**The Battle of Neighborhoods**

--- Where to Build New Pet Store in Houston, Texas, USA

Dan Wang

August 21, 2020

## Introduction

### Background

Pets are human’s best friends. However, lots of pets are abandoned every year due to different reasons: pet’s health issues, owner’s lack of life stability, financial issues, abusing, and so on. If the abandoned pet is lucky, it will be adopted in a pet center or taken care of in different pet related organizations or stores. How those abandoned pets can be taken care of is something that we can consider.

### Problem

According to Houston BARC, they take in lots of abandoned dogs every year. Dogs are mostly picked up in downtown Houston, North Houston, Chinatown, etc. dense neighborhoods. Dogs abandoned in rural areas or rich neighborhoods in Houston are much less.

However, pet related organizations or stores are not evenly distributed in neighborhoods compared to the abandoned location distributions. So if a new pet store will be built, careful consideration about the location needs to be taken into place.

### Interest

Pet store owners would be very interested in the location choices. And any pet related organizations and pet owners would also be interested in the topic.

## Data

### Data sources

List of Houston Super Neighborhoods can be found in the[Wikipedia page](https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods)*.* Open Data Houston published data on [BARC Dog Intakes for Calendar Years 2011 and 2012](http://data.houstontx.gov/dataset/barc-dog-intakes-for-calendar-years-2011-and-2012).Venues data can be accessed through [Foursquare](https://foursquare.com/).

### Data cleaning

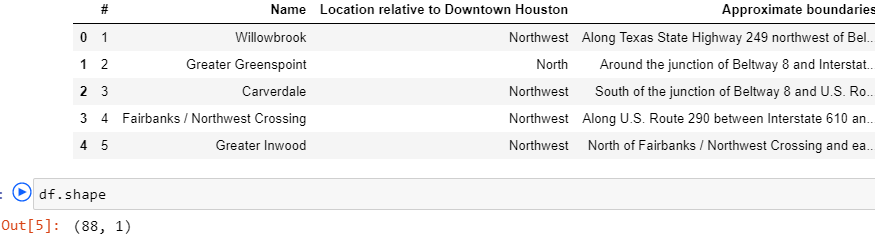
Data downloaded or scraped from multiple sources were combined into one table.

Missing values: for location (latitude and longitude), any missing values rows are removed from analysis due to locations are critical in this topic.

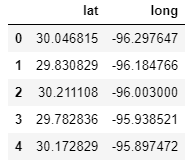
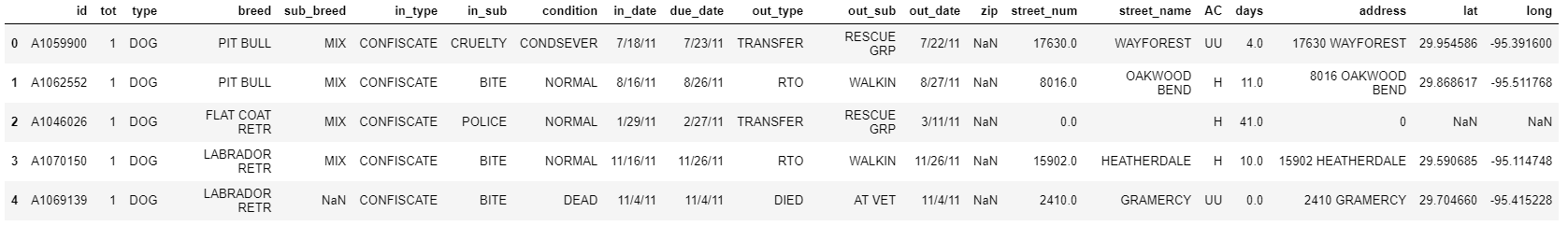
Outliers: any location points that are obvious outside of Houston are removed. There’re in total 5 outlier points.

Unrelated columns: unrelated columns like, dog take in/out time, breed, condition, dates, etc. are not chosen for analysis.

### Feature selection

For Houston Super Neighborhoods data, only the total number of neighborhoods is used for the following analysis (to cluster dog intake locations into 88 groups). 

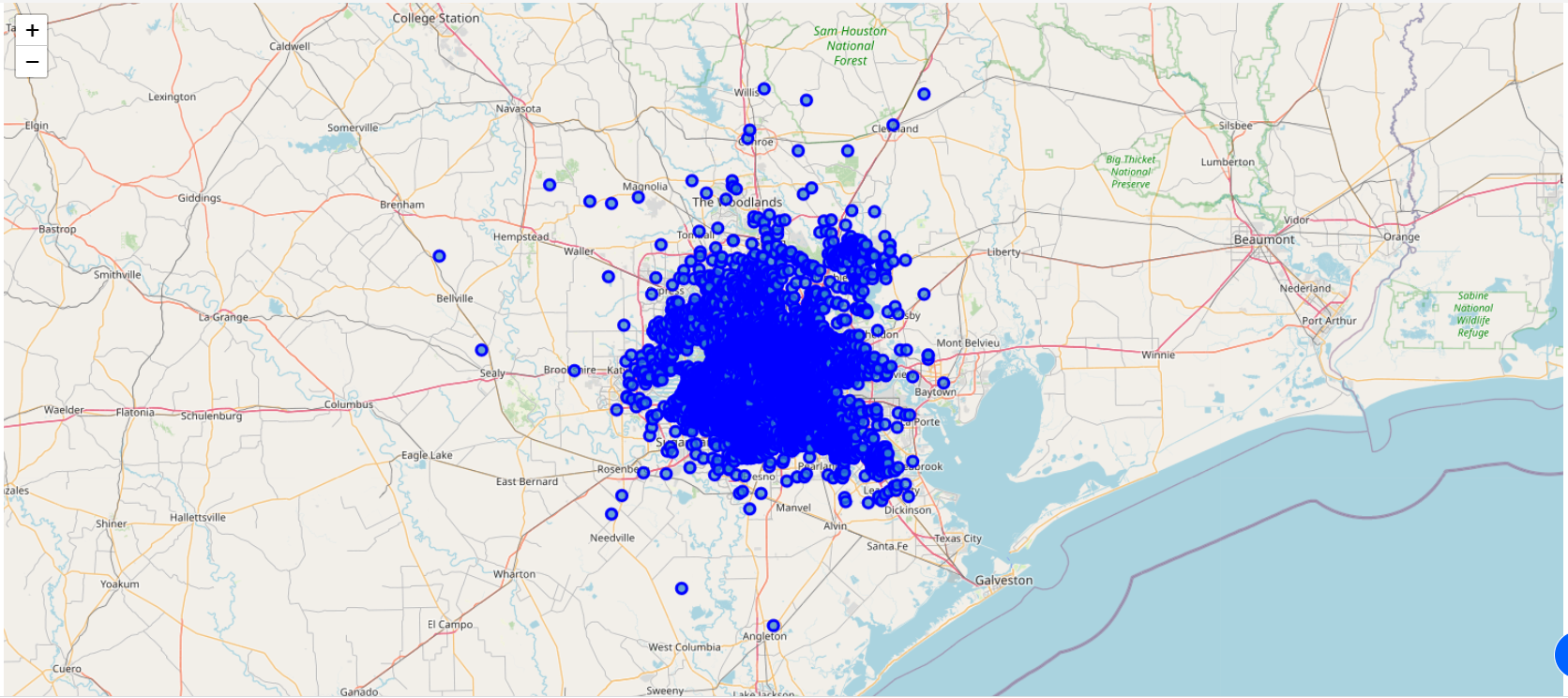
For [BARC](http://data.houstontx.gov/dataset/barc-dog-intakes-for-calendar-years-2011-and-2012) Dog intake data, only latitude and longitude columns are chosen for the following Exploratory Data analysis (EDA). All other columns are dropped off due to this topic is focused location.



## Methodology

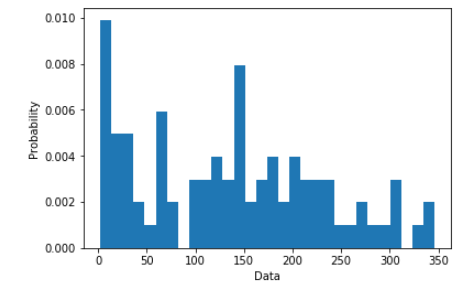
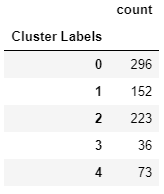
### 3.1 Exploratory Data Analysis

The cleaned [BARC](http://data.houstontx.gov/dataset/barc-dog-intakes-for-calendar-years-2011-and-2012) Dog intake location data is shown in the following map. Huge amounts of intake are distributed in Houston, larger numbers in denser neighborhoods, smaller numbers in rural or richer neighborhoods.

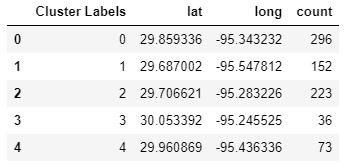


Since the intake location data is a mass, they are clustered into 88 groups based on location (like 88 super neighborhoods) using KMeans modeling.

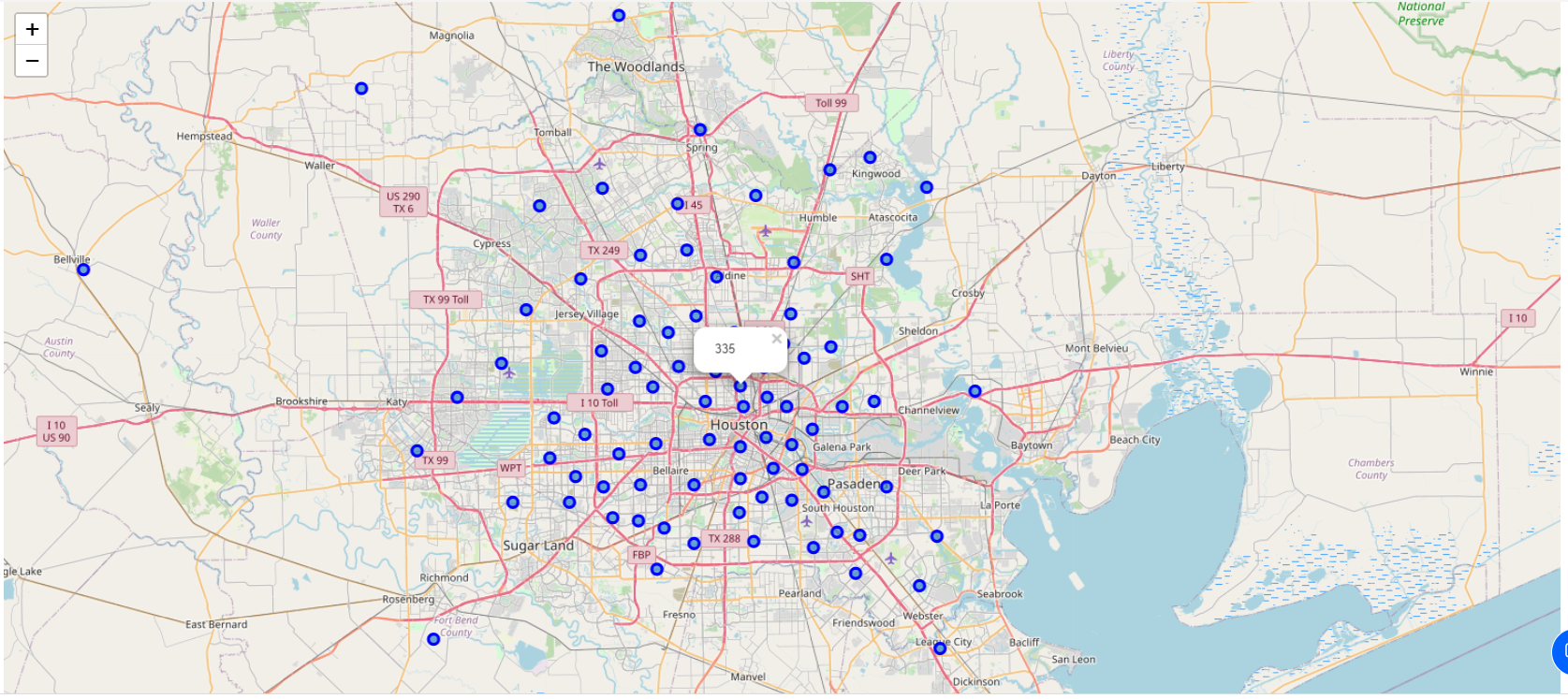
For each group, the total number of intakes is calculated as the “count” column. A histogram chart is also shown below. This dataset is Right-skewed, which means the larger amount of intakes, the smaller the chance is.



Then, the dog intake location is combined with number of intakes in that location. The combined dataframe is shown below.



For each “Neighborhood”, dog intakes are shown below with the intake number as popup.



### 3.2 Modeling

4 groups are preliminary and reasonable to start: large amount of intake areas with little pet stores, small amount of intake areas with enough pet stores, large amount of intake areas with enough pet stores, small amount of intake areas with little pet stores.

It’s clear that the 1st and 4th groups are the choices to build new pet stores. The 1st group is urgent in needing a pet store, which is the top choice.

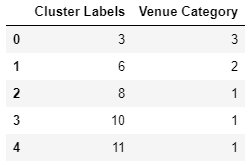
Segmentation and clustering is used to separate neighborhoods into groups. The underlying algorithms are KMeans modeling. Number of clusters is set to 4.

### 3.3 Foursquare Venues

Through access to Foursquare, venues and their category are collected. Then filter out only “Pet” related venues shown as below.

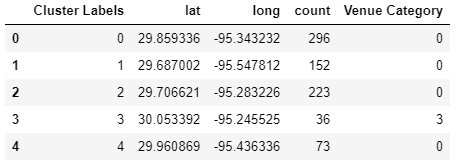


Assume that all kinds of pet stores will help with saving abandoned dogs. So the more detailed types of pet stores are not discussed. Only the total number of pet stores in a neighborhood is considered as below.

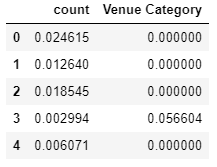


### 3.4 Segmentation and Clustering

Here’s the combined dataframe with “neighborhood” location, intake numbers (count) and pet store count (venue category).

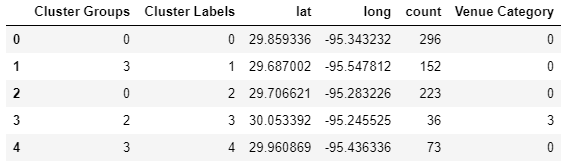


2 features are used to fit the model: intake numbers (count) and pet store count (venue category), which are corresponding to the demand and supply relationship: dog abandon situation and pet store supporting condition. The normalized dataframe is shown below.

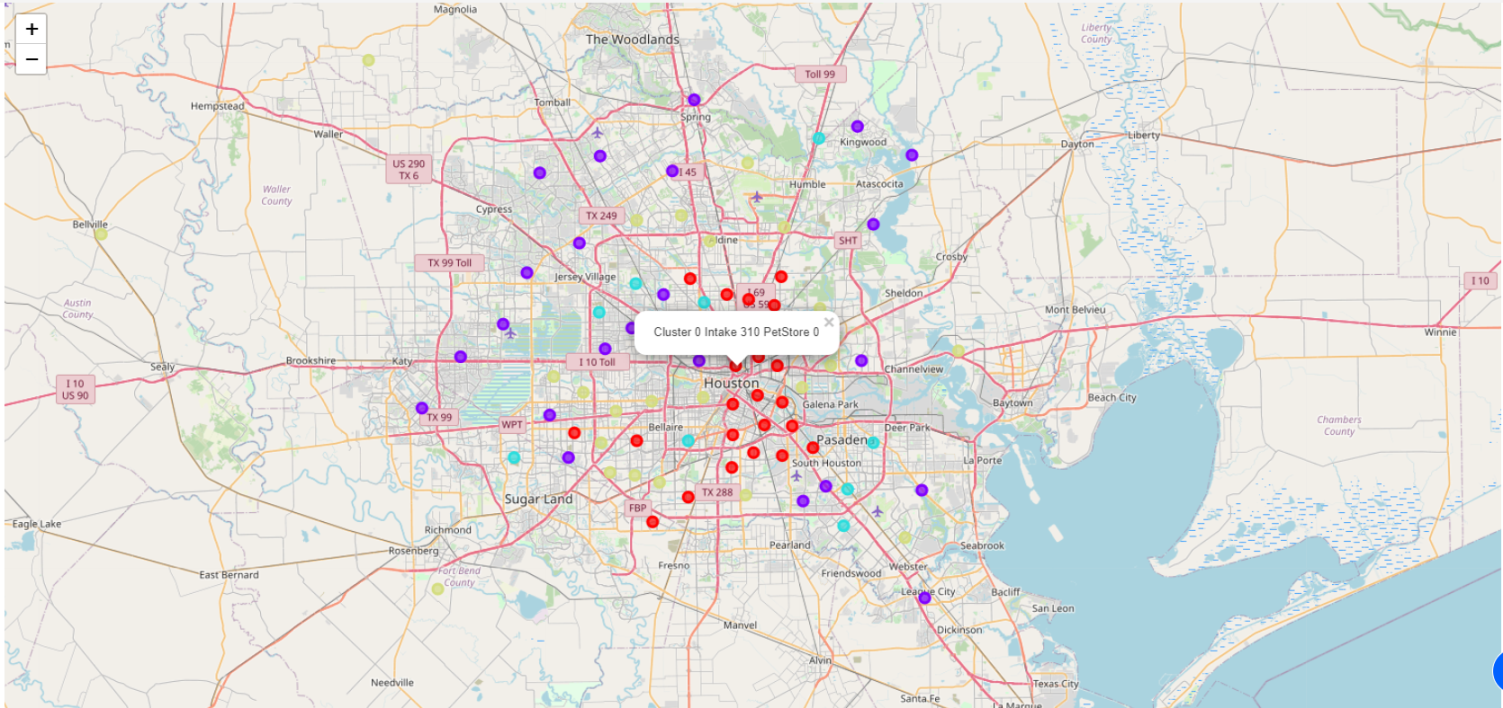


## Results

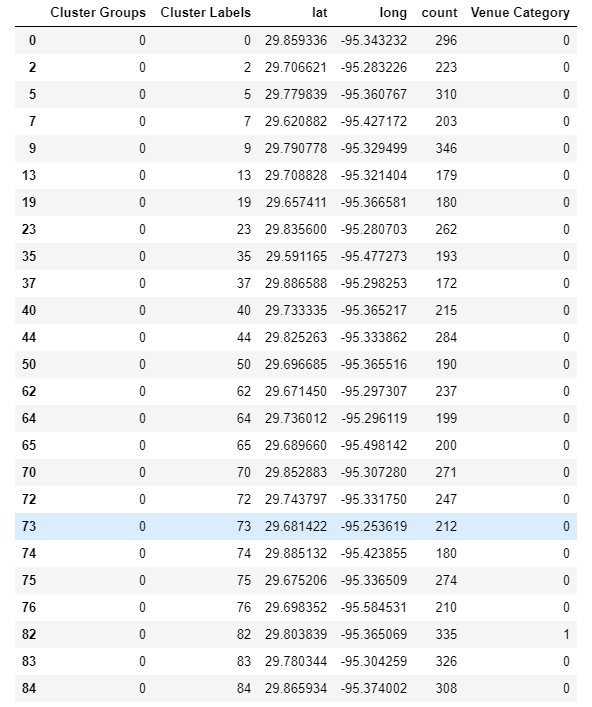
After model training, the clustering result is shown as below with 4 cluster groups and corresponding location, intake count and pet venue count.



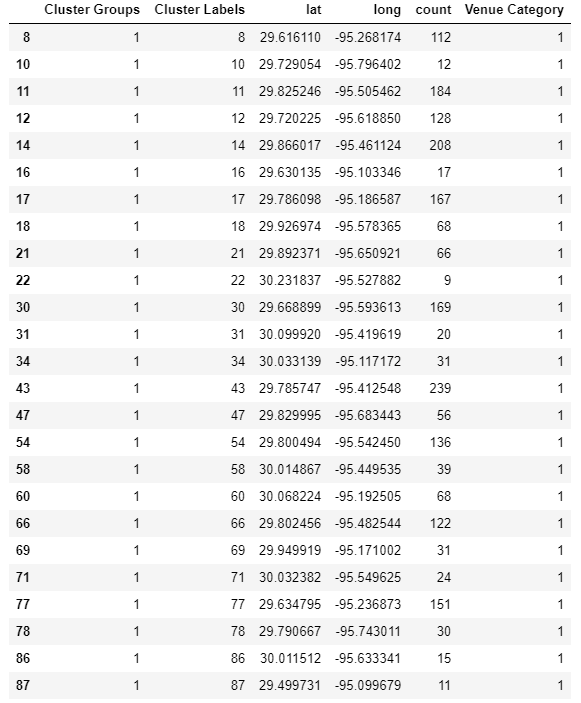
The 4 groups are shown in the map below.



Below are 4 clusters that are generated from KMeans algorithm:

**Cluster 0 Large amounts of intakes areas, small amount of pet stores**

**Cluster 1 Small amount of intakes areas, large amount of pet stores**



#### *Cluster 2* Large amount of intakes *areas*, large amount of pet stores

#### 

#### *Cluster 3* Small amount of intakes *areas*, small amount of pet stores

#### 

## Discussions

Dog intake data is not grouped precisely on Houston super neighborhoods; instead they are grouped from clustering models, which results in 88 “neighborhoods”-like groups. Some extra work can be done to group the intake data strictly on neighborhoods. And results can be compared.

Pet stores are simply summarized based on “neighborhoods”. The specific types of pet stores are not discussed here. However, different types of pet stores may contribute differently about saving abandoned dogs. Extra work can be done to explore the different pet store types.

## Conclusions

In this study, Houston areas are clustered into 4 groups: large amount of intake areas with little pet stores, small amount of intake areas with enough pet stores, large amount of intake areas with enough pet stores, small amount of intake areas with little pet stores.

Areas in the 1st group are best places to build new pet stores. Because of the large amounts of dog intakes and little numbers of pet stores, any new built pet store in this group will help ease the demand and supply relationship and contribute to the community.

## References

1. Wikipedia, List of Houston Super Neighborhoods, <https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods>
2. Open Data Houston, [BARC Dog Intakes for Calendar Years 2011 and 2012](http://data.houstontx.gov/dataset/barc-dog-intakes-for-calendar-years-2011-and-2012), <https://opendatahouston.s3.amazonaws.com/2013-05-13T20:00:59.773Z/barc-dog-intakes-for-calendar-years-2011-and-2012.csv>
3. Foursquare, <https://foursquare.com/>
4. Factors that Influence Intake to One Municipal Animal Control Facility in Florida: A Qualitative Study, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5532563/>