PREDICTING TIME TO ADOPTION FOR SHELTER ANIMALS

Project by Matt Williams

September 8th, 2020

PROBLEM STATEMENT

Meet Milo. He is one of the millions of animals that end up in shelters worldwide each year. In the US alone, an estimated 6.5 million pets are surrendered to shelters each year, and nearly 25% of them are not adopted.

Our goal will be to identify factors that contribute to quicker adoption speeds, and attempt to build a model to predict how quickly shelter animals are adopted.

DATA

Source: Kaggle - Petfinder.my Adoption Prediction

PetFinder collaborates closely with animal lovers, media, corporations, and global organizations to improve animal welfare.

Target: Adoption Speed

- 0: adopted same day
- 1: adopted in 1 - 7 days
- 2: adopted in 1st month
- 3: adopted in 30 – 90 days
- 4: adopted in more than 100 days

Training/Testing Data – nearly 19,000 animals

- Images
- Metadata
- Sentiment

Data Fields

- **PetID** - Unique hash ID of pet profile
- **AdoptionSpeed** - Categorical speed of adoption. Lower is faster. This is the value to predict. See below section for more info.
- **Type** - Type of animal *(1 = Dog, 2 = Cat)*
- **Name** - Name of pet *(Empty if not named)*
- **Age** - Age of pet when listed, in months
- **Breed1** - Primary breed of pet *(Refer to BreedLabels dictionary)*
- **Breed2** - Secondary breed of pet, if pet is of mixed breed *(Refer to BreedLabels dictionary)*
- **Gender** - Gender of pet *(1 = Male, 2 = Female, 3 = Mixed, if profile represents group of pets)*
- **Color1** - Color 1 of pet *(Refer to ColorLabels dictionary)*
- **Color2** - Color 2 of pet *(Refer to ColorLabels dictionary)*
- **Color3** - Color 3 of pet *(Refer to ColorLabels dictionary)*
- **MaturitySize** - Size at maturity *(1 = Small, 2 = Medium, 3 = Large, 4 = Extra Large, 0 = Not Specified)*
- **FurLength** - Fur length *(1 = Short, 2 = Medium, 3 = Long, 0 = Not Specified)*
- **Vaccinated** - Pet has been vaccinated *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Dewormed** - Pet has been dewormed *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Sterilized** - Pet has been spayed / neutered *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Health** - Health Condition *(1 = Healthy, 2 = Minor Injury, 3 = Serious Injury, 0 = Not Specified)*
- **Quantity** - Number of pets represented in profile
- **Fee** - Adoption fee *(0 = Free)*
- **State** - State location in Malaysia *(Refer to StateLabels dictionary)*
- **RescuerID** - Unique hash ID of rescuer
- **VideoAmt** - Total uploaded videos for this pet
- **PhotoAmt** - Total uploaded photos for this pet

IMAGE METADATA

Images processed using Google Vision API

Returns JSON containing features of each image

Retained dominant feature from each category


```
"cropHintsAnnotation": {
    "cropHints": [
        {
            "boundingPoly": {
                "vertices": [
                    {},
                    {
                        "x": 359
                    },
                    {
                        "x": 359,
                        "y": 479
                    },
                    {
                        "y": 479
                    }
                ]
            },
            "confidence": 0.79999995,
            "importanceFraction": 1
```

```
"labelAnnotations": [
    {
        "mid": "/m/0kpmf",
        "description": "dog breed",
        "score": 0.93895864,
        "topicality": 0.93895864
    },
    {
        "mid": "/m/0bt9lr",
        "description": "dog",
        "score": 0.9127391,
        "topicality": 0.9127391
    },
```

```
"imagePropertiesAnnotation": {
    "dominantColors": {
        "colors": [
            {
                "color": {
                    "red": 205,
                    "green": 198,
                    "blue": 192
                },
                "score": 0.2881473,
                "pixelFraction": 0.27667227
```

FEATURE ENGINEERING

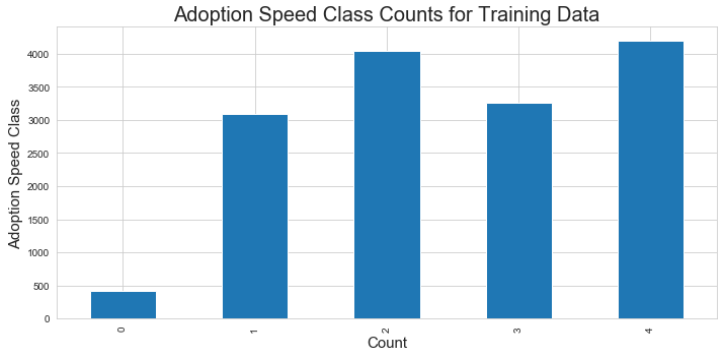
Coat Combinations

Solid Color?

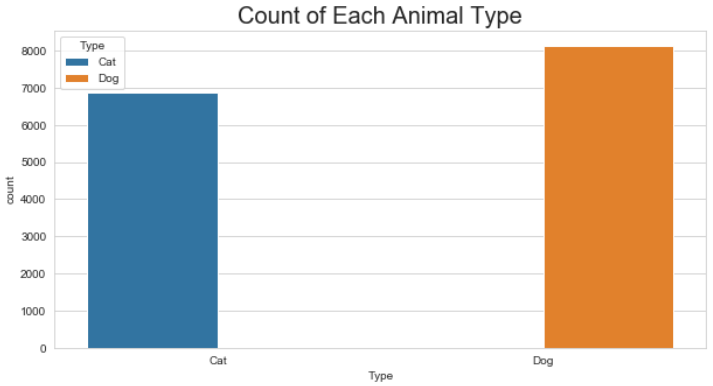
Description Word Count

Description Sentiment

Description Point of View

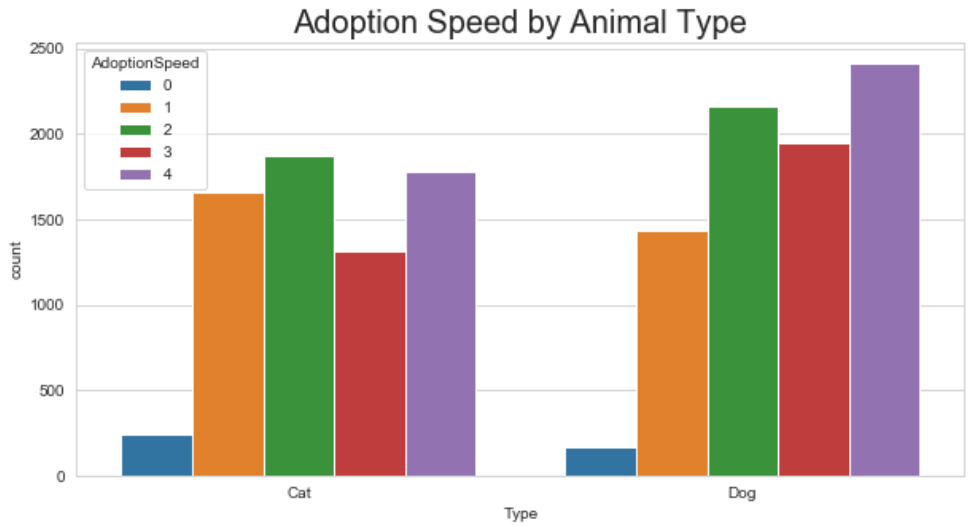


	Adoption Speed Count	Adoption Speed %
4.0	4197	0.279931
3.0	3259	0.217368
2.0	4037	0.269259
1.0	3090	0.206096
0.0	410	0.027346



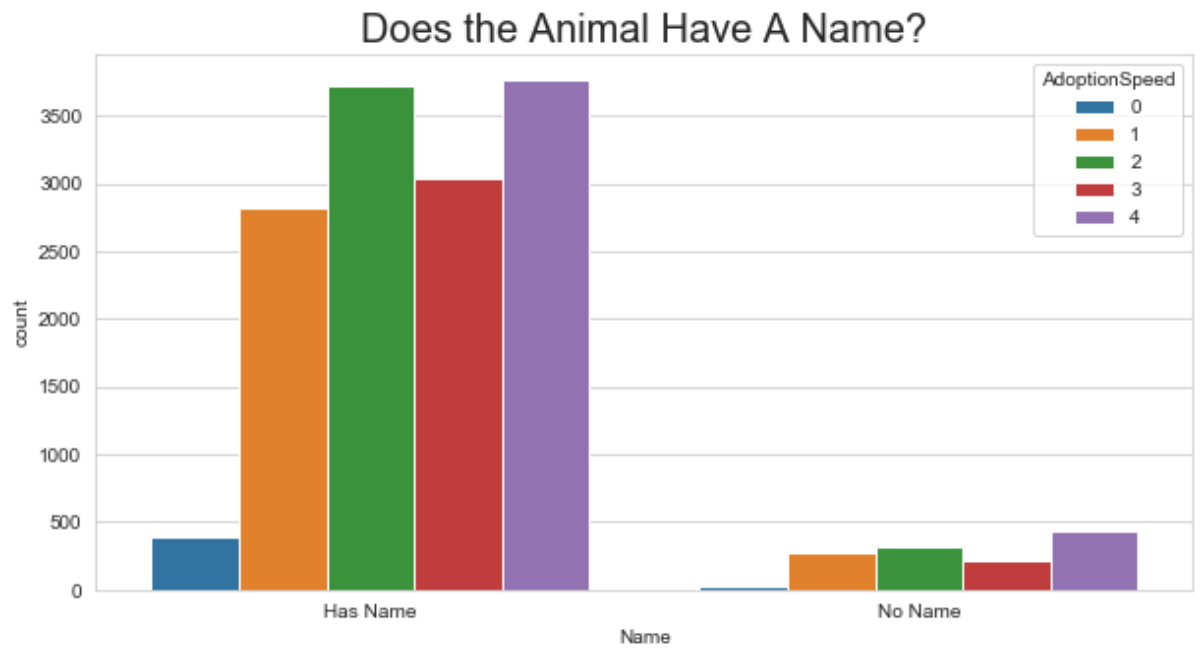
GENERAL OBSERVATIONS

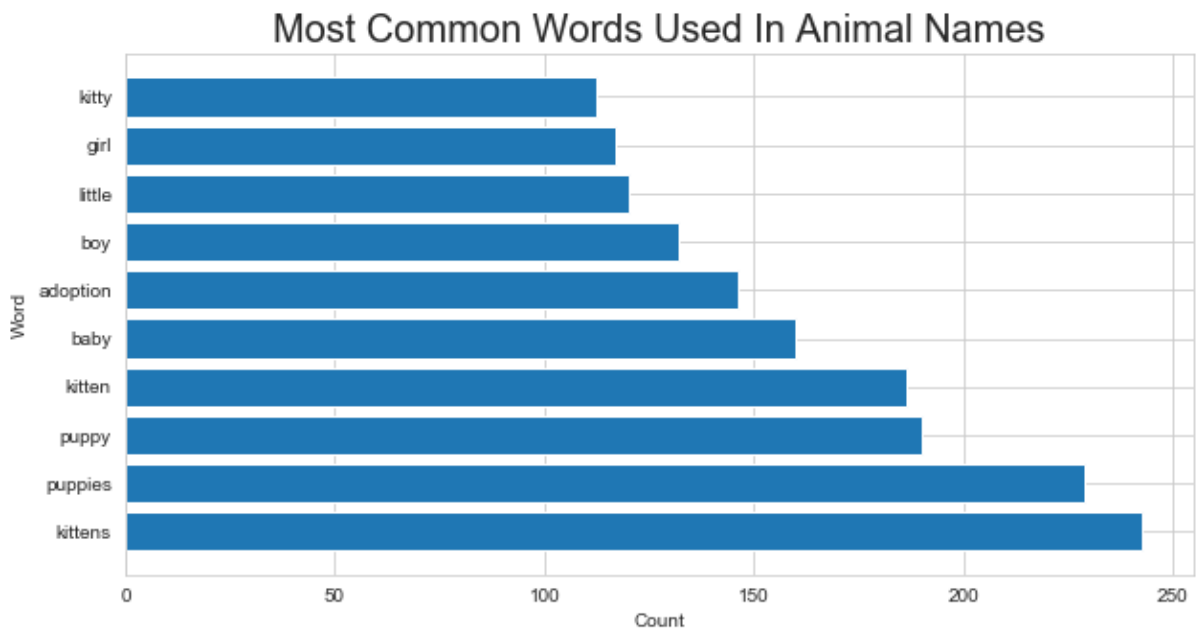
ANIMAL TYPE



AdoptionSpeed	0	1	2	3	4
Type					
Cat	0.034980	0.241218	0.272992	0.190934	0.259875
Dog	0.020905	0.176463	0.266109	0.239670	0.296852

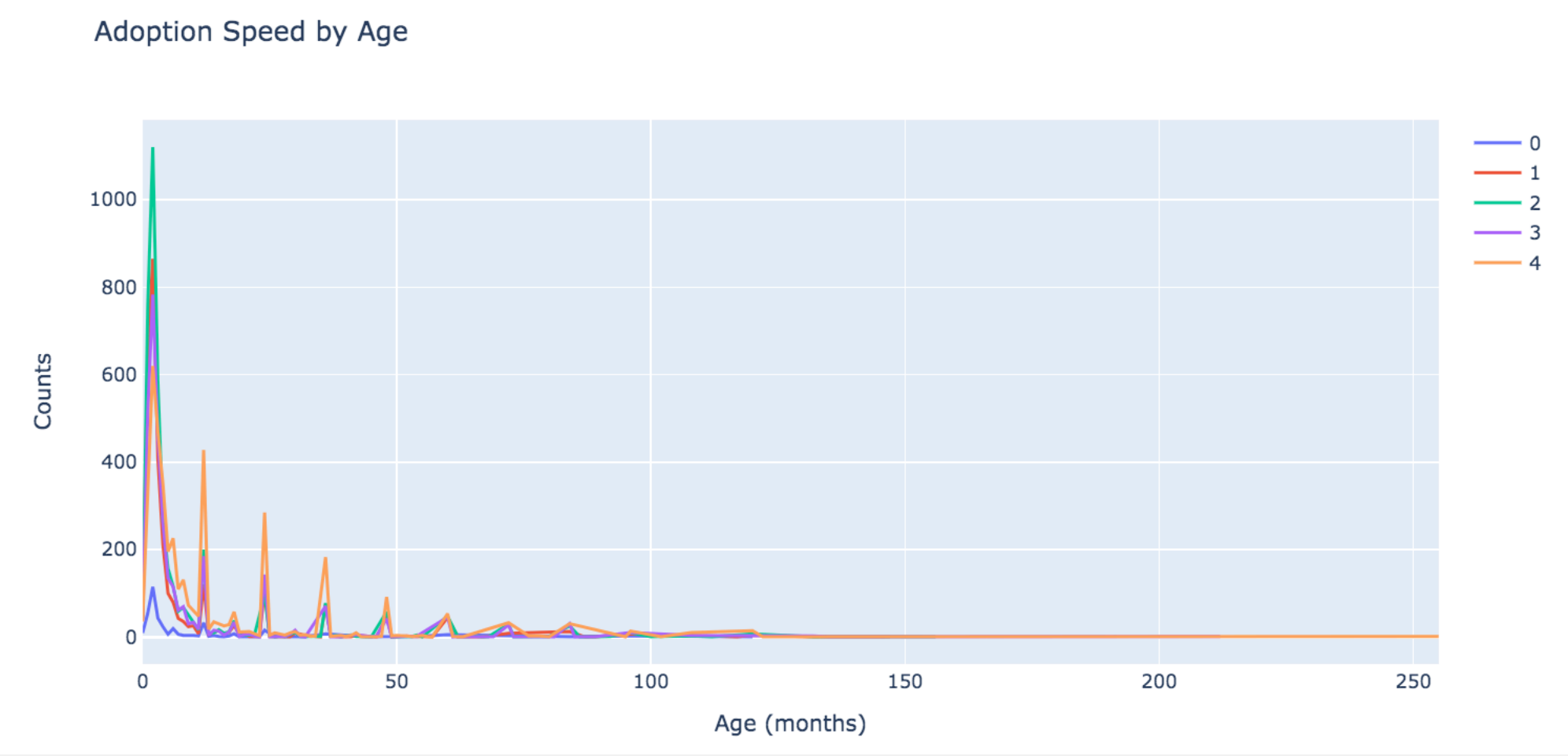
NAMES



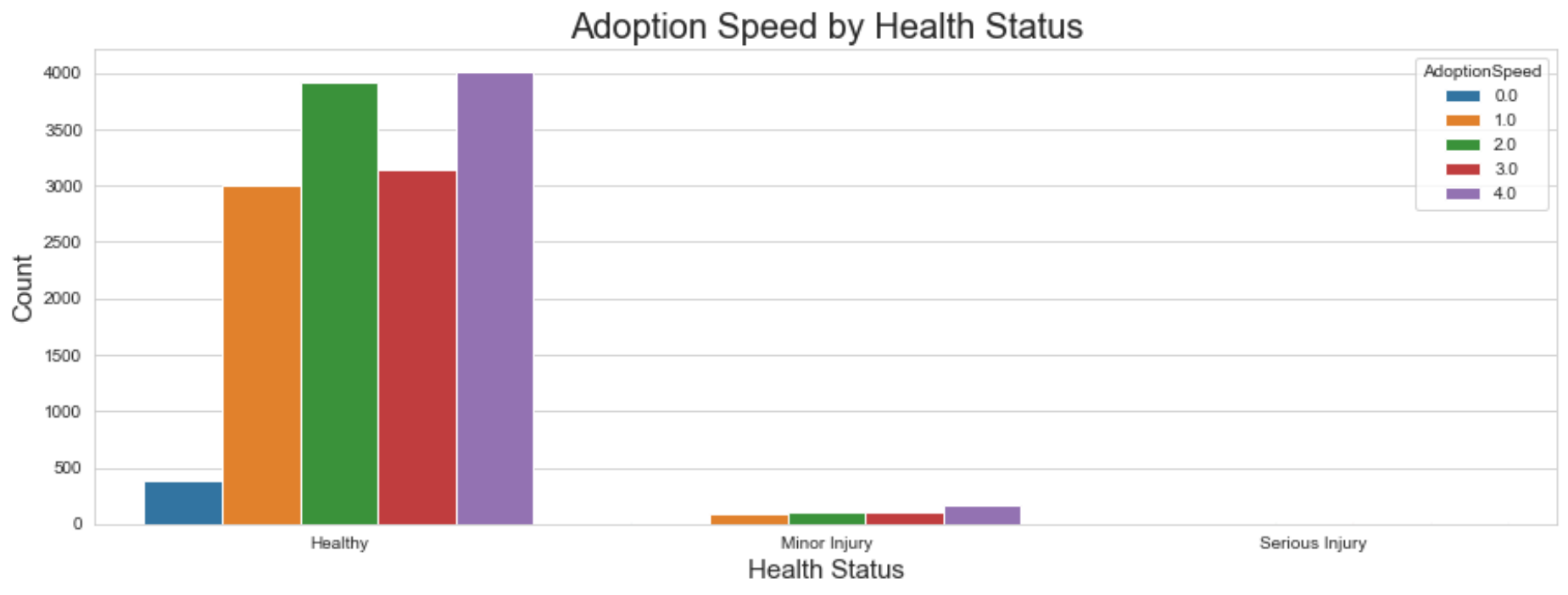


AdoptionSpeed	0.0	1.0	2.0	3.0	4.0
Name					
Has Name	0.027956	0.205227	0.271185	0.221535	0.274097
No Name	0.020684	0.215593	0.248210	0.171838	0.343675

AGE


Adoption Speed by Age
Counts
1000
800
600
400
200
0
0
50
100
150
200
250
Age (months)
0
1
2
3
4

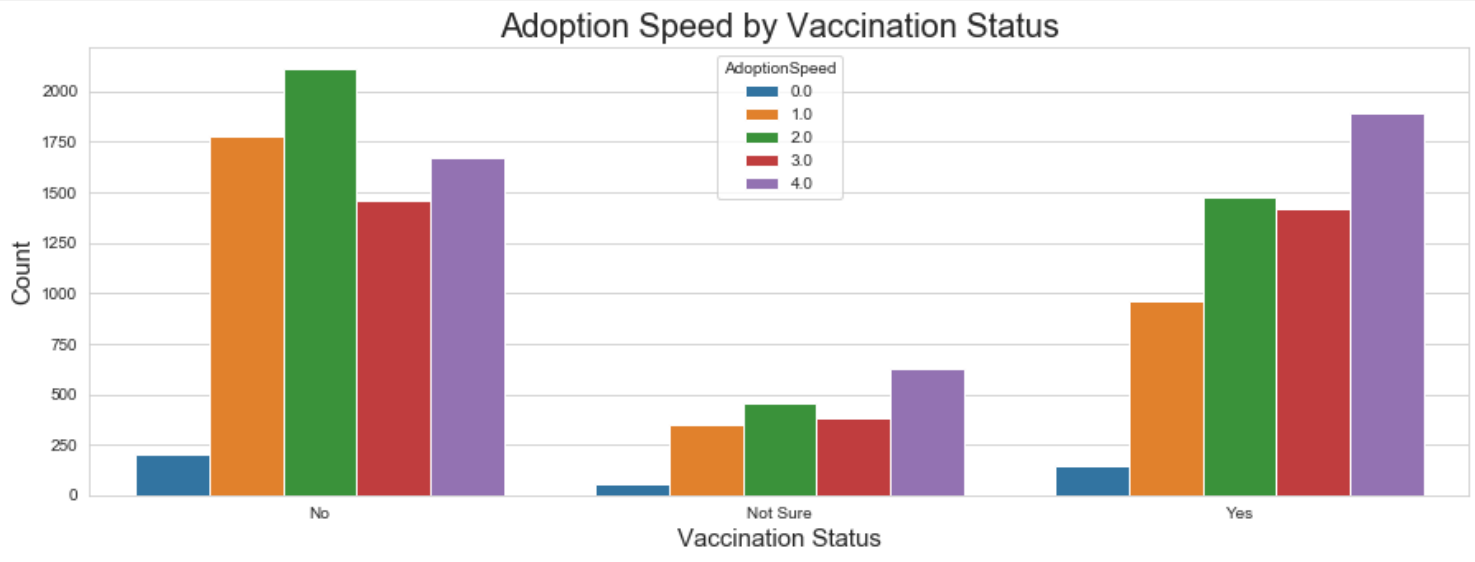
HEALTH STATUS

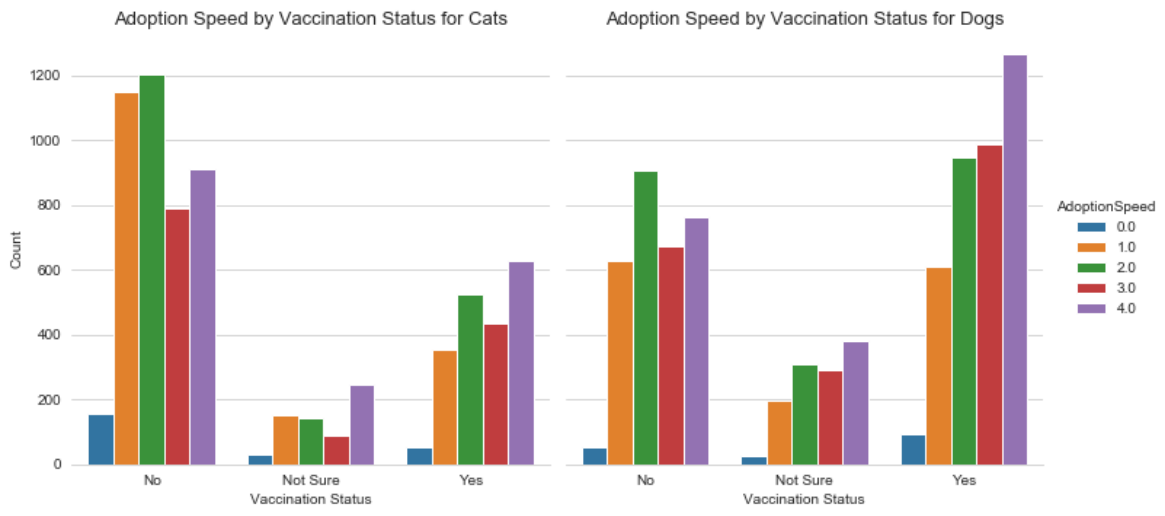


AdoptionSpeed	0.0	1.0	2.0	3.0	4.0
Health					
Healthy	0.027076	0.207142	0.271101	0.217571	0.277110
Minor Injury	0.035343	0.185031	0.220374	0.203742	0.355509
Serious Injury	0.029412	0.058824	0.176471	0.323529	0.411765

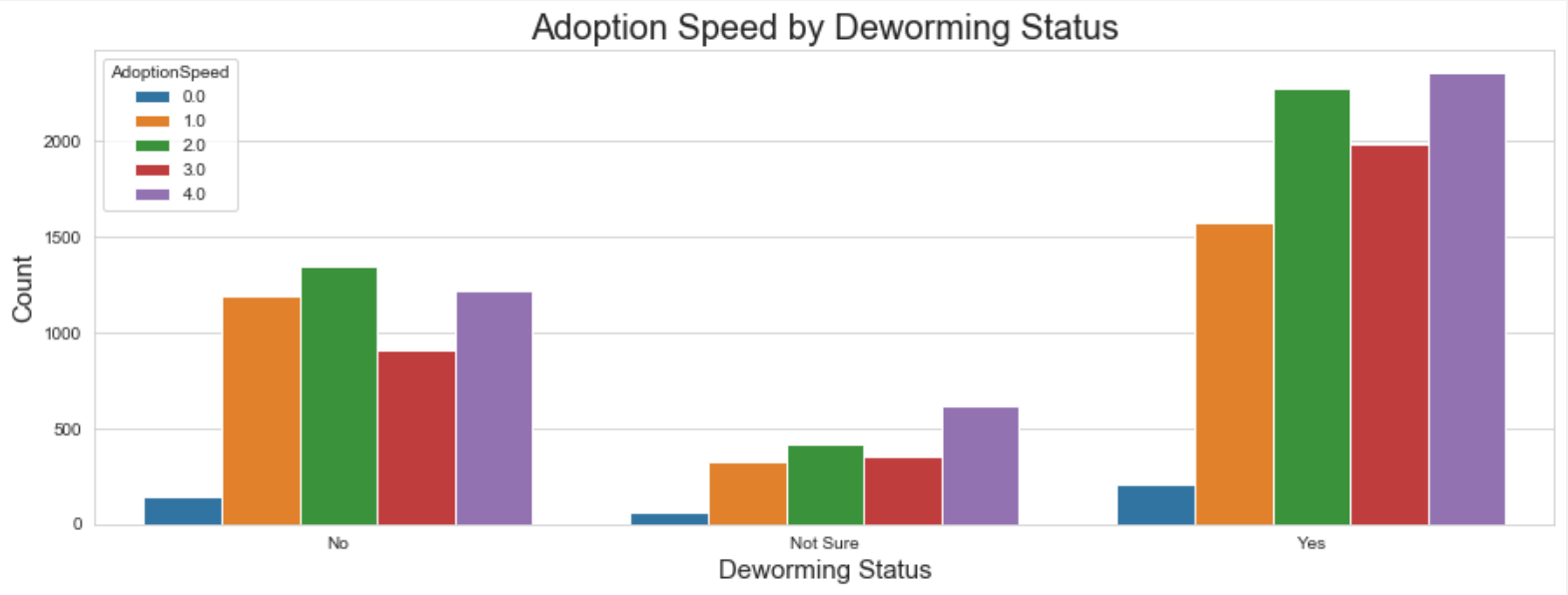
```
Healthy            14478
Minor Injury         481
Serious Injury        34
Name: Health, dtype: int64,
Healthy          0.965651
Minor Injury     0.032082
Serious Injury   0.002268
```

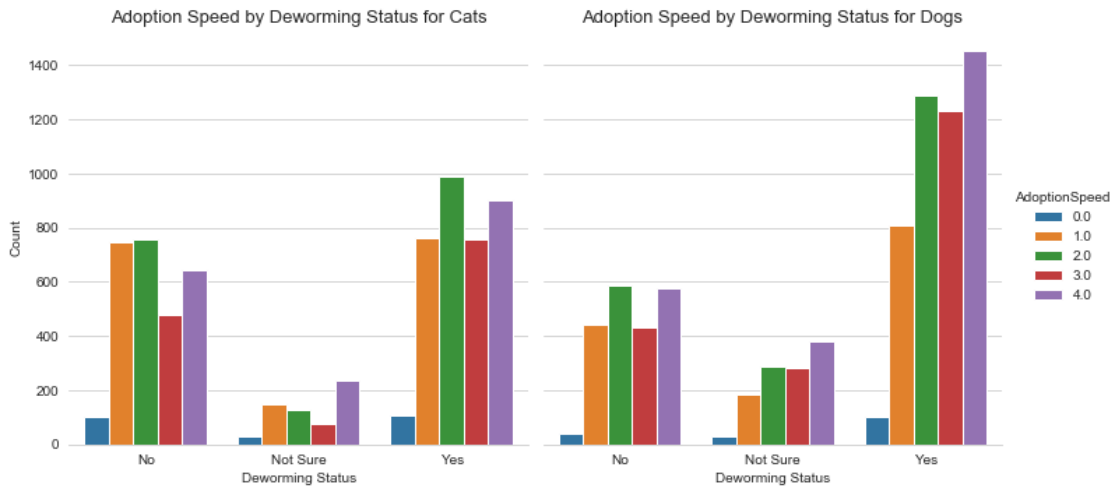
VACCINATION STATUS



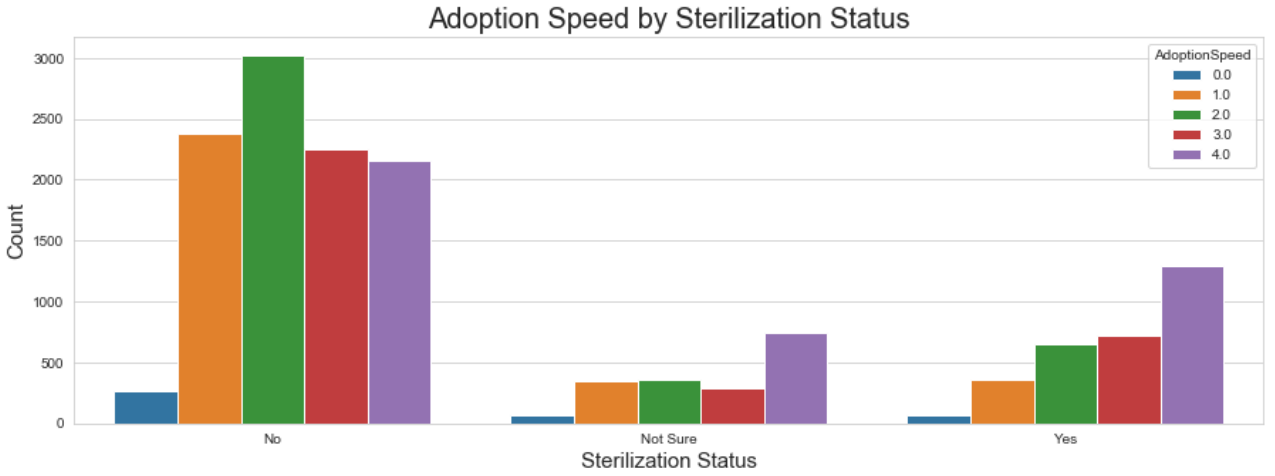


DEWORMING STATUS



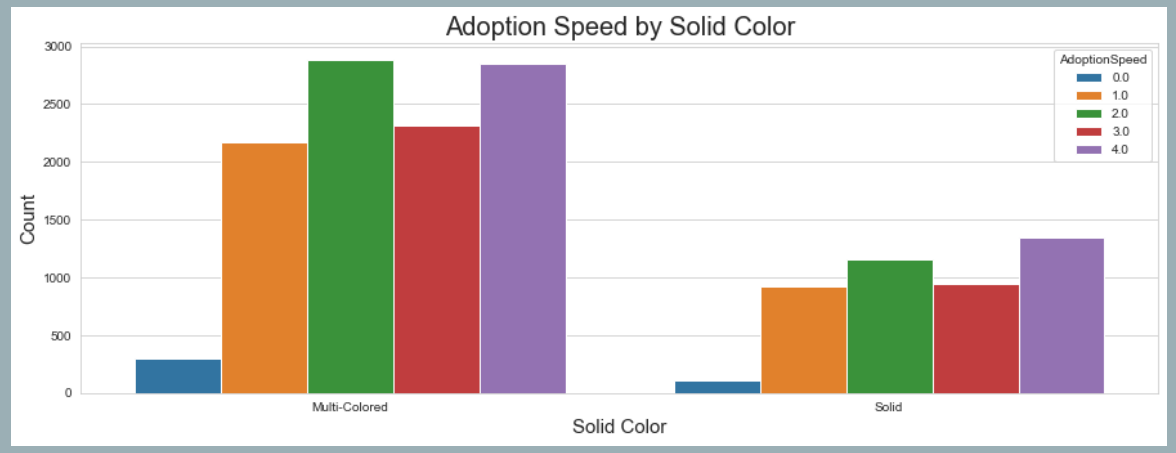


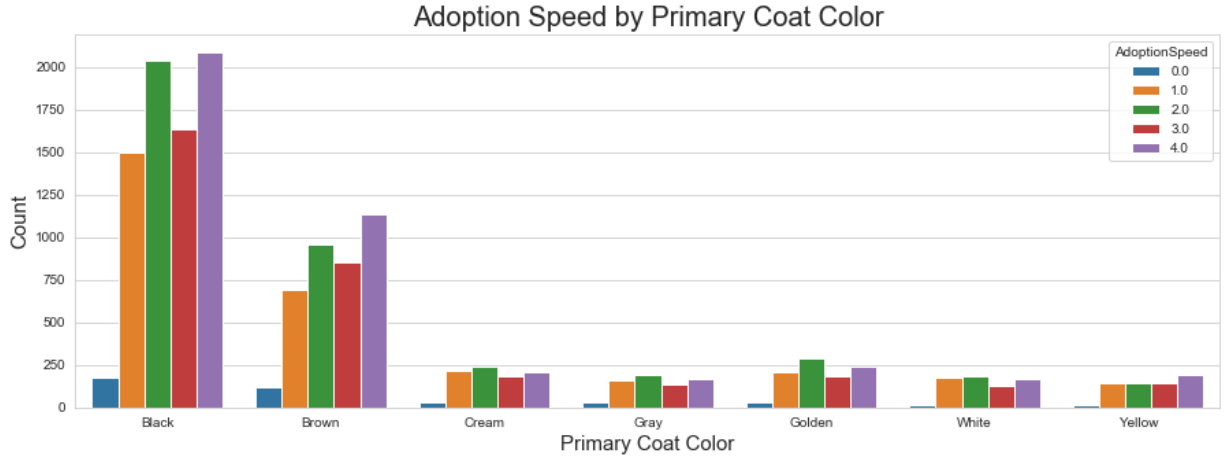
STERILIZATION STATUS

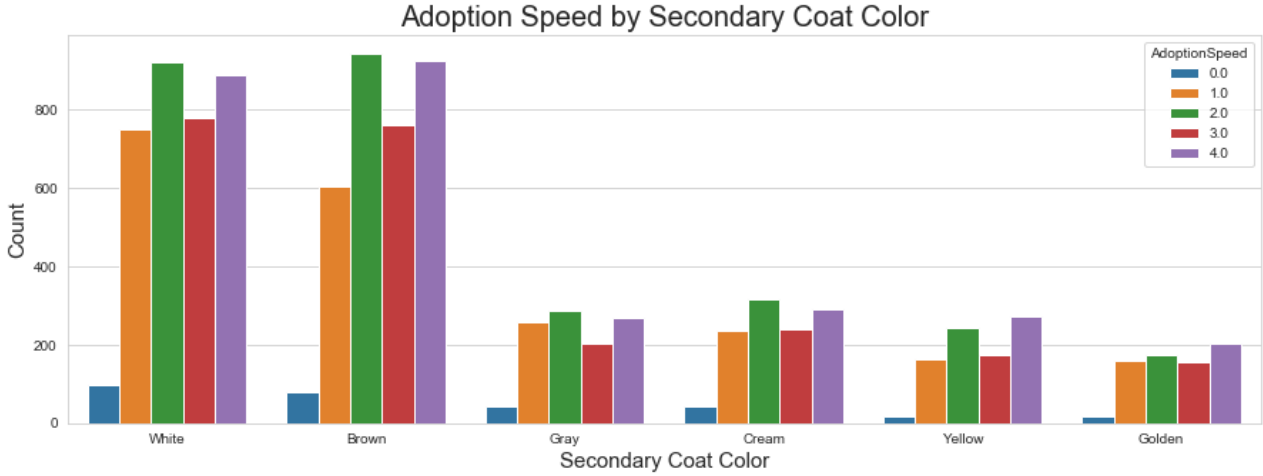


AdoptionSpeed	0.0	1.0	2.0	3.0	4.0
Sterilized					
No	0.026794	0.235983	0.299891	0.223380	0.213953
Not Sure	0.038567	0.191736	0.197796	0.158678	0.413223
Yes	0.022573	0.117381	0.211545	0.232183	0.416317

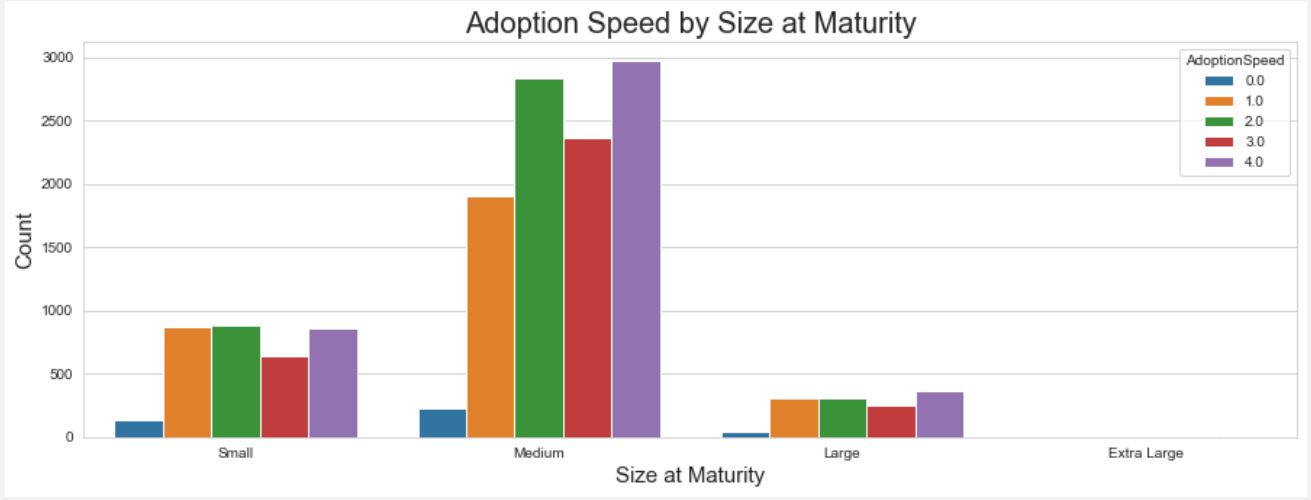
COAT COLORS

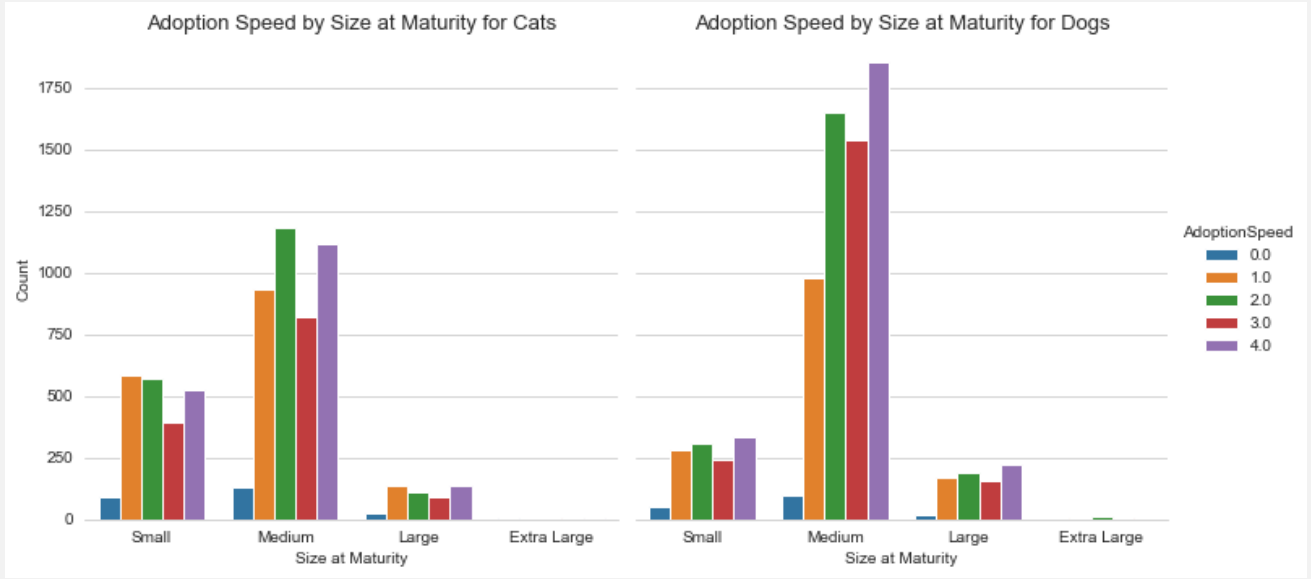

Adoption Speed by Solid Color
AdoptionSpeed
0.0
1.0
2.0
3.0
4.0
Count
3000
2500
2000
1500
1000
500
0
Multi-Colored
Solid
Solid Color


Adoption Speed by Primary Coat Color
AdoptionSpeed
0.0
1.0
2.0
3.0
4.0
Count
2000
1750
1500
1250
1000
750
500
250
0
Black
Brown
Cream
Gray
Golden
White
Yellow
Primary Coat Color

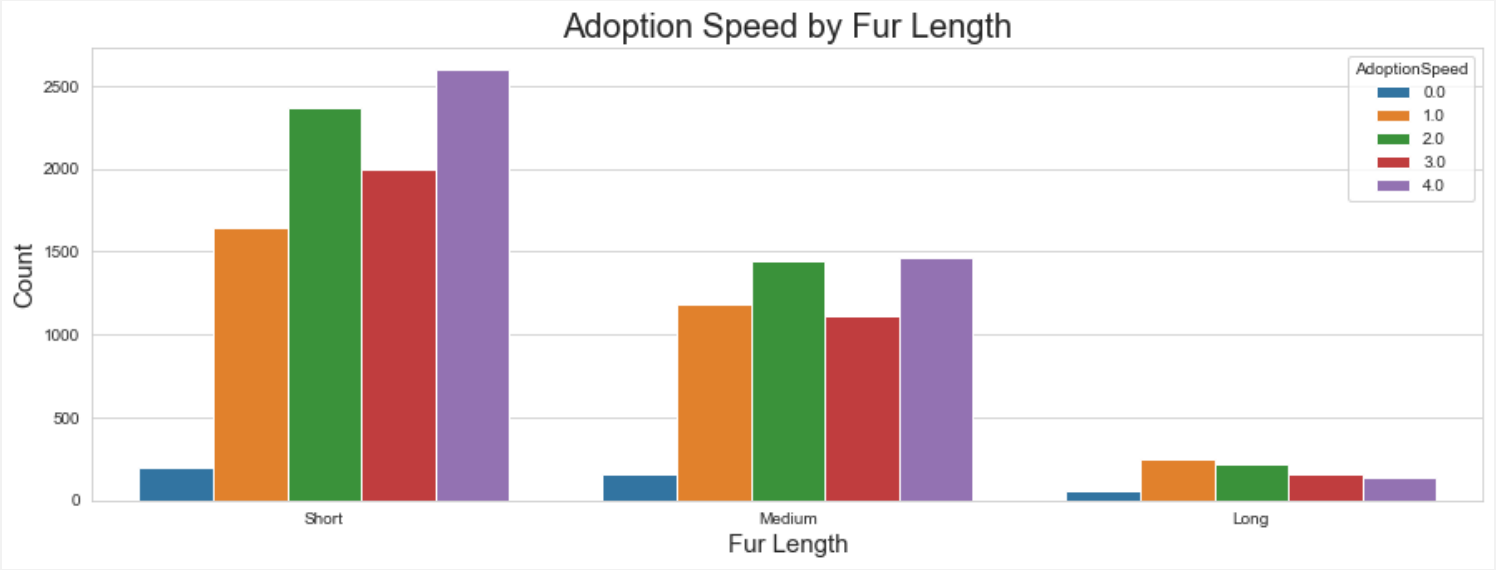

Adoption Speed by Secondary Coat Color
AdoptionSpeed
0.0
1.0
2.0
3.0
4.0
Count
800
600
400
200
0
White
Brown
Gray
Cream
Yellow
Golden
Secondary Coat Color

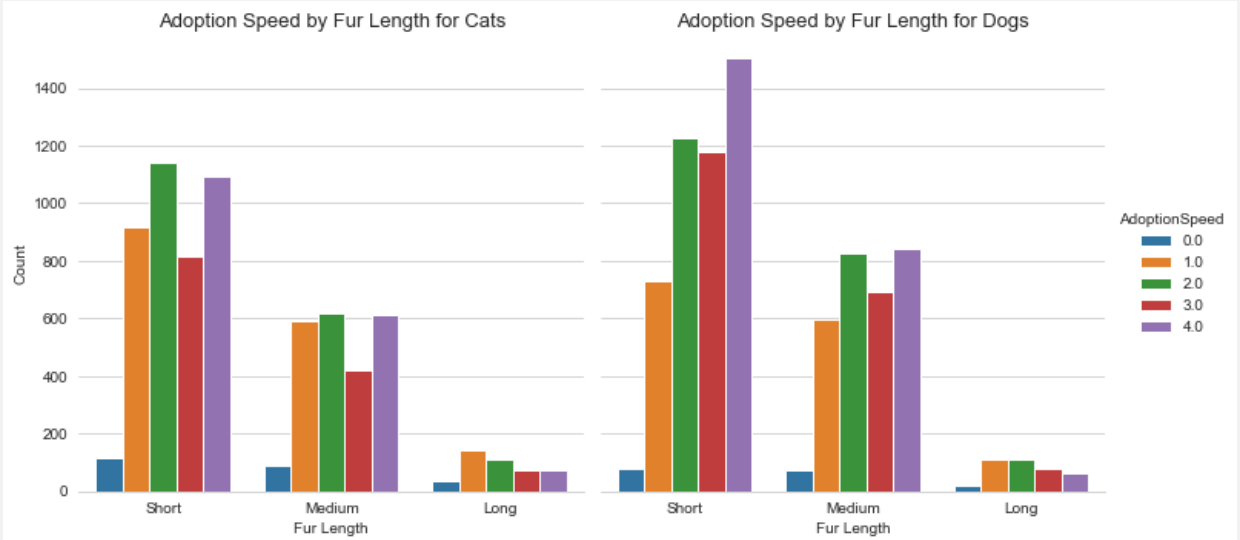
MATURITY SIZE



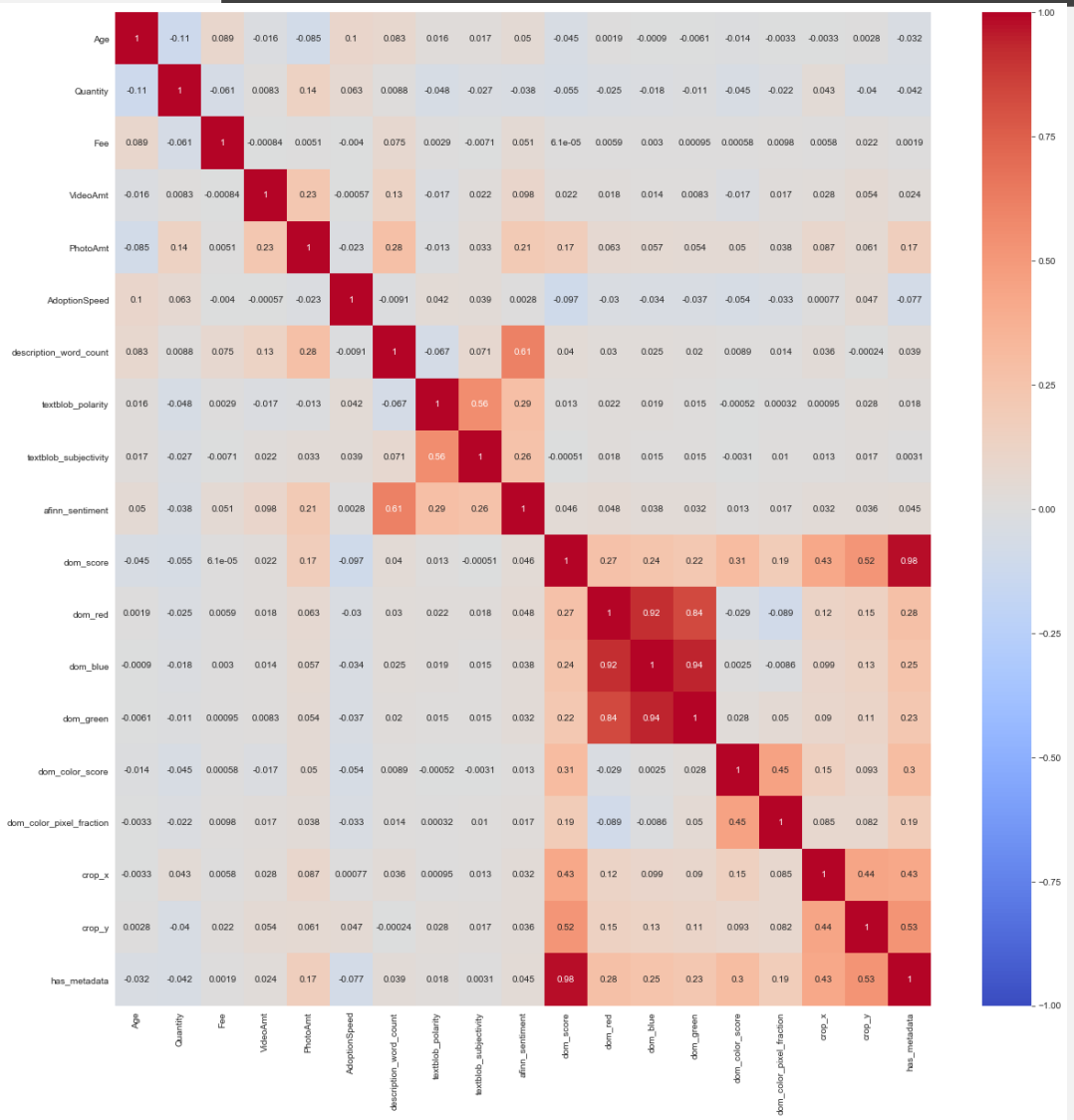


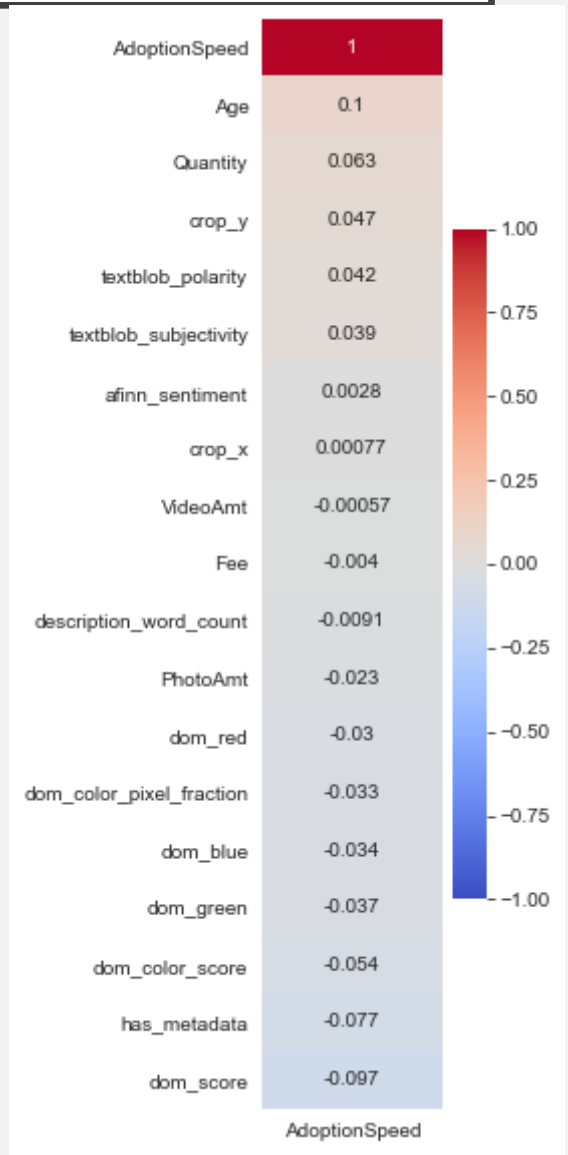
FUR LENGTH


Adoption Speed by Fur Length
Count
2500
2000
1500
1000
500
0
Short
Medium
Long
Fur Length
AdoptionSpeed
0.0
1.0
2.0
3.0
4.0


Adoption Speed by Fur Length for Cats
Adoption Speed by Fur Length for Dogs
Count
1400
1200
1000
800
600
400
200
0
Short
Medium
Long
Fur Length
AdoptionSpeed
0.0
1.0
2.0
3.0
4.0

NUMERICAL VARIABLES





PREPROCESSING & BASELINE MODEL

- Preprocessing:
  - Dummied categorical columns
  - Used TFIDF-Vectorizer on description column
  - Scaled numeric columns using standard scaler
  - 79/21 Train-Test Split on the Training Data

- Baseline Model (metric = accuracy):

4.0	0.279931
2.0	0.269259
3.0	0.217368
1.0	0.206096
0.0	0.027346

LOGISTIC REGRESSION

Best Train Score:

0.37419790611279974

Best Test Score

0.31533820260400125

- Using numeric columns only produced "best" results – lowest bias and variance
- Adding dummied columns and text data produced very high training scores (.98 +) but test scores that were worse than our baseline accuracy.

NEURAL NETWORKS

Best Train Score:

0.3772

Best Test Score:

0.3671

- Best results from using numeric columns only
- Regularization techniques used:
  - Early Stopping
    - Created more overfit models
  - Dropout
    - A little less overfit, with lower test accuracy
  - L2
    - Produced best results

CONCLUSIONS & NEXT STEPS

Adoption speeds are difficult to predict.

Next Steps:

- Incorporate more metadata
- Account for unbalanced classes
- Explore different evaluation metrics

Happy Ending: Milo was adopted the same day he arrived at a shelter!

QUESTIONS?