**capstone project – battle of neighborhoods**

**Background**

Toronto has a population of 2.9 million, being the most populous cities in Canada. As one of the most diverse places in the planet, Toronto has been permeated by foods from different regions and countries. Today, Toronto has over 8,000 restaurants, and people can find Chinese Restaurants, French restaurants, Indian Restaurants, and any other cuisine they can imagine.

Considering the advantages of cultural diversity and census, some individuals seek their fortunes by starting a restaurant. Although cuisines and service are important to a restaurant, other factors like residents' preference and income, transportation and competition exert great influence on the revenue of a restaurant.

**Business Problem and Target Audience**

Can we make an ideal decision on the location of a brand-new restaurant by analyzing data of neighborhoods and other restaurants?

For the reason, the capstone is aiming at finding promising sites for an Asian restaurant. In this capstone, an ideal site will be in a neighborhood with a large census, high average income, a high proportion of Asian and accessible parking and transportation. Lastly, the distribution of similar restaurants should be low. These advantages increase the likelihood of attracting affordable consumers to an Asian restaurant. The list of ideal sites can be valuable to any individual or company who want to start running an Asian Restaurant.

**Data**

1. the latitude and longitude coordinates of each neighborhood
2. the neighborhood profile including population of working age, the number of people whose mother tongue is Japanese, Chinese Languages, and Korean
3. the number and distribution of parking, bus stops and metro stations of each neighborhood
4. the number and distribution of Asian restaurants (Chinese, Japanese, Korean)

* The geographical information of neighborhoods can be scraped by using BeautifulSoup from Wikipedia <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* geographical information of neighborhoods can be provided by coursera. <https://cocl.us/Geospatial_data>
* The 2016 Neighbourhood Profilea data set can be download from the City's [Open Data Portal,www.toronto.ca/open](https://www.toronto.ca/city-government/data-research-maps/open-data/)
* With the coordinates of neighborhoods, the transportation information and the number of Asian restaurants is possible to get via Foursquare API.

**Analyzing Methodology**

Step1: collect clean data

Step2: merge neighborhoods’ features of population and the proportion of Asian residents with the transportation information and show on map

Step3: cluster neighborhoods into 3 groups, aiming at finding one cluster (cluster 1) with a large number of affordable Asian people and convenient transportation.

Step4: find out the number of Asian Restaurants and analyze their cuisines and popularity in neighborhoods in cluster 1

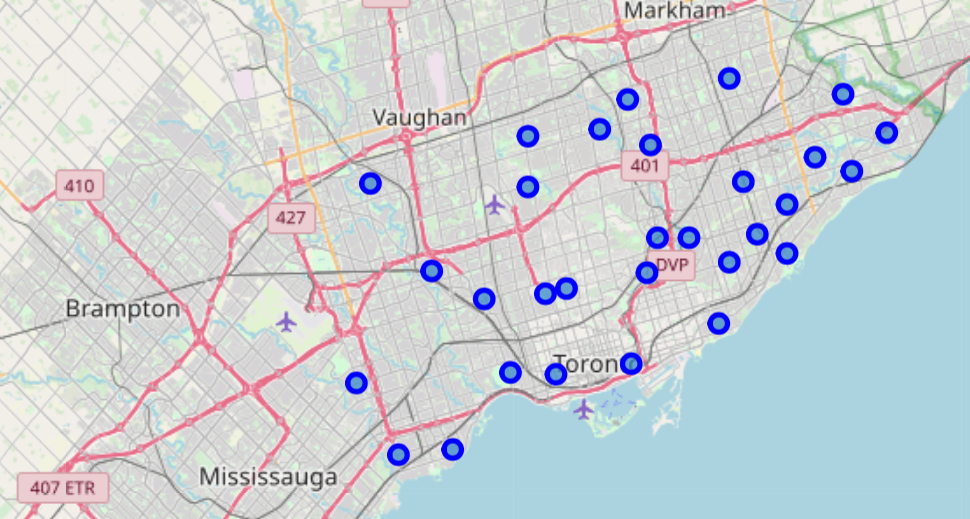
Step5: plot the distribution of popular Asian Restaurants and transportation information, to find potential site of a new Asian restaurant

Language: python (anaconda jupyter notebook)

Data Scraping Libraries: Request, BeautifulSoup, etc.

Data preprocessing Libraries: pandas, numpy, etc.

Plotting Libraries: maplotlib, folium, etc.

Machine learning methods: clustering methods like sciki-learn k-means

figure

Figure 1 shows the distribution of neighborhoods. Boroughs includes Scarborough, North York, East Toronto, Central Toronto, Downtown Toronto, York, West Toronto, Etobicoke