# **Introduction**

One of my friends and her husband live in Toronto. They want to do some investments. They prefer to start a small business, because they don't want invest too much money. They all love food and have good cooking skills for Chinese foods. It seems like that opening a Chinese restaurant is a good idea for them.

Location is extremely important for opening a restaurant. But Toronto is so big that investigating all the neighborhoods is time consuming and expensive. It’s not possible to investigate all of them on site.

The purpose of my job is to find out some neighborhoods that have the potential to be a good choice, so that can reduce their workloads.

# **Data Sources**

Demographics Information for neighborhoods in Toronto comes from website. I decide to scrape data from Wikipedia page (<https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods>)

Geographic information comes from package ‘geopy.geocoders’, I use neighborhoods’ name to find their latitude and longitude.

Venues’ information is from Foursquare. The required data is as follows:

1. Chinese restaurants information in each neighborhood

2. Total restaurants information in each neighborhood

3. Information about all venues in each neighborhood

4. Population Density for each neighborhood

# **Methodology**

There are three main factors for location selecting.

First, the neighborhood must have a large population density. It's very straightforward that larger population density means more potential customers. I will scrape population from Wikipedia page (https://en.wikipedia.org/wiki/Demographics\_of\_Toronto\_neighbourhoods)

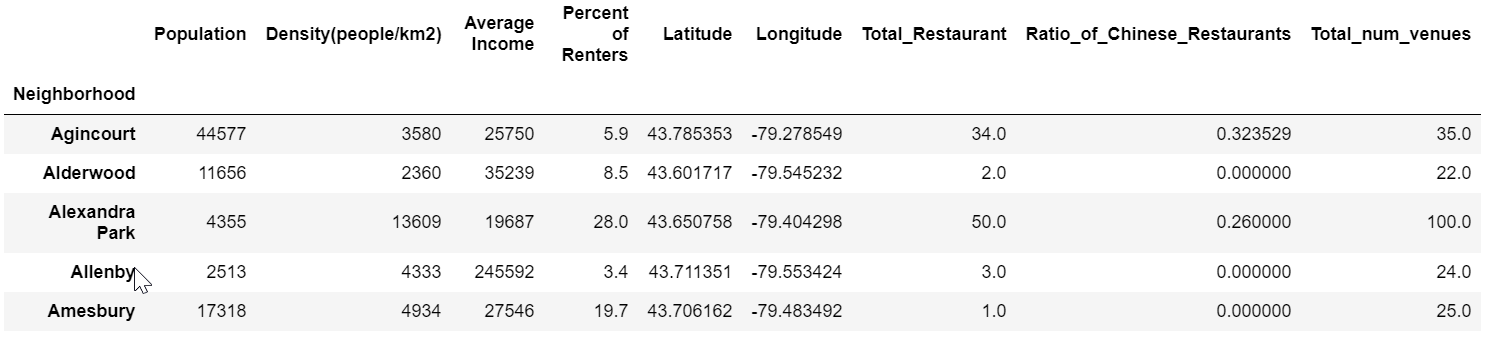
Second, the neighborhood needs to have strong commercial atmosphere. I will use the amount of total venues within 1000 meters radius of each neighborhood to represent business atmosphere. The higher the amount, the stronger the atmosphere.

Third, there can't be too many competitors in the neighborhood. I will use ratio of the number of Chinese restaurants over the amount of total restaurant. If the ratio is small, the competition in that neighborhood might be small.

After getting all the information I need, I will do a clustering to classify all the neighborhoods. Then, I’ll find out which cluster has high population density, strong commercial atmosphere, low competition. Neighborhoods in this cluster are the candidate for on-site investigation.

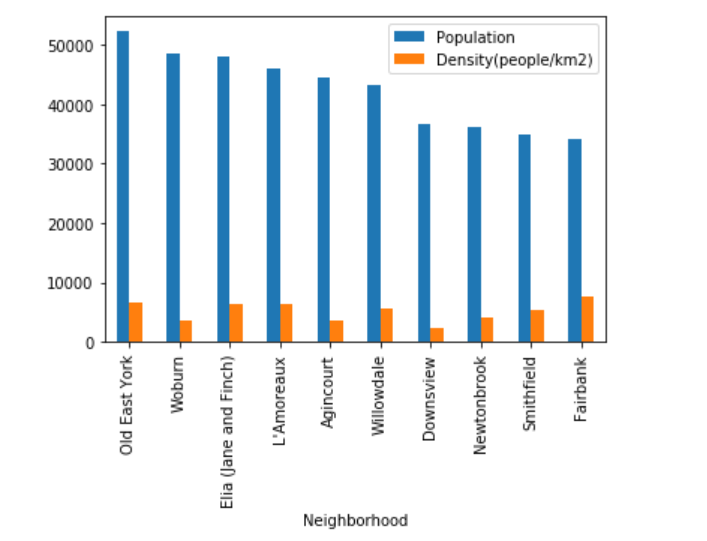
## **Data set**

After scraping the internet, querying forsquare, and doing some transformation, I get a 159\*9 dataset. The first 5 rows of my dataset are as fellow:

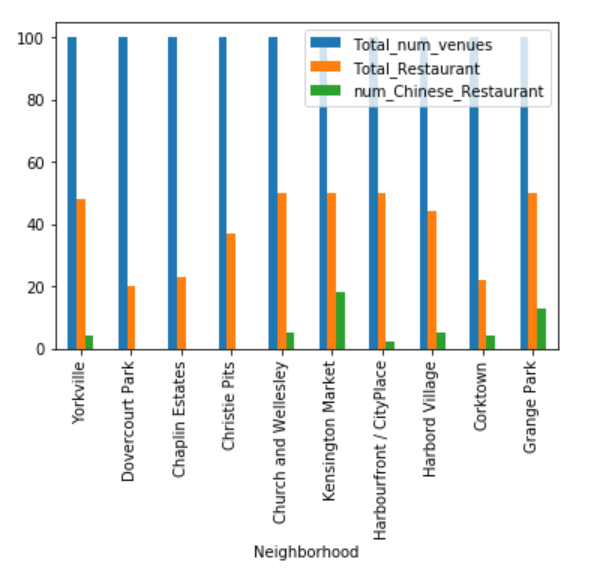


## **Bar Chart**

I also draw some bar charts to get a first impression of the dataset.



We can see that high Population doesn’t mean high population density.



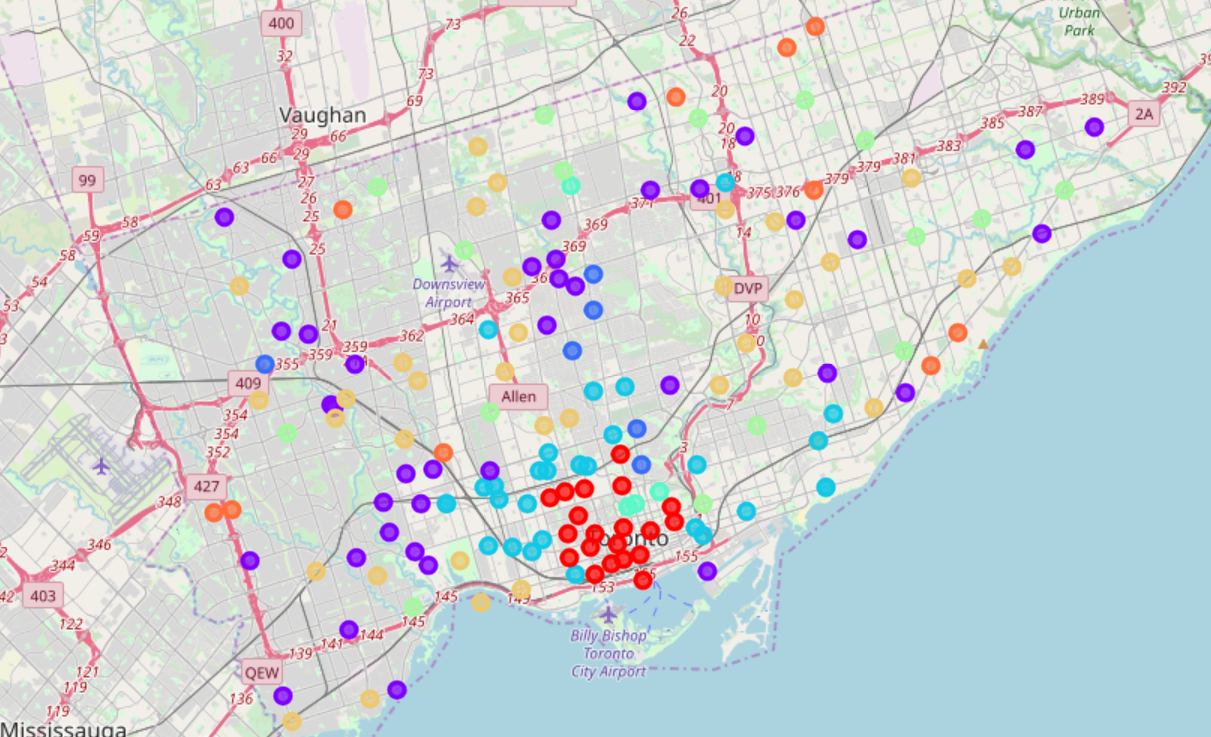
From this graph, we can see that restaurant always make up a large port of all venues. More venue always means more customer. So, the ideal neighborhood maybe is similar to Christie Pits which has many venues but doesn’t process any Chinese restaurant. But we still need to take population into account.

## **Clustering**

I choose K-means clustering. Because I don’t what one cluster contains too many neighborhoods, I set the number of cluster equal to 8.

# **Results**

The map for clustering is as follows:



# **Discussion**

It very hard to decide whether a neighborhood is suitable, but if our intuitive criterion make sense, we can find out which cluster is best for opening a restaurant.

# **Conclusion**

I find that neighborhood in cluster 5 have the same attributes: have high population, high density, and high number of venues. The ratio of Chinese restaurant is relatively low.

All the attributes are good for opening a Chinese restaurant.

So, I recommend neighborhoods in cluster 5 which are 'Agincourt', 'Bendale', 'Don Valley Village', 'Downsview', 'Fairbank', 'L'Amoreaux', 'Newtonbrook', 'Old East York', 'Richview', 'Riverdale', 'Scarborough Junction', 'Smithfield', 'West Hill', ' Willowdale', 'Woburn', 'York University Heights' as the promising places for opening a Chinese restraurant.