

ALLEGHENY COLLEGE
DEPARTMENT OF COMPUTER SCIENCE

Senior Thesis

Developing Data Standardization and Analysis Strategies for Animal Shelters

by

Mariah Gammon

ALLEGHENY COLLEGE

DEPARTMENT OF COMPUTER AND
INFORMATION SCIENCE

Project Supervisor: **Professor Kapfhammer**
Co-Supervisor: **Professor Bonham-Carter**

3 May 2024

Abstract

This thesis provides an in-depth discussion on and the production of an interactive tool focusing on the benefits of data driven analysis, specifically within animal shelters. There isn't currently a standardized data collection methodology in the shelter sector. Therefore, nonprofit organizations, such as animal shelters, typically lack data based decision making. I am providing a well-defined data analysis system that can be used to take raw data and put it into a form that is easily understood and can be leveraged by users to make decisions related to their marketing practices. This project is designed to help increase animal adoption rates, especially for under-adopted breeds and species. It creates a standardized industry-wide data collection methodology to be used in marketing and promotion by the shelters — especially addressing their challenges with breeds and species at particular risk. This project results in the development of a two part project — a standardized data collection form and a dashboard reactive to file input.

Table of Contents

Introduction	4
Motivation	4
Overpopulation	6
Data/Lack Thereof	8
Sense of Urgency	9
Problem Stated and Research Questions	9
 Current State of the Art	 11
Methodology	11
Goals of the Project	13
Ethical Implications	13
Thesis Structure	15
 Related work	 16
Using Data in Animal Shelters	16
Overcrowding	18
Rightsizing & Capacity Limits	19
Targeted Marketing Strategies	21
Pet Selection Preferences and Marketing Strategies	24
 Method of approach	 28
Study Design	30
 Experiments	 36
Experimental Design	36
 Threats to Validity	 45
Validity Checks	45

Data Collection Form	45
Dashboard	47
Conclusion	50
Future Work	50
Future Ethical Implications and Recommendations	52
Conclusions	52
References	54

List of Figures

1	Shelter Data Ex 1	5
2	Shelter Data Ex 2	6
3	Research Development	28
4	Collection Form	37
5	Month Comparison	39
6	Domestic Comparison	40
7	Months Visualization	41
8	Dashboard	42

Introduction

Motivation

Less than half of the number of animals that shelters (in the United States) each year are adopted out, about 6.5 million entering with only 3.2 million adopted [9]. About 1.5 million animals in shelters are euthanized [9]. Millions of animals per year are killed due to overpopulation in shelters [9]. Cats that enter shelters have a 45% chance of eventually being euthanized, even though 99% of animals in shelters are healthy enough to be adopted [9].

Animal shelters are (typically) non-profit businesses that constantly deal with understaffing and overcrowding issues. What if there was a way to help these shelters utilize consolidated data to make connections and better understand how to care for and adopt out their animals?

This project is designed to do that, focusing on creating a means for standardized data collection and data-driven analysis, resulting in improved decision making for animal shelters. My research has been conducted researching relevant and related literature and data, using publicly available data sources, as well as through communication directly with individual animal shelters, gaining insight and data from them. The main source of publicly available data was gathered from the Austin Animal Center, publicly available on their website. They contain intake and outtake data representing various data values from October 2013 to the present. Additional publicly available data can be found in the Shelter Animal Counts dataset. Shelter Animal Counts has free tools such as a state-level dataset, annual data from various organizations reporting intake and outtake data, and industry wide reports. They also have a de-identified organization-level dataset that is available for purchase if needed. Local shelters gave me some examples of the ways in which they collect data. Below are some examples of data that they use.

Director's Board Report - February 2024						
Animal Activity for the Month of January 2024						
Animal Type	In	Adopted	Reclaimed	Transferred	Animals in Care	Notes
Cats	56	56	1	7	0	
Cats in from Foster	23	0	0	0	0	9 Cats Transferred Into Foster
Cats in from Other Rescues	4		0	0	0	4 Transferred In from Love4Satos
Dogs	37	15	9	13	0	
Dogs in from Foster	1	0	0	0	0	5 Dogs Transferred Into Foster
Dogs repeat strays	0	0	0	0	0	
Dogs in from other shelter	3	0	0	0	0	3 dogs in from Love4Satos
Ferrets	1	2	1	1	0	
Guinea Pigs	4	3	0	0	0	
Chinchillas	0	0	0	0	0	
Rabbits	7	0	2	6	0	
Chickens	0	0	0	0	0	
Fish	4	0	0	0	0	
Birds	5	1	1	2	0	
Reptiles	2	0	0	1	0	
Snake	0	0	0	0	0	
Rats	0	0	0	0	0	
Gerbils	0	0	0	0	0	
Totals	147	77	14	30	99	"Record Breaking" Volume of Animals Brought Into the Shelter In One Month
Note: As of February 1, 2024 there are 99 Animals in Care						

Figure 1: Shelter Data Ex 1

Admissions & Transfer breakdown of numbers						
	Dogs	Cats	Rabbits	Other	Monthly Total	YTD Total
Owner Surrender	6	55	2	-	63	503
Return	1	-	-	-	1	31
Stray	1	6	-	-	7	147
Service In	16	1	-	-	17	217
Transfer	10	1	-	-	11	75
Born in Care	-	-	-	-	-	12
Emergency Board	-	-	-	-	-	7
Monthly Totals	34	63	2	-		
YTD Totals	284	662	43	3		

Total number of animals admitted in September 2023: 99

Total number of animals admitted year-to-date 2023: 992

Admissions Applications received		
	September	YTD
Dogs	54	460
Cats	136	887
Rabbits	9	102
Home-to-Home	5	85
Total	204	1534

**The numbers above reflect just the total number of applications received. Many of the applications include multiple animals per application.*

Figure 2: Shelter Data Ex 2

Relevant related research consists of work that discusses a wide range of relevant topics in the shelter sector, such as using data in animal shelters, overcrowding, rightsizing & capacity limits, and pet selection preferences and marketing strategies. The related works and methodology sections will delve more deeply into the relevant research.

Overpopulation

Many animal shelters are faced with capacity and overpopulation concerns. An animal shelter being overcrowded can lead to multiple challenges and harmful effects. Negative results of overpopulation include diseases spreading quickly, decreased focus on the individual wellbeing of each animal, and challenges with animal behavior and health adjustments. These concerns can

contribute to increased costs for the shelters and animal deaths, including through euthanization. These factors also result in decreased adoption rates and happy and healthy animals. Behavior fluctuations might include increased aggressiveness, increased stress, fear and anxiety, destructive actions as well as decreased interest in playing, eating, and grooming. All of these traits are dangerous for the mental and physical well being of the animals and possibly decrease their chances of being adopted.

When animal shelters become overcrowded, they have to resort to turning animals away and/or increased euthanization depending on their policies. There are two main types of animal shelters, no-kill vs. a more traditional one.

No-Kill Shelters [4]:

- Do not euthanize healthy or treatable animals, only resorting to euthanization if animals are terminally ill or if their behavior is too dangerous
- Privately run
- Accept animals on a conditional basis
- Have to turn away animals when they are at capacity or overcrowded
- Deal with animals ending up in cages and confined for very long periods of time when accepting animals that struggle to be adopted

Traditional Shelters [4]:

- Accept all animals, regardless of the state they are in, either by choice or because they are governmentally owned and required
- Become severely overcrowded or euthanize animals that did not necessarily need to be for other reasons
- Receive many animals turned away from no-kill shelters that aren't left on the streets
- End up euthanizing animals when dealing with animals that struggle to be adopted and are there for too long

My aim is to help both types of shelters because all animal shelters deal with challenges involving overcrowding and animal length of stay. The fact that less than 50% of the number of animals that go into shelters yearly are adopted from shelters is heartbreaking [1]. Around 3.2 million cats enter shelters and only 1.3 million are adopted each year [1]. In addition, it's estimated that shelters on average have to turn away 25-60 dogs per month due to lack of space, with some shelters having to turn away up to 200 dogs per month [1].

Data/Lack Thereof

Collecting data is highly imperative for shelters to be able to perform analyses and view insights into how their shelter works to serve their animals. Data can be used to improve shelter operations and measure current success rates as well as areas for improvement [8]. Quality data collection is important, but the animal shelter sector currently lacks an overall standardized aspect of collection [8]. Standardized analysis is difficult to achieve if standardized collection does not exist. Research discusses how important it is for shelters to have proper data collection and analysis, but roadblocks exist that lead to their insufficiency, such as limited training and resources [8].

Additionally, it seems that this lack of resources hinders animal shelters' abilities to help them make informed decisions. [8] From the research that I've done and all of the animal shelters that I've spoken to, while they may keep and collect intake and outtake data, they are fully not utilizing or analyzing it, if at all. When I proposed my idea to them, they all seemed to agree that it would be useful. They tend to have data to varying degrees revolving around intake and outtake and characteristics of animals, but they don't actually do anything with that data. When making marketing decisions, especially concerning social media posts and which animals they promote and when, they said they do not have any particular strategy, just that they post animals and hope for the best, basically. They either do not consider a particular strategy, or they do not have the resources and time to devote to it. Some examples of feedback from animal shelters that I've spoken to are:

"Our marketing director doesn't necessarily have a strategy in writing; so I'm not able to provide much info in regards to that." - Homeward Animal Shelter

"The Humane Rescue Alliance, like most shelters nationwide, is dealing with a surge in canine influenza as well as an exploding animal population. The latter is a direct result of issues pertaining to the pandemic." "our staff is working 110 percent to deal with the aforementioned issues" - Humane Rescue Alliance

Animal shelters and literary sources acknowledge the importance of data collection and analysis to aid in shelter operations, decision making, and animal outcomes, but obstacles are present that contribute to their lack of ability to utilize data effectively. Due to many individual animal shelters struggling already with limited resources, time, and technical abilities, incorporating standardized practices involving data use can be difficult, which

leads to data driven decision making to go unused. Resources in this field are greatly needed, and these challenges are addressed in this thesis project in the development of user friendly collection and analysis tools.

Sense of Urgency

Animal shelters also face the challenge of having a sense of urgency [8]. As mentioned previously, overpopulation is a high concern, so being able to decide which animals to target first to market or be accepted into shelters would be helpful for improving that amount of open space [3]. When saying “accept into shelters,” I mean that these no-kill shelters that have the ability to turn away animals can take their data into account and decide which animals to accept/turn away based on which animals they know they can adopt out effectively.

Problem Stated and Research Questions

This thesis provides an in-depth discussion on, and the proposal for the production of an interactive document focusing on the benefits of data driven analysis, specifically within animal shelters. There isn’t currently a standardized data collection methodology in the shelter sector. Therefore, nonprofit organizations, such as animal shelters, typically lack data based decision making. I am providing a well-defined data analysis system that can be used to take raw data and put it into a form that is easily understood and can be leveraged by users to make decisions related to their marketing practices. This project is designed to help increase animal adoption rates, especially for under-adopted breeds and species. Its aim is to create a standardized industry-wide data collection methodology to be used in marketing and promotion by the shelters - especially addressing their challenges with breeds and species at particular risk. Research questions that can be asked to validate the project’s goals are: Can a tool be developed that helps the most at risk animals be identified and presented toward a higher likelihood of being adopted? Does my tool, which consists of a collection form and dashboard, incorporate enough data to answer questions that are important to make marketing decisions. Questions such as: Does it output relevant data? Can it identify at risk breeds, ages, etc. based on factors such as average length

of stay? Can it identify trends in time (such as months) for intake and outcome? Can this data be used, coinciding with research, to determine animals in need of optimal spacing and promotion in the shelters? Is this tool able to be utilized by an animal shelter that doesn't know much about computer science?

Many reasons exist as to why data is important for animal shelters, such as transparency, evaluation of progress, and the ability to make informed decisions and plans. Data allows for animal shelters to be transparent with themselves and with the general public and potential adopters. Seeing data allows them to identify where they are succeeding and where different strategies might be necessary, as well as pointing out when and where those strategies can be implemented.

Current State of the Art

Methodology

To develop research ideas, I used primary and secondary sources.

I reviewed scholarly articles to gain an understanding of the current research, methodologies, and best practices in animal shelter data collection and usage. I also gathered information revolving around adoption rates and animal welfare in animal shelters, marketing and adoption strategies, and data analysis.

I found publicly available data and resources to use and show standards for analysis of trends and useful data values to include when collecting data. Publicly available data is useful for setting templates, increasing public knowledge, and comparing trends.

I reached out to multiple individual shelters, hearing back from some about their insights and intake, outtake, and marketing data. I listened to their feedback and opinions. Most said that they do not have a current marketing strategy, but they do collect intake and outtake data to some degree. However, as my research also showed, there is no current standardized collection method that is recommended as best practice, and this was proven further by speaking with shelters directly.

I began by developing an interactive Quarto document in RStudio, incorporating graphs and Shiny applications. Originally, the project consisted of more of an insight into public data trends and a step by step process on how to develop and run the code as an animal shelter.

I have since developed the project further to include a data collection form through a Shiny app and a Shiny Dashboard to make the project much more user friendly and reusable for animal shelter staff that are not familiar with computer science themselves and want an efficient and easy to understand way to collect and analyze their data.

I designed and created a structured data collection form that helps to gather information that is relevant and helpful to animal shelters and their analysis. Based on my research and conversations, I have identified key points of data and standardized field names. I then developed a Shiny application that works to create a form for users to input these key points of data to create a CSV (comma-separated values) file that can then be used for further analysis and review. A CSV file is a text file that saves data in a table format. CSV files are compatible, effective files that are popularly used. This form allows for standardization and consistency.

I then used R programming to create a Shiny application that works as an interactive platform for analyzing and exploring data revolving around an animal shelter with visualizations that can offer insights into key factors. R and Shiny are helpful tools because they are easily interactive and flexible. I used reactive programming to allow for updates based on user input, so users are able to upload their standardized csv file to analyze their own data.

I evaluated the validity and effectiveness of this tool by testing it on representative samples from shelters and representative test data through utilizing the data collection form. This analysis and test process shows that the form and analysis is valid and usable. The tool is accompanied by documentation to explain what is required for use from an animal shelter and can be used with minimal technical knowledge.

This research works to contribute several aspects to the existing body of knowledge that is the animal shelter community and data management.

Creating a thoroughly structured and standardized data collection form and user friendly Shiny app addresses the gap that exists in data collection methodology in animal shelters currently. It gives shelters a way to uniformly gather and analyze consistent information that can enhance their ability to make data driven decisions.

Utilizing R and Shiny to provide interactive technology is a useful strategy to aid in analysis, making data collection and analysis tools more accessible and usable for animal shelter staff that have limited technical knowledge. Data driven decision making can be made easier and more accessible for shelters this way, creating a more streamlined and standardized approach for shelters. They can examine their challenges, successes, and opportunities.

Providing documentation on how this tool works and is effective, as well as explaining what data values are required for use, allows for shelter workers to both have the knowledge to use the tool and increase their awareness of the importance of data driven analysis. Animal shelters lack standardized best

practice methods when it comes to data collection and analysis, so this tool helps to increase awareness, knowledge, and support resources. Shelter staff can be more secure in their knowledge of key data to track. This research can help staff to recognize the feasibility and usefulness of utilizing resources such as R and Shiny apps that they may not have the time and resources to investigate themselves, as research shows how understaffed and busy shelters typically are.

This work adds value to the animal shelter field by addressing gaps and developing effective tools and strategies to impact the way that animal shelters work and make decisions. Implementing tools and documentation that are user friendly and easy to follow helps animal shelters improve their ability to use and analyze data, which can help them further to enhance their decision making strategies.

Goals of the Project

The key goals of this project are to create standardized data collection and analysis tools using RStudio and Shiny apps that animal shelters can utilize to review data on factors such as average length of stay, adoption trends, and intake and outcome timelines. Additionally, this project works to give shelters the resources and ability to make informed data driven decisions regarding marketing and operations strategies. They can use these tools to help identify animals that need promotion and increased strategies to improve animal well being and adoption rates.

Overall, increased knowledge of the benefits of collection and analysis of data can be provided for animal shelters and the general public. Providing evidence and transparency is always important to help people understand.

Ethical Implications

Various ethical factors must be considered when handling this project. Dealing with animal shelters, which are typically local, small, nonprofit businesses, as well as discussing animal lives, are both tricky subjects that must be dealt with carefully.

To start, information privacy and data collection is a concern. In some cases, animal shelters can be hesitant to share their data, at least in a public

way, as to protect themselves and their privacy. In at least one case, I had a shelter reach out and ask their name not to be disclosed and their data not publicly shown, although it could be shown to those overseeing my work for the sole sake of maintaining their privacy. Working directly with people, especially small, locally owned businesses, can always be a challenge. Local animal shelters like this one want to ensure that their data is protected and are skeptical to use it publicly. Even though they are willing to share data and communicate with me through this project, and they have expressed desire to see how this work could help them, they are still cautious and would like to remain virtually anonymous. While factors like this one can be challenging, ensuring that the people I communicated with throughout this project were respected and their information was kept confidential when desired is extremely important.

I also practiced thorough and open communication concerning the goals and purpose of my project. Not only is it ethically responsible, it also helps build trust with the animal shelter community. This factor is easily mitigated by keeping shelters anonymous that request to be, referring to them as “local animal shelters,” and only publicly sharing information that is allowed. Handling data with care, confidentiality, and privacy throughout the project has been ensured. Additionally, publicly available data can be utilized for the purpose of the secondary audience previously described, as a public source of knowledge and learning. As for the project’s purpose to be used by animal shelters to plug in their data, that work would be done internally, with shelters having access to the code they need and able to keep that information to themselves to be used as they desire, restricting access to authorized personnel only.

Data collection would be another concern when dealing with small, local businesses because shelters might usually be in charge of manually updating their own data, which can always lead to errors and misinformation. Animal shelters should ensure that their data is accurate when using this tool.

Another ethical factor that is of concern is unintended consequences. When handling a project that involves using data to view trends in animal adoption, the question that can arise is if animal shelters begin focusing their efforts on different types of animals, could unintended consequences result that could lead to other animals being mistreated? This research aims to aid the overall standardization process of data collection and analysis in animal shelters, allowing for better use and handling of data. This research and work is to be utilized to a) help with public awareness and shelter knowledge

and promote overall increases in adoption rates from and donation rates towards animal shelters and b) help animal shelters use their data to help with promotion strategies and give attention equitably to their animals. This research helps to increase equitable and standardizes access to and use of data collection and analysis for animal shelters.

Thesis Structure

This thesis is structured into multiple sections that guide the reader through an in depth overview of this project.

Introduction Section:

Provides a high level overview of the purpose and development of this project and its key contributions

Related Work's Section:

Provides a scholarly review of related literature that aid in the necessary background and development of this research and tools. This section reviews literature that involves animal shelter adoption and operations, data collection and analysis and why they are important in the animal shelter field, using interactive technology.

Approach/Methods Section:

Provides a more in depth description of the tools that were created and what they do. This section will explain the process of developing the research idea and tools. This section delves into the mechanics of how the interactive data collection form and dashboard work.

Results Section:

Provides evaluation of the tools with experiments. This section shows how the interactive platform was tested to ensure that it works properly and answers the research questions.

Conclusion and Future Work:

Provides an overview of final thoughts and future work that can be done to continue to advance the development of this tool and its goals.

Related work

Using Data in Animal Shelters

Data usage in animal shelters is important because data plays a pivotal role in improving the way that shelters operate, aiding in animal welfare and data driven decision making. Data can ensure that shelters keep track of the way that they work, such as intake and outcome information, animal length of stay, and distribution of resources. Data can even be used to see more in depth about animal welfare, health, and behavior. Data helps give insight into adoption trends, successes, progress, and barriers. Adoption strategies can be developed based on analysis of these factors to identify animals in need of care and promotion. Strategic planning and budgeting can be made in accordance with data insight.

Yadhunath discusses the practice he refers to as “Animal Shelter Analytics,” how it can help animal shelters operate, and his personal experiences with implementing it [8]. He discusses the development of animal shelters through time and how the establishment of “animal welfare” and the desire to care more for the animals came to be [8]. Animal shelters serve the purposes of both taking care of animals and controlling animal populations so as not to harm humans as well.

Yadhunath deliberates the importance of data use in animal shelters [8]. Data is one of the most important components of animal shelters. Intake and outtake collection, information about all of the animals, financial and cost data, adopter data, etc. are all valuable aspects of a shelter, and this data can be taken advantage of to continue to help the shelter function and perform well. This kind of data can be utilized to help save lives, increase awareness and adoption rates, prioritize urgency, evaluate costs and performance, and communicate with potential adopters or donors.

Yadhunath discusses how data driven practices contribute to animal shel-

ters and their importance [8]. Data is quantifiable and can allow for animal shelters to examine their progress, both their successes and areas for improvement, and set goals that can be evaluated. It also allows for validity and transparency, which is especially useful for a small, local, nonprofit organization that relies heavily on donations from the community [8]. Data also can help to make informed decision making and future plans [8].

Data can help to acknowledge not only where improvements and differing strategies might be necessary, but it can also help to see where they can be implemented, what the root of the problem is, and overall better insight. If influxes of animals are coming and going with no evaluation of the data surrounding it, shelters have no basis to go off of when caring for and advertising animals and their strategies will not be well informed or personalized. However, if they can recognize that x characteristic of cats have the longest length of stay or x amount of this breed or this age have higher adoption rates or x amount of cats are taken in during these months, they can examine this data and try to make decisions and strategies that can hone in on issues or take advantage of promotion. Yadhunath describes the various stakeholders involved in and the type of data collected by shelters before delving into how Animal Shelter Analytics can be utilized proficiently [8]. Data can provide key insights into how a shelter works and how well it performs, specific animals and characteristics, geographic trends, and improve public relations and community.

Yadhunath also discusses the challenges that Animal Shelter Analytics faces, which include inconsistent or poor quality data, varying or lack of software, and hesitancy to change, especially with small, overworked businesses [8]. Yadhunath also supplies suggestions to help mitigate these challenges, such as specific guidelines for data collection and developing and teaching a data driven atmosphere [8]. Yadhunath describes his personal experiences with the practice that he refers to as “Animal She/Iter Analytics” [8]. He discusses why data collection and analysis is important in animal shelters and various ways in which it can be used to measure success and provide key insights that can help animal shelters plan and promote effectively [8]. He discusses how he worked as an intern with one shelter [8]. Yadhunath’s article provides good insight and background into the importance of data usage in shelter management [8]. My work provides ways in which these practices can be put into use. Yadhunath helps to prove the importance of shelters having a tool such as mine to help them utilize their data [8].

Overcrowding

Turner et al. discusses the complications associated with the amount of stray or unowned animals, specifically cats, that exist. The amount of stray cats is continuing to increase, especially during spring and summer, which is peak kitten season. Cat stray and euthanization rates are typically higher than dog numbers [6].

Turner et al. consider the implications of funding shortfalls and the costs associated with running a shelter and caring for animals [6]. With many shelters being small and locally run, relying heavily on donations or dealing frequently with fluctuations, keeping up with costs can be hard. Different times of year might have different amounts of available funds to work with. When funds are strained, animal care may also be strained, with less money for food, supplies, maintenance, staffing, and medical care. Turner et al. also debate the use of expired pharmaceutical supplies leftover from veterinary centers that are donated to them, as these animals deserve equitable care, possibly even higher quality.

Overcrowding provides significant risks to animal welfare. Animal welfare becomes endangered with overcrowding [6]. Stress, negative behavior, spread of disease (especially to vulnerable communities), and unsanitary conditions are all at risk of being increased [6]. Turner et al. highlight that feline respiratory disease complex is a particular concern. Turner et al. also point out the potential for increased exercise and community engagement to help with animal well-being and adoption rates [6]. Even just having toys around for cats or adopters to see can help with adoption rates.

Turner et al. make a good point that data driven approaches to address overcrowding and enhance shelter operations are important. Analyzing adoption trends, environmental factors, and community engagement is valuable to optimize the allocation of resources and promote animal wellbeing [6].

Turner et al. advocate for increased spread of awareness and public knowledge of the struggles of animal shelter management, animal well being concerns, and encouragement to adopt from shelters [6]. This project can help to add to this notion, being able to create tools that can increase standardized data collection and analysis that provides evidence as to why these things are important. Additionally, with shelters dealing with fund and animal adoption concerns, being able to utilize data to show trends can help in that category. For one, shelters can see the typical trends in funding fluctuations and plan accordingly in advance to optimize their use of funds. They can

also use the trends in adoption rates to see how and where their promotion and well being strategies work as well as how to plan their future promotion strategies. Animals costing money, having longer lengths of stay, or not getting along with other animals in communal spaces might be targeted for more urgent promotion. Evaluating effective promotion strategies is also important moving forward.

Rightsizing & Capacity Limits

Karsten et al., in their research “An observational study of the relationship between Capacity for Care as an animal shelter management model and cat health, adoption and death in three animal shelters”, explore the concept of “rightsizing” in animal shelters, developing a “Capacity for Care” management model [3]. They analyzed the capacity for animal housing and the effects that the amount of animals in a shelter have and explored the “optimal daily population” that “meets the health and welfare requirements of all animals, while maintaining or improving on the shelter’s goals for adoption or other positive outcomes” (called the C4C) [3].

Karsen et al. describe the harm from having too few or too many animals [3]. They determine that too many animals in a shelter could lead to increased spread of disease and medical care costs, decreased animal well-being and mental health, less optimal care from staff, fewer adoptions, and more euthanization [3]. Housing quality is another factor that Karsten et al. take into careful consideration, specifically with shelter cats, because when the quality of housing is poor, cats suffer, resulting in stress and anxiety, lack of eating, playing, and grooming, and higher aggressive tendencies [3]. These attributes are not only harmful to the well-being of the cats, but they could also lead to them being less likely to be adopted. The model that Karsten et al. describe the British Columbia Society for the Prevention of Cruelty to Animals established to examine Capacity control and rightsizing of shelters in 3 animal shelters [3]. They created the concept of pass-through housing for cats [3]. Their C4C model led to increased adoption rates (by 15%), lower average length of stay (from 40-22 days), and decreased estimated total of cats in isolation [3].

Karsten et al. describe how the Canadian Federation of Humane Societies (CFHS) developed a similar model that works to optimize the number of animals housed at one time, manage length of stay, and provide high quality

housing, aiming to improve animal well-being overall [3]. They determine that “Strategies to achieve these goals will vary by shelter but can include managing intake, utilizing daily population rounds, implementing changes to medical treatment and sanitation protocols, and altering the adoption process” [3]. The research and work done here are significant and show the importance of optimizing animal population and care in animal shelters. Animal shelters can utilize these strategies and my project to help influence these decisions.

Karsten et al. provide research and evidence as to why increased adoption rates and optimal capacity are important, and how finding ways to contribute to “rightsize” can improve the function of animal shelters overall [3]. Poor quality housing is detrimental to the animals being kept and can lead to declines in health, energy eating, play and increases in stress and fear and aggression, which are all factors that could lead to animals being less likely to be adopted and more likely to be euthanized [3]. Space guidelines exist for shelters to optimize space for cats, which can be provided either by lowering population to help with the existing housing available or by expanding their space. However, shelters often don’t have the space and funds to perform the latter suggestion, so finding ways to increase adoption rates and lower population of shelters is a necessary route to increase overall wellbeing.

Long length of stay, low adoption rates, and overpopulation are problems that animal shelters face, leading to lower quality of well-being for animals. Additionally, research and trends show that some characteristics of animals can lead to them being less likely to be adopted or have longer average lengths of stay or higher euthanization rates or higher costs to the animal shelter. Reasons for these trends could be attributed to various reasons, such as negative stereotypes, lack of public knowledge about the animals or the importance of adopting from shelters, or even poor public perception due to negative animal behavior in shelters (due to non optimal capacity and housing conditions) [3]. Regardless of the reason, these disparities exist and are proven to be detrimental to the functioning of shelters and the wellbeing of animals.

My project aligns with the notions that Karsten presents with the C4C model by developing a standardized data analysis system for animal shelters. By giving shelters a means to effectively collect and analyze data, this project can enhance driven decision making and optimize marketing strategies for promoting animal adoption. Karsten et al. emphasize how data driven strategies are important and my project aims to utilize data analysis

to point out trends and patterns related to adoption rates, breed and other characteristic-specific challenges, and length of stay. Animal shelters can analyze their successes and areas for improvement. They then can establish what characteristics and kinds of animals need more help with promotion. My tool aids in shelters being able to make actionable insights from the data provided. They can determine what advertising strategies are most and least effective to help spread awareness and equitable adoption rates for all animals.

The C4C model implementation stressed importance on the overall physical housing and capacity to maintain in animal shelters [3]. This project was very useful and insightful. My project aims to analyze specific trends and narrow down to the root of the problems. Both projects aim to optimize population in shelters and decrease average lengths of stay but with different strategies. The C4C research shows how looking at data and implementing strategies can have positive effects on the functioning of shelters and provide more opportunities for animal happiness and adoption [3].

Targeted Marketing Strategies

Karsten et al. also describe different advertising strategies that can help lead to specific animals being adopted, such as placing the animals you desire to more urgently be adopted directly into adoption/in public space as opposed to keeping them isolated or inaccessible to the public [3]. This information can be useful to my tools, as shelters can use the data and trends that they see to determine which animals they want to promote at a given time and ensure they are more likely to be seen. For example, my tools can be utilized to identify animals with longer lengths of stay and/or lower adoption rates and provide informed insight into developing targeted promotion strategies for these animals. Highlighting specific animals can increase their chances of getting adopted. Data tools allow animal shelters to make informed decisions about what animals are optimal to promote and how to promote them effectively, which can help to reduce overcrowding and improve housing, which in turn reduces average length of stay and spread of diseases like URI [3]. Additionally, limiting the choices for potential adopters helps to increase the likelihood and ease of making a decision to adopt.

Choosing to highlight and promote specific animals can actually help adopters make their decisions by simplifying the process. Karsten et al. also

provide evidence that helps to mitigate the ethical fears associated with the unintended consequences of increased euthanization to some animals [3]. They acknowledge that fears that lowering capacity (or highlighting specific animals) could lead to the increase of euthanization, but they point out that at each shelter they implemented their strategies in, the probability of shelter death decreased [3]. Higher adoption rates, lower average length of stay, and healthier populations, which are all directly related to improved strategies and lower capacities, are contributors to less deaths. Specifically, when length of stay is lower, more animals are able to be cared for over time with a lower daily population. Karsten et al. discuss how lower length of stay is a major contributing factor to positive results in animal shelters, which is what my project aims to help [3].

Much research discusses how important it is for shelters to have proper data collection and analysis, but roadblocks exist that lead to their insufficiency, such as limited training and resources.

Swoyer discusses a presentation done by Jill Dyche at the Pacific Northwest BI Summit, in which she presents on the importance of digitization in animal shelters [5]. Dyche communicated that shelters she has worked with don't tend to utilize the data that they have, leading them unable to hone in on trends or individual cases [5]. One shelter in particular has some of the highest euthanasia rates in the country, but without looking into that data, the numbers just show rates of animals leaving the system, not how they left [5]. If data isn't carefully examined and the adoption rates more clearly shown, key insights can be taken away. Lack of data usage hinders shelters' abilities to see individual cases, examine potential adopter data, or utilize predictive or social media analytics. Dyche shares a personal story of a dog who was put to sleep after being advertised on social media and receiving high levels of interest, but the communication of data was too slow to be of service [5]. Data use in shelters should be standard practice, and Dyche's examples show how data and social media should go hand in hand, as this communication and analysis can save lives. Dyche attempts to prove the value that using data to find trends and insights can have, discussing how different factors and characteristics of animals can affect their likelihood of being euthanized or their length of stay [5]. These arguments coincide with the arguments and insights that I make in my work and can show through visualizations that animal shelters can plug their data into. Dyche argues into how digitization can help animal shelters thrive both with animal well-being and adoption and with costs, allowing them to allocate resources more

efficiently [5]. Dyche thinks that by emphasizing digitization in shelters, funds can be better allocated and more money and time can be spent on promotion, advertising, and training for staff [5]. With my project and the insights that I provide, areas in which more funds are being spent, as well as where increased advertising can be implemented are shown and can be used by shelters.

In *Critical Problems for Research in Animal Sheltering, A Conceptual Analysis*, Horecka discusses the critical position of data collection in animal shelters and the difficulties they experience in utilizing it to enhance the way they operate, such as limited resources [?]. There are some systems in place to help recognize data management, but a widespread standardized system is still difficult to achieve. Horecka also stresses the importance of safeguarding and handling data with care, especially when sharing it. Best practices when it comes to data involving animal shelters are still shaky and there is a lack of consistent data strategies, which make it difficult for overall trends to be acknowledged. Defining and keeping tabs of KPIs can be difficult as well, especially with limited knowledge and resources.

By giving animal shelters user friendly tools like mine for data collection and analysis purposes, my project can help animal shelters surpass these difficulties they are facing and enhance their ability to make informed decisions and improve animal welfare and outcomes. These methods can overall improve the efficiency with which animal shelters run and make decisions.

In her book, *Strategies for Successful Animal Shelters*, Laura A. Reese discusses the many challenges that animal shelters face [?]. She states that maintaining records in standardized ways should be practiced and outlines some basic and essential records to keep track of: “a unique identifying number, description, microchip number if present, source of animal, dates of entry and departure, outcome, species, age, gender,” and some specific medical or behavioral information if needed [?]. She discusses the challenges that are involved in research and data practices within animal shelters and how increased focus on standardization of these aspects can result in the ability to identify best practices when it comes to the operations of animal shelters. My project coincides with this research by helping to aid in the standardization of the data collection process and provides data values that agree with Reese’s findings.

Pet Selection Preferences and Marketing Strategies

Weiss et al. describes a study in which an adopter survey was conducted to examine various factors behind pet adoption, such as why pets were chosen, the type of information provided, and the behavior of the animal and how it was viewed [7]. They express that this study can be used to help shelters market their animals, develop matches, utilize resources, and possibly increase adoptions [7]. Weiss et al. describe how characteristics of animals can play a role in their likelihood of adoption [7]. In some cases, personality, appearance, and temperament can be reasons behind adoption as well.

Weiss et al. describe a study in which cats that were perceived as friendly, happy, chill, and outgoing were preferred to cats seen as sad or fearful [7]. This data coincides with Karsten et al.'s that nonoptimal conditions in shelters can lead to detrimental effects on the behavior and level of fear in cats, which can lead to potential adopters not perceiving them as "good" pets [3]. If my data can be used to identify and target animals that need more urgent promotion and then utilize effective marketing strategies, better advertising and more adoptions can be put into motion. Implementing posts and promoting animals in ways that highlight their unique characteristics and personality traits, as well as giving potential adopters limited but personalized choices, can help the promotion of any given animal in a shelter. Allowing for adopters to connect with the personality of an animal can help.

Weiss et al. describe how certain behaviors can affect adoption likelihood and choices, such as active cats receiving more attention, animals that approach, and animals at the front of kennels/cages/shelters as opposed to animals farther back [7]. All of these are factors that can be manipulated according to the animals that shelters determine they want to promote most urgently based on the data they examine in the tools I created. For example, if a shelter analyzes that medium sized, Russian Blue, mid-aged cats, are facing challenges of longer length of stay or low adoption rates, they can alter these factors to ensure that these kinds of cats are more likely to be receiving attention by potential adopters.

As Karsten et al. determine, limited options and increased exposure can help to ensure that potential adopters are likely to adopt, so by examining their data and acting accordingly, animal shelters can help to promote animals that need it [3]. Weiss et al. also discuss the implications of the

information about animals provided to potential adopters [7]. The ways in which animals are talked about is important in the perception that potential adopters have. This research acknowledges that shelters should be cautious and thoughtful about the way that they promote animals that they want to [7]. Improving strategies and advertisements can help to improve adoptions, especially of animals that have characteristics and behaviors that potential adopters might have preconceived notions and stereotypes about (leading to lower adoption rates of certain breeds and types of animals). As Weiss et al. express, some characteristics of animals such as breed, color, and age can not be changed, factors exist that can be altered, such as behavior, information provided, context of viewing, and marketing strategies [7]. My work can help with these expressions because although some factors cannot be changed, trends can be acknowledged in certain situations and the types of animals that require more modification to the factors that can be changed can be noted and strategies can be implemented accordingly. Animals can be trained and promoted to help their chances of receiving attention and adoption.

Weiss et al. describe that this study was performed in both traditional/-municipal shelters and private, limited intake, no-kill shelters [7]. Adopters were asked to fill out a survey to determine various factors of pet selection after they chose their pet. Questions ranged from general, quantifiable, demographic and behavioral ones to open ended questions about the most important reason for adoption and what they did upon meeting them, to which categories were established and frequencies were counted [7]. Information coming from staff members seemed to be an important source for adopters, so animal shelters could ensure that staff are well versed about all of their animals, paying special attention to those most urgently needing adopted and characterizing their personalities in a unique, fun, yet accurate manner [7]. Direct interaction with the animals was another key aspect highlighted, and Weiss et al. discusses how less popular or more frequently passed over animals could benefit from this tactic [7]. Therefore, shelters can utilize my work to determine which animals are more frequently passed over and pay special attention to advertise them via interactions with the public and potential adopters, especially those with less noticeable/ “adoptable” behaviors and characteristics. Appearance can also be a factor that potential adopters take into heavy consideration, so in these cases Weiss et al. propose shelters highlighting behavior and personality traits highly to help find matches. These findings can also help shelters prioritize their resources.

Gouveia et al. discuss research done on the behavior of cats in a shelter, based on various factors including length of stay [2]. Long lengths of stay might contribute to negative behavioral effects on cats' activity, eating, and aggression [2]. The study that Gouveia et al. describes expresses the necessity of earlier adoptions and shorter lengths of stay as well as better quality of life for those animals that are staying longer (which could lead to increases in costs of sustaining them) [2]. Gouveia et al. describe the struggles that adult and elderly cats go through to get adopted quickly and the stress that is endured the longer an animal is kept in a shelter [2].

This study acknowledges research that suggests boredom can grow in animals kept in shelters due to an inability to experience the environment. My work could also be used to help shelters acknowledge which animals have been experiencing boredom and behavioral issues in association with longer lengths of stay and it can help to create a system to ensure animals do not go too long without interacting with the environment. For example, if the charts show that certain animals have been in the shelter for x amount of time, they are then given priority to either a) be in a foster program or b) be involved in a program to give them environmental interaction. My work can be used to help track the length of stay, along with other factors, of animals to help make decisions and set up schedules accordingly.

Another interesting aspect that Gouveia et al. describe is that adjustment to the animal shelter takes time and "aggression and fear in cats" begins to decrease after a few months of being there [2]. Additionally, they note that the sex and age of the cat may also contribute to differing behaviors [2]. The study that Gouveia conducted, in which animals were placed in shelters and observed on a consistent basis, showed that time spent in inactivity became higher with time spent in the shelter, most significantly noted when animals were housed for at least 7 years as compared to 1-6 [2]. Longer resting periods, disinterest, and "learned helplessness" were all noted characteristics, which deplete animal wellbeing [2]. These factors can also impact the risk of disease and death because this inactivity and lack of emotion can create confusion as to behavior versus health concerns [2]. Gouveia et al. also express that overeating, particularly in cats, is a concern with increased stress levels [2]. Overeating can cause negative effects on the wellbeing of animals as well as cost more spending on food for the shelters. Animals with longer lengths of stay also had high rates of negative encounters and aggression [2].

Another interesting aspect that Gouveia et al. examined was the effect of different factors and characteristics of animals (such as sex, age, if the

animal is spayed/neutered) involving group behavior [2]. They discuss how these findings could be useful when determining groups to create, and they suggest to look into how “group composition” affects “individual behavior” to determine “the most compatible group characteristics that will contribute to improve the quality of life of cats that are rendered to a shelter, particularly those that are destined for long term confinement” [2]. My work contributes to this suggestion and helps with these concerns. My work can help to navigate the different characteristics when determining group composition. This way, my work can be used to help urgently promote animals to increase their likelihood of adoption and also help to improve the quality of life for the animals that continue to have longer lengths of stay. This study also helps to prove that the need for early adoptions and improved quality of life are necessary along with means to help make that happen.

Method of approach

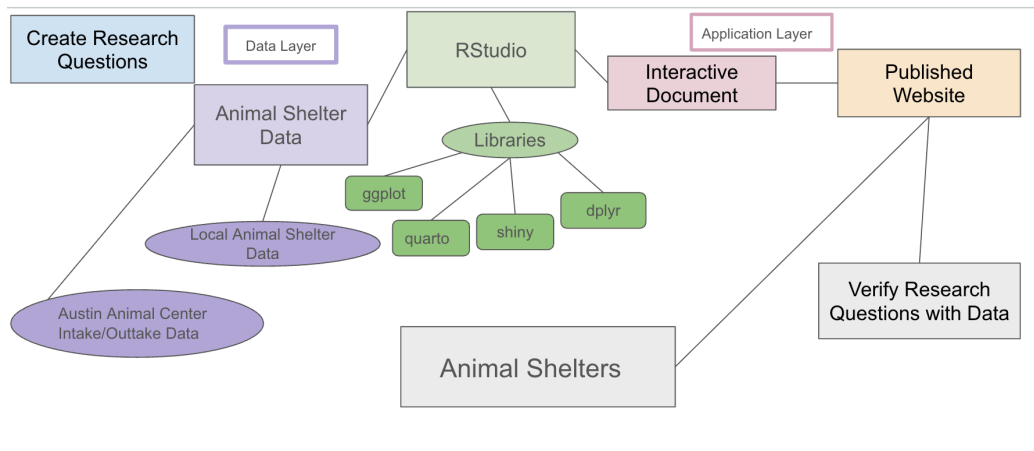


Figure 3: Research Development

The architecture diagram above frames the process and layers of this project.

The first step was to determine the research questions that I wanted to investigate in my project and answer with my tool.

The overall questions are:

Can a tool be developed that helps the most at risk animals be identified and presented toward a higher likelihood of being adopted?

Does my tool, which consists of a collection form and dashboard, incorporate enough data to answer questions that are important to make marketing decisions.

Is this tool able to be utilized by an animal shelter that doesn't know much about computer science?

These questions were established by developing the overall research gap that is the lack of an overall standardized data collection methodology in the shelter sector, which results in there lack of data based decision making.

Once these questions were established, the next layer involved the process of seeking information and acquiring data. To complete this project, I have done both online research and seeking publicly available data and research by reaching out to various animal shelters. Pros and cons exist for publicly available vs. local shelter data.

Pros of publicly available data sources online include:

- Consistency
- Efficiency
- Easily accessible
- High attention to detail/transparency
- Helpful if local data did not work out
- Useful for setting templates, public knowledge increase, showing national trends

Cons of publicly available data sources include:

- Less personal
- Inability to ask for further information or explanation
- Less flexible

Pros of working directly with shelters:

- More personal
- Local/specific trends as opposed to national

Cons of working directly with shelters:

- Not always reliable with response & data
- Not always consistent
- Varying degrees of data collection/specifics
- Having data and utilizing both kinds of resources was an important aspect of this project.

I have reached out to numerous animal shelters, both locally and throughout the nation. While some have been unable to help, most notably due to having to deal with overcrowding, understaffing, and/or disease influx, I was able to discuss with some others. To be more specific, the effects of this level of being busy affected many different aspects of their data. Part of it meant that they were too busy to exert energy to work with me, part of it meant that they were too busy to keep sufficient and proper/uniform data, part of it meant they were too busy to organize data in an effective and usable manner, and part of it meant that they did not use the data for analysis purposes or have an organized/standardized way to do it and were too busy/preoccupied to do it. We discussed the importance of data collection and analysis in shelters, what their current methods and forms are, what I was looking for, and what their thoughts were. Each conversation varied depending on the shelter, but that was the general outline. Most animal shelters have intake and outtake data, but admitted that they don't necessarily utilize it to examine trends or make decisions. Overall, speaking with animal shelters directly mainly resulted in furthering the evidence that useful, standardized, consistent data that can be used to make informed decisions in the shelter sector is severely lacking.

I also researched and reviewed scholarly articles to understand current research methodologies and best practices involving animal shelter data collection and usage. The literature review assisted in finding research that supported and discussed the challenges that I identified. This research further supported the notion of the need for a standardized data collection process for animal shelters to increase their ability to make data based decisions. I also reviewed methods of marketing strategies that are proven to be effective that can be used in tandem with a data analysis process.

Study Design

After this mixed-methods approach of in depth research collection, I developed the idea to help animal shelters standardize the collection of their data and use it for analysis purposes.

Originally, the tactic I took was relying more heavily on shelters having more consistent data and better technical knowledge. I had begun creating an interactive document on RStudio that required shelters to be able to work more closely with code, as it was made up of various code block sections and

a step by step process of how to run it locally themselves and alter their code and/or data to adhere to the standards I had developed during my test code that was publicly available. When this document was rendered and published publicly, it contained static graphs and visualizations as well as Shiny applications that were able to be utilized by anyone who had access to the website. However, the data I had used was contributed by a publicly available dataset. This document was helpful in that it could allow for public knowledge increase by examining how data could be used to seek trends, but shelters would have to be able to locally use R and follow a step by step process, changing values as necessary, if they wanted to use their own data. Therefore, this tool would not have been easily used by shelter staff with very limited technical knowledge. However, the graphs and visualizations I created were beneficial in the transformation to a more user friendly development.

The next step was to develop a method of standardized data collection to ease this process and then create a reactive dashboard based on file input that could represent trends and visualizations. If animal shelters have csv files with data values that coincide with the data values necessary for the dashboard to work, or if they have the ability to easily change their data value names, they can use their own file. However, if not, they can easily use the data collection form I have developed to create a standardized form. The data values that are necessary are listed and provided.

I designed a structured data collection form to gather relevant information from animal shelters. Data values included are: `animal_id`, `animal_type`, `breed`, `age`, `sex`, `color`, `name`, `size`, `outcome_type`, `intake_condition`, `intake_month`, `outcome_weekday`, `outcome_month`, `intake_date`, `outcome_date`. The form was created using Shiny apps, which is a web application framework in R. Figure 1 illustrates the user interface of the data collection form.

```
$ # UI definition
  ui <- fluidPage(
    titlePanel("Animal Shelter Data Entry"),
    sidebarLayout(
      sidebarPanel(
        textInput("file_path", "File Path", placeholder = "Enter file path"),
        textInput("animal_id", label = "Animal ID", placeholder = "Enter Animal ID"),
        selectInput("animal_type", label = "Animal Type", choices = c("Select Animal Type", "Dog", "Cat", "Other")),
        textInput("breed", label = "Breed", placeholder = "Enter Breed"),
```

```

    numericInput("age", label = "Age", value = NULL),
    selectInput("sex", label = "Sex", choices = c("Select Sex", "Intact Male", "Intact Female", "Neutered Male", "Neutered Female")),
    textInput("color", label = "Color", placeholder = "Enter Color"),
    textInput("name", label = "Name", placeholder = "Enter Name"),
    selectInput("size", label = "Size", choices = c("Select Size", "Small", "Medium", "Large")),
    selectInput("outcome_type", label = "Outcome Type", choices = c("Select Outcome Type", "Adopted", "Euthanized", "Returned to Owner", "Other")),
    selectInput("intake_type", label = "Intake Type", choices = c("Select Intake Type", "Volunteer", "Surrender", "Stray", "Other")),
    selectInput("intake_condition", label = "Intake Condition", choices = c("Select Intake Condition", "Good", "Fair", "Poor", "Other")),
    selectInput("intake_weekday", label = "Intake Weekday", choices = c("Select Intake Weekday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")),
    selectInput("intake_month", label = "Intake Month", choices = c("Select Intake Month", "January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December")),
    selectInput("outcome_weekday", label = "Outcome Weekday", choices = c("Select Outcome Weekday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")),
    selectInput("outcome_month", label = "Outcome Month", choices = c("Select Outcome Month", "January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December")),
    dateInput("intake_date", label = "Intake Date", value = NULL),
    dateInput("outcome_date", label = "Outcome Date", value = NULL)
  ),
  mainPanel(
    actionButton("submit", "Submit"),
    br(),
    tableOutput("recent_entry_table")
  )
)
)
)

```

The algorithm works to: Use Shiny's `fluidpage()` function to define the UI layout. The sidebar panel layout includes input fields for the relevant data values about the animals. The main panel includes an action button for submitting data and a table output to display the most recent entry. The next figure displays a sample of the server logic.

```

# Server logic
$ server <- function(input, output, session) {
  # Store the Data
  data <- reactiveValues(entries = data.frame())

  # Show the most recent entry
  output$recent_entry_table <- renderTable({
    data$entries
  })
}

```

```

    # Submit button
observeEvent(input$submit, {
  # Show loading
  showModal(modalDialog(
    title = "Processing",
    "Please wait while the data is being processed...",
    footer = NULL,
    size = "m",
    backdrop = "static",
    keyboard = FALSE
  ))

  # validate fields
  if (input$animal_id == "" || input$animal_type == "Select Animal Type" || input$
    # close loading
    removeModal()

    # validation error
    showModal(modalDialog(
      title = "Validation Error",
      "Please fill out all required fields.",
      easyClose = TRUE
    ))
    return()
  }
}

```

The server logic works to: Create the “data” object to store the data that is entered. `renderTable()` serves to update with the most recent entry When the submit button is pressed, the observer validates to make sure that all of the fields have been filled, or an error message will appear. Then, a new entry will be created. The length of stay by days is calculated based on the intake and outcome dates and stored as an additional data field in the table and csv file that is created or appended by submitting the form. The input fields are cleared after the form is submitted. `ShinyApp()` is used to run the Shiny application. This application works to provide users with a form that they can use to enter data about the animals in their shelters and store their data in a csv file with standardized, consistent, and relevant data.

The next step was to create an interactive data analysis platform using R and Shiny. This platform provides a tool for users to upload their standardized csv file that was created by the data collection form, or a file with the same data values of their own, and perform interactive analysis. This next figure displays a sample of code used in the creation of this platform.

This Shiny application works to: Create a dashboard with a header and sidebar menu. Set up different tabs for different analyses. The tabs contain plots and interactive elements that aid in exploring data. Take in a csv file and react to it while also allowing for user interactivity to explore the data further. Utilize ggplot2 and plotly. The dashboard provides an interactive platform for users to explore and analyze data through visualizations and analyses, incorporating data such as breed counts, average length of stay, breed by animal type, intake and outcome distributions, etc.

This tool can be utilized, and it can be tested using a standardized dataset that already exists or is created by the data collection form. Additionally, shelter staff can be guided through the process of using the tools effectively. These tools can be used with minimal technical knowledge, as they are applications that are published and can be publicly used.

The github repository that contains the data for this tool can be found : <https://github.com/ReadyResearchers-2023-24/mariahgammonartifact.git>

However, there are also steps you can take if you want to be able to view and run the program locally on your device.

This project was created using R Studio. R Studio can be downloaded on Windows, Mac OS, and Linux and requires a 64 bit operating system. It can be downloaded for free by searching it and clicking on the link to RStudio Desktop.

Select the version that matches your computer's operating system.

Over 12,000 packages are available to download for R that can be downloaded to help create projects, perform analyses, and utilize the software. The libraries that you would need to download in the console to perform my project are: `library(ggplot)` `library(dplyr)` `library(shiny)` `library(quarto)`.

R is the programming language that I used to complete the interactive document/webpage portion of the project. R is free and open source software that is beneficial for statistical analysis and visualization. Because it is open source and issued under the General Public License, anyone can use it for free and without restrictions, and it is easily shared with and collaborative with the public.

Data visualization is beneficial for reproducibility and transparency. Us-

ing code to create the visualizations, plots, and view trends, allows for that code to be modified and used in the future as opposed to starting from scratch each time. R visualizations allow for the user to have a sense of control over their work and what they want to represent and how they want to represent it. R gives the user access to customization and outputs that can be highly beneficial to contribute to making data based decisions. Influential companies and news outlets support the use of R visualization as their primary source, such as BBC and The New York Times. The pieces that make up the graphics are known as a grammar of graphics. In my thesis, I used ggplot2, which is a piece of tidyverse.

Ggplot2 is a popular package in R that aids with data visualization and analysis. Features of note in ggplot are highly customizable and layered graphics, faceting, themes, and tidyverse integration.

The Shiny package in R works to help to create interactive features and web applications that can run R code. Shiny is useful for many different tools, such as embedding in a webpage, integrating into R Markdown documents, and creating dashboards. Shiny is very useful to be able to work effectively with users and user input. Shiny applications utilize a definition in a user interface and a server script. It also effectively incorporates reactive programming based on user inputs and which R code will be executed. The dplyr package is useful for data manipulation purposes.

Experiments

Experimental Design

Various steps were taken to perform experimentation and review evaluation when it comes to this project. As stated previously, there are steps that can be taken to run the experiment and play with the code directly on your local device by downloading R and ensuring that the proper libraries are installed. Once you have the proper programs and codes input, you can either edit the code further to modify it as you please and account for changes you want to make, then render it or publish it yourself.

However, if you want to reproduce the experimental work once the artifact is already established, that is possible too. When examining the interactive document created through R, documentation is provided to explain the purpose and include a guided tour on what to do to utilize the applications.

Experiments can be run by utilizing the tools, using existing and test data, to ensure that they run properly, efficiently, and as expected, and that they can be used to answer the research questions thoroughly. The gap found acknowledges a lack in a standardized data collection and analysis process for animal shelters to utilize and even further, utilize to help drive decision making. This tool creates access to that. When reaching out and speaking to animal shelters directly, as well as when doing online research, I realized that there is no uniform way to collect and analyze data. It's hard to analyze and identify trends if you don't take a step back and create a way for shelters to be able to collect data in a standardized way. Incorporating a data collection form helps to mitigate the issues of having so many varied values and types of datasets. Especially because many animal shelters either don't have the computer skillsets or time/resources to create a way to efficiently collect data in a standardized way, my taking the time and effort to create such a form in a centralized, uniform way, helps immensely to ease and standardize the

collection and analysis process, which aids in wedging this gap. So, to further run experiments and ensure the tools work effectively, the tools had to be used. Experiments were conducted by utilizing the tools and verifying the conclusions via the research questions. Real, established animal shelter data with standardized values in a csv file was run through the dashboard to ensure that it ran and showed effective visualizations that could be used to identify at risk animals. Additionally, this real, established data and testing data was input into the data collection form and then the CSV file created from that form was run through the dashboard to ensure that the tools can both be used together to create visualizations and that the data values align. Also, the data collection form was tested to ensure it ran smoothly and properly.

A screenshot of a web browser window displaying a form titled "Animal Shelter Data Entry". The form is located on the left side of the browser window and contains several input fields and dropdown menus. The fields are: "File Path" (text input), "Animal ID" (text input), "Animal Type" (dropdown menu with "Select Animal Type" as the placeholder), "Breed" (text input), "Age" (text input with a spinner), "Sex" (dropdown menu with "Select Sex" as the placeholder), "Color" (text input), "Name" (text input), and "Size" (dropdown menu with "Select Size" as the placeholder). A "Submit" button is located to the right of the form. The browser's address bar shows "127.0.0.1" and the system clock indicates "Fri Mar 22 10:20 PM". The browser's menu bar includes "Safari", "File", "Edit", "View", "History", "Bookmarks", "Window", and "Help". The browser's toolbar includes various icons for navigation and development tools. The browser's sidebar shows a list of tabs, including "Animal..." and "update...".

Figure 4: Collection Form

To think about how this experiment can help to answer and verify the research questions, let's review them. Can a tool be developed that helps the most at risk animals be identified and presented toward a higher likelihood of being adopted?

To start, the tools were definitely able to be developed. The data collection tool asks for specific and relevant information that is relevant for identifying at risk animals. Breed, age, health condition, length of stay, outcome type, and intake days/months are all relevant characteristics that contribute to identifying at risk animals. These characteristics are established to be relevant for this task by a) doing online research of suggested best practices b) deducing basic variables that can determine trends that would establish distinguishable factors such as breed by length of stay and c) speaking with animal shelters and determining the most typical factors to collect and selections within those factors, such as types of outcomes. These data values that are requested by the form are useful and relevant to be able to create visualizations and identify trends that are relevant to identifying animals at risk. This evaluation answers the first part of how this question is verified.

Additionally, this question can be answered by viewing the visualizations created by the dashboard. Does the dashboard create visualizations that can be used to identify at risk animals and aspects that contribute to animals that need attention to be focused on them?

For example, graphs are created that show average lengths of stay by breed and age, and it also allows for drill down to examine specific breeds and ages up close. It also shows average length of stay by animal type. It also shows intake and outcome months. These visualizations are all able to be used to identify at risk animals that should be presented for a higher likelihood of being adopted because these visualizations allow for someone to examine which animals have been in the shelter for longer periods of time, and this factor in comparison to other animals and breeds, to see which animals should be examined to be highlighted. Additional graphs and visualizations can be incorporated in the future with the same established data values to increase the effectiveness of this aspect. But, having these data values exist, be standardized, and be collected in an efficient manner and easily accessible file is already an immense step. Proving that visualizations that are useful can be created and utilized is also a big step.

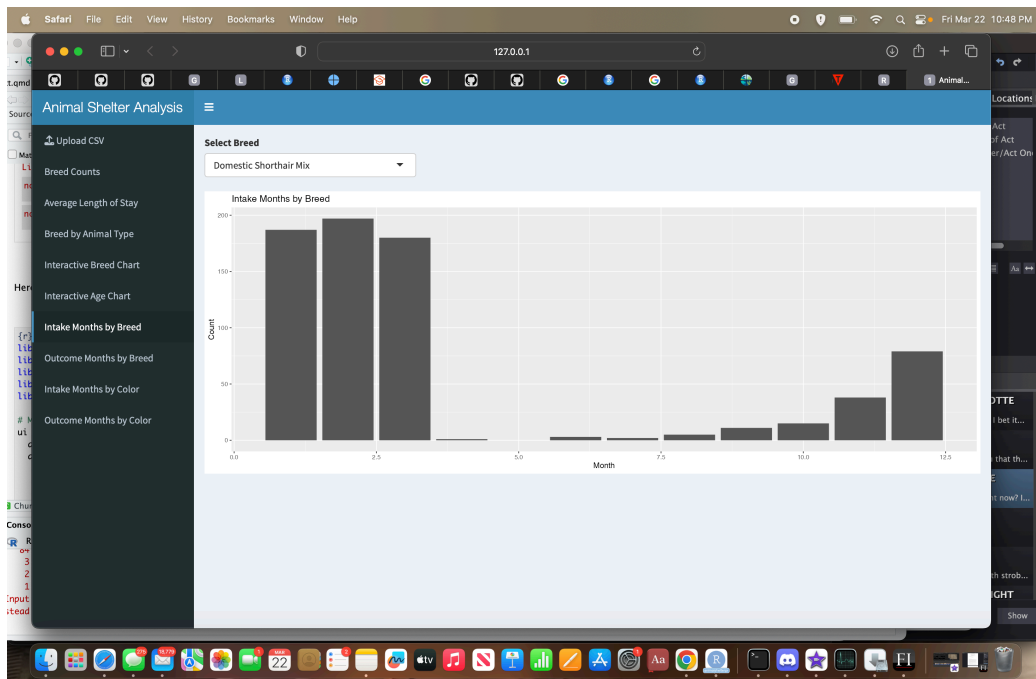


Figure 5: Month Comparison

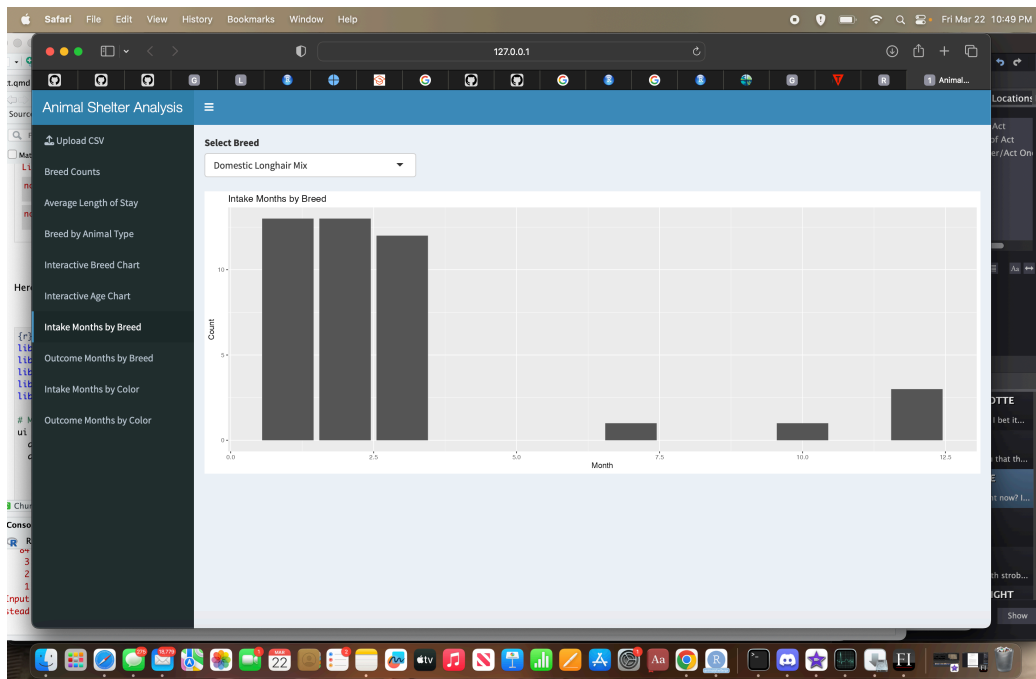


Figure 6: Domestic Comparison

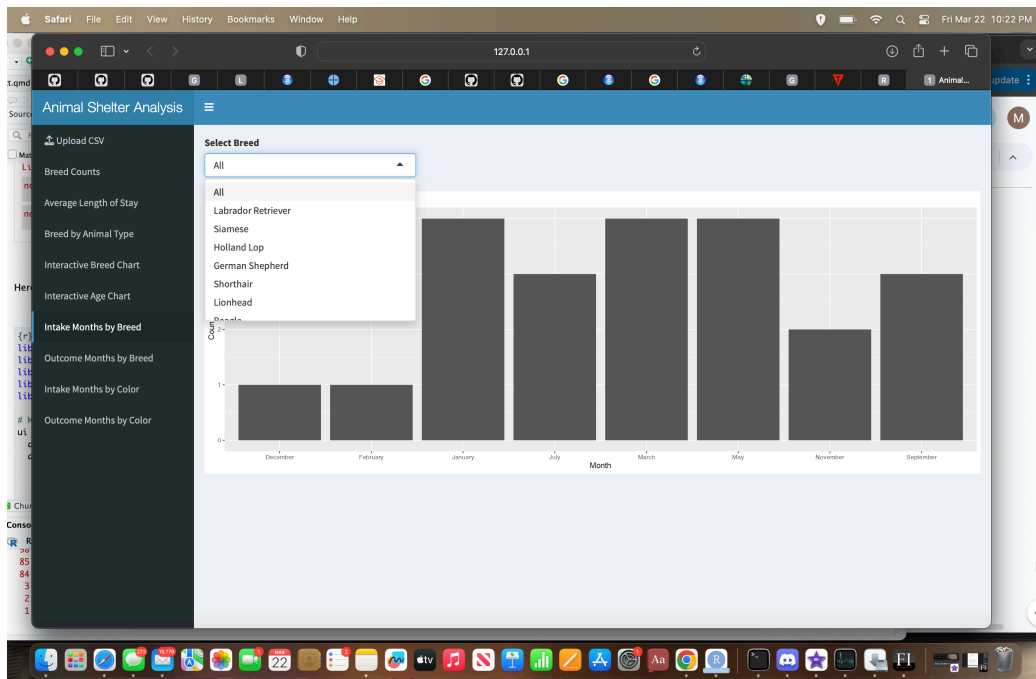


Figure 7: Months Visualization

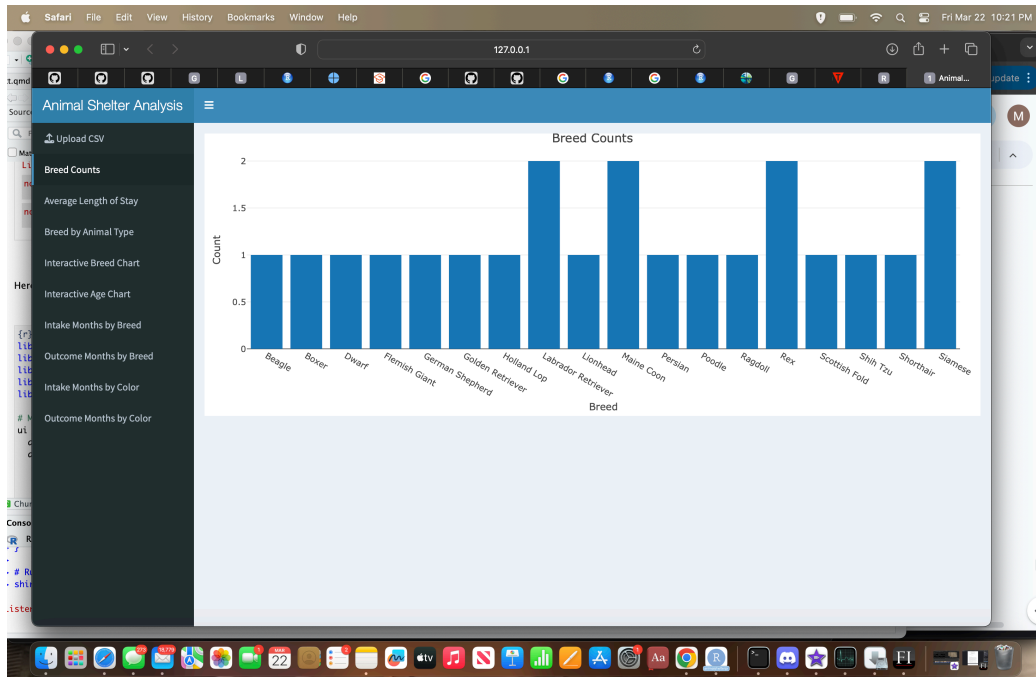


Figure 8: Dashboard

Does my tool, which consists of a collection form and dashboard, incorporate enough data to answer questions that are important to make marketing decisions?

This question is already partially answered above, but can continue to be answered in depth here. The data values were carefully determined based on standardization and useful purposes. These data values that are to be input in the collection form can all be used to provide useful insights and answer questions that can contribute to marketing decisions. For example, the intake and outcome days and months, which are then represented in the dashboard and can be drilled down by breed and age, etc. can show which days of the week and which months are more popular for income and outcome, both on a broad perspective and drilled down into specific characteristics and animals. When viewing these visualizations, marketing decisions can be made accordingly, keeping in mind which times of the year or week are best to promote which kinds of animals. Additionally, research suggests that the positioning of animals physically in the shelter when visitors are there also

has an impact on whether they get adopted, so by identifying the at risk animals, shelters can determine where to spatially place them and how many animals to include in a room at a given time, to help aid in their adoption likelihood, which is a marketing decision.

Is this tool able to be utilized by an animal shelter that doesn't know much about computer science?

When researching what it takes to work at an animal shelter, both through expectations found online and by speaking directly with animal shelters, basic computer skills are typically expected to work. So, it should be expected that an animal shelter worker with basic computer skills, but still limited technical skills, especially concerning computer science, should be able to use this artifact. Various factors verify that this is possible. To begin with, documentation accompanies the artifact, with explanations as to what its purpose is and how to use it/fill it out. The documentation clearly states what data values are required for the dashboard to work without any modifications, and it also clearly states that the data collection form is for use if the shelter does not have a standardized file with the values already. The data collection form itself is a Shiny application that provides inputs for the various data fields that are needed. It states what the value being asked for is above each input, and each input contains a placeholder to describe what type of data goes inside each piece. As explained in the Validity Checks section, there are various validity enforcement tools in place, such as validation checks, error messages if failed, and selection restrictions, to make the filling out of the form easy to use and understand. Once the form is filled out, either a new CSV file is created, or a previously established CSV file is updated based on the new inputs into the specified location. The user then has a file with standardized values that can easily be uploaded into the data dashboard.

The dashboard has validity enforcement tools in place as well to ensure that it is easy to use and understand by the user. All the user has to do to be able to utilize the dashboard is upload their premade CSV file, and the graphs and visualizations will be made available for them to review. The graphs are all efficiently labeled and selection options are limited and labeled.

This process is much more efficient than the previous idea, with the more static document, as it would have involved a lot more coding involvement for the user, where they would have to run the code themselves and change data value names in their files or throughout the code. In this case, the data collection form helps to ease the process to make it more efficient and usable

and the user will not have to do any code themselves if they do not desire to. The data collection form is available for them to directly fill out and the CSV file is created/modified with their submissions. They then can upload that file into the dashboard that is already running and available to them. The input that is requested from the user is clearly labeled and directed and the dashboard only requests the file to be uploaded before the visualizations are created. The visualizations are clearly labeled and able to identify trends. Therefore, the tools are usable and easy to follow for an animal shelter worker that has limited computer science knowledge.

The experimentation and evaluation prove this question because I ran multiple tests to ensure usability. I took a file provided by an established animal shelter with these values and was able to run it through the dashboard, which produced visualizations as expected with clearly labeled points, tables, and values. I also took sample data, both established and testing data, and used it to fill out the data collection form to ensure that the data collection form ran smoothly, efficiently, and accurately. The input fields were easy to follow and the validation checks ran as expected. If all of the fields were not filled, an error message was displayed with clear instructions on how to fix it, and if each piece of data was filled out correctly, the submission was added correctly and the file was created. Then, I used that created file to upload it into the dashboard, and the dashboard ran as expected, developing visualizations based on the files that were uploaded. Running these experiments and ensuring that everything worked as expected in a thorough and guided process, not having to use any computer science skills, shows that an animal shelter could use it, and these tools can help standardize and simplify the data collection process for an animal shelter.

Threats to Validity

Validity Checks

Data Collection Form

In creating the data collection form, I have incorporated validity checks to mitigate room for error when a user is filling out the form, which helps to eliminate error in data collection and future analysis. Before processing the submission, I included a form submission handler that runs checks before accepting the submission and processing the data.

The “observe\$event” checks for entries into the input variable, which means that when the form is submitted it will run. A modal dialogue will pop up to let the user know that the data is loading and being processed. This action occurs before the validation checks, but it is necessary to allow for increased transparency and usability for users, so they are more informed about the data collection and submission process.

Next, field validation occurs. This code is run to check all of the input fields are not empty or set to default values such as “Select Animal Type” or “Select Intake Condition.”

These conditions are checked, and if any of them are not met, a an error message pops up, saying “Validation Error: Please fill out all required fields.” The return statement will cause the function to end in that case, so if the validation check fails, further code will not be executed. The form will not submit in this case. The user will then have to ensure they fill out all of the fields before they can submit the form.

This validation check is necessary to ensure that there are no empty or invalid responses in the form, which would lead to insufficient or inaccurate fields, which would hinder the data collection and analysis process. After the validation process, the loading modal goes away.

This code to check validation is necessary and useful to guarantee that no null values are represented and every field is filled out before allowing the form to be submitted. The user receives a guided feedback process when validating so they can understand the process. This process is made usable and helpful for the user, as they are kept informed throughout the data collection process. This process helps to mitigate any confusion and aid in user transparency and expectations. It is also useful to decrease the chances of having any invalid or inaccurate data before submitting the form, and this precaution helps to prevent further inaccurate results in future analysis.

When the validation check is run, it displays an error message that aids in the user's understanding of what went wrong and how to fix it before they can proceed. That step enhances the usability of the form. The integrity of the collected data is improved, as incomplete submissions will not be allowed. Overall, the process of validation helps to guarantee complete data and an improved user experience that is easy to understand and follow along. Validation checks are necessary to aid in the process of preventing issues.

Additional validity enforcement aspects are made during the Input process itself. A variety of questions are displayed, as shown above. Some allow for users to input text in the form, while others make users to select from a variety of options. One check for validity in this process is that each question states what it is looking for and contains a placeholder with further instructions on what to enter or select in the box. Placeholders supply the type of input that is expected from the users in each field. It helps serve as instruction to ensure users know what type of information to include where.

Some selections, such as age, intake date, and outcome date have the default set to NULL, to guarantee that users are selecting a value that is valid as opposed to accidental default values that could end up in data being inaccurate. The age input also requests that the value input is numeric, ensuring that the age is displayed as a number and decreasing the chance of the data being inaccurate due to other characters being placed there.

Furthermore, the use of select inputs require predefined choices for the user to select from. This way, users have to select a value from one of the choices given to them, which guarantees the answer is valid as well as standardized, as there is less room for user input error and/or misspell. Inputs utilizing select choices include: animal type, sex, size, outcome type, intake type, intake condition, intake weekday, outcome weekday, and outcome month. Intake date and outcome data utilize date input, which ensures the user selects a date for the value entry. Some of the choices do still allow

for text input, such as file path, animal id, breed, color, and name, as these values are broader and have a wider variety of answers that could be input. However, the future work section covers this topic and provides various steps that could be taken moving forward to mitigate room for error here.

After the validation check/form submission, there are additional validity enforcement aspects. In this section, the `new_entry` data frame is created, and values from the input fields are assigned to columns. Data type conversion is automatically utilized by R. This data type conversion is seen with the age input, where if a string were input, it would automatically convert it to a numeric value because of the age column's format. Values that are input will be formatted to the expected data type and help to ensure the uniformity and validity of the data. The calculation for `time_in_shelter_days` finds the difference between outcome and intake date. This calculator serves to check validity because the values would have to be valid dates in order for the calculation to work. The code also reviews if the file path that was input exists before creating the CSV file. This check examines whether the file path is valid and writable, which is very useful to save the data correctly.

Additionally, if the file already exists, it will add the new data to that file, which will help the user keep track of all of their data. After the data is submitted and checks and operations are done, the code works to clear the input fields, which aids in validating data for future submissions because users are starting with clear fields and there is less room for user and input error, with an increased chance of valid and accurate data. All of these checks and operations serve to help ensure that the data collection form is usable, reliable, thorough, and in good condition. This process is made as simple and guided as possible as to enhance the process and make it usable for staff with limited technical knowledge.

The data collection form contains various checks to ensure the validity of the process and the data collected. These checks are also highly important for the created CSV file to be valid, usable, and accurate in its use in the following dashboard.

Dashboard

In creating the dashboard, several validity enforcement aspects also take place.

To begin, in the server function, a reactive expression “dataset” is set to

read the CSV file that is input. A file has to be inserted in order for it to be read, which guarantees that the application will not continue without a valid file. When rendering plots with “renderPlotly” and “renderPlot” functions, the code checks to see if the required dataset is there before it will continue to create the plot. When the data needed is there, then the plot will be rendered. This check makes sure that the data required to create the plot is available for plotting.

Dropdown menu selections are made so that only values in the dataset are shown and able to be selected. For example, with selecting breeds and colors to be shown in the plot, “breed_selection,” “breed_selection_intake,” “breed_selection_outcome,” “color_selection_intake,” and “color_selection_outcome,” only breeds and colors that are actually included in the dataset will be available for selection. This process guarantees that only valid values will be able to be chosen. The use of reactive expressions also helps with ensuring validity because they work to filter the dataset based on selections, which guarantees that only valid values will be included in the plots. Additionally, creating a filter for outcome types guarantees only outcome types that are valid will be included.

These checks are useful for various reasons. Only rendering plots when the dataset is available allows for the avoidance of errors and empty plots due to invalid or missing data, which aids in upholding data integrity and lessening the chances for misleading visualizations and analysis. Filtering options make it easier to focus on specific pieces of data and aid in meaningful insights and analyses in an effective manner for users that they can have input in. Additionally, dropdown menus that only include valid choices, helps to give users a clear outline of the choices they can make to drill down into what will be acceptable. Only including relevant and valid data allows for the visualizations to be clear and accurate. These checks can help to improve the usability, transparency, effectiveness, and reliability of the dashboard, helping to allow for informed data driven decisions to be made from it.

In general, various aspects of code in the creation of the dashboard are included to guarantee that valid actions are able to be performed, given the dataset that is input. These checks, combined with the validity checks in the data collection form that creates the CSV file that will be used to be input in the dashboard application, work to make sure that the operations run smoothly and accurately, the data and plots are valid and accurate, and users can be easily guided through the process.

By using the collection form first, users can be more confident that the

data and visualizations displayed in the dashboard will be accurate and trustworthy because it is coming from a validated, consistent, and standardized source. The data has gone through checks to aid in its accuracy and validity before being put through the dashboard process and only valid data is accepted. Both applications work separately and together to help ensure a verified and effective data collection and analysis process.

Conclusion

The tools that I have created and experimentation that I have run show that these tools can aid in the process of usability and understanding of data for animal shelters. The results show that a data collection form to create a uniform CSV file is something that the animal shelter sector currently lacks, so developing one that can be easily used is beneficial. The results show that this form is able to be used and verified to ensure that the data is valid and accurate. The results also show that once this data file is created, it can be uploaded and correctly used to create a dashboard that utilizes accurate visualizations to accurately represent the data that is uploaded that can be used to identify factors that can be used in the future to drive decisions. The data collection form creates uniform and relevant fields for future use and is helpful in guiding users through an understanding of what is expected and why. Animals and their characteristics can be drilled down and inspected to determine which animals have longer lengths of stays than others.

Future Work

This project provides a good framework for the beginning of setting up a standardized, uniform process for data collection and analysis in animal shelters.

However, many ways to expand upon this project can be examined.

To begin with, aspects of the data collection form can be modified and expanded on in order to improve the data validity and reliability, decreasing the room for error. For example, currently, the form includes a combination of user inputs, select, text, numeric, and date inputs. The select inputs allow for a more specified, uniform, and accurate collection of data. However, some of the inputs still request text input, which is more broad and varied. In the future, I would hope to be able to implement stricter tools for inputting these values, at least breed and color, as names is much more difficult to

standardize, to allow for less room for error and to increase the validity of the data. A couple of different methods for this can be implemented. Either a very extensive select input can be incorporated, after doing research to ensure a valid and verified list of breeds that are proven to be a cohesive list, with an “Other” option available as well, in case. Additionally, another avenue that can be looked into would be to incorporate fuzzy matching or AI that can be utilized to help ensure that the breed and color names are standardized and uniform. This tool would allow for a much more detailed approach that would help guarantee less room for error and uniformed values. Similar words and breeds that were input would be able to be recognized and included under a standard name. For example, if a breed was input incorrectly due to user error, it would be discovered and the name changed to the standard value so that the data would be accurate and valid and there would not be random misnamed values. Or, if breeds were referred by different names, such as a Lab vs. Labrador, the fuzzy matching would allow for this discrepancy to be caught and these values combined to be included under the standard name. These additional modifications would be really beneficial to expand this project and improve its usability and validity. The data would be more likely to be accurate and valid and it would allow for less thought and room for user error, making it more understandable and easy to use for the users.

The impact of this research project is that it has proven the concept for creating a portal for helping animal shelters improve their data-driven marketing. The completed vision includes an online form to enable animal shelters to enter data they can then leverage to create valuable visuals and graphs that will enable them to see where they can improve their marketing efforts. My goal is to eventually turn this into an online portal accessible to all animal shelters. Ideally, I would be able to implement some AI technology that could act as a translator between the data required in the form I have created and the data that an animal shelter might already be producing. Some sort of “fuzzy” match technology could be utilized.

The vision would be to host input sheet and the marketing graphs in a cloud portal. I would like to develop a series of related pre-packaged marketing programs, initially social media posts, that would be recommended based on the data input by the animal shelter. There could potentially be free and premium levels for accessing these marketing programs. In addition I could see sponsorships and advertising to businesses that sell supplies and other services to animal shelters. We could also host Webinars and blog posts designed to help animal shelters with their marketing.

Future Ethical Implications and Recommendations

Further ethical implications for the public release of this software would revolve around data privacy and unintended consequences. Data, especially data collected from private, small, locally run businesses such as animal shelters allows should be ensured that data privacy is respected and taken care of. If animal shelters access this tool, they should be able to use it locally and on their own devices, without the invasion of others or others being able to access it. However, one future work that could be incorporated would be the ability for shelters to opt in to allow their data and trends to be publicly shared, in efforts to increase transparency and standardization and trend visualization on a national level. However, this idea would be voluntary opt in only. In the meantime, the data that is used is for personal use and would not be shared with anyone other than themselves. The shelter can save the images and visualiations that are presented in the dashboard to their local device without sharing it.

The other ethical implication would be unintended consequences in the sense of promoting animals above others. This project and these tools in themselves promote the standardization and improved process of collecting and analyzing data. It creates an effecient and useful tool to help shelters better collect and view the data surrounding their animals, which is proven to have benefits. It helps make shelters' operations more equitable and allow for increased accesibility into shelter data and how their shelter operates, especially for shelters with limited knowledge and resources for computer science and data collection. This project in itself is aimed more at improved data collection than valuing animal lives over one another. However, future work with marketing tactics could stumble into this concern if not considered carefully. This process is aimed at ensuring equitable treatment and marketing tactics aimed at improving adoption rates overall and aiding at risk animals, not at the cost of others. The goal is to improve the operations and adoption rates of animal shelters in general and at a wide aspect.

Conclusions

These tools are effective in bridging the gap in the animal shelter sector that involves the standardization of data collection and analysis. After speaking

directly with animal shelters and examining their current data firsthand, the current data utilization process is extremely varied and hard to understand and analyze. There is no uniform way, and animal shelters have trouble being able to know how to handle and collect their data. This project bridges this gap and answers relevant research questions by creating a standardized and uniform data collection form that requires relevant fields to aid in the identification of at risk animals to create a csv file that can be uploaded into the creation of a data dashboard with relevant visualizations. This tool, while having room for future work and modifications, begins the process of creating a standardized way of utilizing data for animal shelters that has been previously unavailable to them. It requires limited computer science knowledge to utilize and is easy to follow, use, and understand. With the help of this tool, animal shelters can hopefully improve the way that they operate, collecting and analyzing data to drive their decisions and ensure that their adoption rates, specifically of at risk, underappreciated animals, are increased.

References

- [1] 2023. Pet facts. (2023).
- [2] Magalhães Gouveia Kelly and Liliana de Sousa. 2011. The behaviour of domestic cats in a shelter: Residence time, density and sex ratio. *Applied Animal Behaviour Science* 130, 1-2 (2011), 53–59.
- [3] Wagner Karsten C. L. and K. F. Hurley. 2017. Research, re: Search and re-search. *The Veterinary Journal* 227, (2017), 15–22.
- [4] SENCER. 2017. Do no-kill shelters really benefit animals? *NQED* (2017).
- [5] Steve Swoyer. 2016. Analytics, digital records could save money and animals at pet shelters. *TDWI* (2016).
- [6] Patricia et al. Turner. 2012. Animal shelters and animal welfare: Raising the bar. *The Canadian veterinary journal* 53, 8 (2012), 893–6.
- [7] Miller Weiss Emily and Carla Vela. 2012. Why did you choose this pet?: Adopters and pet selection preferences in five animal shelters in the united states. *Animals* 2, 2 (2012), 144–159.
- [8] Ramshankar Yadhunath. 2021. Saving Animal Lives with Data. *Towards Data Science* (2021).
- [9] The Zebra. 2023. Pet adoption statistics. *The Zebra* (2023).