Using Foursquare location data to determine clusters for potential new business locations

Alvin Jin

Introduction

In Canada, small businesses make up over 97% of the total businesses[1]. Businesses in the service sector are created and fail at greater rates than other industries. Restaurants and bars have been hit especially hard during 2020 due to the COVID-19 pandemic[2]. It is not all bad news and some businesses have thrived during the virus, those related to the pet industry have seen strong gains due to people spending more time in their homes[3]. In 2018 it was estimated that 41% of households in Canada owned a dog, with the population of dogs up to 8.2 million from 7.6 million in 2016[4], a number that will likely be even higher when the data for 2020 is tallied.

The closing of bars and increase in pet ownership suggest that once restrictions are lifted, there may be a potential new market for a new business that combines features of the two, a bar and dog park. With the high rates of turnover in service sector businesses, it is important to take as many steps as possible to ensure that any new business is in a position to survive. A good location is one thing that is essential for the success of any business. This project will utilize data taken from Foursquare to determine potential areas around Toronto that may be a good location for a bar and dog park.

Data

Data was collected from two different sources. Postal code and neighborhood location names for Toronto were scraped from the Wikipedia page displaying all postal codes in Canada starting with the letter M. Venue name, category, and venue location information were extracted from the Foursquare developer API. Venue data was limited to 100 entries per neighborhood due to Foursquare limits.

Methodology

Data was prepared by limiting venue categories to those related to alcohol (Pub, Beer

Store, Wine Shop, Beer Bar, Bar, Wine Bar, Cocktail Bar, Irish Pub, Liquor Store, Sports Bar,

Brewery, Hotel Bar, Gay Bar, Sake Bar), pets (Pet Store, Dog Run), or outdoor activities (Park,

Lake, Field, Trail, Beach, River). The elbow method was used to determine the ideal number of

clusters. Within-cluster sum of squares (WSS) from one to ten was used to create a graph, and

the optimal value of k was determined by examination of the graph. K-means clustering was then

conducted using the results of the previous graph to identify cluster centers of related businesses,

providing potential locations for a bar and dog park.

Results

A total number of 215 venues were selected based on the previously listed criteria. The

graph below shows the WSS of k from one to ten. Based on the graph, the slope reduces at a k of

six and that was selected as the optimal number of clusters.
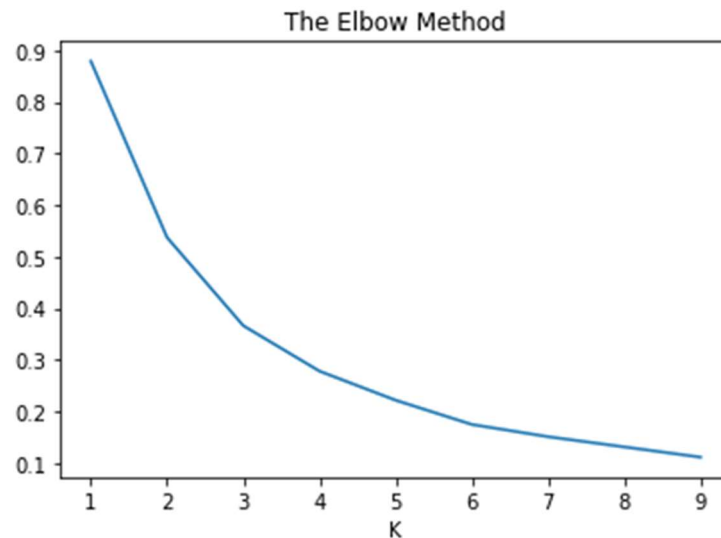
The six cluster centers were found to be at the following latitude and longitude. These points were then graphed on a map of Toronto using Folium.

Discussion

Unsupervised machine learning was able to identify six potential clusters that evenly distribute venues that related to alcohol, pets, and outdoor activities. These can be starting points to determine where a potential new business of serving alcohol and dog park could be located. This is just a first step and more information could be added to the model. For one, different weights could be given to the different categories presented in these analyses.

Future studies could also benefit from adding additional information, such as neighborhood income and wealth, rate of foot traffic through the areas, and even more detailed information about pet ownership based on sales data to improve location identification.

Conclusion

Most small businesses fail and it is important to have every advantage when starting a new business. Identifying potential locations to start the business is just one of many ways that machine learning can help improve the chance of success.

References

1. https://www.ic.gc.ca/eic/site/061.nsf/eng/h_03126.html

2. https://nowtoronto.com/food-and-drink/toronto-restaurants-closed-2020-during-covid-19

3. https://www.retail-insider.com/retail-insider/2021/01/pet-focused-retailers-in-canada-see-sales-gains-during-pandemic-amid-work-from-home/

4. https://www.cahi-icsa.ca/press-releases/latest-canadian-pet-population-figures-released