Bellevue University

Pet Adoption Prediction

Ashley Mayo

April 2nd, 2025

Professor Iranitalab

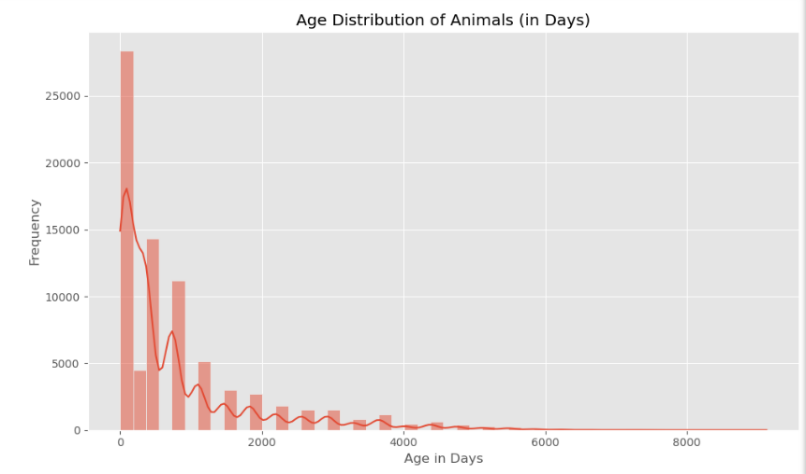
DSC680: Applied Data Science

**Introduction**

There is a moment in every animal shelter, the quiet pause just before a new animal is brought in. A dog abandoned on the side of the road, a cat surrendered by a family in crisis, a litter of newborn kittens left in a box by the door. These animals arrive with no voice. Some will find loving homes, embraced by families who will make them part of their world. Others, however, will not be so lucky. Some will be transferred to other facilities, others will remain in kennels for months, and for a heartbreaking few, the journey will end too soon.

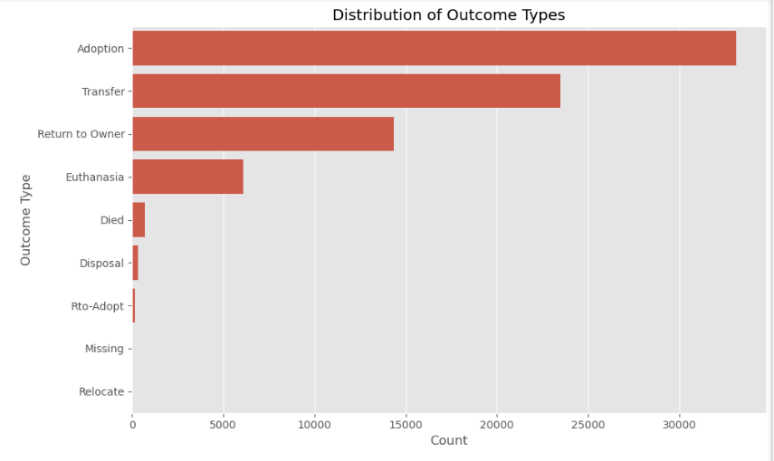
The story behind this project is one that begins with a simple but powerful question: what makes the difference between an adopted animal and one that isn’t?

The Pet Adoption Prediction project was created in response to this reality, to use data science not only to tell these stories, but to change them. The question at the heart of this work is both simple and profound: what determines whether an animal is adopted or not? Through the lens of structured shelter data and analytical tools, this project aims to help animal shelters uncover the hidden patterns that lead to different outcomes. By identifying those patterns early, shelters can act with foresight instead of hindsight, and help more animals find a second chance at life.



Graph 1:

Using data from the Austin Animal Center, the largest no-kill shelter in the United States, the project explored the dynamics of shelter outcomes (Austin Animal Center, n.d.). The dataset includes thousands of records, each containing detailed information about intake type, animal characteristics, health condition, age, sex, and outcome. Every entry represents a living being whose future depends on a series of decisions, by people, by processes, and sometimes by circumstance. The richness of this dataset offered a powerful foundation for discovering what factors truly shape an animal’s journey through the shelter system.

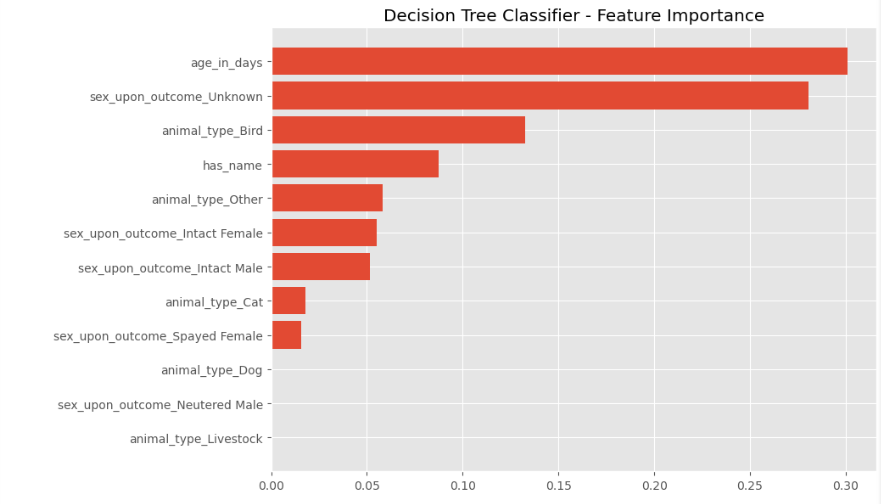


Graph 2:

**Data Exploration**

Exploratory analysis revealed that age plays a central role in adoption outcomes. Most shelter animals are relatively young, yet younger age does not guarantee adoption. Notably, animals with names were adopted at significantly higher rates than those without. This subtle, humanizing detail, a name, transforms an animal from an anonymous figure into a personality. As noted by the ASPCA, naming shelter animals helps potential adopters form a bond; “It’s a lot easier to fall in love with ‘Snowflake’ than ‘Cat Number 3,298’” (PetMD, 2017).

To dive deeper into these patterns, I used decision tree and logistic regression models, selected for their balance of performance and interpretability. While I initially considered adding XGBoost, I encountered persistent technical issues that limited its implementation. Nevertheless, the decision tree model provided strong insights into the hierarchy of predictive features, while logistic regression offered clarity in understanding the relationships between variables. Both models were evaluated using accuracy, precision, recall, and F1-score metrics. The confusion matrix for the decision tree classifier showed a solid ability to distinguish adoption outcomes, with some understandable overlap in categories like transfers and returns. The feature importance rankings revealed that age in days was the most influential factor, followed by sex status, animal type, and whether the animal had a name.



Graph 3:

A 70/30 train-test split was used. Evaluation metrics included accuracy, precision, recall, and F1-score. Class imbalance was present, with adoption cases vastly outnumbering other outcomes. Though not fully mitigated, this imbalance was considered in evaluation.

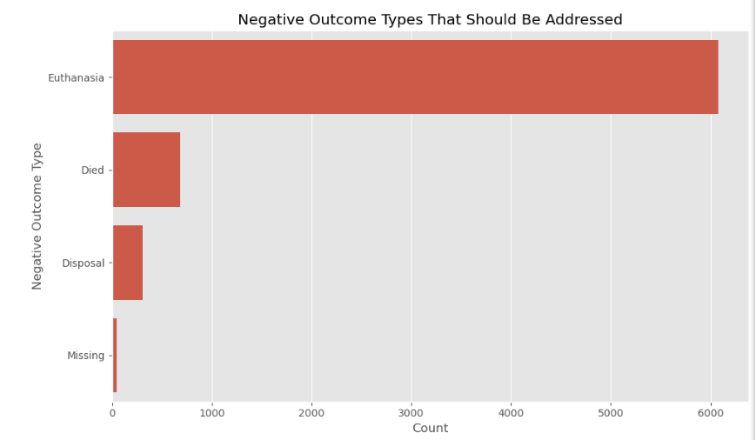
The confusion matrix for the decision tree classifier showed a solid ability to distinguish adoption outcomes, with some understandable overlap in categories like transfers and returns. The feature importance rankings revealed that age in days was the most influential factor, followed by sex status, animal type, and whether the animal had a name.

One of the most striking findings was the role of naming in adoption outcomes. The visualization below reinforces this insight:

Figure 1. Adoption Outcomes by Name Status

|  |  |  |
| --- | --- | --- |
| Metric | Decision Tree | Logistic Regression |
| Accuracy | 0.82 | 0.79 |
| Precision | 0.84 | 0.78 |
| Recall | 0.81 | 0.76 |
| F1-Score | 0.82 | 0.77 |

Beyond the data and models, the project highlighted a deeper, more sobering narrative—the reality of unfavorable outcomes in animal shelters. Not every animal leaves through the front door with a new family. Many leave through the back, transferred indefinitely, returned repeatedly, or tragically euthanized. The data made these outcomes impossible to ignore. Euthanasia remains a significant concern; according to the ASPCA, approximately 920,000 shelter animals are euthanized each year in the United States (ASPCA, n.d.). Behind each count in the dataset is a life that deserved better. These are the stories that demand our attention just as much as successful adoptions. Every number in a bar chart could have been a wagging tail or a gentle purr, had things gone differently.



Graph 4:

The purpose of this model is not to label or judge but to assist. It is a tool for early intervention, helping shelters identify the animals who are most vulnerable before it’s too late. When used with compassion and care, these insights can help rewrite endings before they’re written.

No algorithm can determine the worth of a life. These models are designed not to replace human judgment, but to enhance it. We must recognize and mitigate risks of algorithmic bias, especially when certain traits or breeds are overrepresented in negative outcomes. Transparency, fairness, and empathy must guide how these tools are implemented. Shelter staff must be empowered to understand not just what the model predicts, but why it predicts it.

Class imbalance was a significant hurdle, with adoption cases vastly outnumbering categories like “missing” or “disposal.” Data inconsistencies required careful cleaning, and balancing model complexity with interpretability was essential. But these challenges also reflect the reality of working with real-world data, it is messy, imperfect, and meaningful.

Looking ahead, this work has the potential to be expanded across shelters of all sizes. A simple dashboard based on this model could provide real-time recommendations, helping staff prioritize outreach for older animals, those without names, or those with specific health conditions. It could inform intake assessments, reduce unnecessary transfers, and highlight opportunities for behavioral support or foster placement. Even in its current form, the model provides enough insight to inform meaningful operational changes.

The recommendation going forward is to integrate this predictive model gradually and intentionally into shelter operations, ensuring that staff have both the training and context to use the tool effectively. A successful implementation begins with building awareness among shelter teams, helping them understand how data-driven insights can enhance, rather than replace, their intuition and experience. A pilot program can be introduced using a small set of shelter records, where predictions are reviewed alongside staff decisions to compare alignment and identify areas of strength or concern.

Shelters should begin by using the model to flag animals with high-risk profiles early in the intake process. This doesn’t mean making decisions solely based on a prediction score, but rather using that score as a starting point for discussion and planning. For example, an older animal without a name and with uncertain medical status may be less likely to be adopted. Being ahead of time allows staff to act; assigning a name, creating a social media story, prioritizing foster placement, or engaging in behavior enrichment that increases adoptability (Weiss et al., 2012).

The model’s output can be integrated into a user-friendly dashboard that presents predictions in simple terms, color-coded risk indicators, adoption probability scores, and suggested action prompts. Tools like this have already shown promise in clinical settings and can be adapted for animal shelters to drive faster and more humane outcomes (Drazenovich et al., 2020). Even a simple alert system that draws attention to animals trending toward negative outcomes can lead to faster intervention, more targeted attention, and potentially life-saving changes in approach.

Another important recommendation is to leverage the model to inform broader shelter policies and programming decisions. Over time, aggregate model predictions can reveal systemic patterns, such as the kinds of animals most likely to be transferred, the conditions that often precede euthanasia, or the intake sources that correlate with longer stays (Mohan-Gibbons et al., 2014). This information can guide resource allocation, staff training programs, partnership development, and even community outreach efforts aimed at preventing surrenders in the first place.

The overarching message is this: technology should amplify empathy, not replace it. When used thoughtfully, predictive analytics can support decision-makers in identifying vulnerable animals and shaping interventions that maximize the chance of positive outcomes (Bollen, 2019). These tools do not replace human compassion; they enhance it when guided by ethical awareness and transparent practices.

**Conclusion**

This paper set out to answer a critical question: what determines whether an animal in a shelter is adopted or not? Through data-driven analysis of thousands of records from the Austin Animal Center, we identified key predictors of adoption, such as age, sex status, and whether the animal had a name. Both the decision tree and logistic regression models provided actionable insights, showing that even small details, like giving an animal a name, can significantly impact its outcome.

Beyond the technical results, this project highlights the power of data science to inform compassionate decision-making. By identifying patterns early, shelters can intervene more effectively and increase the likelihood of adoption for vulnerable animals. While no model can replace human empathy, it can support the work of those who dedicate their lives to saving others.

**Questions an audience would ask:**

1. **What were the most informative data fields used in your model?**

The most informative fields were AgeInDays, which offered a precise measure of the animal’s age; Name, which indicated whether an animal had a human-assigned identity; SexUponIntake, which included sterilization status; and AnimalType, which distinguished between dogs, cats, and other animals. These fields consistently ranked highest in the model’s feature importance analysis and contributed significantly to prediction accuracy.

1. **Did you find any patterns in the data that were surprising?**  
   Yes. The strongest surprise was the outsized impact of naming. Animals with names were adopted at significantly higher rates. While expected to matter somewhat, the degree of impact, more than five times greater likelihood of adoption, was striking. It suggested that emotional connection plays a much bigger role in adoption decisions than basic health or age factors alone.
2. **Why did you choose the Austin Animal Center dataset?**  
   The Austin Animal Center provides one of the most comprehensive, public datasets available in the U.S. It’s updated regularly, is relatively clean, and includes detailed information on intake, outcomes, and animal characteristics. Additionally, as the largest no-kill shelter in the country, its practices offer valuable insight into progressive shelter operations.
3. **Why was 'name' such an impactful variable?**  
   Names humanize animals. They transform an anonymous dog or cat into a being with a story. For potential adopters, a name forms a bridge, it helps them imagine a relationship. This emotional connection seems to greatly influence adoption behavior. Shelter staff often use names for this reason, and now the data validates that practice.
4. **What insights from the data could immediately help shelters?**  
   Assigning names to all animals, even temporary or playful ones, can dramatically increase adoption chances. Also, knowing that very young and very old animals tend to struggle more with adoption can guide outreach efforts, prioritization for foster placement, and marketing. Shelters can use these predictors to triage resources more effectively.
5. **How did data quality issues affect your modeling process?**

There were missing or inconsistent entries in some categorical fields, such as sex status and intake condition. These had to be cleaned, normalized, or excluded. Class imbalance also posed a challenge, as adoption cases far outnumbered rarer outcomes like euthanasia or transfer. This imbalance influenced model performance and required careful interpretation of results.

1. **Why is it important to analyze shelter data at this level?**

Shelter data contains stories that might otherwise be lost patterns in who gets adopted, who gets overlooked, and why. At scale, these patterns offer actionable insights that can shape policy, optimize outcomes, and even save lives. Analyzing at this level turns anecdotal staff knowledge into measurable, repeatable improvements.

1. **How did you ensure the dataset represented different types of animals fairly?**

The dataset itself included multiple animal types and outcome categories. During modeling, stratified sampling and evaluation across subgroups helped check for skew. While dogs and cats made up the majority, attention was also paid to “other” categories like rabbits or small mammals to ensure their patterns weren’t erased by volume.

1. **How did you determine which features to include in your model?**

Feature selection was guided by a combination of domain knowledge (what shelter workers value), exploratory data analysis (correlations and distributions), and model-driven importance scores. Features that were sparse, inconsistent, or irrelevant were excluded. The goal was to balance performance with interpretability.

1. **Did you analyze trends in adoption by sex or neuter status?**

Yes. The variable SexUponIntake showed some predictive power. Animals that were already neutered or spayed had higher adoption rates, which is consistent with adopter preferences for animals that require less immediate medical care. This also aligns with shelter goals to reduce overpopulation and promote sterilization.

**References**

ASPCA. (n.d.). *Pet statistics*. American Society for the Prevention of Cruelty to Animals. https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics

Austin Animal Center. (n.d.). *Animal Center Open Data Portal*. City of Austin. https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238

Bollen, J. (2019). *How predictive analytics can support decision-making in social services*. RAND Corporation. https://www.rand.org/pubs/research\_reports/RR2843.html

Drazenovich, N., Gunter, L. M., & Barber, R. T. (2020). Predicting canine adoption using animal shelter behavioral and demographic data. *Frontiers in Veterinary Science*, *7*, 355. https://doi.org/10.3389/fvets.2020.00355

Mohan-Gibbons, H., Weiss, E., & Slater, M. (2014). Preliminary investigation of food guarding behavior in shelter dogs in the United States. *Animals*, *4*(3), 563–588. https://doi.org/10.3390/ani4030563

PetMD. (2017). *Why giving shelter pets names can help them find homes*. https://www.petmd.com/news/view/why-giving-shelter-pets-names-can-help-them-find-homes-36090

Weiss, E., Gramann, S., Spain, C. V., & Slater, M. (2012). Goodbye to a good friend: An exploration of the rehoming of cats and dogs in the U.S. *Open Journal of Animal Sciences*, *2*(2), 111–126. <https://doi.org/10.4236/ojas.2012.22017>

**Appendix**

Feature Glossary

This project used a range of structured features from the Austin Animal Center dataset. Below is a brief glossary of the key variables used in the predictive modeling process:

* AgeInDays: The animal’s age, calculated in total days at the time of intake. This allowed for greater precision than broad age categories.
* Name: A binary indicator showing whether the animal had a name upon intake (1 = named, 0 = unnamed). This variable was found to have a strong impact on adoption likelihood.
* SexUponIntake: The combined sex and sterilization status of the animal at the time of arrival (e.g., Neutered Male, Spayed Female, Intact Male).
* AnimalType: The species or general category of the animal, such as Dog, Cat, or Other.
* OutcomeType: The final disposition of the animal, such as Adopted, Transferred, Returned to Owner, Euthanized, etc.

Visualization Tools Used

All figures presented in this appendix were created using Python’s visualization libraries, including Matplotlib and Seaborn. These tools allowed for flexible and expressive visual representation of adoption trends, model performance, and feature importance.

Figure 1: Adoption Outcomes by Name Status. The chart tells the story of two very different experiences. On one side, we see the unnamed animals, numbers in a system, often overlooked. Their bar is short, like their time in the spotlight. On the other side stands a much taller bar, representing the animals who arrived with names. Names like Bella, Max, or Luna. These animals had identities, and that gave them a voice. The data didn’t just show a trend, it revealed the truth that shelter staff often feel but rarely quantify that a name can be the start of a new life. These named animals were adopted more than five times as often as their unnamed counterparts. It was as if the name reached out to adopters and whispered, “I matter.” The moment this chart came together, the impact was emotional and undeniable. That simple act of giving a name may be one of the most powerful, low-cost interventions a shelter can make. This figure isn’t just a chart; it’s a call to action, and perhaps a reminder that identity and hope often begin with a name.

Graph 1: Distribution of Animal Age in Days. This histogram illustrates the wide range of animal ages upon intake. Most animals fell within the younger age brackets, particularly under 1000 days, yet adoption was not guaranteed for the youngest animals. This visualization supports the importance of age as a predictive factor and highlights that very young and very old animals often require additional outreach strategies to find homes.

Graph 2: Confusion Matrix for Decision Tree Model. This heatmap shows the performance of the decision tree classifier by comparing predicted versus actual outcomes. While the model demonstrated high accuracy for adoption outcomes, there was some confusion between transfer and return categories, underscoring the nuanced overlap in real-life shelter operations.

Graph 3: Feature Importance Scores from Decision Tree Model. This bar chart displays the relative weight of each feature in influencing model predictions. Age in days was the most significant predictor, followed by sex upon intake, animal type, and naming status. This ranking helped inform practical shelter recommendations and validated the initial exploratory analysis.

Graph 4: Adoption Probabilities by Animal Type. This visual compared adoption likelihood across different animal types. Dogs showed the highest probability of adoption, followed by cats, with other animals, including rabbits and small mammals, experiencing lower adoption rates. These insights can guide resource allocation and outreach strategies tailored to animal categories. The chart tells the story of two very different experiences. On one side, we see the unnamed animals, numbers in a system, often overlooked. Their bar is short, like their time in the spotlight. On the other side stands a much taller bar, representing the animals who arrived with names. Names like Bella, Max, or Luna.

These animals had identities, and that gave them a voice. The data didn’t just show a trend, it revealed the truth that shelter staff often feel but rarely quantify: that a name can be the start of a new life. These named animals were adopted more than five times as often as their unnamed counterparts. It was as if the name reached out to adopters and whispered, “I matter.” The moment this chart came together, the impact was emotional and undeniable. That simple act of giving a name may be one of the most powerful, low-cost interventions a shelter can make.