

Introduction section:

Background and problem description:

Imagine I am a Data Science Professional working in Toronto. On a fine day, an entrepreneur comes to my office. She says that she is thinking of opening a restaurant in Toronto. She asks my help in deciding which place in Toronto would be appropriate for the restaurant so as to get a good business. I gladly offered to help her with my knowledge in data science and come up with a suitable location for her restaurant in Toronto.

I asked the client about the category of the restaurant that she wants to set up. She said that she would like to set up an Italian restaurant. To help her, I need to find out with the help of my data science skills, which region in Toronto might be suitable for establishing an Italian restaurant so as to get a good business. Now, I need to find out external factors affecting a restaurant business and explore the neighborhoods of Toronto to find the location.

Data Section:

Description of data and the sources that will be used to solve the problem:

I need the neighborhood names in Toronto and their coordinates. For that I plan to use the Toronto neighborhood data available on Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). I can obtain the coordinates of the neighborhood location by mapping the postal codes for the neighborhoods from the CSV file that can be obtained from https://cocl.us/Geospatial_data.

Now based on the coordinates and names of the neighborhoods, I can generate a Folium Map of Toronto with the neighborhood names as pop ups. I plan to use the Foursquare API to find out the restaurant categories in each neighborhood around Toronto. Now I can cluster the neighborhoods to find out an appropriate cluster suitable for the establishment of an Italian restaurant. The cluster in which there are few Italian restaurants might serve such a purpose.