

Data Wrangling Project

Summary and conclusions

Insight 1 :

The prediction model has been most confident in predicting golden retriever by 0.72 confident ratio in the second most confident went for the pembroke breed by 0.71.



A golden retriever rated 13/10.

Insight 2 :

The results of the multiple linear regression shows us that the model predicts that the numerator rate for pembroke is 11.44 and the predicted numerator rate for the labrador retriever is less than pembroke's rate by 0.3092 and for golden retriever is higher by 0.1322 .

Dep. Variable:	rating_numerator	R-squared:	0.017			
Model:	OLS	Adj. R-squared:	0.011			
Method:	Least Squares	F-statistic:	2.815			
Date:	Sun, 19 Sep 2021	Prob (F-statistic):	0.0614			
Time:	11:45:47	Log-Likelihood:	-586.04			
No. Observations:	331	AIC:	1178.			
Df Residuals:	328	BIC:	1189.			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
intercept	11.4432	0.152	75.184	0.000	11.144	11.743
labrador_retriever	-0.3092	0.210	-1.471	0.142	-0.723	0.104
golden_retriever	0.1322	0.193	0.686	0.493	-0.247	0.511
Omnibus:	111.863	Durbin-Watson:	1.411			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	397.050			
Skew:	-1.464	Prob(JB):	6.05e-87			
Kurtosis:	7.497	Cond. No.	4.18			

OLS Regression Results



A Pembroke female dog studies international doggo communication.



A labrador_retriever named Gus he has gotten over 25K Favourites.

Insight 3 :

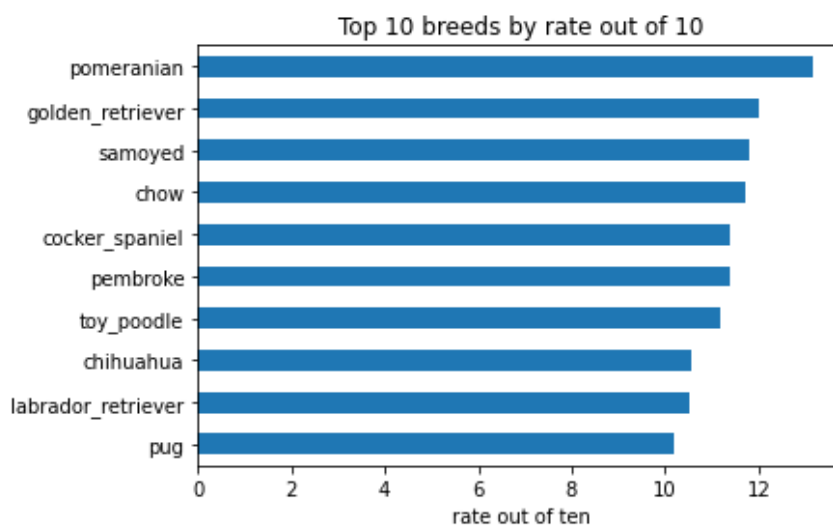
From the linear regression model results we would predict that for each number increases in the rating_numerator the favourites count would increase by 4.7 .

Dep. Variable:	favorites	R-squared:	0.000			
Model:	OLS	Adj. R-squared:	-0.000			
Method:	Least Squares	F-statistic:	0.5038			
Date:	Sat, 18 Sep 2021	Prob (F-statistic):	0.478			
Time:	12:51:22	Log-Likelihood:	-22240.			
No. Observations:	2055	AIC:	4.448e+04			
Df Residuals:	2053	BIC:	4.449e+04			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
intercept	8522.5673	278.878	30.560	0.000	7975.655	9069.480
rating_numerator	4.7441	6.684	0.710	0.478	-8.363	17.852
Omnibus:	1671.609	Durbin-Watson:	1.183			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	46369.150			
Skew:	3.691	Prob(JB):	0.00			
Kurtosis:	25.069	Cond. No.	43.5			

OLS Regression Results

Visualization :

This bar chart illustrates the percentage of the top ten breeds by numerator rate that has denominator of ten for the top ten predicted type of dogs , the information is divided into ten groups , we can see here that pomeranian breed has gotten the highest ratio by 13.16 , in contrast the pug breed has gotten the lowest of the top ten by 10.2 , while cocker_spaniel and pembroke have gotten same proportion by 11.4 for each , and for labrador and chihuahua breeds they also have gotten almost equal average of ratio by 10.5 for each .





A dog named Comet of the Pomeranian breed.



A dog named Waffles of the Pug breed.