A Home for All

Milestone Report
January 2020
C. Jansson

Agenda

- Project Purpose
  - Customer Overview
  - Data Set Overview
- Data Wrangling Methods
- Exploratory Data Analysis and Statistical Findings
  - Feature by feature findings
- Summary

Project Purpose

- Local area animal shelters are struggling
  - Pet intake is at an all time high with
    - 950 dogs coming into the shelter in April 2019 alone
  - Adoption rates are not keeping up
- Specialty rescues are especially struggling
- The question:
  - What if there were a way to increase the likelihood that an animal gets adopted?
    - Can targeted animal “marketing” be leveraged to increase odds of adoption for at risk animals?

Customer Overview

Big Bones Canine Rescue

- Specializes in adoption of oversized breed dogs
- Mission: “save dogs from death, illness, and misfortune regardless of their size, age, breed, or health status.”
    - Accept dogs via owner relinquishment, surrounding kill-shelters, and as deemed necessary by the shelter owner

Dataset Overview

- Data utilized will be drawn from the Kaggle PetFinder Competition
  - https://www.kaggle.com/c/petfinder-adoption-prediction
  - Over 10,000 animals from the Malaysian PetFinder database.
  - 9 tables with feature data
  - Includes a predefined test and train dataset
  - Includes test images and videos which are not addressed in the scope of this project
- By leveraging “market” trends in adoptable dog data sets
  - Target features for dogs that are high-risk for going unadopted
  - Identify potential areas of improvement for internet listings
  - Identify which dogs - based on appearance alone- will need more foster support and additional involvement in adoption events

Data Wrangling

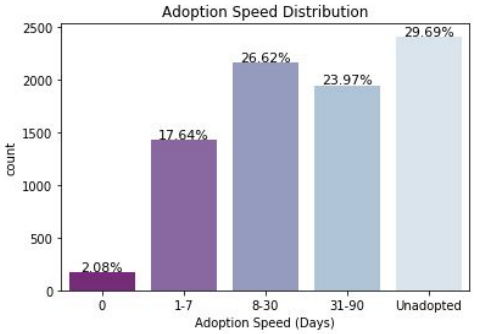
- Data was relatively clean
- Almost all features have been converted to ordinal values in place of categorical data
    - Null values were replaced with 0’s for balanced distributions during statistical analysis
- “Description” data was not targeted as the customer’s focus is more targeted at animal traits specifically
- “Name” information had to be heavily cleaned:
    - Several non-null values that were non-useful data
    - Examples include: ““Lost Dog”, “Cute Puppies”, “Boy”, “Girl”, Miscellaneous breed names (labrador, terrier, etc.), “No Name Yet”, “Please Name Me”, “Save me or I'll Die”, “Urgent home needed”, “Puppy”, and various descriptions of puppies (happy puppy, big eyes, sad puppy, bouncy puppy, etc.)

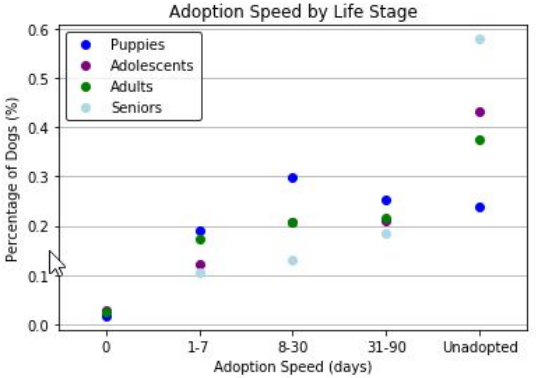
EDA & Statistical Methods

- Data was explored graphically via Exploratory Data Analysis (EDA)
- Statistical methods used:
  - Chi-Square testing
    - More than 90% of data is categorical
  - Cramer’s V Method:
    - A method of categorical correlation for data sets with more than 2 features

General Information on Dataset

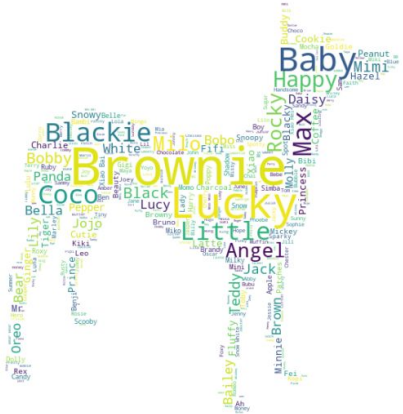
- Approximately 30% of dogs are unadopted
  - As compared to 50% in US
- Only ~2% are adopted on the same day they're listed
- Younger dogs are adopted more quickly than more senior dogs

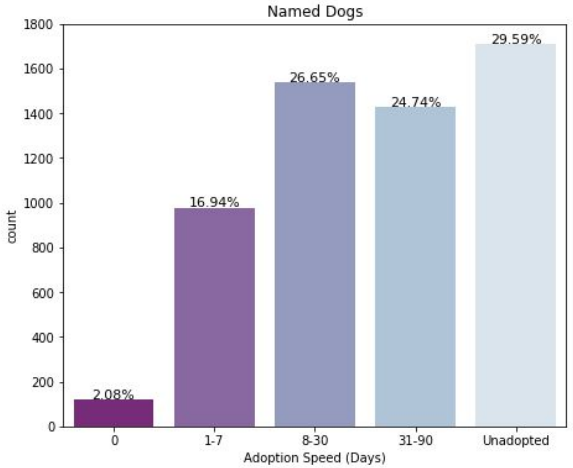


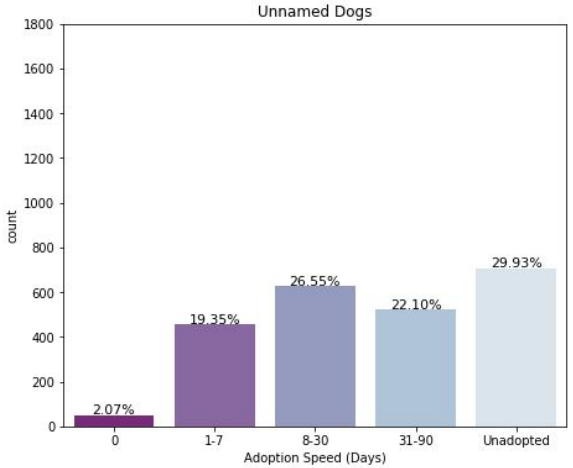


Animal Name

- Having a name is more important than the specific name
- Most common names are descriptive

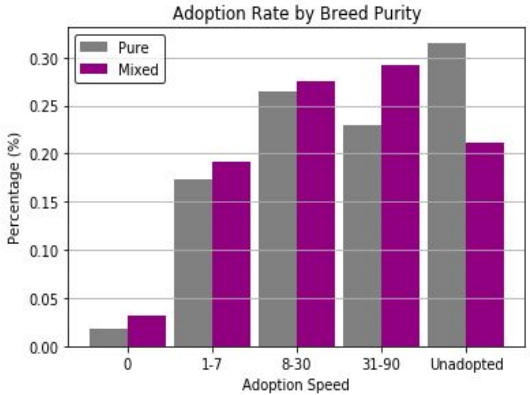


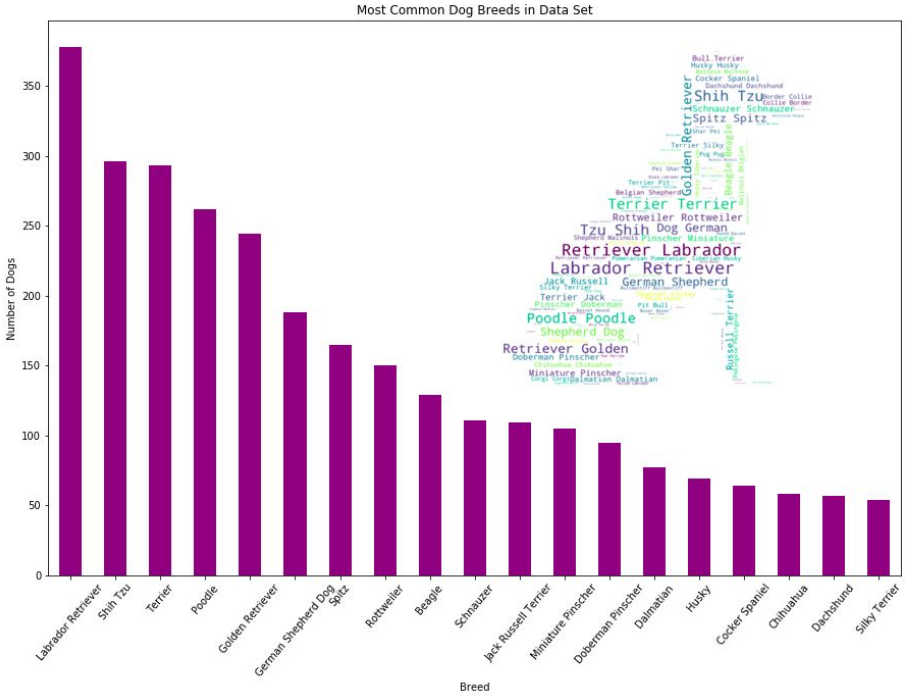




Breed

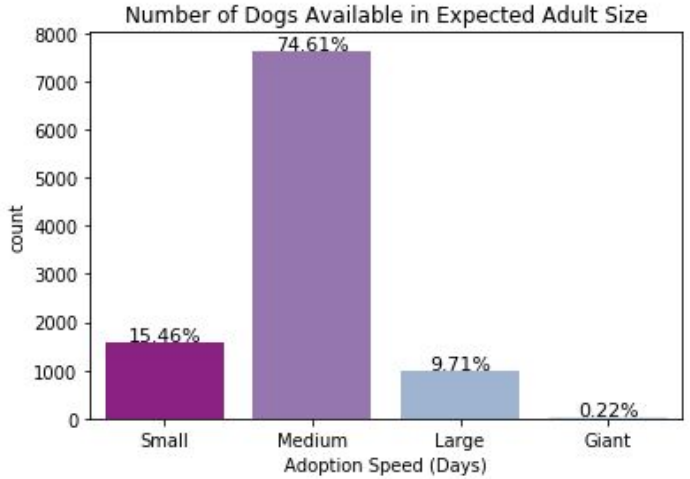
- Mixed breed dogs are adopted more quickly on average
- Labradors, Shih Tzu's, and Terriers are the most common dogs available

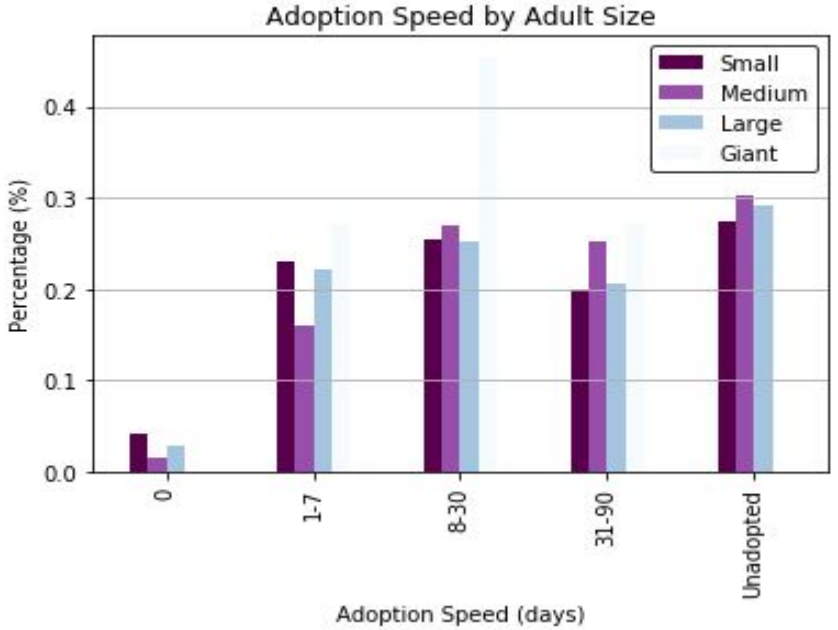




Adult Size

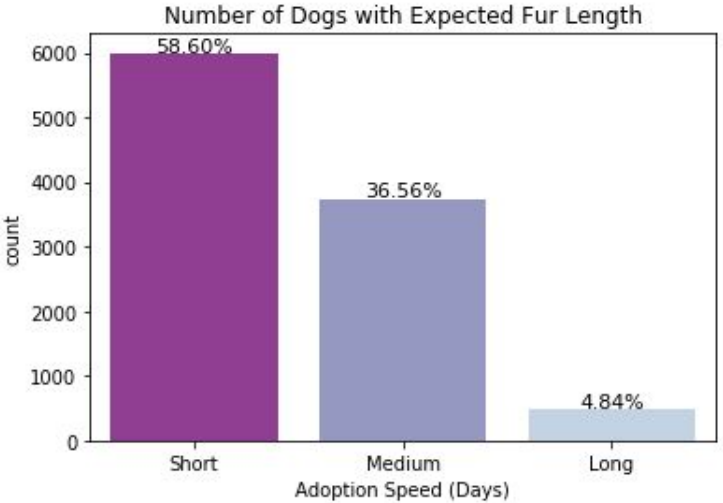
- Only 0.22% of the data set are “Giant” breed animals
- Giant breed animal trends toward fastest adoption - likely due to scarcity

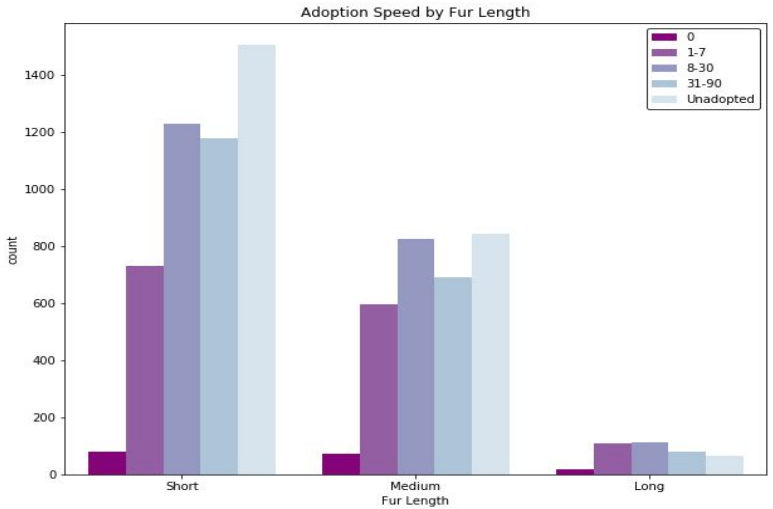




Fur Length

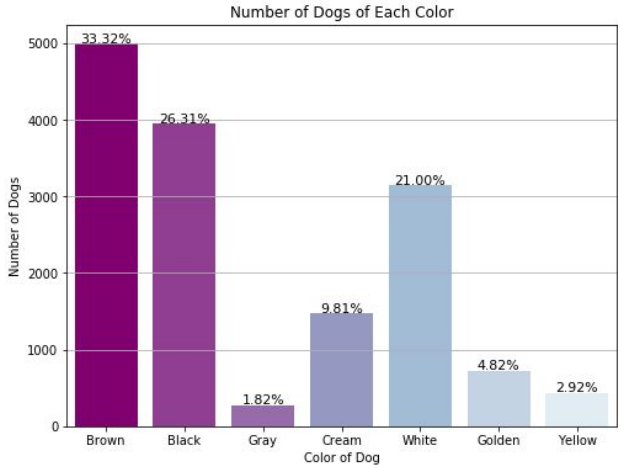
- Dogs in Malaysia trend toward short-haired
- Fur length has little impact on adoption speed

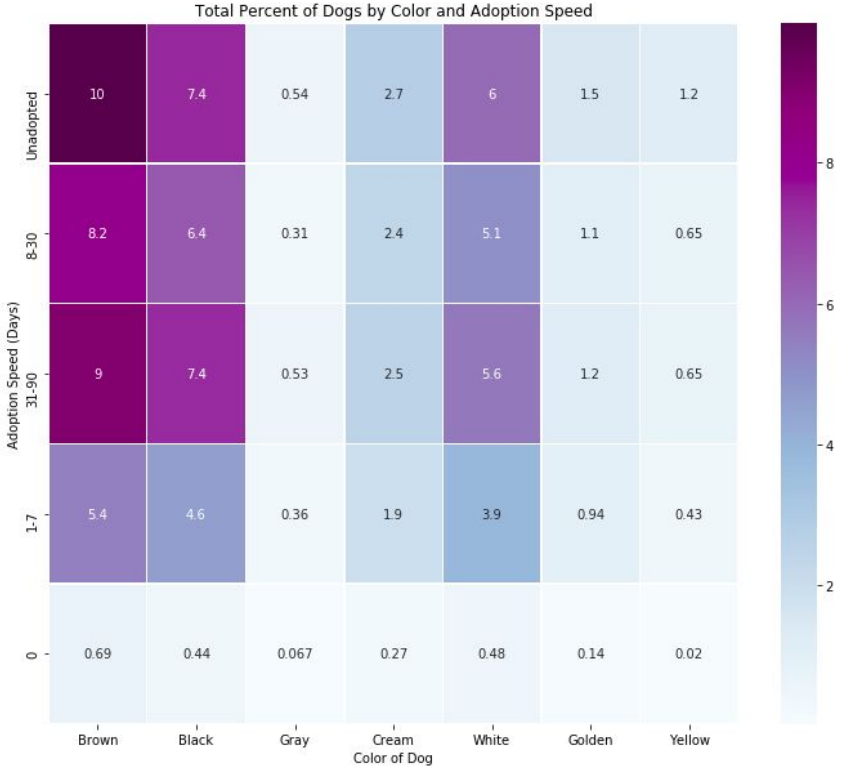




Color

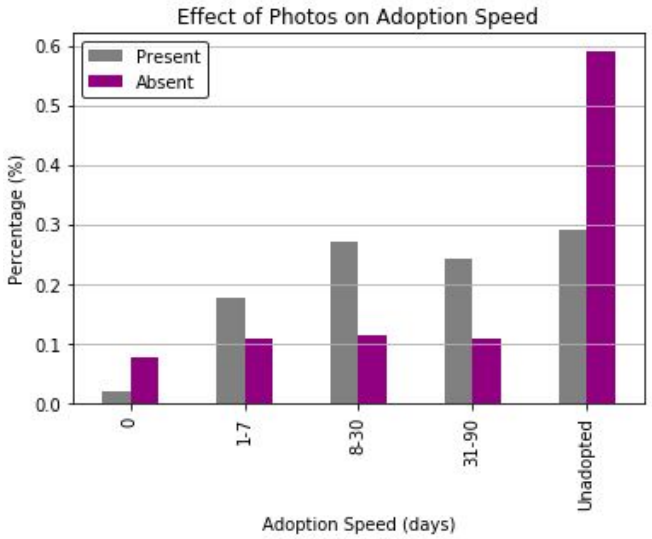
- Darker colored dogs are less likely to be adopted
- Number of colors is not a significant indicator of adoption speed

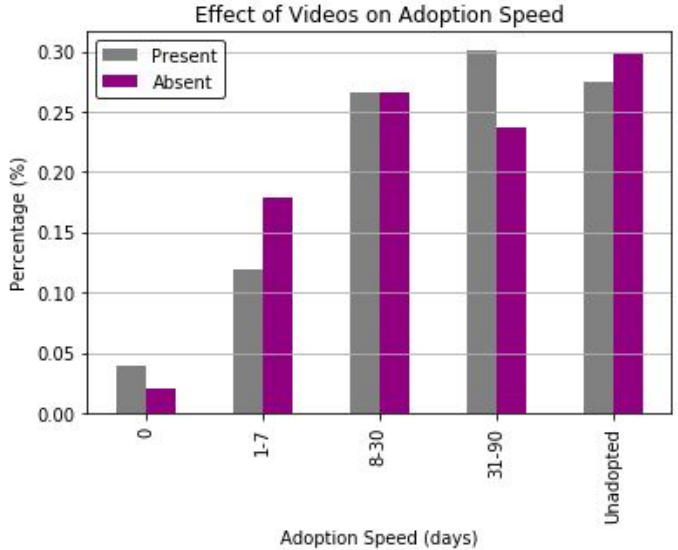




Photos/Videos

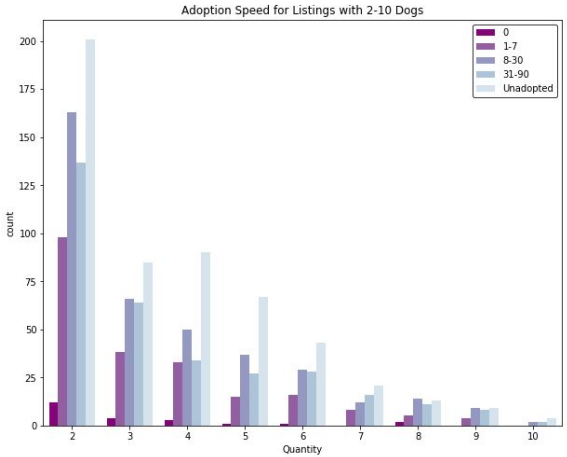
- After >2 photos, there is no benefit to adoption speed
- Videos have very little impact on adoption speed

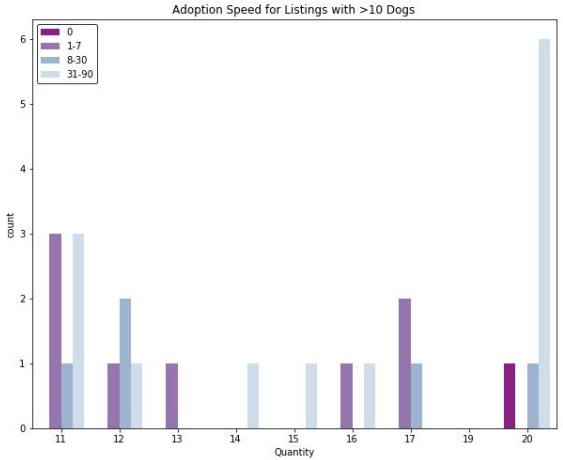




Number of Dogs Per Listing

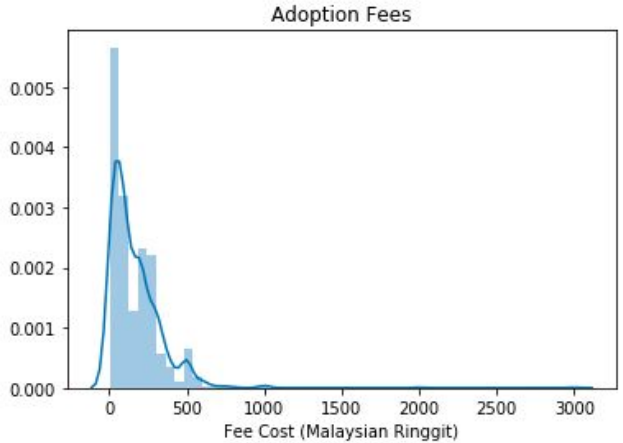
- The greater the # of animals per listing the less likely adoption is
- This is considered to be an artifact:
  - Actual adoption times are obscured because there is no way to record adoption speed for each animal
  - The longest adoption time, if any, is al that is captured

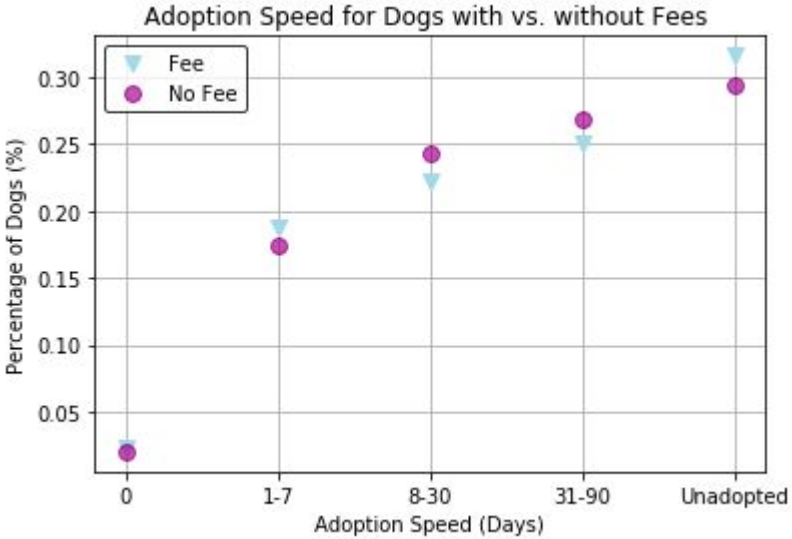




Adoption Fee

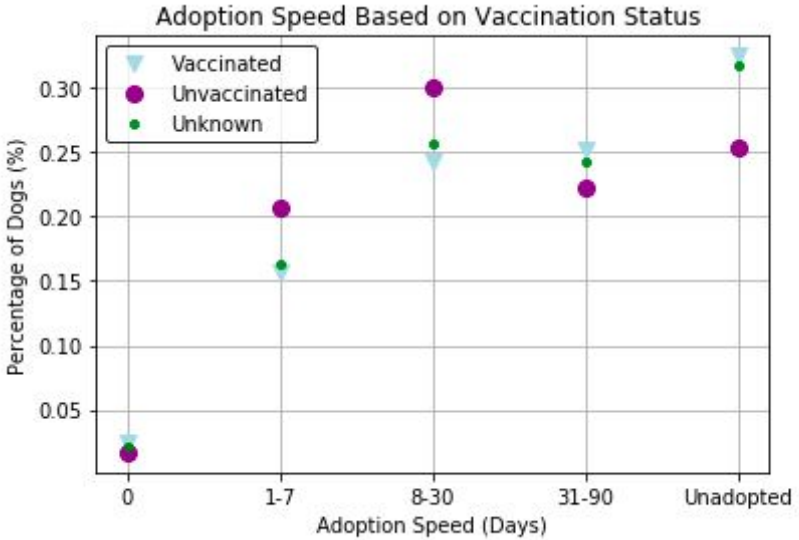
- Adoption fee does have some impact on adoption speed but it is very low
- Correlation of -0.03 showed that dogs with a fee are adopted faster

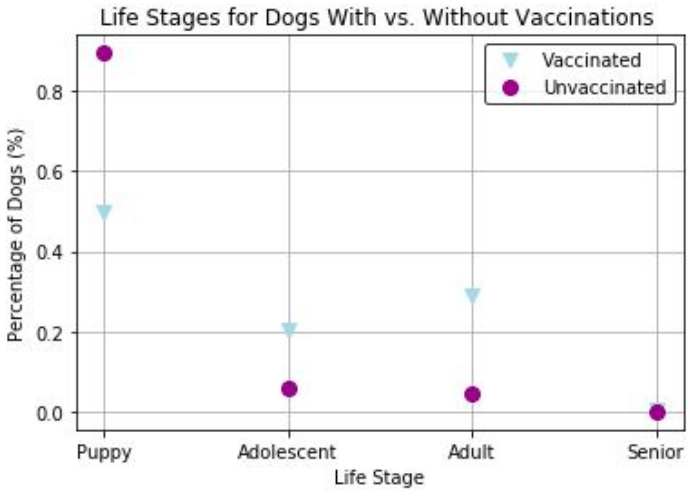




Animal Health: Vaccination

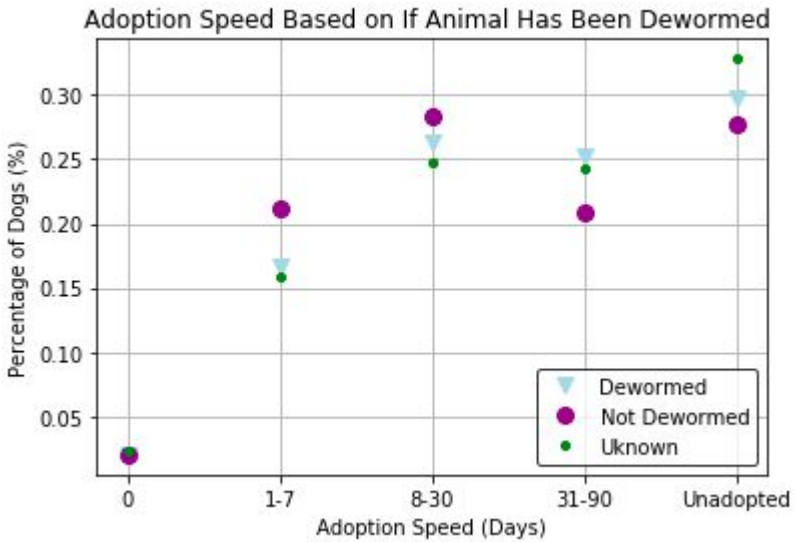
- Vaccination status only matters in dogs that are not puppies
  - Vaccinated dogs are adopted only slightly faster after puppy stage

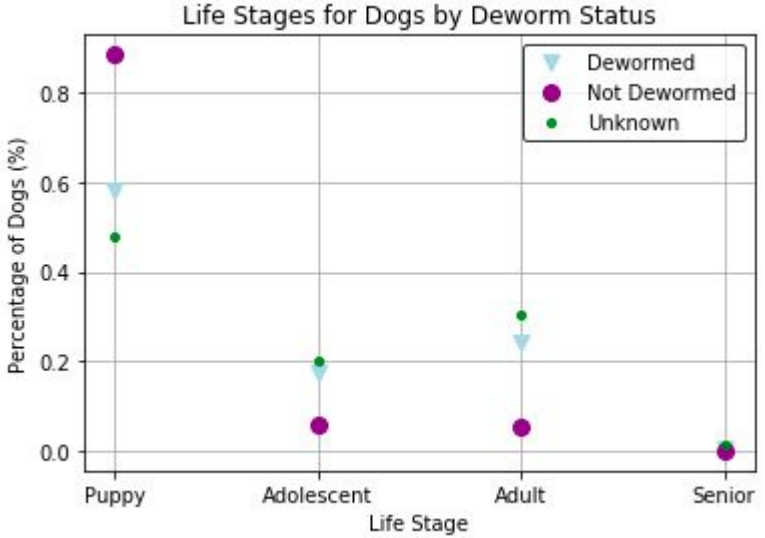




Animal Health: Wormed/Dewormed

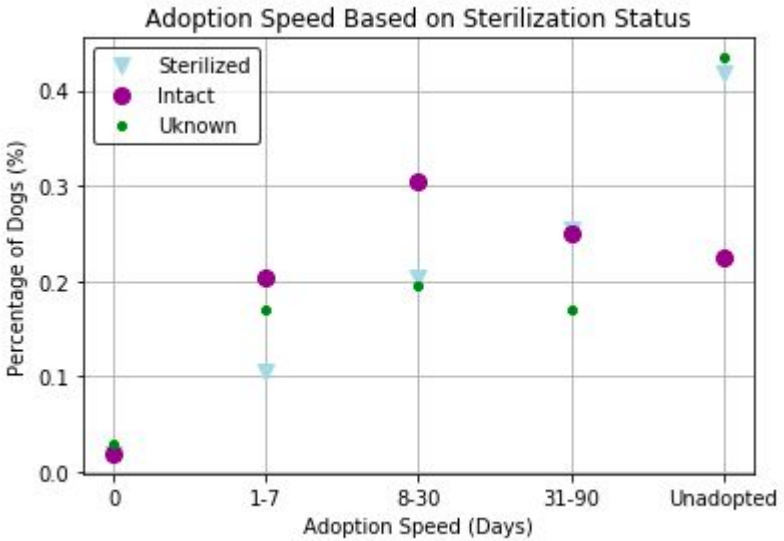
- Animals that are not wormed are adopted fastest…. But again this is due to puppies!
- Minimal difference in adoption speed after removing the puppies





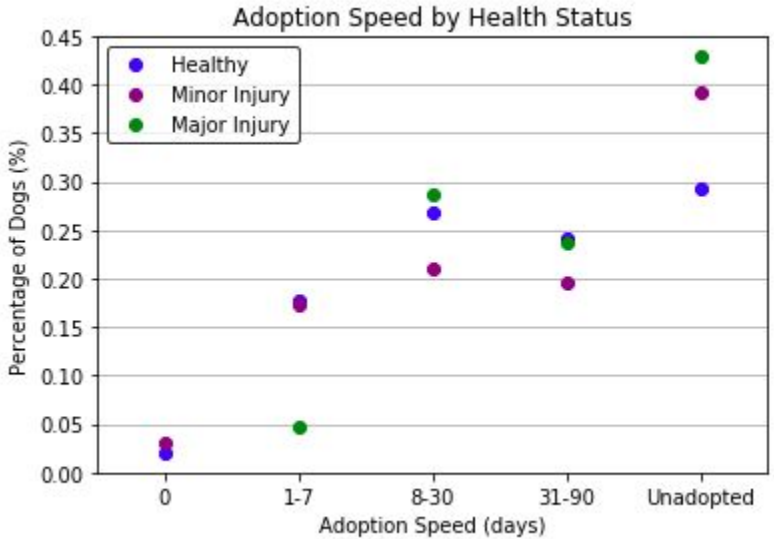
Animal Health: Sterilization

- Intact dogs are adopted fastest
- May be due to desire to breed animals
- Often shelters charge a fee to leave animals intact, it’s possible animals are being sterilized post adoption

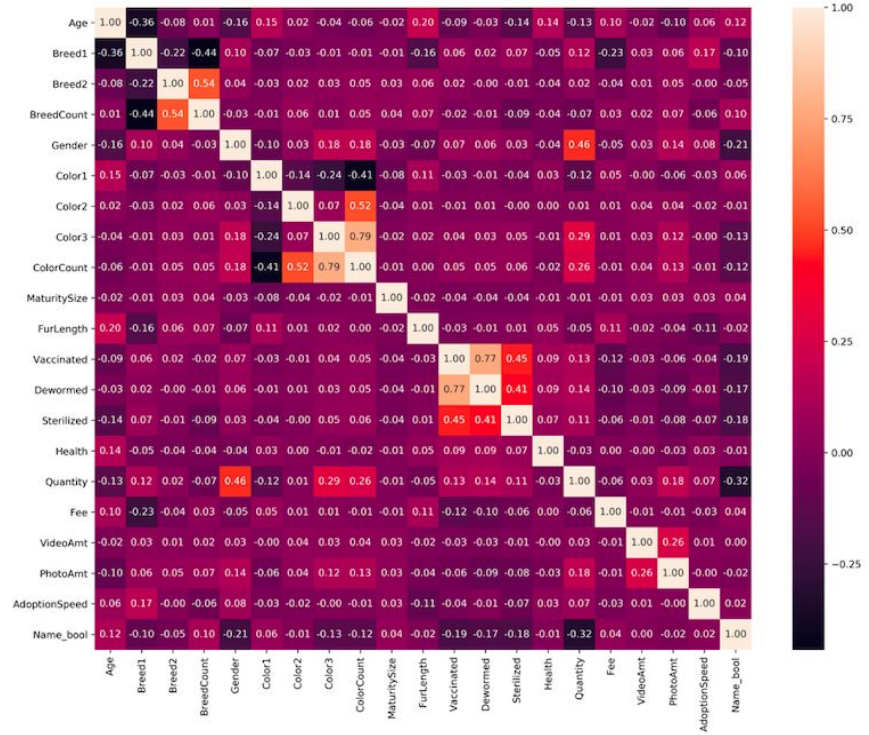


Animal Health: Pre-existing Conditions

- Healthier animals have a better chance of being adopted quickly
- Dogs with major injuries have best adoption rate within the first week
    - May be due to shelter to rescue relocation



All Feature Adoption Speed Correlation



Summary

- Overall the data showed the results expected with a few exceptions.
- Malaysian dog trends are slightly different than those in the U.S.
- Most “at-risk” dogs include those that are:
  - Dark colored
  - Seniors (>9 years old)
  - Dogs with health issues.
- Surprising findings:
  - The insignificance of videos
  - The impact of age on the importance of other features
- Dogs with higher than expected chances of adoption
  - Unsterilized puppies

Thank you!