EDA: Animal Shelter Effectiveness

Nicole Buros

Dataset Description

▶ The dataset I chose is a set of key data points for the animal care and control shelter in Austin, Texas from 2013-2021.

▶ I'm wanting to find the relationship between the variables and what relationship they have to the number of adoptions each year, and if there is any way to predict animal outcome success (adoption vs euthanasia) from the other factors, like breed, sex, or age.

Variable 1: Age

▶ Age of animal at time of outcome, in years

▶ The older ages tend to be less frequent, likely because the life span of animals are typically less than 15 years on average and most animals brought into the shelter are young

▶ The mean age is 1.84 years old
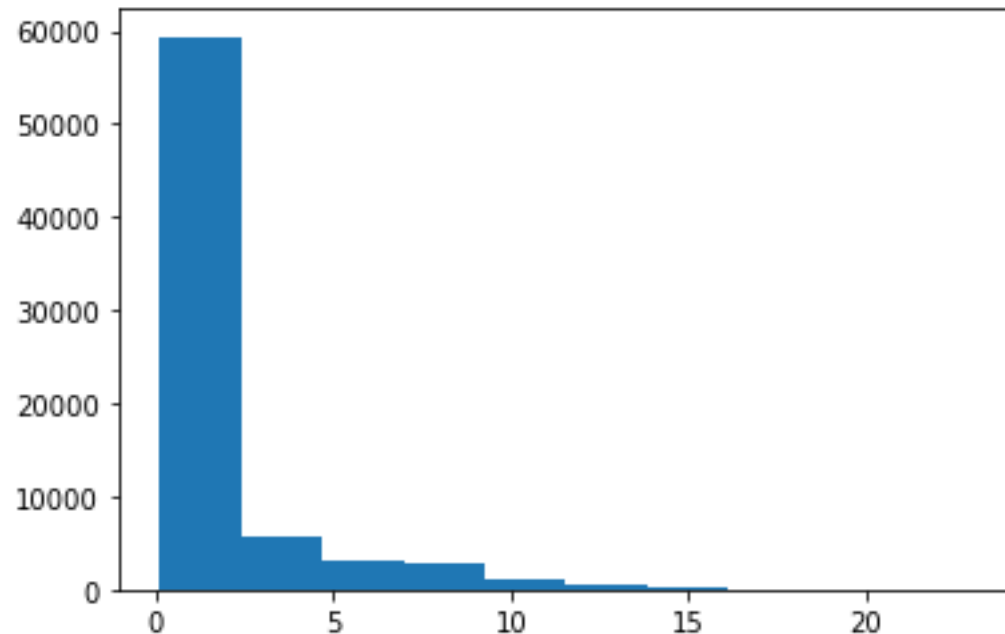
Variable 2: Breed

▶ This represents the breed type of each animal

▶ Outliers and the chart in general I was having difficulty with, as there were too many options
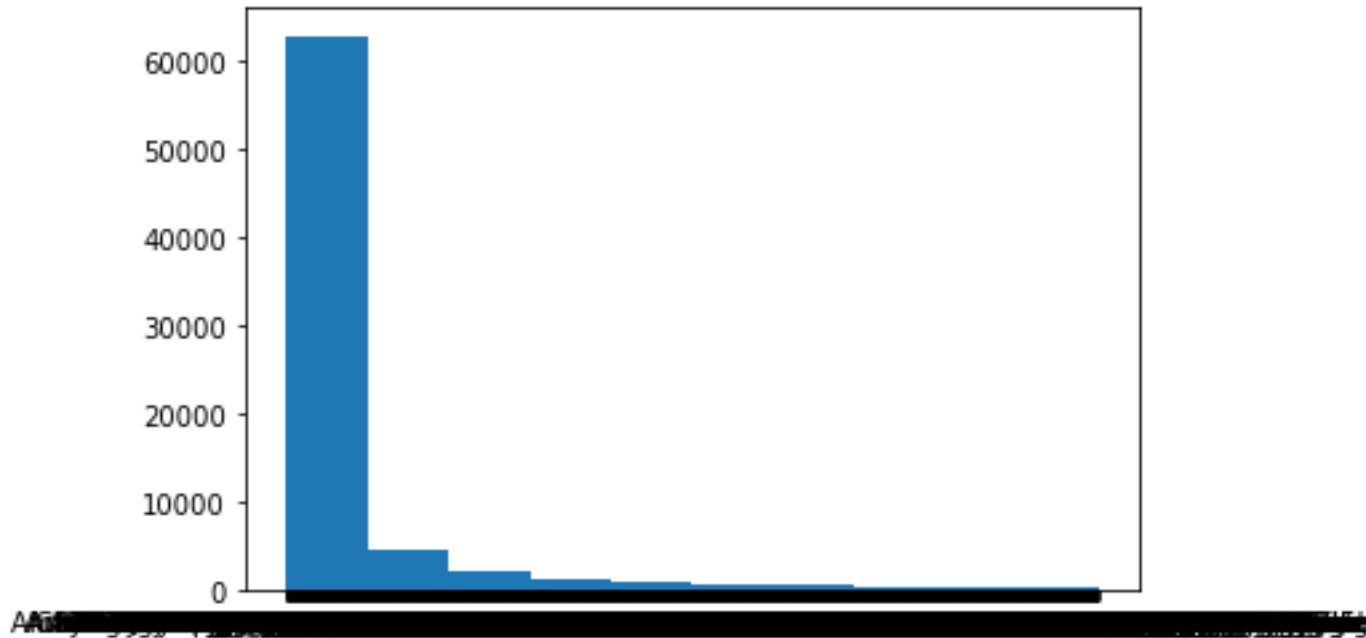
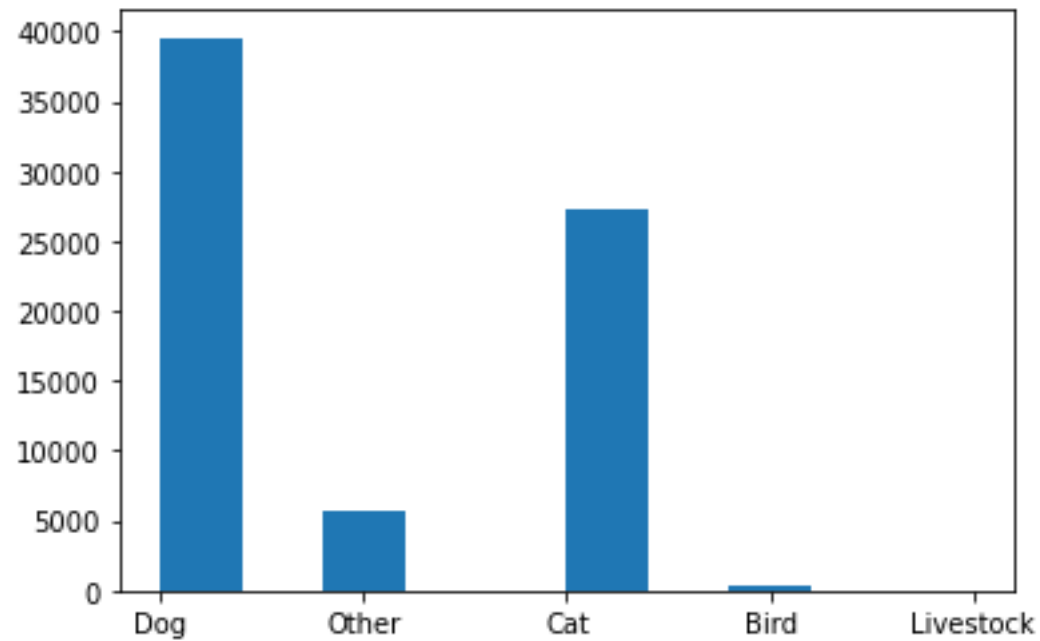# Variable 3: Outcome

- The outcome for the animal, narrowed to adoption or euthanasia

- Euthanasia is a much less frequent outcome, likely due to education of the public and the demand for pets remaining stable year over year
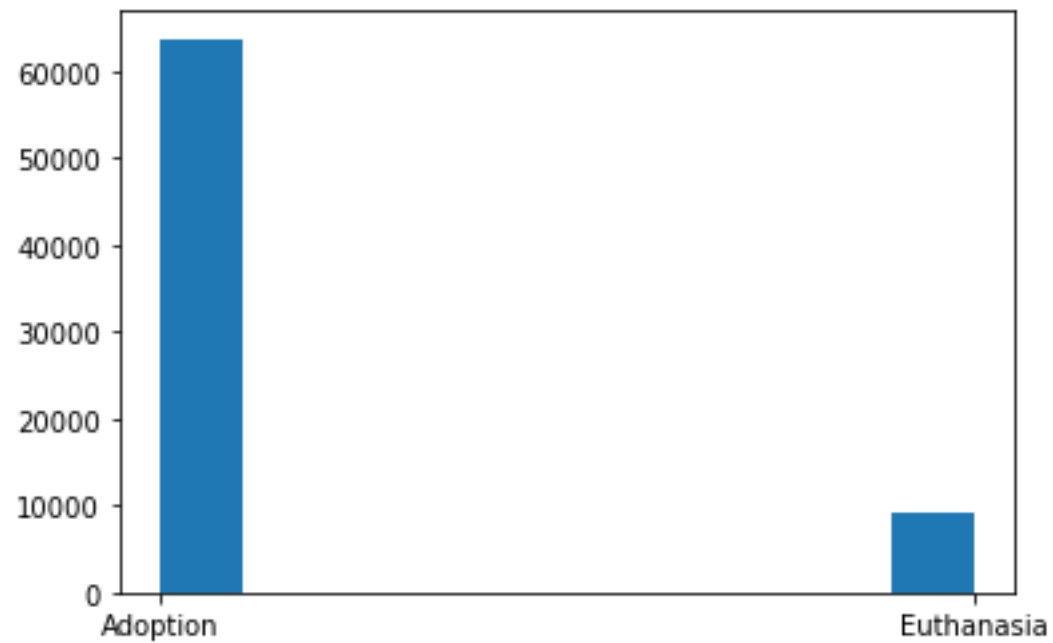
# Variable 4: Animal Type

▶ This variable notes the type of animal, such as cat, dog, bird, livestock, and other

▶ It seems that Other, Bird, and Livestock are outliers in how infrequently they appear in the data overall

Variable 5: Sex

- The sex of the animal at time of outcome.

- It is typically shelter policy to spay/neuter animals before adoption will be allowed, which likely explains the infrequency of the other categories.
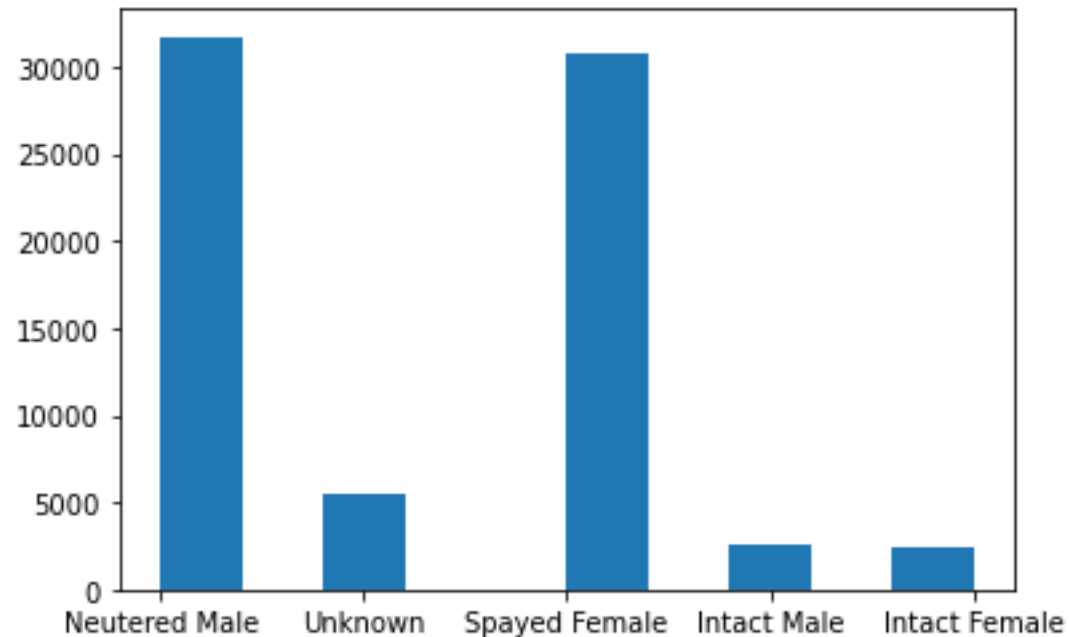
Comparing Two Scenarios Using PMF

▶ Comparing the probability of having a young versus an older dog in the shelter
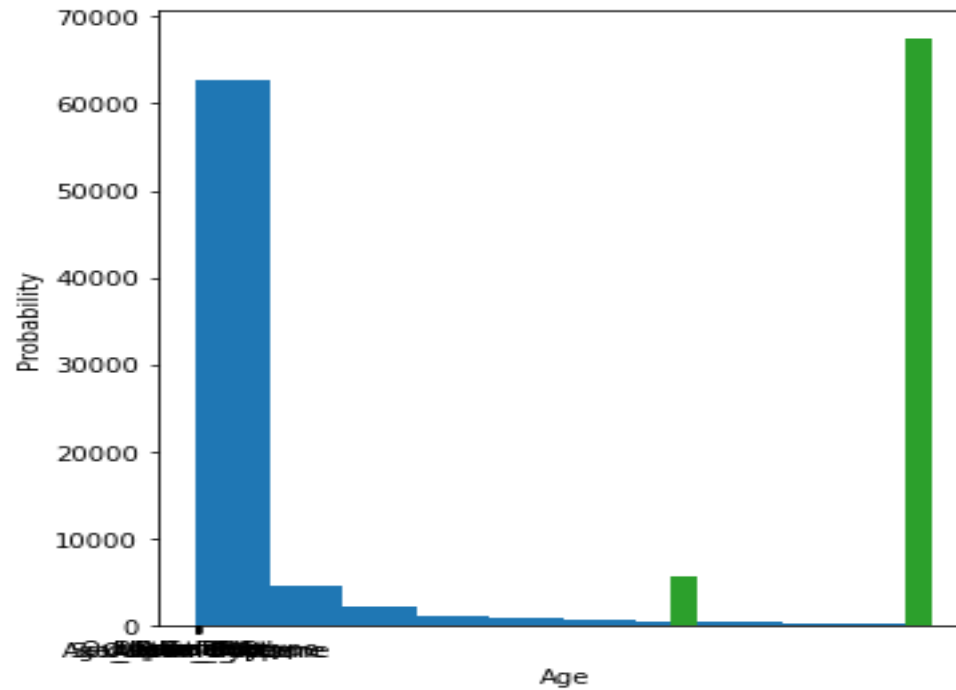
CDF with Variable 1

▶ I did a CDF with the age variable, seen below. I'm not sure what to take from the results or if it is even accurate, as calculating the probability by hand gave me a different answer but I can't find what I'm doing wrong.

```
Cdf([ 0.1    0.25  0.5    1.     2.     3.     4.     5.     6.     7.     8.     9.
 10.    11.    12.    13.    14.    15.    16.    17.    18.    19.    20.    21.
 22.    23.  ], [0.00174075 0.01034856 0.48805461 0.66599504 0.81416451 0.86439958
 0.89253944 0.91778719 0.93360473 0.94743479 0.96284113 0.97127075
 0.98194827 0.9858958  0.99122771 0.99469551 0.99668298 0.9983552
 0.99915018 0.99952027 0.99976699 0.99984923 0.99994517 0.99995888
 0.99998629 1.         ])
```

Scatterplots: Correlation & Causation

▶ It seems there is a concentration of younger animals across the board, and the sex of the animal is mostly evenly distributed around age. Attempting a scatterplot with my other variables proved difficult due to oversaturation.
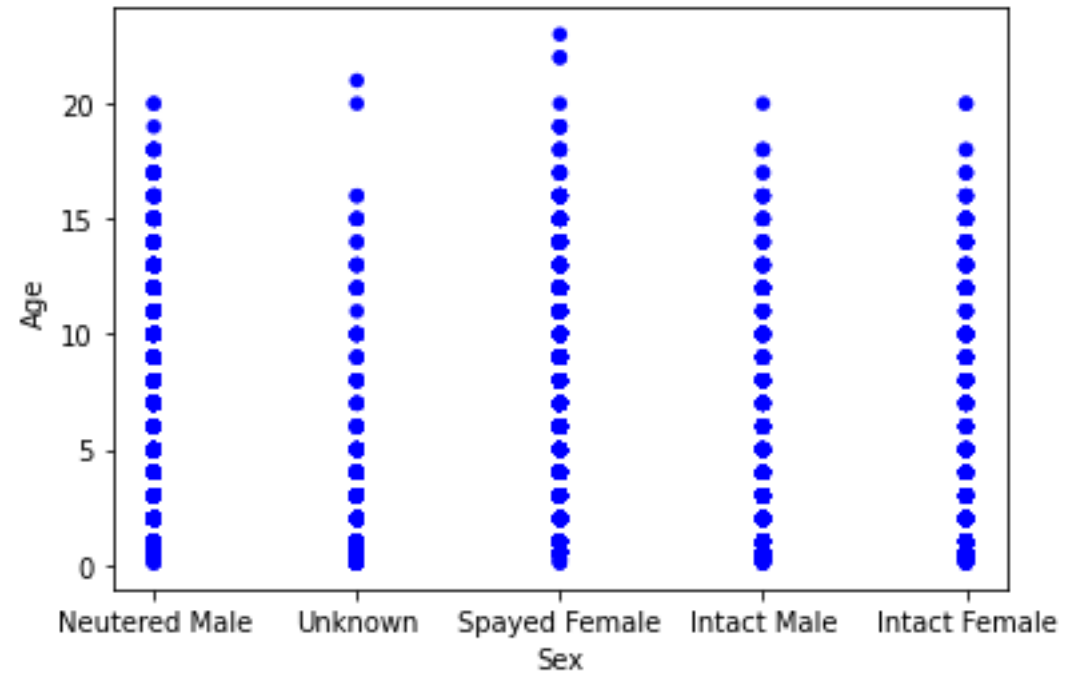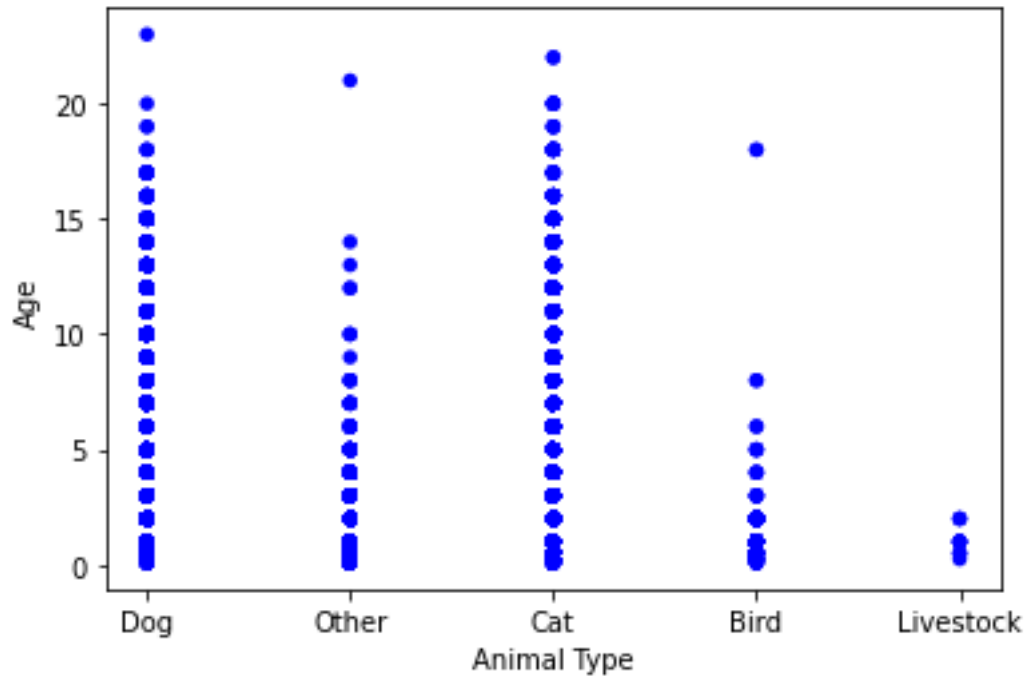
Regression Analysis

```
                          OLS Regression Results
==============================================================================
Dep. Variable:          Outcome_Type   R-squared:                       0.009
Model:                           OLS   Adj. R-squared:                  0.009
Method:                Least Squares   F-statistic:                     696.7
Date:               Wed, 01 Jun 2022   Prob (F-statistic):           8.09e-153
Time:                       19:45:58   Log-Likelihood:                -22635.
No. Observations:              72957   AIC:                         4.527e+04
Df Residuals:                  72955   BIC:                         4.529e+04
Df Model:                          1
Covariance Type:            nonrobust
==============================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept         1.1018      0.002    724.155      0.000       1.099       1.105
Age_upon_Outcome  0.0130      0.000     26.396      0.000       0.012       0.014
==============================================================================
Omnibus:                   29515.496   Durbin-Watson:                   1.972
Prob(Omnibus):                 0.000   Jarque-Bera (JB):            89070.307
Skew:                          2.230   Prob(JB):                         0.00
Kurtosis:                      6.067   Cond. No.                         3.99
==============================================================================
```