act_report

Executive Summary

Within this document the analysis undertaken and insights found within the *'We rate dogs'* twitter page analysis are discussed. The following tests were undertaken:

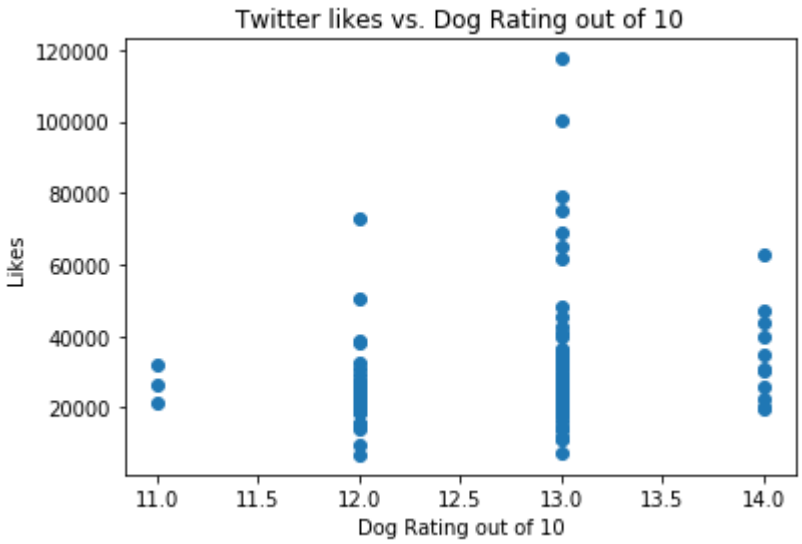
- Relationship between the dog rating and likes
- Frequency that the dog predicter AI guesses something that is not a dog on the first guess
- Relationship between dog type and the number of retweets

Relationship between the dog rating and likes

Within this section I was investigating if the rating given by *'We rate dogs'* had of the number of likes a tweet received. In the previous section the ratings had been factorised to all be out of 10, therefore the denominators were not considered in the analysis.

Because some rating had extreme numbers, e.g. 1776/10 (based on an American Independence Day tweet) I further cleaned the data set by removing the extreme outliers, the bottom and top one percent.

Once completed I set up a simple linear regression model to predict the number of likes based on rating. Based on the results it appeared that there was a correlation between the two with the p-value being 0, thus showing that this correlation is statistically significant. Below is a plot showing the relationship between dog rating and likes:



Frequency that the dog predicter AI guesses something that is not a dog on the first guess

The second thing I looked at was the frequency within which the AI image predicter would predict something that was not a dog based on the confidence. I.e, how was confidence related to whether or not the image was a dog. In order to this I set up a logistical regression model. The results unsurprisingly showed that the confidence was strongly linked to whether or not a dog bread has been selected. The p-value was 0 in this instance showing that it was statistically significant.

Relationship between dog type and the number of retweets

The final area I looked into was the relationship between the dog type and the number of tweets. For this I set up a multiple linear regression model, first creating dummy variables for each of the dog types. Setting 'None' as the baseline all other dog_types were put into the model. The results were as followed, showing that *'doggo, doggo/floofer, doggo/pupper'* appeared to be most significant in affecting whether or not a tweet would be retweeted.

OLS Regression Results

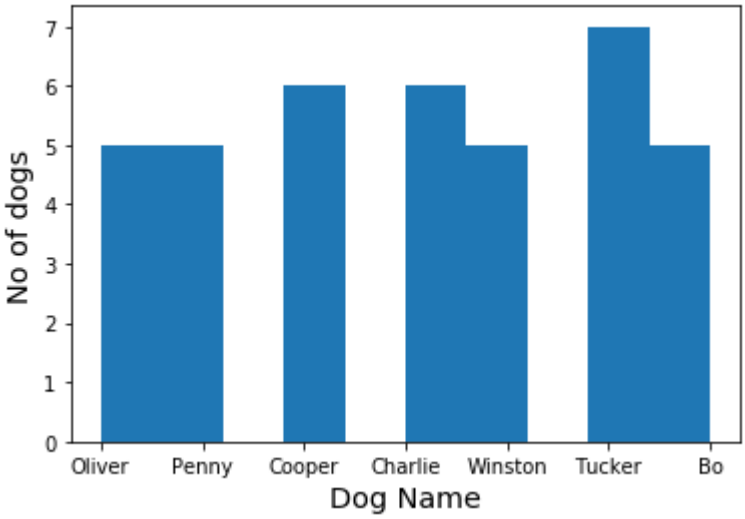
Dep. Variable:	retweet_count	**R-squared (uncentered):**	0.109
Model:	OLS	**Adj. R-squared (uncentered):**	0.103
Method:	Least Squares	**F-statistic:**	18.04
Date:	Thu, 27 Feb 2020	**Prob (F-statistic):**	8.03e-20
Time:	10:27:41	**Log-Likelihood:**	-9156.0
No. Observations:	893	**AIC:**	1.832e+04
Df Residuals:	887	**BIC:**	1.835e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
doggo	6614.3433	841.503	7.860	0.000	4962.773	8265.913
puppo	3088.0000	6888.002	0.448	0.654	-1.04e+04	1.66e+04
pupper	3932.0000	2296.001	1.713	0.087	-574.228	8438.228
floofer	4617.0000	2603.420	1.773	0.076	-492.582	9726.582
doggo/floofer	4169.5000	874.777	4.766	0.000	2452.626	5886.374
doggo/pupper	6004.3043	1436.248	4.181	0.000	3185.464	8823.145

Omnibus:	890.604	**Durbin-Watson:**	1.502
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	51614.670
Skew:	4.556	**Prob(JB):**	0.00
Kurtosis:	39.113	**Cond. No.**	8.19

Most Common Dog Name

Within the final section I reviewed which is the most common dog name using the .value_counts function. Joint top was Tucker with Charlie and Cooper a close second. In order to visually assess the data I plotted a histogram with only the most popular dog names, see below:


No of dogs
0
1
2
3
4
5
6
7
Oliver
Penny
Cooper
Charlie
Winston
Tucker
Bo
Dog Name