

---

# PREDICTING TIME TO ADOPTION FOR SHELTER DOGS



---

## BACKGROUND

- ▶ Over 4 million dogs enter animal shelters in the U.S. each year
- ▶ Shelters often have limited financial resources
  - ▶ Funded by government or donations
- ▶ How can we help shelters to spend limited resources effectively?

---

## WHAT CAN WE DO WITH DATA SCIENCE TO HELP THEM?

- ▶ Try to help them identify which dogs might take longer to be adopted, and therefore cost more to shelter
- ▶ Shelters can spend their limited resources focused on speeding up time to adoption for these dogs

---

## GOALS

- ▶ Predict, upon its arrival, how long a given dog will take to be adopted
- ▶ Explore whether the dog's name affects this time

---

## WHY NAME?

- ▶ Shelters can't change most of the characteristics of a given dog (size, colour, breed, coat)
- ▶ Shelters can change a dog's name

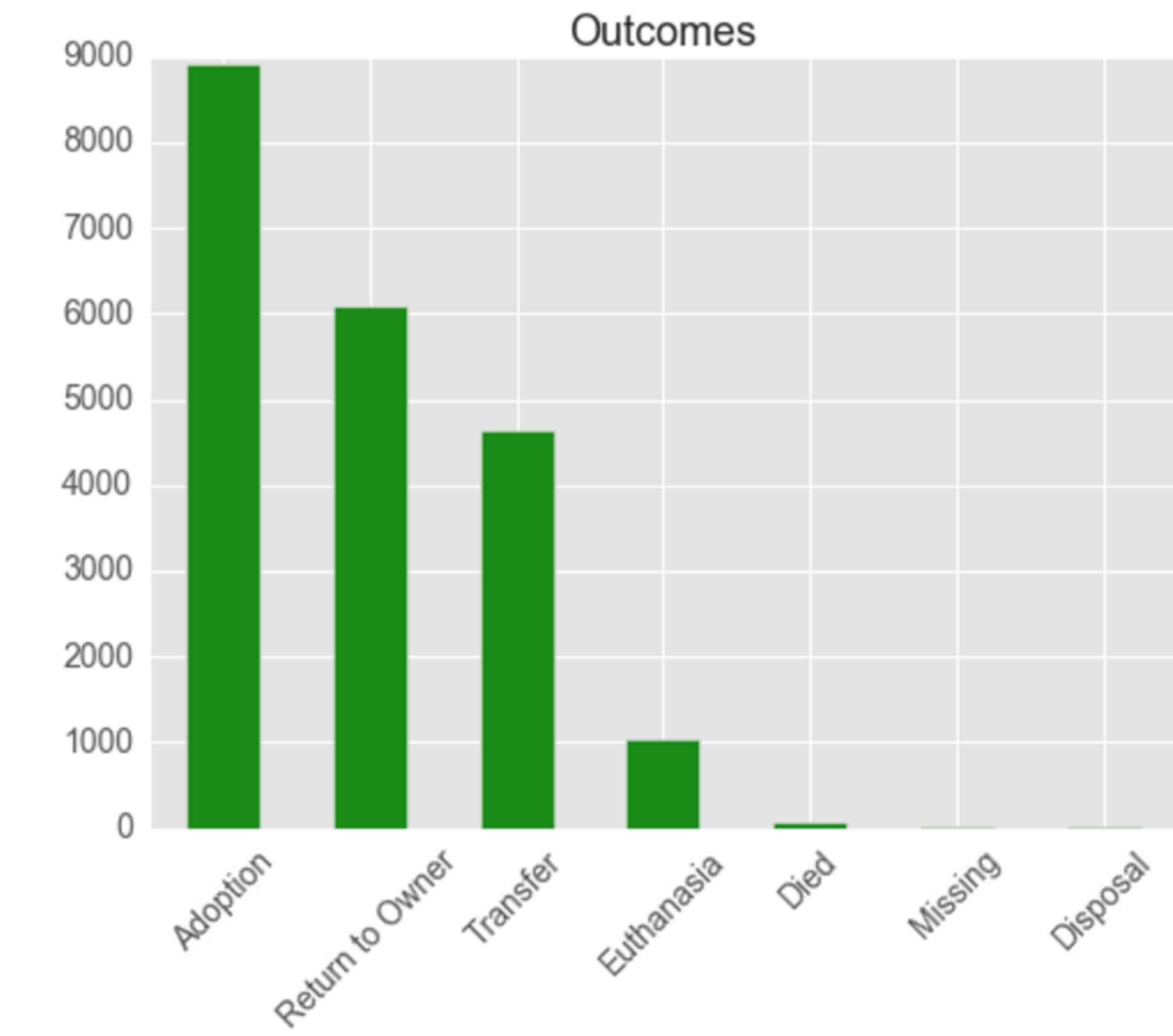
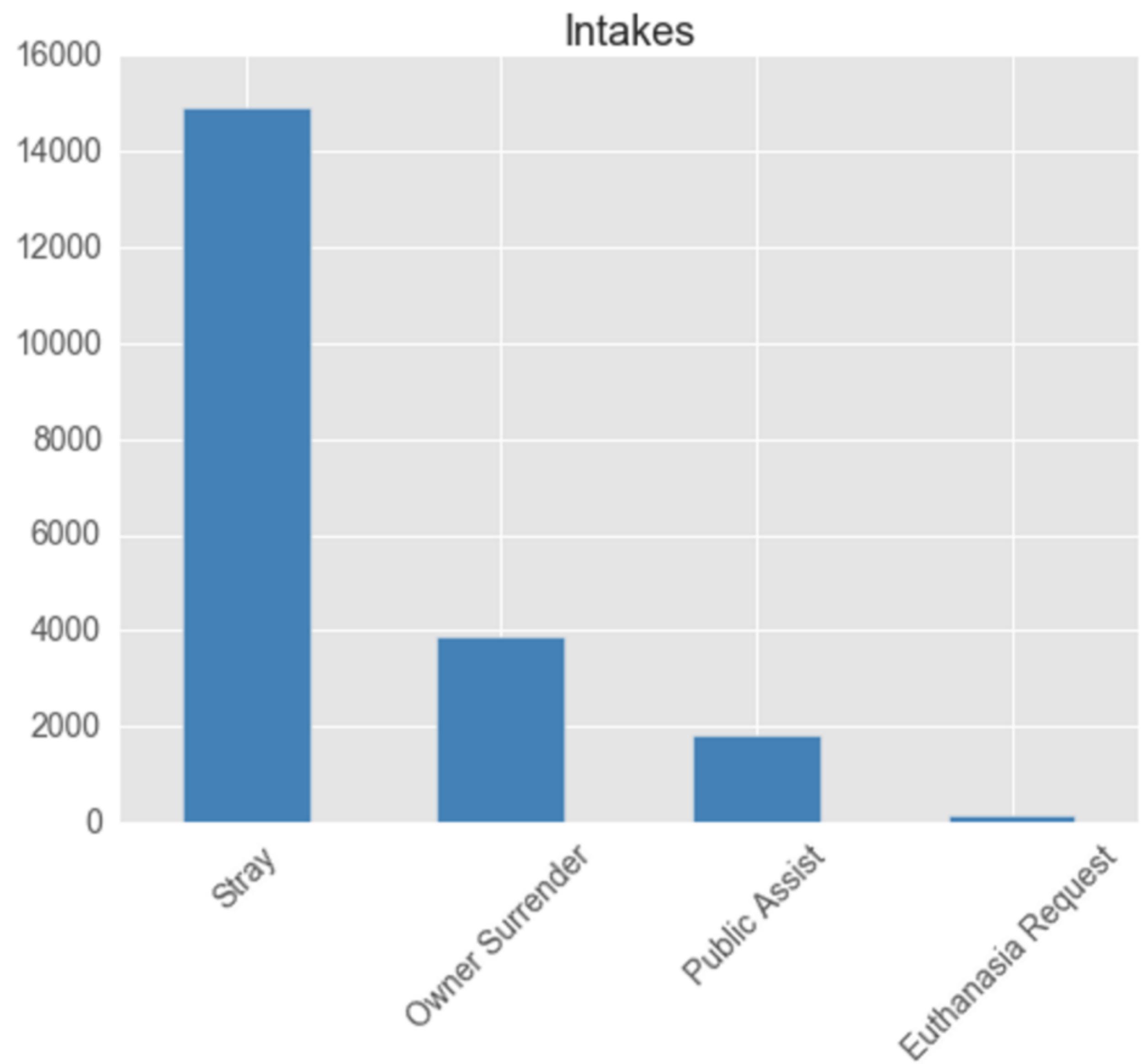
# DATA

---

- ▶ Data from Austin Animal Center
- ▶ No-kill shelter
- ▶ Intakes / Outcomes
  - ▶ Restricted to dogs entering in 2014 / 2015
- ▶ Includes data on:
  - ▶ Dog characteristics (breed, colour, condition, age, name)
  - ▶ Intake circumstances / Outcome type

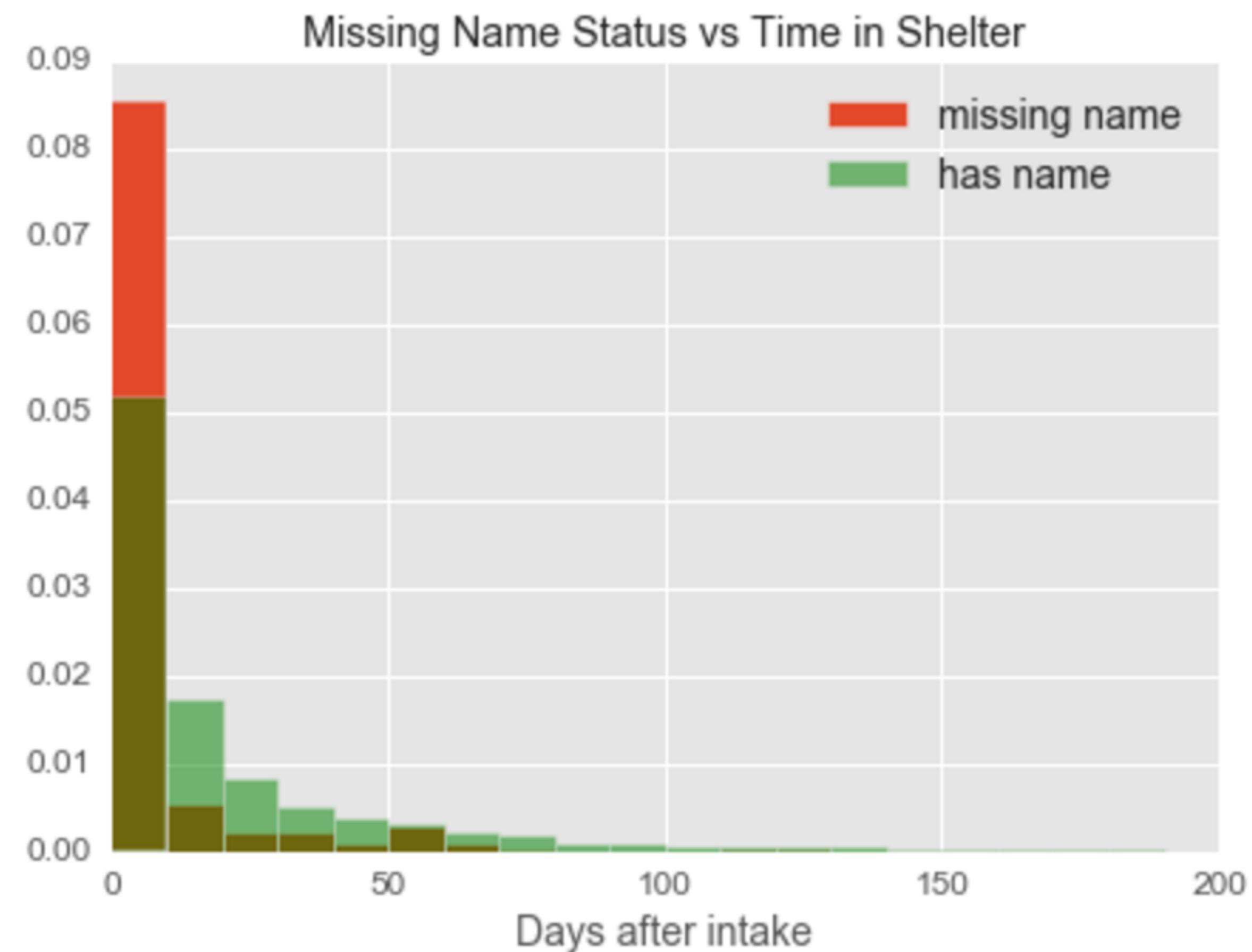


# INTAKES / OUTCOMES

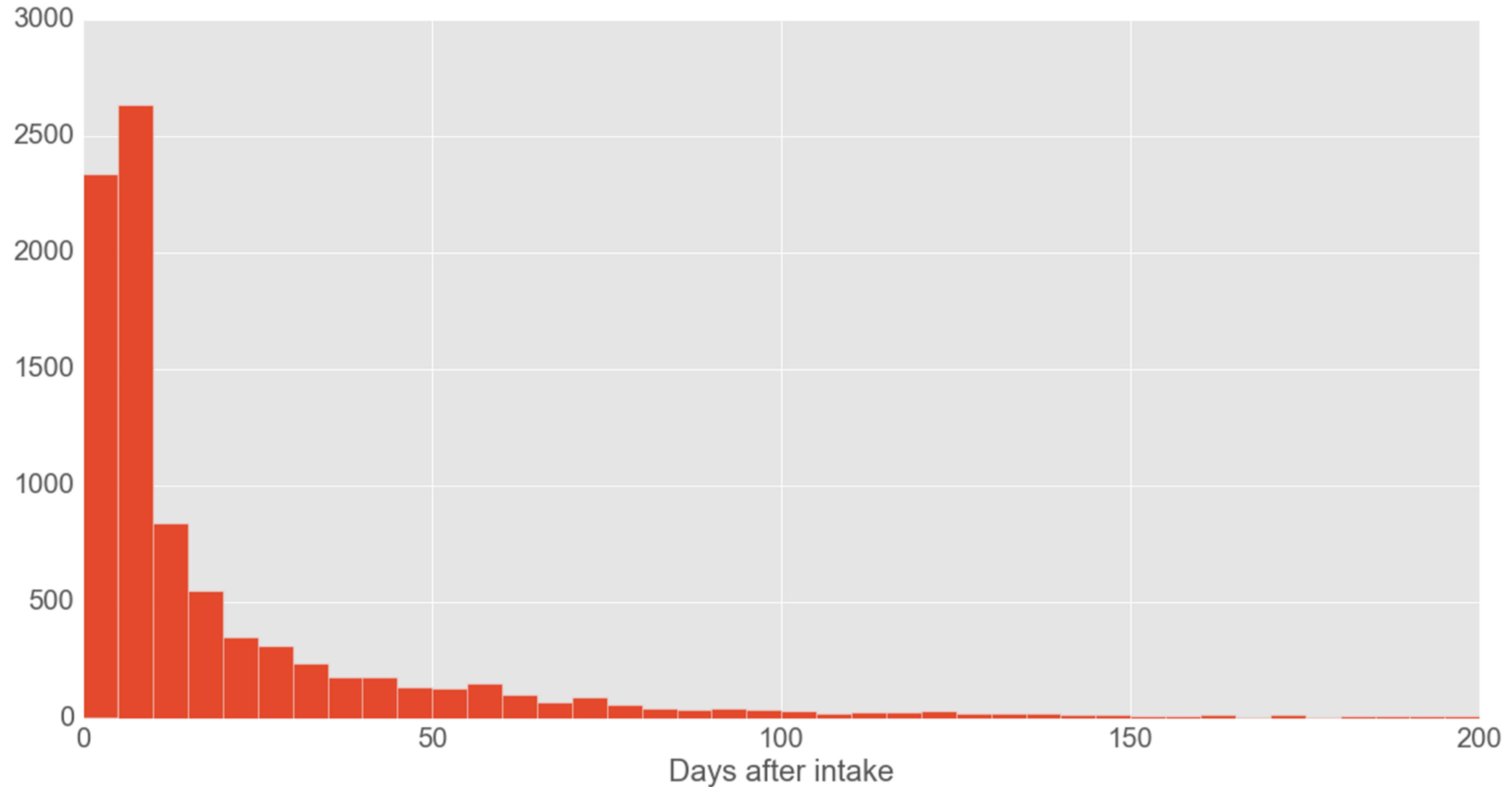


# MISSING DATA: NAMES

- ▶ 14% adopted dogs missing name
- ▶ Not a function of time
- ▶ Higher proportion of dogs with missing names adopted within 1 week
- ▶ Missingness function of outcome?
- ▶ Restricted data just to dogs with names (~7,500)



# TIME TO ADOPTION



---

## FEATURE ENGINEERING: COLOUR

- ▶ Colour feature had little consistency in labelling
  - ▶ *White, Sable/White, Black Brindle/White, Tan*
- ▶ Classified colour words into four groups: dark, medium, light, mixed
- ▶ Bucketed descriptions into six groups:
  - ▶ dark / dark mix
  - ▶ medium / medium mix
  - ▶ light / light mix

---

## FEATURE ENGINEERING: BREED

- ▶ Breed, or breed mix, listed for each dog
  - ▶ *Chihuahua Shorthair Mix, Pit Bull/Labrador Retriever, Pit Bull*
- ▶ Marked each dog as mix / not mix
- ▶ How can we map breed back to dog characteristics, especially for mixed breed dogs?

---

## FEATURE ENGINEERING: BREED

- ▶ Used web scraping to collect breed information from American Kennel Club
  - ▶ AKC Group, coat length, grooming / exercise requirements, energy level, size
- ▶ Matched listed breeds to characteristics (programmatically corrected misspellings)
- ▶ Calculated score for each dog
  - ▶ Average for breed mixes

# EXAMPLE



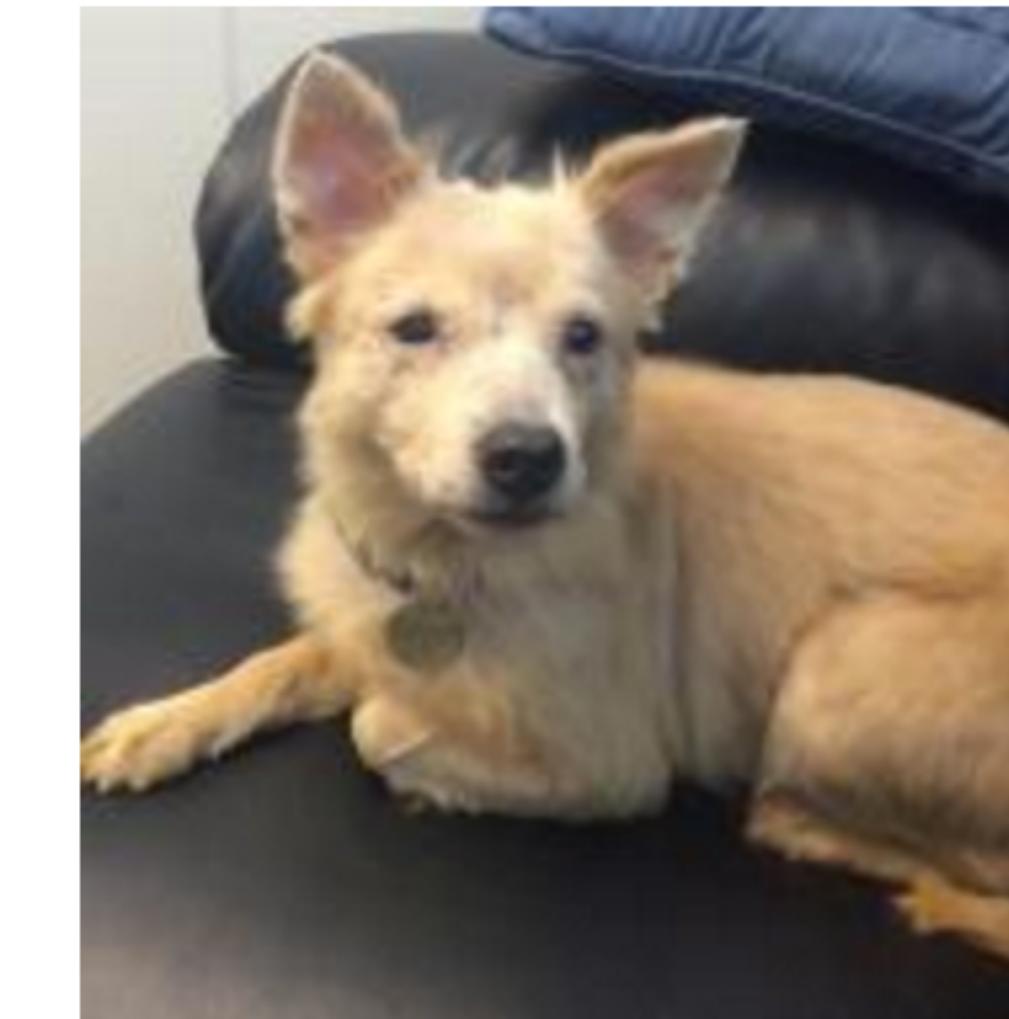
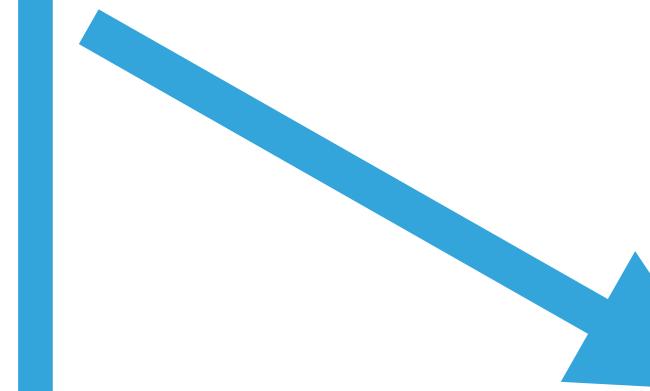
## Shetland Sheepdog

Size:	2
Energy:	2
Exercise Req:	2
Coat Length:	3
Grooming Req:	1



## American Eskimo Dog

Size:	2
Energy:	2
Exercise Req:	2
Coat Length:	1
Grooming Req:	1



BRIANNA

Shetland Sheepdog Sheltie &  
American Eskimo Dog Mix

Size:	2
Energy:	2
Exercise Req:	2
Coat Length:	2
Grooming Req:	1

# FEATURE ENGINEERING: NAMES



TEENSIE

Is it an English word?

VELVET



BUDDY

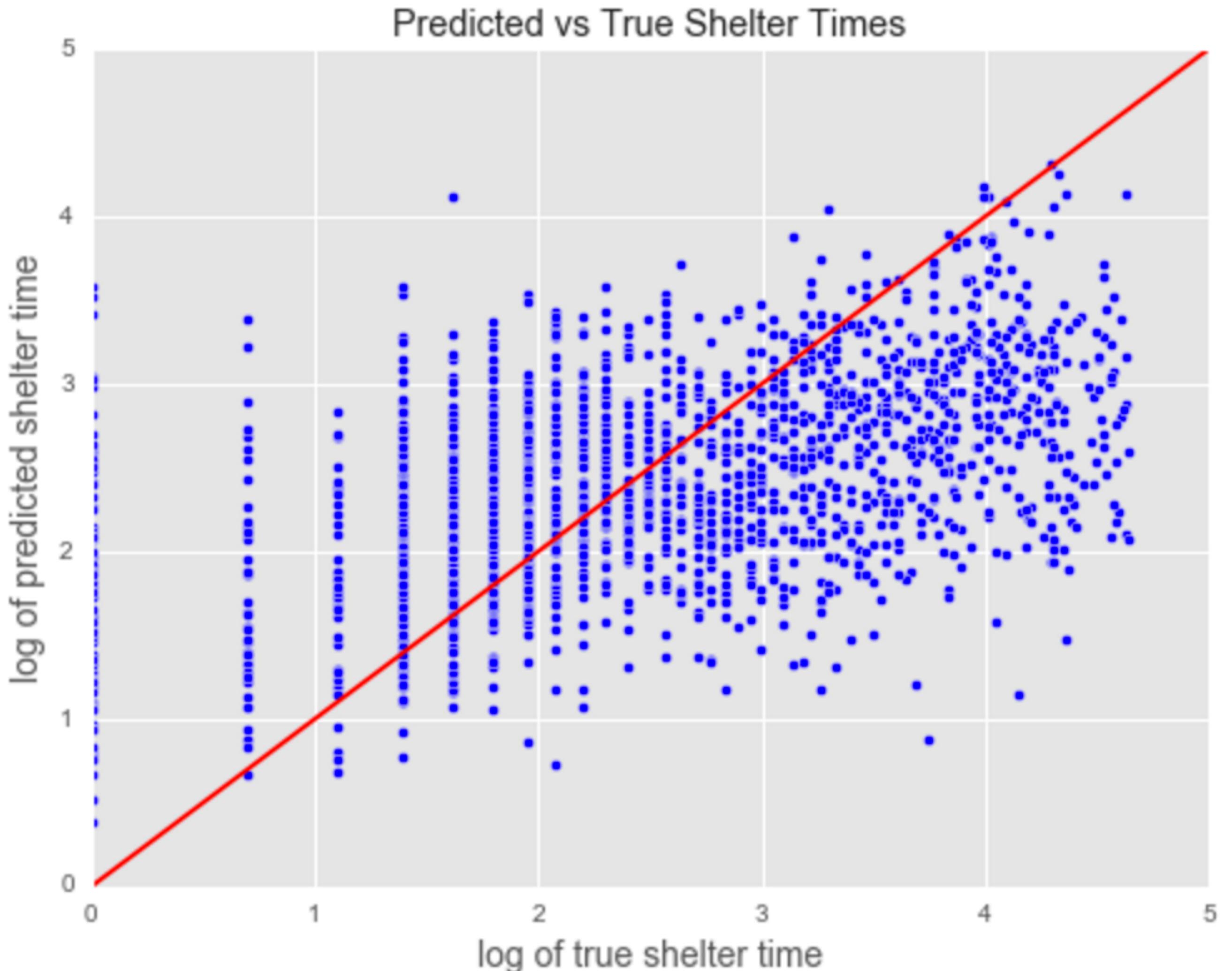
Is it a human name?  
Is it a popular human name?

EMMA



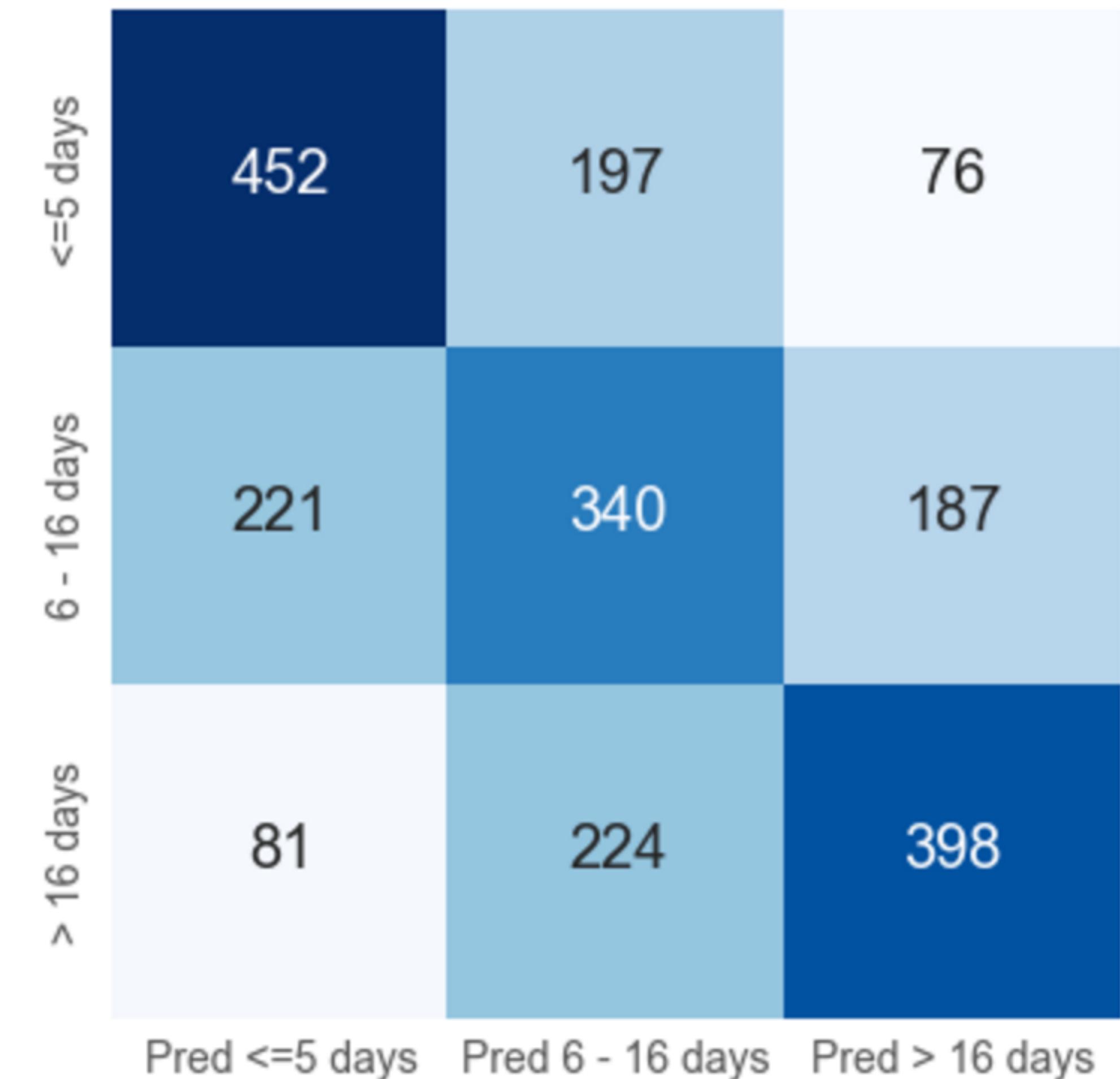
# REGRESSION ANALYSIS

- ▶ Goal to predict actual number of days spent in shelter until adoption
- ▶ Gradient boosting regressor performed best
- ▶ Predicted  $\log(\text{time in shelter})$
- ▶ Average R<sup>2</sup> of 0.32



# CLASSIFICATION ANALYSIS

- ▶ Divide time to adoption into three class:
  - ▶  $\leq 5$  days, 6-16 days,  $> 16$  days
- ▶ Gradient Boosting Classifier
- ▶ Precision for classes 1 / 3 = 60%
- ▶ Precision for class 2 = 45%



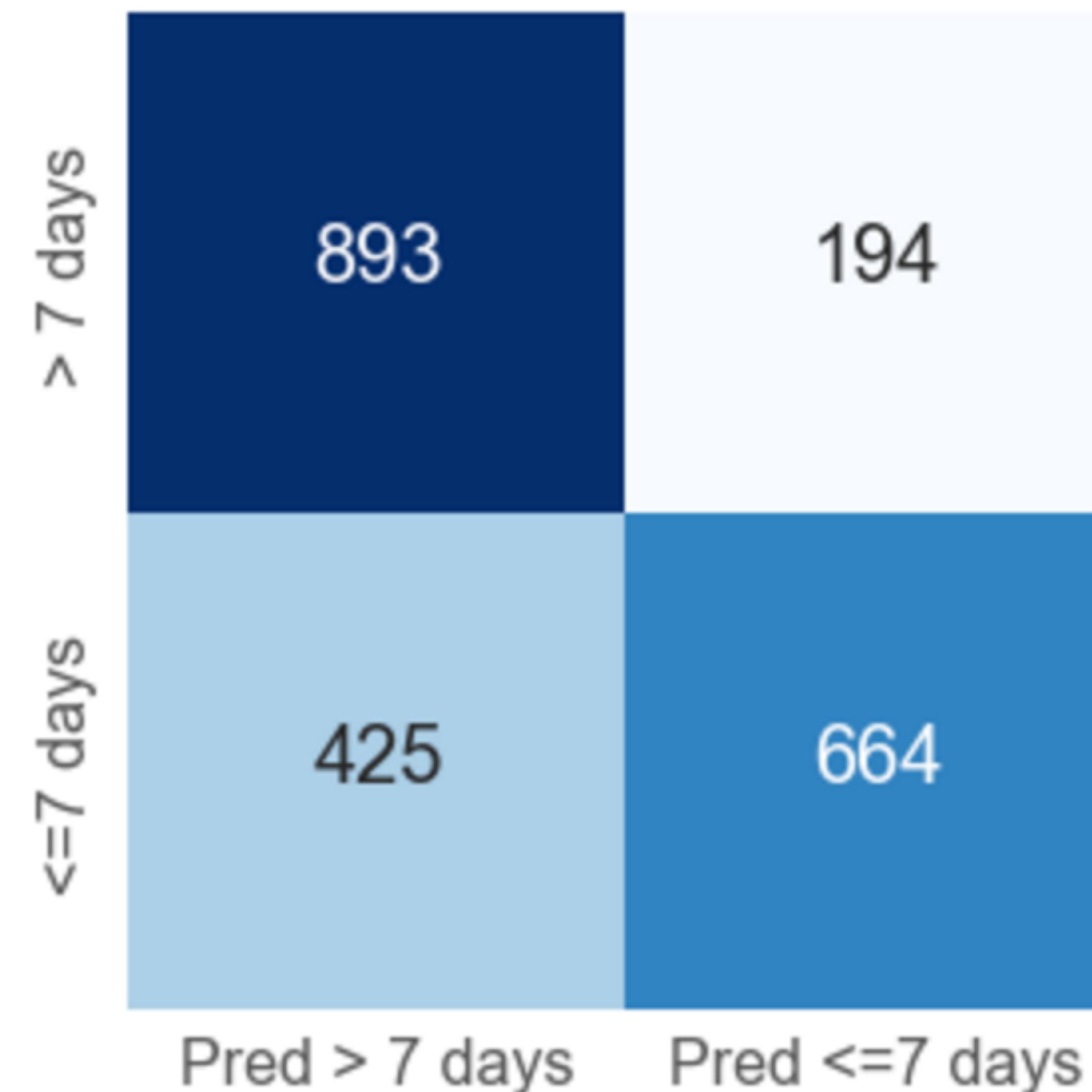
# CLASSIFICATION ANALYSIS

- ▶ Divide outcome into 2 class:
  - ▶ Within 8 days
  - ▶ Longer than 8 days
- ▶ 72% accuracy
- ▶ 71% recall
- ▶ 73% precision

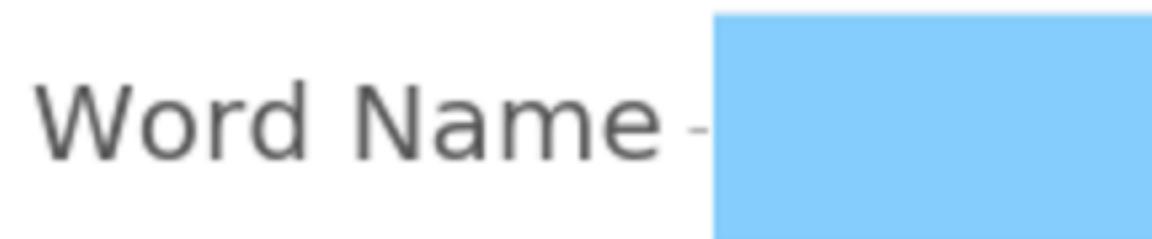
	<=8 days	> 8 days
Pred > 8 days	296	776
Pred <=8 days	311	793

## CLASSIFICATION ANALYSIS: FURTHER TUNING

- ▶ Is identifying more “high-risk” dogs important? Can we sacrifice some overall accuracy?
- ▶ Tune the threshold to 40%
- ▶ Recall increased to 82%
- ▶ But precision dropped to 68%



## FEATURE IMPORTANCE: TOP 5

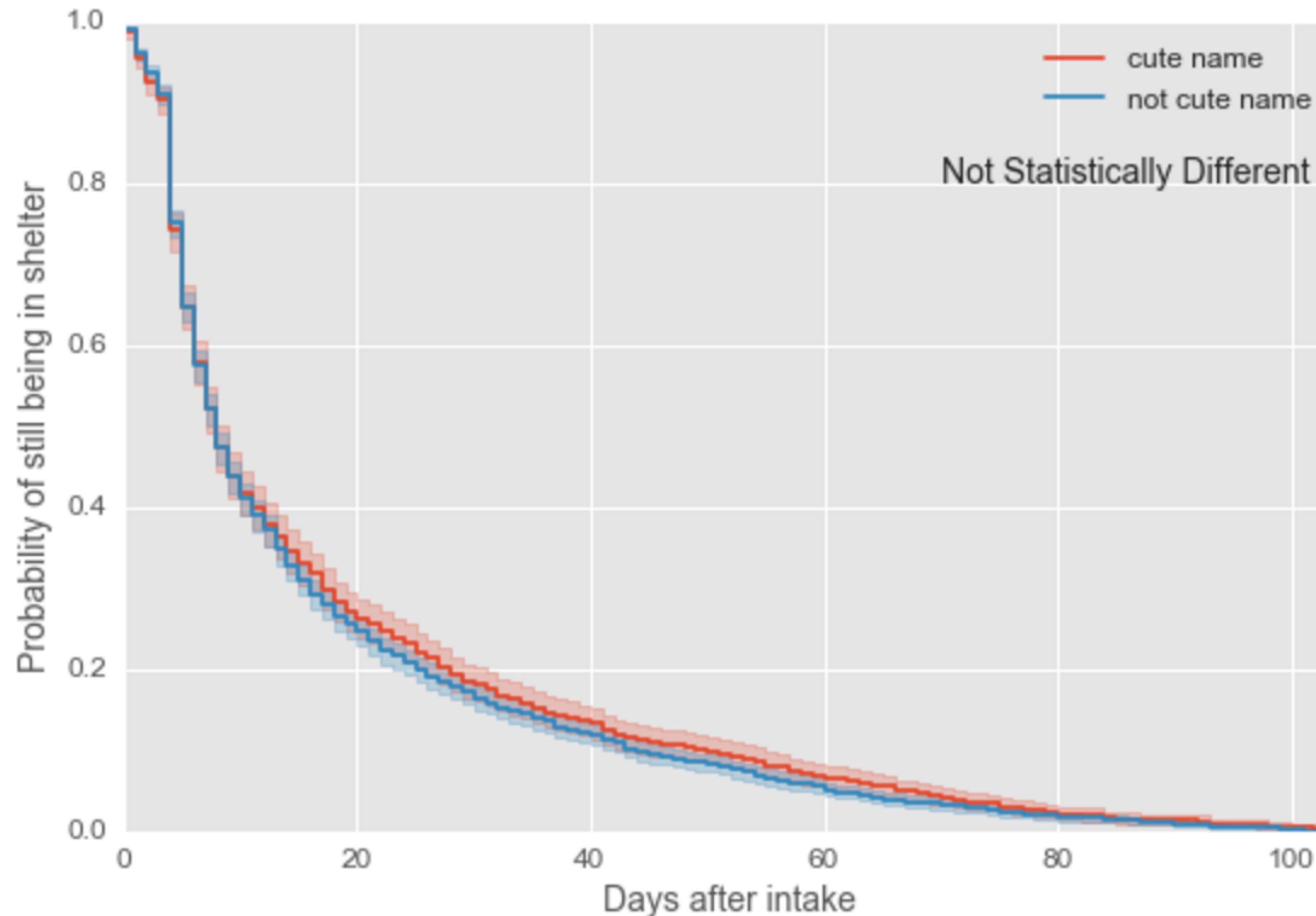


---

## FEATURE IMPORTANCE: NAME

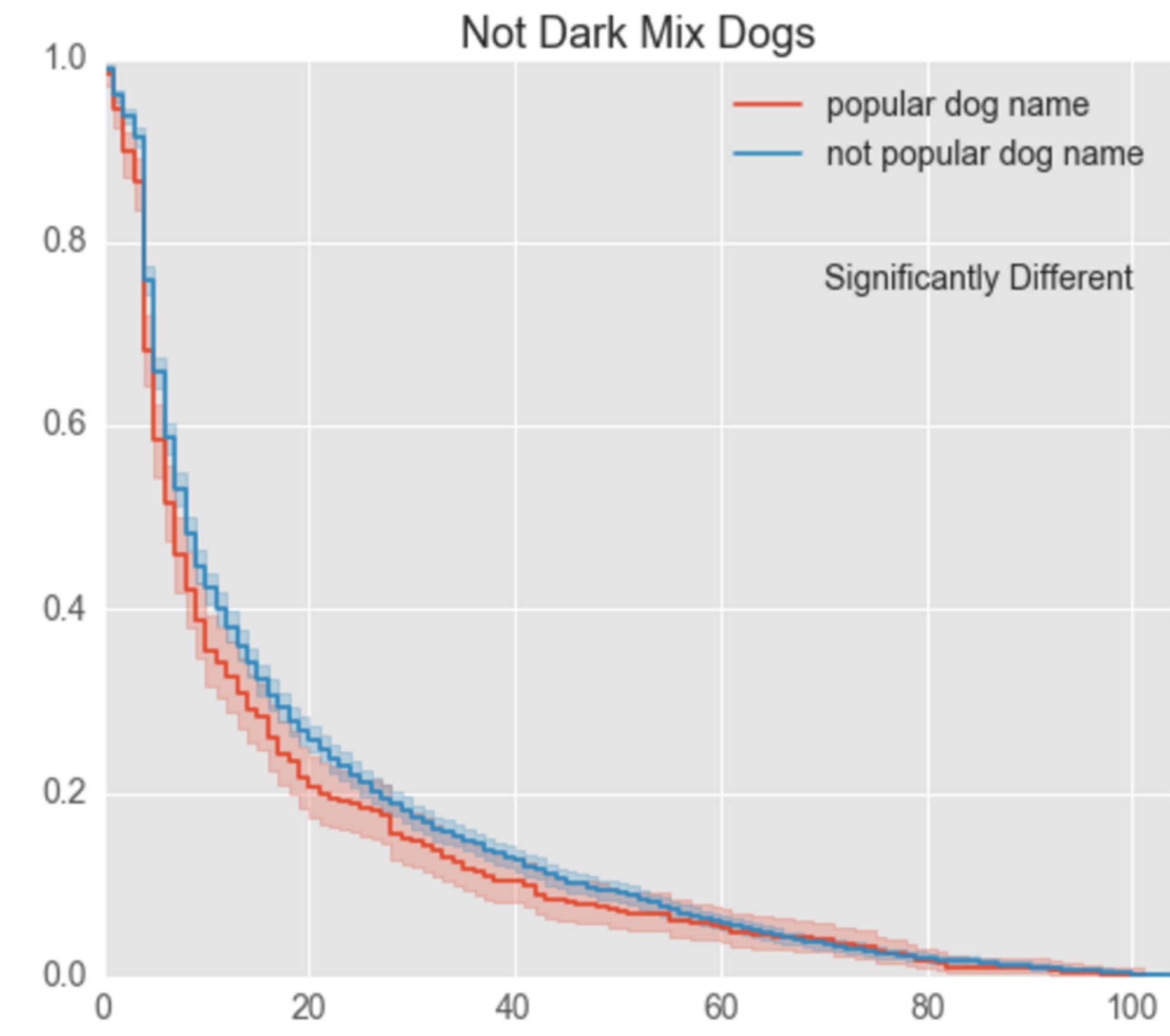
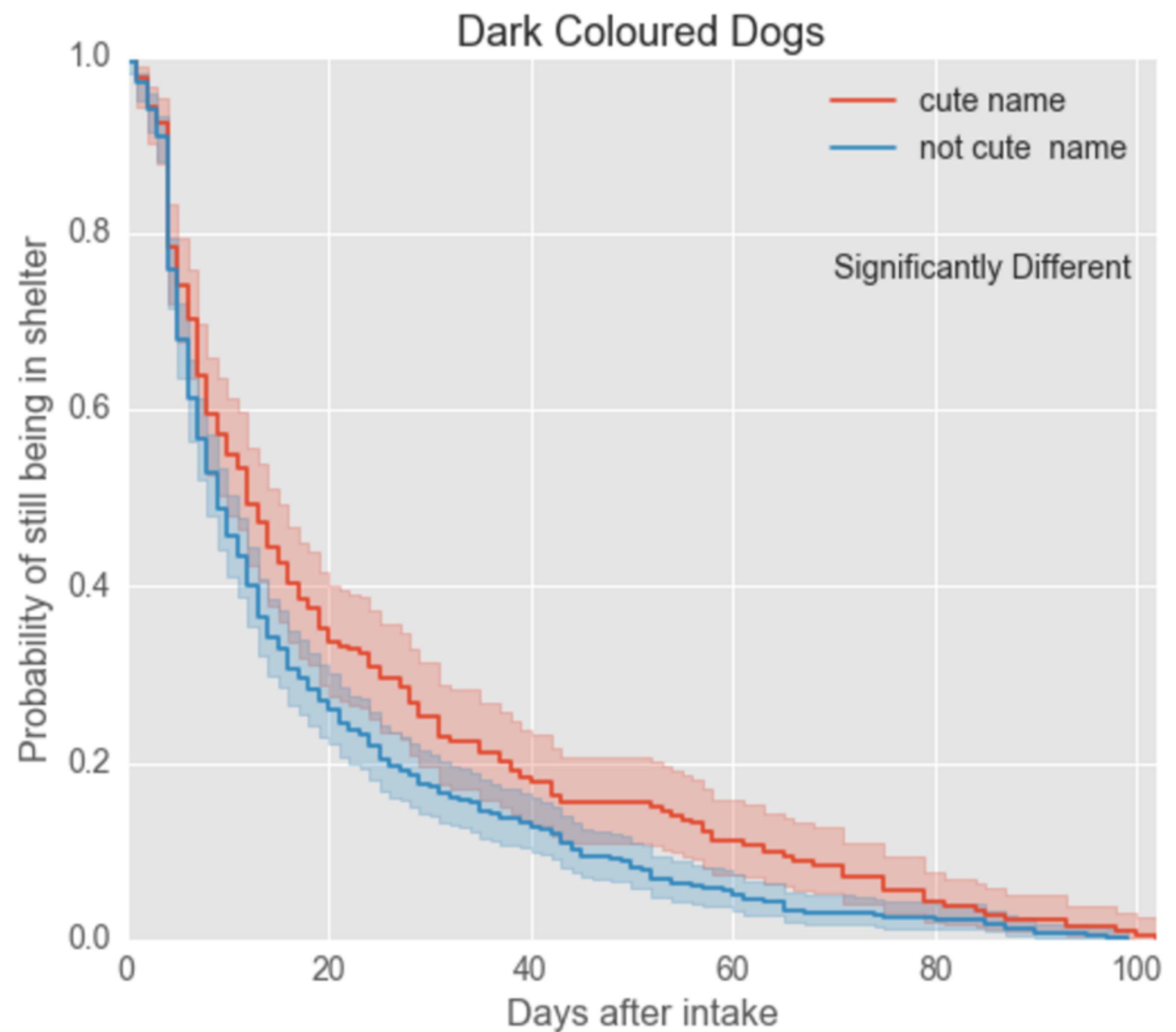
- ▶ Three name characteristics in top 15
  - ▶ English Word Name
  - ▶ Cute Name
  - ▶ Human Name
- ▶ But effect of name is complicated

# COMPARING ACROSS ENTIRE SAMPLE



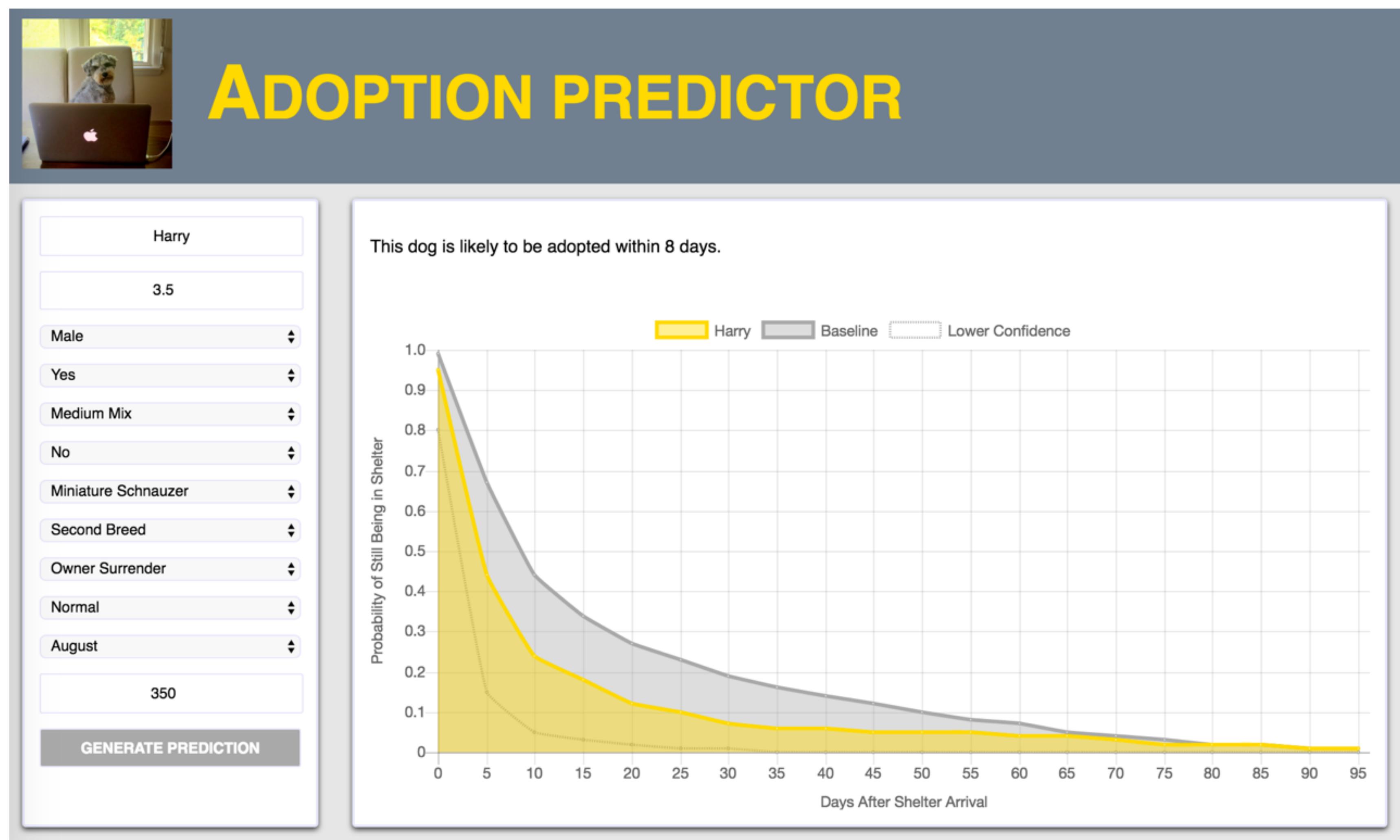
- ▶ Survival analysis shows no difference between dogs with cute names and not cute names
- ▶ Same result for other name features

# WITHIN CERTAIN GROUPS



# DEPLOYMENT TO SHELTERS

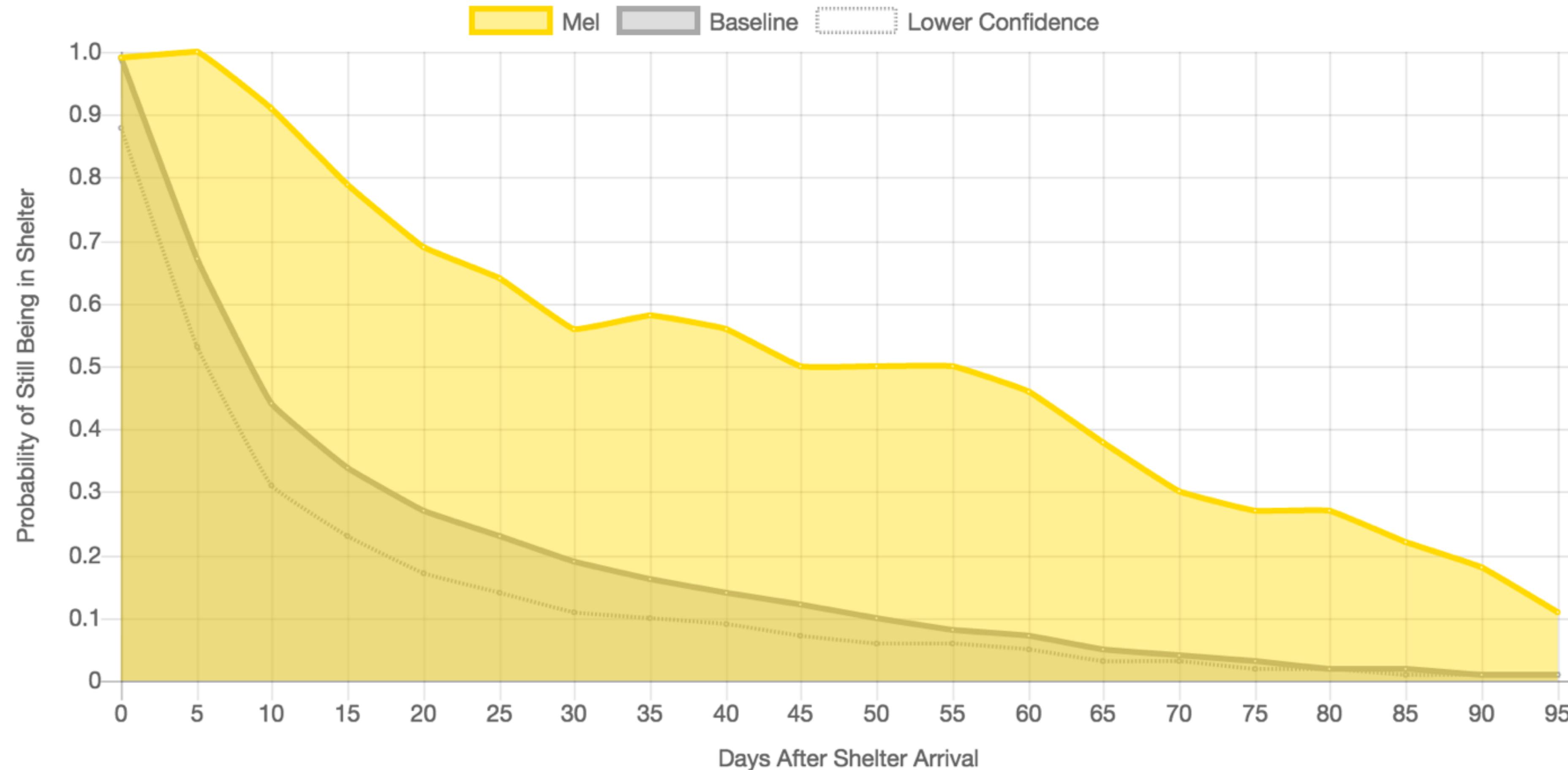
- ▶ User inputs information upon intake
- ▶ Classification model predicts whether dog will be adopted within 8 days
- ▶ Survival regression shows probability over time



# NAME RECOMMENDATIONS WHERE APPROPRIATE

Mel
15
Female
No
Light Mix
No
Beagle
None
Owner Surrender
Sick
April
250
<b>GENERATE PREDICTION</b>

This dog is likely to take longer than 8 days to be adopted. How about you call it Brownie?



---

## NEXT STEPS

- ▶ Continue to update model with new data
- ▶ Use data from other shelters
  - ▶ Limited generalizability currently