve:

- Areas in Toronto which have good Chinese restaurants

- Areas in Toronto which have Chinese communities
- Best option to start a new restaurant in Toronto according to Chinese community
- Best locality to live where Chinese community is around.

Data Needed

Following data shall be needed:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to Chinese restaurants. This will help us find the neighborhoods that are more suitable to open a Chinese Restaurant.

Methodology

- Scrapping of Toronto neighborhoods using Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API like done in course

Here is list of steps to do this project:

- Get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Do the web scraping by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame.

- However, it is only a list of neighborhood names and postal codes. We need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, we use CSV file provided by IBM team to match the coordinates of Toronto neighborhoods.

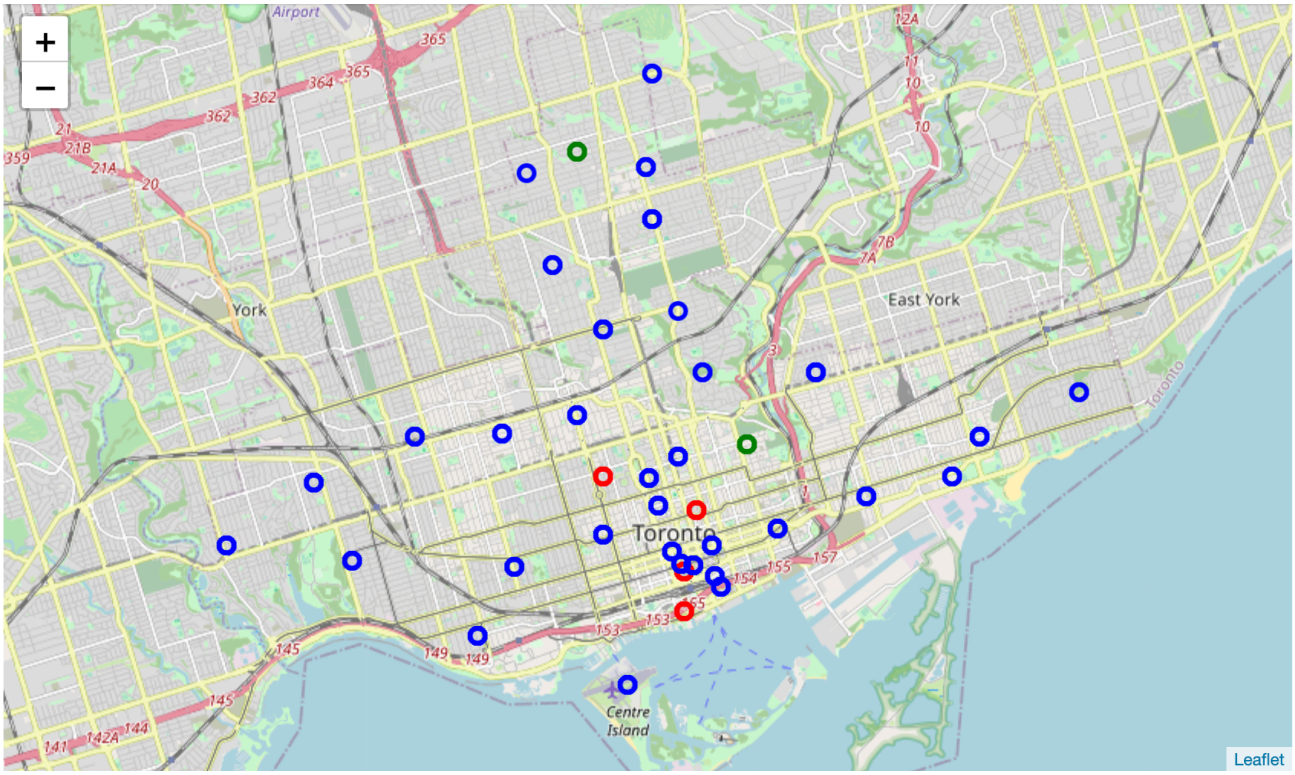
- After gathering these coordinates, we visualize the map of Toronto using Folium package to verify whether these are correct coordinates.

- Now we use Foursquare API to pull the list of top 100 venues within 500

meters radius. For this we have created a Foursquare developer account in order to obtain account ID and API key to pull the data.

- From Foursquare, we pull the names, categories, latitude, and longitude of the venues.
- With this data, we can also check how many unique categories that we can get from these venues.
- Then, we analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.
- Now we specifically look for “Chinese restaurants”.
- Lastly, we perform the clustering method by using k-means clustering.
- Now we cluster the neighborhoods in Toronto into 8 clusters based on their frequency of occurrence for “Chinese food”. Based on the results (the concentration of clusters), we will be able to recommend the ideal location to open the restaurant.

Results



The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Chinese restaurants are in each neighborhood:

- Cluster: Neighborhoods with a smaller number of Chinese restaurants.
- Cluster: Neighborhoods with no Chinese restaurants.
- Cluster: Neighborhoods with a greater number of Chinese restaurants

The results are visualized in the above map.

Discussions

Here are the outcomes of the analysis:

- Most of the Chinese restaurants are in Cluster 0 and Cluster 2
- This means indirectly that there is enough market of Chinese food in these particular clusters
- This also means that these areas might have more Chinese population compared to some other areas

Conclusion

Therefore in order to either open a new Chinese restaurant or have living in Chinese community, these areas are better

- Garden District, Ryerson
- Harbourfront East
- Toronto Dominion Centre
- North Toronto West
- St. James Town