Opening a new Chinese restaurant in Toronto Canada

Introduction

Toronto is one of the most densely populated areas in Canada.Being the largest city in Canada with an estimated population of over 6 million, there is no doubt about the diversity of the population.Toronto is well known for its great food.The cuisine of Toronto reflects Toronto's size and multicultural diversity.Different ethnic neighbourhoods throughout the city focus on specific cuisines,such as authentic Chinese and Vietnamese found in the city's Chinatowns, Korean in Koreatown, Greek on The Danforth, Italian cuisine in Little Italy and Corso Italia, and Indian/Pakistani in Little India.In addition to ethnic cuisines, Toronto is also home to many fine dining establishments and chain restaurants ranging from fast food to casual or upscale dining.

The aim of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the best neighbourhood in Toronto to open a restaurant. Through this project, we will find the most suitable location for an entrepreneur to open a new Chinese restaurant in Toronto, Canada.

The target audience of this project is Entrepreneurs or Business owners who want to open a new Chinese Restaurant or grow their current business.

Dataset

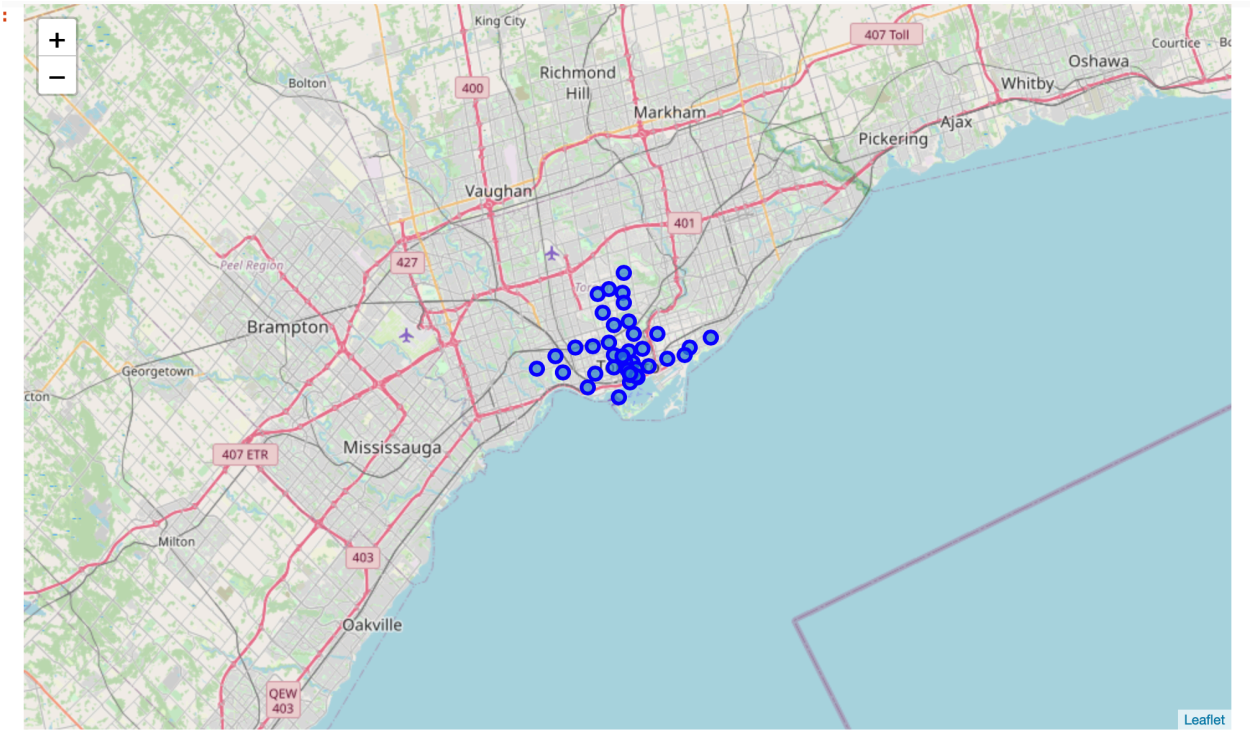
The data was collected from multiple sources which will provide the list of neighbourhoods in Toronto (via Wikipedia), the Geographical location of the neighbourhoods (via Geocoder package) and Venue data pertaining to Chinese restaurants (via Foursquare). The Venue data will help find which neighbourhood is best suitable to open a restaurant.

Methodology

First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting
the list of neighborhoods from wikipedia page
(“ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ”).

I did the web scraping the Wikipedia page by utilizing BeautifulSoup library scraping method as it is easier and more convenient to pull tabular data directly from a web page into dataframe and have done data pre processing and data cleaning.
However, it is only a list of neighborhood names and postal codes. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder package to import the

csv file conatining the latitudes and longitudes for various neighbourhoods in Canada. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.



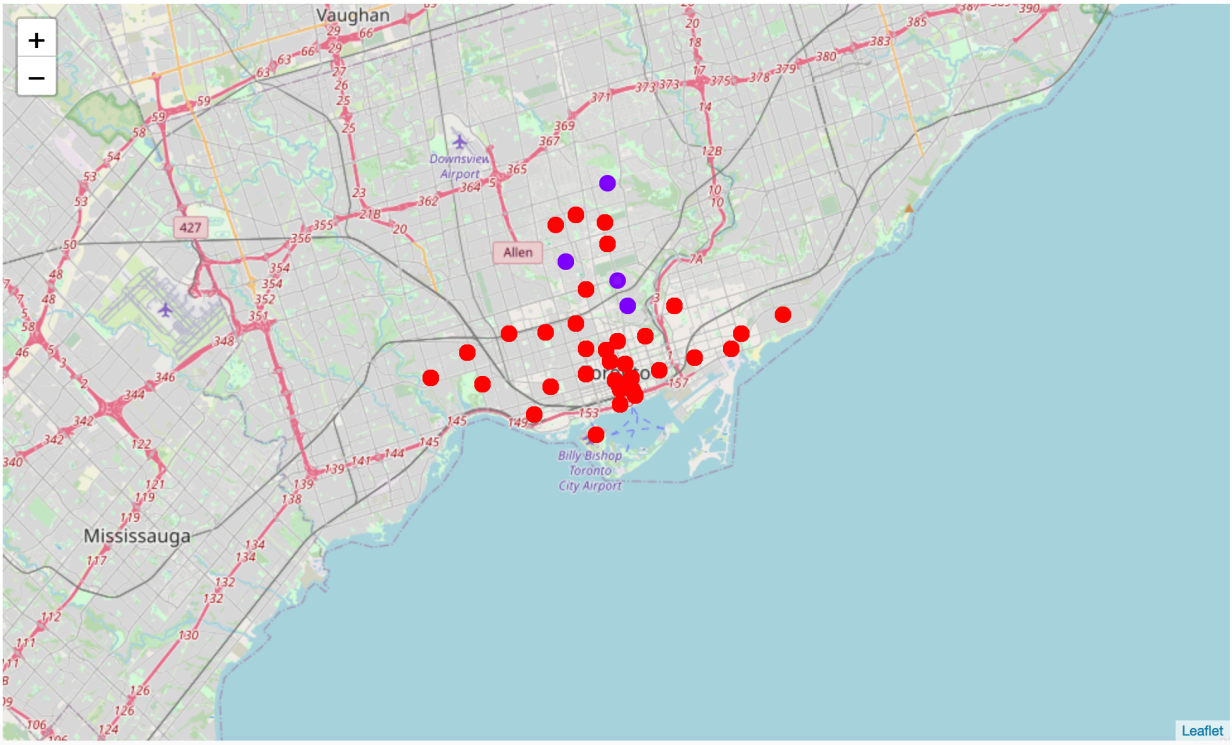
Then I used the Foursquare API to explore the neighborhoods where I passed the coordinates of each neighborhood to the Foursquare API, which returned a list of venues within a given radius of 500 meters and up to 100 venues max. The code requested and returned venues for all 103 neighborhoods in Toronto. The size of the resulting dataframe was 1639 rows by 7 columns, meaning we have 1639 venues in our dataset with neighborhood name, latitude, longitude, venue name and geocoordinates and its category.

With this data, I also check how many unique categories that I can get from these venues. Then, I 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.

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly

suited for this project as well. I have clustered the neighborhoods in Toronto into 2 clusters based on their frequency of occurrence
for “Chinese food”. Based on the results (the concentration of clusters), I 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 2 clusters based on how many Chinese restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with high number of Chinese restaurants
- Cluster 1: Neighborhoods with no Chinese restaurants

The results are visualized in the above map with Cluster 0 in red color and Cluster 1 in purple color.

Discussions

Most of Chinese restaurants are in Cluster 0 which is around Toronto Dominion Centre, Design Exchange, St. James Town, Cabbagetown areas and lowest in Cluster 1 areas

which are Rosedale ,Forest Hill North & West, Forest Hill Road Park and Lawrence Park areas. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Chinese restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Chinese restaurant in these locations with little to no competition. Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

References

List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Foursquare Developer Documentation: https://developer.foursquare.com/docs