# MATH 463 Project

What Factors Impact How Long Animals Stay in Shelters?

Catie Lamberson, Lola Sneddon, Claire Wagner

2023-05-02

## Abstract

This paper seeks to explore the factors that influence the amount of time spent by animals in shelters. Using data from the Austin Animal Center, we trained a multiple regression model to predict the duration of a given animal's stay at the shelter based on information known before that animal's release. While we found that a linear model was not suitable, we were still able to draw key conclusions about predictors, two of which we then verified by hypothesis testing. Notable predictors for length of stay include type of animal and age group of animal, with dogs spending less time in the shelter than cats and animals older than 12.5 years spending less time in the shelter than younger animals.

## Introduction

Approximately 6.3 million animals enter shelters each year in the United States. Only about 4.1 million of these are adopted each year.[1] This places a significant burden on animal shelters, which must utilize limited resources to house animals for varying periods of time while attempting to ensure the best possible outcome for each animal upon release.

We have chosen to investigate data concerning animals taken in by a shelter with the goal of identifying factors that affect the length of an animal's stay in a shelter. By evaluating various predictor variables and their respective impacts on length of stay, we can gain insights into what should be done to facilitate shelters' decisions about how to best care for each animal, from fostering or adoption to longer-term housing at the shelter.

The data used for our analysis was sourced from the Kaggle dataset "Austin Animal Center Shelter Intakes and Outcomes."[2] This dataset was compiled by Aaron Schlegel using data published by the Austin Animal Center (AAC), the largest animal shelter in the United States with a no-kill policy, which houses more than 18000 animals per year.[3] The AAC collects data on each animal that comes through the shelter at the time of intake as well as at the time of outcome (release from the shelter).

## Exploratory Data Analysis

The original AAC dataset contained 41 features and 79672 observations. To clean the AAC dataset and make it more suitable for our analysis, we applied the following modifications and feature engineering:

[1] https://www.aspca.org/helping-people-pets/shelter-intake-and-surrender/pet-statistics
[2] https://www.kaggle.com/datasets/aaronschlegel/austin-animal-center-shelter-intakes-and-outcomes
[3] https://www.austintexas.gov/department/austin-animal-center-0

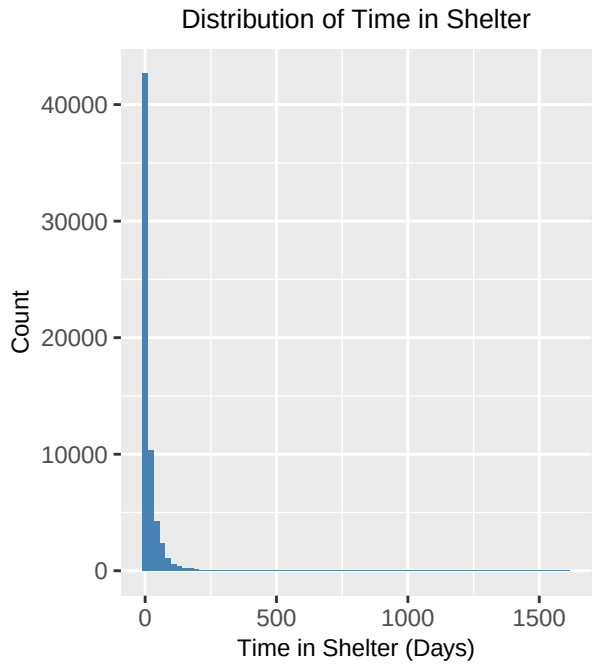
- We removed 6938 observations with missing or unknown values.
  - 6785 of these removed observations had missing or unknown values for `sex_upon_intake`, `sex_upon_outcome`, or both.
  - 147 of these observations had missing or unknown values for `intake_condition`.
  - 10 of these removed observations had missing or unknown values for `outcome_type`.
- For each animal that was taken into the shelter on more than one occasion, we kept only the observation corresponding to the first intake and outcome for that animal (resulting in the removal of 7668 additional observations).
- We removed features that had too many unique values (such as `breed`) or were redundant (such as `animal_id_outcome`).
- We removed features in datetime format regarding animals' birth dates and ages because these must often be estimated and are unlikely to be precise. We retained the categorical feature `age_upon_intake_age_group` since lack of precision for this attribute is less problematic, but we collapsed all age groups with ages greater than 15 into the new category `(15.0, 25.0]`.
- We created a new ternary feature, `black_coloring`, recording whether each animal was partially, fully, or not at all black in coloring, using the preexisting feature `color`. We removed 2484 observations for which our parsing algorithm was unable to determine whether the animal was black-colored or not.
- We replaced the preexisting features `sex_upon_intake` and `sex_upon_outcome` with four new binary features — `sex`, `neutered_or_spayed_upon_intake`, `neutered_or_spayed_upon_outcome`, and `neutered_or_spayed_during_time_at_shelter` — to divide the information encoded in the original features into separate, more easily analyzable attributes. We removed 3 observations with inconsistent neutered/spayed status (where an animal was recorded as spayed or neutered on intake but intact on outcome).

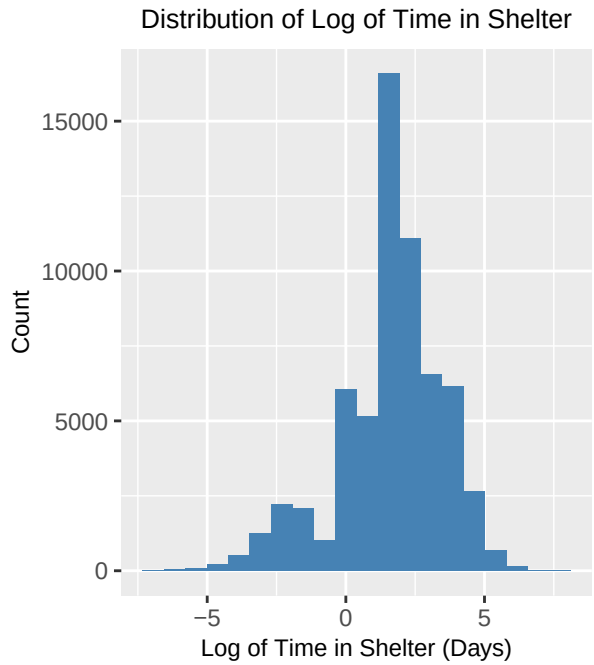
The cleaned dataset, which spans the years 2013 to 2018, contains 20 features and 62579 observations (each of which corresponds to a unique animal). There are no missing values. See Appendix A for descriptions of the features in the dataset.

As we can see by examining the histograms below, `time_in_shelter_days` is skewed to the right with some significant outliers. We can reduce the impact of these outliers by taking the log of `time_in_shelter_days`. As the table of summary statistics shows, this log transformation moves the mean and median closer together and reduces both the standard deviation and maximum value.

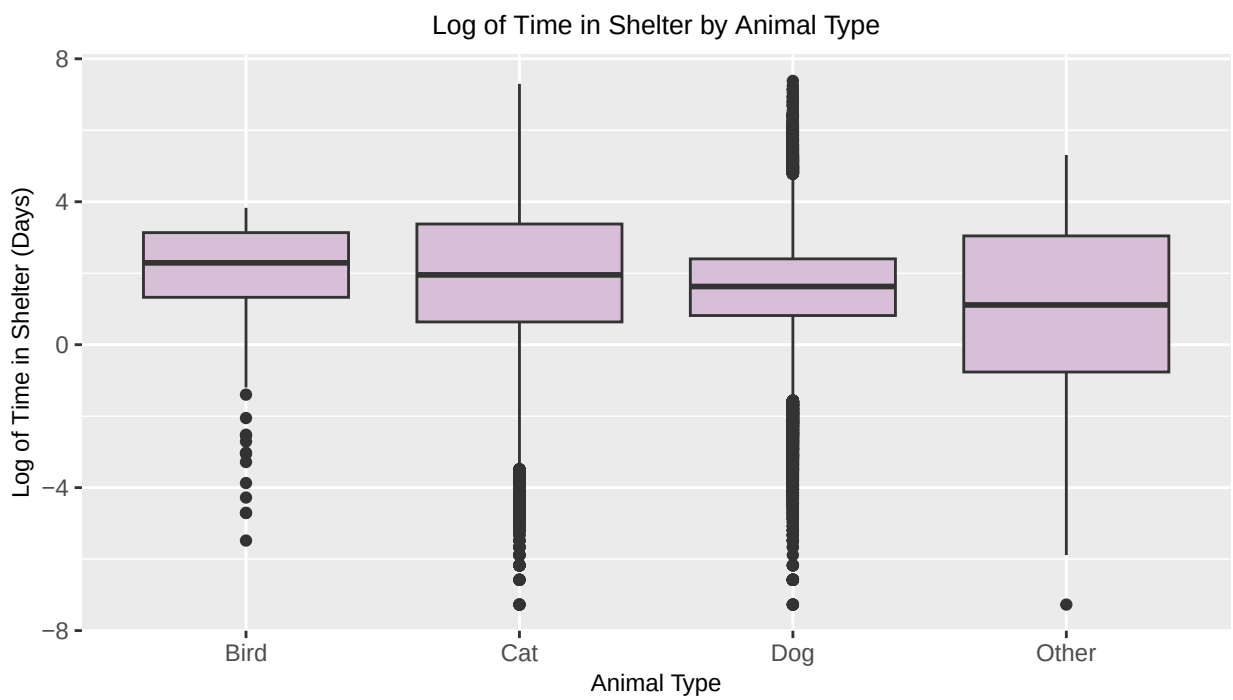
Table 1: Summary Statistics for Time in Shelter and Its Log

| | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|---|---|---|---|---|---|---|---|---|---|
| time_in_shelter_days | 0.0006944 | 2.1152778 | 5.272917 | 15.818403 | 1606.194444 | 18.199746 | 43.099804 | 62579 | 0 |
| log_time_in_shelter_days | -7.2723984 | 0.7491861 | 1.662584 | 2.761174 | 7.381623 | 1.579691 | 1.902436 | 62579 | 0 |



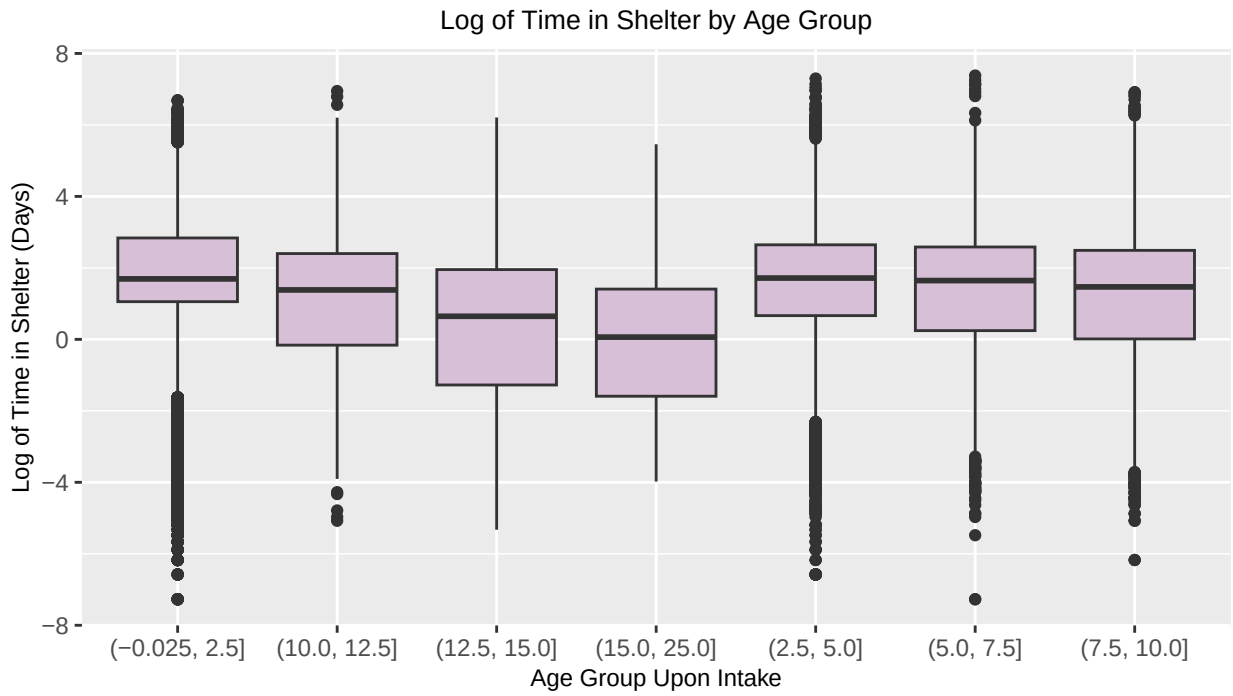




Even using the log of `time_in_shelter_days`, there are significant outliers in the amount of time that various types of animals spend in the shelter, as shown by the box plot below. The median log of time in shelter differs from group to group, suggesting, for example, that the median log of time in shelter may be larger for cats than for dogs. However, further investigation is needed to tell if these differences are significant because of the variability in each group.




Similarly, the next box plot shows that most age groups have both a considerable degree of variability and a considerable number of outliers in the log of time in shelter. The exceptions to this rule are the two age groups with animals older than 12.5, which lack the outliers seen in the other groups. The median log of

time in shelter appears to be smaller for these two older age groups than for the younger age groups, but further investigation is needed to determine if this difference is significant.




## Statistical Analysis

To analyze our data, we will train multiple regression models and perform hypothesis tests on relevant predictors.

### Multiple Regression

With the goal of helping shelter staff identify which animals are most likely to spend the longest amount of time in the shelter, we trained a multiple regression model, Model 1, to predict the duration of an animal's stay. To reduce the impact of the outliers in `time_in_shelter_days` on our model, we used the log of `time_in_shelter_days` as the target feature.

The predictor variables for Model 1 consist of all attributes in the dataset (with the exception of the unique identifier `animal_id`) that would be known to shelter staff after a given animal's intake into the shelter but before its release. This means that we did not include outcome-related features such as `outcome_type` or `outcome_year` as predictor variables, since the values of these features are only known after an animal's release from the shelter.

In total, Model 1 uses twelve predictor variables: `animal_type`, `black_coloring`, `intake_condition`, `intake_type`, `sex`, `neutered_or_spayed_upon_intake`, `neutered_or_spayed_during_time_at_shelter`, `intake_month`, `intake_year`, `intake_weekday`, `intake_hour`, and `age_upon_intake_age_group`.

Model 1 has an $R^2$ value of about 0.2655, suggesting that it explains about 26.55% of variation in the log of `time_in_shelter_days`. A summary of the model is given in the table below; see Appendix B for a summary of the coefficients of the model.

Table 2: Summary of Model 1

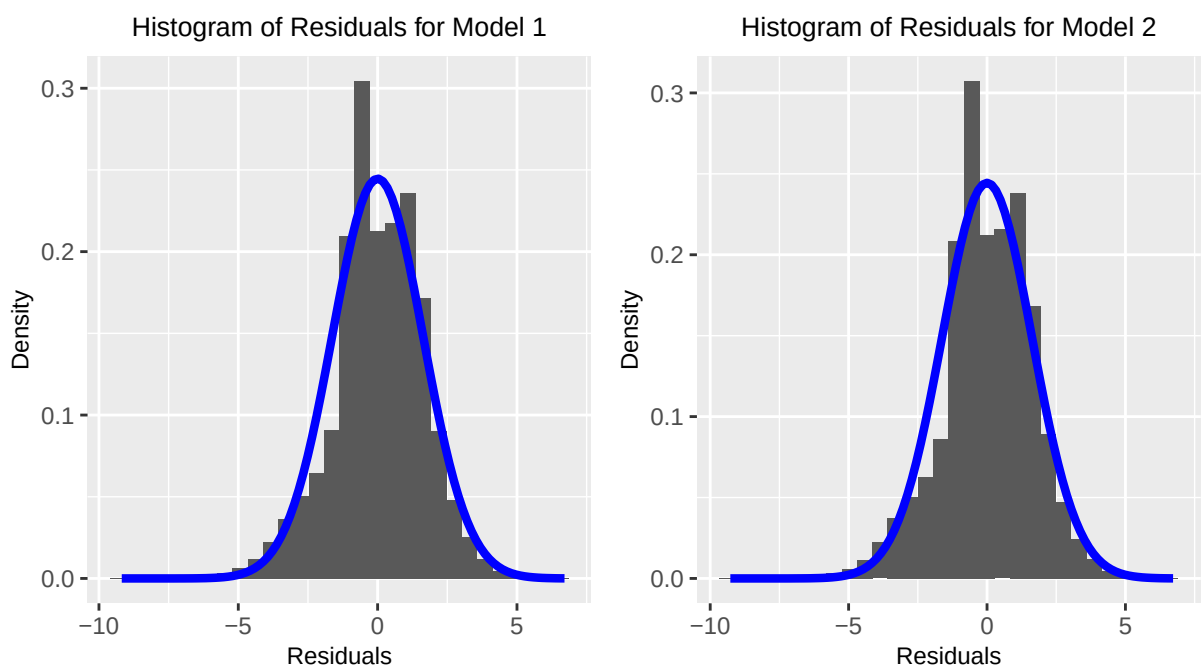
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2655066 | 0.2646958 | 1.631336 | 327.4771 | 0 | 69 | -119386.9 | 238915.7 | 239557.9 | 166352.6 | 62509 | 62579 |

For comparison, we then created a slightly simpler multiple regression model, Model 2, which also uses the log of `time_in_shelter_days` as the target feature. This second model includes most of the same predictors as Model 1, but it omits predictors for which the p-values for all categorical levels were greater than 0.05. Notably, only the feature `intake_hour` met this criterion for removal. A summary of Model 2 is given in the table below; see Appendix C for a summary of the coefficients of the model.

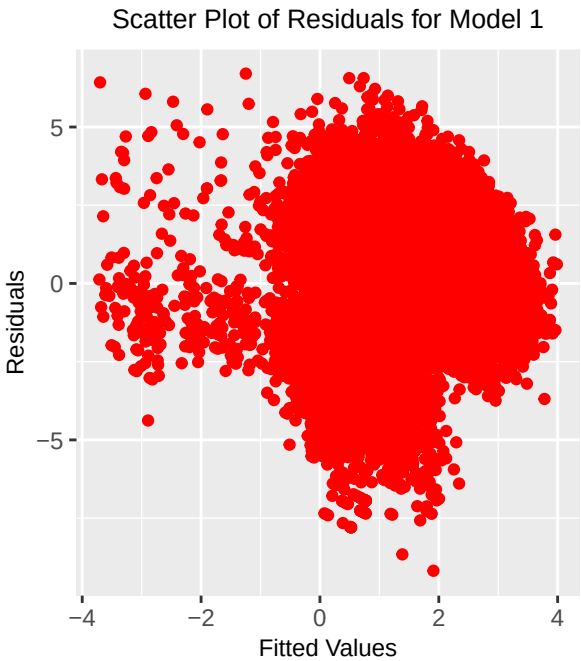
Table 3: Summary of Model 2

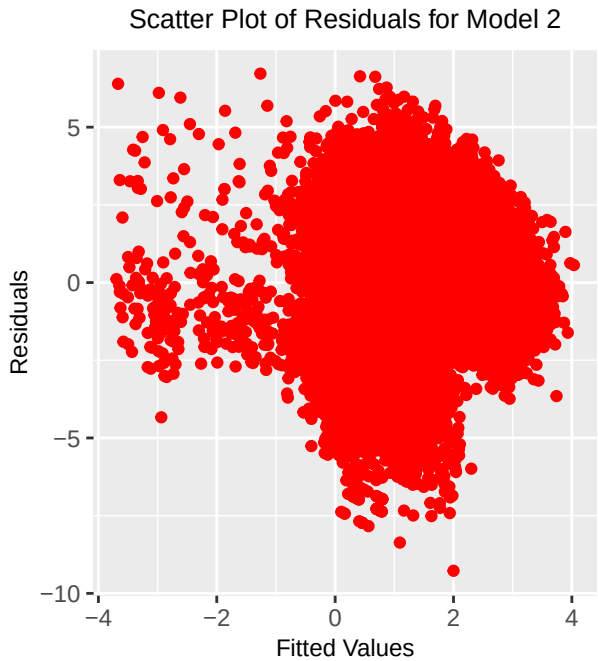
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2641708 | 0.2636296 | 1.632519 | 488.0366 | 0 | 46 | -119443.7 | 238983.4 | 239417.6 | 166655.1 | 62532 | 62579 |

To verify the assumptions behind Model 1 and Model 2, we can examine a histogram of the residuals for each model, overlaid with a normal distribution curve (in blue). As shown in the plot below, we see that the residuals appear to be generally normally distributed for both Model 1 and Model 2. This suggests that the random error in both models follows a normal distribution, satisfying the nearly normal residuals condition.




However, if we look at a scatter plot of residuals vs. fitted values for each model, we see a clear pattern with an uneven spread. This suggests that variability in the residuals is not constant and the relationship between `time_in_shelter_days` and the predictor variables is not linear for either model, violating both the constant variability and the linearity conditions. As a result, we conclude that Model 1 and Model 2 are not appropriate linear models.




However, we would still like to ascertain if either model achieves better performance than the other. Using ANOVA, we determined that there is a statistically significant difference (p-value $< 1\times10^{-8}$) between the performance of Model 1 and the performance of Model 2, despite the fact that Model 2 omits only a single predictor variable from Model 1. The ANOVA test can detect this statistically significant difference because the large sample size lends it high power to detect differences. It should be noted that we have not fully satisfied the condition for ANOVA that requires both models to be valid linear models. As a result, we should interpret the results of this test with some caution. The results of the ANOVA are given in the table below.

Table 4: ANOVA between Model 2 and Model 1

| res.df | rss | df | sumsq | statistic | p.value |
|---|---|---|---|---|---|
| 62532 | 166655.1 | NA | NA | NA | NA |
| 62509 | 166352.6 | 23 | 302.5307 | 4.94259 | 0 |

Since Model 1 has an adjusted $R^2$ value of 0.2647, compared to Model 2's adjusted $R^2$ value of 0.2636, we conclude that Model 1 is a better model than Model 2 (at least statistically, even if not practically).

In light of the violated constant variability and linearity conditions for Model 1, as well as the model's relatively low $R^2$ value, we conclude that a linear model is unsuitable for predicting the duration of an animal's stay in the shelter. Nevertheless, an analysis of the coefficients of Model 1 allows us to make a few general but helpful observations about the significance of the predictor variables. As we do so, we must keep in mind that interpretations must be made with caution because the linearity of the model is in question. (Note that all interpretations of coefficients are made while holding the other variables constant.)

- On average, dogs spend less time at the shelter than other types of animals (p-value $< 1\times10^{-8}$).
- On average, animals that are not black-colored spend less time at the shelter than partially or fully black-colored animals (p-value = 0.0000920).

- On average, pregnant animals spend more time at the shelter than non-pregnant animals (p-value = 0.0045030), while sick animals spend less time than healthy animals (p-value = 0.0000038).

- On average, animals that are surrendered by their owners, are brought in by public assist, or are strays spend more time at the shelter than wildlife or animals brought in for euthanasia (p-value $< 1\times10^{-8}$ for all categories).

- On average, male animals spend less time at the shelter than female animals (p-value = 0.0000046).

- On average, animals that are neutered/spayed upon intake into the shelter or during their time at the shelter spend more time at the shelter than animals that are not (p-values $< 1\times10^{-8}$).

- On average, animals taken in by the shelter between May and November spend more time at the shelter (p-values range from $< 1\times10^{-8}$ to 0.0030472).

- On average, animals between the ages of 12.5 and 25.5 spend less time at the shelter than younger animals (p-values $< 1\times10^{-8}$).

### Hypothesis Testing

To verify our observations from the multiple regression model, we will conduct hypothesis tests for differences of means with regards to animal type and age group. We will use the original `time_in_shelter_days` attribute instead of its log in our hypothesis tests for the sake of interpretability.

**Difference Between Types** Firstly, we will test whether there is a significant difference between the mean lengths of stay in shelter for cats and dogs. Our goal is to confirm or refute our observation based on Model 1 that dogs spend less time in the shelter on average. Our null hypothesis is that there is no such difference, while our alternative hypothesis is that there is such a difference. We will begin by testing the assumptions for applying a t-test for difference in means.

Table 5: Sample Statistics for Cats vs. Dogs

| Group | Count | Variance of Time in Shelter in Days |
|---|---|---|
| Cat | 23598 | 1725.418 |
| Dog | 38214 | 1944.082 |

The sample variations of time in shelter for cats and dogs are not equal (although their ratio, at about 1.13, is less than 3), so we will apply Welch's Two Sample t-test. The number of cats and the number of dogs are both much greater than 30. Therefore, our assumptions for applying the t-test hold.

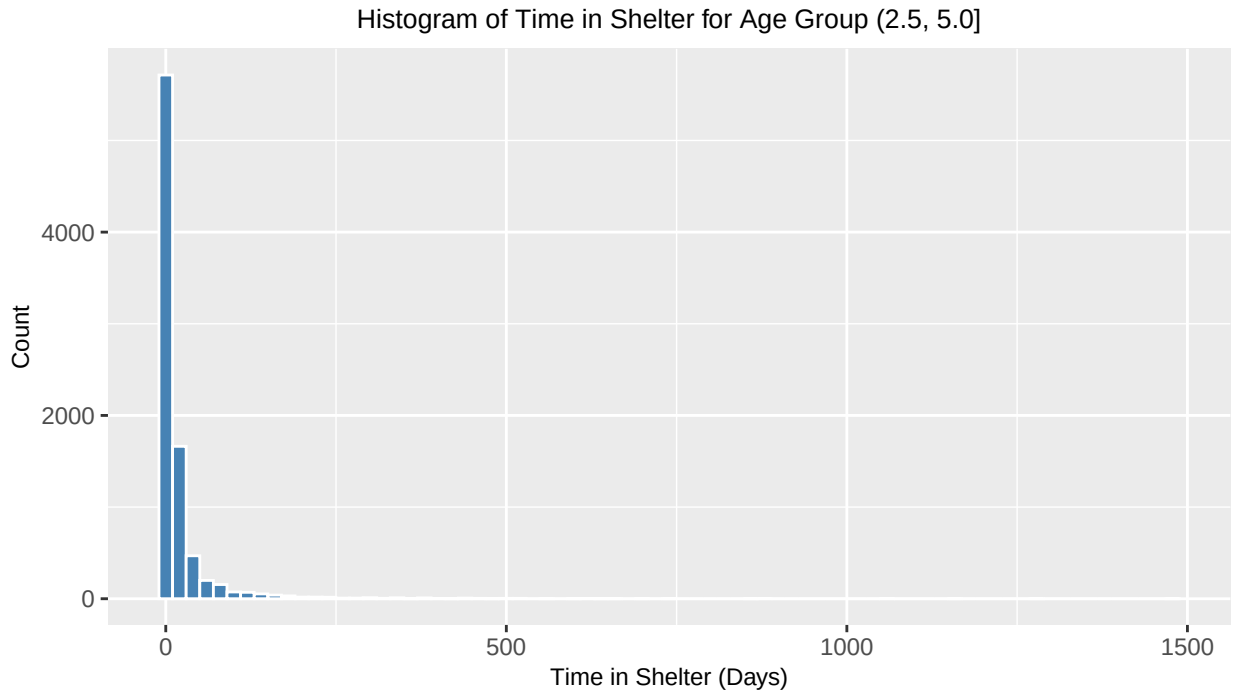
The results of the t-test are given in the table below. We have evidence (p-value $< 1\times10^{-8}$) to reject the null hypothesis in favor of the alternative hypothesis that there is a statistically significant difference in the mean time in shelter in days between cats and dogs. We have 95% confidence that the true difference is in the interval (5.785792, 7.166121). Since all values in this interval are positive, we conclude that, on average, cats spend more time in the shelter than dogs.

Table 6: T-Test for Mean Time in Shelter (in Days) for Cats vs. Dogs

| mean_diff | mean_cats | mean_dogs | t | p.value | df | conf.low | conf.high | method | alternative |
|---|---|---|---|---|---|---|---|---|---|
| 6.475956 | 22.22318 | 15.74722 | 18.39117 | 0 | 52240.33 | 5.785792 | 7.166121 | Welch Two Sample t-test | two.sided |

**Difference Between Ages** Secondly, we will test whether there is a significant difference between lengths of stay in shelter for different age groups. Our goal is to confirm or refute our observation based on Model 1 that animals older than 12.5 years spend less time in the shelter on average than younger animals.

We cannot use ANOVA because the time in shelter for a given age group is not necessarily normally distributed. For example, as shown in the histogram below, the time in shelter spent by the `(2.5, 5.0]` age group is not normally distributed.




Instead, we will use the non-parametric Kruskal-Wallis rank sum test. Our null hypothesis is that there is no difference in median time in shelter between age groups, while our alternative hypothesis is that there is such a difference. The only assumption we need to satisfy for this test is that all sample sizes are large enough. Since all of our sample sizes are much larger than 5, as shown in the table below, this assumption is satisfied.

Table 7: Counts of Age Groups

| Age Group | Count |
|---|---|
| (-0.025, 2.5] | 46345 |
| (10.0, 12.5] | 892 |
| (12.5, 15.0] | 735 |
| (15.0, 25.0] | 196 |
| (2.5, 5.0] | 8611 |
| (5.0, 7.5] | 2821 |
| (7.5, 10.0] | 2979 |

The results of the Kruskal-Wallis rank sum test are shown in the table below. We have evidence (p-value $< 1\times10^{-8}$) to reject the null hypothesis in favor of the alternative hypothesis that there is a statistically significant difference in median time in shelter between age groups.

Table 8: Kruskal-Wallis Test for Median Time in Shelter (in Days) by Age Group

| statistic | p-value | df | method |
|---|---|---|---|
| 589.4827 | 0 | 6 | Kruskal-Wallis rank sum test |

Since our sample sizes are large enough, we can determine which age groups have statistically significant differences in median duration of stay in shelter by using the non-parametric pairwise Wilcoxon rank sum test, adjusting p-values with the conservative Bonferroni method. The results of this test are shown in the table below.

Table 9: Pairwise Wilcoxon Test for Median Time in Shelter (in Days) by Age Group

| Age Group 1 | Age Group 2 | P-Value |
|---|---|---|
| (10.0, 12.5] | (-0.025, 2.5] | 0.0000000 |
| (12.5, 15.0] | (-0.025, 2.5] | 0.0000000 |
| (12.5, 15.0] | (10.0, 12.5] | 0.0000000 |
| (15.0, 25.0] | (-0.025, 2.5] | 0.0000000 |
| (15.0, 25.0] | (10.0, 12.5] | 0.0000000 |
| (15.0, 25.0] | (12.5, 15.0] | 0.4501885 |
| (2.5, 5.0] | (-0.025, 2.5] | 0.0000010 |
| (2.5, 5.0] | (10.0, 12.5] | 0.0000000 |
| (2.5, 5.0] | (12.5, 15.0] | 0.0000000 |
| (2.5, 5.0] | (15.0, 25.0] | 0.0000000 |
| (5.0, 7.5] | (-0.025, 2.5] | 0.0000002 |
| (5.0, 7.5] | (10.0, 12.5] | 0.0000239 |
| (5.0, 7.5] | (12.5, 15.0] | 0.0000000 |
| (5.0, 7.5] | (15.0, 25.0] | 0.0000000 |
| (5.0, 7.5] | (2.5, 5.0] | 0.5410556 |
| (7.5, 10.0] | (-0.025, 2.5] | 0.0000000 |
| (7.5, 10.0] | (10.0, 12.5] | 0.6485470 |
| (7.5, 10.0] | (12.5, 15.0] | 0.0000000 |
| (7.5, 10.0] | (15.0, 25.0] | 0.0000000 |
| (7.5, 10.0] | (2.5, 5.0] | 0.0000000 |
| (7.5, 10.0] | (5.0, 7.5] | 0.0024003 |

Based on these p-values, we can make the following observations:

- The median time in shelter spent by age group `(-0.025, 2.5]` is statistically significantly different from the median times spent by all other age groups (p-values range from 0.0000010 to $< 1\times10^{-6}$).

- The median times in shelter spent by age groups `(2.5, 5.0]` and `(5.0, 7.5]` are statistically significantly different from the median times spent by all other age groups except each other (p-values range from 0.0024003 to $< 1\times10^{-8}$).

- The median times in shelter spent by age groups `(7.5, 10.0]` and `(10.0, 12.5]` are statistically significantly different from the median times spent by all other age groups except each other (p-values range from 0.0024003 to $< 1\times10^{-8}$).

- The median times in shelter spent by age groups `(12.5, 15.0]` and `(15.0, 25.0]` are statistically significantly different from the median times spent by all other age groups except each other (all p-values $< 1\times10^{-8}$).

Now that we have confirmed that there are indeed differences in median time in shelter between age groups, we will investigate whether there is a difference in mean time in shelter between animals who are older than 12.5 years and animals who are 12.5 years old or younger. Our null hypothesis is that there is no such difference, while our alternative hypothesis is that there is such a difference. We will begin by testing the assumptions for applying a t-test for difference in means.

Table 10: Sample Statistics for Animals Older than 12.5 vs. Animals 12.5 or Younger

| Group | Count | Variance of Time in Shelter in Days |
|---|---|---|
| 12.5 or Younger | 61648 | 1865.131 |
| Older than 12.5 | 931 | 1308.250 |

The sample variations of time in shelter for animals older than 12.5 and younger animals are not equal (although their ratio, at about 1.43, is less than 3), so we will apply Welch's Two Sample t-test. The number of animals older than 12.5 and the number of younger animals are both much greater than 30. Therefore, our assumptions for applying the t-test hold.

The results of the t-test are given in the table below. We have evidence (p-value $< 1\times10^{-8}$) to reject the null hypothesis in favor of the alternative hypothesis that there is a statistically significant difference in the mean time in shelter in days between animals older than 12.5 years and animals 12.5 years old or younger. We have 95% confidence that the true difference is in the interval (-9.590813, -4.88845). Since all values in this interval are negative, we conclude that, on average, animals older than 12.5 years spend less time in the shelter than younger animals.

Table 11: T-Test for Mean Lengths of Stay for Animals Older than 12.5 vs. Animals 12.5 or Younger

| mean_diff | mean_older | mean_younger | t | p.value | df | conf.low | conf.high | method | alternative |
|---|---|---|---|---|---|---|---|---|---|
| -7.239632 | 11.06782 | 18.30745 | -6.042552 | 0 | 970.4706 | -9.590813 | -4.88845 | Welch Two Sample t-test | two.sided |

**Summary of Hypothesis Test Results** Our hypothesis test regarding cats vs. dogs found that cats spend more time in the shelter on average than dogs. This supports our observation from Model 1 that dogs spend less time in the shelter on average.

Our hypothesis tests regarding age groups found that there are statistically significant differences in median stay in shelter between various age groups. Additionally, we found that animals older than 12.5 spend less time in the shelter on average than animals who are 12.5 or younger. The latter result supports our observation from Model 1 that animals between the ages of 12.5 and 25.5 spend less time in the shelter on average.

## Conclusion

We undertook our analysis of the Austin Animal Center dataset with the goal of identifying predictors that shorten or prolong an animal's stay at the shelter (the outcome variable `time_in_shelter_days`). We used the log of the outcome variable in our regression models to reduce the impact of outliers, but we used the original attribute in our hypothesis tests for the sake of interpretability.

To predict the duration of an animal's stay in the shelter, we created an initial multiple regression model, Model 1. As predictors, we used all attributes that would likely be known to shelter staff before that animal's release from the shelter. Model 1 achieved an $R^2$ value of 0.2655, suggesting that it explains only 26.55% of the variability in the number of days spent in the shelter. Using an ANOVA test (p-value $< 1\times10^{-8}$) and comparison of adjusted $R^2$ values, we concluded that Model 1 achieved statistically significantly better

performance than a simplified model, Model 2, that omitted the least significant predictor variable in Model 1.

Since some of the assumptions behind Model 1 were violated, we determined that a linear model was ultimately unsuitable for predicting the length of an animal's stay in the shelter. Nevertheless, analyzing the coefficients of Model 1 allowed us to make some key observations about the apparent impacts of various predictors on the target feature. In particular, we noted that the average amount of time spent in the shelter appeared to vary depending on an animal's type and age group upon intake.

To verify these two observations, we conducted hypothesis tests to detect differences in means and medians. We concluded that, on average, dogs spend less time in the shelter than cats and animals older than 12.5 years old spend less time in the shelter than younger animals (both p-values $< 1\times10^{-8}$).

Based on these findings, shelters such as the AAC may wish to reexamine their policies to better care for animals that may have to stay in the shelter for a longer amount of time and to expedite the process of finding these animals a more permanent home. For example, knowing that cats currently spend a longer amount of time in the shelter than dogs on average, the AAC could take steps to reduce the disparity by improving the process of fostering and adoption for cats.

It is worthwhile to note that our findings do not necessarily provide a clear idea of how practically significant the differences in mean and median length of stay at the shelter are (especially since using the log of the amount of time spent in the shelter as the outcome variable for our models reduced the interpretability of the model results). Furthermore, because we limited our analysis to data from a single animal shelter, our conclusions are not necessarily applicable to other shelters. Future research could improve the generalizability of our findings by comparing the Austin Animal Center dataset to data from other shelters.

Additionally, our multiple regression model highlighted multiple predictor variables as significant for predicting duration of stay in shelter. However, we limited our in-depth analysis with hypothesis testing and confidence intervals to only two of these predictors (animal type and age group on intake). A closer examination of the other variables suggested to be significant by our model (such as animals' sex, health status, and time of intake) might yield additional insights into how shelters can improve the efficiency and efficacy of their care for different categories of animals.

## Appendix A: Explanation of Features in Dataset

The features in the cleaned dataset that we used are described in the table below.[4]

Table 12: Dataset Features, Types, and Descriptions

| Feature | Type | Description |
|---|---|---|
| animal_id | character | unique identifier |
| animal_type | character | type of animal |
| black_coloring | character | whether animal has full, partial, or no black coloring |
| time_in_shelter_days | numeric | number of days spent by animal in shelter |
| intake_condition | character | condition of animal upon intake into shelter |
| intake_type | character | type of intake into shelter |
| outcome_type | character | type of outcome for animal |
| sex | character | sex of animal |
| neutered_or_spayed_upon_intake | character | whether animal was neutered/spayed upon intake into shelter |
| neutered_or_spayed_upon_outcome | character | whether animal was neutered/spayed upon outcome |
| neutered_or_spayed_during_time_at_shelter | character | whether animal was neutered/spayed during its stay in the shelter |
| intake_month | integer | month when intake into shelter occurred |
| intake_year | integer | year when intake into shelter occurred |
| intake_weekday | character | weekday when intake into shelter occurred |
| intake_hour | integer | hour when intake into shelter occurred |
| outcome_month | integer | month when outcome occurred |
| outcome_year | integer | year when outcome occurred |
| outcome_weekday | character | weekday when outcome occurred |
| outcome_hour | integer | hour when outcome occurred |
| age_upon_intake_age_group | character | age group of animal upon intake into shelter |

All features in the dataset except for `time_in_shelter_days` are categorical. The levels of these variables are given in the table below.

Table 13: Summary of Categorical Features

| Feature | Levels |
|---|---|
| animal_type | Bird, Cat, Dog, Other |
| black_coloring | black only, not black, partly black |
| intake_condition | Aged, Feral, Injured, Normal, Nursing, Pregnant, Sick |
| intake_type | Euthanasia Request, Owner Surrender, Public Assist, Stray, Wildlife |
| outcome_type | Adoption, Died, Disposal, Euthanasia, Missing, Relocate, Return to Owner, Rto-Adopt, Transfer |
| sex | Female, Male |
| neutered_or_spayed_upon_intake | False, True |
| neutered_or_spayed_upon_outcome | False, True |
| neutered_or_spayed_during_time_at_shelter | False, True |
| intake_month | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 |
| intake_year | 2013, 2014, 2015, 2016, 2017, 2018 |
| intake_weekday | Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday |
| intake_hour | 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 |
| outcome_month | 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 |
| outcome_year | 2013, 2014, 2015, 2016, 2017, 2018 |
| outcome_weekday | Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday |
| outcome_hour | 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 |
| age_upon_intake_age_group | (-0.025, 2.5], (10.0, 12.5], (12.5, 15.0], (15.0, 25.0], (2.5, 5.0], (5.0, 7.5], (7.5, 10.0] |

[4]The descriptions are based on our interpretations of the feature names, since no descriptions were provided by the data source.

## Appendix B: Coefficients for Model 1

Table 14: Coefficients for Model 1

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -0.0840727 | 0.2382602 | -0.3528609 | 0.7241939 |
| animal_typeCat | -0.9761306 | 0.1379239 | -7.0773154 | 0.0000000 |
| animal_typeDog | -1.3602547 | 0.1377171 | -9.8771677 | 0.0000000 |
| animal_typeOther | -0.6725400 | 0.1542582 | -4.3598324 | 0.0000130 |
| black_coloringnot black | -0.0888576 | 0.0227198 | -3.9110137 | 0.0000920 |
| black_coloringpartly black | -0.0606165 | 0.0257003 | -2.3585960 | 0.0183473 |
| intake_conditionFeral | 0.1321259 | 0.2066357 | 0.6394146 | 0.5225556 |
| intake_conditionInjured | 0.0045037 | 0.1055153 | 0.0426827 | 0.9659546 |
| intake_conditionNormal | 0.1001664 | 0.1014639 | 0.9872122 | 0.3235424 |
| intake_conditionNursing | 0.0635013 | 0.1110706 | 0.5717202 | 0.5675136 |
| intake_conditionPregnant | 0.7713819 | 0.2715472 | 2.8406916 | 0.0045030 |
| intake_conditionSick | -0.4952117 | 0.1070794 | -4.6247163 | 0.0000038 |
| intake_typeOwner Surrender | 2.0522289 | 0.1150577 | 17.8365187 | 0.0000000 |
| intake_typePublic Assist | 2.4634096 | 0.1170444 | 21.0467890 | 0.0000000 |
| intake_typeStray | 1.6714676 | 0.1144054 | 14.6100405 | 0.0000000 |
| intake_typeWildlife | -1.9488370 | 0.2050624 | -9.5036285 | 0.0000000 |
| sexMale | -0.0600737 | 0.0131064 | -4.5835244 | 0.0000046 |
| neutered_or_spayed_upon_intakeTrue | 0.8206796 | 0.0202035 | 40.6206063 | 0.0000000 |
| neutered_or_spayed_during_time_at_shelterTrue | 2.1150446 | 0.0162360 | 130.2686653 | 0.0000000 |
| intake_month2 | -0.0149280 | 0.0340615 | -0.4382673 | 0.6611940 |
| intake_month3 | 0.0074598 | 0.0335198 | 0.2225488 | 0.8238874 |
| intake_month4 | 0.0561969 | 0.0342939 | 1.6386854 | 0.1012838 |
| intake_month5 | 0.2498813 | 0.0321551 | 7.7711322 | 0.0000000 |
| intake_month6 | 0.2623862 | 0.0326176 | 8.0443037 | 0.0000000 |
| intake_month7 | 0.2146083 | 0.0335314 | 6.4002183 | 0.0000000 |
| intake_month8 | 0.1440423 | 0.0336635 | 4.2788848 | 0.0000188 |
| intake_month9 | 0.1552513 | 0.0335740 | 4.6241471 | 0.0000038 |
| intake_month10 | 0.0977259 | 0.0329815 | 2.9630491 | 0.0030472 |
| intake_month11 | 0.1717597 | 0.0338943 | 5.0675108 | 0.0000004 |
| intake_month12 | 0.0472648 | 0.0343573 | 1.3756829 | 0.1689248 |
| intake_year2014 | -0.1871514 | 0.0325967 | -5.7414253 | 0.0000000 |
| intake_year2015 | -0.0064177 | 0.0327310 | -0.1960729 | 0.8445538 |
| intake_year2016 | 0.0708050 | 0.0330290 | 2.1437223 | 0.0320590 |
| intake_year2017 | -0.0064412 | 0.0331453 | -0.1943310 | 0.8459174 |
| intake_year2018 | -0.1370814 | 0.0487684 | -2.8108672 | 0.0049424 |
| intake_weekdayMonday | -0.0493124 | 0.0242397 | -2.0343610 | 0.0419194 |
| intake_weekdaySaturday | 0.0309750 | 0.0241049 | 1.2850074 | 0.1987945 |
| intake_weekdaySunday | -0.0194853 | 0.0251605 | -0.7744414 | 0.4386727 |
| intake_weekdayThursday | -0.0204042 | 0.0247253 | -0.8252375 | 0.4092399 |
| intake_weekdayTuesday | -0.0636502 | 0.0244955 | -2.5984395 | 0.0093671 |
| intake_weekdayWednesday | -0.0026402 | 0.0243348 | -0.1084952 | 0.9136033 |
| intake_hour1 | 0.1292293 | 0.2295381 | 0.5629969 | 0.5734390 |
| intake_hour2 | -0.2762738 | 0.2427891 | -1.1379168 | 0.2551596 |
| intake_hour3 | -0.1064210 | 0.2890152 | -0.3682195 | 0.7127108 |
| intake_hour4 | -0.3742544 | 0.3432694 | -1.0902645 | 0.2756009 |
| intake_hour5 | 0.0032457 | 0.4692117 | 0.0069174 | 0.9944808 |
| intake_hour6 | -0.1459750 | 0.1898165 | -0.7690323 | 0.4418771 |
| intake_hour7 | -0.1598798 | 0.1334840 | -1.1977452 | 0.2310208 |

Table 14: Coefficients for Model 1 *(continued)*

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| intake_hour8 | -0.1708334 | 0.1331363 | -1.2831466 | 0.1994455 |
| intake_hour9 | -0.0739685 | 0.1309097 | -0.5650346 | 0.5720523 |
| intake_hour10 | -0.2431382 | 0.1271429 | -1.9123218 | 0.0558395 |
| intake_hour11 | -0.2155753 | 0.1243070 | -1.7342163 | 0.0828846 |
| intake_hour12 | -0.1924224 | 0.1245056 | -1.5454917 | 0.1222324 |
| intake_hour13 | -0.1536504 | 0.1246464 | -1.2326899 | 0.2176961 |
| intake_hour14 | -0.1159201 | 0.1248003 | -0.9288449 | 0.3529731 |
| intake_hour15 | -0.1764306 | 0.1249334 | -1.4121966 | 0.1578971 |
| intake_hour16 | -0.0932321 | 0.1249195 | -0.7463372 | 0.4554665 |
| intake_hour17 | -0.0853282 | 0.1258650 | -0.6779344 | 0.4978158 |
| intake_hour18 | -0.1029884 | 0.1260698 | -0.8169158 | 0.4139797 |
| intake_hour19 | 0.0945482 | 0.1390240 | 0.6800854 | 0.4964529 |
| intake_hour20 | 0.1713043 | 0.1641774 | 1.0434095 | 0.2967627 |
| intake_hour21 | 0.1895845 | 0.1538518 | 1.2322542 | 0.2178588 |
| intake_hour22 | 0.1546674 | 0.1521003 | 1.0168777 | 0.3092155 |
| intake_hour23 | 0.0839206 | 0.1529053 | 0.5488405 | 0.5831168 |
| age_upon_intake_age_group(10.0, 12.5] | 0.1055906 | 0.0568753 | 1.8565277 | 0.0633831 |
| age_upon_intake_age_group(12.5, 15.0] | -0.4442426 | 0.0628216 | -7.0714912 | 0.0000000 |
| age_upon_intake_age_group(15.0, 25.0] | -0.7370962 | 0.1194477 | -6.1708696 | 0.0000000 |
| age_upon_intake_age_group(2.5, 5.0] | 0.1397725 | 0.0207093 | 6.7492491 | 0.0000000 |
| age_upon_intake_age_group(5.0, 7.5] | 0.1548569 | 0.0334582 | 4.6283688 | 0.0000037 |
| age_upon_intake_age_group(7.5, 10.0] | 0.1184638 | 0.0330971 | 3.5792845 | 0.0003448 |

## Appendix C: Coefficients for Model 2

Table 15: Coefficients for Model 2

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -0.2287489 | 0.2040292 | -1.1211574 | 0.2622252 |
| animal_typeCat | -0.9929970 | 0.1378700 | -7.2024171 | 0.0000000 |
| animal_typeDog | -1.3683047 | 0.1376647 | -9.9394034 | 0.0000000 |
| animal_typeOther | -0.6946585 | 0.1542052 | -4.5047679 | 0.0000067 |
| black_coloringnot black | -0.0882282 | 0.0227289 | -3.8817664 | 0.0001038 |
| black_coloringpartly black | -0.0610976 | 0.0257088 | -2.3765276 | 0.0174794 |
| intake_conditionFeral | 0.1291498 | 0.2062000 | 0.6263325 | 0.5310992 |
| intake_conditionInjured | 0.0063477 | 0.1053625 | 0.0602466 | 0.9519594 |
| intake_conditionNormal | 0.0921092 | 0.1014430 | 0.9079901 | 0.3638869 |
| intake_conditionNursing | 0.0579908 | 0.1110387 | 0.5222577 | 0.6014927 |
| intake_conditionPregnant | 0.7791662 | 0.2716644 | 2.8681207 | 0.0041306 |
| intake_conditionSick | -0.4952813 | 0.1070714 | -4.6257109 | 0.0000037 |
| intake_typeOwner Surrender | 2.0617288 | 0.1150726 | 17.9167578 | 0.0000000 |
| intake_typePublic Assist | 2.5046645 | 0.1169059 | 21.4246281 | 0.0000000 |
| intake_typeStray | 1.6898114 | 0.1143927 | 14.7720175 | 0.0000000 |
| intake_typeWildlife | -1.9196763 | 0.2050663 | -9.3612493 | 0.0000000 |
| sexMale | -0.0600972 | 0.0131141 | -4.5826395 | 0.0000046 |
| neutered_or_spayed_upon_intakeTrue | 0.8230972 | 0.0202090 | 40.7291876 | 0.0000000 |
| neutered_or_spayed_during_time_at_shelterTrue | 2.1153369 | 0.0162442 | 130.2206906 | 0.0000000 |
| intake_month2 | -0.0169590 | 0.0340747 | -0.4977023 | 0.6186956 |
| intake_month3 | 0.0065958 | 0.0335374 | 0.1966689 | 0.8440873 |
| intake_month4 | 0.0568636 | 0.0343086 | 1.6574183 | 0.0974400 |
| intake_month5 | 0.2488980 | 0.0321652 | 7.7381265 | 0.0000000 |
| intake_month6 | 0.2604395 | 0.0326297 | 7.9816621 | 0.0000000 |
| intake_month7 | 0.2087563 | 0.0335367 | 6.2247193 | 0.0000000 |
| intake_month8 | 0.1397359 | 0.0336748 | 4.1495733 | 0.0000334 |
| intake_month9 | 0.1506802 | 0.0335800 | 4.4872056 | 0.0000072 |
| intake_month10 | 0.0979216 | 0.0329988 | 2.9674319 | 0.0030041 |
| intake_month11 | 0.1719153 | 0.0339068 | 5.0702366 | 0.0000004 |
| intake_month12 | 0.0473455 | 0.0343715 | 1.3774632 | 0.1683740 |
| intake_year2014 | -0.1877536 | 0.0325981 | -5.7596417 | 0.0000000 |
| intake_year2015 | -0.0072492 | 0.0327371 | -0.2214376 | 0.8247524 |
| intake_year2016 | 0.0689122 | 0.0329958 | 2.0885153 | 0.0367554 |
| intake_year2017 | -0.0097141 | 0.0331080 | -0.2934056 | 0.7692131 |
| intake_year2018 | -0.1425456 | 0.0487733 | -2.9226148 | 0.0034723 |
| intake_weekdayMonday | -0.0469381 | 0.0242451 | -1.9359831 | 0.0528743 |
| intake_weekdaySaturday | 0.0261525 | 0.0240860 | 1.0857998 | 0.2775718 |
| intake_weekdaySunday | -0.0203145 | 0.0251235 | -0.8085842 | 0.4187574 |
| intake_weekdayThursday | -0.0201814 | 0.0247324 | -0.8159878 | 0.4145102 |
| intake_weekdayTuesday | -0.0624301 | 0.0245037 | -2.5477799 | 0.0108435 |
| intake_weekdayWednesday | -0.0026130 | 0.0243411 | -0.1073498 | 0.9145119 |
| age_upon_intake_age_group(10.0, 12.5] | 0.1132550 | 0.0568939 | 1.9906366 | 0.0465252 |
| age_upon_intake_age_group(12.5, 15.0] | -0.4455717 | 0.0628409 | -7.0904747 | 0.0000000 |
| age_upon_intake_age_group(15.0, 25.0] | -0.7335378 | 0.1195092 | -6.1379208 | 0.0000000 |
| age_upon_intake_age_group(2.5, 5.0] | 0.1403325 | 0.0207115 | 6.7755773 | 0.0000000 |
| age_upon_intake_age_group(5.0, 7.5] | 0.1564083 | 0.0334723 | 4.6727684 | 0.0000030 |
| age_upon_intake_age_group(7.5, 10.0] | 0.1176916 | 0.0331099 | 3.5545716 | 0.0003789 |