

Il Cane's Camion

IBM Applied Data Science Capstone Project

Jeanna Schoonmaker



March 2020

Table of Contents

Table of Figures.....	2
Abstract.....	3
Introduction	4
Methodology.....	5
Results.....	7
Discussion.....	13
Conclusion.....	14
Works Cited.....	15

Table of Figures

Figure 1 - Five Largest US Cities as scraped from wikipedia	7
Figure 2 - Table of pet owner percentage and normalized population density	8
Figure 3 - Bar charts of pet percentage and population density	8
Figure 4 - Pet store data obtained from Foursquare	9
Figure 5 - Initial Chicago map of all dog-friendly locations	10
Figure 6 - Elbow curve visualization	11
Figure 7 - Lat/long for cluster of each center, and sample of labeled dataset after KMeans	11
Figure 8 - Cluster map	12

Abstract

A hypothetical business team uses US Census Bureau data on population density and pet ownership to choose a major US city in which to pursue opening a food truck aimed at four-legged clientele. Chicago, IL is chosen, and pet-friendly neighborhood locations are identified and mapped. The KMeans machine learning algorithm is utilized to find clusters with a high concentration of dog amenities in downtown Chicago. The center of the clusters are identified by latitude and longitude, as well as mapped, indicating suggested operating locations based on being central to many other dog-friendly locations.

Introduction

According to the American Pet Products Association, spending on pet-related expenditures in the United States was predicted to increase to 75 billion dollars in 2019 (American Pet Products). However, pets and pet ownership are not just big business - pets were regarded as members of the family by 95% of pet owners in a 2015 Harris Poll. The same poll showed that 71% of the pet owning respondents owned dogs, and that 68% of pet owners reported their household income as 100k+ (The Harris Poll).

Another trend in the past decade: food trucks. The “mobile food” industry was estimated to grow to 900 million in 2019, and nearly 25% of food truck operators cited staying ahead of changing trends as a reason for launching a mobile business (Saxe).

To capitalize on this pet-friendly trend, as well as the popularity of food trucks, world famous restaurateurs Fido and Max have decided to launch a new venture: Il Cane's Camion, which is Italian for "The Dog's Truck". This dog-centric food truck plans to cater to the most discriminating of poochy palates by offering fresh, organic doggie treats, made especially for use by families who want to lavish love and attention on their dogs (with an option for further pet products to be included in the future).

In order to consider good business locations for Il Cane's Camion, Fido and Max are looking for a major US city with a high population of pet owners and many dog-friendly venues that would increase the likelihood of local foot traffic. Research will be done on the 5 largest major metropolitan areas within the United States, by population, as possible business sites. The key indicators to be considered are a) the percentage of the city's population who are pet owners, and b) population density, since high foot traffic is considered one of the most important features of food truck success (Murray). Once a city has been chosen, suggested city locations, identified as those centralized in a cluster of businesses catering to dogs, will be mapped as possible daily or weekly parking spots for Il Cane's Camion.

The business questions to be considered are:

1. What is the population density of the 5 largest US cities?
2. What percentage of the population are pet owners in the 5 largest cities (by population) in the United States?

Once a city has been recommended, research will include:

1. Which areas of the chosen city have veterinarians within close proximity?
2. Where are the most dog parks located in the chosen city?
3. Which areas of the chosen city have nearby pet stores and pet services?
4. Taking all of the above into consideration, use machine learning to find several suggestions for Il Cane's Camion to park and operate.

Methodology

Data gathering began with using BeautifulSoup to scrape a Wikipedia site (U.S. Cities by Population) with tabulated population density data for the largest US cities. Extraneous string characters were removed so the population density column would be recognized as numeric and could be normalized for plotting.

After review, consideration was only given to the top five cities by population, as both the overall population as well as the population density dropped precipitously beyond that. Additionally, the United States Census Bureau's American Housing Survey (Census Bureau) only gathered information about pet ownership according to which cities were largest in 2017. Due to the variance in the years when data was collected, researching cities in the top 6-10 range by population would have meant missing pet ownership data for some cities. With only two key factors (population density and pet ownership percentage) being considered, missing data would severely limit analysis, so only the top five cities were researched further.

Once a dataframe was created from the web scraping of population density data, the American Housing Survey Table Creator offered at census.gov was utilized. The table criteria used was:

- Select area: Metro Area: {New York, Los Angeles, Chicago, Houston, or Phoenix}
- Select year: 2017
- Select a table: Emergency and Disaster Preparedness

Each table was modified to eliminate unnecessary rows and columns and calculate a percentage of pet owners based on the total estimated population of the city and the total number of survey respondents who, when asked if an evacuation plan was in place for pets, answered "No pets."

Population density was normalized for exploratory data visualization purposes, then plotted as a bar chart with seaborn, along with percentage of pet ownership. The visualization helped make both the similarities and differences in the datasets clear.

After Chicago was chosen as the city in which the venture would proceed, latitude and longitude was obtained through the Geocod.io API for downtown Chicago (geocod.io).

The Foursquare API was then used, with Chicago latitude and longitude numbers obtained in the previous step, to search for pet and dog-friendly business locations in downtown Chicago. Search terms included a limit of 100 results, along with a radius of 8000 m, or roughly 5 miles. The category ids used to query the Foursquare API were

found in the site's developer docs (Venue Categories), and four categories were found to be relevant: pet stores, pet services, veterinarians, and dog runs/parks. A call to the Foursquare API was made for each category id. The information was returned in json format and reformatted as a dataframe listing the venue's name, category, and latitude and longitude. All four categories were then marked on a folium map of downtown Chicago.

In choosing which machine learning clustering algorithm to use, both DBSCAN and KMeans clustering were considered. With such a small radius of area to consider, and with the locations already being filtered as pet-friendly, minimal 'noise' was found in the data, so KMeans clustering was chosen as the better option.

Once the map was rendered, all dog-friendly business locations were concatenated into one dataframe which was then iterated through to plot an elbow curve graph for use in finding the optimal k value for the KMeans algorithm. When a k of 3 was found, the KMeans algorithm was utilized on the dataset and the list of dog-friendly businesses were separated into 3 clusters. The latitude and longitude of the centers of each cluster were found.

Finally, another folium map was rendered, which shows downtown Chicago and the dog-centric businesses, separated by color into the cluster they are in. The centers of each cluster are also marked, representing suggested locations for Il Cane's Camion to operate.

Results

The first questions to be addressed were:

1. What is the population density of the 5 largest US cities?
2. What percent of the population are pet owners in the 5 largest cities (by population) in the United States?

Once data was obtained from the population table from Wikipedia, the following dataframe was created (Fig. 1):

Figure 1 - Five Largest US Cities as scraped from wikipedia

Out[2]:

	City	2018estimate	2016 population density
Id			
1	New York[d]	8398748	28,317/sq mi
2	Los Angeles	3990456	8,484/sq mi
3	Chicago	2705994	11,900/sq mi
4	Houston[3]	2325502	3,613/sq mi
5	Phoenix	1660272	3,120/sq mi

It is immediately noticeable that the population density is much higher in the top three cities, with Chicago having a population density nearly 30% higher than that of Los Angeles.

The Census Bureau American Housing Survey results provided the data needed to calculate the percentage of pet owners in each of the top five US cities. The table below (Fig. 2) shows that New York's population was estimated to have 33% pet ownership, while Los Angeles' pet owners comprise 41% of the population and Chicago's is at 40%. Over 47% of the population of Houston is estimated to own a pet, with Phoenix topping that at 55% pet ownership among the population.

Figure 2 also shows the normalized population density data, with New York with the highest value of 1.0 and Phoenix with the value of 0.00000.

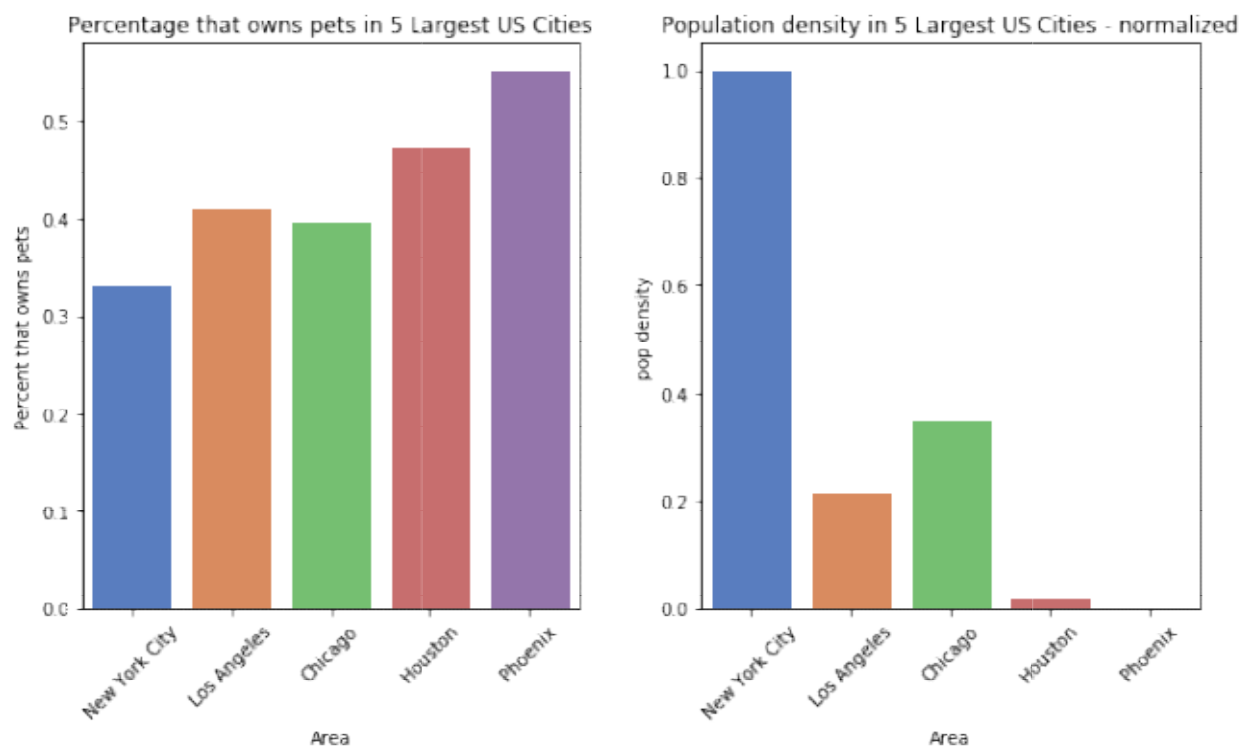
Figure 2 - Table of pet owner percentage and normalized population density

Out[4]:

	Area	Percent that owns pets	pop density
Id			
1	New York City	0.330479	1.000000
2	Los Angeles-Long Beach	0.408922	0.212882
3	Chicago	0.396257	0.348454
4	Houston	0.473508	0.019566
5	Phoenix	0.552741	0.000000

The graph below (Fig. 3) is the result of plotting a bar chart for the data shown in Figure 2. New York's percentage of pet ownership is lowest, while its population density is highest. The opposite is true for Phoenix! Visualizing the data helped rule out New York, with its low pet ownership percentage, and Houston and Phoenix, with their low population density. Between Los Angeles and Chicago, the higher population density combined with similar pet percentage ownership for the population means Chicago will be suggested for the base of operations for Il Cane's Camion!

Figure 3 - Bar charts of pet percentage and population density



Now that a city has been chosen, the remaining questions to be examined are:

1. Which areas of downtown Chicago have veterinarians within close proximity?
2. Where are the most dog parks located in downtown Chicago?
3. Which areas of Chicago have nearby pet stores and pet services?
4. Taking all of the above into consideration, which areas of downtown Chicago would be optimal for Il Cane's Camion to park and operate?

The Foursquare API was used to gather data for locations of veterinarians, dog parks, pet stores, and pet services, and the json response was parsed into data frames like this:

Figure 4 - Pet store data obtained from Foursquare

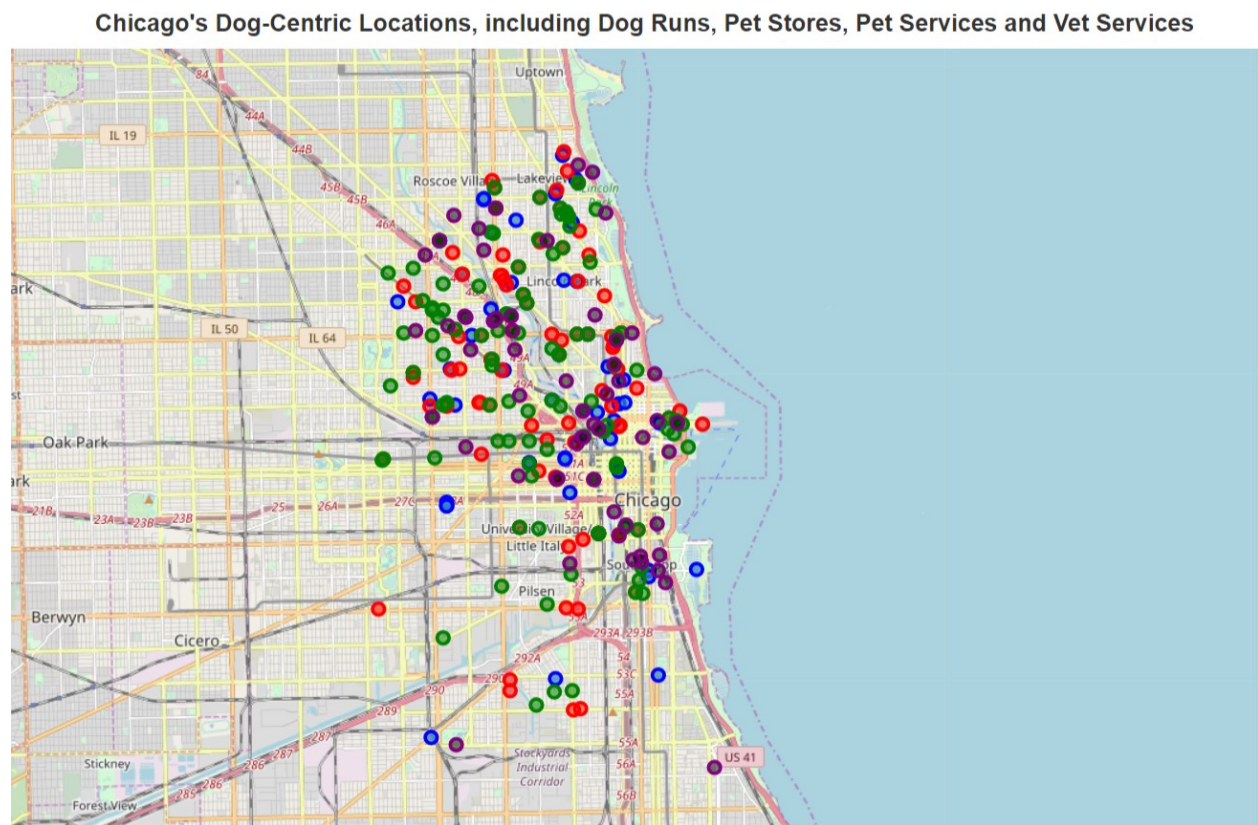
Out[27]:

	Name	Category	Latitude	Longitude
0	Kriser's Natural Pet	Pet Store	41.869137	-87.627229
1	The Anti-Cruelty Society	Pet Store	41.891375	-87.632548
2	Paw Naturals	Pet Store	41.880418	-87.650930
3	Kriser's Natural Pet	Pet Store	41.892732	-87.618315
4	PetSmart	Pet Store	41.906464	-87.650007
5	Old Town Aquarium	Pet Store	41.910252	-87.634857
6	Doggy Style Pet Shop	Pet Store	41.903182	-87.678284
7	Bark N' Bites	Pet Store	41.830842	-87.643930
8	Kriser's Natural Pet	Pet Store	41.911733	-87.679972
9	Petco	Pet Store	41.890087	-87.637159

Though the radius for results was relatively small at only ~5 miles, there were 60 dog runs/parks listed, 100 businesses categorized as pet services, 87 pet stores, and 53 veterinarians. This appears to be ample data for clustering, and is a further confirmation that Chicago has many areas that would be considered “dog friendly” and that should provide ample foot traffic from people taking their dogs to the vet, to the dog park, etc.

To explore the visualization of these locations on a map, Folium was used to map downtown Chicago, with different points for each of the locations in each of the tables. Different colored markers were used to represent the different categories.

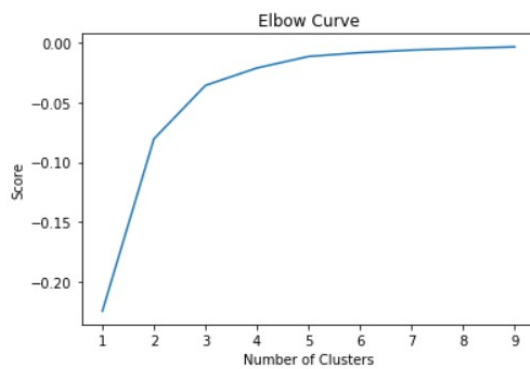
Figure 5 - Initial Chicago map of all dog-friendly locations



At first glance, while it is easy to identify that there are many dog-friendly businesses in the heart of Chicago, it is less obvious if there are any clear clusters, or centers of businesses, where a food truck catering to dogs might find the most success. The pet friendly locations do seem to be more concentrated at the top of the dataset area, though, with more distance variance for locations on the central and south part. It is worth noting, however, that all of these datapoints fit within a ~5 mile radius, so even the points that might look spread out are likely in closer proximity to other dog-friendly locations than they seem at first glance.

Next, all dog-friendly business locations were joined into one dataset, which was iterated through with the KMeans algorithm for each value of k in the range 1-10. This was for the purpose of creating an elbow curve visualization, which is shown below (Fig. 6):

Figure 6 - Elbow curve visualization



On reviewing the elbow curve graph, it was determined that the most obvious curve occurs at $k=3$.

The KMeans clustering algorithm was utilized on the dataset with all of the dog-centric business locations within 5 miles of downtown Chicago using $k=3$ and $k++$ initialization. Each dog-centric business location was assigned a cluster label and the center of each cluster was found and assigned a latitude and longitude, as shown in figure 7.

Figure 7 - Lat/long for cluster of each center, and sample of labeled dataset after KMeans

```
print(centers_df)
x.tail(5)
```

	Latitude	Longitude
0	41.926190	-87.652723
1	41.876270	-87.638224
2	41.905702	-87.680656

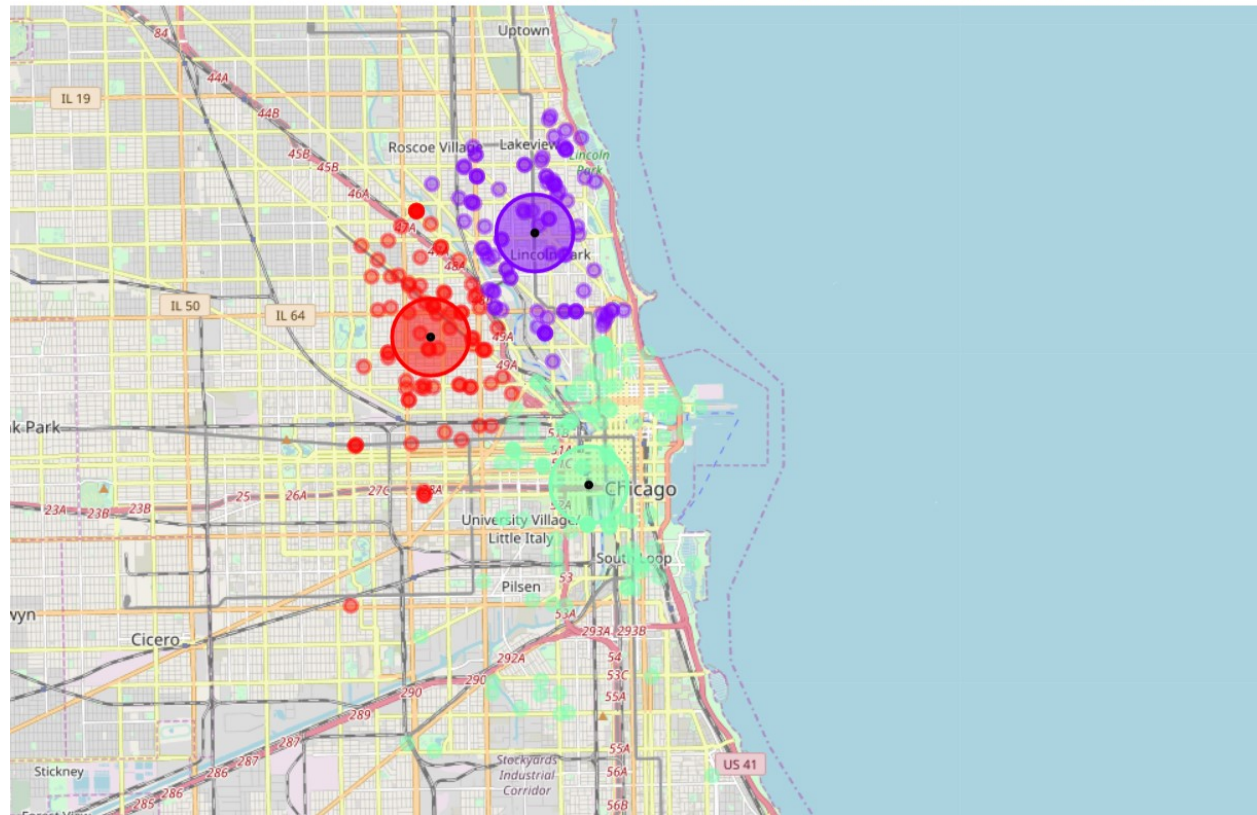
Out[21]:

	Name	Latitude	Longitude	cluster_label
48	Pets First Veterinary Clinic	41.943755	-87.644989	0
49	Roscoe View Animal Clinic	41.939317	-87.671517	0
50	Higgins Animal Clinic	41.939650	-87.671462	0
51	Wrigleyville Veterinary Center	41.948889	-87.649093	0
52	Brighton Park Animal Hospital	41.824741	-87.686716	1

Finally, another Folium map of downtown Chicago was rendered, each cluster with its own color and a black dot representing its center. Each location from the original dataset is also marked and distinguishable by cluster according to color. As was expected after the first map with dog-friendly locations was rendered, two of the three clusters are on the north side of the location covered by the dataset, while the third covers the more widely spread locations on the central and south part of the dataset.

Figure 8 - Cluster map

Possible Il Cane's Camion Locations - 3 Main Clusters of Dog-Centric Locations in Chicago, IL



Discussion

The food truck industry has been associated with innovation and creativity since its inception. Meanwhile, pets have been a part of Americana since before Norman Rockwell's paintings of boys and dogs were on the cover of *The Saturday Evening Post* (Rockwell). Joining the two together capitalizes on both!

The analysis showed that of the five largest cities in the United States, by population, New York (1st) had the highest population density but the lowest percentage of the population that owned pets. Conversely, Phoenix (5th) had the lowest population density but the highest percentage of pet ownership. Both were excluded from consideration, along with Houston (4th), which was similar to Phoenix, but with a lower percent of the population owning pets and slightly higher population density. Los Angeles (2nd) and Chicago (3rd) both had similar percentages of their population who were pet owners, but Chicago had higher population density, so Chicago was suggested as the launch pad for *Il Cane's Camion*.

With 50+ locations for each of the four dog-related venue categories within 5 miles of downtown Chicago, there was ample data for identifying areas where dogs and their owners will often be out and about. With foot traffic being found as especially critical to the success of food trucks (Murray), it is ideal that dogs and their owners need to take regular trips outside, as well as going on walks. Using only a 5 mile radius means it wouldn't be uncommon for a dog and its owner to visit or pass several of the dog-friendly locations during one outing.

KMeans clustering was done on the combined dataset of all dog-related businesses in order to identify 3 clusters and their central points, so that the operators of *Il Cane's Camion* have three data-driven location suggestions to park and open for business. Latitude and longitude for the center point of each cluster were given, so *Il Cane's Camion's* driver will need nothing but google maps or a similar program to input the coordinates and find the suggested locations!

While this data is compelling, further analysis is recommended before making a final decision on locations. The cost of licensing, owning, and operating a food truck in the city of Chicago or other possible city options was not taken into account, nor were any restrictions on food truck parking researched. Another interesting dataset to add to this analysis would be the availability and location of dog-friendly housing in Chicago.

Depending on the success of *Il Cane's Camion*, future research could also identify which other pet treats might sell well, with cats, rabbits, guinea pigs, etc. as options. Other cities could also be considered as areas for expansion, either by having additional dog food trucks in those cities, or possibly by commuting to another city in the winter when Chicago weather might decrease foot traffic and affect sales.

Conclusion

Mobile food operators are often on the front lines of setting new trends. Fido and Max, the stakeholders in Il Cane's Camion, want to twist the traditional food truck concept by serving treats to pooches instead of people.

In this project, the five biggest cities in the United States were analyzed to determine the estimated percentage of pet owners from each city's population. Since foot traffic is an important feature for the success of a food truck, population density was also considered. The city of Chicago is recommended for the dog/food/truck hybrid venture Il Cane's Camion.

Through gathering category-specific venue location data from Foursquare, a map was created of dog-related businesses within a radius of 5 miles in downtown Chicago. Clustering was applied, using the KMeans algorithm, and three central points were found, along with their latitude and longitude coordinates, and can be recommended as places for Il Cane's Camion to stop, park, and open up shop for a few hours. Further analysis is suggested on food truck regulations in downtown Chicago, as well as research into pet-friendly housing in the area. In the future, Il Cane's Camion could expand into selling treats to other pet types, or possibly franchising to have multiple pet food trucks in multiple US cities.

Works Cited

Retrieved March 2020, from American Pet Products:

https://www.americanpetproducts.org/press_industrytrends.asp

Retrieved March 2020, from The Harris Poll: <https://theharrispoll.com/whether-furry-feathered-or-flippers-a-flapping-americans-continue-to-display-close-relationships-with-their-pets-2015-is-expected-to-continue-the-pet-industrys-more-than-two-decades-strong/>

Retrieved March 2020, from Saturday Evening Post: <https://www.saturdayeveningpost.com/wp-content/uploads/satevepost/9290928.jpg>

AHS. Retrieved March 2020, from <https://www.census.gov/programs-surveys/ahs.html>

geocod.io. Retrieved March 2020, from <https://www.geocod.io/docs/?python#authentication>

Murray, P. *Optimizing Food Truck Locations*. Retrieved March 2020, from <https://carto.com/blog/optimizing-food-truck-locations/>

Saxe, L. Retrieved March 2020, from Forbes: <https://www.forbes.com/sites/lizzysaxe/2018/12/12/want-to-know-the-future-of-food-trucks-in-2019-read-this-report/#4728fd25398f>

U.S. Cities by Population. Retrieved March 2020, from wikipedia.org: "https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population"

Venue Categories. Retrieved March 2020, from <https://developer.foursquare.com/docs/resources/categories>