

# The Battle of Neighborhoods

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

Dan Wang  
August 21, 2020

## 1. Introduction

### 1.1 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.

### 1.2 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.

### 1.3 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.

## 2. Data

### 2.1 Data sources

List of Houston Super Neighborhoods can be found in the [Wikipedia page](#). Open Data Houston published data on [BARC Dog Intakes for Calendar Years 2011 and 2012](#). Venues data can be accessed through [Foursquare](#).

### 2.2 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.

## 2.3 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).

#	Name	Location relative to Downtown Houston	Approximate boundaries
0 1	Willowbrook	Northwest	Along Texas State Highway 249 northwest of Bel..
1 2	Greater Greenspoint	North	Around the junction of Beltway 8 and Interstat..
2 3	Carverdale	Northwest	South of the junction of Beltway 8 and U.S. Ro..
3 4	Fairbanks / Northwest Crossing	Northwest	Along U.S. Route 290 between Interstate 610 an..
4 5	Greater Inwood	Northwest	North of Fairbanks / Northwest Crossing and ea..

```
: df.shape
```

```
Out[5]: (88, 1)
```

For BARC 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.

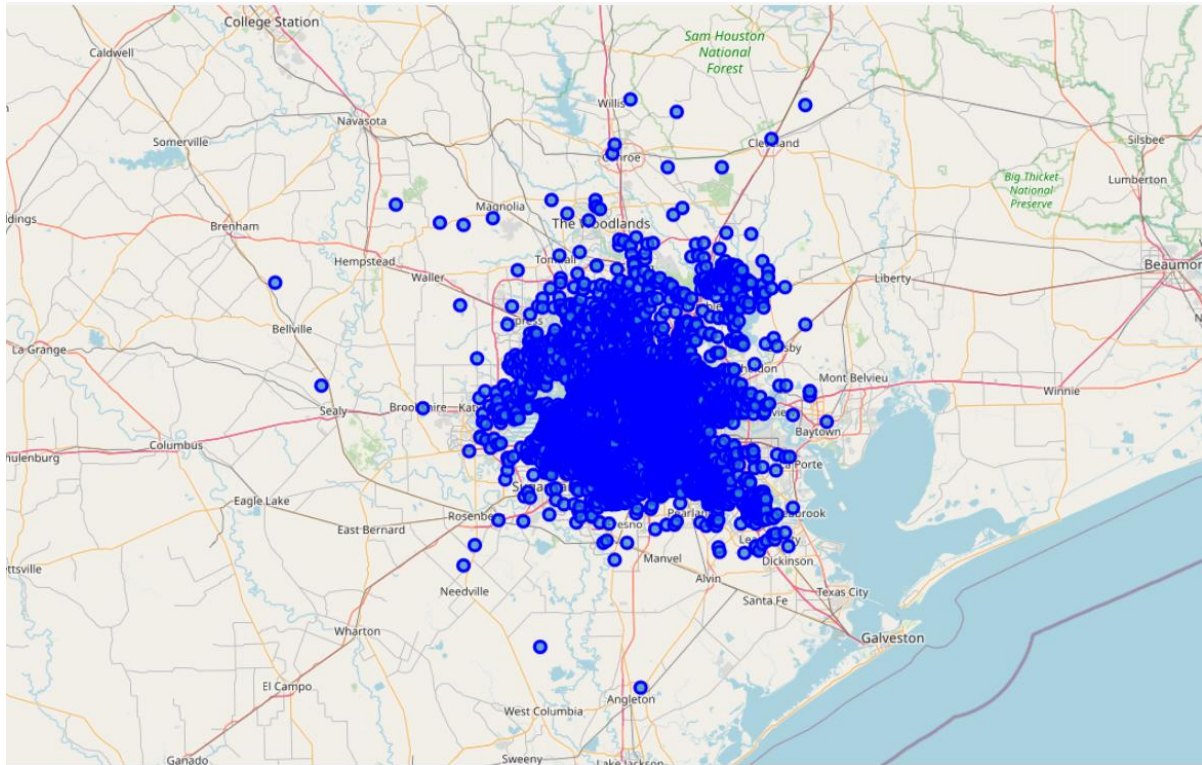
	id	tot	type	breed	sub_breed	in_type	in_sub	condition	in_date	due_date	out_type	out_sub	out_date	zip	street_num	street_name	AC	days	address	lat	long
0	A1059900	1	DOG	PIT BULL	MIX	CONFISCATE	CRUELTY	CONDSEVER	7/18/11	7/23/11	TRANSFER	RESCUE GRP	7/22/11	NaN	17630.0	WAYFOREST	UU	4.0	17630 WAYFOREST	29.954586	-95.391600
1	A1062552	1	DOG	PIT BULL	MIX	CONFISCATE	BITE	NORMAL	8/16/11	8/26/11	RTO	WALKIN	8/27/11	NaN	8016.0	OAKWOOD BEND	H	11.0	8016 OAKWOOD BEND	29.868617	-95.511768
2	A1046026	1	DOG	FLAT COAT RETR	MIX	CONFISCATE	POLICE	NORMAL	1/29/11	2/27/11	TRANSFER	RESCUE GRP	3/11/11	NaN	0.0		H	41.0	0	NaN	NaN
3	A1070150	1	DOG	LABRADOR RETR	MIX	CONFISCATE	BITE	NORMAL	11/16/11	11/26/11	RTO	WALKIN	11/26/11	NaN	15902.0	HEATHERDALE	H	10.0	15902 HEATHERDALE	29.590685	-95.114748
4	A1069139	1	DOG	LABRADOR RETR	NaN	CONFISCATE	BITE	DEAD	11/4/11	11/4/11	DIED	AT VET	11/4/11	NaN	2410.0	GRAMERCY	UU	0.0	2410 GRAMERCY	29.704660	-95.415228

	lat	long
0	30.046815	-96.297647
1	29.830829	-96.184766
2	30.211108	-96.003000
3	29.782836	-95.938521
4	30.172829	-95.897472

## 3. Methodology

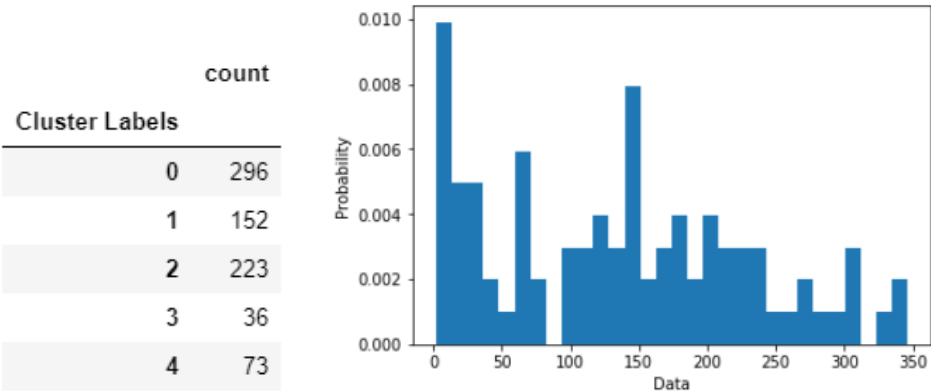
### 3.1 Exploratory Data Analysis

The cleaned BARC 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.

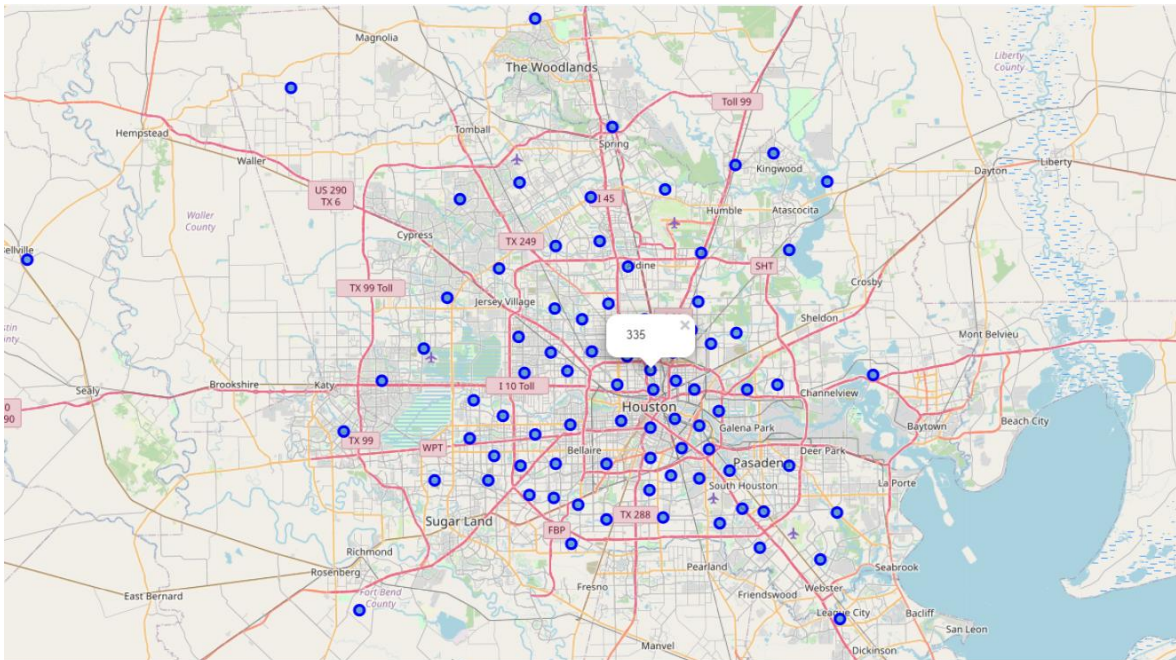
For each group, the total number of intakes is calculated as the “count” column. A histogram chart is also shown below. This dataset is Right-skewed, which means the larger amount of intakes, the smaller the chance is.



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

	Cluster Labels	lat	long	count
0	0	29.859336	-95.343232	296
1	1	29.687002	-95.547812	152
2	2	29.706621	-95.283226	223
3	3	30.053392	-95.245525	36
4	4	29.960869	-95.436336	73

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 1<sup>st</sup> and 4<sup>th</sup> groups are the choices to build new pet stores. The 1<sup>st</sup> 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.

	Cluster Labels	Intake Latitude	Intake Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
329	3	30.053392	-95.245525	Petco	30.049813	-95.244298	Pet Store
387	3	30.053392	-95.245525	PetSmart	30.014236	-95.260582	Pet Store
391	3	30.053392	-95.245525	Animal Ark	30.067745	-95.196330	Pet Store
639	6	29.844780	-95.550552	Aquarium World	29.846201	-95.504434	Pet Store
660	6	29.844780	-95.550552	PetSmart	29.855734	-95.513606	Pet Store

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.

Cluster Labels	Venue Category
0	3
1	6
2	8
3	10
4	11

### 3.4 Segmentation and Clustering

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

	Cluster Labels	lat	long	count	Venue Category
0	0	29.859336	-95.343232	296	0
1	1	29.687002	-95.547812	152	0
2	2	29.706621	-95.283226	223	0
3	3	30.053392	-95.245525	36	3
4	4	29.960869	-95.436336	73	0

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.

	count	Venue Category
0	0.024615	0.000000
1	0.012640	0.000000
2	0.018545	0.000000
3	0.002994	0.056604
4	0.006071	0.000000

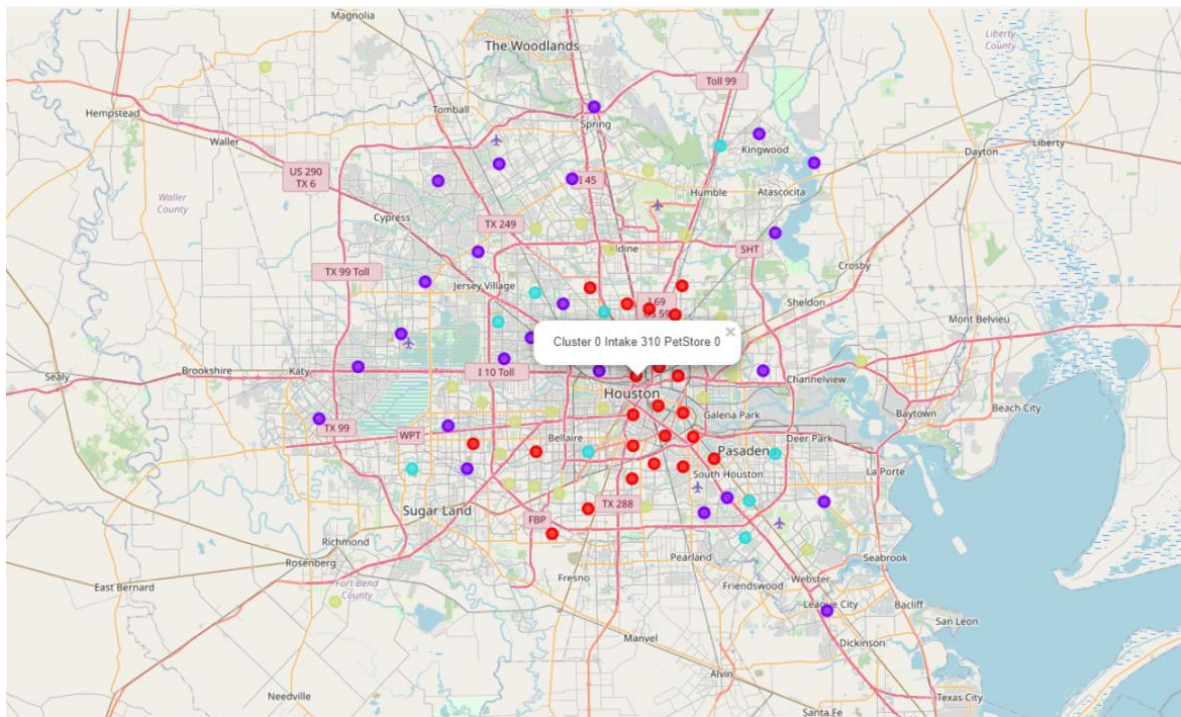


## 4. Results

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

Cluster Groups	Cluster Labels	lat	long	count	Venue Category
0	0	29.859336	-95.343232	296	0
1	3	29.687002	-95.547812	152	0
2	0	29.706621	-95.283226	223	0
3	2	30.053392	-95.245525	36	3
4	3	29.960869	-95.436336	73	0

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 Groups	Cluster Labels	lat	long	count	Venue Category
0	0	0	29.859336 -95.343232	296	0
2	0	2	29.706621 -95.283226	223	0
5	0	5	29.779839 -95.360767	310	0
7	0	7	29.620882 -95.427172	203	0
9	0	9	29.790778 -95.329499	346	0
13	0	13	29.708828 -95.321404	179	0
19	0	19	29.657411 -95.366581	180	0
23	0	23	29.835600 -95.280703	262	0
35	0	35	29.591165 -95.477273	193	0
37	0	37	29.886588 -95.298253	172	0
40	0	40	29.733335 -95.365217	215	0
44	0	44	29.825263 -95.333862	284	0
50	0	50	29.696685 -95.365516	190	0
62	0	62	29.671450 -95.297307	237	0
64	0	64	29.736012 -95.296119	199	0
65	0	65	29.689660 -95.498142	200	0
70	0	70	29.852883 -95.307280	271	0
72	0	72	29.743797 -95.331750	247	0
73	0	73	29.681422 -95.253619	212	0
74	0	74	29.885132 -95.423855	180	0
75	0	75	29.675206 -95.336509	274	0
76	0	76	29.698352 -95.584531	210	0
82	0	82	29.803839 -95.365069	335	1
83	0	83	29.780344 -95.304259	326	0
84	0	84	29.865934 -95.374002	308	0

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

Cluster Groups	Cluster Labels	lat	long	count	Venue Category
8	1	8	29.616110 -95.268174	112	1
10	1	10	29.729054 -95.796402	12	1
11	1	11	29.825246 -95.505462	184	1
12	1	12	29.720225 -95.618850	128	1
14	1	14	29.866017 -95.461124	208	1
16	1	16	29.630135 -95.103346	17	1
17	1	17	29.786098 -95.186587	167	1
18	1	18	29.926974 -95.578365	68	1
21	1	21	29.892371 -95.650921	66	1
22	1	22	30.231837 -95.527882	9	1
30	1	30	29.668899 -95.593613	169	1
31	1	31	30.099920 -95.419619	20	1
34	1	34	30.033139 -95.117172	31	1
43	1	43	29.785747 -95.412548	239	1
47	1	47	29.829995 -95.683443	56	1
54	1	54	29.800494 -95.542450	136	1
58	1	58	30.014867 -95.449535	39	1
60	1	60	30.068224 -95.192505	68	1
66	1	66	29.802456 -95.482544	122	1
69	1	69	29.949919 -95.171002	31	1
71	1	71	30.032382 -95.549625	24	1
77	1	77	29.634795 -95.236873	151	1
78	1	78	29.790667 -95.743011	30	1
86	1	86	30.011512 -95.633341	15	1
87	1	87	29.499731 -95.099679	11	1

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

Cluster Groups	Cluster Labels	lat	long	count	Venue Category
3	2	3	30.053392 -95.245525	36	3
6	2	6	29.844780 -95.550552	78	2
20	2	20	29.820762 -95.397759	232	2
24	2	24	29.688653 -95.427607	124	2
27	2	27	29.631130 -95.205630	66	2
36	2	36	29.687055 -95.170188	19	2
41	2	41	29.835043 -95.366564	308	2
42	2	42	29.879173 -95.499495	148	2
52	2	52	29.855720 -95.405465	220	2
55	2	55	29.668575 -95.668380	40	3
56	2	56	29.826635 -95.447863	220	3
81	2	81	29.586764 -95.212040	99	2

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

Cluster Groups	Cluster Labels	lat	long	count	Venue Category
1	3	1	29.687002 -95.547812	152	0
4	3	4	29.960869 -95.436336	73	0
15	3	15	29.754077 -95.268651	106	0
25	3	25	29.647812 -95.501272	141	0
26	3	26	30.147539 -95.871130	5	0
28	3	28	29.741210 -95.406112	153	0
29	3	29	29.747170 -95.572974	122	0
32	3	32	29.946167 -95.293732	60	0
33	3	33	29.725610 -95.527895	139	0
38	3	38	29.929674 -95.397383	138	0
39	3	39	29.938822 -96.241207	2	0
45	3	45	29.766504 -95.614003	104	0
46	3	46	29.235670 -95.529312	2	0
48	3	48	30.352693 -95.102322	3	0
49	3	49	29.509574 -95.774990	4	0
51	3	51	30.024424 -95.345320	33	0
53	3	53	29.798058 -95.052581	12	0
57	3	57	29.848608 -95.244459	142	0
59	3	59	29.650986 -95.535997	114	0
61	3	61	29.571889 -95.126620	105	0
63	3	63	29.736851 -95.477553	150	0
67	3	67	29.780400 -95.230067	147	0
68	3	68	29.623322 -95.347853	150	0
79	3	79	29.639170 -95.466612	148	0
80	3	80	30.304163 -95.345575	5	0
85	3	85	29.954709 -95.499169	60	0

## 5. 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.

## 6. 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 1<sup>st</sup> 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](https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods)
2. Open Data Houston, 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/>