

Predicting which country to choose for residence

Data

Data Sources

Chicago,USA data contains Chicago city's Borough, Neighbourhood, Latitude and Longitude. Chicago data can be found [here](#). Toronto,CA data contains Toronto city's Postcode, Borough, Neighbourhood, Latitude and Longitude. Toronto data can be found [here](#) and [here](#).

Data Cleaning

For chicago data, scrapped table containing Neighbourhood and Community area using pandas read_html() method. Renamed column names 'Community area' as 'Borough'. It contained only Borough and Neighbourhood names. Grouped 'Neighbourhood' values on unique 'Borough' values and listed them as one value separated by commas. To get venue details, we first need latitude and longitude coordinates of each borough. I used google search method to obtain gps coordinates of all boroughs and saved them as a .csv file. Extracted it as a new dataframe.

Both dataframe have Borough column as common, so merged both dataframe on Borough values to obtain gps coordinates. The final dataframe *chicago_data* looks like this.

	Borough	Neighbourhood	Latitude	Longitude
0	ALBANY PARK	Albany Park,Mayfair,North Mayfair,Ravenswood M...	41.9683	-87.7280
1	ARCHER HEIGHTS	Archer Heights	41.8079	-87.7236
2	ARMOUR SQUARE	Armour Square,Chinatown,Wentworth Gardens	41.8408	-87.6340
3	ASHBURN	Ashburn,Ashburn Estates,Beverly View,Crestline...	41.7479	-87.7072
4	AUBURN GRESHAM	Auburn Gresham,Gresham	41.7434	-87.6562

[Chicago city data](#)

After cleaning there were 81 records and 4 features(Borough, Neighbourhood, Latitude, Longitude).

For Canada data, scrapped the wikipedia page with postcode, borough and neighbourhood column values using pandas read_html() method. There are few rows that has "Not assigned" value in Borough column. Delete those rows using drop() function. One entry has "Not assigned" value in Neighbourhood column. Replace its value with the value in its Borough column. Apply groupby() method, to group neighbourhoods with same postcode as one row.

Create a new dataframe to read .csv file that contains Toronto city's postcode and gps coordinates. Merging both dataframe on Postcode, to obtain gps coordinates of each postcode. The final dataframe *canada_data* looks like this.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Maivern	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

[Toronto city data](#)

After cleaning there were 99 records and 5 features(Postcode, Borough, Neighbourhood, Latitude, Longitude).

Foursquare API

Foursquare Places API is used to obtain data related to 'venues' in any given location using its gps coordinates and a specific url created for API call. It is important to note that Foursquare defines a 'venue' as a place that one can go to, or check-in to, and that a 'venue' is not necessarily a specific venue but can be any establishment such as a restaurant or type of retail shop, etc. Each Foursquare 'venue' is assigned a 'category' and each 'category' is associated with a particular 'categoryID'. After every API call, use *requests library get()* function to return a json file with all venue information for entire location specified in the url.

First we create a url with our client id, client secret, version for using Foursquare API. Extract venues list within specified radius for first neighbourhood of chicago dataframe.

```

LIMIT = 100
radius = 750
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{&radius={}&limit={}'.format(
    CLIENT_ID, CLIENT_SECRET, VERSION, n_latitude, n_longitude, radius, LIMIT)
url
'https://api.foursquare.com/v2/venues/explore?&client_id=3UMNFX0SRVQJMVVCJ3B2C5KONK5FB1ETLMHFII8N0B3X0PN5&client_secret=QW2CZDT4IFFE1ZZBMM1LC
VDSUV40IOIDYBG3PUVHQ3HSY2H&v=20180605&ll=41.9683,-87.728&radius=750&limit=100'

```

[Example Foursquare API url](#)

We define a function and extract venue details for all Neighbourhoods in Chicago dataframe and Canada dataframe. A dataset of all venues associated with each Chicago city neighbourhood and Toronto city neighbourhood was created by recursively sending *get requests* to the Foursquare API.

The function *getNearbyVenues()* recursively sends a *get requests* to Foursquare for each neighbourhood that requests all nearby venues. While looping through each neighborhood from the chicago and canada dataset, the function appends each venue entry to a list and, after looping through each neighborhood, creates a DataFrame of all of the results. Included for each entry in the dataset are neighborhood name and gps coordinates, and venue name, gps coordinates and category.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        result = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name'] for v in result])
```

Extracting venue list for each neighbourhood

The resulting data frame for chicago neighbourhoods looks like:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Albany Park,Mayfair,North Mayfair,Ravenswood M...	41.9683	-87.728	Starbucks	41.968911	-87.728817	Coffee Shop
1	Albany Park,Mayfair,North Mayfair,Ravenswood M...	41.9683	-87.728	Lawrence Fish Market	41.968280	-87.726250	Seafood Restaurant
2	Albany Park,Mayfair,North Mayfair,Ravenswood M...	41.9683	-87.728	Marie's Pizza & Liquors	41.968132	-87.731533	Pizza Place
3	Albany Park,Mayfair,North Mayfair,Ravenswood M...	41.9683	-87.728	Ssyal Korean Restaurant and Ginseng House	41.968172	-87.733207	Korean Restaurant

The canada neighbourhoods data frame looks like:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Scarborough Historical Society	43.788755	-79.162438	History Museum
3	Guildwood,Morningside,West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place