IBM Data Science Capstone Project Report

Identifying The Most Dog-Friendly Neighborhoods in Manhattan, New York City

Van Anh Nguyen

7 September 2019

Figure 1: An illustration of dogs in Manhattan by Julia Rothman, New York Times[1]

1   __Introduction__

Whether you are a dog owner or a business in the dog care industry, identifying locations that are friendly to dogs in the city is always an important mission. Among various metropolitans across the world, New York City ("NYC") is perceived as one of the friendliest cities to dogs with a number of parks, dog-run spaces and walkable neighborhoods. According to New York Open Data, an online data platform of the NYC Government, as of 26 July 2019, there are 345,727 dogs with an active license in the city. This report aims to identify the most dog-friendly neighborhoods in Manhattan, the most dog-dense borough in NYC.
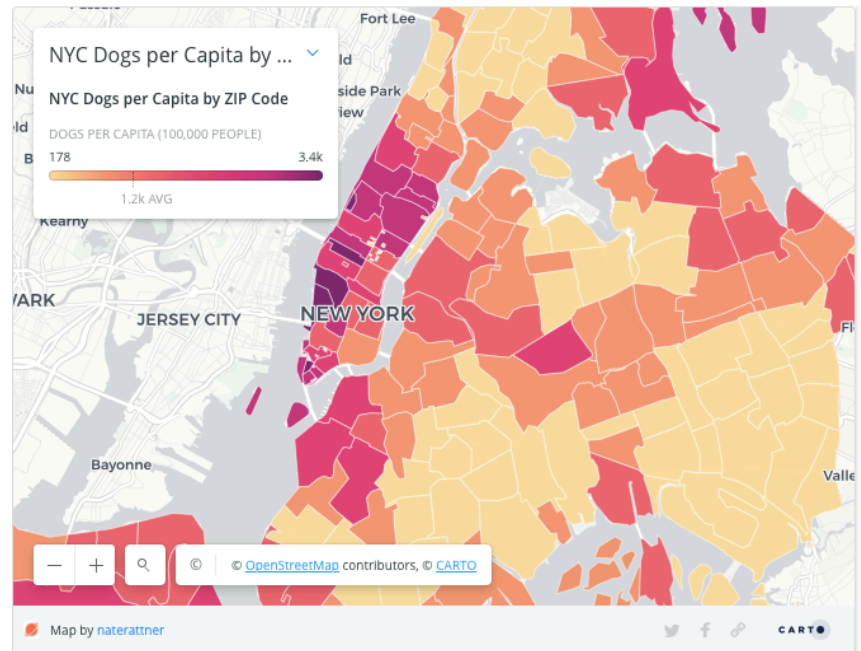

Figure 2: A heat map of New York's dog population per capita by Nate Rattner showing that Manhattan is the most dog-dense borough in New York City.[2]

2   __Objective__

This report evaluates the friendliness of Manhattan's neighborhoods and thus identify the top neighborhoods for dog owners and dog care businesses in this metropolitan. To this end, we mainly focus on analyzing the availability of public facilities necessary to dog care across Manhattan's neighborhood, namely parks, dog run and pet facilities. The report's findings will serve as a helpful source of information for dog owners and businesses in the dog care industry in New York City in general and Manhattan.

3   __Data__

To achieve its objective, the report uses the following data:

- The list of neighborhoods in New York City by New York University containing 306 neighborhoods across 5 boroughs: https://geo.nyu.edu/catalog/nyu_2451_34572.
- Longitude and latitude of each of the neighborhood using Python Geocorder package

---

- Venue data of each of the community areas using Foursquare API. In particular, Foursquare API provides location based experiences with diverse information about venues, users, photos, and check-ins. It supports real time access to places that assigns users to specific locations.

4   **Methodology**

First, we need to create a data frame with a full list of all neighborhoods in Manhattan as well as their coordinating latitude and longitude. This step requires:

- Loading the 2014 New York City Neighborhood Names dataset from the Spatial Data Repository of New York University and transforming it into a panda data frame.
- Retrieving longitude and latitude of each of the neighborhood using Geocoder, a package that allows us to convert a location address into geographical coordinates.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Figure 3: Manhattan's neighborhood data frame

Next step, we explore venues location in neighborhoods of Manhattan using Foursquare API. Foursquare API is a social location service that allows users to explore the world around them. With a Foursquare developer account, we use our user credentials to search for nearby venues of a specific type, explore a particular venue, and search for trending venues around a location.

For the purpose of this project, we are going to focus mainly on public parks and dog runs, two facilities essential to any dog-friendly neighborhood. Using Foursquare API, we can easily check the mean of the frequency of occurrence of parks and dog runs among various neighborhood of Manhattan. From here, our task of assessing dog-friendliness of Manhattan neighborhoods boils down to identifying neighborhood with the highest mean of the frequency of occurrence of parks and dog runs. To make the results visually engaging, we will also group neighborhoods into clusters and assign each of the clusters with a color and label. This step will be done with the employment of k-means clustering, one of the simplest and popular unsupervised machine learning algorithms that groups data points to the nearest cluster with the center being a centroid.

The final results will be color-coded maps of various clusters representing different levels of park and dog run concentration in Manhattan.

5   **Results and Observations**

Our data analysis shows that the 3 Manhattan neighborhoods with the highest concentration of parks are:

- Stuyvesant Town
- Battery Park City
- Morningside Heights

Cluster Labels		Neighborhood	Park	Borough	Latitude	Longitude
0	2	Battery Park City	0.090	Manhattan	40.711932	-74.016869
25	2	Morningside Heights	0.075	Manhattan	40.808000	-73.963896
30	2	Stuyvesant Town	0.100	Manhattan	40.731000	-73.974052

Figure 4: Manhattan's neighborhoods with the highest concentration of parks
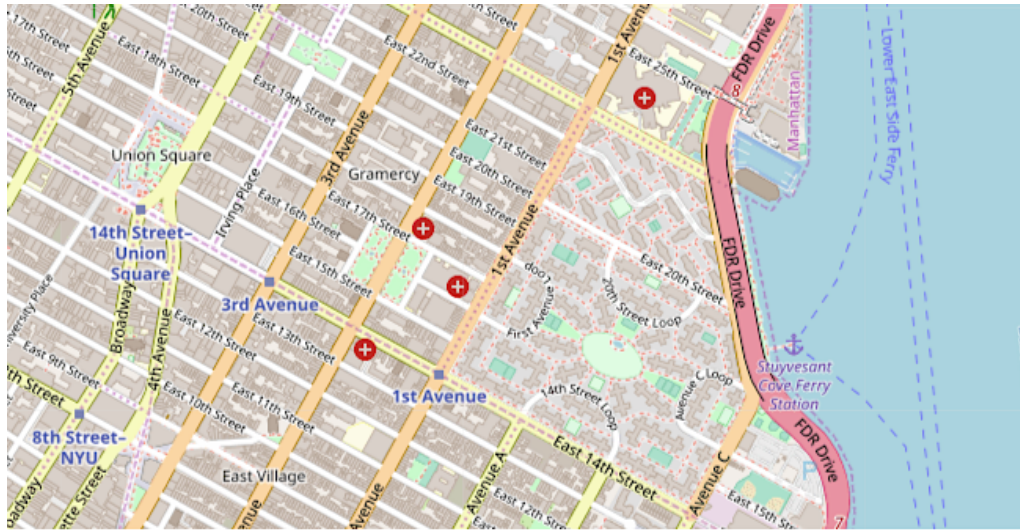

Figure 5: Map of Stuyvesant Town showing that the neighborhood has plenty of parks and open spaces
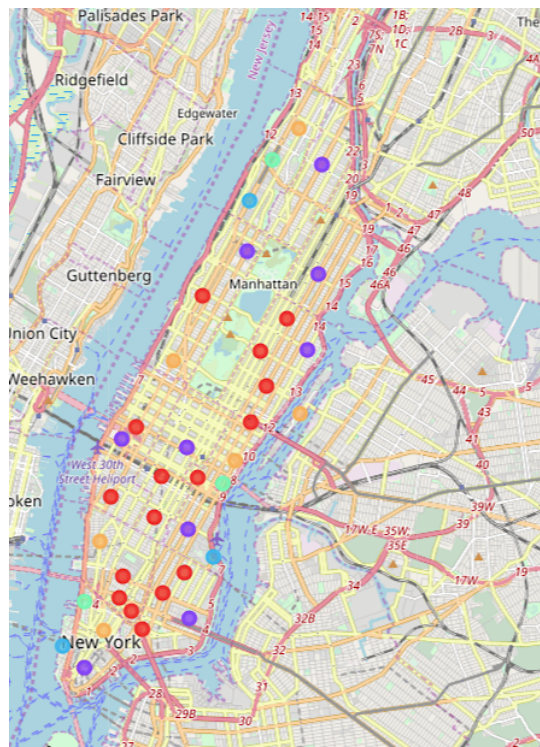
Figure 6: Map of Manhattan showing clusters of neighborhood by the concentration of parks

Notably, a further examination of venues in Stuyvesant Town, the most park concentration neighborhood, shows that among the 30 top venues within 1km of the neighborhood, there are not only various parks but also a trail, a scenic lookout, a pet cafe and even a pet service store! This indicates that Stuyvesant Town is indeed a neighborhood with a variety of options or the enjoyment of active dogs.

Regarding dog-runs, the following 3 neighborhoods are the most concentrated area:

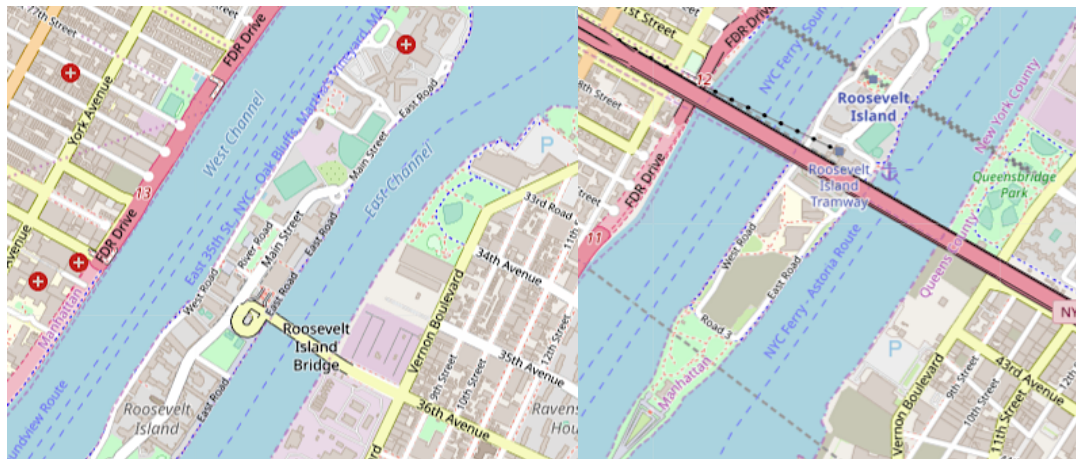| Rank	Neighborhood	Mean of mean of the frequency of occurrence of dog runs
1	Roosevelt Island	0.04000
2	Hudson Yards	0.02564
3	Tudor City	0.02439

Figure 7 and 8: Maps of Roosevelt Island showing that the neighborhood has plenty of parks and open spaces

A further examination of popular venues near Roosevelt Island also shows that *7 out of top 30 venues nearby the island are open spaces*, including 4 parks, a scenic lookout, a recreation park and a water front! What a good candidate for a dog-friendly neighborhood in a concrete jungle like New York city!

Figure 9: Map of Manhattan showing clusters of neighborhood by the concentration of dog runs

6   **Discussion**

Interestingly, Battery Park City, the second most concentrated park neighborhood from our analysis, was also selected as the most dog-friendly neighborhood in NYC in 2017 by Street Easy, an US-based real estate marketplace. The analysis of Street Easy was based on the following key variables in each neighborhood: (1) the share of pet-friendly rental units (2) the share of dog sitters (3) the number of dog parks, dog runs and off-leash areas and (4) the share of veterinarians. [3] The similarity between our analysis and Street Easy's indicates that high concentration of open spaces like parks and dog runs is indeed an important factor to decide the dog-friendliness of any neighborhood.

However, we also note in addition to these open spaces, there are quite a few other facilities that are equally critical to the enjoyment of dogs that we do not have the capacity to include in this report, such as walkable sidewalks, pet-friendly restaurants and coffee shops, vet clinics and pet care stores. Furthermore, on top of all the necessary facilities, a neighborhood can only be truly dog-friendly if it also has a meaningful number of other healthy dogs and welcoming owners. It is in fact not easy to find a neighborhood with all of these elements in a big city like NYC. Hopefully with more and more useful information like the one provided in this report, dog parents will be well informed in choosing the best home for themselves and their fluffy children.

---

[3]   https://streeteasy.com/blog/nycs-most-dog-friendly-neighborhoods/

7   **Limitations and Suggestions for Further Research**

We are aware that during clustering, Murray Hill, a neighborhood in Queens, was mistakenly included in both the used data frame and resulting maps as a neighborhood in Manhattan. Even though we have checked and double checked the data and codes multiple times, the reason for this mistaken is still unknown to us.

In addition, as mentioned in the discussion, in this research, we only examined parks and dog runs as the two essential facilities of a dog-friendly neighborhood. We believe that there are quite a few other variables that should be included to improve the comprehensiveness of the research. These variables include (but not limited to) walkability, safety, availability of pet-friendly restaurants and coffee shops, vet clinics and the number of dogs with active license. To the best of our knowledge, data of such variables in NYC/Manhattan are only available in Zip Code unit. As Zip Code is not a universal geographic area but a collection of routes and addresses created by the US Postal Service for mail delivery, it is generally more difficult to work with.[4] Future research could find a way to transform this unit into corresponding geographic area(s) so that the above mentioned data can be put into use easily.

---

[4]   https://medium.com/@naterattner/new-york-citys-most-dog-dense-zip-codes-3697e2993908