Visualization and Analysis Report

After I gather all the data, assessed it and cleaned it I went on to analyze and visualize it to gain some insights from the data.

First I used the describe() method and I discovered that-: a. 75% of the dogs received a rating of 12 and below, and b. The most favorited post received 165 031 likes.

	rating_numerator	rating_denominator	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count	followers_count
count	2068.000000	2068.000000	2068.000000	2.068000e+03	2.068000e+03	2068.000000	2068.000000	2.068000e+03
mean	12.263056	10.513056	0.594944	1.345462e-01	6.028703e-02	2824.509671	8462.479207	7.418640e+06
std	40.749075	7.189152	0.271201	1.007553e-01	5.094828e-02	4909.535881	12736.435294	2.283842e+02
min	0.000000	2.000000	0.044333	1.011300e-08	1.740170e-10	12.000000	0.000000	7.418512e+06
25%	10.000000	10.000000	0.364570	5.352722e-02	1.616932e-02	601.750000	1612.000000	7.418540e+06
50%	11.000000	10.000000	0.588620	1.181350e-01	4.933745e-02	1329.500000	3727.500000	7.418628e+06
75%	12.000000	10.000000	0.845599	1.955617e-01	9.198323e-02	3262.250000	10604.750000	7.418686e+06
max	1776.000000	170.000000	1.000000	4.880140e-01	2.734190e-01	84221.000000	165031.000000	7.420830e+06

Secondly I used the value_counts() method to find the most common dog name:

```
frequent_names = df_combined['name'].value_counts().nlargest(10)
frequent_names
```

```
Tucker     10
Penny      10
Lucy       10
Charlie    10
Cooper     10
Oliver     10
Winston     8
Sadie       8
Lola        8
Bo          8
```

The result came out with six dog names with 10 appearances each.

Next I visualized the favourite and retweet counts for each dog stage, and the chart shows that `puppo` were favorited a lot. However this is because the total count of `puppo` is greater than any other dog stage. Another noticeable thing , is that favorites are always greater than retweets for any dog stage.

4. Favourites and retweets for each dog stage

```
categories =['pupper','doggo','puppo','floofer']
retweet_count_average=df_combined.groupby('stage')[['retweet_count', 'favorite_count']].mean()
ax = retweet_count_average.loc[categories].plot(kind="bar")
plt.xticks(rotation =360)
plt.title('The Number of Favorites and Retweets at Each Stage')
```

```
Text(0.5, 1.0, 'The Number of Favorites and Retweets at Each Stage')
```