IBM
Capstone Project

IBM Capstone Project (Week 4). This is the main project for the IBM Data Science Professional certification. IBM Data Science Project Week 4

Intro. One of my friends is an oriental man who owns a Chinese food restaurant in the area. He is well versed in Chinese food, knows how to cook and loves his business, but I am sure that if he started a business in a place where there are a noticeable percentage of Chinese immigrants than here, his business would be more effective. I will imagine that he decided to move to a large multinational city, let it be Toronto(Canada). Let me remind you that Toronto has a large percentage of immigrants from China (https://worldpopulationreview.com/world-cities/toronto-population)

Business problem. For those who would try to open a business in a new place, there is always an acute problem of locating their point of the service sector or institution. A solution to this problem will be found using data science tools.

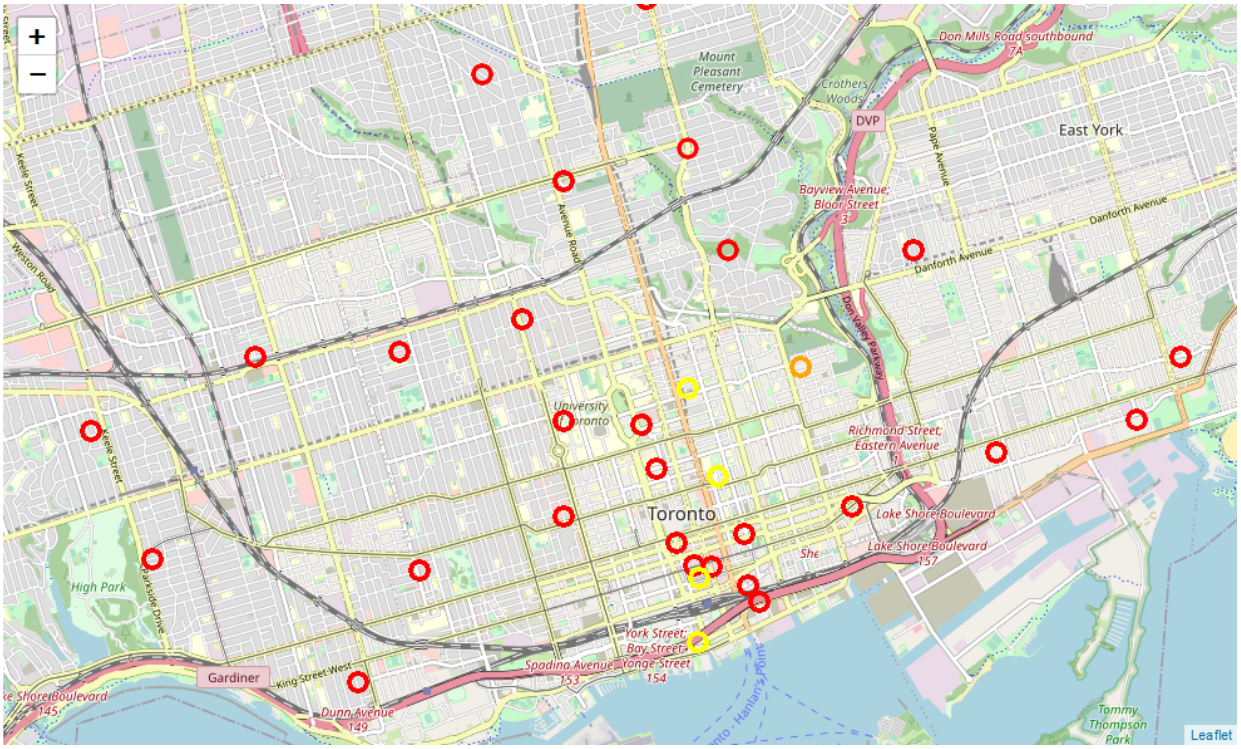
A person who would be interested in the result obtained is an entrepreneur who is going to open a Chinese restaurant.

Data and sources. Next data will help to solve this problem:

-List of neighborhoods in Toronto(Canada).(Wikipedia)
-Latitude and Longitude ofneighborhoods.(Geocoder)
-Chinese restaurantvenuedata.(Foursquare API)

Solution. So first of all we need to get a list of neighborhoods in Toronto Canada. To do this, we refer to the article containing the relevant information on Wikipedia (originally the link was https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, but we had to refer to the previous version of the table because someone had changed it radically). I used Beautiful Soup to extract the table from the page and parse it to turn it into a data frame. The next step was to assign coordinates to the neighborhood using the Geocoder package.

We now turn to using the Foorscuare API to compile a list of the top 100 establishments within a 500 meter radius. It is important here that in addition to the name and coordinates, we can understand the category of the institution. The frequency of occurrence of locations by category in each neighborhood. Now we apply the k-means clustering method, which distributes each data point to the closest cluster. As a result, after applying the method, we group the districts into 3 types of clusters according to the frequency of occurrence of restaurants with Chinese food.



So on the map we can see a breakdown by clusters, where:

- Cluster 0 (red) area without Chinese restaurants
- Cluster 1 (orange) with few Chinese restaurants
- Cluster 2 (yellow) with a prominent presence of Chinese restaurants

As you can see, most of the Chinese restaurants are located along the Church and Wellesley axis towards the Harbourfront. The presence in the Cabbagetown area is also noticeable.

We consider clusters 0 as a recommended area for opening a Chinese restaurant, here you can advise the direction from Berczy Park to Regent Park