

Finding the best place in Toronto to open an Indian restaurant

Amit Kumar

April 13, 2019

1. Data acquisition and cleaning

1.1 Data Source

First we need the data of Toronto neighborhoods which can be found on https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The extracted data contains the columns namely PostalCode, Borough and Neighborhood, where PostalCode column contain zip code of all the neighborhoods, Borough column contains the borough name to which a particular neighborhood belongs and Neighborhood column contains name of all the neighborhoods in the Toronto.

| | PostalCode | Borough | Neighborhood |
|---|------------|------------------|------------------|
| 1 | M1A | Not assigned | Not assigned |
| 2 | M2A | Not assigned | Not assigned |
| 3 | M3A | North York | Parkwoods |
| 4 | M4A | North York | Victoria Village |
| 5 | M5A | Downtown Toronto | Harbourfront |

Fig.1 Extracted data from Wikipedia

We are using **Foursquare** location data To extract the detailed data about restaurants in each neighborhood, but to get detailed information from Foursquare we need to provide coordinates longitude and latitude of each neighborhood and this data can be extracted from https://cocl.us/Geospatial_data.

| | PostalCode | Latitude | Longitude |
|----------|-------------------|-----------------|------------------|
| 0 | M1B | 43.806686 | -79.194353 |
| 1 | M1C | 43.784535 | -79.160497 |
| 2 | M1E | 43.763573 | -79.188711 |
| 3 | M1G | 43.770992 | -79.216917 |
| 4 | M1H | 43.773136 | -79.239476 |

Fig.2 Neighborhood coordinates

1.2 Data Cleaning

The data extracted from the Wikipedia page contains 'Not assigned' boroughs and neighborhood. All the rows which contains not assigned borough will be dropped from the dataset. In case the neighborhoods are having not assigned value but having assigned borough, the borough will be assigned to the corresponding neighborhood.

| | PostalCode | Borough | Neighborhood |
|----------|-------------------|------------------|---------------------|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Harbourfront |
| 3 | M5A | Downtown Toronto | Regent Park |
| 4 | M6A | North York | Lawrence Heights |

Fig.3 Data after dropping not assigned boroughs

1.3 Data Preprocessing

After the data is cleaned we will preprocess it before feeding it to the machine learning model. First we will combine the data from Wikipedia and coordinates in to a single dataset.

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|------------|-------------|--------------------------------------|-----------|------------|
| 0 | M1B | Scarborough | Rouge,Malvern | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Highland Creek,Rouge Hill,Port Union | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood,Morningside,West Hill | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | 43.773136 | -79.239476 |

Fig.4 Combined data

1.3.1 Extracting information about neighborhoods using Foursquare

Now we will feed the combined data to the Foursquare to extract the detailed information about each neighborhood. The figure just shows snapshot not an entire details.

| | PostalCode | Borough | Neighborhood | Borough Latitude | Borough Longitude | Venue Name | Venue Latitude | Venue Longitude | Venue Category |
|---|------------|--------------|-----------------------------|------------------|-------------------|------------------------------------|----------------|-----------------|-------------------|
| 0 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | The Big Carrot Natural Food Market | 43.678879 | -79.297734 | Health Food Store |
| 1 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Grover Pub and Grub | 43.679181 | -79.297215 | Pub |
| 2 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Starbucks | 43.678798 | -79.298045 | Coffee Shop |
| 3 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 | Upper Beaches | 43.680563 | -79.292869 | Neighborhood |
| 4 | M4K | East Toronto | The Danforth West,Riverdale | 43.679557 | -79.352188 | Pantheon | 43.677621 | -79.351434 | Greek Restaurant |

Fig.5 Foursquare location data