Predicting Length of Stay for Shelter Dogs

Rachael Latimer

As people stayed home for most of the early days of the coronavirus pandemic, the demand for goods and services increased. Some of the increased demand was not surprising and likely expected; items such as home workout equipment, trampolines, and lumber. However, some of the demand took the industry by surprise: yeast for baking, and pets. In fact, the interest in pet adoption increased so much that shelters were regularly reporting empty kennels and sifting through dozens of adoption applications for a single puppy.

Unfortunately, as vaccines were rolled out and people began returning to work and school, shelters and foster groups filled up with animals that were no longer compatible with people's lifestyles. The decrease in demand for dogs means that people can be more selective in the kind of dog they adopt. However, the information provided by animal shelters and rescue groups are often based on a short period of time with the animal and the animal's appearance. One shelter's Border collie mix might be another's spaniel or shepherd mix. This best guess breed identification can have significant impacts on a dog's future and could be the difference between adoption and euthanasia. Objectively identifying the impact of a dog's listed breed on the length of stay in an animal shelter could provide shelters with the information needed to shift away from listing a dog's breed as the primary information for potential adopter and toward a more holistic evaluation of a dog's temperament and future needs.

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Data

The data were obtained from [kaggle] (https://www.kaggle.com/aaronschlegel/austin-animal-center-shelter-intakes-and-outcomes?select=aac_intakes_outcomes.csv). It was originally provided by the Austin Animal Center in Austin, Texas. The data include information about the intake and outcome of the animal, and details on the type and condition of the animal. A brief examination of the data revealed that the animal shelter takes in animals in addition to typical domestic pets (cats and dogs). For the purposes of this study, the following types of animals were excluded: cats, birds, and animals that were classified as other, including rabbits, bats,

snakes, raccoons, ferrets, reptiles, and other wild animals that live in close proximity to humans. Additionally, dog breeds with sample sizes less than 20 were excluded as this small sample made it difficult to accurately model the length of stay for the breed. The final data set included variables of the animal (breed, age on intake, sex, condition of the animal), circumstances of the animal arriving at the shelter (type of intake, month of intake), and specifics of the outcome of the animal (outcome, month of outcome, time spent in the shelter, measured in days).

Initial data visualization was performed to understand the data available. This included visualizing the number of animals taken into the shelter each month (Figure 1) and further exploring the number of each type of animal taken in each month (Table 1).

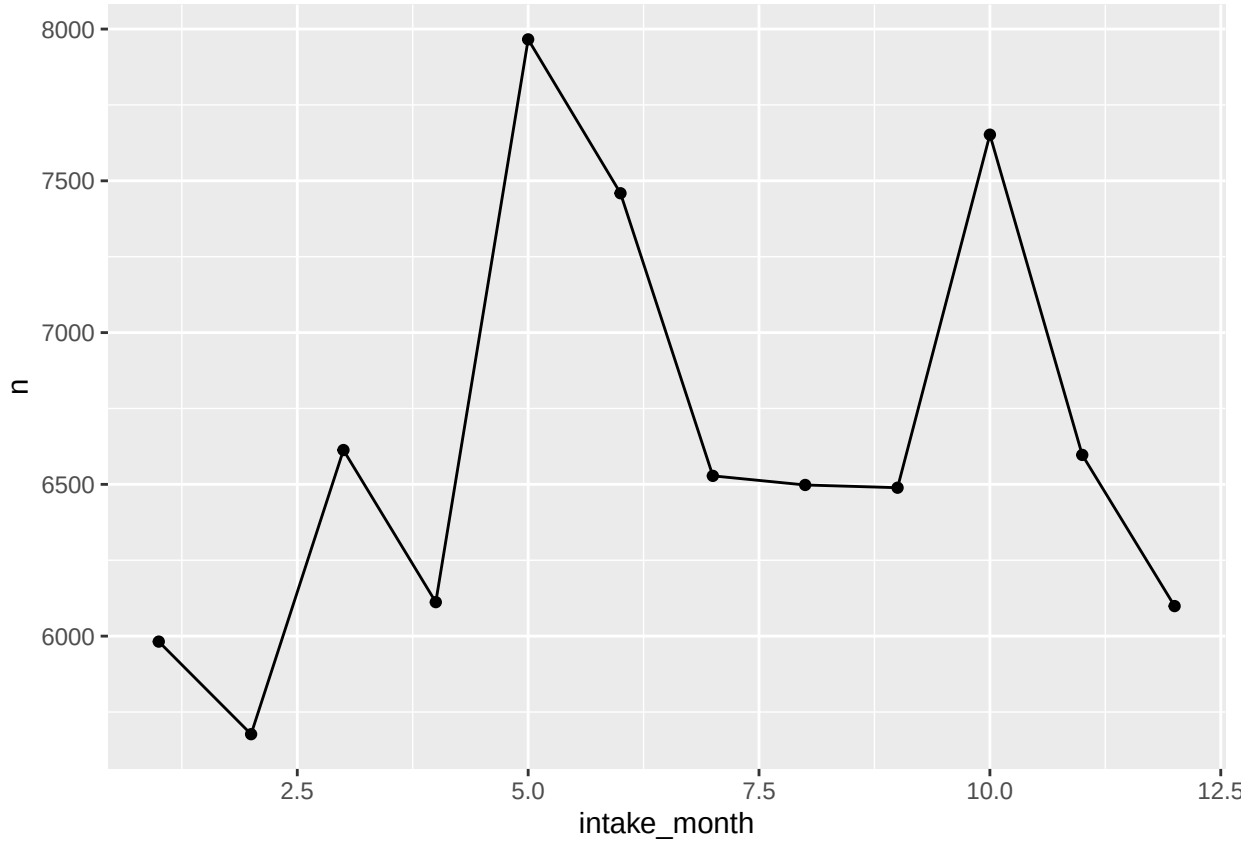

Figure 1. *Shelter Animals Taken in Each Month*

Table 1. *Type of Shelter Animal Taken in Each Month*

	Animal Type			
Intake Month	Bird	Cat	Dog	Other
1	19	1529	4181	253
2	61	1320	3970	326
3	26	1600	4096	891
4	35	2330	3353	394
5	38	3741	3867	320

6	24	3519	3556	360
7	25	2893	3312	298
8	22	2819	3271	386
9	23	2756	3420	290
10	29	3003	4215	405
11	19	2285	4038	255
12	18	1744	4087	250

Models

```
##    alpha lambda
## 35     0   1.71
```

```
## [1] 0
```

```
## [1] 163
```

```
## $everything
##    user  system elapsed
##   36.09    0.61   13.16
##
## $final
##    user  system elapsed
##    3.36    0.02    1.05
##
## $prediction
## [1] NA NA NA
```

Three types of modeling approaches were explored: linear regression, linear regression with ridge penalty, and bagged trees.

Results

Table 2. *Linear Regression with Ridge Penalty Predictors*

```
##                                    [,1]
## na_ind_outcome_type            -29.54081
## breed_American.Bulldog          24.37111
## breed_Bulldog                   20.68927
## (Intercept)                     19.10515
## breed_Collie.Smooth             18.35328
## breed_Tibetan.Spaniel          -18.22182
## breed_English.Coonhound         17.58684
## breed_American.Pit.Bull.Terrier 17.40489
## breed_Silky.Terrier            -16.77498
## breed_Flat.Coat.Retriever       14.29413
```

Comparing the three models, reveals that the (model) has the lowest RMSE (Table 3).

Table 3. *Model Comparison*

```
##                                     Model     RMSE      MAE       Rsq
## 1                     Bagged Trees Model 34.61954 12.44347 0.2594215
## 2 Linear Regression with Ridge Penalty 37.66347 16.57493 0.1068479
```

Discussion

Conclusion

References

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.