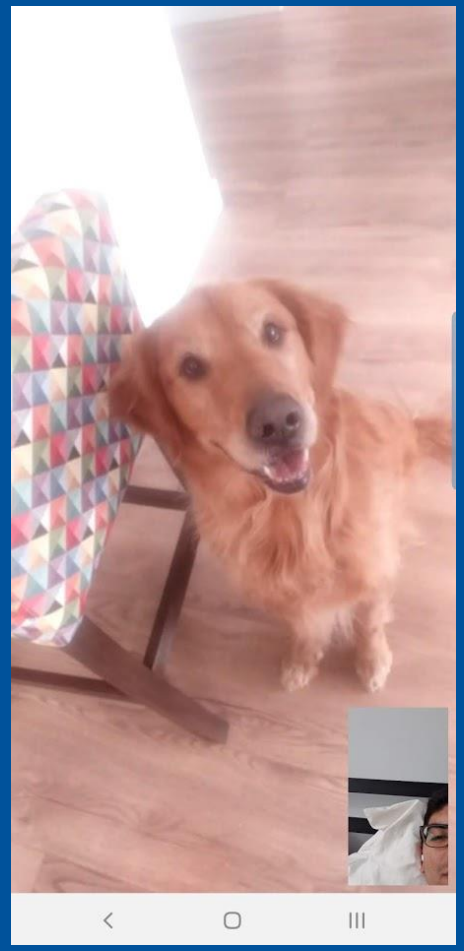
Getting our hairy friends a nice home

Factors that impact pet adoption

Juan David Martínez Gordillo

Duke MIDS
Master in Interdisciplinary Data Science

My motivation: Bingo!



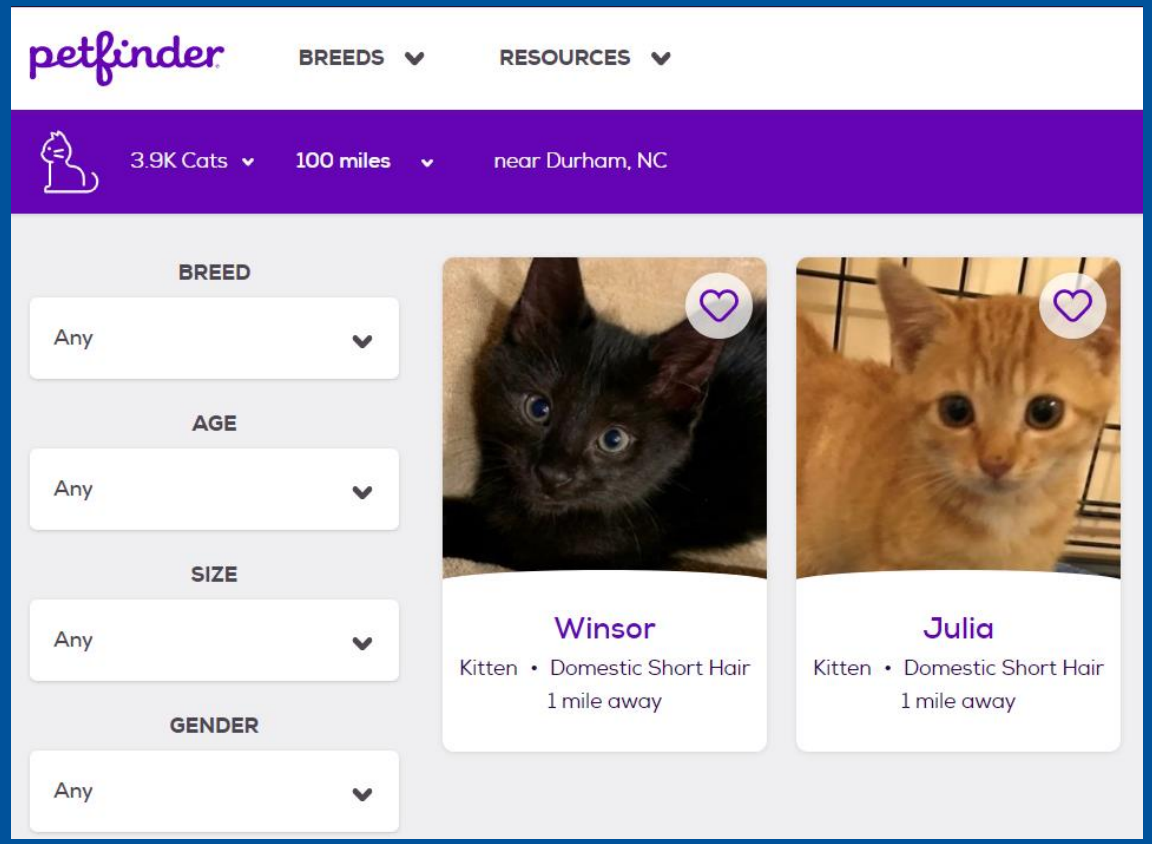
Duke MIDS
Master in Interdisciplinary Data Science

Some context

- Between 2010 and 2016, **33 thousand pets were abandoned** just in Colombia
- Is estimated that in Colombia there are about **one million stray pets**
- The Humane Society of the United States (HSUS) estimates that **three to four million animals die in American shelters each year**

Duke MIDS
Master in Interdisciplinary Data Science

A novel solution: Petfinder.com



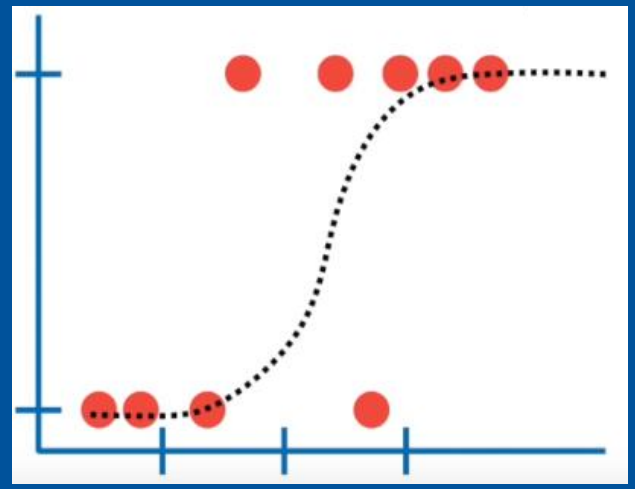
- The largest pet Web site on the Internet
- Connecting homeless animals with the people who want them
- Petfinder has been behind the placement of more than 25 million animals.

Duke MIDS
Master in Interdisciplinary Data Science

Objective

An interpretable modeling technique:

- Increase the likelihood of adoption for any new pet coming into the site



An accurate classification method:

- If we know which types of pets are less likely to adopt, we can increase the site's efforts on those to ensure they are adopted

Duke MIDS
Master in Interdisciplinary Data Science

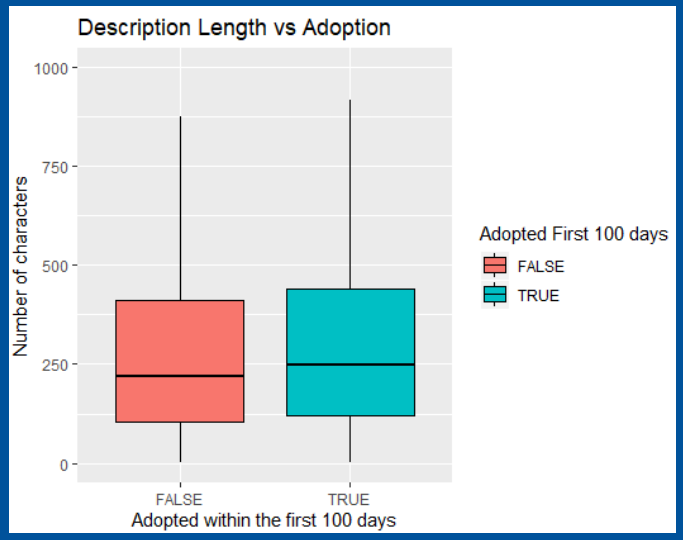
Data

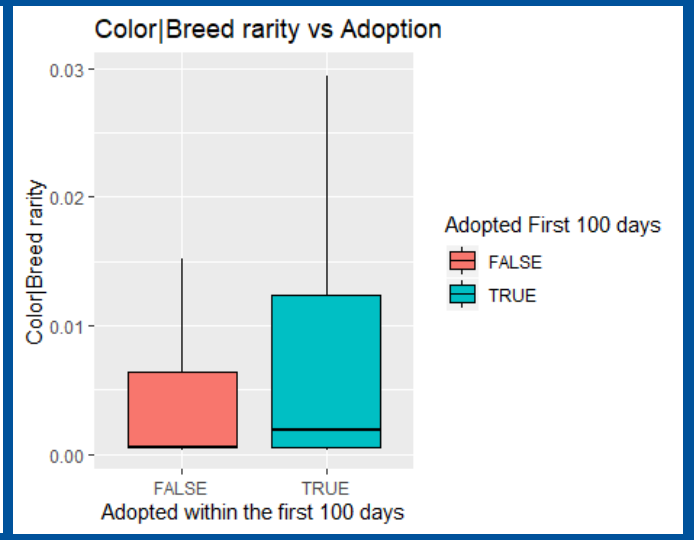
- Observations: 14,993 pet listings (Petfinder.my, 2018)
- Original variables: 24
  - Age, Gender, Color, Health, Type, Breed, Maturity Size, Fur length, Among others

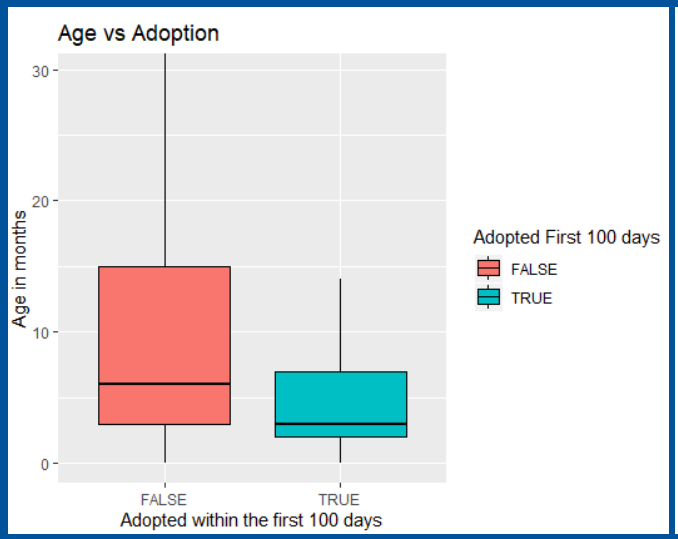
Exploratory Data Analysis

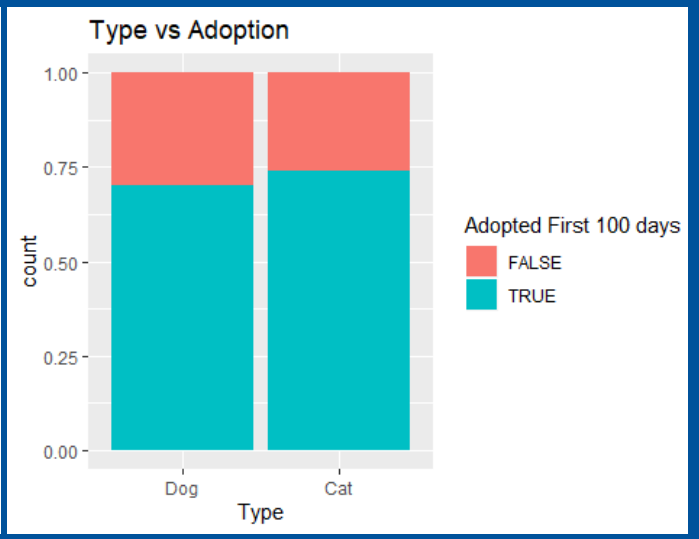
- **Target variable:** Pet adopted in 100 days or less.

- Feature engineering:
  - Mixed Breed
  - Multicolor
  - Life Stage
  - Description length
  - Fee per capita
  - Num rescues
  - Color/Breed rarity
  - Color rarity

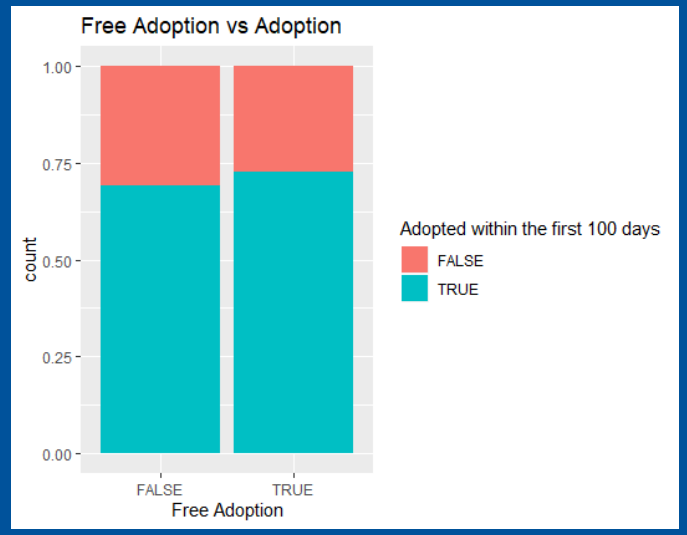


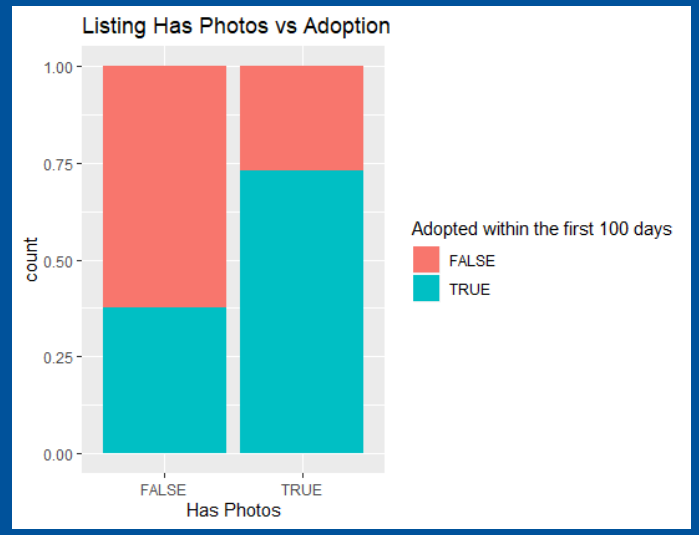


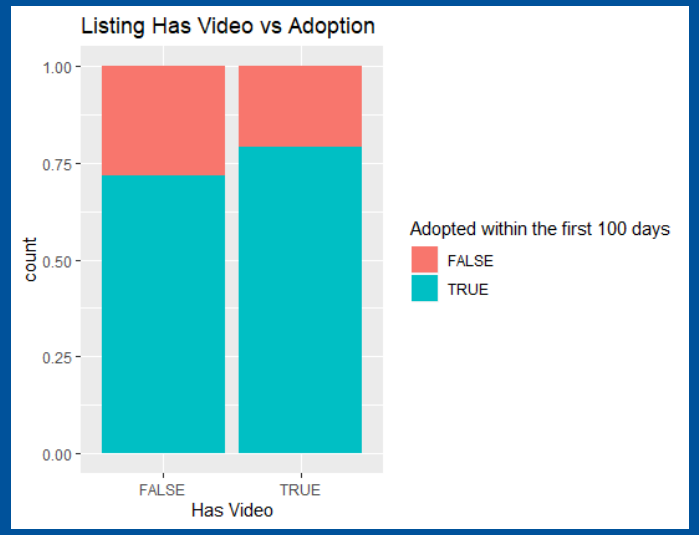


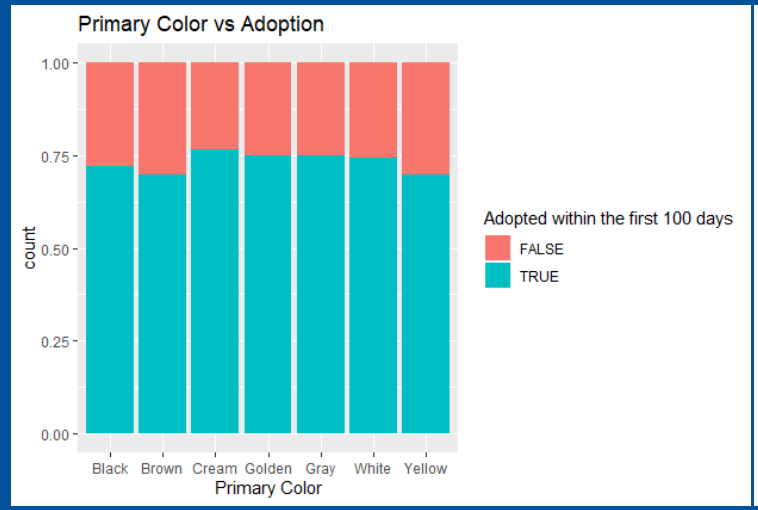


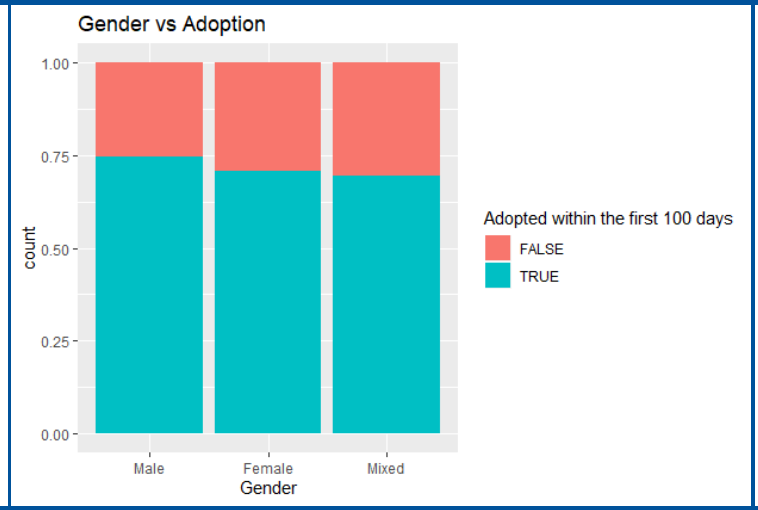
Exploratory Data Analysis

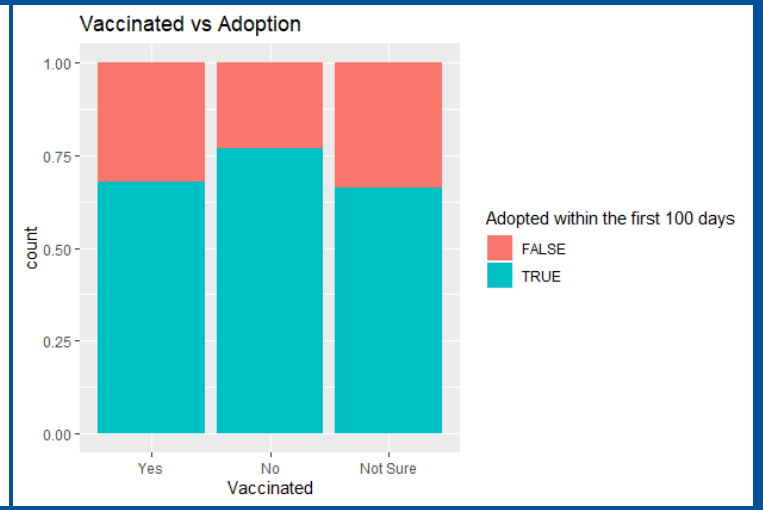












Model: Logistic regression

- Main significant results:

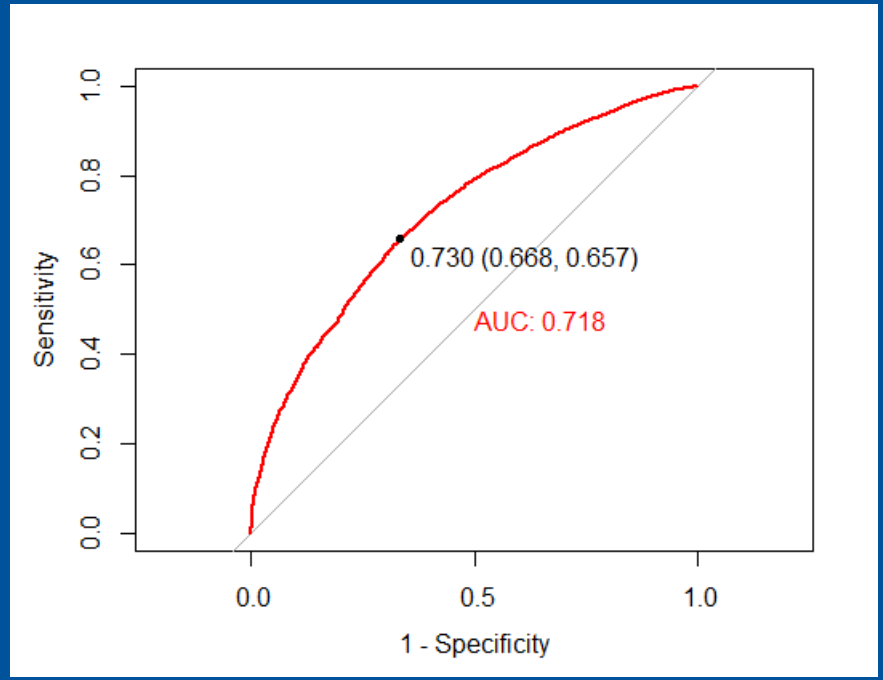
Variable	Type of Change	Effect on likelihood of being adopted within 100 days
Age	A shift in the life stage (Puppy, Junior, Mature, Senior)	↓
Group	If the listing is a group of pets	↓
Free Adoption	If the adoption fee is zero	↑
Has Photos	If the listing has photos	↑
Breed/Color Rarity	The breed and color of the pet is not frequent	↑

Variable	Type of Change	Effect on likelihood of being adopted within 100 days
Type	Being a Cat	↑
Gender	Being a female	↓
Multicolor	If the pet has more than one color	↑
Description Length	Higher number of characters in the listing description	↑
Number of rescues	Rescues made by the listing creator	↑

Duke MIDS
Master in Interdisciplinary Data Science

Model: Logistic regression

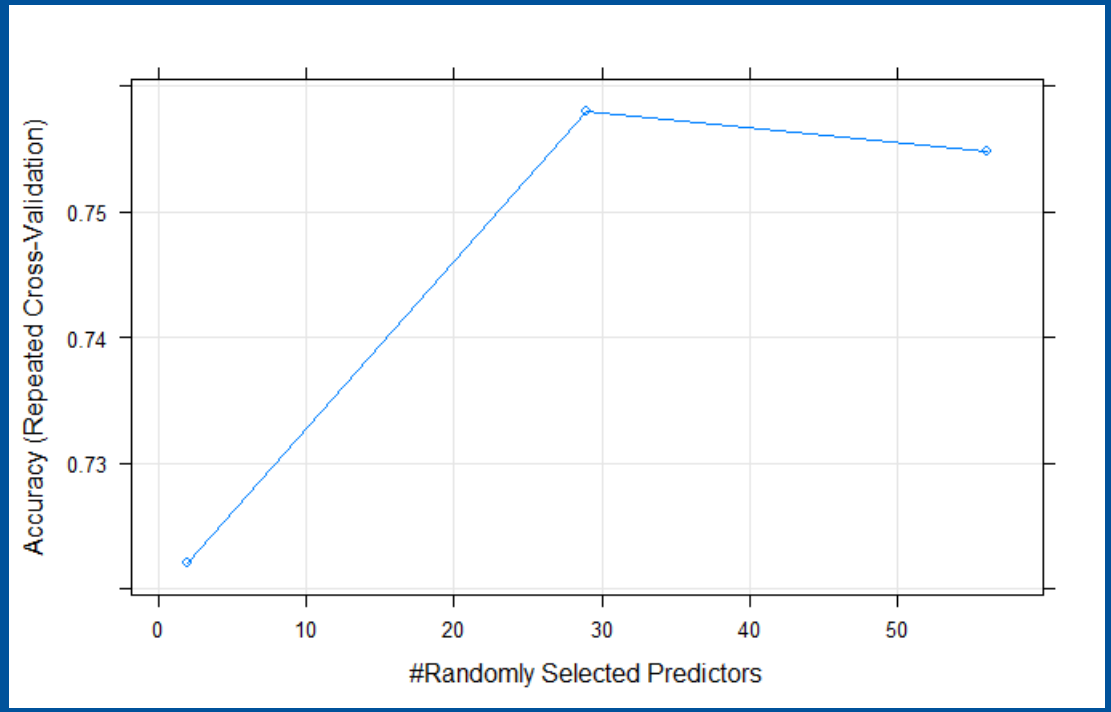
- Baseline: 44 variables
- Backward BIC Stepwise selection: 14 variables:
  - Age
  - Group
  - FreeAdoption
  - HasPhotos
  - PrimaryColor
  - Type
  - Gender
  - MaturitySize
  - FurLength
  - Sterilized
  - Multicolor
  - Breed/Color Rarity
  - DescriptionLength
  - NumRescues
- CV Accuracy: 74.05%



Duke MIDS
Master in Interdisciplinary Data Science

Model: Random Forest

- Repeated CV accuracy: 75.8%
- Random Variables selected for each tree: 29 (out of 42)
- Number of trees: 500



Duke MIDS
Master in Interdisciplinary Data Science

Conclusions & Action Items

- To increase the owner-pet match the listings should require to upload a picture
- The system should not allow the creation of Group listings, instead, there should be created a separate listing for each pet
- Zero adoption fee should be implemented for pets that have a lower likelihood of being adopted within 100 days
- The rescuers should be motivated to write longer descriptions to increase the chances of adoption for the pet
- Experienced rescuers could play a role in the adoption of unusual pets

Duke MIDS
Master in Interdisciplinary Data Science

Thank you!

Duke MIDS
Master in Interdisciplinary Data Science