Team Iris Final Presentation

Understanding and Predicting Dog Adoption Dynamics

Bula Ha, Minjie Shen,Yuqing Bian, Qingjuan Dan, Tianyiru Chen

12/2023

Introduction of our team

Meet The Team and Roles
Bula Ha,
Minjie Shen,
Yuqing Bian,
Qingjuan Dan,
Tianyiru Chen

The Roles of Our Members

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.
- Organized the final report and PPT files.

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.
- Contributed to the final report and presentation of relevant parts.

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.

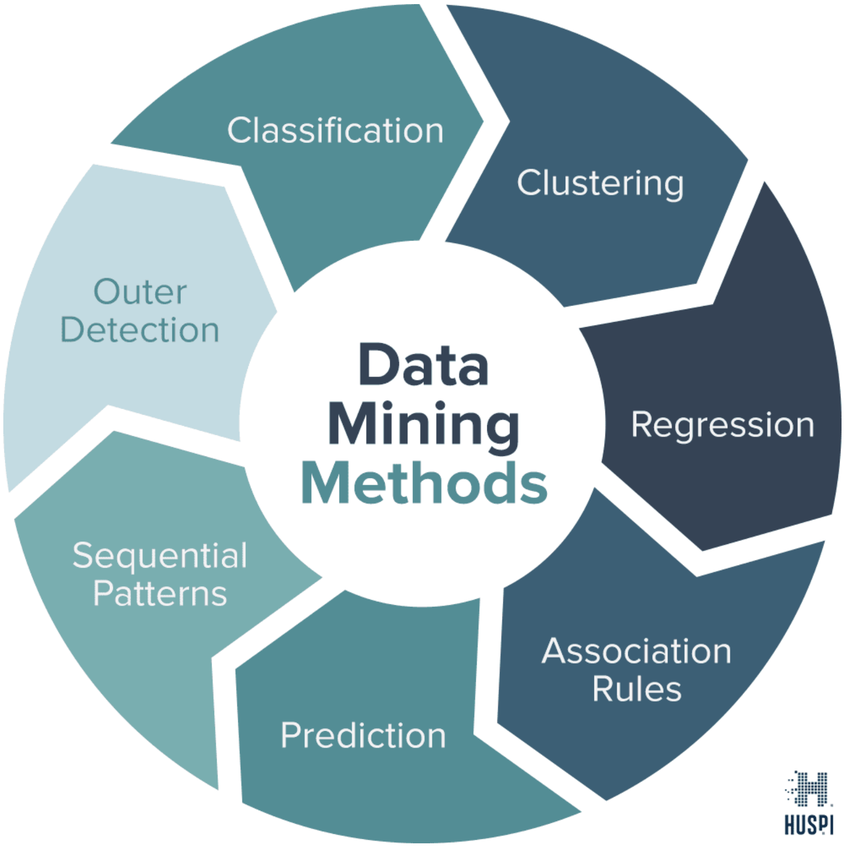
Project Overview

- Project Focus: Understanding Factors Impacting Dog Adoption
- Datasets: Austin Animal Center, Petfinder API, Dog Breed Information
- Objective: Develop predictive models for adoption probabilities and shelter stay durations

Project Significance

Understanding the intricacies of dog adoption is vital for shelters and animal welfare organizations. Our project aims to provide in-depth insights that can optimize adoption strategies, refine resource allocation, and potentially mitigate shelter overcrowding, thus reducing associated costs and enhancing operational efficiency.

Methodologies



- Data Collection and Cleaning
- Data Preparation and Normalization
- Unsupervised Learning Techniques(Association Rules)
- Supervised Learning: Regression(Linear Regression) and Classification(Decision Tree, k-Nearest Neighbors (kNN) Classifier)

Code Review and Findings

Data Pre-processing and Standardization

Minimizing Data Volume
- Instance Reduction
- Feature Reduction

Data Transformation
- Aggregation
- Generalization
- Normalization
- Feature Construction

Data Cleaning
- Addressing Missing Values
- Removing outliers
- Correcting Inconsistent Data
- Removing Duplicate Data

Data Integration

Code Review and Findings

Analysis of Adoption Patterns

Trends in Adoption

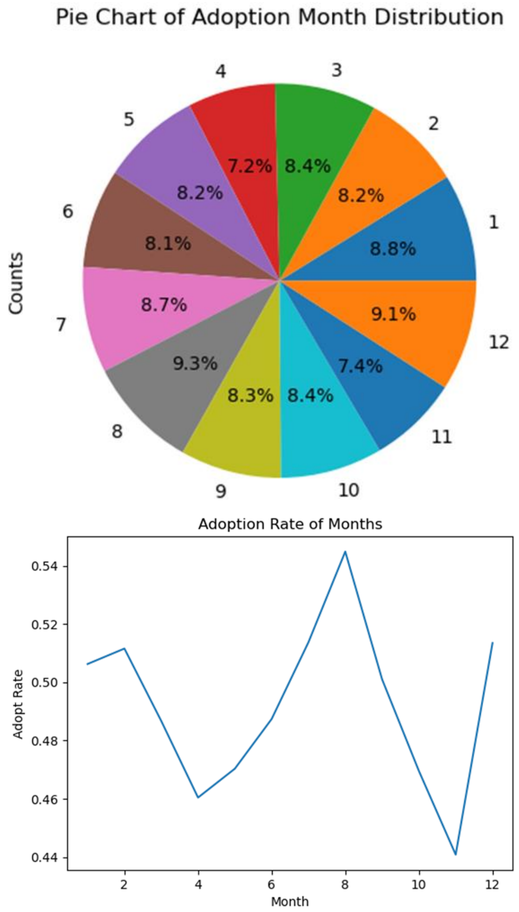
- Analyzed a decade's data (2013-2023), examining seasonal and annual adoption rate trends, and breed preferences trends.

Influencing Factors of Adoption

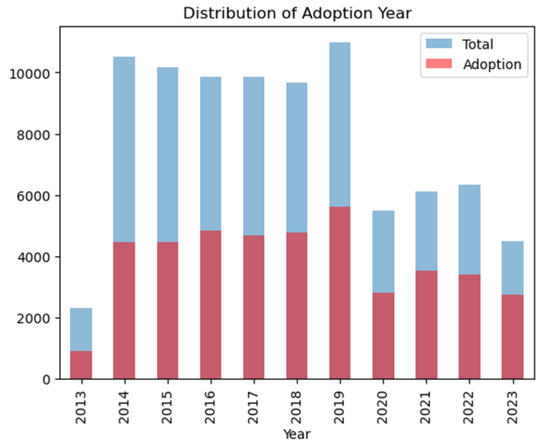
- Utilized data visualization tools and statistical methods with Pandas & Numpy & Matplotlib
- Applied the Apriori algorithm for Association Rule mining to discern patterns in adoption

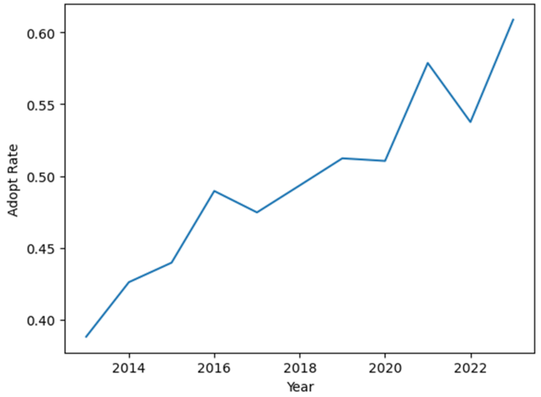
Code Review and Findings

Analysis of Adoption Patterns - Trends

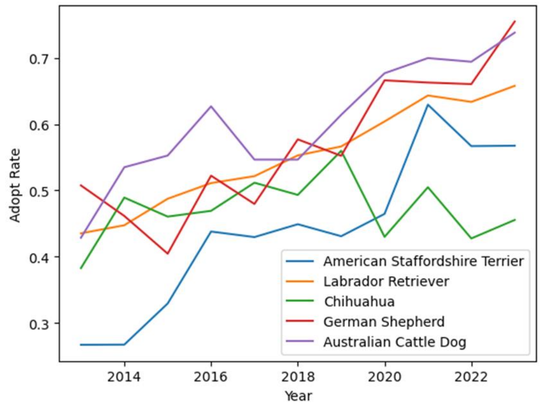


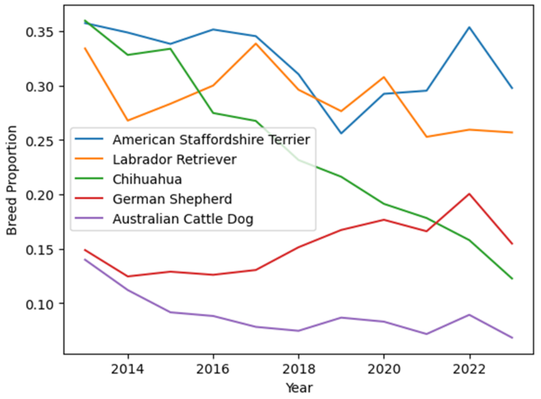
summer and winter holidays being popular





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



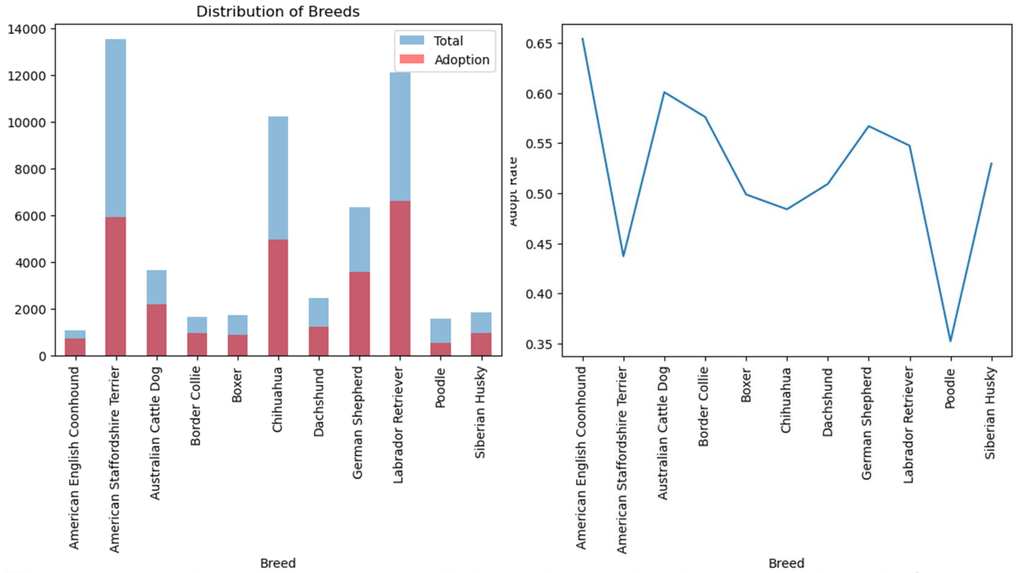


Adoption rates for these breeds are generally on the rise, with the exception of the Chihuahua. There's an upward trend in the adoption of German Shepherds, while the adoption rate for Chihuahuas has notably decreased.

Code Review and Findings

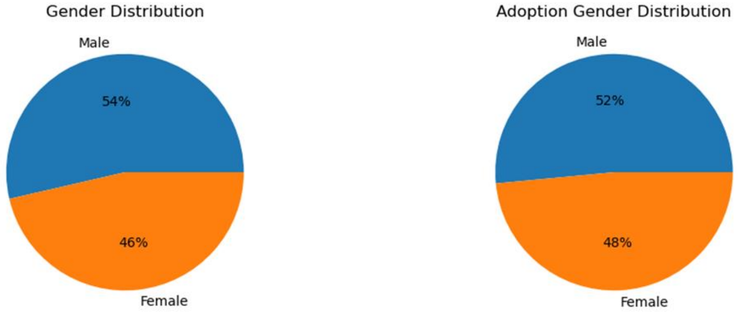
Analysis of Adoption Patterns – Influencing Factors

- Breed



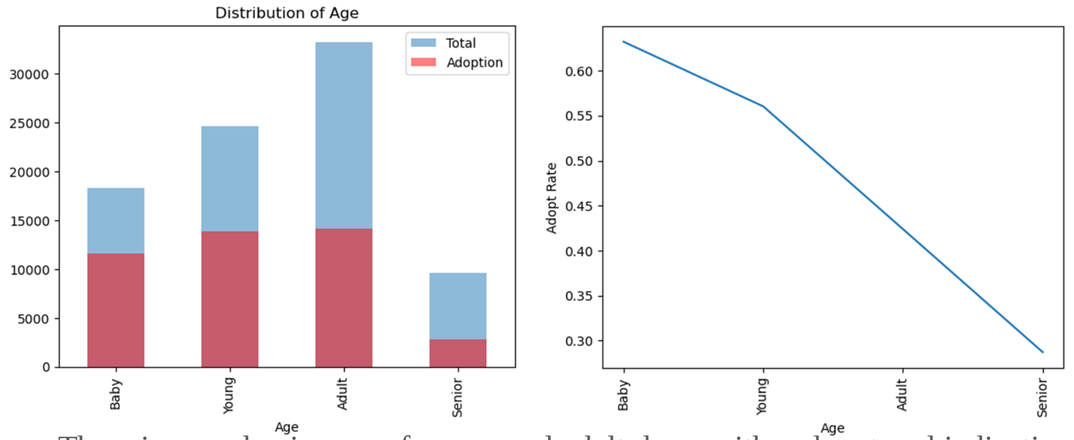
The most popular breeds among all dogs align with the top ten breeds for adoption. Additionally, breeds like the American English Coonhound, Australian Cattle Dog, Border Collie, and German Shepherd are notable for having the highest adoption rates.

- Gender



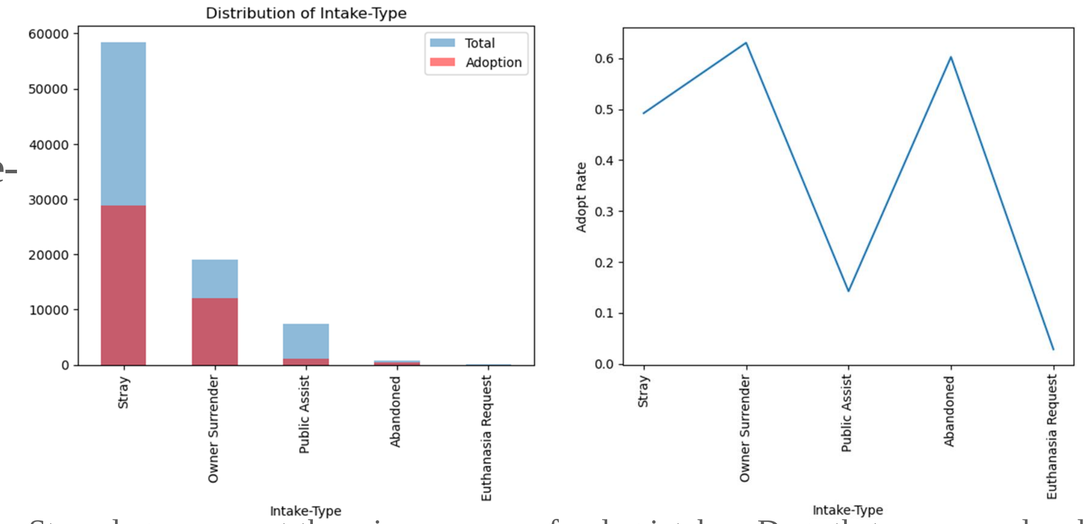
There is a higher ratio of female dogs being adopted.

- Age



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

- Intake-Type

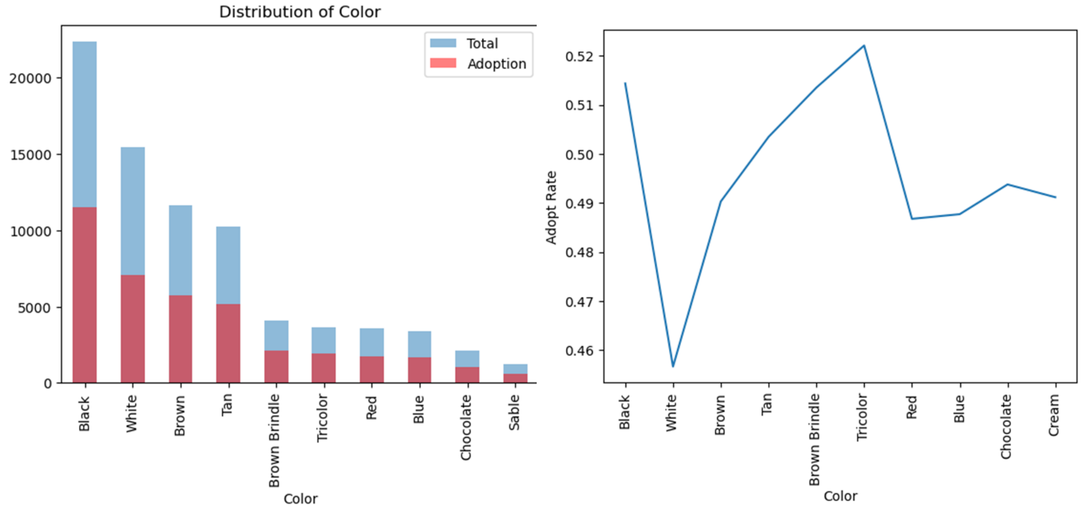


Stray dogs represent the primary reason for dog intakes. Dogs that are surrendered by their owners, abandoned, or stray tend to have higher adoption rates.

Code Review and Findings

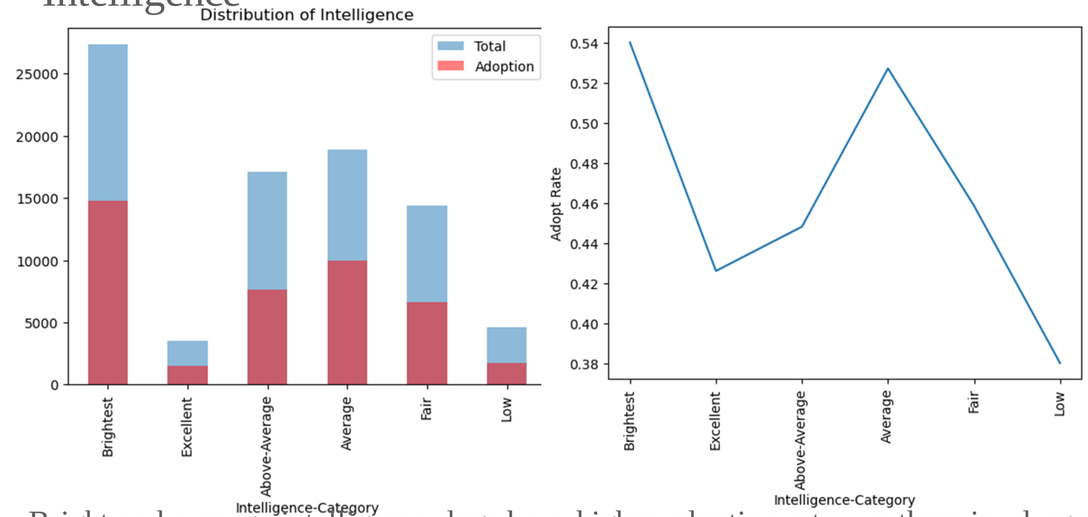
Analysis of Adoption Patterns – Influencing Factors

- Color



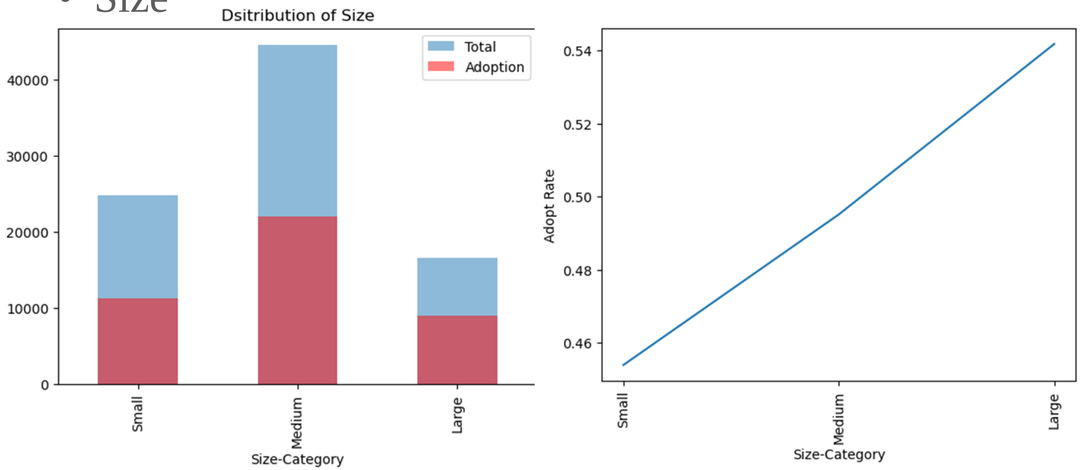
Dark and mixed colors like black/tricolor/brindle are preferred.

- Intelligence



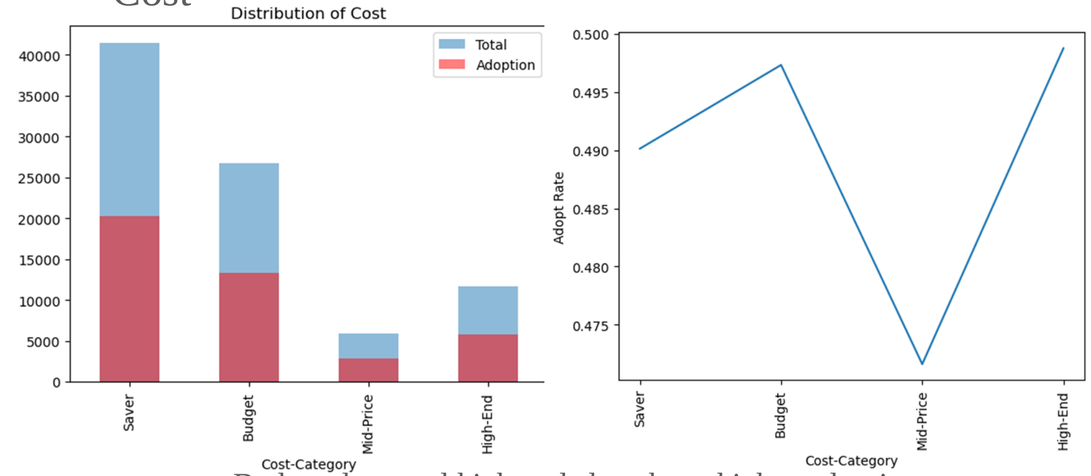
Bright and average intelligence dogs have higher adoption rates, as there is a larger presence of these dogs. Generally, smarter dogs tend to be preferred for adoption.

- Size



Larger dogs are more likely to be adopted.

- Cost



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

Code Review and Findings

Analysis of Adoption Patterns – Influencing Factors

- Association Rules

	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

Age Factor - Younger dogs, specifically in the baby/young age categories, show a higher likelihood of adoption.

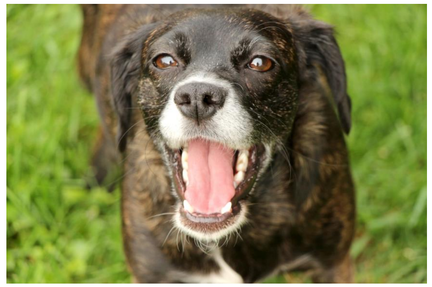
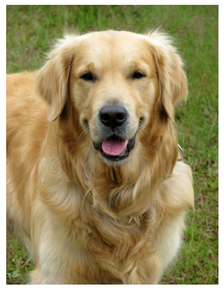
Breed Preference - Dogs belonging to the Terrier category exhibit a greater propensity for being adopted.

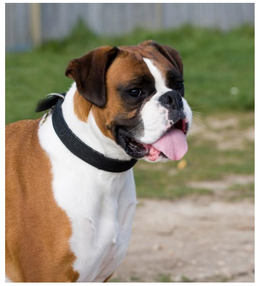
Size Influence - Large dog breeds tend to be more favored in adoption decisions.

Gender Bias - Female dogs appear to have a higher Confidence compared to male dogs.

Intelligence Perception - Dogs perceived as 'bright' are preferred for adoption.

Color Selection - Black dogs demonstrate a significant adoption rate, indicating a color preference.



Code Review and Findings

Supervised Learning

We focus on supervised learning techniques.
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 its 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.

- Target features: *IsAdopted*, *ReFoundByShelterAfter1stAdoption*

ADOPTION , ABANDON

Code Review and Findings

Supervised Learning: Linear Regression Model

- 8 features: *gender*, *intelligence*, *size*, *age*, *longevity*, *total-cost*, *food-cost*, *purchase-cost*
- Training and test sets: (*80*, *20*)
- Target features: *IsAdopted*, *ReFoundByShelterAfterAdoption*

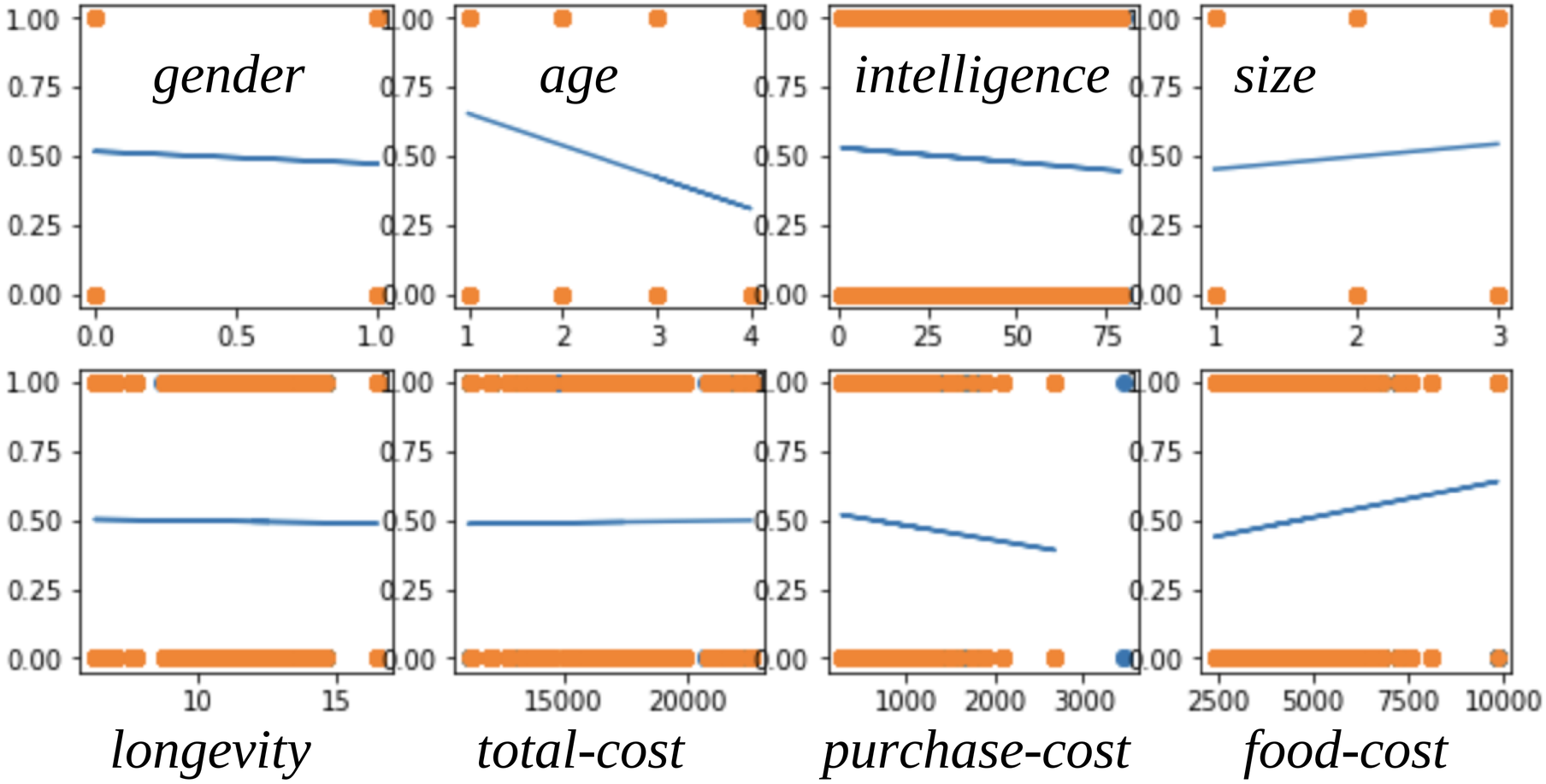
ADOPTION , ABANDON

Code Review and Findings

Supervised Learning: Linear Regression Model

what will affect a dog to find a new home?

ADOPTION ?

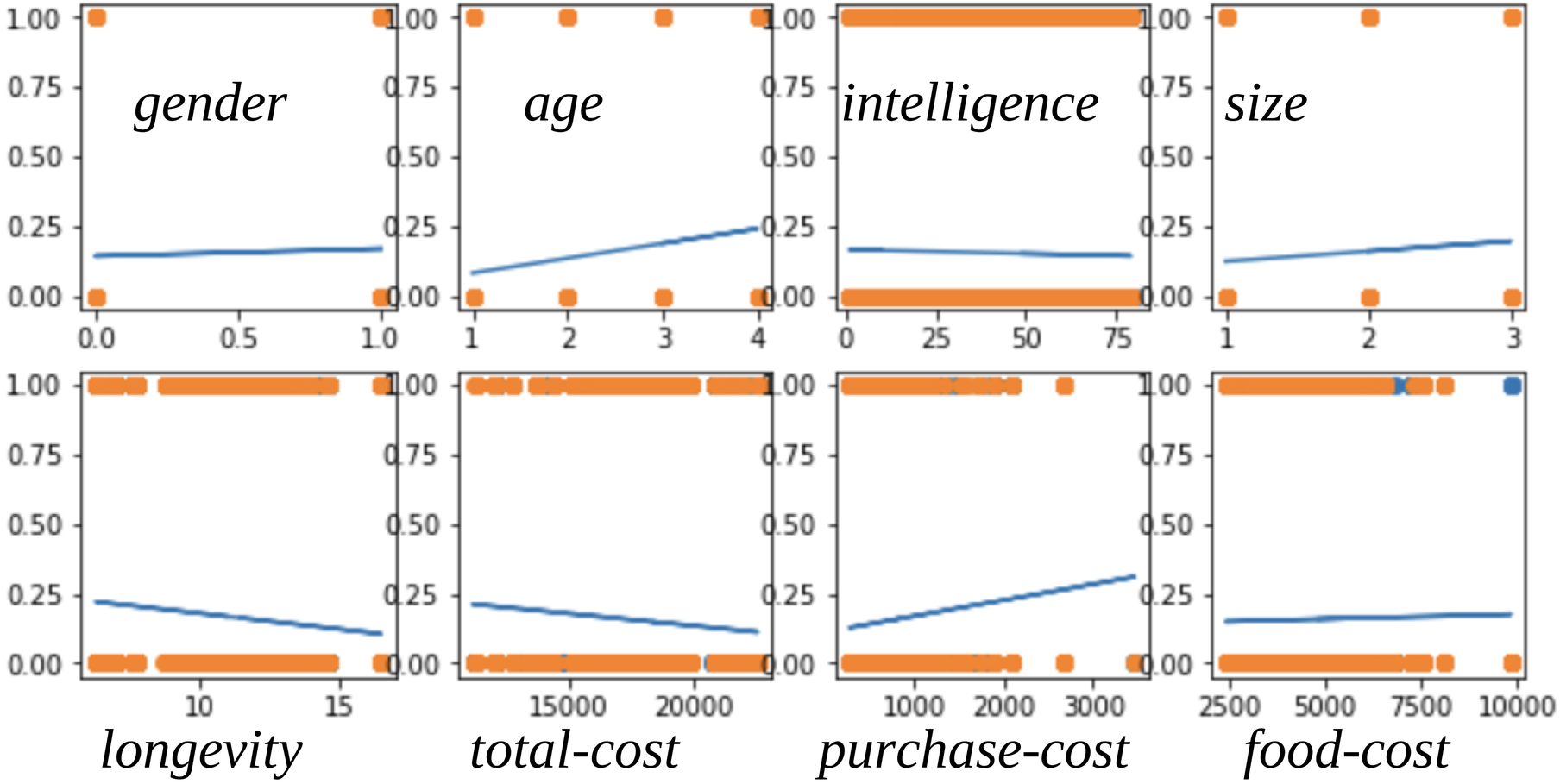


Code Review and Findings

Supervised Learning: Linear Regression Model

What will affect a dog is found again by shelter after adoption?

ABANDON ?



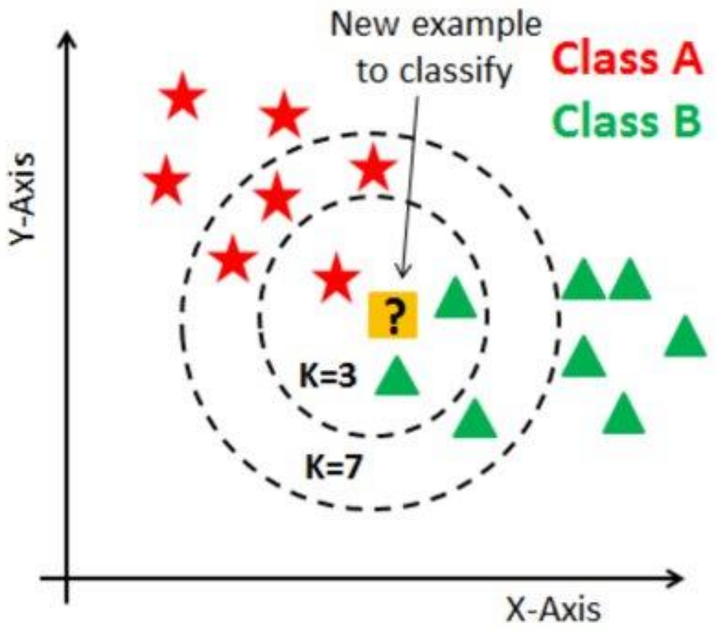
Code Review and Findings

Nearest Neighbor (kNN) Classifier Model

Utilizing supervised learning methods, particularly a classifier model, our objective is to anticipate the factors influencing a dog's likelihood of finding a new home and understand how these variables operate when a shelter rediscovers a dog post-adoption.

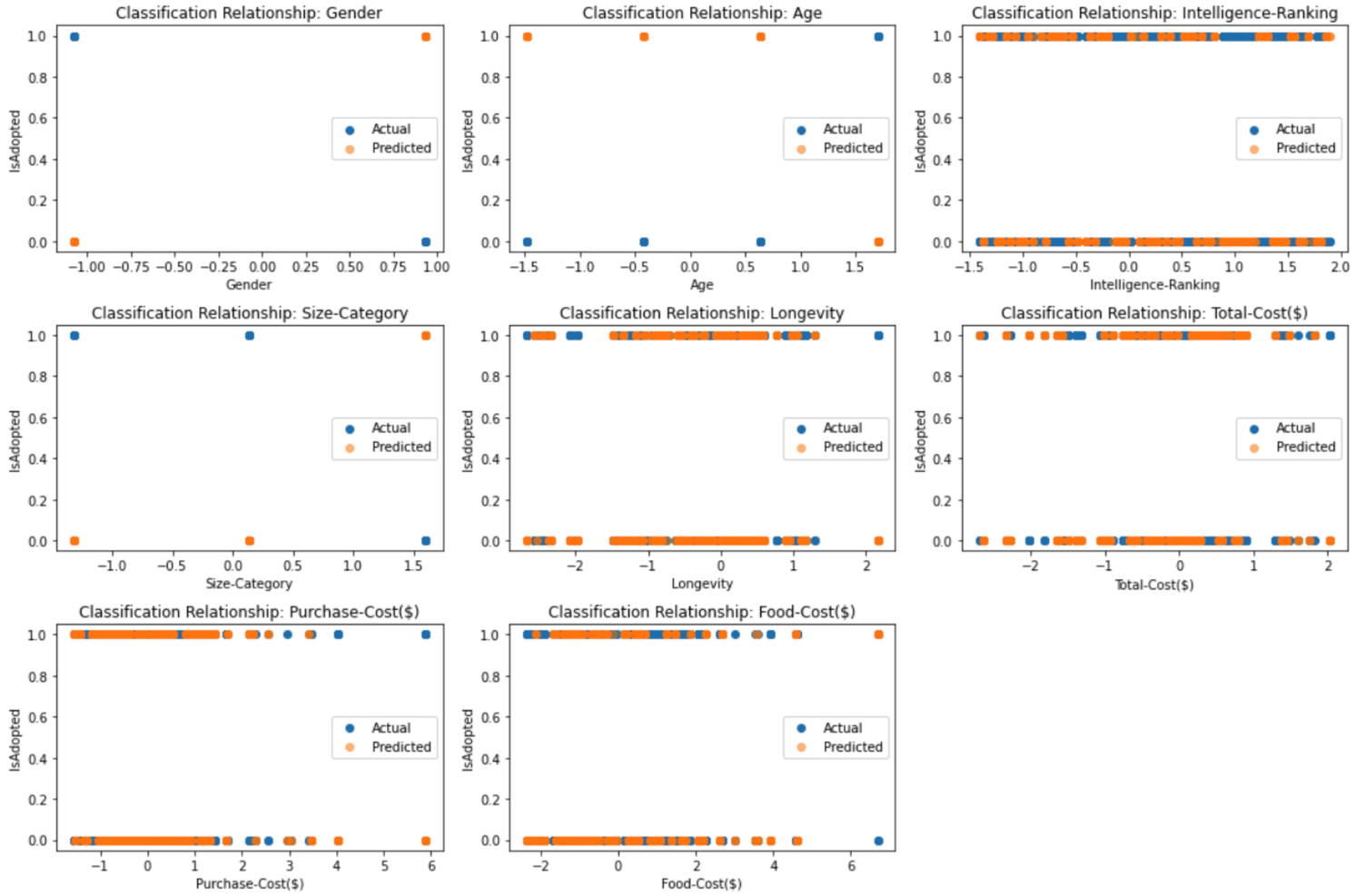
By establishing correlations between different factors and the duration a dog spends in a shelter, our goal is to develop a predictive model that assists in estimating this vital metric, whether this dog will be adopted or not in the near future.

The 8 features we used for thee kNN model: gender, intelligence, size, age, longevity, total-cost, food-cost, purchase-cost



Code Review and Findings

Nearest Neighbor (kNN) Classifier Model



For IsAdopted decision

Evaluation for the feature: Gender
Accuracy: 0.47884615384615387
Precision: 0.47140834331772136

Evaluation for the feature: **Age**
Accuracy: 0.5372377622377622
Precision: 0.5156137479541735

Evaluation for the feature: Intelligence-Ranking
Accuracy: 0.5288461538461539
Precision: 0.5202974521516518

Evaluation for the feature: Size-Category
Accuracy: 0.5238927738927739
Precision: 0.5378279438682123

Evaluation for the feature: Longevity
Accuracy: 0.5286130536130537
Precision: 0.5237065859633953

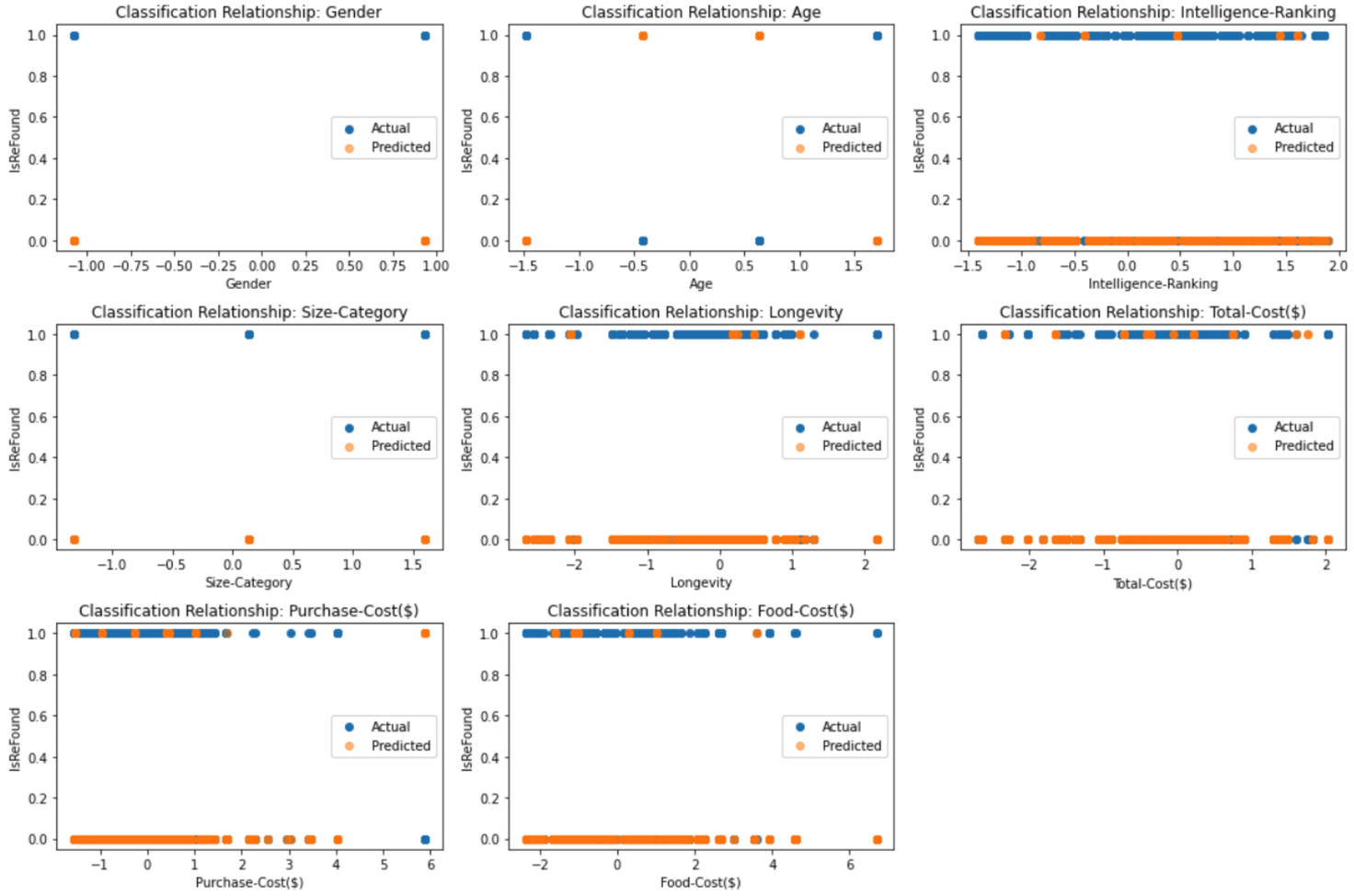
Evaluation for the feature: **Total-Cost($)**
Accuracy: 0.5336247086247087
Precision: 0.5287038317886291

Evaluation for the feature: Purchase-Cost($)
Accuracy: 0.5228438228438228
Precision: 0.5114564654313608

Evaluation for the feature: **Food-Cost($)**
Accuracy: 0.5308857808857809
Precision: 0.5313886045718185

Code Review and Findings

Nearest Neighbor (kNN) Classifier Model



Evaluation for the feature: **Gender**
Accuracy: 0.8474941724941725

Evaluation for the feature: Age
Accuracy: 0.4244172494172494

Evaluation for the feature: Intelligence-Ranking
Accuracy: 0.8251165501165502

Evaluation for the feature: **Size-Category**
Accuracy: 0.8474941724941725

Evaluation for the feature: **Longevity**
Accuracy: 0.8374125874125874

Evaluation for the feature: Total-Cost($)
Accuracy: 0.8136363636363636

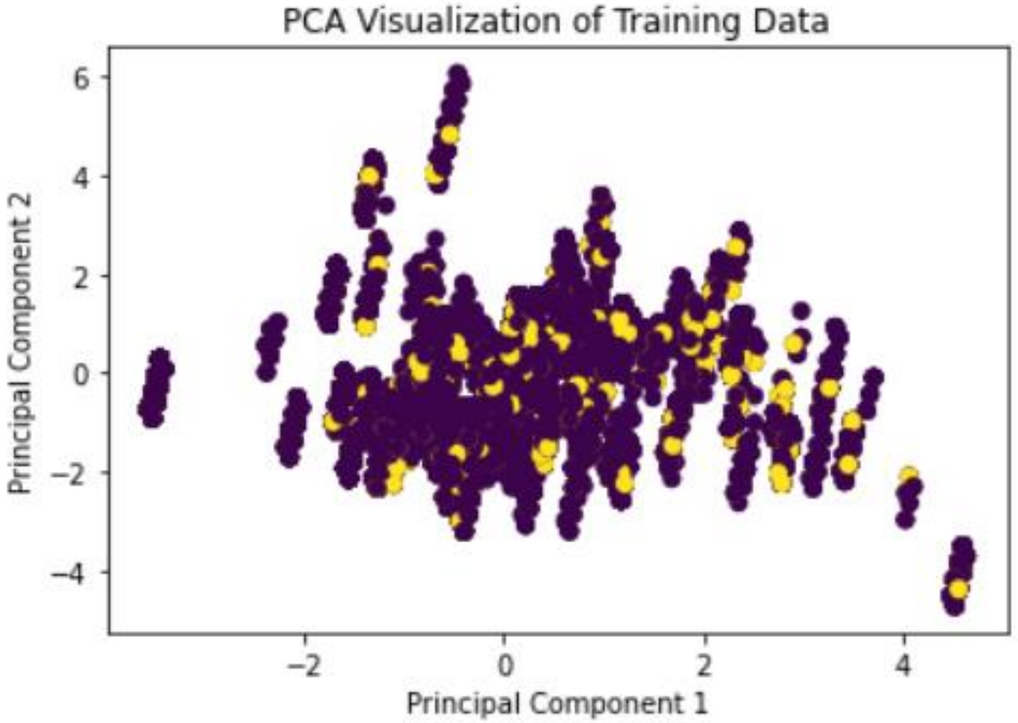
Evaluation for the feature: Purchase-Cost($)
Accuracy: 0.8057692307692308

Evaluation for the feature: Food-Cost($)
Accuracy: 0.7314685314685314

For isReFoundByShelter decision

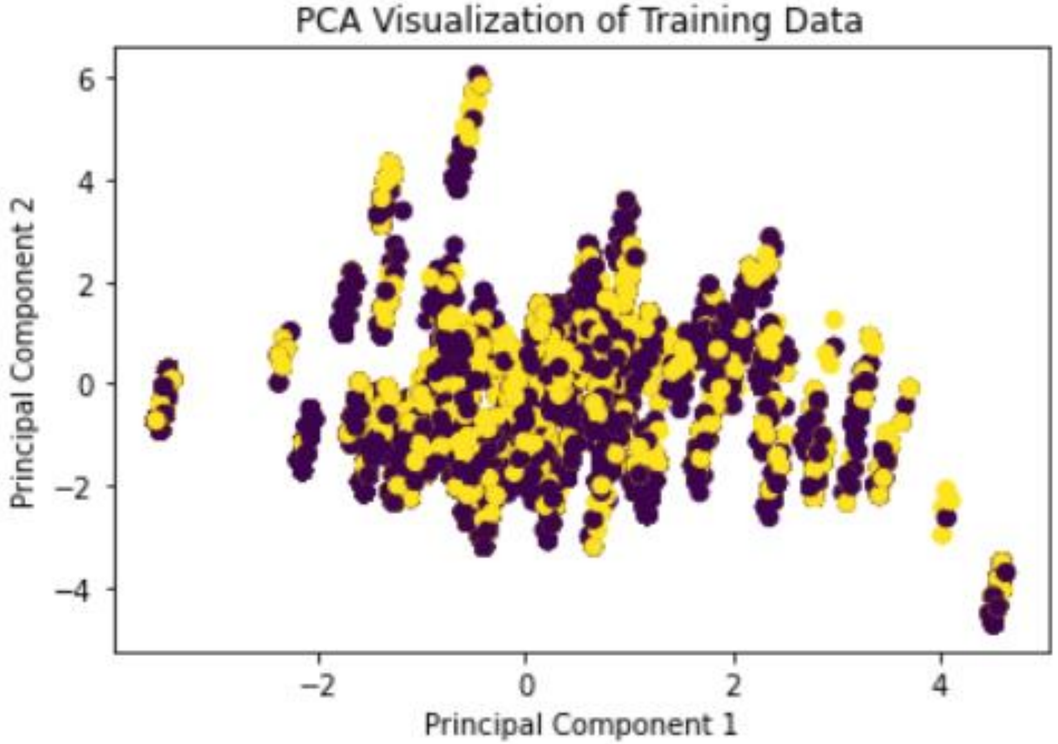
Code Review and Findings

Nearest Neighbor (kNN) Classifier Model



All 8 features model Accuracy:
0.5979020979020979

For IsAdopted decision

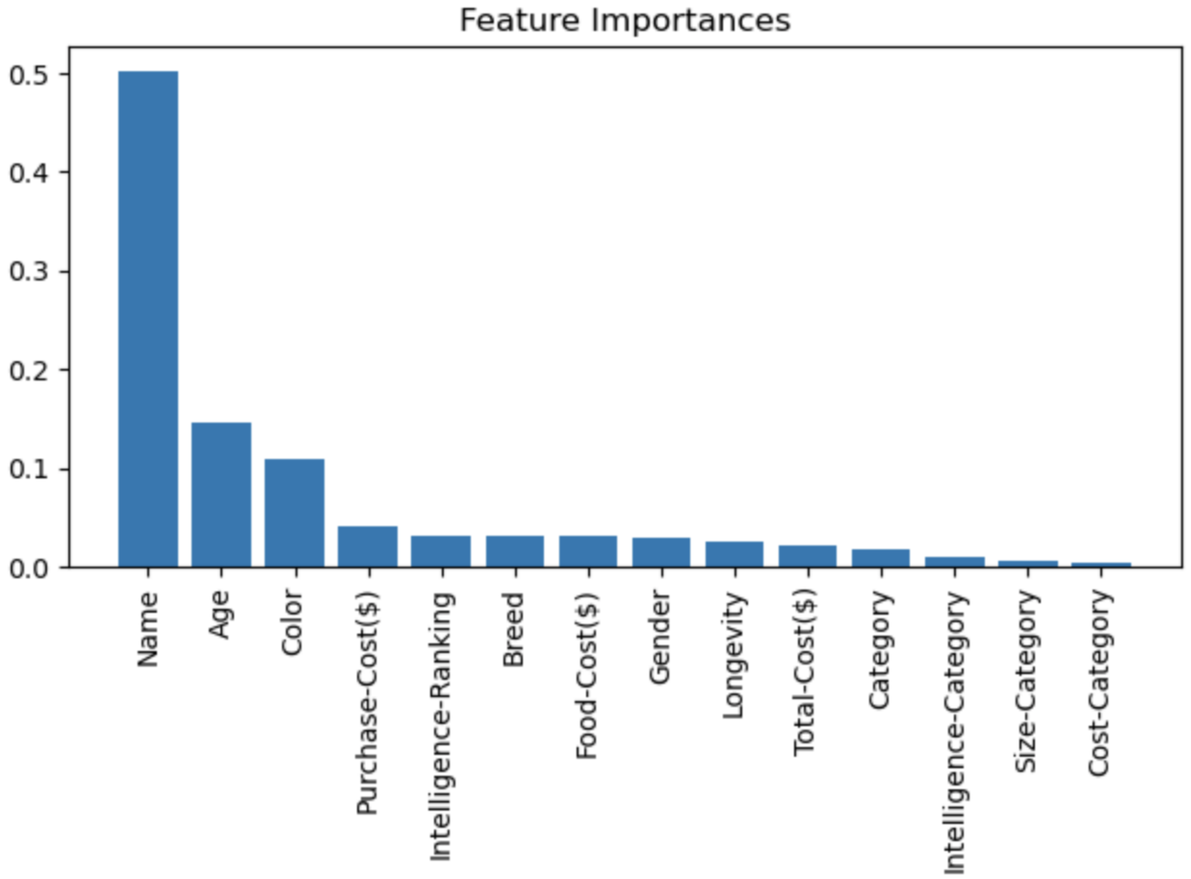


All 8 features model Accuracy:
0.8476689976689976

For isReFoundByShelter decision

Code Review and Findings

Decision Tree Classifier Model



We used decision tree classifier to train a prediction model. The model identified several key features that significantly impact adoption outcomes.

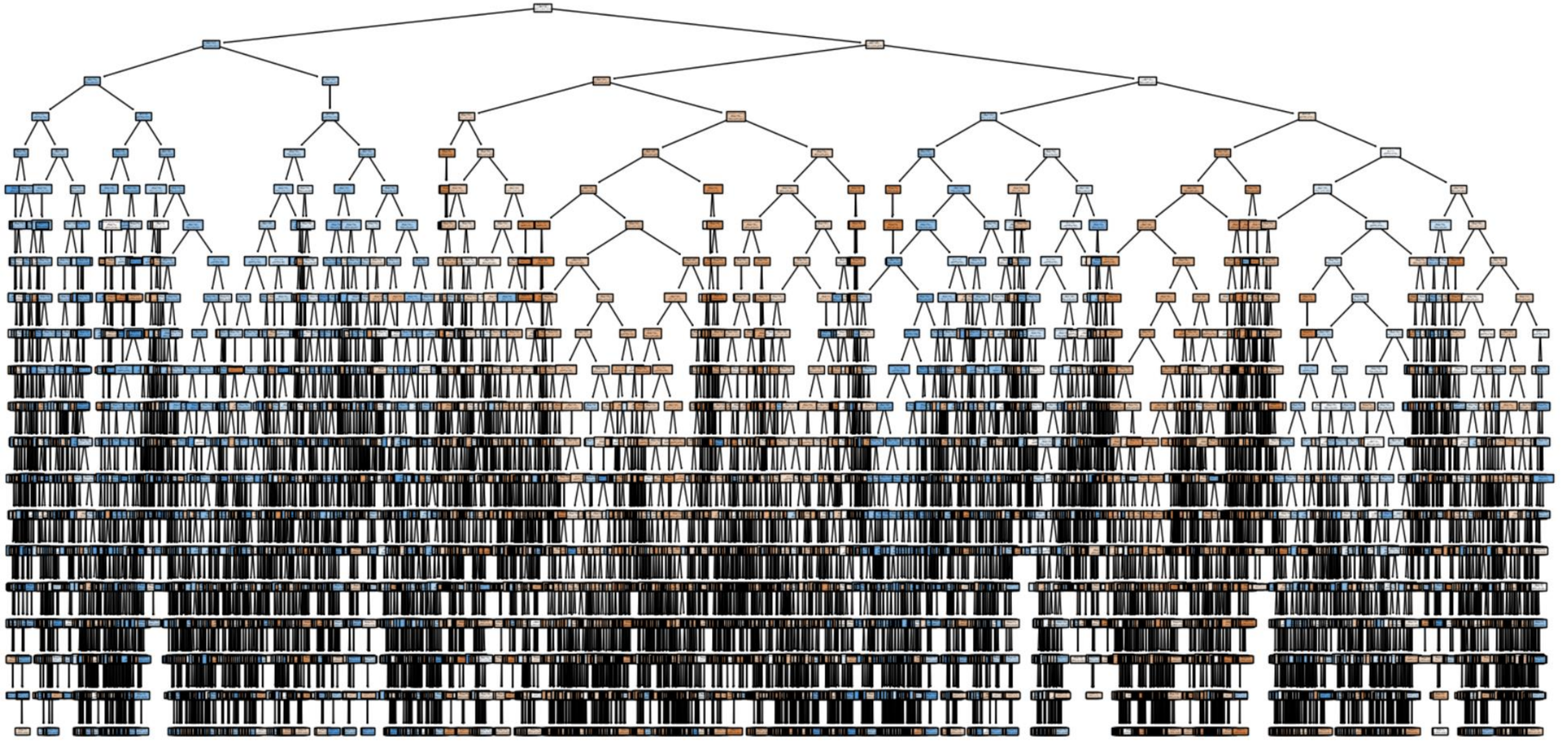
- Notably, the **age** and **color** of the dog are among the most influential factors.

- Other features like purchase cost, intelligence ranking, breed, food cost, gender, longevity are also impacting the adoption decision.

- Younger dogs and certain breeds are more likely to be adopted quickly, while dogs with health issues or those who have been in the shelter for an extended period face lower adoption rates.

Code Review and Findings

Decision Tree Classifier Model



The implications of these findings suggest that shelters could benefit from targeted marketing strategies for older and long-term resident dogs, as well as enhanced healthcare interventions. By addressing the identified factors, shelters can implement more effective policies to increase the likelihood of adoption for all dogs. This model serves as a starting point for deeper analysis and continual improvement in strategies to find homes for sheltered dogs.

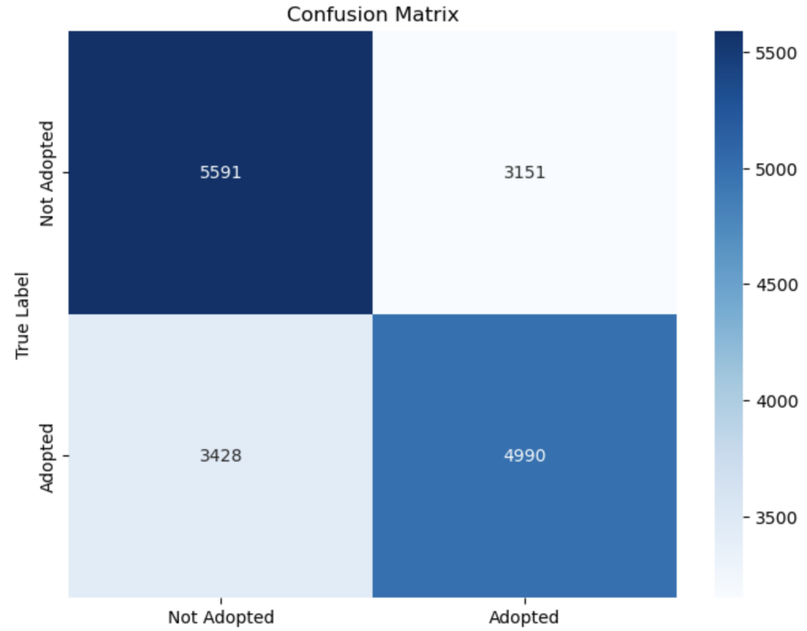
Code Review and Findings

Decision Tree Classifier Model – Accuracy

```
Accuracy: 0.6166083916083916
              precision    recall  f1-score   support

       False       0.62      0.64      0.63      8742
        True       0.61      0.59      0.60      8418

    accuracy                           0.62     17160
   macro avg       0.62      0.62      0.62     17160
weighted avg       0.62      0.62      0.62     17160
```



- We used decision tree classifier model to train a prediction model with the accuracy of 62%. It shows

True Positives (Bottom Right): The model correctly predicted 5,032 instances where dogs were adopted.

True Negatives (Top Left): The model correctly predicted 5,636 instances where dogs were not adopted.

False Positives (Top Right): The model incorrectly predicted 3,106 instances where dogs were not adopted, but the model predicted they would be.

False Negatives (Bottom Left): The model incorrectly predicted 3,386 instances where dogs were adopted, but the model predicted they would not be.

Future Work

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

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.

Thank you!