Capstone Project – Applied Data Science | Coursera

Introduction/Business Problem

Problem Description

A friend will be moving to a different city Bangalore, India from her home city of New Delhi, India for a job change.

She wants to find a neighborhood to stay in the new city which is the most like her home neighborhood.

Problem Background

I think this is an interesting problem to solve using Data Science as many people move between cities for various reasons such as job change, studies etc.

Doing this kind of analysis will give them a more familiar/similar environment in which they can adjust to their new home more easily.

Target Audience

- People moving to a new city
- Property rental websites
- Real estate agents

Data

Foursquare API

https://developer.foursquare.com/places

```
{'type': 'Feature',
 'id': 'nyu_2451_34572.1',
 'geometry': {'type': 'Point',
  'coordinates': [-73.84720052054902, 40.89470517661]},
 'geometry_name': 'geom',
 'properties': {'name': 'Wakefield',
  'stacked': 1,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'annoangle': 0.0,
  'borough': 'Bronx',
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661]}}
```

Geocoder

https://geocoder.readthedocs.io/

```
address = 'Toronto, CA'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of {} are {}, {}.'.format(address, latitude, longitude))

The geograpical coordinate of Toronto, CA are 43.653963, -79.387207.
```

Wikipedia

Delhi neighborhoods

https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi

Bangalore neighborhoods

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Methodology

- Get co-ordinates of home neighborhood i.e. Vasant Kunj, Delhi using GeoCoder.
- Get co-ordinates of away neighborhoods in Bangalore using GeoCoder.
- Verify co-ordinates from Wikipedia
  https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore
- Clean co-ordinates for neighborhoods
- e.g. Koramangala and Electronic City
- Get details of venues in the neighborhoods using FourSquare API
- Convert categorical values by one-hot encoding
- Use K-Means to cluster the neighborhoods by similarity.
- Plot the neighborhoods on Folium map
- Find if there are away neighborhoods in the same cluster as the home neighborhood.

Results

- The home neighborhood of Vasant Kunj did not match with any away neighborhood in Bangalore as suggested by the K-Means algorithm.
- Home neighborhood of Vasant Kunj is not very diverse with only 5 unique venue categories.

[61]:	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Vasant Kunj, DL	28.529249	77.154134	The Biker's Cafe	28.529229	77.155501	Café
1	Vasant Kunj, DL	28.529249	77.154134	Domino's Pizza	28.525263	77.156484	Pizza Place
2	Vasant Kunj, DL	28.529249	77.154134	Sagar ratna	28.525225	77.155648	Vegetarian / Vegan Restaurant
3	Vasant Kunj, DL	28.529249	77.154134	Nirula's	28.532362	77.151272	Fast Food Restaurant
4	Vasant Kunj, DL	28.529249	77.154134	Fab India	28.532398	77.151259	Clothing Store

```
[64]: print('There are {} uniques categories.'.format(len(vk_venues['Venue Category'].unique())))
      There are 5 uniques categories.
```

- Away neighborhoods in Bangalore Vasant Kunj are far more diverse with 71 unique venue categories.

```
71]: print('There are {} uniques categories.'.format(len(combined_venues['Venue Category'].unique())))
     There are 71 uniques categories.
```

- Within Bangalore 5/10 neighborhoods fall in the same cluster with Restaurants and Cafes being a significant percentage.
- The presence of pubs and drinking areas in Bangalore seems to skew the results away from Vasant Kunj.

Discussion

The user seems to have selected neighborhoods in Bangalore which are more non-residential localities as compared to her home neighborhood of Vasant Kunj which is a typical residential society in Delhi

The user needs to broaden her search in Bangalore and select more residential neighborhoods there if she wants to find a similar environment as in her home neighborhood.

[61]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Vasant Kunj, DL	28.529249	77.154134	The Biker's Cafe	28.529229	77.155501	Café
1	Vasant Kunj, DL	28.529249	77.154134	Domino's Pizza	28.525263	77.156484	Pizza Place
2	Vasant Kunj, DL	28.529249	77.154134	Sagar ratna	28.525225	77.155648	Vegetarian / Vegan Restaurant
3	Vasant Kunj, DL	28.529249	77.154134	Nirula's	28.532362	77.151272	Fast Food Restaurant
4	Vasant Kunj, DL	28.529249	77.154134	Fab India	28.532398	77.151259	Clothing Store

[64]:
```python
print('There are {} uniques categories.'.format(len(vk_venues['Venue Category'].unique())))
```
There are 5 uniques categories.

Conclusion

The user seems to have selected neighborhoods in Bangalore which are more non-residential localities as compared to her home neighborhood of Vasant Kunj which is a typical residential society in Delhi

The user needs to broaden her search in Bangalore and select more residential neighborhoods there if she wants to find a similar environment as in her home neighborhood.