WeRateDogs - Twitter Data

The dataset that I have wrangled (and analyzed and visualized) is the tweet archive of Twitter user **@dog_rates**, also known as **WeRateDogs**. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. 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're good dogs Brent**." WeRateDogs has over 4 million followers and has received international media coverage.

My goal: wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning is required for "*Wow!*"-worthy analyses and visualizations.

Analysis and Visualization

- The most common rating found is 11/10 i.e. 1.1 , maximum is 1776/170 and minimum is 0.

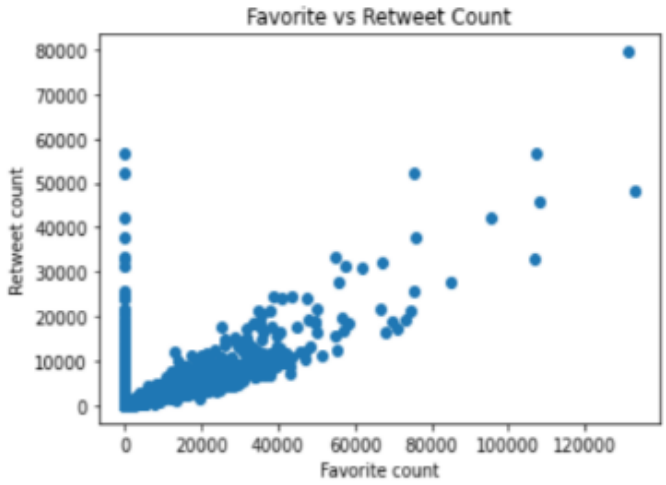
In [53]:
```
merge.describe()
```

Out[53]:

	tweet_id	retweet_count	favorite_count	rating_numerator	rating_denominator	rating
count	2.354000e+03	2354.000000	2354.000000	2354.000000	2354.00000	2354.0
mean	7.426978e+17	3164.797366	8080.968564	13.084919	10.45582	inf
std	6.852812e+16	5284.770364	11814.771334	45.889374	6.74809	NaN
min	6.660209e+17	0.000000	0.000000	0.000000	0.00000	0.0
25%	6.783975e+17	624.500000	1415.000000	10.000000	10.00000	1.0
50%	7.194596e+17	1473.500000	3603.500000	11.000000	10.00000	1.1
75%	7.993058e+17	3652.000000	10122.250000	12.000000	10.00000	1.2
max	8.924206e+17	79515.000000	132810.000000	1776.000000	170.00000	inf

- Relation between Favourite & Retweet count

```
In [54]: plt.scatter(merge['favorite_count'], merge['retweet_count'])
         plt.xlabel('Favorite count');
         plt.ylabel('Retweet count');
         plt.title('Favorite vs Retweet Count');
         plt.show()
```




```
In [55]: merge['favorite_count'].corr(merge['retweet_count'])
Out[55]: 0.7028813596145037
```

The above two results show that there is a strong correlation between the favorite count and retweet count.

- Maximum Favourite and Retweet count:

Max favorite count:

```
In [56]: fav_pop=merge.sort_values(by=['favorite_count'],ascending=False)
         fav_pop.head(10)
```

Out[56]:

	tweet_id	full_text	retweet_count	favorite_count	timestamp	text	rating_numerator	rating_denominator	name	dog_class	rating
412	822872901745569793	Here's a super supportive puppo participating ...	48265	132810	2017-01-21 18:26:02	Here's a super supportive puppo participating ...	13.0	10	NaN	puppo	1.3
1037	744234799360020481	Here's a doggo realizing you can stand in a po...	79515	131075	2016-06-18 18:26:18	Here's a doggo realizing you can stand in a po...	13.0	10	NaN	doggo	1.3
65	879415818425184262	This is Duddles. He did an attempt. 13/10 some...	45849	107956	2017-06-26 19:07:24	This is Duddles. He did an attempt. 13/10 some...	13.0	10	Duddles	NaN	1.3

Max retweet count:

```
In [57]: retweet_pop=merge.sort_values(by=['retweet_count'],ascending=False)
         retweet_pop.head(10)
```

Out[57]:

	tweet_id	full_text	retweet_count	favorite_count	timestamp	text	rating_numerator	rating_denominator	name	dog_class	rating
1037	744234799360020481	Here's a doggo realizing you can stand in a po...	79515	131075	2016-06-18 18:26:18	Here's a doggo realizing you can stand in a po...	13.0	10	NaN	doggo	1.3
259	842892208864923648	RT @dog_rates: This is Stephan. He just wants ...	56625	0	2017-03-18 00:15:37	RT @dog_rates: This is Stephan. He just wants ...	13.0	10	Stephan	NaN	1.3
533	807106840509214720	This is Stephan. He just wants to help. 13/10 ...	56625	107015	2016-12-09 06:17:20	This is Stephan. He just wants to help. 13/10 ...	13.0	10	Stephan	NaN	1.3
		RT @dog_rates:			2016-08-	RT @dog_rates:					

Has the site become popular over the years?

Has the site become more popular over the years?

```
In [ ]: merge['year'] = pd.DatetimeIndex(merge['timestamp']).year
        merge.head(2)
```

```
In [59]: x=merge.year.sort_values()
         yr=x.unique()
```

```
In [60]: re=merge.groupby(['year'])['retweet_count'].mean()
         re
```

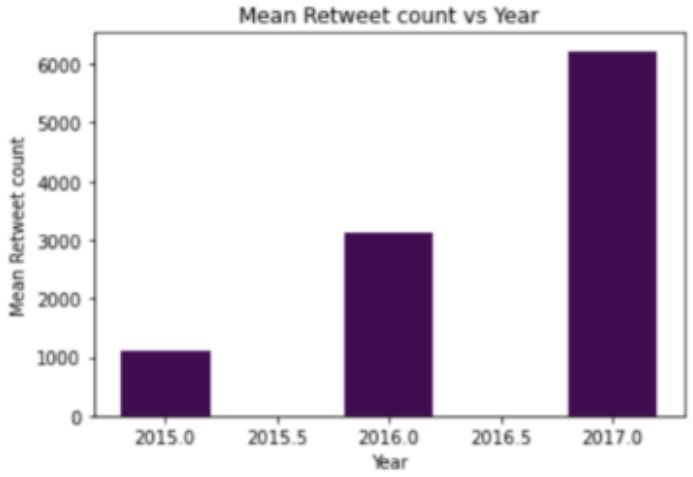
```
Out[60]: year
         2015    1097.304348
         2016    3123.785956
         2017    6225.058091
         Name: retweet_count, dtype: float64
```

```
In [61]: fav=merge.groupby(['year'])['favorite_count'].mean()
         fav
```

```
Out[61]: year
         2015     2519.078261
         2016     6997.131134
         2017    18700.885892
         Name: favorite_count, dtype: float64
```

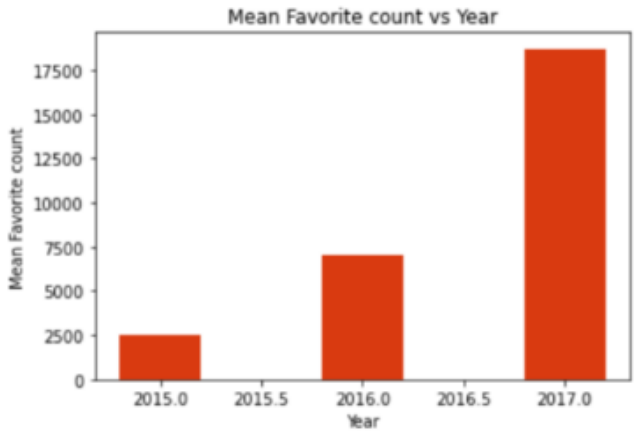
- Mean Retweet count per year

```
In [65]: plt.bar(yr,re,width=0.4,color='#410d51')
         plt.xlabel("Year")
         plt.ylabel("Mean Retweet count")
         plt.title("Mean Retweet count vs Year")

Out[65]: Text(0.5, 1.0, 'Mean Retweet count vs Year')
```



- Mean Favourite count per year

```
In [66]: plt.bar(yr,fav,width=0.4,color='#d93a10')
         plt.xlabel("Year")
         plt.ylabel("Mean Favorite count")
         plt.title("Mean Favorite count vs Year")

Out[66]: Text(0.5, 1.0, 'Mean Favorite count vs Year')
```



From the above two graphs it is clear that the page has gained a lot of popularity over the years