Unlocking the Secrets to Quick Pet Adoptions

Team 7 Project Milestone

Yushan Shi(ys468)
Yuqi Su(ys457)
Yurui Wu(yw695)
Mingxin Liu(ml754)

01

Introduction

Background

Millions of stray animals suffer on the streets or are euthanized in shelters every day around the world. If homes can be found for them, many precious lives can be saved.

PetFinder.my

We will develop algorithms to predict how quickly pets are adopted

150,000

Waiting for a Forever Home annually

02

Exploratory Data Analysis (EDA)

Variable

AdoptionSpeed – the value we want to predict

adopted
0 - Pet was adopted on the same day as it was listed.

1 - Pet was adopted between 1 and 7 days (1st week) after being listed.

2 - Pet was adopted between 8 and 30 days (1st month) after being listed.

3 - Pet was adopted between 31 and 90 days (2nd & 3rd month) after being listed.

unadopted
4 - No adoption after 100 days of being listed. (There are no pets in this dataset that waited between 90 and 100 days).

Variable

Interval Variables
- Age — Age of pet when listed, in months
- Fee — Adoption fee of the pet, where 0 = free
- Quantity — Number of pets represented in profile
- VideoAmt — This is the total amount of videos on record for a pet
- **SentimentScore** — A value between − 1 and 1 regarding how meaningful the Description was

Ordinal Variables
- FurLength — the length of fur on the pet
  - 0 = not specified
  - 1 = short
  - 2 = medium
  - 3 = long
- Health — the health condition the pet is in
  - 0 = not specified
  - 1 = healthy
  - 2 = minor injury
  - 3 = serious injury
- MaturitySize — the size of the pet once it will reach maturity
  - 0 = not specified
  - 1 = small
  - 2 = medium
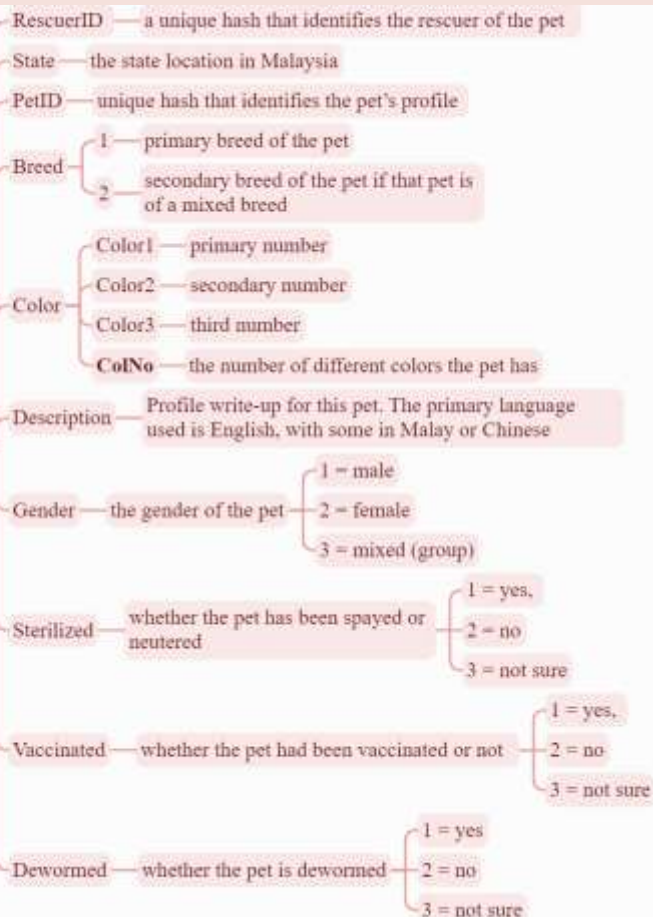 - 3 = large
  - 4 = extra-large

Variable

Binary Variable — Type — the type of animal
- 1 = dog
- 2 = cat

Nominal Variables
- RescuerID — a unique hash that identifies the rescuer of the pet
- State — the state location in Malaysia
- PetID — unique hash that identifies the pet's profile
- Breed
  - 1 — primary breed of the pet
  - 2 — secondary breed of the pet if that pet is of a mixed breed
- Color
  - Color1 — primary number
  - Color2 — secondary number
  - Color3 — third number
  - **ColNo** — the number of different colors the pet has
- Description — Profile write-up for this pet. The primary language used is English, with some in Malay or Chinese
- Gender — the gender of the pet
  - 1 = male
  - 2 = female
  - 3 = mixed (group)
- Sterilized — whether the pet has been spayed or neutered
  - 1 = yes,
  - 2 = no
  - 3 = not sure
- Vaccinated — whether the pet had been vaccinated or not
  - 1 = yes,
  - 2 = no
  - 3 = not sure
- Dewormed — whether the pet is dewormed
  - 1 = yes
  - 2 = no
  - 3 = not sure

Example: Labrador Retriever

important features on profile

- Age = 2
- Breed1 = 141, Breed2 = 218
- Gender = 3
- Color1 = 1, Color2 = 0, Color3 = 0
- MaturitySize = 3
- FurLength = 2
- Vaccinated = 2
- Dewormed = 1
- Health = 1
- Quantity = 2
- Fee = 150
- VideoAmt = 0
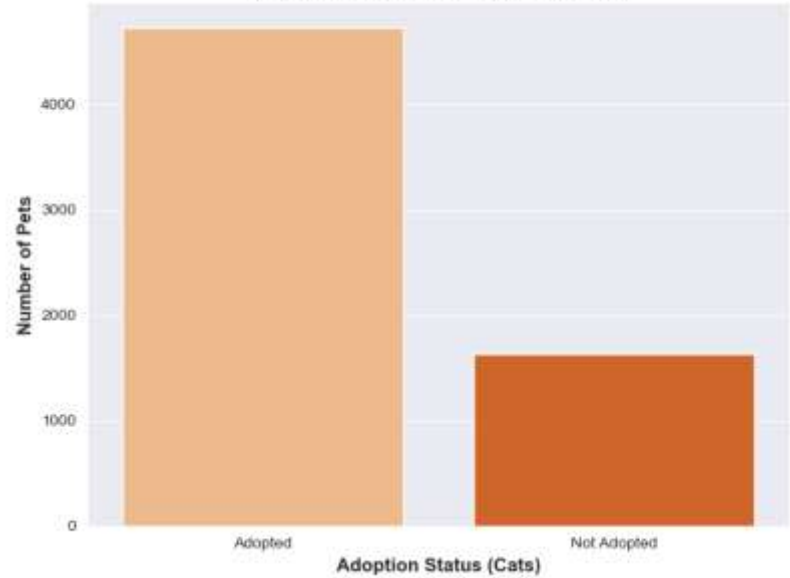- PhotoAmt = 10
- AdoptionSpeed = 2
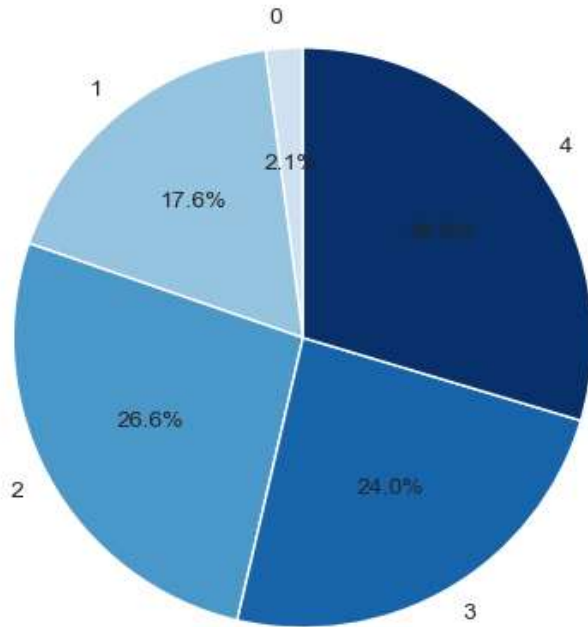
Distribution of adoption rate
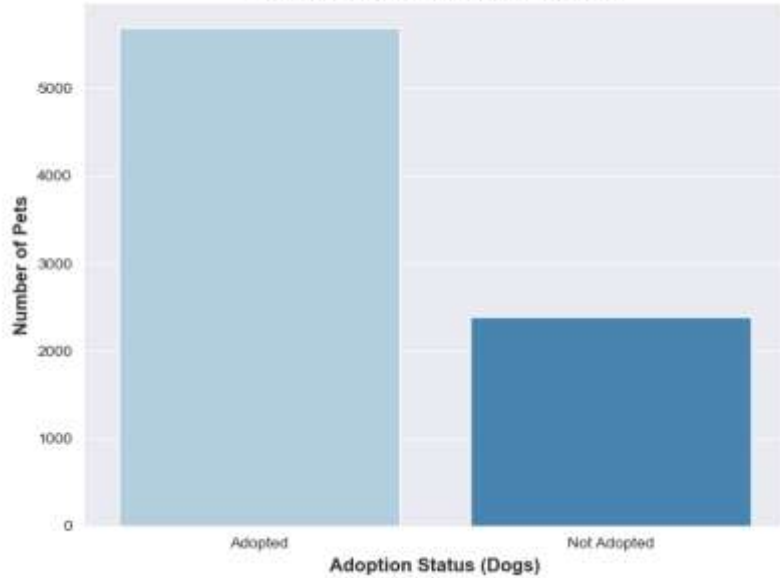
Distribution of adoption speed

Adoption Speed Distribution - Dogs

0
4
1
2.1%
17.6%
2
26.6%
24.0%
3

Adoption Speed Distribution - Cats

0
4
1
3.4%
25.7%
23.9%
2
27.6%
19.4%
3

Relationship

Smoothed Relationship Between Age and Mean Adoption Speed (Dogs vs Cats)
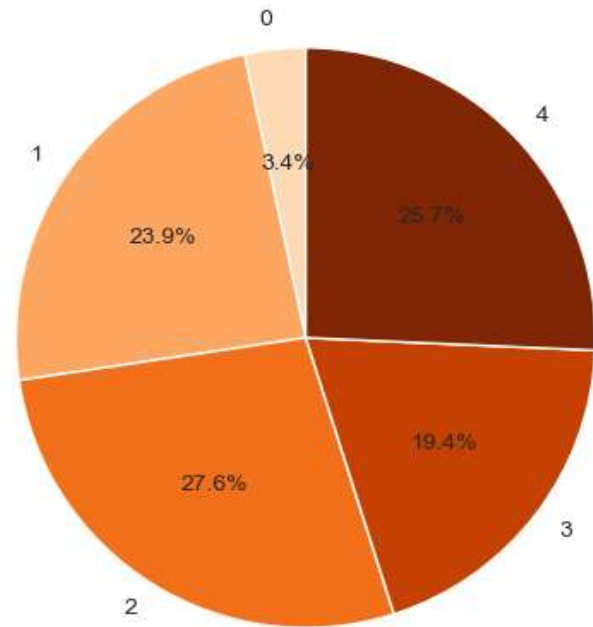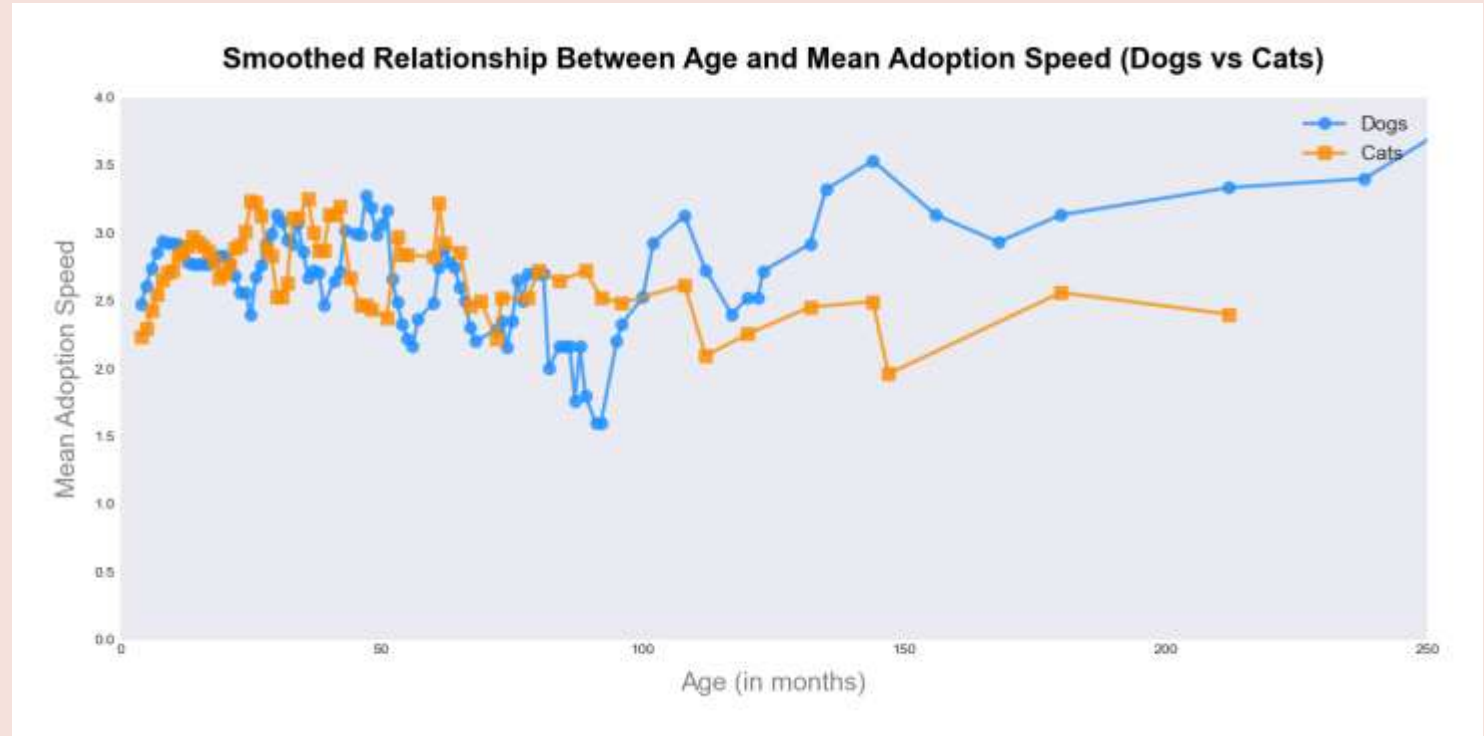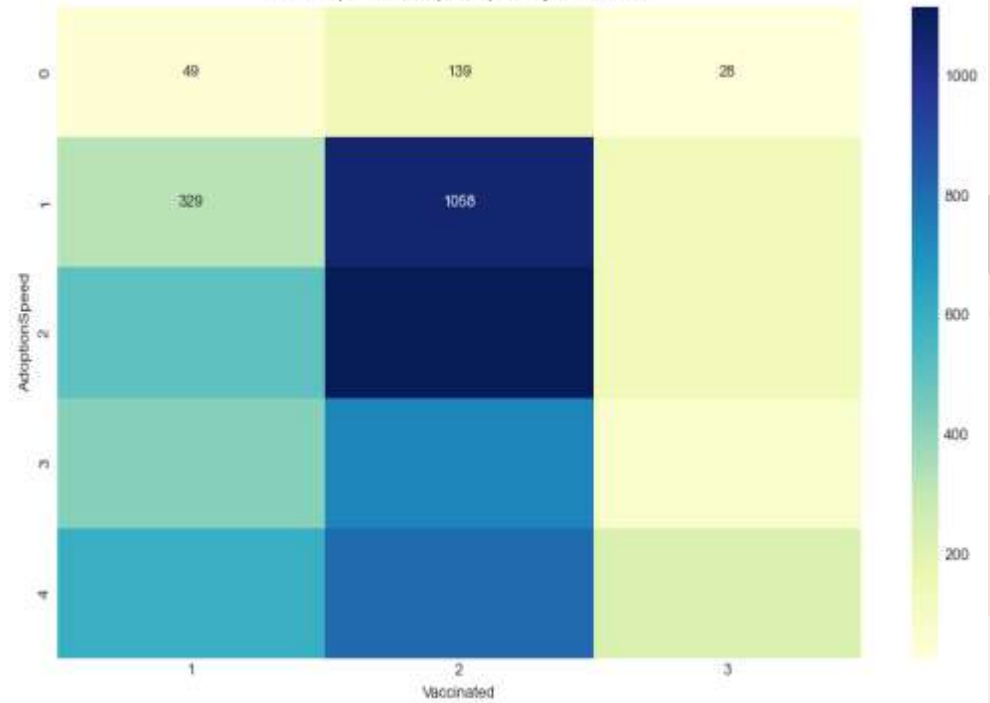
Relationship

Heatmap of Dog AdoptionSpeed by Vaccinated
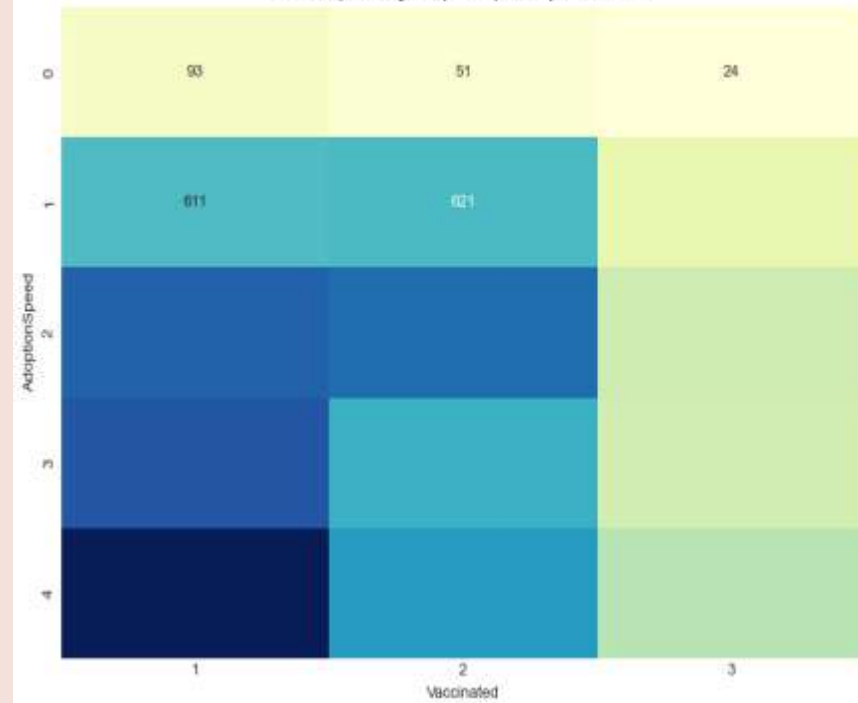
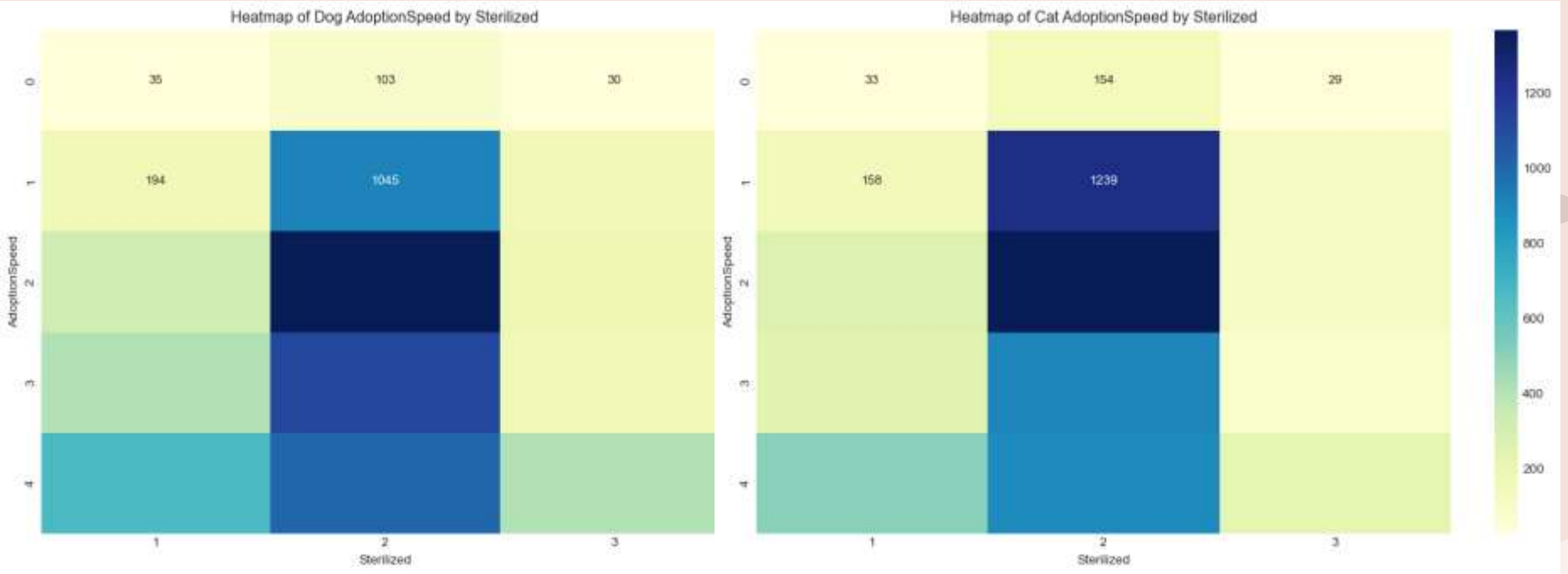Heatmap of Cat AdoptionSpeed by Vaccinated

Relationship

Heatmap of Dog AdoptionSpeed by Dewormed

Heatmap of Cat AdoptionSpeed by Dewormed

Relationship

Heatmap of Dog AdoptionSpeed by Sterilized

Heatmap of Cat AdoptionSpeed by Sterilized
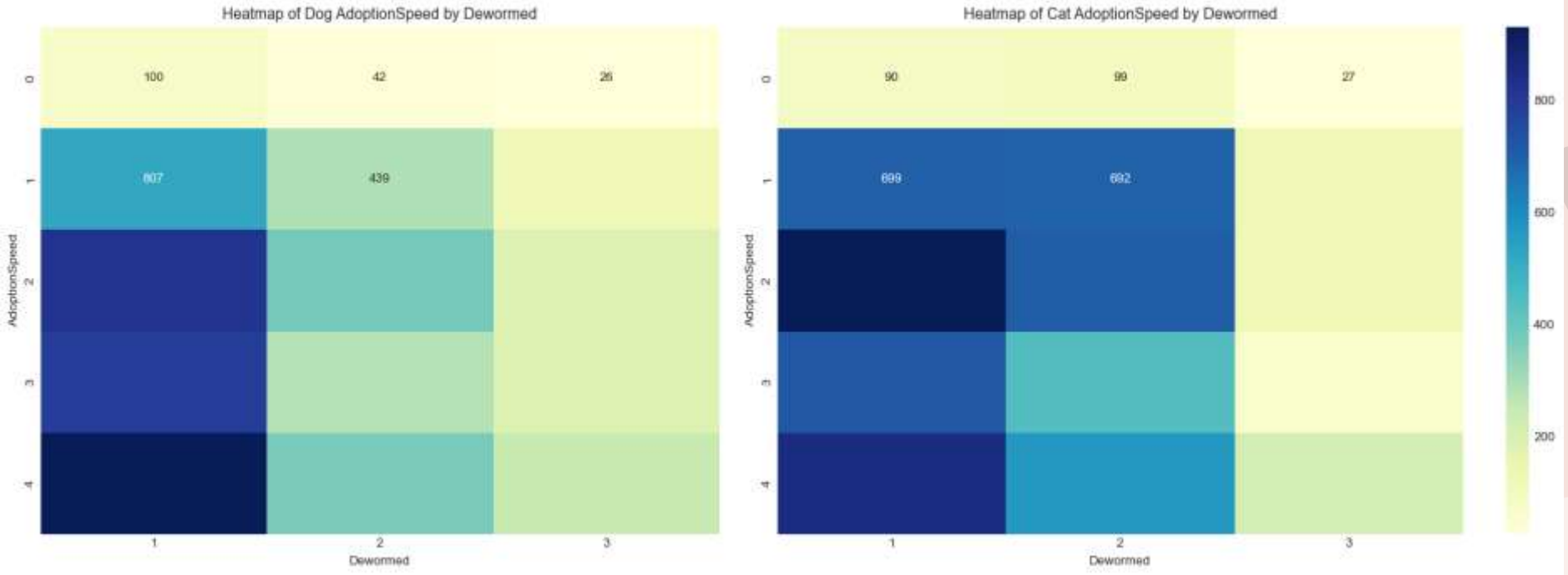
Relationship
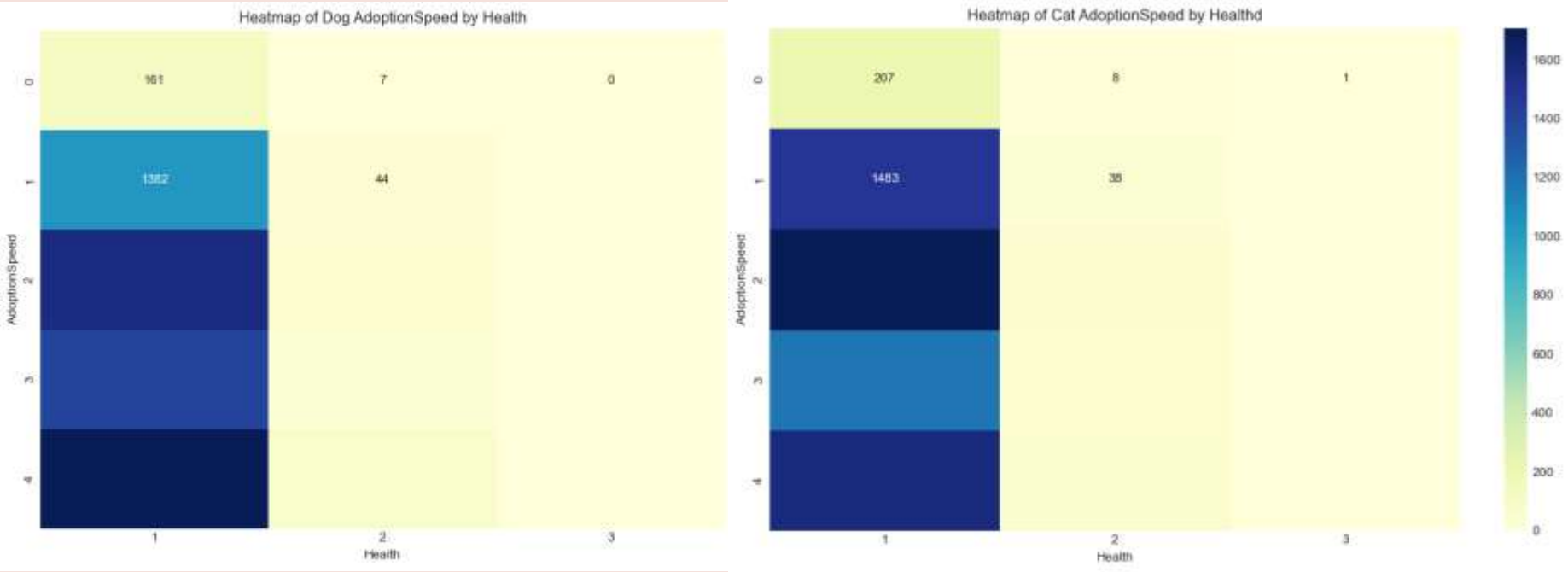
Chi-square Test

Variable	Age	Breed	Color	Maturity Size	Fur Length	Vaccinated	Sterilized	Dewormed	Health
P value	1.84e-109	1.36e-62	2.97e-09	6.58e-24	3.17e-30	2.01e-55	1.36e-168	2.06e-28	0.003

The P value is quite small for all variables!

03

Model Analysis

Random Forest Approach for Pet Adoption Analysis

Methodology:

- Parameter optimization through RandomizedSearchCV exploration
- Enhanced learning using SMOTE balancing technique
- Feature importance calculated through Gini impurity reduction

Key Finding: Fee, video content, and pet characteristics emerged as influential factors, offering actionable insights for adoption strategies.
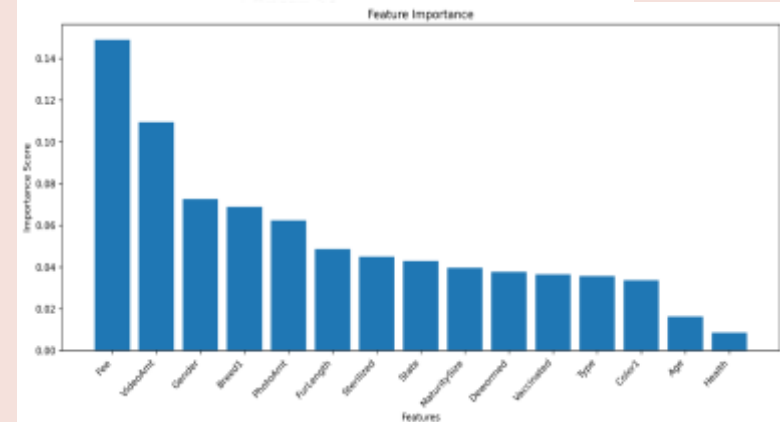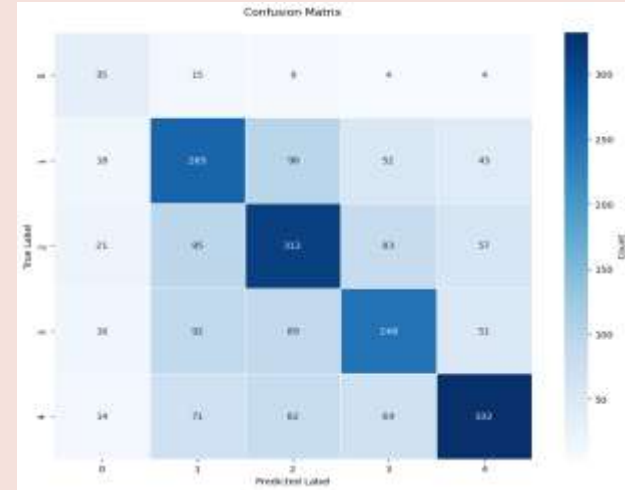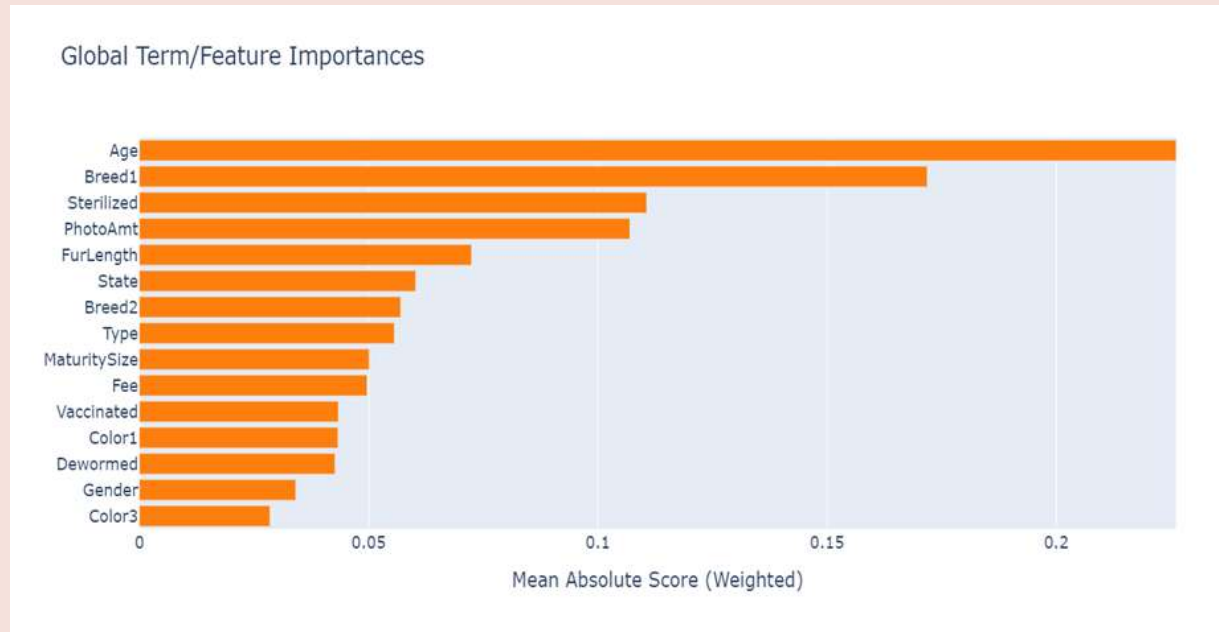
Feature Importance (GAM)
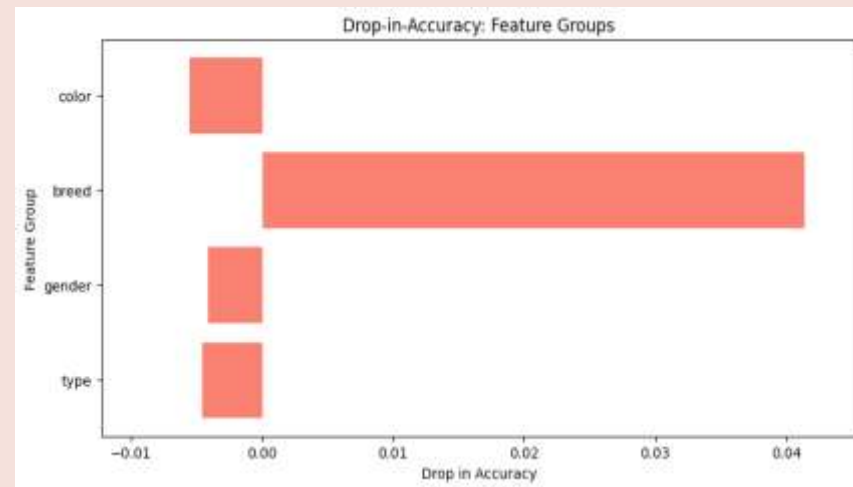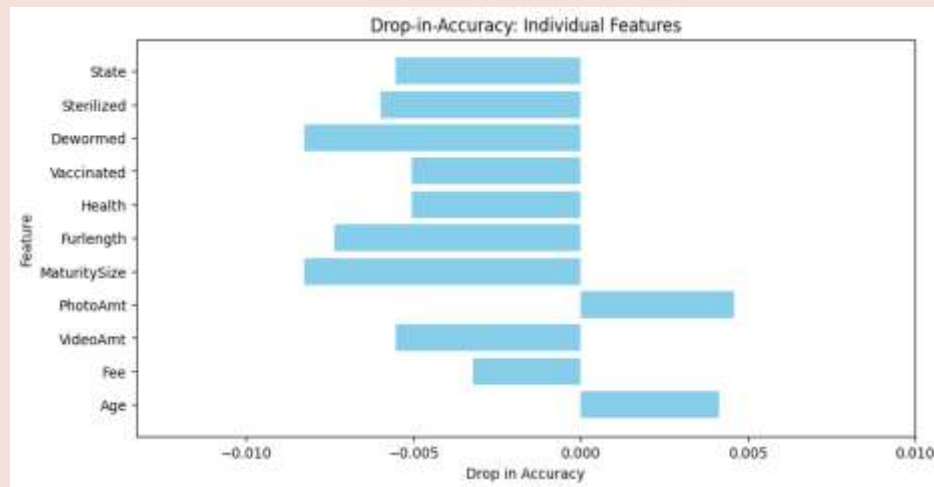
Key Finding: Age, breed and photo Amount are identified as significant factors
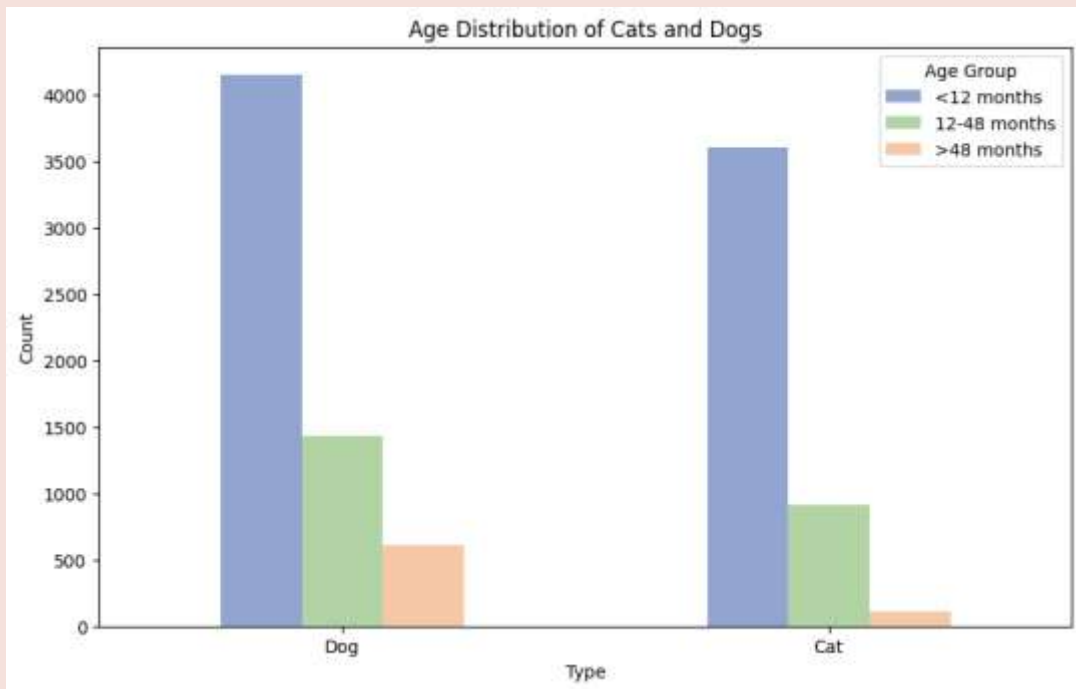
Feature Importance (EBM)

Key Finding: Age, breed and photo Amount are identified as significant factors again

Why age?

Global Term/Feature Importances

Dataset Insights

Age Distribution of Cats and Dogs

Class Imbalance in Age Groups

- Overemphasize patterns in the <12 months group due to its dominance.

Reflects real-world trends

- Younger pets are generally more adoptable, potentially contributing to faster adoption times

Key Takeaways

- **For Rescuers**: Ensuring good health, vaccination, and sterilization is vital, but showcasing animals through photos and videos is equally important.

- **For Animal Lovers**: Shelters are home to so many healthy and adorable puppies and kittens, all waiting for a chance to be loved. With care and affection, they can become lifelong friends.

04

Contributions

Contributions

1. Introduction: Yuqi

2. Dataset description and Analysis: Yuri, Yushan and Mingxin

3. Narrative and insights: Mingxin, Yuqi, and Yuri

4. Project timeline: Mingxin

5. Presentation and video: all team members

THANKS!