act_report

March 30, 2021

1 Data Analyst Nanodegree

2 Project 4 - Wrangling and Analyze Data

2.1 Findings on WeRateDogs Twitter archive

2.1.1 Project outline

The project is centered around **WeRateDogs** Twitter account.
Their usual tweets are dogs photos with **humourous ratings** of x/10, where x is often more than 10, such as **13/10** .
The account is popular and therefore has enough data to explore and attempt learning from.

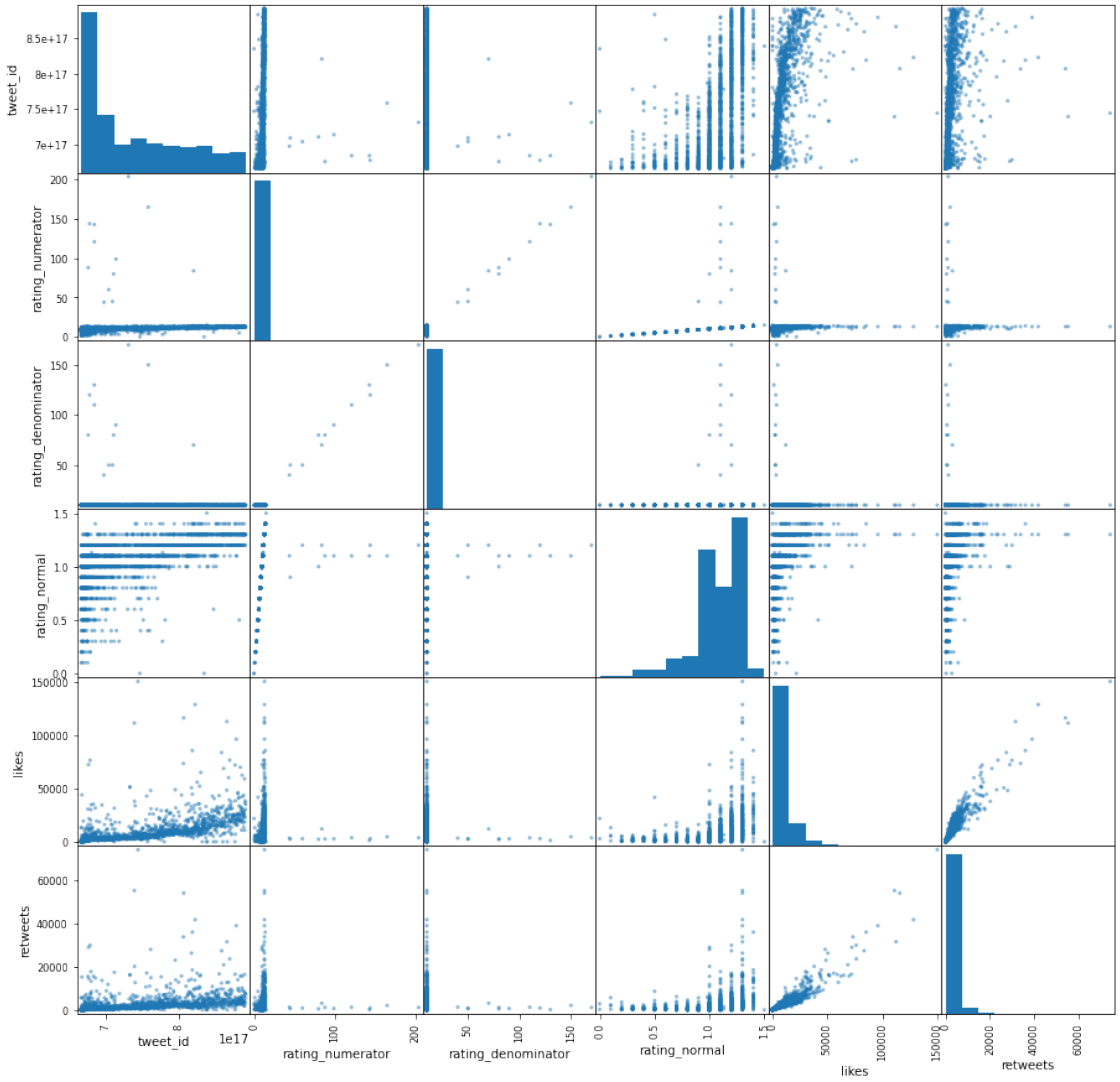
2.1.2 Input

For analysis and visualization, **clean, wrangled** data was used. The data included: - **Dog ratings** issued by WeRateDogs - **Likes** and **retweets** - **Dog names** - **Unique dog stages**: doggo, floofer, pupper, puppo

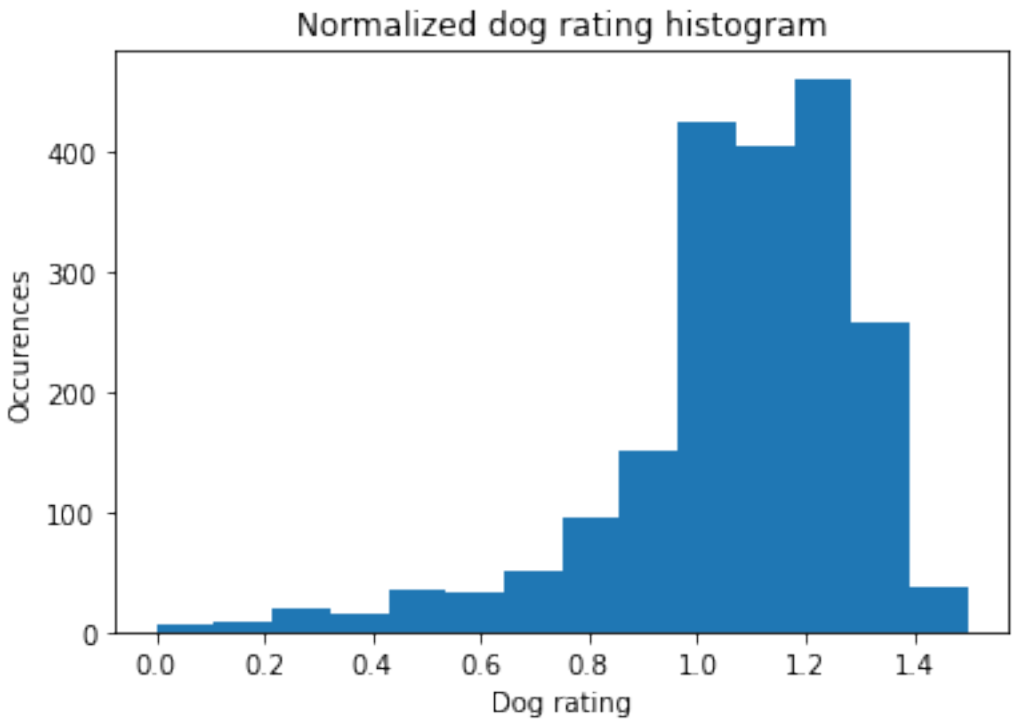
2.1.3 Findings

	tweet_id	rating_numerator	rating_denominator	rating_normal	likes	retweets
count	1.995000e+03	1995.000000	1995.000000	1995.000000	1986.000000	1986.000000
mean	7.365631e+17	11.128962	10.511278	1.055559	7978.730614	2343.596677
std	6.771664e+16	8.569456	7.259293	0.218176	11808.610270	4207.636784
min	6.660209e+17	0.000000	10.000000	0.000000	0.000000	11.000000
25%	6.759938e+17	10.000000	10.000000	1.000000	1667.000000	517.000000
50%	7.092251e+17	11.000000	10.000000	1.100000	3562.000000	1115.500000
75%	7.890232e+17	12.000000	10.000000	1.200000	9890.000000	2671.500000
max	8.924206e+17	204.000000	170.000000	1.500000	150130.000000	74079.000000

Numeric data summary Normalized dog rating (`rating_normal`) is **centered around 1.05-1.10** with mean and median relatively near to one another.
This also confirms that in more cases than not the rating **exceeds 10/10** (ie. 1.00).
Likes and retweets have their **means much higher than medians** which suggests their distributions are heavily right-skewed.



Visual outlook on numeric data Plots in the matrix suggest that **likes** and **retweets** are correlated.



Dog ratings distribution The ratings distribution is left-skewed, meaning **most dogs get high ratings** from WeRateDogs curators.

OLS Regression Results

Dep. Variable:	retweets	**R-squared:**	0.863
Model:	OLS	**Adj. R-squared:**	0.863
Method:	Least Squares	**F-statistic:**	1.254e+04
Date:	Tue, 30 Mar 2021	**Prob (F-statistic):**	0.00
Time:	01:44:37	**Log-Likelihood:**	-17491.
No. Observations:	1995	**AIC:**	3.499e+04
Df Residuals:	1993	**BIC:**	3.500e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
intercept	-297.3139	42.045	-7.071	0.000	-379.771	-214.857
likes	0.3310	0.003	111.999	0.000	0.325	0.337

Omnibus:	1814.247	**Durbin-Watson:**	1.276
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	250413.315
Skew:	3.740	**Prob(JB):**	0.00
Kurtosis:	57.374	**Cond. No.**	1.72e+04

Likes and retweets relation **Retweets are clearly correlated to likes** with R-squared of 86.3%.

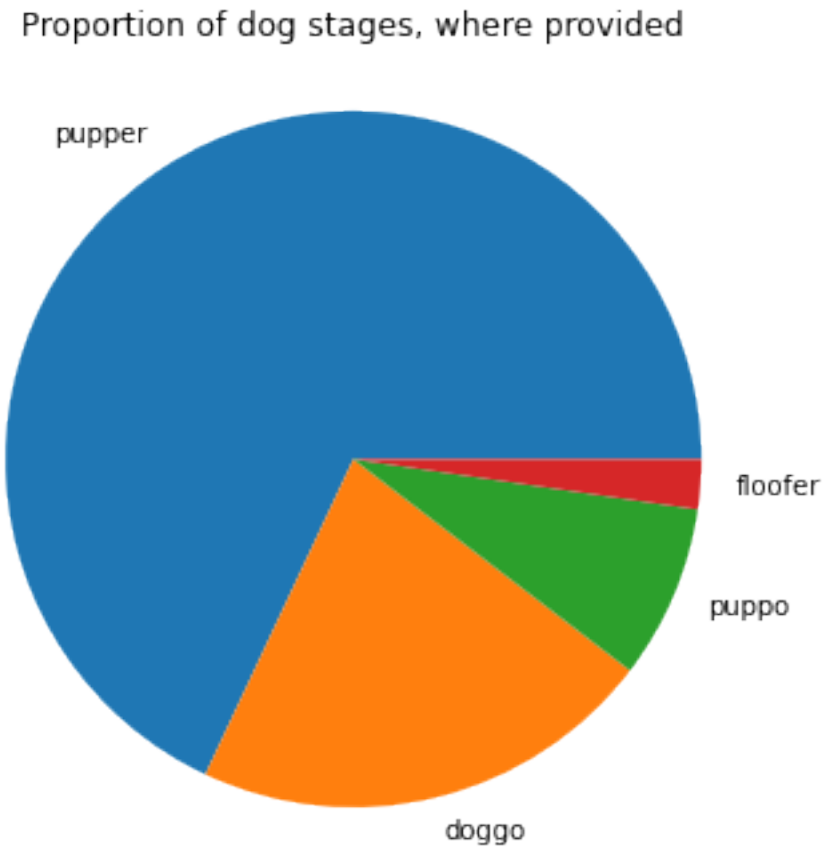
OLS Regression Results

Dep. Variable:	likes	**R-squared:**	0.145
Model:	OLS	**Adj. R-squared:**	0.145
Method:	Least Squares	**F-statistic:**	338.5
Date:	Tue, 30 Mar 2021	**Prob (F-statistic):**	6.14e-70
Time:	02:29:41	**Log-Likelihood:**	-21376.
No. Observations:	1995	**AIC:**	4.276e+04
Df Residuals:	1993	**BIC:**	4.277e+04
Df Model:	1		
Covariance Type:	nonrobust		

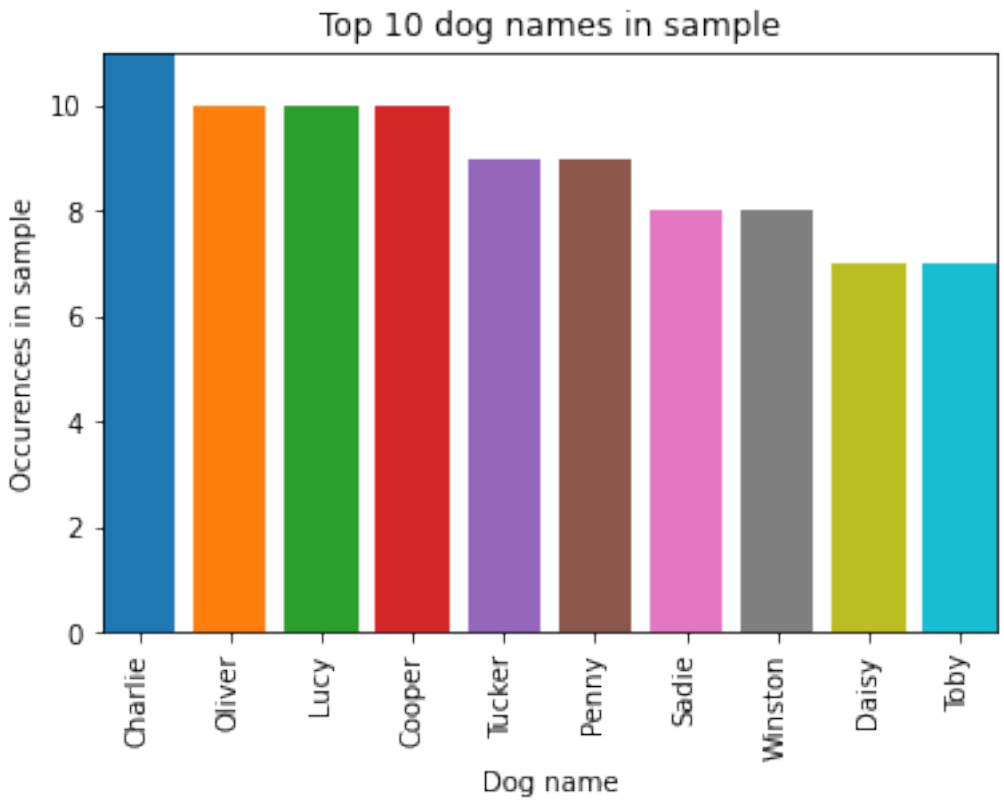
	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
intercept	-1.374e+04	1205.474	-11.398	0.000	-1.61e+04	-1.14e+04
rating_normal	2.058e+04	1118.396	18.397	0.000	1.84e+04	2.28e+04

Omnibus:	1962.214	**Durbin-Watson:**	1.616
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	117625.050
Skew:	4.649	**Prob(JB):**	0.00
Kurtosis:	39.450	**Cond. No.**	9.81

Dog ratings and likes relation The **relation between dog ratings and likes is weak** with R-squared of 14.5%. That is, how many likes a tweet/dog gets is dependent on the WeRateDogs-issued dog rating in 14.5%.



Most popular dog stages Roughly two-thirds of dog 'stages' assigned by WeRateDogs curators are **puppers** !



Dog ratings distribution Self explanatory. My neighbors' got a doggo named Winston, if you doubt the figures.
Still, the winner is… **Charlie** . No wonder - seems as good a name for a girl-doggo as well as a boy-doggo.

2.1.4 Summary

- Dog **ratings distribution is left-skewed**, meaning most dogs get high ratings.
- **Retweets are strongly related to likes** at R-squared of 0.863.
- **Relation of likes to dog ratings is weak** at R-squared of 0.145.
- Above relations are both statistically and practically **significant**.
- Most popular dog names and stages are visible in above plots.
- WeRateDogs' tweets typically get **33 retweets per 100 likes**.
- We can expect that for every **0.1 point increase in normalized dog rating, a tweet gathers 2058 likes**.
- **The WeRateDogs Twitter account is hilarious.**