



# Capstone Project - The Battle of Neighbourhoods

Exploring Toronto  
Neighbourhoods for a New  
Chinese Restaurant Location  
Recommendation



# Introduction

In this project are using web data and Foursquare location data to explore different neighbourhoods of Toronto to figure out the potential neighbourhoods to start a new restaurant business. In this project, we will go through our analysis from problem designing, data preparation to the final conclusion that can be leveraged by the business stakeholders to make their decisions.

# Business Problem & Discussion

## **Possibility of opening an Chinese Restaurant in Toronto, Canada with Recommended Location:**

Chinese population in Canada grew beginning in the 1960s due to changes in immigration law and political issues in Hong Kong. Additional immigration from Southeast Asia in the aftermath of the Vietnam War and related conflicts and a late 20th century wave of Hong Kong immigration led to the further development of Chinese ethnic enclaves in the Greater Toronto Area. The Chinese established many large shopping centres and restaurants in suburban areas catering to their ethnic group.

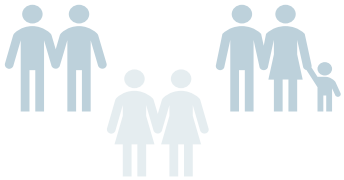
Choose a good location to open a new Chinese restaurant is not easy job without exploring and analysing the ethnic group areas and location of current opened Chinese restaurant.

# Methodology



Restaurants in Toronto

## Foursquare Data



Chinese ethnic group areas  
& Neighbourhood list

## Wikipedia Data



Neighborhoods Coordinates

## Cocl.us Data

1. Clean and correlate data of Toronto neighborhoods which is scrapped from Wikipedia and their coordinates from Cocl.us.
2. Clean scrapped data related to Chinese ethnic groups and where they are concentrated in Toronto districts after scrapping related tables from Wikipedia
3. Identify the Boroughs which containing the Chinese population
4. Get all Chinese restaurant in Toronto per each neighborhood
5. Using Kmean algorithm to clustering the restaurants data into group of different clusters which reflect the density of restaurants per neighborhood/Borough.
6. Identify the neighborhoods that has high Chinese population and less number of Chinese restaurants that are considered potential for our business case



Data Collection, Cleansing, Preparation and EDA



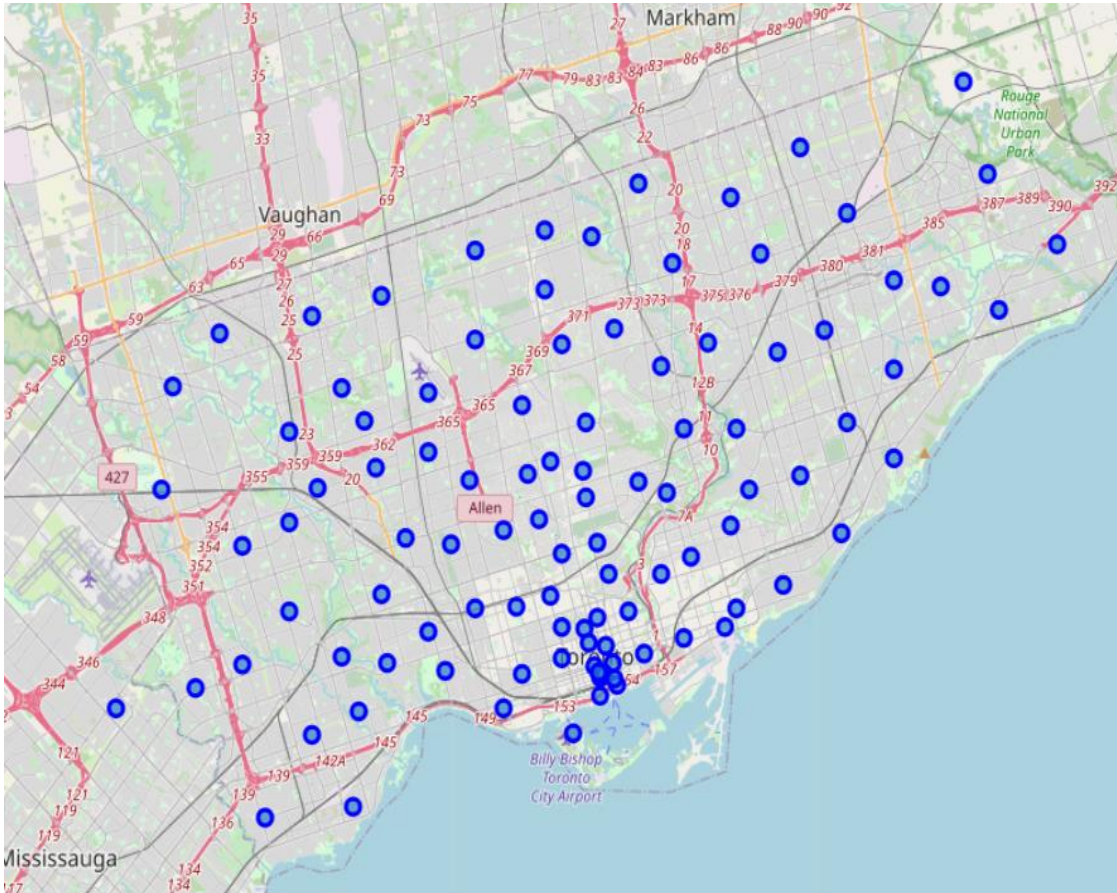
Model Selection, training and deployment



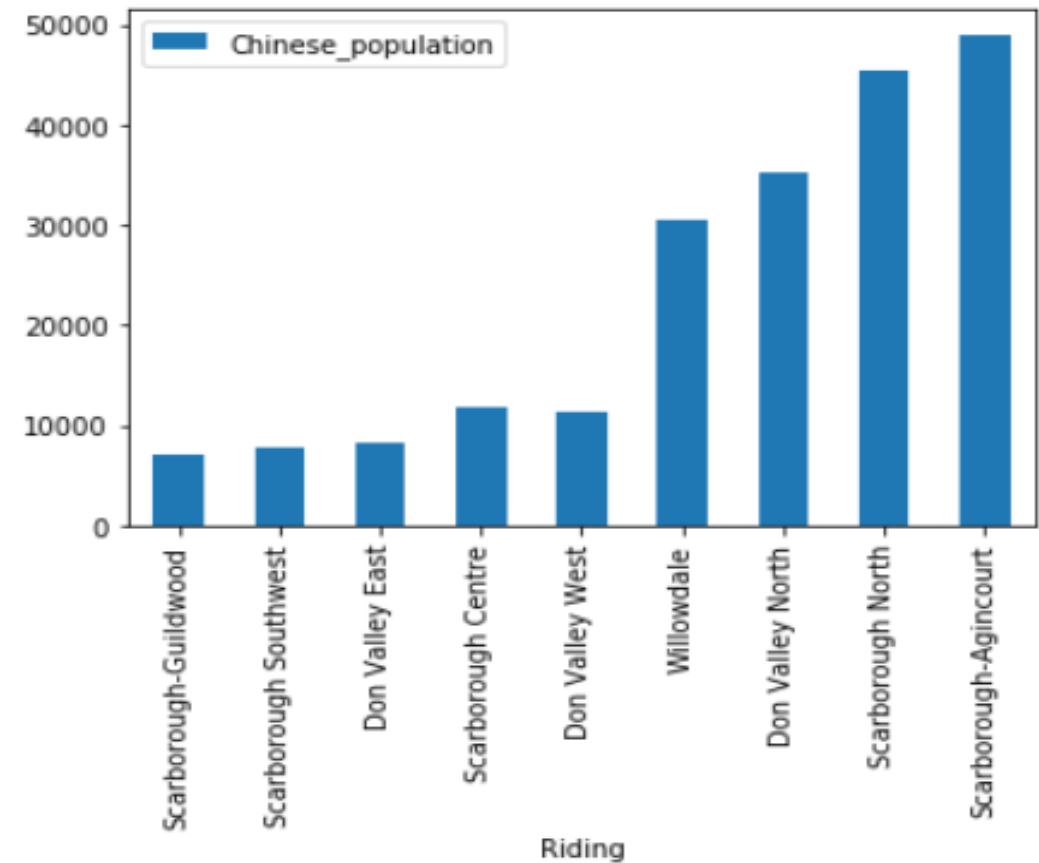
Output and Conclusion

# EDA

## Toronto Neighborhoods



## Chinese Population Per Riding In Toronto



# EDA

## Top 10 Common Venues Per Neighborhood

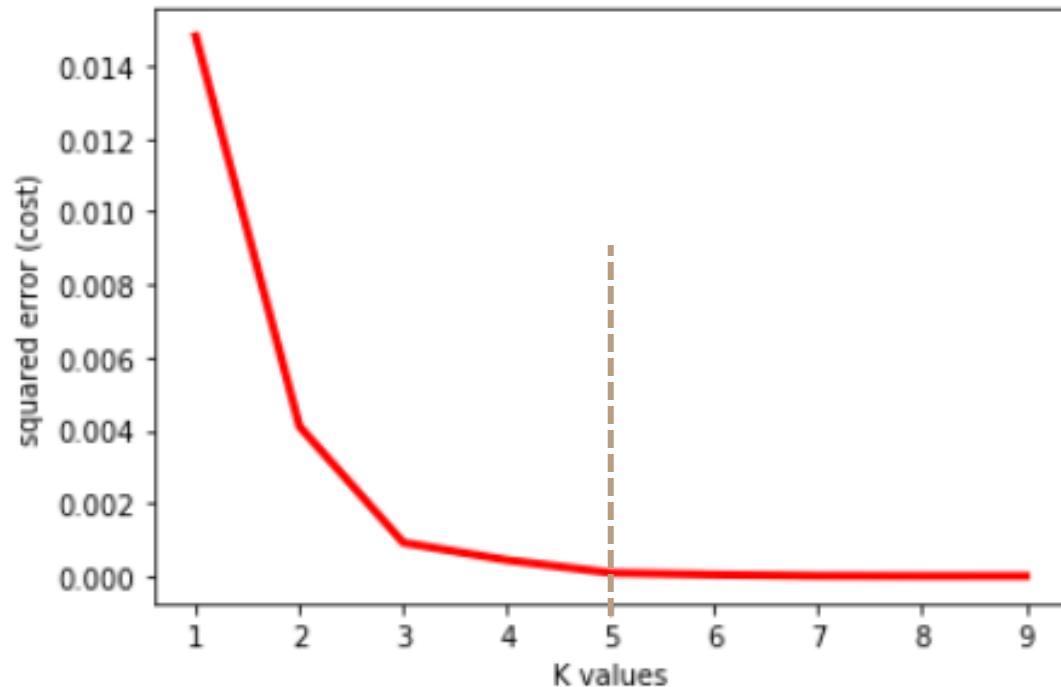
"Sample of 5 rows"

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Latin American Restaurant	Vietnamese Restaurant	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
1	Bathurst Manor / Wilson Heights / Downsview North	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
2	Bayview Village	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
3	Bedford Park / Lawrence Manor East	Restaurant	Italian Restaurant	Sushi Restaurant	Indian Restaurant	Comfort Food Restaurant	Greek Restaurant	American Restaurant	Thai Restaurant	Seafood Restaurant	German Restaurant
4	Berczy Park	Restaurant	Seafood Restaurant	Comfort Food Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Eastern European Restaurant	Portuguese Restaurant	Greek Restaurant	Thai Restaurant

# Modelling

## Clustering Neighbourhoods of Toronto:

First step in K-means clustering is to identify optimum K value to identify the number of clusters in a given dataset. So we will use elbow method on the Toronto dataset with Chinese restaurant percentage to identify K value.

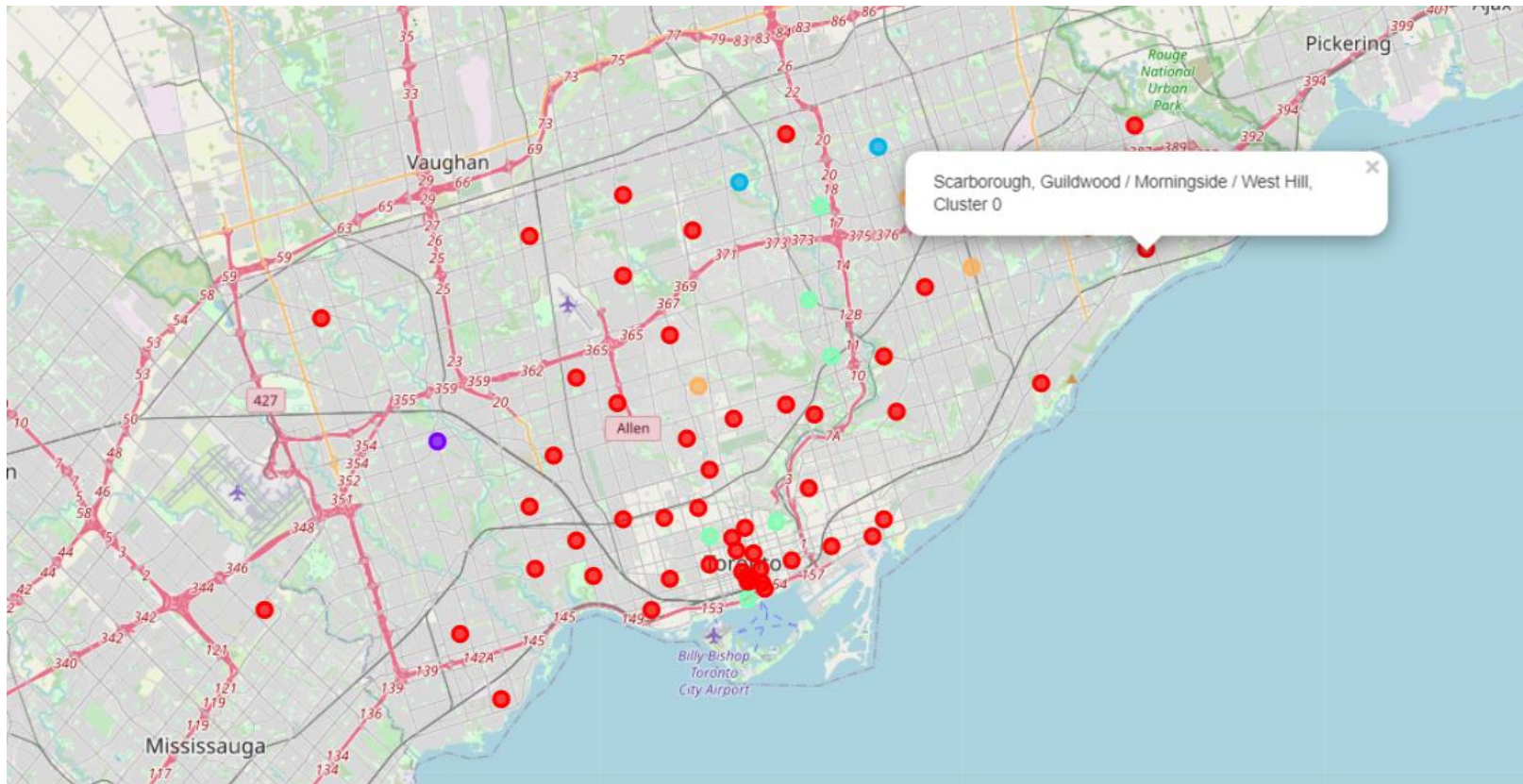


As per shown in the elbow figure 5 is the optimum number of cluster that give optimum squared error value



# Modelling

## Model Output:



## 5 Clusters are shown in the Map

- Cluster 1 represent the least and zero Chinese restaurant density neighborhood
- Cluster 2 represent the highest Chinese restaurant density neighborhood
- Cluster 3 represent above medium Chinese restaurant density neighborhood
- Cluster 4 represent low Chinese restaurant density neighborhood
- Cluster 5 represent medium Chinese restaurant density neighborhood



# Modelling

## Model Output:

	Borough	Neighborhood	Cluster Labels	Chinese Restaurant
1	North York	Victoria Village	0	0.000000
2	Downtown Toronto	Regent Park / Harbourfront	0	0.000000
3	North York	Lawrence Manor / Lawrence Heights	0	0.000000






	Borough	Neighborhood	Cluster Labels	Chinese Restaurant
70	Etobicoke	Westmount	1	1.0

	Borough	Neighborhood	Cluster Labels	Chinese Restaurant
39	North York	Bayview Village	2	0.5
90	Scarborough	Steeles West / L'Amoreaux West	2	0.4

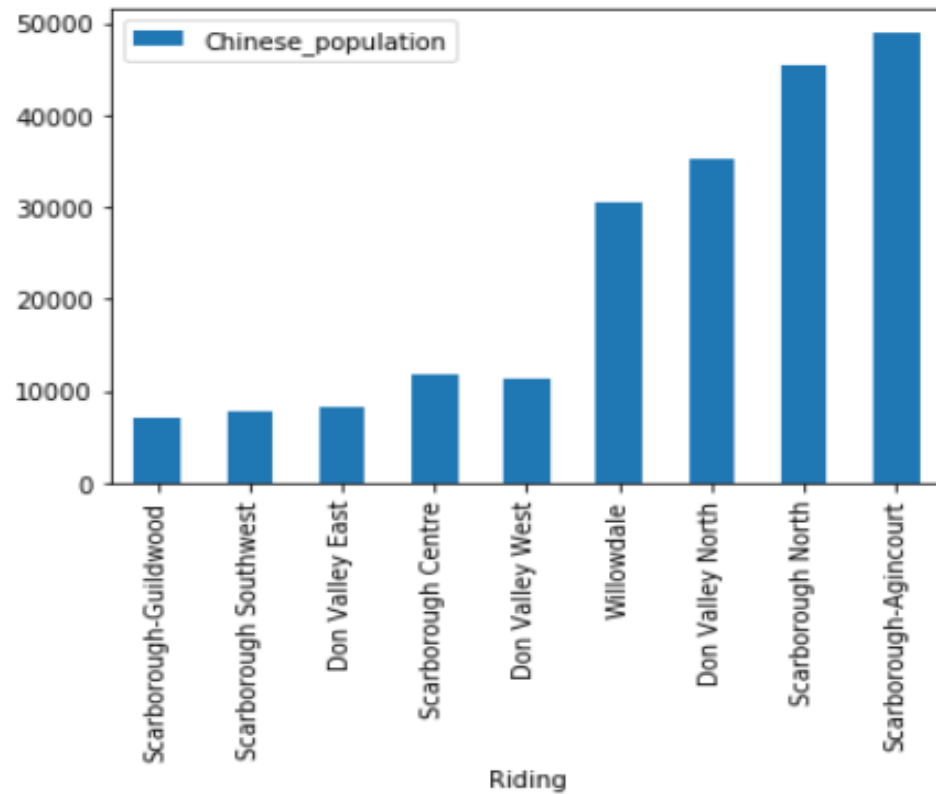
	Borough	Neighborhood	Cluster Labels	Chinese Restaurant
7	North York	Don Mills	3	0.100000
13	North York	Don Mills	3	0.100000
33	North York	Fairview / Henry Farm / Oriole	3	0.076923

	Borough	Neighborhood	Cluster Labels	Chinese Restaurant
65	Scarborough	Dorset Park / Wexford Heights / Scarborough To...	4	0.25
73	Central Toronto	North Toronto West	4	0.25
82	Scarborough	Clarks Corners / Tam O'Shanter / Sullivan	4	0.25

## 5 Clusters are shown in the Map

-  Cluster 1 represent the least and zero Chinese restaurant density neighborhood
-  Cluster 2 represent the highest Chinese restaurant density neighborhood
-  Cluster 3 represent above medium Chinese restaurant density neighborhood
-  Cluster 4 represent low Chinese restaurant density neighborhood
-  Cluster 5 represent medium Chinese restaurant density neighborhood

# Results and Discussion

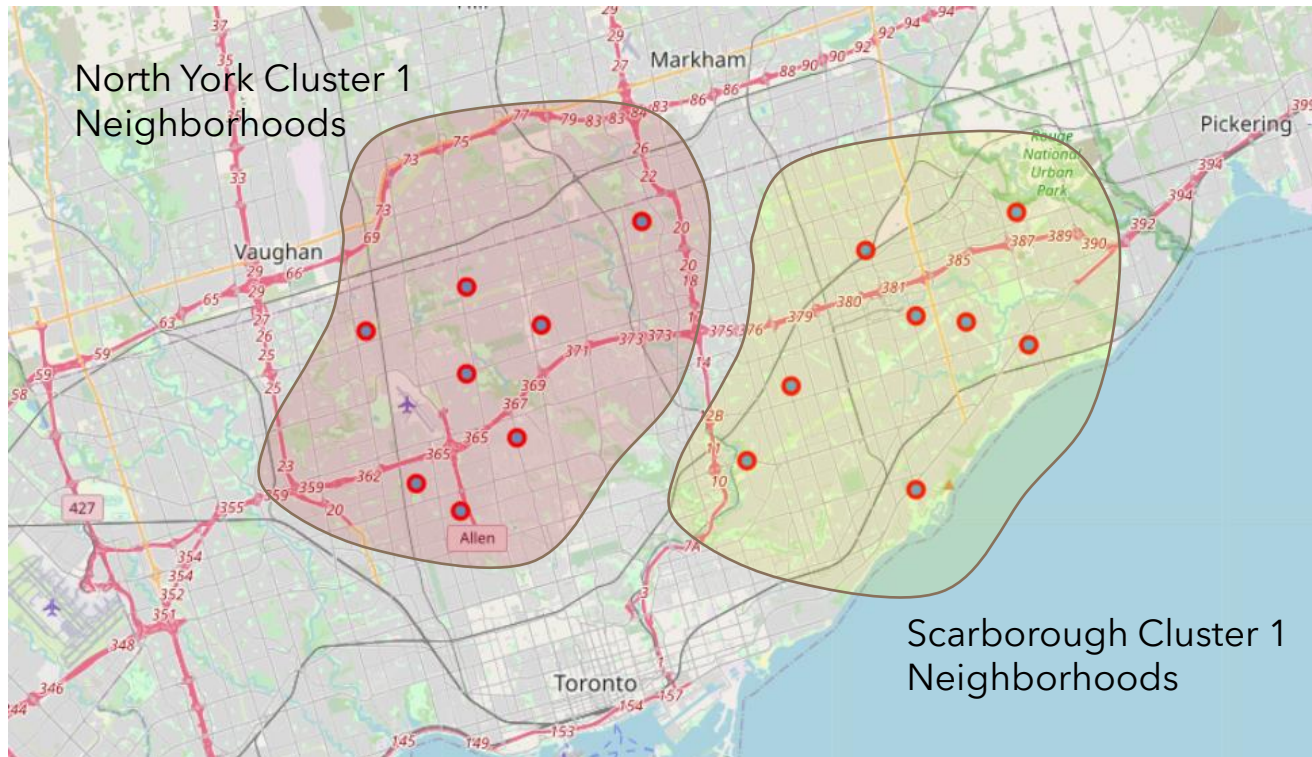


Riding	Population	Chinese_population	Chinese Population %	Riding_Borough
Scarborough-Guildwood	101115	7179.165	7.1	Scarborough
Scarborough Southwest	108295	7797.240	7.2	Scarborough
Don Valley East	93170	8292.130	8.9	North York
Scarborough Centre	110450	11818.150	10.7	Scarborough
Don Valley West	101790	11400.480	11.2	North York
Willowdale	117405	30407.895	25.9	North York
Don Valley North	109060	35335.440	32.4	North York
Scarborough North	97610	45486.260	46.6	Scarborough
Scarborough-Agincourt	104225	48985.750	47.0	Scarborough

Riding_Borough	Population	Chinese_population	Chinese Population %
North York	421425	85435.945	20.273108
Scarborough	521695	121266.565	23.244724

All Shown Ridings are belonging to North York and Scarborough and Scarborough has more Chinese population

# Results and Discussion



Borough	Neighborhood
Scarborough	Malvern / Rouge
Scarborough	Guildwood / Morningside / West Hill
Scarborough	Woburn
Scarborough	Cedarbrae
Scarborough	Cliffside / Cliffcrest / Scarborough Village West
Scarborough	Wexford / Maryvale
Scarborough	Agincourt
North York	Victoria Village
North York	Lawrence Manor / Lawrence Heights
North York	Glencairn
North York	Hillcrest Village
North York	Bathurst Manor / Wilson Heights / Downsview North
North York	Northwood Park / York University
North York	Bedford Park / Lawrence Manor East
North York	Willowdale

**Recommended Neighborhoods where we have low density of Chinese restaurants and high Chinese population**  
**Priority 1 would be given to Scarborough Neighborhood where it has higher population**



# Conclusion

In This Project we have defined a business problem. We have used many python libraries to collect and clean the data , to manipulate the contents and to analyse and visualize those datasets. We have made use of Foursquare API to explore the venues in neighbourhoods of Toronto, then get 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 technique to predict the output given the data and used Folium to visualize it on a map.

we can use this project to analysis any scenario such as opening a different cuisine restaurant or opening of a new gym and etc. Hopefully, this project helps acts as initial guidance to take more complex real-life challenges using data-science.



# Thanks

---