Data Analytics Project: Austin Animal Shelter

Project Goals:

We are analyzing a dataset of intakes and outcomes from Austin Animal Shelter to understand animal adoption trends, including which attributes of animals result in a higher likelihood of adoption. Our final goal is to predict whether or not an animal will be adopted based on characteristics Austin Animal shelter can identify upon intake.

This problem is important because according to the ASPCA, 6.5 million animals enter animal shelters each year. Furthermore, each year 1.5 million shelter animals are euthanized. We chose this problem because as Austin continues to grow and continue to be a pet-friendly city, naturally the number of animal intakes will continue to increase and the Austin Animal Shelter is already experiencing tumult from having to turn away animals. To avoid this problem, the shelter is able to transfer animals to other shelters, partner with fosters, and other mechanisms. We hope our analysis will provide detailed insight into the leading factors of whether an animal will be adopted or not to inform the operations of the Austin Animal Shelter.

Exploratory Analysis:

We used datasets of Austin Animal Shelter intakes and outcomes found on http://data.austintexas.gov. We performed the following steps to clean our data and extract the features needed for our analysis and model:

1. Performed an inner join on intakes/outcomes datasets on Animal ID
2. Extracted the gender and spayed/neutered status for each of the intake and outcome columns
3. Calculated duration of stay, placed animals in "age buckets" and "duration buckets" to simplify analysis of these continuous variables
4. Created boolean fields for specific breed terms (i.e. pit bull chihuahua mix would be given a 1 in both the pit bull and chihuahua columns)

Intake/Arrival

How do animals arrive at the shelter, and how does this relate to outcomes?

68.2% of shelter intakes are strays, 20.9% are owner surrender, and only .35% are euthanasia requests (see *Figure 1*). Fortunately, 89% of animals come into the shelter without being injured or sick. Intake type is actually fairly indicative of outcome. For example, 62.8% of owner surrenders were adopted, 66.6% of public assists were returned to owner and 71.4% of euthanasia requests were euthanized. However, the top 3 reasons for being euthanized were suffering, rabies risk, and aggressive.

Stay

What does the average stay look like for animals in the animal shelter?

The mean duration of stay is 15 days; however, because of a few extreme outliers that highly skewed the data, we believe that the median, which comes out to be slightly under 5 days, is a more accurate representation of average time spent (see *Figure 2*). Looking at the distribution of animals in the shelter, we see that dogs and cats account for over 95% of all intakes (62.5% and 32.6% respectively). Stay for dogs and cats is pretty similar, with around 55% in the 0-7 days

group for both of them. Stay for "Other" and "Bird" is substantially different, with 92.8% and 64%, respectively, for the 0-7 days group. The difference in stay between them is, perhaps, because of the lower number of records. "Other" can also include wildlife which is largely euthanized and therefore decreases their stay in the shelter.

Is there a difference between length of stay for adopted Dogs and Cats?
The mean stay for dogs is 14 days, while for cats the mean stay is 18 days. Again, because of some extreme outliers, the median is a better representation of average stay, which comes out to be 5 days for both dogs and cats (see *Figure 3* and *Figure 4*). When considering only "Adoption" as outcome, we can see that cats spend a significantly longer amount of time in the shelter compared to dogs. When we look at the male vs female breakdown, we see that both male dogs and cats are adopted faster (20 and 30 days on average, respectively) than their female counterparts (22 and 33 days). When we look at the age breakdown, dogs are adopted faster at around early adulthood. The average length of stay for a newborn puppy is slightly over 70 days, which equates to roughly two and a half months, while the shortest average length of stay for dogs was in the 1-3 year old range, with 14.8 days. For cats, we can see that the shortest length of stay was in the 1-3 year old range at nearly 14 days, followed by the 1-6 months range at 27 days. This allows to conclude that adopters of cats tend to go for cats that are in early adulthood, similar to adopter of dogs. The longest stays were for very young cats with less than a week of age (close to 87 days).

How many animals are a euthanasia request and don't get euthanized?
Out of all the intakes that had a request for Euthanasia, 77 (29%) were not euthanized. Majority of these were either transferred or adopted (35 and 17 respectively). Continuing with our findings, we discovered that 35% of all the intakes were neutered or spayed during their time in the shelter. Spaying and neutering animals is necessary to reduce overcrowding in the shelter, as overcrowding can put a strain on the number of resources the shelter can provide to animals while they wait for homes.

Outcome/Result
Are there certain months that people tend to make more adoptions on? What about certain days of the week (weekends)?
The datasets we obtained range from October of 2013 to April of 2017 (see *Figure 5* for a distribution of month and year). Overall, most adoptions occurred during Saturdays and Sundays, with 21.6% and 20.1% respectively (see *Figure 6)*. The most popular month to adopt was July, with 10.5% of all adoptions (see *Figure 7*). There seems to be some seasonality when it comes to adoptions. Holiday months, such as December and January are the next most popular months for adoptions, with 10% and 9.1% respectively.

How does the likelihood of adoption change based on type, gender, and age?
The likelihood of adoption increases if the animal is a dog. A dog's percentage of being adopted is 44.68% vs a cat's 42.5%. Dogs also tend to have lower euthanization percentages. Surprisingly, more dogs are returned to their owners than cats by a substantial margin: 33.3% vs 5.4% respectively. Finally, animals that fall in the 1-6 months and 7-12 months range are more likely to get adopted.

Our Predictive Models

For our predictive analysis, we wanted to predict whether or not an animal will be adopted based on the features we had available. For our predictive models, we used the following:

- K-Nearest Neighbors: We used this model because we thought that there might be distinct clusters around the factors we viewed in our exploratory analysis, but wanted to see if there might be clusters we may have missed with other factor combinations.
- Naive Bayes: We wanted to try a Naive Bayes model because a lot of the features are independent of one another, such as Gender and Animal Intake Type or Intake Condition.
- Classification Tree: We wanted to see if the data could be aligned on certain features like we looked into on our exploratory analysis.
- Random Forest: We wanted to include an ensemble method and we chose random forest because classification trees are easy to explain and we want our findings to be actionable for Austin Animal Shelter.

As we tested each predictive model, we evaluated the predictors that we were using during the process. We started with Gender, Age_Bucket, Animal_Type_intake, Intake_Type, and IntakeCondition; we chose these features because we used many of them in our exploratory analysis. These features allowed us to create a basic model, but the test accuracy was not much better than our baseline prediction accuracy. To improve our accuracy, we added in the boolean features for dog type (ex: “retriever”) and a feature that measured whether or not the animal came in with a name (dog’s name or “No name” under name_intake). We added these features because we wanted to explore whether a dog’s breed affects adoption; additionally, we believe that many dogs who have names (and cute ones at that) would see more interest and we wanted to see if the data supported that idea.

Before diving into our results, we first calculated our baseline prediction accuracy by finding the majority case of adopted divided over the total cases. For our data, the majority case (Not Adopted) is 57.9% so we want our predictive models to have better accuracy than this. Below are the results for our models using all of the predictors previously mentioned.

Classification Tree (Using a max depth of 3)	Training: 65.2%	Test: 65.4%
K-Nearest Neighbors (Using 50 Neighbors)	Training: 67.7%	Test: 67.6%
Naive Bayes	Training: 65.3%	Test: 65.8%
Random Forest (Using 100 estimators and max depth of 3)	Training: 65.4%	Test: 69.2%

Using these results, the random forest model is the predictive model that works best for our predictive task of which animal will get adopted or not. The model, with a test accuracy of 69.2%, adds over 10% to the predictive task over chance. 10% in the context of 70,000 cases is quite large - this will allow the Austin Animal Shelter to more accurately predict whether or not an animal will get adopted for 7,000 more animals.

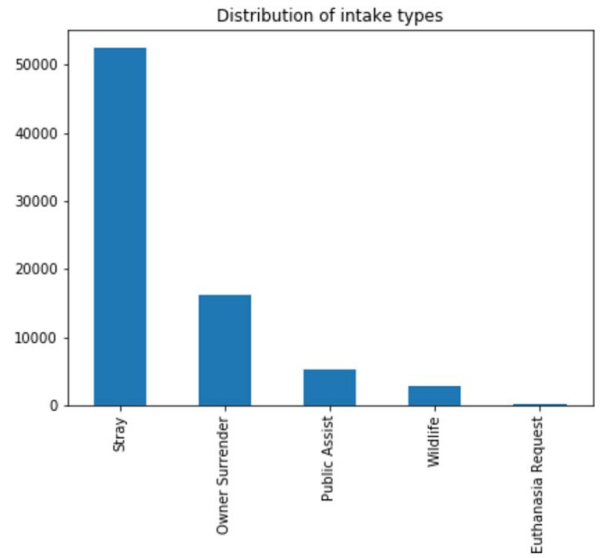
We found a classification tree model with a higher test accuracy (69.8%) but the tree was unpruned, meaning that it was next to impossible to actually use in a real context. Our pruned classification tree (and what one of the random forest trees may look like) can be seen in Figure 8. The relative results can be seen for our predictions in the values included in the leaf nodes as the bottom of the classification tree. As you can see with the classification tree, the most important feature used was the Intake_Type. If the animal was an owner surrender, they are 2x as likely of being adopted, whereas if they were not owner surrendered they are .66x as likely to be adopted. An interesting note is that the name_start feature *does* help with the predictive power; this factor comes into play in the 3rd level of tree depth. Adding this feature to the model increase the test accuracy by ~3% on the classification model.

Implications:

The ability to be able to predict the likelihood of an animal being adopted is powerful. Using our model, the Austin Animal Shelter would be able to better prioritize animals that have a low likelihood of being adopted in terms of featuring these animals on their website, requesting transfers, and setting up foster programs. Furthermore, knowing that an animal having a name is a strong predictor of adoption suggests Austin Animal Shelter should try naming the unnamed animals to see if that increases adoption rates. Lastly, Austin Animal Shelter can use our exploratory analysis findings to prioritize intakes by gender, age and breed when they are limited on space.

Appendix

Figure 1



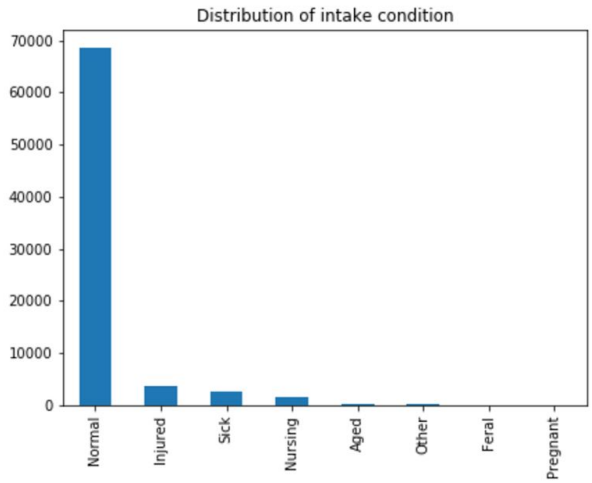


Figure 2

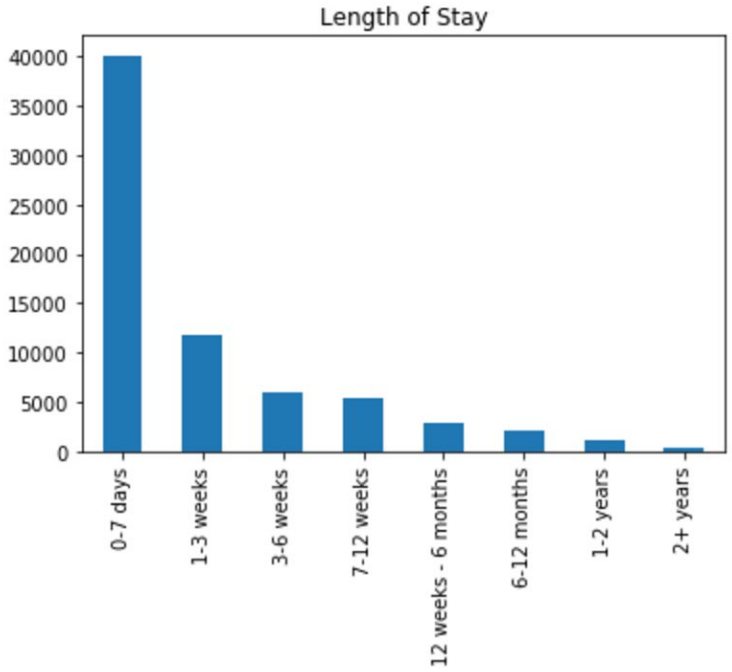

Length of Stay
40000
35000
30000
25000
20000
15000
10000
5000
0
0-7 days
1-3 weeks
3-6 weeks
7-12 weeks
12 weeks - 6 months
6-12 months
1-2 years
2+ years

Figure 3

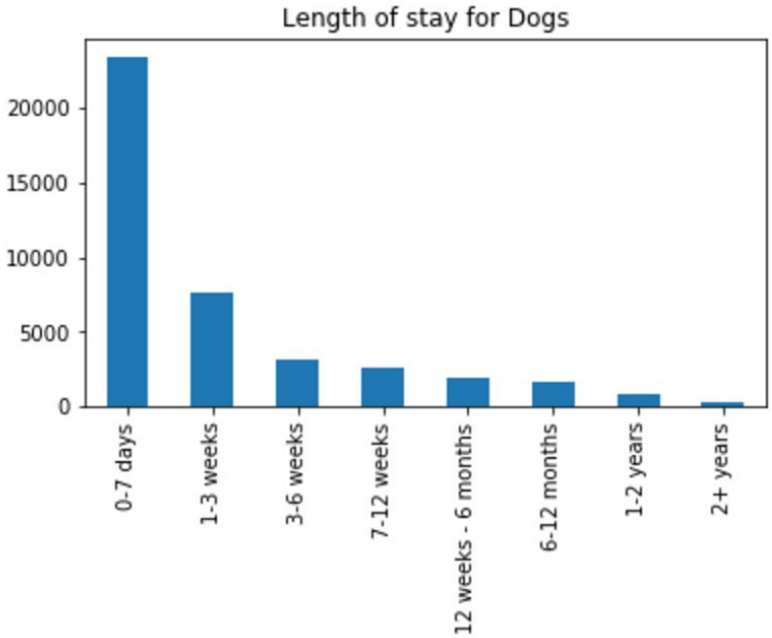

Length of stay for Dogs
20000
15000
10000
5000
0
0-7 days
1-3 weeks
3-6 weeks
7-12 weeks
12 weeks - 6 months
6-12 months
1-2 years
2+ years

Figure 4

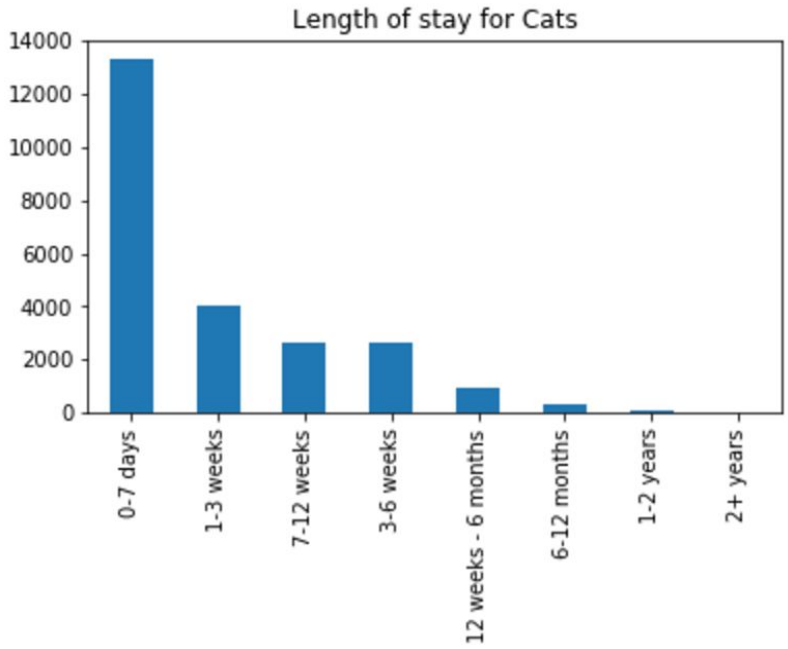


Figure 5

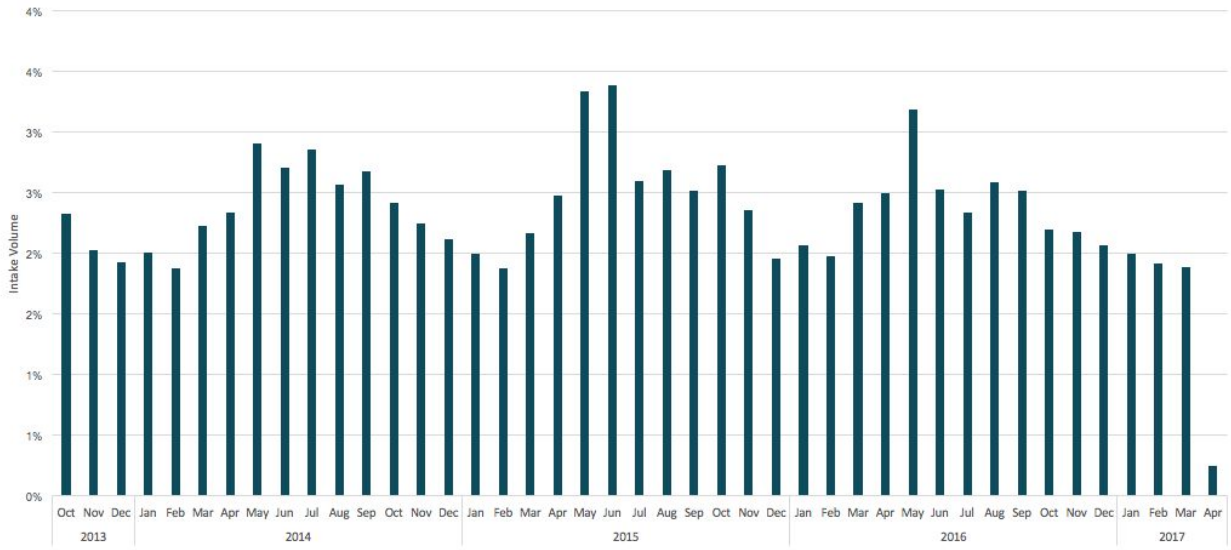


Figure 6

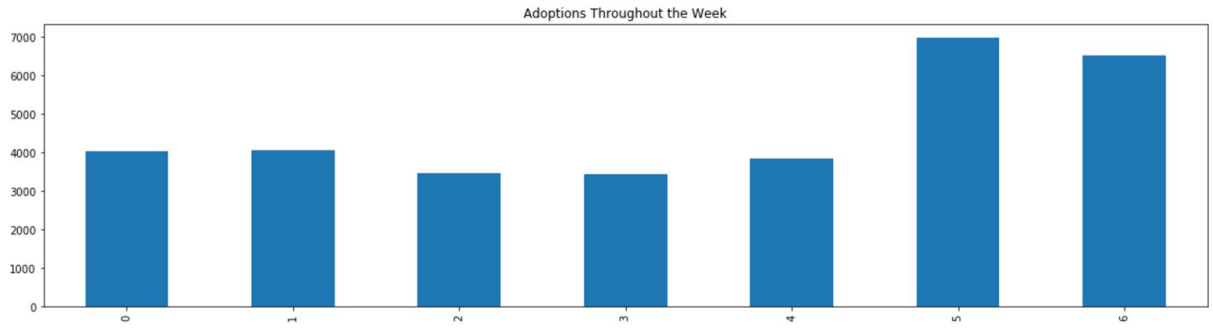


Figure 7

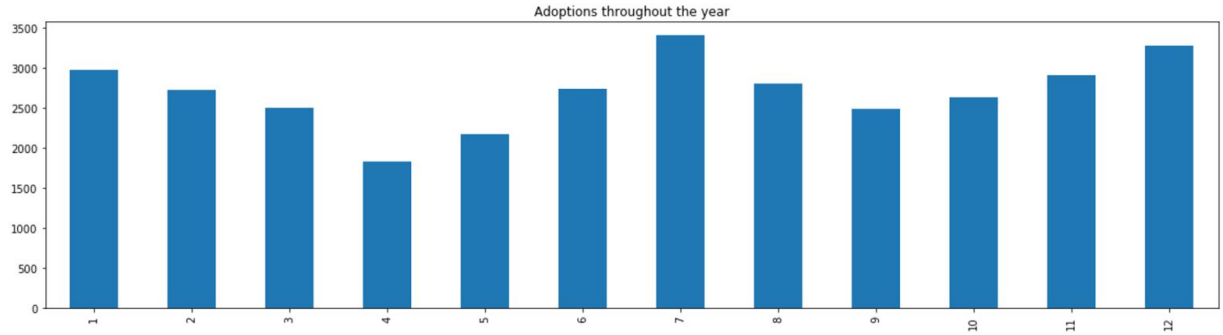


Figure 8