Predicting Timely Adoption

and Optimizing Adoption Channels for Shelter Animals

Rebecca V. Lewis

Bellevue University

DSC 630 - Amirfarrokh Iranitalab

August 8, 2020

Abstract

In areas where pet overpopulation is high, the rate at which animals are entering shelters greatly outweighs the number of animals leaving through permanent adoption, temporary fostering or transport outside of the region. Despite the extraordinary efforts of the animal welfare community, euthanasia is often unavoidable, especially in areas of the South where many pet owners have not adopted the spay and neuter practice for their pets. The objective of this project is to determine if taking a closer look at shelter data and animal characteristics can not only predict the likelihood that an animal will be adopted out quickly, but also direct shelter managers to make subtle changes that will decrease the length of stay for animals and effectively route them through the best adoption channel.

Introduction/Background

According to research estimates from the American Society for the Prevention of Cruelty to Animals (ASPCA), over 6.5 million pets enter shelters each year [3]. Of those, approximately 1.5 million animals are euthanized [3]. Very often when trying to decrease euthanasia, animal welfare agencies with abruptly stop euthanizing animals, which often creates larger problems more detrimental to the health of the entire population in the shelter. In reality, a completely “No Kill” community does not exist; in fact, other countries, such as Italy, have passed this type of legislation and have suffered consequences [2]. They inadvertently created communities where the numbers of animals in shelters and the cost to run those shelters has increased, all while the standard of care for the animals has decreased. This type of legislation can also cause more animals to be abandoned on the street which can turn into a public safety issue. Both municipal and privately owned shelters are always looking for and creating programs to move their adoptable animals faster. Depending on the program, additional resources may be needed to keep them running in an industry where resources, specifically financial,

are already running low. Reducing the length of stay for animals in the shelter through adoption, as opposed to euthanasia, is a cost-effective solution that can make room for new animals coming in and optimize the shelter's resources. Using characteristics about the animals to place them in the appropriate adoption program can increase their likelihood of adoption.

Methods

Data Description

The data used to conduct this analysis comes from the Louisiana SPCA located in Algiers, Louisiana. Input features will be taken from dogs and cats whose intake date was between January 1, 2015 and December 31, 2019 with an adoption date by February 29, 2020. This cutoff date was chosen to avoid any inconsistencies related to changes to operations made in response to the COVID pandemic. The data was downloaded from their shelter management system PetPoint which limited the date range to two years at a time. The primary data file included intake data with the result of adoption. Each observation in this file represents an event where an animal became part of the shelter population and was ultimately adopted. Additional files specific to intake and outcome had to be downloaded for more detailed variables. Dogs and cats will be analyzed separately as they have different factors that can impact adoptability. All intakes happened at the main campus of the Louisiana SPCA. Only animals with an outcome of adoption are included in the datafile. The following table describes the information that was reviewed further for potential feature use.

Column Name	Description	Data Type
AnimalID	Unique identifier for the animal	String
Species	Categorical variable specifying species of the animal.	String
PrimaryBreed	Primary Breed of the animal.	String
SecondaryBreed	Secondary Breed of the animal.	String
Gender	M or F indicating the gender of the animal.	String
DOB	Estimated Date of Birth for the animal	DateTime
IntakeDate	Date the animal was taken into the shelter	DateTime
OutcomeDate	Date the animal was adopted	DateTime

IntakeAgeMonths	Age of the animal in months at the time of intake	Float
OutcomeAgeMonths	Age of the animal in months at the time of outcome	Float
AgeGroup	A grouping to describe the age of the animal.	String
IntakeAsilomar	The Asilomar standard label assigned to the animal at intake.	String
IntakeCondition	A label used to identify the condition of the animal at the time of intake.	String
IntakeType	A label associated with the type of intake being performed.	String
IntakeSubType	The intake subtype further categorizes the type of intake with 42 more labels.	String
FoundZip	The zipcode associated with where the animal was located at intake.	String
IntakeReason	The reason the animal is being turned into the shelter.	String
IntakeAgencyAssociation	Other animal welfare agencies bring animals to the Louisiana SPCA. This field represents if an agency is associate with the intake.	String
InitialStage	The initial stage that the animal was placed in upon intake.	String
OutcomeAsilomar	The Asilomar standard label assigned to the animal at adoption.	String
OutcomeSubType	A category that provides information about the type of adoption	String
Size	The size of the animal in one of four values: Small, Medium, Large, and Extra-Large.	String
LengthOwned	A numeric value associated with the amount of time the animal was owned or in the care of the person bringing the animal to the shelter.	Float
LengthOwnedUnits	The units associated with the length of time the animal was in the care of the person bringing it in. Can be in weeks, months, or years.	String
IntakeLocation	The location where the animal was placed at intake. This will most likely be the room or area of the shelter.	String
IntakeSubLocation	The sublocation where the animal was placed at intake. This will most likely be the kennel number.	String
PreAltered	The spay/neuter status of the animal before it entered the shelter.	String
OutcomeLocation	The location where the animal was located when adopted. This will most likely be the room or area of the shelter.	String
OutcomeSubLocation	The sublocation where the animal was placed at intake. This will most likely be the kennel number.	String

Data Preparation

Preliminary data cleansing steps were performed using Python 3.7 in Jupyter Notebooks with the pandas and numpy libraries. The main data files that included the intake with results for two-year increments were concatenated. Summary statistics were viewed on each variable provided in the files. Variables with many nulls, little variation in values, and personal information related to the employee or client were removed. Observations that resulted in an adoption after 2/29/2020 were removed from the dataset. If an animal was adopted and returned to the shelter, that animal will have more than one record in this raw data file. Returned animals represented a very small percentage of the total number of records so they were filtered out of the data. The location data file was joined to the initial dataset,

and validation was performed to ensure the overall observation count remained the same. The initial dataset prior to exploratory data analysis contained 10490 observations and 32 variables; among which 4617 were cats and 5873 were dogs.

Exploratory Data Analysis

Exploratory data analysis was performed using R in the RStudio IDE. Visualizations were completed the ggplot2 library and dplyr aided in subsetting the data. Key steps taken to prepare the data for analysis include:

- Kittens and puppies below 2 months of age were removed from the dataset because they are often in foster care for weeks until they are old enough to be placed for adoption. 2359 records were removed from the dataset meeting this criterion.
- Through visualization of each variable exploratory data analysis, it was discovered that some variables are specific to intake types. These variables would only be impactful on the analysis if the data was modeled for each intake type separately which is beyond the scope of this project. IntakeSubType, IntakeReason, IntakeAgencyAssociation, LengthOwned, and LengthOwnedUnits were removed from the dataset.

Outliers

The following outliers and data challenges were discovered and resulted in changes to the observations included for analysis:

- Outliers were identified in Intake and Outcome Age variables. The minimum values were negative. These observations were removed through our preparation step of removing animals under 2 months of age.

- Intake Asilomar ratings contained blank and unknown values affecting 49 observations. These were removed.
- Animals with an Intake Type of Return were found in the dataset. Because we decided to exclude returned animals from this analysis, these 7 records were removed.
- Gender contained one unknown value for cats. This observation was removed.

Irrelevant and Redundant Variables

Variables from the original dataset were eliminated based on irrelevance and correlation with other similar variables.

- DOB, IntakeAgeMonths, OutcomeAgeMonths, AgeGroup - The intake and outcome age in months were plotted with grouping by age group because a high correlation was suspected. This correlation was confirmed so only one variable related to the age of the animal was kept as a feature, IntakeAgeMonths.
- SecondaryBreed for both cats and dogs were redundanct with the information that was provided by the PrimaryBreed variable

Feature Selection and Extraction

The data preparation and exploratory data analysis processes resulted in the variables in below to be used for feature extraction.

Numeric Variable Summary

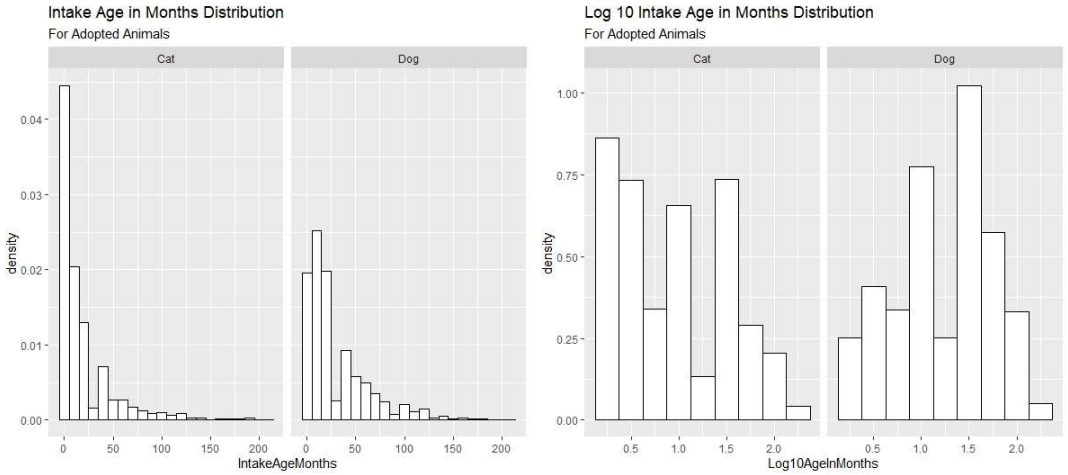
The age of the animal in months is the only numeric variable. The distribution for both cats and dogs is not normal with a positive skew. After a log10 transformation is applied to the variable the distribution is more uniform for cats and close to normal for dogs.

Dogs

Variable	Min	Max	Mean	Std_dev	Skewness	Kurtosis	IQR
IntakeAgeMonths	2	212	30.13829	31.52417	1.73783	6.112354	28

Cats

Variable	Min	Max	Mean	Std_dev	Skewness	Kurtosis	IQR
IntakeAgeMonths	2	204	20.47779	29.36719	2.638367	11.27403	21



Categorical Variable Summary

Most the variables in the dataset are categorical. To comply with most algorithm requirements, they are transformed to a numerical form to be used in models using one hot encoding.

Dogs

Variable	Levels	Mode	Frequency	% of Population
PrimaryBreed	153	Terrier	597	11.02%
Gender	2	Male	2800	51.7%
IntakeDateOnly	1625	2016-08-30	19	.35%
OutcomeDateOnly	1635	2016-07-23	45	.83%
IntakeAsilomar	4	Healthy	3316	61.23%
IntakeCondition	12	Appears Normal	2960	54.65%
IntakeType	4	Stray	2082	38.44%
OutcomeAsilomar	4	Healthy	3258	60.16%
OutcomeSubType	10	Onsite	4444	82.05%
Size	4	Small	2698	49.82%

IntakeLocation	31	10 Dog Healthy Hold	2541	46.92%
IntakeSubLocation	104	Cage 2	494	9.12%
PreAltered	3	N	4168	76.96%
OutcomeLocation	15	00 Dog Adoption Condo	3222	59.49%
OutcomeSubLocation	144	20 Foster Care	611	1128%

Cats

Variable	Levels	Mode	Frequency	% of Population
PrimaryBreed	21	Domestic Shorthair	2247	84.60%
Gender	2	F	1405	52.90%
IntakeDateOnly	1104	2016-04-14	14	.53%
OutcomeDateOnly	1180	2019-08-17	19	.72%
IntakeAsilomar	4	Healthy	1874	70.56%
IntakeCondition	11	Appears Normal	1716	64.61%
IntakeType	4	Stray	1587	59.75%
OutcomeAsilomar	4	Healthy	1904	71.69%
OutcomeSubType	17	Onsite	1243	46.80%
Size	4	Small	1370	51.58%
IntakeLocation	23	04 Cat Feral Room	1901	71.57%
IntakeSubLocation	95	04 Cat Feral Room	358	13.48%
PreAltered	3	N	1646	61.97%
OutcomeLocation	19	00 Cat Adoption Condo	998	37.58%
OutcomeSubLocation	126	SCC	323	12.16%

PrimaryBreed

Dogs contain 153 different levels in PrimaryBreed with Terrier being the most prevalent. There are most likely too many levels here to include all the categories however a few features can be derived. A binary feature to indicate whether the breed is considered a “Pit Bull Type” will be added including the following breeds that are often considered pit bulls: ‘Terrier, Pit Bull’, ‘Terrier, American Pit Bull’, ‘Bulldog, American’, ‘Terrier, American Staffordshire’, ‘Terrier, Staffordshire Bull’, and ‘Bullmastiff’. There are also a few categories that specify no breed for which a ‘Mixed’ variable will be added.

Cats contain 21 different options with Domestic Shorthair being the most frequent. Domestic shorthair, medium hair, and longhair are used when a specific breed is unknown. Including the specific breeds will most likely not be beneficial because of the highly imbalanced weight of the classes. However, a binary feature can be included to indicate whether the breed was specified to see if it plays a role in adoption.

IntakeAsilomar, OutcomeAsilomar

The Asilomar Standard is used to categorize animals to establish a standard when measuring the adoptability of an animal. The Asilomar statuses are Healthy, Treatable-Manageable, Treatable_Rehabilitatable and UnhealthyUntreatable. Overall, the Asilomar rating at the time of the animal's adoption is highly correlated with the rating at intake. There do appear to be some improvement in ratings within the two middle categories. A feature will be derived to indicate whether there was an improvement in Asilomar status before adoption.

IntakeCondition

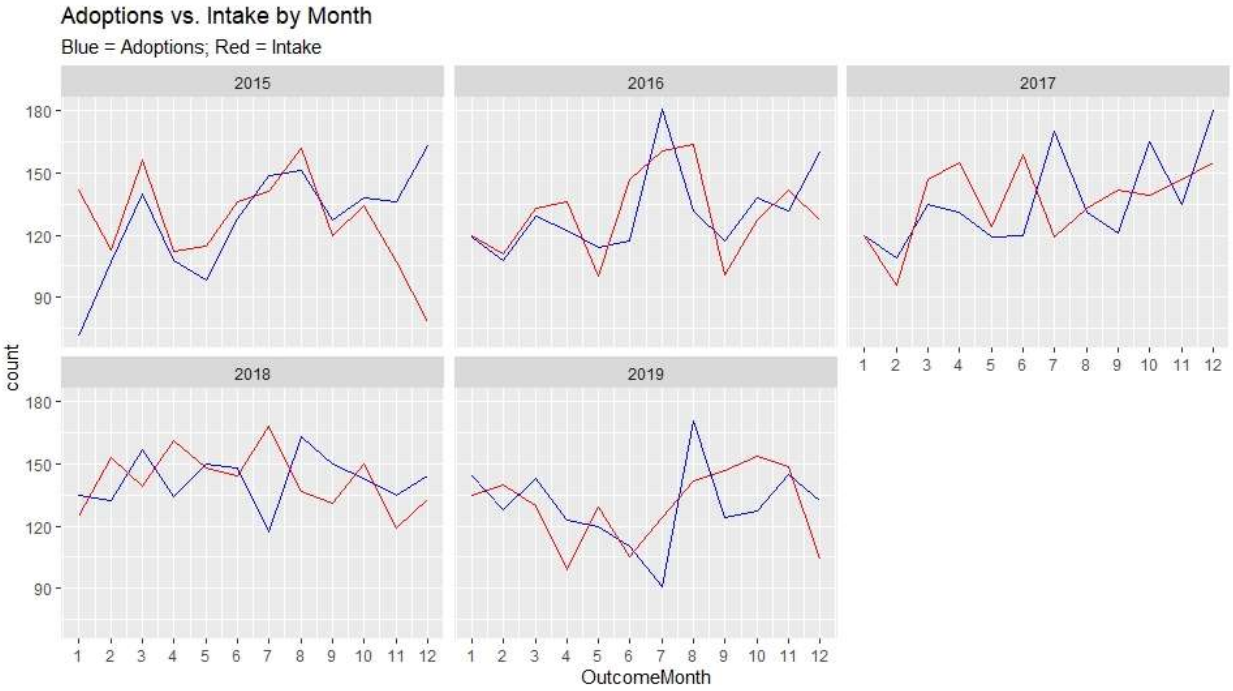
Intake condition can have correlations to a variety of other fields. Because Asilomar status assesses the animal's health in a broader category, the intake conditions of Appears Normal, Normal, and Healthy all correlate directly with the Healthy Asilomar status. The Geriatric intake condition will also correlate highly with age.

Additional information can be derived from this variable to determine if other observed health issues have an impact on length of stay in the shelter. Heartworm positive status, sick or injured (combination of sick, sick and injured and injured/wounds), skin condition, emaciated, and feral are binary features that can be derived from this variable for both cats and dogs.

IntakeDate, OutcomeDate

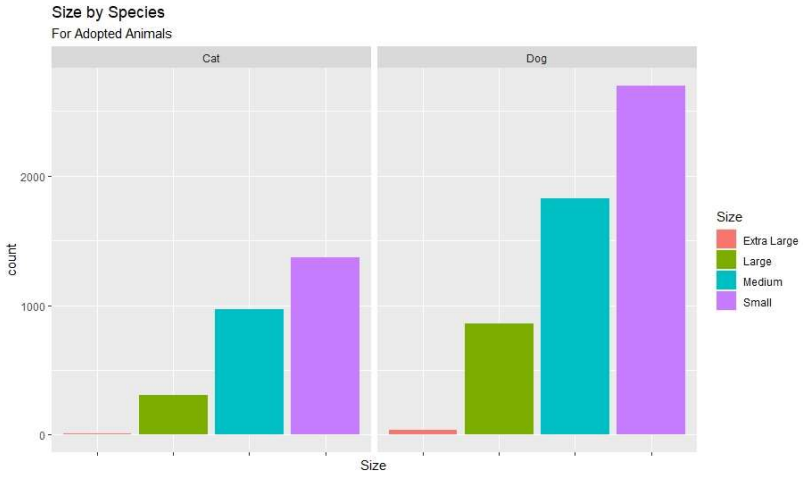
The intake and outcome dates are not useful in their current format and are more helpful as month and year values. The peaks of intake typically happen between July and September each year and appear to have an irregular distribution when separating by year. However, when combined the number of intakes for each month for animals adopted have a uniform distribution.

Adoption outcomes by month were compared to intakes per month for those animals. The peaks for adoptions tend to mirror the peaks for intakes. This is expected since the only intakes we are viewing are those for adopted animals. To prevent future leakage, these values will not be included in features in the model but will be used to derive the target variables.



Size

Size is an assumed influencer in animal adoptability by animal welfare individuals. Looking at the distributions, the classes are highly imbalanced. For analysis, a feature will be created to indicate whether the animal is considered small or not as that appears to be the most frequent class.



PreAltered

Most the animals are not spayed/neutered when they enter the shelter. If an animal is already spayed or neutered at intake, then their time to adoption may be lower since they do not have to wait for surgery. A binary feature is added to represent pre-altered status.

OutcomeSubType

The Outcome Subtype has too many classes to use for a feature as is. For use in predicting the adoption length of stay, it will be used to make binary flags for descriptors about the adoption. For predicting the adoption channel, it will become the target variable as one of the following groups: Barn Cat, Foster, Louisiana SPCA Offsite, Veterinary Partner Offsite, Retail Location Offsite, and Onsite.

An interesting observation about the outcome subtype is that cats seem to have a more diverse distribution among the different adoption channels.

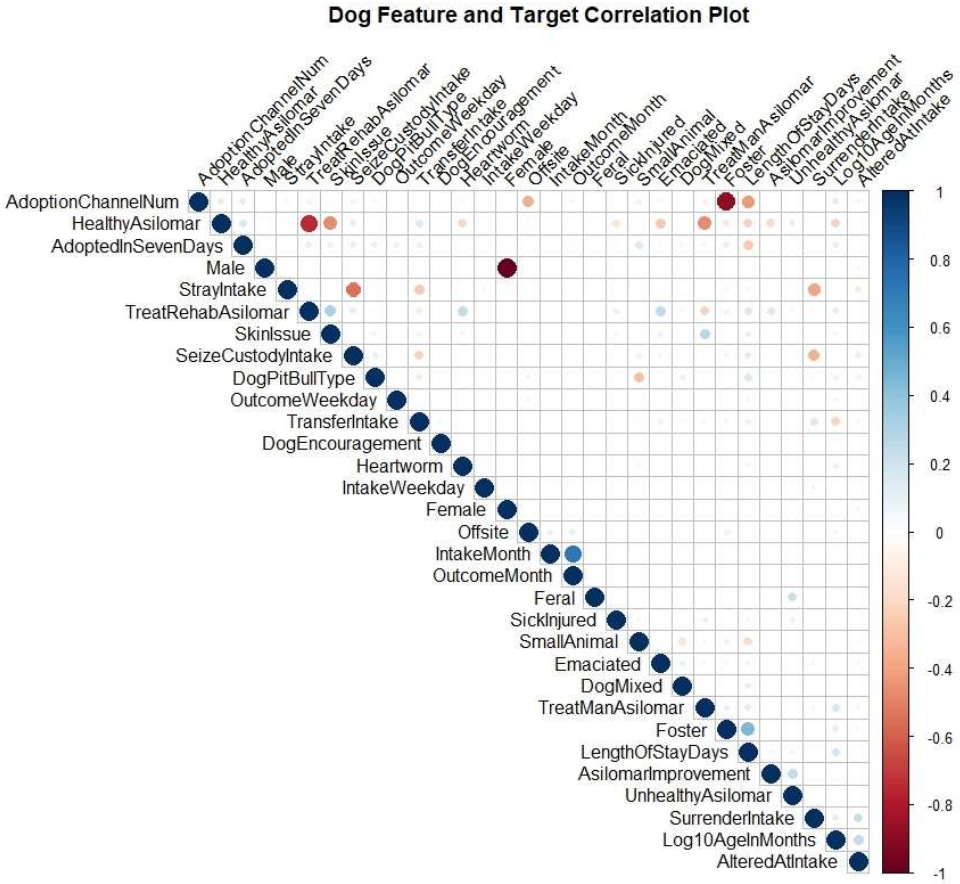
IntakeLocation, IntakeSubLocation, OutcomeLocation, OutcomeSublocation

Much of the information provided by the intake and outcome locations can be derived from other variables. One interesting feature that can be derived from this data is whether the animal spent time in the encouragement room. The encouragement room is used for animals who are not adoptable due to fearful behavior. They are sometimes provided extra time with staff and volunteers to help rehabilitate them. It could be useful to see if spending time in the encouragement room impacts adoptability. This will only apply to dogs.

Correlation Observations

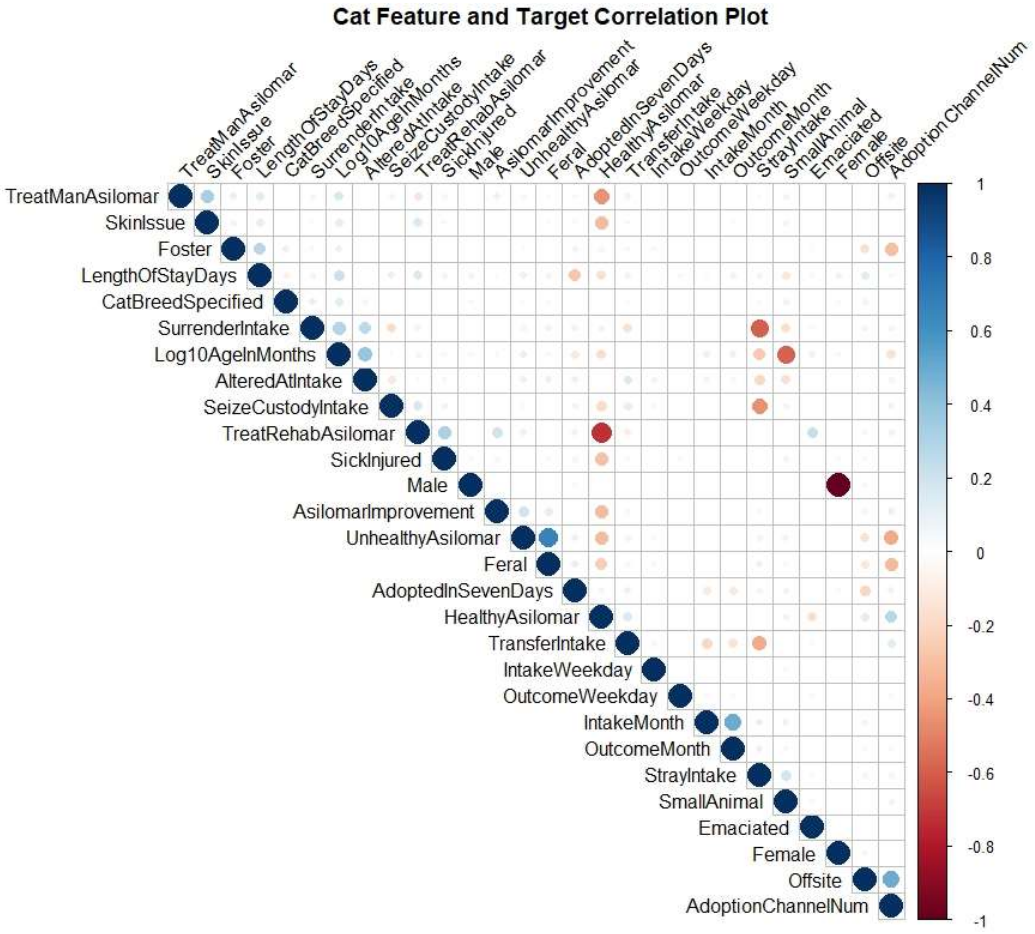
Correlation plots were created for the resulting features for cats and dogs to examine any strong correlations that could result in feature reduction.

- In dogs and cats, the male and female features had a strong negative correlation. Since it is a binary categorical variable, we can remove one of the genders and still derive meaning from the analysis. Female was removed from both datasets.
- Outcome Month was found to have a strong positive correlation with Intake Month which makes sense since our analysis is focused on animals' adoption outcome within the seven days. Outcome Month and Weekday could also introduce future leakage so it was removed from both datasets.



- For cats, the heartworm variable created a null value in the correlation matrix which makes sense since cats rarely contract heartworms. Heartworm was removed from the cat dataset.

- Other moderate correlations were noted in the categorical variables that were one hot encoded. They were left in the analysis to get a measure of influence for each one on the target variables.



Target Variables

The final derived variables to be used as targets are summarized below for cats and dogs individually.

Adopted in Seven Days

The classes for Adopted in Seven Days are highly imbalanced with only 9.1% of cats being adopted in 7 days and 12.4% of dogs being adopted in seven days. A model will need to be chosen for this target with resampling or bootstrapping in mind to balance the classes.

Species	Mode	Frequency	% of Population
Dog	0 - No	4746	12.4%
Cat	0 - No	2413	9.1%

Adoption Channel

Overall, dogs are more likely to be adopted from the Louisiana SPCA main campus while cats have a close second of offsite at a Retail Partner location. Both cats and dogs are just as likely to be adopted from an offsite adoption event sponsored by the Louisiana SPCA. Only cats can be adopted through the Barn Cat program obviously and the least likely place for both cats and dogs to be adopted from is a Veterinary Partner.

Dogs

Class	Frequency	% of Population
Onsite	4444	82.05%
Foster Home/ Fast Track	485	8.95%
Louisiana SPCA Offsite	477	8.81%
Retail Partner Offsite	9	.17%
Veterinary Partner Offsite	1	.02%

Cats

Class	Frequency	% of Population
Onsite	1243	46.80%
Retail Partner Offsite	770	28.99%
Barn Cat	294	11.07%
Louisiana SPCA Offsite	211	7.94%
Foster Home/ Fast Track	113	4.25%
Veterinary Partner Offsite	25	.94%

Modeling

The modeling for this project sought to determine whether we could predict if an animal would be adopted within seven days and whether we could select the appropriate channel based on their

demographics and characteristics. Python 3.7 through Jupyter Notebook was used for this analysis. The numpy and pandas libraries were used for manipulating the data. The scikit-learn library was used for preprocessing, sampling and modeling. Yellowbrick and matplotlib were used for visualizations, and imblearn was used for sampling.

Predicting Timely Adoptions (within 7 days)

For both dogs and cats, the binary target variable indicating whether they were adopted in seven days was highly imbalanced as noted above. A value of 0 represented "Not Adopted in 7 Days" and 1 represented "Adopted in 7 Days." Three models were evaluated on the dataset.

- A Logistic Regression model with upsampling and downsampling on the target class and scaling using StandardScaler.
- A Support Vector Classifier was used with the kernel parameter set to linear, the class weight equal to balanced and probability as true.
- A Random Forest ensemble method was used with the class weight set to balanced and max depths of 3, 5, and 8.

The accuracy, recall and the ROCAUC scores were used to evaluate the models.

Predicting the Optimum Adoption Channel

The target adoption channels were more imbalanced for dogs than cats. The observations with targets of Retail and Veterinary Partner adoption channels were removed from the dataset to focus on the most important channels of Offsite, Onsite, and Foster. The target variables for cats were slightly imbalanced but could be corrected with resampling. The Foster and Offsite variables were removed from the dataset as they provided future leakage to the models creating an accuracy of 100% during preliminary analysis. The following models were used to predict the adoption channel for cats and dogs:

- Three decision tree classifiers were built with depths of 3, 5, and 8 using SMOTE to resample based on the target variable.
- Three random forest ensemble classifiers were built using the class weight of balanced to handle imbalanced classes with maximum depths of 3, 5, and 8.
- K-Nearest neighbor was used with and without SMOTE resampling on the target variable.

Accuracy, recall and the ROC AUC Score were used to evaluate the models.

Results

Because the target variables in both the timely adoption and adoption channel predictions were highly imbalanced, accuracy could not be used as the only metric to evaluate the models. It was used as a consideration but the recall and area under the curve were the primary scores in model selection.

Timely Adoption Prediction Results

Dogs

Timely adoption data was highly biased towards animals that were not adopted within seven days. The model that produced the best accuracy was the Random Forest Classifier with a score of 71.42%. It also had the highest recall for the minority target of Adopted in Seven Days. The ROCAUC score was not the best among the models but there was very little change in this measure for the different models. Overall the balanced Random Forest models performed significantly better for the under represented class than the upsampled or downsampled logistic regression models.

Model Evaluation						
Score	Log Reg (Upsampling)	Log Reg (Downsampling)	SVC	RF (Depth 3)	RF (Depth 5)	RF (Depth 8)
Accuracy	63.08%	61.74%	58.84%	62.08%	68.06%	71.42%
Recall	21.45%	21.10%	75.57%	76.02%	69.23%	63.35%

ROC_AUC	74.79%	74.13%	74.00%	74.12%	74.53%	74.63%

Best Recall Score by Target						
Target	Log Reg (Upsampling)	Log Reg (Downsampling)	SVC	RF (Depth 3)	RF (Depth 5)	RF (Depth 8)
Not Adopted in 7 Days	95%	95%	94%	95%	94%	93%
Adopted in 7 Days	21%	21%	20%	23%	23%	25%

The most influential features that contributed to positive prediction of being adopted within seven days for the Random Forest model (depth of 8) was the age of the dog, a small size, whether the animal was in the foster program, the intake month and weekday, and a healthy asilomar status at intake.

Cats

Contrary to our results for dogs, the model with both the best accuracy (86.32%) and the best recall for the under-represented class (32%) is the Logistic Regression model with downsampled data. The recall for animals adopted in seven days was significantly better than the remaining models attempted at 32%. The ROCAUC score was also significantly better for this model at 85.69%.

Model Evaluation						
Score	Log Reg (Upsampling)	Log Reg (Downsampling)	SVC	RF (Depth 3)	RF (Depth 5)	RF (Depth 8)
Accuracy	60.10%	86.32%	68.07%	75.26%	77.08%	83.24%
Recall	17.07%	31.82%	75.00%	81.25%	72.50%	58.75%
ROC_AUC	83.89%	85.69%	80.16%	82.59%	82.40%	81.87%

Best Recall Score by Target						
Target	Log Reg (Upsampling)	Log Reg (Downsampling)	SVC	RF (Depth 3)	RF (Depth 5)	RF (Depth 8)
Not Adopted in 7 Days	98%	94%	96%	98%	97%	95%
Adopted in 7 Days	17%	32%	19%	24%	24%	29%

The most influential features of accurately predicting whether a cat would be adopted in seven days is whether they were part of the offsite or foster program (negatively), their feral status (positively), whether they were altered at intake (positively), and whether they were sick or injured.

Adoption Channel Prediction Results

The target variable for both dogs and cats were imbalanced and biased towards Onsite adoptions. To manage the class imbalances, the SMOTE (Synthetic Minority Oversampling Technique) was used with the Decision Tree, and KNN Classifiers and the class weight parameter was set to balanced for Random Forest. Using SMOTE with Random Forest, produced the same results as the balanced class weight. Because of the high bias towards predicting Offsite adoptions, the recall of the other classes as well as the ROCAUC were considered in model selection.

A change to the features was made to remove the Foster and Offsite variables because they were creating perfect prediction results which indicated future leakage in both the dog and cat datasets. Feral was also removed from the cat dataset as it is a strong indicator of the barn cat program.

Dogs

The dog adoption channel was evaluated for the Foster, Offsite and Onsite programs only. The best performing model in overall accuracy (78.81%) and recall (79%) was K-nearest neighbors without resampling the data using SMOTE. When looking at the individual recall scores for each class, this model is also the best to predict what animals would be the most suited for the Offsite adoption program. It is not good at accurately predicting animals adequate for foster. The Random Forest models with a depth of 5 or 8 are best at predicting animals for the foster program.

Model Evaluation								
Score	Decision Tree (Depth 3) SMOTE	Decision Tree (Depth 5) SMOTE	Decision Tree (Depth 8) SMOTE	RF (Depth 3) Balanced	RF (Depth 5) Balanced	RF (Depth 8) Balanced	KNN (k=5)	KNN (k=5) SMOTE

Accuracy	42.21%	42.21%	53.64%	54.8%	58.13%	63.28%	78.81%	53.36%
Recall (Wt. Avg)	42%	42%	54%	55%	58%	63%	79%	53%
ROC_AUC	53.40%	53.81%	54.42%	54.85%	55.46%	56.02%	51.68%	56.40%

Best Recall Score by Target								
Target	Decision Tree (Depth 3)	Decision Tree (Depth 5)	Decision Tree (Depth 8) SMOTE	RF (Depth 3) Balanced	RF (Depth 5) Balanced	RF (Depth 8) Balanced	KNN (k=5)	KNN (k=5) SMOTE
Foster Home	12%	11%	14%	13%	15%	15%	6%	14%
Offsite	11%	12%	11%	11%	12%	15%	28%	14%
Onsite	84%	83%	84%	85%	85%	85%	83%	85%

Because K-nearest neighbors does not have a feature importance coefficient measure, random forest was used to determine the most influential features for prediction. The age of the dog, intake month and day, size of the animal and gender of male were the most influential features for this analysis.

Cats

The adoption channels for cats were more evenly dispersed for cats than dogs, but still significantly imbalanced. The best model for overall accuracy and recall was the K-nearest neighbors algorithm without using SMOTE to balance the classes. This model was also the best at predicting the animals most suited for the Barn Cat and Foster programs. The KNN algorithm with SMOTE resampled classes predicted the animals best suited for offsite most accurately. The Decision Tree algorithm with a maximum depth of 8 using SMOTE resampling was the best at predicting the animals that would be well suited to be placed at a retail partner. None of the models could accurately predict Veterinary Partner Placement which could be due to not having enough samples for this class.

Model Evaluation								
Score	Decision Tree	Decision Tree	Decision Tree	RF (Depth 3) Balanced	RF (Depth 5) Balanced	RF (Depth 8) Balanced	KNN (k=5)	KNN (k=5) SMOTE

	(Depth 3) SMOTE	(Depth 5) SMOTE	(Depth 8) SMOTE					
Accuracy	28.28%	31.24%	32.61%	35.23%	37.51%	37.05%	42.03%	31.47%
Recall (Wt. Avg)	28%	31%	33%	35%	38%	37%	42%	31%
ROC_AUC	53.96%	52.40%	53.30%	54.87%	55.83%	55.23%	53.66%	55.35%

Best Recall Score by Target								
Target	Decision Tree (Depth 3)	Decision Tree (Depth 5)	Decision Tree (Depth 8) SMOTE	RF (Depth 3) Balanced	RF (Depth 5) Balanced	RF (Depth 8) Balanced	KNN (k=5)	KNN (k=5) SMOTE
Barn Cat	26%	35%	37%	35%	39%	38%	40%	26%
Foster Home	5%	12%	6%	12%	12%	13%	17%	8%
Offsite	0%	0%	7%	11%	12%	13%	2%	15%
Onsite	57%	55%	55%	58%	58%	54%	52%	56%
Retail Partner	30%	34%	36%	33%	35%	32%	30%	31%
Vet Partner	0%	0%	0%	0%	0%	0%	0%	0%

Because K-nearest neighbors does not have a feature importance coefficient measure, random forest was used to determine the most influential features for prediction. The age of the cat, intake month and day, intake type of stray and healthy asilomar status were the most influential features for prediction the adoption channel for cats.

Discussion

Can we predict if an animal will be adopted in 7 days?

Based on the results of this analysis, the selected models are not very strong predictors of whether an animal can be adopted in seven days. The challenge faced here is that a very low percentage of animals are adopted within that short of a time frame. All of the models were excellent in predicting if an animal was not adopted in seven days because that target represented 80-90% of the dataset which means even if the model classified everything as 0, it would still have an 80-90% accuracy

rate. A better benchmark could be 10 or 14 days. A confounding factor is that some animals require stray holds of three to five days before they can even be placed on the adoption floor which could have contributed to the poor results.

Focusing on the model that performs best for the animals who were adopted in seven days allows us to detect those animals. When decisions are made to pull animals off the adoption floor for a special event or when an animal is selected for a marketing campaign, animals who are not likely to be adopted quickly should be chosen. Basically, an animal that will be identified by this model does not need additional assistance for adoption.

An interesting observation when comparing the results for cats and dogs is that the final model for cats was more accurate at predicting if they were adopted in seven days.

Can we predict the best adoption channel for an animal?

From this analysis, the selected models are not very strong predictors of the adoption channel for cats and dogs. Just like the binary target variable above, the Onsite adoption program represented most of the population so the models would have an inherent bias to predicting that class. Interestingly, it was observed that the best model to predict each class could be different.

Further research could be performed on each program individually to see if animals who were adopted through those programs can be identified. The models observed here as the best for each channel can be a starting point and fine-tuned. Both the Foster Program and Offsite programs are successful ways in which dogs are being adopted which makes sense because they can be taken out to interact with the community easily. Retail Partners and the Barn Cat program scored the highest in true predictions for cats and would be good candidates for further review. Feral cats are in demand in farms and rural areas to control rodent populations. Retail partners like pet stores often have available space

to house cats and cats are preferred because they do not require walks and are typically more independent.

Conclusion

While the results of this study do not solve the problem of reducing the length of stay for shelter dogs and cats, the results indicate that data science can be a helpful tool in improving shelter outcomes and provide a basis for continuing research. The most successful models identified here can be used as detection tools to aid in decision making when placing an animal on the adoption floor, in a program, or as a focus of a marketing campaign. Further development of the best models could be completed to determine if better results are possible. Ideas on additional research to perform based on the results of this analysis:

1. The benchmark for length of stay could be increased to 10 or 14 days for predicting timely adoption.

2. Performing predictive analytics with each intake type separately so intake specific variables excluded in this study can provide further information on what adoption channels those animals would be best for as well as how quickly they are adopted.

3. Performing predictive analytics on the most popular adoption programs for cats and dogs individually can better detect animals suited for those programs.

Glossary

Intake – the event of an animal becoming part of an animal shelter and owned by the organization.

Outcome – the event of an animal no longer being part of the shelter population, either through a live release channel or death.

Asilomar Standards – standards developed by leaders in the field of animal welfare to create a common way to report shelter statistics making it possible to compare their results. [1] These types include:

Healthy: dogs and cats 8 weeks of age or older that have previously shown no behavioral signs that could pose a safety risk or physical signs of disease or injury currently or in the future. [1]

Treatable-Rehabilitatable: dogs and cats who are not healthy but can become healthy if provided medical, foster, behavioral, or standard care typical of pet owners. [1]

Treatable-Manageable: dogs and cats who are not healthy and will not become healthy but could live a satisfactory life if provided medical, foster, behavioral, or standard care typical of pet owners and will not pose a safety risk to humans. [1]

Unhealthy: dogs and cats that have a behavioral trait that poses a safety risk to humans and are not likely to become healthy or treatable if provided medical, foster, behavioral, or standard care typical of pet owners. Dogs and cats that have a disease or injury that adversely affects their quality of life or are under the age of 8 weeks that is not likely to improve if provided medical, foster, behavioral, or standard care typical of pet owners. [1]

Barn Cat Program – adoption program that allows feral cats considered untreatable to be adopted to live in an outdoor rural location

Foster Home / Fast Track Program – a shelter program that places animals who need extra behavioral or medical attention in a volunteers home to help the animal. An animal that is placed in a fast track foster home is usually healthy but has been at the shelter a long period of time. The volunteer takes the animal out into the community and performs an active role in placing the animal with a permanent family.

Louisiana SPCA Offsite Program – programs that bring animals residing in the shelter to an event in the community to find a permanent family.

Onsite – refers to when an animal is adopted from the Louisiana SPCA campus.

Retail Partner Offsite – placement of an animal with a retail location in the community to find a permanent family. The animal stays at that site until adoption.

Veterinary Partner Offsite - placement of an animal with a veterinary office location in the community to find a permanent family. The animal stays at that site until adoption.

References

[1] *A Guide to the Asilomar Accords Definitions*. (n.d.). Animal Humane Society. Retrieved August 8, 2020, from https://www.animalhumanesociety.org/about/guide-asilomar-accords-definitions

[2] Menchetti, L., Mancini, S., Catalani, M. C., Boccini, B., & Diverio, S. (2015). RandAgiamo™, a Pilot Project Increasing Adoptability of Shelter Dogs in the Umbria Region (Italy). *Animals (2076-2615)*, *5*(3), 774–792. https://doi.org/10.3390/ani5030383

[3] *Pet Statistics | Shelter Intake and Surrender | ASPCA*. (n.d.). Retrieved June 13, 2020, from https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics