Chinese Restaurant in Toronto Neighborhood

Predicting the Profitable Neighborhood to open a chinese restaurant

Ravali Chinthala

26.02.2020

1. INTRODUCTION

1.1 Background

Chinese cuisine is an important part of Chinese culture, which includes cuisine originating from the diverse regions of China, as well as from Chinese people in other parts of the world. Because of the Chinese diaspora and historical power of the country, Chinese cuisine has influenced many other cuisines in Asia, with modifications made to cater to local palates. Chinese food staples such as rice, soy sauce, noodles, tea, and tofu, and utensils such as chopsticks and the wok, can now be found worldwide.

1.2 Problem

Though Chinese recipes are served by many restaurants, there will be some authentic flavour missing most of the time. So, people tend to go for dine out or look out to order from chinese restaurants to have all flavours packed in dishes. But, is it profitable to open these restaurants in a random neighborhood ? No, We should consider the neighborhood and interests of people nearby.

1.3 Interest

- Stakeholders or people interested in Food & Beverage business would be interested to find out profitable areas.
- Investors would be interested to invest capital amounts in such restaurants.
- Data Scientists who analyze Neighborhoods would like to know such details.

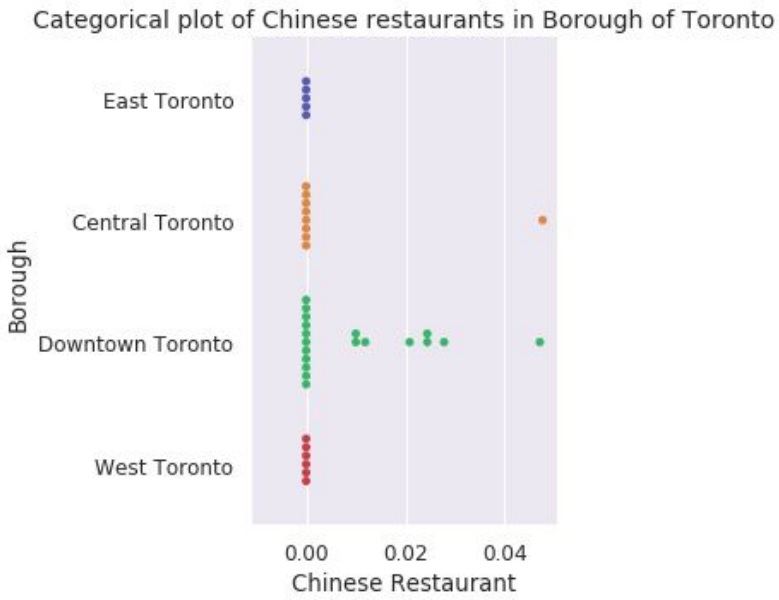
2. DATA SOURCES

- Toronto City data that contains list Boroughs, Neighborhoods from Wiki https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- We get geolocation including latitude, longitude and Postcode from http://cocl.us/Geospatial_data
- Using FourSquare API , we get all venue details in Neighborhood . It helps us in clustering neighborhood.

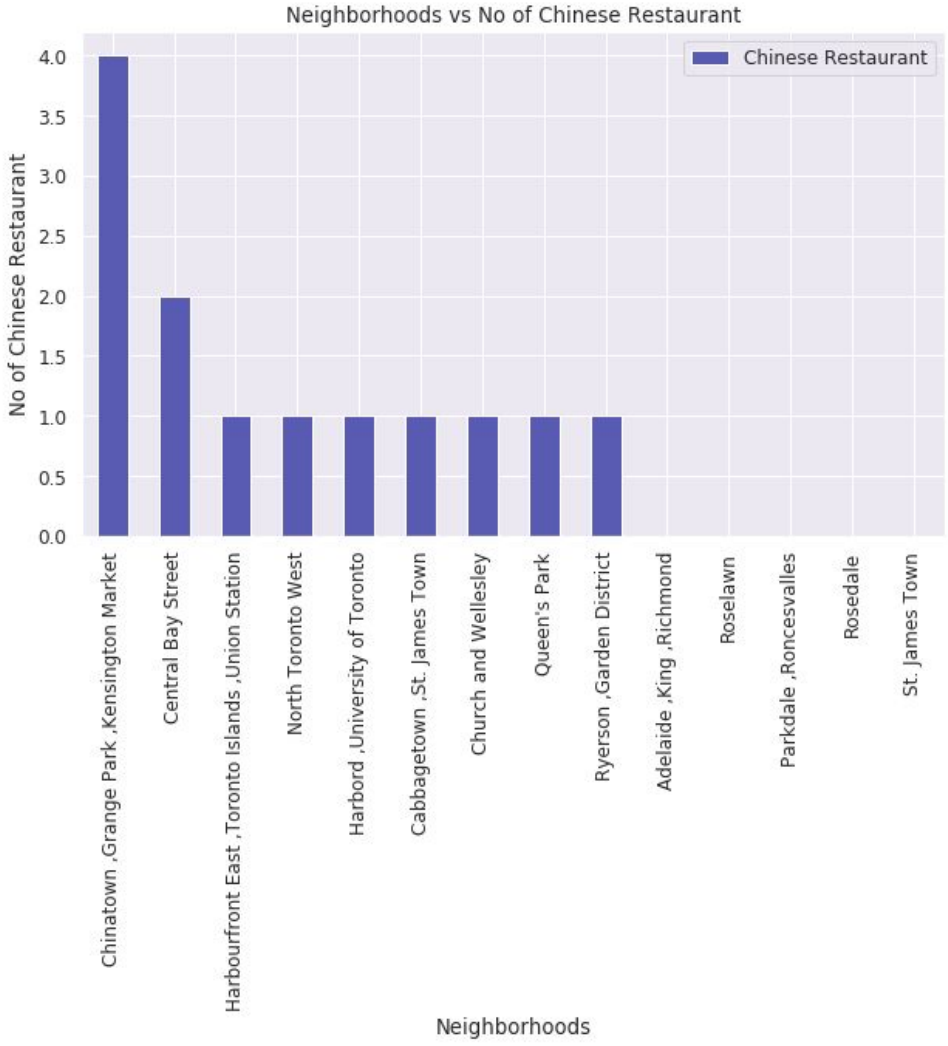
3. EXPLORATORY DATA ANALYSIS

Data analysis helps us to find any inconsistencies around data and thus will not be biased in drawing insights. Let us create a map of Toronto using latitude and longitude values of boroughs & Neighborhood.

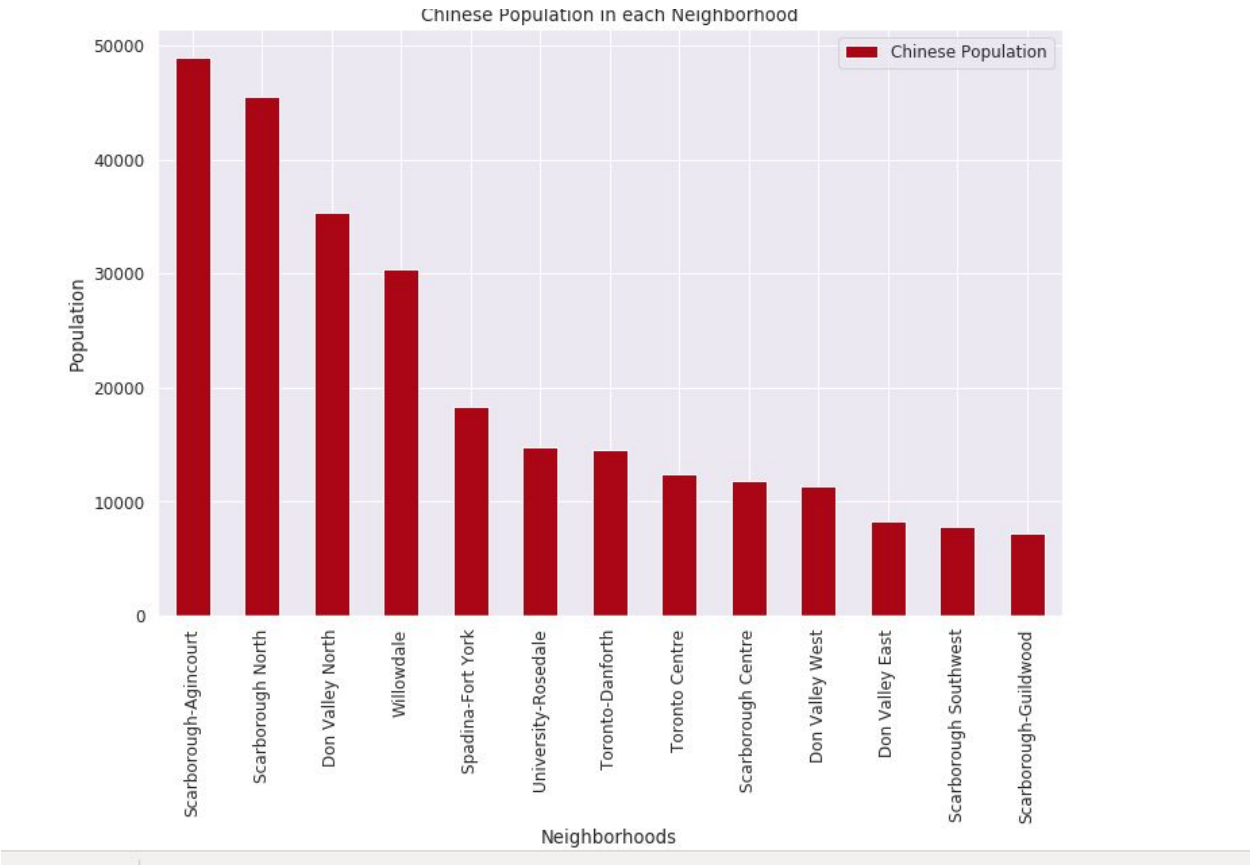
Correlation between variables is one of the important metrics that has to be computed. Let us see the relationship between neighborhoods and Chinese restaurants .



This shows that currently chinese restaurants are dense in the area of Downtown Toronto where they are very sparse in other areas.



This graph shows that Chinese restaurants are densely located in the neighborhoods of Chinatown, Grange park, Kensington market . Also, there are no such restaurants in Adelaide , Richmond ,Roselawn, Rosedale etc., neighborhoods.

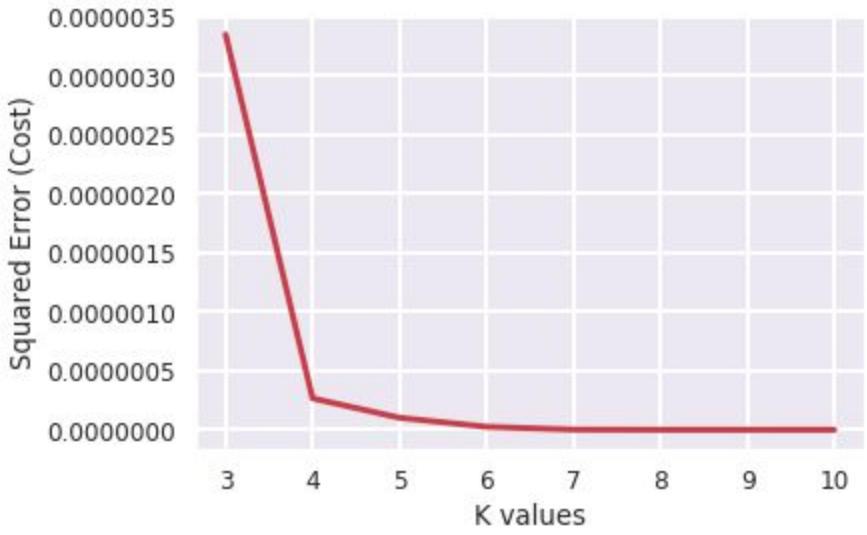


This graph shows the relationship between Chinese ethnic origin Population and neighborhoods.

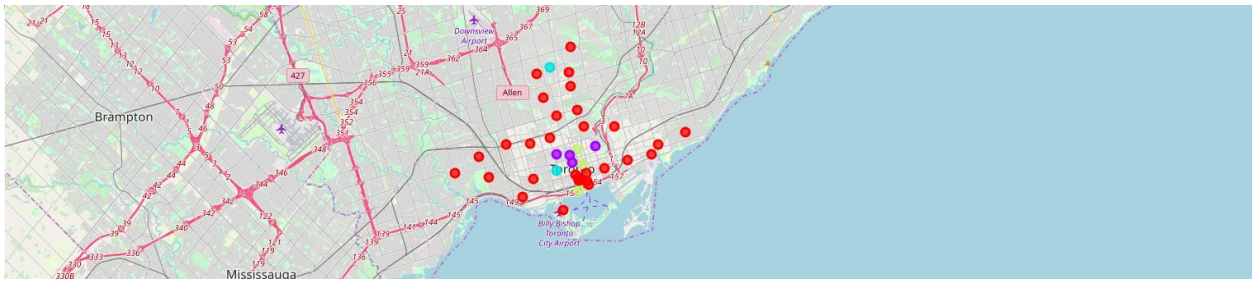
PREDICTIVE MODELING

Since it is unlabelled data, we can do only modeling like segmentation and clustering. In this scenario, we are using k-means clustering. Initially, we need to come up with best k, i.e., best no.of clusters that data can be formed into.

We are using squared error cost against each K value from 3 to 13, then plotting at which K , model has lower error cost.



From the above graph, we have observed that ideal K is 4. We have total of 4 clusters such as 0,1,2,3. Let us examine the graph to see what each cluster refers to.



1. Red colored points belong to cluster 0, where they are neighborhoods with no chinese restaurants.
2. Purple colored points belong to cluster 1, where neighborhoods with sparsely populated chinese restaurants.
3. Blue colored points belong to cluster 2, where neighborhoods with densely populated chinese restaurants.
4. Yellow colored points belong to cluster 3, where neighborhoods with medium population.

RESULTS

From the analysis we have performed till now , from clustering & visualization of the data, as the business problem started with identifying a good neighborhood to open a new Chinese restaurant, we looked into all the neighborhoods in Toronto, analysed the Chinese population in each neighborhood & spread of Chinese restaurants in those neighborhoods to come to conclusion about which neighborhood would be a better spot for opening a new Chinese restaurant. I have used data from web resources like Wikipedia, geospatial coordinates of Toronto neighborhoods, and Foursquare API, to set up a very realistic data-analysis scenario. We have found out that —

In those 11 boroughs we identified that only Central Toronto, Downtown Toronto ,North York & Scarborough boroughs have a high amount of Chinese restaurants with the help of Swarm categorical plots between Number of Chinese restaurants in Borough of Toronto.

In all the ridings, Scarborough-Guildwood, Scarborough-Rouge Park, Scarborough Centre, Scarborough North, Humber River-Black Creek, Don Valley East, Scarborough Southwest, Don Valley North & Scarborough-Agincourt are densely populated with Chinese crowd ridings.

With the help of clusters examining & categorical plots looks like Downtown Toronto, Central Toronto, East York are already densely populated with Chinese restaurants. So it is better idea to leave those boroughs out and consider only Scarborough, East Toronto & North York for the new restaurant's location.

After careful consideration it is a good idea to open a new Chinese restaurant in Scarborough borough since it has a high number of Chinese population which gives a higher number of customers possibility and lower competition since very less Chinese restaurants in the neighborhoods.

CONCLUSION

Finally to conclude this project, We have got a chance to solve a business problem like how a real like data scientists would do. We have used many python libraries to fetch the

data , to manipulate the contents & to analyze and visualize those datasets. We have made use of Foursquare API to explore the venues in neighborhoods of Toronto, then get a good amount of data from Wikipedia which we scraped with help of Wikipedia python library and visualized using various plots present in seaborn & matplotlib. We also applied machine learning techniques to predict the output given the data and used Folium to visualize it on a map. Also, some of the drawbacks or areas of improvements show us that this analysis can further be improved with help more data and different machine learning techniques. Similarly we can use this project to analyze any scenario such as opening a different cuisine or success of opening a new gym, etc. Hopefully, this project helps acts as initial guidance to take more complex real-life challenges using data-science.