Using Machine Learning to find the best housing option in Toronto

Carlos Eduardo Flores Tinoco
IBM Data Science Professional Certificate
November 2020

Migration shapes the society

- Globally, every week 3 million people move into cities.

- Urban population represents 56.2% of the worldwide community.

- The continually increasing urbanization and migration present great opportunities for economic and cultural thriving.

- At the same time, it also presents challenges for the migrants and receiving communities.

- The rate at which newcomers adapt to local rules and costumes will directly impact their integration to the society.

The diversity of the cities can be used to facilitate integration

- We can exploit this diversity as a means to facilitate immigrants with potential communities with similar taste and traditions.

- Devising a personalized recommendation system that considers the needs and preferences of an incoming person will heavily aid in the selection of the best fitting neighborhood upon migration to a new city.

Building a personalized recommendation system for relocation

- We must consider the spatial distribution of the incoming place and match them to the user's needs (e.g., nearby hospitals, restaurants, sports venues).

- Furthermore, it should also consider negative needs, such as dislike for heavily concurred places.

- The classification of businesses has to be as extensive and inclusive as possible.

Developing a neighborhood recommendation system based on Toronto venues

- The City of Toronto, Ontario, Canada is one of the most diverse cities in the world and one with a high immigration rate.

- The distinct types of venues and places are well documented either by public (e.g., Toronto Open Data) and private (e.g., FourSquare) entities.

Mapping of venues and places of interest in each neighborhood of Toronto

- 15180 spots of interest within 24 different categories obtained from Toronto Open Data or Foursqare API.

- Venues within 1.2km from the neighborhood center were considered in range.

Table 1. Number of venues in the city of Toronto

Group[a]	Venue[b]	Description[c]
Banks	113	Financial services
Cafe	422	Coffe, tea, pastry
Cinema	75	Theaters, cinema
Food Afr ME	82	Africa and Middle East Food
Food Asian	392	Eastern asian food
Food Gral	486	General food venues
Food Indo	57	Indian, Pakistani, and Tibet Food
Food Lat Am	103	Latin American and Caribbean Food
Food Mediterrean	296	Iberic, French, Italian, Greek Food
Food Other Europe	13	German, Swiss, Eastern European Food
Health Centers	51	Hospitals and health centers
Hotel	41	Hotels and hostels
Indoor Attr	11	Bowling, Arcade
Large Oudoor Attr	25	Stadiums, arenas, airports
Markets	208	Markets, grocery stores
Medical	90	Pharmacies
Museum	203	Museums, galleries, community centers
Outdoor Attr	381	Parks, zoos, lakes
Pub	250	Bars, pubs, clubs, lounges
Public transport	9472	Public transport stops and terminals
Salon	35	Beauty salon, spa
Schools	1194	Schools
Shops	1012	Shops, boutiques and services
Sports	168	Gym, courts, sports venues

[a] Group name
[b] Number of venues for the group indicated
[c] Examples of venues allocated in each group

Clustering Toronto neighborhoods based on the frequency and diversity of venues

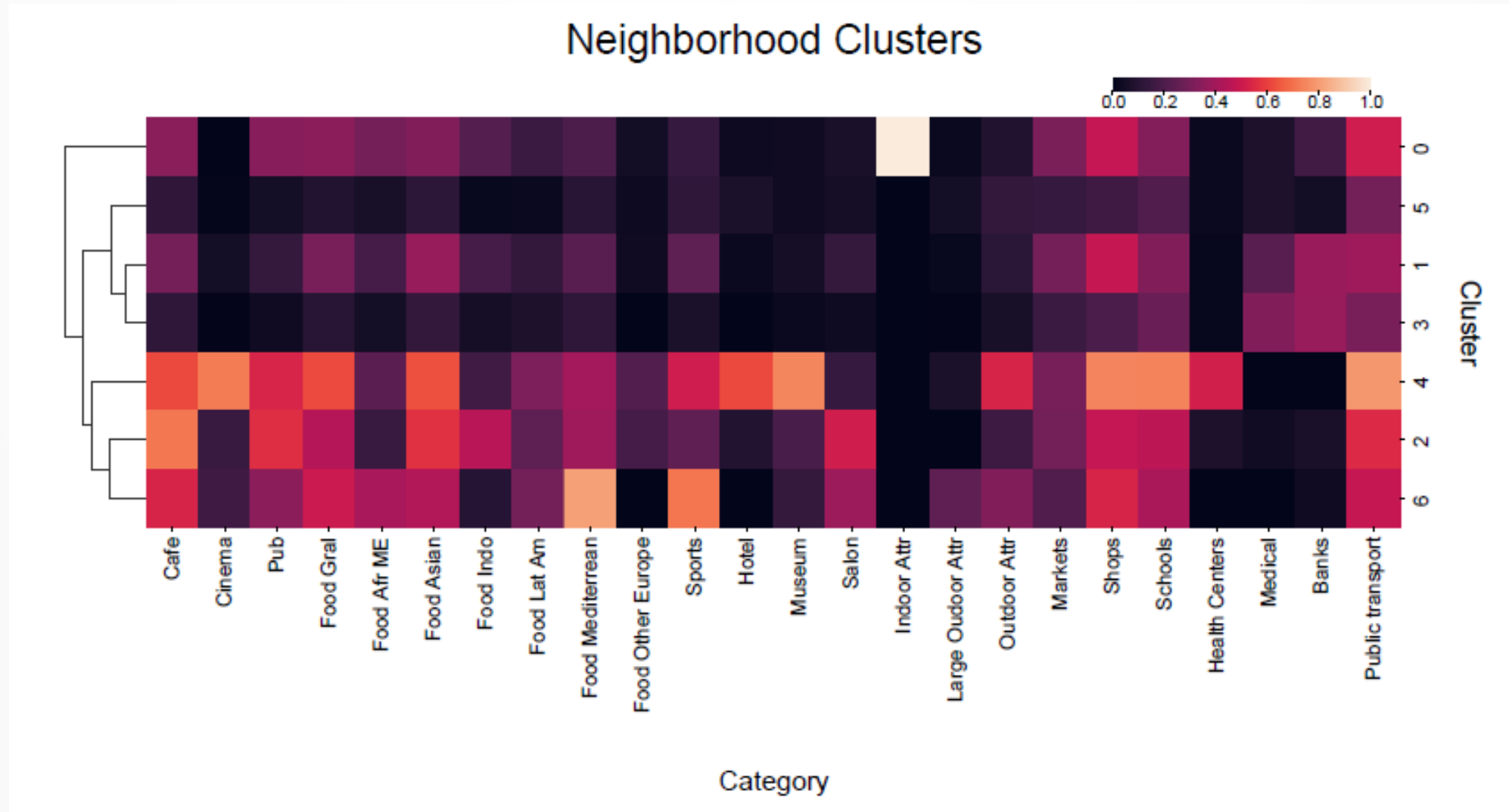- K-Means clustering of the neighborhoods revealed 7 clusters

Table 2. Number of venues in the city of Toronto

Cluster	Neighborhood[a]
0	10
1	24
2	11
3	36
4	7
5	46
6	6

[b] Number of neighborhood for the cluster indicated

Toronto Neighborhoods

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
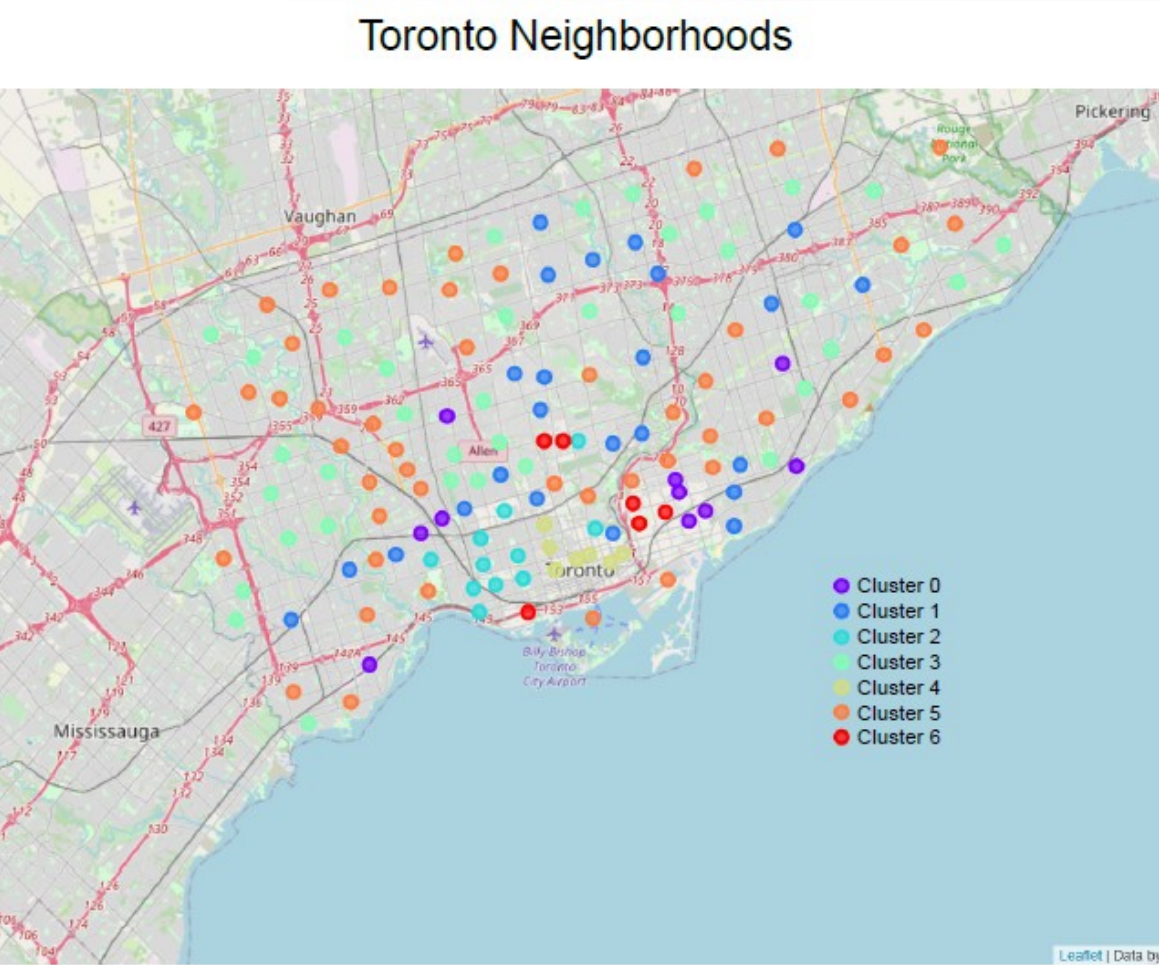
Clusters display differential variety and number of venues

Neighborhood Clusters

Using the clusters to recommend a newcomer the best fitting neighborhoods

- Case 1: A user that wants to be close to commercial venues

- Case 2: A user that prefers open spaces and wants to avoid large gatherings

Table 3. Neighborhood fit to Profile 1

Cluster	2	6	1	0	5	4	3
Similarity[a]	0.53	0.40	0.32	0.16	0.16	0.07	0.02

[a] Correlation between the user profile and the cluster

Table 4. Neighborhood fit to Profile 2

Cluster	5	6	3	0	1	4	2
Similarity[a]	0.30	0.08	0.07	0.06	0.05	-0.05	-0.18

[a] Correlation between the user profile and the cluster

Conclusion

- K-Means clustering of venues provided a functional personalized recommendation system of neighborhood relocation.

- Easy-to-use system used by individuals willing to find the best neighborhood to relocate as well as relocation agencies to provide better offers to their customers.

- This method can be expanded to any other city.

- This system is not limited to the housing market and could be used to forecast the success of business according to the surrounding venues.