REPORT

GETTING FAMILIAR WITH CUSTOMERS FOR A NYC PET STORE

LKA.CO
27/11/2020
DIANA ORJUELA

1. Introduction

1.1 Background

LKA is a company focus on the marketing and distribution of products and services for pets (cats and dogs). It a startup that operates in Colombia and México through an app and a website that allows its customers to create orders, request services and make purchases online. LKA's value proposition is to provide a whole universe of delivery services so that the furry pets of the families have everything they need.

1.2 Problem

In order to improve LKA communication model with its customers and achieve the true omnichannel. It is necessary to find the best place in the city of New York. The company wants to build its first location. The pet store will offer top quality multiservices and will be part of an exciting loyalty program to improve the service experience of its clients.

2. Data acquisition and cleaning

2.1 Data sources

New York has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

Luckily, those datasets exists for free on the web. here are the links to the datasets:

https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

https://data.beta.nyc/dataset/pediacities-nyc-neighborhoods

https://data.beta.nyc/en/dataset/pediacities-nyc-neighborhoods/resource/7caac650-d082-4aea-9f9b-3681d568e8a5?inner_span=True

On the other hand, all dog owners residing in NYC are required by law to license their dogs. The data is sourced from the DOHMH Dog Licensing System (https://a816-healthpsi.nyc.gov/DogLicense), where owners can apply for and renew dog licenses. Each record represents a unique dog license that was active during the year, but not necessarily a unique record per dog, since a license that is renewed during the year results in a separate record of an active license period.

Each record stands as a unique license period for the dog over the course of the yearlong time frame. The dataset is on NYC OpenData page: https://data.cityofnewyork.us/Health/NYC-Dog-Licensing-Dataset/nu7n-tubp . Metadata Last Updated February 7, 2020.

Unfortunately, Foursquare API do not have enough information about pet stores in NYC. Thus, I did not use it.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values related to boroughs and neighborhoods where dogs lives, so I decide to remove those records. Fortunately, I found a dataset with boroughts to merge with the dogs dataset.

There are some problems with the datasets. First, the dog dataset has several records with License Expired year from 2016 until 2020. Giving the fact that one records is one licence for a dog in one year of its life, I had to remove the records within that range.

Furthermore, I realized there are some birth year mistakes where those values were since year 1. Thus, I deleted some records keeping in mind that dogs do not live more than 20 year average.

Although, there are some missing, unknown and unusual categorical values in features as follow: “dog name” and “breed name”, I decide to keep those records for the cluster analysis because those data represent dogs as well.

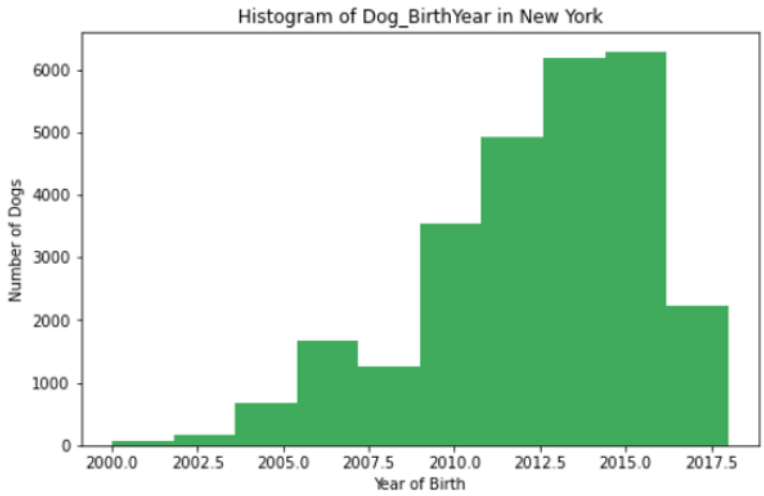
2.3 Feature selection

I have 5 features selected to run the K-Means:
Dog_BirthYear, Dog_Gender, Breed, Neighborhood and Dog_Name.

3. Methodology

In this section, I solve some business questions that will be especially interesting for marketing people.

3.1 How old are NYC dogs?



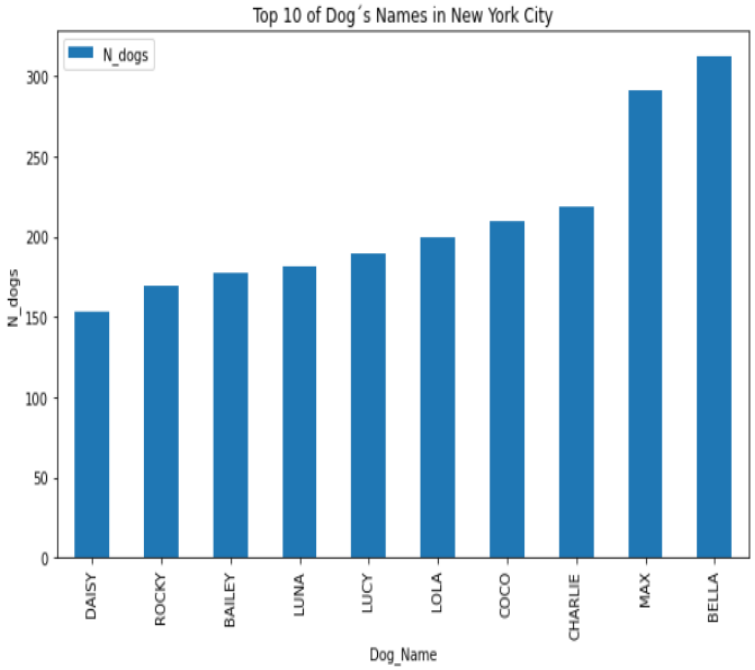
Most of them are in the range of 5 - 10 years old

3.2 How many dogs are there in New York by gender?

There are 12577 females and 14400 males in NYC.

3.3 What is the most popular dog name in New York?

If you own one of the 26,977 -plus dogs that are licensed in New York City, chances are rather good your pup is either named Bella or Max.



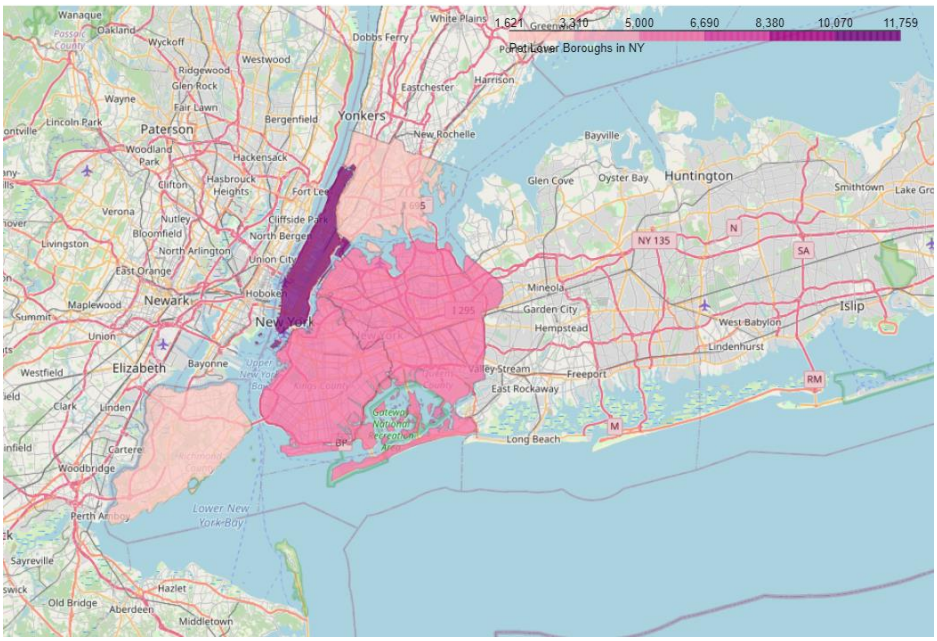
3.4 What is the top 10 dog breed names in New York?

You can find 1239 yorkshire terrier dogs in NYC.

Breed_Name	N_dogs
Havanese	517
American Pit Bull Terrier/Pit Bull	534
American Pit Bull Mix / Pit Bull Mix	736
Maltese	748
Labrador Retriever Crossbreed	842
Poodle	1000
Labrador Retriever	1087
Chihuahua	1140
Shih Tzu	1236
Yorkshire Terrier	1239

3.5 Which are the pet lover boroughs in NYC?

Borough	N_dogs
Staten Island	1720
Bronx	2017
Queens	5129
Brooklyn	6451
Manhattan	11660

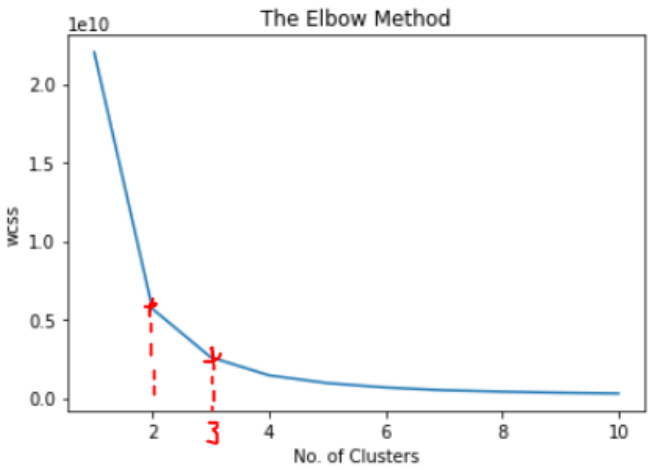


3.6 Which are the Pet lover neighborhoods in NYC?

There are a lot of dogs in Manhattan. That information really surprised me. I found 6 Manhattan neighborhoods where live a lot of dogs. Thus, I decided to analyze it deeper.

4. Clustering Modeling

I decided to use K-means algorithm.
I calculated the sum of squared errors for different number of clusters "K" ranging from 1 to 11. The Elbow Method helped me to choose the number of clusters. According to the plot the number of clusters could be 2 or 3. I ran the model with K= 3.



5. Results and Discussion

Our analysis shows that although there is a great number of dogs (~26000), there are pockets of high dog density fairly close to city center. Highest concentration of dogs was detected on Manhattan, Brooklyn and Queens so we focused our attention to Manhattan and its corresponding 53 neighborhoods which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics which are related with dogs trained for security and a number of pet lovers.
Moreover, we got some answer related to the average dogs age, top 10 dog names, top 10 breed names and the number of dogs by gender. Therefore, in this way we acomplished the objetive "getting familiar with New York dogs".

The pet store can focuses on 3 clusters in Manhattan. You can find the neighborhoods which are part of those clusters on the list below.

Neighborhood
Central Harlem
East Harlem
Lower Manhattan
Inwood and Washington Heights
Greenwich Village and Soho
Gramercy Park and Murray Hill
Lower East Side
Chelsea and Clinton
Upper West Side
Upper East Side

6. Conclusions

Purpose of this project was to identify New York areas close to center with high number of dogs in order to aid stakeholders in narrowing down the search for optimal location for a new Pet Store. By calculating dogs density distribution we have first identified general boroughs that justify further analysis (Manhattan), and then generated extensive collection of neighborhoods which could be called pet lover zones. Clustering of those locations was then performed in order to create major pet lover zones where we can getting familiar with New York dogs in order to offer the best service according to their needs.