# REPORT
WERATEDOGS

Data Source:

The data was gathered from a given CSV, a website, and Twitter's API. Tweepy was used to access the API and gather the JSON data for the tweets. The JSON data was stored in a text file, then loaded what is required into a pandas data frame.

Process:

- The quality and tidiness issues were identified.
- A copy was created for each dataset before cleaning.
- The issues were fixed and data was cleaned.
- The cleaned data was saved.
- The dataset was explored using data visualization.
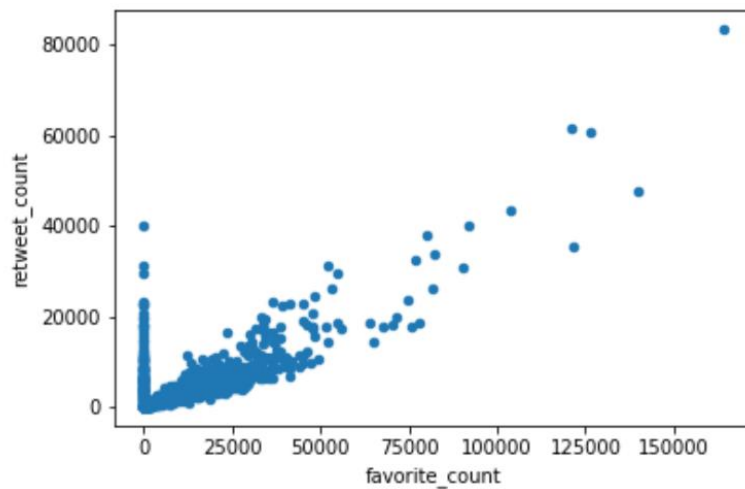
Data visualization:

- A bar plot was plotted to find the number of dogs that belongs to each stage.
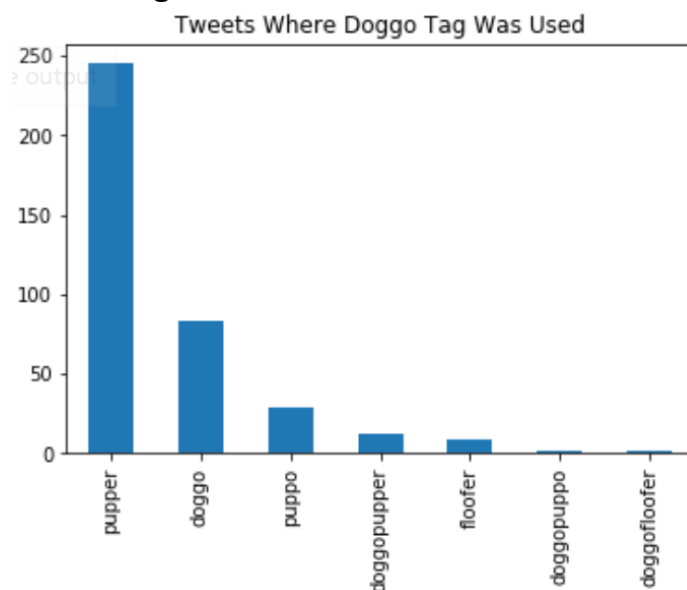
Inference:
Pupper have the highest count in the data and Doggo & Floofer the least.
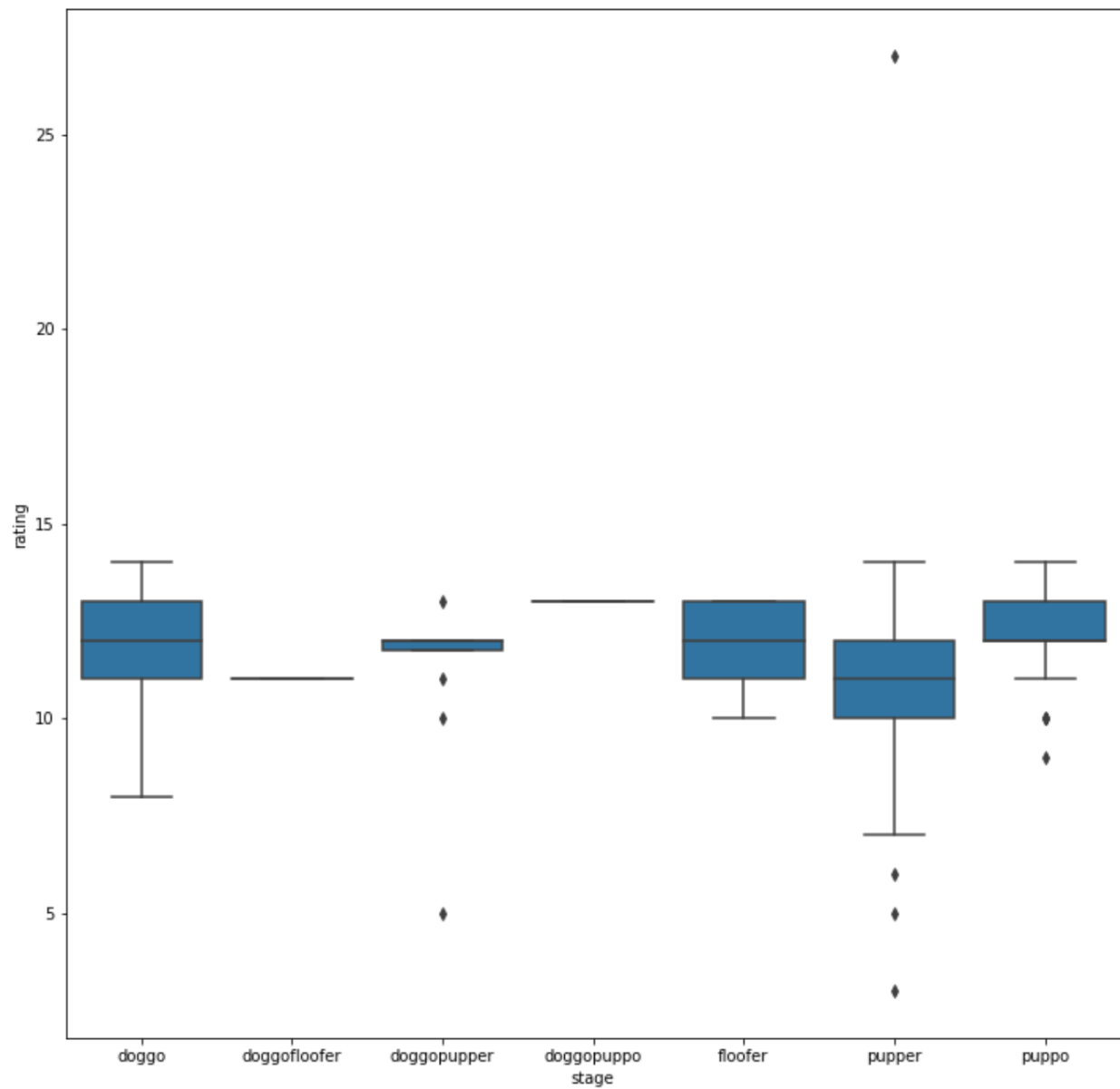
- A scatter plot was plotted to find the correlation between favorite count and retweet count.

Inference:
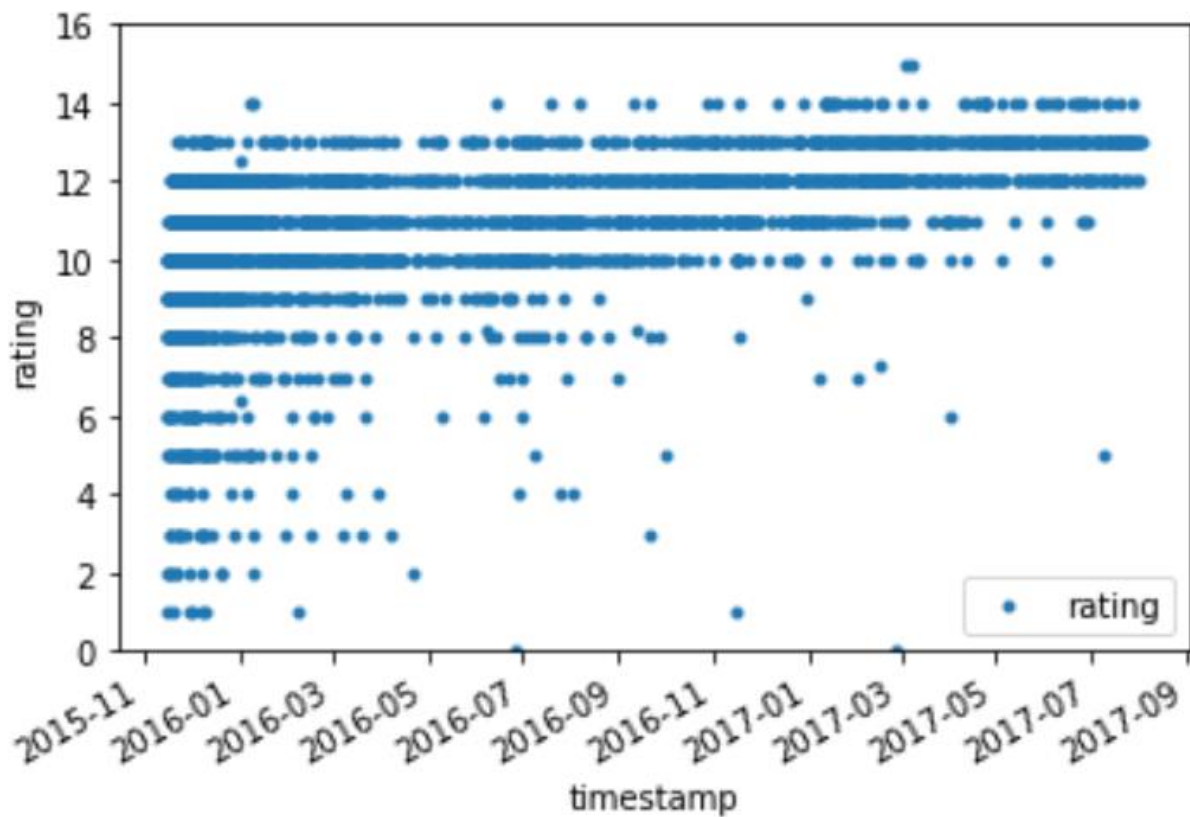retweet_count and favorite_count are positively correlated.

- A boxplot was plotted between rating and dog stage.

Inference:
The overall rating is almost in the same range for all stages with pupper in lower end and doggo & puppo in higher end.

- Rating vs timestamp was plotted to view the change in rating density over time.

Inference:
One could observe that mostly, irrespective of the time the rating given is 12.

Conclusion:

- The required data were gathered.
- The datasets were merged appropriately.
- Tidiness and quality issues were identified, and data was then cleaned.
- The data was analyzed and visualized with following inferences:
    1. Pupper have the highest count in the data and Doggo & Floofer the least.
    2. retweet_count and favorite_count are positively correlated.
    3. The overall rating is almost in the same range for all stages with pupper in lower end and doggo & puppo in higher end.

4. *One could observe that mostly, irrespective of the time the rating given is 12.*