act_report

December 17, 2018

1 Act Report

After process of cleaning data the DataFrame has 1979 observations. The most important informations are: - ragins - dogs stages - retweet counts - favourite counts

1.1 Analyze and Visialise

In [76]:

Out[76]:

	in_reply_to_status_id	in_reply_to_user_id	rating	retweet_count
count	2.200000e+01	2.200000e+01	1979.000000	1979.000000
mean	6.962067e+17	4.196984e+09	1.055028	2645.801415
std	4.391913e+16	0.000000e+00	0.217832	4730.033255
min	6.671522e+17	4.196984e+09	0.000000	12.000000
25%	6.724855e+17	4.196984e+09	1.000000	590.500000
50%	6.756022e+17	4.196984e+09	1.100000	1270.000000
75%	7.002919e+17	4.196984e+09	1.200000	3035.000000
max	8.558181e+17	4.196984e+09	1.400000	83893.000000

	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	1979.000000	1979.000000	1979.000000	1.979000e+03	1.979000e+03
mean	8708.233451	1.204144	0.593248	1.346685e-01	6.036492e-02
std	12804.295222	0.562354	0.271855	1.007429e-01	5.093878e-02
min	80.000000	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	1872.000000	1.000000	0.362715	5.417505e-02	1.622240e-02
50%	3897.000000	1.000000	0.587342	1.180890e-01	4.953060e-02
75%	10883.500000	1.000000	0.843635	1.953115e-01	9.164355e-02
max	164648.000000	4.000000	1.000000	4.880140e-01	2.734190e-01

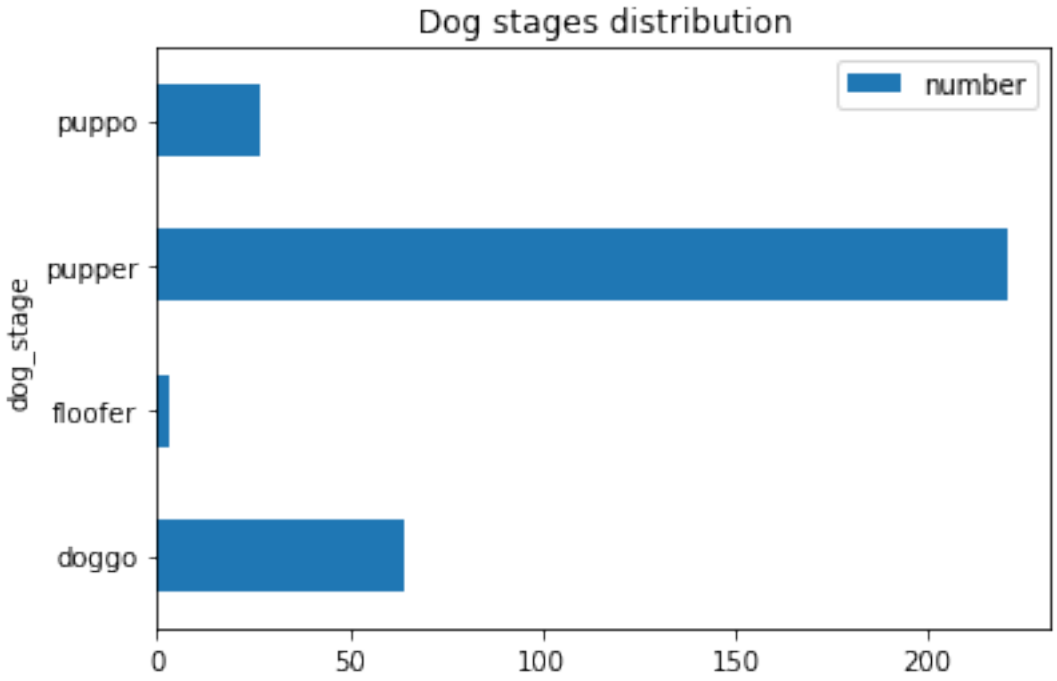
	predicted_conf
count	1674.000000
mean	0.548697
std	0.299368
min	0.000010
25%	0.299772
50%	0.548457
75%	0.819996
max	0.999956

1.2 Visualisations

1.2.1 Graph 1: Dog stages distribution

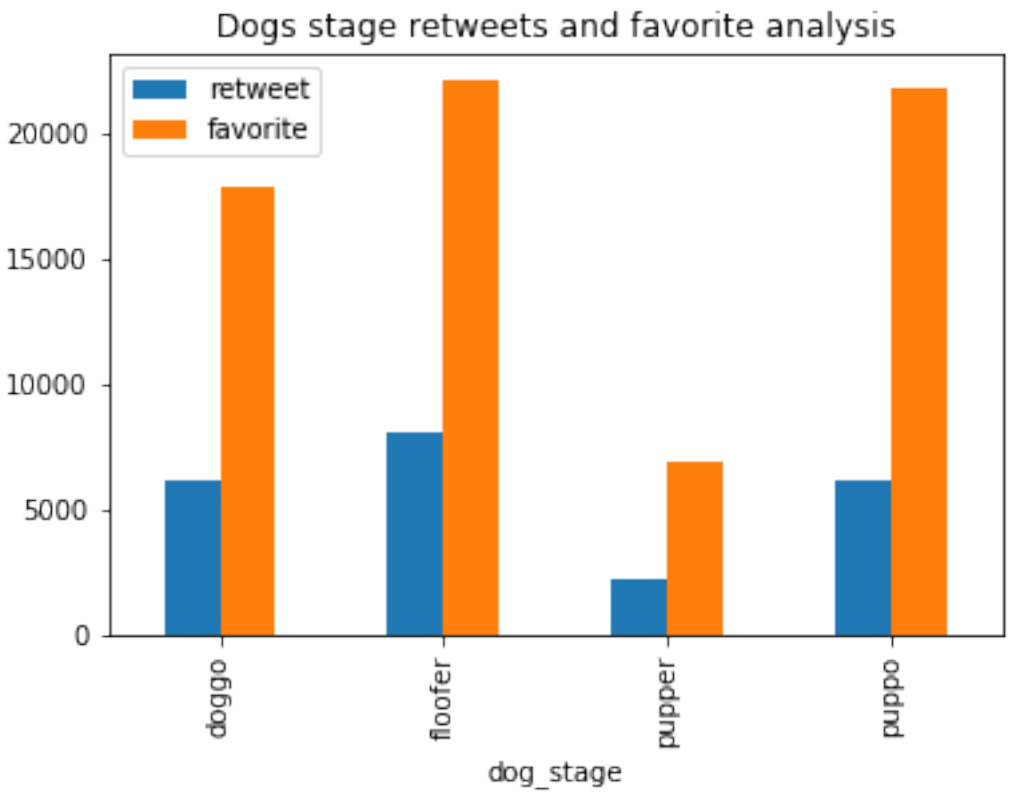
The graph present disribuiton of dog stages among the popoilation

In [79]:



1.2.2 Graph 2: Retweets and favoruites distribuiton for dog stages

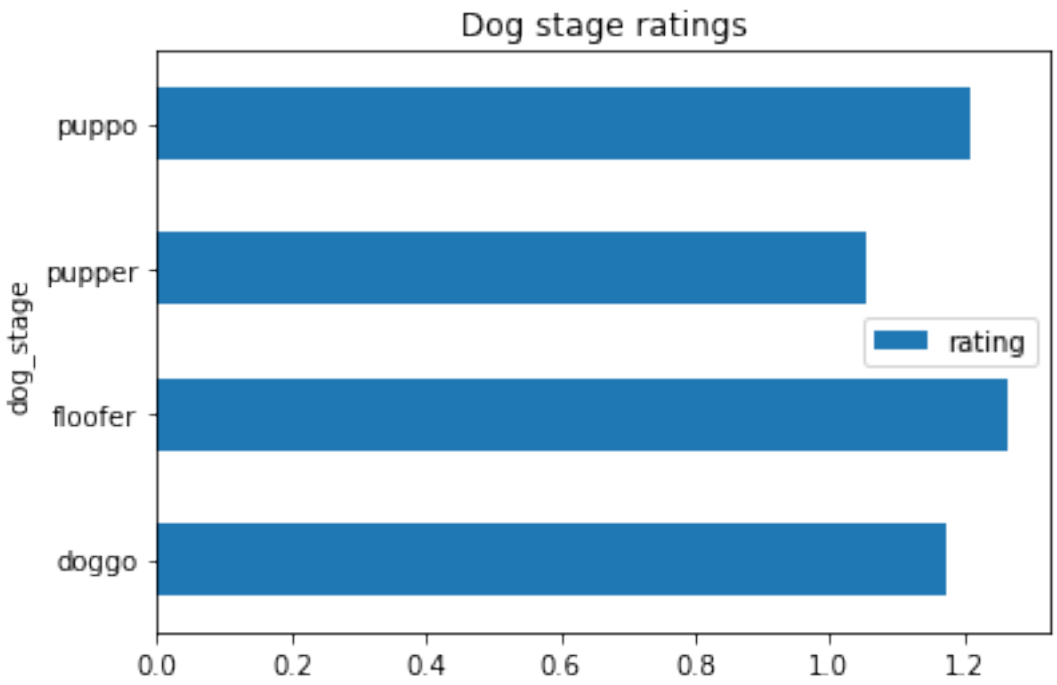
In [80]:



On this chart we can see the relationship between retweets and favourites among different dog stages

1.2.3 Graph 3: Ratings for dog stages

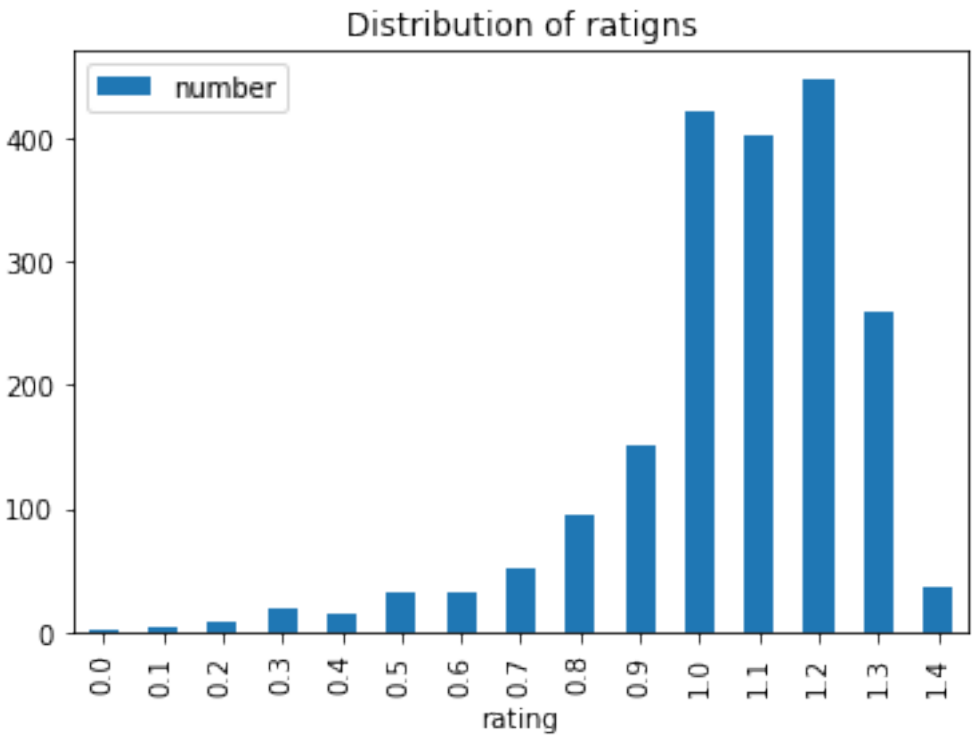
In [81]:



We can see relationship between dog stage and ratings

1.2.4 Graph 4: Ratings distribution

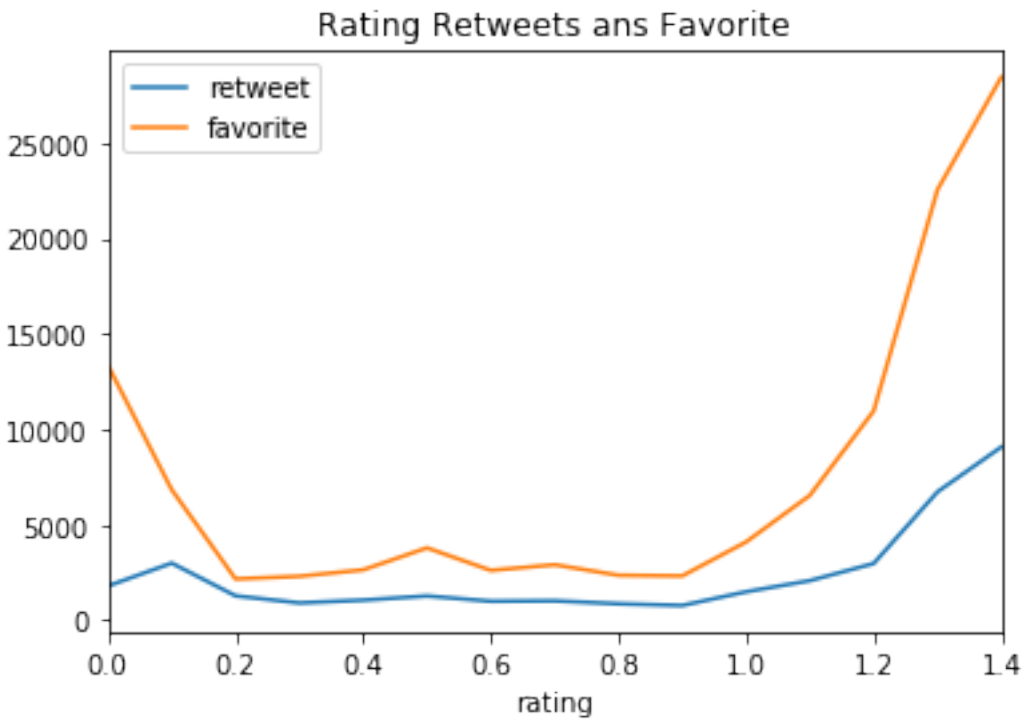
In [83]:



We can observe the rating distribution among all ratings

1.2.5 Graph 5: Rating and retweets relationship

In [84]:



As we can see there is a relationship between ravoirite and rating values

1.3 Insights

- The mean for rating is 1.055 and the thre most common ratings are 1.2,1.0,1.1
- What we can conclude is that the frequency becomes smaller as the rating becomes more extreme.
- The rating distributions is not normal it is left skewed distribution
- Posts with contains higher ratings are supposed to get more favorite ones and theb beeing retweeted.
- Tweets with rating 1.4 gets the highest favorite counts and retweet counts.
- Puppers has the biggest frequency from all gour categories but in the same time they have lowest favourite counts and retweets.

1.4 Limitations

From the DataFrame I got rid off those tweets which didn't have images and also some of those which had bad rating. It was neccesary to get good values for rating. What is more we also needed to calculate the ratings becouse vlueas for rating_numerator and rating_denominators were in two variables and some of theme was useless in puprose of getting wrigt rating