Introduction

1.1 the Problem

The purpose of this investigation is to use population statistics to see if trends can be found to suggest where a producer of meat, dairy, or egg substitute products might want to distribute or promote their products in the city of Toronto.

1.2 Background

According to Vegetarian Resource Group 37% of people in the united states in 2016 avoided eating dishes containing meat, fish, or poultry at least part of the time when dining out[1]. In the UK vegan take out is the fastest growing area of take out meals[2] and the number of vegans in the US increased six times in 2018[3]. While in Canada, the majority of people committing to vegan or vegetarian lifestyles are under 35[4]. This suggests strongly that restaurants and grocery stores will likely want to carry alternatives to meat in the near future as the data suggests a growing number of vegetarians and vegans.

Data

2.1 Data Sources

In this project we will look at the neighborhoods of Toronto and using the foursquare API find out which have the highest density of Vegan and Vegetarian restaurants. We will add this data to data pulled from the Canadian census bureau to find the fractions of: males, females, people of high income (top five deciles), people on public assistance, people employed, people of European descent, people of Asian descent, Canadian Citizens, and people aged fifteen to thirty-four. Population density will be used in addition. The population data will be taken from Toronto Open Data project and is sourced from the Canadian Census Bureau (https://open.toronto.ca/dataset/neighbourhood-profiles/). Geographic data is sourced from the same project (). The specific details were chosen as they are often accessible to the public in for major cities in Canada and the United States making any trends found desirable.

2.2 Data Cleaning

The total fraction of each of the fractional categories listed above was not readily available in the census and the census columns were listed by neighborhood with data categories as a column. I made the characteristic the head of a new transposed data frame with the neighborhoods as the index. The numbers over 1000 in the survey contained commas I.E. 1,000 and not 1000. This meant that extraneous symbols had to be removed using string manipulation. For the geographic data I had to use string manipulation to make the names of the neighborhoods match those of population data dataframe. Finally I had to sort the geography dataframe by neighborhood name, set the names as the index and merge the frames. For the venues I chose to get geographic coordinates and write a function to match the venue to closest neighborhood in distance from the center.

The fractional values needed were not readily available to I made lists of sums of population of males and females over all age brackets and then divided by the population column. I did this with income brackets in the top five income deciles and the age brackets for people of a young age total. The two racial groups merely needed to be divided by total population element-wise. The percent employed and percent on public assistance were readily available and simply set to fractional values.

Methodology

For this project k-means clustering was used to divide the neighborhoods into clusters. The relevant data includes the fraction of males in the area, the fractions of females, the fraction people of European descent, the fraction of people of Asian descent, the fraction of people in the top five income deciles, the fraction of people employed, the fraction of people receiving public assistance, and the fraction of the population between fifteen and thirty-five years of age. Subsequently the number of vegan or vegetarian restaurants in close proximity to each neighborhood would be ascertained. The data would then be clustered and the individual clusters examined. The best k values for clustering would be assessed via the elbow test method. After this it would be assessed which neighborhoods in the desired cluster had the highest number of grocery stores in order to see what locations might be best for promotion or distribution.

Results

After the clustering was complete it was found that the greatest density of vegan and vegetarian restaurants was found in the cluster of neighborhoods near center city Toronto. There were four total clusters formed. The number was chosen using the elbow method with results showing cluster efficiency versus number below. A value of five was chosen for the clustering.
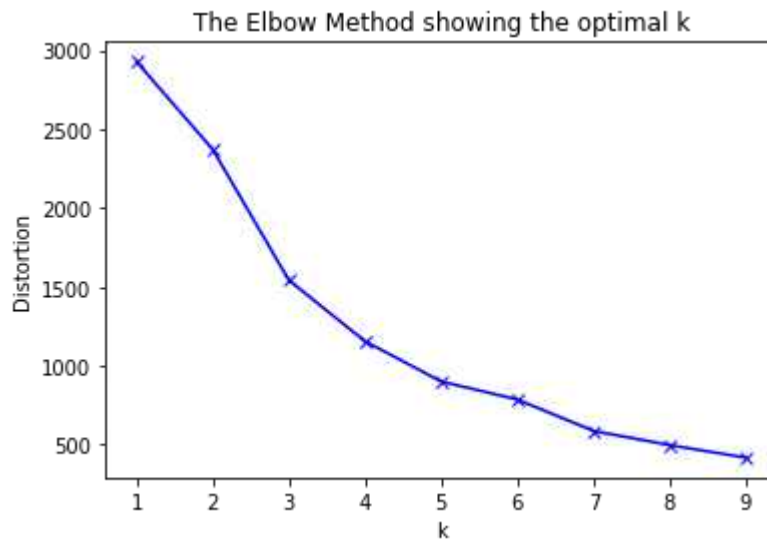

Fig. 1 Data collected versus cluster number.

On forming the clusters it became apparent that two of the clusters contained all of the top ranked vegan and vegetarian restaurants in the Toronto area. These clusters had the highest fraction of young people and also had high average population densities.

As expected it was found that these neighborhoods had a large portion of young people living in them compared to the other clusters. It was found also that a high average population density could indicate the presence of many vegan/vegetarian restaurants but was not necessarily decisive. For example the highest density of these venues is found in a neighborhood with a high population density, while another North St. James Town has none. Below figure two will show the number of restaurants and young people for neighborhoods with more than one restaurant and figure three will do the same for population density. In both case the venue count is in red and the other value in blue.
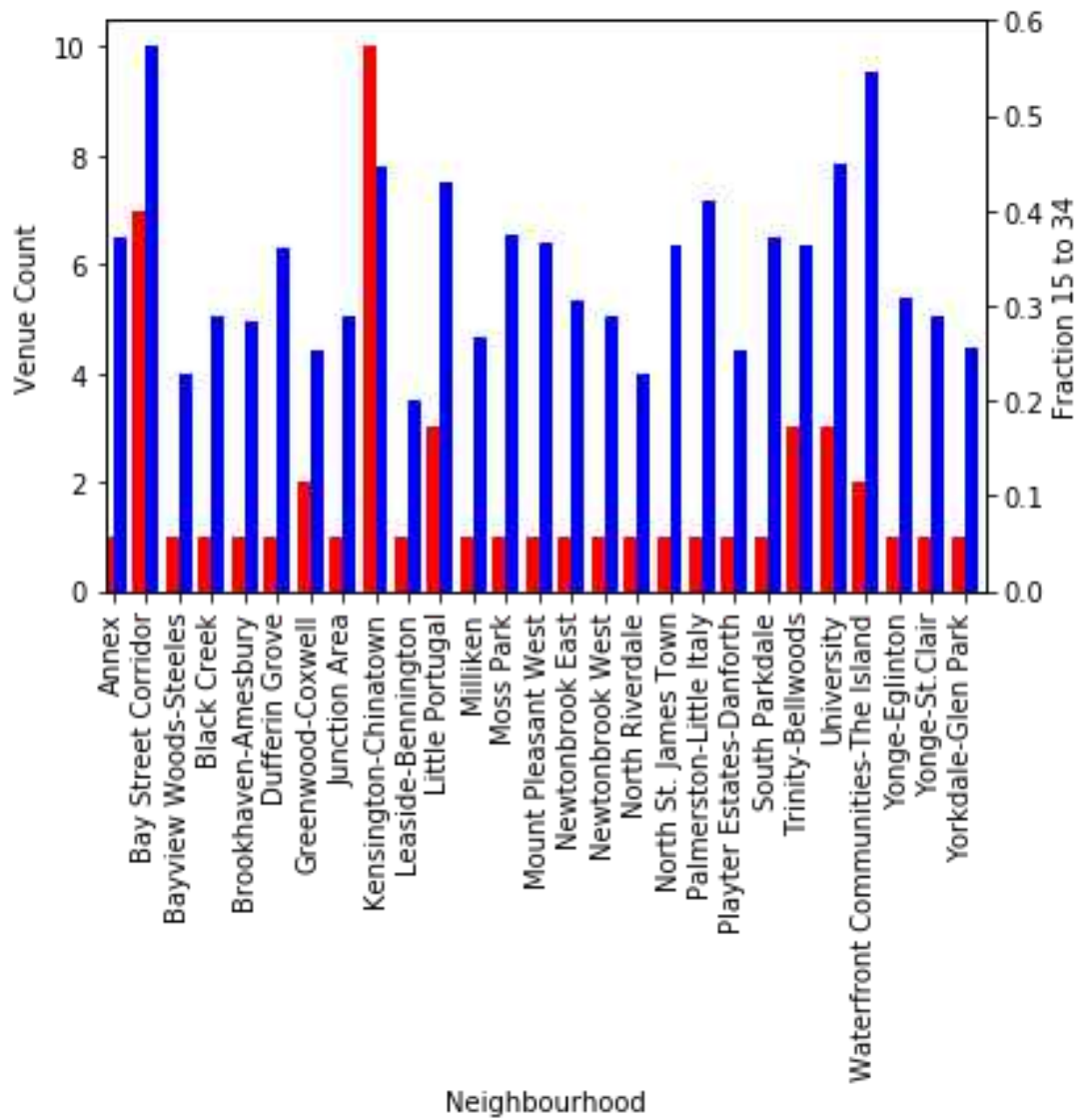
Figure 2. The number of venues (red) and the fraction of young people (blue) for neighborhoods with more than one venue.
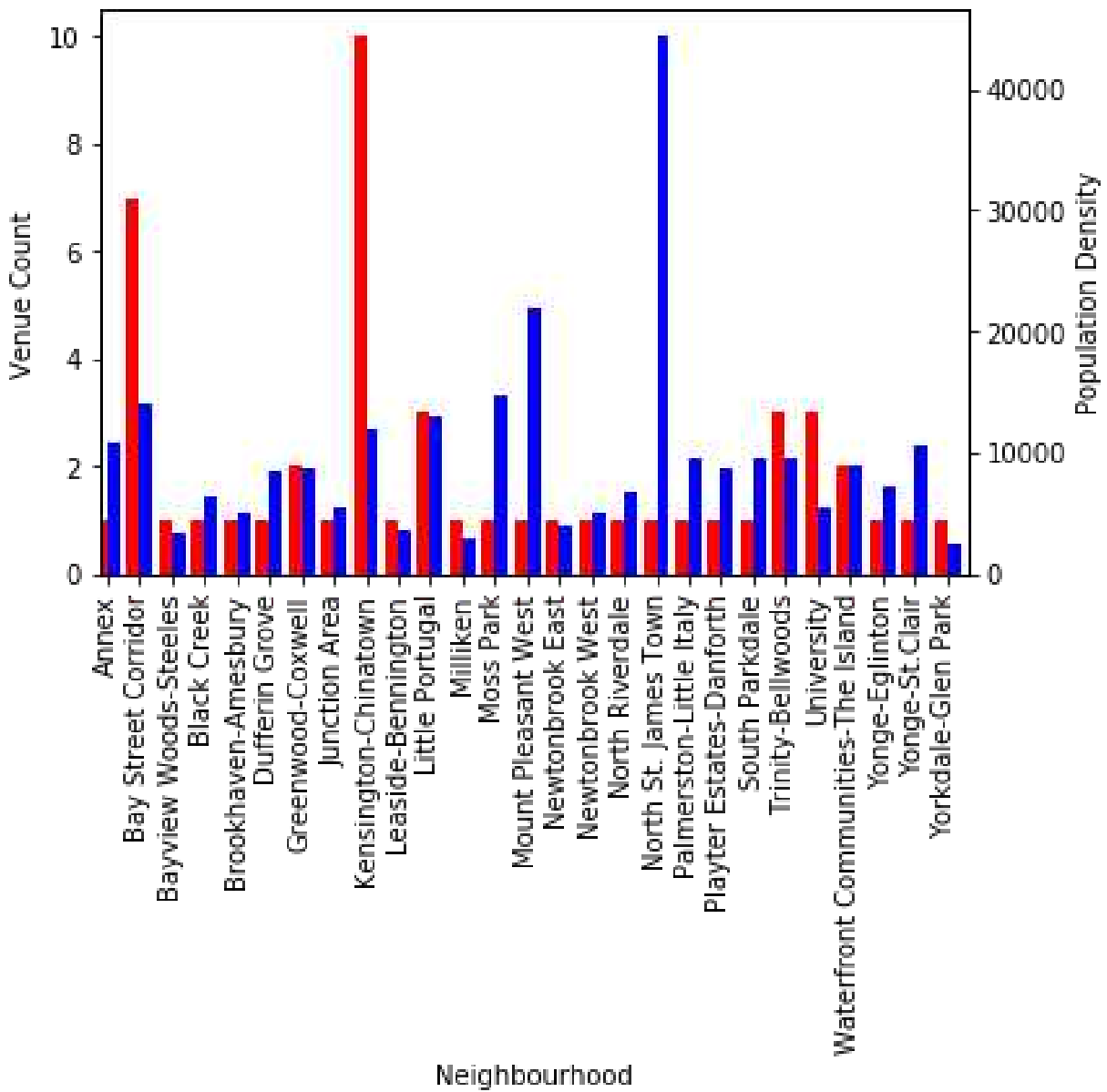
Figure 3. The number of venues (red) and the population density (blue) in neighborhoods with more than one venue.

Figure 2 shows some consistency to the relationship of large numbers of young people and the presence of a vegan or vegetarian establishment. There does not however seem to be a similar relationship in population density from figure 3.

Examination of the descriptions of the two densely populated clusters showed no similarity in terms of the average fraction of people employed, of people with high income levels, or of people receiving public assistance. Both cluster have slightly more women than men and both cluster seem to be racially diverse. In the most densely populated cluster the majority of the population was Asian while in contrast in the second most densely populated cluster the majority of the population is of European origins

Discussion

Due to the large number of young people and high density of vegan restaurants it seems plausible that efforts towards marketing or distribution of animal product substitutes should be targeted towards the neighborhoods in the two clusters most densely populated with the desired restaurants. In general it seems like areas with high population density and a large fraction of young people are the best areas for marketing or distribution of animal product substitutes. However it is unclear what other, if any, relationships might exist between the other variables suggested and the presence of vegans or vegetarians in those areas.

The lack of data on the actual number of vegans or vegetarians living within these neighborhoods makes the drawing of any conclusions about the impact of particular data impossible. However the trend towards racial diversity and the fact that the total number of high income people does not appear to be a factor are both promising leads for future research. It should also be noted that restaurants in densely populated areas may not be catering directly to people who are from that neighborhood but to people outside of it. Again this would require further study.

Conclusion

It is recommended that if one were to market or distribute products that are substitutes for animal products one should do so in the neighborhoods contained in the two clusters most densely populated with vegan and vegetarian restaurants. However it should be noted that no conclusions on where the customers of these restaurants come from, or why they are vegan can be drawn from analysis here. It is recommended that further research on this topic be done. This would likely require data on where customers that eat at these restaurants reside and such data is outside the scope of the data used here. Some information on how commutes are made in each neighborhood is available through the census bureau and may be useful in further study,

Citations

1. Oberst, Lindsay. "Statistics." *The Vegan Society*, www.vegansociety.com/news/media/statistics.

2.Oberst, Lindsay. "Why the Global Rise in Vegan and Plant-Based Eating Isn't A Fad (600% Increase in U.S. Vegans + Other Astounding Stats)." *Food Revolution Network*, 18 Jan. 2018, foodrevolution.org/blog/vegan-statistics-global/.

3.Oberst, Lindsay. "Canada's Surprising New Proposed Health Guide Favors Plant-Based Foods and Ignores Industry Input." *Food Revolution Network,* 21 July 2017, foodrevolution.org/blog/canada-plant-based-food-guide/.

4. Thomson, Aly. "Most Vegans, Vegetarians in Canada Are under 35: Survey." *The Globe and Mail*, 13 Mar. 2018, www.theglobeandmail.com/canada/article-most-vegans-vegetarians-in-canada-are-under-35-survey/.

https://www.toronto.ca/city-government/data-research-maps/open-data/