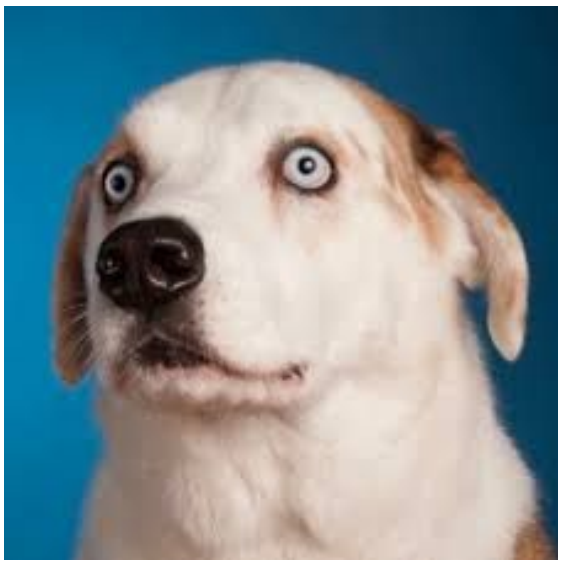
WRANGLE AND ANALYZE DATA

WE RATE DOGS!

LET YOUR DOG BE FAMOUS



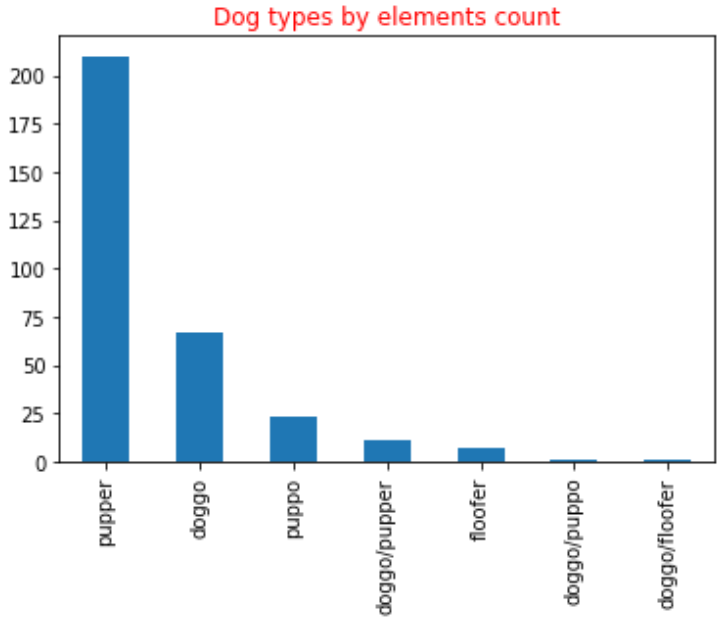
Dogs are so cute, we can easily fall in love with their little cute body. If you're like me you will love this Twitter_account WeRateDogs because they rate dogs ! They are not very though indeed, they gave a mark on 10 range. You send them your dog they assign him a cute dog type like 'puppo' or 'floofer'

A RISING STAR

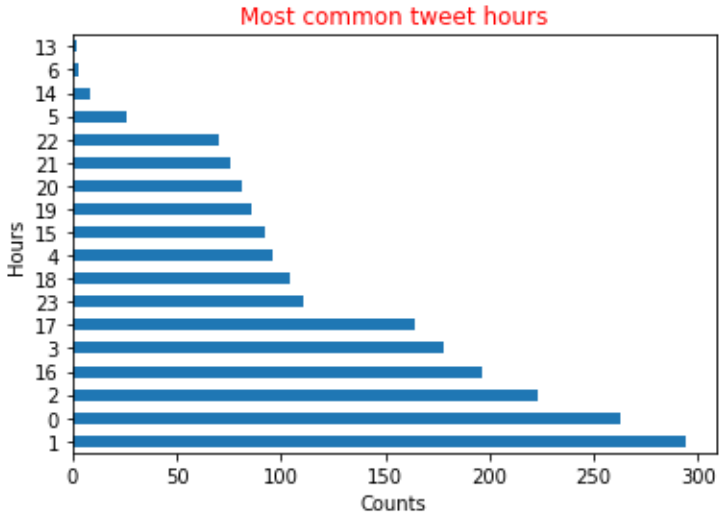
This account have around 8.7 million followers, that's a lot. People follow each tweet, retweet It, add It to favorite. In sum all the actions you can do on Twitter and It generate what we all know as data.

And I'm very kind I analysed those data for you (I was a little bit pushed by Udacity to do It in fact 😝). Let's do It !

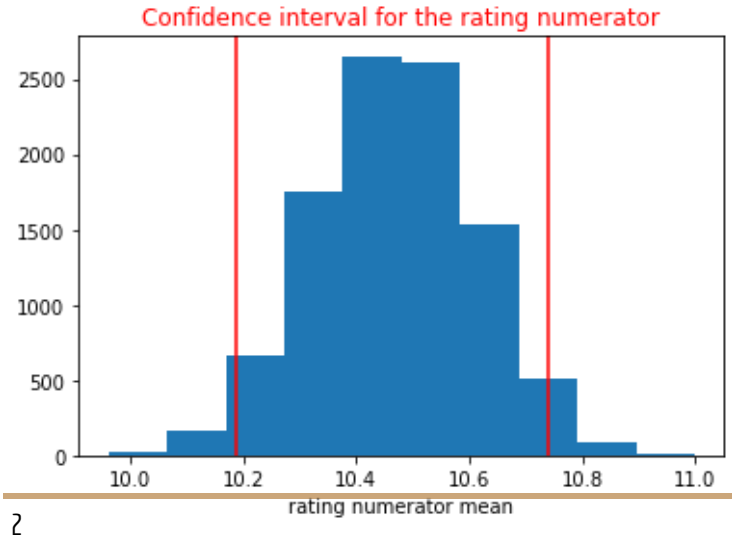
We rate Dog analysis



Those are the types of dogs you meet on the account I created an histogramme to visualize the most common type of dog.

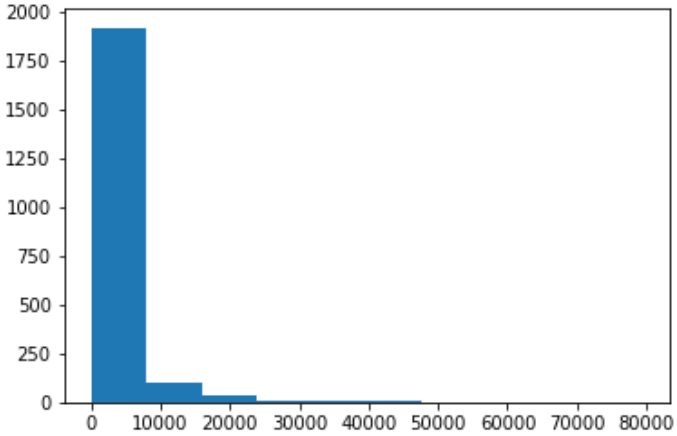


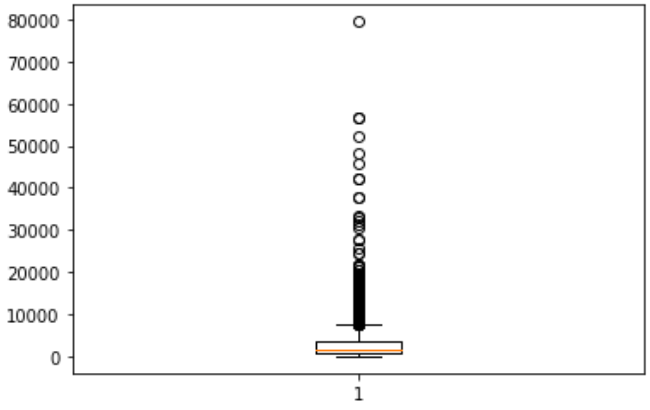
This account is a night owl. The majority of the post are sent between midnight and 2 am. Maybe this is a time when people will react quickly ?



This is the mean of the rating you can get for a dog. On a scale of 10 It’s pretty good, they’re good boys I guess ?

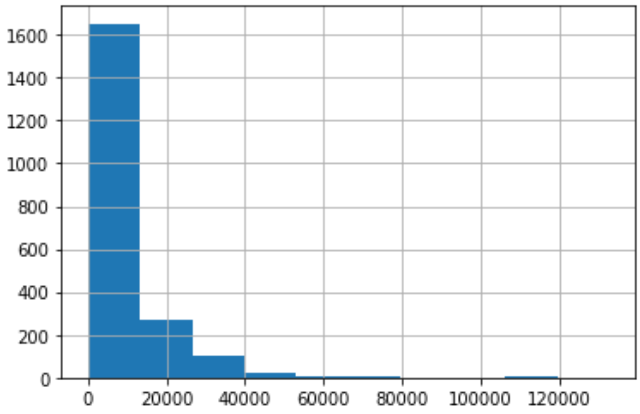
RETWEET COUNT

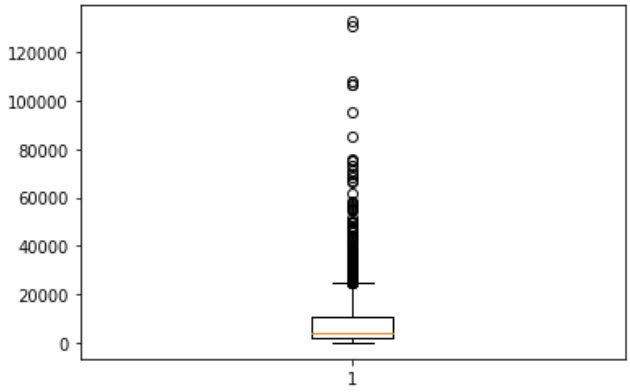




Those tweets are retweeted around a 30000 times. But one was retweeted 80000 one time. The dog must have been very cute !

FAVORITE COUNT





This is pretty much the same, there are around 60000 retweets for each and there are some outliers around 130000 retweets.

Favorite and retweet are in fact correlated: There is a little correlation of **0,625 (R^2)**.

OLS Regression Results

Dep. Variable:	favorite_count	**R-squared:**	0.625
Model:	OLS	**Adj. R-squared:**	0.625
Method:	Least Squares	**F-statistic:**	3449.
Date:	Wed, 25 Mar 2020	**Prob (F-statistic):**	0.00
Time:	14:40:23	**Log-Likelihood:**	-21413.
No. Observations:	2073	**AIC:**	4.283e+04
Df Residuals:	2071	**BIC:**	4.284e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
retweet_count	1.8920	0.032	58.732	0.000	1.829	1.955
intercept	2926.0633	188.927	15.488	0.000	2555.557	3296.570

Omnibus:	1515.027	**Durbin-Watson:**	1.330
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	150199.520
Skew:	-2.670	**Prob(JB):**	0.00
Kurtosis:	44.357	**Cond. No.**	6.81e+03

A question we could ask ourselves : **Is there an other correlation between the number of retweet and the rating ?**

OLS Regression Results

Dep. Variable:	rating_numerator	**R-squared:**	0.093
Model:	OLS	**Adj. R-squared:**	0.092
Method:	Least Squares	**F-statistic:**	211.9
Date:	Wed, 25 Mar 2020	**Prob (F-statistic):**	9.03e-46
Time:	14:47:27	**Log-Likelihood:**	-4456.6
No. Observations:	2073	**AIC:**	8917.
Df Residuals:	2071	**BIC:**	8929.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	10.1934	0.053	192.455	0.000	10.090	10.297
retweet_count	0.0001	9.03e-06	14.556	0.000	0.000	0.000

Omnibus:	674.347	**Durbin-Watson:**	1.633
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2131.276
Skew:	-1.646	**Prob(JB):**	0.00
Kurtosis:	6.720	**Cond. No.**	6.81e+03

R^2 = 0,093 so there is **no** correlation between those two.