The dataset our team decided to analyze centered around Austin Animal Shelter intakes and outcomes, and 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 for lost and abandoned animals. 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/audience 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 animals at their local shelters. These informed decisions can help curb overpopulation of strays in shelters and community as a whole and offer societal and psychological benefits to members of the Austin community. It should be noted that this dataset only contains data from late 2013 through early 2018.

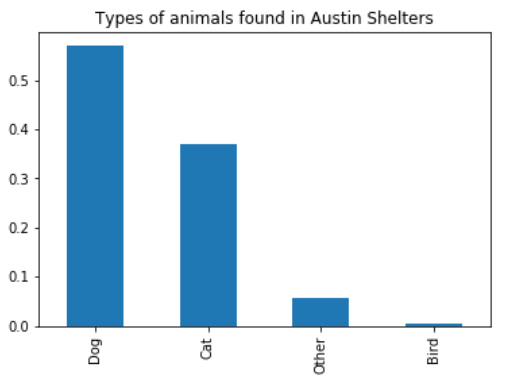
Filtering the initial dataset became an important focus, as the data we were originally working with contained erroneous data about bats, possums, and racoons that didn’t fit the traditional adoption framework. These variables greatly affected our ability to identify important features that impacted adoption rates, as our regressions and classifications would conclude that animals that were not either a cat, dog, or bird had extremely low adoption rates. Thus, we broke the column down into simply the popular categories of animals and ‘other’. This methodology was also used when filtering for breed and color of animal. With over 79,000 animals recorded in the dataset, there were bound to be outliers for breeds/colors that only appeared once or twice and had high adoption rates as a result. For these, we simply grouped breeds/colors into those that appeared more than 100+ times in the data and ‘other’. Another issue that came up when examining summaries of the data was the ‘outcome type’. While the Austin Animal Shelter takes in every lost animal, a significant portion of these are returned to their owner after being lost. Unfortunately, this distorted the dataset in that when running regressions and using dummy variables, those whom were returned to their owners (after grouping into the ‘not adopted category’, since adoption is our target) would result in nonsensical results. Finally, there were certain columns (days in shelter, gender) that required us to correct the datatype to a format that would allow us to apply classifications and functions more efficiently.

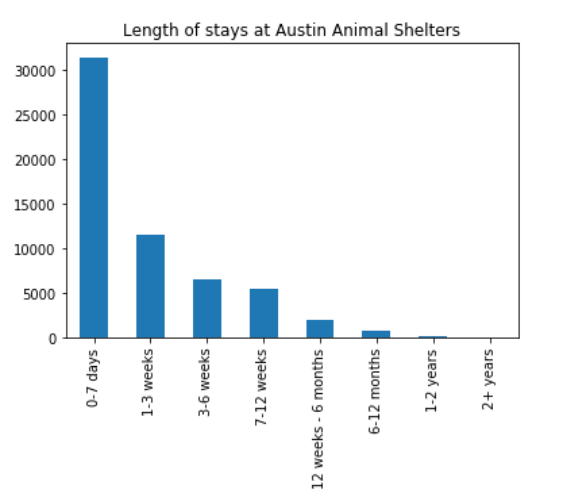
Some basic insights that our team found fit well with rather intuitive assumptions about what adoption rates should look like in Austin. Dogs are the largest percentage of local shelter animals (57%), with over 45,000 coming into Austin shelters over a five-year period[[1]](#footnote-1). Another useful insight was that the length of stays at animal shelters around Austin seemed mostly to be less than a week even after accounting for returns to owners[[2]](#footnote-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 (intuition, people have higher affinity to adopt a puppy than a 7-year old dog), we wanted to see statistically what that looked like for 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.[[3]](#footnote-3)

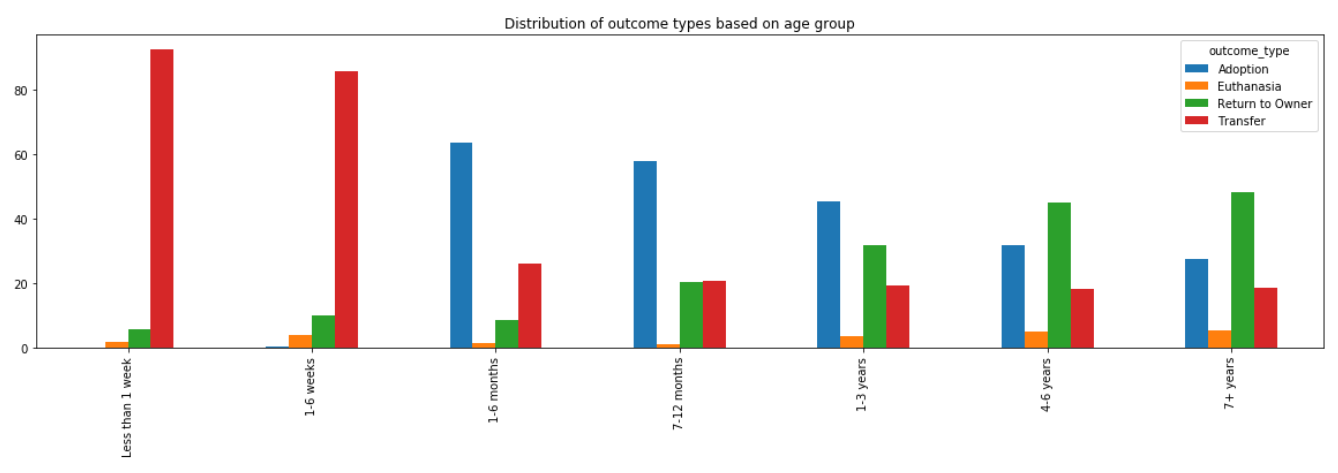
Another valuable analysis would be to look at which months seem to be more popular for adoption. The three most popular intake months are May, October and June (in order). The most popular outcome months are July, October and August. This seems pretty in line with the fact that the average time in the 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. The average time spent in the shelters seems relatively consistent over the course of the past 5 years[[4]](#footnote-4). Finally, it looks like, on average, adoption rates for popular animals (dogs, cats) have steadily increased in Austin over the past 5 years.[[5]](#footnote-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.[[6]](#footnote-6)

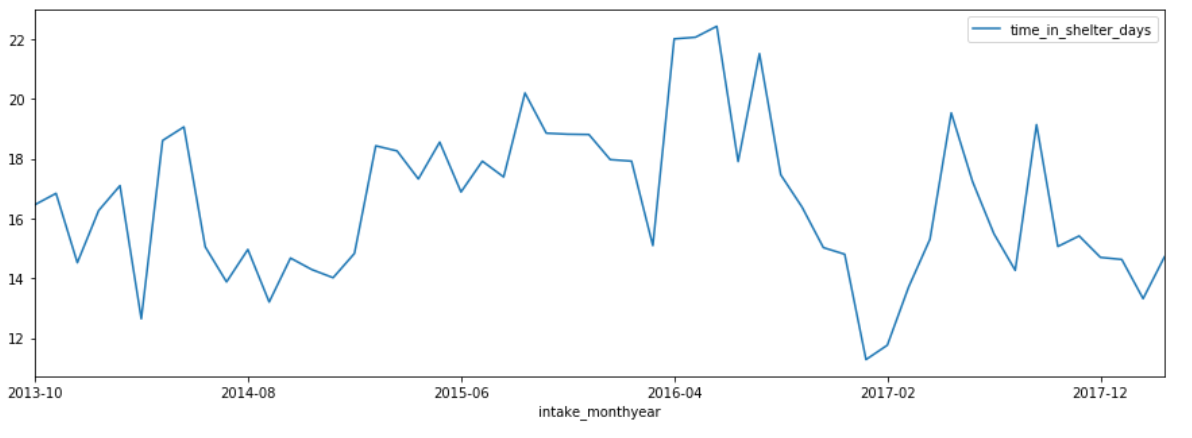
In order to have a holistic interpretation of the data, the team decided to use decision trees, nearest neighbors, naïve bayes, and logistic regression to classify the data and draw our conclusions. Since some of the original features of the dataset included redundant information (multiple columns classify length of time in shelters differently), the first task was to identify the relevant features to use. Ultimately, all the classifications used age, sex upon outcome (neutered or not), animal type, time in shelter) as the most important, which was based on a mix of intuition and series of trials when running regressions.

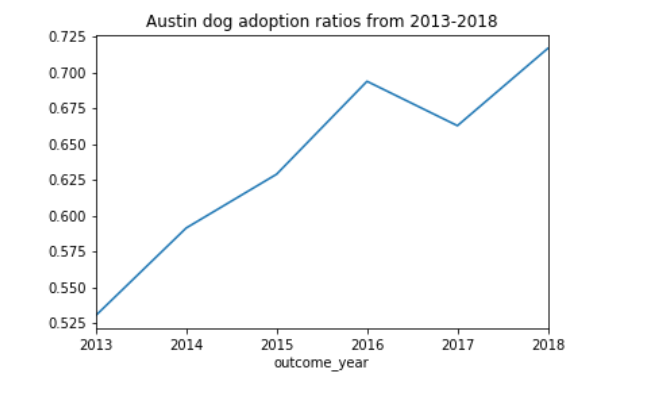
For the naïve bayes classification, we found that the classification gave us an 80.4% accuracy when determining whether a pet was adopted, which is about 23% higher than a random guess. An important factor to look into for naïve bayes is the importance of certain factors in the positive or negative class (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, and it seems that 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[[7]](#footnote-7).

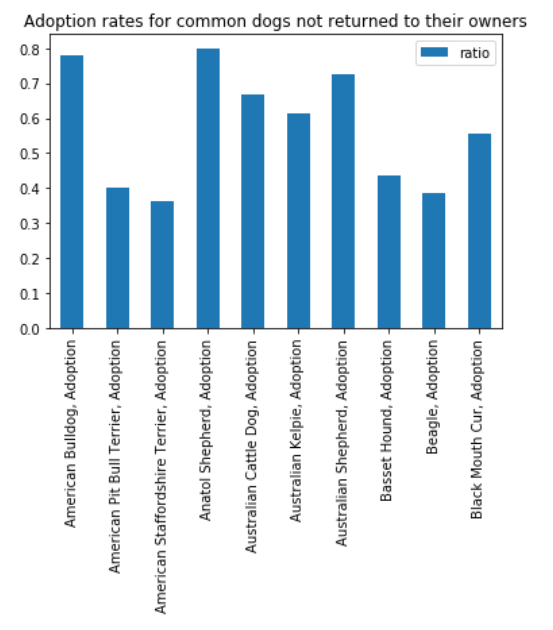


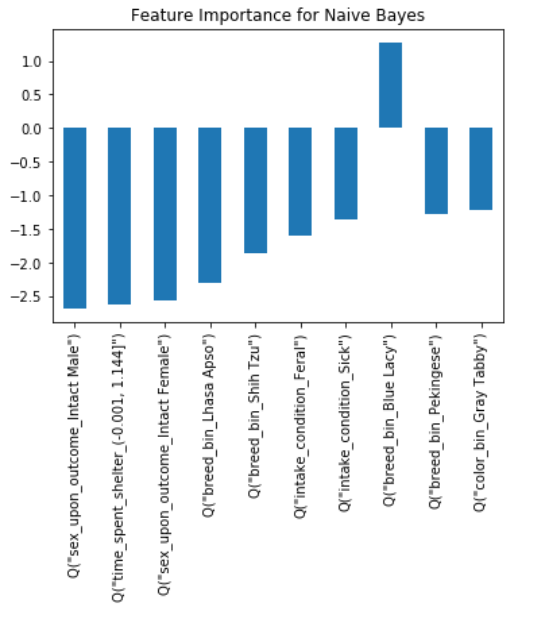












1. Figure 1: ‘Types of Animals found in Austin Shelters’ [↑](#footnote-ref-1)
2. Figure 2: ‘Length of stays at Austin Animal Shelters’ [↑](#footnote-ref-2)
3. Figure 3: ‘Distribution of outcome types based on age group’ [↑](#footnote-ref-3)
4. Figure 4: ‘Average time in shelter over time’ [↑](#footnote-ref-4)
5. Figure 5: ‘Austin dog adoption ratios from 2013-2018’ [↑](#footnote-ref-5)
6. Figure 6: ‘Adoption rates for common dogs not returned to their owners’ [↑](#footnote-ref-6)
7. Figure 7: ‘Feature Importance for Naïve Bayes’ [↑](#footnote-ref-7)