

1. Introduction

In this project, we are going to find a suitable place to open a new Chinese restaurant in Toronto. It is also targeted to those business people who want to find the best location for their company. As many Asian people have immigrated to the country. The marketing report provides that Chinese cuisines have become extremely popular because they offer a healthy alternative to regular native eating habits. That's why the owner of the new Chinese restaurant can have great success and profit. However, as with any business, opening a new restaurant requires serious considerations and is complicated. In particular, the location of the restaurant plays a crucial factor that will influence whether it will have success or a failure. Therefore, our project will attempt to answer the questions.” Where should the investor open a Chinese Restaurant?” and “Where should I go If I want great Chinese food?”

2. Data Description

For this project we need these following data:

1. Toronto City data that contains Borough, Neighbourhoods along with their latitudes and longitudes
 - Data Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
 - Description: This Wikipedia page contains all the information we need to explore and cluster the neighbourhoods in Toronto. We will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas data frame so that it is in a structured format like the Toronto dataset.
2. Geographical Location data using Geocoder Package
 - Data Source: https://cocl.us/Geospatial_data
 - Description: The second source of data provided us with the Geographical coordinates of the neighbourhoods with the respective Postal Codes.
3. Venue Data using Foursquare API
 - Data Source: <https://foursquare.com/developers/apps>
 - Description: All the data related to locations and quality of Chinese restaurants will be collected via the Foursquare API utilized via the Request library in Python.

3. Methodology

In this section, we are going to use Python environment to finish three of stages.

- Collecting the Data.
- Exploratory Data Analysis (EDA)
- Machine Learning Model implying.

In the first stage, the BeautifulSoup library is used to scrape the Toronto neighborhood data on the Wikipedia page. Next, we are going to solve some problems on the collecting data, for example, ignore cell with a borough that is not assigned, more than one neighborhood exists in one postal code area (M5A was listed on Harbourfront and Regent Park). In the not assigned problem, if a cell has a borough but not an assigned neighborhood then it will be the same as its borough. As for the second, those two rows will be combined into one row with the neighborhood separated with a comma as shown. The last step of this stage, adding the latitude and longitudes of each location by merging the geographical dataset and get the top 100 venues that are in Lawrence Park within a radius of 500 meters. In the EDA stage, we are going to dummy each venue category in every neighborhood and convert it to numerical values. In the last stage, as the type of questions we want to solve is unsupervised learning therefore K-Mean clustering of the Machine Learning model will be going to be used. To optimize the cluster in the algorithm, the Elbow Method was used to find the best K.

Figure1. This map visualizes all the boroughs present in Toronto

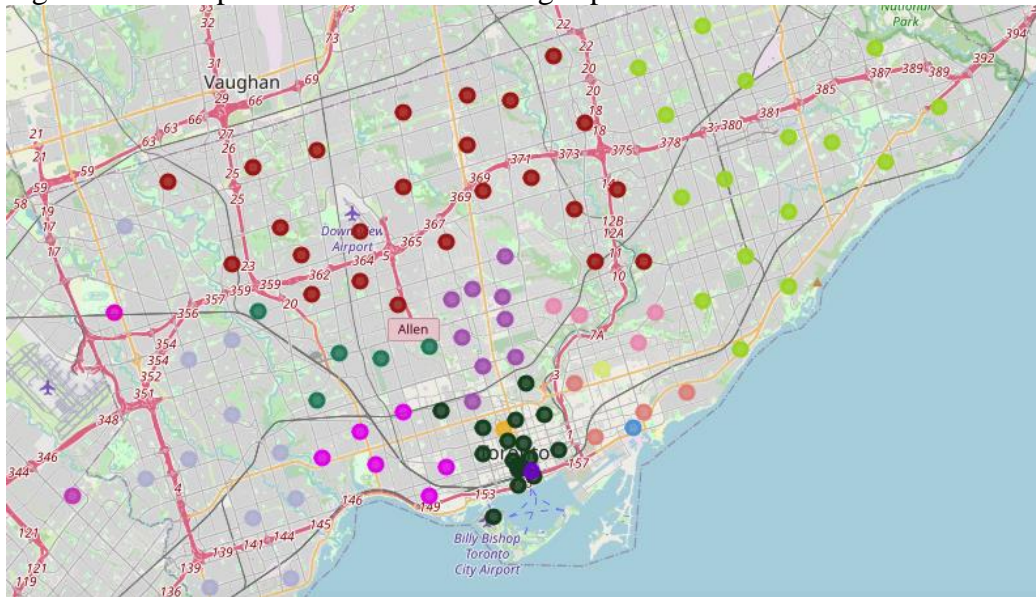
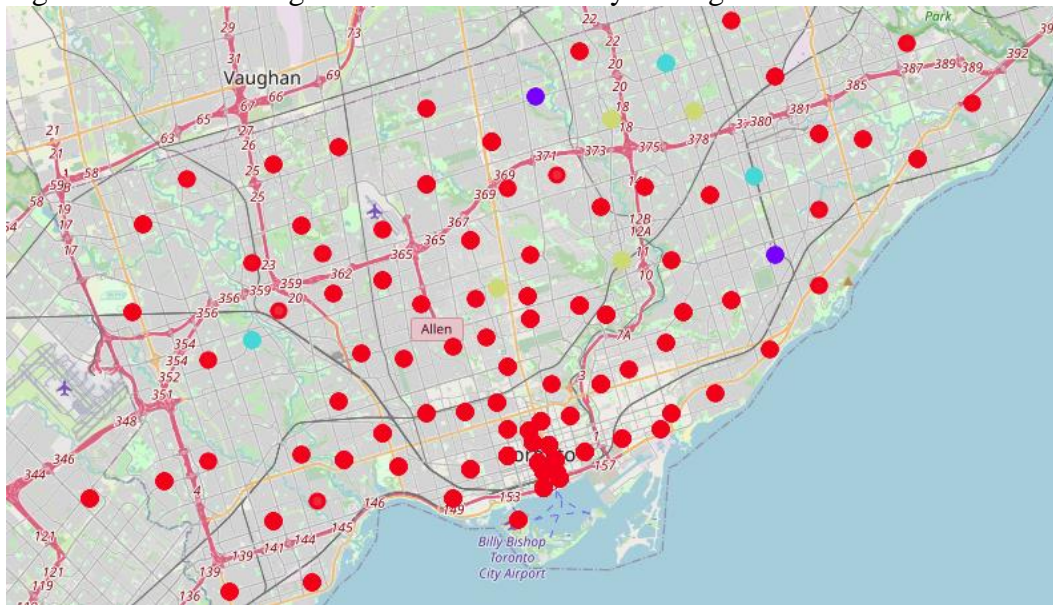
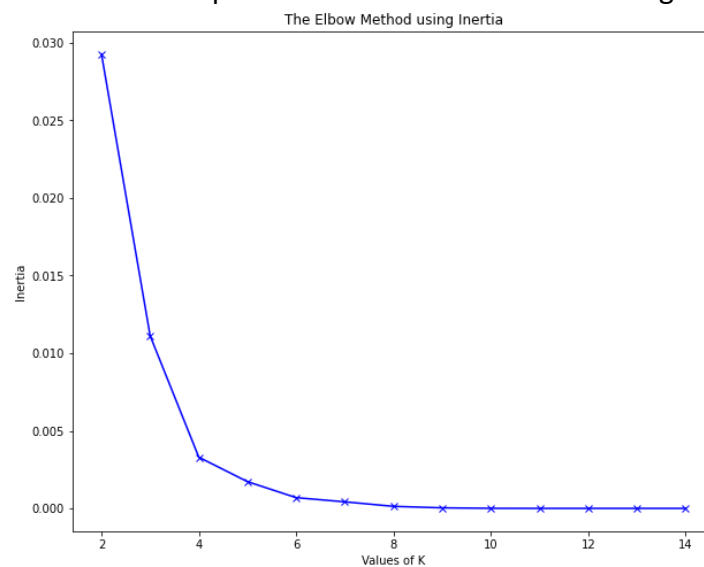


Figure2. The clustering of Chinese restaurant by boroughs in Toronto.

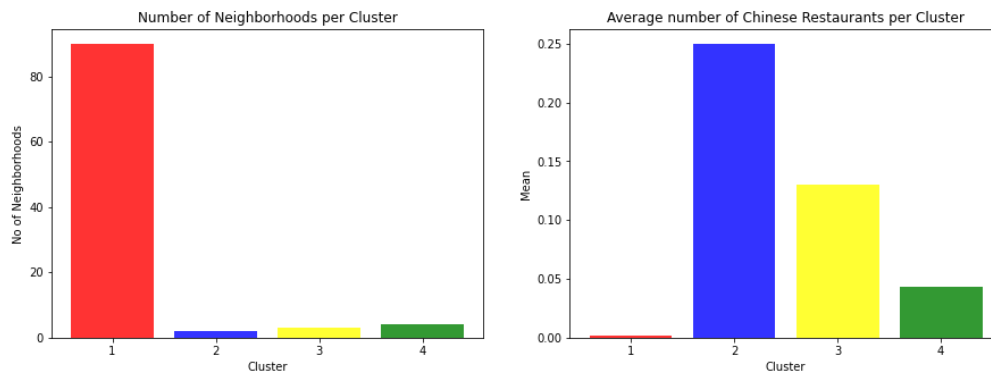


4. Result

Figure3. Elbow Method show that K equal 4 is the best K on our clustering.



Elbow Method show that K equal 4 is the best K on our clustering.



5. Discussion section

The bar charts indicate that cluster 2 (Scarborough and North York) of neighborhoods has the highest average number of Chinese restaurants. The second-highest number is in cluster 3 (Etobicoke and Scarborough). The best place to start a Chinese restaurant is in Parkwoods, North York (cluster 1) because not only do they have many neighborhoods in the borough but there are fewer Chinese Restaurants in the area. Don Mills South, North York (cluster 4) is the second suitable location for opening the venue category.

6. Conclusion

In the project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, exploratory data analysis and imply machine learning to find the answer and give the recommendations to the people who might interest. The findings of this project will help the relevant inventor better understand the advantages and disadvantages of different Toronto neighbourhoods/ boroughs in terms of opening a Chinese restaurant.