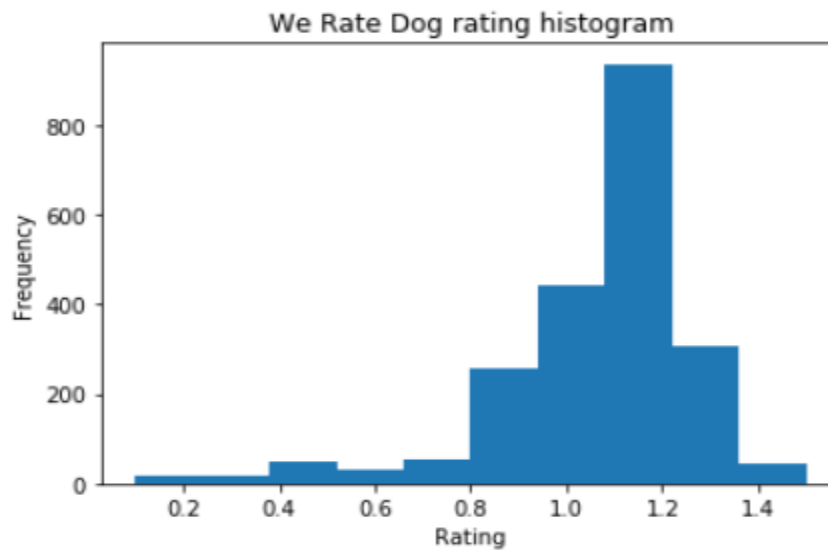# ACT REPORT

This report has worked on We Rate Dog twitter Data with rating people's dogs.


Insight 1

The histogram below describes the rating of dog overall. We can see it is a left skew distribution which major data falls on right side of histogram from 0.9 to 1.4 point. Most of dogs were rated over 10 (denominator). Why? Because "they're good dogs Brent.".

Insight 2

Most of breed tweet in We Rate Dog are Golden Retriever (156/1678). Does it mean Golden Retriever has more rating point?
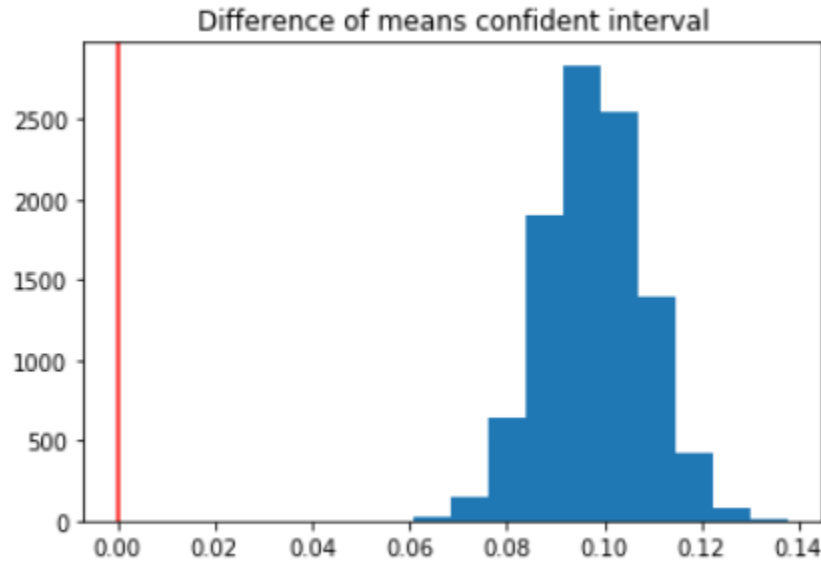
To answer this question, we create hypothesis for it.

$$H_0 : \mu_{golden\ retriever} - \mu_0 = 0$$

$$H_1 : \mu_{golden\ retriever} - \mu_0 > 0$$

H_0 : The null hypothesis which is the mean rating of golden retriever and the mean rating of all dogs.

H_1: The alternative hypothesis which is the mean rating of golden retriever is higher than the mean of all dogs.

$\mu_{golden\ retriever}$: the mean rating of golden retriever.

μ_0: the mean of all dogs.

After using Bootstrapping technique and confident interval approach, we can sample the difference from the mean rating of Golden Retriever and the mean rating of overall breed of dogs. As we can see on the confiden interval above, it is clearly that all of the data are far from 0 which is mean the difference of means is larger than 0. Moreover, by a simple calculator, we have the p-value of difference of mean higher than 0 is 1. Therefore, we can say we are 100% confident to reject the null hypothesis.

Insight 3

We Rate Dog twitter rate people's dog with their own point with a humorous comment about the dog. However, does its rating effect on people clicking favorite? The OLS regression results below show the correlation of rating by We Rate Dog and favorite count for each tweet.

OLS Regression Results

Dep. Variable:	favorite_count	R-squared:	0.146
Model:	OLS	Adj. R-squared:	0.145
Method:	Least Squares	F-statistic:	367.5
Date:	Mon, 25 Mar 2019	Prob (F-statistic):	9.16e-76
Time:	05:38:26	Log-Likelihood:	-23274.
No. Observations:	2160	AIC:	4.655e+04
Df Residuals:	2158	BIC:	4.656e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
rating	2.201e+04	1148.217	19.170	0.000	1.98e+04	2.43e+04
intercept	-1.489e+04	1246.141	-11.947	0.000	-1.73e+04	-1.24e+04

Omnibus:	2116.075	Durbin-Watson:	1.573
Prob(Omnibus):	0.000	Jarque-Bera (JB):	130977.437
Skew:	4.616	Prob(JB):	0.00
Kurtosis:	40.014	Cond. No.	9.94

The formular:

$$\textit{Favorite count} = \beta_0 - \beta_{Rating}\textit{Rating}$$

The p>|t| of coefficient of rating and intercept are 0.0 which means the coefficients are significant statistic and coefficient of rating valuable is positive 2201. Therefore, we can say for each rating We Rate Dog has, the favourit counted increases 2201clicks.