**Choosing the best place to stay in NYC**

Leo Pustovalov

February 18, 2019

1. **Introduction**
   1. **Background**

Everyday dozens of thousands of people are taking and planning their business trips. Some of them take their pet with themselves. Travelling with a pet can be a challenging endeavor – there are lots of problems and difficulties, and the selection of a hotel is one of them.

* 1. **Problem**

I will visit NY City this summer for a business trip. I will take with me my beloved mini pig Sammy. I adore Sammy, but I am afraid that because of stress and jet lag Sammy may catch a cold or some kind of “mal de mer”. So, I need to make a list of vet clinics in the areas (boroughs) of NYC and put them on a map to help me to identify the best place for a room rent, where I will stay for the trip.

* 1. **Interest**

The main stakeholder here is my pet and me.

1. **Data acquisition and cleaning**
   1. **Data sources**

All the data about NYC clinics, hotels, and their ratings can be taken via Foursquare API. AT first, we need to get the coordinates of NYC centre. According to <https://www.latlong.net> the latitude of New York City, NY, USA is 40.730610, and the longitude is -73.935242. Then, according to Wikipedia (<https://en.wikipedia.org/wiki/New_York_City#Geography>) the area of NYC is 1213.37 square km, so if we assume that 1213.37 is the square of the circle, we can easily count its radius - 8.921 km. Of course NYC is not circular by its shape, so we will need to make sure that vet clinics are situated in NYC, but not in New Jersey or other places. Using Foursquare API we will get (collect) the data frame of vet clinics.

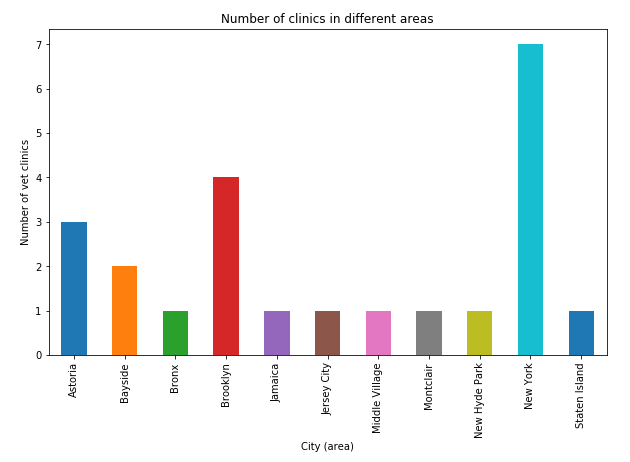
Table 1. Simple vet clinics data frame taken via Foursquare API



* 1. **Data cleaning**

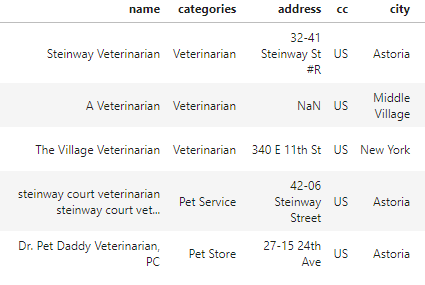
Data collected from Foursquare needs cleaning, for example, data frame from table 1 contains vet clinics not only from NYC, but also from New Jersey and other areas around NYC, this can be seen in the fig 1.

Fig 1. Bar chart of the data frame areas (cities)



In addition, we need to drop the data about unnecessary venues such as "Port Veterinarian Airport Terminal" or “Dr. Pet Daddy Veterinarian, Pet Store”.

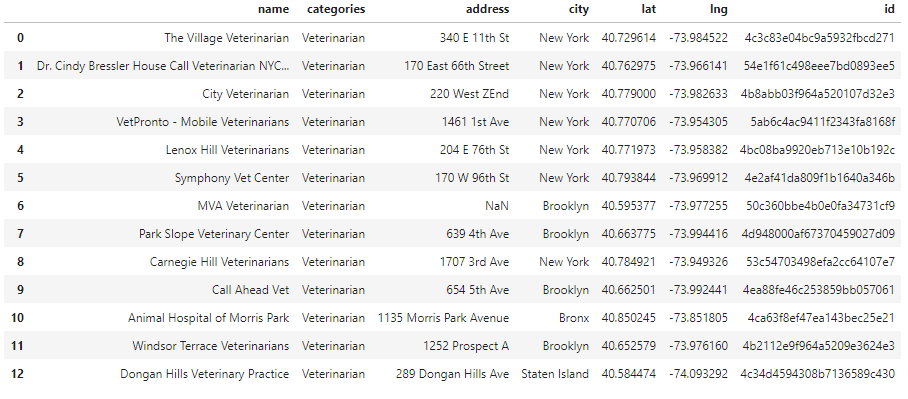
Fig 2. Data frame taken via Foursquare API with unnecessary venues



* 1. **Feature selection**

After data cleaning, there were 12 samples and 16 features in the data. Upon examining the meaning of each feature, it was clear that there was some unnecessary features. After discarding redundant features I got the final vet clinics data frame.

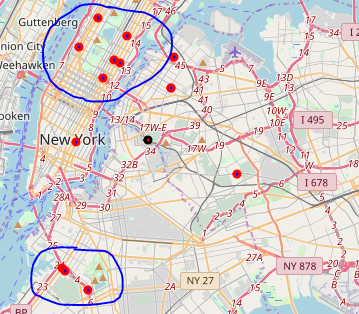
Table 2. Vet clinics data frame after data cleaning



1. **Exploratory Data Analysis**

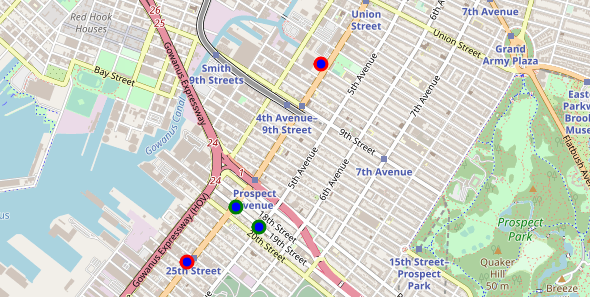
After I generated the map of NYC with vet clinics markers I found 2 clusters of clinics – the first in Manhattan and the second in Brooklyn.

Fig 3. Two clusters of vet clinics in NYC



Manhattan is too expensive and noisy for me personally, so I decided to examine the second cluster of clinics in Brooklyn. Using Foursqure API I got another data frame, with the Brooklyn hotels data. After visualizing the locations of Brooklyn hotels and Brooklyn vet clinics on map I found out that only two hotels are situated near the Brooklyn vet clinics cluster, they are: "Brooklyn Way Hotel, BW Premier Collection" and "Hotel le bleu".

Fig 4. Hotels near vet clinics in Brooklyn



After the comparison of these hotels ratings, I decided to choose “Hotel le bleu” as my place of residence during my future business trip.

1. **Conclusions**

In this study, I analyzed the data about NYC hotels and vet clinics. I identified the best hotel to stay in NYC according to my tastes and opportunities.

1. **Future directions**

I am convinced that this data science approach to business trip planning can be used in other similar cases and needs. Foursquare provides great opportunities for further researches and investigations.