Final Project: Austin Animal Shelter

Introduction

The dataset our team decided to analyze centered around Austin Animal Center intakes and outcomes; we specifically wanted to determine what factors have the most critical impact on adoption rates for pets. This issue is important to the Austin community, as the city is a no-kill zone and non-profits such as Austin Pets Alive are consistently trying to promote adoption/foster care. If we can find certain trends in the data such as adoption rates over a given time period or the popularity of certain breeds, these organizations can better fit their message and help these lost animals find a home. This would also be a great resource for people looking to adopt an animal from a shelter, as better information about what types of pets are in need can help them target specific types and breeds. These informed decisions can curb overpopulation of strays in shelters and Austin as a whole, offering societal and psychological benefits to the Austin community. It should be noted that this dataset only contains data from late 2013 through early 2018.

Filtering the initial dataset became an important focus:

- breaking the animal type column down into simply the popular categories of animals and 'other' (dog, cat, bird, other)
- grouping breeds/colors into those that appeared more than 100+ times in the data and 'other'
- Excluding 'Return to Owner' category from 'outcome type' column in regressions as it would distort dataset
- Formatting certain columns (days in shelter, gender) to condense the string or alter the data type, which allows for more efficient classification and regression.
- Creating bins for duration at shelter and for age; as variables are continuous, organizing them into bins allows for simplified analysis

Exploratory Analysis

Some basic insights that our team found fit well with rather intuitive assumptions about what adoption rates should look like in Austin. Dogs are the largest percentage of local shelter animals (57%), with over 45,000 coming into Austin shelters over a five-year period (Figure 1). Another useful insight was that the length of stays at animal shelters around Austin were mostly less than a week even after accounting for returns to owners (Figure 2). Since our problem's focus is centered around commonly adopted pets, we also wanted to have summaries of length of stays. When only accounting for gender and adoption status, we found that while there was not a significant discrepancy between genders in an animal group, dogs were adopted at a faster rate than cats with an average time spent in shelter before adoption of 12 days less than their feline counterparts. Another interesting aspect of the data was the adoption rate of animals in regards to their age. While puppies and other younger animals receive a large amount of attention and care from the general population, we wanted to see statistically what that looked like for dog adoption rates in Austin. By grouping results into four main categories (adoption, euthanasia, return to owner, transfer), we found that adoption rates do match our intuition when based on age of the pet (Figure 3). Puppies that were between 1-3 months old were the most highly adopted, and age seems to have an inverse relationship with adoption rates.

Another valuable analysis would be to look at which months seem to be more popular for adoption. The three most popular intake months are May, October and June (in order). The most popular outcome months are July, October and August. This seems pretty in line with the fact that the average time in the shelter is 16 days, because all the popular outcome months are close to their intake month. The popular months also (mostly) correlate with summer time, when families may have more time to care for and own a pet, which could explain the trend. This was proven when breaking the data up by season; the summer had the highest adoption rates over the 5-year period of 28.9%. This can encourage the shelter to put more resources into promoting adoption during slower months. The average time spent in the shelters seems relatively consistent over the course of the past 5 years (Figure 4). Finally, it looks like, on average, adoption rates for popular animals (dogs, cats) have steadily increased in Austin over the past 5 years (Figure 5). The ratios presented in Figure 5 are the ratio of dogs that are adopted from total dogs in shelters for that year. Among these dogs, certain breeds have higher adoption rates, which could be explained by general community affinity towards certain breeds (Figure 6).

Solution and Insights

In order to have a holistic interpretation of the data, the team decided to use nearest neighbors, naïve bayes, decision trees and logistic regression to classify the data and draw our conclusions. Through intuition and testing, we identified important features, such as age, sex upon outcome (neutered or not), animal type, and time in shelter.

When looking into nearest neighbors classification, we focused in on an animal’s gender, type, age, breed, intake condition, and time in shelter. Some of these variables were put into bins (age, breed, time in shelter) to make the analysis more efficient; there was only a 1% accuracy difference between 50 neighbors and 200 neighbors. We got a 69% accuracy, 12% higher than a random guess, which will allow 12% AAC to better predict whether or not an animal will be adopted.

For the naïve bayes classification, we got an 80.4% accuracy when determining whether a pet was adopted, about 23% higher than a random guess. An important factor to look into for naïve bayes is the importance of certain factors in the positive or negative class (adopted or not). These are indicated by the 'importance score' after fitting training data to the classification model. The top indices for this particular model were gender upon outcome; an intact male/female pet (not spayed or neutered) had a higher likelihood of not being adopted. This could be explained by general Austin pet health codes, as well as the danger and responsibility of unwanted animal reproduction after adoption. Aside from certain breeds having a higher likelihood of no adoption, there is also a higher likelihood of a pet not being adopted if it is feral or sick when being initially taken into the shelter. This could be explained on one side from people avoiding sick/unhealthy pets, and on the other side from the possibility that shelters are not equipped with the proper resources to heal these sick animals (Figure 7). This is further explained when we looked at animals with a "Euthanasia Request" intake_type. Out of these animals, the percentage of with a Euthanasia outcome_type was 78.1% (compared to the overall average of only 7.8%). Examining these differing percentages gives more validity to our observation that the condition the animal comes to the shelter in and the requests of the owner/person turning in the animal affects adoption rates.

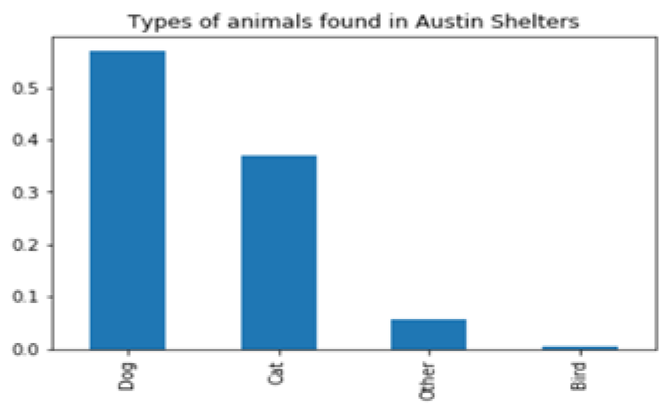
For our decision tree classification, we ran a cross-validation to find the optimal max_depth of 4 and created a tree included below (Figure 8). Some features included in the classification were the shelter outcome, age, intake condition and reason for admittance, animal type and breed. Our goal was to determine if the animal had been in the shelter for more (negative class) or less time (positive class) than the average stay of 16.9 days. While several of the paths led to high entropy, meaning the split was not pure, there were a few paths that led to low entropy/a significant reduction. The first significant path was that the animal was not adopted, was returned to the owner but not an owner surrender and was not a Pitbull. This path had 0.078 entropy and 99.05% of theses samples were in the shelter for less than 16.9 days. The second significant path was a cat that was not adopted and 4-weeks old upon intake, leading to entropy of 0.025. 99.76% of these animals were in the shelter for more than the average time, which makes sense because they were too young to be taken out of the shelter within 16 days.

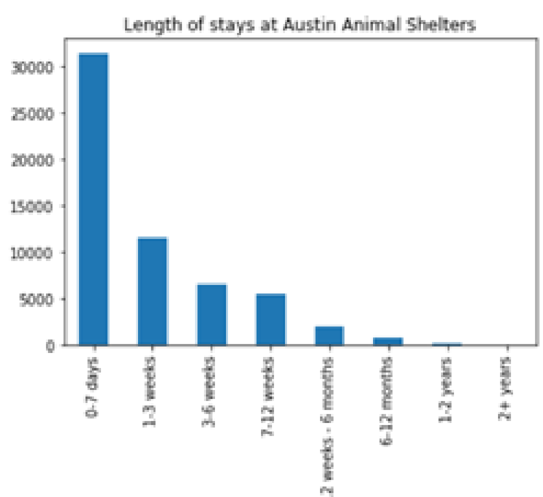
Our final classification was logistic regression, we used all the provided variables except for obvious cases of dependence between features. For each categorical variable, the model uses a binary dummy variable for each feature. The regression classified whether cats and dogs were adopted or not with ~84% accuracy (27% higher than a random guess). By far, the most important variable was the sex of the animal upon outcome. Neutered animals were found to be much more likely to be adopted than intact males and females. The results here mirror the solutions when applying the naïve bayes model. Aside from sex upon outcome, breed and outcome hour also had relatively high weights in the logistic regression. The outcome hour could be significant and possibly explained by standard transfer/euthanasia schedules that the shelter may have, while adoption typically would occur only during daytime hours.

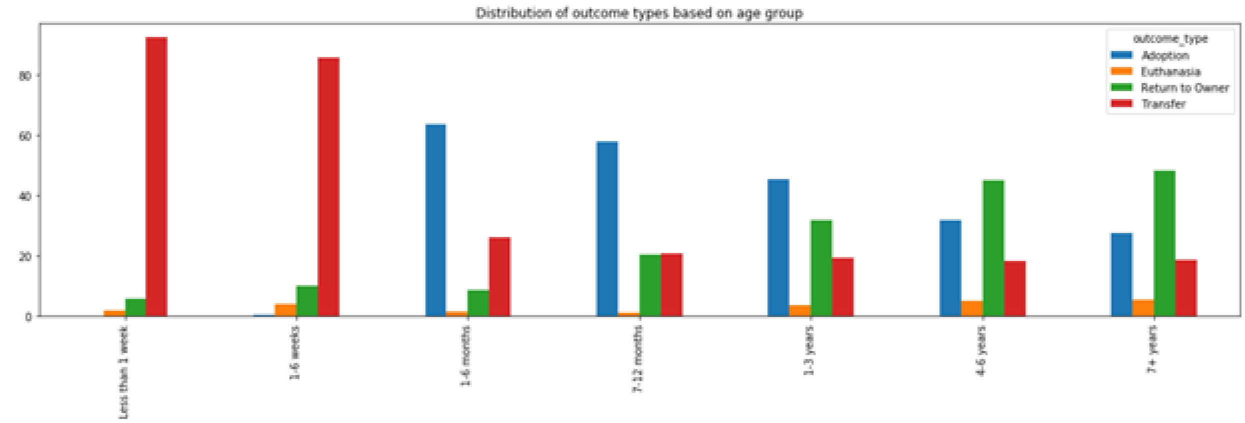
Final Thoughts

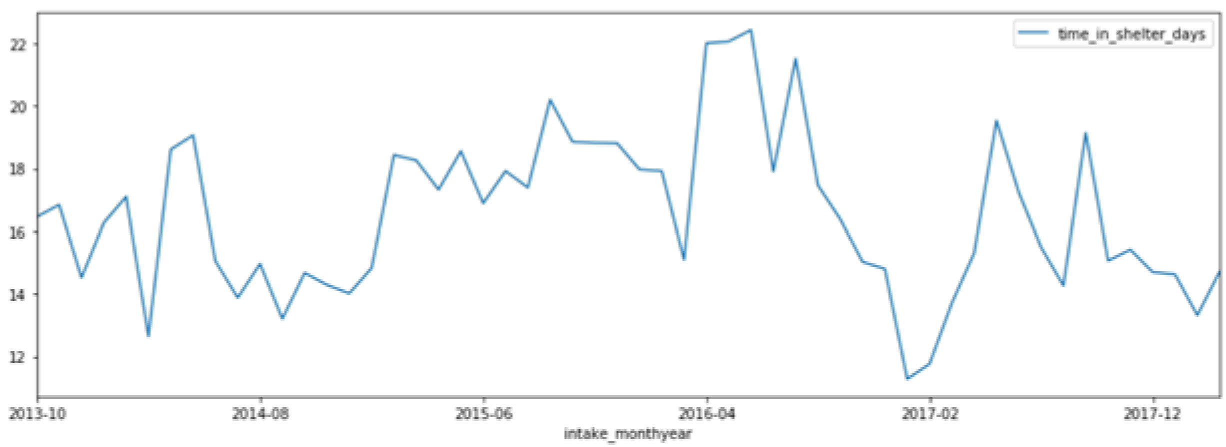
After running four types of classifications on the data, we can say with about 80% confidence that sex upon outcome of the pet is the dominant feature in determining adoption rates. As explained previously, this intuitively makes sense when compared to animal welfare and regulation policies in Austin. We also observed that the type of animal (dog versus cat), age and time of year all affect adoption rates. Furthermore, while there are many variables that could potentially effect adoption rates, we find that other important features include breed, time of outcome, and age. The City of Austin could potentially use this information to promote adoption initiatives for less popular breeds, as well as implementing health policies (spaying and neutering) to potentially bolster adoption rates. With the city's current no-kill policy, it becomes imperative that we keep the population of strays in check and direct funds to shelters that need them the most

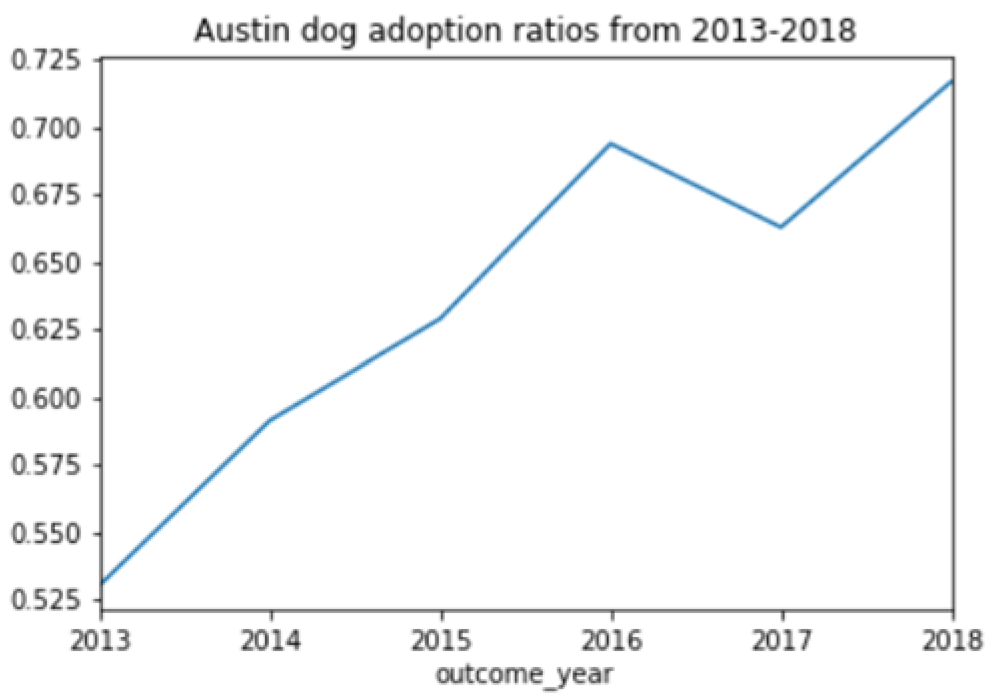
Graphs and Charts

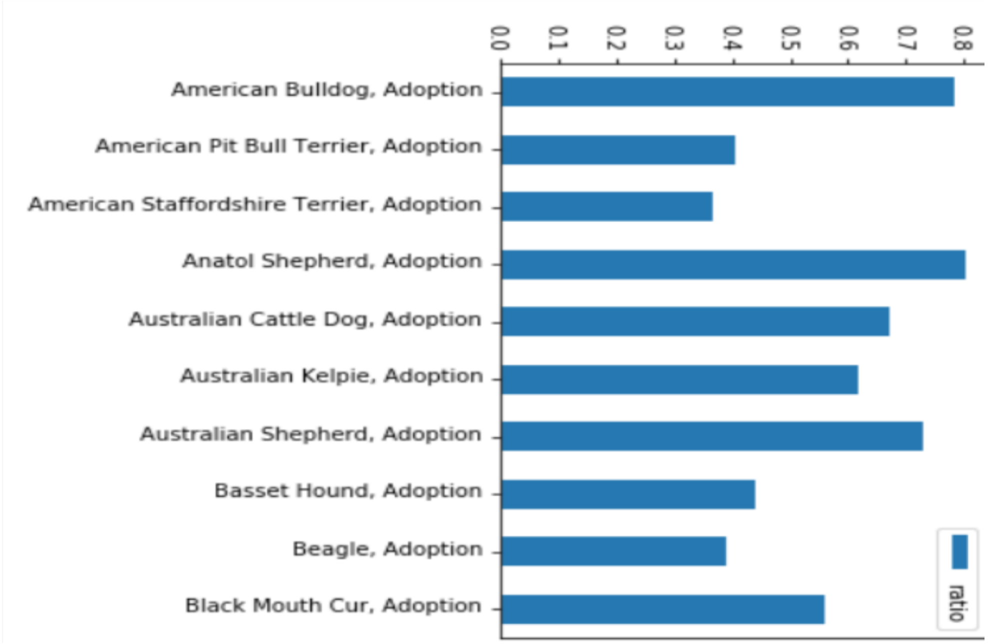




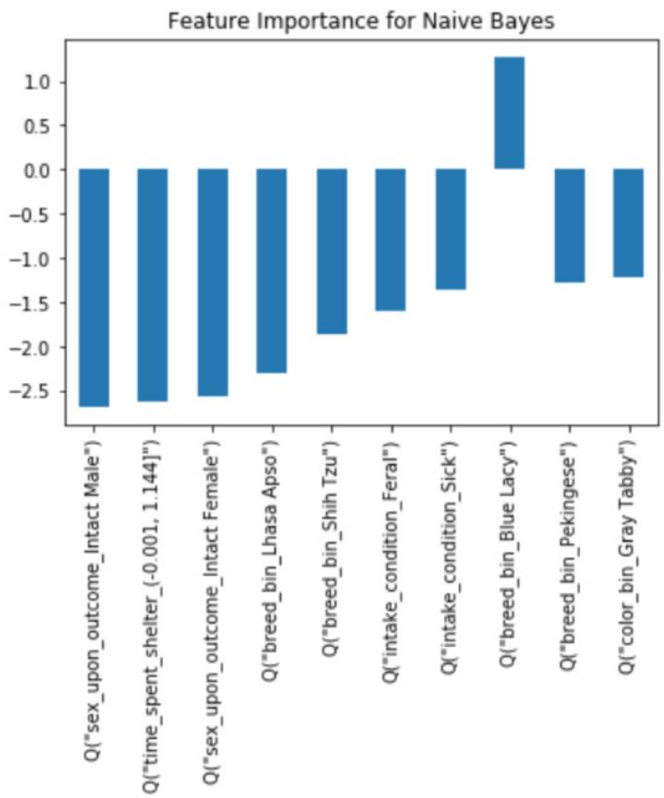

Distribution of outcome types based on age group
outcome_type
Adoption
Euthanasia
Return to Owner
Transfer
0
20
40
60
80
Less than 1 week
1-6 weeks
1-6 months
7-12 months
1-3 years
4-6 years
7+ years


time_in_shelter_days
12
14
16
18
20
22
2013-10
2014-08
2015-06
2016-04
2017-02
2017-12
intake_monthyear

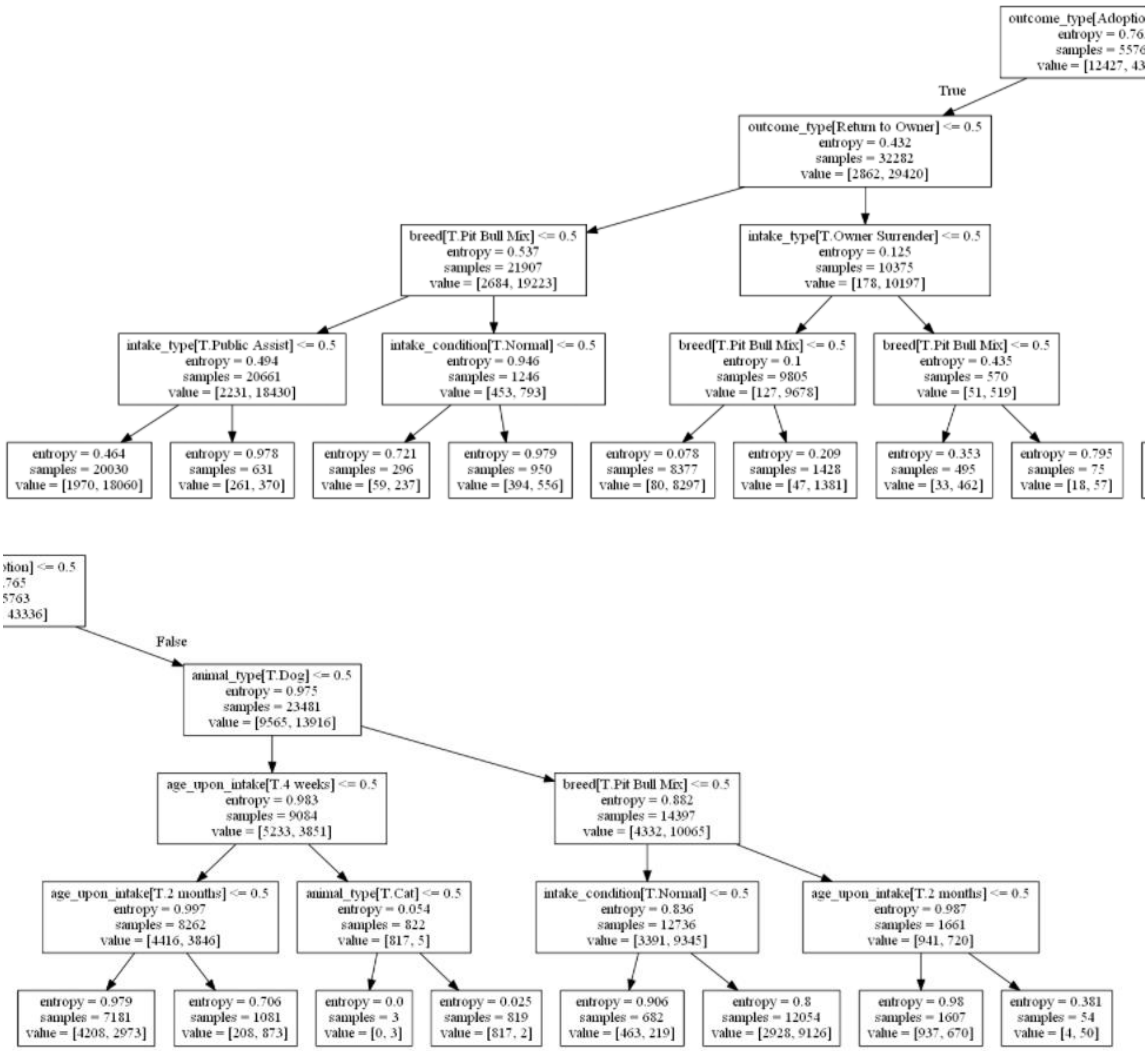

Austin dog adoption ratios from 2013-2018
0.525
0.550
0.575
0.600
0.625
0.650
0.675
0.700
0.725
2013
2014
2015
2016
2017
2018
outcome_year



American Bulldog, Adoption
American Pit Bull Terrier, Adoption
American Staffordshire Terrier, Adoption
Anatol Shepherd, Adoption
Australian Cattle Dog, Adoption
Australian Kelpie, Adoption
Australian Shepherd, Adoption
Basset Hound, Adoption
Beagle, Adoption
Black Mouth Cur, Adoption
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
ratio



Feature Importance for Naive Bayes
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
Q("sex_upon_outcome_Intact Male")
Q("time_spent_shelter_(-0.001, 1.144]")
Q("sex_upon_outcome_Intact Female")
Q("breed_bin_Lhasa Apso")
Q("breed_bin_Shih Tzu")
Q("intake_condition_Feral")
Q("intake_condition_Sick")
Q("breed_bin_Blue Lacy")
Q("breed_bin_Pekingese")
Q("color_bin_Gray Tabby")



outcome_type[Adoptio
entropy = 0.76
samples = 5576
value = [12427, 43
True
outcome_type[Return to Owner] <= 0.5
entropy = 0.432
samples = 32282
value = [2862, 29420]
breed[T.Pit Bull Mix] <= 0.5
entropy = 0.537
samples = 21907
value = [2684, 19223]
intake_type[T.Owner Surrender] <= 0.5
entropy = 0.125
samples = 10375
value = [178, 10197]
intake_type[T.Public Assist] <= 0.5
entropy = 0.494
samples = 20661
value = [2231, 18430]
intake_condition[T.Normal] <= 0.5
entropy = 0.946
samples = 1246
value = [453, 793]
breed[T.Pit Bull Mix] <= 0.5
entropy = 0.1
samples = 9805
value = [127, 9678]
breed[T.Pit Bull Mix] <= 0.5
entropy = 0.435
samples = 570
value = [51, 519]
entropy = 0.464
samples = 20030
value = [1970, 18060]
entropy = 0.978
samples = 631
value = [261, 370]
entropy = 0.721
samples = 296
value = [59, 237]
entropy = 0.979
samples = 950
value = [394, 556]
entropy = 0.078
samples = 8377
value = [80, 8297]
entropy = 0.209
samples = 1428
value = [47, 1381]
entropy = 0.353
samples = 495
value = [33, 462]
entropy = 0.795
samples = 75
value = [18, 57]
ption] <= 0.5
.765
5763
43336]
False
animal_type[T.Dog] <= 0.5
entropy = 0.975
samples = 23481
value = [9565, 13916]
age_upon_intake[T.4 weeks] <= 0.5
entropy = 0.983
samples = 9084
value = [5233, 3851]
breed[T.Pit Bull Mix] <= 0.5
entropy = 0.882
samples = 14397
value = [4332, 10065]
age_upon_intake[T.2 months] <= 0.5
entropy = 0.997
samples = 8262
value = [4416, 3846]
animal_type[T.Cat] <= 0.5
entropy = 0.054
samples = 822
value = [817, 5]
intake_condition[T.Normal] <= 0.5
entropy = 0.836
samples = 12736
value = [3391, 9345]
age_upon_intake[T.2 months] <= 0.5
entropy = 0.987
samples = 1661
value = [941, 720]
entropy = 0.979
samples = 7181
value = [4208, 2973]
entropy = 0.706
samples = 1081
value = [208, 873]
entropy = 0.0
samples = 3
value = [0, 3]
entropy = 0.025
samples = 819
value = [817, 2]
entropy = 0.906
samples = 682
value = [463, 219]
entropy = 0.8
samples = 12054
value = [2928, 9126]
entropy = 0.98
samples = 1607
value = [937, 670]
entropy = 0.381
samples = 54
value = [4, 50]