# Optimal Chinese Restaurant locations in Toronto

**Charlse Jiang**

*2020.5.3*

# 1. Introduction

In this project, we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Chinese restaurant in the city of Toronto, Canada.

In order to achieve our goal, we will try to detect locations with low density of restaurants. Besides, locations with less or even no Chinese restaurants will seem to be great choices. Finally, population of a district will also have a great impact on restaurants operating.

We will use our data science skills to get and analyze restaurants and neighborhoods data in Toronto and finally get some advices on choosing the optimal location.

# 2. Data

## 2.1 Data source

Our data mainly come from Toronto Open Data Portal.
In our analysis, we mainly get use of Boundaries of Toronto Neighborhoods from https://open.toronto.ca/dataset/neighbourhoods/, which we can get a districts distribution of every neighborhood in Toronto and their coordinates information
And another data we used is Toronto Neighborhood Profiles from https://open.toronto.ca/dataset/neighbourhood-profiles/, which we can get population information of every neighborhood.

## 2.2 Data wrangling

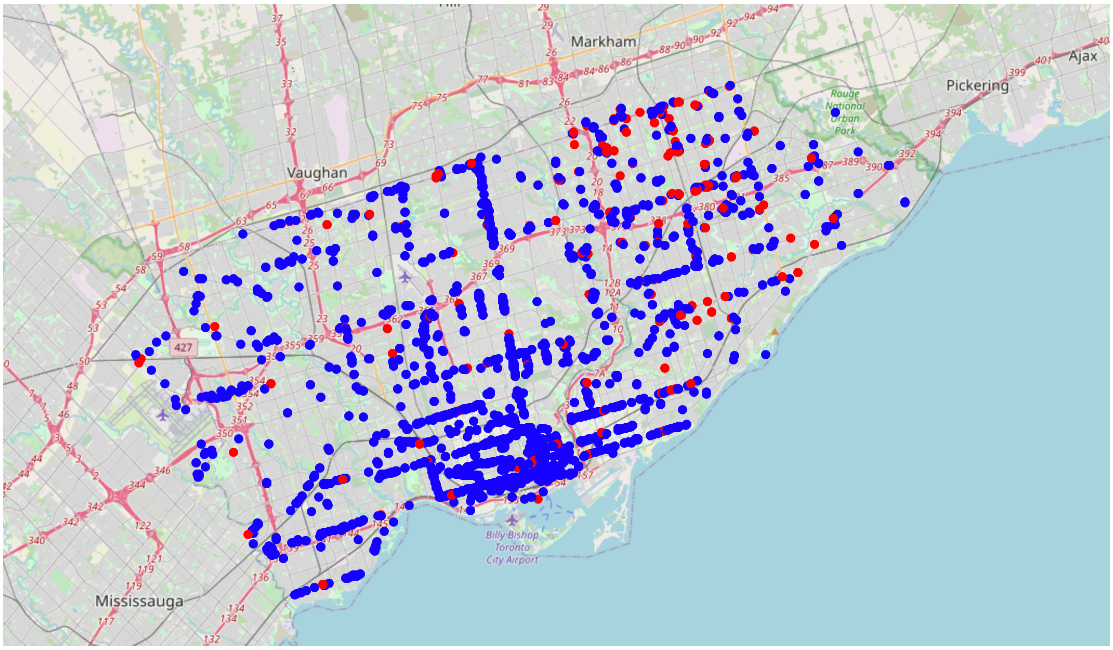
First we get a straight view of Toronto neighborhoods from Toronto Open Data Portal, and we only seize those we need for analyzing:

| AREA_SHORT_CODE | AREA_NAME | LATITUDE | LONGITUDE | Shape__Area |
|---|---|---|---|---|
| **94** | Wychwood (94) | 43.676919 | -79.425515 | 3.217960e+06 |
| **100** | Yonge-Eglinton (100) | 43.704689 | -79.403590 | 3.160334e+06 |
| **97** | Yonge-St.Clair (97) | 43.687859 | -79.397871 | 2.222464e+06 |
| **27** | York University Heights (27) | 43.765736 | -79.488883 | 2.541821e+07 |
| **31** | Yorkdale-Glen Park (31) | 43.714672 | -79.457108 | 1.156669e+07 |

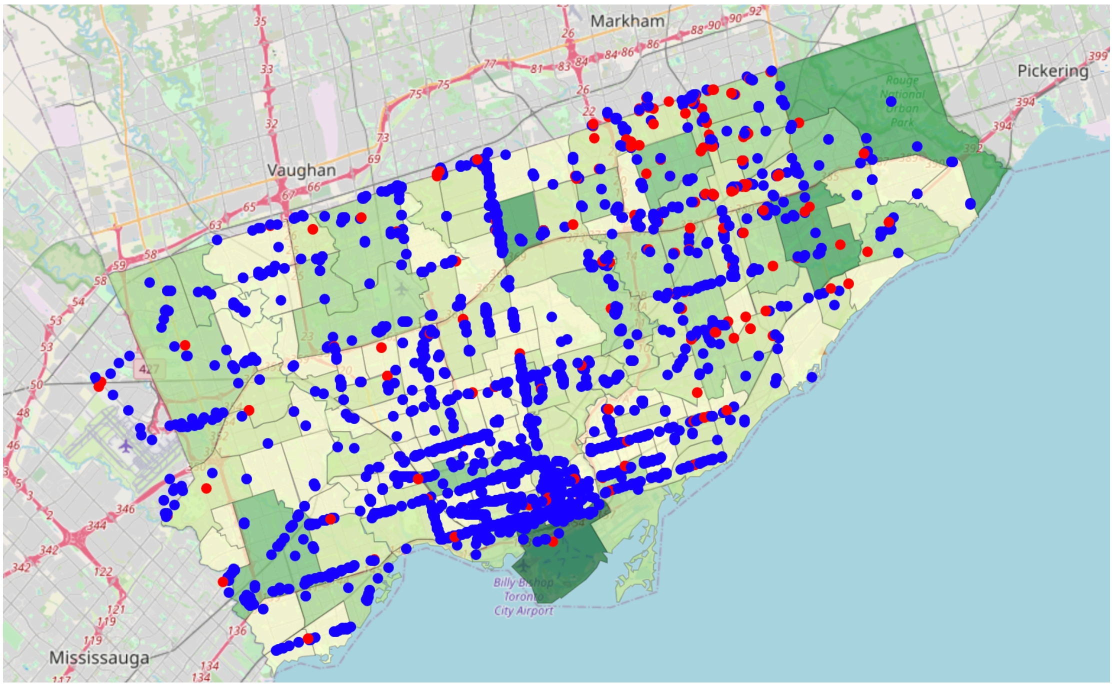
And then we rename the dataframe and create a radius column for exploring neighborhood with:

| Neighborhood_id | Neighborhood | Latitude | Longitude | Area | Radius |
|---|---|---|---|---|---|
| **94** | Wychwood (94) | 43.676919 | -79.425515 | 3.217960e+06 | 1012.081201 |
| **100** | Yonge-Eglinton (100) | 43.704689 | -79.403590 | 3.160334e+06 | 1002.978341 |
| **97** | Yonge-St.Clair (97) | 43.687859 | -79.397871 | 2.222464e+06 | 841.089976 |
| **27** | York University Heights (27) | 43.765736 | -79.488883 | 2.541821e+07 | 2844.445038 |
| **31** | Yorkdale-Glen Park (31) | 43.714672 | -79.457108 | 1.156669e+07 | 1918.799364 |

After that, we utilize Foursquare api to explore restaurant venues near every neighborhood:




Finally, we add population of every neighborhood into our dataframe and draw a choropleth map to mark every restaurant venues:




# 3. Mothedology

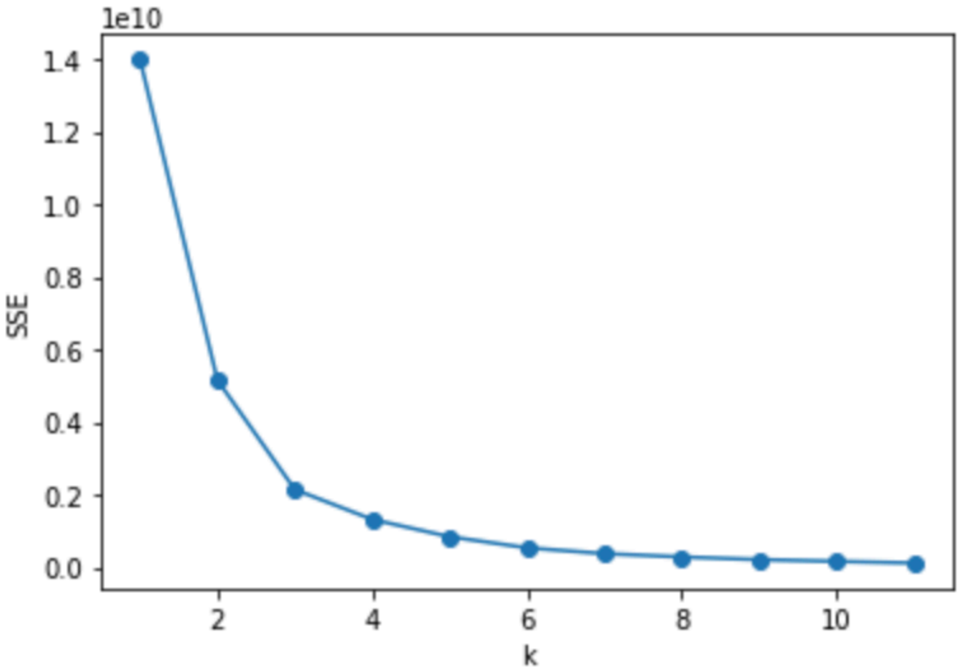
We will use KMeans Clustering algorithm to get our final results.
First, let's merge our neighborhoods data and restaurants data together.

| Neighborhood_id | Neighborhood | Latitude | Longitude | Area | Radius | Population | Restaurants | Chinese Restaurants |
|---|---|---|---|---|---|---|---|---|
| 94 | Wychwood (94) | 43.676919 | -79.425515 | 3.22 | 1.01 | 14349 | 43 | 2 |
| 100 | Yonge-Eglinton (100) | 43.704689 | -79.403590 | 3.16 | 1.00 | 11817 | 63 | 1 |
| 97 | Yonge-St.Clair (97) | 43.687859 | -79.397871 | 2.22 | 0.84 | 12528 | 31 | 2 |
| 27 | York University Heights (27) | 43.765736 | -79.488883 | 25.42 | 2.84 | 27593 | 58 | 5 |
| 31 | Yorkdale-Glen Park (31) | 43.714672 | -79.457108 | 11.57 | 1.92 | 14804 | 51 | 2 |

And then we will build two columns by dividing restaurant counts by area to get density of restaurants of every neighborhood.

| Neighborhood_id | Neighborhood | Latitude | Longitude | Area | Radius | Population | Restaurants | Chinese Restaurants | Restaurants/km2 | Chinese Restaurants/km2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 94 | Wychwood (94) | 43.676919 | -79.425515 | 3.22 | 1.01 | 14349 | 43 | 2 | 13.354037 | 0.621118 |
| 100 | Yonge-Eglinton (100) | 43.704689 | -79.403590 | 3.16 | 1.00 | 11817 | 63 | 1 | 19.936709 | 0.316456 |
| 97 | Yonge-St.Clair (97) | 43.687859 | -79.397871 | 2.22 | 0.84 | 12528 | 31 | 2 | 13.963964 | 0.900901 |
| 27 | York University Heights (27) | 43.765736 | -79.488883 | 25.42 | 2.84 | 27593 | 58 | 5 | 2.281668 | 0.196696 |
| 31 | Yorkdale-Glen Park (31) | 43.714672 | -79.457108 | 11.57 | 1.92 | 14804 | 51 | 2 | 4.407952 | 0.172861 |

After that, we can fit our model now. And let' s first choose a proper k value.




From the figure above we can see that when k = 3 we get an elbow, so we will fit a 3 means clustering models.

Let' s see descriptive information of different clusters:

Cluster 0:

| | Latitude | Longitude | Area | Radius | Population | Restaurants | Chinese Restaurants | Restaurants/km2 | Chinese Restaurants/km2 | labels |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.000000 | 85.0 |
| **mean** | 43.700922 | -79.411510 | 5.766118 | 1.296471 | 13257.352941 | 20.529412 | 1.329412 | 5.293562 | 0.277900 | 0.0 |
| **std** | 0.045213 | 0.099258 | 3.586378 | 0.395621 | 3137.136079 | 17.187058 | 2.072428 | 5.976644 | 0.389955 | 0.0 |
| **min** | 43.592362 | -79.580445 | 0.810000 | 0.510000 | 6577.000000 | 1.000000 | 0.000000 | 0.202224 | 0.000000 | 0.0 |
| **25%** | 43.670886 | -79.485589 | 3.200000 | 1.010000 | 10732.000000 | 7.000000 | 0.000000 | 1.336898 | 0.000000 | 0.0 |
| **50%** | 43.694526 | -79.415342 | 4.750000 | 1.230000 | 13362.000000 | 16.000000 | 1.000000 | 2.947368 | 0.161031 | 0.0 |
| **75%** | 43.728489 | -79.351260 | 7.850000 | 1.580000 | 15873.000000 | 28.000000 | 2.000000 | 6.885246 | 0.358423 | 0.0 |
| **max** | 43.802988 | -79.150843 | 16.940000 | 2.320000 | 18675.000000 | 64.000000 | 15.000000 | 23.706897 | 1.970443 | 0.0 |

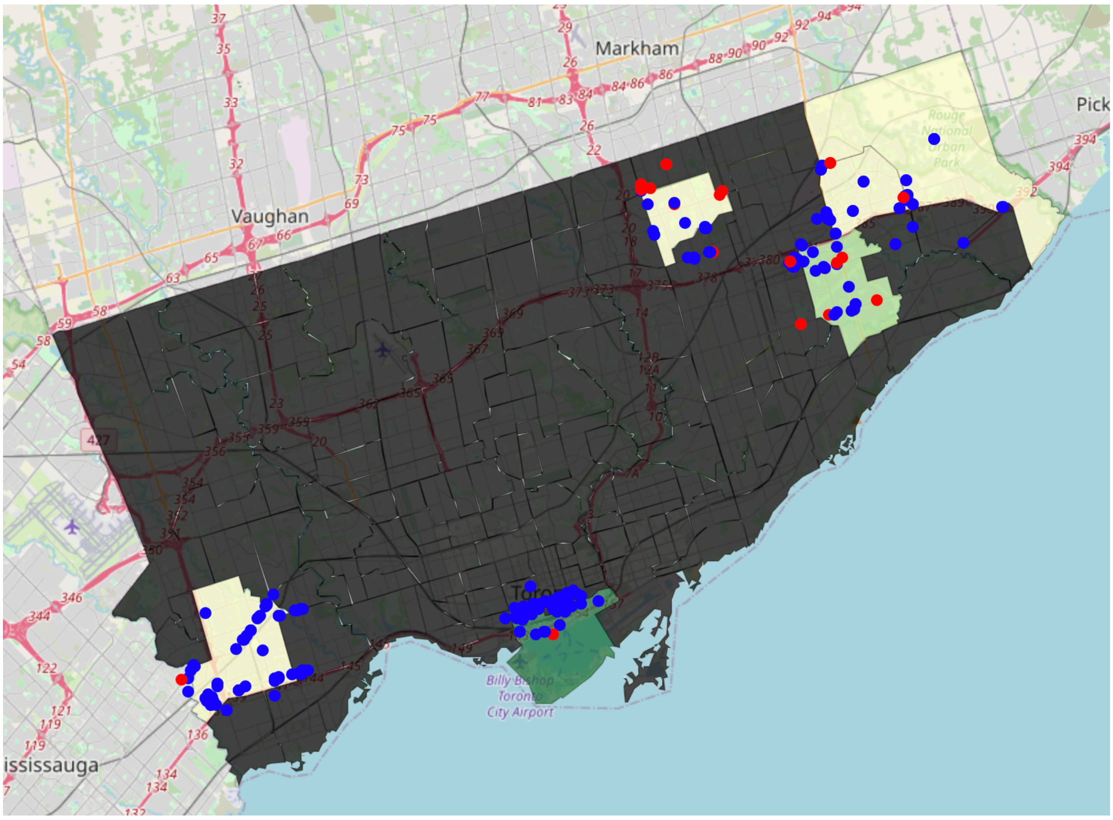
Cluster 1:

| | Latitude | Longitude | Area | Radius | Population | Restaurants | Chinese Restaurants | Restaurants/km2 | Chinese Restaurants/km2 | labels |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.0 |
| **mean** | 43.746466 | -79.324790 | 27.562857 | 2.815714 | 49725.714286 | 39.714286 | 3.714286 | 2.160997 | 0.219281 | 1.0 |
| **std** | 0.079289 | 0.126082 | 20.971600 | 0.990603 | 8044.959575 | 19.206894 | 3.093773 | 1.880402 | 0.227235 | 0.0 |
| **min** | 43.633463 | -79.543317 | 9.710000 | 1.760000 | 43794.000000 | 19.000000 | 1.000000 | 0.263377 | 0.027724 | 1.0 |
| **25%** | 43.700310 | -79.389343 | 15.385000 | 2.210000 | 43979.000000 | 22.500000 | 1.500000 | 1.377337 | 0.035622 | 1.0 |
| **50%** | 43.770602 | -79.314084 | 23.660000 | 2.740000 | 46496.000000 | 36.000000 | 3.000000 | 1.746725 | 0.176160 | 1.0 |
| **75%** | 43.799687 | -79.225551 | 28.330000 | 3.000000 | 51959.500000 | 58.000000 | 5.000000 | 2.091503 | 0.302409 | 1.0 |
| **max** | 43.821201 | -79.186343 | 72.140000 | 4.790000 | 65913.000000 | 62.000000 | 9.000000 | 6.179197 | 0.655022 | 1.0 |

Cluster 2:

| | Latitude | Longitude | Area | Radius | Population | Restaurants | Chinese Restaurants | Restaurants/km2 | Chinese Restaurants/km2 | labels |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.000000 | 48.0 |
| **mean** | 43.717376 | -79.391126 | 11.418542 | 1.803958 | 26179.500000 | 37.020833 | 4.187500 | 4.989689 | 0.409700 | 2.0 |
| **std** | 0.054042 | 0.099979 | 9.039544 | 0.621117 | 4446.639022 | 21.163487 | 6.289661 | 5.630044 | 0.512000 | 0.0 |
| **min** | 43.615924 | -79.596356 | 2.570000 | 0.900000 | 20506.000000 | 4.000000 | 0.000000 | 0.378788 | 0.000000 | 2.0 |
| **25%** | 43.671452 | -79.466695 | 6.187500 | 1.400000 | 22160.500000 | 17.250000 | 1.000000 | 1.736254 | 0.086862 | 2.0 |
| **50%** | 43.716055 | -79.382695 | 9.695000 | 1.760000 | 25424.000000 | 39.000000 | 2.000000 | 3.578019 | 0.217467 | 2.0 |
| **75%** | 43.761497 | -79.302089 | 14.202500 | 2.122500 | 29249.250000 | 54.250000 | 5.000000 | 5.222148 | 0.509082 | 2.0 |
| **max** | 43.820691 | -79.176676 | 57.750000 | 4.290000 | 36625.000000 | 76.000000 | 31.000000 | 25.287356 | 2.622577 | 2.0 |

Now with a descriptive information of our clusters above, we can obviously find that data in cluster 2 have lowest average density of both restaurants and chinese restaurants, and still a high level of average population, which perfectly meet our demands
Finally let's show our final data on a chorolepth map.




# 4. Results and Discussion

We have analyze all 140 neighborhoods in the city of Toronto with venues around them, specially focusing on restaurants and chinese restaurants. we use KMeans Clustering algorithm to determine which neighborhoods have lowest density of both restaurants and chinese restaurants and also a higher level of population. And finally we've got a cluster of 7 neighborhoods which meet our demands. Within those 7 neighborhoods we find there being a outlier so we dropped it and finally we got 6 neighborhoods which can be optimal locations for opening a new restaurants.
Yet this doesn't mean that these 6 neighborhoods are definately most ideal location for a new restaurant to open because we just focus on 3 main elements we thought to be critical.

Therefore, there could be some other neighborhoods can also be a proper loction for a new restaurant.
Besides, in this project radius we used to explore a neighborhood is somehow not that adequate so it may lead to some biases.

## 5. Conclusion

Purpose of this project was to identify Toronto neighborhoods with low number of restaurants (particularly Chinese restaurants) and high population in order to aid stakeholders in narrowing down the search for optimal location for a new Chinese restaurant. By exploring restaurant venues distribution of every neighborhood of Toronto, we use KMeans Clustering algorithm to find out neighborhoods with lowest average density of restaurants (and Chinese restaurants) and also population in that neighborhood, and finally get 6 optimal locations for opening a new chinese restaurant.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.