FINDING TORONTO NEIGHBOURHOOD FOR OPENING OF AN INDIAN RESTAURANT

Introduction/Business Problem:

The exercise is for the hospitality industry, which is very dependent on specific locality because a good neighbourhood can drive the required quality and quantity of traffic/footfall to the establishment. The area also helps in minimizing/ avoid category competition.

Objective:

The objective here is to analysis and provide a list of the suitable neighbourhood to start an Indian Restaurant. The selected city is Toronto. Toronto, being the most populous city and a major financial hub of Canada, attracts residents from all over the world both via employment opportunities and tourism. With nearly half, 46%, of its population coming from other countries. More than 200 different ethnic groups call Toronto home, and they bring with them over 140 languages. It is also one of those cities housing a wide selection of restaurants from cuisines all over the world, as we are going to see in this study.

Here, we deal with a hypothetical situation, of solving queries for planning to start an Indian restaurant in Toronto. The preliminary study of feasible neighbourhoods for this new venture. Here, we use the business concept of proximity to competition, to shortlist ideal areas. Competition can be useful in industries where comparison shopping is popular. Shopping clusters and various food and eating joints around can help the new business to get the initial traction. Therefore, in this project, we try to find the neighbourhoods around Toronto, which has a clustering of Restaurants. However, none of them being an Indian Restaurant. This step will avoid competing with the same cuisine, one that's hard to pull off with a new restaurant.

The course of this project would involve accessing the data of different venues existing within a set radius from each neighbourhood in the list. The collection of data about the different kinds of sites will be collected using Foursquare API, and then only the categories containing the word "Restaurant "will be extracted from them. After this, we remove the data of Indian Restaurants from the dataframe, and then aggregate the other restaurant values together, to represent the relative presence of restaurants near each neighbourhood. Then we add the Indian restaurant column back into the dataframe beside the aggregate units of restaurants of all other cuisines for comparison.

Finally, we sort the dataframe in the decreasing order of aggregate restaurant value, and then pick the neighbourhoods at the top, without an Indian restaurant.

Target Audience

Any entrepreneur wishing to get into the restaurant business by starting an Indian restaurant in Toronto.

Data:

To solve this problem, we will use the following data:

• List of neighbourhoods in Toronto, Canada.

• Latitude and Longitude of the neighbourhoods. This data will be collected in collaboration with FOURSQUARE API.

• Venue data of restaurants. This will help us find the neighbourhoods that have a cluster of restaurants located in them. This data would also help us identify the cuisine of the restaurant, including Indian.

Data Sourcing:

1. The Toronto neighbourhoods details were collected from Wikipedia page:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Venue & Lat-Long data:
FOURSQUARE API

Methodology:

STEP 1 - Accessing the data of different venues existing within a set radius from each neighbourhood in the list.

STEP 2 - The collection of data about the different kinds of venues would be done using Foursquare API. Then only the categories containing the word "Restaurant "will be extracted from them.

STEP 3 - We then remove the data of Indian Restaurants from the dataframe, and then aggregate the other restaurant values together, to represent the relative presence of restaurants near each neighbourhood. Then we add the Indian restaurant column back into the dataframe beside the aggregate value of restaurants of all other cuisines for comparison.

STEP 4 - We sort the dataframe in the decreasing order of aggregate restaurant value, and then pick the neighbourhoods at the top, with a "0" value for Indian restaurants column.

Analyzing the data:

We now examine in detail, the methodology followed for the project and its different stages

The data from the Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M was read into a pandas dataframe using read_html. After that, the rows having "Not assigned" for Boroughs were removed from the dataset. Multiple neighbourhoods having the same postal code were grouped with a comma separator. For rows which had missing neighbourhood field but a non-empty Borough field, the Borough was used as the neighbourhood.

The latitude and longitude data for each neighbourhood was read from http://cocl.us/Geospatial_data into a separate dataset and later merged with the first dataset using the merge function.

We chose to work with only Neighborhoods that were in Toronto, so we selected the Boroughs with the name Toronto in it from the dataframe and marked all of them into a map of Toronto to visualize them.

As a next step, we used the Foursquare API, and the explore endpoint to get the details of venues which were in a radius of 1 km from each neighbourhood. 3189 such venues were acquired via the Foursquare API. We then performed one-hot encoding for the venue categories. We then restricted the dataframe to show only the restaurants among the venues. This included a comprehensive list of cuisines all around the world.

We have captured 53 different cuisines of restaurants, which are active around 39 Neighborhoods of Toronto. Then, we dropped the column containing the Indian restaurants, did a horizontal sum to get an aggregate value of all other cuisines and then added the Neighborhood column and Indian Restaurant column back to the dataframe to generate the following

The next step was to sort the dataframe in the decreasing order of the aggregate Restaurant score, and then look for rows at the top having a 0 value in the Indian restaurant column. This gave us the following result

Of the top 5 rows, we see that Little Portugal is tied with the 2nd spot in the aggregate restaurant spot, and there are no Indian restaurants near the neighbourhood.

Results

From the above data frame, let's consider the top 5 rows, we see that that Davisville has the highest clustering of restaurants around, and it also has the highest number of Indian restaurants around. All neighbourhoods in the top 5, except Little Portugal, has an Indian restaurant operating nearby.

We can also see that Little Portugal and the Danforth West has the same value for aggregate restaurants, making it a tie for second place, but makes it a better option to start our Indian restaurant. Hence, we can focus more on the Little Portugal neighbourhood

Since we have found that Little Portugal has no Indian restaurants nearby, we would like to verify whether the neighbourhood is welcoming for restaurants of different cuisine. Unless the neighbourhood and the residents are open to trying out new cuisines, we won't benefit much from opening our restaurant in this neighbourhood.

For knowing this, we retrieve every restaurant in Little Portugal region by their cuisine type. This is done by calling the column names of restaurant cuisines of Little Portugal which had a non-zero value. This operation resulted in the following output

We can see that the cuisines around Little Portugal are cosmopolitan and we have a good chance of securing business if we select Little Portugal to open our Indian restaurant.

Discussions and Recommendations

This analysis is purely based on the proximity of other restaurants and their cuisine types. Although this analysis provides a preliminary idea of the selection of neighbourhood, further study has to be carried out before finalizing. Factors like rent, availability of space, traffic and busy hours can influence the business of a restaurant to a great extent, and these have to be taken into consideration before finalizing. Further analysis can include the population density, spending capacity of residents of each neighbourhood etc. to gain more leverage of the available data. Additionally, the study was limited to a radius of 1 km, and this can be tweaked for varying the results.

Conclusion

In this project, we utilized the data obtained through the Foursquare API to decide a Neighborhood in Toronto for starting an Indian Restaurant. Based on the analysis, we have selected Little Portugal as a viable option for further study and a potential location for the starting of the new Indian restaurant.

References

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

http://cocl.us/Geospatial_data

https://developer.foursquare.com/docs/places-api/endpoints/