

# Paw Predictors

Predicting Adoption Speed  
for Shelter Pets



Capstone Presentation  
neuefische Data Practitioner Bootcamp  
19th September 2023

# Who we are



**Anna Stein**

Sociologist (M.A.)



**Qurratulain Khaleeq**

Bioinformatician (M.Sc.)



**Jan Schlautmann**

Conservation ecologist  
(M.Sc.)



**Lana Casselmann**

Mathematician (PhD)

# Background

- Many pets don't have a home
- Caring for pets in shelters is costly and the facilities are chronically underfunded
- If shelters run out of resources, pets end up strays or even are euthanized
- Not neutered strays worsen the problem

# Goals

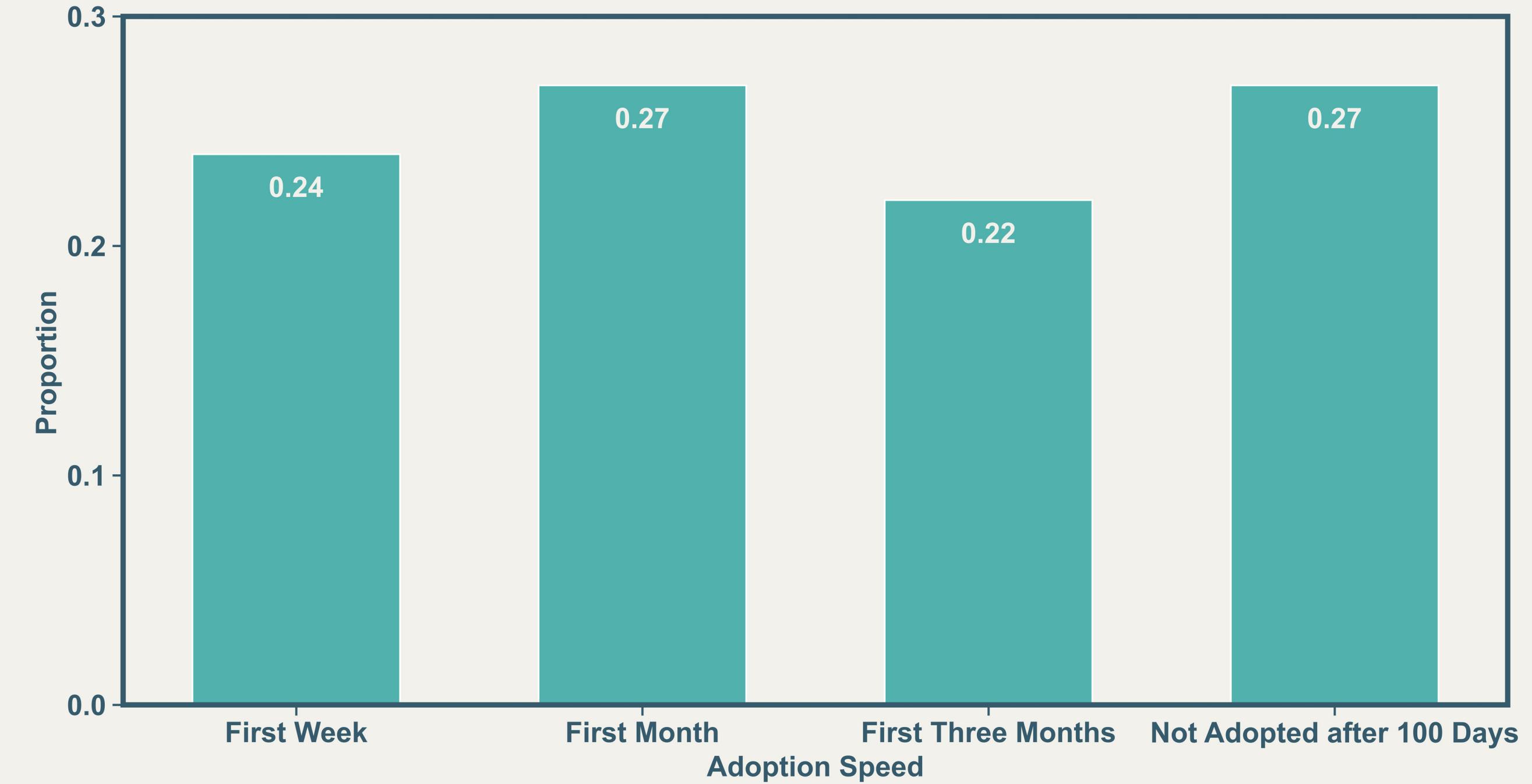
- We identify factors that influence adoption speed and derive recommendations on how to increase it
- We develop an algorithm to predict the time an animal will spend in shelters, thus allowing to plan ahead

# Data

- Roughly 15.000 animals listings (54 % dogs, 46 % cats) from PetFinder.my - Malaysia's leading animal welfare platform
- Over 300 breeds, 16 % pure, 84 % mixed
- Listings include features as age, gender, fur color and length and animal size, as well as health information, photos and text descriptions of the pets

# Target: Adoption Speed

- 4 ordered classes
- evenly distributed



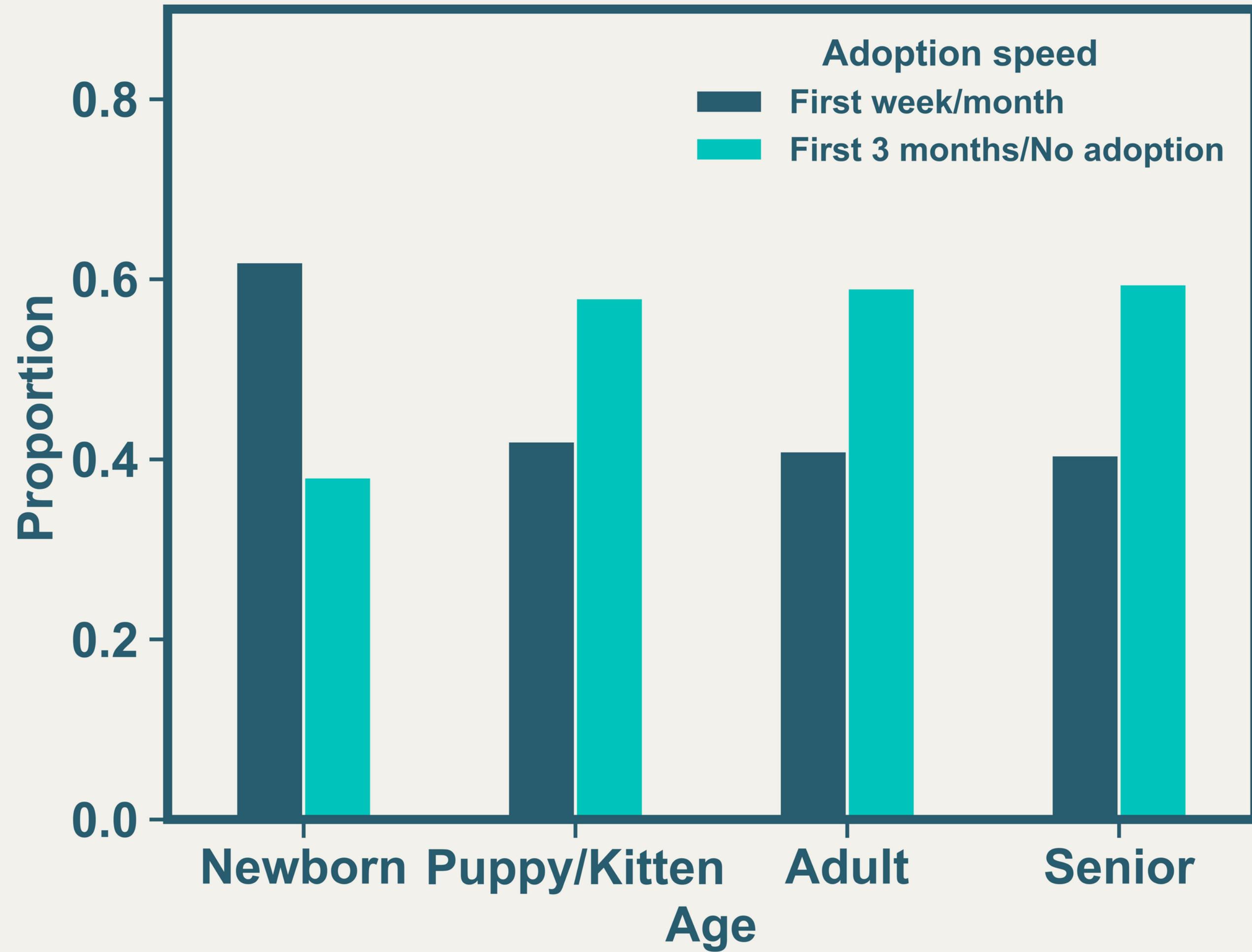
# Feature Engineering

- Age: newborn (0-3 mos), puppy/kitten (4-12 mos), adult (13-72 mos), senior (73 mos and older)
- Amount of Photos: 0-10 photos, 11+ photos (very few instances)
- Description character count
- Breed: purebred, mixed
- ...

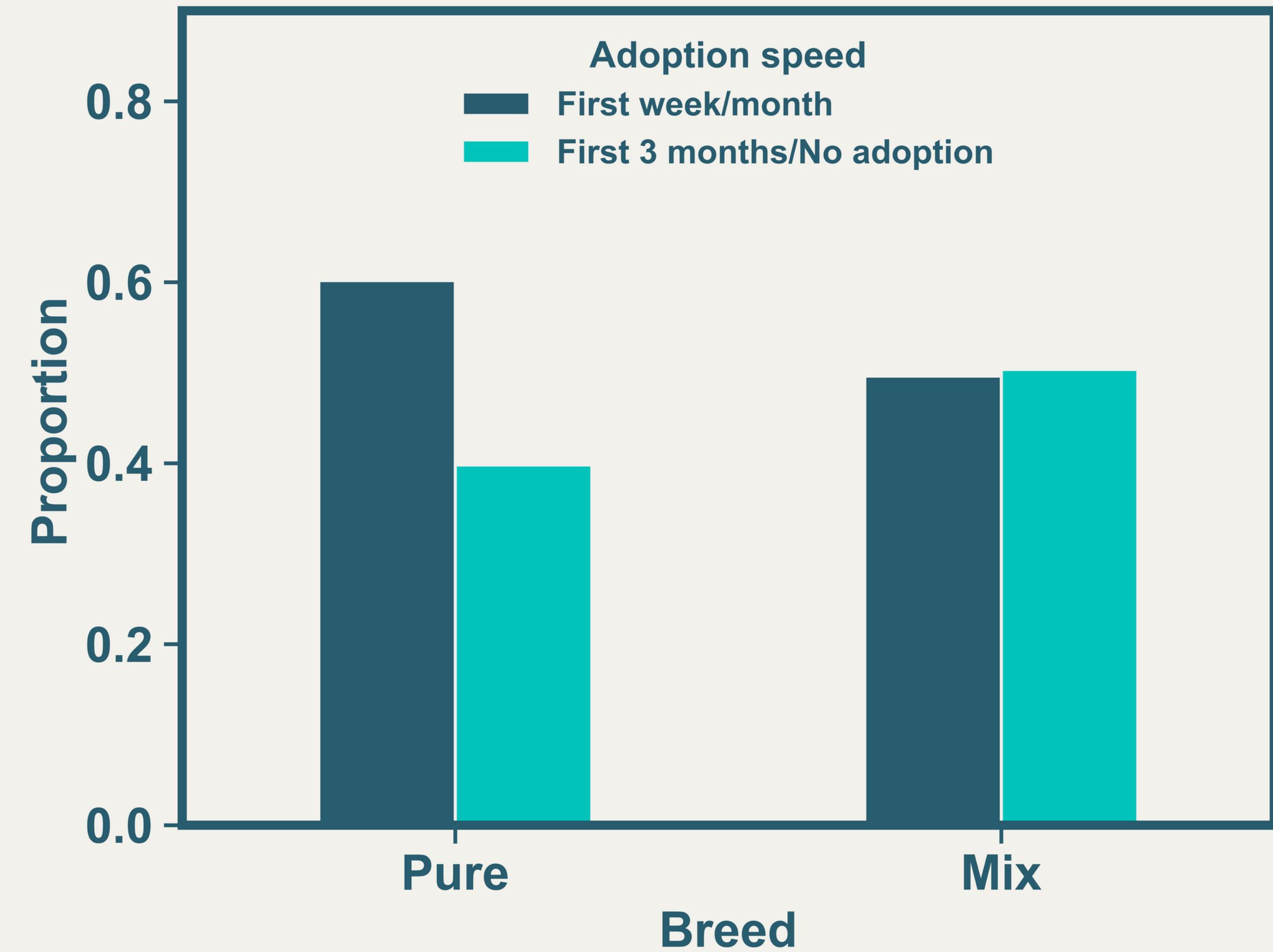
# Hypotheses

- Older animals are adopted slower than younger ones
- Purebred animals are adopted faster than mixed animals
- The amount of animals' photos increases adoption speed

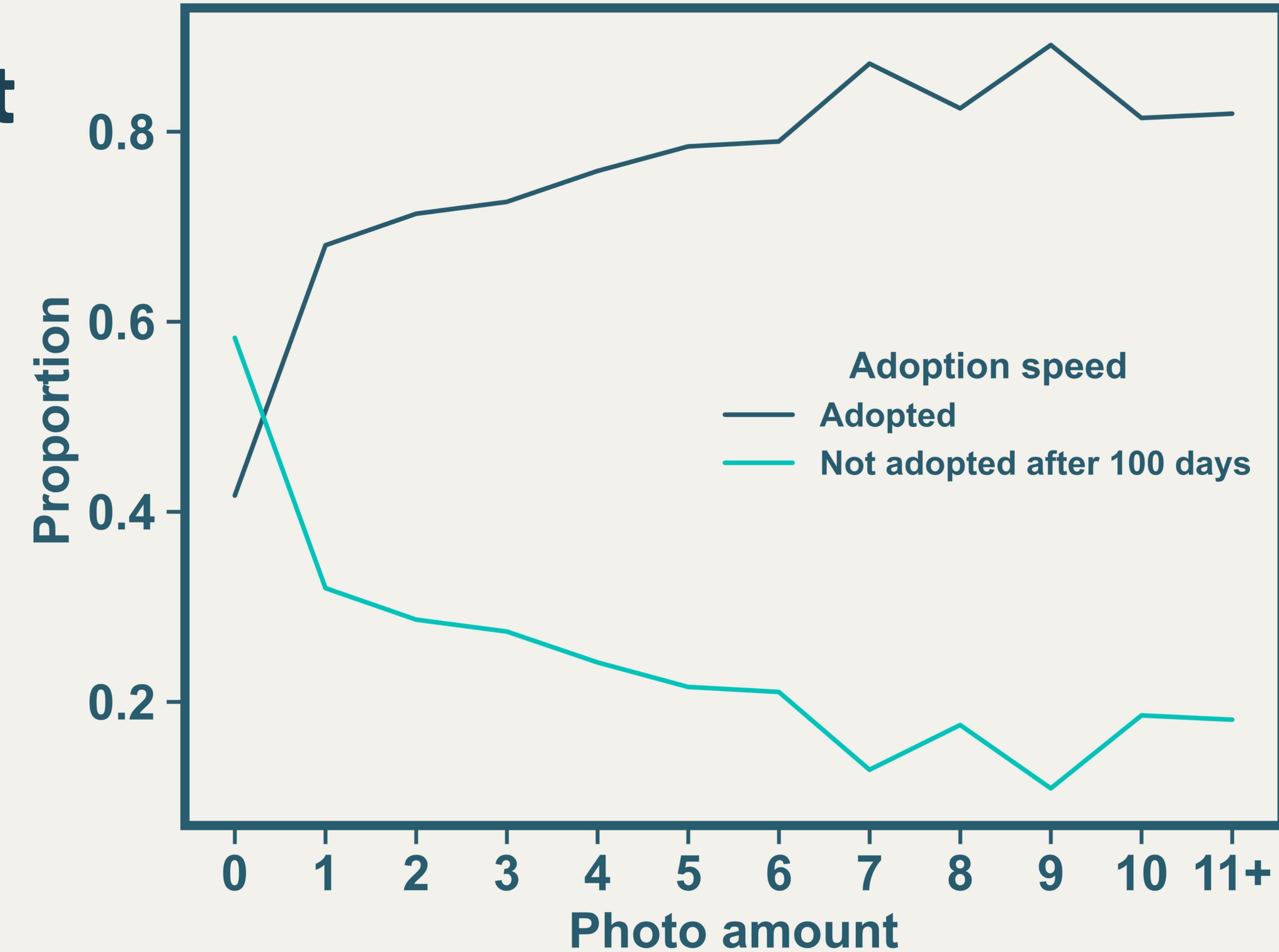
**Younger  
Animals are  
Adopted  
Faster**



# Purebreeds are Adopted Faster

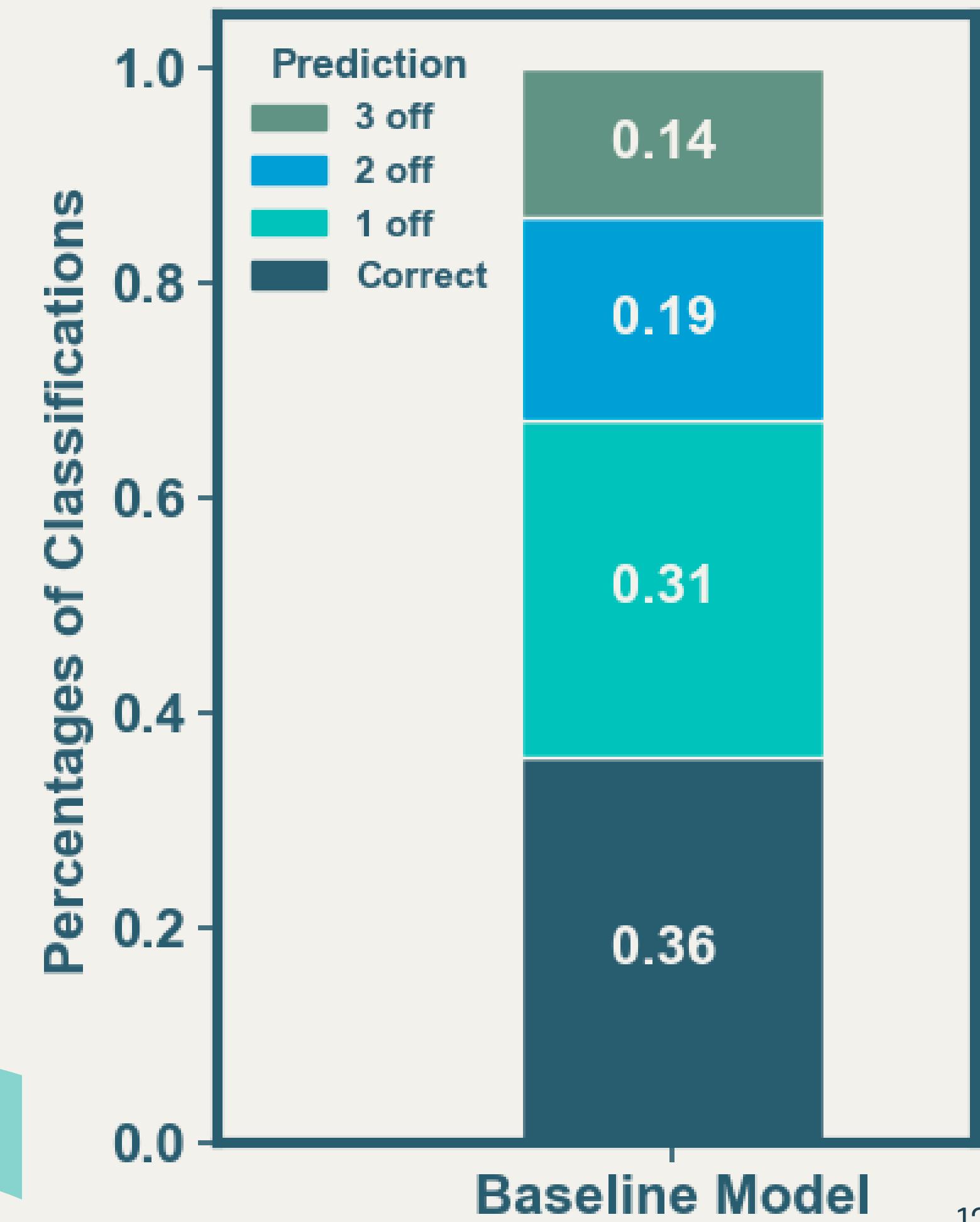


# Photo Amount Influences Adoption Speed



# Baseline Model

- Based on features age, photo amount
- Corresponding to findings of EDA
- Basic decision tree

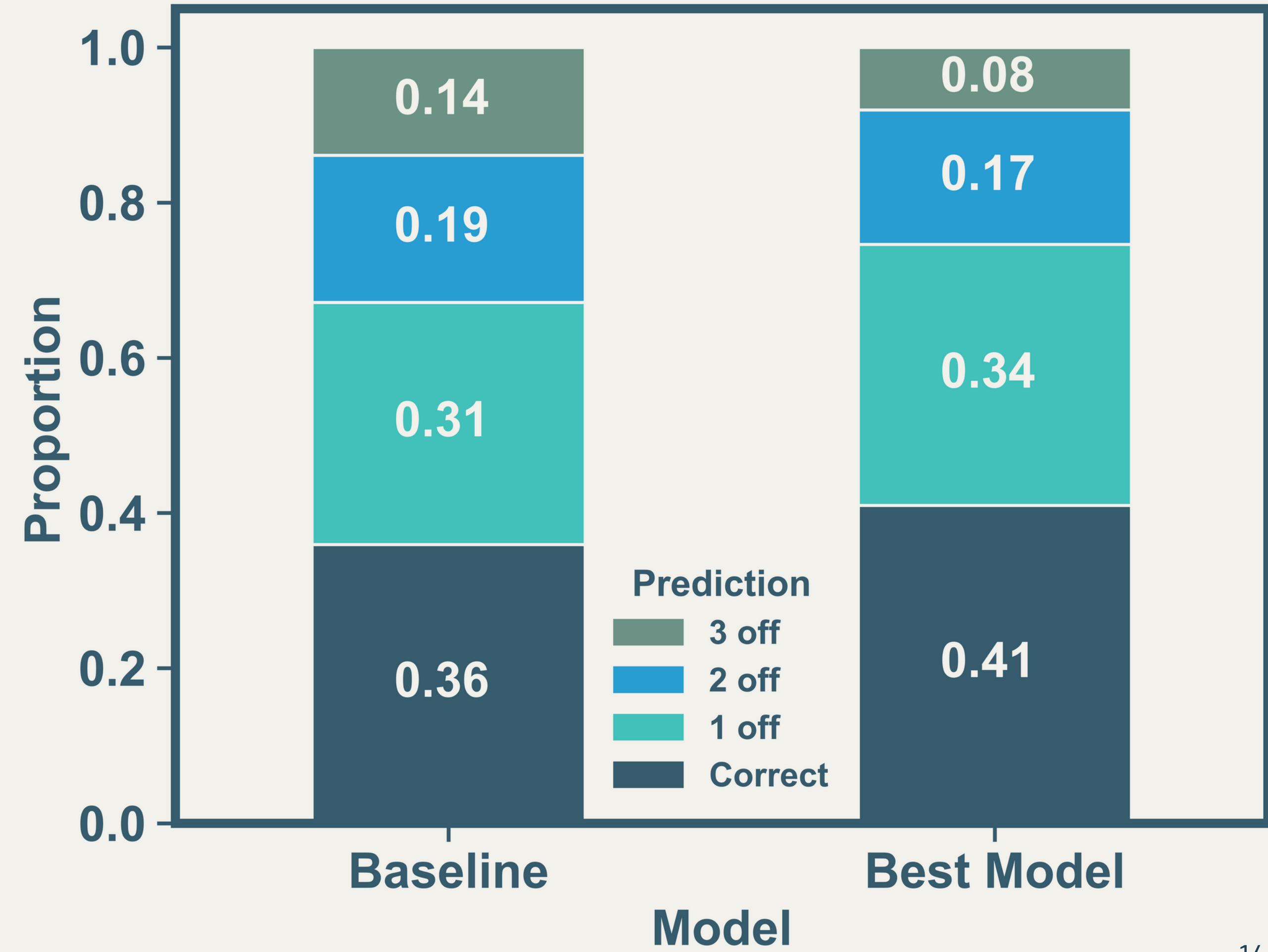


# Decision Metric and Best Model

- **Cohen's Kappa, or the Kappa score:** Measure of agreement between predictions and actual classes in **multi-class** classification.  
-1 = No agreement, 0 = Chance agreement vs 1 = Perfect agreement.
- It penalizes predictions that are further from the actual class labels more.
- Several models were tested and **best model: Gradient booster**, based on trees.
- **Best model's Kappa score** is 0.35 as compared to 0.22 of the baseline model.

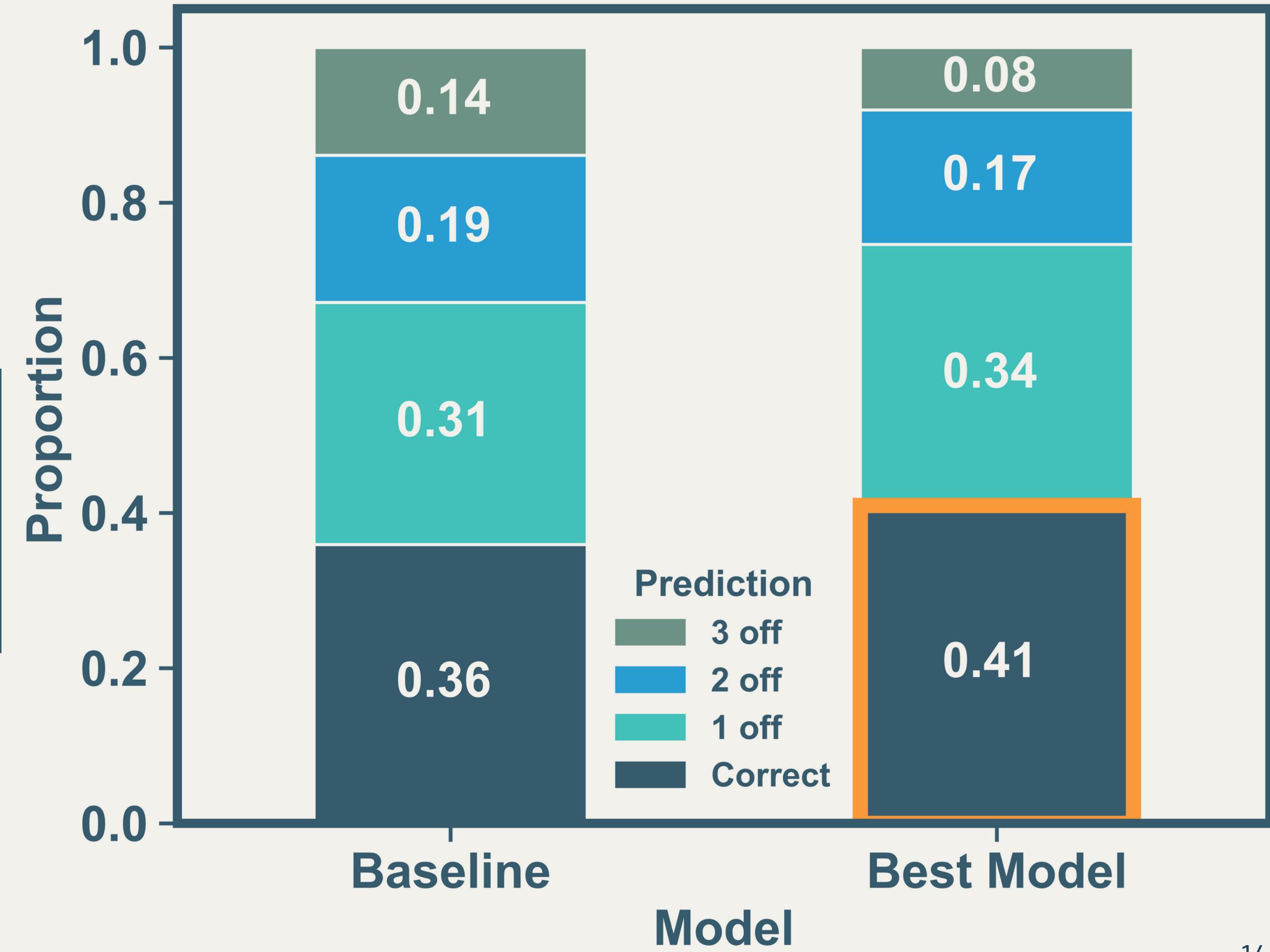
# Model Comparison

The kappa score increased from 0.22 to 0.35

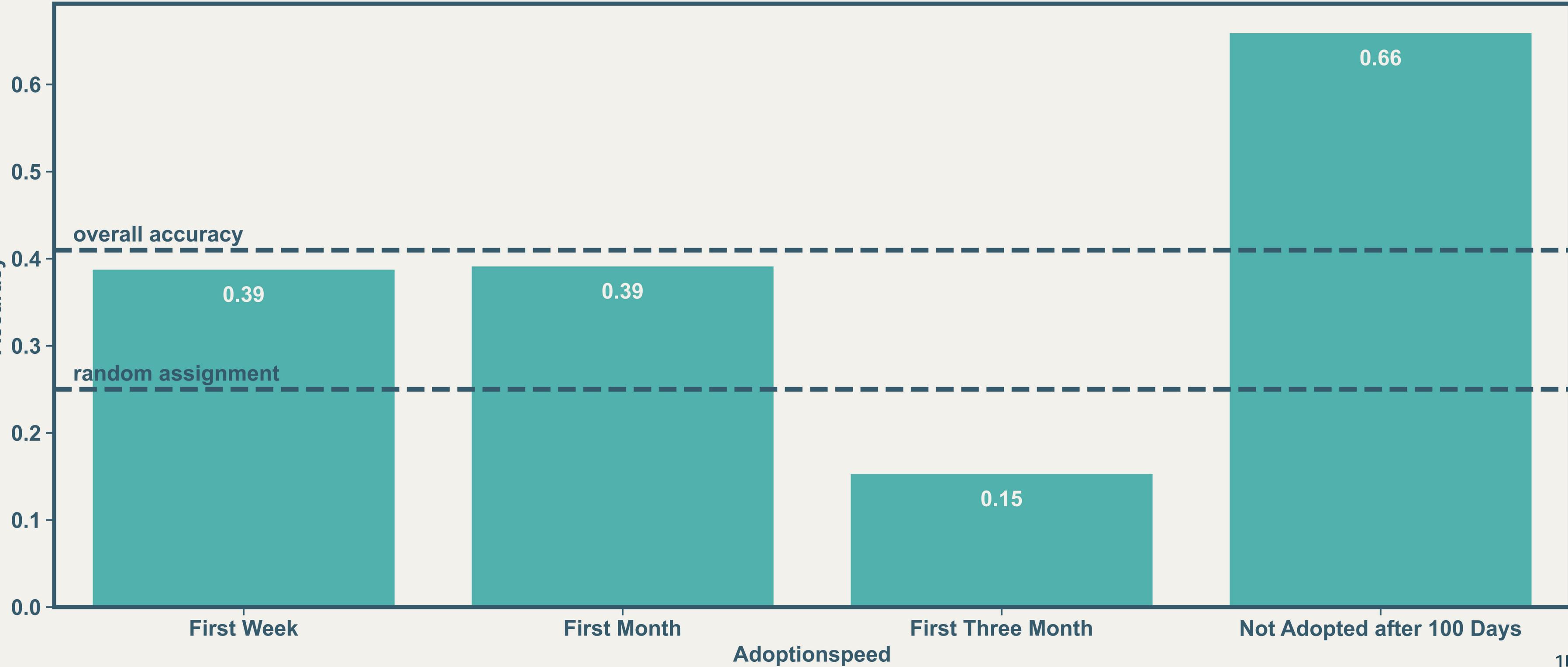


# Model Comparison

The kappa score increased from 0.22 to 0.35



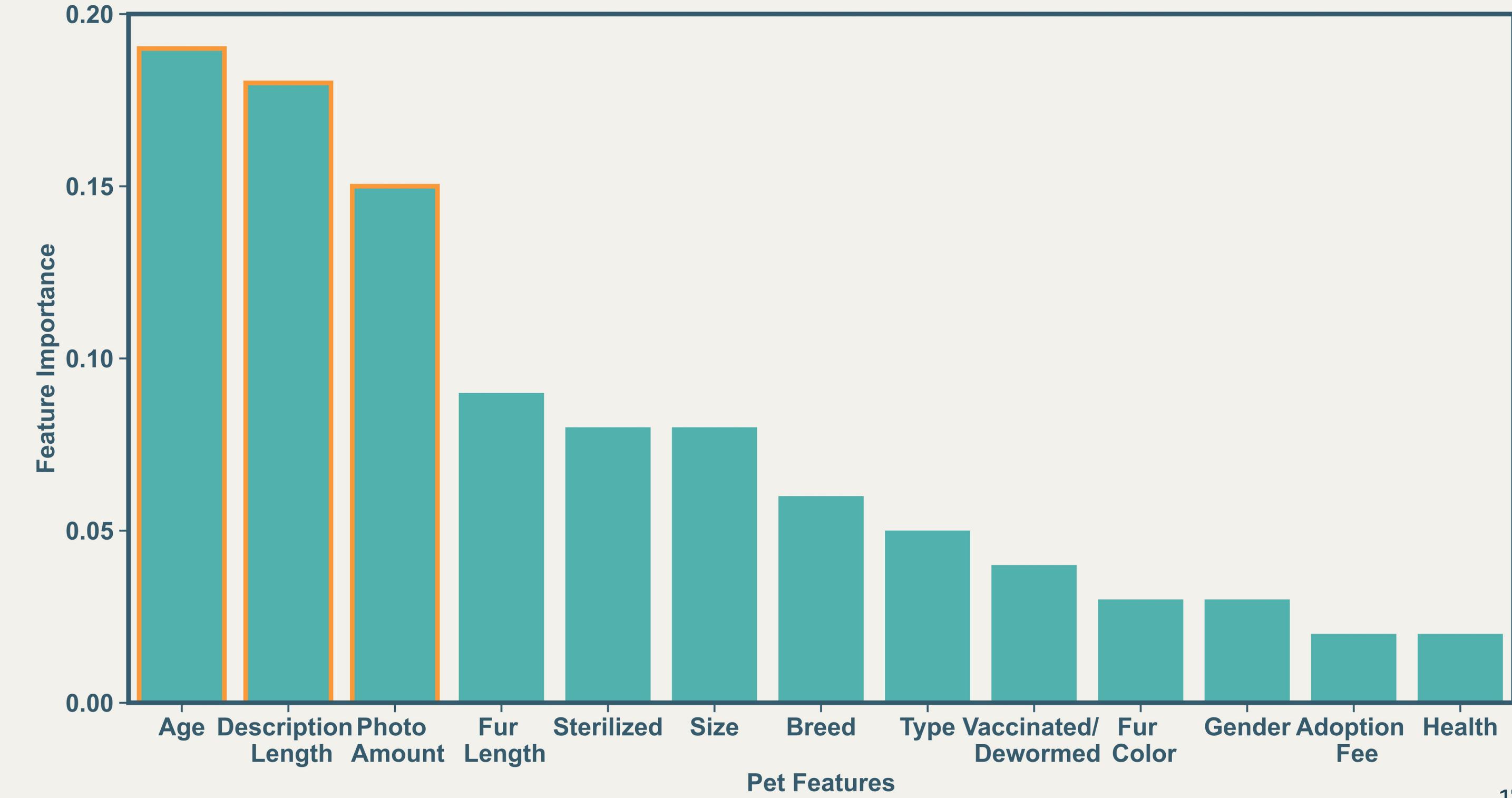
# Correct Predictions of Best Model



# Error analysis

- The best model predicts dogs' adoption speed more precisely than cats'
- Additionally, cats' prediction is more often '3 off'
- Separate models for cats and dogs performed worse than the combined model
- There are no clear patterns in the misclassifications, our model is not biased by the train data

# Feature Importance Best Model



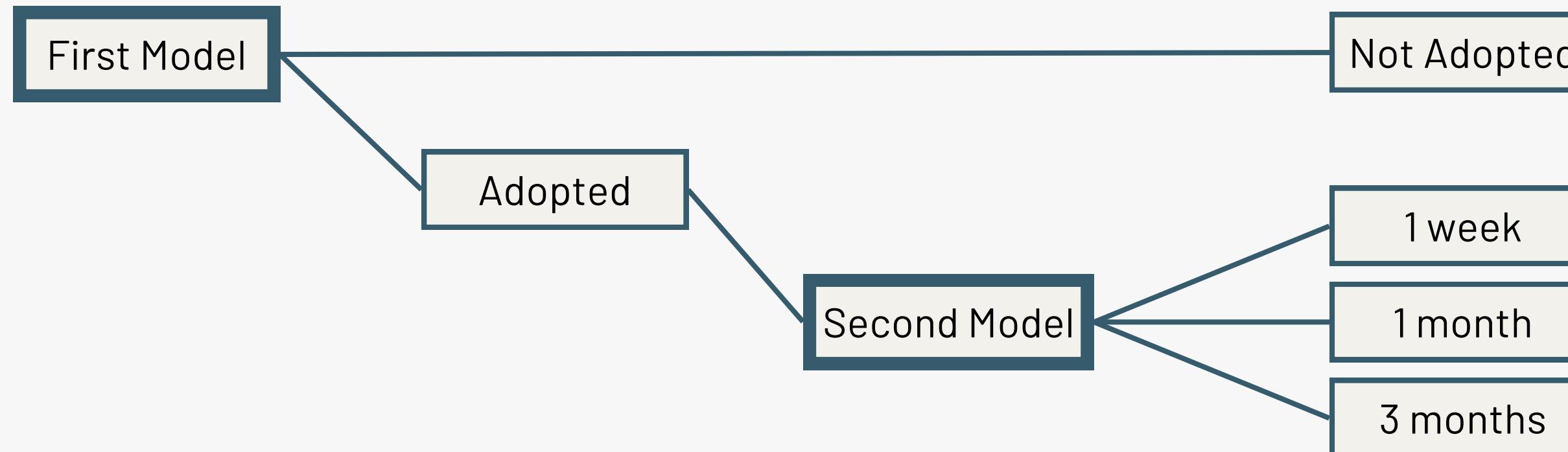
# Conclusions and Recommendations

- The model mostly relies on pets' age, the amount of photos and length of description - traits that were also revealed during the EDA
- Certain features are inherent and can not be influenced, but we can derive some recommendations to increase adoption speed:

- Include at least 4 photos
- Provide a meaningful description, at least 500-1000 characters long
- If a pet is predicted to stay for long, shelters can trust the model and allocate resources accordingly

# Future Work

- Include additional features that might influence adoption speed (chronic illnesses, behavioral traits etc.)
- Include AI-created metadata of pets' description and pictures to the training data to improve the model
- Create a two-step model:





# Thank you



 PetFinder.my