WeRateDogs Twitter Archive and Image Processing Analysis

For this analysis and visualization, I used data from twitter archives from WeRateDogs, a popular twitter account which posts images and text describing different dogs, along with a rating.

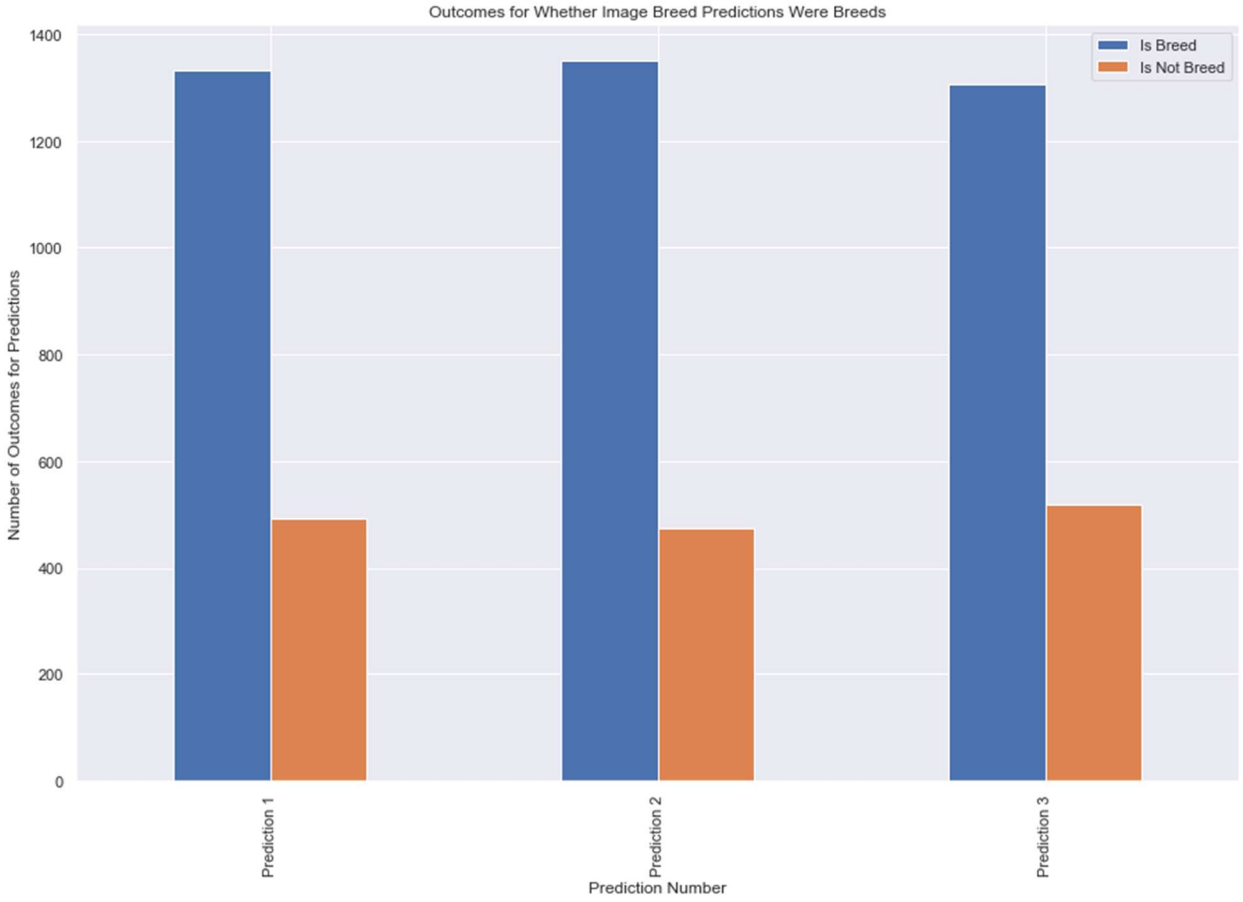
The data included a dataset which used image processing to predict what breeds the dogs were based on the images. The data also included retweet and favorite counts for each tweet in the datasets.

The following insights were gained which analyzing this data:

Insight #1 –

How often were the breed predictions from the image processor correctly predicting a dog breed (instead of something completely different like monitor, envelope, or screen)?

Each tweet had three dog breed predictions per image. From the analysis, I found predictions 1 and 2 correctly predicted a dog breed approximated 73% of time and prediction 3 correctly predicted a dog breed approximately 71% of the time. The graph below depicts the prediction outcomes for predictions 1, 2 and 3.



Insight #2 –

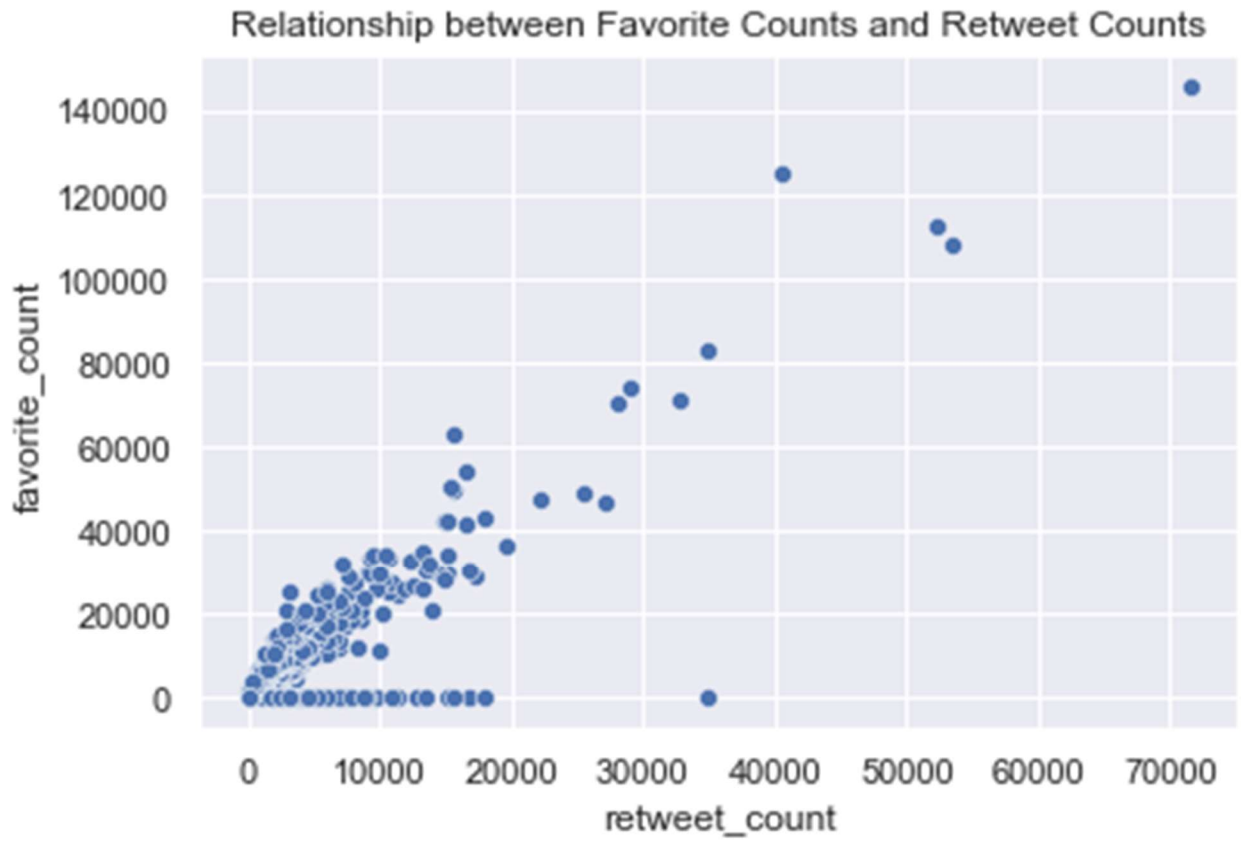
Is there a correlation between the retweet count and favorite count of each tweet?

To answer this, I used the pearsonr method from the scipy library to find the correlation coefficient and the p-value of the retweet count and favorite count. The pearsonr method returns a tuple with (correlation coefficient, p-value). A correlation is considered strong the closer the coefficient is to -1 (perfect negative correlation) or positive 1 (perfect positive correlation).

The p-value tells us whether the correlation is statistically significant. It is considered statistically significant if the p-value is lower than 0.05.

The result from the pearsonr method was (0.8610996769919058, 0.0)which indicates a statistically significant strong positive correlation between the retweet count and the favorite count.

The scatterplot below shows this positive correlation between the retweet count and favorite count of each tweet.



Insight #3 –

What was the #1 predicted dog breed for the highest favorited tweets?

I looked at the top favorite tweets with the breed which prediction 1 from the image processer provided. Here is the dataframe with the information:

	tweet_id	p1	favorite_count
824	744234799360020481	labrador_retriever	146184
311	822872901745569793	lakeland_terrier	125305
412	807106840509214720	chihuahua	112816
107	866450705531457537	french_bulldog	109948
859	739238157791694849	eskimo_dog	108200

Here we see the breeds predicted from the top five favorited tweets were Labrador Retriever, Lakeland Terrier, Chihuahua, French Bulldog, and Eskimo Dog.

Conclusion:

These are a small number of insights gained from the dataset. There are several variables which still have not been analyzed deeper. I look forward to spending more time with the data to find additional insights from the WeRateDogs twitter archives.