Where to build Chinese Restaurant?

Faidhi Hakim

Aug 1, 2021

1. Introduction

1.1 Background
In this project we will try to find the most optimal place to build a restaurant. A lot of businesses always missed the essential key points of a good business location. Good location can significantly boost a company's long term performance. Poor ones can cost millions in lost talent, productivity and capital. Thus, it will be a huge value if the business can identify an optimal location for a startup (specifically a restaurant in our case). Specifically this report will focus on opening a Chinese restaurant in Toronto, Canada.

1.2 Problem
What are the characteristics of a good business location? Is there any empty spaces or land for the business to be started?

1.3 Interest
We will be targeting areas that do not have Chinese restaurants yet. Obviously, business managers would be interested in knowing which would be a great area. Others who might be interested may be the government for census, or the council for taxes.

2. Data Acquisition

The postal code, borough and neighbourhood can be found in the [Postal Code of Toronto Ontario - Canada Postal Code List](#) website. It is the second table and we need to scrape it using bs4 packages.

3. Data Cleaning

Firstly, the data contains empty rows. Secondly, we need to remove the data because every string contains " \n", " ". Thirdly, we noticed that some of the data have the unicode '\xa0', so we had to remove that one as well.

Next, we also had to join neighbourhoods under the same postal code because the dataframe that contains latitude and longitude does so. However, before we do that, some of the neighbourhoods; Willowdale (M2M, M2N, M2R), Downsview (M3M, M3L, M3N) and Islington

Avenue (M9A, M9B) which contains different postal codes in one row. Thus, we fix this by creating additional rows depending on the number of the multiple postal codes.

Then, we moved on to make the postal codes (every row) that contain all neighbourhoods which belong to that postal code and name it 'gu'.

After that, we merged the two dataframe; one from the website (gu), another from the given dataframe from Coursera (hi).

3.1 Feature Selection

After data cleaning, only one row from 'hi' was omitted. This is M3C postal code, because it is not present in 'df'.

4. Exploratory Data Analysis

We use the Foursquare API for this section, to call on nearby venues. We are interested in the category of the venues, especially chinese restaurants. Before that it's good for us to visualize the 102 postal codes using folium on the map.
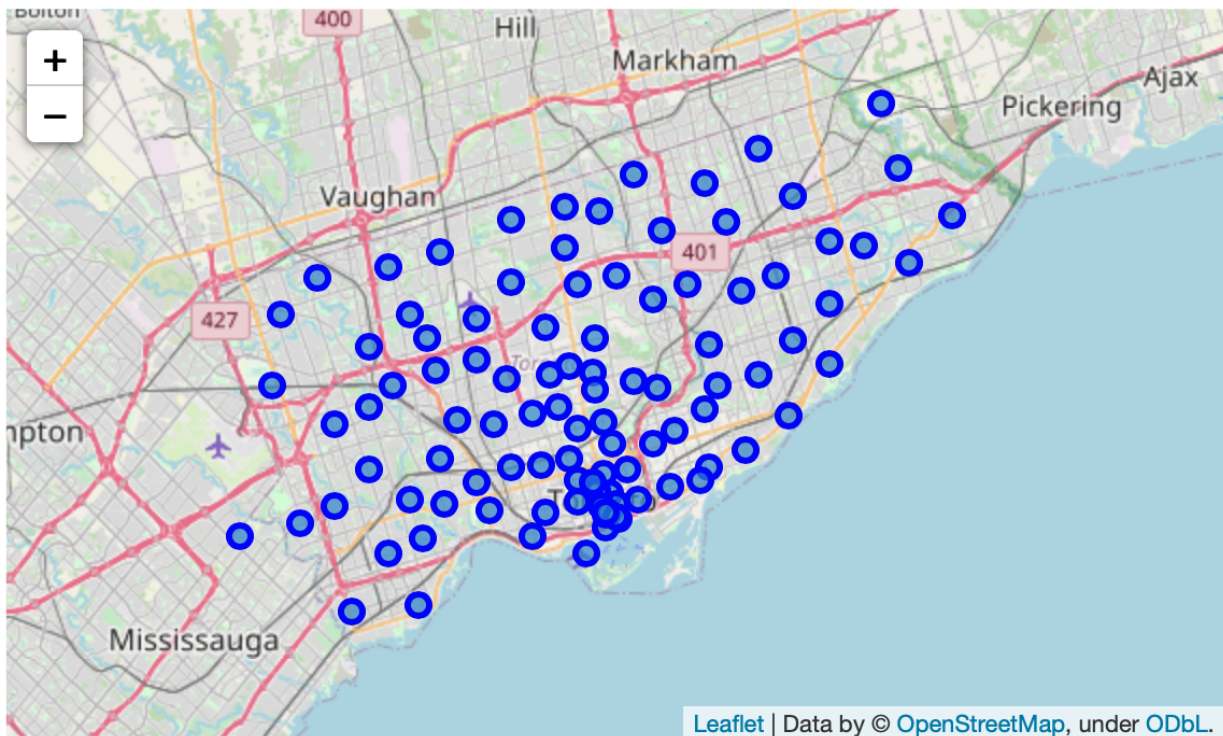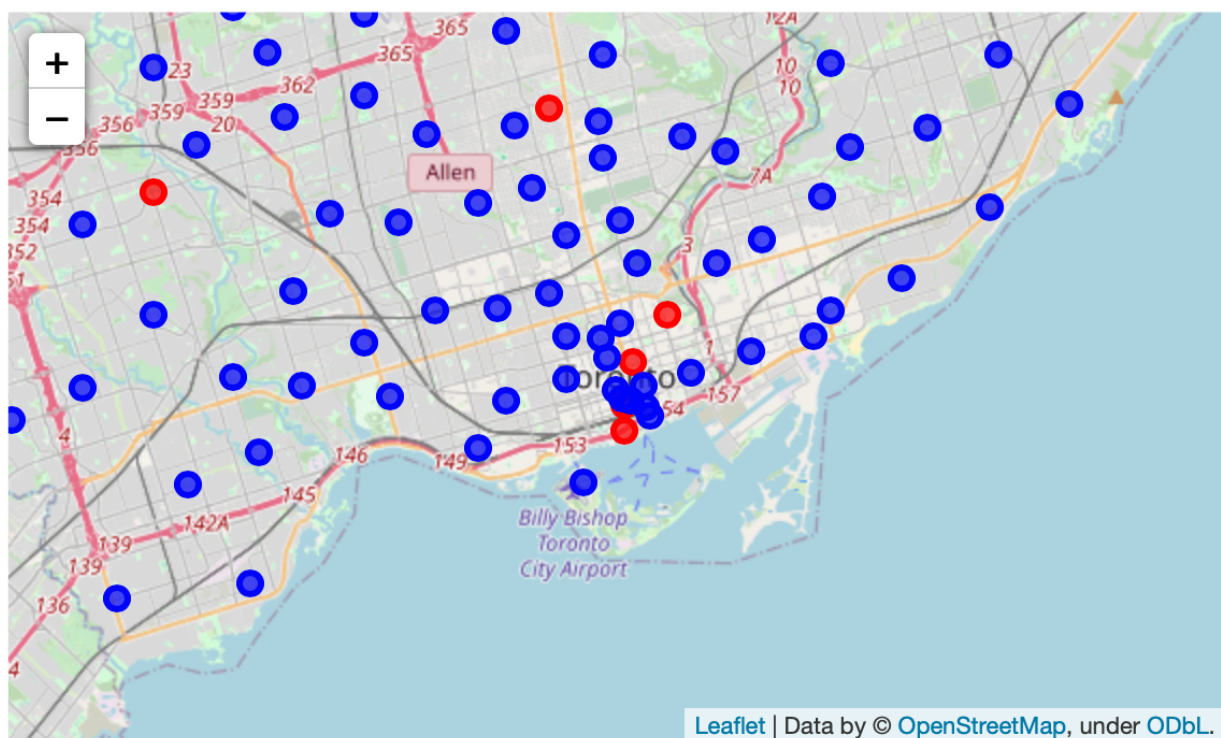

After retrieving the data using the Foursquare API, we can plot the neighbourhoods that still don't have Chinese restaurants.

Here is the neighbourhood with Chinese restaurants.

Here is with the red dots (that contains the Chinese restaurants) and the dots (without the Chinese restaurants).

5. Conclusion

In this study, I analyzed the neighbourhoods in Toronto, Canada that do not have Chinese neighbourhood using Foursquare API. We noticed that there are 190 neighbourhood or 91 boroughs that are possible to open a Chinese restaurants only 11 of those already have Chinese restaurants.