

EDA report

1. Insights given

- Most identified dog breed by the classifier.
- The predicted dog breed with the highest average rating.
- The overall trend of posts on the WeRateDogs page.

2. Thoughts

-I found this part very enjoyable, since my data was clean and organised. This made it easy to navigate and rearrange the columns to get the data I needed.

-For assessing the number of posts per month, I first created a new row containing the year and month of the tweet ('my'). To do this, I had to iterate through the rows of the DataFrame, extract the timestamp and split it to only get the year and month values for each tweet. I then grouped by 'my' and summed the rating denominator. Since all the rating denominators are 10, I could then infer the number of tweets posted every month by dividing the summed denominator by 10.

3. Results

- Several interesting results were found from this exploratory data analysis.
- The mostly identified dog breed was none other than the Golden Retriever!
We can infer that it is one of the most popular dog breeds for people to have as pets.



A cute golden retriever.

- The dog breed which received the highest ratings was the Clumber breed:



- Overall, tweets on the WeRateDogs pages have been steadily increasing since November 2015 (around 150-200 posts a month), but the numbers began to skyrocket in March 2017, where the number of tweets per month reached numbers of up to 700 around July 2017. It's safe to say that this page has gotten a lot more popular over time.