Coursera



Report

Pet Lover area in NYC

Better Understand about pet lovers

Suitable places for pet store to expand

1. **Introduction**

In this project, I would like to find out a suitable location for a pet company to expand its business in New York City. The demand of pet food, pet toy and other pet supplies will be the concern of the pet company. Therefore, to discover the most

suitable location, we need to figure out the places of the greatest concentration of dogs. Also, we need to understand what are the pet lovers needed, so that we can meet their requirements

1. **Data**

New York has a total of 5 boroughs and 306 neighborhoods. It is necessarily to have a dataset which contains all 5 boroughs and the 306 neighborhoods that exist in each borough, also the latitude and longitude coordinates of each neighborhood.

We can find the dataset from following links:

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

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

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

In addition, we need a dataset about the license of dogs in NYC. We are able to find the dataset from NYC OpenData page: https://data.cityofnewyork.us/Health/NYC-Dog-Licensing-Dataset/nu7n-tubp

The original datasets called "df\_merged" include data as follow:

RowNumber, AnimalName, AnimalGender, AnimalBirthMonth, BreedName, Borough, ZipCode, LicenseIssuedDate, LicenseExpiredDate, Extract Year, borough, post\_office, neighborhood, population, density.

Another dataset called "df\_segmentation2" is the one we need to getting familiar with the customers and will include data as follow:

Dogs\_Name, Dogs\_Gender, Dogs\_BirthYear ,Breed\_Name, Borough, Neighborhood.

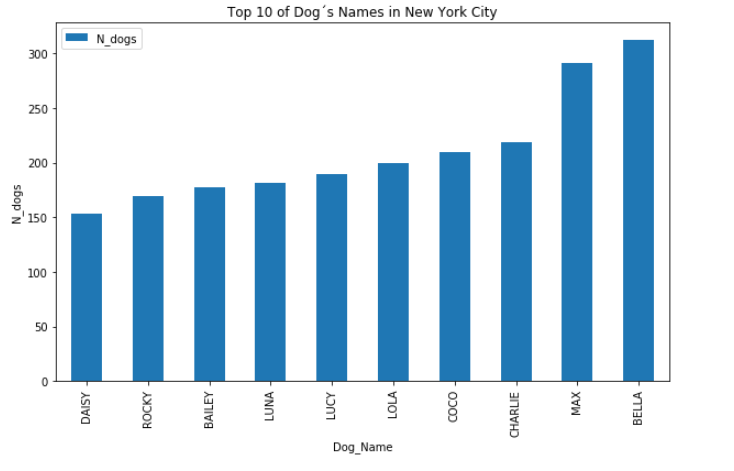
We will also use these 5 features for running K-means

We need to find out the place which is the greatest concentration of dogs, there be the place for pet company to expand.

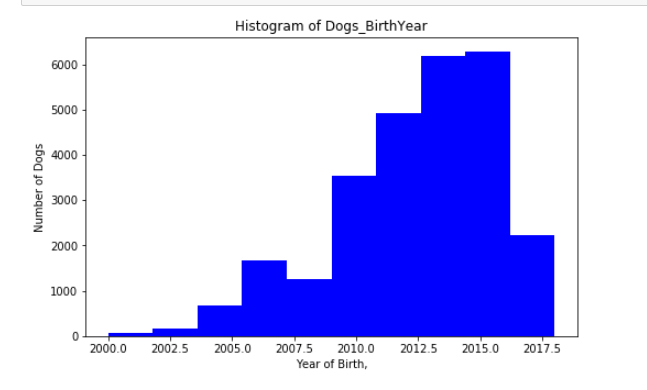
1. **Methodology**

After the analysis, we can have the following result.

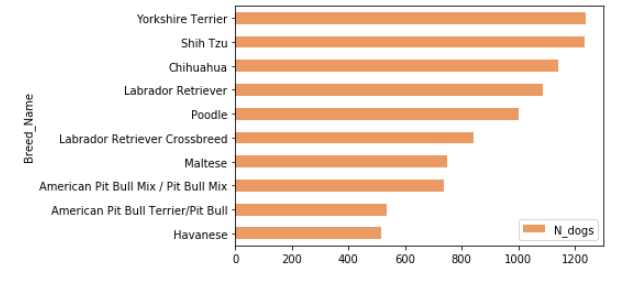
The popular dog names in NTC:



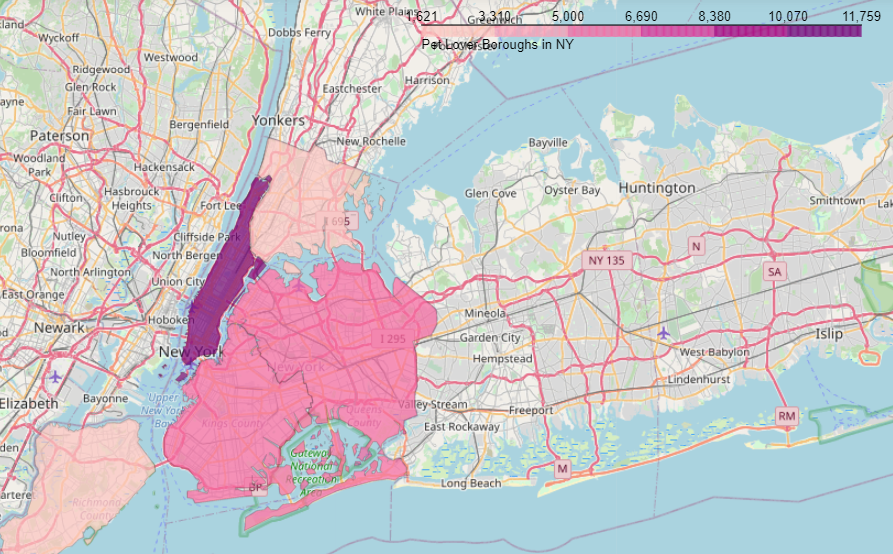
The age of the dog in NYC:

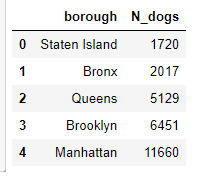
****

Top breeds of dog in NYC:



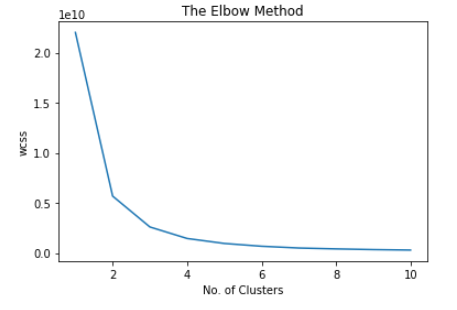
Pet lover area in NYC





In this project, I use K-means algorithm for clustering modeling

After calculated the sum of squared errors for different number of clusters K from the range 1-10. With the help of the Elbow Method, I decided the K value is 3 and run the K-means algorithm.



1. **Results and Discussion**

The result show that the greatest concentration of dogs was detected on Manhattan, Brooklyn and Queens. These three places are potential locations for pet company to expand their business. With the high number of potential customers, pet company able to gain great success in these areas.

In attrition, we also perform some analysis to the average dogs age, top 10 dog names, top 10 breed names and the number of dogs by gender. Therefore, we can better understand our customers and their dog.

1. **Conclusions**

The goal of this project was to identify suitable location for a pet company to expand its business in New York City. We first calculate the dog’s density distribution and located three potential areas, Manhattan, Brooklyn and Queens. Then we generated collection of neighborhoods which could be called pet lover zones. Pet company would consider these places for their business expanding.