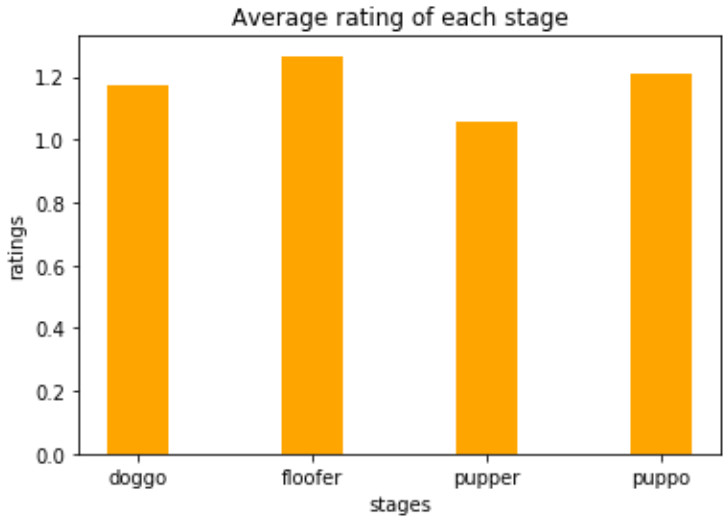
Act Report

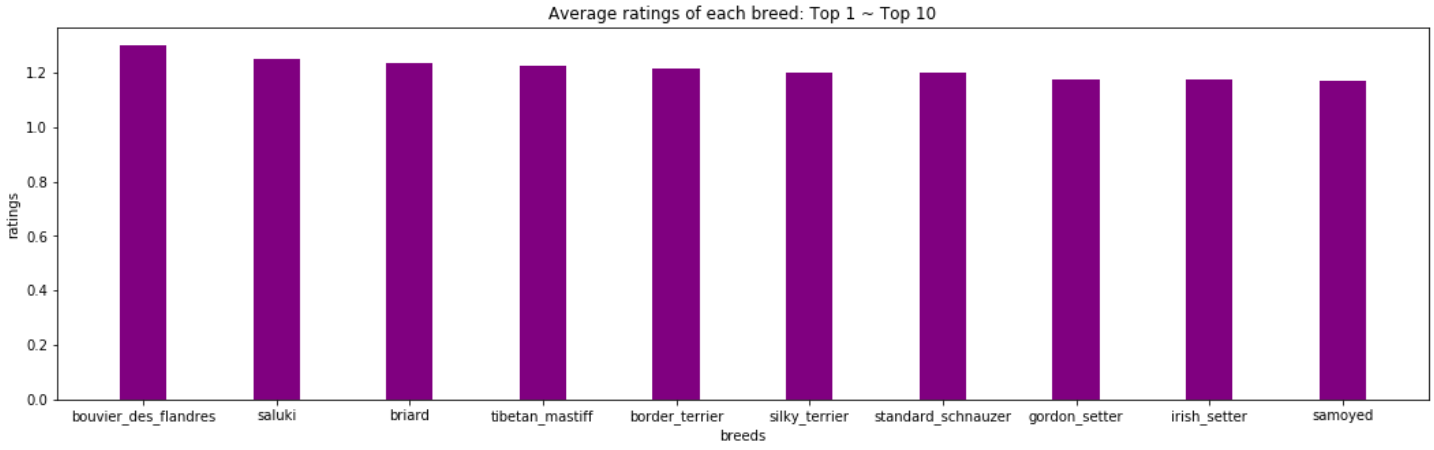
Research Question 1: What stage of a dog has higher ratings on average?

Below is the bar chart of the average ratings of all stages. We can see that dogs of floofer stage have higher ratings than those of other stages.



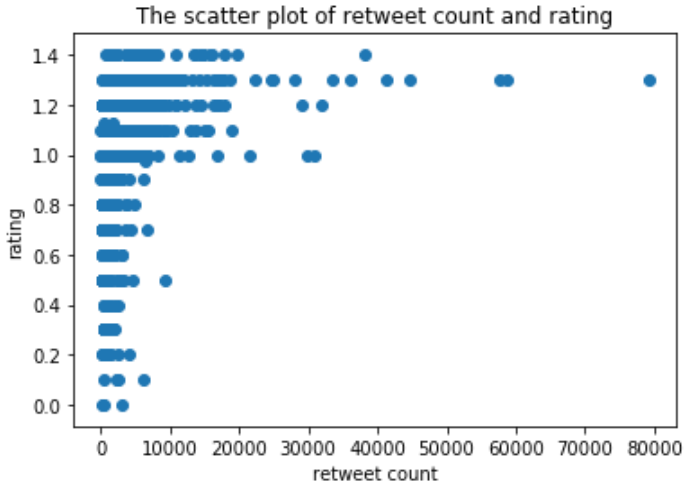
Research Question 2: What breed of a dog gets higher ratings on average.

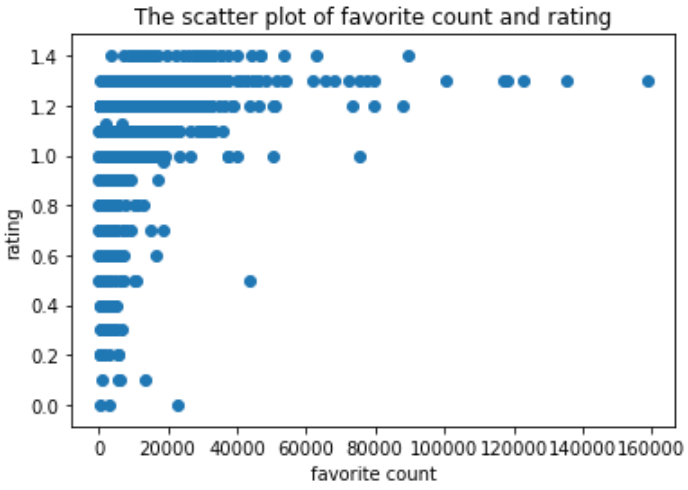
Below is chart of average ratings of top 10 breeds, we can see that bouvier des flandres has a higher rating over other breeds on average.



Research Question 3: What is the relationship between dog ratings and their favorite counts and retweet counts?

The scatter plots below show a very weak linear relationship between retweet_count & rating and favorite_count & rating. The correlations between retweet_count & rating and favorite_count & rating, which are 0.2916 and 0.3841 respectively, also prove the weak linear relationships.



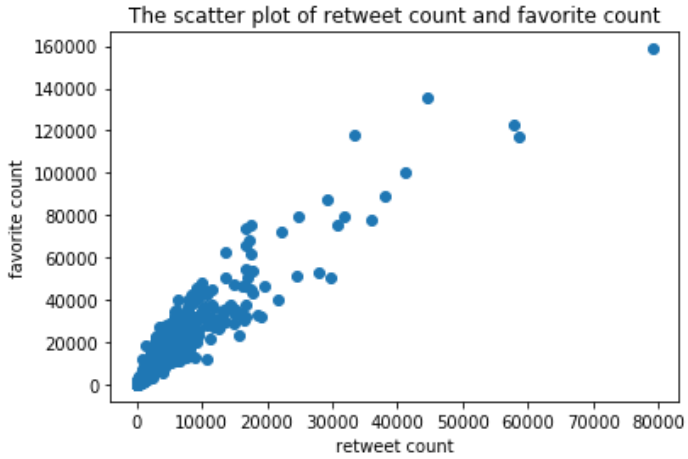


Below is the result table of the linear regression of ratings with retweet_count and favorite_count. The R-squared is very low, indicating the model is not well fit. However, the negative coefficient of retweet_count caught my eye. The scatter plot of retweet_count and favorite_count shows a strong positive linear relationship, which might cause the coefficient of retweet_count flipped.

Dep. Variable:	rating	**R-squared:**	0.179
Model:	OLS	**Adj. R-squared:**	0.178
Method:	Least Squares	**F-statistic:**	215.6
Date:	Thu, 09 Jan 2020	**Prob (F-statistic):**	1.99e-85
Time:	18:14:32	**Log-Likelihood:**	389.64
No. Observations:	1983	**AIC:**	-773.3
Df Residuals:	1980	**BIC:**	-756.5
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	0.9888	0.005	180.675	0.000	0.978	1.000
retweet_count	-2.347e-05	2.7e-06	-8.683	0.000	-2.88e-05	-1.82e-05
favorite_count	1.472e-05	9.79e-07	15.038	0.000	1.28e-05	1.66e-05

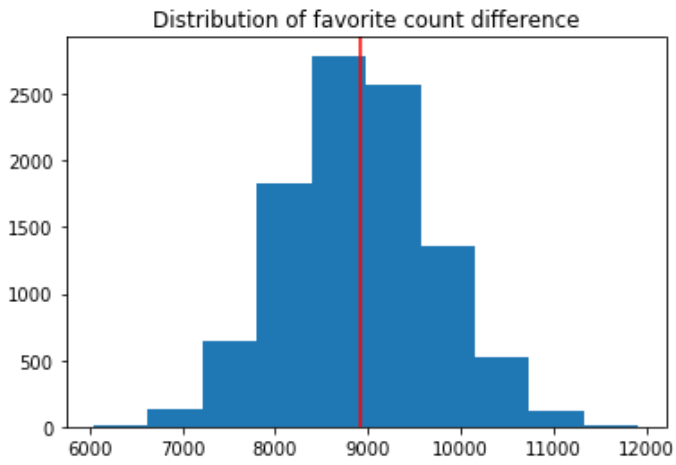
Omnibus:	708.861	**Durbin-Watson:**	1.846
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2686.010
Skew:	-1.737	**Prob(JB):**	0.00
Kurtosis:	7.522	**Cond. No.**	1.92e+04

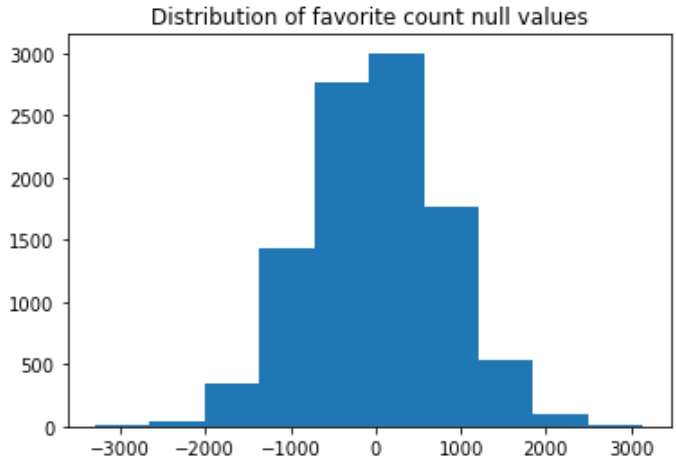


Research Question 4: Are tweets with higher rating dogs more likely to have more favorite count and retweet count?

Null hypothesis: Tweets have higher rating dogs have no more favorite counts than those have lower rating dogs.
Alternative hypothesis: Tweets have higher rating dogs have more favorite counts.

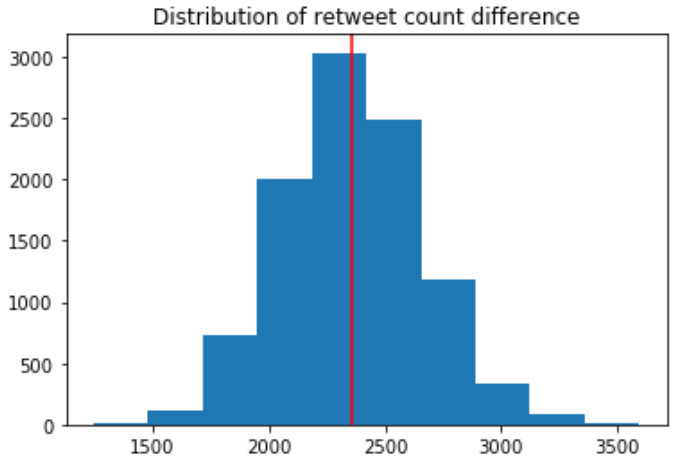
The two charts below showed a normal distribution. The p-value is 0, which means we can reject the null hypothesis. Therefore, tweets with higher rating dogs are more likely to have more favorite counts.

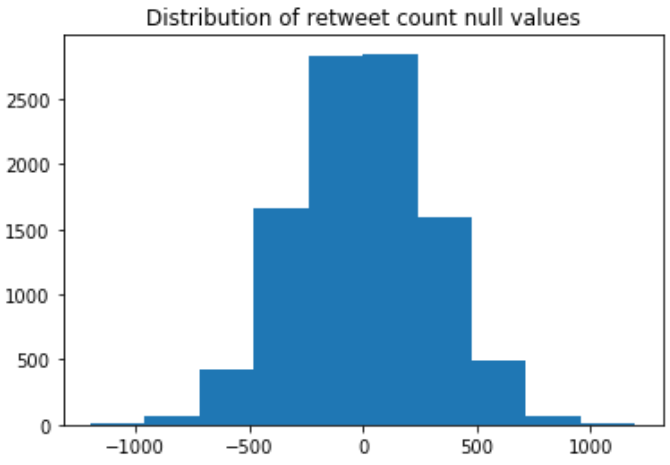




Null hypothesis: Tweets have higher rating dogs have no more retweet counts than those have lower rating dogs.
Alternative hypothesis: Tweets have higher rating dogs have more retweet counts.

The two charts below showed a normal distribution. The p-value is 0, which means we can reject the null hypothesis. Therefore, tweets with higher rating dogs are more likely to have more retweet counts.





Summary:

Dogs at floofer stage have higher ratings on average, bouvier des flandres dogs have higher rating on average. Retweet counts and favorite count has strong and positive linear relationship, but retweet count & rating, favorite count & rating have relatively weak positive linear relationships. Tweets with higher rating dogs are more likely to have more favorite counts and tweets with higher rating dogs are more likely to have more retweet counts.