The Battle of Neighborhoods

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

Dan Wang

August 21, 2020

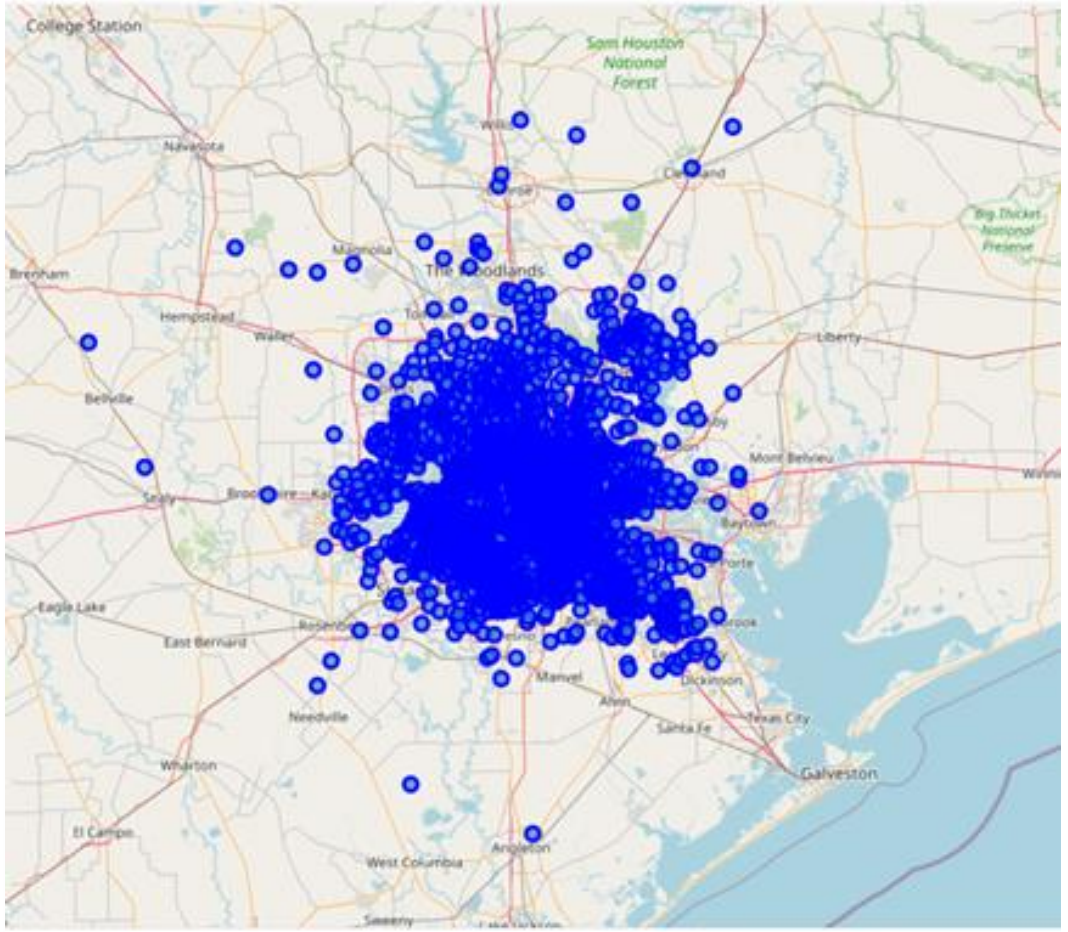
1 Introduction

- Where to Build New Pet Store in a city is valuable to pet store owners, pet organizations, pet owners, etc.
- pet related organizations or stores are not evenly distributed in neighborhoods compared to the abandoned location distributions.
- Datasets include Houston BARC dog intake table and Foursquare pet venues.
- Segmentation and custering is used to separate Houston areas into groups, based on which to choose new pet store locations.

2 Data Acuisition, Cleaning and EDA

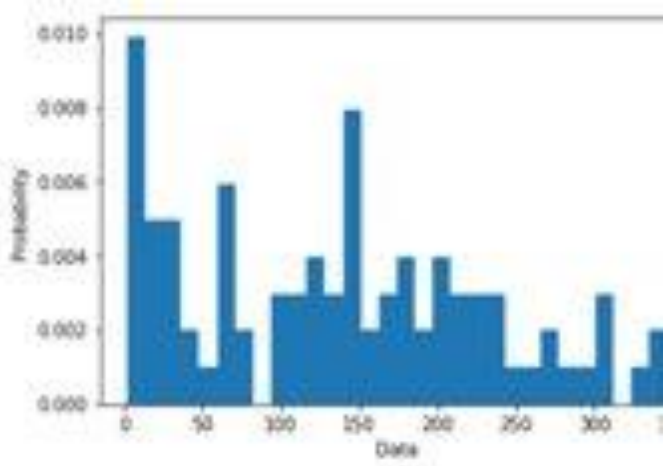
- List of Houston Super Neighborhoods can be found in the Wikipedia page.
- Open Data Houston published data on BARC Dog Intakes for Calendar Years 2011 and 2012.
- Original table includes more than 30000 instances.
- 2 features are kept: latitude and longitude

	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



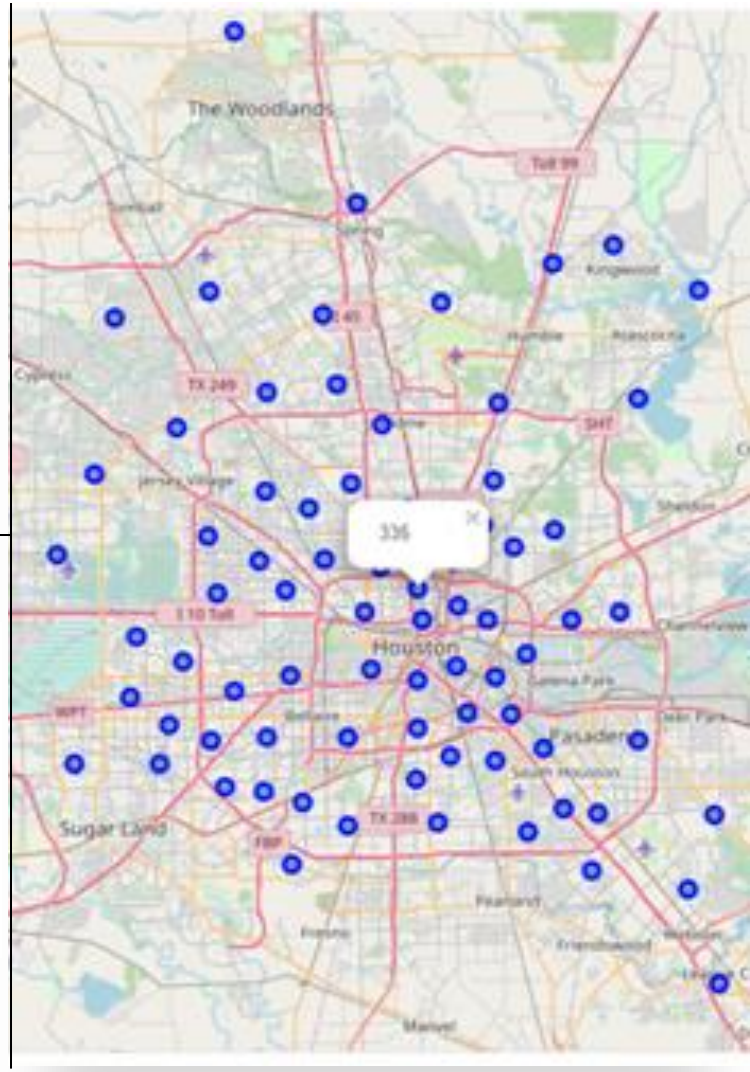
Original Houston BARC Dog intake location data distributions

2 Data Acuisition, Cleaning and EDA – cont'd



- Original intake instances are clustered into 88 groups based on location
- the total number of intakes is calculated as the “count” column
- A histogram chart is shown
- For each “Neighborhood”, dog intakes are shown with the intake number as popup

	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



3 Clustering

	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

Foursquare venues

	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

combined dataframe

	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

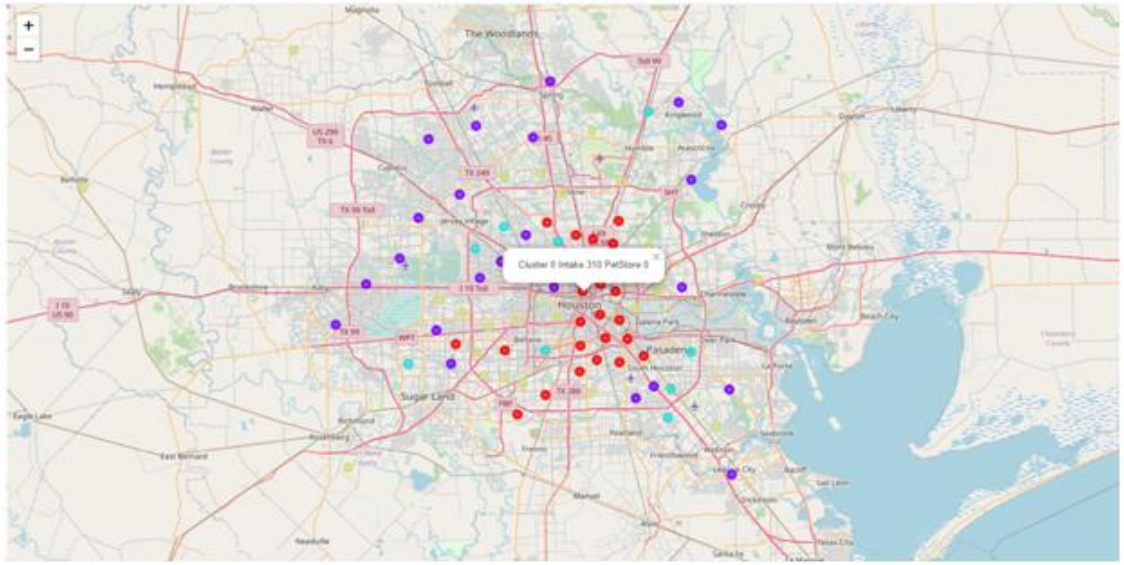
normalized training dataframe

- Segmentation and clustering is used to separate neighborhoods into groups
- Algorithms: KMeans modeling. Number of clusters: 4
- 2 features are used to fit the model: intake numbers (count) and pet store count (venue category)

4 Results

4 clusters:

- *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*



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

5 Conclusions

- Houston areas are clustered into 4 groups.
- Areas in the 1st group (Cluster 0 Large amounts of intakes areas, small amount of pet stores) 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.