Clustering Toronto's neighborhoods to find out where to open a restaurant.

From: Korbinian Nitz

Table of Contents
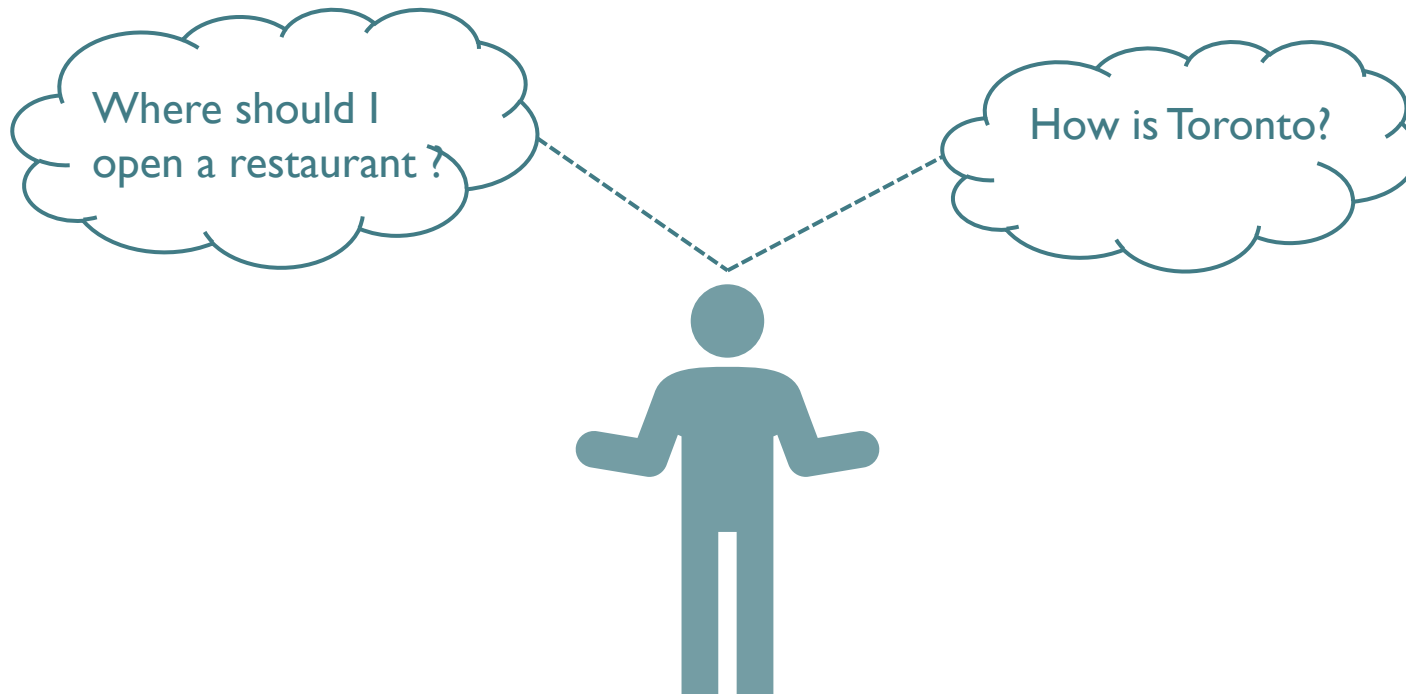
1. Introduction / Business Problem

- Toronto is one of the most trendy and famous cities in Canada.
- A big opportunity for every entreprenuer.
- But which neighborhoods suits my business idea e.g. opening a new restaurant.

→ Lets find out by clustering the neighborhoods based on their similarities.

Where should I open a restaurant ?

How is Toronto?

2. Data Aquisition And Cleaning (I)
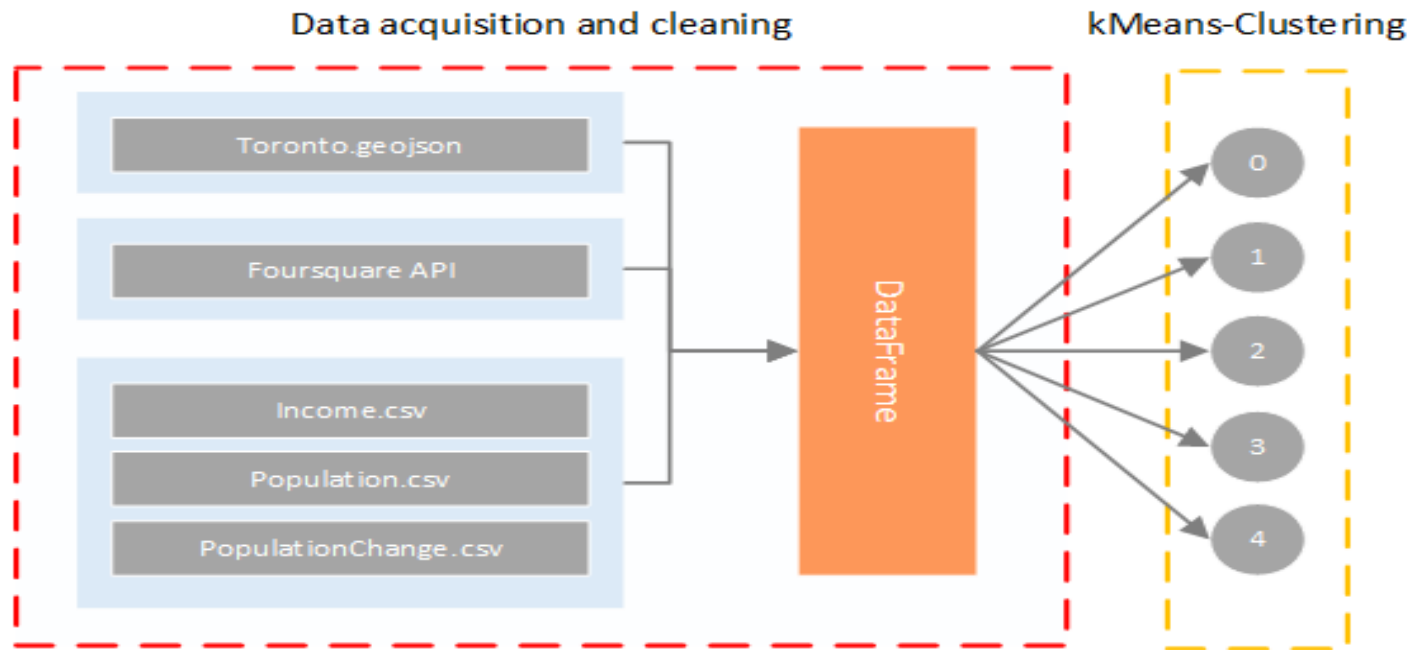
- Data acquisition to get informations about income, population and change in population to determine trendy, rich with a lot of restaurants.

- Use Toronto.geojson to get the coordinates of each neighborhood

- Link them with the Foursquare API to count the restaurants in each neighborhood and transform them into a final dataset for clustering

2. Data Aquisition And Cleaning (II)

- Final dataset for further observations including the demography, income and amount of restaurants.
- Table 1 shows an example of 4 neighborhoods in Toronto.

Neighborhood	Young [%]	Middle Age [%]	Old [%]	Total Population	Change in Population [%]	Income / Household [$]	Restauran ts
West Humber-Clarville	45.86	33.55	20.59	32880	-2.32	63977	2
Mount Olive-Silverstone-Jamestown	51.86	32.80	15.38	33090	0.51	49601	14
Thistletown-Beaumond Heights	43.14	33.41	23.45	10235	2.2	54910	19
Rexadale-Kipling	41.30	35.60	23.09	10350	0.39	53779	4

3. Exploratory Data Analysis (I)

- First evaluate the data for whole Toronto to see how Toronto is like.
- Binning demography into 3 groups → mostly young and middle age people.

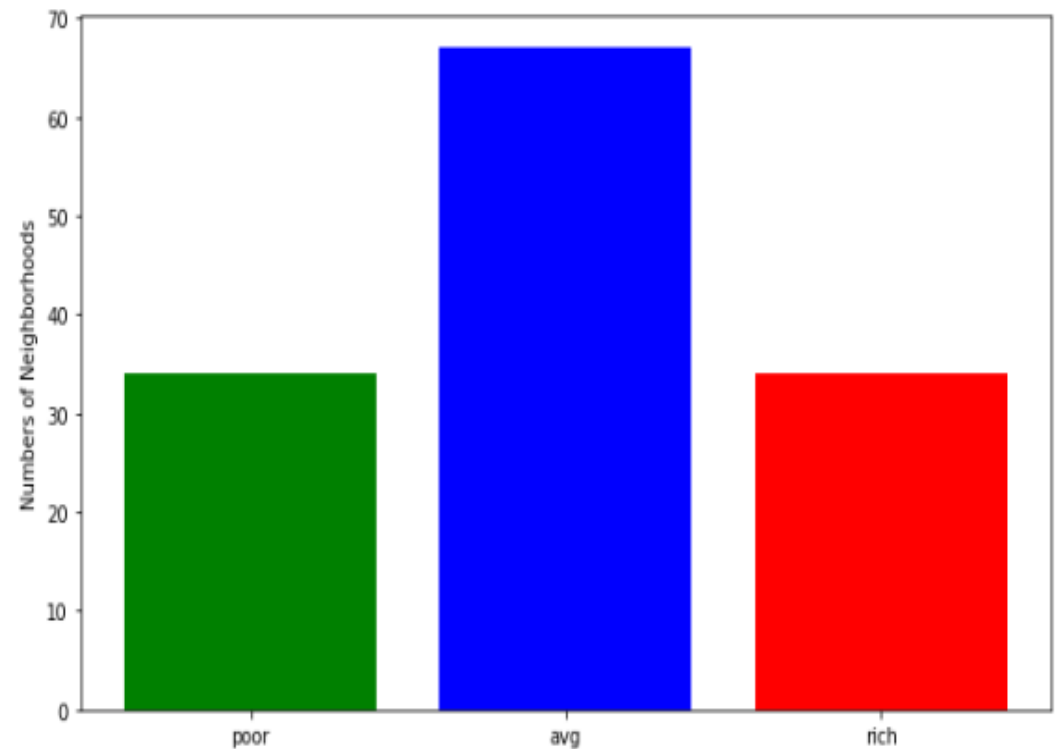| Category	Condition
Young	$< 30\ years$
Middle Age	$30\ years \leq x$ $< 60\ years$
Old	$x \geq 60\ years$
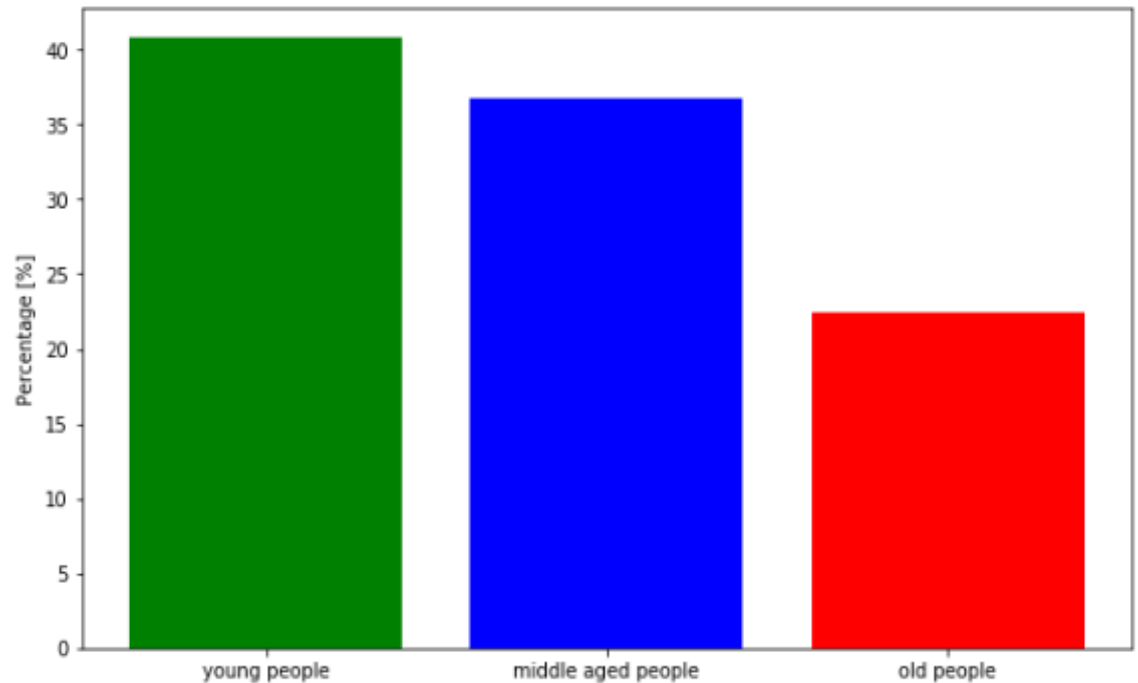
3. Exploratory Data Analysis (II)

- Also interesting to see the amount of neighborhoods which are poor, avg or rich.

- Mostly avg income.

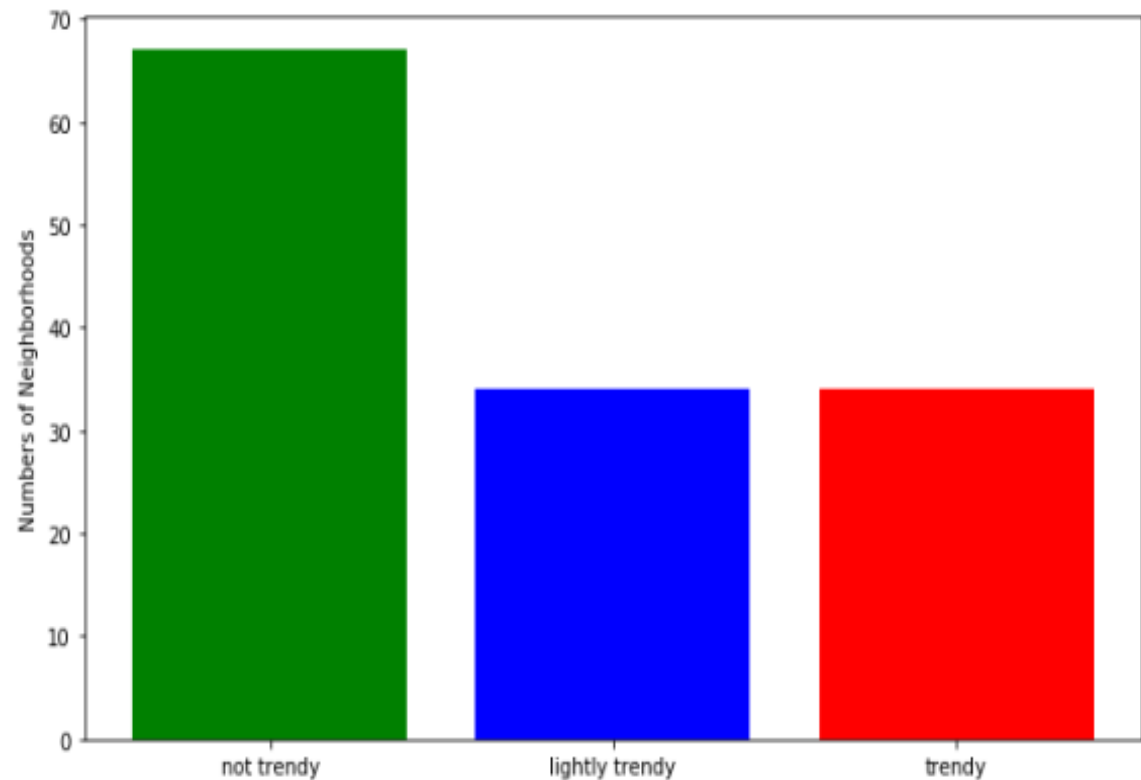| Category	Condition
Poor	$< 48921\ \$$
Avg	$48921\ \$ \leq x$ $< 64768\ \$$
Rich	$x \geq 64768\ \$$

3. Exploratory Data Analysis (III)

- Evaluating the change in population within the last 4 years to find out trendy areas.

- Mostly not trendy areas

Category	Condition
Not trendy	$< 1.4\,\%$
Lightly trendy	$1.4\,\% \leq x$ $< 4\,\%$
Trendy	$x \geq 4\%$

3. Exploratory Data Analysis (IV)

- Connecting to Foursquare and count the amount of restaurants in each neighborhood to find out which neighborhood are known for a lot of restaurants by using following conditions.

Category	Condition
Less	< 4
Medium	$4 \leq x < 19$
A lot	$x \geq 19$

4. Cluster Modeling (I)

- First transform the values into dummy variables so we can use the kMean cluster algorithm.

- By using the elbow method we find out that 5 clusters are the best.

- Goal: cluster neighborhoods into cluster based on their similarities so the entrepreuner can find easily pattern and get an overview where to open a restaurant.

4. Cluster Modeling (II)
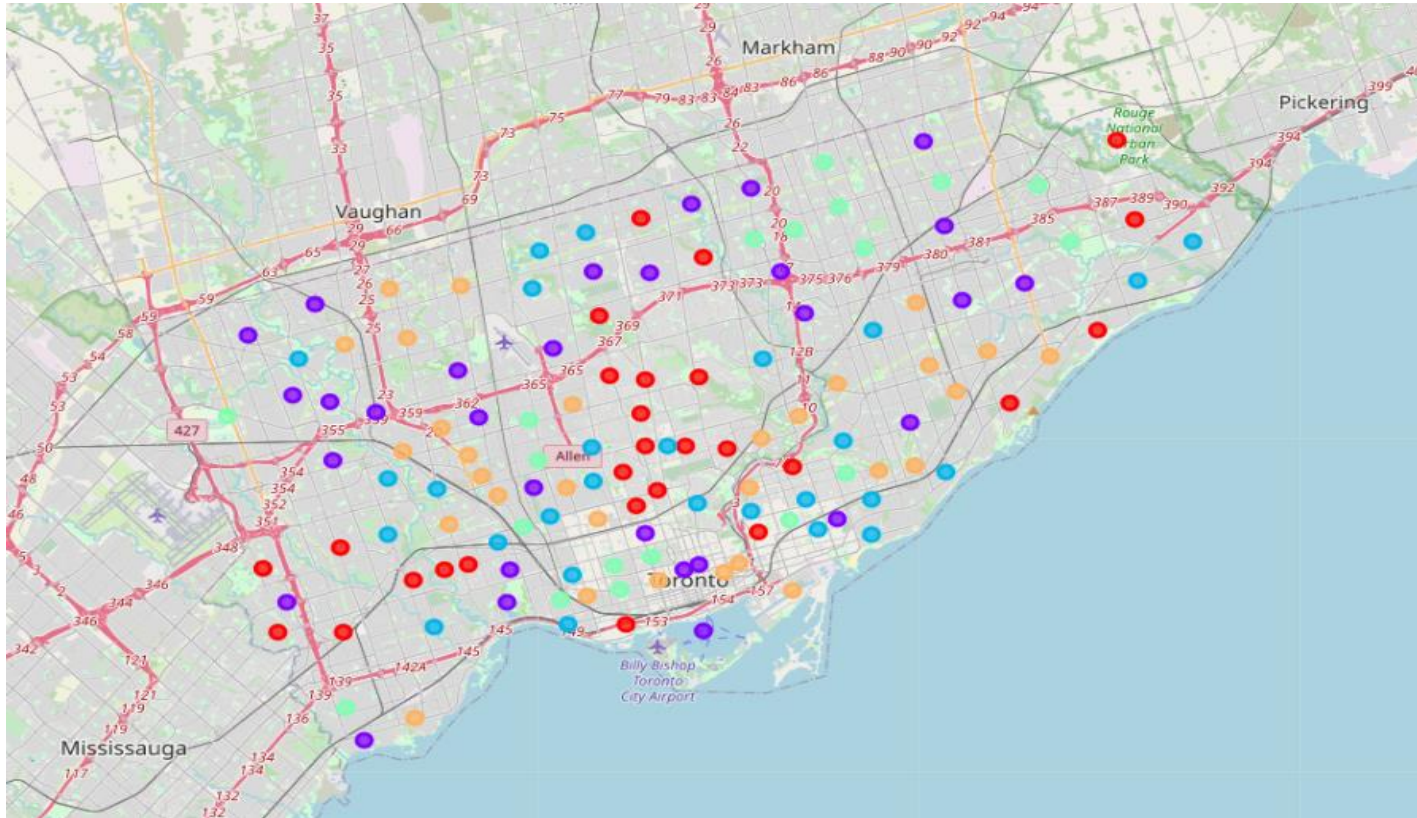
Map of Toronto with marked neighborhood depending of the cluster (red = Cluster 0, purple = Cluster 1, blue = Cluster 2, light green = Cluster 3, orange = Cluster 4.

4. Cluster Modeling (III)

- Amount of Neighborhoods in each category.

Category	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Poor	0	7	18	0	9
Average	0	17	0	21	18
Rich	27	10	0	0	3
Not trendy	16	0	18	21	0
Lightly trendy	1	0	0	0	30
Trendy	10	34	0	0	0
Less venues	10	11	7	0	10
Medium venues	10	12	10	16	9
A lot venues	7	11	1	5	11

5. Conclusions

- Cluster 0 shows trendy areas where only rich people live and already have a lot of restaurants which means a big opportunity for opening a restaurants. As it looks people in that cluster likes to go to restaurant.

- Cluster 1 are super trendy areas with a good mix of all different population types and a lot of restaurants. Here we also have a big opportunity to open a restaurant.

- Cluster 4 is maybe also interesting since all neighborhoods are lightly trendy and could become a new hotspot.

- Neverless, cluster 2 and 3 are not recommended for opening a restaurant since only poor people, not trendy areas and less restaurants are based in. Neverless, still maybe a good idea for a cheap pizza place.

- All in all, we could find neighborhoods based on their similarity which makes it easier for the entreprenuer to find neighborhoods which suits his business idea.