

REPORT

Data Analysis: Insights and Conclusions

By: Nacer KROUDIR

2023

Introduction:

The data used for this project is the tweet history of the Twitter user *@dog_rates*, also known as **WeRateDogs**. **WeRateDogs** is a Twitter account that evaluates dogs and adds a humorous comment. Once issues have been identified and handled programmatically, we will analyze the refined data and present our findings through visualization.

Data Analysis:

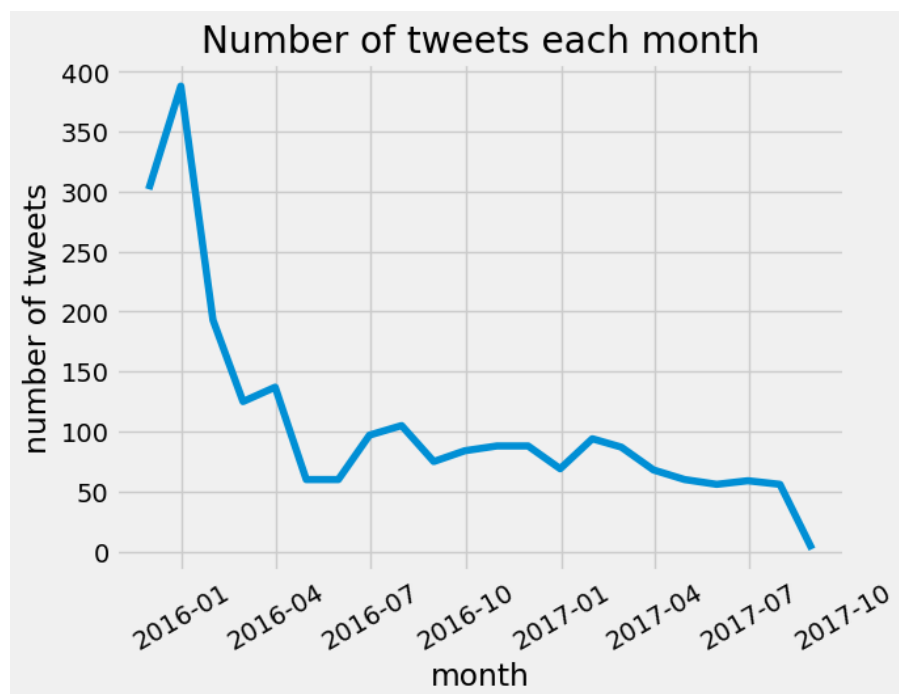
Questions to address

- Has the number of tweets increase or decrease over time?
- What are the most common dog stages in WeRateDogs tweets?
- What are the top 20 dog breeds with the highest average rating by WeRateDogs?
- What are the top 15 dog breeds mentioned in the dataset?
- What are the most popular dog names in the tweets?
- Is there a relationship between favorite count and retweet count?
- What are the most repeated words in the tweets?

The analysis

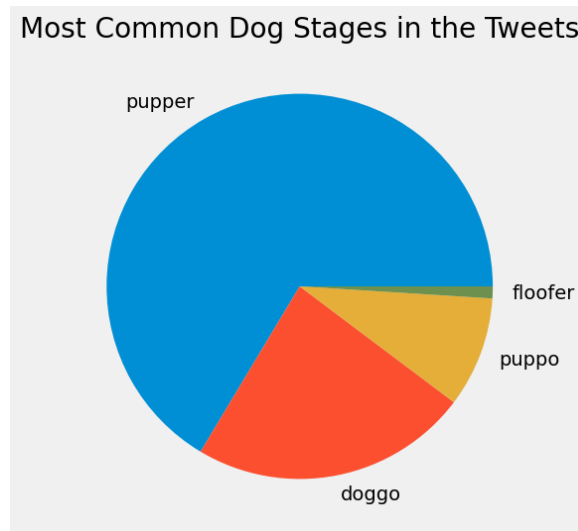
Has the number of tweets increase or decrease over time?

- The number of tweets decreased drastically with time. falling from over 300 tweet a month to around 60 a month.
- The peak of number of tweets by month is 388 on December, 2015.
- The biggest decrease in the number of tweets in two consecutive months is on December, 2015 to January, 2016; with a fall of 195 tweets.



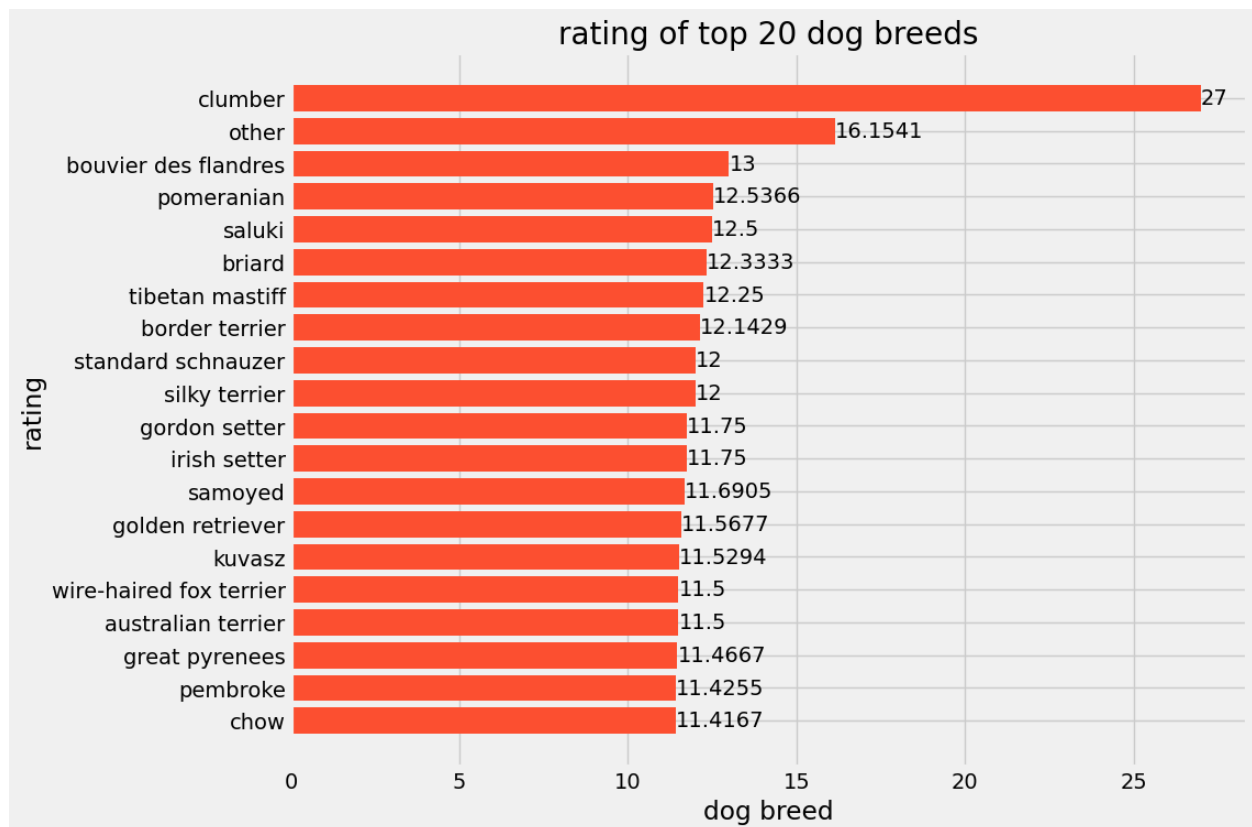
What are the most common dog stages in WeRateDogs tweets?

- The most common dog stage is pupper with over 65%; followed by doggo.
- The least mentioned dog stage is floofer.



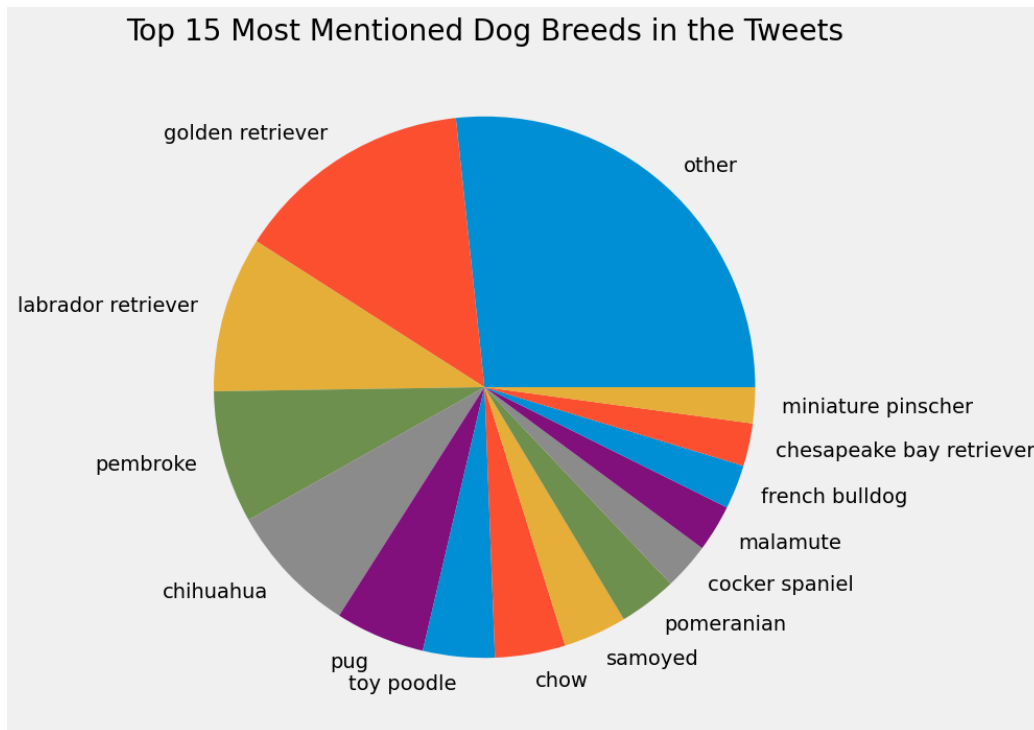
What are the top 20 dog breeds with the highest average rating by WeRateDogs?

- The highest rated dog breed is clumber (That might be due to its appearance once in the dataset).



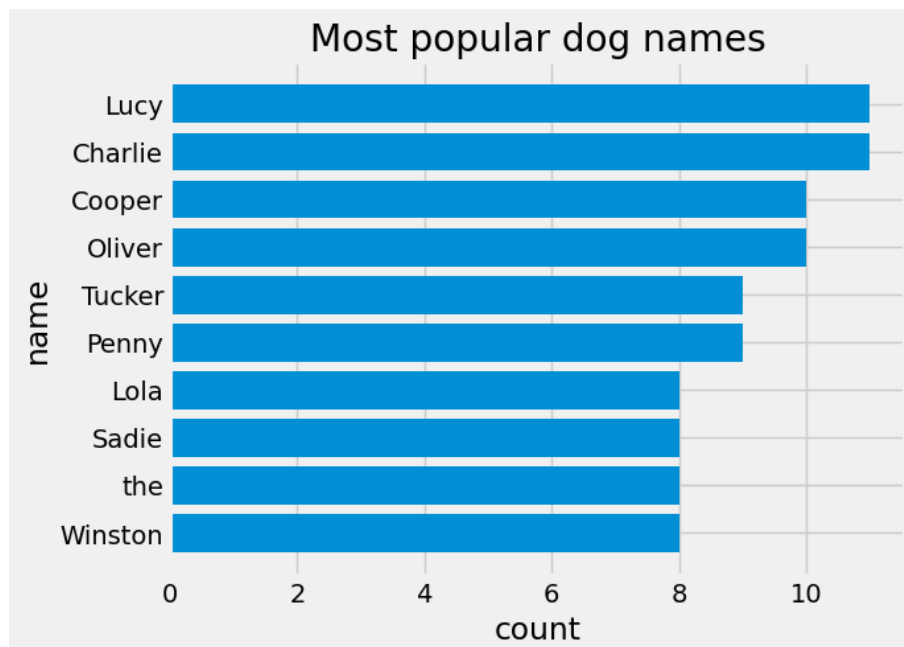
What are the top 15 dog breeds mentioned in the dataset?

- Golden retriever and labrador retriever are the most common dog breeds in the tweets.
- Pembroke and Chihuahua come in third and fourth place respectively.



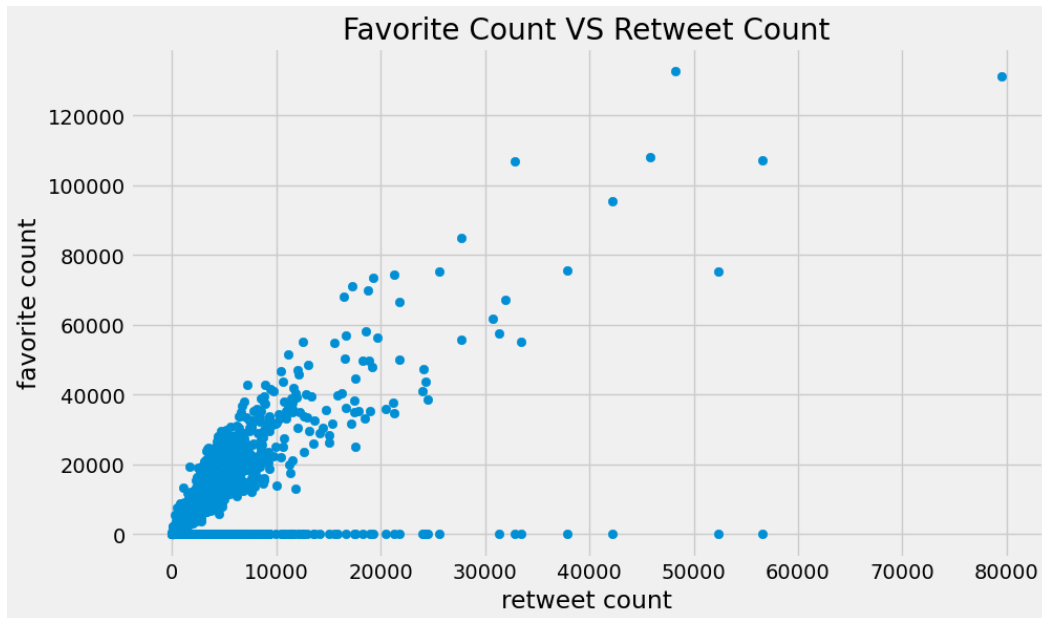
What are the most popular dog names in the tweets?

- The most common dog name are Lucy and Charlie with 11 dogs holding that name.
- Oliver and Cooper come next with 10 mentions each.



Is there a relationship between favorite count and retweet count?

- There is a positive linear correlation between favorite_count and retweet_count.



What are the most repeated words in the tweets?



Conclusion

The insights gathered in the analysis are interesting. Further analysis can be done on this dataset using other advanced techniques.