

Exploring Neighborhoods of Toronto to Open an Indian Restaurant

2. Data

2.1 - Data Sources

- Neighborhoods Data - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
This Wikipedia Page contains Postal Code, Borough & Neighborhood Names in Toronto
- Geographical Coordinates - https://cocl.us/Geospatial_data
This CSV File contains Geographical coordinates of the Neighborhoods
- Demographics Data - https://en.m.wikipedia.org/wiki/Demographics_of_Toronto
This Wikipedia Page contains information on the distribution of population by ethnicity
- Venues Details - Foursquare's Explore API (<https://developer.foursquare.com/>)
Foursquare API can be used to retrieve details for each of the venues in Toronto

2.2 - Data Cleaning

Neighborhood data scrapped from Wikipedia has three features; Postal Code, Borough and Neighborhood. The records with no assigned Borough (< 1%) were excluded from the analysis.

It was observed that more than one Neighborhoods can exist in the same Postal Code Area. These Neighborhoods were combined into one row separated by comma to remove duplicate Postal Codes in the dataset.

For Boroughs with unassigned Neighborhoods, the Neighborhood name has been considered same as the Borough to avoid missing values in the dataset.

Demographics Data extracted from Wikipedia was filtered to keep only the data points related to Indian Population.

For this analysis Venue Details from Foursquare's Explored API was filtered to retain only the top 100 venues within 1 KM radius of each neighborhood.

The following four features from the Venue Details were retained for further analysis:

- Name - Name of the Venue
- Category - Category Type as defined by the API
- Latitude - Latitude value of the Venue
- Longitude - Longitude value of the Venue