

# WeRateDogs Analysis Report

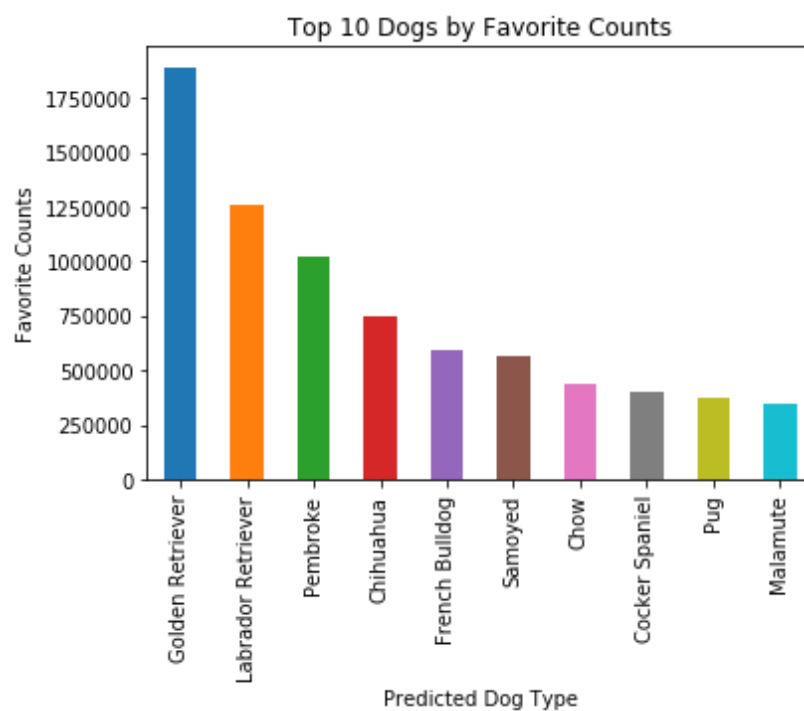
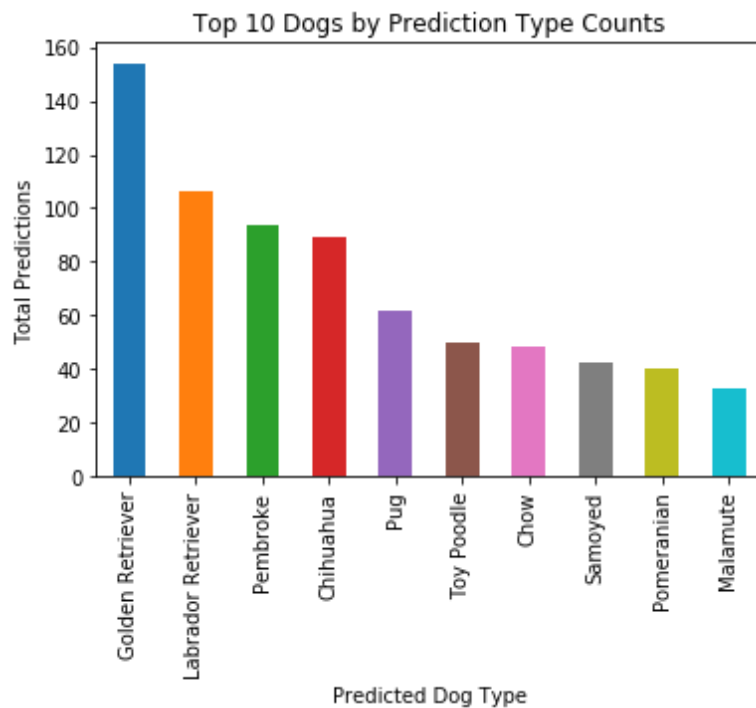
This document contains my insights into the WeRateDogs data that was gathered, cleaned, and assessed.

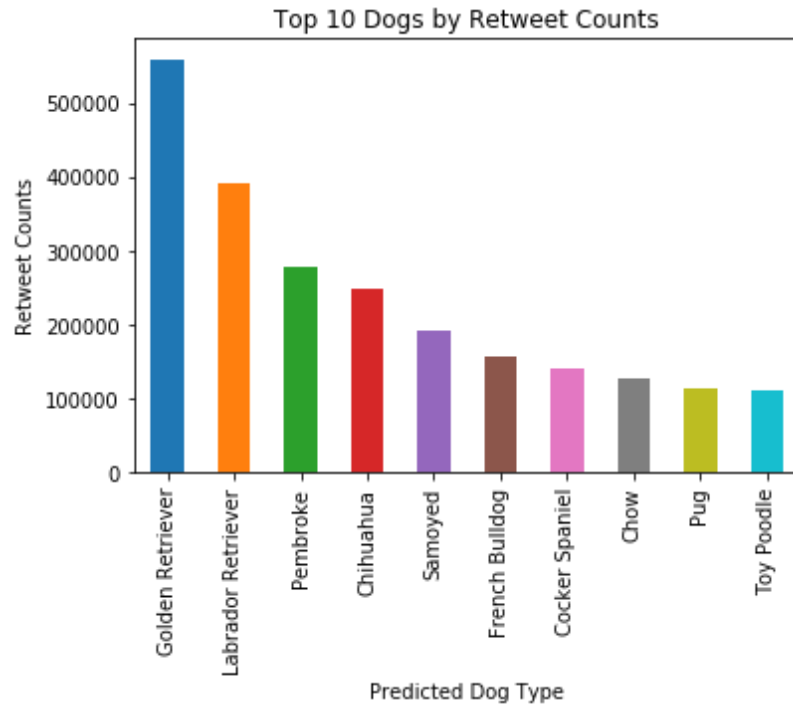
WeRateDogs is a Twitter account setup by Matt Nelson to rate people's dogs. People post pictures of their dogs with a funny comment for others to rate. The ratings follow a scale of 1 to 10, however most of the ratings are typically over 10. Over the years the site has grown in popularity and achieved international media coverage. [\[1\]](#)

What type of dogs are the most popular on the site? We can break this into three different categories: type of dog, most favorited, and most retweeted.

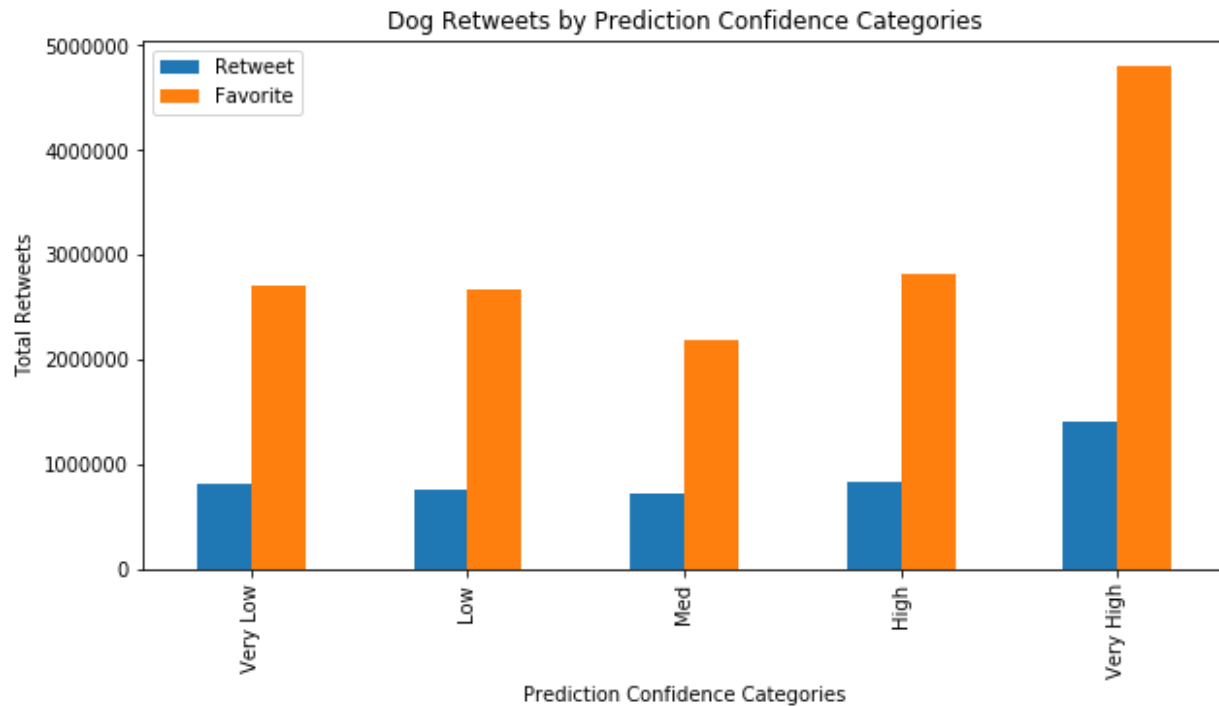
It is clear by the charts below that the winner for all three categories by a dog type prediction algorithm are: the Golden Retriever followed by the Labrador Retriever, and Pembroke.

Dog Breed	Retweet Count	Favorite Count	Dog Prediction Count
Golden Retriever	561,591	1,896,894	155
Labrador Retriever	392,047	1,255,976	106
Pembroke	278,983	1,019,732	94





With any prediction algorithm there is a confidence level for the predicted results. We wanted to know if the confidence level had any relation to the number of retweets or favorites of the dogs? In the below chart we can see the results broken up into 5 confidence levels. Right away we can see that highest confidence category has a larger number of retweets (1,410,919) and favorites (4,827,354). Looking at it further we can see some discrepancies with the other categories. The lowest counts belong to the middle category (724,632 - 2,191,100) while the lowest two categories have higher counts (812,402 - 2,702,878 and 756,040 - 2,668,863). This should be investigated further before drawing any conclusions.



Categories	Retweet	Favorite
Very Low	812,402	2,702,878
Low	756,040	2,668,863
Med	724,632	2,191,100
High	843,549	2,842,052
Very High	1,410,919	4,827,354

Over time the user community has evolved the types of pictures they take of their dogs. Would that change show up in the results from the prediction algorithm? Looking at the chart below we can see the average confidence level over time. No trends immediately jump out. The overall average is 3.225. Without further investigation we do not know this has any impact overall.

