Act Report

We use the cleaned data to create analysis and visualizations. We start our analysis by having a look at the descriptive statistics of our numerical variables.

```
df_data.describe()
```

	rating_numerator	rating_denominator	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	1875.000000	1875.000000	1875.000000	1875.000000	1875.000000	1875.000000	1.875000e+03	1.875000e+03
mean	12.574933	10.501867	2629.077333	8827.301867	1.211200	0.596811	1.345796e-01	6.045325e-02
std	42.653153	7.025747	4666.814263	12775.316263	0.570703	0.271063	1.004998e-01	5.112660e-02
min	0.000000	2.000000	11.000000	75.000000	1.000000	0.044333	1.011300e-08	1.740170e-10
25%	10.000000	10.000000	575.000000	1957.000000	1.000000	0.367430	5.370120e-02	1.624950e-02
50%	11.000000	10.000000	1280.000000	4002.000000	1.000000	0.594467	1.175660e-01	4.934910e-02
75%	12.000000	10.000000	2995.500000	11063.500000	1.000000	0.848438	1.954815e-01	9.281770e-02
max	1776.000000	170.000000	80729.000000	161154.000000	4.000000	1.000000	4.676780e-01	2.734190e-01

From the above statistics results we note the following

- The average retweet count of the tweets was 2629 and the maximum retweet count is 80729
- The average favorite count of the tweets was 8827 and the maximum favorite count is 161154
- Rating has an outlier 1776

These notes lead us to ask the following:

- What kind of dog that has this outlier rating?
- Does this outlier rating has maximum retweet count and favorite count?
- What is the highest frequency of rating?
- Is there a correlation between retweet count and favorite count?
- Does the number of retweets and favorites increase by time at WeRateDog account?

Let's find the tweet with the outlier rating

```
outlier_record = df_data.query('rating_numerator == 1776')
```

The tweet that has the outlier rating has the below information:

```
This is Atticus. He's quite simply America af. 1776/10
https://t.co/GRXwMxLBkh
https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg
Retweet count 2545
Favorite count 5280
```

The tweet with the outlier rating doesn't have the maximum retweet and favorite counts.

```
max_fav_cnt = df_data.favorite_count.max()
favorite_dog = df_data.query('favorite_count == ' + str(max_fav_cnt))
idx = favorite_dog.index[0]
```

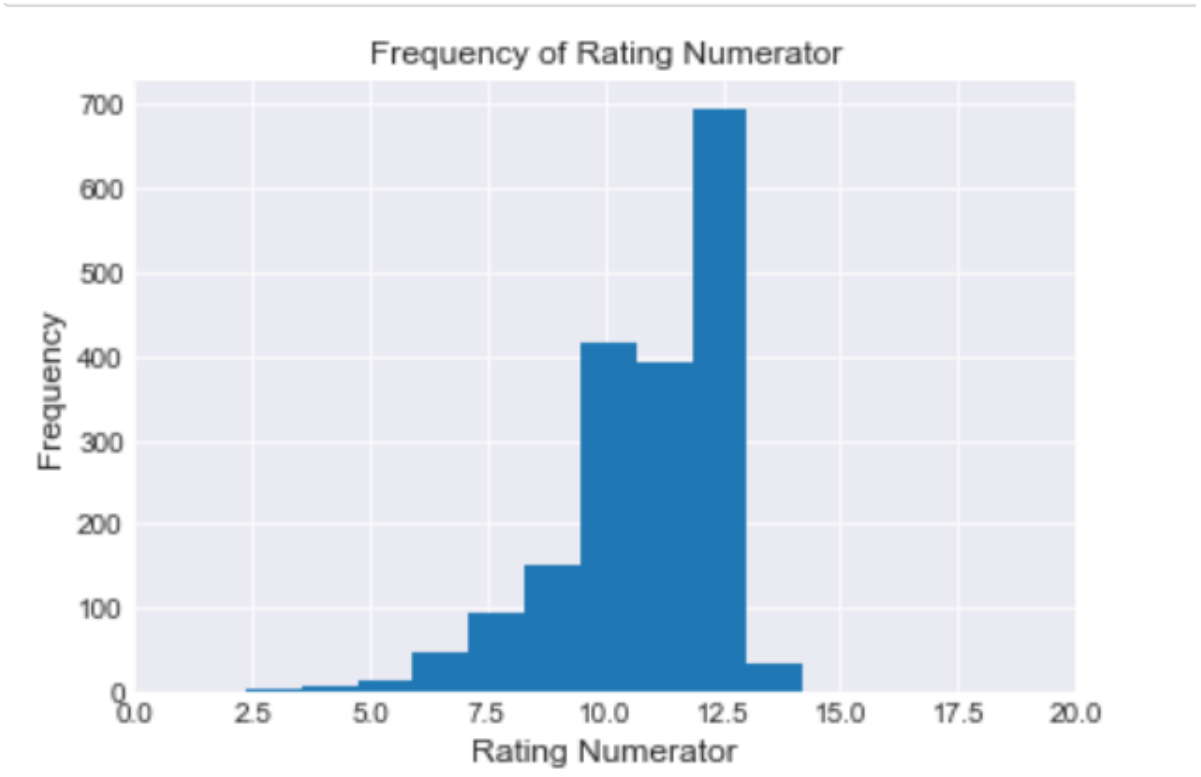
Output:

```
Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid
by Tina Conrad) https://t.co/7wE9LTEXC4
The dog type of the most favorite is doggo
The most favorite dog has 80729 tweets
```

As we see the most favorite tweet has the highest retweet count.

```
df_data.rating_numerator.value_counts()

plt.hist(df_data.rating_numerator,bins=1500)
plt.title('Frequency of Rating Numerator')
#We are able to set x-axis limits by using resutls from
value_counts() method
plt.xlim(0,20)
plt.xlabel('Rating Numerator', fontsize=12)
plt.ylabel('Frequency',fontsize=12)
plt.show()
```



From the above chart, it looks like the highest frequency rating are from 7 - 14

```
x = df_data['favorite_count']
y = df_data['retweet_count']

df_data.plot.scatter('favorite_count','retweet_count',
figsize=(8,6), marker='+',c="black",s=50,alpha=0.6,linewidth=1)
```

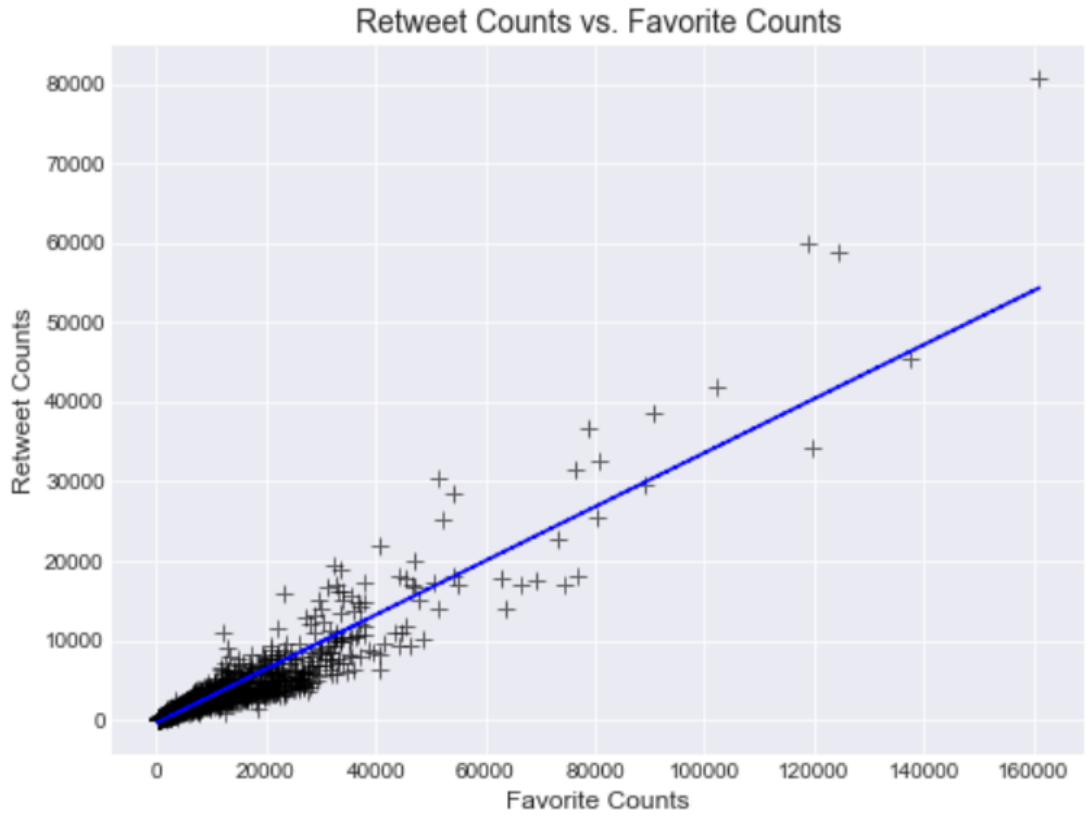
```
z = np.polyfit(x, y,1)
p = np.poly1d(z)

plb.plot(x, p(x), 'b--')


plt.title("Retweet Counts vs. Favorite Counts", fontsize=14)
plt.xlabel("Favorite Counts", fontsize=12)
plt.ylabel("Retweet Counts", fontsize=12)

correlation_r = np.around(x.corr(y), decimals=2)

plt.show()
print("Correlation: " + str(correlation_r))
```



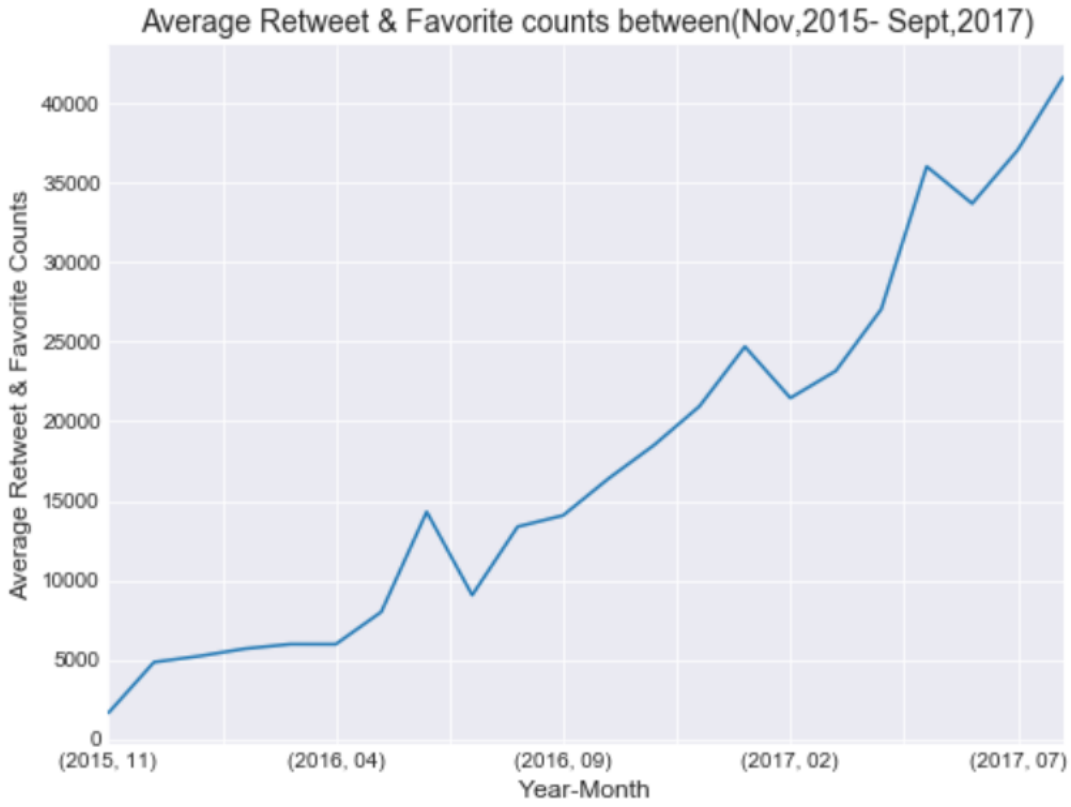

```
Correlation: 0.93
```

There is a strong positive relationship between retweets and favorites. Therefore, we are going to sum retweets and favorite counts for each tweet and use this summation to check retweets and favorite behavior during time.

```
df_data['sum_retweet_fav'] = pd.Series(list(range(len(df_data))))

for idx, row in df_data.iterrows():
    df_data.loc[idx,'sum_retweet_fav'] = row.retweet_count +
row.favorite_count
```

```
temp_df = df_data.groupby([ 'year','month']).sum_retweet_fav.mean()
temp_df.plot(kind='line',figsize=(8,6))
plt.title("Average Retweet & Favorite counts between(Nov,2015-
Sept,2017)", fontsize=14)
plt.xlabel("Year-Month", fontsize=12)
plt.ylabel("Average Retweet & Favorite Counts", fontsize=12)
plt.show()
```



The line chart displays that retweet and favorite counts of tweets increase by time. This gives us an insight that WeRateDog popularity increase by time.