

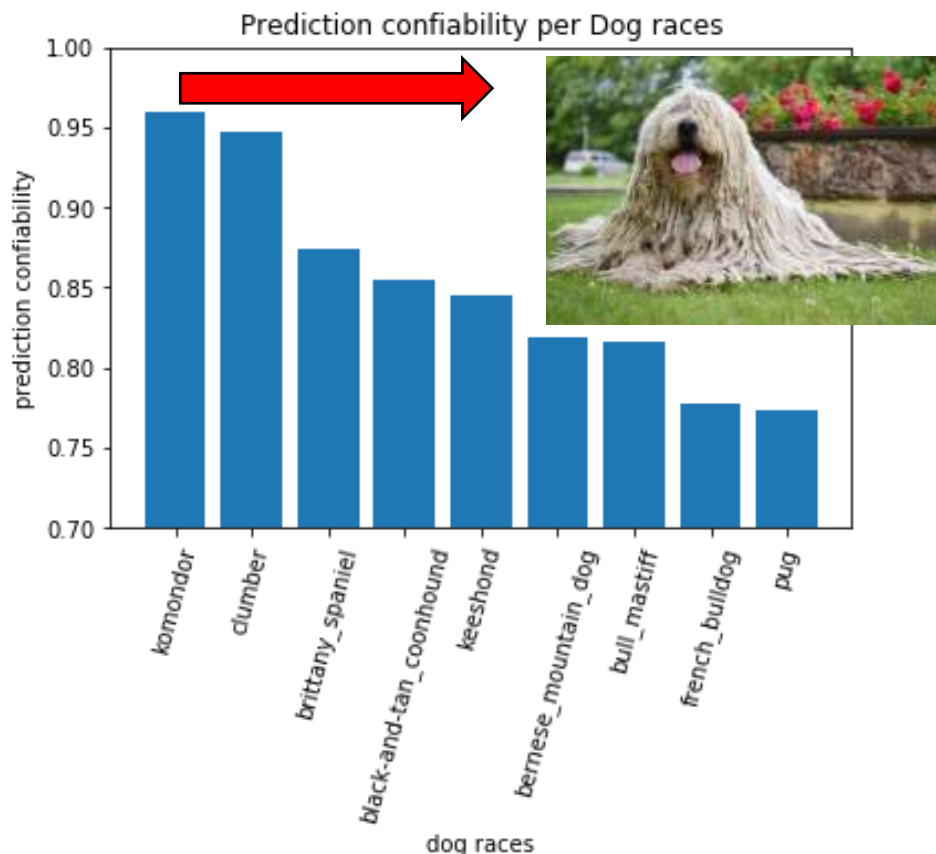
ACT_REPORT: Data Wrangling (ND111)

By Vítor Almeida

After the data wrangling process, three visualizations were performed:

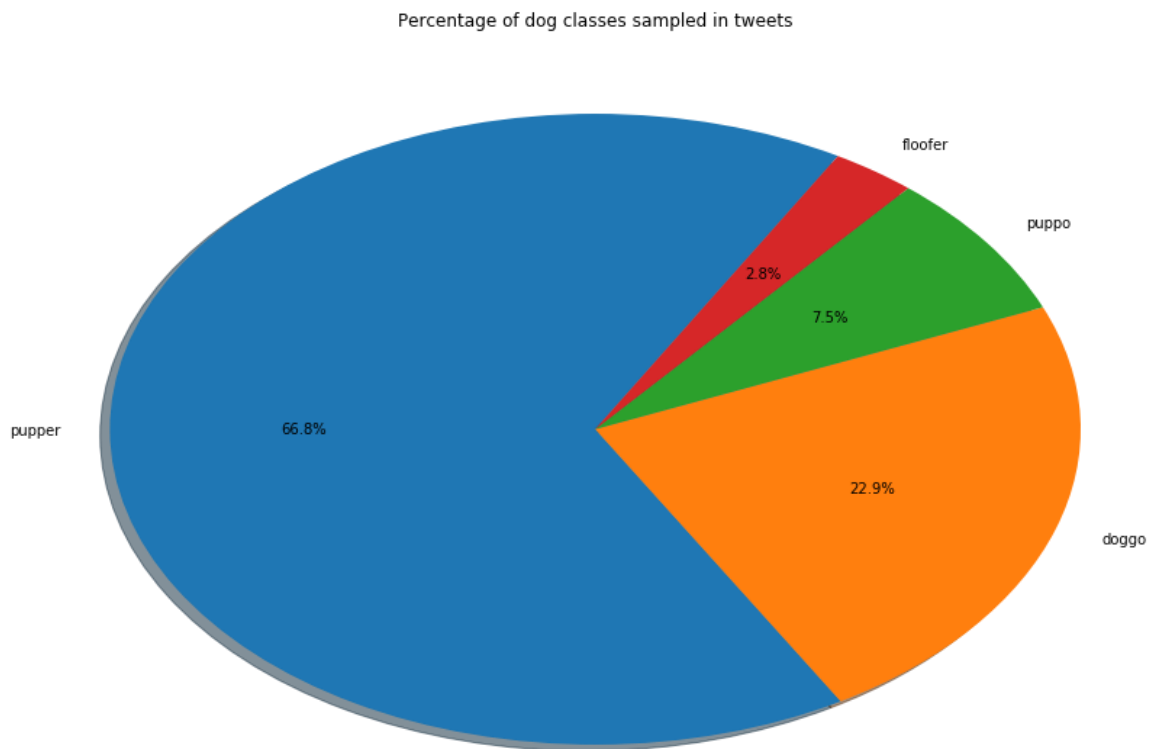
- ➔ Prediction confiability per dog race
- ➔ Percentage of dog stages represented in tweets
- ➔ Retweet and favorite counts per month

Among all 111 dog races, only 9 (about 8%) were predicted with more than 75% of confiability by the algorithm. The highest confiability achieved came from *komodor* race, which is reasonable – the race has a very prominent visual pattern.



Additionally, the most well represented class in the tweets is the **pupper** (66.8%) dogs, followed by the **doggo** ones (22.9%), beign the latter a sequel of the

former, as points the “dogtionary”. Data from transitional stages (e.g., **doggofloofer**) were omitted from pie plot (<1%). Since nowadays many people live in small apartments, the pupper community must have been favoured as the main kind of dog pet. Or it just because they are potentially more cute, who knows. Additionally, it appears that people are avoiding the furred ones.



Finally, during the period between October 2015 and August 2017, **both retweet and favorite counts have increased**, what is to be expected as WeRateDogs popularity grown. Nevertheless, two findings were observed. The favorite counts increased in a greater fashion than did the retweets - its curve have a greater slope. In the other hand, mean of retweets counts stabilized next to 5000 per month, after January 2017.

It appears that most of the followers and general people prefer to favorite and receive more from the tweets, rather than share it. This could be a possible outcome for most of the twitter accounts that gets popular over time.

Another finding was **a peak of both retweet and favorite counts** around June 2016. Some code were made to evaluate what was so special about these tweets. Among the 5 dog races best represented in tweets from June 2016, the first two are tied having both 5 tweets (*shetland_sheepdog* and *labrador_retriever*).

The most retweeted (82209) and favorited (161805) tweet coming from this period was about a *labrador_retriever*. This dog race is one of the most docile and friendly. It is hard to simple avoid a tweet picture about it. As the data from the "peak event" shows, WeRateDogs should invest in more labrador retrievers if popularity turns bad.

