

The best location choice for opening a Pizza restaurant in Toronto.

Haowen Yang

1. Introduction

Our clients hope to open a Pizza restaurant in Toronto, and then aim to maximize the profits. What we can help is to choose a location in Toronto which tends to have a better future by using machine learning method.

2. Data source

The data set we used is all from foursquare, which is nearly the location information of different kinds of venues. To begin with, we focus our potential objectives on Central Toronto, the region representing the very center of this city. There are mainly two part in this report. The first part is identifying the competitors in Central Toronto, who also sells Pizza. To get better turnover, we should try our best to avoid high density of the existed competitors. The second part is to cluster all the venues in city center, which aim to cluster the city center into different sub-centers.

3. Methodology

Initially, we find the location of central Toronto, and zoom in. After which, we can get the location of all the Pizza restaurants (the blue points in figure1). From this plot, our client can clearly identify the distribution of the competitors.

The second part is to mark the venues in center Toronto. Generally, we can consider the region with high density of different venues is prosperous region and is suitable to open a Pizza, and vice versa. The density distribution can be shown in figure2, where the red point is the location of city center and little blue points are venues. What we want to do is clustering these points and forms four sub-regions, using K-Means. We set the parameter is 4 clusters in k-means and get the figure3, where we can see the conclusion of clustering: the four different colors mean different clusters.

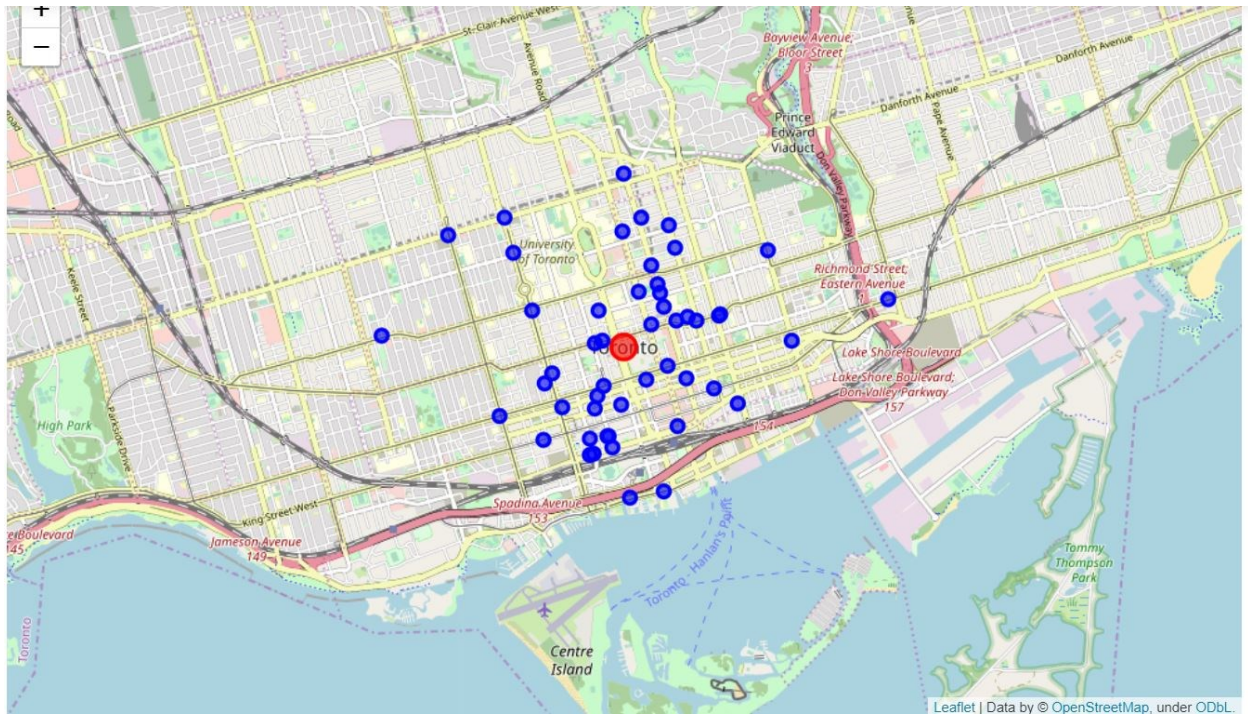


Figure 1

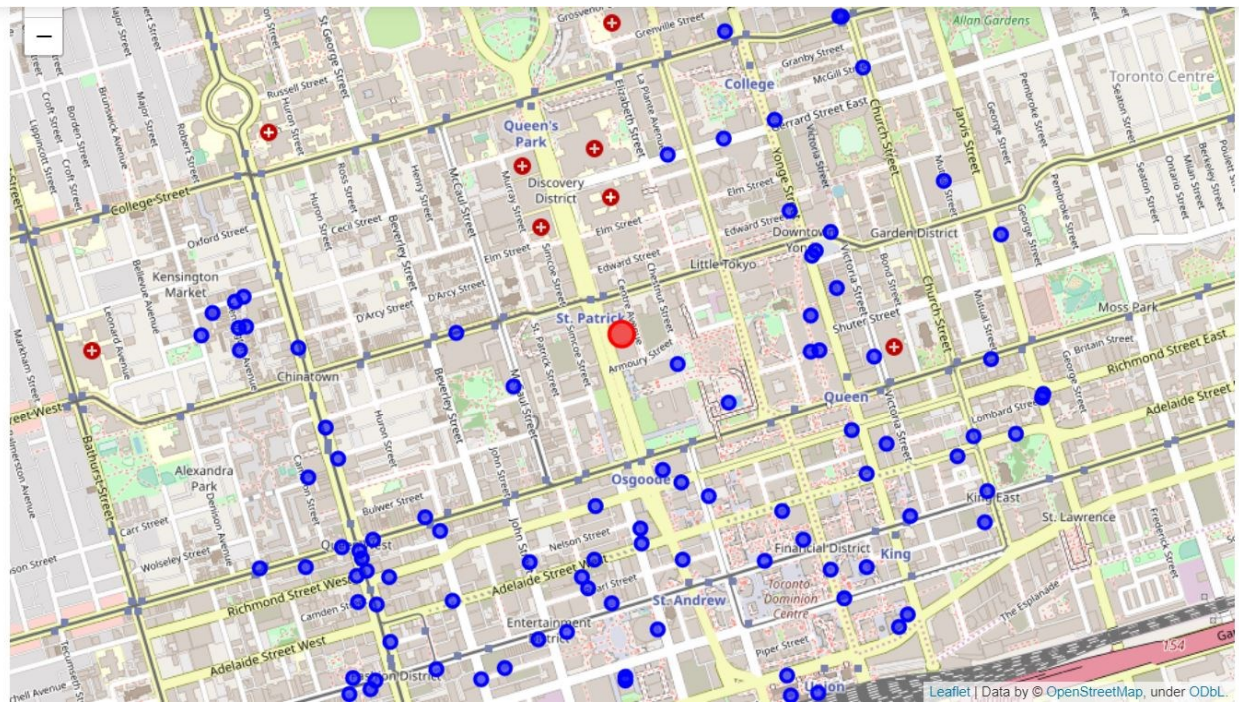


Figure 2

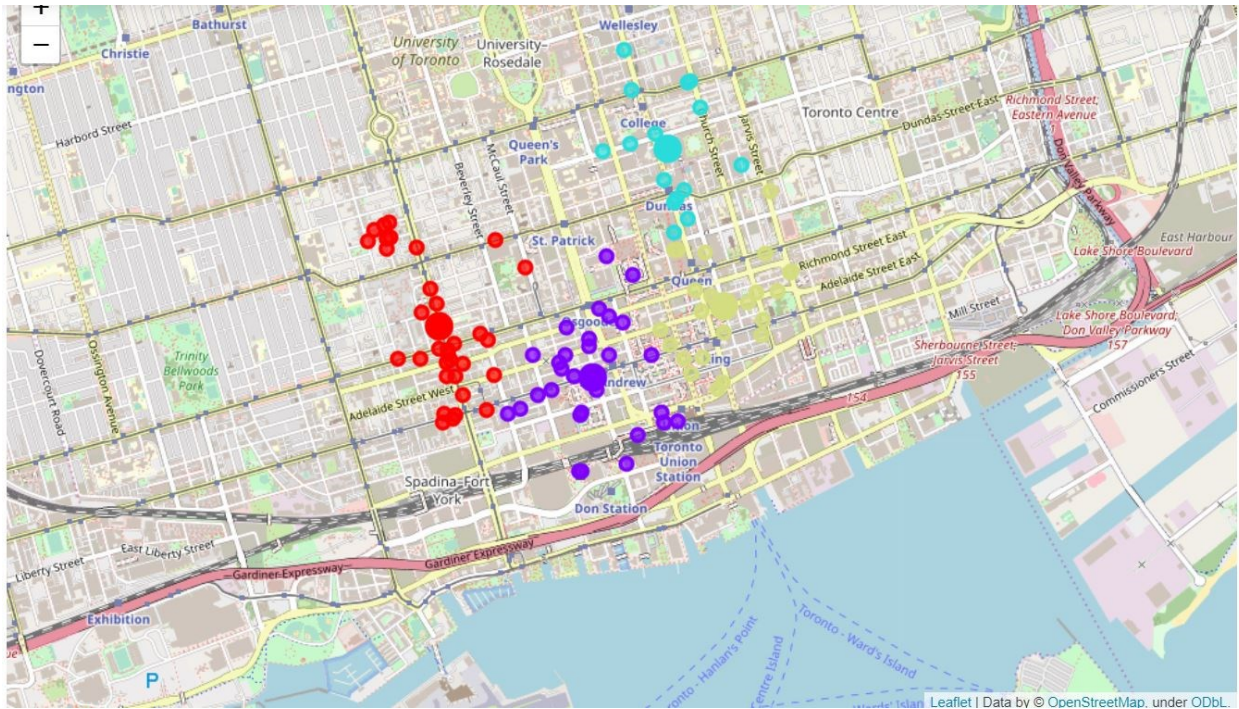


Figure 3

4. Results

To decide the best location, we need merge the two parts. In figure 4, we can notice the relationship between four sub-centers and our competitors. The little blue points are Pizza restaurants in Central Toronto. And the colorful points are candidates we want to choose since they are all the center of the clusters, which are thought to be more prosperous than other location.

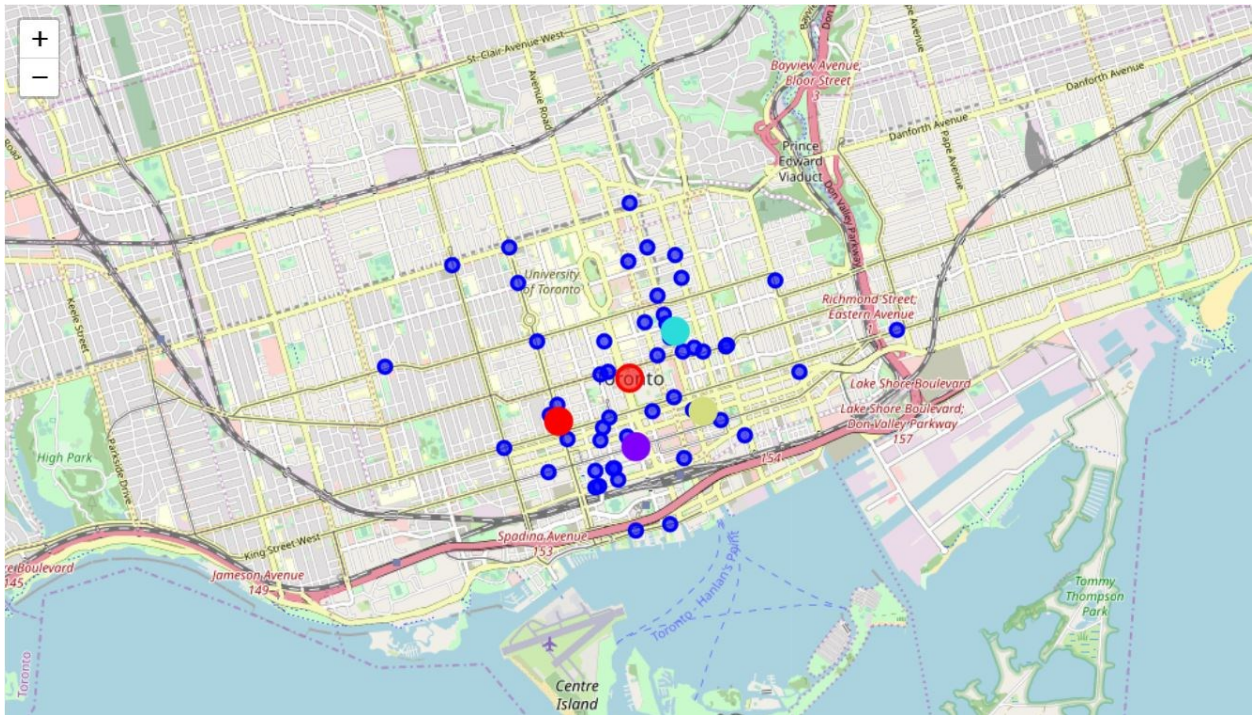


Figure 4

5. Discussion

To discuss the final decision, we need some statistical information.

Let's first decide which point located in the most popular region, that is red point with 33 venues, purple point with 29 venues, blue point with 15 venues and yellow one with 23 venues.

Next, it is important to know the low-density region of competitors, that is yellow point with least competitors, and is considered the best location.

We can also get the distance between them and city center. Since the difference is small, we can just ignore it.

6. Conclusion

The concrete information is not enough in this report, for instance, we don't consider the location which may induce the cost. In a nut shell, this is a prototype of the location choice problem and after considering more information, we can get a better model.