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 and

