Applied Data Science Capstone Project:

Cluster and Segment Neighborhoods in Major Cities for Expats

1. Introduction/Business Problem

As many people leave their home countries to move and work abroad they become expats to that country. To define an expat or expatriate, it is any person who lives temporarily or permanently in a country other than their country of citizenship. Leaving your neighborhood behind and moving to a new neighborhood in a new country can be quiet challenging and someone can still get confused while transitioning into a new culture/social etiquette. Moving into a new country but a similar neighborhood can help expats to quickly adapt to their new environment. Therefore, the approach proposed is to segment and cluster neighborhoods of two major cities. Providing this type of guidance can help expats find a suitable neighborhood in a new country and adjust much faster.

The problem is clearly addressed to expats and that is the target audience. As an expat, before moving into a new country you are expected to do research. This research would consist of recognizing which neighborhood is most suitable to your needs and life style. The easier you make it for you to settle in, meet people with similar hobbies and start to feel at home, the better.

2. Data

To solve this problem, multiple datasets will be used in combination with the Foursquare location data. Data will be used to cluster and segment neighborhoods in two major cities. The two major cities to be taken as an example are Toronto, Canada and New York City, U.S.

The first step in data collection is to extract the list of Toronto and New York City neighborhoods. Luckily, the datasets exist for free on the web. The New York City neighborhoods dataset is published by the New York (City). Department of City Planning and can be found on geo.nyu.edu website which is a spatial data repository maintained by New York University (NYU). The Toronto neighborhoods dataset can be scraped online from Wikipedia which includes the Postcode, Borough, and Neighborhood name.

Next, the Geocoder library can be used to fetch latitude and longitude coordinates for each of the neighborhoods. Adding the geographical coordinates (latitude and longitude) allows to map these neighborhoods using the folium API. For example, mapping these coordinates provides a better visual to understanding the distribution in each city.

Finally, the Foursquare location API will be used to extract the list of venues surrounding each of the neighborhoods and this list, which contains venues like restaurants/gym/coffee shops/ parks, will be used to cluster and segment neighborhoods in Toronto and New York City. The data will be merged, and further analysis will be performed to clean and prepare it for modeling.

Toronto Neighborhood data with latitude and longitude:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Neighborhood data merged with Venues data:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Glen Stewart Ravine	43.676300	-79.294784	Other Great Outdoors
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood

Sample Foursquare response to extract venues list:

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '5d23a06ba6ec98002c2ccada'},
 'response': {'warning': {'text': "There aren't a lot of results near you. Try som
area."},
  'headerLocation': 'Malvern',
  'headerFullLocation': 'Malvern, Toronto',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 2,
  'suggestedBounds': {'ne': {'lat': 43.8111863045, 'lng': -79.18812958073042},
   'sw': {'lat': 43.80218629549999, 'lng': -79.2005772192696}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4bb6b9446edc76b0d771311c',
       'name': "Wendy's",
       'location': {'crossStreet': 'Morningside & Sheppard',
        'lat': 43.80744841934756,
        'lng': -79.19905558052072,
        'labeledLatLngs': [{'label': 'display',
          'lat': 43.80744841934756,
          'lng': -79.19905558052072}],
        'distance': 387,
        'cc': 'CA',
        'city': 'Toronto',
        'state': 'ON',
        'country': 'Canada',
        'formattedAddress': ['Toronto ON', 'Canada']},
```