Project #4 Act Report

Data Wrangling **(WeRateDogs.)**

Introduction:

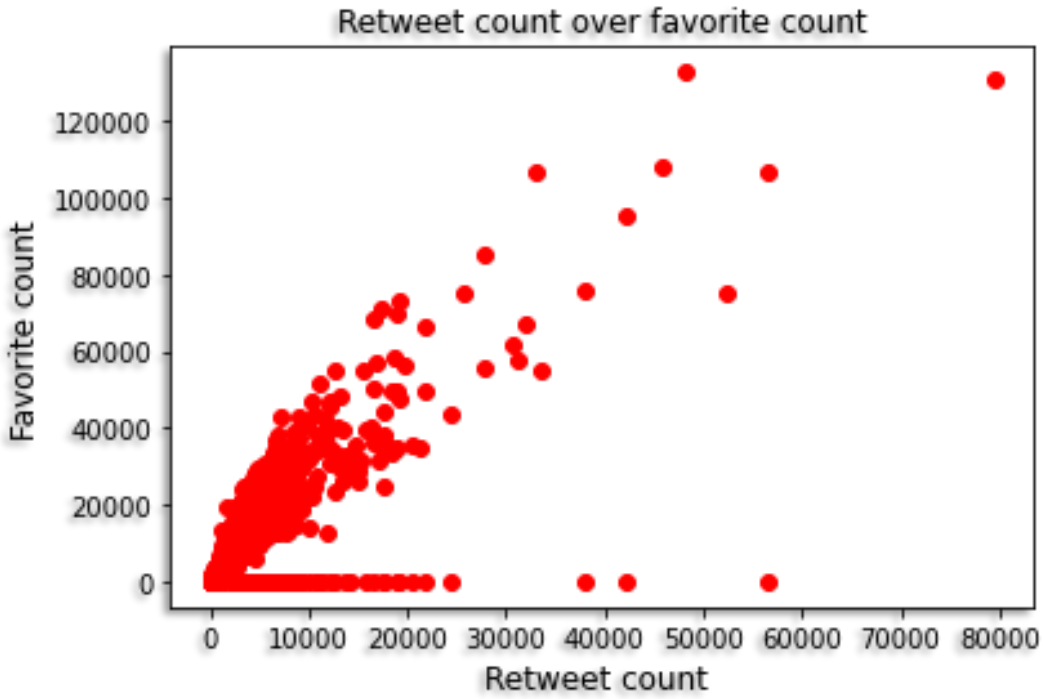
Wrangling and assessing data of WeRateDogs has reached us to get knowledge and information from the data, and also we found out interesting and trustworthy insights and visualization, the visuals shown in this report are in specified columns which are:

- retweet count
- favorite count
- stages
- timestamp
- rating.

In this report, I will illustrate the questions that was on my head with its extracted answers from the analysis and visualization.

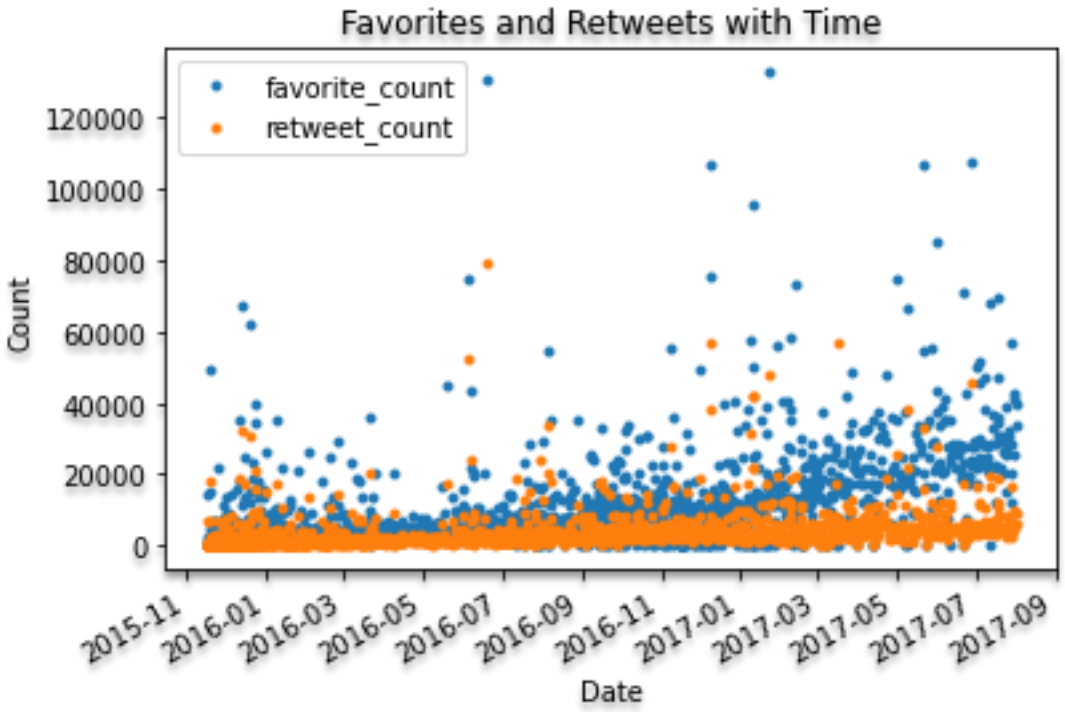
Q1. Does retweet_count over favorites_count?

Answer: As a result shown from below plot, we can figure out that retweets_counts and favorite_counts affect on each other positively and they have strong positive correlation, that means, more retweets leads to more favorites and vice versa. Also there is overlap between (0,0) to (15000,5000).



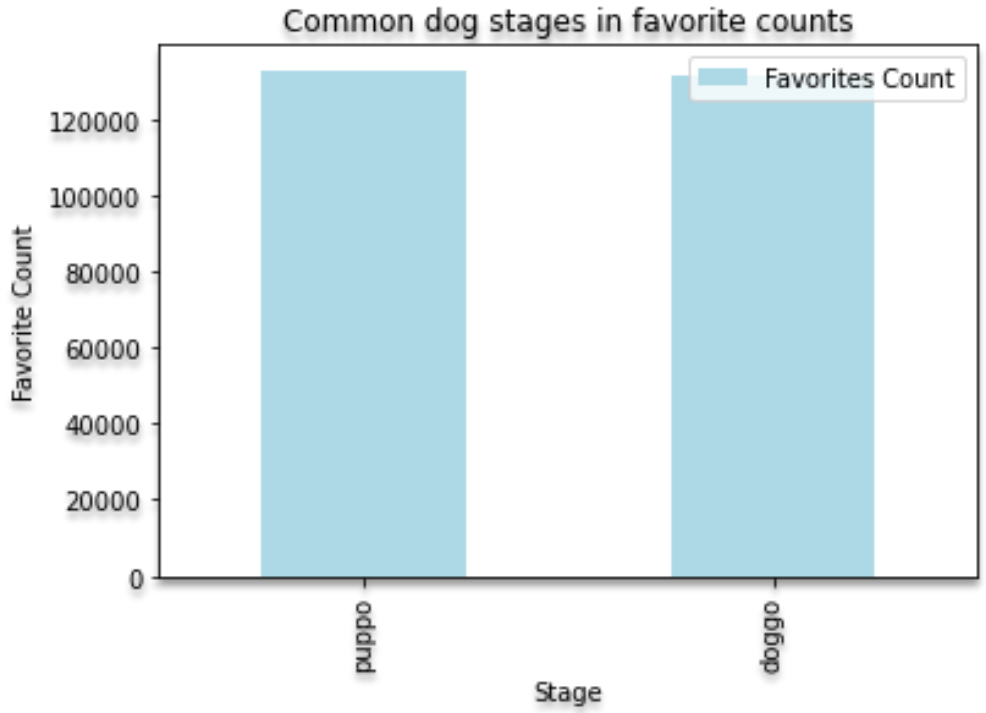
Q2. Does the time affect on favorite_count and retweet_count?

Answer: From the below plot, we can find that all of the three features has positive relationship (it increases with time). Most of the dots are very near to each other, so the favorite_count and retweet_count have strong positive correlation.



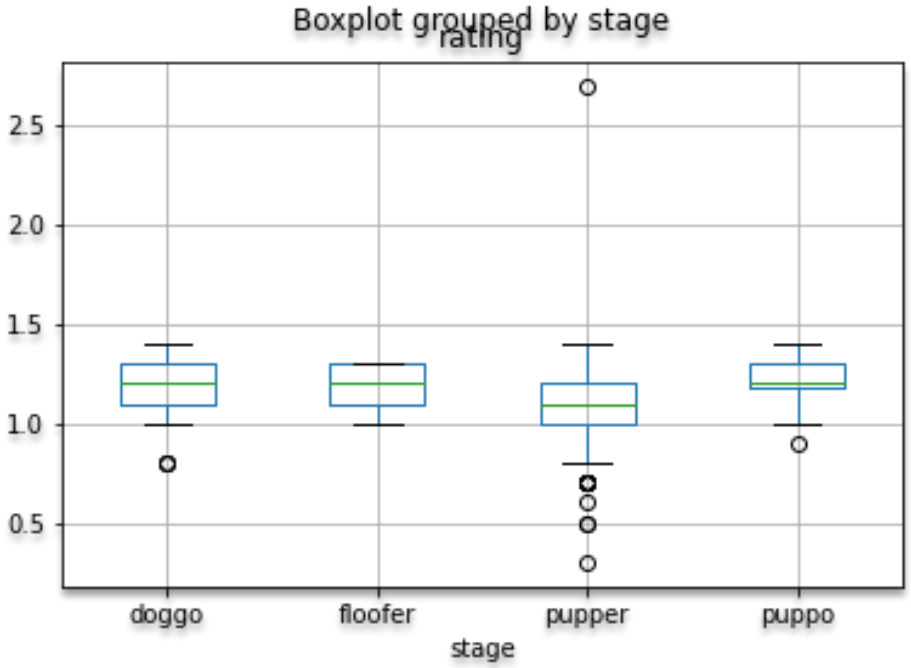
Q3. What are the most common or popular dog stages?

Answer: From the below plot, we can figure that Puppo and Duggo are the two most common dog stages based on favorite counts.



Q4. Which dogs get lowest rating?

Answer: The boxplot above shows the ratings and dogs stages, and we can find that 'puppers' gets lowest rating comparing to other digs, also their is many outliers that affects its mean negatively. Also the description of each dog stage ensures that.




```
In [39]: twitter_df.groupby('stage')['rating'].describe()
```

Out[39]:

stage	count	mean	std	min	25%	50%	75%	max
doggo	67.0	1.197015	0.147679	0.8	1.100	1.2	1.3	1.4
floofer	8.0	1.187500	0.112599	1.0	1.100	1.2	1.3	1.3
pupper	221.0	1.077376	0.206781	0.3	1.000	1.1	1.2	2.7
puppo	24.0	1.204167	0.126763	0.9	1.175	1.2	1.3	1.4