Pet Adoption Outcome Prediction and Data Analysis

Group 14

Agenda

- Motivation
- Methodology
- Data extraction
- EDA and visualization
- Regression models and analysis
- Impact of Project

Motivation

- Animal shelters are at maximum capacity and are lacking resources they need to expand
- We want to perform data analysis to understand the outcomes of animals that are taken into the shelter to figure out how shelters could increase adoption rates of all types of animals

Methodology

- Filtered and pre-processed the dataset to retrieve relevant information
- Used a specific shelter's data and used exploratory data analysis on animal data
    - Analyzed different animal species, types of animal outcomes, and animal characteristics
- Performed linear and nonlinear regression to identify significant features in adoption likelihood for the two most frequently adopted animal species

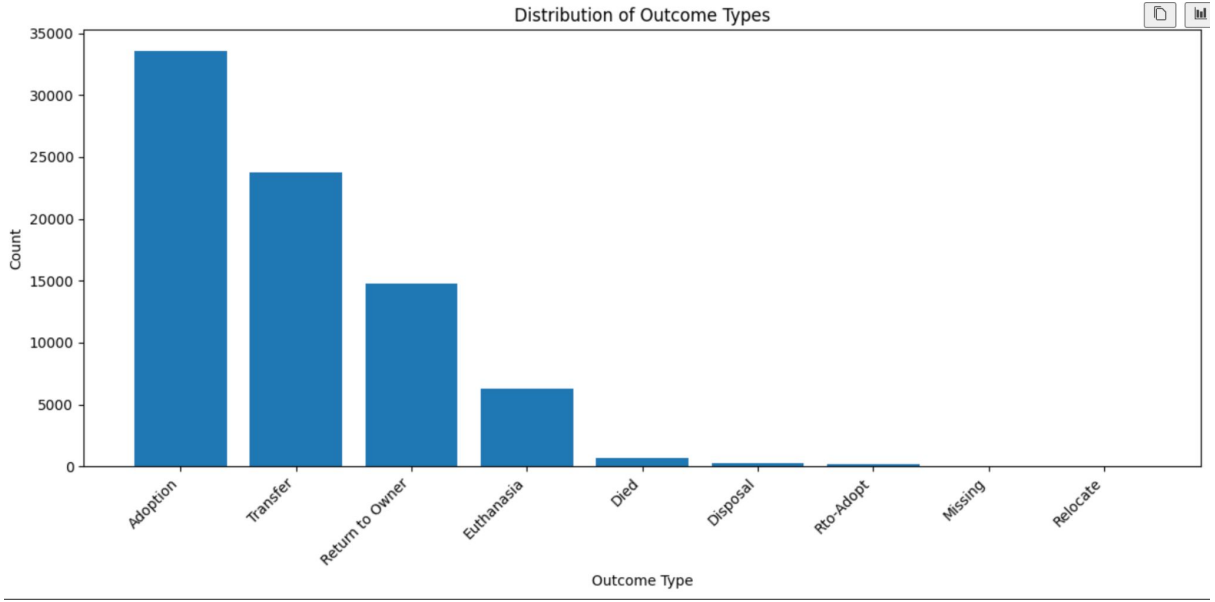
Dataset

- Austin Animal Center Shelter Intakes and Outcomes
- 79000 outcomes of animals starting from 2013
- Features include age, date, outcome type, animal type, location, etc.
- Data cleaning involved importing csv to polars, removing null outcomes, removing redundant data

Exploratory Data Analysis

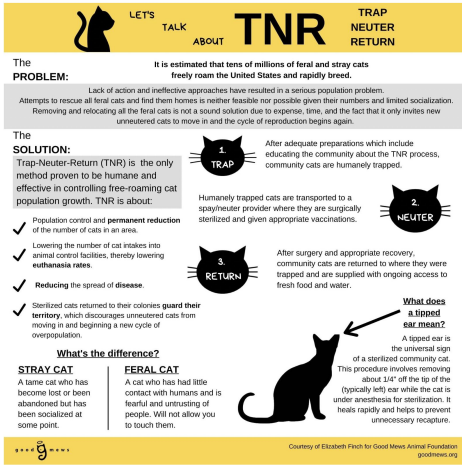
Overall outcome trends

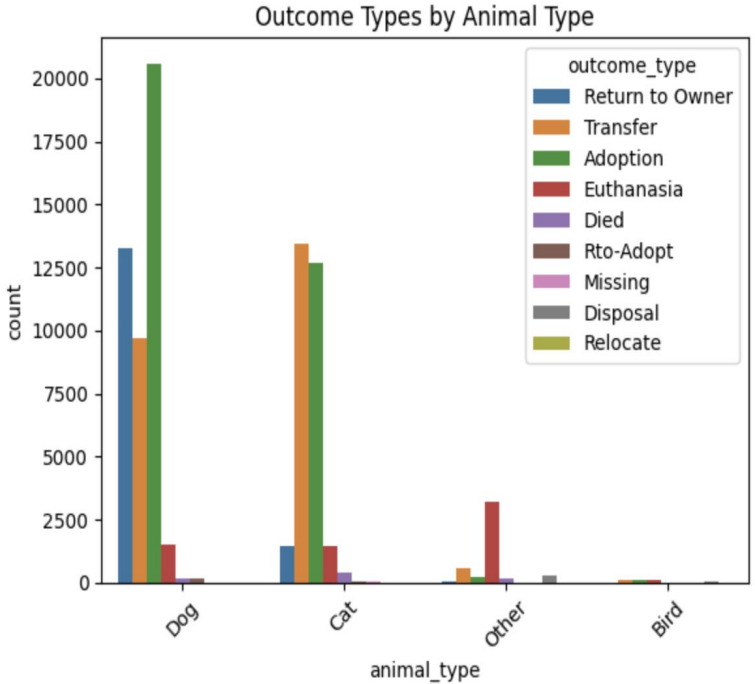
- High adoption rates indicate strong community engagement.
- High return-to-owner rate reflects effective reunification.
- Frequent transfers suggest limited capacity.



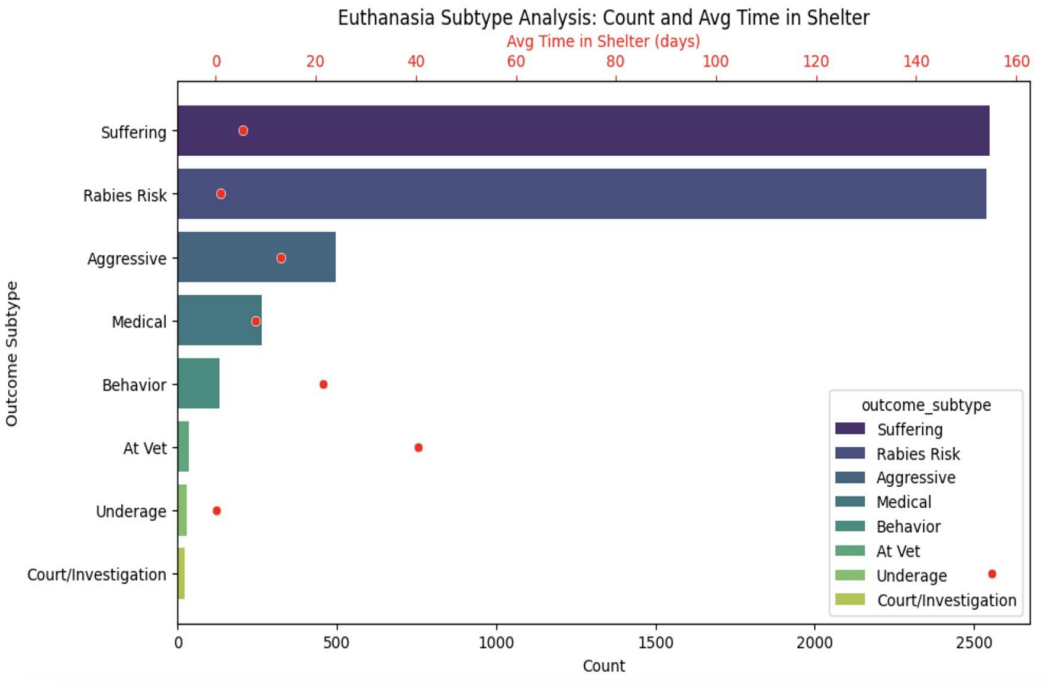
Outcome trends by animal type

- Dogs have the highest adoption rates, followed closely by cats.
- Dogs are much more likely to be returned to their owners compared to other animals.
  a. Most dogs are microchipped and licensed
- Cats are more likely to be transferred than dogs.
  a. Trap-Neuter-Return programs to reduce the number of transfers and the risks of exhausting shelter capacity



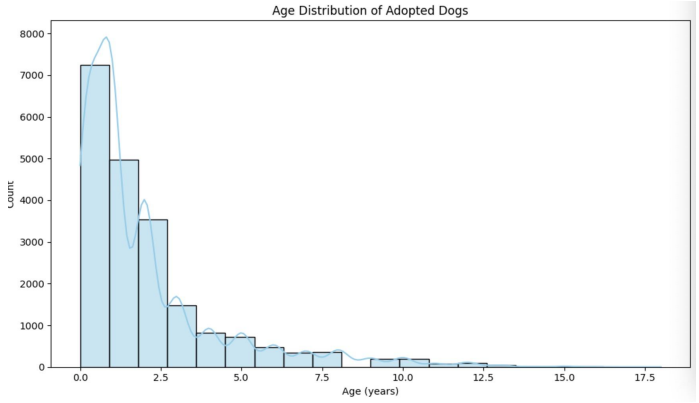


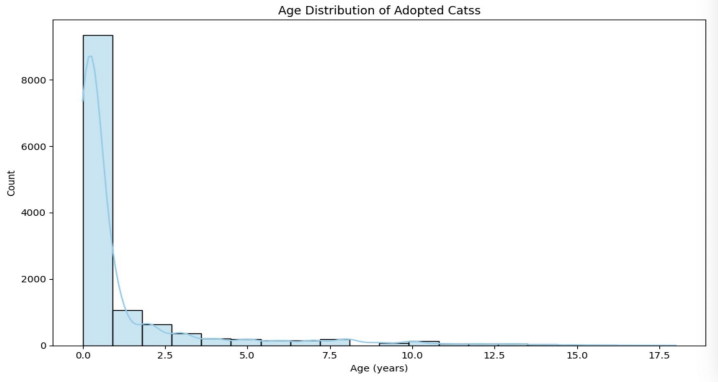
Euthanasia analysis



- Suffering and rabies risk are two main reasons that cause shelter to euthanize animals
- Animals with rabies risk were in the shelter for a period of 0 days, indicating that they were euthanized immediately
- Time for animals that are suffering to be euthanized is also short which could indicate limited resources

Analysing trends for dog and cat adoption

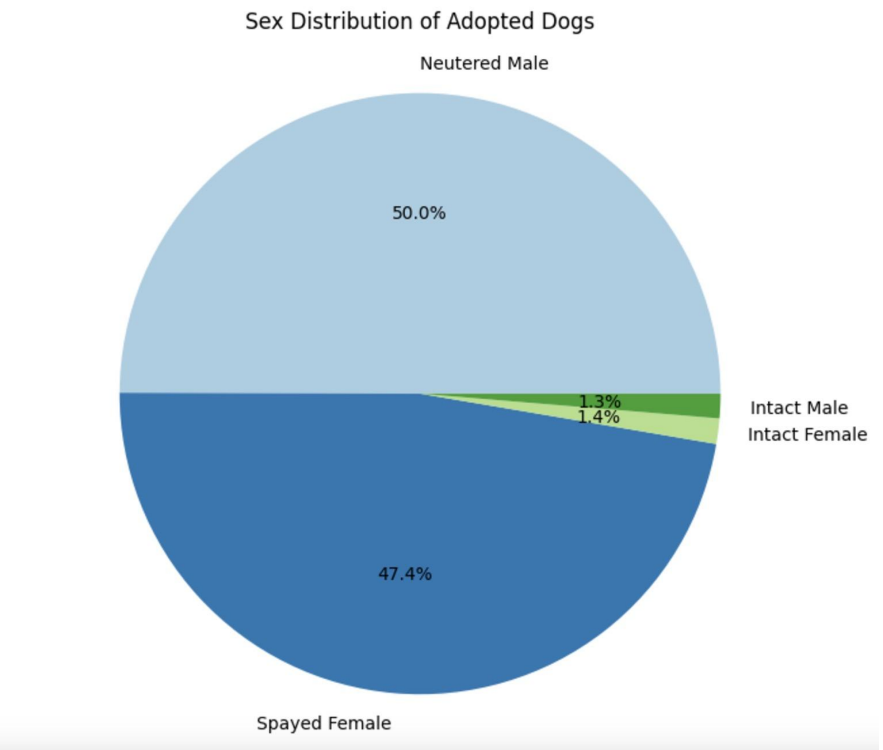




- We analyze four factors that may influence animals' adoption results: breeds, age, color and sex
- Age distribution over adopted dogs and cats
  - as dogs and cats getting older, their chances of being adopted decrease
  - compared to dog adoption trend, the results of cat adoption based on age factor shows a huger disparity
  - kittens aged 0-8 months are about 10 times more likely to be adopted than other age groups
  - help shelter make better promotion adoption plan: focus on posting updates about cats over one year old on social media to draw the attention of potential adopters, helping them get adopted more effectively.

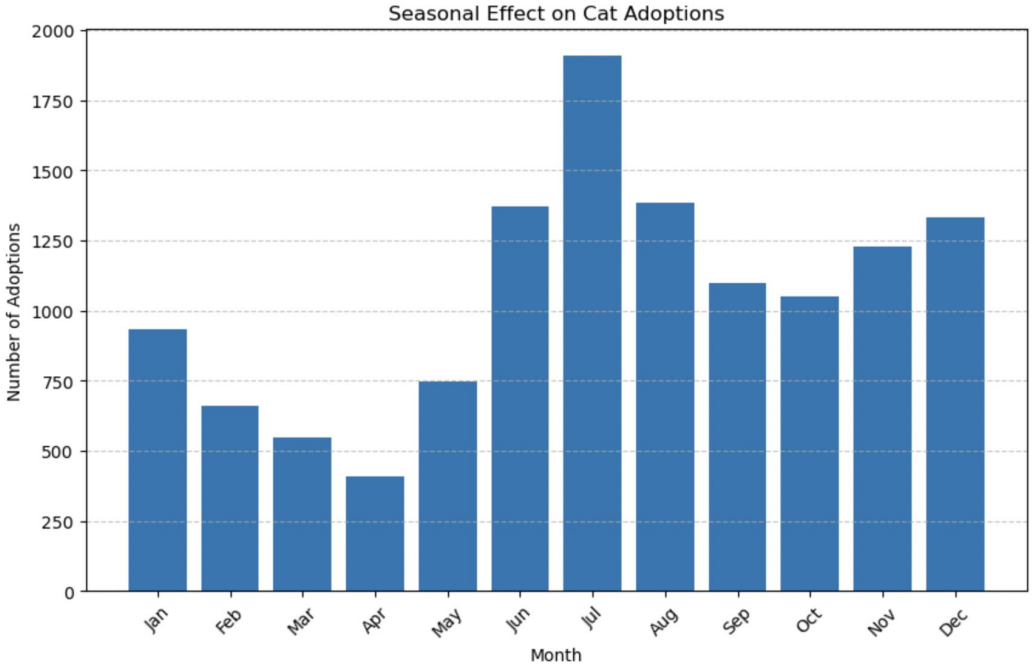
Sex Analysis for dog and cat

- Male and female dogs have nearly equal adoption rates
- Spayed dogs are 30 times more likely to be adopted compared to intact dogs
- Sex analysis of cat has nearly the same pattern as dog’s
- Implication for shelter: have the animals neutered to increase their chances of adoption



Seasonal effect on cat adoptions

Interestingly, this shelter observes more adoptions in the later half of the year.



Regression Models and Analysis

Significant Features Impacting Dogs' Adoption Likelihood

- Given that dogs have the highest adoption rates, we wanted to see which features are more statistically significant in impacting adoption rates
- Features observed: Age, Sex, Color
- Used only a few colors for an exploratory regression model, to expand on this, can use all the other colors
- Performed linear regression with adoption likelihood as the dependent variable

OLS Regression Results

Dep. Variable:	encoded_adoption	R-squared:	0.001
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	16.70
Date:	Sun, 01 Dec 2024	Prob (F-statistic):	4.40e-05
Time:	19:13:58	Log-Likelihood:	-6246.4
No. Observations:	13768	AIC:	1.250e+04
Df Residuals:	13766	BIC:	1.251e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.1871	0.004	44.830	0.000	0.179	0.195
age_upon_intake_years	-0.0042	0.001	-4.087	0.000	-0.006	-0.002

Omnibus:	3568.966	Durbin-Watson:	1.491
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7043.745
Skew:	1.696	Prob(JB):	0.00
Kurtosis:	3.883	Cond. No.	5.32

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
results_age.pvalues[1]
```

```
4.395518899237089e-05
```

Significant Features Impacting Dogs' Adoption Likelihood

OLS Regression Results

Dep. Variable:	encoded_adoption	R-squared:	0.001
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	2.375
Date:	Sun, 01 Dec 2024	Prob (F-statistic):	0.0270
Time:	19:13:58	Log-Likelihood:	-6247.6
No. Observations:	13768	AIC:	1.251e+04
Df Residuals:	13761	BIC:	1.256e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.1783	0.004	50.158	0.000	0.171	0.185
color_Apricot[T.True]	-0.1783	0.092	-1.928	0.054	-0.360	0.003
color_Black[T.True]	-0.0101	0.014	-0.736	0.462	-0.037	0.017
color_Blue[T.True]	-0.0886	0.043	-2.047	0.041	-0.173	-0.004
color_Brown[T.True]	-0.0297	0.016	-1.890	0.059	-0.061	0.001
color_Gray[T.True]	-0.0487	0.052	-0.937	0.349	-0.151	0.053
color_Tan[T.True]	0.0176	0.015	1.172	0.241	-0.012	0.047

Omnibus:	3567.728	Durbin-Watson:	1.489
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7040.148
Skew:	1.695	Prob(JB):	0.00
Kurtosis:	3.880	Cond. No.	28.6

Apricot: 0.053827495708603394
Black: 0.46184628233347846
Blue: 0.040720637438249316
Brown: 0.05882841232278112
Gray: 0.34869685739200995
Tan: 0.24132890356203648

OLS Regression Results

Dep. Variable:	encoded_adoption	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.002
Method:	Least Squares	F-statistic:	11.95
Date:	Sun, 01 Dec 2024	Prob (F-statistic):	6.56e-06
Time:	19:13:58	Log-Likelihood:	-6242.8
No. Observations:	13768	AIC:	1.249e+04
Df Residuals:	13765	BIC:	1.251e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.133e+11	5.07e+10	2.232	0.026	1.38e+10	2.13e+11
gender_classified_F[T.True]	-1.133e+11	5.07e+10	-2.232	0.026	-2.13e+11	-1.38e+10
gender_classified_M[T.True]	-1.133e+11	5.07e+10	-2.232	0.026	-2.13e+11	-1.38e+10

Omnibus:	3544.272	Durbin-Watson:	1.488
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6965.254
Skew:	1.686	Prob(JB):	0.00
Kurtosis:	3.876	Cond. No.	3.32e+13

```
results_gender.pvalues[1]
```

0.025636496988339808

```
results_gender.pvalues[2]
```

0.02563649698828926

Conclusions

- Age of the dog upon intake was the statistically significant feature in comparison to gender and color
- Therefore, it had more of an impact on adoption likelihood
- Given the positive correlation value, this means that the younger the dog is the more likely they are to be adopted than an older dog

Non-linear Regression Analysis for Cats

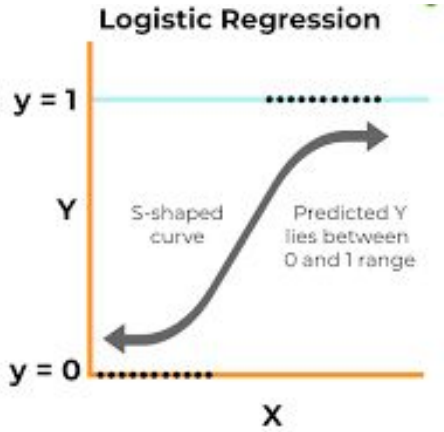
- Coming to cats, we want to:
  - Predict the likelihood of a cat being adopted.
  - Estimate how long a cat will stay in the shelter.
  - Utilize features like age, sex, breed, color, and intake condition.

Generated by GPT-4o

Classification

- Used logistic regression with features: age, sex, breed, color, intake condition
- Significantly better than random guess

	precision	recall
negative	0.62	0.55
positive	0.54	0.61



Predicting Time in Shelter

- Initial Random Forest model performed poorly (negative R^2).
- Improved by:
  - Performing OLS to select significant features ($p < 0.05$).
  - Removing low variance features (variance < 0.01).
  - Removing multicollinearity (VIF < 10).
- Gradient Boosting Regressor showed better performance ($R^2 = 0.05$).

$$VIF_i = \frac{1}{1 - R_i^2}$$

OLS Regression Results

Dep. Variable:	time_in_shelter_days	R-squared:	0.023
Model:	OLS	Adj. R-squared:	0.011
Method:	Least Squares	F-statistic:	1.972
Date:	Tue, 03 Dec 2024	Prob (F-statistic):	6.91e-22
Time:	01:08:33	Log-Likelihood:	-1.3796e+05
No. Observations:	26958	AIC:	2.766e+05
Df Residuals:	26642	BIC:	2.792e+05
Df Model:	315		
Covariance Type:	nonrobust		

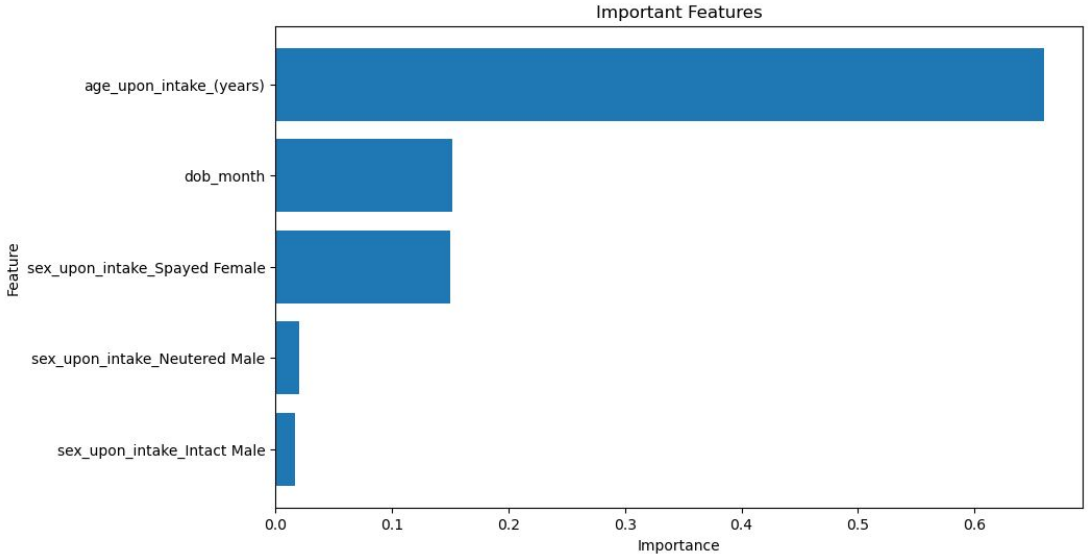
	coef	std err	t	P>|t|
const	25.2796	31.842	0.794	0.427
age_upon_intake_(years)	0.7849	0.116	6.756	0.000

```
Significant Features:
Index(['age_upon_intake_(years)', 'dob_month', 'sex_upon_intake_Intact Male',
       'sex_upon_intake_Neutered Male', 'sex_upon_intake_Spayed Female',
       'breed_Chartreux Mix', 'breed_Cymric Mix',
       'breed_Pixiebob Shorthair Mix', 'color_Blue Tabby/Cream',
       'color_Blue/Tortie', 'color_Brown Tabby/Agouti',
       'color_Lynx Point/Tortie Point', 'color_Orange Tabby/Brown',
       'color_Tortie/Black Smoke'],
      dtype='object')

Variance Inflation Factor (VIF) after removing near-zero variance features:
                         feature       VIF
0        age_upon_intake_(years)  1.941191
1                    dob_month  2.039314
2    sex_upon_intake_Intact Male  1.579943
3  sex_upon_intake_Neutered Male  1.565108
4  sex_upon_intake_Spayed Female  1.546945
```

Feature Importance

- Most important features: Age, sex, and month of birth.
- Insights:
  - Older cats are less likely to be adopted.
  - Spayed female cats are preferred.
  - Month of birth indicates potential seasonal effects or dataset biases.



Chartreux

Cymric

Pixiebob Shorthair

Case Study

- Created scenarios with varied features.
- Observations:
  - Older and sick cats stay longer and are less likely to be adopted.
  - Spayed females are more adoptable among young, healthy cats.

Scenario	Age	DoB Month	Sex	Breed	Color	Intake Condition	Predicted Time in Shelter (days)	Probability of Adoption
1	2	5	Neutered Male	Chartreux Mix	Blue/Tortie	Normal	18.74	0.73
2	2	3	Spayed Female	Cymric Mix	Lynx Point/Tortie Point	Normal	20.42	0.82
3	2	3	Intact Female	Cymric Mix	Lynx Point/Tortie Point	Normal	18.00	0.79
4	15	3	Spayed Female	Cymric Mix	Lynx Point/Tortie Point	Normal	28.57	0.58
5	5	10	Neutered Male	Pixiebob Shorthair Mix	Orange Tabby/Brown	Injured	22.58	0.62

Conclusions

- Shelters should promote the importance of microchipping and licensing amongst pet owners
- Raise awareness about the specific needs and care requirements of less common pets (birds, reptiles, etc.)
- In order to increase adoption among older animals, shelters can advertise older animals for older owners, since older animals tend to be calmer and less maintenance than younger animals