Project 2 : Wrangle and analyze data
Reports
and Visualizations

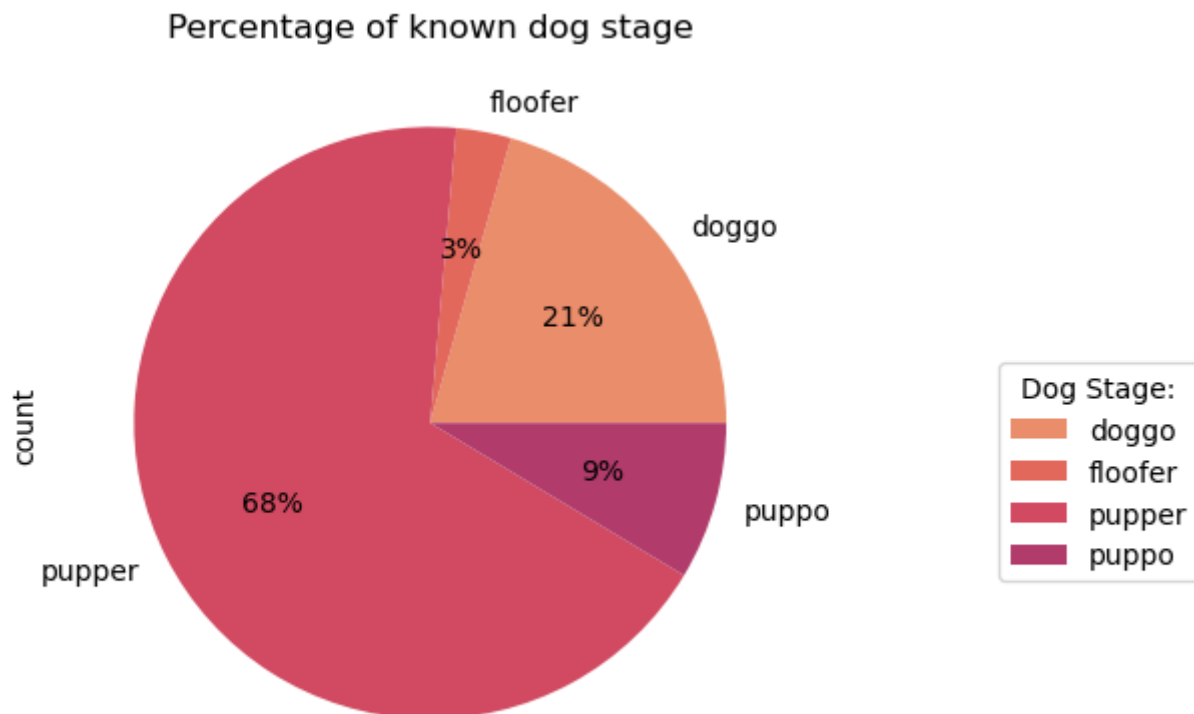# Udacity Data Analyst Nanodegree

Mame Diarra NIANE

Introduction

WeRatesDogs is a twitter account that rates people's dogs with humorous comments about the dog.

After gathering,assessing and cleaning the data ,the next step is to do insights and visualization. i ask myself 4 questions that I'm about to respond to.

- 1.What is the repartition of the type of dog in the dataset

- 2.what is the most retweeted tweet

- 3.What is the most liked tweet (based on the favorite count)

- 4.What is the total number of tweets over the months

1.What is the repartition of the type of dog in the dataset ?

Since there are a total of 4 dog stages . I used a pie chart to display the percentage of each dog stage in the dataset . Note that most of the tweets do not provide the stage of the dog . so the pie chart is about the dog stages that are present in the dataset .
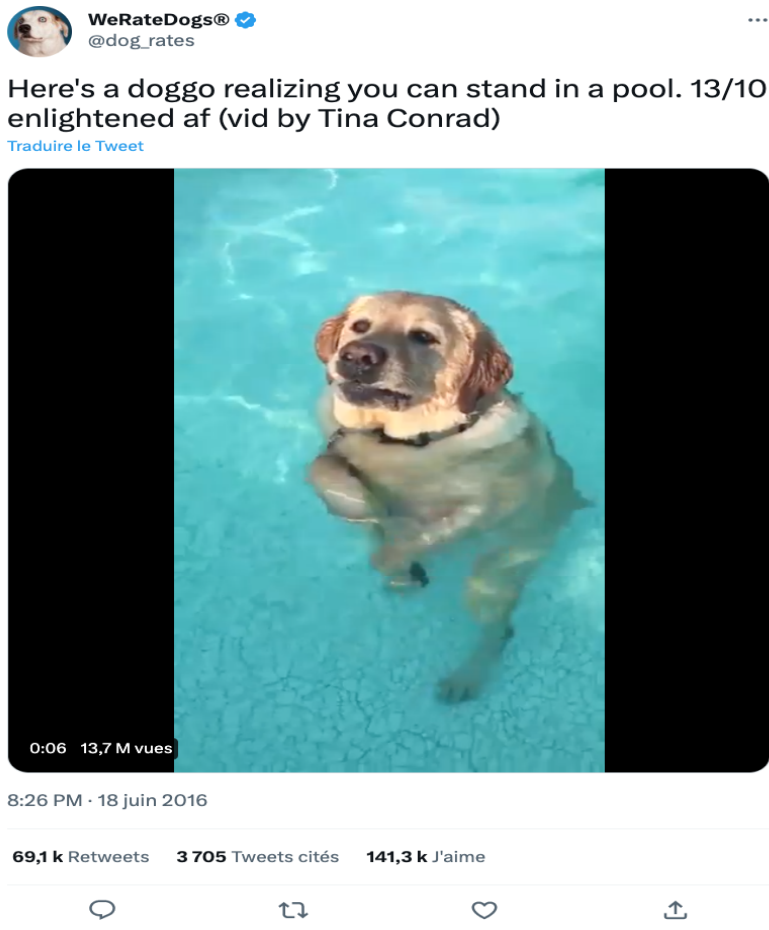

Percentage of known dog stage

We can see that among the known dog stages pupper is the most present with 68% followed by the doggo with 21% and puppo with 9% and and the lowest rated dog stage on the WeRateDogs twitter account is the floofer about 3%.

2.what is the most retweeted tweet ?

tweet_id	timestamp	source	text	rating_numerator	rating_denominator	name	dog_stage	retweet_count	favorite_count
744234799360020481	2016-06-18 18:26:18	Twitter for iPhone	Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad) https://t.co/7wE9LTEXC4	13.0	10	None	doggo	69163	141325

I did a sorting with the sort_values function to have the most retweeted tweet of all the dataset .

We can see that the most retweeted tweet is from June 2016 and has 69163 retweets. The dog has a rating of 13/10.

https://twitter.com/dog_rates/status/744234799360020481?s=20&t=SFtjjsBir-O7Xk_PboEuuw

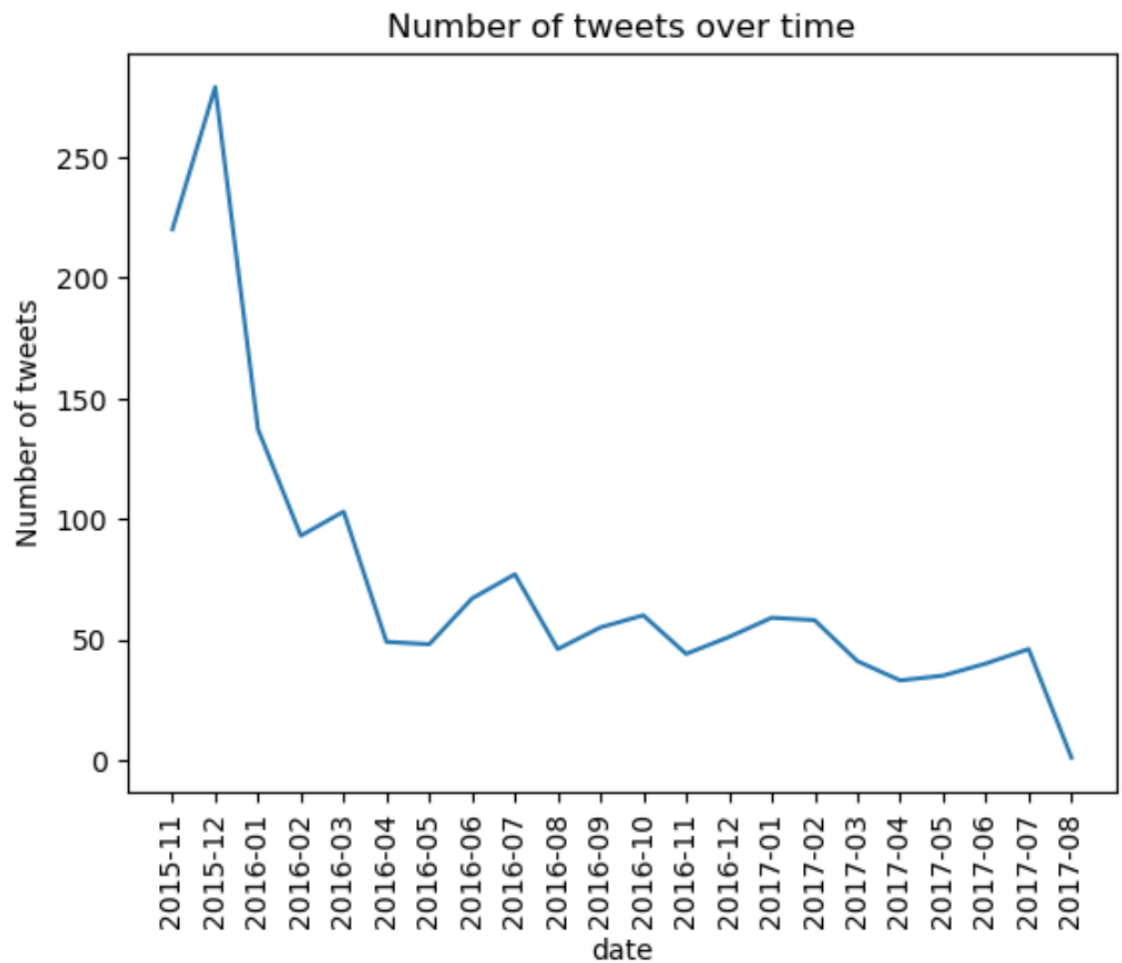## 3.What is the most liked tweet (based on the favorite count)

The same method that I've used for the retweet is used for the likes .
We can see after analysis that the tweet that has the most likes is the same one that has the most retweets and it shows a dog swimming in a pool .the tweet has 141325 likes and the dog has a rating of 13/10

tweet_id	timestamp	source	text	rating_numerator	rating_denominator	name	dog_stage	retweet_count	favorite_count
744234799360020481	2016-06-18 18:26:18	Twitter for iPhone	Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad) https://t.co/7wE9LTEXC4	13.0	10	None	doggo	69163	141325

4.What is the total number of tweets over the months ?

	date	tweet_id
1	2015-12	279
0	2015-11	220
2	2016-01	137
4	2016-03	103
3	2016-02	93
8	2016-07	77
7	2016-06	67
11	2016-10	60
14	2017-01	59
15	2017-02	58
10	2016-09	55
13	2016-12	51
5	2016-04	49
6	2016-05	48
9	2016-08	46
20	2017-07	46
12	2016-11	44
16	2017-03	41
19	2017-06	40
18	2017-05	35
17	2017-04	33
21	2017-08	1

Number of tweets over time

Here is the line plot that shows the variation over the months . After analysis ,we can see that the twitter account WeRateDogs did 279 tweets in December 2015 .

We can see that the number of tweets drops at February 2016 and start to slightly differ from month to month since March 2016.It might be because the account is now quite popular and doesn't necessarily need to post a lot compared to their debut .