Best Neighbourhood to Start a Cafe in Toronto

For Coursera DataScience Final Assignment

By

Ryan Tam

Introduction

Starting a new cafe has always been a challenge. Very often, we see a lot of them located very close to each other, thus creating huge competition that can affect the success of the business. Other times, we may see a populated area, but with very few shops available. To maximize the chance of the success of any business, some of the biggest factors are the location and the potential market. In this project, we would like to analyze location data to find the best location with the most potential customers in Toronto to start a new cafe. At the same time, we want a location with a lower number of other similar venues to avoid competition. Specifically, we are interested in finding out the number of competitive venues per capita within each neighborhood and decide which is the best location to start the new cafe. For this project our target audience would be any new entrepreneur or any cafe franchise such as Tim Hortons, Starbucks, and McDonalds, who may be interested in opening a new cafe or looking for an area to expand.

Data

To begin our analysis, we must look at the factors that can increase our potential market and the number of other similar cafes around the area. Let's define the cafe we have in mind first. We want to start a cafe that will sell caffeinated beverages, snacks and some baked goods to go with our favourite drinks. Consider items that we may see from a Starbucks, or Tim Hortons menu. The two main things we want for our new cafe is to have less competition in the area, and more potential customers.

In order to find the best potential neighborhood to start the cafe in Toronto, we will need the location data from foursquare as required by the capstone final project. More specifically, we want to see how many other similar cafes are concentrated within each neighborhood. This will

give us an idea of the level of competition for each neighborhood when we start our new cafe. When we use the foursquare data, we want to look for other venues with tags such as ‘cafe’, ‘coffee shop’, potentially ‘fast food restaurants’ since a lot of them also sell caffeinated beverages, and ‘bubble tea shop’.

Since our cafe is aimed at a neighborhood with higher population. To find out about that, we turn to the toronto.ca website to find the neighborhood with the most population to maximize the chance of our business at success. The data from toronto.ca is from 2016, we can project the number of population to the current year using the population growth percentage for each neighborhood to see if the potential change in population will give us different results.

Once we have the information of those two potential factors, we can use them to determine which neighborhood would be the best for our new cafe with the highest potential customers and the least amount of competitions. Essentially, we are looking for the neighbourhood with the lowest number of similar venues per capita in order to determine the best location for our new cafe. To complete our analysis, we can use our venue data to determine our biggest competitors within Toronto.

Methodology

Using python, we obtained the population data of each neighbourhood in Toronto from toronto.ca website. As we looked through the information included in the population data, we realized it does not include coordinates of the neighbourhood. To obtain the latitude and longitude, we looked for the location data of the neighbourhoods in Toronto from toronto.ca website again and combined those information together into our dataframe. Both of those files from toronto.ca were saved for the ease of use after. We output a map of Toronto with each neighbourhood marked with blue dots as shown in our first map.

The next step we did was to get the venue data from foursquare.com. We wanted to know the number of similar venues around each neighbourhood to see the number of competition our new cafe will face. We selected a 2km radius of each neighbourhood for our search since the focus of our cafe is the local area. We defined the main tags such as ‘cafe’, ‘coffee shop’, ‘fast food restaurants’, ‘bubble tea shop’ to narrow down our search. Since foursquare limits the number of search results for each call at 50, with a limit of 500 calls each hour, we spreaded out our calls over a day. After every run, we saved the result to a csv file and continued until we had covered enough potential competitions. In case of potential duplicate results since we are separating the calls, we ran a command to delete any doubled result.

Finally, to obtain the lowest competition per capita in each neighbourhood, we calculated the ratio of venues to each local neighbourhood population and arranged them accordingly. Since the population data is from 2016, in an attempt to project the population adjustment for 2020, we used the population growth rate to estimate the number of population for this year to see if it may have any noticeable impact to our result.

Result

Based on the result, it appears that the neighbourhood around Rouge area in Toronto is the most suitable for our new cafe, with only 6 venues serving the locals within a 2km radius. The other neighbourhoods that are also suitable for our cafe include Glenfield, West Hill, L'Amoreaux, and Malvern. All of those are ranked as top 5 in terms of competition per capita. Since I used the population growth to project the population change in each neighbourhood from 2016, I arranged the result to see if it may have an impact. It appears that the change has little to no noticeable impact to the competition of our new cafe. I also displayed 2 maps at the end to show the location of the top 2 neighbourhoods to start the cafe. Finally, I included a table

to show the venue counts within Toronto to get an idea of the venues we may be competing for business. The table at the end shows that our biggest competitors are Tim Hortons, Starbucks, Second Cup, Coffee Time and Country Style.

Discussion

There may be a number of factors that can affect our result. In the process of searching for potential competitors for the new cafe, we discovered that foursquare will return results that did not have the same category as what we were searching for. For example, when we were searching for the tag ‘Cafe’, McDonald’s showed up even though its main category is ‘fast food restaurant’. On top of that, due to the limited number of results per call set by foursquare, I chose the most relevant tags that are direct competitors to the new cafe. However, other venues can also sell caffeinated beverages as well. I estimate that the impact of that can be mitigated if the focus of the cafe is takeout.

Upon reviewing the maps of the top location in the Rouge neighbourhood, it appears to be set in the middle of a zoo! This is potentially the main reason there are so few venues returned during the search within a 2km radius on foursquare. The next best neighbourhood is Glenfield. The map shows that it is in the middle of a residential area, which can potentially be a great location for the new cafe. For this reason, I believe the Glenfield neighbourhood is the better location to start the new cafe.

Due to the fact that the cafe is very localized, and the area covered for this study is to find the best location within a huge city, any generalization by using machine learning may not be effective. For this reason, I decided not to use machine learning for this project. The main factors we needed are population and number of venues, which we can easily obtain. However, a more specialized study using machine learning can be used if the target location is smaller.

We may even factor in movement and traffic data to see what other places are better for the new cafe.

Conclusion

After completing the analysis, we can conclude that one of the best neighbourhoods to start our new cafe is Glenfield in terms of competition per capita. There are other great places to start the new cafe as well such as West Hill, L'Amoreaux and Malvern . The result shows the number one location to start the cafe is in Rouge. However, due to the fact that the coordinates obtained for that neighbourhood from toronto.ca website points to the middle of the zoo, it may not be the most suitable location to have our new cafe, as the coordinates given may throw off our result when searching for other competitors. However, if traffic and movement information are factored in, it may still be a viable location. But that will be a study for the next project.