# Team Iris Final Project Report

**project partners' names**:
Bula Ha,
Minjie Shen,
Yuqing Bian,
Qingjuan Dan,
Tianyiru Chen

## Abstract

Our project delves into the intricate landscape of dog adoption dynamics, utilizing datasets sourced from the Austin Animal Center and supplementary data on dog breeds obtained from online sources. The primary objective is to unravel the intricate determinants that influence the likelihood of dog adoption. We aim to construct robust predictive models capable of estimating both the probability of adoption and the duration a dog spends within a shelter before finding a new home. Additionally, our project aims to comprehensively assess the economic aspects of dog adoption by integrating data on raising costs.

By harnessing various sophisticated data mining techniques and leveraging a diverse array of datasets—comprising the Animal Center Intakes and Outcomes data spanning from October 2013 to the present, the Petfinder Public API, and supplementary information on dog intelligence and breed size—we aspire to significantly enrich the understanding of dog adoption dynamics. Our methodology entails employing a wide spectrum of analytical methodologies, including Data Quality and Standardization, Association Rules, Clustering, Regression, and Supervised Learning. Through these techniques, our aim is to uncover the underlying factors influencing adoption rates and predict the duration of shelter stays for dogs.

The conclusive deliverables of our project are envisaged to encompass in-depth descriptive insights into the characteristics of dogs associated with higher adoption rates. Furthermore, we aim to develop sophisticated predictive models capable of accurately forecasting both the likelihood of adoption and the duration a dog may spend in a shelter. These insights hold the potential to provide invaluable assistance to shelters by optimizing adoption strategies, refining resource allocation, and potentially mitigating shelter overcrowding, thus curtailing associated costs.

# Introduction

**(i) Statement of the problem**
In addressing the complexities inherent in the process of dog adoption, our project seeks to examine and elucidate the multifaceted determinants that impact the likelihood of dogs being adopted. By delving into the realm of shelter dynamics and employing sophisticated data analysis techniques, our primary aim is to develop predictive models that can provide crucial insights into the adoption process.

**(ii) Importance of solving this problem**
The significance of comprehending the factors governing dog adoption cannot be overstated for animal shelters and welfare organizations. By identifying the attributes that contribute to higher adoption rates, shelters can tailor their strategies, potentially reducing the duration of shelter stays and enhancing the well-being of sheltered dogs. Moreover, the ability to predict adoption likelihood and shelter duration aids in effective resource allocation, precise cost estimation, and meticulous budget planning for shelters, thereby enhancing operational efficiency.

**(iii) Background and literature survey**
The realm of animal shelter dynamics, particularly dog adoption, has been a subject of study in various research domains. Literature suggests that factors such as breed, size, age, and behavior significantly impact adoption rates. However, comprehensive predictive models incorporating diverse datasets and advanced data mining techniques are still evolving in this field. The inclusion of external datasets, such as dog intelligence and breed size, holds promise in uncovering nuanced correlations between dog features and adoption outcomes. Our project aims to contribute novel insights to this domain by leveraging comprehensive datasets and innovative methodologies.

The amalgamation of diverse data sources and advanced analytics techniques presents an opportunity to bridge gaps in the understanding of dog adoption dynamics, potentially revolutionizing the strategies employed by shelters and enhancing the overall adoption process.

# Methodology

Our approach involves a multi-faceted methodology, leveraging various techniques to address the identified problem in understanding and predicting dog adoption dynamics.

**1. Data Collection and Cleaning**

Data Quality and Standardization:
The initial phase of our methodology revolves around the critical aspect of data cleaning and standardization. Given that our training and test datasets originate from different sources with varying formats and features, we will employ Data Quality and Standardization techniques. This step aims to homogenize the datasets, ensuring consistency and compatibility for subsequent analysis.

**2. Unsupervised Learning**
Association Rules:
To explore potential associations between dog size, intelligence, raising cost, and adoption rate, we will utilize Association Rules. This unsupervised learning technique enables us to identify interesting patterns or relationships within the data, thereby revealing underlying connections between these variables and adoption rates.

**3. Supervised Learning**
a. Regression
Prediction of Shelter Duration:
Leveraging supervised learning techniques, particularly linear regression, our aim is to predict what will affect a dog to find a new home as well as how these feathers work when a dog is found again by shelter after adoption. By establishing a relationship between various factors and shelter duration, we aim to create a predictive model that aids in estimating this crucial metric.

b. Classification
i. Decision Tree
Prediction of Adoption Likelihood:
The Decision Tree classifier, will be employed to predict the likelihood of a dog being adopted. This classification technique allows us to create a tree-like model of decisions based on various features, aiding in predicting adoption outcomes.

ii. k-Nearest Neighbors (kNN) Classifier
Prediction of Adoption Likelihood:
The k-Nearest Neighbor (kNN) Classifier, will also be utilized to predict the probability of a dog getting adopted. This classification method relies on the similarity of features between instances to make predictions, contributing to our comprehensive analysis of adoption likelihood.

Our solution encompasses a comprehensive approach, integrating techniques from data cleaning to supervised and unsupervised learning. By standardizing and cleaning datasets, identifying association rules, and employing regression and classification models, our methodology aims to unravel the intricate determinants influencing dog adoption dynamics. This approach facilitates a deeper understanding of adoption patterns and aids in the creation of predictive models essential for animal shelters and welfare organizations.

# Code Review

**Data Acquisition**

- Procured Intakes & Outcomes datasets for Austin Animal Center from data.austintexas.gov.
- Obtained Dog Breeds information from the Best-In-Show website.
- Accessed active canine data from Petfinder using their open API.

**Data Pre-processing and Standardization**

Minimizing Data Volume

- Instance Reduction:
    - In the Animal Center dataset, isolated dog-specific records, decreasing the count from 158K to 87K for both Intakes and Outcomes.
- Feature Reduction:
    - In the Austin Animal Center datasets, omitted redundant and irrelevant features like ‘Animal-Type’, ‘MonthYear’, ‘Location’.
    - In the Dog Breed Data, refined features to align with research needs.
    - For the Petfinder dataset, selected key attributes and discarded non-essential ones (e.g., ‘Description’, ‘Organization’, ‘Emails’).

Data Cleansing

- Addressing Missing Values:
    - In the Animal Center Intakes and Outcomes datasets, substituted missing values of ‘Name’ and ‘Outcome-Subtype’ with ‘Unknown’. Drop rows with missing values of other features.
    - In the Dog Breed dataset, filled missing Intelligence, Longevity, and Cost data with category averages or majority values.
- Removing outliers and artifacts:
    - Corrected irregularities like abnormally low longevity or negative age values.
- Correcting Inconsistent Data:
    - Standardized breed names and gender names across all datasets.
- Removing Duplicate Data:
    - Eliminated redundant features like MonthYear which duplicated with DateTime.

Data Transformation

- Aggregation:
    - Used Aggregation in unsupervised learning analysis.
- Generalization:
    - Simplified the numeric feature “Age(days)” into 4 higher levels - ‘Baby’, ‘Young’, ‘Adult’, ‘Senior’.

    - Used Agglomerative Clustering to place various dog breeds into 4 price buckets – 'Saver', 'Budget', 'Mid-Price', 'High-End' according to numeric features of 'Total-Cost', 'Purchase-Cost', 'Food-Cost'.
- Normalization:
    - Normalized the feature ‘Age(days)’ from “x days”, “x months”, “x years” to numeric days.
    - Normalized mix breed to primary breed.
    - Normalized mix color to primary color.
- Feature Construction:
    - Derived a new boolean feature “IsAdopted” from “Outcomes-type”.

Data Integration

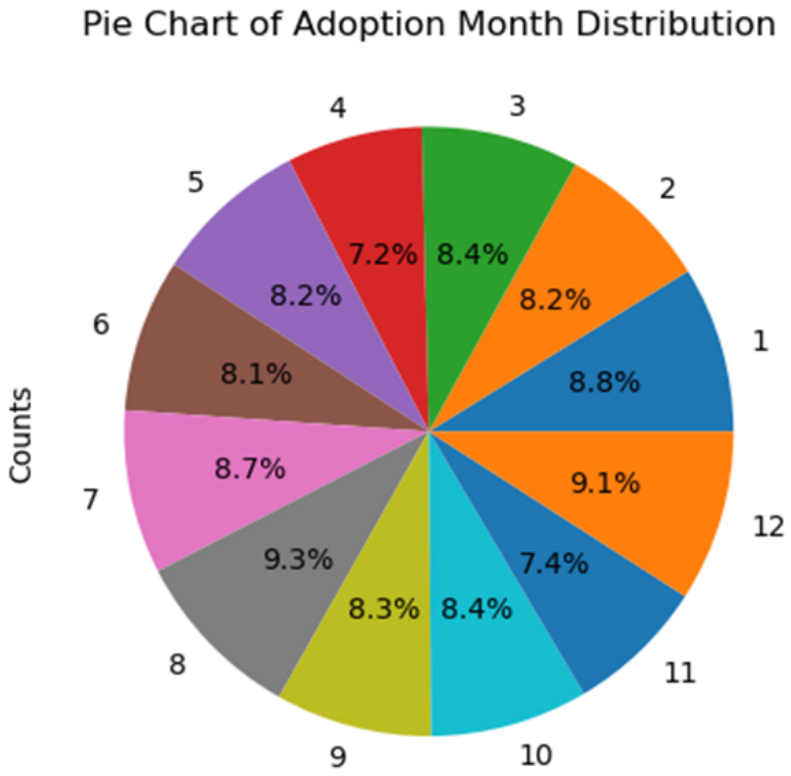
- Merged datasets to form a comprehensive view of each dog's journey and characteristics.

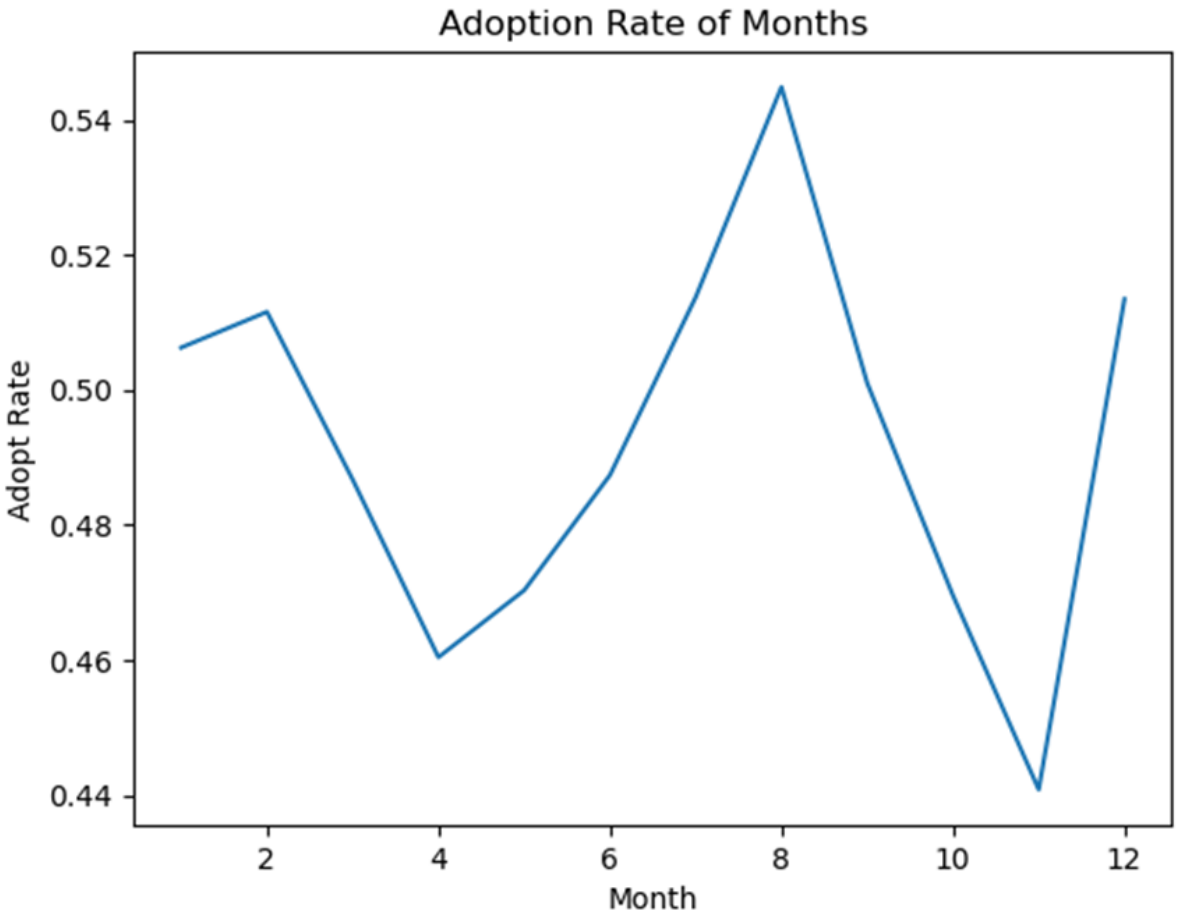
**Analysis of Adoption Patterns**

- Trends in Adoption:
    - Analyzed a decade's data (2013-2023), examining seasonal and annual trends, and breed preferences in adoptions.
- Influencing Factors of Adoption:
    - Utilized data visualization tools and statistical methods with Pandas & Numpy & Matplotlib to assess how breed, age, gender, color, intelligence, size, cost, and intake type influence adoption rates.
    - Applied the Apriori algorithm for Association Rule mining to discern patterns in adoption.

# Results

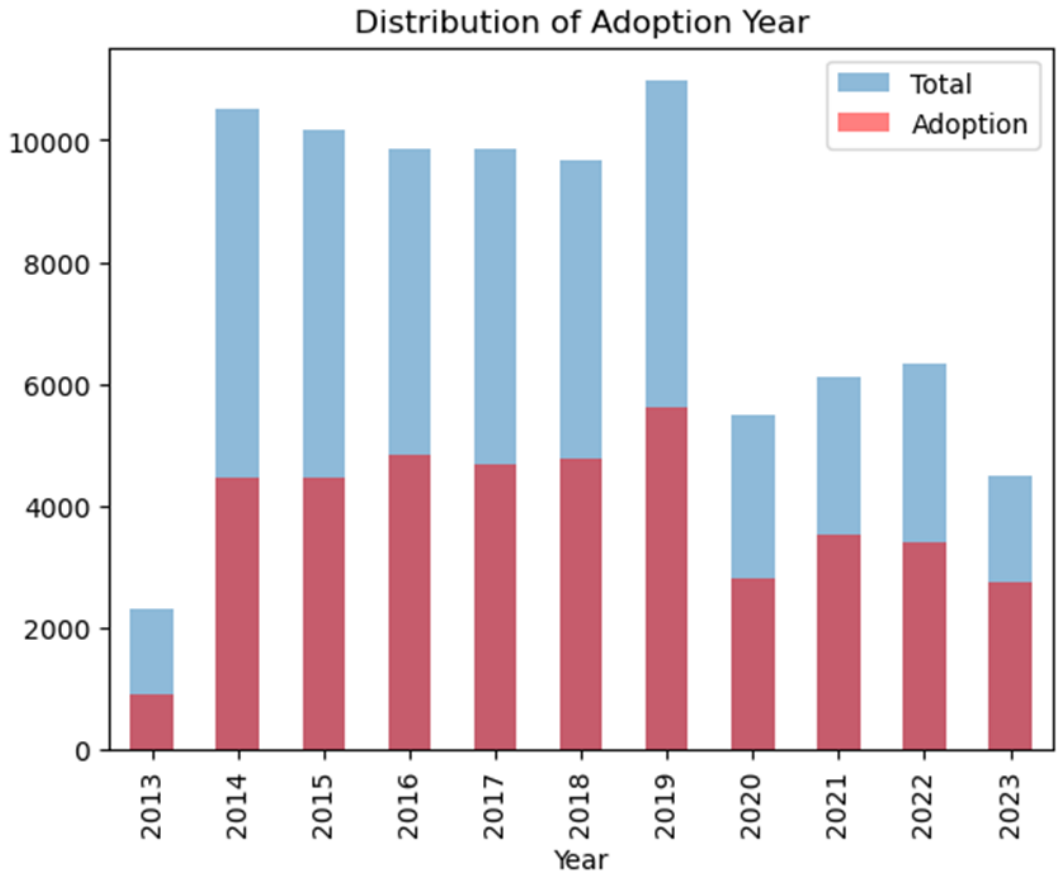
1. Adoption Patterns
    1.1. Adoption Trends
- Popular Months

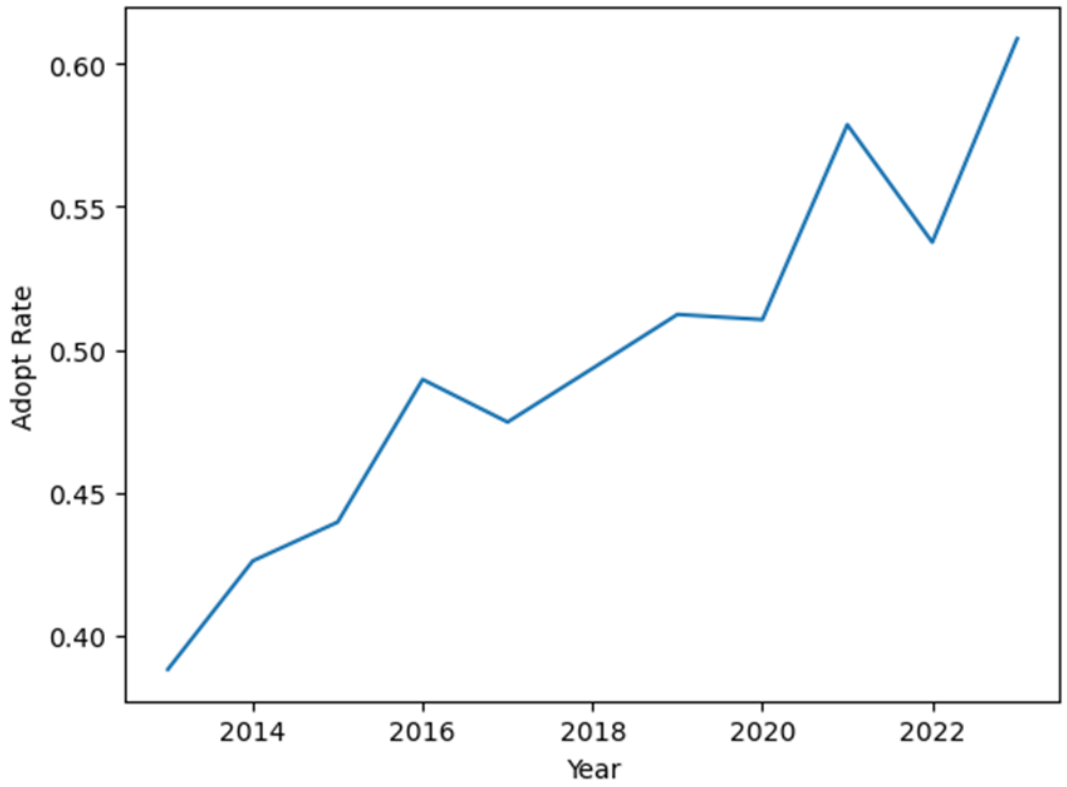

Pie Chart of Adoption Month Distribution
Counts
1
2
3
4
5
6
7
8
9
10
11
12
8.8%
8.2%
8.4%
7.2%
8.2%
8.1%
8.7%
9.3%
8.3%
8.4%
7.4%
9.1%


Adoption Rate of Months
Adopt Rate
0.54
0.52
0.50
0.48
0.46
0.44
2
4
6
8
10
12
Month

Identified seasonal preferences, with summer and winter holidays being popular for adoptions.The most popular month to adopt is August with 9.3% of all adoptions and 55% adoption rate. Holiday months, such as December and January are the next most popular months for adoptions with proportion 9.1% and 8.8% respectively.

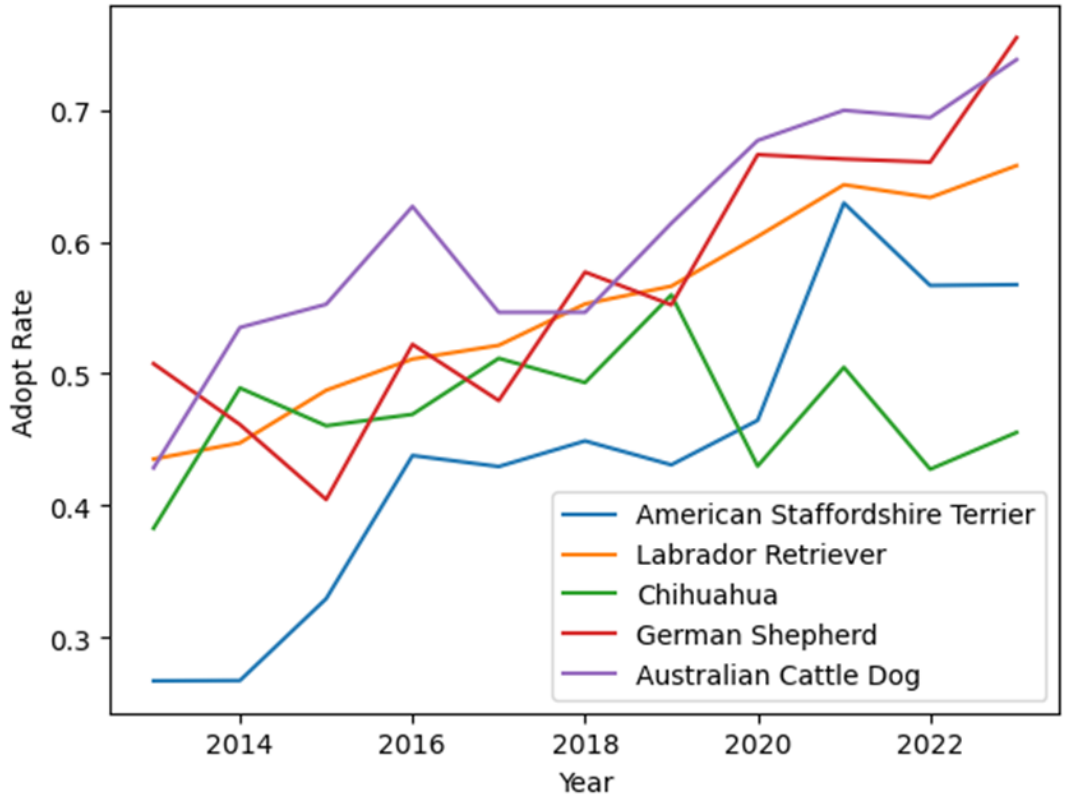
- Yearly Trend

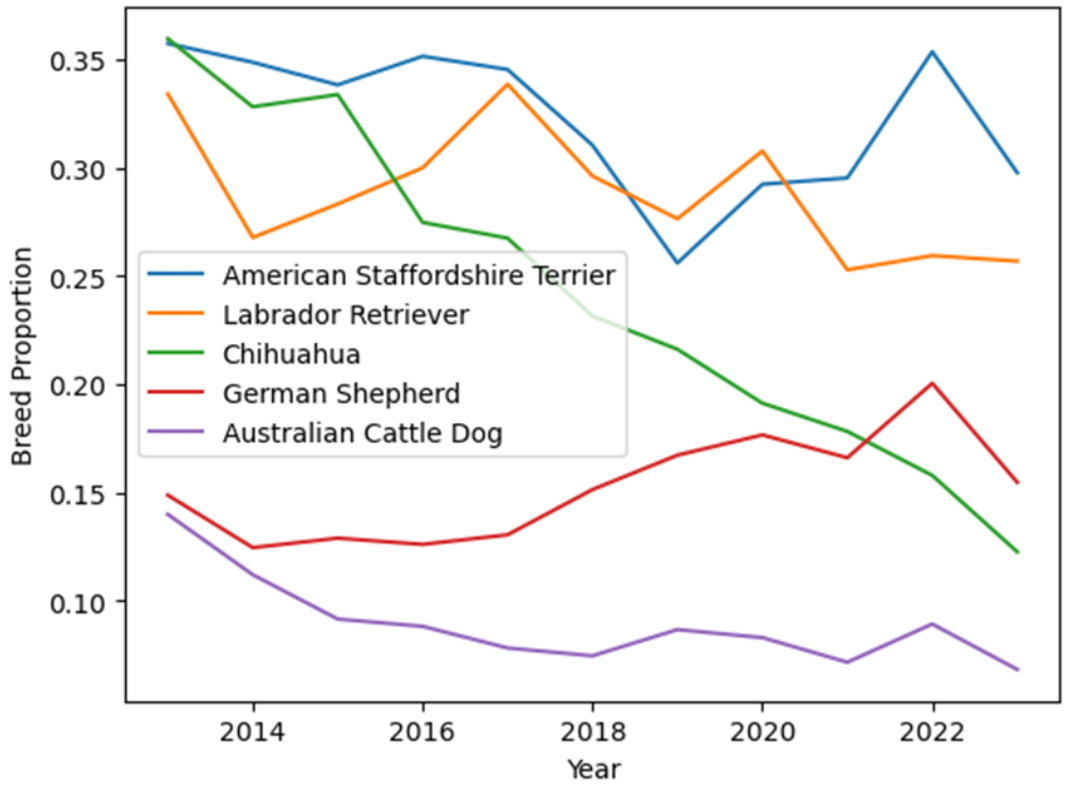





Noted a decline in total and adopted dogs amount post-2020. However, the adoption rate has been climbing steadily.

- Breed Trend






The five most common dog breeds at the Austin Animal Center are the American Staffordshire Terrier, Labrador Retriever, Chihuahua, German Shepherd, and Australian Cattle Dog. Adoption rates for these breeds are generally on the rise, with the exception of the Chihuahua. Specifically, the adoption rates for the American Staffordshire Terrier and Labrador Retriever have remained stable with minor variations. There's an upward trend in the adoption of German Shepherds, while the popularity of Australian Cattle Dogs has been declining. The adoption rate for Chihuahuas has notably decreased.

## 1.2. Adoption Influencers

1.2.1 With analysis and visualization of each dog feature, find specific breeds, age groups, colors, sizes and other characteristics to be more favorable for adoption.

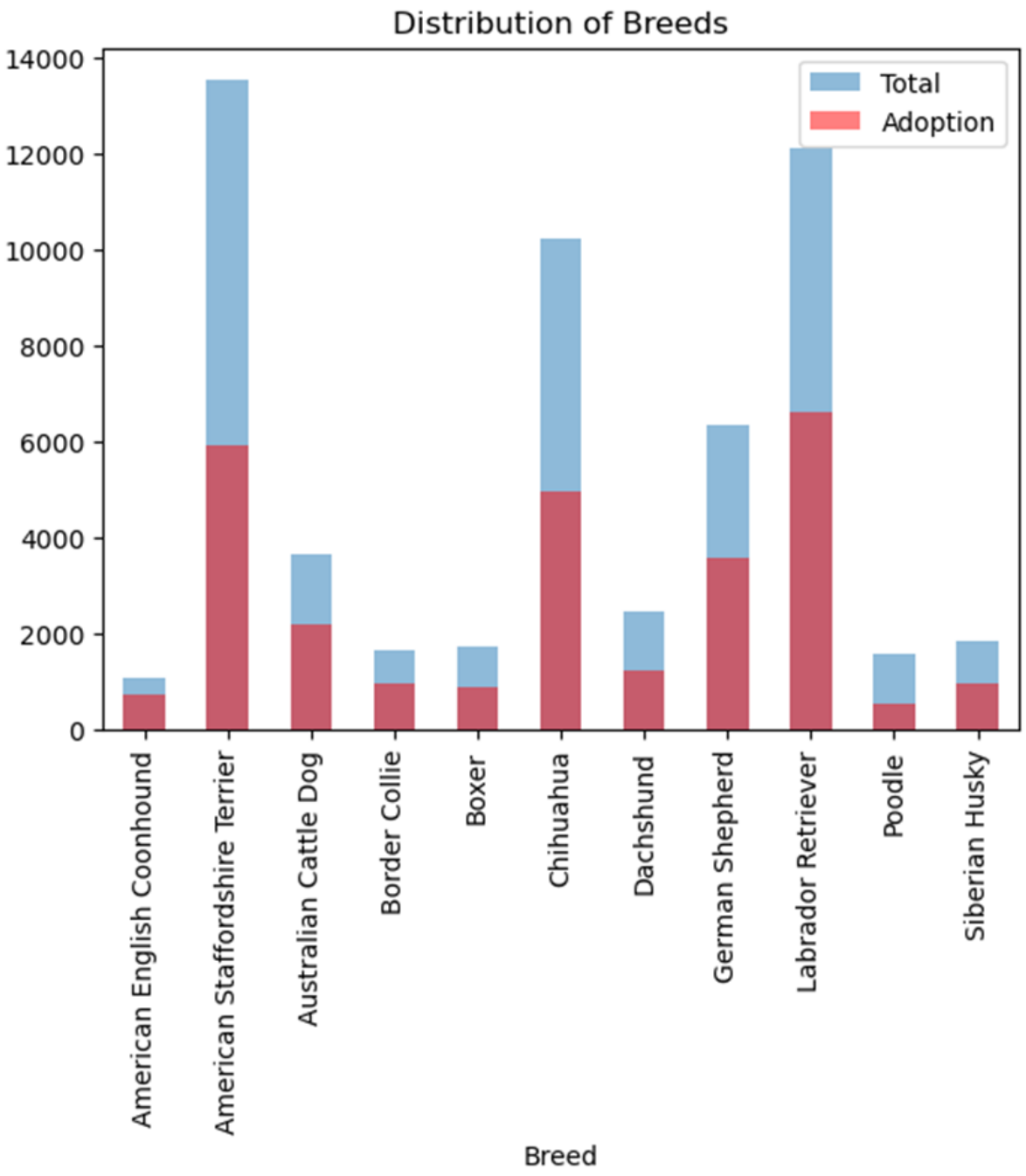
- Breed

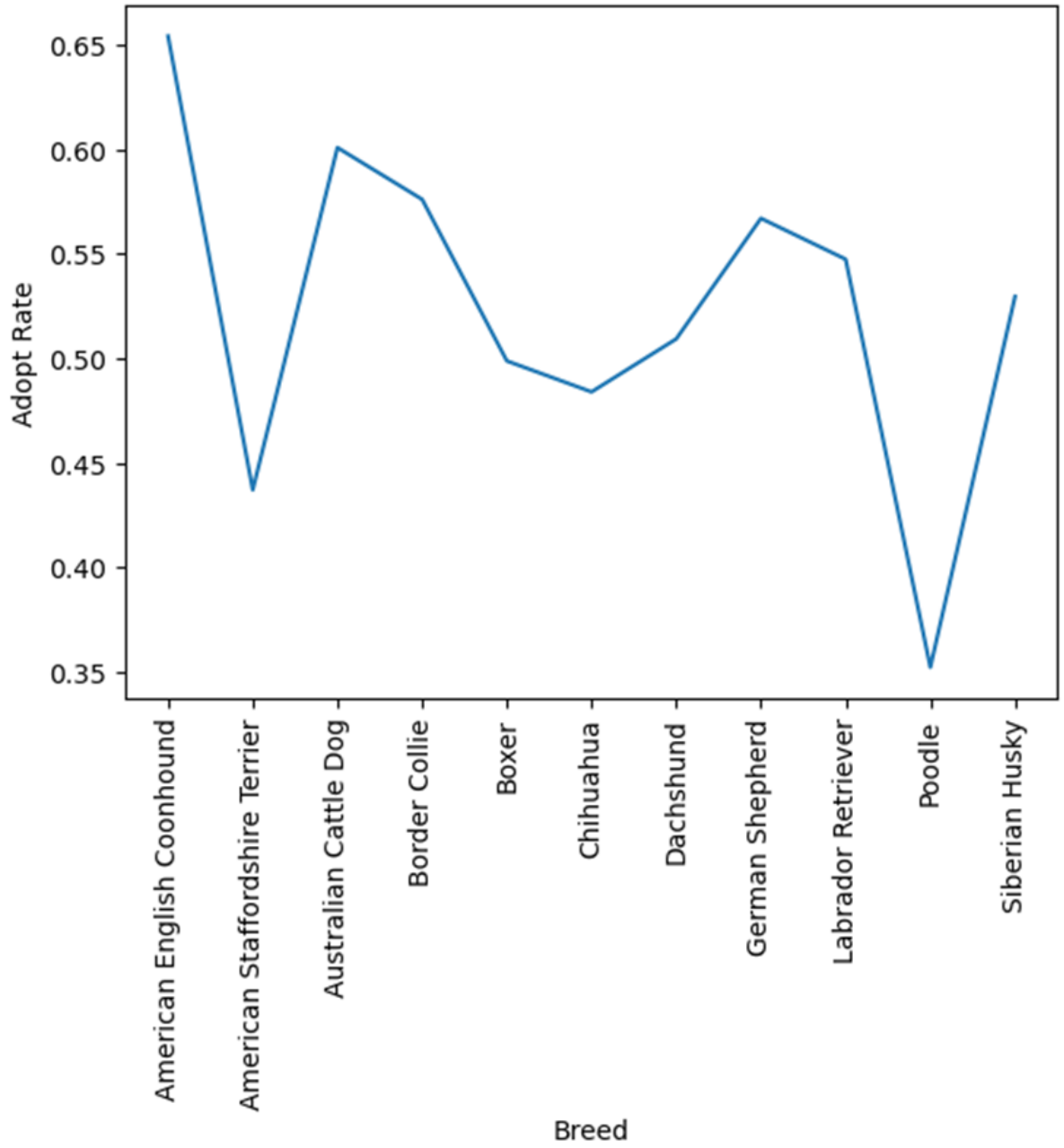
Top 10 breeds of all dogs:

| | Breed | Counts |
|---|---|---|
| 0 | American Staffordshire Terrier | 13512 |
| 1 | Labrador Retriever | 12109 |
| 2 | Chihuahua | 10222 |
| 3 | German Shepherd | 6336 |
| 4 | Australian Cattle Dog | 3646 |
| 5 | Dachshund | 2448 |
| 6 | Siberian Husky | 1837 |
| 7 | Boxer | 1754 |
| 8 | Border Collie | 1661 |
| 9 | Poodle | 1581 |

Top 10 breeds of adopted dogs:

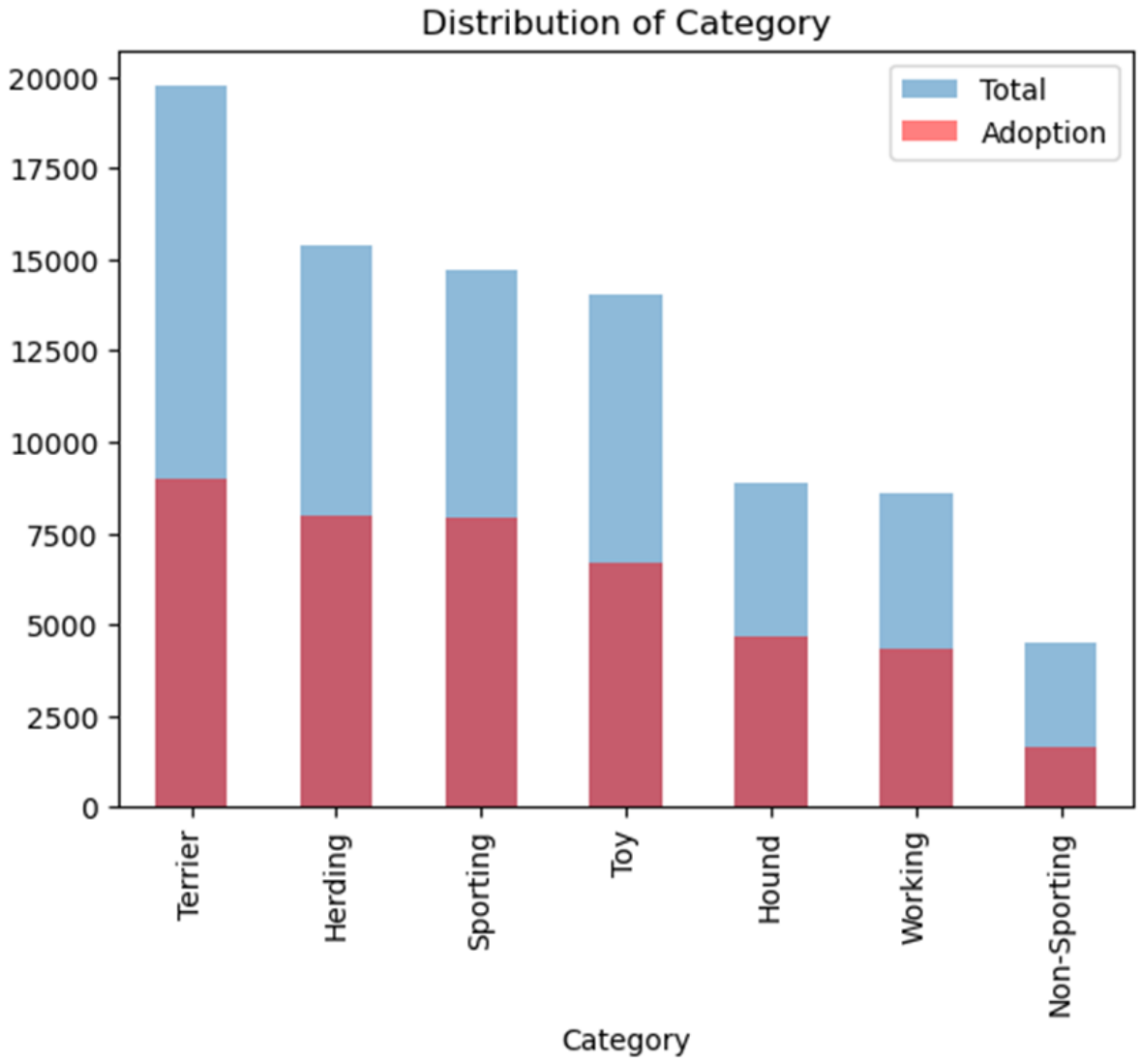
| | Breed | Counts |
|---|---|---|
| 0 | Labrador Retriever | 6630 |
| 1 | American Staffordshire Terrier | 5907 |
| 2 | Chihuahua | 4948 |
| 3 | German Shepherd | 3593 |
| 4 | Australian Cattle Dog | 2191 |
| 5 | Dachshund | 1247 |
| 6 | Siberian Husky | 973 |
| 7 | Border Collie | 957 |
| 8 | Boxer | 875 |
| 9 | American English Coonhound | 721 |


Distribution of Breeds
Total
Adoption
14000
12000
10000
8000
6000
4000
2000
0
American English Coonhound
American Staffordshire Terrier
Australian Cattle Dog
Border Collie
Boxer
Chihuahua
Dachshund
German Shepherd
Labrador Retriever
Poodle
Siberian Husky
Breed

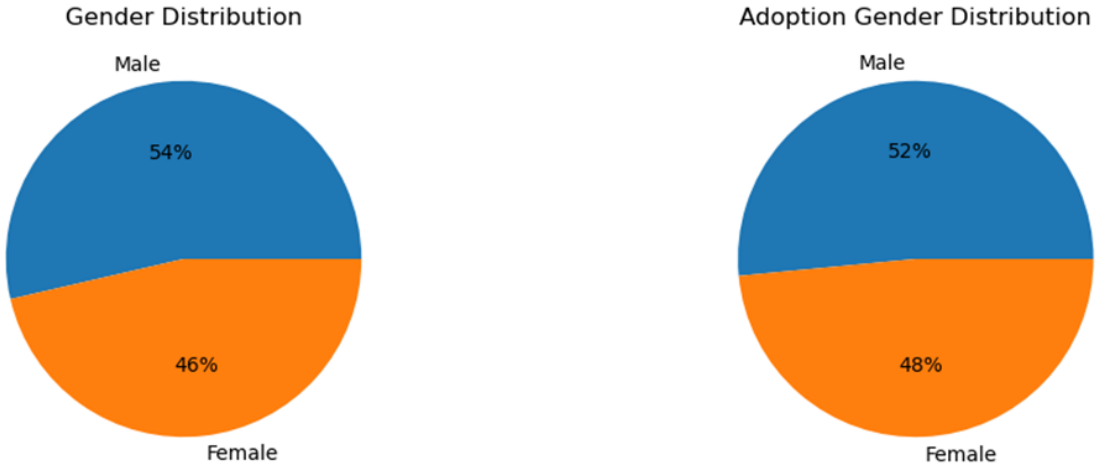



The most popular breeds among all dogs at the Austin Animal Center, which align with the top ten breeds for adoption, include the American Staffordshire Terrier, Labrador Retriever, Chihuahua, and Australian Cattle Dog. Additionally, breeds like the American English Coonhound, Australian Cattle Dog, Border Collie, and German Shepherd are notable for having the highest adoption rates.

- Category

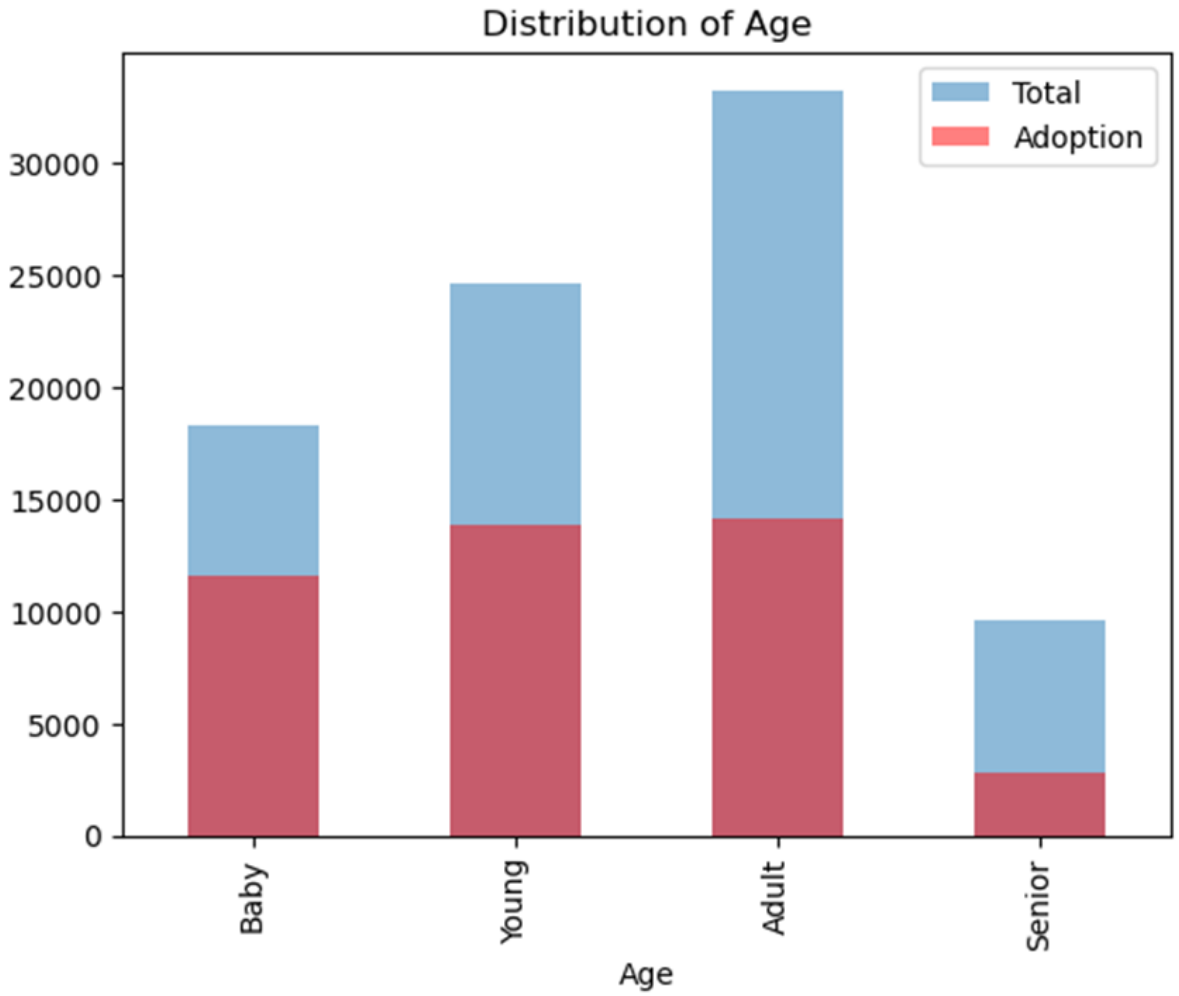


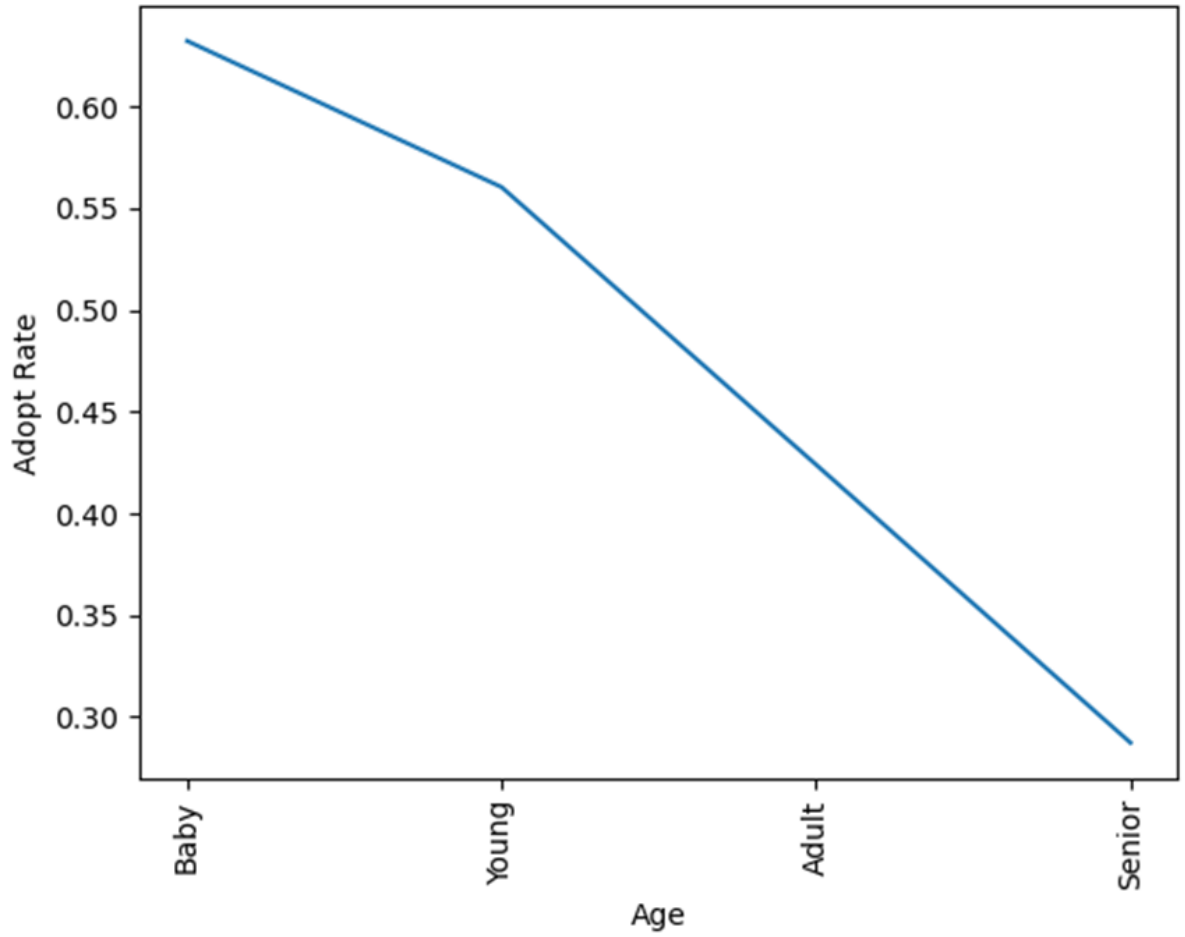

- Gender




There is a higher ratio of female dogs being adopted.

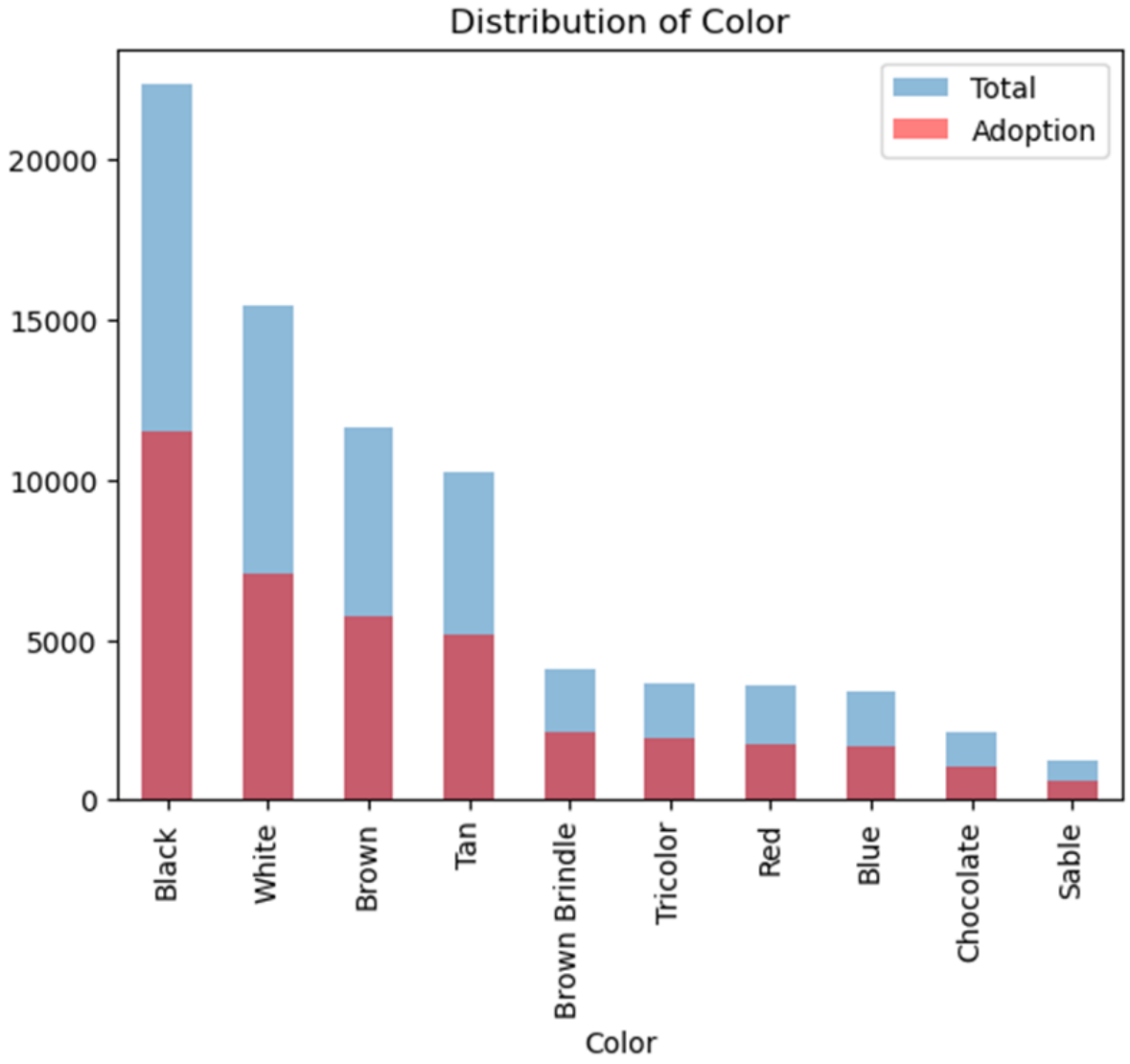
- Age

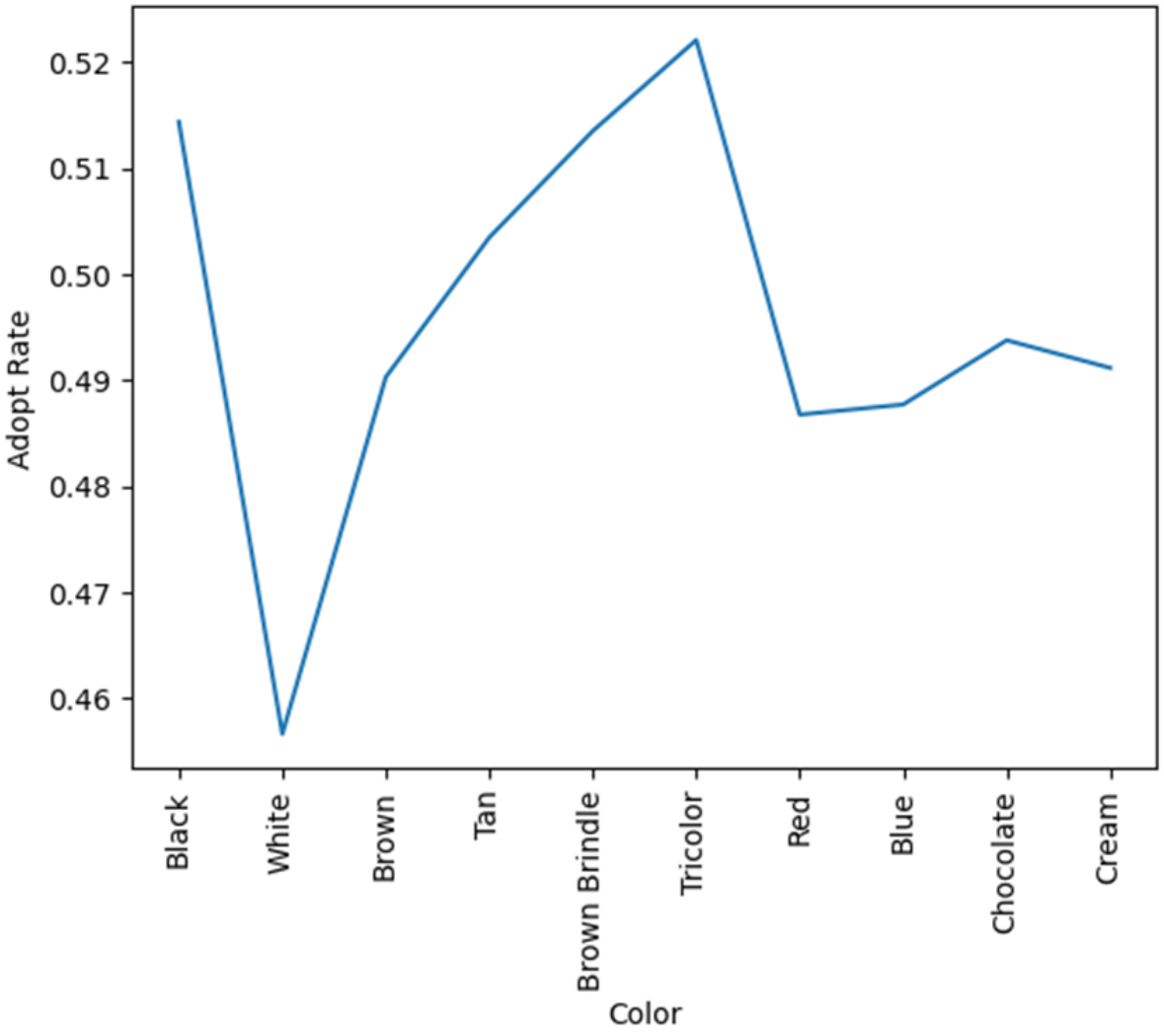

Distribution of Age
Total
Adoption
30000
25000
20000
15000
10000
5000
0
Baby
Young
Adult
Senior
Age




There is a predominance of young and adult dogs, with a clear trend indicating that younger dogs are more frequently adopted.

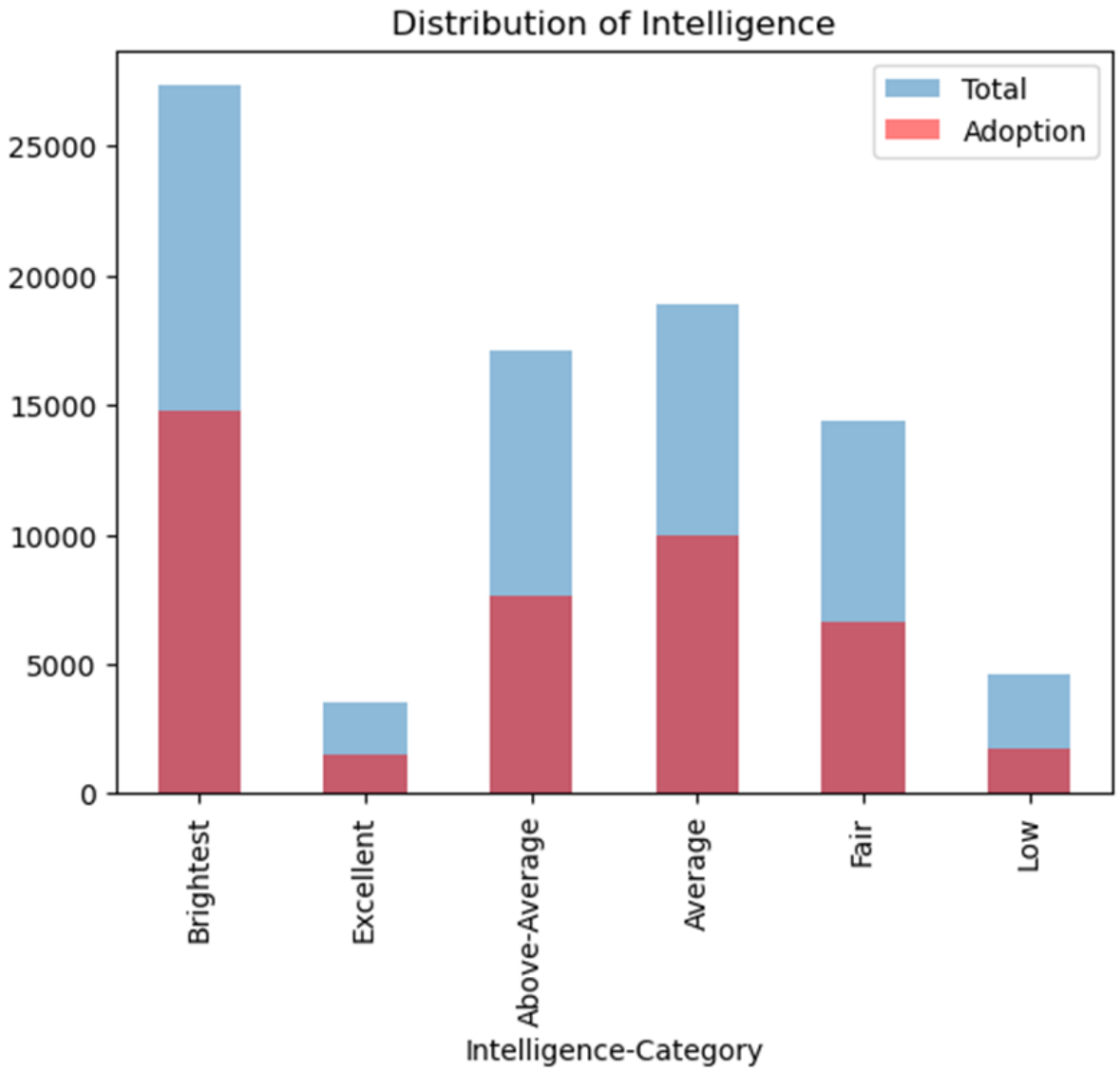
- Color

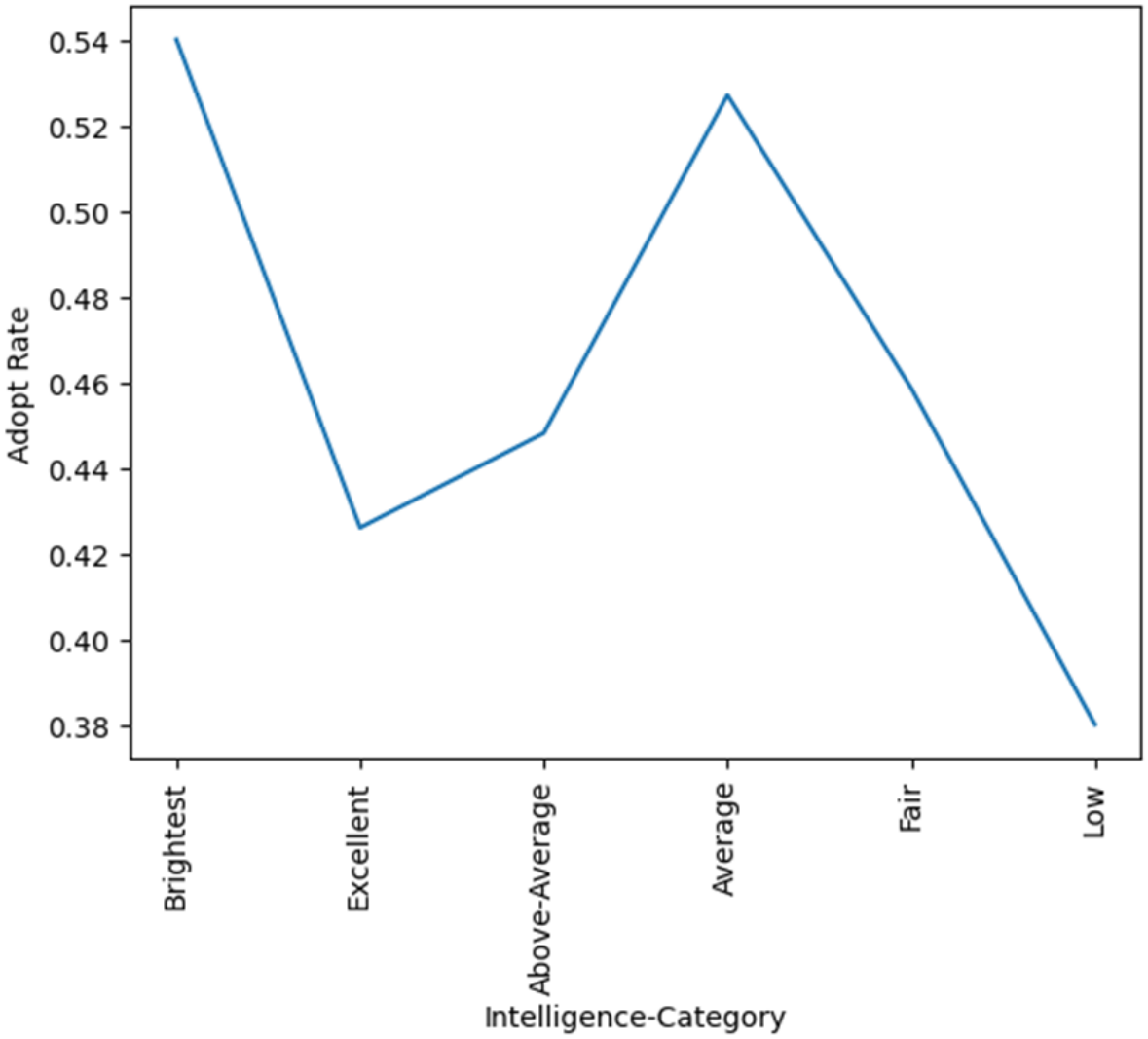

Distribution of Color
Total
Adoption
20000
15000
10000
5000
0
Black
White
Brown
Tan
Brown Brindle
Tricolor
Red
Blue
Chocolate
Sable
Color




Dark and mixed colors like black/tricolor/brindle are preferred.

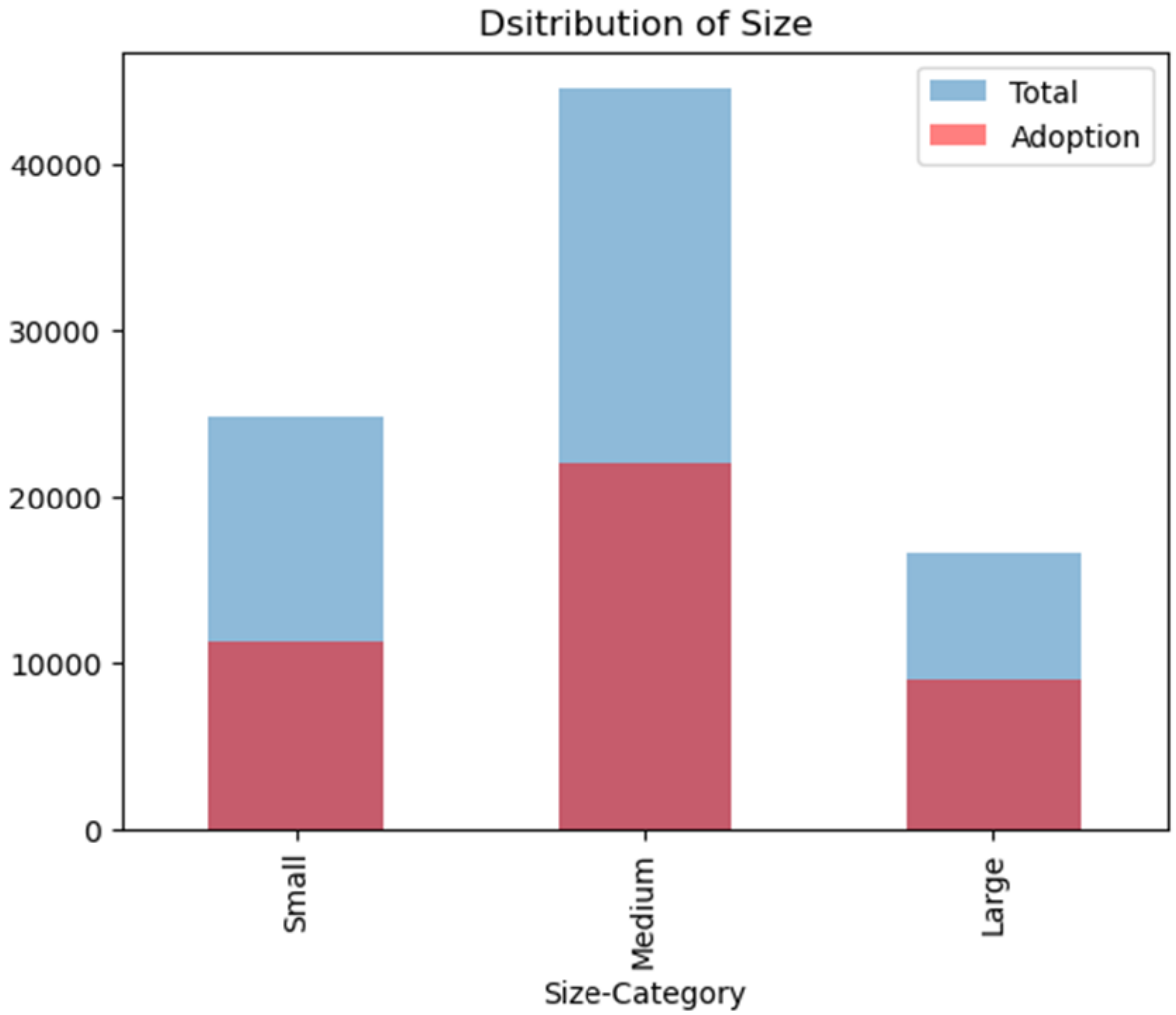
- Intelligence

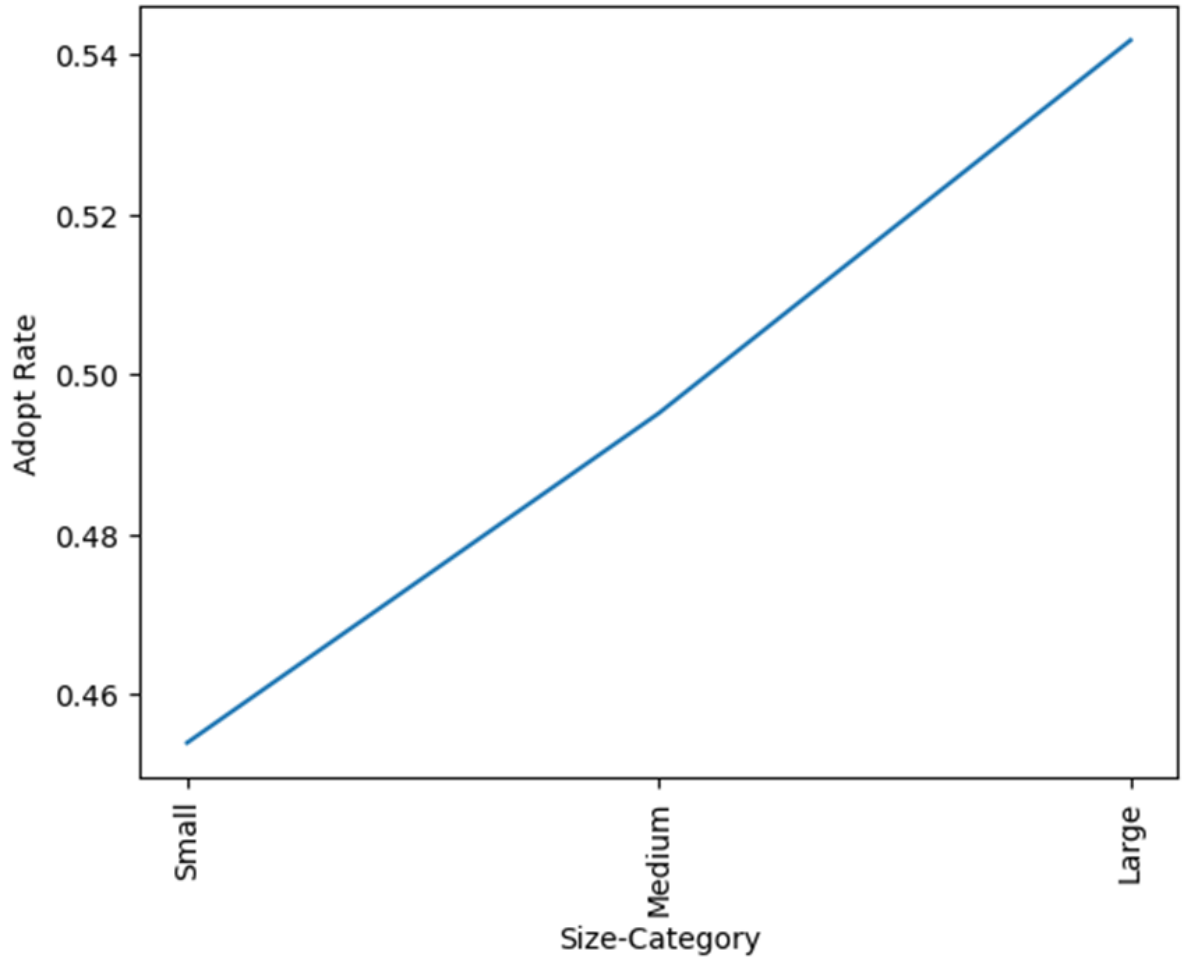





Bright and average intelligence dogs have higher adoption rates, as there is a larger presence of these dogs at the Austin Animal Center. Generally, smarter dogs tend to be preferred for adoption.

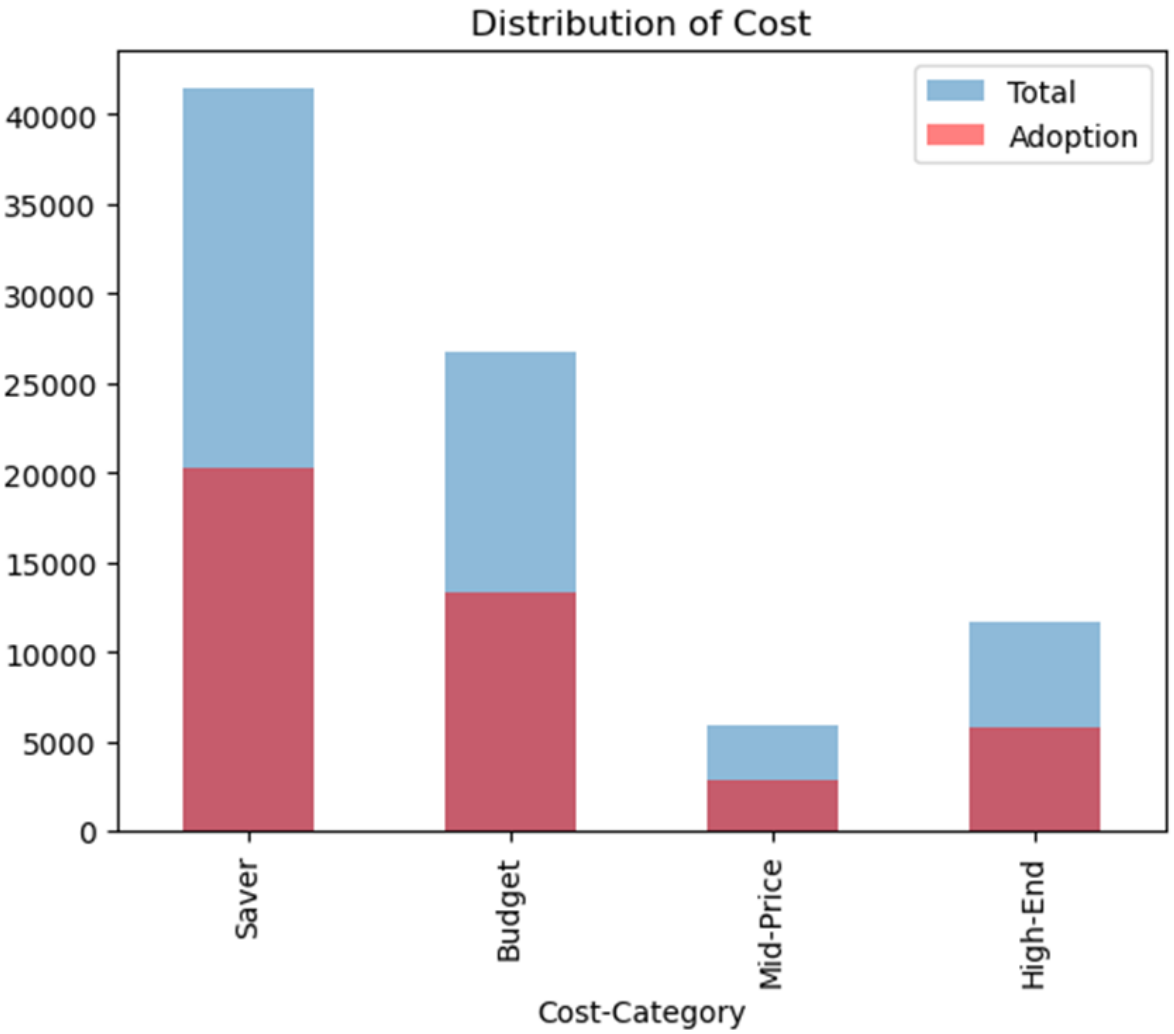
- Size

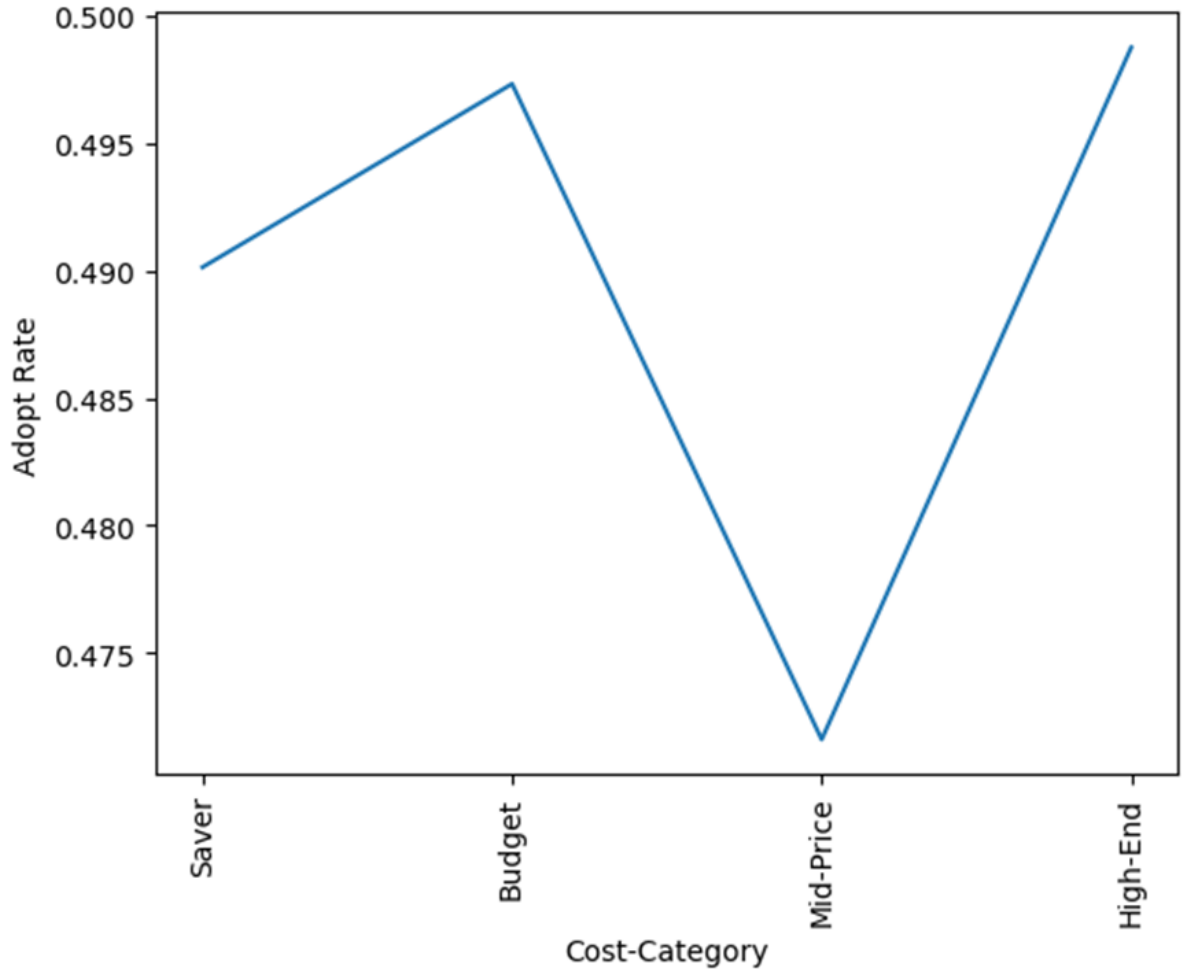

Dsitribution of Size
Total
Adoption
40000
30000
20000
10000
0
Small
Medium
Large
Size-Category




Larger dogs are more likely to be adopted.

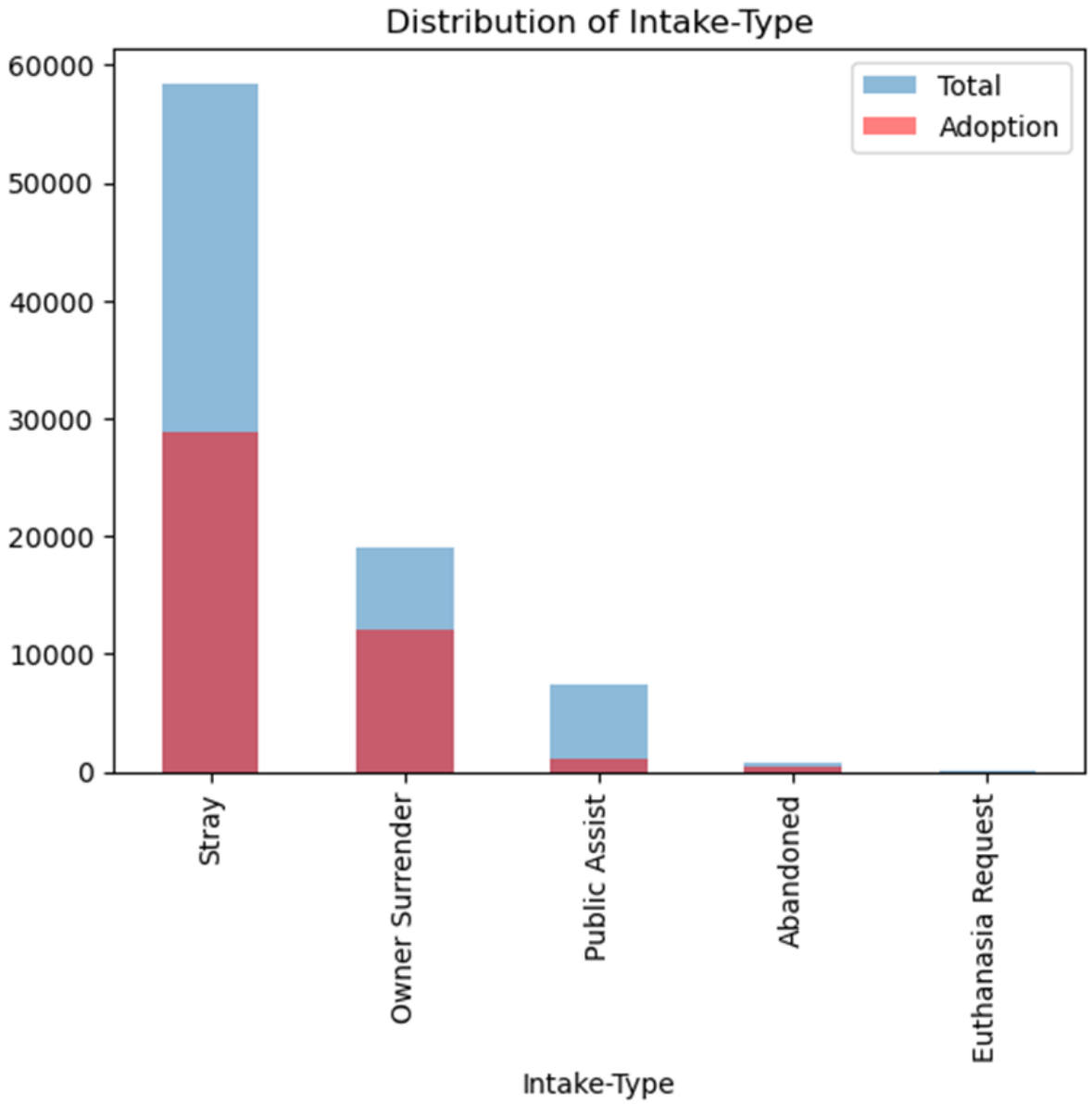
- Cost

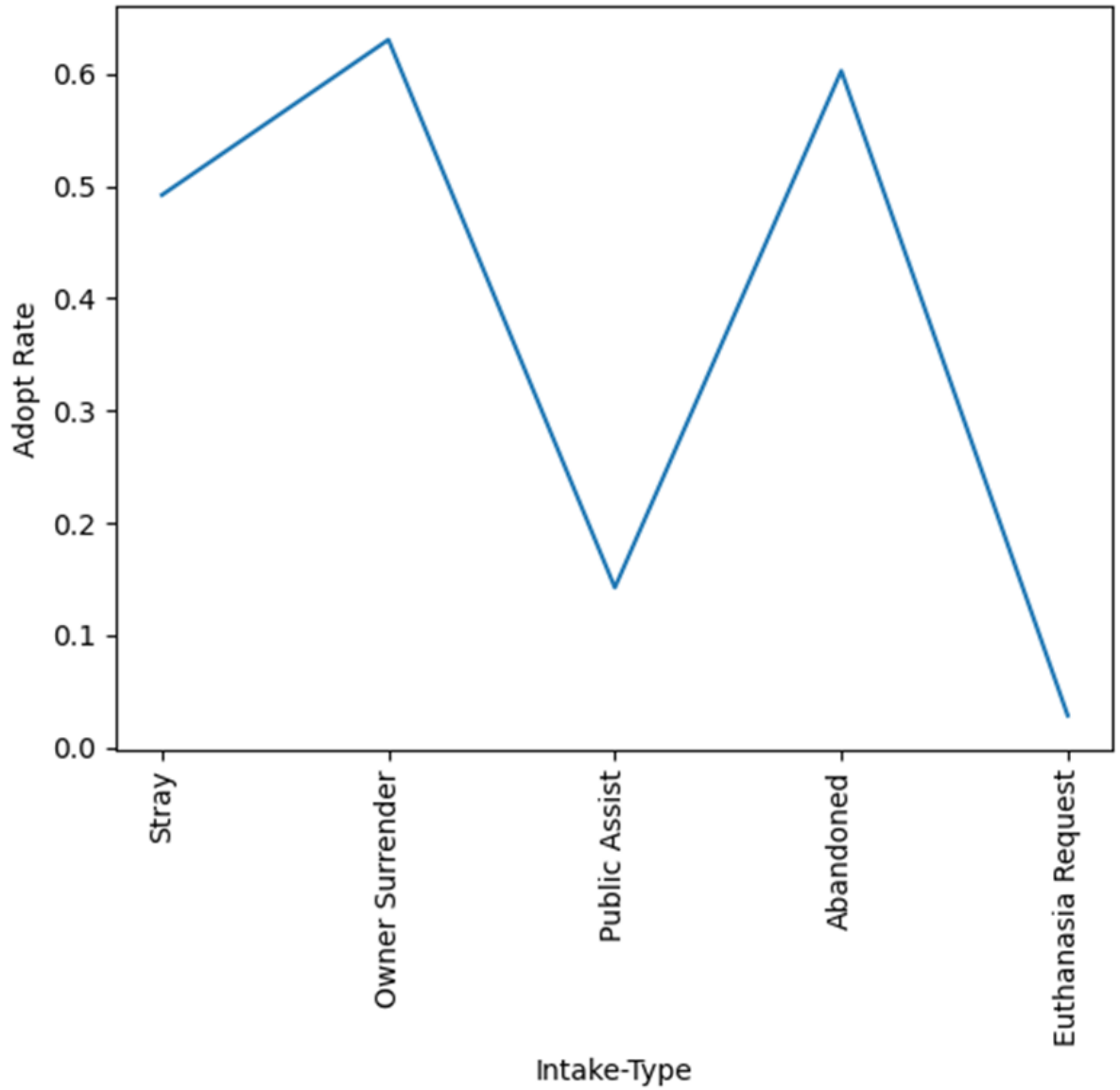

Distribution of Cost
Total
Adoption
40000
35000
30000
25000
20000
15000
10000
5000
0
Saver
Budget
Mid-Price
High-End
Cost-Category




Budget dogs and high-end dogs have higher adoption rates.

- Intake-Type


Distribution of Intake-Type
Total
Adoption
60000
50000
40000
30000
20000
10000
0
Stray
Owner Surrender
Public Assist
Abandoned
Euthanasia Request
Intake-Type




Stray dogs represent the primary reason for dog intakes. Dogs that are surrendered by their owners, abandoned, or stray tend to have higher adoption rates.

1.2.2 Correlation Rules further illuminate preferences for certain dog attributes in adoption decisions.

| | From | To | Confidence | Support | Interest-Factor |
|---|---|---|---|---|---|
| **4** | (Baby) | (True) | 0.632120 | 0.135 | 1.284 |
| **12** | (Young) | (True) | 0.560307 | 0.161 | 1.138 |
| **24** | (Adult) | (True) | 0.423960 | 0.164 | 0.861 |
| **40** | (Terrier) | (True) | 0.454965 | 0.105 | 0.924 |
| **64** | (Large) | (True) | 0.541831 | 0.104 | 1.101 |
| **66** | (Female) | (True) | 0.514601 | 0.239 | 1.045 |
| **85** | (Budget) | (True) | 0.497325 | 0.155 | 1.010 |
| **87** | (Stray) | (True) | 0.492011 | 0.335 | 0.999 |
| **89** | (Brightest) | (True) | 0.540447 | 0.172 | 1.098 |
| **91** | (Medium) | (True) | 0.495178 | 0.257 | 1.006 |
| **135** | (Male) | (True) | 0.472937 | 0.254 | 0.961 |
| **137** | (Saver) | (True) | 0.490134 | 0.237 | 0.996 |
| **139** | (Small) | (True) | 0.453959 | 0.131 | 0.922 |
| **141** | (Black) | (True) | 0.514352 | 0.134 | 1.045 |
| **143** | (Average) | (True) | 0.527434 | 0.116 | 1.071 |

Analysis of correlation rules from the datasets reveals several trends in dog adoptions:

Age Factor - Younger dogs, specifically in the baby/young age categories, show a higher likelihood of adoption.
Breed Preference - Dogs belonging to the Terrier category exhibit a greater propensity for being adopted.
Size Influence - Large dog breeds tend to be more favored in adoption decisions.
Gender Bias - Female dogs appear to have a higher adoption rate compared to male dogs.
Intelligence Perception - Dogs perceived as 'bright' or highly intelligent are preferred for adoption.
Color Selection - Black dogs demonstrate a significant adoption rate, indicating a color preference.

These insights corroborate the findings obtained from statistical analysis and data visualizations, confirming the patterns and preferences observed in the dog adoption process.

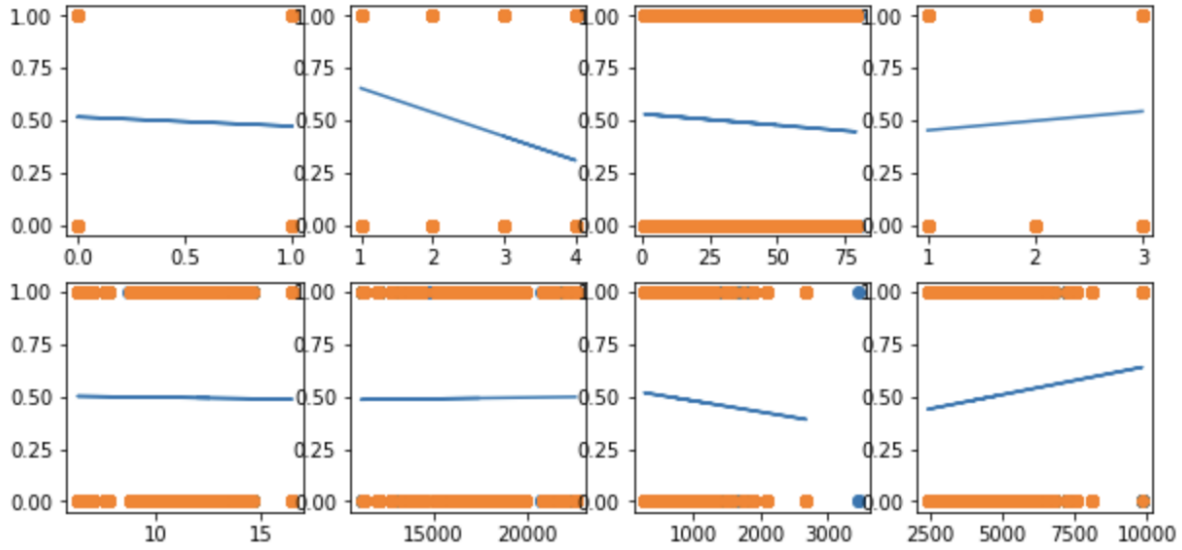
**Insights for Shelter Strategies:**
The analysis presents actionable insights for shelters aiming to optimize adoption strategies and resource allocation:

- Tailoring promotion efforts towards smaller breeds or dogs with high intelligence ratings may expedite their adoption.
- Implementing targeted marketing or reduced adoption fees for breeds with lower raising costs might enhance their adoption rates.
- Prioritizing the adoption of younger or smaller dogs could potentially reduce shelter overcrowding.

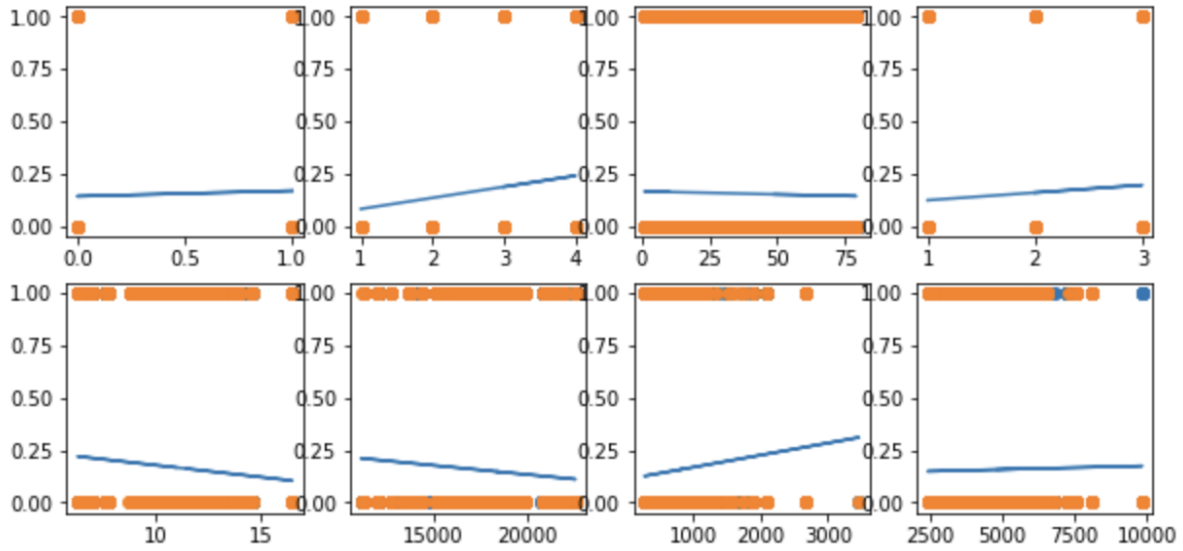
2. Linear Regression Model

We select 8 features (*gender, intelligence, size, age, longevity, total-cost, food-cost, purchase-cost*) with training and test sets (*80, 20*) to fit the linear regression model for 2 target features (*IsAdopted and ReFoundByShelterAfter1stAdoption*).

linear regression model for IsAdopted:




linear regression model for ReFoundByShelterAfter1stAdoption:

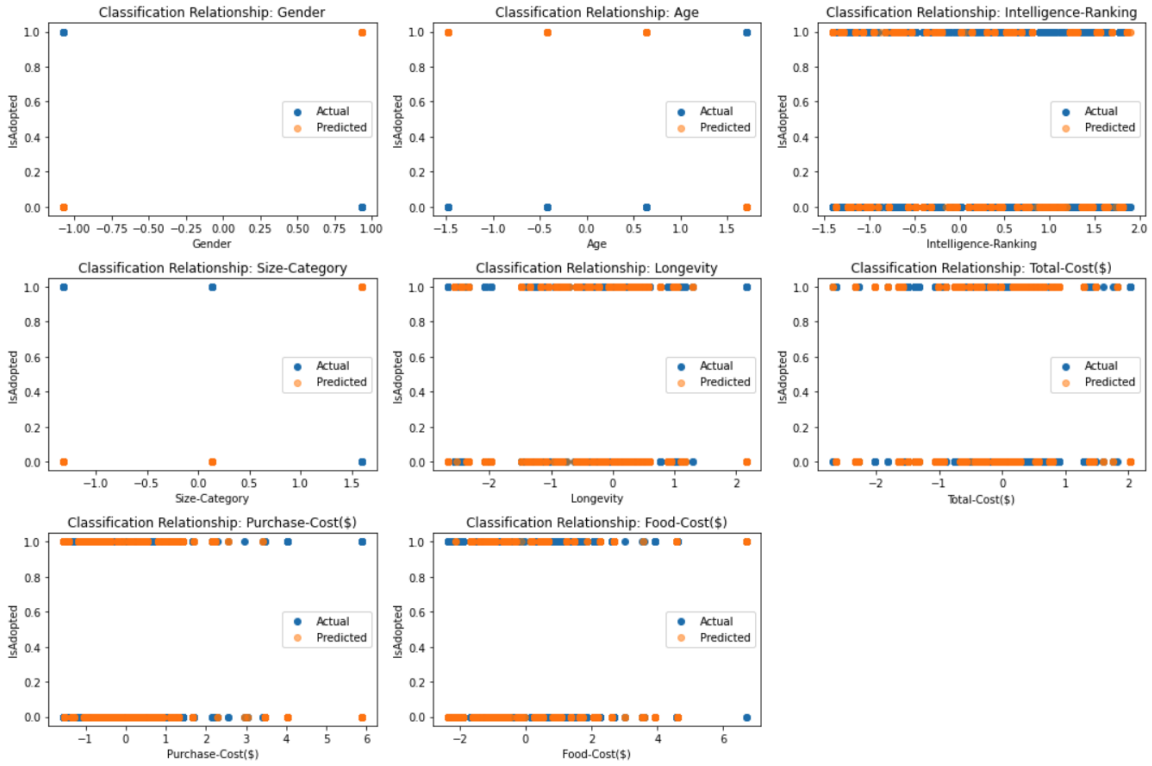


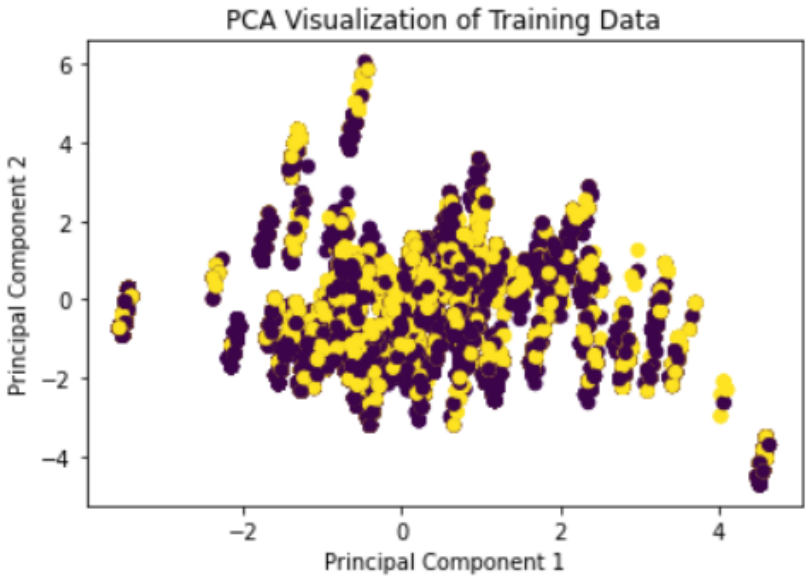

From the linear regression model, we can see:
1. Gender, Longevity, Total-Cost do not affect if the pet will be adopted. The line is quite flat.
2. Intelligence-Ranking, Size-Category, Purchase-Cost have tiny effects on adoption. Intelligence-Ranking and Purchase-Cost have negative effects, meaning if a pet is less-intelligent and costs more, it has less chance to be adopted. Size-Category has the opposite positive effect.
3. Age and Food-Cost have a dramatic influence on adoption. If a pet is younger and costs less on food, it has a much bigger chance to be adopted.
4. Age and Purchase-Cost have tiny effects on if a pet is re-found. Aging and more purchase-cost pets seem to have a trend to be re-found more than one time.
5. Age is the most predictive for adoption and re-found. It has a low average of mean squared error and high average of variance score.
7. The model performs much better with all features rather than any individual ones. This quite makes sense since the model is better trained. Some features fit the model quite well though some others do not have a quite relatively linear relationship.

3. Nearest Neighbor (kNN) Classifier Model

We select the same 8 features (gender, intelligence, size, age, longevity, total-cost, food-cost, purchase-cost) with training and test sets (80, 20) to fit the kNN classifier model for 2 target features (IsAdopted and ReFoundByShelterAfter1stAdoption).

- Nearest Neighbor (kNN) Classifier Model for IsAdopted:

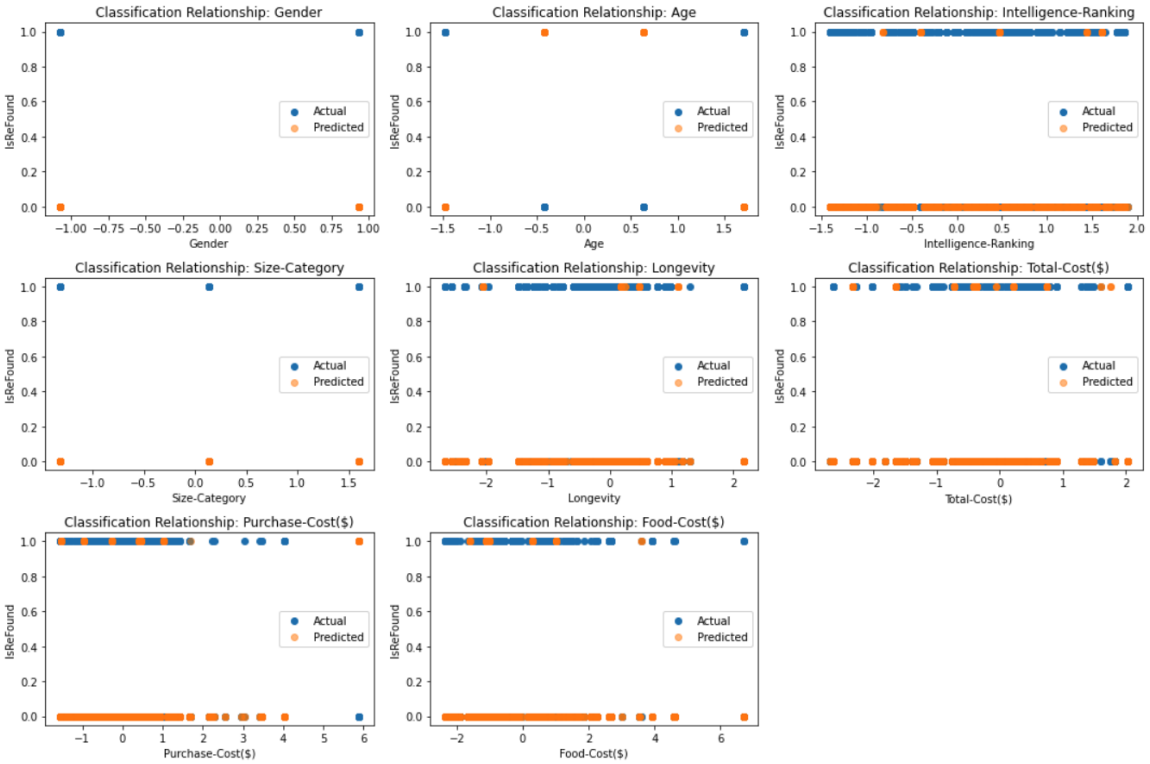


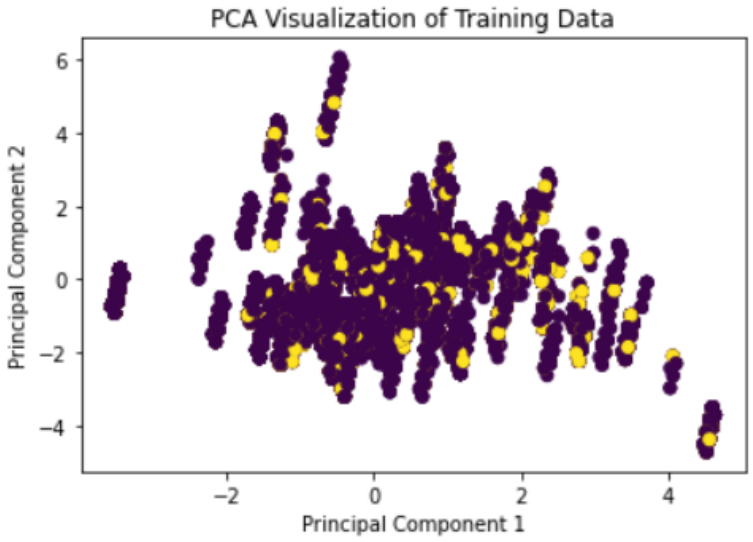




From the evaluation of the kNN model results, we know that:

1. Age, Gender, Size, Total-Cost do not affect if the pet will be adopted.
2. Total-cost, Intelligence level and Food-Cost have a high influence on adoption.
3. The model performs much better with all features rather than any individual ones. This quite makes sense since the model is better trained for classification.
4. The PCA plot suggests that adoption is a complex decision that can't be easily predicted by one or two features alone.

- Nearest Neighbor (kNN) Classifier Model for ReFoundByShelterAfter1stAdoption:




From the evaluation of the kNN model results, we know that:

1. Age, Purchase cost, Food cost, Total-Cost do not affect if the pet will be refound by shelter.
2. Size, Gender, Longevity and Intelligence level have a high influence on whether this dog will be refound by shelter.
3. The model performs much better with all features rather than any individual ones. This quite makes sense since the model is better trained for classification.
4. The PCA plot suggests that to be refound by shelter is a complex decision that can't be easily predicted by one or two features alone.

Conclusion:

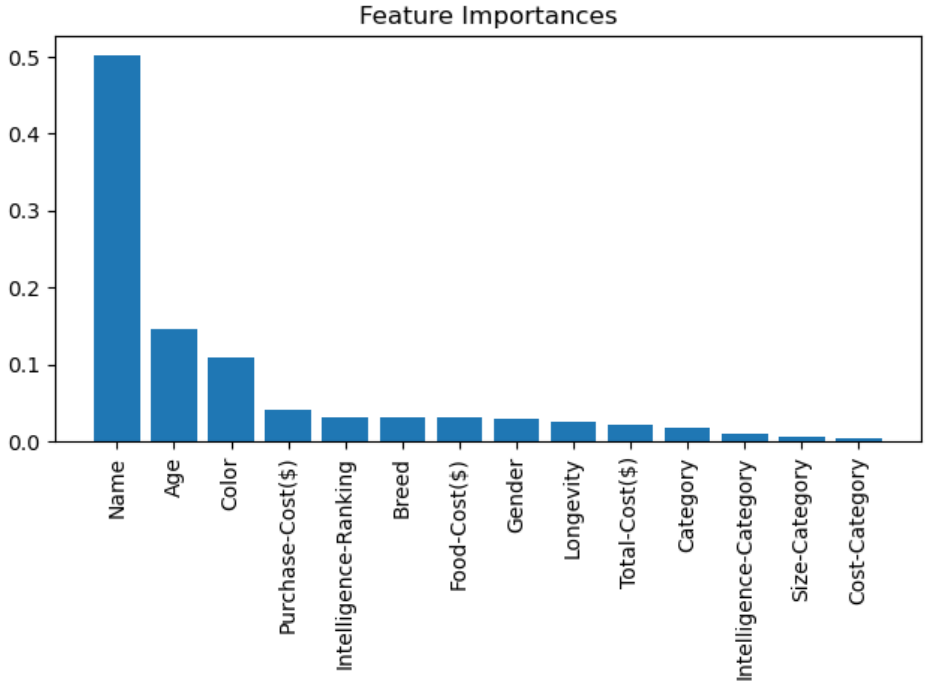
In conclusion, the k-Nearest Neighbor (kNN) Classifier analysis sheds light on the factors influencing both dog adoption and the likelihood of being refunded to the shelter. The model emphasizes that age, purchase cost, food cost, would impact the adoption decision. At the same time, features such as size, gender, longevity, and intelligence level play a crucial role in determining whether a dog will be refunded to the shelter.

The evaluation of the kNN model underscores the importance of considering a comprehensive set of features, as opposed to individual ones, to achieve better predictive performance. The complexity of both adoption and refunding decisions is highlighted by the PCA plot, emphasizing the intricate nature of these processes that cannot be easily distilled to one or two isolated features. As such, the model suggests that addressing factors like size, gender, longevity, and intelligence level is essential for understanding and potentially influencing the outcomes, be it successful adoptions or dogs being refunded to the shelter.

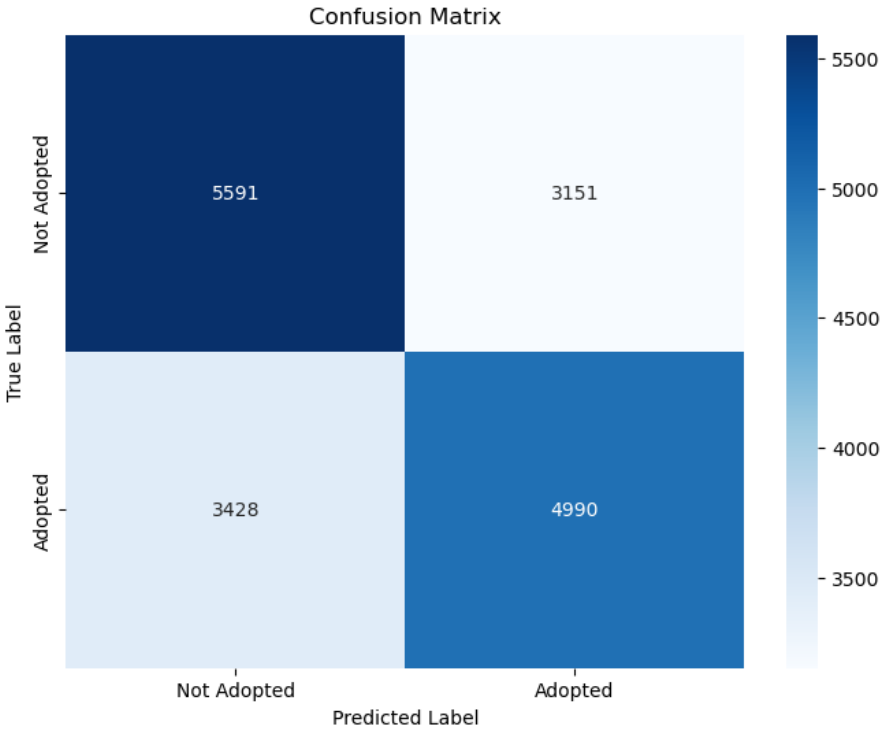
4. Decision Tree Classifier Model

We select 13 features (age, color, purchase cost, intelligence-ranking, breed, food-cost, gender, longevity, total-cost, category, intelligence-category, size-category, cost-category) with training and test sets to fit the decision tree classifier model for IsAdopted.

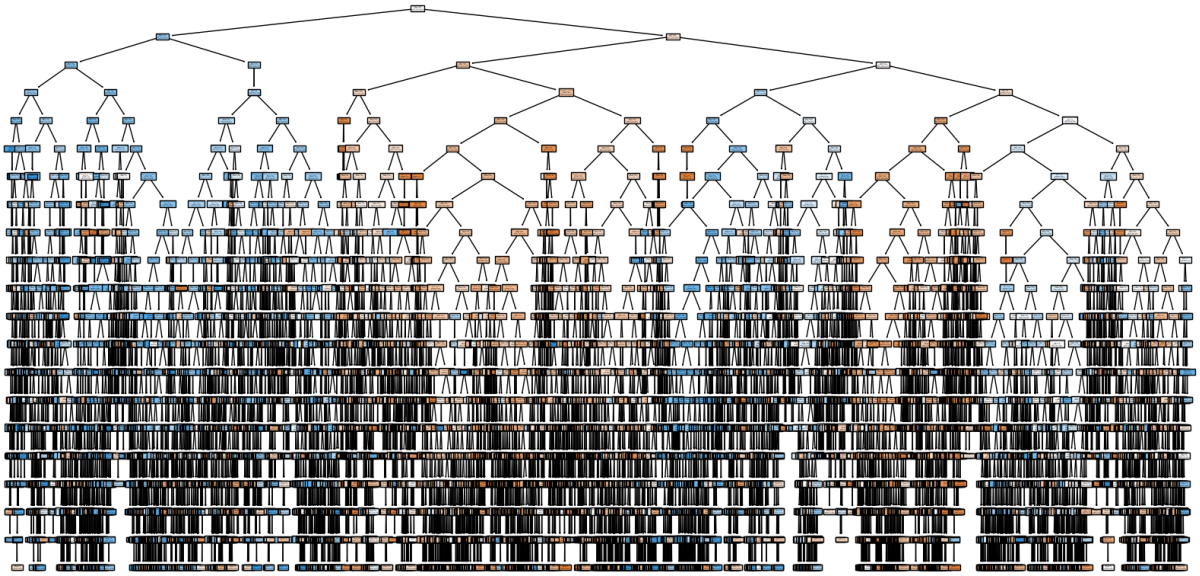
For the IsAdopted features, we rank some features in the descending sequence.




This is the confusion matrix for evaluating how accurate the model is.




This is the decision tree model we build.



Conclusion for IsAdopted:
The decision tree model highlights that age and color are pivotal in determining dogs' adoption from the shelter. Other features like purchase cost, intelligence ranking, breed, food cost, gender, longevity are also impacting the adoption decision. Younger dogs and certain colored dogs are favored. The model suggests prioritizing healthcare and promoting long-term residents to improve adoption rates.

# Discussion

# 1. Adoption Patterns

## 1.1. Adoption Trends

During summer and winter holidays, when people often go on vacations, especially those with children, there's a tendency to delay pet adoption until after these periods. This is a practical decision, considering the commitment involved in taking on a new pet. Consequently, animal shelters, including the Austin Animal Center, typically see an increase in their populations during these months.

In terms of the annual trend in dog adoptions, there was a notable decrease in the total number of dogs from 2020, likely due to the impact of the COVID-19 pandemic. However, paradoxically, the pandemic also spurred a surge in pet adoptions. As people found themselves quarantined and isolated, the demand for companionship led to a significant increase in pet adoption, with many seeking the emotional support and connection that pets provide.

Regarding breed-specific adoption trends, there has been a noticeable decline in the popularity of Chihuahuas. While they are cherished by some, Chihuahuas may not be as sought-after as other breeds for several reasons, such as their small size and the perception of their temperament. Some people may prefer larger dogs, and Chihuahuas are known to be one of the smallest breeds. Additionally, they can sometimes be high-strung, nervous, and prone to anxiety, traits that might not align with the preferences of all potential adopters.

## 1.2. Adoption Influencers

Younger dogs are often preferred by adopters. A common reason for this preference is the desire to bond with the dog from its puppyhood and to ensure proper socialization and training from an early age.

Big dogs are also favored by many. These larger breeds are known for their quick learning abilities, making them adept at picking up various tricks and behaviors. Despite some misconceptions about their temperament, big dogs like boxers and bullies are often more emotionally stable, well-mannered, and child-friendly compared to some smaller breeds.

Female dogs tend to be adopted more frequently due to the myth that they are calmer, easier to train, friendlier, and easier to control. Their generally smaller size compared to males can also make them more appealing, as they are perceived to be less costly and easier to handle.

Black dogs hold a special appeal as well. Breeds like Labrador Retrievers, Spaniels, Shepherds, and Terriers, which are often highly intelligent, may have black fur. Their appearance is seen as versatile, looking great with various accessories and in any color. The preference for smarter dogs is understandable, as they are often seen as easier to train and more responsive.

The cost of owning a dog varies significantly based on breed, location, potential health issues, and other factors. Budget-friendly dogs are more likely to be adopted due to the lower costs associated with their care. However, higher-end breeds might also be popular

due to their other attractive characteristics, despite the greater financial commitment they entail.

**Association between Dog Characteristics and Adoption Rate**

The revealed associations between dog attributes and adoption rates align with certain intuitive expectations while providing novel insights into adoption dynamics. The prolonged shelter stays for larger breeds seem logical, as their adoption might require more space or specific living conditions. Similarly, the positive correlation observed between higher intelligence ratings and increased adoption likelihood aligns with potential adopters' preferences for trainable and intelligent dogs. Moreover, the link between lower raising costs and higher adoption rates suggests that affordability plays a role in adoption decisions, highlighting an actionable area for shelters to consider in their adoption strategies.

**Predictive Modeling for Shelter Duration**

The predictive model's ability to estimate the duration a dog spends in a shelter before adoption exhibits practical implications for shelter management. The identified factors, such as age, size, and breed type, influencing shelter durations, resonate with the challenges shelters face in accommodating and finding suitable homes for different dog categories. The shorter stay durations for younger and smaller dogs imply potential adopters' preferences for specific age and size categories, influencing adoption timelines.

**Prediction of Adoption Likelihood**

The efficacy of classification models, the Decision Tree and kNN Classifier, in predicting adoption likelihood signifies the relevance of various dog features in determining adoption outcomes. The Decision Tree's ability to discern patterns based on breed, age, and behavior underscores the importance of these attributes in adopters' decision-making processes. Additionally, the kNN Classifier's reliance on feature similarity in predicting adoption probabilities indicates that dogs with similar traits are likely to have comparable adoption chances, providing insights into adopters' preferences for specific dog characteristics.

**Insights for Shelter Strategies**

The actionable insights derived from the analysis offer tangible strategies for shelters seeking to improve adoption rates and alleviate shelter overcrowding. Tailoring promotion efforts or reducing fees for certain dog categories based on size, intelligence, or raising costs could potentially expedite their adoption. The focus on younger or smaller dogs aligns with the observed shorter shelter stays for these categories, suggesting a need for proactive measures to expedite their rehoming.

In conclusion, the numerical outcomes and identified patterns from our analysis provide a nuanced understanding of dog adoption dynamics. These findings offer shelters practical strategies to enhance their adoption processes, effectively allocate resources, and ultimately improve the welfare of sheltered dogs.

# Future Work

While our study endeavors to provide valuable insights into dog adoption dynamics, there remain avenues for future exploration and enhancement:

**Enhanced Predictive Models**
Refinement of existing predictive models by incorporating more nuanced features.
Exploration of advanced machine learning algorithms for improved accuracy.

**Real-time Prediction Models:** Developing real-time predictive models that continuously update based on new intake data could offer more dynamic and accurate estimations of adoption likelihood and shelter durations.

**Collaboration with Shelters:** Collaborating directly with animal shelters to implement and validate the models in real-world scenarios, gathering feedback, and refining the models based on practical insights and experiences

# Conclusion

In conclusion, our study endeavors to shed light on the intricate dynamics of dog adoption by leveraging comprehensive datasets and data mining techniques. Through our analysis, we identified key factors influencing adoption rates, including dog size, intelligence, and raising costs. Our predictive models demonstrated efficacy in estimating adoption probabilities and shelter durations, offering valuable insights for shelters to optimize their strategies and resource allocation.

However, our study is not exhaustive and leaves room for further exploration and refinement. Despite the achievements in predictive modeling and insights gained, there exist opportunities for future work, such as delving into enhanced predictive models, real-time prediction models, and practical implementation in collaboration with shelters.

Ultimately, our study represents a significant step toward understanding and enhancing dog adoption processes. The insights and models developed pave the way for future endeavors aiming to improve animal welfare, optimize shelter operations, and facilitate successful dog adoptions.

# References

Our project gathered data from multiple sources to conduct a comprehensive analysis of dog adoption dynamics. The following datasets were utilized:

**Data Sources**

Austin Animal Center Intakes & Outcomes Datasets:

- Austin Animal Center Intakes: Link
- Austin Animal Center Outcomes: Link
- These datasets were obtained from the data.austintexas.gov website, providing comprehensive records of animal intakes and outcomes from the Austin Animal Center.

Dog Breeds Dataset from Best-In-Show:

- Dog Breeds dataset: Link
- This dataset was sourced from the Best-In-Show website, offering valuable insights into various dog breeds and serving as supplementary information for our analysis.

Petfinder Open API:

- Petfinder Open API Documentation: Link
- The Petfinder Open API was utilized to access real-time data on adoptable dogs available on the Petfinder website. This API enriched our analysis by providing up-to-date information on active dogs for adoption.

**Interpretation sources**

https://communitynewspapers.com/doraltribune/why-the-animal-shelter-population-soars-during-summer-months/

https://www.rover.com/blog/pandemic-pet-adoption-boom/

https://www.townandcountrymag.com/society/tradition/news/a2996/10-reasons-big-dogs-are-the-best-dogs/

https://be.chewy.com/pet-parenting-pet-lovers-13-reasons-to-love-black-dogs/

https://www.caninejournal.com/cost-of-owning-a-dog/

# Annex-A: Individual Contributions

Bula Ha:

- Contributed to the interpretation of results and discussion sections.
- Write the final project proposal and final project report
- Assisted in the final presentation PowerPoint.

Minjie Shen:

- Focused on kNN model creation and implementation
- Used kNN model to predict the isAdopted and reFoundByShelter feature
- Contributed to the final report and presentation of relevant parts

Yuqing Bian:

- Focused on data collecting; data pre-processing and standardization; adoption patterns analysis.
- Employed data statistical & visualization methods and Association Rules to analyze adoption patterns.
- Contributed to the final report and presentation of relevant parts.

Qingjuan Dan:

- Implemented the Decision Tree classifier for predicting adoption likelihood.
- Assisted in the presentation.

Tianyiru Chen:

- Led the overall supervised learning developments, covering the original studies of adoption prediction and refound-after-adoption prediction.
- Developed and implemented linear regression models (code, result and discussion).
- Organized coding files.
- Contributed to the final report and presentation of relevant parts.

This summary encapsulates the key contributions made by each team member, showcasing their specific roles and responsibilities within the project. It highlights the diverse skill sets utilized and the collective efforts made towards achieving the project objectives.