IDENTIFICATION OF A DESIRED LOCALITY IN TORONTO, CANADA (CHINESE LOCALITY)

K S KARAN

October, 2019.

TABLE OF CONTENTS:

Introduction:

The particular project is based upon the selection of an ideal locality for a migrating audience who wish to settle in an area that panders to those with a taste for the Chinese culture. Due to the large influx of migration into Canada, such assistance would be valuable to help the target audience have an idea of their desired locality for residence. This has been done by utilising particular unsupervised machine learning algorithms and Foursquare API.

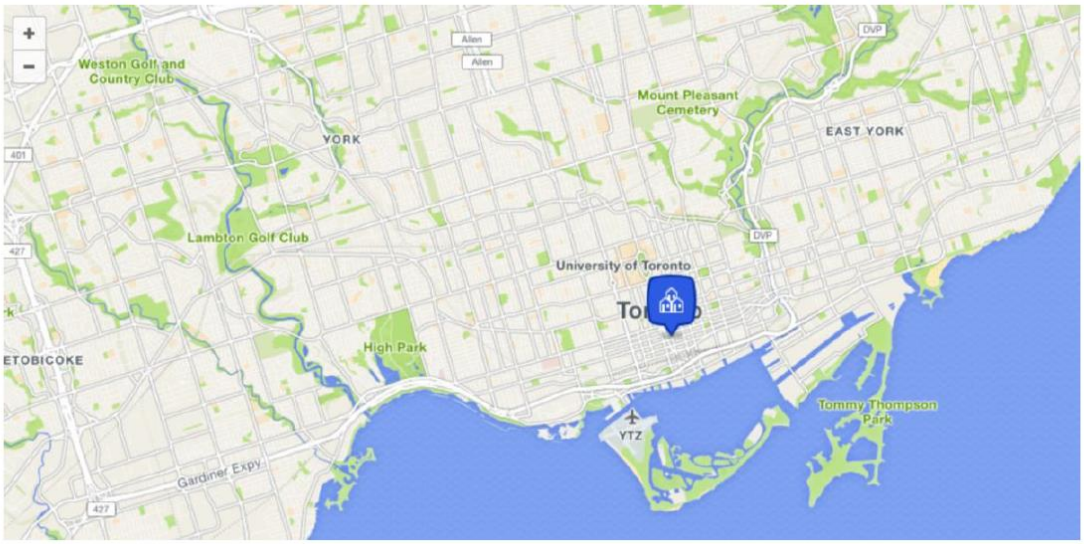


Fig: 1. Map of Toronto, Canada from a foursquare map from the web.

Background:

Currently due to the migration levels in the world, it is not surprising to see people moving from one corner of the world to another for various reasons such as professional careers, education etc. In such cases, the first and foremost issue that arises would be that of housing. Every individual would prefer to stay in a locality that is slightly similar to that of their hometown.

Problem:

With regards to my above observations, I have decided to study the data and identify a neighbourhood that is preferable to a person coming from a Chinese background. This includes a locality having restaurants, education institutes etc that propagate the Chinese

culture. To help in determining such an area, I decided to take the number of Chinese restaurants to be the main criteria for my project studies.

To get started right away with this project, we need some data.

• Data from websites

• Location data

Description of the data:

- Scraping of Toronto neighborhoods via Wikipedia.
- Getting latitude and longitude data of the various neighborhoods via Geo-coder packages/ with a .csv file.
- Using Foursquare API to get venue data related to neighbors.

NOTE: Here for finding the neighbourhoods that are most favourable, as stated above, I have selected the number of Chinese restaurants as the main criteria as that would allow me to identify which area has the most concentrated amount of people preferring the Chinese culture and would likely have a larger density of people preferring the Chinese culture.

Discussing the problem:

- Getting the list of neighbourhoods in Toronto by extracting said list from Wikipedia page.
- Web scraping by pandas HTML table scraping, pulling tabular data directly from the web page into the data frame.
- Utilizing Foursquare API to pull list of venues near the neighbourhood with the CSV file provided by IBM.
- Visualize Toronto map using folium package to verify the coordinates.
- Use Foursquare API to pull the venues, names, categories, latitudes and longitudes.
- Check the unique categories from the venues.
- Analyze each neighborhood by grouping the rows by neighborhood and checking frequency of occurrence of each venue category.
- After clustering venues, perform clustering methods by k-means clustering, an unsupervised machine learning algorithm.
- Based on the result, recommend the ideal neighbourhood having highest Chinese population density.

Target Audience:

- Chinese nationals
- Professionals wanting to open Chinese businesses such as restaurants or schools.
- Citizens who want to study/experience Chinese culture.

Methodology:

1) **Business Understanding** : Comprehending the business problem and obtaining the appropriate solution to it is a challenging phase, identifying clearly the features and the factors that are acceptable to the solution that we present itself with. Evaluate all the targets and make a decision according to the desired problem.

2) **Analytical Approach:** For the given business problem, we need to present a solution as to which algorithm can be used to solve the given problem. For this project, I have used k-means clustering algorithm which is used to group the neighbourhoods into clusters.

3) **Data Requirements:** The given project is based upon location data, hence I scraped the Toronto neighbourhoods data set and the location coordinates data set.

4) **Data Collection :**

- Obtaining the list of neighbourhoods in Toronto, Canada by extracting the list of neighbourhoods from Wikipedia page. ("https://en.wikipedia.org/wiki/list_of_postal_codes_of_Canada:_M")

- Web scraping with a pandas HTML table, scraping method is easier to pull the tabular data promptly from the webpage into the data frame.

- The second data set is the location data of Toronto, Canada neighbourhoods, collected the data set as.CSV file from IBM.

5) **Data Understanding:**

- By viewing the data frame, checking all the columns and rows, filling any missing values in the data frame and then we are ready for data cleaning.

- For the location data sets, we use the Foursquare API to pull a list of venues near these neighbourhoods.

6) **Data Preparation:**

- After collecting all the data sets, I cleaned the data and merged the two data frames.
- Subsequently gathering all these coordinates, I visualised the map of Toronto to check the coordinates of the Toronto data.

- Following this, I utilized the Foursquare API to pull the list of venues within a 500 metre radius.

7) **Data Analysis:**

- Next I analysed each neighbourhood by grouping the rows by neighbourhood and taking the frequency of occurrence of each category.

- Here, I have opted for one single category i.e Chinese restaurants so as to make the studies a little easier.

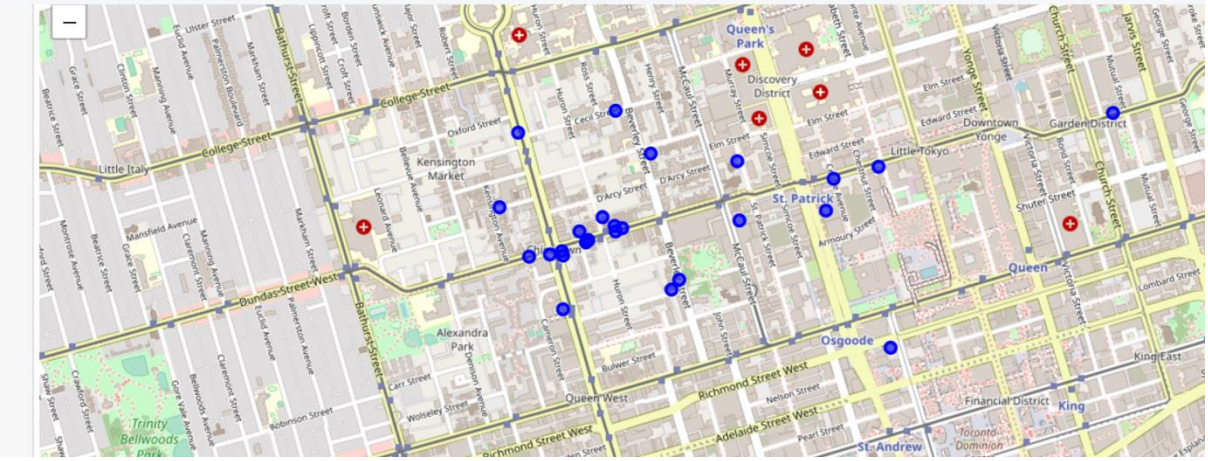
- Next, I used k-means clustering to find all the clusters of all related venues in the neighbourhood.

8) Data Visualisation:

- After running the clustering method, I visualised clusters on the map with the help of folium software.

- Examined the number of clusters and selected the locality that would most likely have a high Chinese population density.

Result:



As per my findings above, I have determined the neighbourhood that likely has the highest Chinese population density due to the presence of various amenities such as restaurants, educational institutes etc. to be "Chinatown, Grange Park, Kensington Market."

Conclusion:

I used Foursquare API to identify all the Chinese restaurants in Toronto and grouped them by the neighbourhood in which they're located. By doing that I identified "Chinatown, Grange Park, Kensington Market" as the most suitable neighbourhood in regards to my original purpose. By displaying the results on a map I can see "Chinatown, Grange Park, Kensington Market" is an area with a large number of Chinese amenities and has the largest number of Chinese restaurants.

NOTE: For a business analysis on the place to open a Chinese restaurant as well, we can use the data obtained above. The best location for such a venture would be a locality that has few or no Chinese restaurants. From the table above, we can determine that a neighbourhood such as North Toronto West could be ideal for such a venture as there will be less competition when compared to a neighborhood such as “Chinatown, Grange Park, Kensington Market.”