Report

-

Location Analysis for Veggie Grill in Toronto, CA

Image by lucasgeorgewendt from Pixabay

Ramin Hassanzadeh

6/15/20

Disclaimer: Please be aware that this is a mock project and decisions for Veggie Grill to expand to Toronto and the customer segmentation are not based on facts, but are made up for the purpose of this project.

Table of Contents

1. Introduction

Veggie Grill is a fast-casual vegan restaurant chain that operates in California, Oregon, Washington, Illinois, and Massachusetts. The first restaurant opened in 2006 in Irvine, California, which has since grown to be the largest vegetarian and vegan restaurant company in the U.S. The chain focuses on offering only plant-based food, with no meat, dairy, eggs, cholesterol, animal fat or trans-fat.

The owners of Veggie Grill have decided to expand into the Canadian market. Management of Veggie Grill have chosen Toronto as the city where the first restaurant should be opened in Canada.

To gain foothold in Toronto fast, a suitable location in the center of Toronto for the first restaurant should be chosen. The management of Veggie Grill now approached Capstone Data Science to analyze the ideal location, based on two criteria:

1. **Attractiveness:** Attractiveness of the neighborhood based on potential customers. Based on a customer segmentation done by Veggie Grill some months ago they identified that their customers are between the age of 15 to 45. Furthermore, their main revenue stream comes from employees who go out for lunch during lunch break.
2. **Competition:** Overall competition of vegan/vegetarian restaurants in the area.

2. Data

To determine a suitable location mainly 4 data sources will be used:

1. **Foursquare Venue Data:** For competition data Foursquare will be used to build competition clusters based on density of vegan/vegetarian restaurants located in a neighborhood. The data will be retrieved via the Foursquare API.
2. **Toronto Economic Data:** Toronto economics data will be used to determine how many businesses are located in a neighborhood, since Veggie Grill's main revenue stream comes from employees who go out for lunch. The data will be retrieved via the website of the city of Toronto: https://open.toronto.ca/dataset/wellbeing-toronto-economics/.
3. **Toronto Demographics Data:** Toronto population data will be used to determine how many of the customers in the identified age group between 15 to 45 years are located in a neighborhood. The data will be retrieved via the website of the city of Toronto: https://open.toronto.ca/dataset/wellbeing-toronto-demographics/
4. **Toronto Location Data:** To plot the high potential neighborhoods on a map, location data for longitude and latitude values will be downloaded from the Toronto website as well: https://open.toronto.ca/dataset/neighbourhoods/. Generating the data via geocoder is not possible without bigger efforts, since the neighborhood names cannot be easily found via geocoder.

Basically two analyses were conducted, based on clustering, which will result in a competition clustering and an attractiveness clustering. Both clustering exercises were then combined to derive a conclusion which neighborhoods have the highest potential to open a first restaurant.

All relevant data of the above-mentioned data sources will be combined into one data frame and then plotted on a Toronto map with meaningful color coding so that the relevant neighborhoods can be instantly identified.

3. Methodology

3.1 Data Cleaning

Since not all the information from the above-mentioned data sources was needed, the data was cleaned and aggregated to prepare it for the analysis.

Foursquare Venue Data: Via the Foursquare API venues per neighborhood were downloaded into a data set. The data set contains the venue name, the venue latitude and longitude and the venue category. For the analysis only an aggregated level on venue category is needed to determine competitiveness of a neighborhood.

Toronto Economics Data: This data set contains information per neighborhood on economic indicators such as businesses, child care, debt risk score, housing prices, local employment, and social assistance recipients. For the purpose of this analysis only the number of businesses was needed and the rest of the columns were neglected.

Toronto Demographics Data: This data set contains demographic information per neighborhood. The population is listed in 5-year intervals (0-4 years, 5-10 years, etc.). For this analysis only the population per neighborhood which is between 15-45 years of age was relevant. Therefore, all population figures for this age group were summed up to one aggregated number.

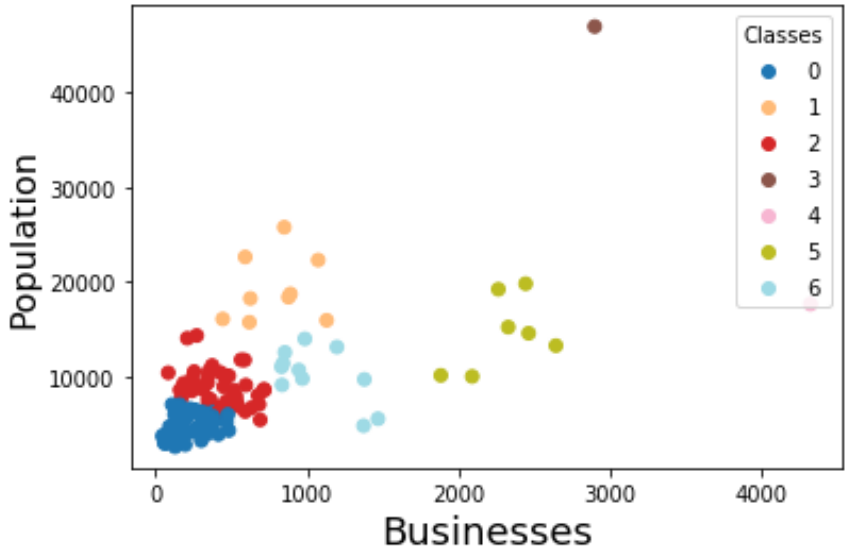
Toronto Location Data: For the Toronto location data information on area ID, area name and description, longitude and latitude values, etc. was available. From this data set only the longitude and latitude values were relevant and therefore the other information was neglected.

3.2 Attractiveness Analysis

To analyze the attractiveness of a neighborhood the number of business per neighborhood and the population data of residents between 15-45 years were combined into one data set for further analysis.

To be able to compare both input variables, the standard scaler from the scikit-learn library was used to standardize the data.

Finally, the k-means algorithm from the scikit-learn library was used to build 7 clusters from the data. The data was then plotted on a scatter plot to show the relationship between population and number of business and the 7 clusters.



Attractiveness Classification

Together with the management of Veggie Grill text labels for the 7 different clusters were defined. Each cluster was assigned to either having "high", "medium", or "low" attractiveness.

1. **High Attractiveness:** As you can see in the scatterplot above cluster 3 and 4 mark exceptional neighborhoods, which were therefore classified as "high" potential. Cluster 5 was also labeled as "high", since it has a significant amount of businesses present which means more employees looking for lunch opportunities.
2. **Medium Attractiveness:** Cluster 1, although higher in population than cluster 5 was classified as "medium" since it was lower in businesses located in these neighborhoods. The management of Veggie Grill prioritized number of businesses over population, since they see more potential to serve workers during lunch time, than the local population going out for lunch or dinner. Cluster 6 was also labeled as medium since it has an adequate amount of businesses and population.
3. **Low Attractiveness:** The rest was classified as not attractive for opening the first restaurant.

The labels were then applied to the original data set. Below you can see an extract of the generated data set.

Neighborhood	NeighbourhoodID	Businesses	15 to 45 years	Attractiveness
West Humber-Clairville	1	2463	14590.0	high
Mount Olive-Silverstone-Jamestown	2	271	14350.0	low
Thistletown-Beaumond Heights	3	217	4080.0	low
Rexdale-Kipling	4	144	4070.0	low
Elms-Old Rexdale	5	67	3705.0	low

3.3 Competitive Analysis

To analyze the competition in the different neighborhoods data from Foursquare was used to extract the venues per neighborhood, focusing on restaurants offering vegan and vegetarian food. Then the neighborhoods were classified into high, medium, and low competitive neighborhoods.

The resulting data frame from the Foursquare API request was transformed via one hot encoding for summing all vegan and vegetarian restaurants in a neighborhood. All other categories besides the Vegetarian / Vegan Restaurant category were dropped.

In the next step a k-means classification algorithm was applied to the data set using only the sum of Vegetarian / Vegan Restaurants in a neighborhood as an input variable. The data was classified into three competition buckets: high, medium, and low competitive neighborhoods.

Extract from the data frame:

NeighbourhoodID	Vegetarian / Vegan Restaurant	Competition
78	3	high
84	1	medium
104	1	medium
95	1	medium
65	1	medium
80	1	medium
1	0	low

3.4 Location Identification

After classifying the neighborhoods based on attractiveness and competition both data sets were merged together and a logic was defined to evaluate the neighborhoods. The following table shows the logic of the location classification.

Attractiveness		Competition		Location Classification
high		low		Excellent
high		medium		Okay
high		high		Poor
medium		high		Poor
medium	+	low	=	Good
medium		medium		Poor
low		high		Very Poor
low		medium		Poor
low		low		Poor

The classification was applied to the original data set as you can see in the extract below.

Neighborhood	N_ID	Businesses	15 to 45 years	Attractiveness	Vegetarian / Vegan Restaurant	Competition	Location Choice
West Humber-Clairville	1	2463	14590.0	high	0	low	Excellent
Mount Olive-Silverstone-Jamestown	2	271	14350.0	low	0	low	Poor
Thistletown-Beaumond Heights	3	217	4080.0	low	0	low	Poor
Rexdale-Kipling	4	144	4070.0	low	0	low	Poor
Elms-Old Rexdale	5	67	3705.0	low	0	low	Poor

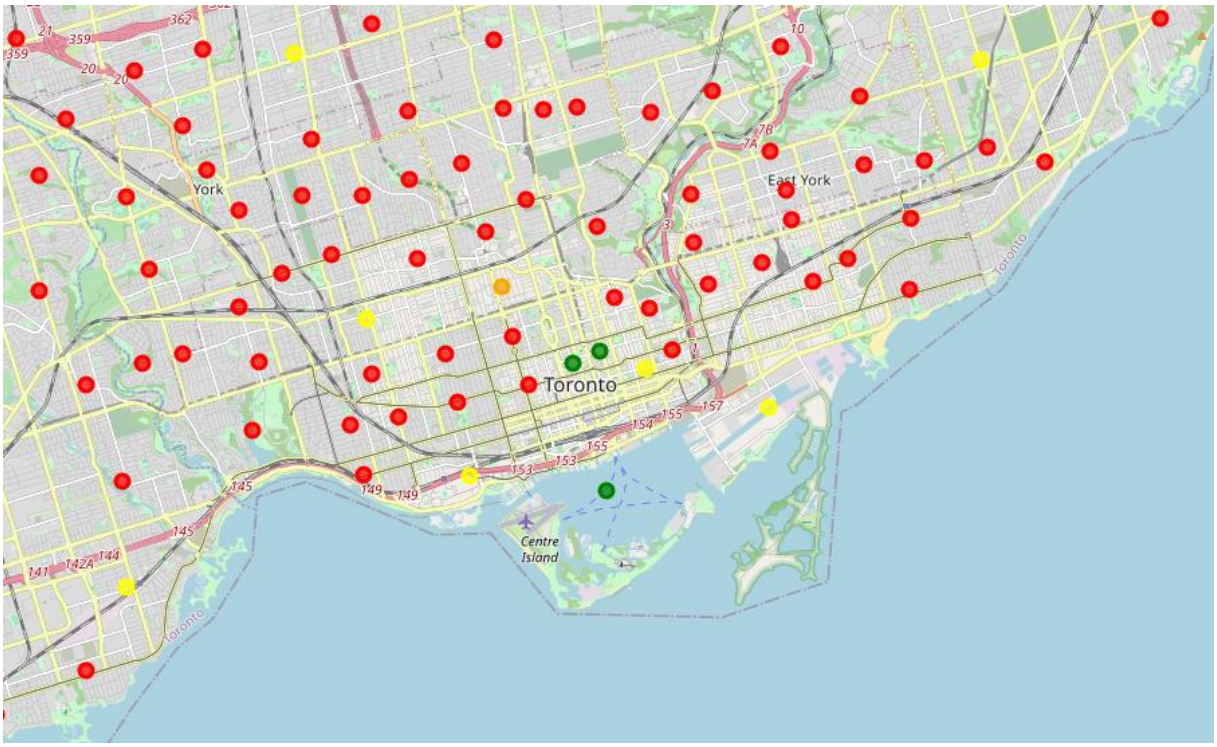
Extract of consolidated data frame.

4. Results & Discussion

The results of the analysis were plotted on a graph for visual representation. This was done using folium and a color coding.

Below you can see the map with the neighborhoods based on the location choice classification. The locations are colored coded by the following logic:

- **Excellent** is highlighted in **green**
- **Good** is highlighted in **yellow**
- **Okay** is hihglighted in **orange**
- **Poor** and **very poor** are highlighted in **red**



Let's now explore the 3 neighborhoods marked in green in the Toronto center further. These are the neighborhoods:

- Waterfront Communities-The Island
- Bay Street Corridor
- Church-Yonge Corridor

Neighborhood	Businesses	15 to 45 years	Location Choice
Church-Yonge Corridor	2443	19815.0	Excellent
Bay Street Corridor	4324	17665.0	Excellent
Waterfront Communities-The Island	2899	46945.0	Excellent

As you can see above Bay Street Corrider offers the most amount of businesses with a high population number of habitants between 15 to 45 years. Since it is located in the heart of Toronto, compared to Waterfront Communities-The Island, it should be preferred.

Limitations of this analysis

There are a couple limitations to this analysis, which will be discussed in the following:

1. First of all, the data that was provided by the city of Toronto is dated between 3 to 6 years back. So, the analysis is not dealing with the most up to date data regarding the population and business information.
2. Secondly, we are only looking at explicitly vegan or vegetarian restaurants from Foursquare but neglecting restaurants offering vegetarian or vegan options. For a further deep dive this could be considered.
3. Last but not least, availability of a suitable restaurant location and rental fee levels of neighborhoods were not considered which could also have an impact on location choice

5. Conclusion

To conclude Capstone Data Science suggests, due to the above analysis to open a first restaurant in the Bay Street Corridor neighborhood. Since both Church-Yonge Corridor and Bay Street Corridor are located next to each other, a location in between those neighborhoods could be chosen as well.

Image by mark.watmough from Wikipedia