Pet Adoption

Adoption Timeframe Prediction

By Jessica Lewis

The Problem

Can we predict how long until an animal is adopted based on their location and the characteristics listed on their online profile?

Using these predictions and extrapolating the important features, can we fine-tune an animal's online profile to increase their rate of adoption?

The Data

Petfinder API

(using the PetPy library)

US Census

(basic population demographics)

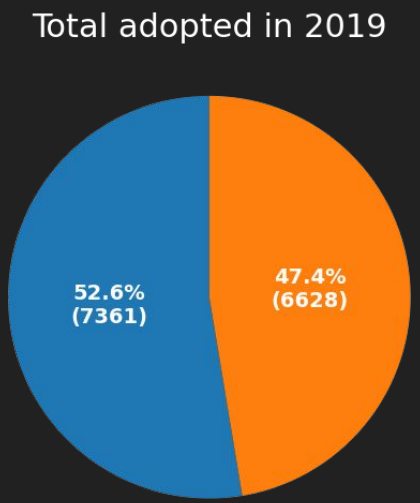
Data Cleaning and Preprocessing

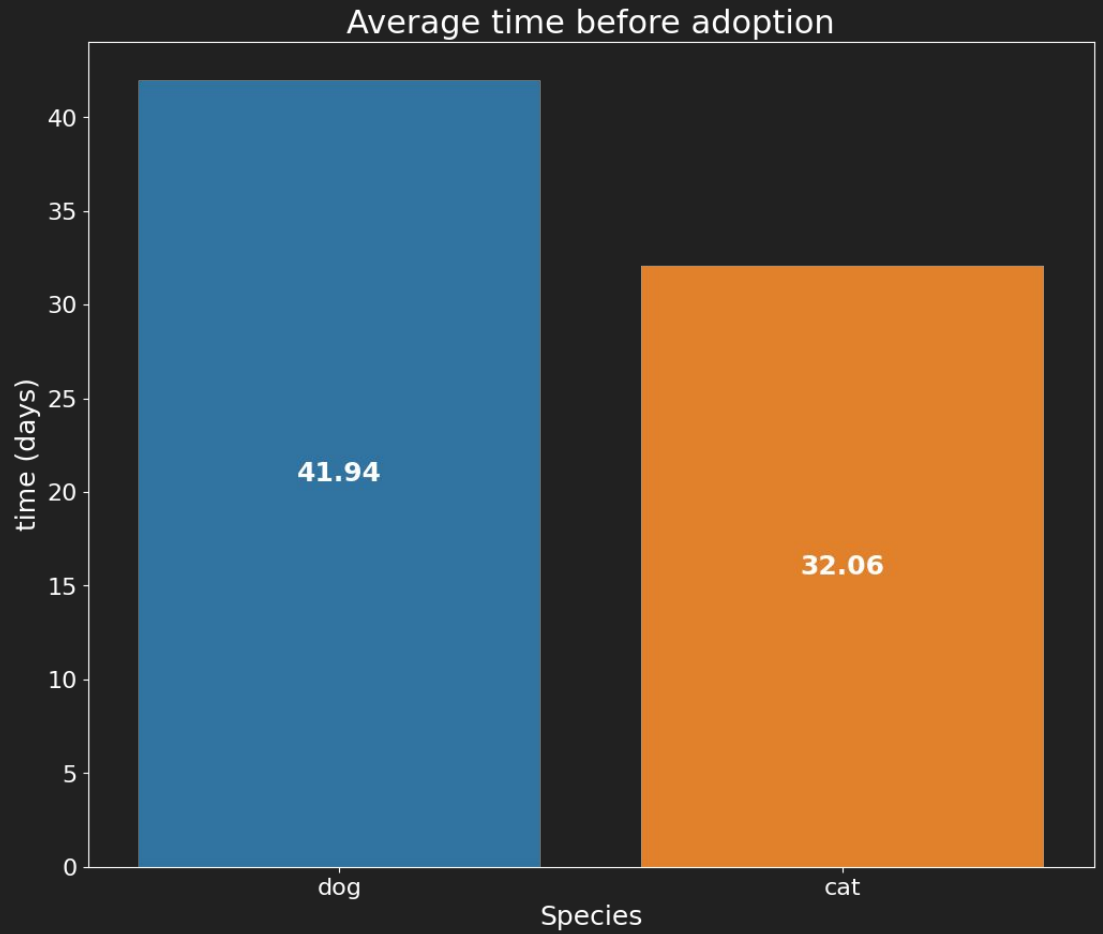
- Cats and dogs are processed separately
- Dictionaries are pulled into their own fields
- Unnecessary columns are dropped
- Population data is added from census import
- Duplicate rows are removed
- Missing data is replaced or dropped
- Trimmed outliers
- Data is run through a scaler

EDA

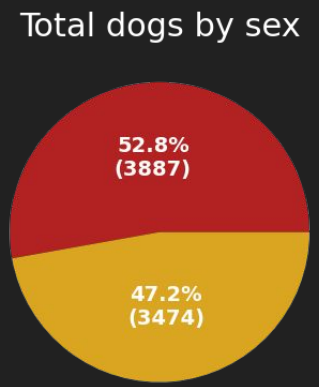
Let’s take a look at some of the relationships between our dependent variable (duration_as_adoptable) and some other features.

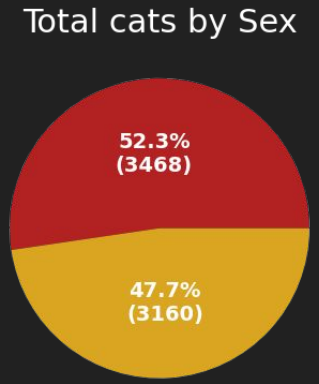
Adoption Time Comparison

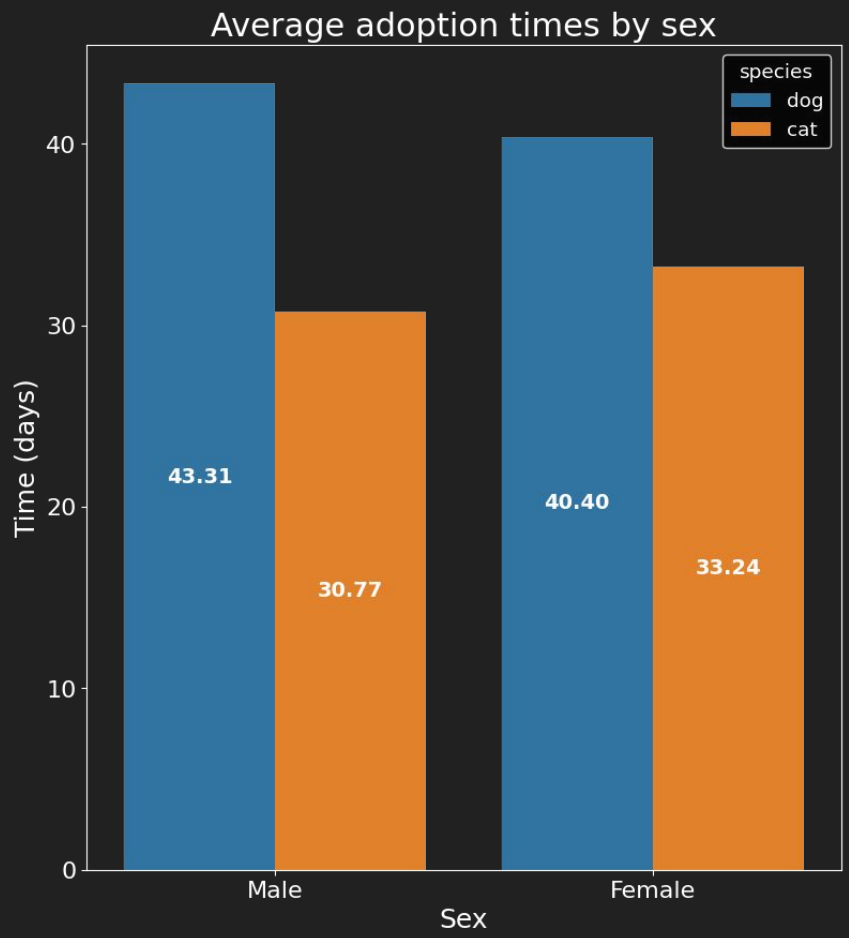




Adoption Comparison by Sex

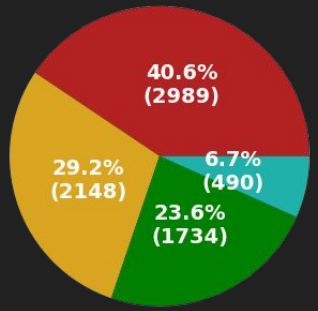




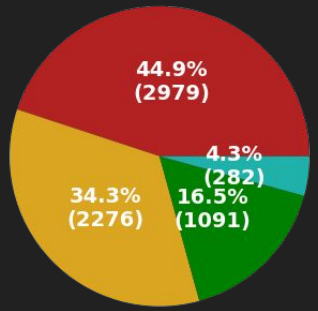


Adoption Comparison by Age

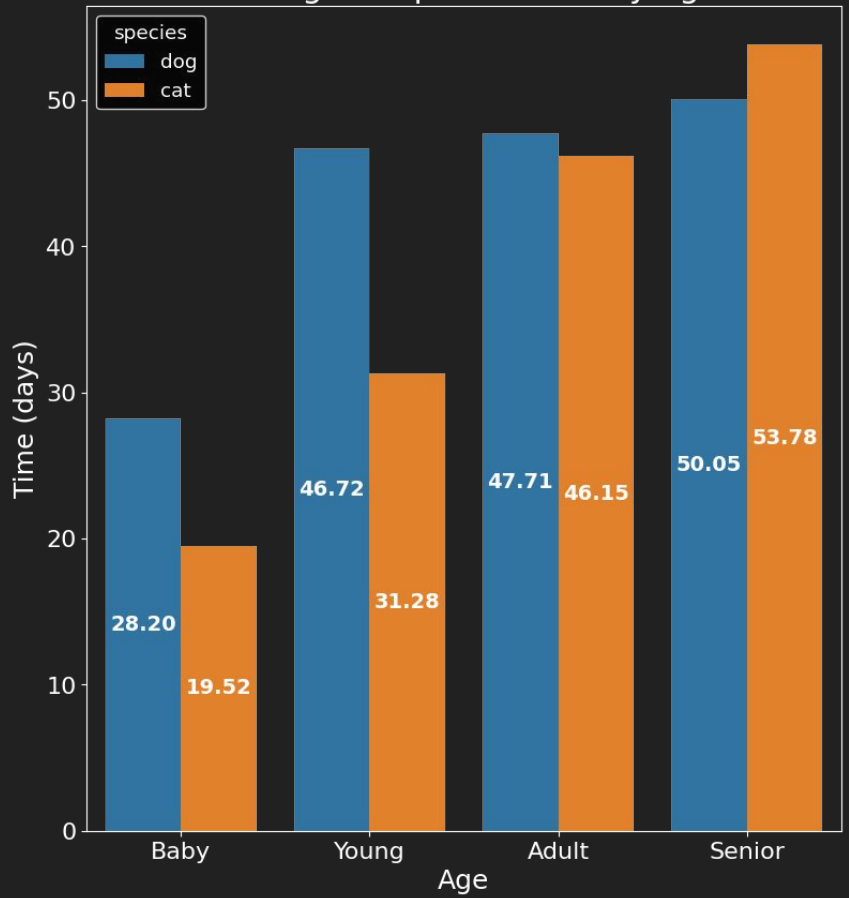
Total dogs by age



Total cats by age

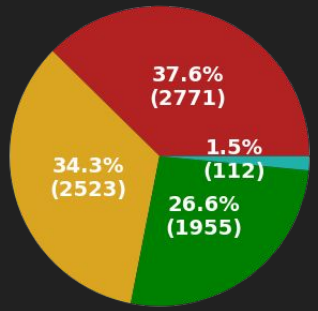


Average adoption times by age

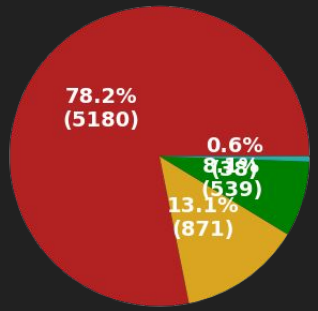


Adoption Comparison by Size

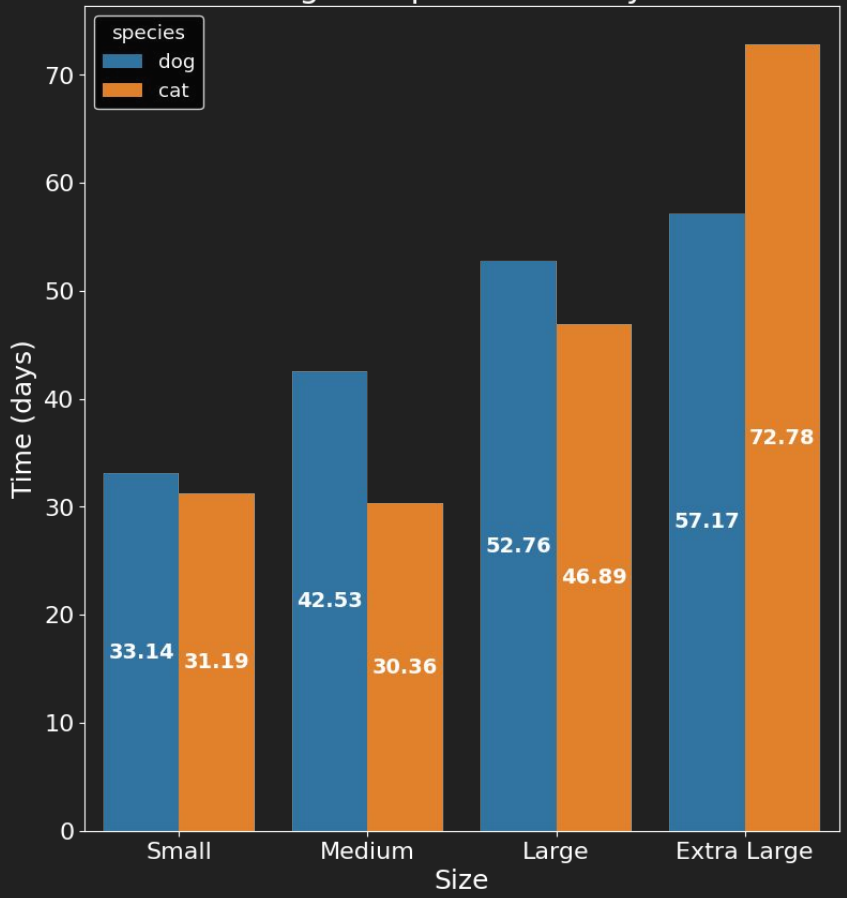
Total dogs by size



Total cats by size

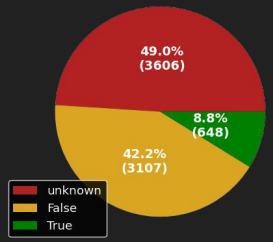


Average adoption times by size

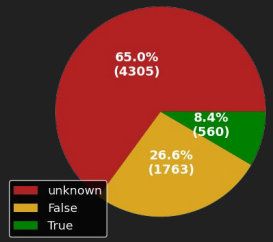


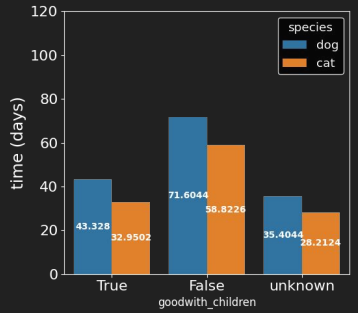
Average adoption times by compatibility

Total dogs good with children

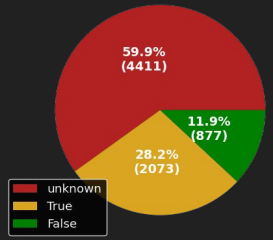


Total cats good with children

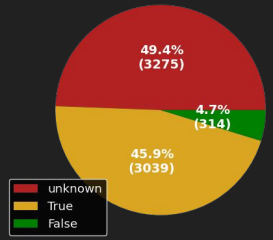




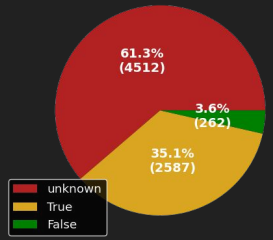
Total dogs good with cats



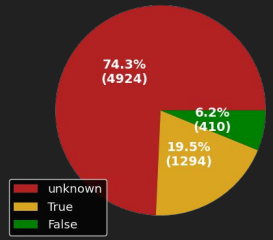
Total cats good with cats



Total dogs good with dogs



Total cats good with dogs



Feature Selection

- Used 3 models for feature selection
  - RandomForest
  - XGBoost
  - F_regression
- All initial features ended up being included in the model.

Feature Selection: Dogs

- gender
- size
- coat
- distance
- spayed_neutered
- house_trained
- special_needs
- shots_current
- breed_primary
- breed_mixed
- color_primary
- goodwith_children
- goodwith_dogs
- goodwith_cats
- hasimage
- hasvideo
- city
- population

Feature Selection: Cats

- age
- breed_mixed
- breed_primary
- city
- coat
- color_primary
- declawed
- distance
- gender
- goodwith_cats
- goodwith_children
- goodwith_dogs
- hasimage
- hasvideo
- house_trained
- population
- shots_current
- size
- spayed_neutered
- special_needs

Model Selection

I tested four models and assessed their performance on each dataset

Dogs

	R^2	RMSE
XGBoost	0.173	0.888
GradientBoosting	0.18	0.891
KNNeighbors	0.135	0.896
RandomForest	0.162	0.908

Cats

	R^2	RMSE
GradientBoosting	0.123	0.906
RandomForest	0.0997	0.909
XGBoost	0.1026	0.912
KNNeighbors	0.0469	0.934

Hyperparameter Tuning

Tuned parameters and their values:

Dogs

Model: XGBoost

- 'objective': 'reg:squarederror'
- 'n_estimators': 26

Cats

Model: GradientBoostingRegressor

- 'learning_rate': 0.1
- 'max_depth': 2
- 'n_estimators': 233

I probably should have done more tuning, but by this point I'd spent way longer than I'd scheduled for this project and honestly the results weren't amazing.

Model Performance

Average time before adoption

	Predicted	Actual
Dogs	31 days	32 days
Cats	24 days	24 days

Average prediction is actually pretty good, but you'll see in the next slide that individual predictions are not great.

Model Metrics

	R^2	RMSE	Mean Absolute Error
Dogs	.11	41.89	26.97
Cats	0.04	34.78	22.83

Neither model performs exceptionally well.

Considering the mean absolute error is essentially the margin of error, we're looking at an average of 31 days for a dog to be adopted plur or minus 27 days. Likewise, it's 24 days for cats plus or minus 23 days.

Client Recommendations

Improve the Models

- Add more features
  - Feature engineering
  - Demographic data
  - Animal profile image data
- Further hyperparameter tuning in final models
- Change dependent variable from continuous to finite
  - Make date ranges to classify records into