

The Best Location for a Vegetarian Restaurant in Toronto

1. Introduction

1.1. Background Information

Canadians are going meatless. In September 2018, Charlebois, Somogyi, and Music from Dalhousie University conducted a study, which consisted of an online survey administered to Canadian consumers to determine their attachment to meat consumption and willingness to adopt a plant base diet. The survey was carried out in 3 days and they gathered 1027 samples. Based on this study, Charlebois, Somogyi, and Music estimated that over 6.4 million Canadians prefer to either reduce or eliminate meat consumption. The reasons for Canadians going meatless were found to be related to Health, Animal Welfare, the Environment and Taste.

The study revealed that the typical vegetarian or “to-be-vegetarian” is a woman, who is less than 38 years old, is highly educated, and her income is more than \$75,000. Even though, health benefits gained by being a vegetarian seem to be important for both genders, women appear to be more concerned about animal welfare than men. Men, regardless of their age, consider eating meat as a great pleasure in life. Moreover, the less educated men love meat in general.

Here is the link for the [Dalhousie study](#) for further information.

1.2. Problem

As Canadians are going meatless, the new restaurants are needed that provide vegetarian meals. Toronto is a very diverse city and full of restaurants, café shops, and fast food places. However, vegetarian restaurants are hard to find, and the goal of this study of “The Best Location for Vegetarian Restaurant in Toronto” is to answer the question: what is the best neighbourhood to open a new vegetarian restaurant? This study uses the profile of the typical vegetarian provided by the Dalhousie study as a base line. First, this study tries to find out

the neighbourhoods, where the typical vegetarians live. Once the neighbourhoods have been found, the next step is to find out the existing venues in those neighbourhoods and determine the potential neighbourhood for establishing a vegetarian restaurant.

1.3. Interest

Establishing a new vegetarian restaurant or even many of them, is a great business opportunity for an individual(s), who has a spirit of an entrepreneur and may even have the same concerns (e.g. health, animal welfare, environment, taste) as the consumers in the Dalhousie study. Therefore, the end result of this study is an enormous help when figuring out what's the best neighbourhood for a vegetarian restaurant is.

2. Data Acquisition and cleaning

2.1. Data Sources

In addition to the information of the Dalhousie study, this project uses two datasets to find out where the potential vegetarians live. City of Toronto uses the census data of population in Canada to compile their own datasets to help government and community organizations with local planning. These datasets are open for everyone and can be found at City of Toronto's [Open Data portal](#). This project utilized two of those datasets:

1. Neighbourhoods

The City of Toronto has divided Toronto in 140 social planning neighbourhoods and they can be found in the [Neighbourhoods](#) dataset.

2. Neighbourhood Profiles

City of Toronto has compiled the demographic, social and economic characteristics of the people and households in each City of Toronto neighbourhood of those 140 neighbourhoods and they can be found at the [Neighbourhood Profiles](#) dataset.

Also, the Foursquare was utilized to pull the list of venues near those neighbourhoods.

2.2. Feature Selection

The Neighbourhoods dataset includes 16 features and 140 rows, one row for each neighbourhood. The features needed from this dataset were AREA_SHORT_CODE, AREA_NAME, LONGITUDE, and LATITUDE. The AREA_SHORT_CODE was used for matching neighbourhoods' data with neighbourhoods' profile data. The AREA_NAME, LONGITUDE, and LATITUDE were needed in order to the Foursquare to pull the list of venues near the neighborhoods, mainly for clustering purposes.

The Neighbourhood Profiles dataset includes 163 features and 2383 rows. In this dataset, the neighborhood names are as columns. The first row has "Neighbourhood Number" which was used when matching data with the neighbourhoods' data as it is the same as the AREA_SHORT_CODE in the neighbourhoods' data. This study was interested in data related to the number of populations that matches with the vegetarian profile in Dalhousie study (e. g. gender, age, income, education). The data, which needed to be extracted are stored under the features of CATEGORY, TOPIC, DATA SOURCE, CHARACTERICS, and all those 140 neighbourhoods, which each has an own column. The purpose is to find out in which neighbourhoods the potential vegetarians and future customers live.

2.3. Data cleaning

The end result was a combined dataset, which only had the information needed. The Neighbourhoods dataset didn't have any missing values or inaccurate data. From this dataset, the unnecessary columns were dropped. The Neighbourhood Profiles dataset required more work. It had lots of unnecessary data, which was not needed in this study. Several columns were dropped and only the rows that were necessary in order to find out vegetarian profiles and their locations were extracted. The dataset had missing data, which were filled in with zeros. Also, most of the datatypes were converted from object to int64 to

allow calculations. Finally, the cleaned dataset was transposed and combined with the Neighbourhoods dataset.

The final dataset looked like this below:

| Feature Name | Description |
|-----------------------------|---|
| AREA_SHORT_CODE | Neighborhood identification number |
| Neighborhood | Name of the neighborhood |
| Longitude | Geographical coordinates which specifies east-west position of the neighborhood on earth surface. |
| Latitude | Geographical coordinates which specifies north-south position of the neighborhood on earth surface. |
| Female_Total | Total number of 25 to 39 years old females in a neighborhood. |
| 70K+_Total | Total number of individuals, who earns \$70, 0000 in a year or more. |
| High_Education_Total | Total number of highly educated individuals in a neighborhood. |

3. Methodology

3.1. Exploratory Data Analysis

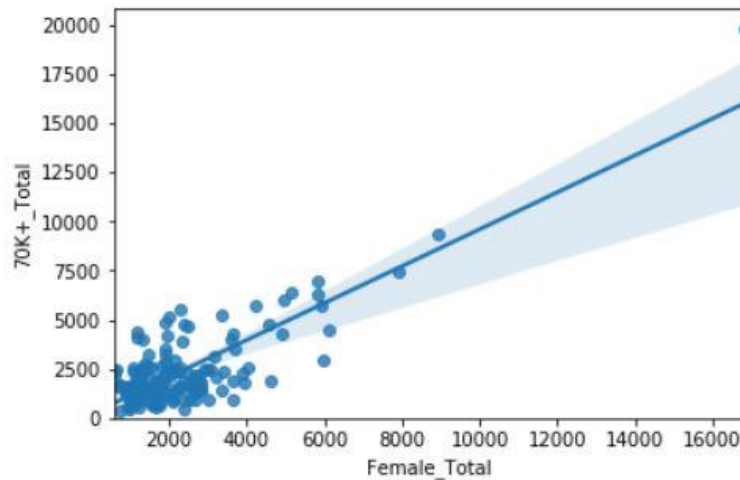
3.1.1. Analyzing Individual Feature Patterns using Visualization

To verify the representation of the desired subjects related to profile, some relationships were investigated using `corr()` – function and plotting with `regplot` from `seaborn` library.

3.1.1.1. Relationship Between Female and Income

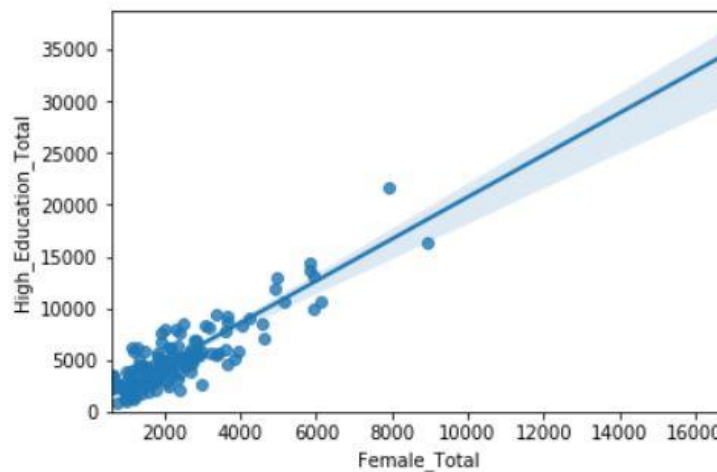
The correlation between female and income was 0.81 and the correlation was positive. It means that the more females in the neighborhood, the

more individuals who earn more than \$70,000 per year live in the neighborhood.



3.1.1.2. Relationship Between Female and Education

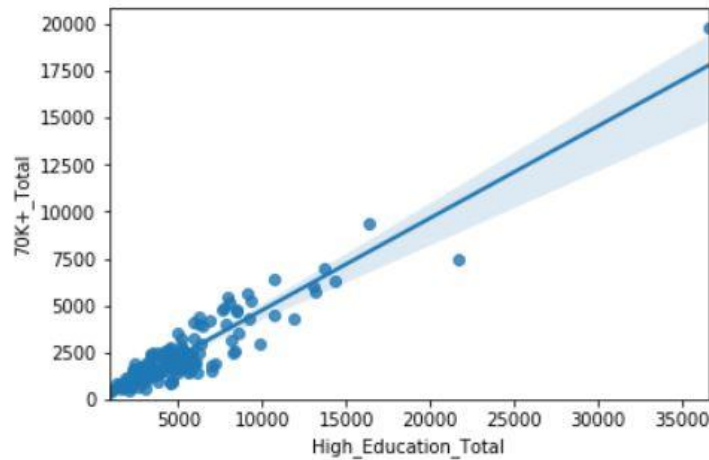
The correlation between female and income was 0.93 and the correlation was positive. It means that the more females in the neighborhood, the more individuals who have high education live in the neighborhood.



3.1.1.3. Relationship Between Education and Income

The correlation between education and income was also 0.93 and the correlation was positive. It means that the more highly educated

individuals live in the neighborhood, the more individuals who earn more than \$70,000 in a year live in the neighborhood.



3.1.2. Descriptive statistical analysis

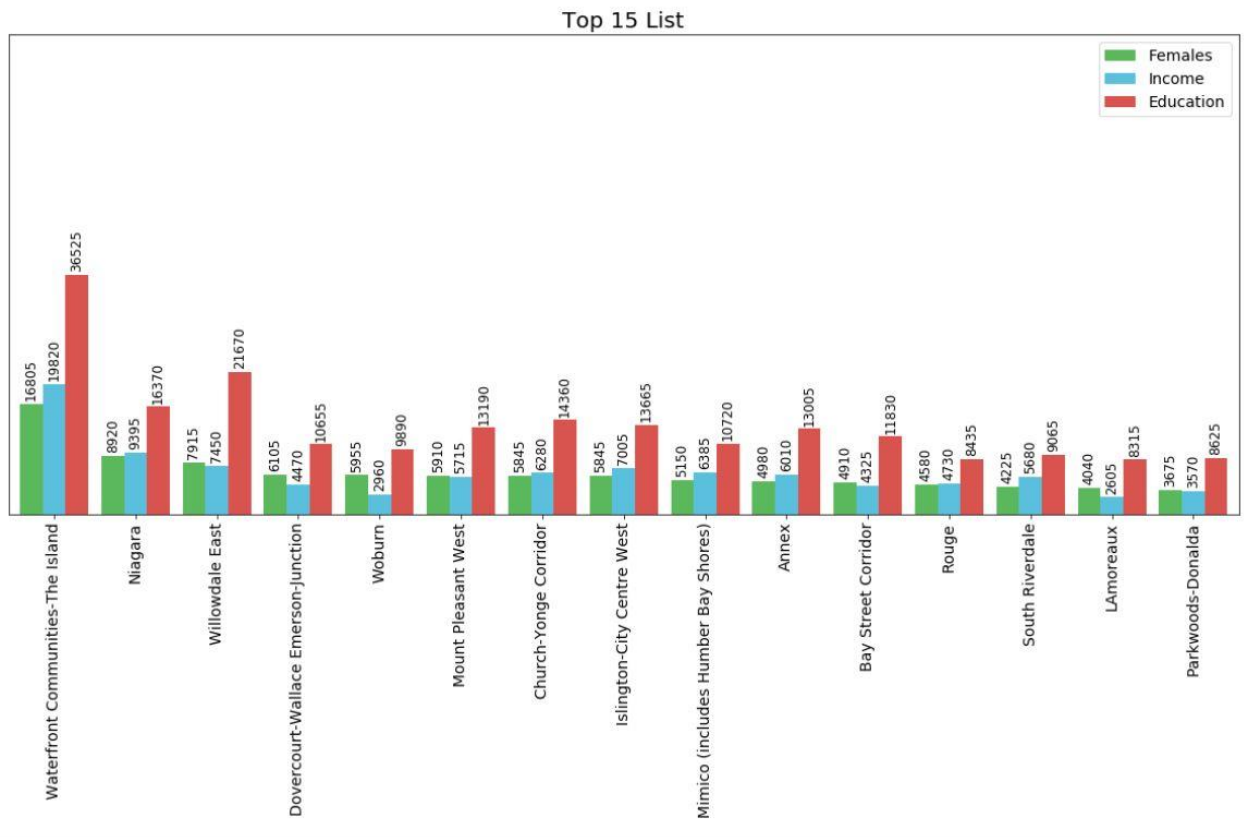
First, some basic statistics were calculated, mainly using percentage calculation. The population in Toronto is 2 731 571. The results showed that:

- Females from 25 to 39 years old form 12.3% of Toronto's total population (Female count= 336 960).
- App. 12.7% of Toronto population earns \$70,000.00 or more in a year.
- App. 27.4% of Toronto population are very well educated.

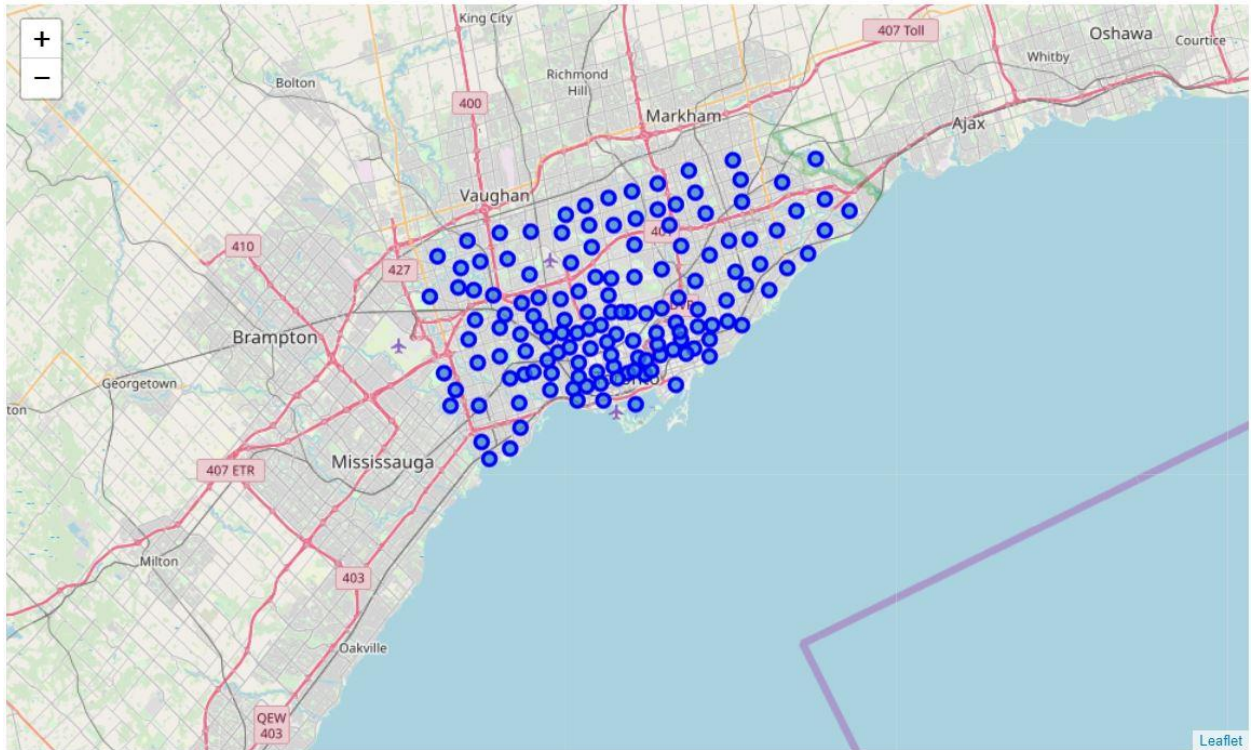
In sum, based on the demographic information there is a good market window in Toronto for a new vegetarian restaurant.

Second, out of curiosity, I wanted to figure out the best way to come up with the top 15 list of neighborhoods that fit the best with the Dalhousie profile, and what I was able to extract from the datasets. I came up with the idea where I sorted each feature (female/age, income, education) and put the top 40 of them in three different data frames. Then, I added them in the same data frame, sorted again by female (as it was the most important factor in this study), and kept only duplicates, one row of each. I saved the top 15 list

in a separate dataset for future use. The final top 15 list looks like this below created by using matplotlib.pyplot:



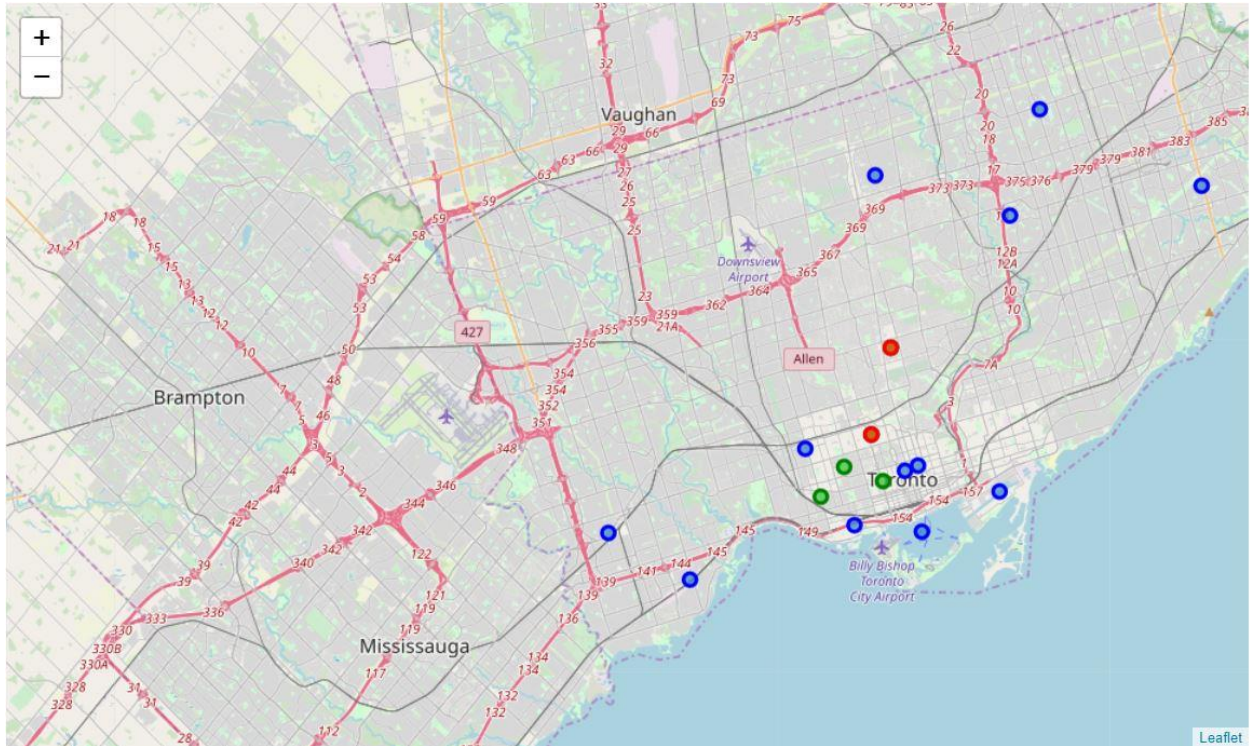
Third, I used geopy library and folium map to create a map of Toronto with all 140 neighborhoods. It looks like this below:



Fourth, I wanted to see how many vegetarian restaurants there are in Toronto. Getting this information, I used Foursquare API. The result showed that there are only 5 locations in Toronto, where we can eat in a Vegetarian restaurant:

| | Neighborhood | Vegetarian / Vegan Restaurant | AREA_SHORT_CODE | Longitude | Latitude | Female_Total | 70K+_Total | High_Education_Total |
|---|------------------------------|-------------------------------|-----------------|------------|-----------|--------------|------------|----------------------|
| 0 | Annex (95) | 1 | 95 | -79.404001 | 43.671585 | 4980 | 6010 | 13005 |
| 1 | Kensington-Chinatown (78) | 3 | 78 | -79.397240 | 43.653554 | 2850 | 1825 | 5345 |
| 2 | Little Portugal (84) | 1 | 84 | -79.430323 | 43.647536 | 3415 | 2405 | 5585 |
| 3 | Mount Pleasant West (104) | 1 | 104 | -79.393360 | 43.704435 | 5910 | 5715 | 13190 |
| 4 | Palmerston-Little Italy (80) | 1 | 80 | -79.418409 | 43.659157 | 2800 | 2195 | 5545 |

Then, I used the folium map again, this time to draw the following map, where **blue** markers denote the locations, where the potential customers live from the top 15 list, which I created earlier. The **green** markers show the locations, where we have vegetarian restaurants, and **red** markers show locations that have both criteria: they are on the top 15 list and they have a vegetarian restaurant.

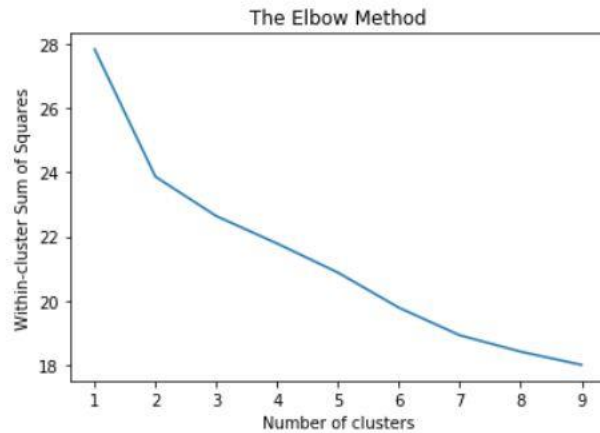


Interestingly, most of these markers are close to each other, which made me reassured that I'm on the right path for figuring out the best place for a vegetarian restaurant.

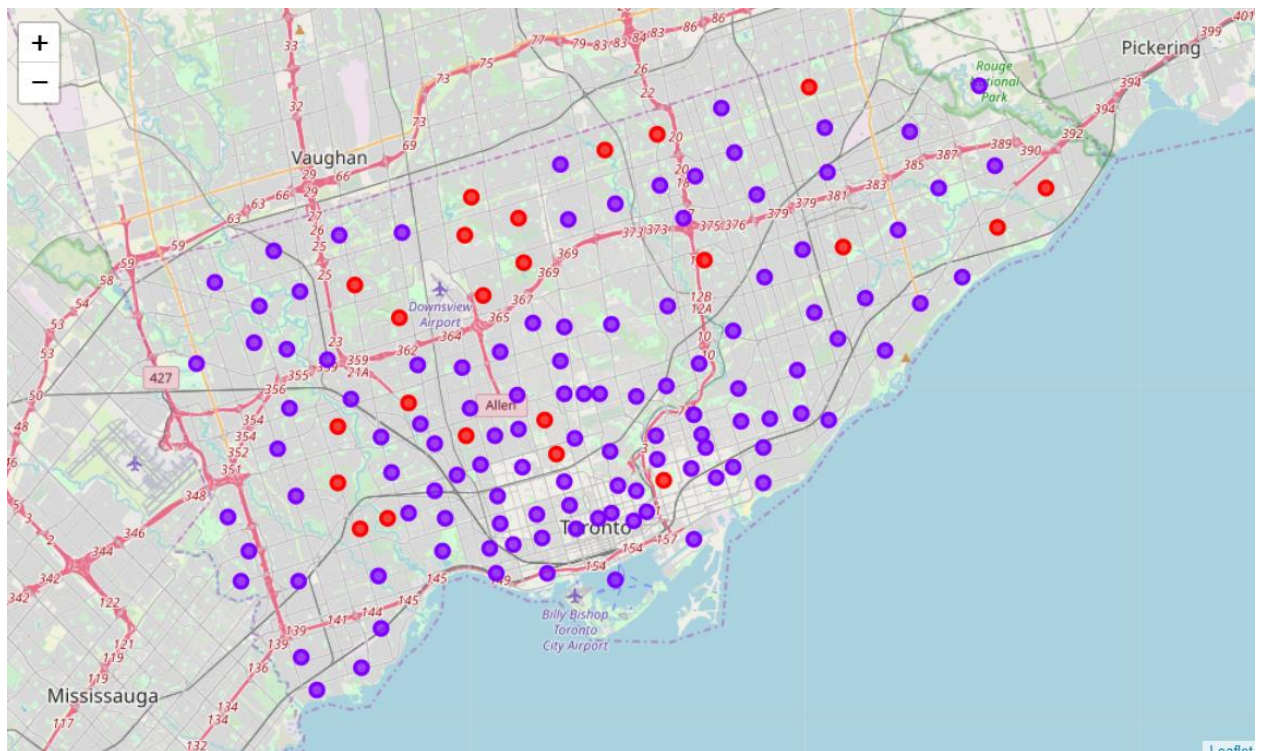
3.2. Machine Learning

I used K-Means for clustering neighborhoods based on venues that Foursquare API provided me from each neighborhood. First, I needed to figure out the best k , and then perform K-Means, and then put clusters on the map using folio map, and finally list the clusters for further analysis.

I used Elbow method for finding out the best k . The Elbow method draws an "elbow" that helps to select the optimal number of clusters by fitting the model with a range of values for K . If the line chart resembles an arm, then the "elbow" (the point of inflection on the curve) tells the number of clusters.



The Elbow Method suggested clustering Toronto in 2 different clusters. This cluster suggestion was used when performing K-Means. The cluster map is below.



4. Results

I named the two clusters with names “Sleeping Communities” and “Heartbeats of Toronto”. The difference between these two locations is apparent.

Cluster 1: “Sleeping Communities” are marked with red color.

The most common venues in this cluster seem to be: Park, Eastern European Restaurant, Egyptian Restaurant, Ethiopian Restaurant. There don't appear to be room for variety as much as in the other cluster. I programmed a piece of code that checked how many vegetarian restaurants are in this cluster and how many of the top 15 potential locations can be found in this cluster. The result showed that none of the existing vegetarian restaurants reside in the cluster one, and only one of the top 15 potential locations list can be found in this cluster.

Cluster 2: “Heartbeats of Toronto” are marked with purple color.

It can be seen immediately that this cluster is more versatile and multicultural than the cluster 1. Cluster 2 seems to have something for everybody. I also checked how many vegetarian restaurants are in this cluster and how many of the top 15 potential locations can be found in this cluster. The result showed that all of the existing vegetarian restaurants reside in the cluster 2, and 14/15 locations on the top 15 potential locations list can be found in this cluster.

5. Discussion

Already intuitively, I could say that locations in the cluster 1 “Sleeping Communities” are not the best choices for a new vegetarian restaurant. There must be a reason why venue offerings there are so unilateral. The results of the K-Means clustering are in accordance with my intuition.

I chose Waterfront Communities-The Island to be the best location for the new vegetarian restaurant for the following three reasons:

1. Most of the females from 25 to 39 years old live there (n=16 805) and the gender was the most important factor in the Dalhousie study. Also, individuals with greatest income (n=19 820) and education (n=36 525) live in the Waterfront Communities-The Island neighborhood.
2. Waterfront Communities-The Island neighborhood resides in the same cluster with the existing vegetarian restaurants i.e. "Heartbeats of Toronto". It is a busy area which helps attracting customers, who may be just walking on the waterfront.
3. Waterfront Communities-The Island neighborhood is far enough from competition i.e. it is not too close any of the existing vegetarian restaurants.

6. Conclusion

While there are many locations in Toronto, where a vegetarian restaurant could succeed very well, this study showed that Waterfront Communities-The Island neighborhood is the best choice for a new one. However, before putting this plan into reality, I would like to perform a survey in this location to ask questions about interest in vegetarian food, attitude regarding animal welfare, and populations health habits.