

act_report

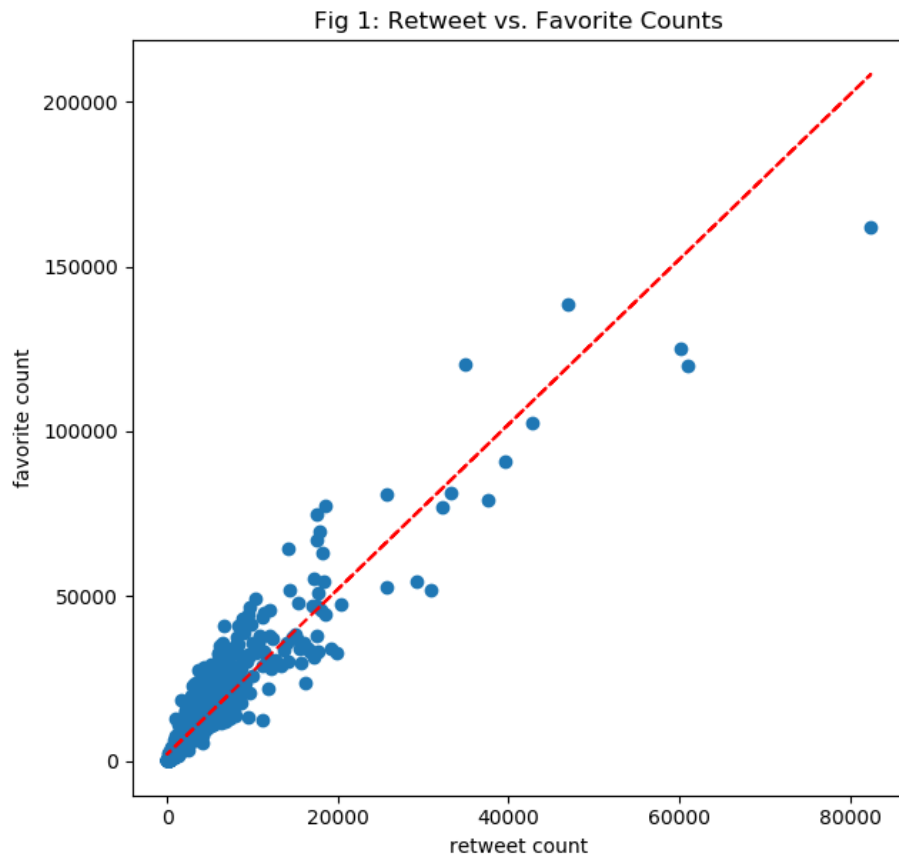
June 5, 2019

1 Project: Wrangle and Analyze Data

This is a summary of analysis and visualizations of Wrangle and Analyze data project. The methods, codes and more details for these analysis can be found in Jupyter Notebook: `wrangle_act.ipynb`.

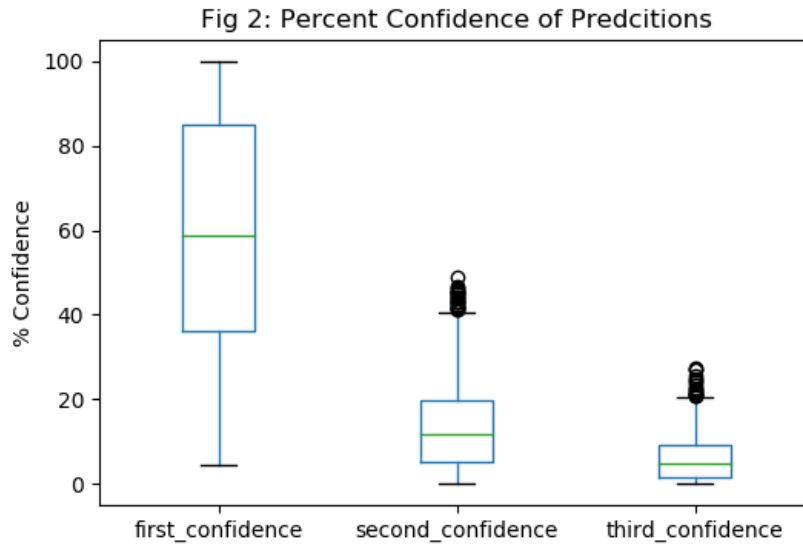
1.0.1 Is there a correlation between retweet and favorite counts?

Couple of interesting details about these dog tweets are retweets counts and favorite counts which can be found in `df_tam` (a data frame from `twitter_archive_master.csv` file). According to my analysis these two have a strong positive correlation with a correlation coefficient of **0.93**. A scatter plot for this relationship is shown below. However neither of these two features have significant correlation with dog ratings.



1.0.2 How confident are the image predictions?

In the table `df_tam`, there are three predictions for each twitter image. And each of these predictions have a percent confidence of the prediction and also whether the prediction is true or false. As can be seen from the box plot below, the first prediction of the images have higher confidence than the second and third predictions. The percent confidences of the first predictions ranged between 4% - 100% with an average and a standard deviation of 59% and 27% respectively. The confidence range for second predictions is 0% - 49% with an average of 13.5%. The confidence range for third predictions is 0% - 27% with an average of 6%. The second and third predictions are not as confident as first predictions according to these data.



1.0.3 What are the average confidences for true and false predictions of each predictions?

For the first prediction, true predictions have an average confidence of 61.3% while for false prediction, it is 53.6%. For the second prediction, true predictions have an average confidence of 11.7% while for false prediction, it is 14.1%. For the third prediction, true predictions have an average confidence of 5.7% while for false prediction, it is 6.2%.

1.0.4 What percentage of each prediction is being true or false?

73.98% of first prediction are true and 26.02% are false. 74.92% of first prediction are true and 25.08% are false. 72.52% of first prediction are true and 27.48% are false.

1.0.5 How many of second predictions with a percent confidence over 40% are true?

There are total of 31 predictions with above 40% confidence. Out of this 26 are true predictions.

1.0.6 How many of third predictions with a percent confidence over 20% are true?

There are total of 23 predictions with above 40% confidence. Out of this 16 are true predictions.