Using foursquare API and Clustering to Identify a suitable location to open a new Bar in Toronto, Canada

Aishik Mukherjee

April 24th, 2020

1. INTRODUCTION:

For this Capstone project, I am creating a hypothetical scenario for a concept manager/owner who wants to open a new Bar in Toronto, Canada. With the purpose in mind, finding the location to open a Bar is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location. In order to do so, I shall do some analysis on the location data using foursquare API and clustering algorithm.

1.1 Business Problem:

The idea behind this project is that there may be enough number of Bars in Toronto and it might present a great challenge for this entrepreneur who is based in Canada, to choose a suitable location in Toronto to sustain his business where the competition might be low. The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Bar in Toronto, Canada. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question: In Toronto, if an entrepreneur wants to open a new Bar, where should they consider opening it?

1.2 Target Audience :

The owner/manager is the sole target client who seeks a good location to open a new Bar

2. DATA:

In order to solve this problem, I will be needing some relevant data on which we can work on. Specifically, the list of neighborhoods in Toronto, Canada. Then we would need to analyze and plot the neighborhoods in a map, we will be needing the location and longitude values of the neighborhoods. Then ultimately, we would require venue data related to Bars. This will help us find the neighborhoods that are most suitable to open a new Bar.

3. EXTRACTING DATA:

We will use the Data of neighborhoods in Toronto from wikipedia, which is available here at -
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
The table in the given site provides us with the Postal Codes, Boroughs and Neighborhoods in Toronto, Canada.

Next, we will be needing location data of the neighborhoods we got from wikipedia, specifically the longitude and latitude values. For this, we can use Geocoder package or a file which contains the latitude and longitude values of these neighborhoods(http://cocl.us/Geospatial_data)

Next, we will be needing the venue data related to these neighborhoods. For this we shall be using the foursquare API.