

---

# Opening a food and supply store for dogs in Manhattan

---

## IBM Applied Data Science Capstone Project

Carlotta Brunetti - 6 December 2020



*Image taken from images.fineartamerica.com*

---

# 1. Introduction

## 1.1 Background

### *The pet industry expenditures*

According to the statistics provided by the American Pet Products Association (APPA) [1], the pet industry expenditures in the United States have been characterised by a steady increase over the past 30 years. In this period of time, Americans have increasingly spent money on pets as they are more and more humanised and treated as members of the family [2]. The total U.S. pet industry expenditures are estimated to reach \$99 billions by the end of 2020 that is \$9 billions dollar more than in 2018 [1]. High-quality food and treats for pets make up the most profitable and the highest selling portion of the pet market (i.e., estimated at \$38.4 billions).

In US, there are more households with dogs than with children. Indeed, 67% of US households own a pet and, of all types of pets, dogs are the most popular. Around 63.4 million US households own a dog compared to 42.7 million who own a cat [1].

### *The pet industry and the COVID-19 pandemic*

The COVID-19 pandemic has negatively affected many industries and caused unprecedented disruptions to economies. However, the pet market industry has long been proved to be recession-resistant [2][3]. That is particularly true for the economic crisis caused by the COVID-19 spread. According to the U.S. Pet Market Outlook, 2020-2021 [4], pet adoptions have increased as they provide emotional support and companionship during the lockdown periods. Moreover, owners spend more time with their pets and, hence, invest more money on them [2]. For this reason, we can expect that 2020 and 2021 will be particularly successful years for the pet market.

## 1.2. Business problem

According to a study conducted by SmartAsset, New York City is among the 25 most pet friendly cities in America [5] with 8.5 million residents and more than 600,000 dogs. For this reason, opening a food and supply store for dogs would be a profitable business as well as a service very appreciated by the residents who own a dog. However, the location of the store need to be chosen accurately as it is the key factor in the success of the business. In this project, I combine Foursquare APIs location data and machine learning to help the stakeholders make decisions by answering the following business questions:

- 
- 1) Which are the best neighbourhoods to live with a pet dog in Manhattan?
  - 2) Where should an entrepreneur open a food and supply store for dogs in Manhattan?

### 1.3. Target stakeholders

Since the pet industry is one of the most stable and recession-resistant industries, this project is of interest for all those dog-lover entrepreneurs seeking to open a successful business during the COVID-19 pandemic. The outcome from this project is also relevant for dog owners moving to Manhattan who are interested in finding the best neighbourhood to live with their furry-friends.

## 2. Data

For this project, we need the following data:

- **Zip codes and corresponding neighbourhood names in Manhattan**

*Data source:* <https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>

*Data description:* The data are contained in a .csv file which consists in two columns: neighbourhood names and zip codes, respectively. Note that more than one zip codes is associated to a neighbourhood.

- **Latitude and longitude of the zip code areas in Manhattan**

*Data description:* The latitude and longitude of each zip code area in Manhattan are found using the pgeocode library in python that allows to use the zip code as input to get the corresponding geographical coordinates.

- **NYC Dog Licensing Dataset**

*Data source:* <https://data.cityofnewyork.us/Health/ NYC-Dog-Licensing-Dataset/nutn-tubp>

*Data description:* This is a public database of licensed dogs in New York City updated on February 8, 2020. The dataset contains different information (dog name, breed, date of birth etc.) but we are mainly interested in the zip code area in which the dog live. Indeed, from this data, we can extract the total number of dogs living in each neighbourhoods of Manhattan

- **Venue data and venue category ID**

*Data source:* Foursquare APIs, <https://developer.foursquare.com/docs/build-with-foursquare/categories/>

**Data description:** By using this API, I will get all the venues in each neighbourhood filtered by category of interest (e.g., veterinarians, dog runs, pet service, pet store etc.). The categories are selected based on their ID code that can be found at the link above.

- **Polygons of the zip code areas in New York City**

**Data source:** [https://jsspina.carto.com/tables/nyc\\_zip\\_code\\_tabulation\\_areas\\_polygons/public/map](https://jsspina.carto.com/tables/nyc_zip_code_tabulation_areas_polygons/public/map)

**Data description:** The data are contained in a json file and consists in the polygons coordinates that delimit each zip code area in New York City. Among them, I select those located in Manhattan.

## 3. Methodology

### 3.1 Zip code areas and dog count in Manhattan

The first step is to derive the latitude and longitude coordinates from the zip code of each area in Manhattan using the pgeocode library. There are in total 43 zip code areas in Manhattan. I also count the number of licensed dogs living in each area. I can visualise these data by plotting the polygons coloured based on the number of dogs using the folium library (Figure 1).

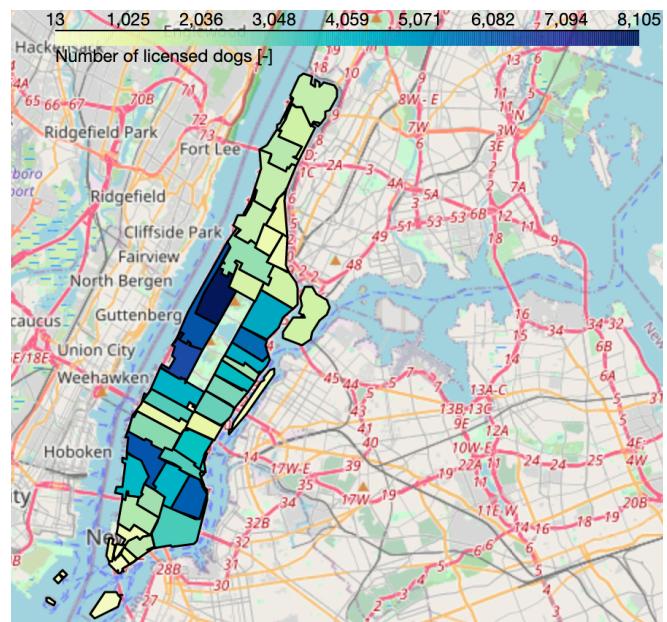


Figure 1: Number of licensed dogs in each zip code area of Manhattan

### 3.2 Venue data from Foursquare APIs

I explore the venues within an area of 500 m around the latitude and longitude coordinates of each zip code area. I limit the search to specific venue category that are relevant to answer the business questions. I have selected the following 9 categories:

- 1 - Supermarket
- 2 - Pet service (e.g., pet wash, spa, hotel, daycare)

- 
- 3 - Pet store
  - 4 - Veterinarian
  - 5 - Dog run
  - 6 - Park
  - 7 - Trail
  - 8 - Pet café (i.e., café completely dog friendly)
  - 9 - Animal shelter

### 3.3 Clustering the zip code areas in Manhattan

I apply the k-means clustering method to cluster the zip code areas in Manhattan based on the venue data from Foursquare APIs. The method minimizes within-cluster variances that are quantified by the squared Euclidean distances. The first thing to do is to choose the appropriate number of clusters. At this purpose, I use the elbow method which consists in plotting the sum of squared distances as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use (Figure 2). According to the elbow method, 3 clusters should be chosen.

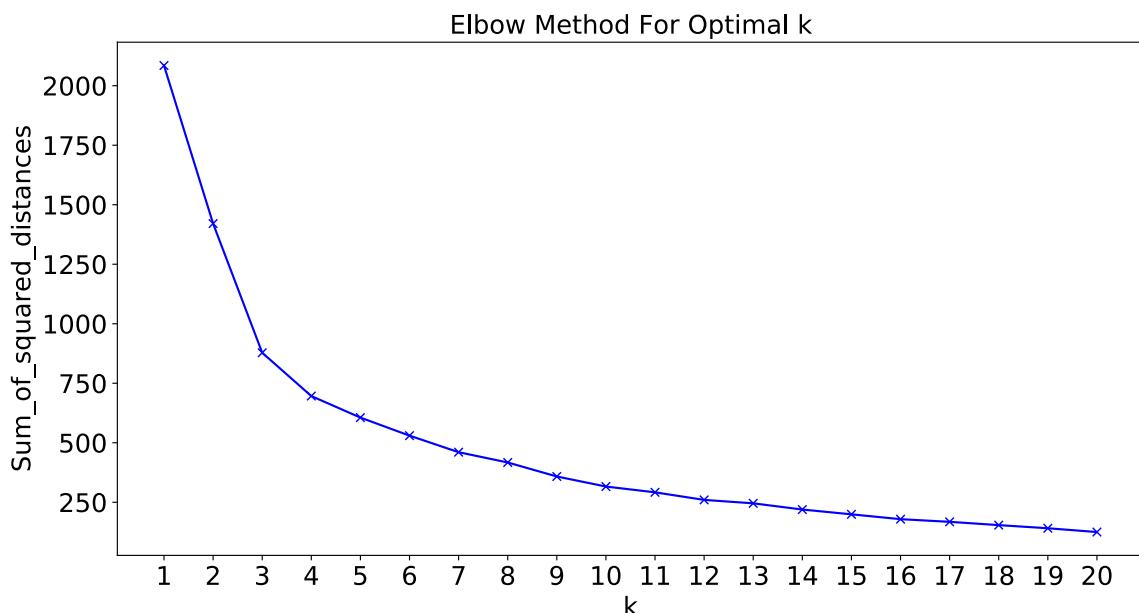


Figure 2: The elbow method used to choose the appropriate number of clusters, in this case 3.

---

### 3.4 Data analysis to answer the first business question

In this section, I explain the data analysis that I have performed in order to find an answer to the question: Which are the best neighbourhoods to live with a pet dog in Manhattan?

Firstly, I analyse the clustering results and select the clustered area(s) that is(are) more relevant to answer the business question. Secondly, I investigate in more details the zip code areas in the selected cluster(s). In particular, I compute the total number of venues (excluding the category "Animal Shelter" that is not relevant to answer the first business question) for each zip code area in the cluster. I then look at the distribution of the venues (stacked bar chart) for the first 5 zip code areas with the highest total number of venues. I will select the best zip code area in cluster 2 and 3 in which to live with a furry friends by considering the highest number of venues as well as a reasonably good distribution of venues in each category.

### 3.5 Data analysis to answer the second business question

In this section, I explain the data analysis that I have performed in order to find an answer to the question: Where should an entrepreneur open a food and supply store for dogs in Manhattan?

Firstly, I want to know the total number of venues where food and supplies for dogs are sold. At this purpose, I compute the total number of pet stores and supermarkets ( $N_{fs}$ ) in each zip code area. Secondly, I want to know the number of pet stores and supermarkets relatively to the number of licensed dogs ( $N_d$ ) that live in each zip code area. At this purpose, I compute their ratio:

$$\alpha = \frac{N_{fs}}{N_d} \quad (1)$$

The zip code area with the lowest ratio (few food and supply stores compared to the number of dogs) can be considered as a good spot in Manhattan where to open a food and supply store for dogs.

A fact to be considered is that the dogs in animal shelters are not licensed and, consequently, they are not included in the dog count. Therefore, I have also identified the zip code area with the highest number of animal shelters.

## 4. Results

### 4.1 Clustering the zip code areas in Manhattan

Based on the clustering results (Figure 3), I compute and display the total number of venues in each category and for each cluster (Figure 4).

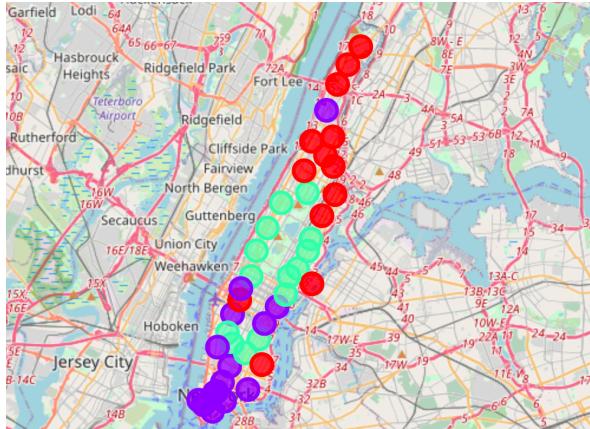


Figure 3: Map of the three clustered areas in Manhattan

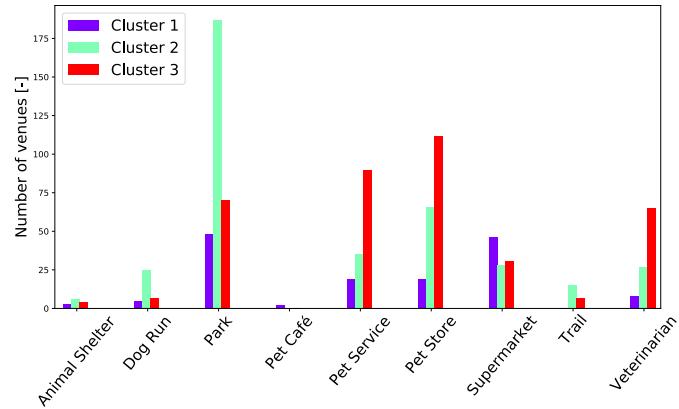


Figure 4: Distribution of venues in each category and cluster

Cluster 2 is mainly characterized by green areas (i.e., dog runs, parks, trails) whereas cluster 3 is mainly characterized by various pet services (i.e., pet services, pet stores, veterinarians). Cluster 1 is where most of the supermarkets are located but, overall, is not so relevant to answer the first business questions. As a consequence, I investigate more in detailed the distribution of the venues for the first 5 zip code areas with the highest total number of venues in cluster 2 (Figure 5) and cluster 3 (Figure 6).

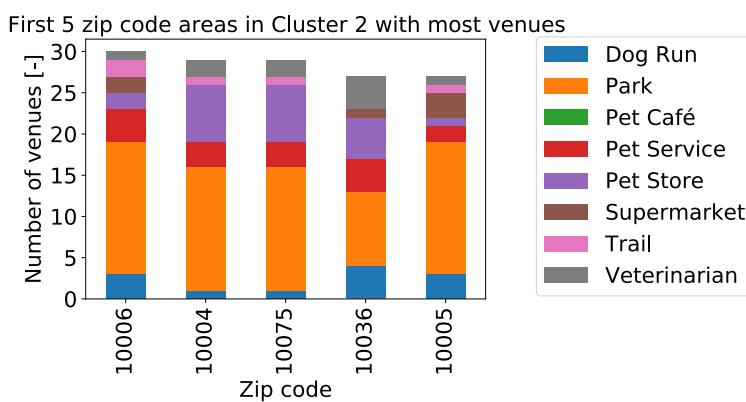


Figure 5: Distribution of the venues for the first 5 zip code areas with the highest total number of venues in cluster 2

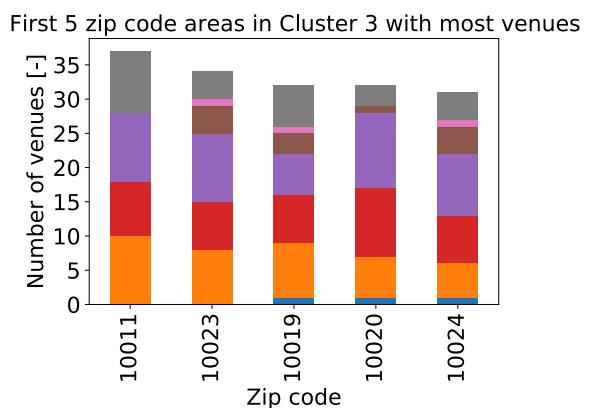


Figure 6: Distribution of the venues for the first 5 zip code areas with the highest total number of venues in cluster 3

The zip code area 10006 is the one with the highest number of green areas (dog runs+parks+trails). It has also some pet stores and pet services as well as one veterinarian. This area is suggested for active dogs and dog owners who love to spend most of their time outdoor.

The zip code area 10011 is the one with the highest number of various pet services (pet services+pet stores+veterinarians). It has also some parks but none dog runs. This area is suggested for dog owners who love to have a wide choice of pet services and vet care facilities.

## 4.2 Number of food and supply stores and dog count

I compute the ratio (Equation 1) presented in Section 3.5 for each zip code area. I display them together with the number of licensed dogs (Figure 7).

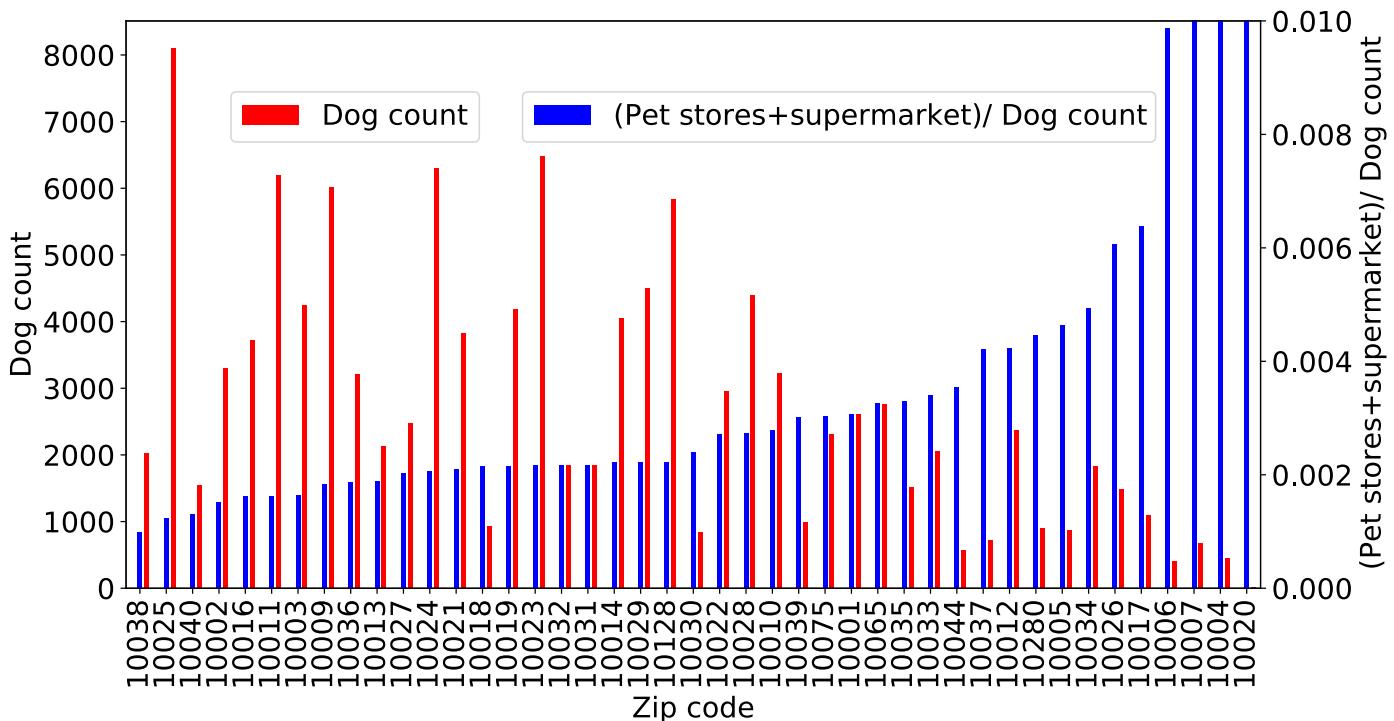


Figure 7: Ratio and number of licensed dogs for each zip code area of Manhattan. Note that the upper limit of the left y-axis was set to 0.01 to facilitate the visualization of the results.

The zip code area 10038 is the one with the smallest number of pet stores and supermarkets per number of licensed dogs. Therefore, it might be a successful area in

which to open a supply and food store for dogs. Also, zip code area 10025 might be a good spot. Indeed, it is the zip code where most of the dogs lives and the ratio is still quite low.

As already pointed out in Section 3.5, the dogs in animal shelters are not licensed and, as a consequence, they are not included in the dog count. There are 13 animal shelters in Manhattan distributed over 43 zip code areas. Indeed, most of the areas have none animal shelter and very few has one. However, in the zip code area 10013 there are 4 animal shelters and, hence, the ratio (Equation 1, Figure 7) is likely overestimated for this area.

The main results are summerized in Figure 8.

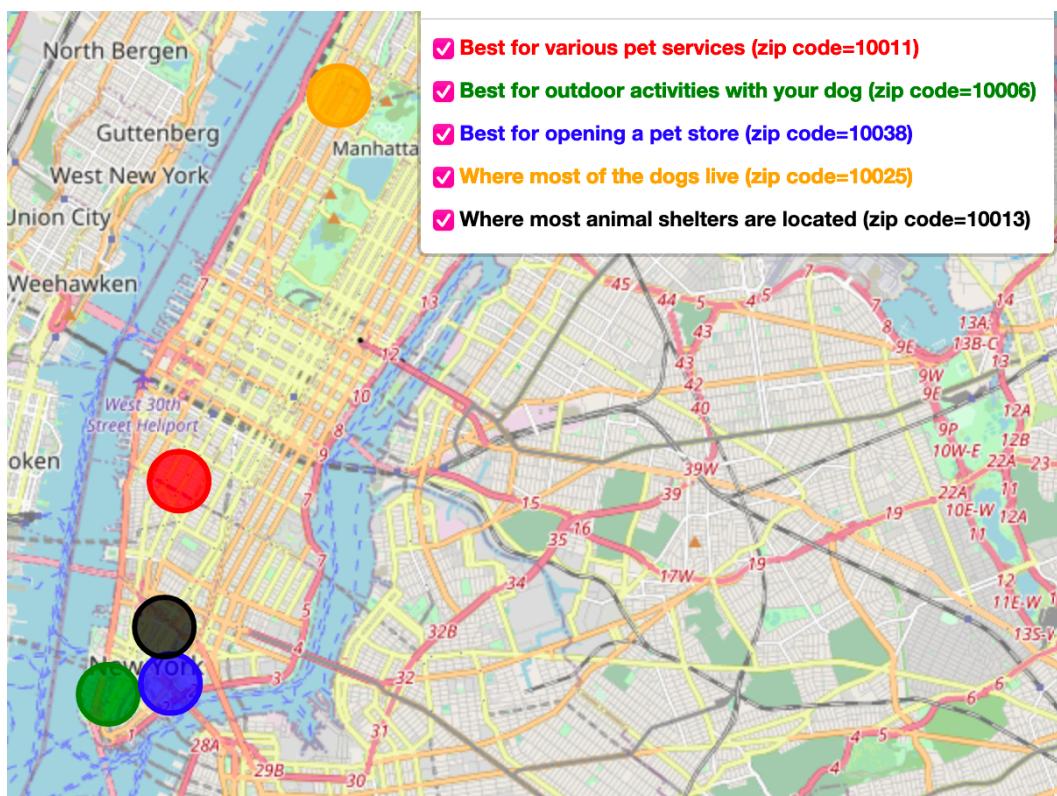


Figure 8: Main areas of interest to address the business questions

The zip code area 10038 other than having the lowest ratio (Equation 1) is also located next to the zip code area 10013 where there are the highest number of animal shelters that might need a lot of food and supplies. Moreover, the zip code area 10038 is also next to the zip code area 10006 where most of the green areas are located and likely very frequented by dog owners.

---

## 5. Discussion

Some observations need to be done to properly evaluate the results in the correct prospective:

- The size of the areas used to explore the Foursquare venues (circle with a 500 m radius, Figure 3) does not exactly match the polygons representing the zip code areas (Figure 1). For instance, in my analysis, the total number of pet stores and supermarkets is computed in the 500 m radius circles whereas the number of dogs is related to the entire zip code polygon.
- The Foursquare database is prone to incorrect classification of the venues since it is based on the input from the users.
- Not all the dogs living in Manhattan are licensed.
- Other datasets could be taken into consideration to even better address the business questions such as the average rent prices, average gross incomes and crime statistics in each zip code area.

## 6. Conclusion

In this project, I have combined Foursquare APIs location data and machine learning to help the stakeholders make decisions by answering the following business questions:

- 1) Which are the best neighbourhoods to live with a pet dog in Manhattan?
- 2) Where should an entrepreneur open a food and supply store for dogs in Manhattan?

According to the clustering results on the Foursquare APIs location data, the best neighbourhoods to live with a pet dog in Manhattan are Lower Manhattan (zip code 1006) and Chelsea (zip code 10011). Specifically, the zip code area 10006 is suggested for active dogs and dog owners who love to spend most of their time outdoor. The zip code area 10011 is instead recommended for dog owners who love to have a wide choice of pet services and vet care facilities.

According to additional exploratory analysis on the Foursquare APIs location data, the best spot in Manhattan in which to open a food and supply store for dogs is in Lower Manhattan (zip code 10038). The zip code area 10038 is the one with the smallest number of pet stores and supermarket per number of licensed dogs. Moreover, it is next to the zip code area 10013 where there are the highest number of animal shelters that might need a lot of food and supplies. Zip code area 10038 is also next to the zip code area 10006 where most of the green areas are located and likely very frequented by dog owners.

---

## 7. References

- [1] <https://www.americanpetproducts.org/>
- [2] <https://www.entrepreneur.com/article/352418>
- [3] <https://www.prnewswire.com/news-releases/packaged-facts-55-billion-pet-products-market-is-proving-recession-defiant-301082066.html>
- [4] <https://www.packagedfacts.com/Pet-Outlook-13135569/>
- [5] <https://smartasset.com/mortgage/most-dog-friendly-cities-in-america-2020>