Dennis Zaidi

Abstract

The NYC Store Doggy Style NYC, would like to open a store in Toronto, Canada. What is the best location to open a pet store in Toronto, Canada. In terms of customer quantity and quality.

Doggy Style NYC expansion

Leveraging data and machine learning to help a New York pet store, Doggy Style NYC, expand to Toronto

Contents

[1. Introduction 1](#_Toc2626247)

[2. The Data 1](#_Toc2626248)

[3. Methodology 1](#_Toc2626249)

[K Mean Clustering 4](#_Toc2626250)

[Phase 1 4](#_Toc2626251)

[Phase 2 5](#_Toc2626252)

[Discussion 6](#_Toc2626253)

[It got good results 6](#_Toc2626254)

[Improved data needed 6](#_Toc2626255)

[More data needed 6](#_Toc2626256)

[Conclusion 7](#_Toc2626257)

# 1. Introduction

**The business problem:**

The NYC Store Doggy Style NYC, would like to open a store in Toronto, Canada. What is the best location to open a pet store in Toronto, Canada. In terms of customer quantity and quality.

http://www.doggystylenyc.com/

**Who would be interested in this project:**

For 2018, it estimated that $72.13 billion was spent on pets in the U.S.

Expansion into Canada is a massive business opportunity, with Toronto being the largest market in Canada.

By using data, it is hoped that a more informed approach to neighbourhood selection could be utilized.

It is worth noting that this approach could be extended almost any other retail sector.

# 2. The Data

Statistical data on distribution of pets in the city as well as existing potentially competitive retail stores and services are critical factors in determine the suitability of a location for a pet store. Data on distribution of pets and potentially competitive stores is needed.

The following data sets will be utilized:

* The City of Toronto’s Open Data Catalogue for data on pets Toronto by area code
* Wikipedia for a list of Toronto's Areacodes by Borough and Neighbourhood and area code
* Forsquare API to get the store and services in a given Borough and Neighbourhood of Toronto and the store in NYC
* Long and Lat corrdinates by area code provided by <https://cocl.us/Geospatial_data>

All of the datasets are available via url or API.

# 3. Methodology

To select the ideal location for a neighbourhood, we will use two phases, with both leveraging the power of machine learning to help us un our search.

**Phase 1 - Quantity - Clustering the neighbourhoods based on dog and cat populations - i.e. our customers.**

* **Data Needed:** Toronto Open Data, Toronto Wikipedia, Log and Lat Coortinates.
* **Machine Learning Algorithm Used**: K Means - popular and effective for unsupervised clustering.

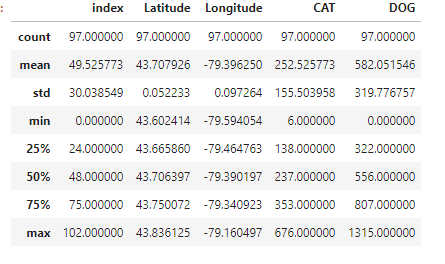
**Phase 2 - Quality - From the best neighbourhoods, selecting the ones that more most like the one our shop is in in NYC - Doggy style knows its neighbours love the store in New York so they want the most similar neighbours in Toronto!**

* **Data Needed:** Foursquare API - neighbourhood data
* **Machine Learning Algorithm Used:** K Means - popular and effective for unsupervised clustering.

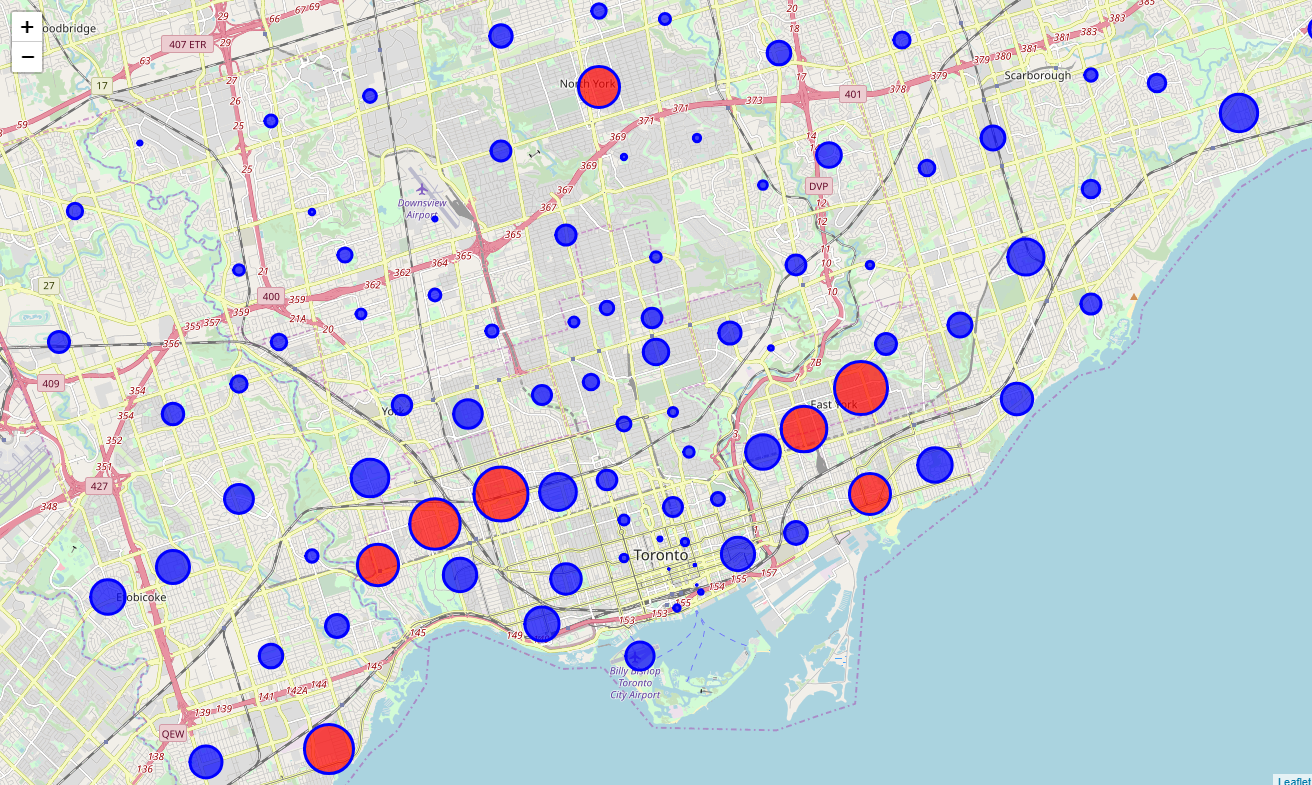
By identifying the neighbourhoods with the most customers and the best customers, Doggy Style NYC can mitigate the risks of opening a new store in a bad location in a new city.

3.1 Importing and Cleaning the Data

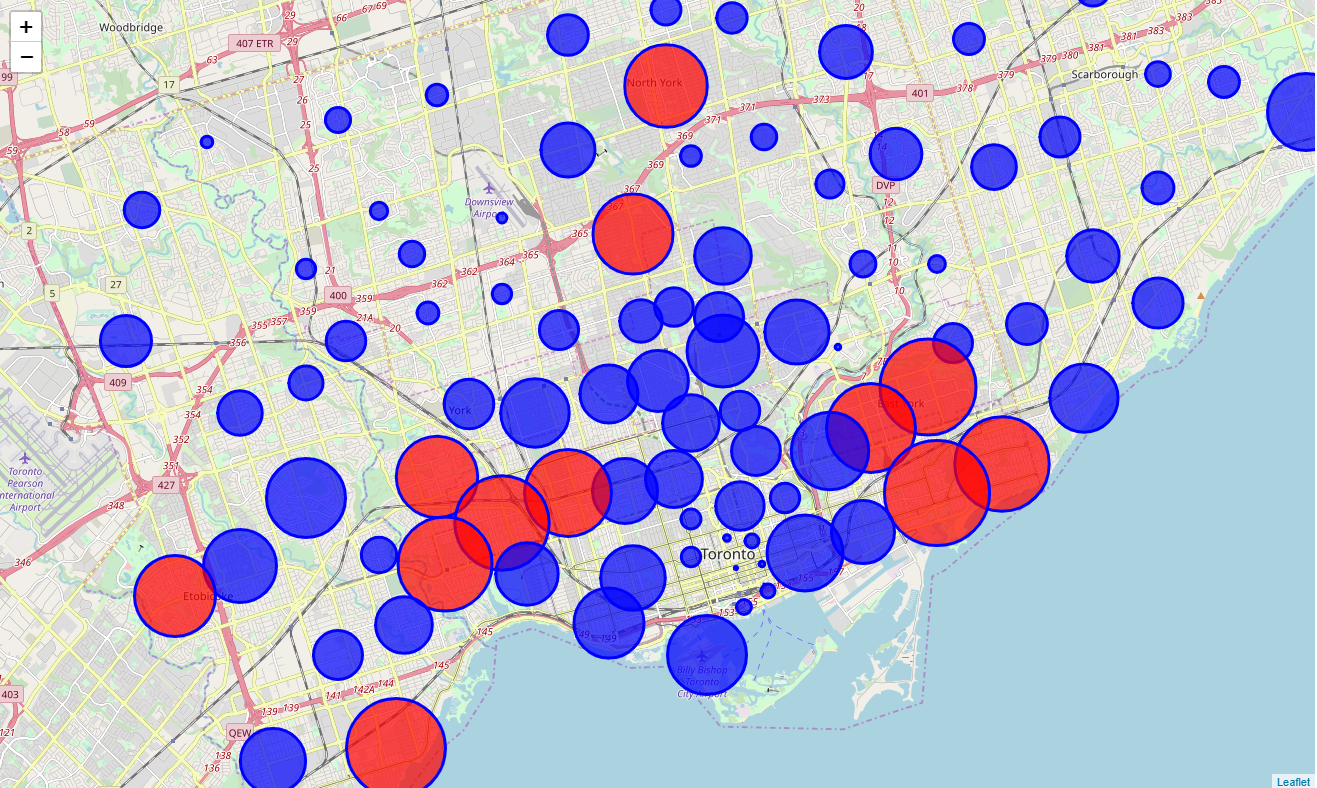
3.2 Understanding the Data



Bubble Plot Distribution of Registered Cats in Toronto



Bubble Plot of Registered Dogs in Toronto



## K-Means Clustering

One of the algorithms that can be used to do cluster segmentation is K-means clustering.

It divides the data into k non-overlapping subsets or clusters without any cluster internal structure or labels.

This means, it's an unsupervised algorithm.

Objects within a cluster are very similar and objects across different clusters are very different or dissimilar.

This approach was applied to the combined data of the cat and dog populations to find the areas similarly highly populated by both dogs and cats.

When applied to the data set it returned the follow results:

To select the ideal location for a neighbourhood, we will use two phases, with both leveraging the power of machine learning to help us un our search.

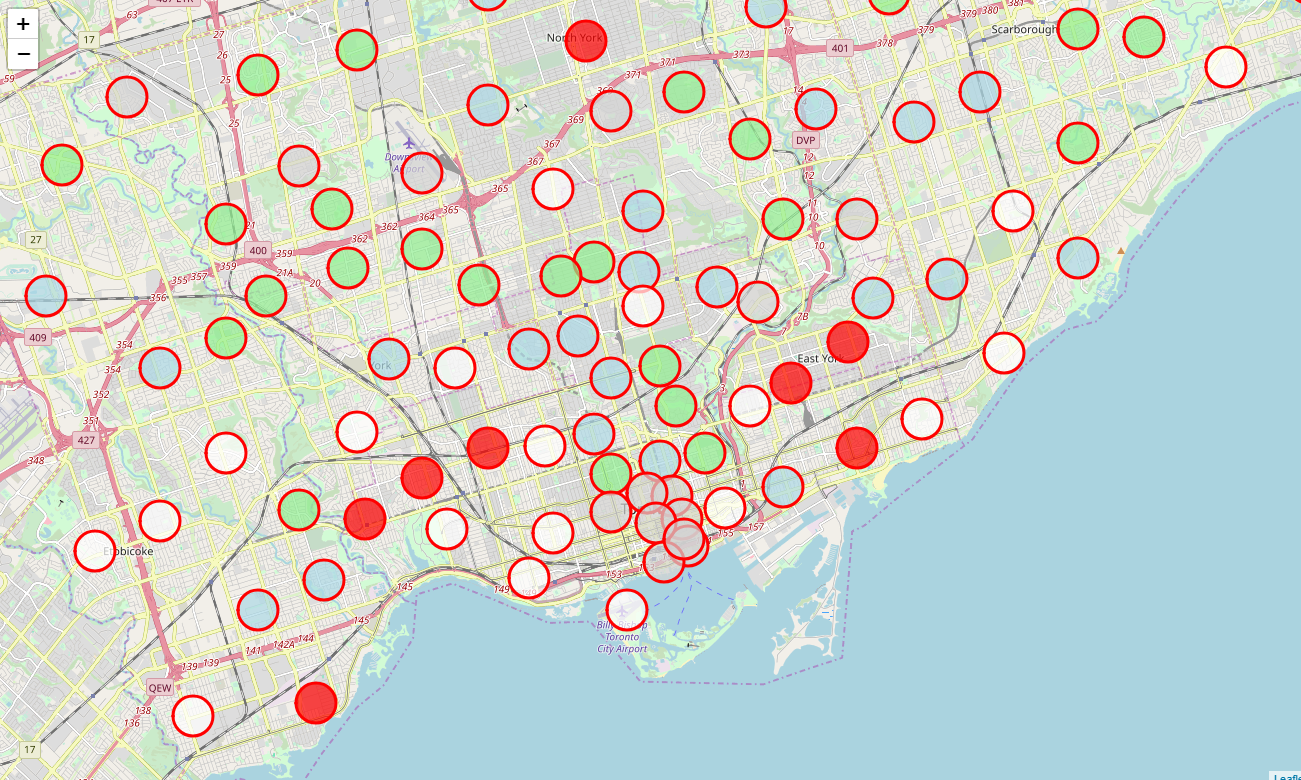
## Phase 1

Quantity - Clustering the neighbourhoods based on dog and cat populations - i.e. our customers.

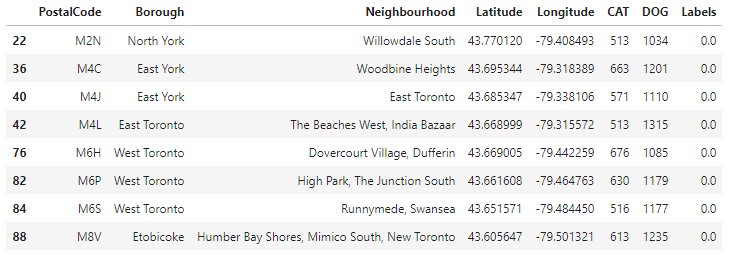
Data Needed : Toronto Open Data, Toronto Wikipedia, Log and Lat Coortinates.

Machine Learning Algorythmn used : K Means - popular and effective for unsupervised clustering.

RED NEIGHBOURHOODS ARE HIGHEST IN DOGS AND CATS.



The following neighbourhood were in the phase 1 cluster:



## Phase 2

Quality - From the best neighbourhoods, selecting the ones that more most like the one our shop is in in NYC - Doggy style knows its neighbours love the store in New York so they want the most similar neighbours in Toronto!

Data Needed : Foursquare API - neighbourhood data

Machine Learning Algorithm used : K-Means - popular and effective for unsupervised clustering.

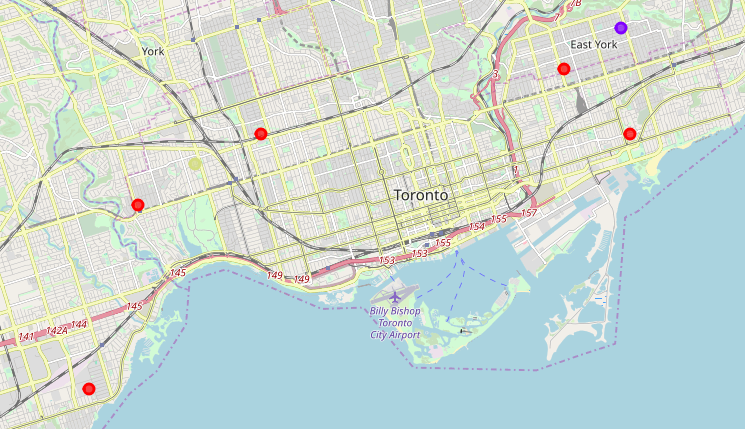
By identifying the neighbourhoods with the most customers and the best customers, Doggy Style NYC can mitigate the risks of opening a new store in a bad location in a new city.

Results:

5 Best locations for the Doggy Style NYC expansion:

* East Toronto
* The Beaches West, India Bazaar
* Dovercourt Village, Dufferin
* Runnymede, Swansea
* Humber Bay Shores, Mimico South, New Toronto

Locations of best neighbourhoods for Pet Store are located in RED



# Discussion

## It got good results

It is worth noting that different quantities for clusters and numbers of features were used but the model, but these numbers seem to produce the best results.

## Improved data needed

This model and dataset could be futher improved by looking at the following data

- competition in the areas from other pet stores

- cost of rental in the areas selected by the model

to improve the viability of each location.

## More data needed

Furthermore, additional features in other datasets not explored by this report could help to further refine the models and produce better results.

# Conclusion

According to the data and the machine learning models used to cluster the data, the 5 Best locations for the Doggy Style NYC expansion are:

- East Toronto

- The Beaches West, India Bazaar

- Dovercourt Village, Dufferin

- Runnymede, Swansea

- Humber Bay Shores, Mimico South, New Toronto