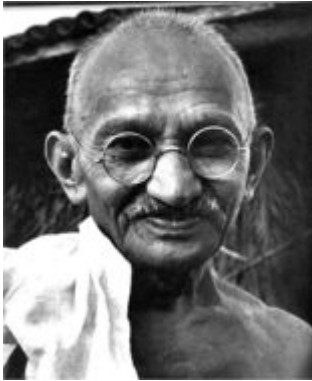# Faster Pet Adoption

Data science application: analysis of the online presentation factors that influence how fast a pet gets adopted from a rescue shelter.

https://www.goodreads.com/quotes/340-the-greatness-of-a-nation-and-its-moral-progress-can

Is an animal's online presence a predictor of its adoptability?

IMG Classification - ML

By classifying animal shelter data using image processing and tracing an animal's dates of arrival and departure from a shelter, it can be decided if we should fail to reject the null hypothesis and determine that pets are too adorable to be left, no matter how they pose for the cameera.
In the following image, it is clear to see the image classification categories, as pets are either:

- Standing
- Sitting
- Laying
- Clothed

It is also
evident in our JSON files that we are able to determine an animal's length of stay.

The length of stay is the key feature in this analysis, barring potential bias given the age of the animal. The suspected age of the animal is also a included in the data set. It is accounted for by grouping using age categories, as deemed necessary.

We can use the IDs given by PetFinder as unique identifiers in the data set to track the animals. As is seen in this JSON file, the URL is associated with said pet ID.

NLP

Aside from image classification, the dataset also contains descriptions of the animals, which can be analyzed using bag of words or other NLP methods.

Minimum Viable Product

Data Set

The minimum viable product for Capstone 1 will be a dataset of images downloaded using RESTful API calls. The src folder contains the stub used to generate the images.

 Web scraping will be necessary to determine if an animal is already adopted, given permission is granted. The web scraping will be done only if the API owners, who grant public access to their API, do not track arrival vs. adoption dates. When a pet is adpoted, their JSON no longer displays the term "status":"adoptable," and instead their page provides a list of other animals that are up for adoption.

Data Collection

To approach the project methodically and to limit bias, a JSON file is printed per month for the last 2-3 years. Each JSON file contains over 6,000 records, depending on how many new pets were available at that time.
The data was quite cumbersome to handle and difficult to find, so there remains much to look forward to in terms of EDA.

Data Structures

For Capstone I, the underlying data structure will be a dictionary of dictionaries, as can be seen in the image below. The JSON data is cumbersome to work with, but once turned into dictionaries, it will prove to be more efficient to work with. The images themselves are rather blurry, which is why the classification algorithm will be rather basic (body position and in dog garments or not in dog garments).

Data Access

The data is gathered using a curl command in the command terminal, using an API issued by PetFinders.

My Motivation

> I am passionate about this project because I am a disabled veteran, and my dog saved my life. Surely more dogs can do more, for more vetearns and more people in general, and we can continue to be proud and grateful to this great nation!