

Relocation Neighborhood Recommendations

Trinath Sahu
3rd November 2019

Introduction

Background:

- As per report, between 2012 and 2013, about 36 million Americans moved to a new home. That is about one in every 8 people move to a new location.
- Top reasons of these movements include new job opportunity, Long term onshore travel, Own home, and job transfer
- Advantages of relocation might include:
 - ❖ Add on to job security
 - ❖ Can be an opportunity for career advancement
 - ❖ Increase your standard of living and quality of life
 - ❖ Personal development and new experiences
 - ❖ New friends
 - ❖ Better climate

Introduction (contd...)

Challenges in Relocation:

- Find housing in an unfamiliar area
- Right schooling and child care can be tough
- Transportation system
- And many more...

Result:

Road block to relocation benefits including new job opportunity.

Introduction (contd...)

Interests:

- An automated system recommending neighborhoods in new place similar to current neighborhood .
- Ex: If the system says neighborhood x, y and z in Toronto are like your current neighborhood 'a' in Newyork then you can imagine how much it will be helpful to them.

Data acquisition and cleaning (contd...)

Data Source:

- Location venue data: Foursquare API calls
- Toronto neighborhoods:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Latitudes and Longitudes of Toronto neighborhoods:
Geospatial_Coordinates.csv
- Latitudes and Longitudes of Newyork neighborhoods:
nyu_2451_34572-geojson.json

Data acquisition and cleaning (contd...)

Data cleansing:

- Scrap the Wikipedia website page which has Toronto neighborhoods
- Remove the rows having Borough as 'Not assigned'
- Merge multiple neighborhoods of same Borough into single row.
- Neighborhood with 'Not assigned' value will be updated to make it same as corresponding Borough name.

Data acquisition and cleaning (contd...)

Feature Selection:

- Venue categories around all neighborhoods of Toronto
- Venue categories around current neighborhood in Newyork
- Some examples of Features:
Park, School, College, Restaurant, Hospital,
Pharmacy, Metro station, Shopping Mall,
Bank, Gym, Grocery store, Bus station etc.

Exploratory Data Analysis

- Set current neighborhood in Newyork to 'Wakefield'
- View the current neighborhood of Newyork

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
5	Wakefield	40.894705	-73.847201	SUBWAY	40.890656	-73.849192	Sandwich Place
6	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898276	-73.850381	Caribbean Restaurant
7	Wakefield	40.894705	-73.847201	Koss Quick Wash	40.891281	-73.849904	Laundromat

Exploratory Data Analysis (contd...)

- Scrap the Wikipedia page and view the Toronto neighborhoods along with their Latitudes/Longitudes. Also, view the number of neighborhoods in Toronto. 103 neighborhoods were found in Toronto.

Out[42]:

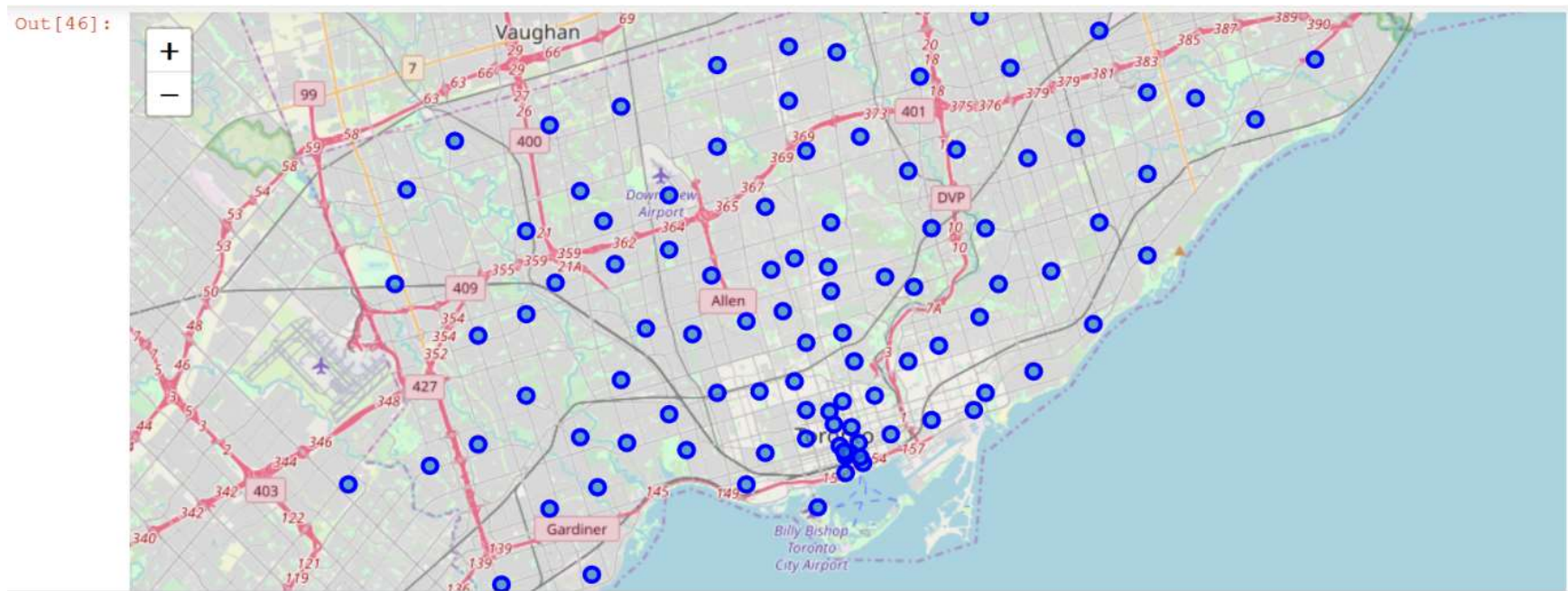
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

In [43]: `downt_data.shape`

Out[43]: (103, 5)

Exploratory Data Analysis (contd...)

Display Toronto neighborhoods in Folium map.



Exploratory Data Analysis (contd...)

- Call Foursquare API to get the nearby venues around Toronto neighborhoods and view them. Do the same thing for Newyork current neighborhood and merge data into Toronto data. Total 2252 venues and 274 unique venues were found.

```
In [50]: print(downt_venues.shape)
downt_venues.head()
```

```
(2252, 7)
```

```
Out[50]:
```

	Neighborhood	Neighborhood Latitude	Neighborhood 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	Guildwood,Morningside,West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
3	Guildwood,Morningside,West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood,Morningside,West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant

Exploratory Data Analysis (contd...)

Display top 10 venues nearby neighborhoods in Toronto.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide,King,Richmond	Coffee Shop	Café	Bar	Sushi Restaurant	Hotel	Steakhouse	Asian Restaurant	Thai Restaurant	American Restaurant	Bakery
1	Agincourt	Sandwich Place	Lounge	Skating Rink	Breakfast Spot	Print Shop	Women's Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	Park	Playground	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore
3	Albion Gardens,Beaumont Heights,Humbergate,Jam...	Grocery Store	Fried Chicken Joint	Pharmacy	Pizza Place	Fast Food Restaurant	Beer Store	Sandwich Place	Dog Run	Dessert Shop	Dim Sum Restaurant

Model development

- The problem is to find similarities between neighborhoods.
- Unsupervised Clustering algorithm is used to solve this kind of problem.
- K-means algorithm fits best compared to Hierarchical and DBSCAN clustering algorithm.

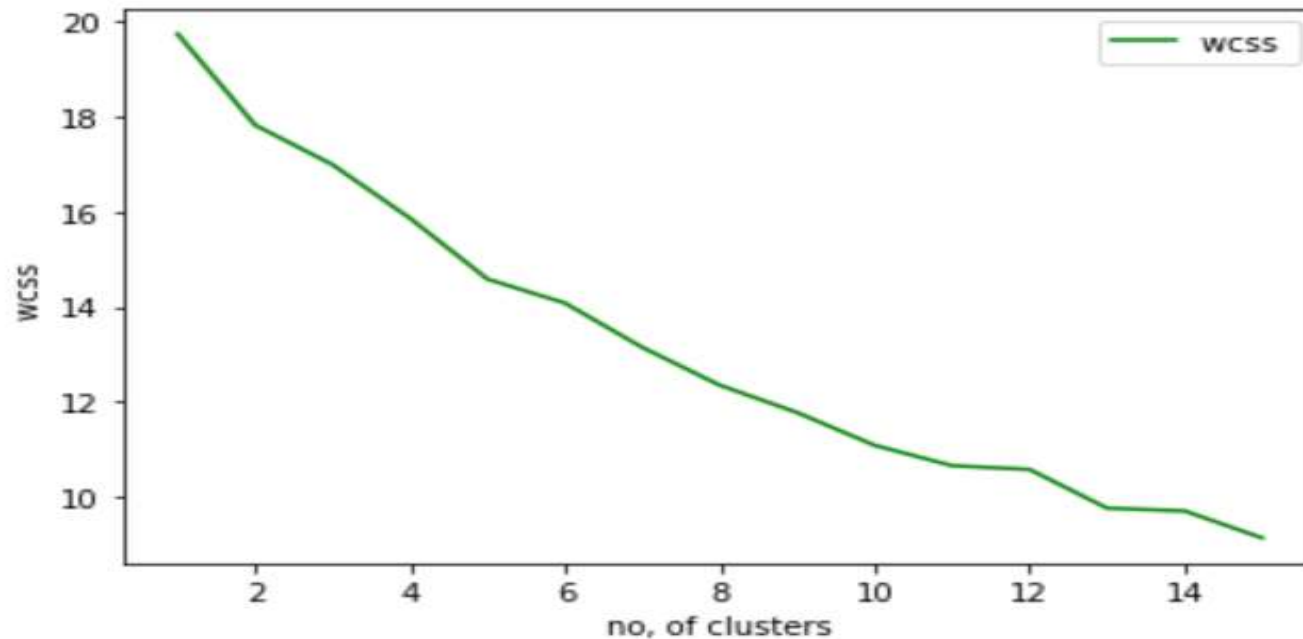
Model development (contd...)

- Use Onehot coding to convert Venue categories into 0's and 1's.
- Merge multiple rows into 1 row per neighborhood with average no. of venues nearby the neighborhood

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum
0	Adelaide,King,Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.00	0.010000	0.010000
1	Agincourt	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
2	Agincourt North,L'Amoreaux East,Miliken,Steel...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
3	Albion Gardens,Beaumont Heights,Humbergate,Jam...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
4	Alderwood,Long Branch	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
5	Bathurst Manor,Downsview North,Wilson Heights	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
6	Bayview Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
7	Bedford Park,Lawrence Manor	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	0.00	0.000000	0.000000

Model Evaluation and refinement

- Value of k needs to be pre-specified in K-means
- Select optimum k-value. Best k as per given plot is 15, which is based on error value.



Results

- The clustering algorithm assigned a cluster to each neighborhood based on similarities, as shown below:

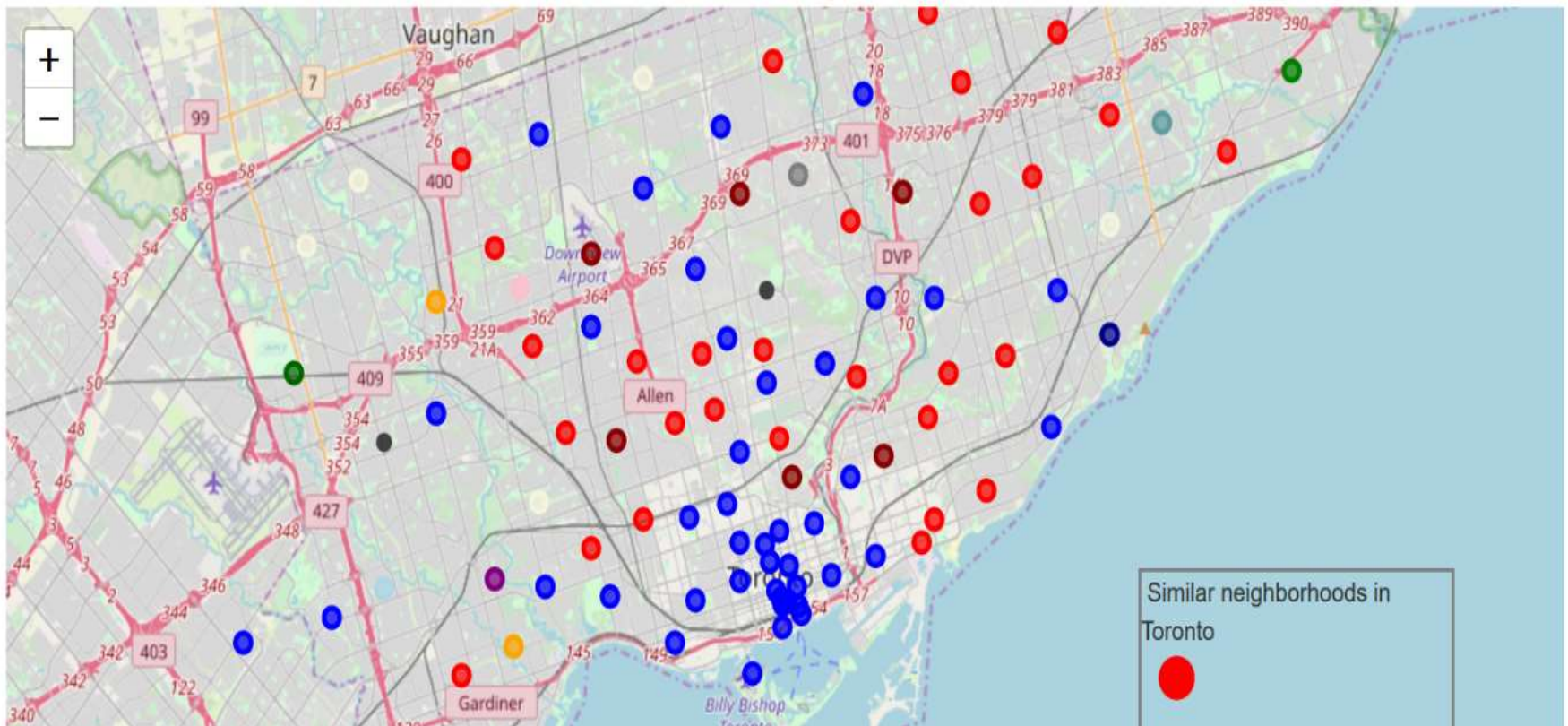
	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	3.0	Fast Food Restaurant	Drugstore	Dim Sum Restaurant	Diner	Discount Store
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	6.0	Bar	Women's Store	Drugstore	Diner	Discount Store
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	8.0	Electronics Store	Mexican Restaurant	Moving Target	Rental Car Location	Pizza Place
3	M1G	Scarborough	Woburn	43.770992	-79.216917	14.0	Coffee Shop	Indian Restaurant	Korean Restaurant	Women's Store	Drugstore
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	8.0	Hakka Restaurant	Bakery	Bank	Athletics & Sports	Caribbean Restaurant

Results (contd...)

- Cluster# assigned to current Newyork neighborhood is 8. So, display all Toronto neighborhoods with cluster# 8

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	8.0	Electronics Store	Mexican Restaurant	Moving Target	Rental Car Location	Pizza Place	Brea
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	8.0	Hakka Restaurant	Bakery	Bank	Athletics & Sports	Caribbean Restaurant	Restai
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	8.0	Bakery	Bus Line	Soccer Field	Bus Station	Fast Food Restaurant	Interse
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford He...	43.757410	-79.273304	8.0	Indian Restaurant	Pet Store	Chinese Restaurant	Latin American Restaurant	Vietnamese Restaurant	Ga
11	M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849	8.0	Auto Garage	Vietnamese Restaurant	Sandwich Place	Shopping Mall	Middle Eastern Restaurant	Brea
12	M1S	Scarborough	Agincourt	43.794200	-79.262029	8.0	Sandwich Place	Lounge	Skating Rink	Breakfast Spot	Print Shop	Won
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302	8.0	Pizza Place	Bank	Italian Restaurant	Rental Car Location	Noodle House	Pharr

- Visualize the clusters of Toronto neighborhoods via Folio map. Highlight the ones which are similar to current Newyork neighborhood.



Conclusion

- Just provide a Newyork neighborhood and get similar neighborhoods in Toronto.
- As said in problem statement, 1 in every 8 people relocate. So, the project will help them by reducing their manual effort by around 70%.
- No more a road blocker from house search perspective for career growth.
- Future goal is to enable the program to recommend neighborhoods to the people relocating from any location to any other location in world.

References

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare APIs
- Geospatial_Coordinates.csv
- nyu_2451_34572-geojson.json