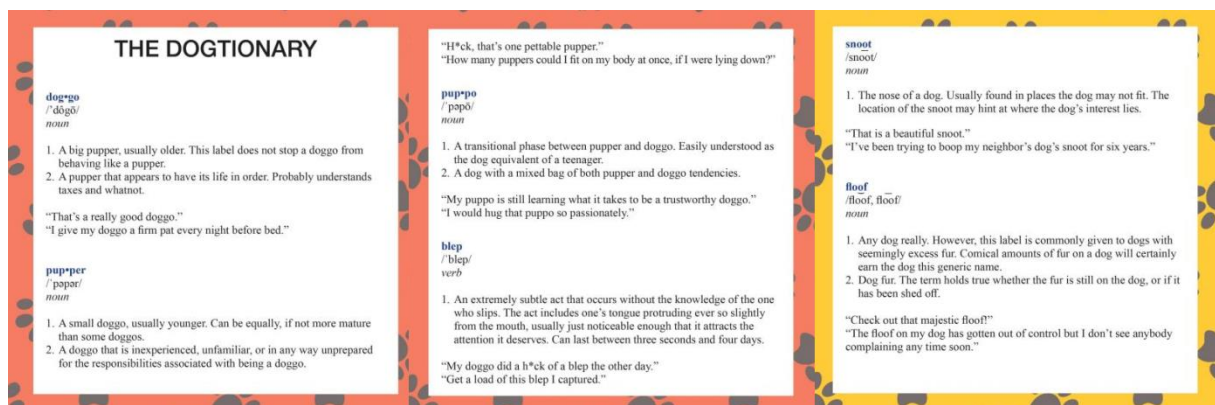# Data Wrangling – Final Project - Act Report

We were very fortunate and got our hands on all the tweets from WeRateDogs until August 2017! And off course we dig into the data to give you some insights and cool stuff. If you haven't hear about WeRateDogs, then let me tell you that you have missed something – unless you do not like dogs, in which case, I am sorry.

The WeRateDogs community tweet the cutest and funniest photos of dogs in the internet.

And because all of the dogs in this picutres are good boys, they all deserve great ratings. Mostly ratings are out of 10 points but sometimes it gets crazy! Finally, let me tell you about the dog stages. According to the "Dogtionary" there are four stages of a dog: the doggo, the pupper, the puppo and the floofer. Feel free to read more about these in the following screenshot of the Dogtionary:

Now that you know more about the dogs and this awesome community, let's talk about data. We got all the information we needed from twitter, but we also let our prediction algorithm tell us what race of dog is in the pictures. We were interested in three questions:
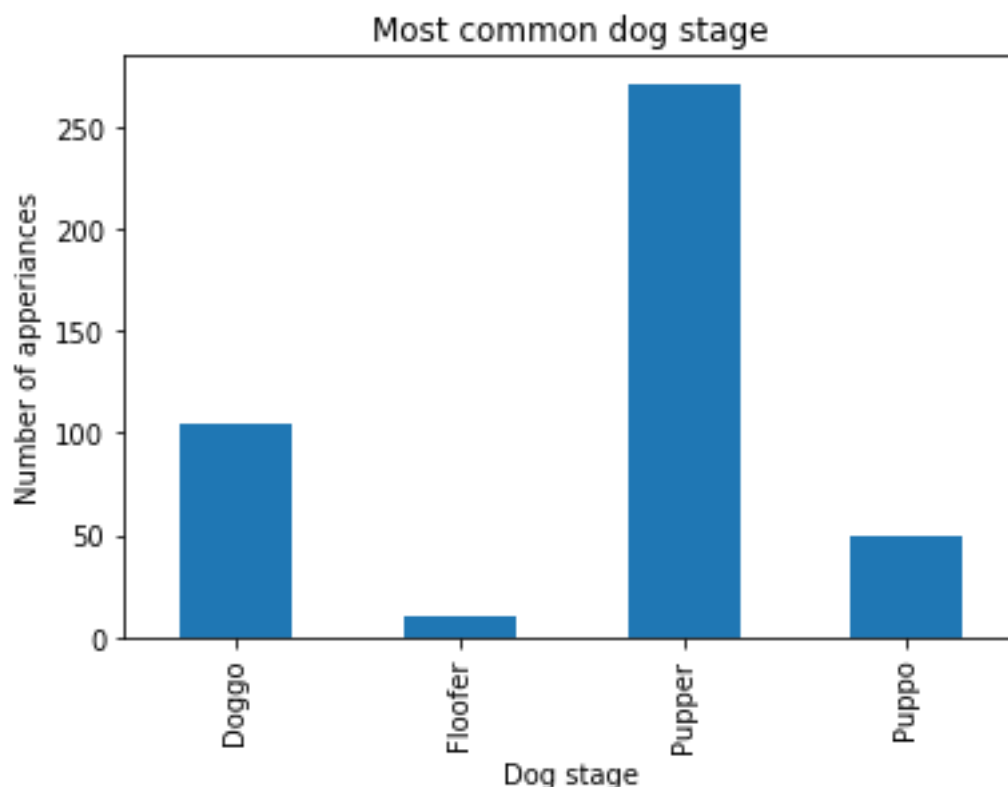
1. What is the most common race according to our prediction model?
2. What is the most common stage of a dog?
3. How are the number of retweets affected by the given rating and stage of a dog?

What is the most common race according to our prediction model?

It is the Golden Retriever! Did you expect that? It is followed by Labradors, Pembrokes, and Chihuahuas.

What is the most common stage of a dog?

It is the Pupper!! How about the others?

We have to say that unfortunately most of the tweets did not include this information, so we do not think this information is representative. However, it is fun to see that from those tweets with a dog stage, the Pupper was the most common of them all!

How are the number of retweets affected by the given rating and stage of a dog?

It turns out that the rating the person gave to the dog was positive correlated with the number of retweets. However, do not start immediately giving your tweets the highest ratings… this could only explain around 10% of the variation of the number of retweets.

But we did not stop here! We were curious and look also if the race of the dog was a good predictor for the number of retweets. Unfortunately, our linear model showed, that the race of the dog had no effect in the number of retweets a tweet gets.

Thank you!