Act_Report

September 8, 2020

0.1 Analyzing and Visualizing the WeRateDog Twitter Account

By Sherif Sakr Date 8 September 2020

Introduction WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson,As of December 2018, the Twitter account has nearly 7.6 million followers, and Nelson sees 30,000 likes on a post as being viral. His most popular post was of a dog marching in the 2017 Women's March, which was retweeted more than 50,000 times and favorited 134,000 times.

The dataset that are wrangled, analyzed, and visualized is the Tweet archive of Twitter user@dog_rates, also known as WeRateDogs.These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. why? Because "they are good dogs Brent."

#####

Analyzing and Visualizing data

Descriptive Statistics, the Five-number summary

[29]:
```
# Use Pandas.DataFrame.describe() function for general descriptive statistics
analyze_df[['rating','retweet_count','favorite_count']].describe()
```

[29]:

	rating	retweet_count	favorite_count
count	2086.000000	2086.000000	2086.000000
mean	12.191755	2480.712848	8251.186002
std	40.471225	4342.006894	11992.381486
min	0.000000	11.000000	70.000000
25%	10.000000	550.000000	1820.000000
50%	11.000000	1200.000000	3758.500000
75%	12.000000	2818.750000	10314.000000
max	1776.000000	76323.000000	154053.000000

[30]:
```
print('The maximum value of retweet is:',format(analyze_df.retweet_count.max()))
```

```
The maximum value of retweet is: 76323
```

[31]:
```
print('The maximum value of favorite (like) is:',format(analyze_df.
↪favorite_count.max()))
```

```
The maximum value of favorite (like) is: 154053
```

[32]:
```
print('The maximum value of retweet is:',format(analyze_df.retweet_count.min()))
```

```
The maximum value of retweet is: 11
```

[33]:
```
print('The maximum value of favorite (like) is:',format(analyze_df.
↪favorite_count.min()))
```

```
The maximum value of favorite (like) is: 70
```

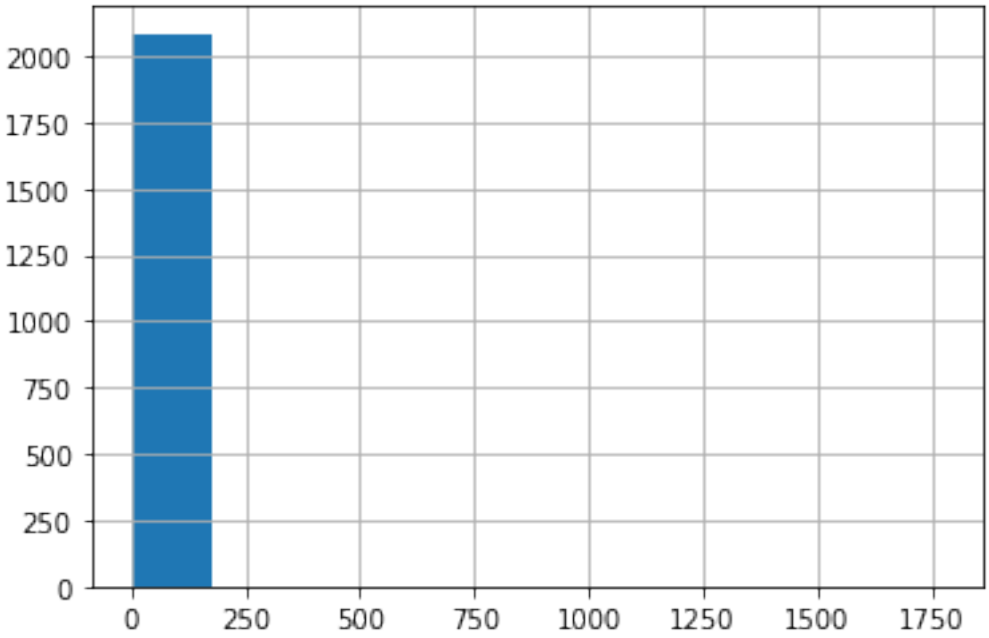
[34]:
```
# Display the descriptive statistics for favorite count on dogs stage category
analyze_df.groupby('dogs_stage')['favorite_count'].describe()
```

[34]:

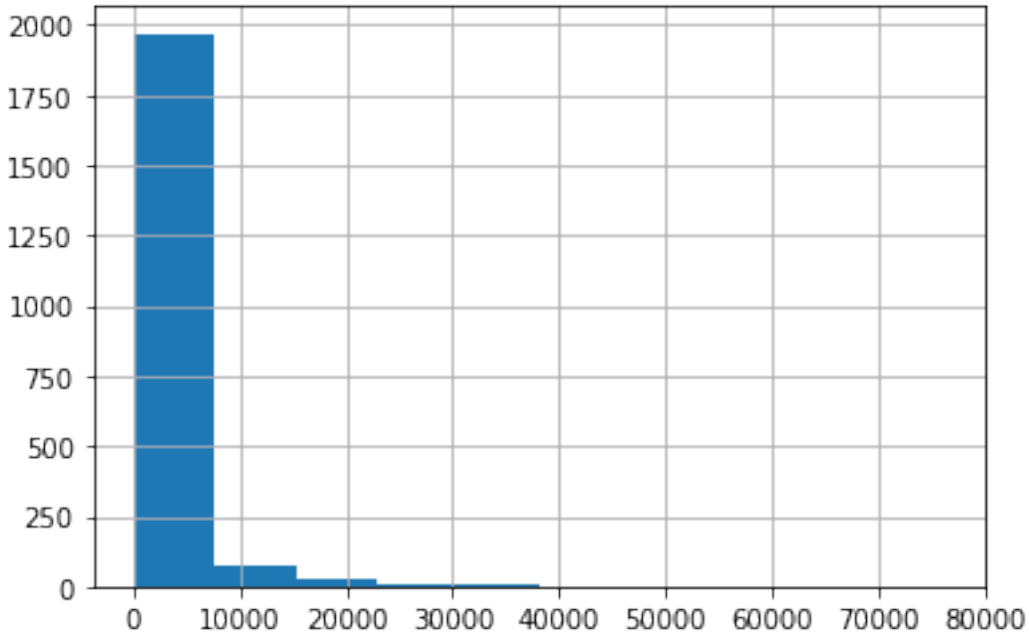
dogs_stage	count	mean	std	min	25%	50%	75%	max
None	1752.0	7870.299658	10807.847508	70.0	1585.50	3551.5	10055.00	118965.0
doggo	71.0	17306.647887	24183.099227	2297.0	6607.50	10635.0	17218.00	154053.0
floofer	10.0	10530.900000	9254.382511	1435.0	3934.25	7750.0	14564.75	29866.0
pupper	229.0	6855.187773	10593.003231	238.0	2198.00	3071.0	7398.00	115071.0
puppo	24.0	21637.083333	27312.858602	2935.0	6336.25	14592.0	20532.25	131363.0

Histograms for various features

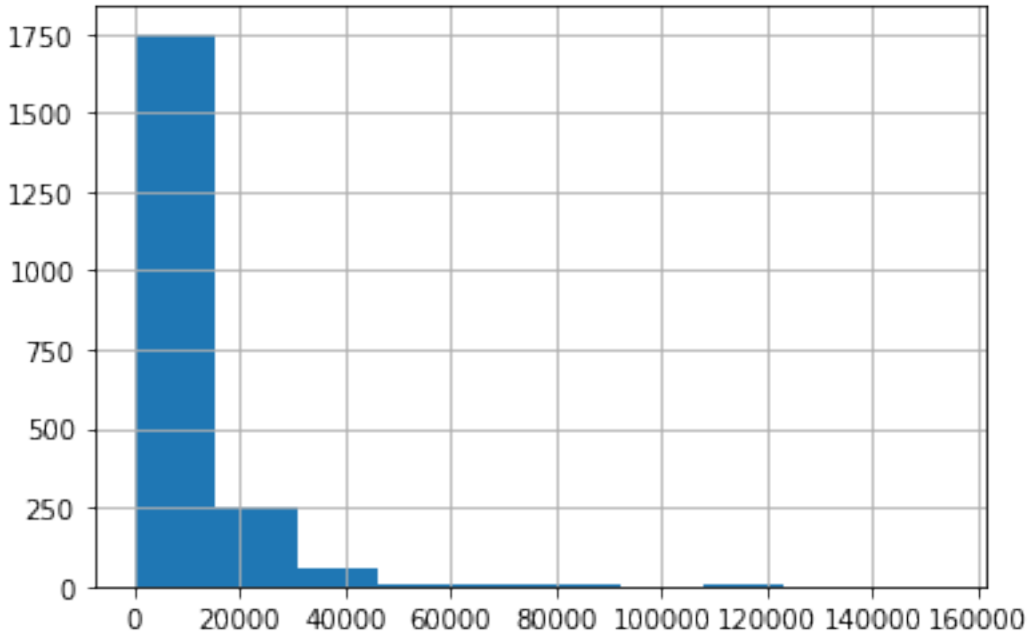
[35]:
```
analyze_df.rating.hist();
```



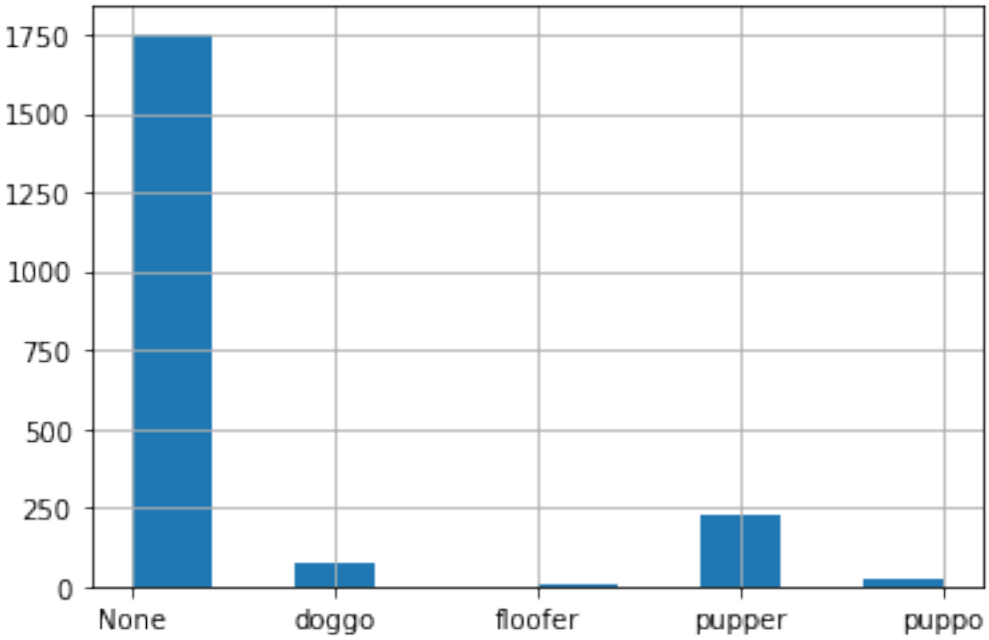
[36]:
```
analyze_df.retweet_count.hist();
```



[37]:
```
analyze_df.favorite_count.hist();
```

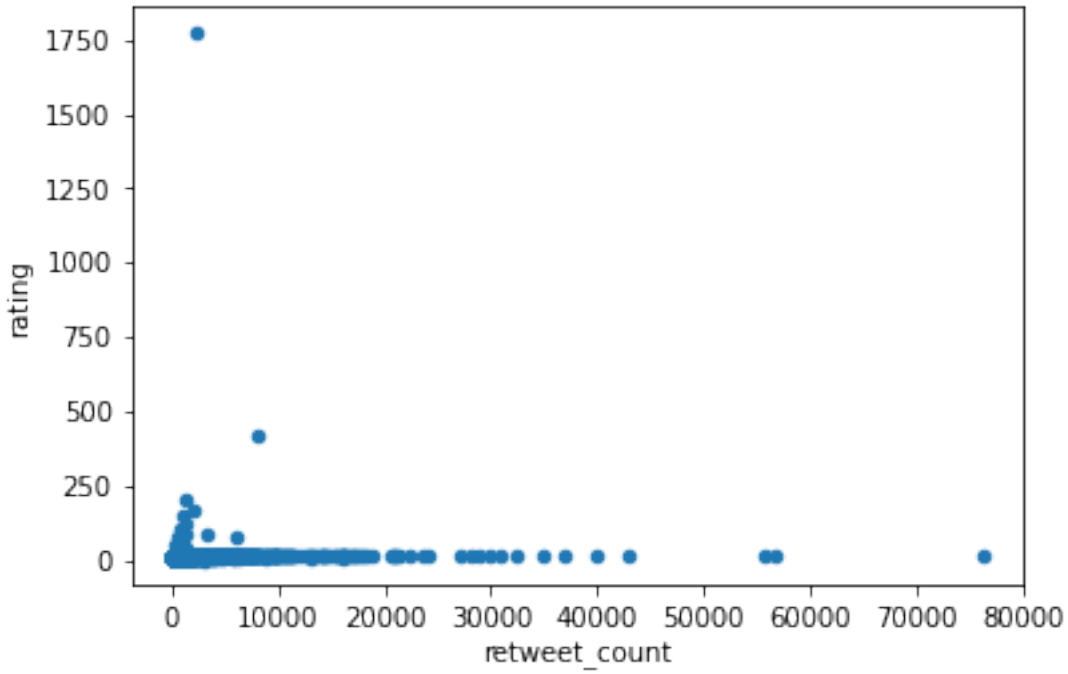


[38]:
```
analyze_df.dogs_stage.hist();
```

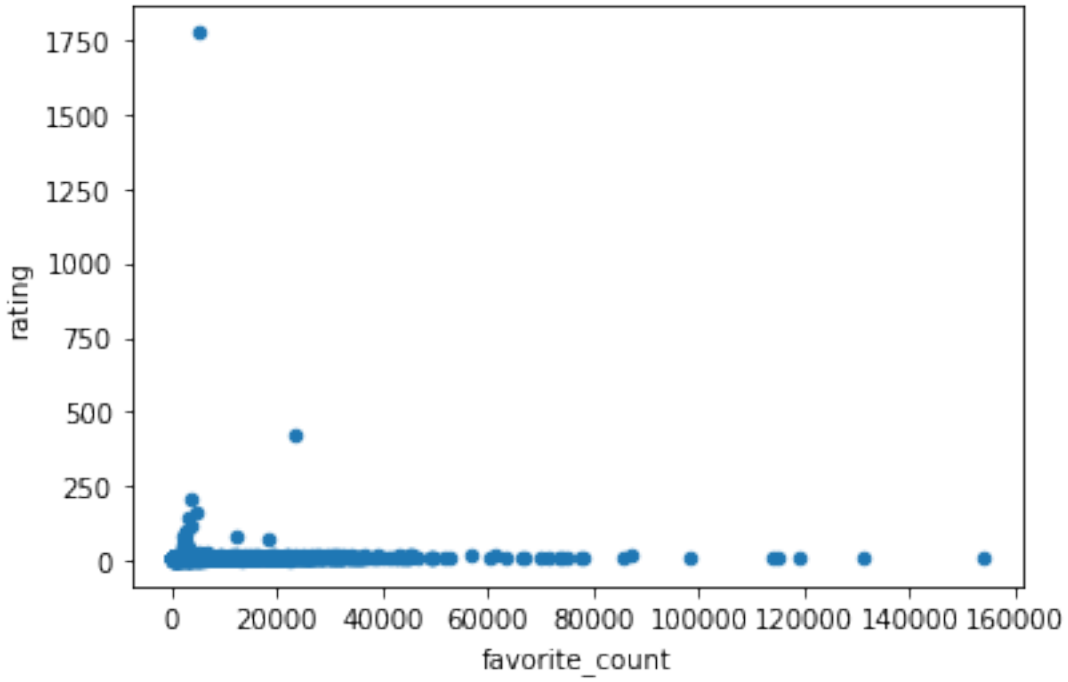


Scatterplots of various features

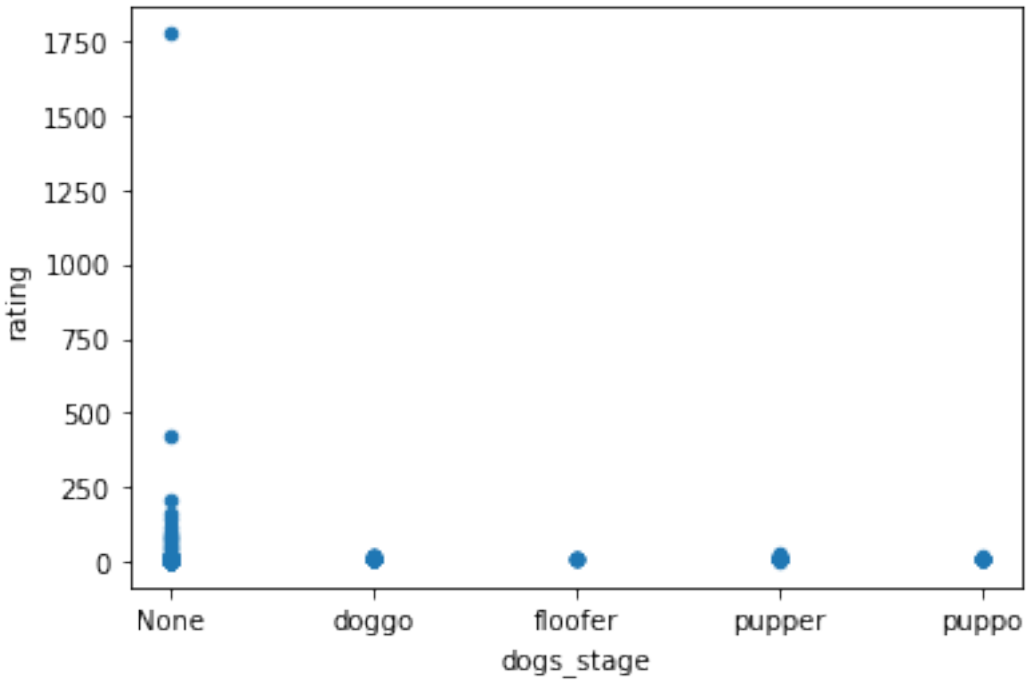
[39]:
```
analyze_df.plot(x='retweet_count',y='rating',kind='scatter');
```



[40]:
```
analyze_df.plot(x='favorite_count',y='rating',kind='scatter');
```



[41]:
```
analyze_df.plot(x='dogs_stage',y='rating',kind='scatter');
```



[42]:
```
# The mean 'ratings' for every dog stage 'dogs_stage'
analyze_df.groupby(['dogs_stage'])['rating'].mean().sort_values(ascending=False)
```

[42]:
```
dogs_stage
None       12.389840
puppo      12.041667
doggo      11.830986
floofer    11.800000
pupper     10.820961
Name: rating, dtype: float64
```

[43]:
```
# The mean 'ratings' for every dog stage 'dogs_stage'
analyze_df.groupby(['dogs_stage'])['rating'].mean().sort_values(ascending=False)
```

[43]:
```
dogs_stage
None       12.389840
puppo      12.041667
doggo      11.830986
floofer    11.800000
pupper     10.820961
Name: rating, dtype: float64
```

[44]:
```
# Display dog breeds are predicted
breeds = analyze_df[analyze_df.breed == True]
breeds.prediction.value_counts()
```

[44]:
```
golden retriever      126
Pembroke               87
Labrador retriever     87
Chihuahua              74
pug                    52
                      ...
Japanese spaniel        1
standard schnauzer      1
clumber                 1
EntleBucher             1
groenendael             1
Name: prediction, Length: 111, dtype: int64
```

[45]:
```
# Display dog breeds are not predicted
breeds = analyze_df[analyze_df.breed == False]
breeds.prediction.value_counts()
```

[45]:
```
seat belt          20
teddy              17
web site           14
dingo               8
tennis ball         8
                   ..
African grey        1
rain barrel         1
radio telescope     1
platypus            1
toilet seat         1
Name: prediction, Length: 257, dtype: int64
```

[46]:
```
# Display dog breeds are predicted with the highest 10 arithmetic mean for␣
 ↪favorite (like) counts
breeds.groupby('prediction')['favorite_count'].mean().sort_values(ascending =␣
 ↪False).head(10)
```

[46]:
```
prediction
conch         43658.0
Angora        42991.5
limousine     42967.0
fountain      41283.0
bubble        37676.5
orange        35729.0
beaver        35217.0
```

```
revolver      32166.0
barbell       32099.0
basketball    31970.0
Name: favorite_count, dtype: float64
```

[47]:
```
# Display dog breeds are predicted with the highest 10 arithmetic mean for␣
 ↪retweet counts
breeds.groupby('prediction')['retweet_count'].mean().sort_values(ascending =␣
 ↪False).head(10)
```

[47]:
```
prediction
conch             17490.0
bubble            15065.5
gondola           14504.0
Angora            14217.0
beaver            14049.0
revolver          12750.0
remote control    12539.0
quilt             10798.0
limousine         10567.0
barbell            9524.0
Name: retweet_count, dtype: float64
```

[48]:
```
# Display the arithmetic mean for dogs stage
analyze_df.groupby(['dogs_stage'])['rating'].mean().sort_values(ascending =␣
 ↪False)
```

[48]:
```
dogs_stage
None       12.389840
puppo      12.041667
doggo      11.830986
floofer    11.800000
pupper     10.820961
Name: rating, dtype: float64
```