Introduction

American flavor of restaurant is very popular in Toronto in Canada. So investors are looking to open an American flavor of restaurant. However, they don't know where he should open it in Toronto in Canada. They hope some suggestions and insights can be given through exploring Foursquare location data.

According to some experienced businessmen who had or have run restaurants for many years, the following three factors should be considered in the selection of location of this flavor of restaurant.

1. The less American flavor of restaurants.

There are the less similar flavor restaurants in a neighborhood, there are the less competitors. Therefore, the less similar restaurants, the less risks of competition.

2. The less other flavor of restaurants.

If there are more other flavor of restaurants in a neighborhood, it is more possible that diners may choose to enter other flavor of restaurants when they cannot find an American flavor of restaurant. If there are less other flavor of restaurants, there are more opportunities for customers to choose the American flavor of restaurants. Therefore, the less other flavor of restaurants, the more opportunities for American flavor of restaurant to be chosen by diners.

3. The more entertainment spots.

A neighborhood has more entertainment spots which don't cater for food and meals means more people could become the customers of restaurants. For example, most people prefer to eating in restaurants after entertainment activities. Therefore, the more entertainment spots, the more potential diners.

Based on the above analysis, investors must find a neighborhood which meets the above three factors at the same time. The problem to be solved is how to find a neighborhood in which there are the least American flavor of restaurants and other flavor of restaurants to reduce competition, and at the same time there are the most entertainment spots which don't cater for food and meals to increase potential diners.

Data

Because the investors are going to open an American flavor of restaurant in a neighborhood in Toronto in Canada. So we use the neighborhood data in the city of Toronto in Canada. To get the Toronto neighborhood data, we will scrape the Wikipedia page which has all the information we need to explore the neighborhoods in Toronto. And we wrangle the data, clean it, and then read it into a pandas dataframe. The dataframe includes the information of Borough, Neighbourhood, Latitude, and Longitude in Toronto in Canada. The following is the first five rows in the dataframe.

After getting the dataframe, we will use Foursquare location data to get 100 venues of each neighborhood. Foursquare location data including all the information of each venue we need to explore the venues in each neighborhood. Then we extract category information of each venue for each neighborhood. We gather all the neighborhoods and compare them through their category information to get the most proper neighborhood.

Methodology

To help investors find the most proper neighborhood, we need to perform the following work:

1. Import and install the required packages

2. Scrape the Wikipedia page to get the Toronto neighborhood information and transform the information into a dataframe

3. Get the 100 venues for each neighborhood throgh FourSquare API

4. Compute the total number of categories of all venues for each neighborhood

5. Generate the dummies for all categories of venues

6. Group according to different neighborhoods and compute the total number of each category for each neighborhood.

7. Compute the total number of category whose name contains 'American Restaurant'

Compute the total number of category whose name contains 'Restaurant'

Compute the total number of category which don't belong to restaurant

8. Add three columns 'totals1','totals2','totals3' . 'totals1' is to compute the total of venues whose category is 'American Restaurant' for each neighborhood. 'totals2' is to compute the total of venues whose category is 'Restaurant' for each neighborhood. 'totals3' is to compute the total of venues whose category is 'not restaurant' for each neighborhood

9. Find those neighborhoods in which there are no American restaurant and restaurants, in other word, find the dataframe which meets 'totals1' equals 0 and 'totals2' equals 0.

10. Based on the above obtained dataframe, find the neighorhoods in which there are the most venues which are not restaurants, in other words, find dataframe which meets 'totals3' equals the max.

11. Get the latitude and longitude of Toronto city

12. Create map of Toronto using latitude and longitude values and add the neighborhoods found by us to the map

Results

The results are obtained through the above steps. The obtained neighborhoods are "CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara" in Downtown Toronto in Toronto city in Canada. In these neighborhoods, there are neither "American flavor of restaurants" nor "other flavor of restaurants" and there are 13 entertainment spots. 13 is the maximum value among all the neighborhoods which have neither "American flavor of restaurants" nor "other flavor of restaurants".

Discussion

The above method only set the maximum value as our results. we can extend the

range of selection through setting the threshold. Those neighborhoods whose total number of entertainment venues are greater than the threshold can also be used as the candidate neighborhoods. In these candidate neighborhoods, we can set other conditions to further filter the neighborhoods such foot traffic. Furthermore, this concept of the method can be used to select other type of venues like gym, shop, bar, store, and so on.

Conclusion

This project can help those investors who are going to open an American restaurant in Toronto city in Canada to look for most proper neighborhoods. The data used includes the Toronto neighborhood data from the Wikipedia page and Foursquare location data through Foursquare API. The method has used data analysis tools , data visualization tools and Foursquare API .