

IBM – COURSERA

Data Science Professional Certificate

Applied Data Science Capstone Week 5

Introduction

This Project is for IBM Data Science Professional Certificate Program's Capstone. In this Project, hypothetical case is created by me that Toronto Area needs more Italian Restaurant. So, I will design this project to help me find the most suitable location for my Italian Restaurant.

Business Understanding

Problem : I want to open a new Italian Restaurant in Toronto, Canada.

Question : If anybody wants to open an Italian Restaurant in Toronto, where should they consider opening it?

Method : Data Science Methodology

Algorithm : Machine Learning Algorithm such as clustering

Data

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to Italian restaurants.

Source

- Toronto neighborhoods according to Wikipedia
- Geocoder Package
- Foursquare API

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

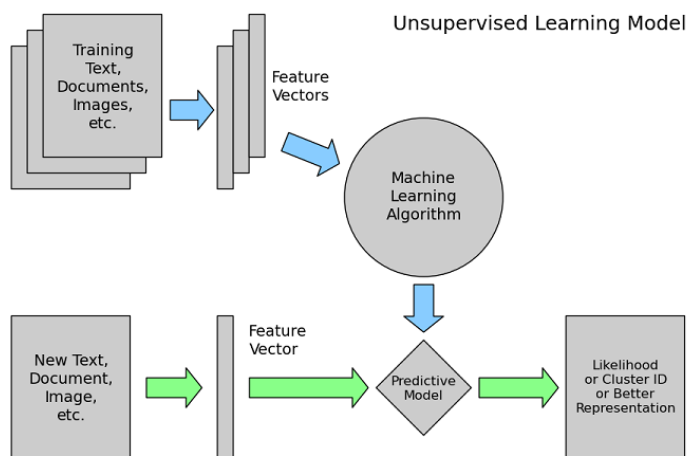
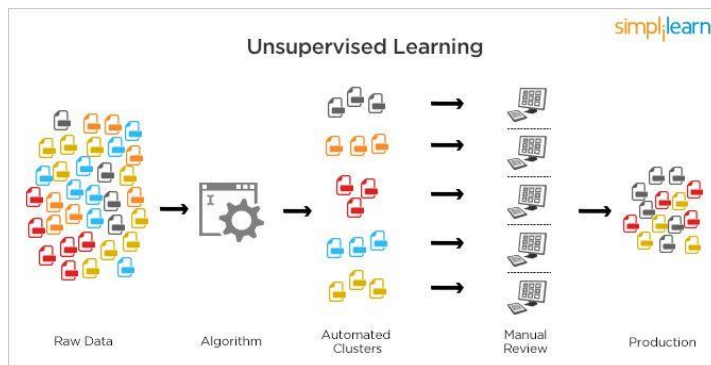
Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant
Adelaide, King, Richmond	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.020000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.020000
Berczy Park	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.018182
Brockton, Exhibition Place, Parkdale Village	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.000000
Business Reply Mail Processing Centre 969 Eastern	0.000000	0.0000	0.0000	0.0000	0.000	0.0000	0.000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.00	0.000000

Methodology

1. Get the list of neighborhoods from Wikipedia.
2. Scraping the data with Pandas HTML.
3. Match the coordinates of Toronto Neighborhoods with Geocoder Package.
4. Visualize the map of Toronto using Folium Package to verify correct coordinates.
5. Create Foursquare developer account.
6. Use Foursquare API to pull the list of top 100 venues within 500 meters Radius.
7. Analyze each neighborhood by grouping the rows by neighborhood and take the mean on the frequency of occurrence of each venue category.
8. Look for Italian Restaurant.
9. Performe the clustering method by using KMeans.
10. Cluster the neighborhoods in Toronto into three clusters based on their frequency of occurrence for Italian Food .
11. Recommend the ideal location to open the restaurant.

Unsupervised Learning



KMeans Clustering Algorithm

Input:

$D = \{t_1, t_2, \dots, t_n\}$ // Set of elements
 K // Number of desired clusters

Output:

K // Set of clusters

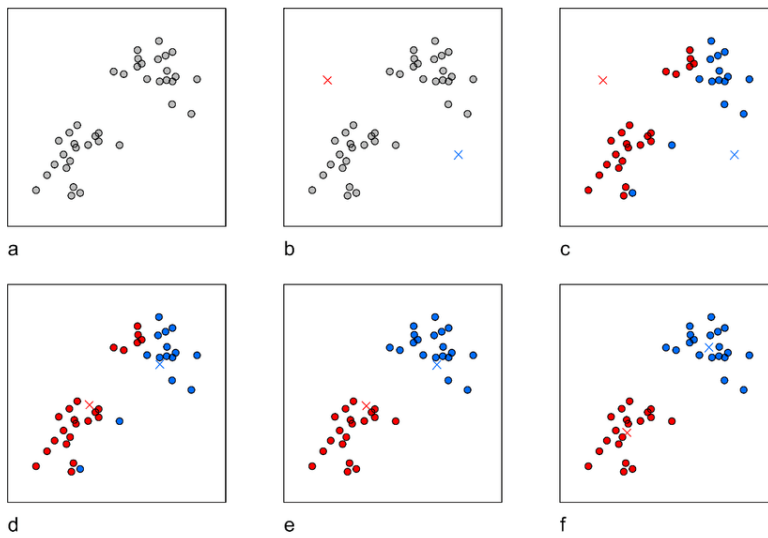
K-Means algorithm:

Assign initial values for m_1, m_2, \dots, m_k

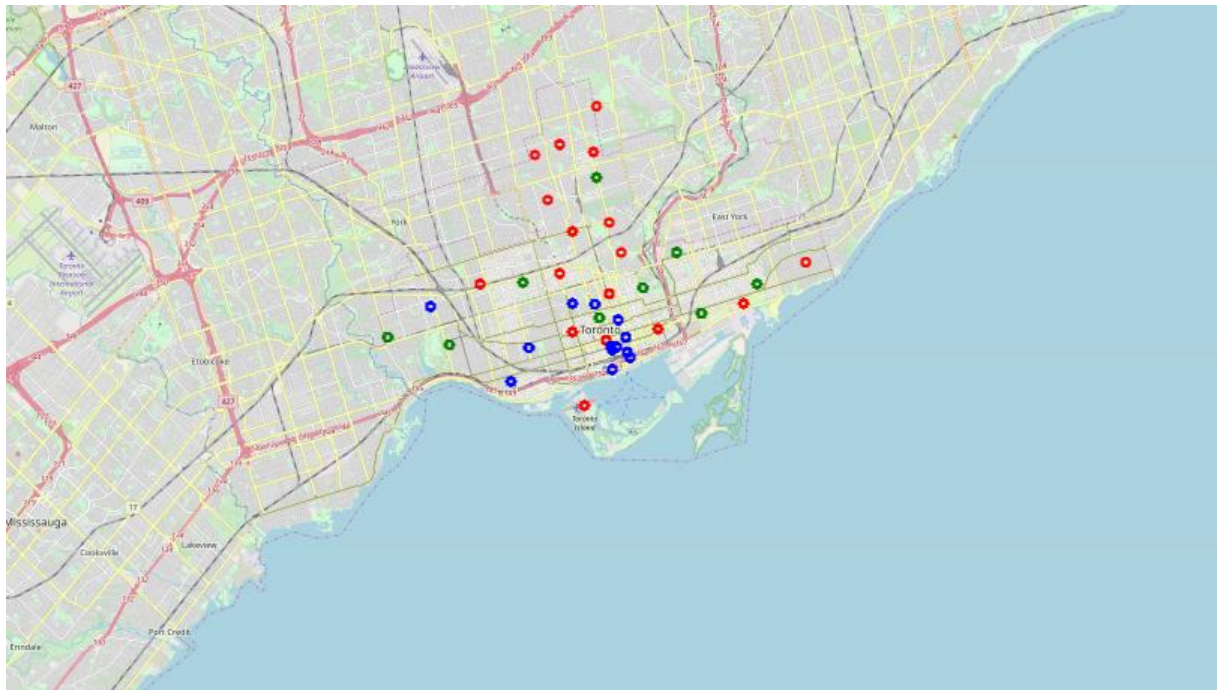
repeat

assign each item t_i to the clusters which has the closest mean;
calculate new mean for each cluster;

until convergence criteria is met;



Result



Cluster 0 : Neighborhoods with no Italian restaurants.

Cluster 1 : Neighborhoods with the more number of Italian restaurants.

Cluster 2 : Neighborhoods with the less number of Italian restaurants.

Recommendations

- Most of the Indian restaurants are in cluster 1.
- Lowest in Cluster 0 which are in North Toronto. Also, there are good opportunities to open.
- Looking at nearby venues it seems cluster 2 might be a good location as there are not a lot of Italian restaurants in these areas.
- Therefore, this project recommends the entrepreneur to open an Italian restaurant in these locations.