Analysis and Insights Report

As a part of the analysis, I kept my focus on understanding how dog ratings, post retweets and likes are influenced by the information the tweets provide on the dogs.

The three questions I tried to answer were:

1. Does information of the dog_stage/name influence the retweet count and favourite count?
2. Is there any correlation between ratings, retweets and favourite counts?
3. Which is the most popular dog_stage (rating_numerator) comparison?¶

Does information of the dog_stage/name influence the retweet count and favourite count?

For answering the first question, I simply used pandas aggregate function - mean.

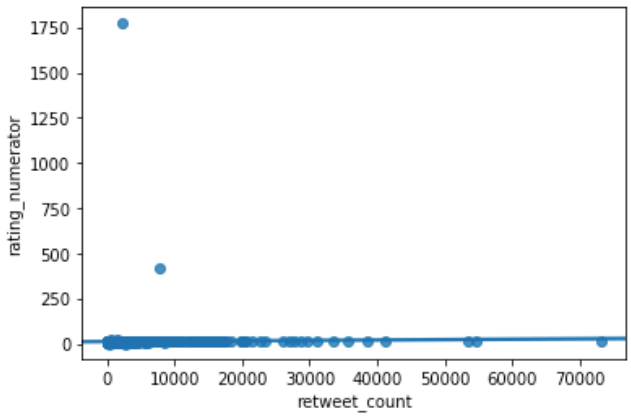
dog_stage	retweet_count	favorite_count
	2218.984650	7613.115975
doggo	5704.845070	16779.464789
doggo,floofer	2895.000000	15247.000000
doggo,pupper	3455.400000	11820.100000
doggo,puppo	16588.000000	42940.000000
floofer	3443.666667	9658.222222
pupper	2062.725225	6381.391892
puppo	5596.260870	20064.086957

dog_has_name	retweet_count	favorite_count
False	2652.212766	7843.898527
True	2262.225589	8034.778451

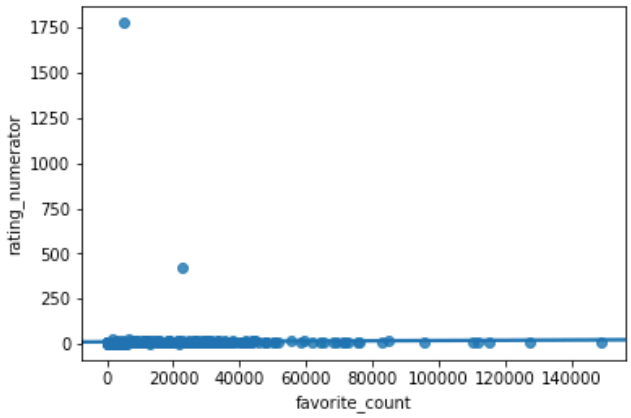
Conclusion: We can conclude that having a dog name or stage mentioned in the tweet does not particularly influence retweets and favourite counts.

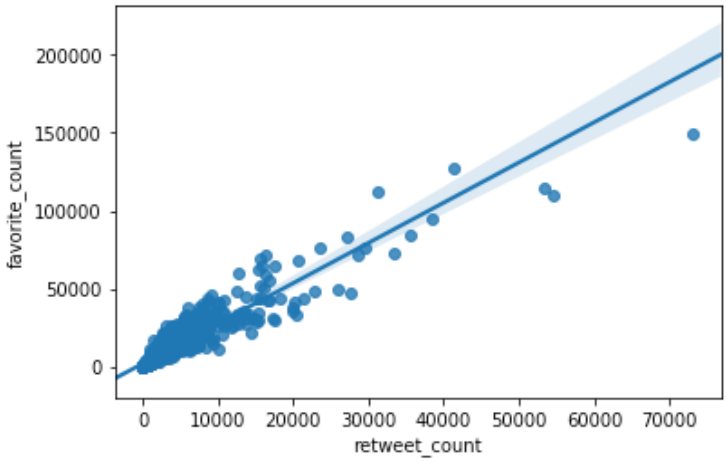
Is there any correlation between ratings, retweets and favourite counts?

For this, I used the seaborn library's scatter plot to plot the correlation between ratings and retweets, ratings and favorite count and retweet and favorite count.




```
sns.regplot(x = "favorite_count", y = "rating_nume
plt.show()
```

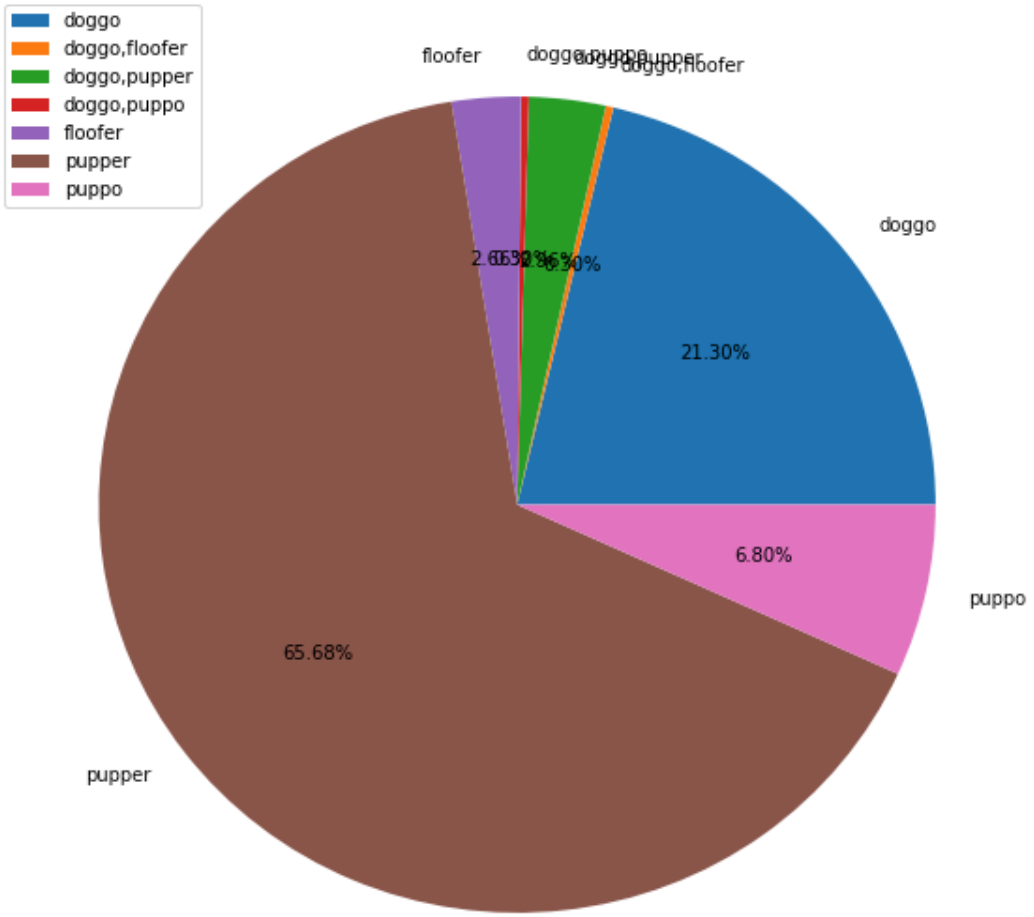




Conclusion: Ratings for dogs do not influence retweets or favourites. But there is a direct correlation between retweets and favourites which is a likely outcome as people usually retweet what they like (favourite)

Which is the most popular dog_stage (rating_numerator) comparison?¶

For this, I used the seaborn library's pie plot to plot the pie chart depicting which dog_stage is the most rated.



Conclusion: Puppers FTW!!¶