

Capstone Project – The Battle of Neighborhoods

Data description:

For the initial data set, I will start by extracting the data from Open Data Soft (ODS) to get a list of the neighborhoods in Chicago. ODS has thousands of datasets available publicly that can be filtered according to state, keywords, etc. For Toronto, the data will be extracted from the Wiki link that was used in the previous exercise. Both links are provided below:

Chicago dataset:

<https://public.opendatasoft.com/>

	State	County	City	Name	RegionID
1	IL	Cook	Chicago	Beverly View	403368
2	IL	Cook	Chicago	Ranch Triangle	403305
3	IL	Cook	Chicago	Schorsch Forest View	403280
4	IL	Sangamon	Springfield	Vinegar Hill	761482
5	IL	Cook	Chicago	Lakewood - Balmoral	403260
6	IL	Cook	Evanston	West Village	764020
7	IL	Rock Island	Moline	Wheelock/Veile	761011
8	IL	Cook	Chicago	Jackson Park Highlands	403354
9	IL	Cook	Chicago	Hyde Park	269586
10	IL	Sangamon	Springfield	West Koke Mill	761475
11	IL	Cook	Chicago	East Beverly	403375
12	IL	Cook	Chicago	Montclare	274579
13	IL	Cook	Chicago	Cragin	28173
14	IL	Cook	Chicago	The Loop	269593
15	IL	Cook	Chicago	South Old Irving Park	403316
16	IL	Cook	Chicago	Wentworth Gardens	403341
17	IL	Vermilion	Danville	Center City	268056
18	IL	Cook	Chicago	Chrysler Village	403336
19	IL	Rock Island	Moline	Karsten's Park/City Line	761016
20	IL	Cook	Chicago	Portage Park	275083

Toronto dataset:

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

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M2B	Not assigned	Not assigned
M3B	North York	Don Mills North
M4B	East York	Woodbine Gardens
M4B	East York	Parkview Hill
M5B	Downtown Toronto	Ryerson
M5B	Downtown Toronto	Garden District
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	Cloverdale
M9B	Etobicoke	Islington

Once these datasets are extracted and read into a pandas data frame, I will first filter for the respective cities and clean the data. This will ensure there are no duplicates or null values and that the data is consistent. To gather the latitude and longitude values, we will be using the **geocoder** package and create a function that will get the coordinates for each neighborhood. We will also need data about the different venues in different neighborhoods for both cities. In order to get this information, we will be using the **Foursquare API** to make calls. Foursquare provides data on location which includes multiple venues and areas of interest in a given radius. This can also include venue names, locations, and photos. After finding the list of neighborhoods and creating the respective data frames, we will connect with the Foursquare API to get information about the location within a 500-meter radius.

The data retrieved from the Foursquare API will contain information such as:

- Name of neighborhood
- Neighborhood latitude and longitude
- Name of venue
- Venue latitude and longitude
- Type of venue

We will be using a clustering approach to explore and segment the neighborhoods. In the end, we can look at the top venues that were visited in each cluster and determine answers to the following questions:

- Comparing Chicago and Toronto, which city has more restaurants as the 1st most visited venue?
- In particular, are there certain neighborhoods that do not have restaurants in their “top 10 visited” list?
- Is there a specific type of restaurant that we can open that isn’t already there?

Doing so can help our potential businessman in choosing the best city and location for opening his/her restaurant. In addition, this could also help future businessmen in determining the different types of business that are already largely available and what type of business they can invest in or open in these locations.