Twitter Data Analysis Report

Author: Simon Thornewill von Essen
Date: 13.02.2018
Visualisation URL: link

Vis 1: Most Popular breeds and names for dogs

This first gives an idea for the popularity and distribution of dog breeds as predicted by the convolutional neural network and names.

From the visualisation, the three most popular breeds are as follows:

1. Golden Retriever
2. Pembroke
3. Chihuahua

It should be noted that Pembroke are Corgis, which are tremendously popular around the internet in my experience.

The three most popular names are below:

1. Lucy
2. Oliver
3. Cooper

I find these names to to be unusual names for dogs. It should be noted that these names appear at most 9 times out of a population of roughly 700, showing that dogs have very diverse names.

Vis 2: Typical Rating for a dog is roughly 12/10

It should go without saying that these are really good dogs, although there are a couple of dogs who have ratings lower than 10, but they still ratings higher than 5/10, which shows that they are still at least better than average dogs.

There are a couple of ratings lower than 5/10, this is something that could be

investigated further to prevent dogs from receiving low ratings. However these could also be submissions in which dogs were being bad or because someone submitted an iguanadon instead of a dog. (Yes, I've seen it.)

Vis 3: Highest Rated Dogs Appear Most Often

Here we can see a relationship that is pretty standard. The more common a dog is, the higher it is rated. This is due to a couple of reasons such as the "mere exposure effect" which makes people prefer people/items they they have encountered before. Furthermore, people buy and take pictures of dogs that they like and will do so more often, increasing this bias even further.

However, it seems that there was one lucky borzoi that got an outrageous rating, good for him!

Vis 4: Popularity follows a log-distribution

Plotting histograms of Retweets and Favourites shows that most posts only receive between zero and a couple thousand retweets and favorites.

Looking at the axes, we can see that people are more likely to favorite than to retweet. This is likely because retweeting means that a user will share this picture with their followers, this means that they are only more likely to share something that is somehow outstanding or better than a simple "favorite".

When a post reaches a certain critical mass, many people tend to get on board and share this post creating a viral effect which social media marketers often vie for.

Vis 5: Log-transformation of Popularity on a scatter plot creates correlation

This effect can be further shown using this scatter plot in which the x and y axes have been log-transformed creating a linear correlation between retweets and favourites.

Then coloring the points in the scatter plot with the rating numerators of the dogs, we can see that although they are good dogs, they are all generally equally good which is contradictory because there certainly are aspects which make one dog more popular than another.