Predicting the outcome of animals in shelters

Animal shelters take in animals from a variety of different sources. Strays make up a large portion of these animals, but a significant number are also brought in by their owners. Bringing your animal to a shelter is an incredibly difficult decision. The goal of this project is to inform people trying to make this decision of what the outcome of them bringing an animal to a shelter might be. I believe there is room to add value to this problem space because people have this idea that bringing an animal to a shelter *might* be sentencing it to euthanasia, but they aren't really sure how probable it is.

This project would aim to eliminate that uncertainty, or at least reduce it.

The project would also be useful to animal shelters that are interested in the outcomes of their intakes.

Data

My dataset is from Kaggle.com. It has info on 80,000 animals from a shelter in Austin, TX.

The variables in my dataset are as follows:

Variable	Description	Type
age_upon_intake	Age of animal on intake	Float
animal_id	Unique id representing animal	string
animal_type	Cat, dog, fish, etc.	string
breed	Animal breed	string
color	Color of animal	string
datetime	Datetime stamp of intake	datetime
found_location	Street address animal was found at	string
intake_type	Normal, injured, sick, etc.	string
name	Name of animal	string
sex_upon_intake	Gender and whether the animal has been spayed/neutered	string

outcome	Adoption, euthanization, etc.	string

Challenges

This is a 5 class classification problem. I feel confident in my ability to do classification problems, but this one, like others, is going to be entirely dependent on my ability to feature engineer. There are huge opportunities for feature engineering here. As I do EDA on the dataset, I'm immediately wondering if black cats have a higher intake rate than others, or a higher euthanasia rate.

This dataset is almost entirely categorical variables. If I explode all of them in to dummy variables, I would be introducing bias in to almost any model. I'll have to pick and choose which variables make the cut carefully.

I plan to build a Flask app and utilize Spark in this project. I believe a Flask app hosted on Heroku could actually be useful for people thinking about giving a pet up to a shelter. As for Spark, I haven't had an opportunity to use it in the bootcamp, and it's a tool I'm very interested in adding to my skill set.

I'm very excited for this project. It's something I have a genuine passion for, and feel comfortable in my ability to deliver a strong product for career day.