Applied Data Science Capstone Project

The Battle of Neighbourhoods



Analysis of Neighbourhoods and Demographics of Calgary, Alberta, Canada.

A description of the data and how it will be used to solve the problem.

Submitted By:

Dhananjay Saxena

About the data

1.**Wikipedia:**

The dataset to be used in this project shall be scraped from the website [Calgary Data](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary). The Wikipedia web page contains a list of all neighbourhoods within the City of Calgary including residential communities, industrial areas, major parks and residual areas by electoral ward. Calgary, Alberta, Canada, as of 2016, has 197 neighbourhoods, which are referred to as "communities" by the City of Calgary, and 42 industrial areas.



2**.Foursquare API:**

Foursquare, Foursquare is a technology company that built a massive dataset of location data. What is interesting about Foursquare is that they were very smart about building their dataset. They actually crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset. Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others, and is currently being used by over 100,000 developers, and this number is only growing.

Please click the here: [**Foursquare**](https://foursquare.com/) to visit their website.



3.**Geopy:**

The geographic co-ordinates for the neighbourhoods/communities has been incorporated into the dataset with the help of Geopy. Geopy is a Python client for several popular geocoding web services.

Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Geopy includes geocoder classes for the OpenStreetMap, Nominatim, Google Geocoding API (V3), and many other geocoding services. The full list is available on the Geocoders doc section. Geocoder classes are located in geopy.geocoders.

Geopy is tested against CPython (versions 3.5, 3.6, 3.7, 3.8) and PyPy3. Geopy 1.x line also supported CPython 2.7, 3.4 and PyPy2.

Please click here: [**GeoPy**](https://geopy.readthedocs.io/en/stable/) to visit the GeoPy documentation page.



The use of data

The table containing Calgary’s neighbourhood data will be scraped from the Wikipedia page and will be combined with GeoPy module’s coordinate to prepare a dataset which will contain data regarding the city’s neighbourhood, its population density and its latitude and longitude.

FoursquareTM API will be used to import the data regarding the venues belonging to all the neighbourhoods. This data will be cleaned and prepared to obtain data to analyse the neighbourhoods of the city.