

The Battle of Neighbourhoods

Prepared By,
Swagat Kumar Behera

Introduction

This is a capstone project for IBM Data Science Professional Certificate. In this project, I am going to showcase a scenario regarding number of Indian restaurants in Toronto and how it is going to benefit for entrepreneurs to open Indian restaurant in Toronto and its neighborhood. Therefore it might be a great opportunity for Canadian based entrepreneurs. More than a million Indian people resides in Canada. So entrepreneurs might think of opening its business in the areas near to Indian communities. With the purpose, finding the best location to open such a restaurant is one of the most important decisions for these entrepreneurs and this project will help them to find the most suitable location

Business Problem

“What is the most suitable location for an entrepreneur to open an Indian Restaurant in Toronto or Canada?” The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Indian Restaurant in Canada, basically near Toronto. So, I will be leveraging the concept of Data Science Methodology along with Machine Learning Algorithms such as “Clustering” and I will be going to suggest some of the possible location and solution to this business problem.

Data Set

To explore into the data and finding a best possible solution, we will need below data:

- List of Postal Code, borough , Neighborhoods in Toronto, Canada
[https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M]
- Latitude and Longitude of these Neighborhoods
[http://cocl.us/Geospatial_data]
- Venue data related to Indian restaurants.

This will help us to find prospect neighborhoods/ location that are more suitable to open an Indian Restaurant.

Methodology

1. Scrapping of Toronto neighborhood details by following Wikipedia page [[https://en.wikipedia.org/wiki/List of postal codes of Canada: M,](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M_)]
2. Getting geographical coordinates of the neighborhoods using the Geocoder package, geographical coordinates of each postal code. [[http://cocl.us/Geospatial data](http://cocl.us/Geospatial_data)]
3. Leveraging Foursquare API to get venue data related to different neighborhood.
4. With this data, unique categories is checked and analyzed each neighborhood by grouping the rows by neighborhood and taking mean on the frequency of occurrence of each venue category and put one constraint as “**Indian Restaurant**” and used in k-means clustering method.
5. Performed clustering method by using k-means clustering. Here, I have set k=4 [number of clusters] and then located data points to nearest cluster while keeping centroid as small as possible.
6. Finally, result shows distribution of Indian Restaurants in each neighborhood in Toronto from k-means clustering.

Conclusion

Cluster 0 : Neighborhoods with no Indian restaurants.

Cluster 1: Neighborhoods with more number of Indian restaurants.

Cluster 2 : Neighborhoods with less number of Indian restaurants

Cluster 3 : Neighborhoods with more number of Indian restaurants

Most of the Indian restaurants are in Cluster 2 and 3, around - The Dan forth West, Riverdale, Davisville, St. James Town, Cabbage town, Harbourfront East, Union Station, Church and Wellesley and Central Bay Street. *Lowest* in the Cluster 0 and 1, areas like The Annex, North Midtown, and Yorkville.

So, cluster 0 might be a good location to open Indian Restaurant as there is not any Indian Restaurant in these areas. Therefore, this project recommends the entrepreneur to open any Authentic Indian Restaurant in these locations.