Increasing adoption rates in rabbit rescues

XAVIER RIOS

B.S. IN STATISTICS

The University of Texas
RioGrandeValley

Background

Rabbits are the 3rd most abandoned pets in the United States behind cats and dogs

Little to no presence for rabbit/small animal specific rescues
- Heavily underfunded
- Hardly known
- Few and far in between (mostly in big cities, if that)

This research project aims to not only increase awareness for these animals, but also to determine who might be most at risk of staying in the rescue for a long period of time
- We reflect on the needs of the organizations rescuing rabbits in a mathematical sense, and by using data science and applied statistics, we hope to create solutions for rabbit rescuers everywhere
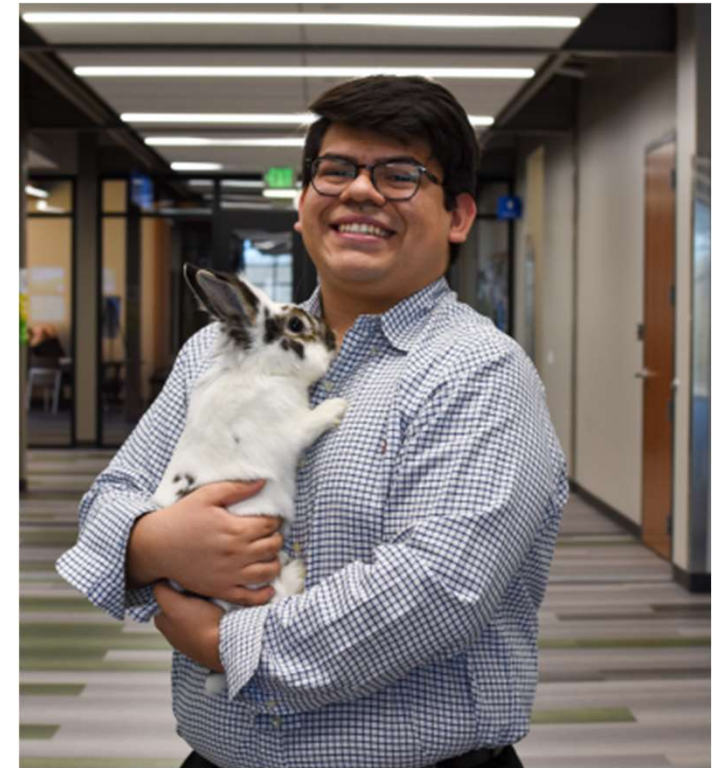
The House Rabbit Society (HRS)

The HRS was founded in 1988 with the mission of:

- ◦ "To rescue abandoned rabbits and find permanent homes for them and

- ◦ Through education, we seek to reduce the number of unwanted rabbits – and to improve bunnies' lives – by helping people better understand these often misunderstood companion animals."

Since then, they have impacted countless lives of rabbits and their humans around the globe

Without them, this project would not be possible

Research Goal / Research Questions

Research Goal:
- Determine types of rabbits who may be most at risk of staying in the rescue based on various characteristics

Research Questions
- What distribution does the stay of rabbits in rescues fall into?
- Does type of rabbit affect their duration of stay in the rescue?
  - Do lop rabbits have statistically significantly shorter stays than other rabbits?
  - Do ruby-eyed white rabbits have statistically significantly longer stays than other rabbits?
  - Do dwarf and regular sized rabbits have statistically significantly different stays at the rescue?
- Do bonded rabbits stay at the rescue significantly longer than single rabbits?
- Can we predict duration of stay of rabbits in shelters based on rabbit demographics, type of rabbit, and previous shelter or rescue affiliated status?

Exploratory Data Analysis

EXPLORING THE WORLD OF RESCUE RABBIT DATA

The Data

Sourced from restricted access ShelterLuv database on HRS rabbits
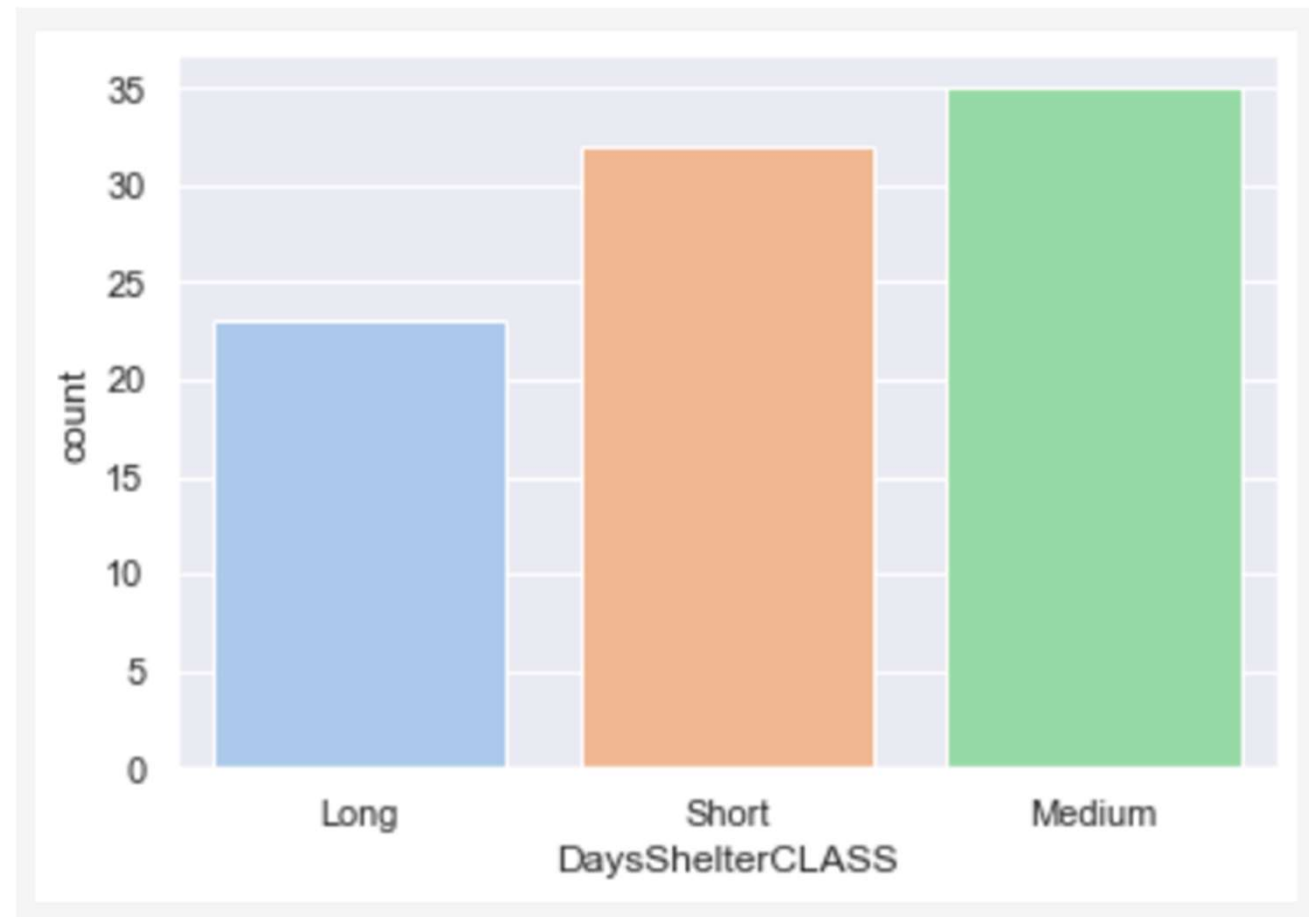
Manually input 90 records of rabbits

Imputed the median value for missing cases since the median is robust to outliers

Rescue Stay Classification

Short stay was less than 60 days (0)
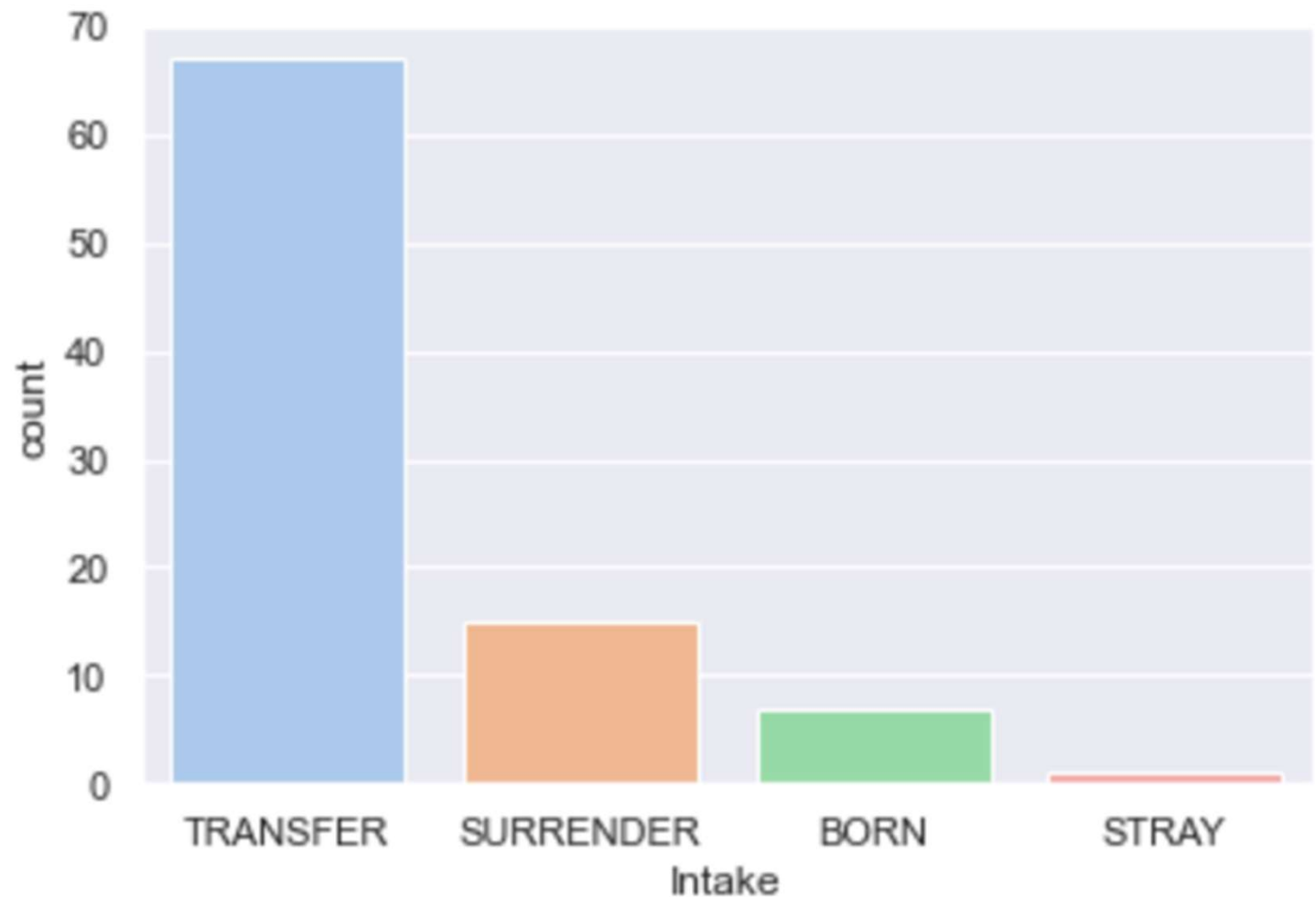
Medium stay cutoff was 270 days (1)

Everything longer than 270 days was considered a long stay (2)

Rabbit Intake Situations

From the figure below, it is evident that the top rescue arrival type is transfer in by a local animal shelter

This makes sense because most rabbit rescues operate in tandem with the local animal shelters, and rarely take in surrenders
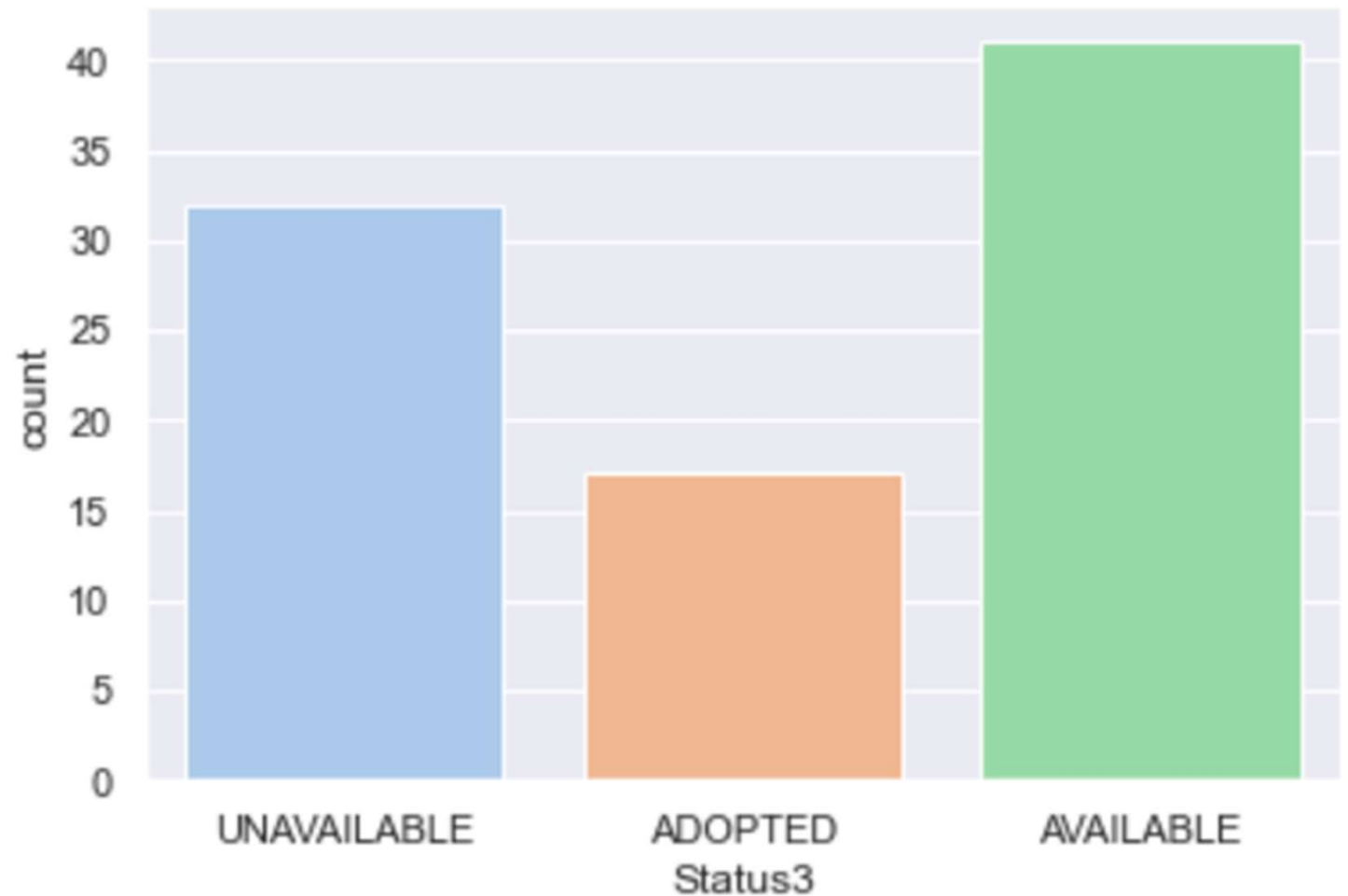
Rabbit Availability Status

The status of rabbits in the rescue is the label put on them for the availability or adoptability

Most are either available or unavailable

Unavailability means that they will probably be in the rescue for a longer time due to some illness

Rescue Stay vs Status

As mentioned before, a short stay is 0, a medium stay is 1, and a long stay is 2

We conducted a χ^2 test for independence with H_0: the variables status and length of stay are independent

With $\alpha = 0.05$, we reject the null hypothesis and conclude that the variables status and length of stay are not independent

$$(\chi^2 = 11.959, df = 4, p = 0.018)$$

DaysShelterClass012	0	1	2
Status3			
ADOPTED	6	8	3
AVAILABLE	10	22	9
UNAVAILABLE	16	5	11

DaysShelterClass012	0	1	2
Status3			
ADOPTED	0.352941	0.470588	0.176471
AVAILABLE	0.243902	0.536585	0.219512
UNAVAILABLE	0.500000	0.156250	0.343750

Rabbit Age and Age Groups

The rabbit rescue in question had age labels, as well as estimated birthdays and age, associated with each rabbit
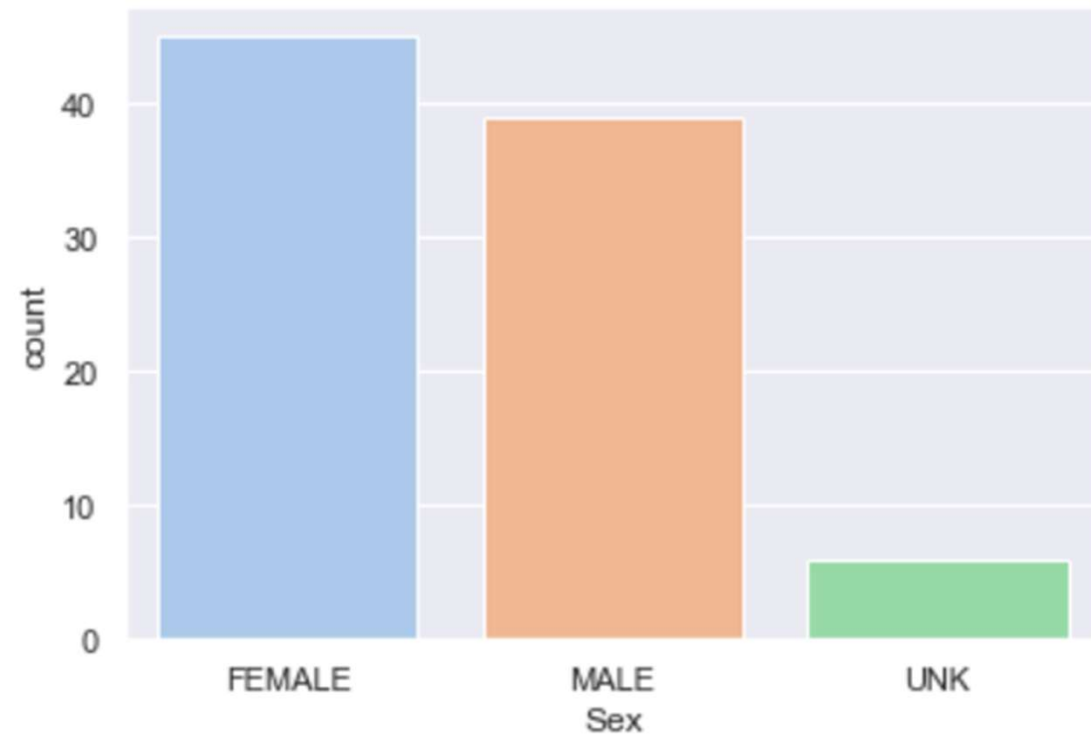
Adults were classified as over 1 year old, seniors were over 6 years old, and juveniles were under 1 year old
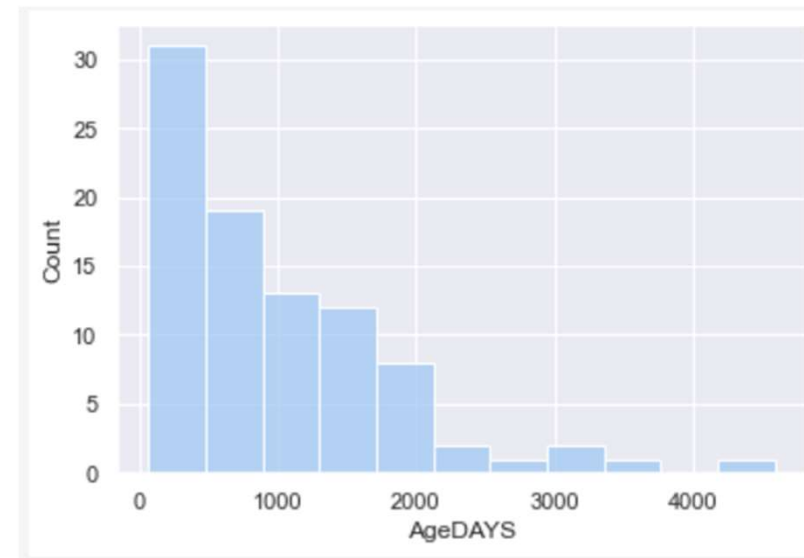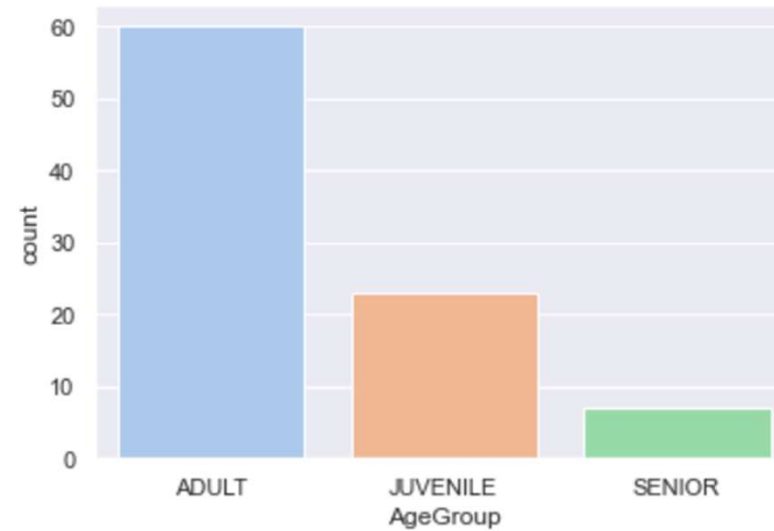
Rabbit Gender

The distribution of gender among the rabbits leans slightly towards females with males being just under 40 and females being around the 45 mark

There are a few cases of unknown gender due to age

Distribution of the Rescue Stay of Rabbits

WHAT DOES IT LOOK LIKE?

Calculating α for the Power Law Distribution

$$\alpha = 1 + x_{min} \left[\sum_{i=1}^{n} \ln \left(\frac{x_i}{x_{min}} \right) \right]^{-1}$$
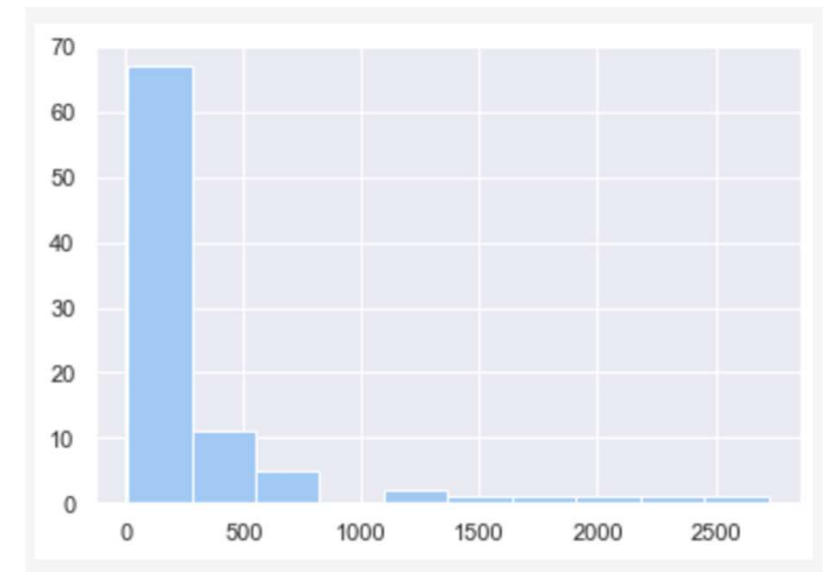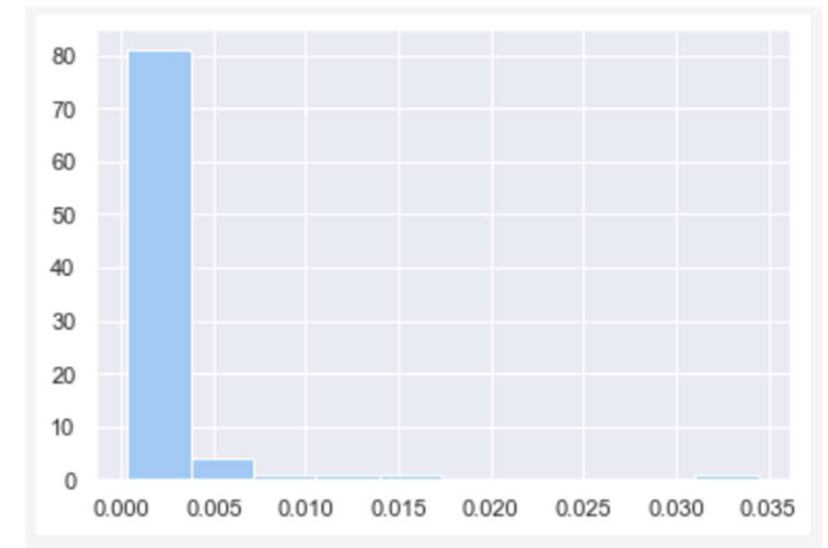
With distribution calculated as:

$$p(x) = \frac{\alpha - 1}{x_{min}} \left[\frac{x_i}{x_{min}} \right]^{-\alpha}$$

Comparing the Simulated Power Law to the Actual Rescue Stay of Rabbits

The top graph represents the simulated power law distribution with $\alpha = 1.0322$

The bottom graph represents the distribution of the actual rescue stay of rabbits

Types of Rabbits and their Stays in Rescues

COMPARING BETWEEN LOP, RUBY-EYED WHITE, AND DIFFERENT SIZED RABBITS

T-test and Cohen's D Statistic

In order to perform the t-tests properly, we conducted Bartlett's test for homoscedasticity

- Ruby-eyed white and lop rabbits, non-lop and lop rabbits, and bonded and single rabbits violated the equal variances assumption
  - An unequal variance t-test was conducted for these differences

H_0: the mean of rescue stay of rabbits are equal

H_A: the mean of rescue stay of rabbits are not equal

Cohen's D Statistic was calculated as $d = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2 + s_2^2}{2}}}$ and used to determine effect size differences

With $\alpha = 0.05$ we found there were statistically significant differences in rescue stays between non-lop and lop rabbits $(t = 2.95, p = 0.008, d = 0.56)$ and between ruby-eyed white and lop rabbits $(t = 2.16, p = 0.043, d = 0.74)$

All other comparisons yielded no statistically significant differences

Classification for Rescue Stay of Rabbits

PREDICTING RESCUE STAY OF RABBITS

Mutual Information Scores

Mutual information is calculated by

$$MI(U, V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \ln \frac{N|U_i \cap V_j|}{|U_i||V_j|}.$$

Once the mutual information was calculated between the dependent and independent variables, to reduce the number of features used, we only considered the top performing variables

Removed AgeGroup due to multicollinearity with AgeDAYS
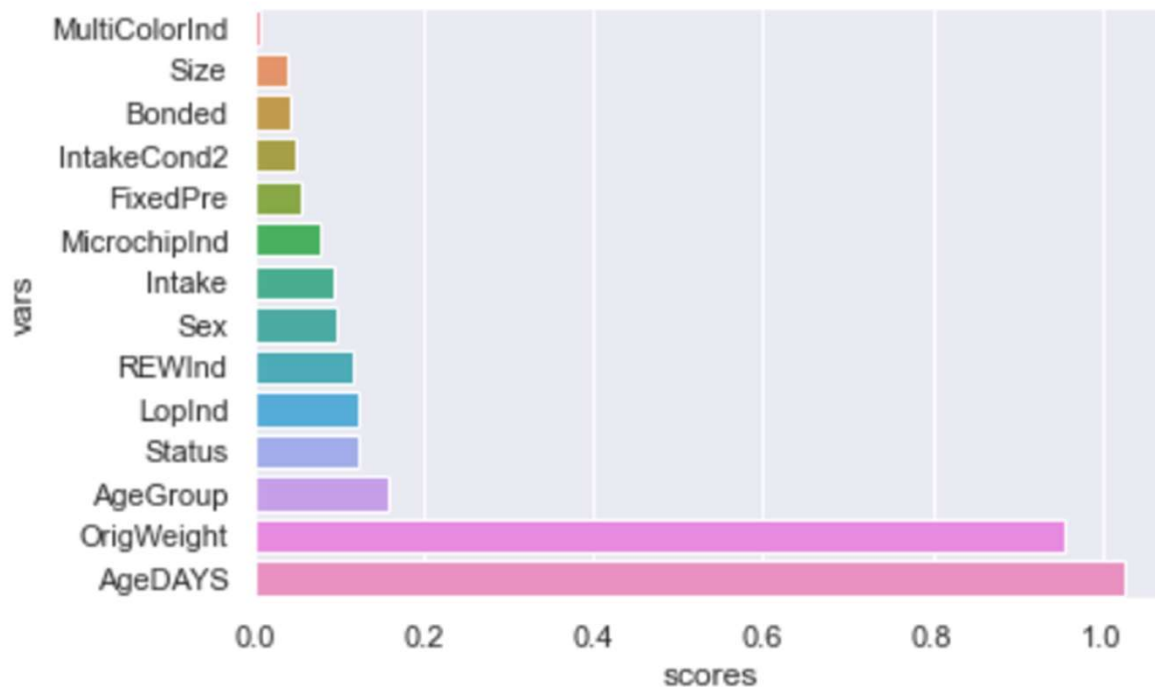
Data Preprocessing

Before training and testing the data, we performed Box-Cox transformations on the estimated age in days and the original weights to normalize the data, which is needed for the Random Forest Classifier and for the Multinomial Logistic Regression

To handle categorical data, we encoded it using the One-Hot Encoder, which creates new features, each being binary variables with a 1 or 0 if the feature contains a specified category
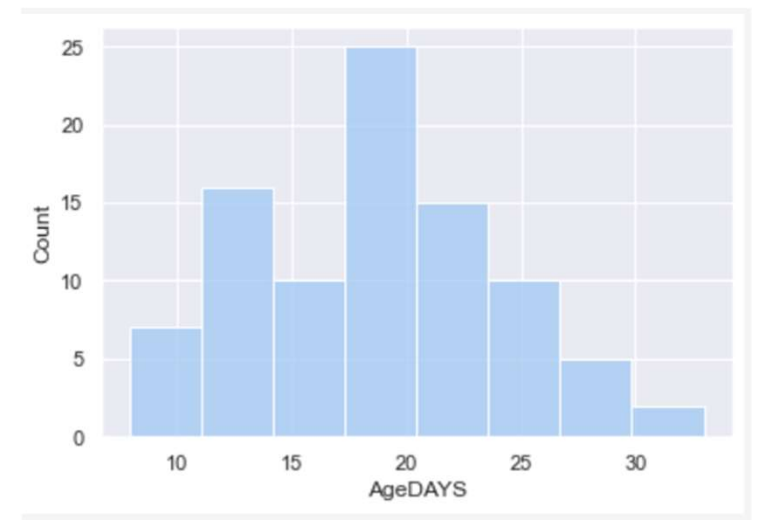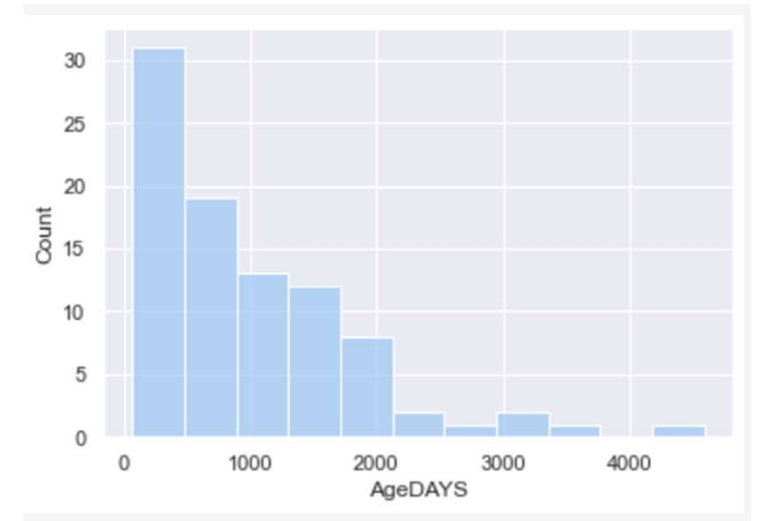
Feature Importances
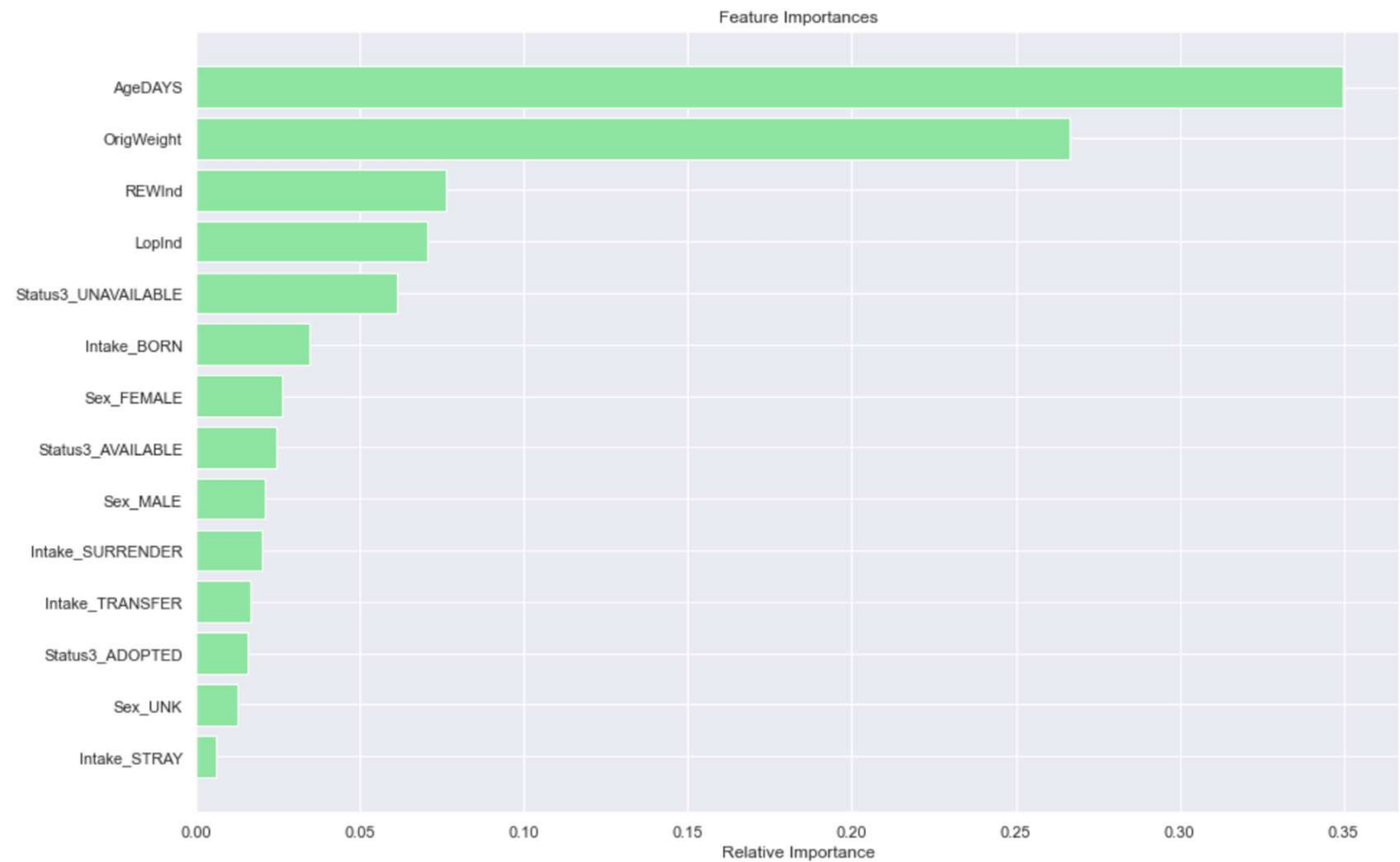
The most important features to the Random Forest Classifier were the age in days, original weight, if they were a ruby-eyed white rabbit, if they were a lop rabbit, and if they're unavailable

Feature Importances

Confusion Matrix and Classification Report (RF)

The 3 by 3 matrix represents the confusion matrix of actual class versus predicted class with correct classifications along the diagonal

Overall, the Random Forest Classifier model had a 67% accuracy for classification

The Random Forest Classifier predicts long stays the best

```
[[3 3 0]
 [1 5 1]
 [0 1 4]]
```

	precision	recall	f1-score	support
0	0.75	0.50	0.60	6
1	0.56	0.71	0.63	7
2	0.80	0.80	0.80	5
accuracy			0.67	18
macro avg	0.70	0.67	0.68	18
weighted avg	0.69	0.67	0.67	18

Confusion Matrix and Classification Report (MLR)

Overall, the Multinomial Logistic Regression model had a 61% accuracy for classification

The Multinomial Logistic Regression model predicts medium stays the best

```
[[3 3 0]
 [0 6 1]
 [1 2 2]]
```

	precision	recall	f1-score	support
0	0.75	0.50	0.60	6
1	0.55	0.86	0.67	7
2	0.67	0.40	0.50	5
accuracy			0.61	18
macro avg	0.65	0.59	0.59	18
weighted avg	0.65	0.61	0.60	18

Model Explanations

SHAP values tell us the direction in which certain features influence the output

Here is an example of one of the bee swarm plots on the long duration stay
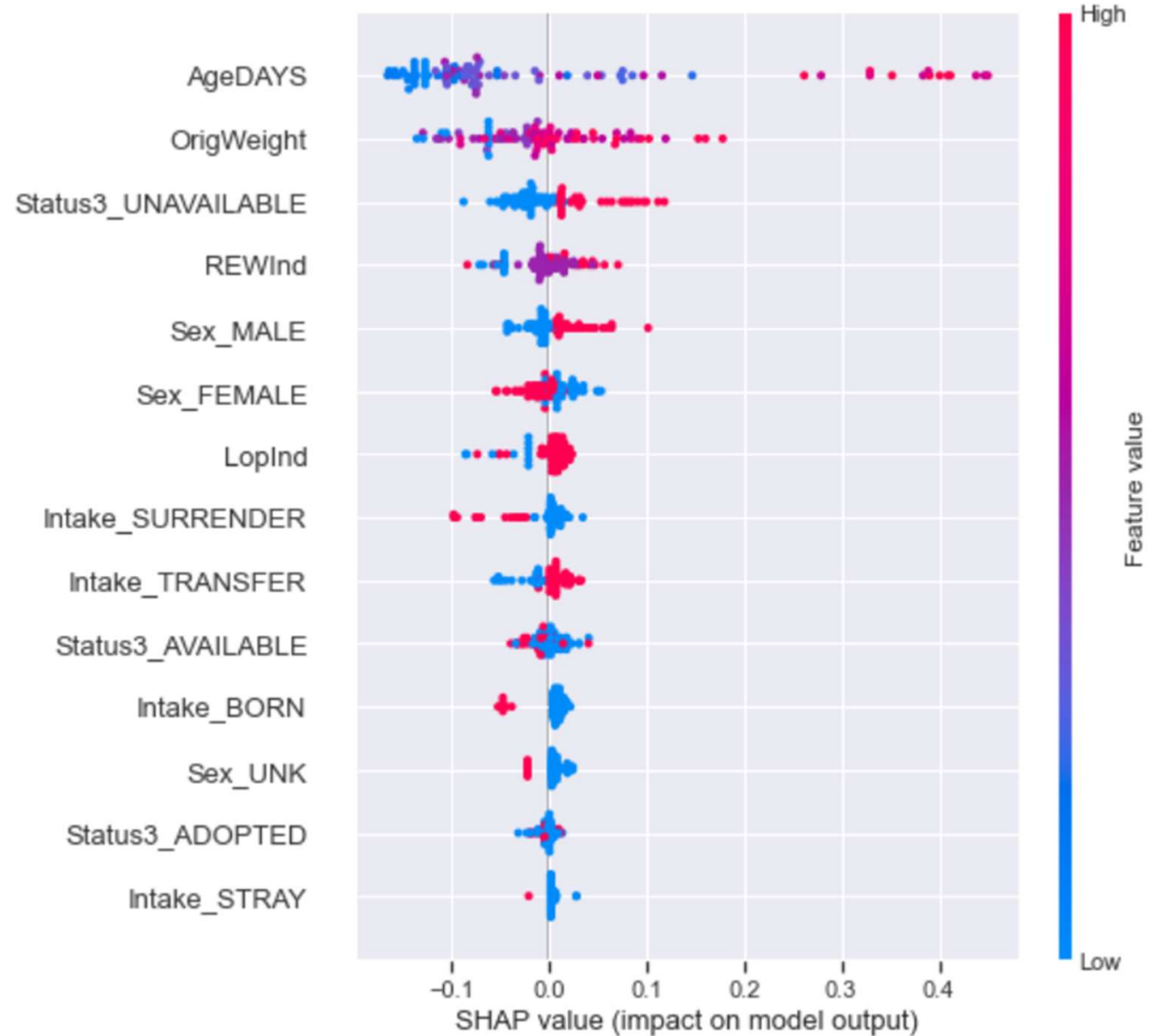
Conclusion

The distribution of the rescue stay of rabbits follows a power law distribution with $\alpha = 1.0322$

Ruby-eyed white rabbits, non-lop rabbits, and older rabbits are most at risk of staying in the rescue for a longer time

Even though we found no significant differences between bonded and non-bonded rabbit stays, the truth is that bonded rabbits are harder to adopt out

We were able to predict the length of stay of rabbits in rescues adequately with the Random Forest Classifier
- This model can be further used to determine the length of stay of rabbits added to the database in the future

Future Research

Future research should be conducted with a larger cohort of rabbits with much more varying backgrounds

In addition, specific breeds should be further investigated since it is known that breeds like lops and lionheads are more likely to be adopted quicker than other types of rabbits
- We had an indicator for lops, but that was easy to tell because each rabbit had a picture

This analysis should also be conducted on other small animals
- The more a pet is misunderstood and handled by children, the more they will end up in rescues or shelters

Time to event analysis (survival analysis) on the rescue stay of rabbits by rabbit demographics to determine the most likely timeframe for a rabbit to be adopted

Hyperparameter tuning for the Random Forest Classifier should be conducted to tweak model performance
- Currently, the parameters used were to avoid overfitting