Animal Shelter Adoption Rate Analysis

Milestone Report

1. Problem Statement

According to the ASPCA, approximately 6.5 million companion animals enter animal shelters in the US each year. Of these, only approximately half are adopted. This low adoption rate compounds into further issues: of the animals that are not adopted, approximately 1.5 million are euthanized[1]. Euthanization is an extremely controversial topic among the animal advocacy community. When faced with overcrowding and limited funding, shelters are left with the choice of whether to allow overcrowding by keeping animals in warehouse shelters with limited care, euthanization, or turning animals away[2]. As a consequence, one of the top reasons animals are turned away from shelters is due to overcrowding.

While there are likely many contributing factors to shelter overcrowding, including mass farming of animals such as in puppy mills and failure to spay and neuter pets, attempting to **maximize existing shelter funds** to improve adoption rates and therefore reduce shelter time is low-hanging fruit that can be completed by shelters without massive effort including buy-in from the public. This maximization may be completed by identifying high-risk adoptive pets that are less likely to be adopted in a timely fashion or by identifying high-adoption times of the year, week, and/or day. These data may ultimately be used for targeted marketing strategies or for reducing unnecessary effort in placing animals with foster families if they are likely to be adopted quickly. Additionally, identifying patterns in animals that are returned to their owners may provide insights into how to replicate success and provide educational opportunities for animals less likely to be returned to their owners.

2. Data Wrangling & Cleaning

2.1 Import and explore dataset

For this analysis, I used the Austin Animal Shelter dataset, which includes intake and outcome data for shelter animals from October 2013 to present.

1. Import dataset using python's Pandas package
2. Explore the shape, structure, and columns of the dataset.
    a. Each row represents a shelter animal, and the dataset contains many columns with descriptive data on each animal.

---

[1] Data from the ASPCA
[2] PETA

    b. I determined during this step that there was already a column that displayed number of days that each animal spent in the shelter; however, it needed to be converted to a datetime value.
3. Split the data into two subanalyses.
   I determined that the shelter times were distributed with an extremely long right tail: 95% of animals were adopted in <71 days, adn 95% of dogs were adopted in <65 days. To answer the question of what affects adoption time in *most* dogs, I used only dogs with adoption times of <65 days. To answer what affects adoption times in dogs that spend a lot of time in the shelter, I looked only at dogs with adoption times of >65 days.

2.2 Adoption times by animal type

4. I used slicing to determine that there were dogs, cats, birds, and 'other' in the dataset, with dog appearing most frequently.
5. I used a boxplot to briefly analyze each group.
    a. The vast majority of animals (75% for all categories) fall into a relatively short adoption time range of less than 20 days.
    b. All groups had a substantial number of outliers beyond the whiskers of the boxplot, making it difficult to interpret this group.
6. I sliced the animals into groups by each animal type.

2.3 Adoption times by dog breed

7. I first looked at dog breeds and determined that the formatting of this column created a lot of complexity due to the presence of mixed breeds. I created a column that returns 1 or 0 if the dog is a mixed breed or pure breed, respectively.
8. I split the breed column into a list of breeds and then used Panda's get dummies function to return 1 or 0 if the list contained a specific breed.
9. I concatenated the dummies dataframe back to the dogs dataframe.
10. I then used slicing and a melt reshape the table to be long with fewer columns instead of wide, with a new column containing a breed name and a corresponding column with 1 or 0 if the dog contained that breed.
11. I used a violin plot and a boxplot to show adoption times by breed, split by purebred and mixed breed. Both plots were difficult to read due to the number of outliers, and for the boxplot, I hid outliers to get a better grasp on the majority of the data.
12. I found that due to the number of breeds, I had to create another column showing only the most common breeds and lumping the others together as "other". I chose breeds that appear >400 times as the cutoff.

2.4 Dog adoption times by AKC Dog Group

13. Since there weren't any clear patterns shown when I looked by breed, I pulled a table that shows breeds by AKC Dog Group in using Pandas and merged this dataframe with the dog breeds dataframe. Once again, notable patterns were difficult to determine with outliers present, so I hid them to see only the majority of the data.

14. I also found that many of the dog breeds were named differently between the shelter dataset. I identified differences and renamed breeds in the AKC dataset to match those in the shelter dataset wherever appropriate.

2.5 Purebred v. Mixed Breed Dogs

15. I looked at all dogs (with outliers hidden) using just purebreds vs mixed breeds and saw only a slight increase overall on adoption times for mixed breeds.

2.6 Adoption time by dog age

16. I also looked at age of dogs using a scatterplot to determine if age affects adoption times, but the age variable was not continuous (rounded by year) and created odd groupings. I decided to instead use age groups as a categorical to look at age as a bar chart.

2.7 Adoption time by coat color

17. I used a very similar analysis to the get dummies for this analysis, as described above.

2.8 Repeat analyses looking only at 'outliers'

I repeated all the above analyses but sliced using only dogs with the highest 5% of adoption times (>65 days) to see if we could learn more about the outlier groups hidden from the initial analyses.

2.9 Analyses by adoption times

I looked at patterns in adoption times by season/time of year, time of day, and day of week.

3. Initial Findings

3.1 Common Trends in Adoption Times

For this analysis, I looked for trends in adoption times for most animals to make determinations about what the shelter can expect for the average animal.

a. **Animal type**
   Dogs are the most common, but cats have longer shelter tenures.

b. **Dog breed**
   Pit Bulls and Chihuahuas are the most common animal types in this shelter. Labrador Retrievers and German Shepherds are also very common. Of these 4, Labs and Pit Bulls have significantly longer adoption times; Chihuahuas actually have significantly shorter adoption times than the rest of the group.

c. **Dog breed group (AKC)**

The Working, Herding, Sporting, and Miscellaneous groups have significantly longer adoption times. Of note, the Sporting group includes Labs, which we know to be both more common and have longer adoption times.

d. Mixed breed v. purebred

Mixed breed dogs were found to have significantly longer adoption times than purebred dogs.

e. Dog age

The youngest dog group actually had the longest adoption times; this shelter seems to have pretty rapid turnover, since the average adoption time for dogs is only 9 days. I would speculate that this result is due to the overall composition of dogs in the shelter being young and the overall adoption time being fast.

f. Black dog syndrome

Blue, brown brindle, fawn, and white colored dogs have significantly longer adoption times. Additionally, black dogs did have longer adoption times, but not by a large margin - the mean adoption time was 8.73 days for black dogs and 9.02 for all other coat colors ($p = 0.008$).

g. Coat color

I looked at whether being single colored or multicolored had an effect on adoption time and found that dogs with multicolor coats do have significantly longer adoption times.

3.2 Trends for animals with long adoption times

For this analysis, I looked at animals with the longest adoption times (the top 5th percentile). The goal is to identify trends in these animals that are more likely to have long adoption times.

a. Animal type

In this extreme group, dogs had significantly longer adoption times and cats had significantly shorter adoption times. Both cats and dogs were nearly equally well-represented in this group.

b. Dog breed

In this group, only American Bulldogs had significantly longer adoption times, with a mean adoption time of 200 days. Additionally, Pit Bulls were by far the most common breed in this group at over 30%.

c. Dog breed group (AKC)

There weren't extreme trends in breed groups for this analysis, although the Hound group did have slightly longer adoption times (mean 180 days, $p=0.03$).

d. **Mixed breed v. purebreed**
   There was no effect of mixed breed v. purebred on adoption times in this extreme group.

e. **Dog age**
   In a reversal from the majority, young dogs had significantly shorter adoption times and dogs 2.5-5 years and senior-aged dogs (15 years+) had significantly longer adoption times.

f. **Black dog syndrome**
   Again, there was no support for black dog syndrome at this shelter. Brown Brindle dogs were the only coat color with a significantly different adoption time: they had a shorter adoption time.

g. **Coat color**
   There was no significant effect of multicolor v single color coats on adoption times in this group.

3.3 Temporal patterns in adoption times

Here, I looked at patterns in when animals are adopted to gain insight into optimal times for the shelter.

a. **Season**
   Overall, there was a clear seasonal trend in adoptions, with more adoptions during the summer. Interestingly, when I parsed out dogs and cats, there was no seasonal trend for dogs, but we did see a seasonal trend for cats.

b. **Time of day**
   There was a clear peak in adoption times flanking a standard business day - between 8-10 AM and 4-6 PM. This trend was more pronounced in dogs than cats, which had a pretty steady adoption time throughout the day.

c. **Day of week**
   I looked at average adoptions per day and found that weekends had significantly higher adoption rates, while late-week days had significantly lower adoption rates (Wednesday-Friday).