act_report

August 28, 2022

0.1 Report: act_report

This project involved wrangling and analyzing Twitter tweet archive of user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent.". This archive contains basic tweet data for all 5000+ of their tweets as they stood on August 1, 2017.

The datasets used in this project are 3 parts: 1. The first is the main Tweeter tweet archive 2. The second is the Tweet Image Prediction 3. An addtional dataset that contained features like retweet count and favorite count

Some analysis were carried out on the merged datasets (saved as twitter_archive_master.csv) and some of the findings were quite intriguing. Also, a few visualizations were made which revealed some interesting patterns and findings. It is important to note that the dataset has not been 100% cleaned, and as such the results may vary, subjective to further wrangling. 1. Firstly, analysis was carried out to see top 10 do breeds having highest rating. The result showed Clumber on the top of the list and Gordon_setter at the 10th position. A study on Clumber breed showed.... Also, other breed that followed in the top 10 include Bouvier_des_Flandres, Saluki, Pomeranian, briard, Tibetan_mastiff, Border_terrier, standard_schnauzer, silky_terrier. It was noted that Clumber stood out with a very high difference of 27, while the next, Bouvier_des_Flandres has a rating of 13. The rest have an approximate rating of 12 and Gordon_setter has 11.7. It shows most of the top 10 breeds have almost similar rating 2. Another analysis was done to see top 10 dog breed with highest favorite count.The result show Saluki coming first and Border_terrier coming 10th. Also some breeds having high rating also appearing as having high favorites. This implies there might be a level of correlation between these columns. 3. Analysis was also done to see which year has the heighest number of retweets. The analysis showed the following:

2017: 5652.979798 2016: 2761.523321 2015 : 1101.363504

The high number of retweets in 2017 is highly suggestive of increased number of Twitter users as at 2017.

Visualizations (seaborn heatmap) show that retweet_count and favorite_count have strong correlations of 0.7 which would be very useful in feature selection for modeling.

More visualization also show that there are more dogs in pupper stage while floofer were the least.

In [ ]: