

Determining the best location to open a vegan restaurant in Toronto based on K-Means clustering

Liam Chalcroft

July 8, 2020

Contents

1	Introduction	4
2	Data and Methods	5
3	Results	7
4	Conclusions	11

List of Figures

1	Clustered data on Folium map for all neighborhoods in Toronto. Clustering performed via k-means clustering with k=20.	7
2	Clustered data on Folium map for all Toronto neighborhoods containing vegan restaurants in their top-20 venues. Clustering performed via k-means with k=20.	8
3	Labels for cluster markers in Figure 2.	8
4	Chloropleth map for home prices in Toronto. Data sourced from Open Toronto Initiative.	9
5	Chloropleth map with overlaid cluster markers.	10
6	Chloropleth map with overlaid clusters for postal codes containing vegan restaurants.	10

1 Introduction

The vegan food industry has been on a steady rise for many years now, and is increasingly becoming a viable expansion for established businesses, as seen by recent vegan additions to the menu at chains such as KFC and McDonald's. The business in question here would like to expand their business by trialling a new vegan chain of their existing franchise, located in Toronto. For this they need to determine which location (specific to the postal code) is most likely to guarantee economic success.

2 Data and Methods

In order to identify the optimal location for the business to open their new vegan chain, data from Foursquare will be gathered on all vegan/vegetarian restaurants and collected into the individual postal codes. The viable location will be determined by finding the ratings of all current vegan restaurants, and identifying areas with a high density of popular restaurants where success is more likely.

Data can be scraped from sites such as Wikipedia on the co-ordinate locations of each Postal Code, and by performing a search on Foursquare for each Postal Code via the Foursquare API it will be possible to identify all vegan/vegetarian restaurants and their ratings for each respective Postal Code.

Initial scraping revealed that only 7 venues in Toronto are categorised as "Vegetarian / Vegan Restaurant", the only category that includes the words 'Vegan' or 'vegan'. To expand the search and find potential other locations, additional categories are included that may be deemed to appeal to a similar demographic as the vegan restaurants. Several categories are expected to appeal to a similar demographic:

- Performing Arts Venue
- Yoga Studio
- Arts and Crafts Store
- Art Gallery
- Salad Place
- Record Shop
- Gluten-free Restaurant
- Gourmet Shop
- College Arts Building
- Indie Movie Theater
- Organic Grocery
- Supplement Shop

This will be validated by performing k-means clustering to evaluate clusters already located in the city of Toronto, and which of these is likely to have large pre-existing footfall from the target demographic due to its proximity to venues of these categories. To choose between these clusters, the rent prices can then be taken into consideration to evaluate economic viability.

Data was sourced from both resources provided by the course instructors, and also from the [Toronto Open Data initiative](#). For the Foursquare API search, Toronto postal codes were obtained via [Wikipedia](#) and neighborhood's co-ordinates obtained using the [GEOJson file provided by Cognitive Class](#). For house pricing analysis, prices for neighborhoods were obtained via a [CSV on the Open Toronto Database](#) and locations of areas determined using a [GEOJson file provided on the same database](#).

After performing a search on the Foursquare API, neighborhoods' top 20 most popular venues were recorded - it is likely that Vegan/Vegetarian restaurants will not be the majority in any region so a wider search is required for regions containing these restaurants. From this search a total of 7 neighborhoods were located.

Further to this, k-means clustering was performed with a high k-value of k=20, in order to discriminate between neighborhoods to a greater degree.

3 Results

K-means clustering was performed on Toronto postal codes using a high k-value of 20. The resulting clusters are visualised in a Folium map as seen in Figure 1. From this we can see that the two dominant clusters are 1 (purple) and 8 (light blue).

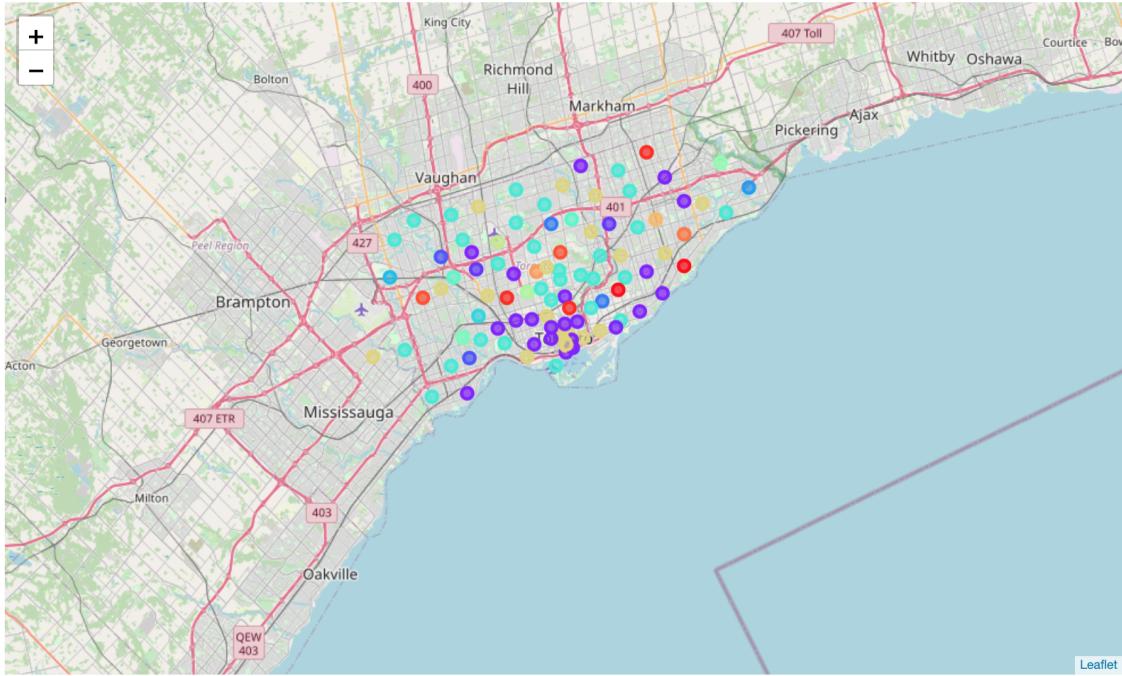


Figure 1. Clustered data on Folium map for all neighborhoods in Toronto. Clustering performed via k-means clustering with $k=20$.

Filtering the data to those relevant to the business' aims is done by only including those that have a vegetarian/vegan restaurant in their top 20 venues. This is plotted on a Folium map in Figure 2 and reveals that five of the seven areas of interest are located in downtown Toronto, with two others further out from the city centre. Interestingly, the majority of these seven locations are classified under two clusters of 1 (purple) and 14 (yellow) as shown in Figure 3.

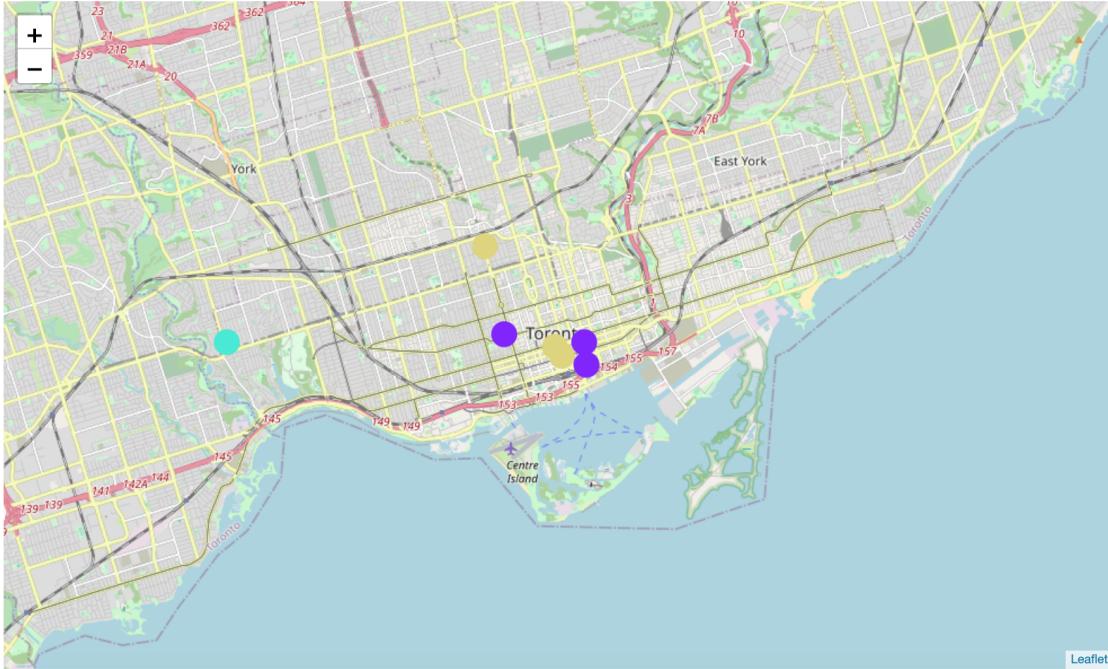


Figure 2. Clustered data on Folium map for all Toronto neighborhoods containing vegan restaurants in their top-20 venues. Clustering performed via k-means with k=20.

Cluster	Colour	Frequency
1.0	Purple	3
14.0	Yellow	3
8.0	Blue	1

Figure 3. Labels for cluster markers in Figure 2.

This indicates that locations under clusters 1 and 14 are most likely to already attract the demographic that a vegan restaurant would appeal greatest to. It is possible that by opening the restaurant in a location that already has similar restaurants could lead to competing for customers, and so it may be better to locate other postal codes that fit these two clusters.

Another factor that is critical for maximising the business' success is the cost of maintenance. A simple way to gauge this is through the average house prices. The average prices for homes in Toronto were obtained via the Open Toronto Initiative, and visualised using a chloropleth in Folium as seen in Figure 4.

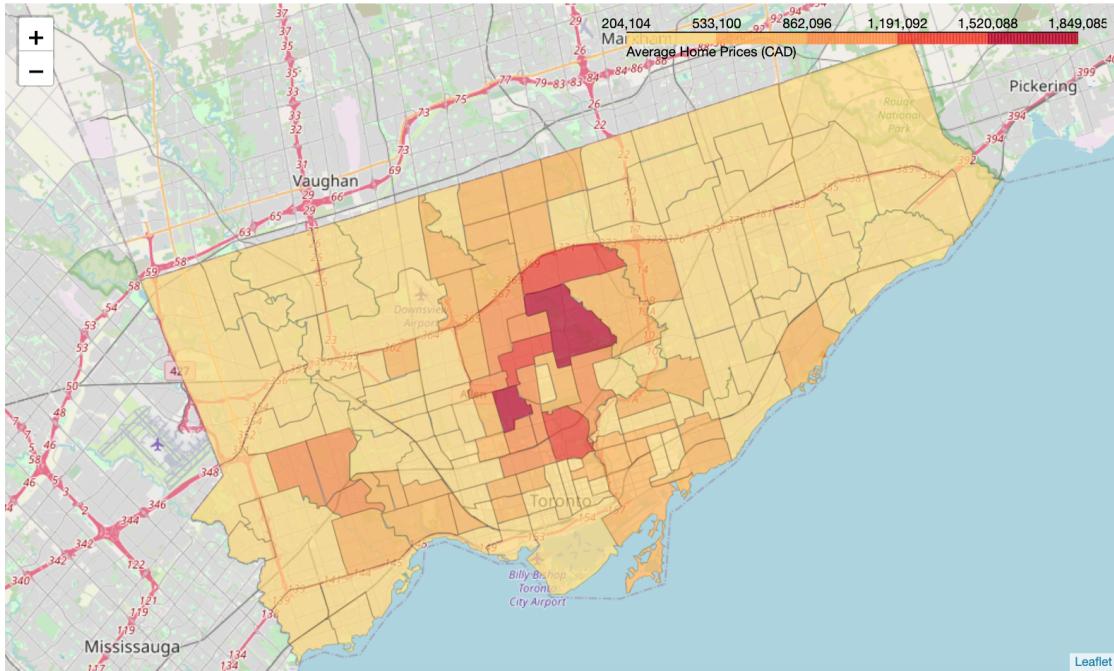
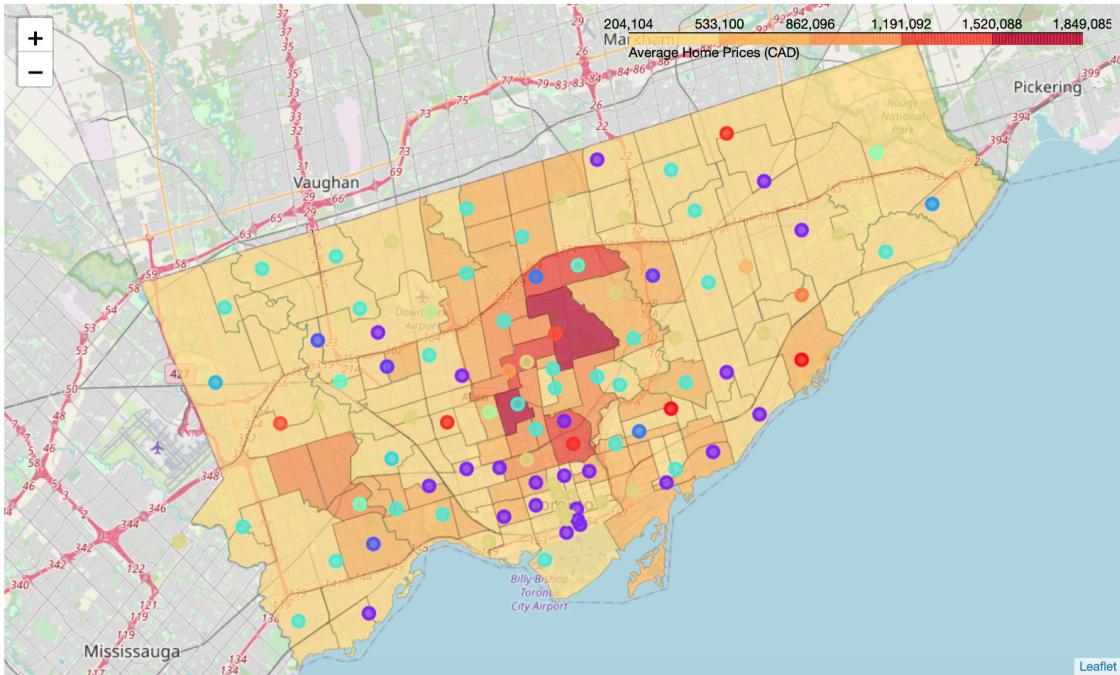


Figure 4. Chloropleth map for home prices in Toronto. Data sourced from Open Toronto Initiative.

The cluster markers are then overlaid onto the chloropleth to compare the two better. A view of all areas is shown in Figure 5 and for the areas containing vegan restaurants already another graph is shown in Figure 6. These figures indicate that the majority of cheaper properties are located in clusters 1 (purple) and 8 (blue). Purple is joint as the most popular location for vegan restaurants currently so cluster 1 appears to be the most appropriate choice for opening a new restaurant. There is a possibility that the prices are skewed for downtown properties, due to many of these being smaller properties. From cluster 1, a number of viable postal codes can be identified that come under the low-price shading in chloropleths. These include:

- M8V
- M6H
- M3M
- M1L
- M1N
- M1H
- M1S
- M5J
- M5E
- M4Y



4 Conclusions

From the data analysis detailed in the above section, it has been shown that there are a number of postal codes with affordable property prices that fit the cluster most popular currently with restaurants within the vegan/vegetarian category, as determined by a k-means clustering algorithm. This list will provide a basis for areas that the client can explore further to determine the most appropriate of these based on criteria such as expected footfall, proximity to suppliers and cost of maintenance bills.