

Coursera Applied Data Science Capstone

Recommendations of locations for a gourmet coffee shop business in Toronto

Catherine Tam

April 25, 2019

Introduction and Business Problem

A client needs some recommendations on where to open her business serving gourmet tea, coffee and desserts in the Greater Toronto Area (GTA).

- Given the exquisite quality of food and drinks served and to cover the costs, she is looking for an affluent neighbourhood where people are willing to splurge.
- The population targeted is also the younger age group under 40 years old who tends to value such lifestyle and experience.
- The client wants to avoid potential competitions in the same area.

Data

- Average Income
 - Sourced from the Canada Revenue Agency Total Income dataset based on Total Income of the tax filers classified by FSA as of December 31, 2015.
- Population under 40
 - Sourced from the Canada Revenue Agency Total Income dataset based on age of the tax filers classified by FSA as of December 31, 2015.
- Neighbourhood names
 - Sourced from a Wikipedia page on Toronto
- Geospatial data
 - Sourced from a csv file posted on http://cocl.us/Geospatial_data
- Existing competitions
 - Sourced from performing a search call on venues to the Foursquare API

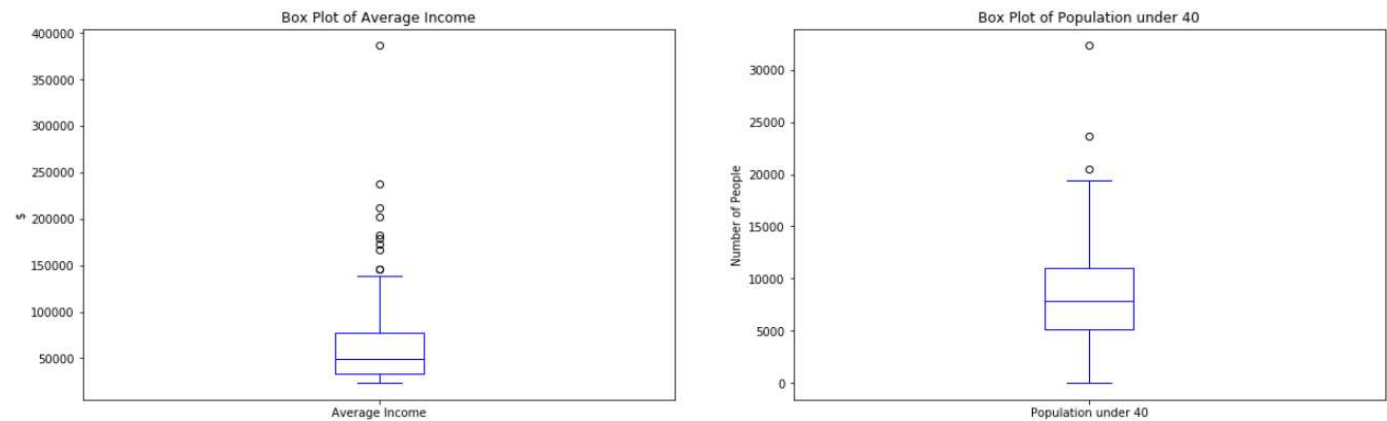
Data

- Dataframe:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude	Average Income	Population under 40	Coffee Shop	Chocolate Shop	Dessert Shop	Café	Number of Coffee Venues
42	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576	172630.555556	150.0	12.0	0.0	0.0	8.0	20.0
48	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	237900.000000	10.0	13.0	0.0	0.0	7.0	20.0
36	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	94132.249071	5810.0	13.0	0.0	0.0	4.0	17.0
24	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	26468.434604	5720.0	13.0	0.0	1.0	3.0	17.0
97	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	386127.272727	20.0	8.0	0.0	0.0	7.0	15.0

Methodology

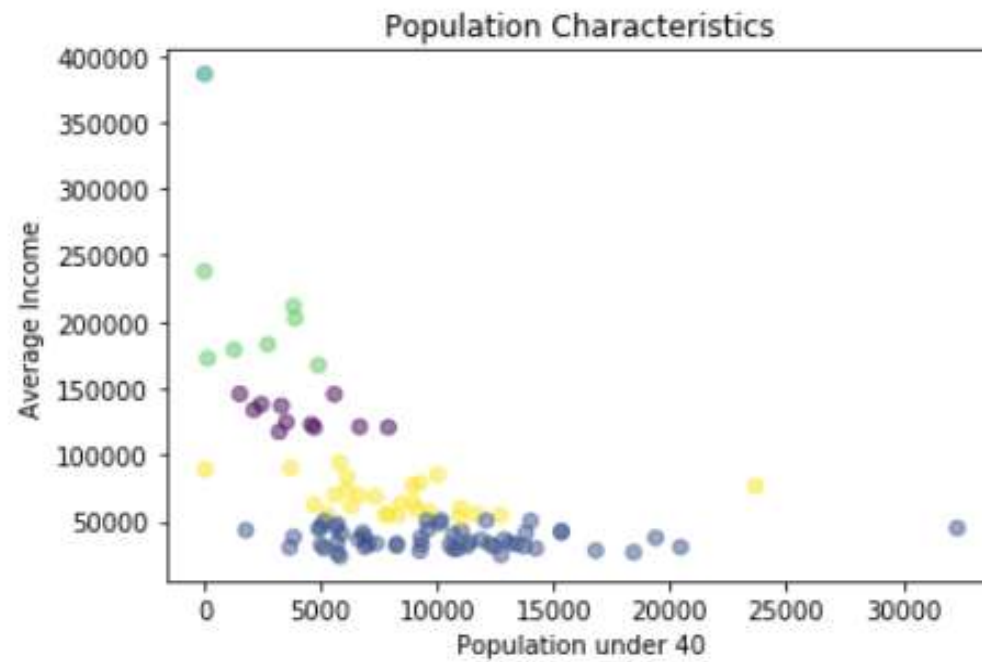
- Box plots of Average Income and Population under 40



- K-means was used for clustering, giving insights on the unlabeled neighbourhood data. The number of clusters was set to 5 for fitting.

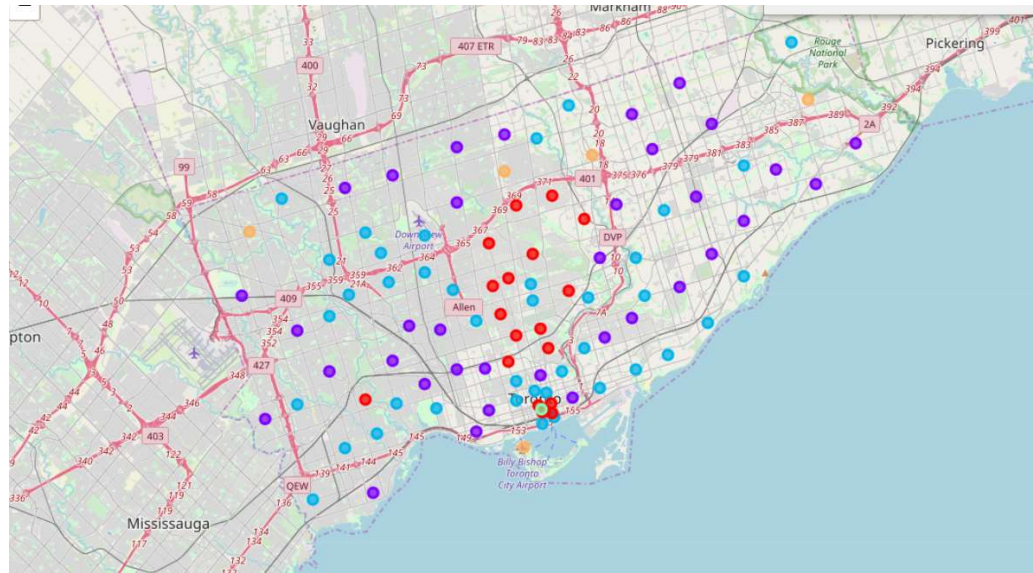
Results

- Results from K-means return the following clustering *in colour*.



Results and Discussions

- Visualization of Results on a map:



- Since neighbourhoods with higher average income are targeted for the business, neighbourhoods in Cluster 1 (shown in red above) would likely represent some good candidates for considerations.

Results and Discussions (cont'd)

- Cluster 1 characteristics:

Postal Code	Borough	Neighbourhood	Latitude	Longitude	Average Income	Population under 40	Coffee Shop	Chocolate Shop	Dessert Shop	Café	Number of Coffee Venues
M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.64718	-79.38158	\$ 172,630.56	150	13	0	0	8	21
M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.6482	-79.37982	\$ 237,900.00	10	13	0	0	7	20
M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.64644	-79.37485	\$ 89,029.41	40	11	0	0	4	15
M5C	Downtown Toronto	St. James Town	43.65149	-79.37542	\$ 145,786.91	1,530	8	0	0	5	13
M5H	Downtown Toronto	Adelaide, King, Richmond	43.65057	-79.38457	\$ 178,935.89	1,280	6	0	0	5	11
M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.67271	-79.40568	\$ 120,758.48	7,910	3	0	0	3	6
M4G	East York	Leaside	43.70906	-79.36345	\$ 122,920.00	4,600	3	0	1	0	4
M5M	North York	Bedford Park, Lawrence Manor East	43.73328	-79.41975	\$ 121,093.92	6,680	2	0	0	1	3
M4R	Central Toronto	North Toronto West	43.71538	-79.40568	\$ 136,834.64	3,320	2	0	1	0	3
M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	43.68641	-79.40005	\$ 167,263.31	4,910	2	0	0	0	2
M3B	North York	Don Mills North	43.74591	-79.35219	\$ 124,592.32	3,530	0	0	0	1	1
M8X	EtoBicoke	The Kingsway, Montgomery Road, Old Mill North	43.65365	-79.50694	\$ 137,988.81	2,450	0	0	0	0	0
M4W	Downtown Toronto	Rosedale	43.67956	-79.37753	\$ 202,622.07	3,910	0	0	0	0	0
M4T	Central Toronto	Moore Park, Summerhill East	43.68957	-79.38316	\$ 183,044.70	2,730	0	0	0	0	0
M5P	Central Toronto	Forest Hill North, Forest Hill West	43.69695	-79.41131	\$ 145,550.40	5,590	0	0	0	0	0
M2L	North York	Silver Hills, York Mills	43.75749	-79.37471	\$ 117,362.20	3,220	0	0	0	0	0
M2P	North York	York Mills West	43.75276	-79.40005	\$ 133,867.64	2,130	0	0	0	0	0
M4N	Central Toronto	Lawrence Park	43.72802	-79.38879	\$ 211,828.79	3,840	0	0	0	0	0
M5N	Central Toronto	Roselawn	43.71169	-79.41694	\$ 120,636.36	4,730	0	0	0	0	0

- High average income
- Decent population under 40
- However, a few neighbourhoods, especially those in the Downtown Toronto area, seem to have a number of existing coffee-related venues already, pointing to potential competitions. The client may want to avoid these neighbourhoods.

Conclusion

Using a combination of statistics and machine learning tools, we are able to address the client's problem and arrive at a list of suggested neighbourhoods that are suitable location candidates for her future gourmet coffee shop business.