



# PREDICTING ADOPTION SUCCESS IN ANIMAL SHELTERS

TEAM 2: Andrew Gatchalian, Davidson Rajasekar,  
Devin Xiang Tian, Kaidi Huang, Yuanhui Yao

BANA 288

# TABLE OF CONTENTS

01.



BACKGROUND  
AND PROJECT  
PURPOSE

02.



DATA DICTIONARY &  
PRE-PROCESSING

03.



EXPLORATORY  
DATA ANALYSIS

04.



MODELING AND  
FINDINGS

05.



TIME SERIES +  
FORECASTS

06.



CONCLUSION AND  
IMPLICATIONS



# BACKGROUND AND PROJECT PURPOSE



# BACKGROUND AND PURPOSE

- Approximately 6.3 million companion animals enter U.S. animal shelters nationwide every year. Of those, approximately 3.2 million are cats.
- Each year, around 920,000 shelter animals are euthanized (530,000 being cats). - ASPCA



## How can we reduce euthanasia rates and improve adoption outcomes in animal shelters?

### PROJECT SCOPE & VALUE

- Develop models to predict and forecast cat adoptions
- Identify factors that significantly contribute to adoption success
- Assist animal shelters in strategic planning and resource allocation to improve adoption rates





## AGE

Younger cats may be more desirable for adoption

Longer expected life span and “cuteness”



## BREED & COAT

Certain breeds associated with perceived temperament  
Specific coat colors or patterns may be more attractive



## SPAY/NEUTER

Preference for spay/neutered cats  
Long term health and cost implications



# DATA DICTIONARY & PRE-PROCESSING



# DATA DICTIONARY

Key variable:

## **outcome\_type**

- All possible outcomes of the cat
- Measure of adoption success
- Converted to a binary (1=Adoption, 0>All other outcomes)

outcome_type	
Ultimate outcome for this animal. Possible entries include transferred, [mercy] euthanized, adopted.	
Transfer	45%
Adoption	43%
Other (3366)	11%
Transfer	
Adoption	
Adoption	
Return to Owner	
Transfer	
Adoption	

- Dataset from Kaggle
- Austin Animal Center (largest no-kill shelter in the US)
- 5-Year span (2013-2018)
- 28,209 observations
- 37 variables -> 172!
  - Outcome Type
  - Spay/Neuter
  - Sex (male or female)
  - Outcome Age
  - Outcome Month
  - Outcome Hour
  - Breed (categorical)
  - etc....

Variable	Data Type	Categories	Description
has_name	Binary	0=No, 1=Yes	Whether cat has a name or not
is_adopted	Binary	0=Not Adopted, 1=Adopted	Whether cat was adopted or not <i>**dependent variable**</i>
sex_male	Binary	0=Female, 1=Male	Sex of the cat
spay_neuter	Binary	0=No, 1=Yes	Whether cat was spayed/neutered
is_kitten	Binary	0=No, 1=Yes	If cat or kitten (cats < 6 months old)
is_weekend	Binary	0=No, 1=Yes	Whether outcome occurred during weekend or not
is_shorthair	Binary	0=No, 1=Yes	If cat breed is shorthair or other
is_solid_pattern	Binary	0=No, 1=Yes	If can has solid coat pattern or not
season	Categorical	(Fall, Spring, Summer, Winter)	Season the outcome occurred
time_of_day	Categorical	(Morning, Afternoon, Closed)	Time the outcome occurred
color	Categorical	(Black, White, Other)	Color of the cat

# DATA DICTIONARY

- Kept unique animal IDs
- Removed NA rows
- Dropped unnecessary variables
- Transformed:
  - is\_adopted** \*dependent\*
  - is\_weekend**
  - is\_shorthair**
  - is\_solid\_pattern**
  - season**
  - time\_of\_day**
  - color**
- 28,208 observations
- 13 variables -> 20!

# FINAL DATASET

dat.test	21954 obs. of 16 variables
dat.train	2000 obs. of 16 variables

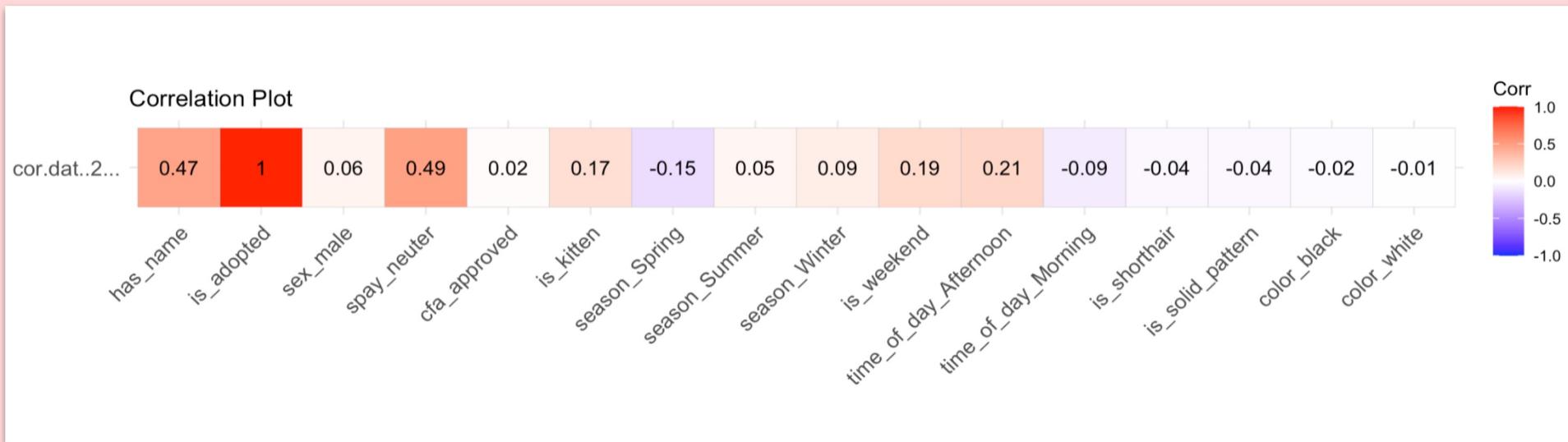
- Created dummies for all categorical variables
- 50/50 training dataset (2000 obs.)
- Balanced testing dataset

#	animal_id	has_name	is_adopted	sex_male	spay_neuter	cfa_approved	is_kitten	season_Fall	season_Spring	season_Summer
1	A684346	0	0	1	0	0	0	1	0	0
2	A685067	1	1	0	0	0	0	1	0	0
3	A678580	1	1	0	1	0	0	1	0	0
4	A675405	1	0	0	1	0	0	0	0	1
5	A670420	0	0	1	0	0	0	1	0	0
6	A684460	1	1	0	1	0	0	1	0	0
7	A673952	0	0	0	1	0	0	0	0	1
8	A686497	1	1	1	1	0	1	0	0	1
9	A687965	1	1	1	1	0	1	1	0	0
10	A668547	1	0	1	1	0	0	0	0	0
11	A682393	1	0	0	0	1	1	0	0	1
12	A681039	0	0	0	1	0	1	0	0	0
13	A682522	0	1	0	1	0	1	0	0	0

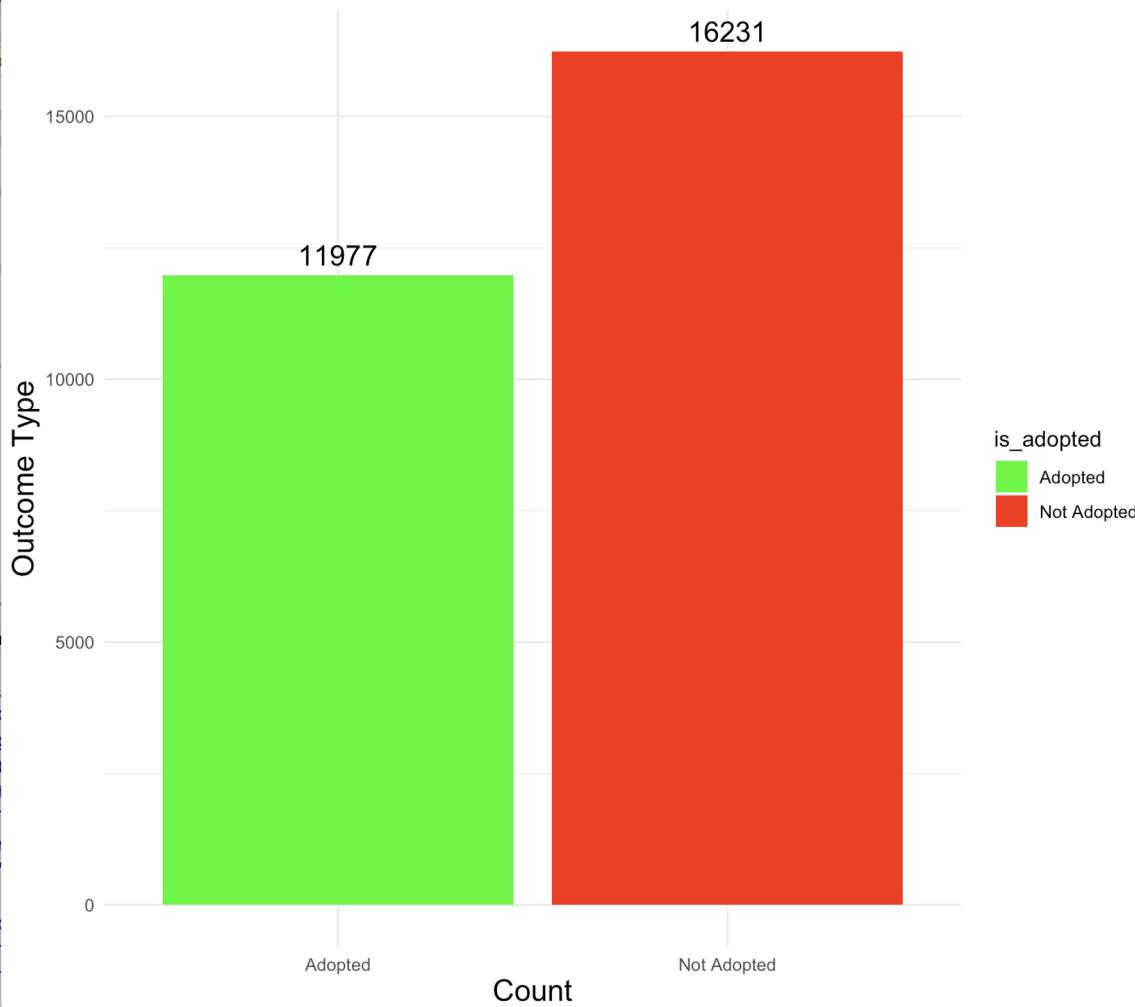
Showing 1 to 13 of 28,208 entries, 20 total columns

# EXPLORATORY DATA ANALYSIS

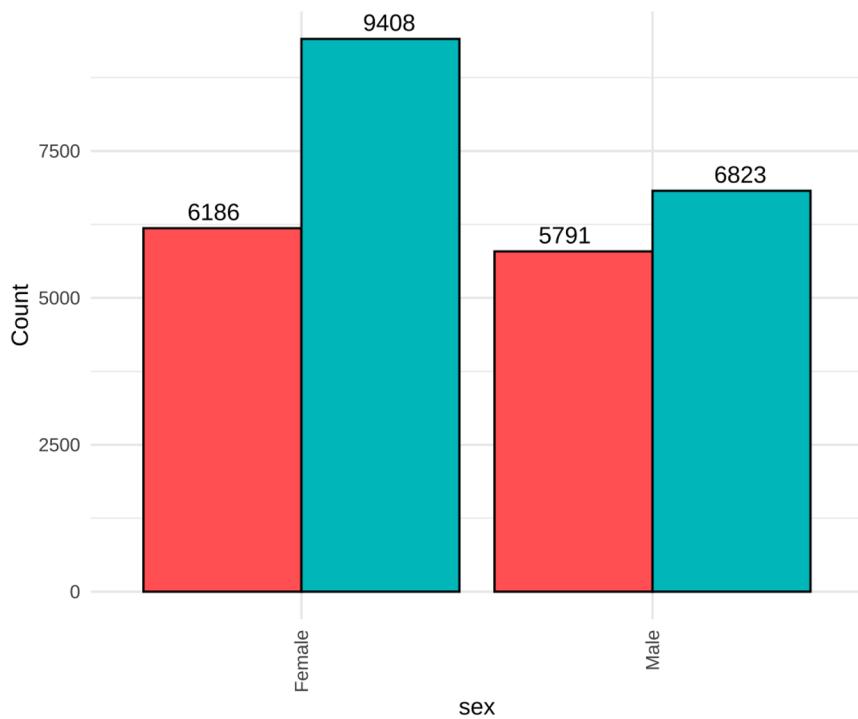




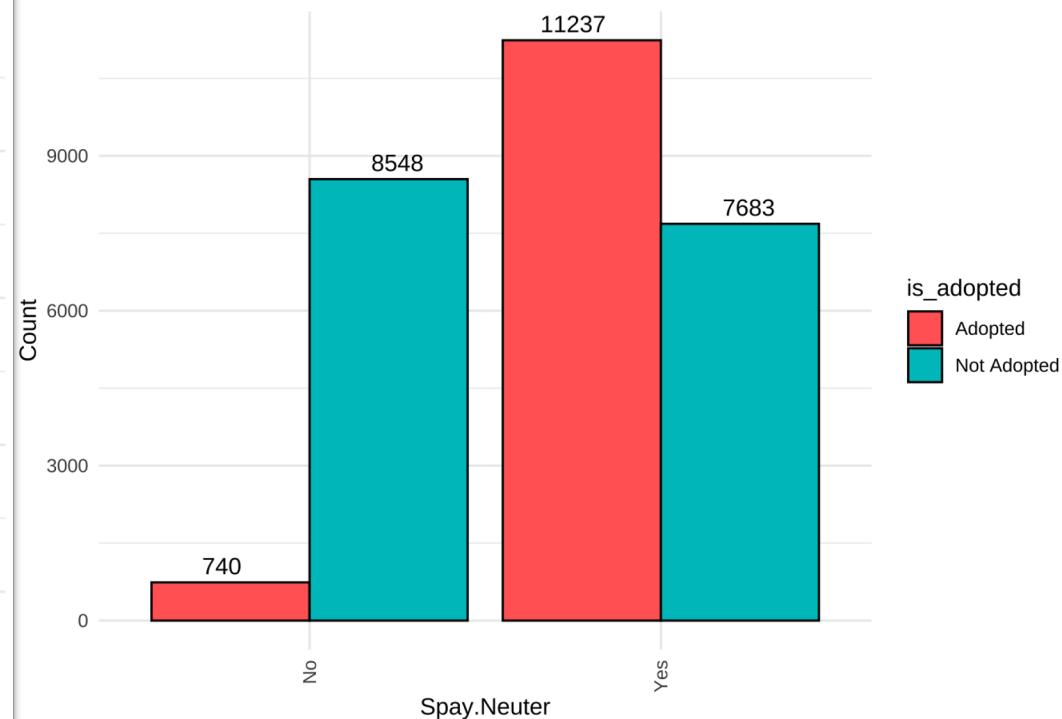
## Outcome Type Counts



Adoption Status by sex



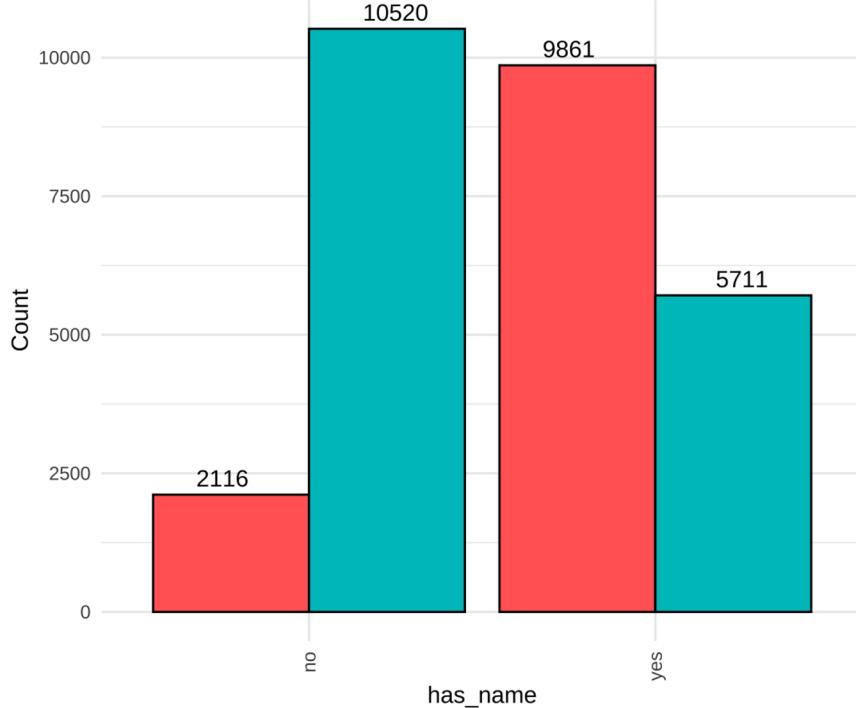
Adoption Status by Spay.Neuter



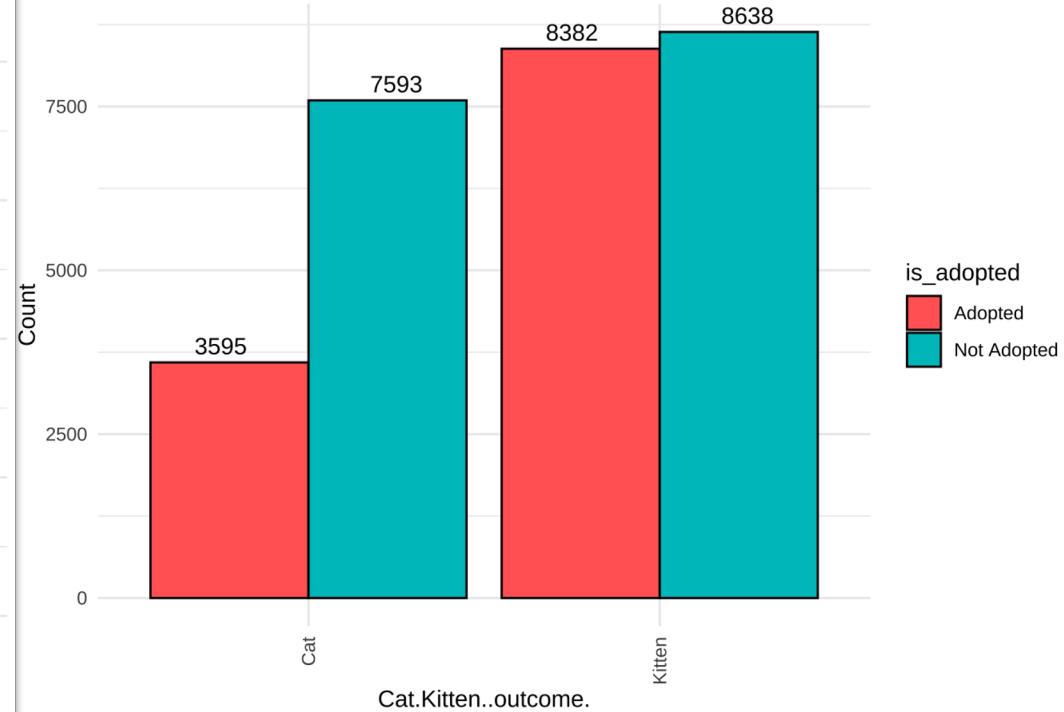
is\_adopted

- Adopted
- Not Adopted

Adoption Status by has\_name



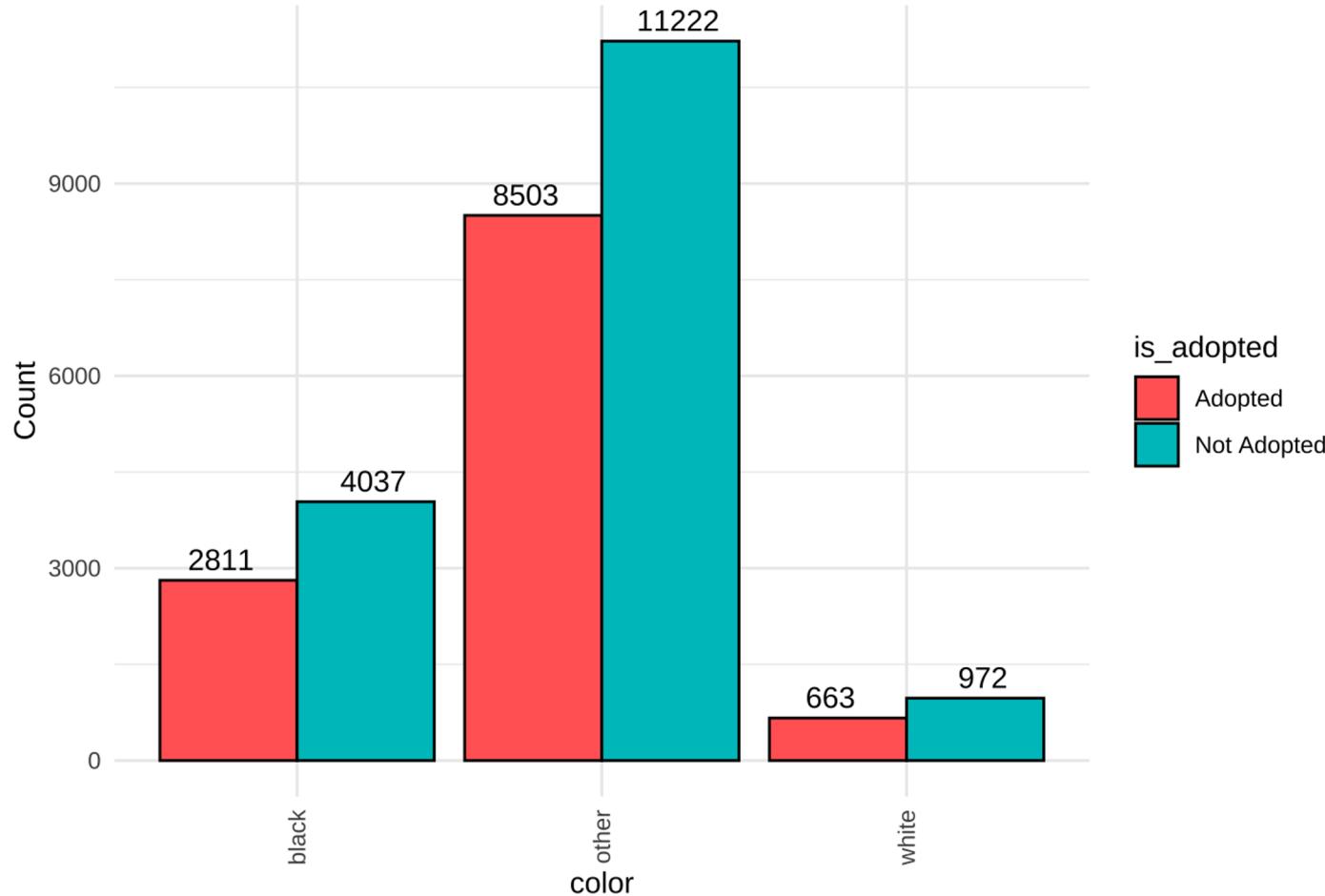
Adoption Status by Cat.Kitten..outcome.



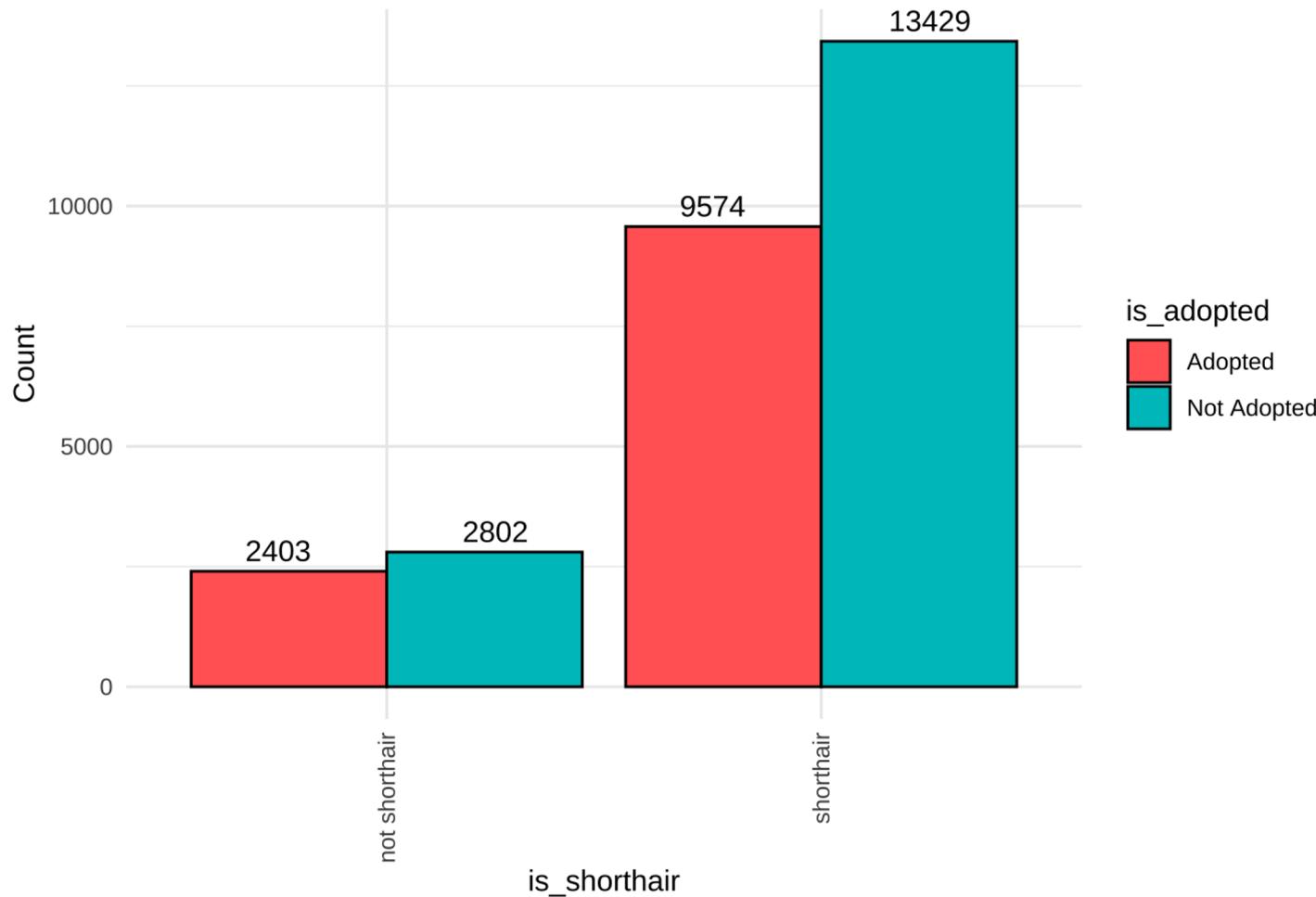
is\_adopted

- Adopted
- Not Adopted

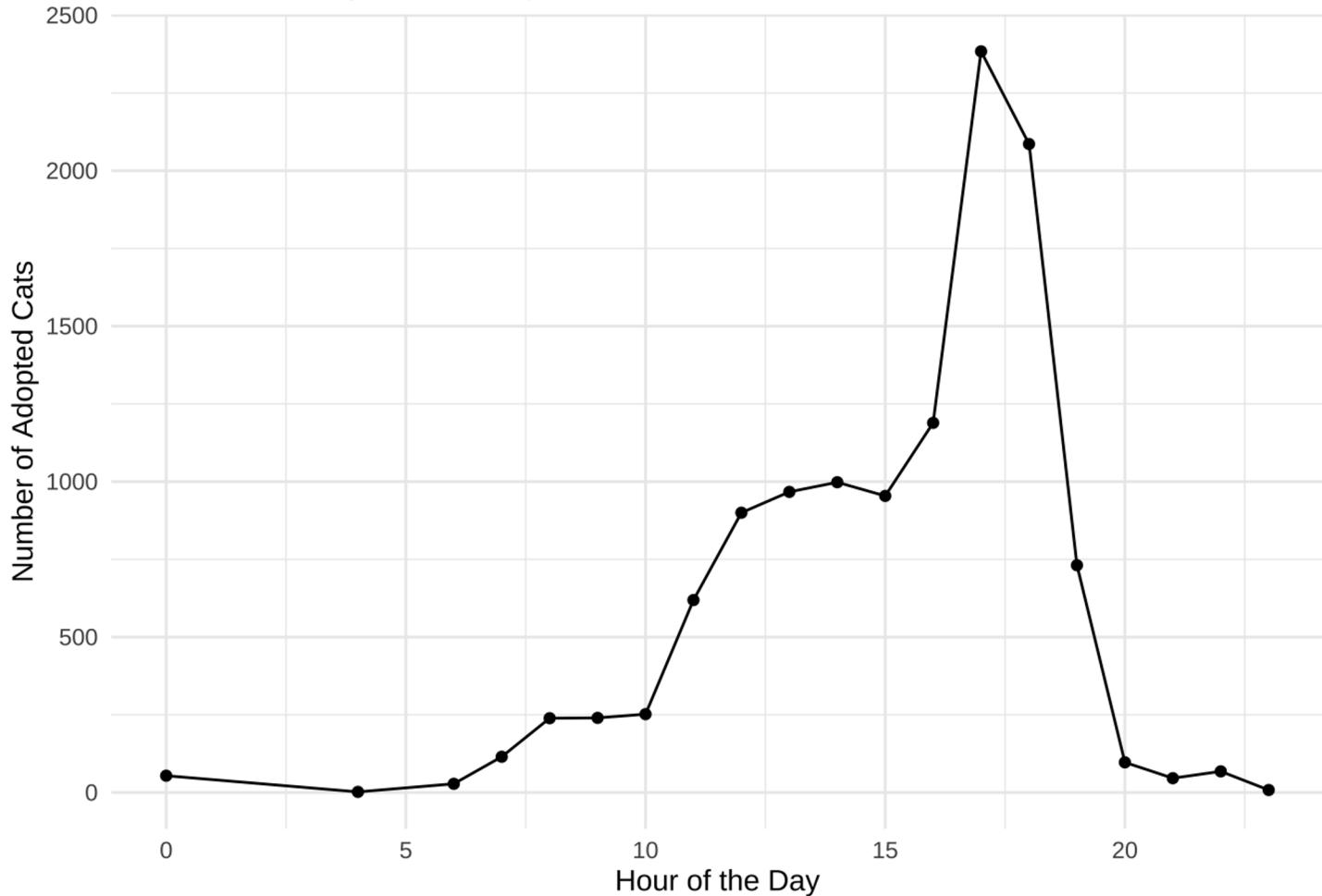
## Adoption Status by color



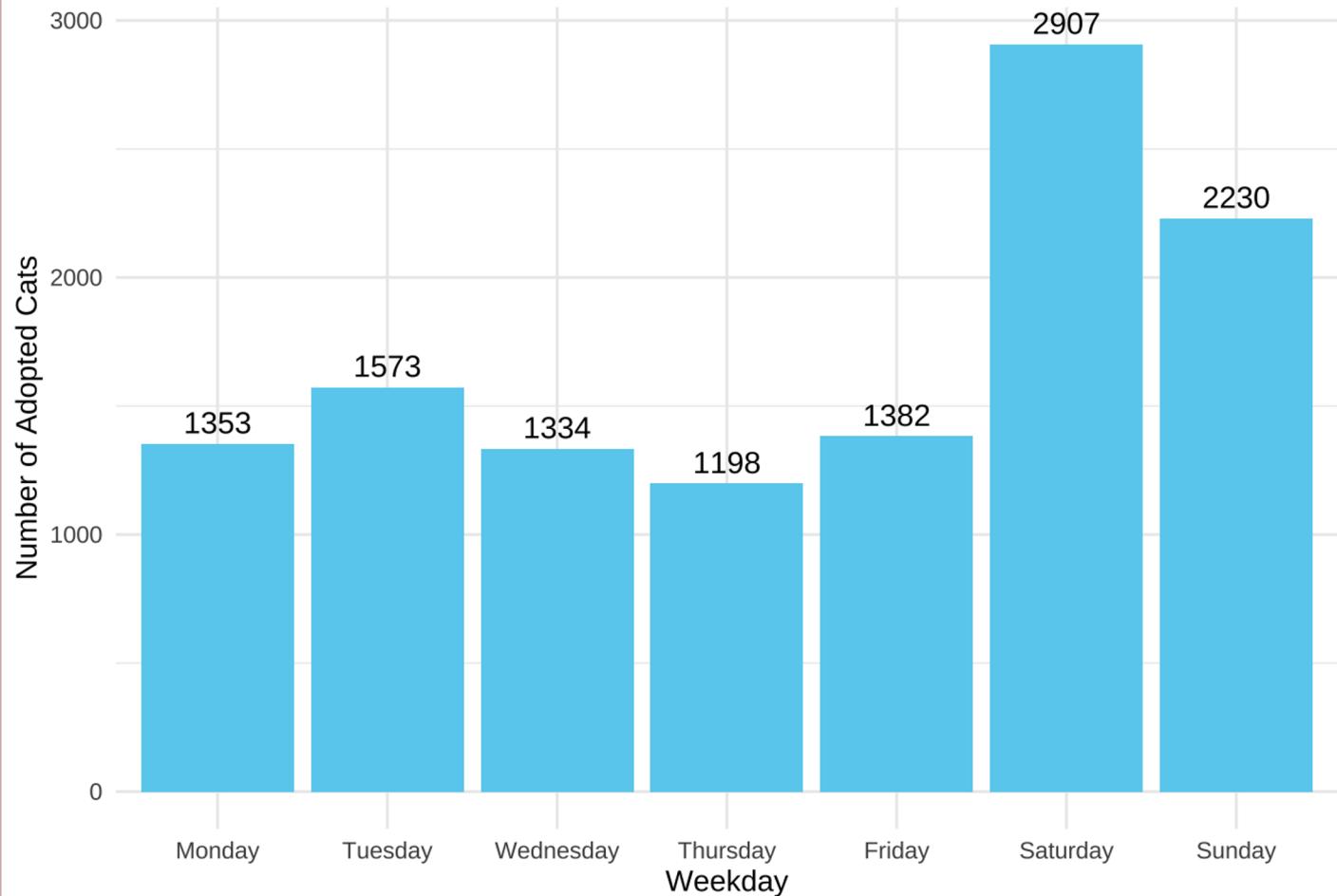
## Adoption Status by is\_shorthair



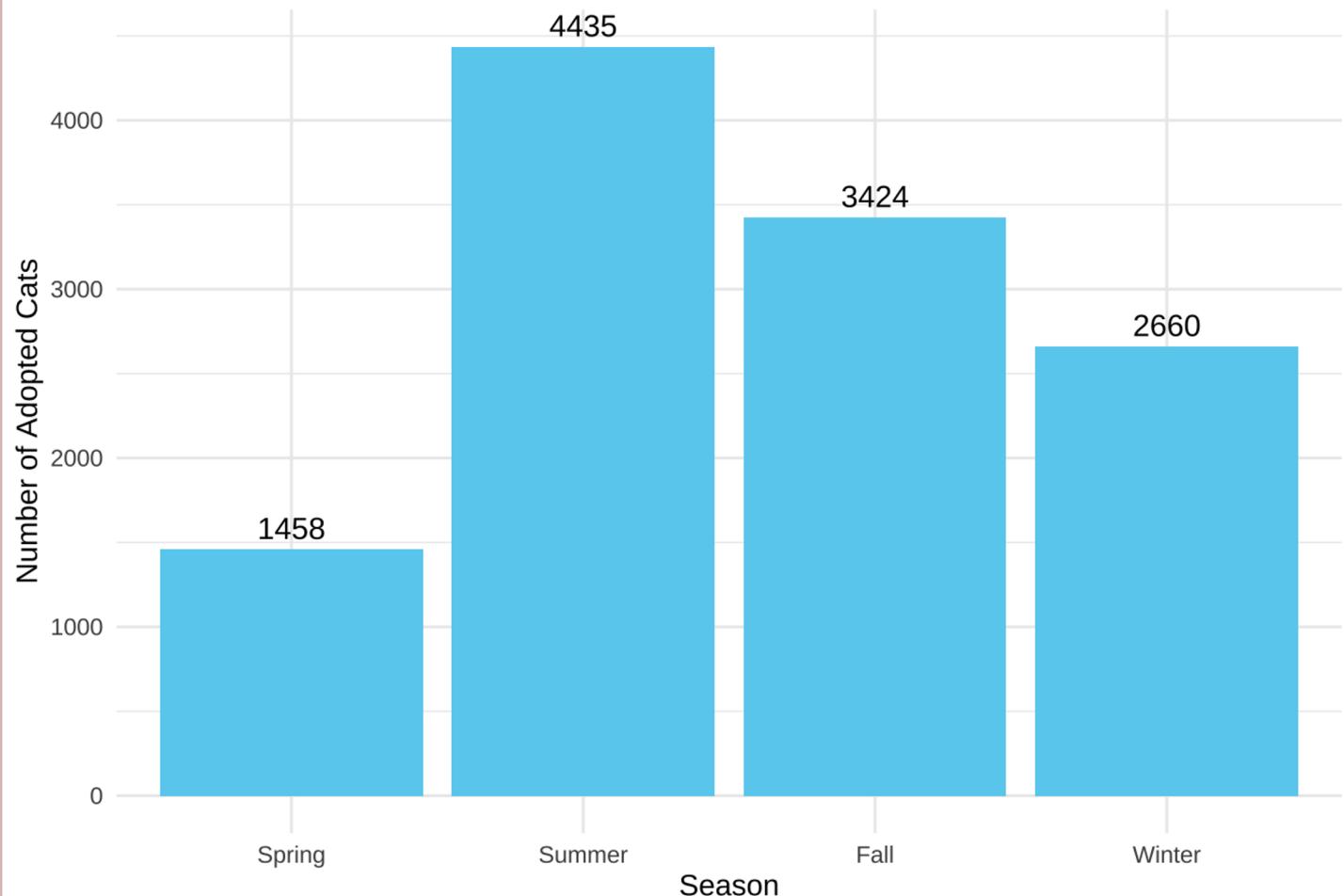
## Number of Adopted Cats by Hour



## Number of Adopted Cats by Weekday



## Number of Adopted Cats by Season



# MODELING AND ANALYSIS



# MODELS EXPLORED

Model	Testing Error	Testing Accuracy
Random Forest	0.1755033	82.45%
Logistic Regression (Model 2)	0.1861164	81.39%
Support Vector Machine	0.1875285	81.25%
Logistic Regression (All-in)	0.1886217	81.14%
Ridge Regression	0.1888039	81.12%
LASSO Regression	0.1902615	80.97%
Decision Tree	0.1932677	80.67%
Pruned Tree	0.1932677	80.67%

# LOGISTIC REGRESSION



Logistic Regression (All-in) Training		Predicted	
		Not Adopted	Adopted
Actual	Not Adopted	773	227
	Adopted	162	838

Logistic Regression (All-in) Testing		Predicted	
		Not Adopted	Adopted
Actual	Not Adopted	8626	2351
	Adopted	1790	9187

Logistic Regression (Model 2) Training		Predicted	
		Not Adopted	Adopted
Actual	Not Adopted	776	224
	Adopted	163	837

Logistic Regression (Model 2) Testing		Predicted	
		Not Adopted	Adopted
Actual	Not Adopted	8698	2279
	Adopted	1807	9170

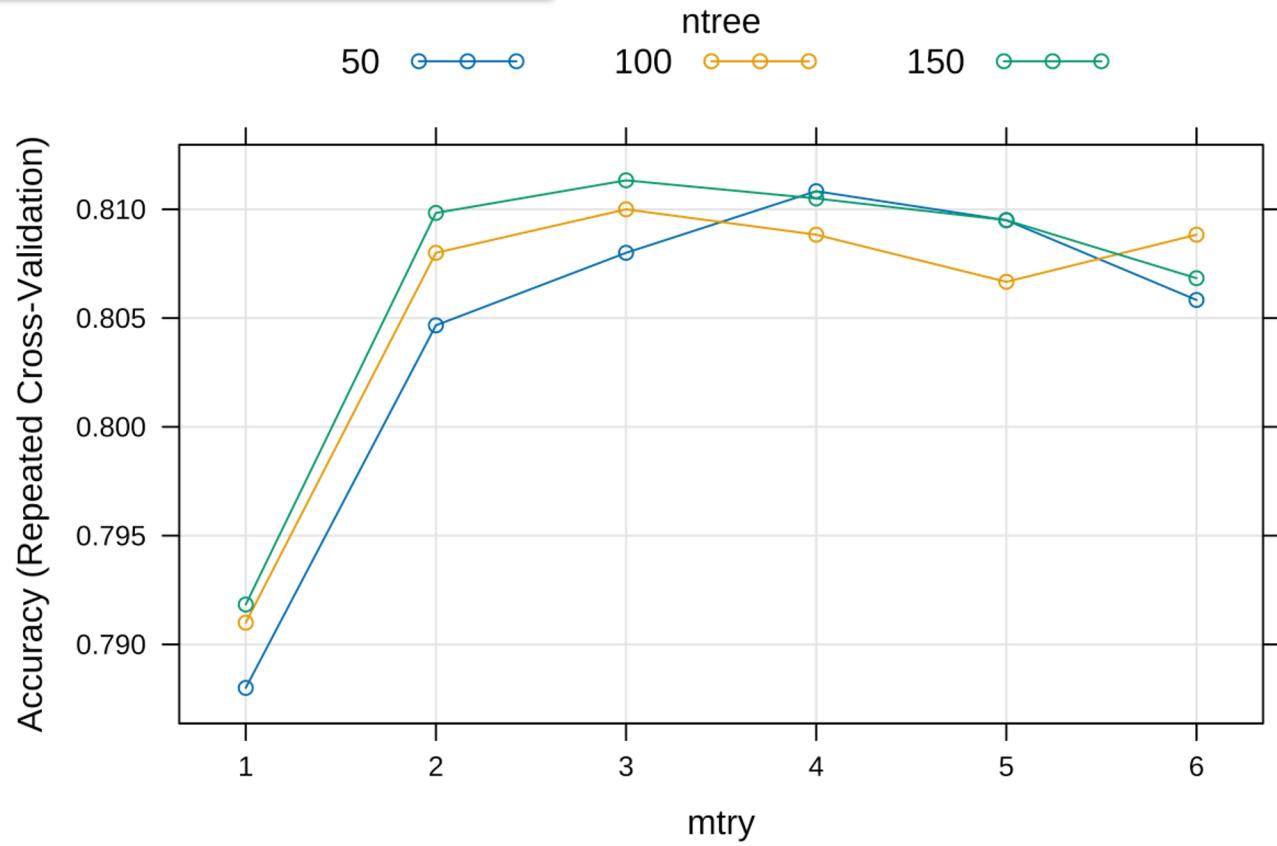
## Model 1 (All-in) Accuracy

- Training: 80.55%
- Testing: 81.14%

## Model 2 (Simplified) Accuracy

- Training: **80.65%**
- Testing: **81.39%**
- Variables removed:  
cfa\_approved,  
season\_Summer,  
is\_shorthair, color\_black,  
color\_white

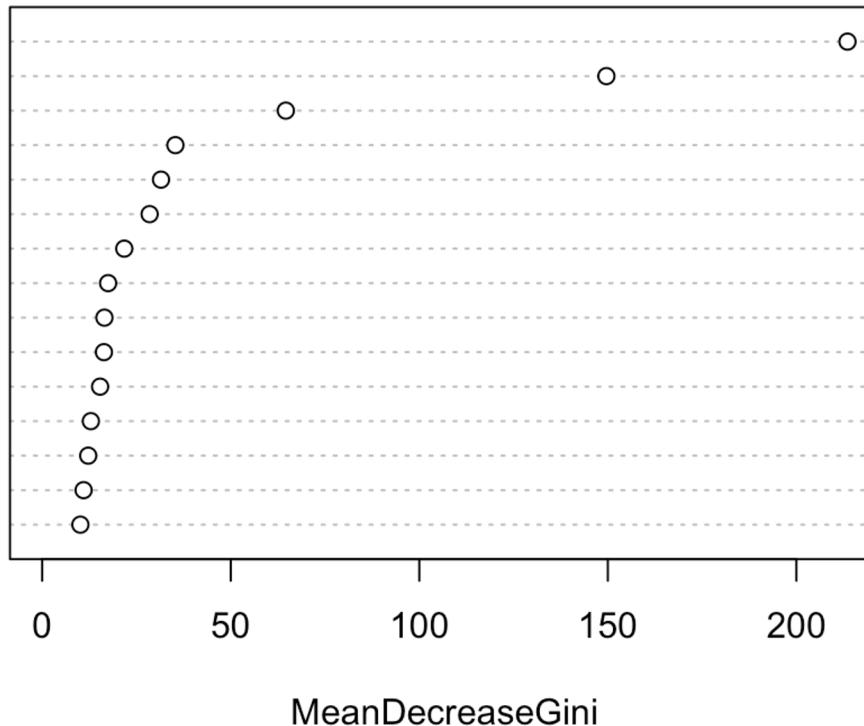
# RANDOM FOREST



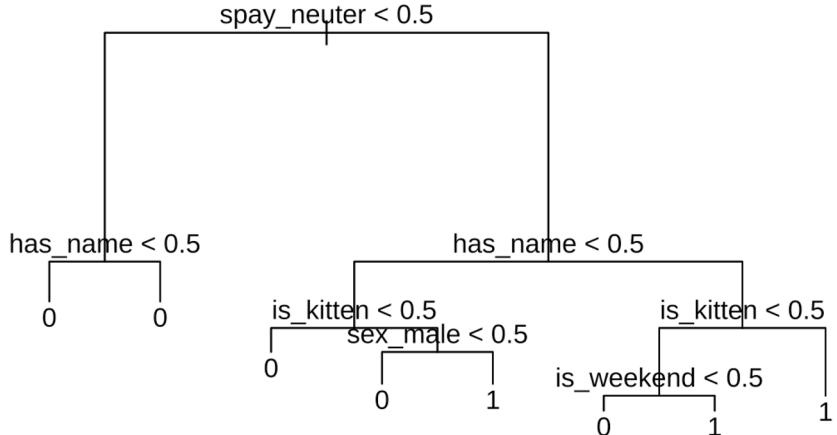
# RANDOM FOREST

**rf.model**

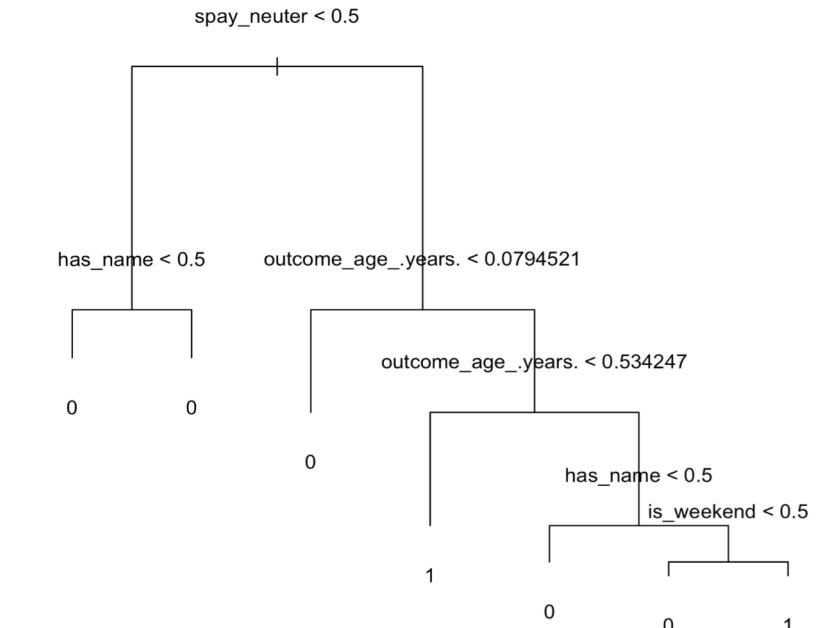
spay\_neuter  
has\_name  
is\_kitten  
sex\_male  
is\_weekend  
time\_of\_day\_Afternoon  
season\_Spring  
is\_solid\_pattern  
time\_of\_day\_Morning  
is\_shorthair  
season\_Winter  
season\_Summer  
color\_black  
cfa\_approved  
color\_white



# DECISION TREE



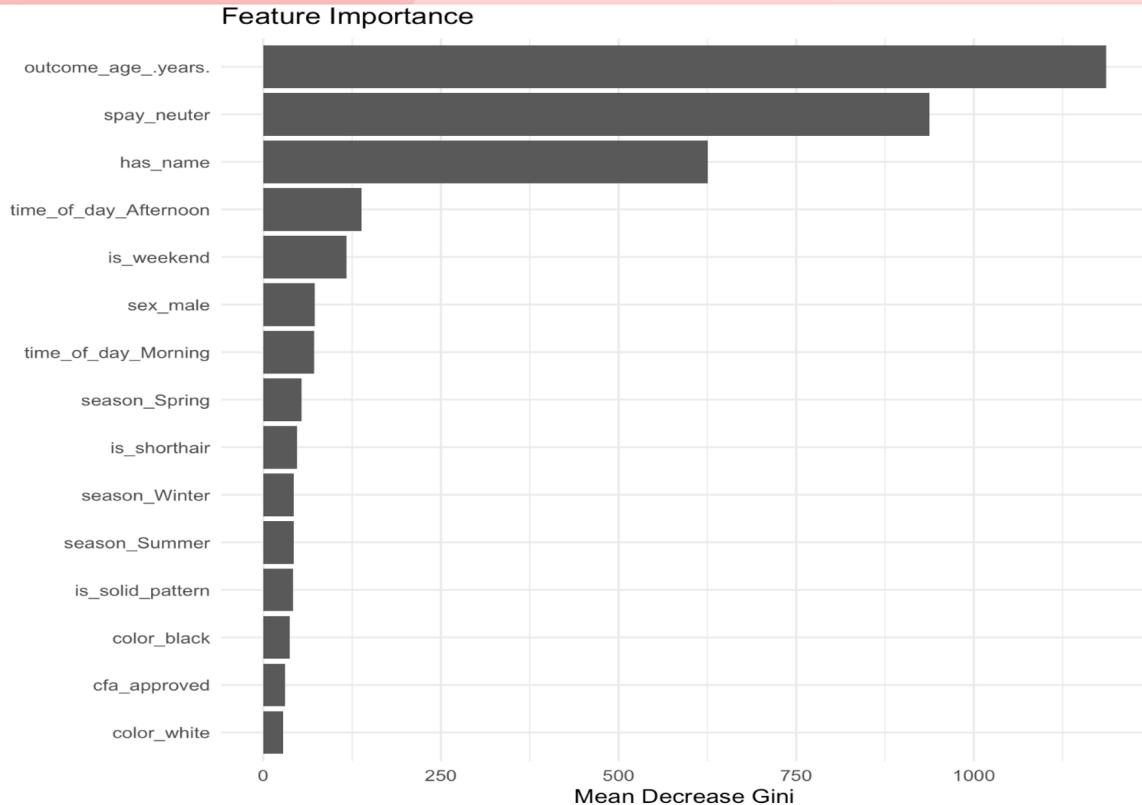
Age factor as  
binary variables  
**Accuracy Rate: 80.67%**



Age factor as  
continuous variables.  
**Accuracy Rate: 84.5%**



# RANDOM FOREST



- **Accuracy : 85%**
- parameters
  - mtry=4 (number of features selected in each model)
  - ntree=1000 ( there are 1000 tree models)
- **“Mean Decrease Gini”** = average decrease in impurity caused by splits over a given feature across all trees in forest.
- The model heavily relies on the age of the outcome for making predictions
- Factors related to the animal's physical condition(spay\_neuter, has\_name) also play essential roles
- temporal factors(time of day, whether its a weekend) and the animal's sex have moderate importance



# MODELS EXPLORED (2)

Model	Testing Error	Testing Accuracy	Previous Testing Accuracy
Random Forest w/ continuous Age	0.1497221	85%	82.45%
Decision Tree w/ continuous Age	0.1612007	83.88%	80.67%



# FINDINGS

- All models perform relatively similar (around 80%)
- Key variables: spay\_neuter, has\_name, is\_kitten, sex\_male, is\_weekend
- Continuous Age variable improves model performance
- Cat colors and breeds are **NOT** good predictors
- Adoption season matters:
  - Cats are less likely to be adopted during Spring
  - Cats are more likely to adopted during Winter
- Day of the week matters:
  - Cats are more likely to be adopted during weekends
- Time of day matters:
  - Cats are more likely to be adopted during afternoon hours (2pm-7pm)



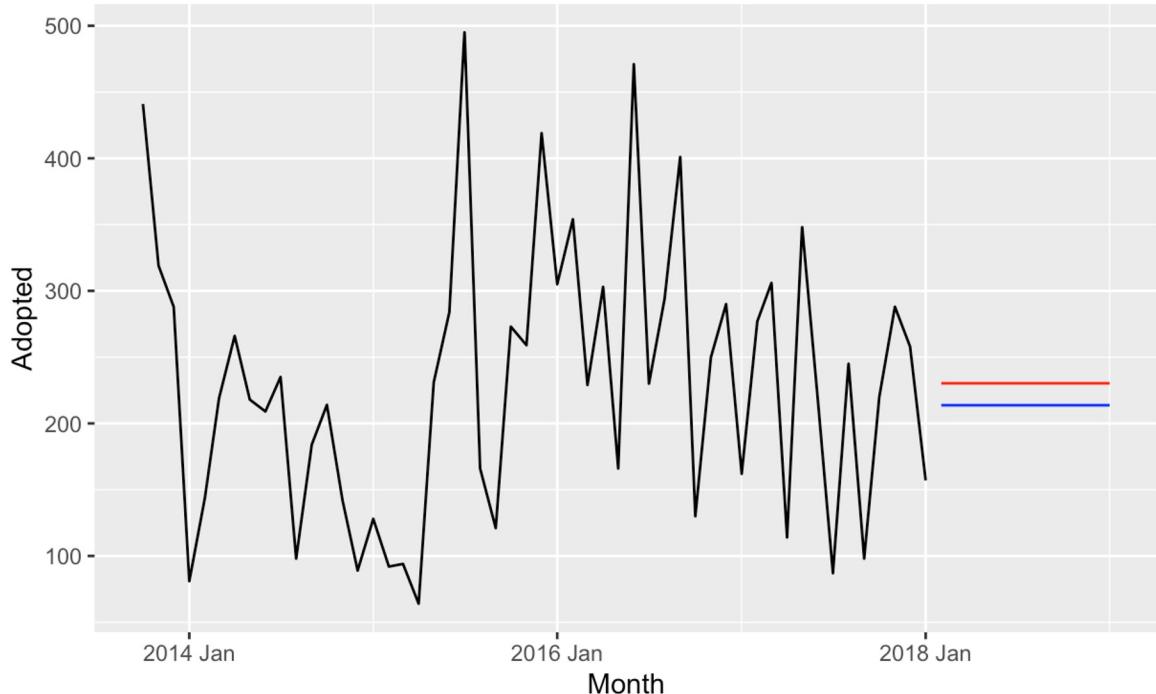
# TIME SERIES + FORECASTS



# 1-YEAR FORECAST



Cat Adoptions 1 Year Forecast  
ARIMA (Red) & ETS (Blue)



Left graph shows the time series of cat adoptions at animal shelters from October 2013 through early 2018.

- **ARIMA model (in red) and an ETS model (in blue).**
- The ARIMA and ETS forecasts are quite close to each other, indicating that both models have captured the underlying pattern in the data similarly
- **RMSE(ARIMA-104, ETS-105)**
- **MAE(ARIMA-82.4, ETS-81.6)**

# CONCLUSION AND IMPLICATIONS



# CONCLUSION

- Less is more...
  - Grouping variables seemed to improve model interpretation and accuracy (in some cases)
  - Performance was similar across all selected models
- Variables that **Influence** Adoption:
  - Spay/Neuter
  - Name
  - Gender
  - Age
  - Weekend
  - Time of Day
  - Season
- Variables with **Limited** Predictive Power:
  - Breed
  - Color
  - Coat Pattern



# IMPLICATIONS + ACTION ITEMS



## SPAY AND NEUTER

- Investment in spaying and neutering programs



## CAT NAMES

- Giving cats names improves likelihood of adoption

## WEEKEND + OPENING HOURS

- Extending shelter hours during weekends

## SEASONAL OPERATIONS

- Special adoption events during holiday seasons

# LIMITATIONS + FUTURE RESEARCH

## **Limitations:**

- Data may not capture all factors that influence adoption success (such as personality traits or interactions with potential adopters)
- Findings may not be applicable to shelters outside of the Austin Animal Shelter due to differences in practices or demographics
- Operational practices may have changed since latest update

## **Future Research:**

- Inclusion of behavioral data
- Expansion to shelters across different regions or countries
- Extend research to other types of animals





THANKS !