

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurant in Toronto, Canada

By:

Pranav viswanathan

Data :

To solve the problem, we will need the following data:

- List of neighbourhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto, the city of the country Canada.
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to public places, tourist spot . We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page

(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of neighbourhoods in Canada,. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighbourhoods using (https://cocl.us/Geospatial_data) which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Public places category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).