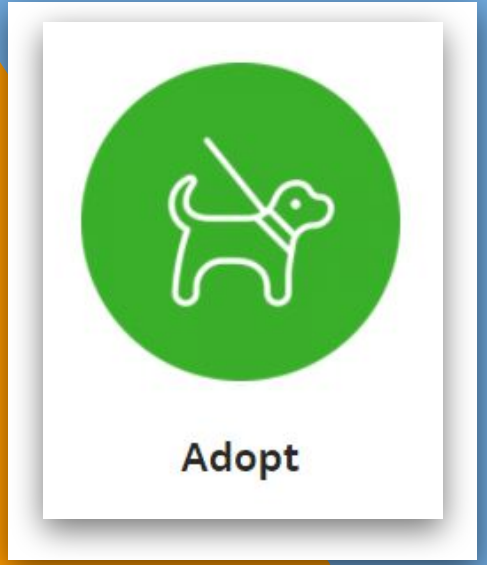
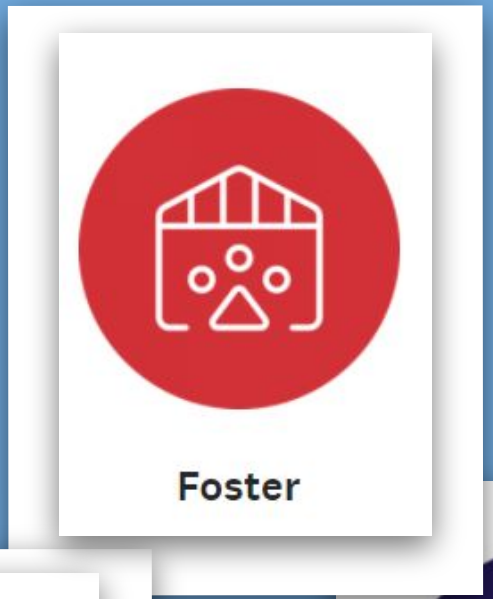
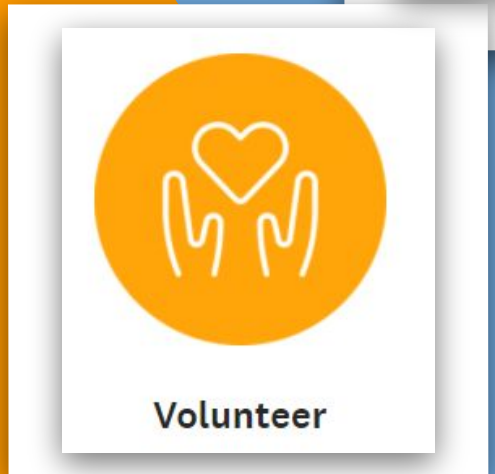
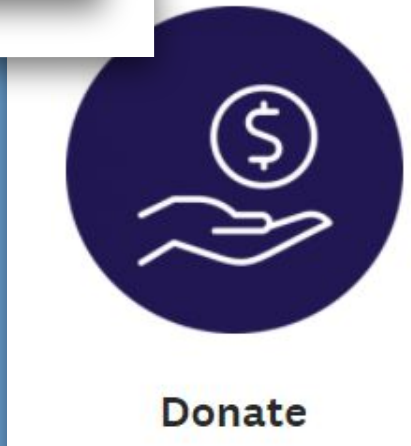
AUSTIN ANIML RESCUE









Topic and Reason

This project studies data from the [Austin Animal Center](http://www.austintexas.gov/content/austin-animal-center) in Austin, TX.

We chose this topic because we would like to help the Austin Animal Center better understand information on their animals so that as many animals as possible have a positive outcome.

Data Source Description

There are two datasets posted by the Animal Center as csv files:

Intakes (information on animals brought to the shelter) and outcomes (information on animals when they leave the shelter).

The information dates from 2013 to present and is updated daily on the Austin Animal Center data website. Variables in the dataset include an Animal ID, Animal Name, Animal Type, Breed, Color, Dates of intake and outcome, Sex, etc.

Links to the raw original datasets:

- [Austin Animal Center Intakes](https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Intakes/wter-evkm)

- [Austin Animal Center Outcomes](https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238)

Questions to answer:

We have two main questions that we are interested in answering, along with related subquestions.

Question 1 is a categorical prediction, and question 2 is a regression prediction. We will start by answering question 1 and if we have time will add a prediction for question 2.

- Question 1: Can we predict the outcome for an animal based on other characteristics?

- What are the possible outcomes that we should consider?

- What factors most influence the determination of the outcome?

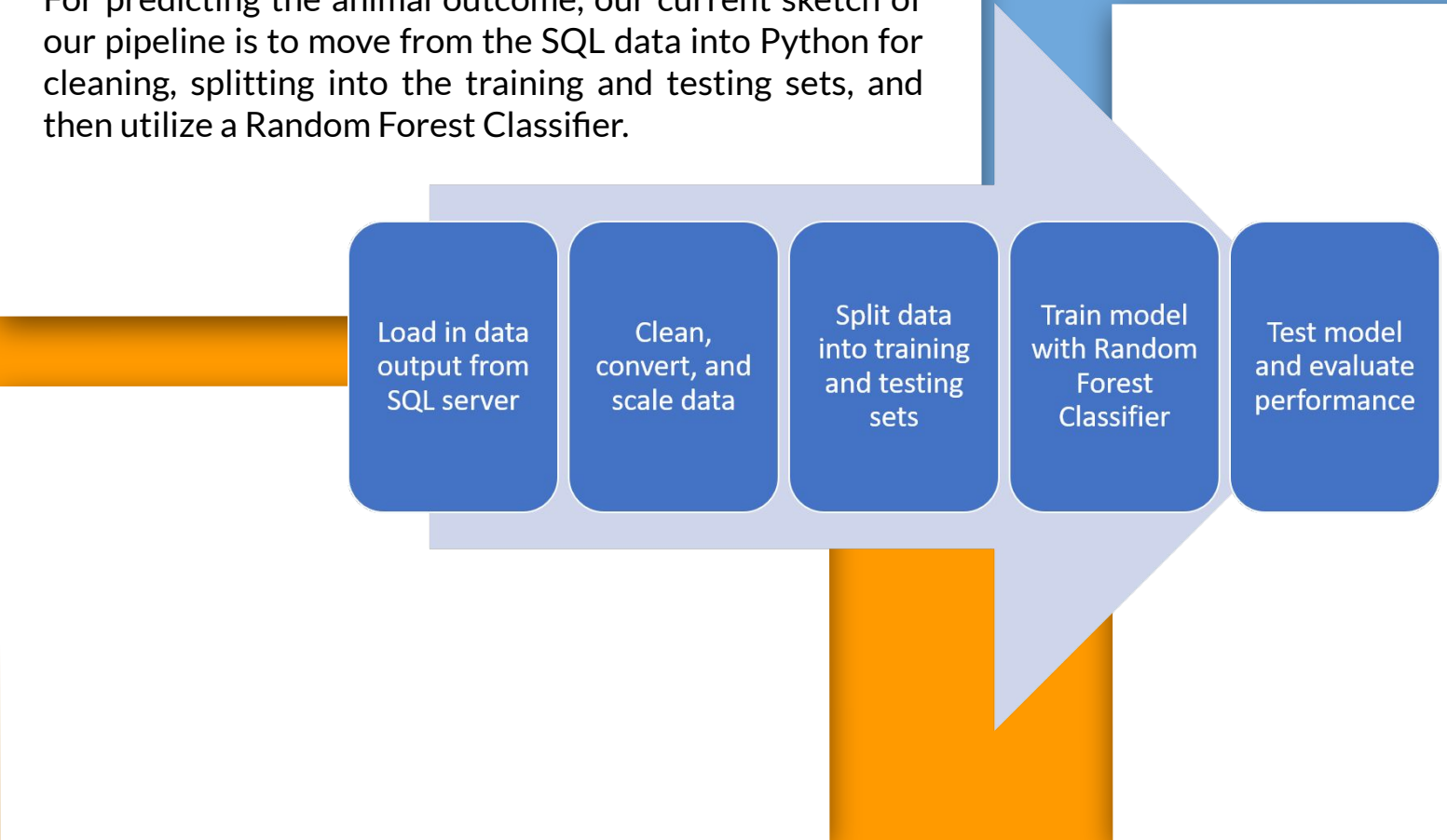
- Question 2: Can we predict the length of stay at the shelter for animals?

- What factors most influence the length of stay?

- Are the factors that influence the length of stay different for different animals?

- Preliminary Machine Learning Model

For predicting the animal outcome, our current sketch of our pipeline is to move from the SQL data into Python for cleaning, splitting into the training and testing sets, and then utilize a Random Forest Classifier.



COMMUNICATION PROTOCOL