



HOW TO FIND THE BEST LOCATIONS TO NEW VET CLINICS

CASE STUDIED (TORONTO)

Rodolfo Hernandez
IBM - Applied Data Science Capstone

Introduction/Business Problem

According to recent statistics (2017), over **57%** of the Canadian homes had pets (mainly dogs and cats) and from 2016 to 2018 the population of dogs had increased while cat population remains stable. In this context a Company specialized in pets care is searching for **locations to new vet clinics** and considers that **Toronto** could be a good option because of its population and the total number of pets (there are over **230,000 dogs** in the city according to Toronto City's statistics).



Parameters to search the best locations:

- 1) The number of potential clients in each area.
- 2) The distance to existing vet clinics and pet stores (our competitors).
- 3) The visibility of clinics is an important factor too, so we can analyze what kind of neighborhoods are the most suitable for new clinics.

Data Sources

Vet clinics:

Sources:

<https://www.toronto.ca/community-people/animals-pets/pet-licensing/bluepaw-partners/vets-hospitals/>

Vets & Hospitals



Title	Description
Bay Cat and Dog Hospital	<ul style="list-style-type: none">• 20 per cent off first exam with free nail trim• 10 per cent off booked dental• 10 per cent off pet toys

Google Earth (.kml files)



Pet stores:

Sources: FOURSQUARE

	Name	Latitude	Longitude
0	Big Al's Pet Supercentre	43.759279	-79.278325
1	PetSmart	43.769139	-79.412522
2	East York Animal Clinic	43.705921	-79.312196
3	PetSmart	43.712682	-79.362636
4	Pet Valu	43.666979	-79.314665

Pets 'population:

Source: <https://open.toronto.ca/dataset/licensed-dogs-and-cats-reports/>

The screenshot shows the City of Toronto Open Data Portal. The URL is https://open.toronto.ca/dataset/licensed-dogs-and-cats-reports/. The page title is "About Licensed Dogs and Cats Reports". It includes a "Data quality score" section (Not available), a "Data last refreshed" section (Dec 31, 2017), and a "Refreshed" section (Annually). There are also sections for "Data type" (Document), "Topics" (Community services), and "Community services". A "DATA PREVIEW" section indicates "Not available for this dataset".

Data Sources

Population / demography:

Sources:

<https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/comprehensive.cfm>

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

Neighbourhoods' population: from the "Statistics Canada" website we can download datasets (in .csv or .tab format) with population and other demographic facts organized by geographical areas.

The screenshot shows the Statistics Canada website for the 2016 Census. At the top, there are links for Subjects, Data, Analysis, Reference, Geography, Census, Surveys and statistical programs, About StatCan, and Canada.ca. A search bar is also present. Below the header, a breadcrumb navigation shows: Home > Census Program > Data products, 2016 Census > Highlight Tables > Population and Dwelling Counts > Comprehensive download files. The main content area is titled "Population and Dwelling Count Highlight Tables, 2016 Census". Under "Download complete geographic level", there are three sections: "Geographic level" (listing Canada, provinces and territories, Census divisions, and Census subdivisions), "File format" (CSV and TAB buttons), and a "Topics" section. The "Topics" section includes "Data quality score" (Bronze), "Data last refreshed" (Oct 7, 2019), "Refreshed" (As available), "Data type" (Document), "Civic issues" (Affordable housing, Poverty reduction, Mobility, Climate change), and "Topics".

The screenshot shows the City of Toronto Open Data Portal. At the top, there is a logo for TORONTO, followed by "OPEN DATA", "Data Catalogue", "Knowledge Centre", "About", "Gallery", and "Contact". A search bar is also present. Below the header, a breadcrumb navigation shows: OPEN DATA PORTAL HOME / OPEN DATA CATALOGUE / NEIGHBOURHOOD PROFILES. The main content area is titled "About Neighbourhood Profiles". It includes a "Details" section with "Data quality score" (Bronze), "Data last refreshed" (Oct 7, 2019), "Refreshed" (As available), "Data type" (Document), "Civic issues" (Affordable housing, Poverty reduction, Mobility, Climate change), and "Topics". The text explains that the Census of Population is held across Canada every 5 years and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labour. City of Toronto Neighbourhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighbourhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighbourhood. A link to the "Neighbourhood Profiles webpage" is provided.

Methodology-Data acquisition & cleaning



VET CLINICS

- 1) Web scraping from Toronto city's webpage (using BeautifulSoup & Regex libraries).
- 2) Get .kml files from Google Earth searching for vet clinics in the surroundings (to complete the initial dataset). We require Geopandas library and drivers to read .kml files.
- 3) Cleaning data (we have duplicated values because of the use of combined sources).



PET STORES

- 1) Searching for related terms in Foursquare (detected 'Pet Stores').
- 2) Get available venues related to 'Pet stores' in Toronto and surroundings (additional query in the surroundings limiting the radius searching limit).
- 3) Cleaning & selecting useful data (name & coordinates).



PET POPULATION

- 1) Available a .xlsx file from Toronto Open Data (contains data about licensed pets in 2016).
- 2) Used pandas.read_Excel function.
- 3) Cleaning & selecting useful data (Postal Code, number of licensed dog, number of licensed cats and total per neighborhood).

Methodology-Data acquisition & cleaning



POPULATION

- 1) Available a .CSV file from Statcan (Statistics Canada). Contains data about population (2016) and additional demographic facts for all regions of Canada.
- 2) Merged Canada's population dataframe with Toronto's neighborhoods dataset to extract valuable data.
- 3) Cleaning & selecting useful data (Population and neighborhood identifiers).



OTHER DATASETS

- 1) Basic neighborhood data (postal codes, names, coordinates) using web scraping (Source: Wikipedia).
- 2) Cleaning & selecting useful data (postal codes, neighborhoods names & coordinates).

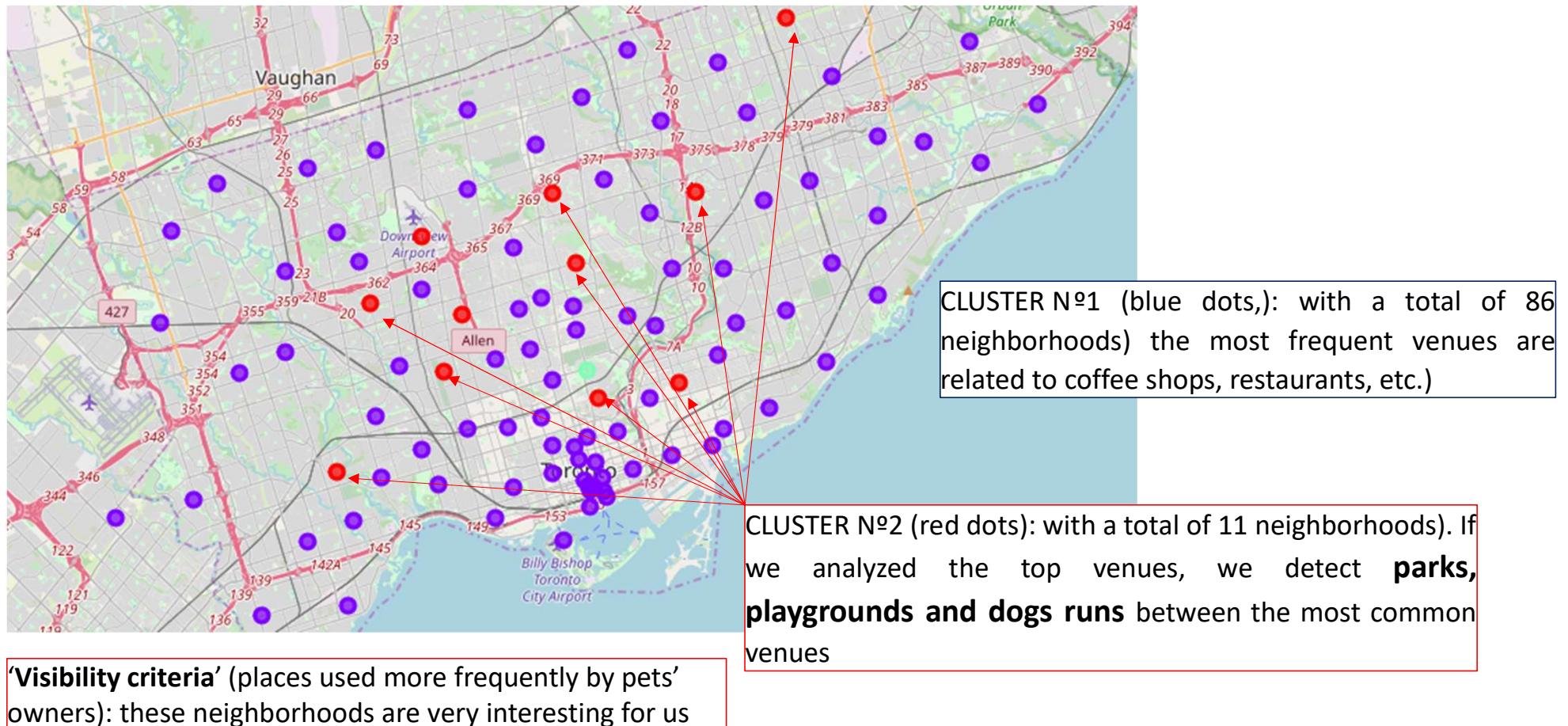


VENUES DATASET

- 1) From Foursquare we can download venues data.
- 2) Cleaning & selecting useful data (neighborhood, venues names, venue category, venue coordinates).

Methodology-Clustering analysis

CLUSTERING ANALYSIS USING VENUE INFORMATION IN EACH NEIGHBORHOOD



Methodology - Descriptive statistics (ratios)

Pets_ratio

$$\frac{\text{Total number of pets} * 1000}{\text{Total number of inhabitants}}$$

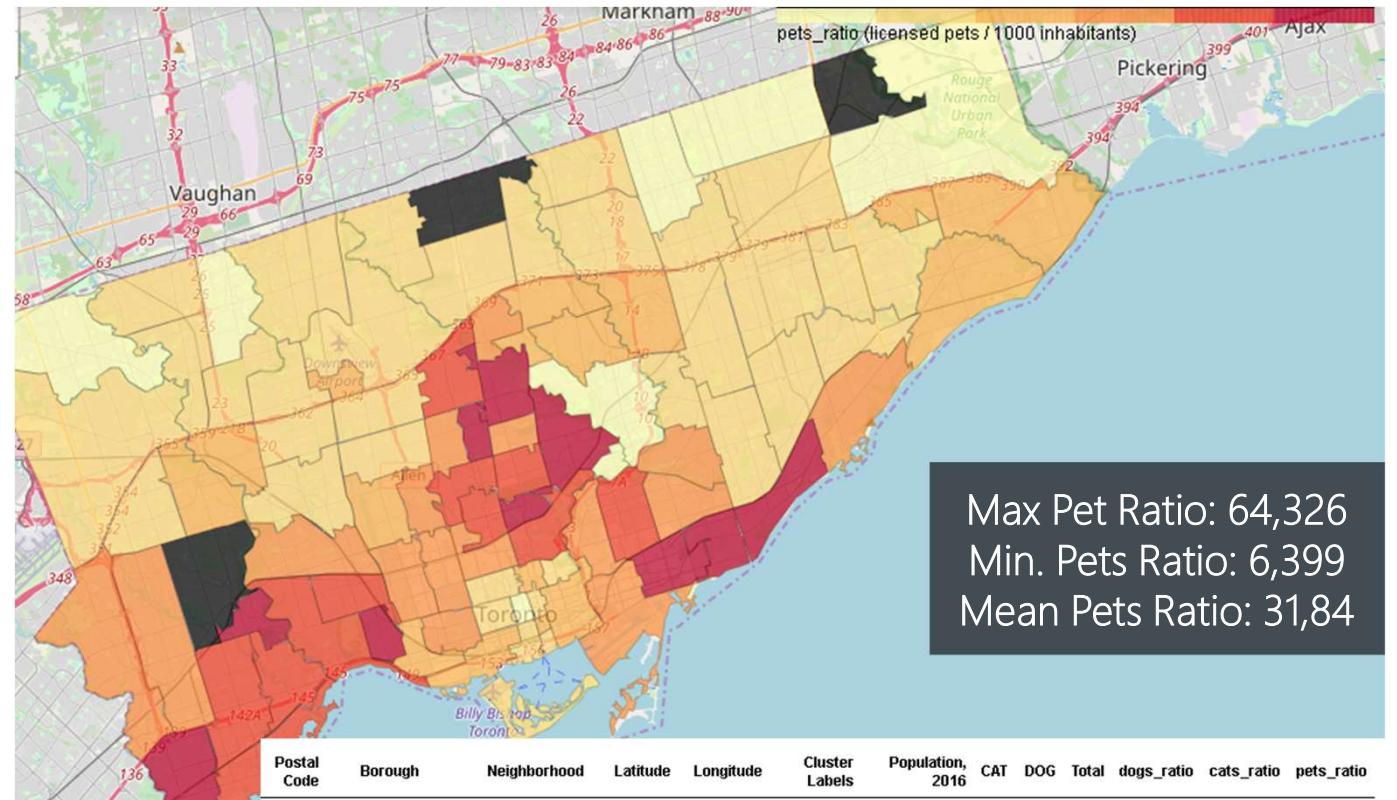
Dogs_ratio

$$\frac{\text{Total number of dogs} * 1000}{\text{Total number of inhabitants}}$$

Cats_ratio

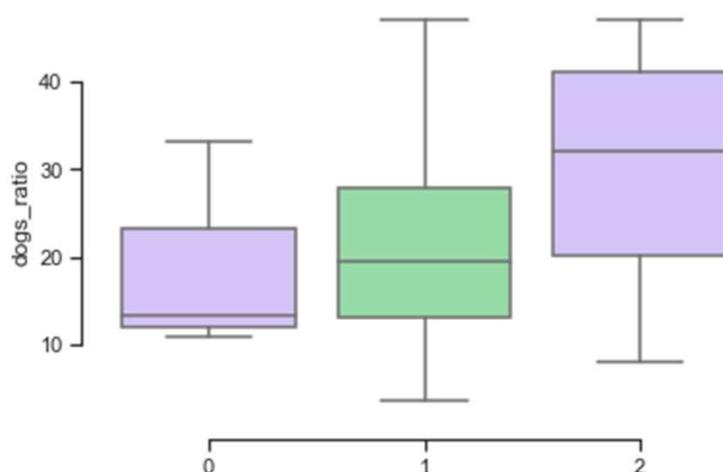
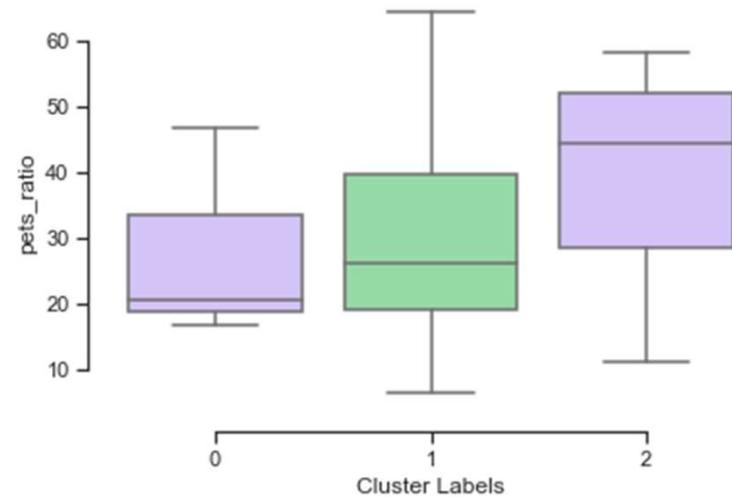
$$\frac{\text{Total number of cats} * 1000}{\text{Total number of inhabitants}}$$

According to statistics: DOGS
NUMBER >2*CATS NUMBER in
Toronto neighborhoods!



Max Pet Ratio: 64,326
Min. Pets Ratio: 6,399
Mean Pets Ratio: 31,84

Methodology - Descriptive statistics (ratios)



These graphs are quite interesting: in neighborhoods included in cluster n°2 (residential areas with parks, dog runs, etc. in the top venues positions), the pets_general ratio is higher than other clusters.

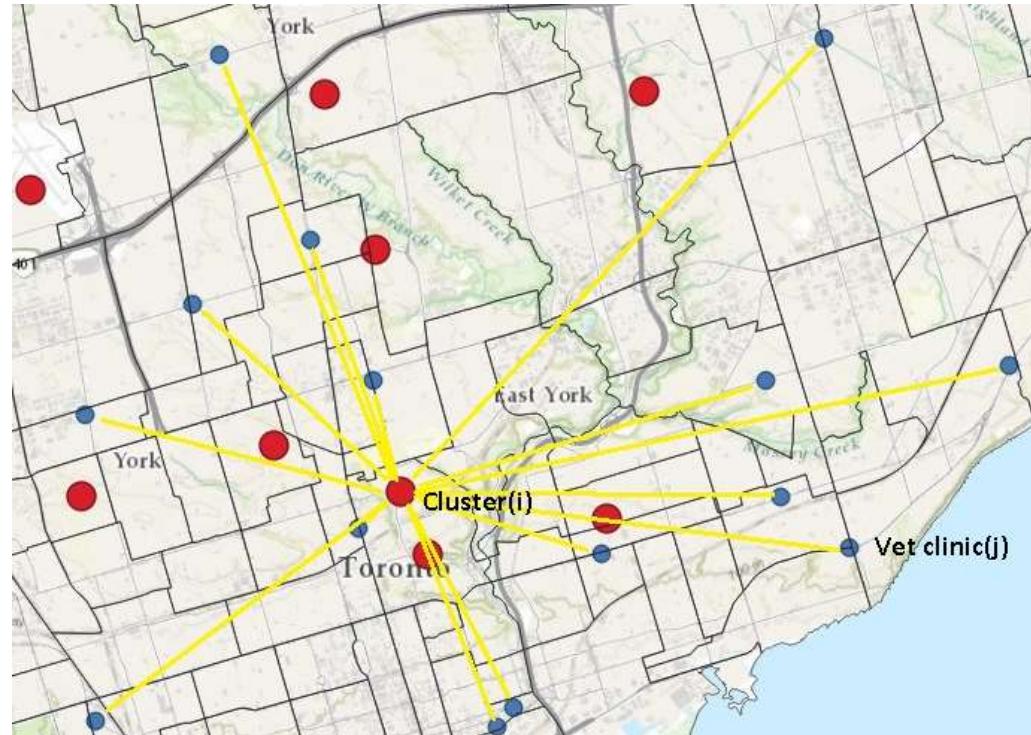
We can observe that the gap between clusters if we choose the dogs_ratio is bigger than general_pets_ratio. This could be interesting for our analysis because the dog's vet spending is higher than cats (according to statistics).

So, the neighborhoods included in the cluster n°2 should be carefully analyzed to our business.

Methodology - Descriptive statistics (distance analysis)

First, we'll calculate the distance between the clusters' position and existing vet clinics (using Geopy library). We are going to consider just the geographical distance (it's a simpler approach than the 'real' distance taking account transport networks but it can be a good reference too).

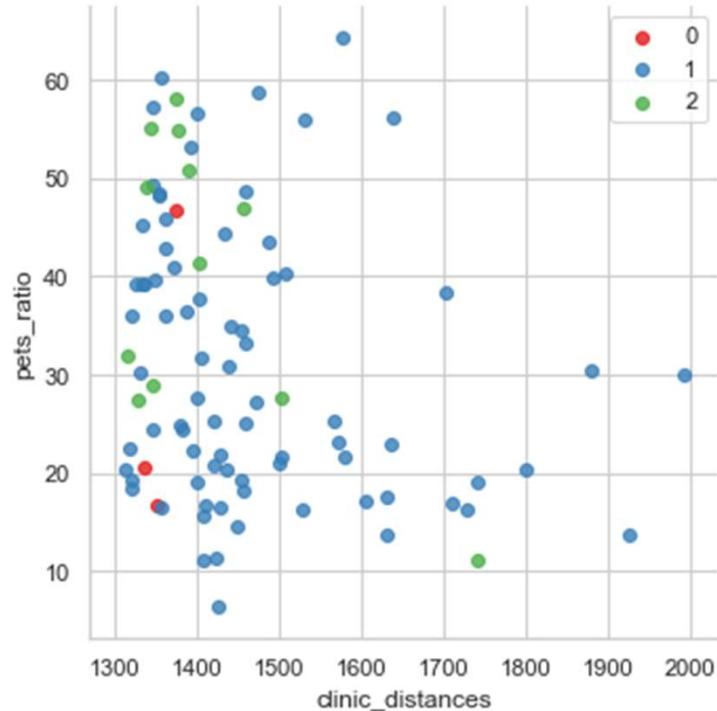
-For each cluster we obtain the accumulated distance respect to all clinics (in kilometers) according to the following scheme:



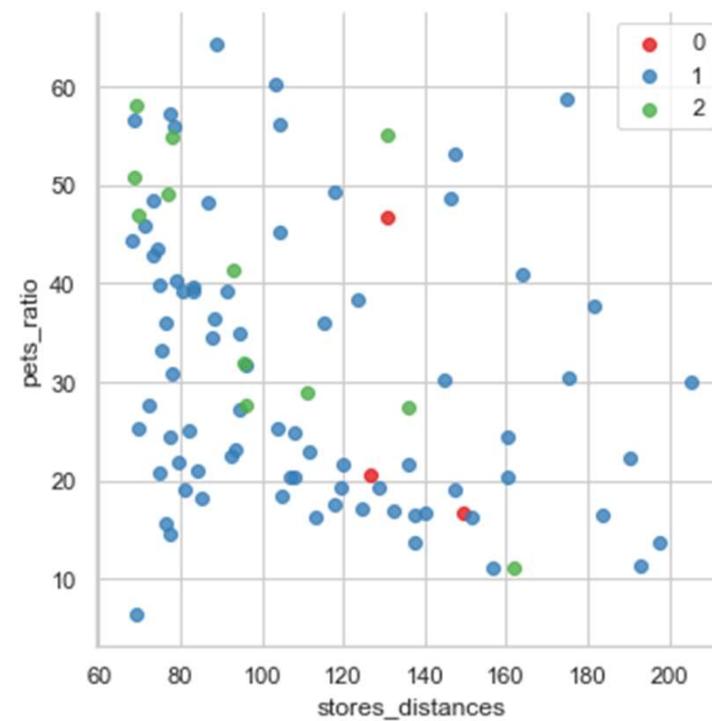
$$distance_{cluster(i)} = \sum_{j=0}^{n_{vets_clinics}} distance (cluster(i) to vet_{clinic(j)})$$

Methodology - Descriptive statistics (distance analysis)

Highest pets_ratios are the most covered by existing vet clinics, although we can observe some remarkable exceptions (see the upper and lower right quadrant of the graph) which can be interesting placements to new clinics.

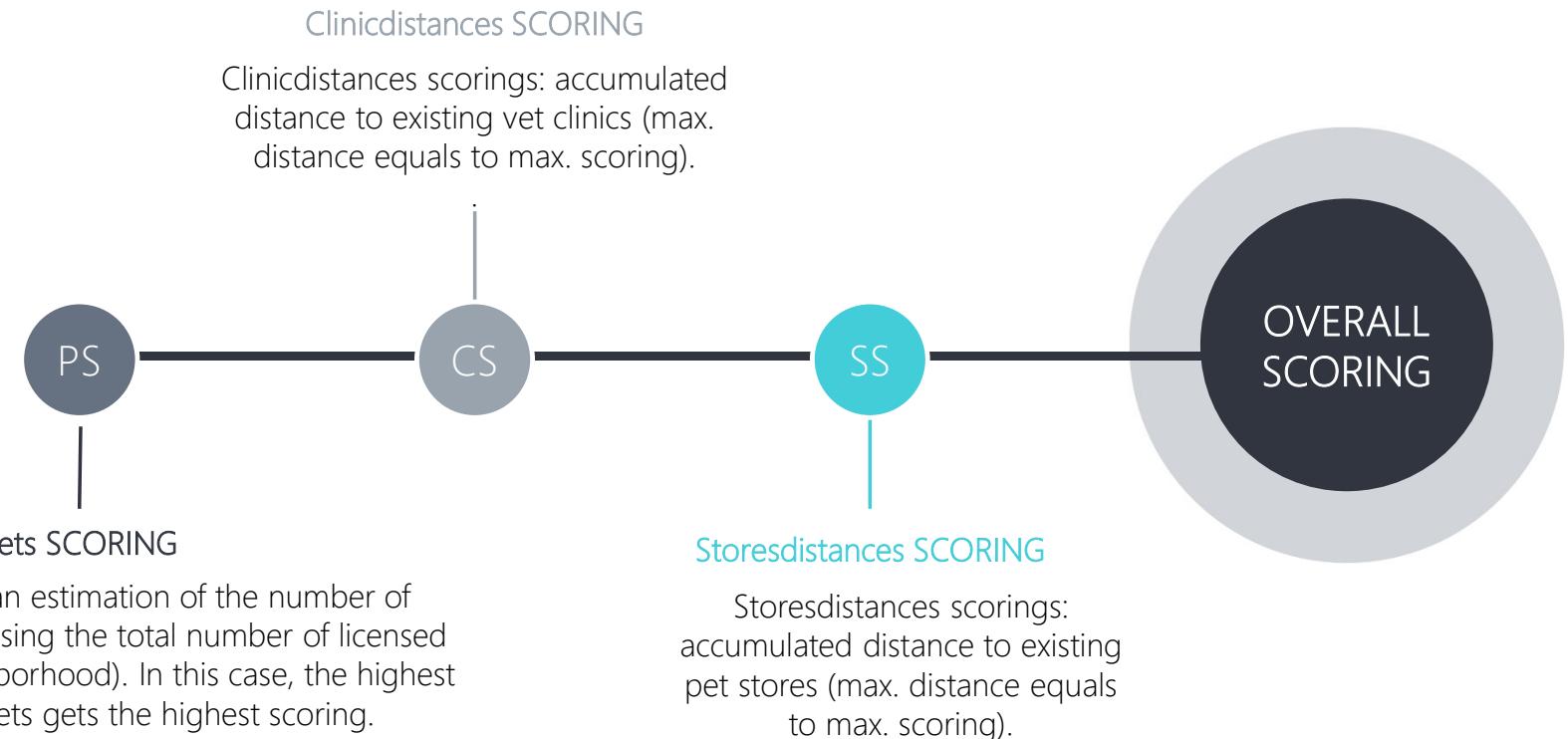


Analyzing pet_stores distances respect to clusters we can reach similar conclusions as vet_clinic_distances: highest pets_ratios clusters are the best covered by existing pet stores.



A scoring approach to localize the best placements

$$Overall_{scoring} = Pets_{scoring} * Clinicdistances_{scoring} * Storesdistances_{scoring}$$



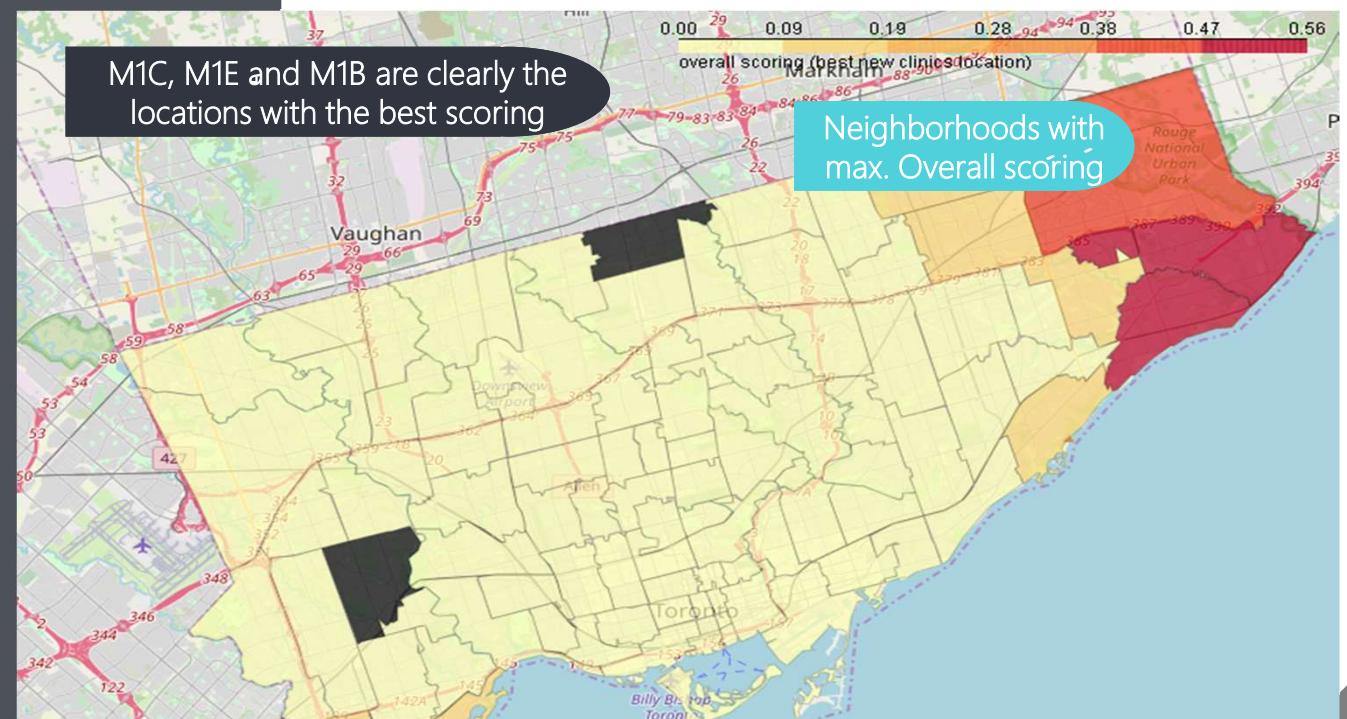
In all cases, we'll standardize the parameters (we'll use a max-min approach), so all factors will work at the same scale (between 0 and 1).

RESULTS

Applying our scoring scheme, we can obtain a sorted dataframe with the best locations for new vet clinics in Toronto (sorted by overall_scoring):

The map shows us an interesting aspect: the best locations are concentrated in a specific area (the east of Toronto) so new clinics can cover a wider area with interesting features (high number of potential clients in a closer radius of influence).

Postal Code	Neighborhood	Cluster Labels	pets_scoring	clinic_distances_scoring	stores_distances_scoring	overall_scoring
M1C	Highland Creek, Rouge Hill, Port Union	1	0.564835	1.000000	1.000000	0.564835
M1E	Guildwood, Morningside, West Hill	1	0.761538	0.833647	0.781212	0.495956
M1B	Rouge, Malvern	1	0.476923	0.904417	0.943867	0.407125
M1G	Woburn	1	0.308242	0.719094	0.672445	0.149051
M1V	Agincourt North, L'Amoreaux East, Milliken, St...	2	0.313736	0.630588	0.682561	0.135037
M8V	Humber Bay Shores, Mimico South, New Toronto	1	0.991209	0.216960	0.569518	0.122477
M8W	Alderwood, Long Branch	1	0.643407	0.237976	0.776452	0.118887
M1S	Agincourt	1	0.313736	0.612370	0.609683	0.117134
M1M	Cliffcrest, Cliffside, Scarborough Village West	1	0.457692	0.574855	0.404906	0.106533
M1K	East Birchmount Park, Ionview, Kennedy Park	1	0.584066	0.476412	0.317487	0.088343



CONCLUSIONS

Procedure to search for the best locations for new vet clinics in Toronto, using information from open data sources and specialized API's such as FOURSQUARE. The starting point of the method is a clustering analysis which it's useful to identify what types of neighborhoods can we find. After, through a combination of different ratios (such as number of pets per neighborhood and accumulated distances from existing vet care venues to the different areas) we've set up a **scoring system** to classify the neighborhoods according to their potential to locate new clinics.

- Proposal of three specific locations situated in the east of Toronto according to our estimation method.
- Our approach it's quite flexible and can be adapted to other areas of study (new cities, also new types of businesses). Besides, we can modify it easily to change our target (for instance, analyze existing vet clinics to invest instead of opening new clinics).