# PetFinder.my - Pawpularity Contest

## Predict the popularity of shelter pet photos

****

**About this Competition**

In this competition, your task is to predict engagement with a pet's profile based on the photograph for that profile. You are also provided with hand-labelled metadata for each photo. The dataset for this competition therefore comprises both images and tabular data.

How Pawpularity Score Is Derived

* The Pawpularity Score is derived from each pet profile's page view statistics at the listing pages, using an algorithm that normalizes the traffic data across different pages, platforms (web & mobile) and various metrics.
* Duplicate clicks, crawler bot accesses and sponsored profiles are excluded from the analysis.

Purpose of Photo Metadata

* We have included optional Photo Metadata, manually labeling each photo for key visual quality and composition parameters.
* These labels are not used for deriving our Pawpularity score, but it may be beneficial for better understanding the content and co-relating them to a photo's attractiveness. Our end goal is to deploy AI solutions that can generate intelligent recommendations (i.e. show a closer frontal pet face, add accessories, increase subject focus, etc) and automatic enhancements (i.e. brightness, contrast) on the photos, so we are hoping to have predictions that are more easily interpretable.
* You may use these labels as you see fit, and optionally build an intermediate / supplementary model to predict the labels from the photos. If your supplementary model is good, we may integrate it into our AI tools as well.
* In our production system, new photos that are dynamically scored will not contain any photo labels. If the Pawpularity prediction model requires photo label scores, we will use an intermediary model to derive such parameters, before feeding them to the final model.

Training Data

* train/ - Folder containing training set photos of the form {id}.jpg, where {id} is a unique Pet Profile ID.
* train.csv - Metadata (described below) for each photo in the training set as well as the target, the photo's Pawpularity score. The Id column gives the photo's unique Pet Profile ID corresponding the photo's file name.

Example Test Data

In addition to the training data, we include some randomly generated example test data to help you author submission code. When your submitted notebook is scored, this example data will be replaced by the actual test data (including the sample submission).

* test/ - Folder containing randomly generated images in a format similar to the training set photos. The actual test data comprises about 6800 pet photos similar to the training set photos.
* test.csv - Randomly generated metadata similar to the training set metadata.
* sample\_submission.csv - A sample submission file in the correct format.

Photo Metadata

The train.csv and test.csv files contain metadata for photos in the training set and test set, respectively. Each pet photo is labeled with the value of 1 (Yes) or 0 (No) for each of the following features:

* Focus - Pet stands out against uncluttered background, not too close / far.
* Eyes - Both eyes are facing front or near-front, with at least 1 eye / pupil decently clear.
* Face - Decently clear face, facing front or near-front.
* Near - Single pet taking up significant portion of photo (roughly over 50% of photo width or height).
* Action - Pet in the middle of an action (e.g., jumping).
* Accessory - Accompanying physical or digital accessory / prop (i.e. toy, digital sticker), excluding collar and leash.
* Group - More than 1 pet in the photo.
* Collage - Digitally-retouched photo (i.e. with digital photo frame, combination of multiple photos).
* Human - Human in the photo.
* Occlusion - Specific undesirable objects blocking part of the pet (i.e. human, cage or fence). Note that not all blocking objects are considered occlusion.
* Info - Custom-added text or labels (i.e. pet name, description).
* Blur - Noticeably out of focus or noisy, especially for the pet’s eyes and face. For Blur entries, “Eyes” column is always set to 0.



This notebook achieved a private rank of 133 (Silver medal) among over 3600 teams.