1. Introduction

In the US and Canada, the rate of dog ownership is increasing at impressive rates. From 2016 to 2018 dog ownership in Canada increased from 7.6 million to 8.3 million, about a 10% increase in dog ownership. An increase in dog ownership allows for new business models to enter the market. The pet grooming and boarding industry has almost doubled over the past decade. Additionally, pet parents have been purchasing a greater range of high value services for their pets. When opening a new business, however, the old saying still holds true. Location, location, location. For a new business to be successful, it needs to be in an area that is convenient for the target market. A machine learning model can help to predict which neighborhoods most likely have dogs. By evaluating proximity to dog runs, parks, and restaurants, we can maximize the return on investment by ensuring high foot traffic.

Problem

A new business owner wants to open a dog grooming and boarding business in Toronto, Ontario. They wish to know which neighborhoods are close to parks, restaurants, and dog runs. By placing their new business is this area, they will maximize their foot traffic and gain a larger share of the market.

2. Data Acquisition and cleaning

To dataset will come from two main datasources. The first data source is scraped from the Wikipedia article defining Toronto neighborhoods, boroughs, and postal codes. The website used for scraping is https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

After the table is scraped, it will be cleaned and formatted. The formatted dataframe is shown in Table 1.

	Borough	Neighborhood	PostalCode
0	North York	Parkwoods	M3A
1	North York	Victoria Village	M4A
2	Downtown Toronto	Harbourfront	M5A
3	Downtown Toronto	Regent Park	M5A
4	North York	Lawrence Heights	M6A

Table 1. Formatted data frame from web scraping

A csv file containing the latitude and longitude coordinates for each postal code will be joined the dataframe. The results are shown in Table2.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Table 2: Data frame showing neighborhood, borough and postal code with latitude and longitude coordinates.

The second dataset comes from the Foursquare API. The foursquare API can be used to map the venues in the area to the Toronto Neighborhoods.

3. Methodology

From the foursquare API, the venues in locations are mapped to the dataframe (table 3)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant

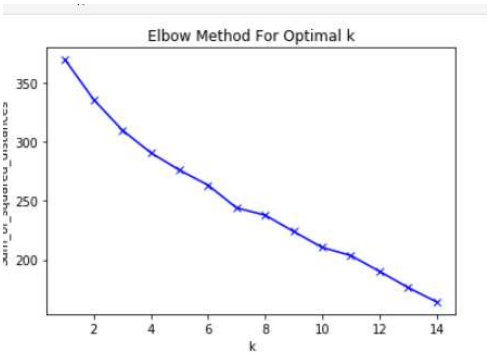
Table 3. All the venues mapped to neighborhoods with venue categories.

A clustering model can be used to determine which neighborhoods are closest to parks, restaurants, and dog runs. The features will need to be modified from a string to a binary classifier to cluster the neighborhoods. The one hot encoding technique can be used, and the results are shown below (Table 4).

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Toy / Game Store	Trail	T S
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	1	0
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	...	0	0	0

Table 4. A dataframe with the venue features to be used for modeling

The elbow method was used to determine the right number of clusters to use for the method. The results can be seen in Figure 1.



The elbow method reveals that the optimal number of k to use is 7. This will cluster the neighborhood into groups of 7, and show which neighborhoods are near the desired venues.

4. Results

K means clustering was used for the model. K Means is a machine learning model for unsupervised learning. This algorithm will categorize like items into groups. The results are shown in Figure 2.

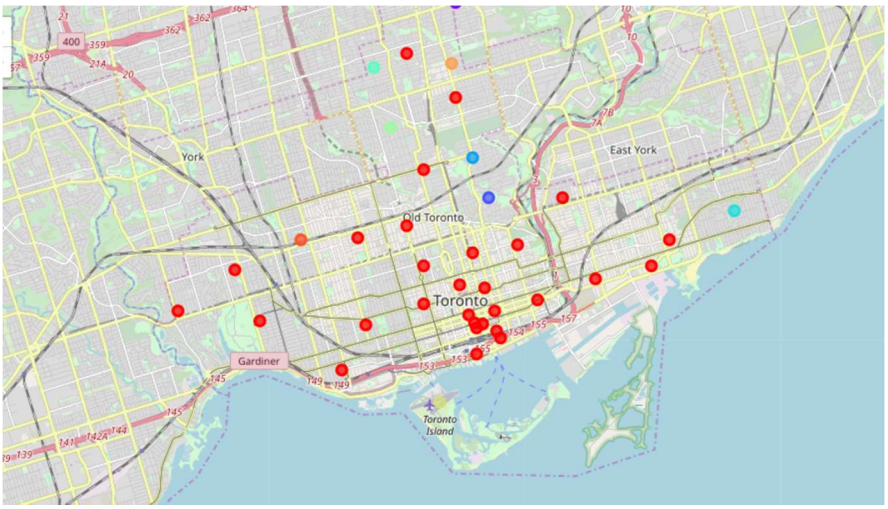


Figure 2. K Means clustering analysis mapped

The neighborhoods in cluster 0, 1, 2 and 6 and are closest to dog walks, parks, and restaurants. Filtering for neighborhoods where the most common venue is parks is summarized in figure 3.

Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
The Beaches West, India Bazaar	43.668999	-79.315572	0	Park	Sandwich Place	Sushi Restaurant	Pet Store	Pizza Place	Movie Theater
Lawrence Park	43.728020	-79.388790	1	Park	Swim School	Bus Line	Yoga Studio	Dog Run	Fast Food Restaurant
Rosedale	43.679563	-79.377529	2	Park	Playground	Trail	Building	Discount Store	Fast Food Restaurant
Cabbagetown, St. James Town	43.667967	-79.367675	0	Park	Restaurant	Coffee Shop	Bakery	Pub	Italian Restaurant
Forest Hill North, Forest Hill West	43.696948	-79.411307	6	Park	Trail	Jewelry Store	Sushi Restaurant	Yoga Studio	Dog Run
Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	0	Coffee Shop	Restaurant	Café	Hotel	Beer Bar	Cocktail Bar

Figure 3. Neighborhoods where the most common venues are parks.

5. Discussion

The results of this study have taken Toronto neighborhoods and clusters them by venues using the 4 square location API. If a person is interested in opening a dog grooming business, the results of this study show they should open the new business in the Foresthill neighborhood or he Lawrence park neighborhood.

6. Conclusion

If a new business owner wanted to open a new dog grooming business in Toronto, the results on this study provide insight to the most common venues in each neighborhood. The next step in this study would be to take into account income level. This can be done by evaluating housing prices in each neighborhood.