The Battle of Neighborhoods

Applied Data Science Capstone

Introduction / Business Problem

Toronto is the biggest and most populated city in Canada. It is the capital of the province of Ontario and the country's financial center. It is considered a global economic and financial hub and is a high-income city in one of the highest income per capita countries in the world.

A good idea seems to be to open a restaurant in a specific area and neighborhood of the city, particularly in Scarborough which is a popular destination for new immigrants in Canada, and one of the most diverse and multicultural areas.

Who would be interested?

This project will be of special interest for anyone looking for an opportunity to open a new restaurant in the city of Toronto and who is aiming to obtain high levels of income and returns from his investment in doing so.

Data Description

1 In order to analyze Toronto's neighborhoods, we need specific data that we can obtain from Foursquare.com.

2 We need information about location and demographics that can help us understand which would be the neighborhoods that will increase the likelihood of success in opening a new restaurant.

3 We analyze population, income and second language information to identify highly populated neighborhoods with high mean income to increase the probability of success for a new restaurant.

4 We use the k-means clustering technique to map, segment and cluster the neighborhoods.

5 We will make a recommendation on opening a specific kind of restaurant in a specific neighborhood or group of neighborhoods

Join the best who trust in Foursquare

Methodology

1   Information about location and demographics of the neighborhoods in Scarborough which is one of the most extensive areas in Toronto.

2   We then, get information of population, income and 2nd language for every neighborhood in Scarborough.

3   Analysis of the information of these three variables in order to identify relevant ethnic groups by their 2nd language, while keeping an eye in those neighborhoods highly populated and with high income as well.

4   List of the different venues in the neighborhoods of Scarborough to get a glimpse of their location and their categories.

5   k-means clustering technique with the considered variables to find statistical similarities between the neighborhoods in Scarborough.

6   Results, gathered information and discussion on the main conclusions in order to make a recommendation.

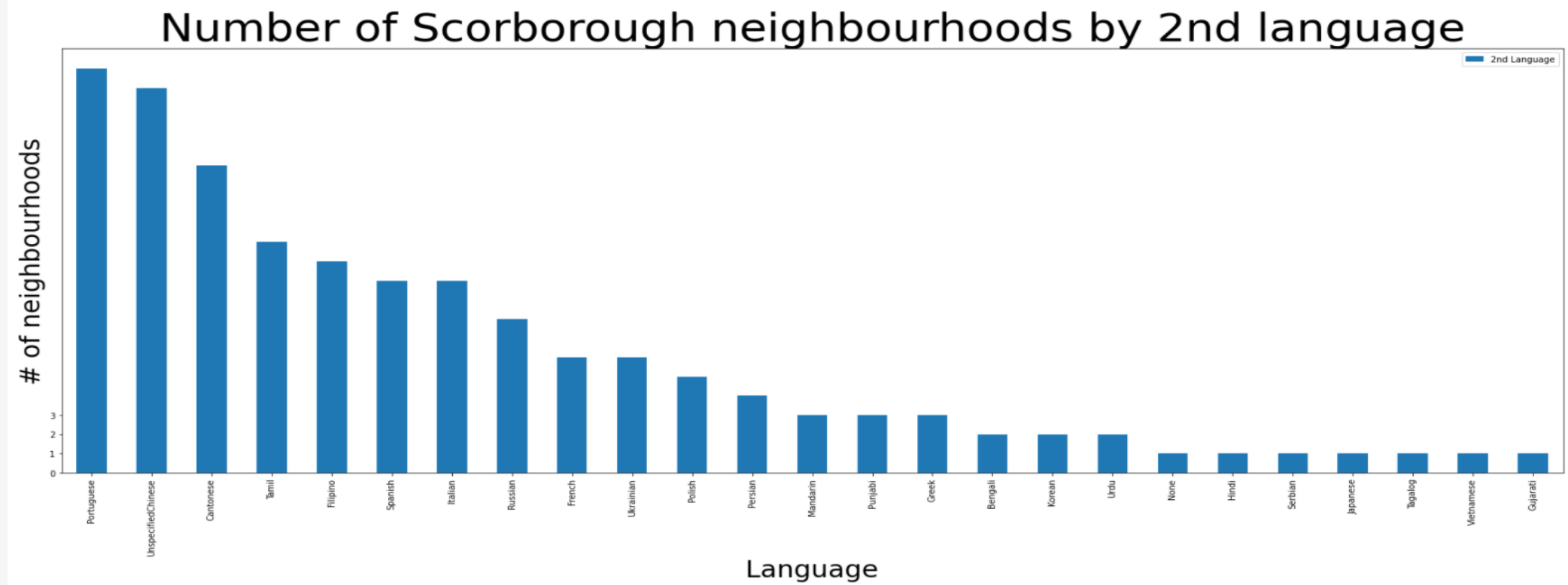# Results

The first interesting result is that the 2nd language of most neighborhoods in Scarborough, Toronto is Portuguese, followed by Chinese, Cantonese and Tamil.

1 Portuguese: Brazil, Portugal, some African countries.

2 Chinese: China, Taiwan, Singapur

3 Cantonese: China

4 Tamil: India, Sri Lanka, Singapur

Number of Scorborough neighbourhoods by 2nd language

Results

A relevant criteria of segmentation for clusters 1,2,3 and 4 is the 2nd language. On the other hand, we have cluster 0 that includes a wide variety of 2nd languages, but they are mostly different from those identified in the other clusters as follows:

Cluster 0:

	Cluster	Neighbourhood	Population	Income	2nd Language
3	0	Allenby	2513	245592	Russian
4	0	Amesbury	17318	27546	Spanish
5	0	Armour Heights	4384	116651	Russian
7	0	Bathurst Manor	14945	34169	Russian
8	0	Bay Street Corridor	4787	40598	Mandarin
...	...	...	...	...	...
146	0	Wilson Heights	13732	37978	Filipino
147	0	Woburn	48507	26190	Gujarati
149	0	York Mills	17564	92099	Korean
150	0	York University Heights	26140	24432	Italian
151	0	Yorkville	6045	105239	French

Cluster 1:

	Cluster	Neighbourhood	Population	Income	2nd Language
6	1	Banbury	6641	92319	UnspecifiedChinese
13	1	Birch Cliff	12266	48965	UnspecifiedChinese
19	1	Cabbagetown	11120	50398	UnspecifiedChinese
34	1	Discovery District	7262	41998	UnspecifiedChinese
36	1	Don Mills	21372	47515	UnspecifiedChinese
37	1	Don Valley Village	29740	30442	UnspecifiedChinese
51	1	Garden District	8240	37614	UnspecifiedChinese
53	1	Grange Park	9007	35277	UnspecifiedChinese
57	1	Harbourfront / CityPlace	14368	69232	UnspecifiedChinese
62	1	Hillcrest	18327	33465	UnspecifiedChinese
63	1	Hoggs Hollow	3123	222560	UnspecifiedChinese
71	1	Lambton	9654	30920	Portuguese
91	1	North York City Centre	10427	34330	UnspecifiedChinese
95	1	Parkway Forest	8498	24333	UnspecifiedChinese
96	1	Parkwoods	26533	34811	UnspecifiedChinese
98	1	Playter Estates	3968	44557	UnspecifiedChinese

Results

Cluster 2:

	Cluster	Neighbourhood	Population	Income	2nd Language
12	2	Bendale	28945	29723	Tamil
27	2	Cliffcrest	14531	38182	Tamil
38	2	Dorset Park	14189	26525	Tamil
44	2	Eglinton East	22387	25307	Tamil
61	2	Highland Creek	12853	33640	Tamil
82	2	Malvern	44324	25677	Tamil
87	2	Morningside	11472	27139	Tamil
109	2	Rouge	22724	29230	Tamil
112	2	Scarborough Junction	25780	25405	Tamil
113	2	Scarborough City Centre	16403	26756	Tamil
114	2	Scarborough Village	12796	24413	Tamil
139	2	West Hill	25632	27936	Tamil

Cluster 3:

	Cluster	Neighbourhood	Population	Income	2nd Language
18	3	Brockton	9039	27260	Portuguese
21	3	Carleton Village	6544	23301	Portuguese
24	3	Christie Pits	5124	30556	Portuguese
31	3	Davenport	8781	28335	Portuguese
39	3	Dovercourt Park	8497	28311	Portuguese
41	3	Dufferin Grove	9875	27961	Portuguese
42	3	Earlscourt	17240	26672	Portuguese
46	3	Fashion District	4642	63282	Portuguese
56	3	Harbord Village	5906	45792	Portuguese
58	3	Harwood	3375	22136	Portuguese
69	3	Junction Triangle	6666	28067	Portuguese
72	3	L'Amoreaux	45862	26375	UnspecifiedChinese
79	3	Little Italy	7917	31231	Portuguese
80	3	Little Portugal	5013	29224	Portuguese
90	3	Niagara	6524	44611	Portuguese

Results

Cluster 4:

	Cluster	Neighbourhood	Population	Income	2nd Language
1	4	Agincourt	44577	25750	Cantonese
2	4	Alexandra Park	4355	19687	Cantonese
9	4	Bayview Village	12280	46752	Cantonese
10	4	Bayview Woods – Steeles	13298	41485	Cantonese
17	4	Bridle Path	1540	314107	Cantonese
28	4	Cliffside	9386	32701	Cantonese
43	4	East Danforth	21440	33847	Cantonese
70	4	Kensington Market	3740	23335	Cantonese
78	4	Leslieville	23567	30886	Cantonese
85	4	Milliken	26272	25243	Cantonese
105	4	Riverdale	31007	40139	Cantonese
120	4	Steeles	24696	26660	Cantonese
126	4	The Beaches	20416	67536	Cantonese
135	4	Upper Beaches	19830	44346	Cantonese
140	4	West Rouge	9300	44605	Cantonese

Cluster	Income
0	$53.836
1	$61.929
2	$28.327
3	$33.122
4	$53.560

Cluster 1 (Chinese) and cluster 4 (Cantonese) the clusters with highest mean income (not considering cluster 0 because of its 2nd language diversity) and also highly populated, and followed by cluster 3 (Portuguese) and cluster 2 (Tamil).

Discussion and Observations

1 There is a huge number of Portuguese neighborhoods in Scarborough, followed by Chinese and then South Asian communities represented in those neighborhoods with Tamil as 2nd language

2 The venue categories are very diverse, with no big concentration in any of the identified categories,

3 Most of the venue categories in Scarborough are food offer related which will be useful when considering locating supply chain units for this industry in Scarborough.

4 a natural and effective segmentation criterion for the different neighborhoods in Scarborough, Toronto is the 2nd language.

5 Chinese and Cantonese clusters are the clusters with highest mean income posing a big opportunity for anyone interested in opening a high-end type of restaurant

Conclusions

1 High demand for Brazilian and Portuguese food restaurants.

Opportunity to open mid to lower end restaurants in neighborhoods belonging to Cluster 3.

2 Similar opportunity for Indian and South Asian food restaurants in neighborhoods belonging to Cluster 2.

3 Huge opportunity for Chinese food restaurants in neighborhoods belonging to Clusters 1 and 4, in high-end or luxury restaurants.

3 Huge opportunities to locate supply chain facilities for almost every kind of food, but specially for the aforementioned such as Chinese, Brazilian, Portuguese, Indian and South Asian.