

# Opening a Vegetarian Gourmet Restaurant in Toronto

## Introduction

In this project it is examined which neighborhoods in Toronto may be best suited for opening a new vegetarian gourmet restaurant. Gastronomy has been one of the hardest hit business sectors by the pandemic. Many restaurants had to close and went out of business for good. As the economy starts to recover, it may be a good time to enter the market with new ideas, concepts and focus. Already in recent years a trend towards high quality food could be observed. There is an increasing awareness for environmental and health issues, which comes along with a rising number of vegetarians. Eating habits are beginning to change beyond the traditional 'green' milieus. Caring more about food and a more pleasant dining culture, is increasingly becoming a matter of aspired lifestyle. As people are rethinking their priorities, there is also a higher willingness to pay for good, high-quality food. Since vegetarian cuisine also has become much more sophisticated and splendid, customers are not just turning to vegetarian meals for health and environmental reasons, but also because they are tasty.

These tendencies provide promising business opportunities for a vegetarian gourmet restaurant. Of course, there already exists a sizable number of vegetarian restaurants in Toronto. A closer look, however, reveals that they can be found predominantly in lower price segments. It has become quite common that also upscale restaurants offer some vegetarian dishes, however, their choice typically is quite limited. Only a specialized vegetarian gourmet restaurant can tap the full richness of vegetarian cuisine. The demand for such a restaurant clearly seems to be given.

Hence, the project addresses prospective new restaurant owners as well as already established players in the market interested in opening a new branch in this promising field. The approach of this study may also very well be applicable to decisions on where to locate a new business generally.

## Description and Background of the Problem

Location matters and this certainly also holds true for restaurants. A flourishing restaurant depends on a healthy mixture of regular customers who come to the restaurant on a regular basis and provide for a steady and predictable flow of income as well as occasional and one-time customers who visit the restaurant for a date, an anniversary, curiosity, or just by chance. The respective shares of these two groups of customers may very well vary for each restaurant, depending on the type of restaurant, individual features and well, also its location.

Hence, there are two underlying and partly interlinked questions that need to be addressed when looking for a restaurant site:

- What are the main target groups and how does location contribute targeting them?
- How well does the planned restaurant and the overall business environment at a specific location fit together?

Regular customers may predominantly be people who live or work relatively close to the restaurant. To target this group of customers it makes sense to have a closer look at some demographic features like age, education and income of people living in the respective neighborhoods. Since the planned restaurant aims at high-quality cuisine, prices will be in the middle to upper price range. This means that regular customers may tend to be relatively well-off, even though not necessarily upper class in order to be able to afford the restaurant as prices will still be lower than in the traditional meat-based haute cuisine. Generally, it can be observed that vegetarians tend to be ecologically minded, relatively well educated, at younger ages and also more female than male. Of course, such characterizations are rather sketchy and also quite transient in light of broad social change that is taking place. Nevertheless, for the time being they may provide some useful indicators for the kind of neighborhood coming into consideration as potential site for the restaurant.

For occasional customers the overall attractiveness of the neighborhood as a hub for night life and for dining out may be the most decisive criterion to come there. To target this group of customers it seems reasonable to look at the local venue structure to get an idea about how busy and lively a neighborhood is and whether there are other attractions and sites nearby that could also draw customers to the restaurant.

Hence, there are many different aspects that flow into the decision-making process and different target groups may also lead to different prioritizations. The challenge thus is to identify those neighborhoods that measure up to these different demands most convincingly.

## **Description of the data and its use**

The study is based on two major sources of data: the Toronto Census dataset available as open data from <https://open.toronto.ca/dataset/neighborhood-profiles/> and the Foursquare global database that can be accessed via PLACES API.

The Census data contains detailed demographic information about 140 Toronto neighborhoods ranging from housing to ethnocultural diversity. For the present study, population, age, personal income and higher education are of particular interest as this data provides the most relevant information about the pool of potential residential customers in each neighborhood.

For the neighborhoods as defined in the census data the latitude and longitude coordinates are retrieved using Geocoder. The dataframe created from this data will then be used for calls to the Foursquare API.

A first call request is made to search for vegetarian restaurants in general as well as for vegetarian restaurants in the upper price range to get an overview of the market and also immediate competitors.

The second request aims at searching for all venues across all neighborhoods. The dataframe resulting from this latter call serves as the main data source for more detailed exploration of various selected venue categories for which the number of venues per neighborhood were counted and added to a new dataframe. The idea is that these numbers and their distribution across neighborhoods provide some valuable insights into the local business structure. Of special interest is also the competitive intensity of a potential location, i.e. in how far already existing venues may either interfere with or rather complement the planned restaurant.

The data retrieved from the Census and the Foursquare databases are then merged into a new dataframe that provides a thorough survey of the 140 neighborhoods based on the previously selected criteria. This new merged dataframe is the fulcrum of this study and serves as basis for the subsequent cluster analysis. The idea is, that based on the resulting clusters of neighborhoods, it will be possible to identify a limited number of neighborhoods as prospective candidates for the planned restaurant. These selected neighborhoods are then compared with each other in more detail in regard to income and age distribution that are visualized in a bar plot. This provides a clearer picture of the pros and cons of each candidate neighborhood and helps narrowing the choice further down.

Caveat: The Toronto Census Data stems from 2016. Since the census is held every 5 years there should be updated data available in the course of this year.

## **Methodology**

The program is written with python 3.7. Several packages and libraries are augmented.

The prime toolkit for data analysis that is applied are pandas. The Toronto Census Data is directly retrieved from the Toronto Open Data webpage (as API) and read into a pandas dataframe. Following, quite a bit of data cleaning had to be performed to obtain a new dataframe with neighborhoods as column index and the selected neighborhood characteristics as columns. Some characteristics can be extracted directly from the source dataframe like population, average income and the number of people with higher education. Some additional, more detailed characteristics like 'individual incomes higher than \$30000' and their share in the overall population have to be specifically calculated. \$30000 annual individual after-tax income has been chosen as lower income limit in consideration of the targeted upper price segment of the restaurant.

People in this income group typically tend to be middle age and older so that the target age group has been set to 30-70 years accordingly. The two sexes were combined to one age group. Even though females may be more likely to visit the restaurant, the sex differences in the

targeted middle age group were too small to justify sex as a separate selection criterion. The absolute number of people in the selected categories were calculated as well as their share of the population which allows a more finely nuanced picture of the socio-economic structure of each neighborhood.

Data wrangling included selecting and renaming columns, removing commas and special characters from column values, transposing columns and rows and doing some math operations like division and summation. While working with the age statistics it turned out that in the source table one row (Females: 10 to 14 years) was wrongly placed in between the age data for males, which also had to be fixed.

Since the Census Data does not contain any geographical data for the 140 neighborhoods, Geocoder was used to retrieve the latitude and longitude coordinates that are needed for using the Foursquare database.

The requests library has been applied to make two kinds of requests to the Foursquare API in order to analyze the business environment for the planned restaurant:

- First, an all-encompassing request for vegetarian restaurants in different price categories in the entire Metropolitan Area based on geodata for Toronto as a whole with a radius of 20000.
  - o Note, that for the search for different price categories the *explore* endpoint is required.
- Second, a broad venue search for each neighborhood with a radius of 1500.

The first request serves the purpose of analyzing the market for vegetarian restaurants in Toronto and to get an idea of its competitive intensity. To pursue this goal, it is indicated to get an as accurate number of already existing vegetarian restaurants as possible.

The starting position for the venue search in the second request is quite different. The goal here is a broad characterization of the business structure and the vibrancy of each neighborhood that flows into the cluster analysis later on. Certain venue categories are selected that are considered meaningful indicators for the attractiveness of the neighborhood as location for the new restaurant, like natural food and health stores, farmers' markets, gyms and yoga studios.

The complete compilation of venues for this kind of analysis is not necessary, as long as the general picture gets not distorted. This aspect is important in regard to the choice of radius. Test trials revealed that even though a larger radius increases the total number of venues retrieved, the category search, however, brings about less differentiation among the neighborhoods as the number of venues shown per category is restricted to 100 so that quite many neighborhoods actually fall into this 100-and-over-bracket. A higher radius also leads to more double counting, diminishing the advantage of missing less venues. Balancing these conflicting effects a radius of 1500 has been chosen for each neighborhood search, which also takes into account that downtown neighborhoods with a relatively high density of venues tend to be rather small.

Technically, the result of the venue search is a json file that is converted to a pandas dataframe by means of the `json_normalize pandas` command. This data table serves as the basis for the more specific category search to determine the number of venues in these categories in each neighborhood. From the obtained data a new dataframe is created with the neighborhoods in the index column and the specific venue characteristics as columns. Eventually, this dataframe is merged with the dataframe retrieved from the Toronto Census. This merged dataframe comprises all 15 selected neighborhood characteristics and serves as the basis for the subsequent cluster analysis. The head of this dataframe is shown here:

	Pop.	Pop. Change (%)	Higher Education	Higher Education (%)	Av. Income (\$)	Av. Ret. Income (\$)	Inc. above \$30000	Inc. above \$30000 (%)	Pop. 30-70 years	Pop. 30-70 years (%)	Venues	Rests	Veg Rests	Compl. Venues	Gyms
Agincourt North	29113	-3.9	10985	37.73	26955	18482	7995	33.0	15425	54.0	41	14.0	0.0	0.0	0.0
Agincourt South- Malvern West	23757	8.0	10285	43.29	27928	18200	6990	38.0	12530	53.0	65	28.0	0.0	0.0	3.0
Alderwood	12054	1.3	5300	43.97	39159	18562	5390	53.0	6950	59.0	68	10.0	0.0	1.0	1.0
Annex	30526	4.6	20430	66.93	80138	42918	14935	58.0	15885	54.0	100	28.0	3.0	3.0	1.0
Banbury-Don Mills	27695	2.9	15950	57.59	51874	29704	13075	59.0	14650	53.0	45	15.0	0.0	1.0	1.0

The essence of the cluster analysis is to segment the neighborhood base into groups of neighborhoods with similar characteristics to narrow down the number of potential candidate neighborhoods. To conduct the cluster analysis, KMeans has been applied. This clustering device is based on numerical and unlabeled values that flow into the algorithm to calculate the mean values that determine the clusters. After various numbers of clusters have been tested, the analysis, eventually, is performed with 10 clusters. This number turned out to be large enough to yield a meaningful differentiation between the neighborhoods and small enough to avoid overly fragmentation so that a meaningful selection of the most promising neighborhoods seemed likely.

To check and further analyze the outcome of the cluster analysis, three candidate neighborhoods are compared with each other in regard to income and age distribution to get a more detailed and more differentiated picture than from the aggregated income and age data used before. Particularly the previous classification of middle age population had not yielded very much differentiation. In regard to income, the shares of the population in 5 income brackets are now calculated for the selected neighborhoods and visualized in a grouped bar chart by means of the Mathplot library. In regard to age, a new dataframe is created from the census data for each selected neighborhood that relates the number of females and males to

the different age groups. These dataframes were used to plot a population pyramid with horizontal bars for each neighborhood separately.

Caveat: As github has some problems displaying text which spans across multiple rows when it is in index, which is the case for some census data, the notebook cannot be retrieved accurately from github, unfortunately. On nbviewer, however, it is displayed properly.

## Results

The Foursquare search for vegetarian restaurants yielded quite an imbalanced distribution in regard to price categories. Of 94 vegetarian restaurants found in Toronto, only 3 restaurants fall into price category 3 (and none into category 4), which means that almost all already existing vegetarian restaurants belong to the low price categories 1 and 2.

The main results of this study refer to the cluster analysis. The clustering of the 140 neighborhoods into 10 clusters brought about a clear pattern that allowed to characterize each cluster by some distinguishing features.

Most clusters, however, could be taken out of consideration right away for being either predominantly residential areas without much business or entertainment going on or for being less suitable for a vegetarian gourmet restaurant due to their social structure. Three clusters are actually shortlisted.

Cluster 9 is interesting as it includes the most affluent downtown neighborhoods with a relatively high density of venues. Besides many restaurants one also finds relatively many venues that indicate high preferences for high quality food and a healthy lifestyle. These neighborhoods certainly belong to the narrow circle of candidate neighborhoods.

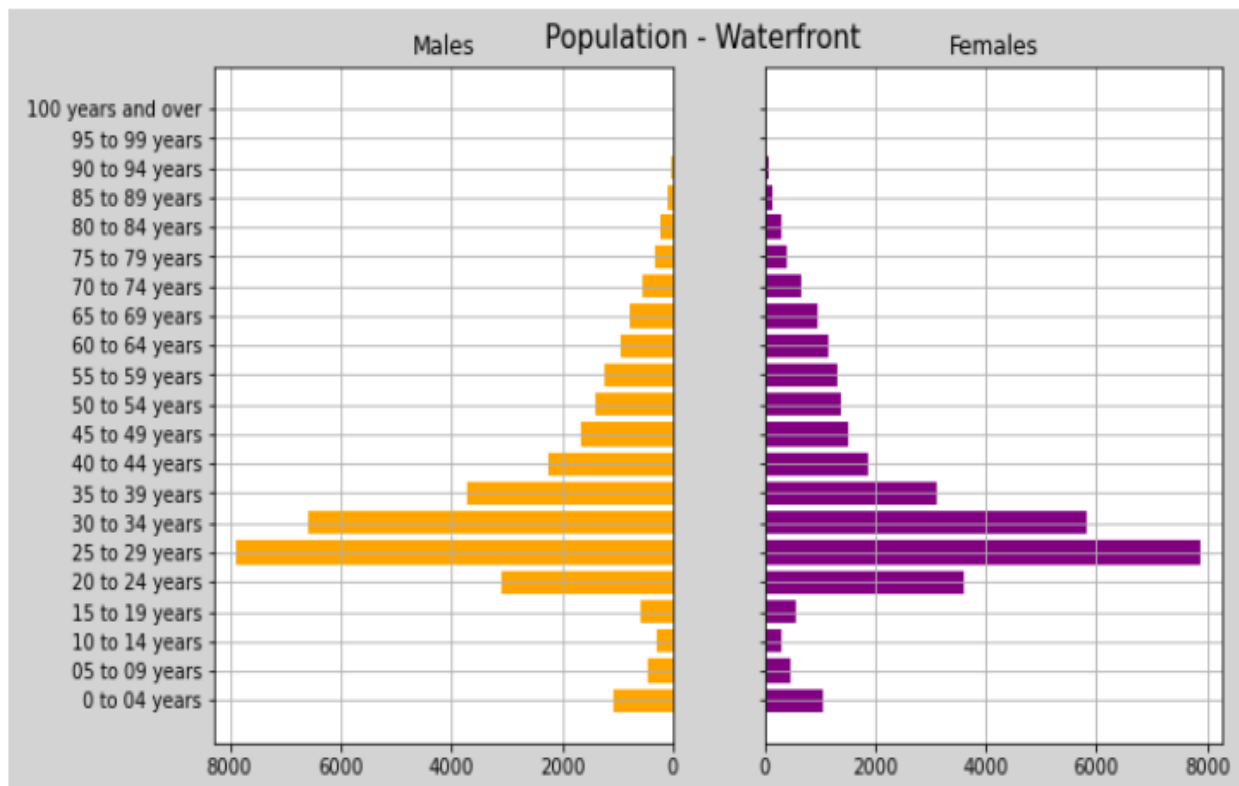
Cluster 2 grabs attention as the cluster with the highest number of already existing vegetarian restaurants and a venue structure that generally suggests high awareness for ecological concerns and an openness to a 'green lifestyle'. From a business clusters perspective, it could make sense to also locate the vegetarian gourmet restaurant in one of these neighborhoods. The rationale, are mutually reinforcing benefits as more and also new customers with respective 'green' preferences may be attracted to the neighborhood also giving even greater distinction to it. Hence, the neighborhoods in this cluster certainly must be seen as likely candidates as well.

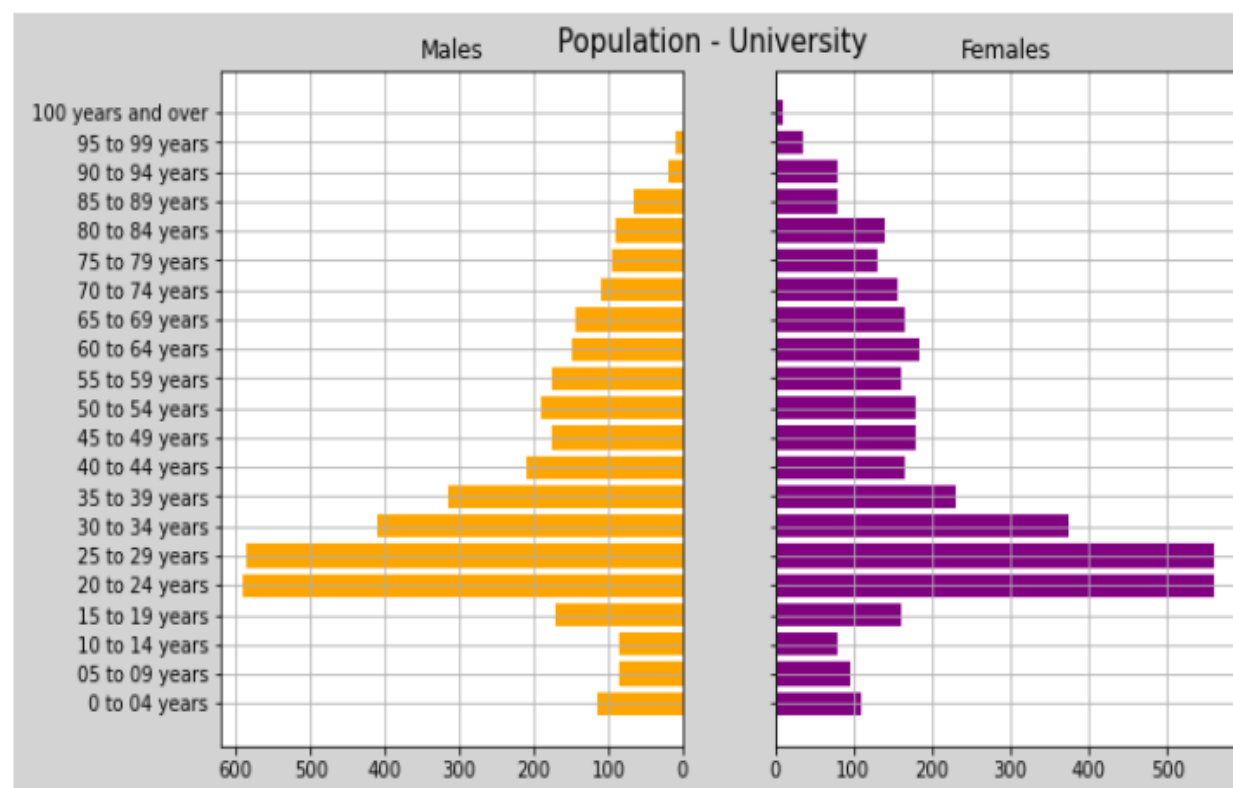
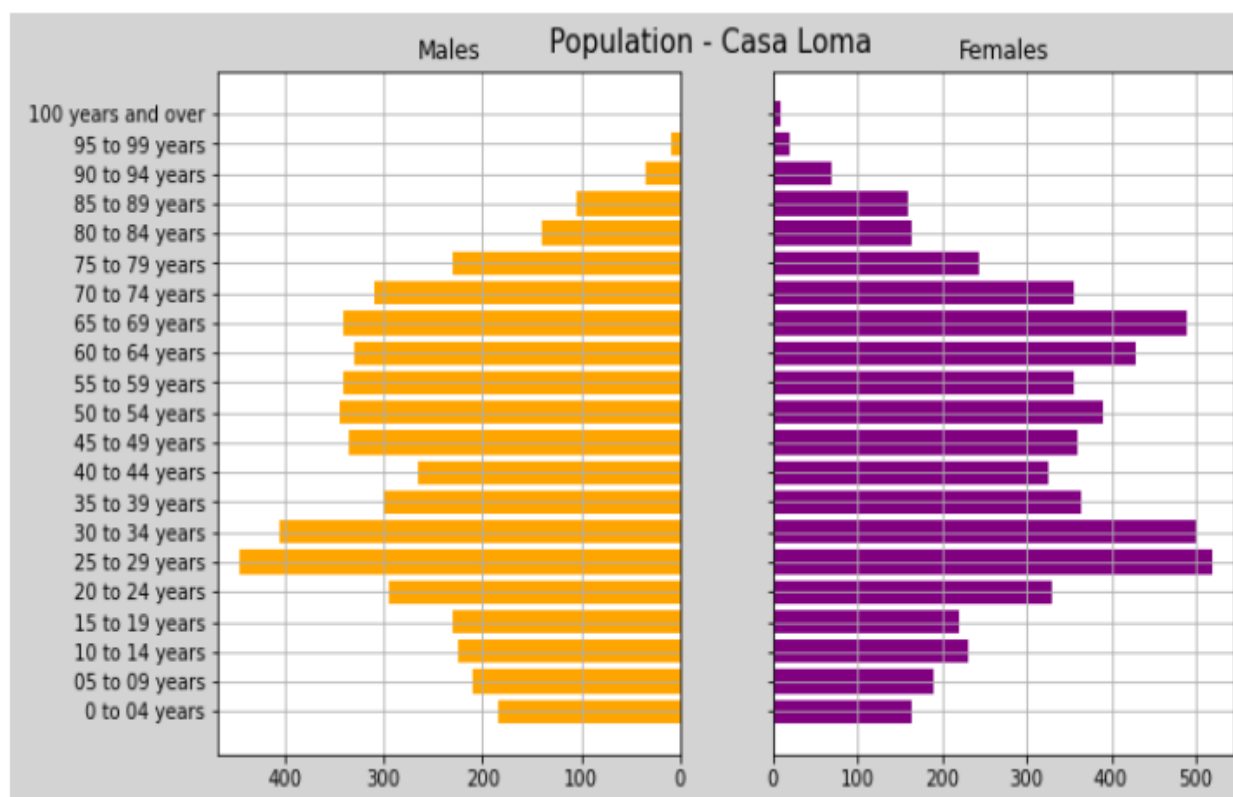
The most promising and selective cluster, however, is cluster 4 that singles out the 'Waterfront Communities – The Island' as the only neighborhood. The Waterfront Communities stand out on various accounts. It is the most populous and fastest growing neighborhood, with the largest share of people with postsecondary education as well as the highest share of individual incomes above \$30000. Also in regard to complementary venues the Waterfront Communities rank first suggesting consumption patterns very well amenable to a vegetarian gourmet restaurant. As

the name of the neighborhood, however already suggests, it is a huge neighborhood that actually comprises several communities, making it difficult to lump all these communities together under one label. In all, however, one can say that it is a vibrant, developing part of Toronto, with new business and venue structures emerging, where a vegetarian gourmet restaurant seems to be a very good match.

To check the plausibility of the selection of the Waterfront Communities – The Island, a comparison of the income and age distribution with two other likely candidate neighborhoods (Casa Loma of cluster 9 and University of cluster 4) has been conducted. In regard to age distribution, it turned out that the Waterfront Communities look more similar to the University with the majority of the population being in their 20s and 30s.

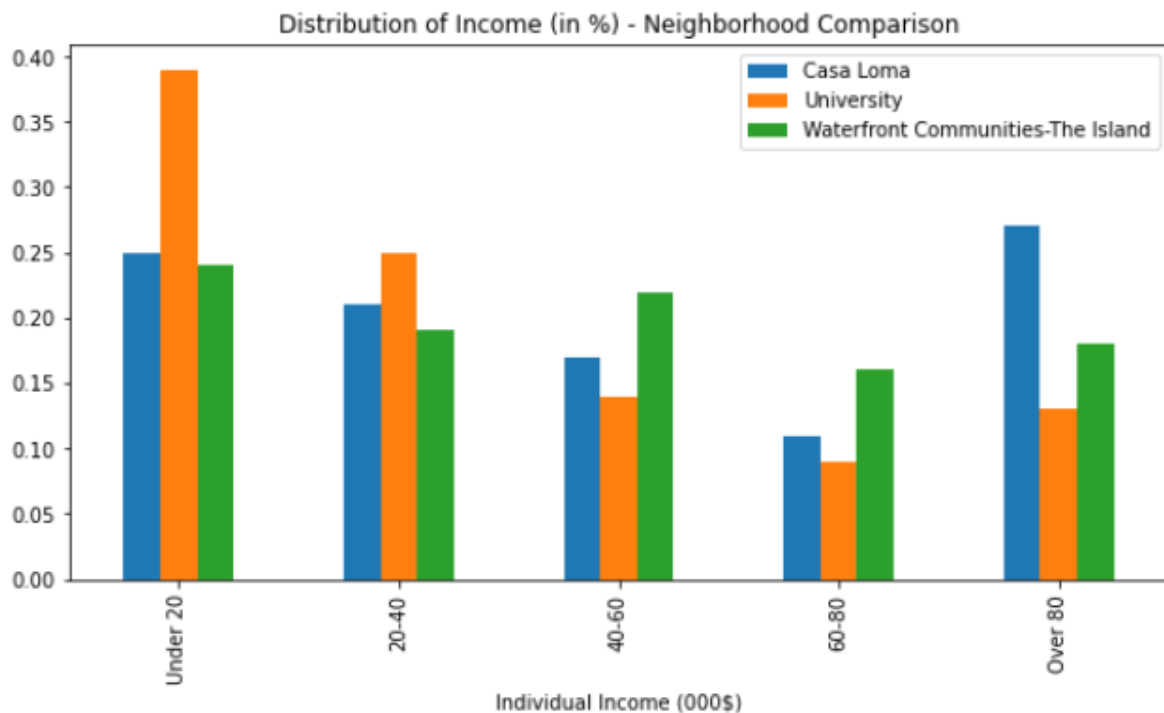
The population pyramids of the three neighborhoods are shown here:







The comparison of the income distribution, on the other hand, revealed a closer similarity with Casa Loma. Casa Loma, however, has a larger share of people in the highest income bracket with individual incomes above \$80000, while the Waterfront Communities show a more balanced income distribution with roughly the same share of incomes in each income bracket. As a consequence, there are many more people in the middle and upper middle income range in the Waterfront Communities than in the other neighborhoods. At the same time, people with incomes within this range may also be the most likely customers for the restaurant.



## Discussion

The market analysis revealed that the market potential for a new vegetarian gourmet restaurant seems to be clearly given. The question is whether the Waterfront Communities are best equipped to tap this potential. It is a neighborhood with a potential customer base that certainly appears to fit a vegetarian gourmet restaurant quite well. The age and income distributions in comparison to University and Casa Loma demonstrate that the Waterfront-Communities seem to provide some kind of synthesis of the respective advantageous aspects of both neighborhoods. University represents the traditional idea of a green constituency of young adults with high values and ideas but little income. Casa Loma has a much more balanced age distribution than the other two neighborhoods with way more middle age and older people who also have much more spending power. They, too, show environmental awareness and high

preferences for a healthy life style and high quality food, which however, does not necessarily imply a commitment to a fully vegetarian diet. Statistical evidence so far suggests that vegetarianism is much more common in younger age groups than among the elderly, which, however, does not mean that this won't change in future as new generations grow older. For the time being, however, the focus on younger people seems justified. The population of the Waterfront Communities of predominantly young, highly educated professionals, seems to share much of the green values, attitudes and outlooks on life prevalent in University, while at the same time having the financial means to live up to these ideas more comfortable, which may also include a more splendid vegetarian cuisine.

Besides residential consumers, the Waterfront Communities and The Island may also attract people from outside the neighborhood, given the proximity to the lakeshore and other attractions including various restaurants that make the neighborhood generally an interesting destination for a visit and for eating out. However, the great size of the neighborhood also needs to be taken into account. As the name already suggests, the neighborhood is comprised of several communities, all with their own characteristics. A thorough location search would demand a further specific analysis of the Waterfront Communities themselves. The most basic question being whether to settle on the Waterfront Communities or on The Island. For the present study, however, it is sufficient to conclude that the Waterfront Communities – The Island is indeed a meaningful and conclusive result.

## **Conclusion and Outlook**

In order to decide on where to locate a new vegetarian gourmet restaurant the present study applied a cluster analysis based on data from the Toronto Census as well as from the Foursquare database. 15 neighborhood characteristics were chosen as selection criteria for the clustering, covering demographic data as well as some business venue counts. Surprisingly, this analysis yielded an even more clear-cut result than expected, which was then also confirmed by some plausibility check. Generally, the conclusiveness of the clustering always depends on the meaningfulness of the selection criteria applied. The analysis in its present form puts relatively much weight on demographics (particularly age) assuming that young, highly educated professionals may be the main customers for the restaurant. This seems reasonable, however, the selection criteria chosen are certainly not exhaustive. Some aspects that are currently not considered may very well be crucial for decision-making. This concerns particularly the costs of premises at a particular location. Accessibility by public transportation may also be an important factor. As long as these aspects can be measured by real estate prices and the proximity of a Metro Station, for example, they can easily be included into the framework. Such additions, however, would also demand additional data sources, which clearly would have gone beyond the scope of this study. In principle, however, the approach is flexible enough to broaden its scope and to serve as a blueprint for a more comprehensive analysis. Even in this limited form, did the study demonstrate that a cluster analysis is well suited for the task at hand.