Daniel Andrade
Matthew Bundas
Shahriar Dipon

Motivation and Proposal
Austin is a unique city because it maintains a 90% survival rate of animals which enter its shelters. Our team proposes to use public data sets from the Austin Animal Shelter to investigate the time an animal will spend in the shelter. This investigation will be based on the animal's attributes provided in the data sets, which include the following: age, color, breed, sex, and outcome type. Investigating these attributes by using machine learning methods may be helpful for other cities or other shelters in occupancy planning or it may provide guidance in determining which animals to transfer to sanctuaries, no kill cities/ shelters, foster homes. etc. The motivation for this project was to choose a meaningful and unique public data set that has not been analyzed excessively on platforms such as Kaggle or UCI Machine learning Repository. Another motivation for studying this data is its potential to reduce the unnecessary killing of animals in shelters.

Problem - Use machine learning algorithms and data mining techniques in an effort to predict the time an animal will spend in a shelter based on its physical attributes.

Solution - Acquire, clean, and merge the two data sets described below and apply various machine learning algorithms. Compare and contrast the algorithms' performance including accuracy and computational costs.

Data
Both data sets are provided by the City of Austin Open Data Portal which is designed to provide high value city data to users interested in finding out more about the city. For this project, we will use two data sets, one which provides information on the animals during intake, and the other detailing information upon outcome. The first dataset, Animal Center Intakes, includes data from October 1st, 2013 to February 24th, 2021. This data set represents the status of animals as they arrive at the Animal Center. All animals receive a unique Animal ID during intake. Annually over 90% of animals entering the center, are adopted, transferred to rescue or returned to their owners. This data set contains 124,000 rows and 12 columns which are Animal ID, Name, Date Time, Month Year, Found Location, Intake Type, Intake Condition, Animal Type, Sex Upon Intake, Age Upon Intake, Breed, and Color. The second data set is the Animal Center Outcomes from October 1st, 2013 to February 24th, 2021. Outcomes represent the status of animals as they leave the Animal Center. The Outcomes data set reflects that Austin, TX. is the largest "No Kill" city in the country. This data set contains 124,000 rows and 12 columns which are Animal ID, Names, Date Time, Month Year, Date of Birth, Outcome Type, and outcome Sub-type, Animal Type, Sex upon Outcome, Age upon Outcome, Breed, Color.

Analysis
Data analysis will be completed using Python and all code and related files will be shared on the team's GitHub repository. Detailed analysis will be included in the final report.