1. Analysis of Rating of Different Dog Stages

In this analysis, the rating means of different dog stages are extracted from the dataset.

The result is:

Dog stage	*Mean Value of Rating*
Doggo	0.589
Floof	0.609
Pupper	0.543
Puppo	0.604

As a consequence, people find dogs in puppo and floof stages more charming. Dogs in pupper stage seem less attractive. Dogs in floof and puppo stages have more chance in rating.

2. Analysis of Rating of Different Dog Breeds

Dog Breed	*Mean Value of Rating*
Clumber	1,000
Bouvier des Flandres	0,650
Saluki	0,625
Briard	0,617
Tibetan Mastiff	0,613
Border Terrier	0,607
Silky Terrier	0,600
Standard Schnauzer	0,600
Great Pyrenees	0,600
Siberian husky	0,588
Gordon Setter	0,588

Some dog breeds seem to get higher mean rating values. Clumber, Bouvier_des_Flandres, Saluki, Briard and Tibetan mastiff are top five dog breeds. As an insight, it can be said that dogs of these breeds will get higher chance in the rating.

3. Linear Relationship between Rating, Favorite count and Retweet count

3 linear regression models based on least square algorithm is built and tested if there are linear relationships between 3 variables. The r-squared values are checked if the models can include most of the dataset with this model.

3.1. Linear Regression Model: Retweet count and Rating

Linear regression model is built using the statmodels library of Python. The regression result is:

OLS Regression Results

Dep. Variable:	retweet_count	R-squared:	0.075
Model:	OLS	Adj. R-squared:	0.075
Method:	Least Squares	F-statistic:	169.8
Date:	Sat, 19 May 2018	Prob (F-statistic):	2.27e-37
Time:	16:29:48	Log-Likelihood:	-20601.
No. Observations:	2093	AIC:	4.121e+04
Df Residuals:	2091	BIC:	4.122e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-3158.0754	467.458	-6.756	0.000	-4074.806	-2241.345
rating	1.115e+04	855.762	13.030	0.000	9471.917	1.28e+04

Omnibus:	2572.102	Durbin-Watson:	1.866
Prob(Omnibus):	0.000	Jarque-Bera (JB):	399040.970
Skew:	6.448	Prob(JB):	0.00
Kurtosis:	69.404	Cond. No.	11.1

The model can fit only 7% of the total data in the linear regression model. Rating and favorite count do not seem to be in linear relationship.

3.2. Linear Regression Model: Retweet count and Rating

Second linear regression model is built:

OLS Regression Results

Dep. Variable:	favorite_count	R-squared:	0.127
Model:	OLS	Adj. R-squared:	0.127
Method:	Least Squares	F-statistic:	304.6
Date:	Sat, 19 May 2018	Prob (F-statistic):	8.66e-64
Time:	16:36:54	Log-Likelihood:	-22578.
No. Observations:	2093	AIC:	4.516e+04
Df Residuals:	2091	BIC:	4.517e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-1.158e+04	1202.387	-9.631	0.000	-1.39e+04	-9221.868
rating	3.842e+04	2201.176	17.453	0.000	3.41e+04	4.27e+04

Omnibus:	1827.779	Durbin-Watson:	1.461
Prob(Omnibus):	0.000	Jarque-Bera (JB):	71024.212
Skew:	3.984	Prob(JB):	0.00
Kurtosis:	30.403	Cond. No.	11.1

The model can fit 13% of the data in this linear regression model. It does not seem to be useful for predicting favorite count when rating is given as input.

3.3. Linear Regression Model: Retweet count and Favorite count

The third model is built for testing the relationship between Retweet and Favorite counts:

OLS Regression Results

Dep. Variable:	retweet_count	R-squared:	0.839
Model:	OLS	Adj. R-squared:	0.839
Method:	Least Squares	F-statistic:	1.091e+04
Date:	Sun, 20 May 2018	Prob (F-statistic):	0.00
Time:	09:46:43	Log-Likelihood:	-18770.
No. Observations:	2093	AIC:	3.754e+04
Df Residuals:	2091	BIC:	3.755e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-294.6897	50.966	-5.782	0.000	-394.639	-194.740
favorite_count	0.3460	0.003	104.469	0.000	0.339	0.352

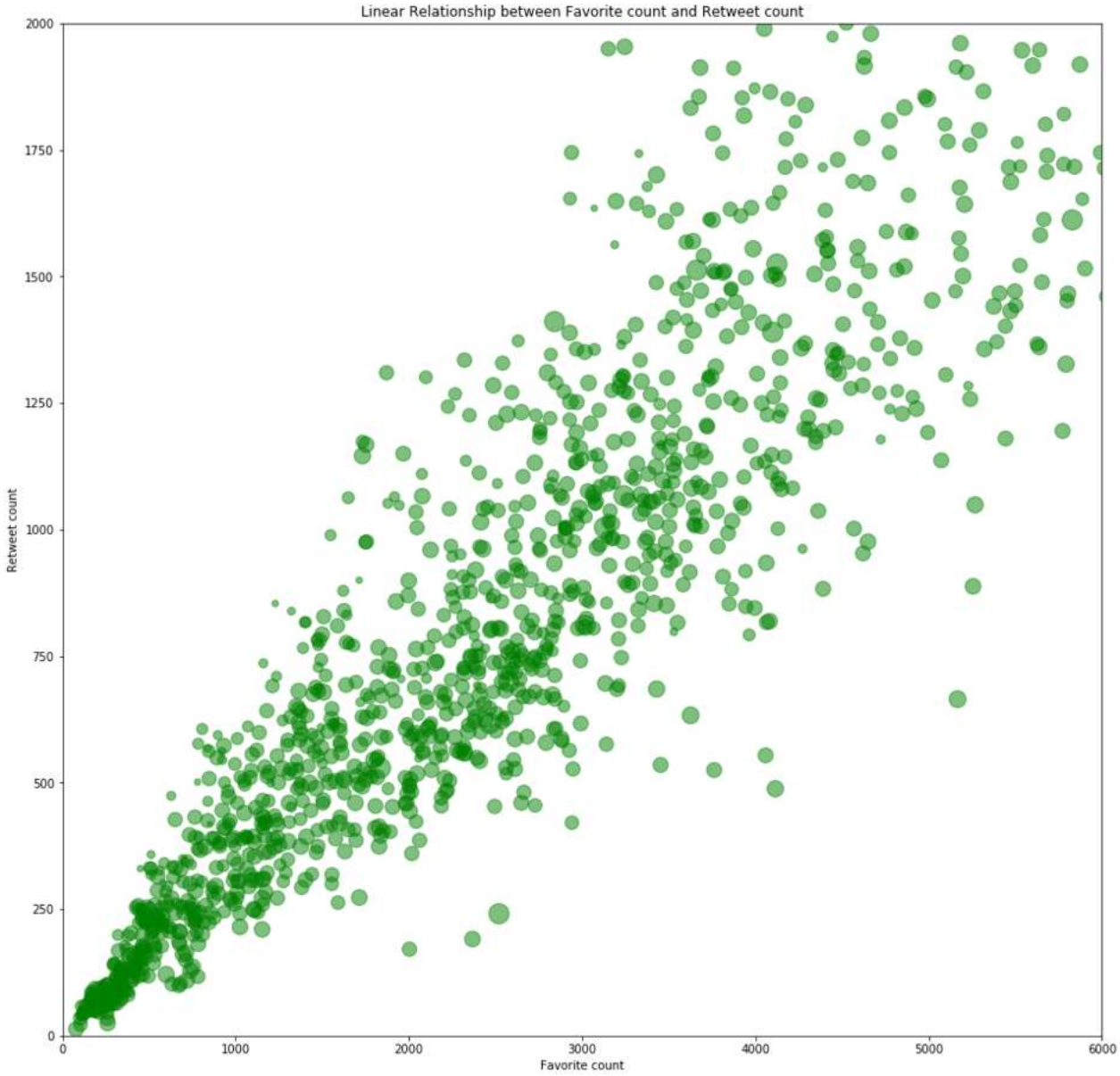
Omnibus:	2219.049	Durbin-Watson:	1.352
Prob(Omnibus):	0.000	Jarque-Bera (JB):	458829.146
Skew:	4.772	Prob(JB):	0.00
Kurtosis:	74.904	Cond. No.	1.89e+04

There is a strong linear relationship between the favorite count and retweet count. R-squared value shows that %84 of the data can be correlated within this linear regression.

The relationship is formulated as: **“Retweet count = 0.35 x Favorite count – 295”**

3.4. Visualisation of Retweet count, Favorite count and Rating data together

Scatter plot function in matplotlib library is used to visualize the relationship between 3 variables together.



Final Results

1. There is strong positive linear relationship between favorite acount and retweet count.

2. There is no linear relationship between rating and retweet count & favorite count: There is no meaningful change in size of dots which are linked to rating value.