Identification of a

suitable neighbourhood in Toronto

to start a retail store

to sell excellent Mediterranean olive oil

- Full report -

Capstone Project – The Battle of Neighborhoods –
 IBM / Coursera

**April 2019** 



## The challenge is to find the most suitable neighbourhood for setting up a small sales outlet in Toronto to sell excellent Mediterranean olive oil

- A trading company specialized in trading with olive oil would like to expand internationally
- Because of
  - the large number of inhabitants with Mediterranean roots (who very much like to use olive oil)
  - the high average income (some olive oil varieties are expensive) and
  - the large number of Mediterranean restaurants (which are potential clients and can be considered multipliers)

Toronto was chosen to open a new retail store.

 However, Toronto is large. For this reason, the trading company has asked a consulting company to find the most suitable neighbourhood for setting up a small sales outlet in Toronto.



 Olive oil is a liquid fat obtained from olives, a traditional tree crop of the Mediterranean Basin. It is commonly used in cooking, whether for frying or as a salad dressing.

 Olive trees have been grown around the Mediterranean since the 8th millennium BC.

 Choosing a cold-pressed olive oil can be similar to selecting a wine. The flavor of these oils varies considerably and a particular oil may be more suited for a particular dish.



Sources:
Wikipedia,
International Olive Council,
https://atasteofolive.com/products/novello-di-gradassi

### Spain is the largest producer of olive oil, followed by Greece and Italy

- Largest Producers of Olive Oil -

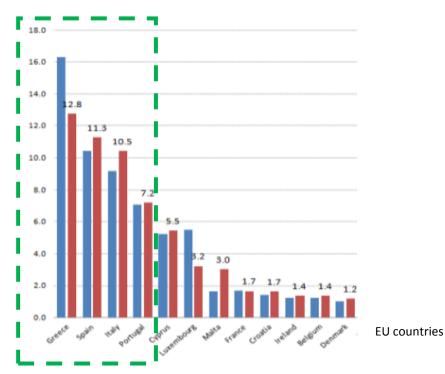
Country	Production [tonnes]	Production [% of global]
Spain	1,290,600	50
Greece	195,000	8
Italy	182,300	7
Turkey	178,000	7
World	2,586,500	100

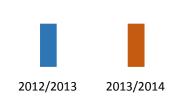
Source: International Olive Council

### Per capita consumption is highest in Greece, followed by Spain, Italy, and Portugal

- Per capita olive oil consumption in EU countries [kg] -

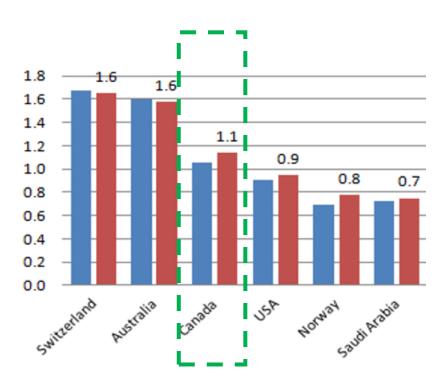
Per capita olive oil consumption [kg]

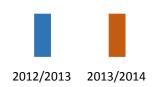




#### Consumption in Canada is far less, but rising

- Per capita olive oil consumption in some non-EU countries [kg] -





# Because of its size, importance within Canada and diverse population Toronto is well suited to be part of the international expansion

- Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016.
- Toronto is a centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.
- The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. Over 200 distinct ethnic origins are represented among its inhabitants.
- While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.



**Business Problem** 

## The trading company is looking for the most suitable neighbourhood for setting up a small sales outlet in Toronto taking into account three main parameters

- The trading company is looking for the most suitable neighbourhood for setting up a small sales outlet in Toronto.
- Important decision criteria are:
  - the **average income** in the neighbourhood as the olive oil varieties with the highest profit margins are expensive and
  - the **number of Mediterranean restaurants** as they are potential customers and can be considered multipliers
  - the **number of inhabitants with Mediterranean roots** as these inhabitants are the consumers who very much like to use olive oil
- The neighbourhood with the best combination of these parameters will be the first choice. However, because of the market entry strategy of the trading company (using restaurants as multipliers) and the quite expansive oils the two first criteria are most important

Approach

### The analysis is done using a six step approach

1	Start with the Toronto neighbourhoods from the assignment of week 3
2	Choose only the neighbourhoods with strong Mediterranean influence (a Mediterranean language as second most important language)
3	Add information on average income and number of inhabitants with Mediterranean roots using information from Wikipedia
4	Add information on the number of Mediterranean restaurants within each Neighbourhood using Foursquare data
5	Cluster the information from all these sources using k-means
6	Final decision

## The starting point of the analysis is the table with the neighbourhoods that has already been used in the assignment of week 3 and has been combined with geospatial data

Postcode +	Borough +	Neighbourhood \$
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
МЗА	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M2B	Not assigned	Not assigned
мзв	North York	Don Mills North
M4B	East York	Woodbine Gardens
M4B	East York	Parkview Hill
M5B	Downtown Toronto	Ryerson
	5	



Postal Code, Latitude, Longitude M1B,43.8066863,-79.1943534 M1C,43.7845351,-79.1604971 M1E,43.7635726,-79.1887115 M1G,43.7709921,-79.2169174 M1H,43.773136,-79.2394761 M1J,43.7447342,-79.2394761 M1K,43.7279292,-79.2620294 M1L,43.7111117,-79.2845772 M1M,43.716316,-79.2394761 M1N,43.692657,-79.2648481 M1P,43.7574096,-79.273304 M1R,43.7500715,-79.2958491 M1S,43.7942003,-79.2620294 M1T,43.7816375,-79.3043021 M1V,43.8152522,-79.2845772 M1W,43.7995252,-79.3183887 M1X,43.8361247,-79.2056361 M2H,43.8037622,-79.3634517 M2J,43.7785175,-79.3465557 M2K,43.7869473,-79.385975 M2L,43.7574902,-79.3747141 M2M,43.789053,-79.4084928 M2N,43.7701199,-79.4084928 M2P,43.7527583,-79.4000493 M2R,43.7827364,-79.4422593 M3A,43.7532586,-79.3296565 M3B,43.7459058,-79.352188 M3C 43 7258997 -79 340923



Pos	stal Code	Borough	Neighbourhood	Latitude	Longitude
0	МЗА	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M5A	Downtown Toronto	Regent Park	43.654260	-79.360636
4	M6A	North York	Lawrence Heights	43.718518	-79.464763
5	M6A	North York	Lawrence Manor	43.718518	-79.464763
6	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
7	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
8	M1B	Scarborough	Rouge	43.806686	-79.194353
9	M1B	Scarborough	Malvern	43.806686	-79.194353

Source:

https://en.wikipedia.org/wiki/List\_of\_postal codes of Canada: M

Source:

 $"http://cocl.us/Geospatial\_data"$ 

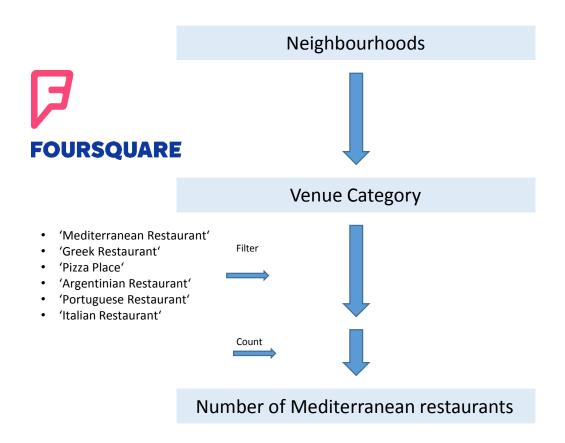
# To filter the neighbourhoods on inhabitants with Mediterranean roots a Wikipedia showing the "second most common language" within a neighbourhood is used

Name	<b>♦ FM ♦</b>	Census Tracts	Population +	Land area ¢ (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after \$ English) by name	Second most common language (after ÷ English) by percentage	Мар
Toronto CMA Average		All	5,113,149	5903.63	866	9.0	40,704	10.6	11.4			
Agincourt	s	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0378.08, 0378.14, 0378.23, 0378.24	44,577	12.45	3580	4.6	25,750	11.1	5.9	Cantonese (19.3%	19.3% Cantonese	Agincourt
Alderwood	Е	0211.00, 0212.00	11,656	4.94	2360	-4.0	35,239	8.8	8.5	Polish (6.2%)	06.2% Polish	
Alexandra Park	OCoT	0039.00	4,355	0.32	13,609	0.0	19,687	13.8	28.0	Cantonese (17.9%	17.9% Cantonese	Sing 5 Sing 5 Sing 5 Sing 6 Si
Allenby	OCoT	0140.00	2,513	0.58	4333	-1.0	245,592	5.2	3.4	Russian (1.4%)	01.4% Russian	
Amesbury	NY	0280.00, 0281.01, 0281.02	17,318	3.51	4,934	1.1	27,546	16.4	19.7	Spanish (6.1%)	06.1% Spanish	Amesbury
Armour Heights	NY	0298.00	4,384	2.29	1914	2.0	116,651	10.8	16.1	Russian (9.4%)	09.4% Russian	Armour Helphis
Banbury	NY	0267.00	6,641	2.72	2442	5.0	92,319	6.1	4.8	Unspecified Chinese (5.1%)	05.1% Unspecified Chinese	
Bathurst Manor	NY	0297.01, 0310.01, 0310.02	14,945	4.69	3187	12.3	34,169	13.4	18.6	Russian (9.5%)	09.5% Russian	Bithart -

# The same table can be used to make a ranking on average income and the number of inhabitants with a certain origin

Name	FM ÷	Census Tracts	Population +	Land area ‡ (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit \$	% Renters	Second most common language (after \$ English) by name	Second most common language (after \$ English) by percentage	Map ‡
Toronto CMA Average		All	5,113,149	5903.63	866	9.0	40,704	10.6	11.4			
Agincourt	S	0377.01, 0377.02, 0377.03, 0377.0 0378.02, 0378.08, 0378.14, 0378.23, 0378.24		12.45	3580	4.6	25,750	1.1.	5.9	Cantonese (19.3%	19.3% Cantonese	Agincourt
Alderwood	E	0211.00, 0212.00	11,656	.94	2360	-4.0	35,239	.8	8.5	Polish (6.2%)	06.2% Polish	
Alexandra Park	OC <sub>0</sub> T	0039.00	4,355	.32	13,609	0.0	19,687	3.8	28.0	Cantonese (17.9%	17.9% Cantonese	Aremandra Bark
Allenby	OCoT	0140.00	2,513	9.58	4333	-1.0	245,592	5.2	3.4	Russian (1.4%)	01.4% Russian	А. О.
Amesbury	NY	0280.00, 0281.01, 0281.02	17,318	3.51	4,934	1.1	27,546	16.4	19.7	Spanish (6.1%)	06.1% Spanish	Amedbury
Armour Heights	NY	0298.00	4,384	:29	1914	2.0	116,651	0.8	16.1	Russian (9.4%)	09.4% Russian	Armout
Banbury	NY	0267.00	6,641	2.72	2442	5.0	92,319	.1	4.8	Unspecified Chinese (5.1%)	05.1% Unspecified Chinese	
Bathurst Manor	NY	0297.01, 0310.01, 0310.02	14,945	4.69	3187	12.3	34,169	3.4	18.6	Russian (9.5%)	09.5% Russian	

### To identify the number of Mediterranean restaurants within a neighbourhood Foursquare information is used and filtered



#### As a last step a k-means clustering and a ranking is carried out

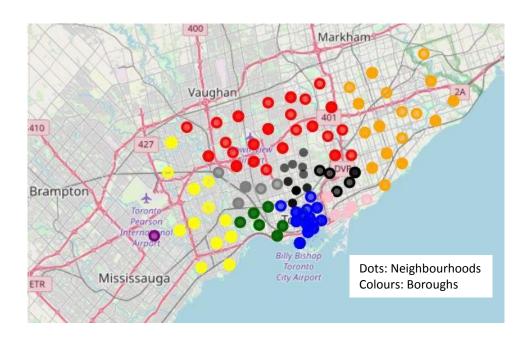
Number of inhabitants with Number of Mediterranean Average income Mediterranean roots restaurants K-means-Clustering Ranking





The starting point of the analysis is formed by the identification of boroughs and neighbourhoods from the assignment of week 3

- Boroughs and neighbourhoods of Toronto -



 Neighbourhoods with strong Mediterranean influence

### Data from Wikipedia show relevant information on the inhabitants of Toronto

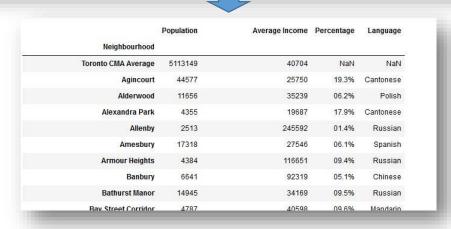
	Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
0	Toronto CMA Average	NaN	All	5113149	5903.63	866	9.0	40704	10.6	11.4	NaN	NaN	NaN
1	Agincourt	S	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0	44577	12.45	3580	4.6	25750	11.1	5.9	Cantonese (19.3%)	19.3% Cantonese	NaN
2	Alderwood	E	0211.00, 0212.00	11656	4.94	2360	-4.0	35239	8.8	8.5	Polish (6.2%)	06.2% Polish	NaN
3	Alexandra Park	OC <sub>0</sub> T	0039.00	4355	0.32	13609	0.0	19687	13.8	28.0	Cantonese (17.9%)	17.9% Cantonese	NaN
4	Allenby	OCoT	0140.00	2513	0.58	4333	-1.0	245592	5.2	3.4	Russian (1.4%)	01.4% Russian	NaN



Neighbourhood

# The dataset on the inhabitants is cleaned and the interesting columns (neighbourhood, population, average income, percentage of inhabitants with a certain language, language) are selected

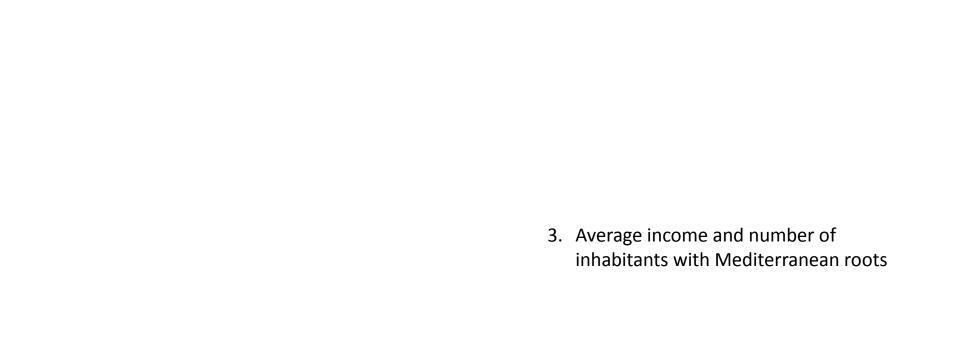
Мар	Second most common language (after English) by percentage	Second most common language (after English) by name	% Renters	Transit Commuting %	Average Income	% Change in Population since 2001	Density (people/km2)	Land area (km2)	Population	Census Tracts	FM	Name	
NaN	NaN	NaN	11.4	10.6	40704	9.0	866	5903.63	5113149	All	NaN	Toronto CMA Average	0
NaN	19.3% Cantonese	Cantonese (19.3%)	5.9	11,1	25750	4.6	3580	12.45	44577	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0	S	Agincourt	1
NaN	06.2% Polish	Polish (6.2%)	8.5	8.8	35239	-4.0	2360	4.94	11656	0211.00, 0212.00	E	Alderwood	2
NaN	17.9% Cantonese	Cantonese (17.9%)	28.0	13.8	19687	0.0	13609	0.32	4355	0039.00	OC <sub>0</sub> T	Alexandra Park	3
NaN	01.4% Russian	Russian (1.4%)	3.4	5.2	245592	-1.0	4333	0.58	2513	0140.00	OCoT	Allenby	4



# After merging the dataframes from the first two steps the neighbourhoods can be filtered on the relevant languages

	Postal Code	Borough	Latitude	Longitude	Population	Average Income	Percentage	Language
Neighbourhood								ı
Lawrence Heights	M6A	North York	43.718518	-79.464763	3769	29867	15.0	Italian
Eringate	M9C	Etobicoke	43.643515	-79.577201	8008	34789	3.4	Portuguese
Little Portugal	M6J	West Toronto	43.647927	-79.419750	5013	29224	23.8	Portuguese
Brockton	M6K	West Toronto	43.636847	-79.428191	9039	27260	19.9	Portuguese
Downsview	M6L	North York	43.713756	-79.490074	36613	26751	11.7	Italian
<b>Humber Summit</b>	M9L	North York	43.756303	-79.565963	12766	26117	15.1	Italian
Bedford Park	M5M	North York	43.733283	-79.419750	13749	80827	0.7	Greek
Mount Dennis	М6М	York	43.691116	-79.476013	21284	23910	7.3	Spanish
Silverthorn	М6М	York	43.691116	-79.476013	17757	26291	11.8	Portuguese
Humberlea	М9М	North York	43.724766	-79.532242	4327	30907	11.1	Italian
Weston	M9N	York	43.706876	-79.518188	16476	27446	7.0	Spanish
The Annex	M5R	Central Toronto	43.672710	-79.405678	15602	63636	1.3	Spanish
nurch and Wellesley	M4Y	Downtown Toronto	43.665860	-79.383160	13397	37653	1.8	Spanish

Only neighbourhoods with the second language "Italian", "Portuguese", "Greek", "Spanish" are chosen



### There is one neighbourhood with distinctly more inhabitants with Mediterranean roots: Downsview

#### - Ranking inhabitants with Mediterranean root -

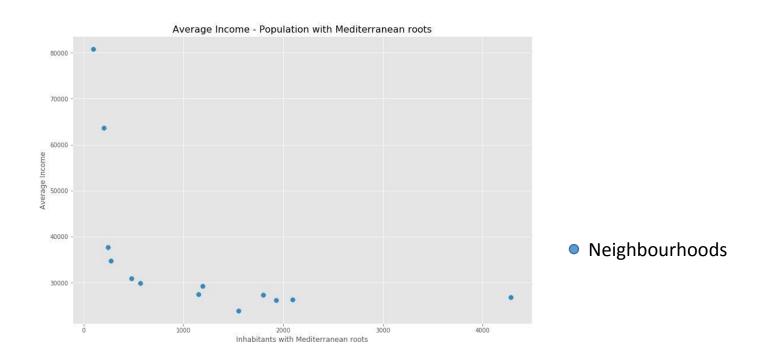
	Neighbourhood	Population	Average Income	Percentage	Inhabitants with Mediterranean roots	Inhabitants Ranking
4	Downsview	36613	26751	11.7	4283	1.0
8	Silverthorn	17757	26291	11.8	2095	2.0
5	Humber Summit	12766	26117	15.1	1927	3.0
3	Brockton	9039	27260	19.9	1798	4.0
7	Mount Dennis	21284	23910	7.3	1553	5.0
2	Little Portugal	5013	29224	23.8	1193	6.0
10	Weston	16476	27446	7.0	1153	7.0
0	Lawrence Heights	3769	29867	15.0	565	8.0
9	Humberlea	4327	30907	11.1	<mark>4</mark> 80	9.0
1	Eringate	8008	34789	3.4	272	10.0
12	Church and Wellesley	13397	37653	1.8	241	11.0
11	The Annex	15602	63636	1.3	202	12.0
6	Bedford Park	13749	80827	0.7	96	13.0

# Downsview, the neighbourhood with distinctly more inhabitants with Mediterranean roots, has only low average income

#### - Ranking average income -

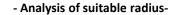
	Neighbourhood	Population	Average Income	Percentage	Average Income Rank
6	Bedford Park	13749	80827	0.7	1.0
11	The Annex	15602	63636	1.3	2.0
12	Church and Wellesley	13397	37653	1.8	3.0
1	Eringațe	8008	34789	3.4	4.0
9	Humberlea	4327	30907	11.1	5.0
0	Lawrence Heights	3769	29867	15.0	6.0
2	Little Portugal	5013	29224	23.8	7.0
10	Weston	16476	27446	7.0	8.0
3	Brockton	9039	27260	19.9	9.0
4	Downsview	36613	26751	11.7	10.0
8	Silverthorn	17757	26291	11.8	11.0
5	Humber Summit	12766	26117	15.1	12.0
7	Mount Dennis	21284	23910	7.3	13.0

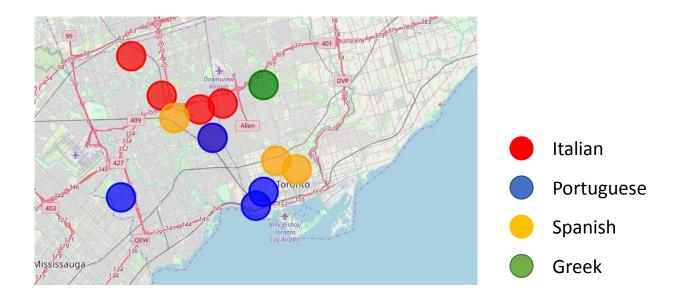
## There is an inverse relationship between average income and number of inhabitants with Mediterranean roots. For this analysis, average income is more important



4. Foursquare analysis to identify the number of Mediterranean restaurants

#### To be able to perform a Foursquare analysis a suitable radius of neighbourhoods has to be identified





A radius of 1250m seems to be appropriate

### A Foursquare analysis identifies the restaurants within each neighbourhood

#### - Foursquare request -

Venue Category	Venue Longitude	Venue Latitude	Neighbourhood Longitude	Neighbourhood Latitude	Neighbourhood	
Boutique	-79.466776	43.718221	-79.464763	43.718518	Lawrence Heights	0
Vietnamese Restaurant	-79.468472	43.721259	-79.464763	43.718518	Lawrence Heights	1
Deli / Bodega	-79.456504	43.721445	-79.4 <mark>64763</mark>	43.718518	Lawrence Heights	2
Sushi Restaurant	-79.456119	43.720610	-79.464763	43.718518	Lawrence Heights	3
Dessert Shop	-79.456219	43.720179	-79.464763	43.718518	Lawrence Heights	4
Burger Joint	-79.456412	43.720730	-79.464763	43.718518	Lawrence Heights	5
Furniture / Home Store	-79.457130	43.724957	-79.464763	43.718518	Lawrence Heights	6
Greek Restaurant	-79.455855	43.718516	-79.464763	43.718518	Lawrence Heights	7
Athletics & Sports	-79.463398	43.724054	-79.464763	43.718518	Lawrence Heights	8
Fried Chicken Joint	-79.455706	43.718309	-79.464763	43.718518	Lawrence Heights	9
Furniture / Home Store	-79.462675	43.719096	-79.464763	43.718518	Lawrence Heights	0

Used to check for duplicates

Relevant column

### By filtering the result of the Foursquare request on the relevant restaurant types and counting the occurrences the number of Mediterranean restaurants are identified

Venue Category	Venue Longitude	Venue Latitude	Neighbourhood Longitude	Neighbourhood Latitude	Neighbourhood	
Boutique	-79.466776	43.718221	-79.464763	43.718518	Lawrence Heights	0
Vietnamese Restaurant	-79.468472	43.721259	-79.464763	43.718518	Lawrence Heights	1
Deli / Bodega	-79.456504	43.721445	-79.464763	43.718518	Lawrence Heights	2
Sushi Restaurant	-79.456119	43.720610	-79.464763	43.718518	Lawrence Heights	3
Dessert Shop	-79.456219	43.720179	-79.464763	43.718518	Lawrence Heights	4
Burger Joint	-79.456412	43.720730	-79.464763	43.718518	Lawrence Heights	5
Furniture / Home Store	-79.457130	43.724957	-79.464763	43.718518	Lawrence Heights	6
Greek Restaurant	-79.455855	43.718516	-79.464763	43.718518	Lawrence Heights	7
Athletics & Sports	-79.463398	43.724054	-79.464763	43.718518	Lawrence Heights	8
Fried Chicken Joint	-79.455706	43.718309	-79.464763	43.718518	Lawrence Heights	9
Furniture / Home Store	-79 462675	43 719096	-79 464763	43 718518	Lawrence Heights	10



#### **Filter**

- 'Mediterranean Restaurant'
- · 'Greek Restaurant'
- 'Pizza Place'
- 'Argentinian Restaurant'
- 'Portuguese Restaurant'
- 'Italian Restaurant'

f Mediterranean restaurar	
	Neighbourhood
	The Annex
	Little Portugal
	Bedford Park
	Church and Wellesley
	Mount Dennis
	Brockton
	Humber Summit
	Eringate
	Weston
	Silverthorn
	Downsview
	Lawrence Heights

5. Clustering using k-means

### A k-means analysis with k=2 shows the importance of the number of restaurants

- k=2 -

	Average Income	Inhabitants with Mediterranean roots	Number of Mediterranean restaurants	Labels	
0	29867	565	1	0	
1	34789	272	2	0	
3	27260	1798	2	0	
4	26751	4283	2	0	Cluster 1
5	26117	1927	2	0	Clustel 1
7	23910	1553	2	0	
8	26291	2095	2	0	
9	27446	1153	2	0	
2	29224	1193	9	1	
6	80827	96	8	1	Cluster 2
10	63636	202	10	1	Cluster 2
11	37653	241	8	1	

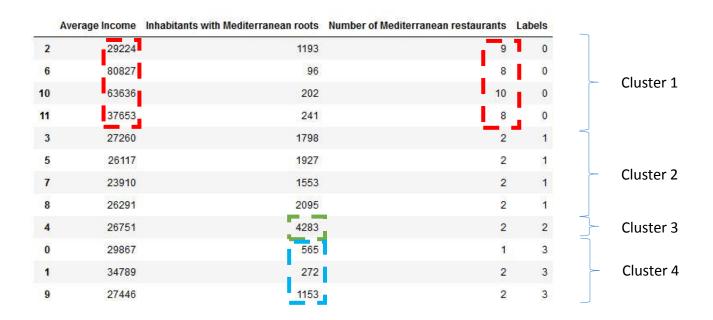
Next to the number of restaurants a k-means analysis with k=3 shows the special relevance of the neighbourhood with a large number inhabitants with Mediterranean roots (Downsview)

- k=3 -

	Average Income	Inhabitants with Mediterranean roots	Number of Mediterranean restaurants	Labels	
0	29867	<b>5</b> 65	1	0	
1	34789	272	2	0	
3	27260	1798	2	0	
5	26117	1927	2	0	Cluster 1
7	23910	1553	2	0	
8	26291	2095	2	0	
9	27446	1153	2	0	
2	29224	1193	9	1	
6	80827	96	8	1	Cluster 2
10	63636	202	10	1	Cluster 2
11	37653	241	8	1	
4	26751	4283	2	2	Cluster 3

The fourth cluster is characterized by a low number of inhabitants and a medium number of restaurants

- k=3 -



6. Final decision

As the number of restaurants, the average income and – to a lesser extend - the inhabitants with Mediterranean roots play an important role these neighbourhoods are chosen for the final recommendation

- k=3 -

	Average Income	Inhabitants with Mediterranean roots	Number of Mediterranean restaurants	Labels		
0	29867	565	1	0		
1	34789	272	2	0		
3	27260	1798	2	0		
5	26117	1927	2	0	_	Clu
7	23910	1553	2	0		
8	26291	2095	2	0		
9	27446	1153	2	0		
2	29224	1193	9	1		
6	80827	96	8	1		Clus
0	63636	202	10	1		Clus
1	37653	241	8	_ 1		
4	26751	4283	2	2		Clus

## To come to a final conclusion the locations of the neighbourhoods relative to one another have to be taken into account

- As stated before, the most important criteria for the new business are
  - Average income
  - Number of Mediterranean restaurants and, to a lesser extent
  - Inhabitants with Mediterranean roots
- The neighbourhoods with the best scores on all of these three criteria have already been identified
- However, to choose one of these the location within Toronto has to be taken into account
- If a neighbourhood is adjacent to other neighbourhoods that score high on the criteria discussed it is more suitable than if it is isolated
- For example, the number of inhabitants with Meditterranean roots is especially high in one neighbourhood (Downsvie).
- This neighbourhood should only be taken into consideration if it is close to neighbourhoods with a high number of Mediterranean restaurants or high average income

#### The Annex is the neighbourhood that offers the best combination of all important characteristics



- The Annex has the most restaurants, the most restaurants in its adjacent neighbourhoods and the second highest average income
- Little Portugal and Church and Wellesley are worse on number of restaurants and average income and have a smaller number of restaurants in their vicinity
- Bedford Park with higher average income ist too far away from the other interesting neighbourhoods
- Downsview is worse because of the low average income

Conclusion

#### Conclusion

- In this assignment neighbourhoods from Toronto were analyzed to identify the neighbourhood that is best suited to start a new retail store for excellent olive oil
- Taken into account
  - the Mediterranean roots of the inhabitants
  - the number of Mediterranean restaurants
  - the average income of the inhabitants
  - the location of the neighbourhoods

The Annex was chosen

 However, for the final decision on the location extra variables regarding the neighbourhood (rent, available store space, competitors, ...) have to be taken into consideration