**Kaggle Shelter Animal Data**

**Project goal:**

The goal of this analysis is to predict shelter animal outcomes, and to find out which animals the shelter may need to give more effort to achieve a good outcome. The outcome can be any one of the following: adopted, died, euthanasia, returned to owner, or transfer. Because the outcome is categorical, I have decided to try two types of models: decision trees, and random forest. Kaggle recommends using something called log loss to judge the goodness of fit, so I will be using that. I also plan to look at the sensitivity, specificity, and accuracy.

**Dataset prep for modeling:**

Most of the variables I can use for prediction are categorical and required a lot of manipulation to get them into a form suitable for modeling. For example, the variable listing each animal’s breed was challenging. I found several possible ways the breed could be listed, for example:

* Golden Retriever Mix
* Labrador Retriever/Golden Retriever
* Golden Retriever

I created one variable with an indicator for the word “Mix” being present in an animal’s breed. I also created two breed variables, a primary breed variable and a secondary variable. One thing I’m still trying to figure out is how to include both breeds in the model. The secondary breed variable is often not populated because many animals are simply listed as a mix, or they are pure bred. As a result, the secondary breeds aren’t very important in the model. I would appreciate any ideas you might have on how to handle this.

**Model Building:**

So far, I’ve built a random forest model, and I plan to also build a decision tree model. Some of the top predictors in my model are whether or not the animal is altered, whether the animal is a dog or cat, whether or not a cat is short haired, if the animal is male or female, and if a dog is a pit bull. So far, the accuracy of my model is only around 0.632, so I need to do more work to raise it.

**Initial Findings:**

I generated some graphs to find out how the top variables are correlated with the outcome. One interesting finding is that dogs are much more likely to be returned to their owners than cats are. It’s possible that this is because cats are more likely to be found feral. However, if that’s not the case, this is one thing the shelter could try to change. Perhaps they could implement a program where they provide more resources and help to individuals who are trying to surrender their cats.

One other interesting finding is that Pit Bulls have very high euthanasia rates compared to other types of dogs. The shelter may want to consider providing more resources to rehabilitate Pit Bulls if the problem is aggression. They could also consider investing in community awareness programs to help educate the public on this frequently misunderstood breed.

I plan to continue to dig into the data to see if I can find any more useful pieces of information.