The best neighborhood in Toronto to open a Greek restaurant

Agenda

- Introduction to the problem
- Data collection and manipulation
- Machine learning application
- Results

Introduction to the problem

- ***Problem*** : Recommendation of the best area in Toronto for somebody who wants to start a Greek restaurant

- ***The idea*** : Find all neighborhoods in Toronto that have Greek restaurants and recommend the neighborhood with the least competition for a new entrepreneur.

- Based on geographical data.

Data collection and manipulation I

Data collected from 3 sources by :

- Scraping Wikipedia table for Toronto postal code
- Using geocoder library to get coordinates of each neighborhood
- Using Foursquare API to detect all the venues on each neighborhood

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Wikipedia data

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Geocoder cords addition to Wikipedia data

Data collection and manipulation II

lecting data from Foursquare
l:

ending a request to API

equesting all the venues in a 500m radius
om each Toronto's neighborhood

eceiving all the venues with their
nformation (i.e. cords, type etc.)

rganizing all the features in one dataframe
order to match the venues with their
eighborhood

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue C
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portugu Restaura
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee S
...	...	...	...	...	...	...	...
2095	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Royal Canadian Legion #210	43.628855	-79.518903	Social C
2096	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Koala Tan Tanning Salon & Sunless Spa	43.631370	-79.519006	Tanning
2097	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Value Village	43.631269	-79.518238	Thrift / V Store

Machine learning application I

Clustering approach:

- Clustering the Neighborhood based on Greek restaurant density
- Each neighborhood might contain more than one Greek restaurants
- The goal is to find the cluster of neighborhoods with small Greek restaurant density
- For the neighborhoods of this cluster to find the one with the smallest density and recommend it

Machine learning application II

Data preparation:

- One-hot encoding the data based on venue category
- Grouping the venues by neighborhoods and calculating mean value of each one
  - Mean value will be the density of each venue in the specific neighborhood
- Filtering to keep only the Greek restaurants
- Keeping only the columns with Neighborhoods and Greek restaurants' densities

	Neighborhoods	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Train Station	Truck Stop	Vegetarian / Vegan Restaurant	Video Game Store	Vietnames Restauran
0	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Parkwoods	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Victoria Village	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

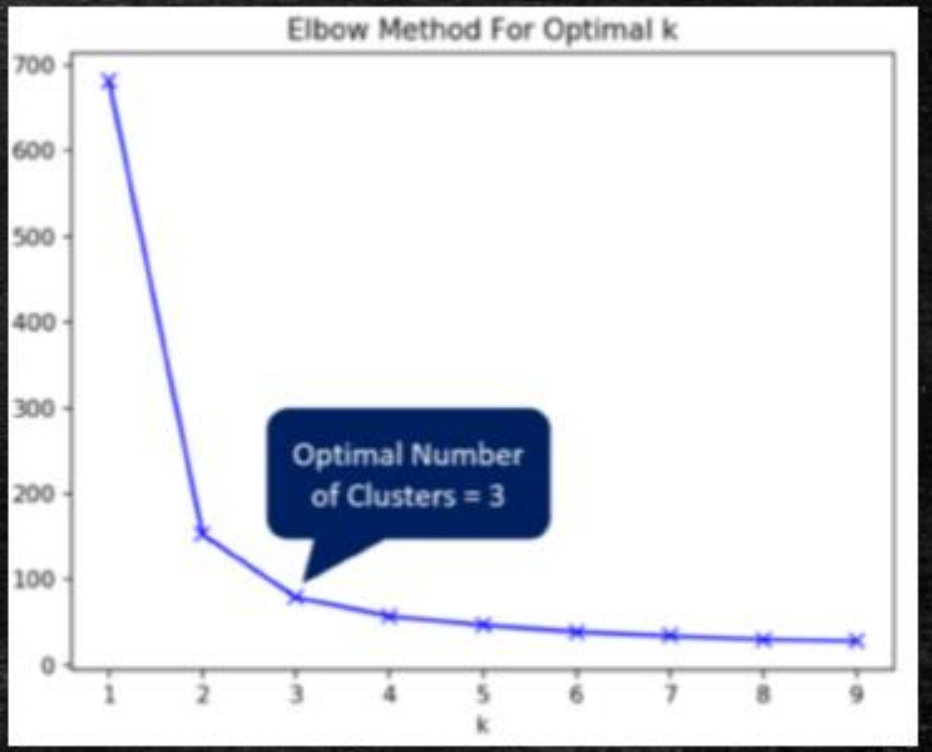
	Neighborhoods	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Train Station	Truck Stop	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0

	Neighborhoods	Greek Restaurant
4	Bedford Park, Lawrence Manor East	0.043478
5	Berczy Park	0.017241
19	Davisville	0.027778
29	First Canadian Place, Underground city	0.010000

Machine learning application III

Clustering:

- Using K-Means ML algorithm
- K values tested (1 to 9)
- Using elbow method to determine that k=3 is the best
- Merging neighborhood density and cluster number with the initial data

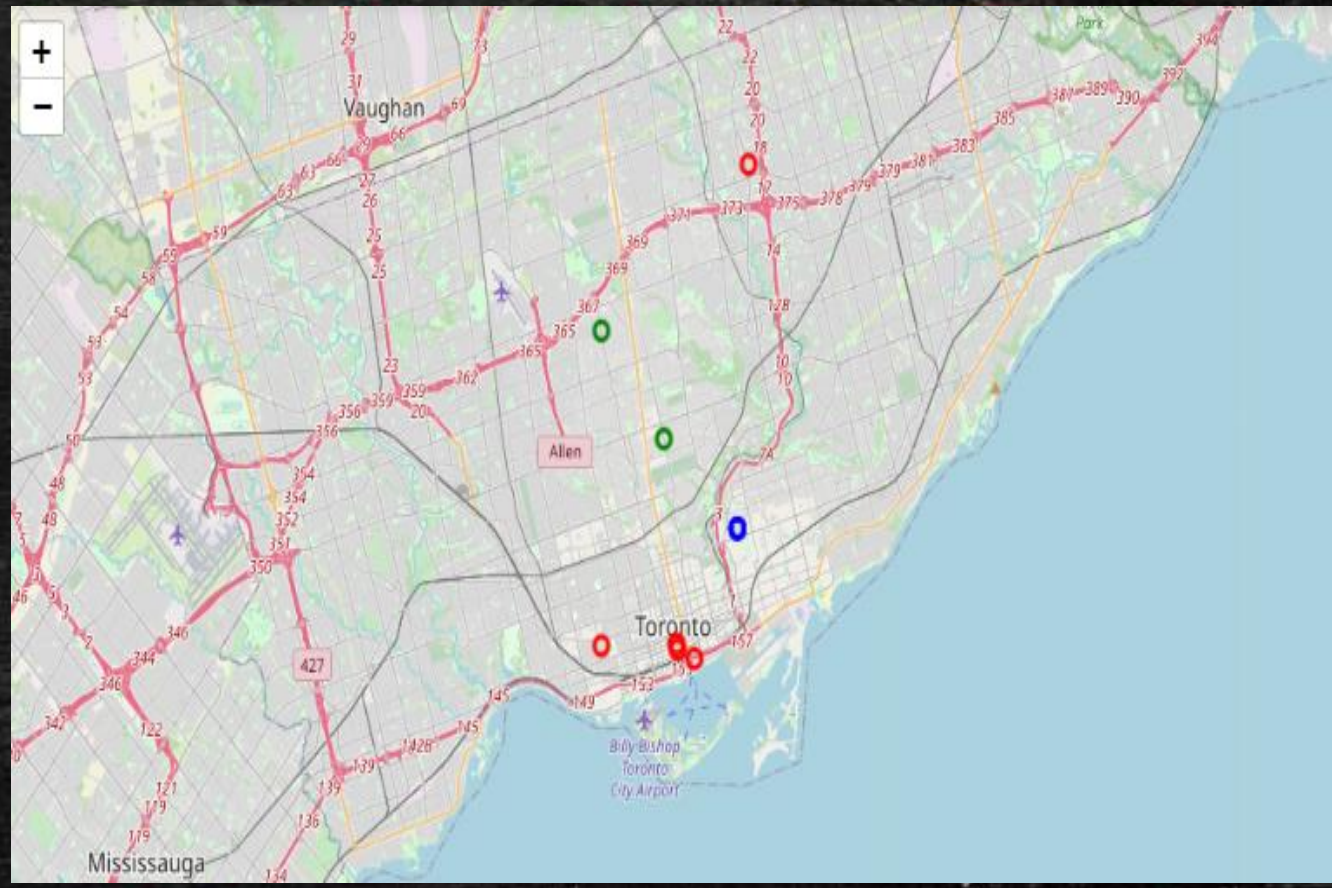


	Neighbourhood	Greek Restaurant	Cluster	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
11	Bedford Park, Lawrence Manor East	0.043478	2	43.733283	-79.419750	Karbouzi Greek Taverna	43.736204	-79.420359	Greek Restaurant
53	Berczy Park	0.017241	0	43.644771	-79.373306	Alexandro's World Famous Gyros	43.641663	-79.375214	Greek Restaurant
101	Davisville	0.027778	2	43.704324	-79.388790	souvlaki express	43.707378	-79.389648	Greek Restaurant
161	First Canadian Place, Underground city	0.010000	0	43.648429	-79.382280	Estiatorio Volos	43.650329	-79.384533	Greek Restaurant

Results I

Clusters

- The neighborhoods organized in 3 clusters
- The creation of each cluster was based on its density
- The next step is to determine which cluster has the neighborhoods with least Greek restaurants



Results II

Counting Greek restaurants:

- Grouping all the rows of the merged data frame by the cluster number and Neighborhood
- Using count() aggregation function to count the Greek Restaurants.
- Choosing the cluster with least Greek restaurants
- Recommending the neighborhoods of this cluster with least Greek restaurants

Results III

Recommendation:

- From the above cluster the one with the least restaurants is Cluster 2 (green)

Cluster	Neighbourhood	Greek Restaurant
0	Berczy Park	1
	Fairview, Henry Farm, Oriole	1
	First Canadian Place, Underground city	1
	Little Portugal, Trinity	1
	Toronto Dominion Centre, Design Exchange	1
1	The Danforth West, Riverdale	7
2	Bedford Park, Lawrence Manor East	1
	Davisville	1

Thank you!