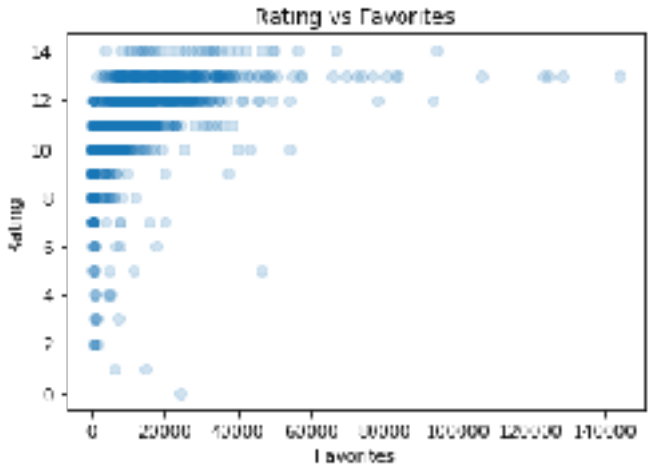
Act Report

Case Study Insights

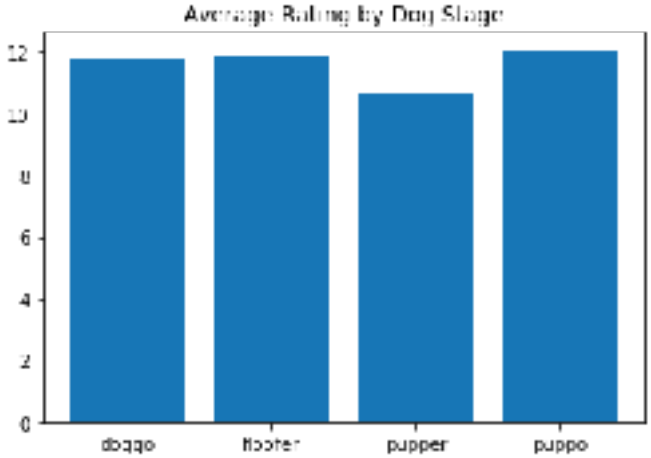
Once the dataset was gathered, assessed, cleaned, and merged, then I could make some investigations into factors driving scores and ratings for dog photos. There were over 30 fields of data, but to keep things simple, I considered the rating score, tweets favorited and retweeted, type of dog, and time of tweet. I wanted to see which known factors demonstrate a correlation to higher ratings. I and then analyzed the effectiveness of our image prediction algorithm.

What factors impacted or helped predict the rating score?

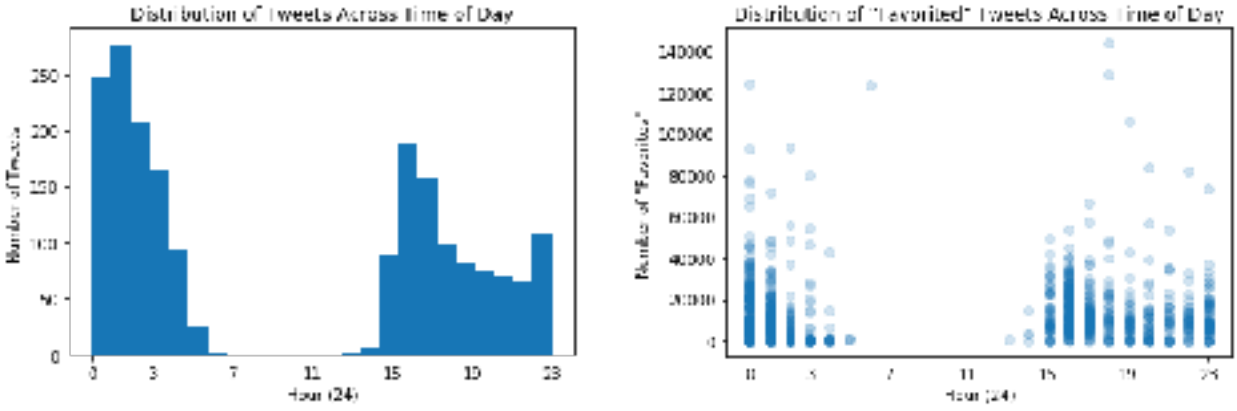
favorites Not surprisingly, tweets that were favorited and retreated more often were more likely to indicate a higher rating.



dog stage Although and doggo, floofer, puppo all had similar mean ratings, the average rating for pupper was notably lower.

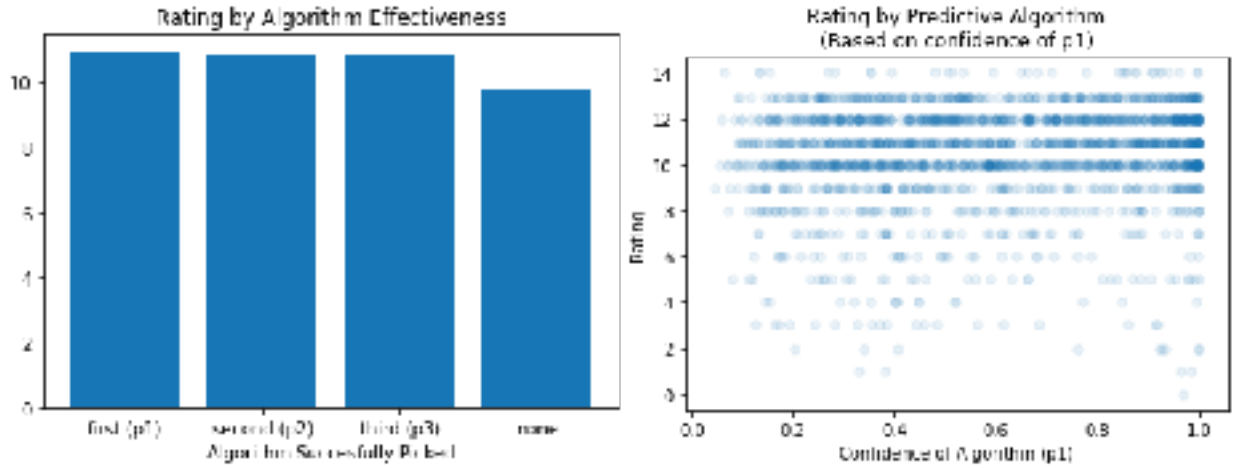


time of day Although there seemed to be more tweets during the hours of 0-5am (unsure of the timezone), there appears to be more favorited tweets during the evening hours.



Does the predictive algorithm help predict the rating?

Before we discuss how accurate the predictive algorithm is, perhaps it can actually have a correlation with the rating of the photo. For example, if the algorithm picked the right dog type in the first guess, does that photo have a likelihood of a higher rating? Similarly, if the algorithm could not determine the dog after 3 attempts, is that photo likely to have a lower rating?



In fact, when analyzing the algorithm's confidence (p1), we do see a correlation with the photo rating. In general, we see a direct relationship between the confidence and the rating; as the confidence increases, so does the rating. And when the algorithm does not make a successful pick, the mean rating is notable lower.

How effective is the predictive algorithm in determining the dog type?

By analyzing the algorithm's confidence for the first guess (p1), we can see that that the algorithm confidence correctly reflects the accuracy of the guess. As the confidence increases, the accuracy percentage would progressively increase at a higher rate.