Wrangling and Analyze Data - insights and visualization report

The objective of this project is to gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then cleaning it using Python and its libraries.

In this project, We will wrangle the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

After gathering, assessing, and cleaning the three datasets as described in wrangle_report.html, I have generated six insights, three of them visualized using python matplotlib library.

Insights:

what is the most retweeted tweet?

The most retweeted tweet was tweet with id 744234799360020481.

what is the tweet that has the highest rating?

The highest rating of dog's is 165/10. it was tweet with id 758467244762497024.

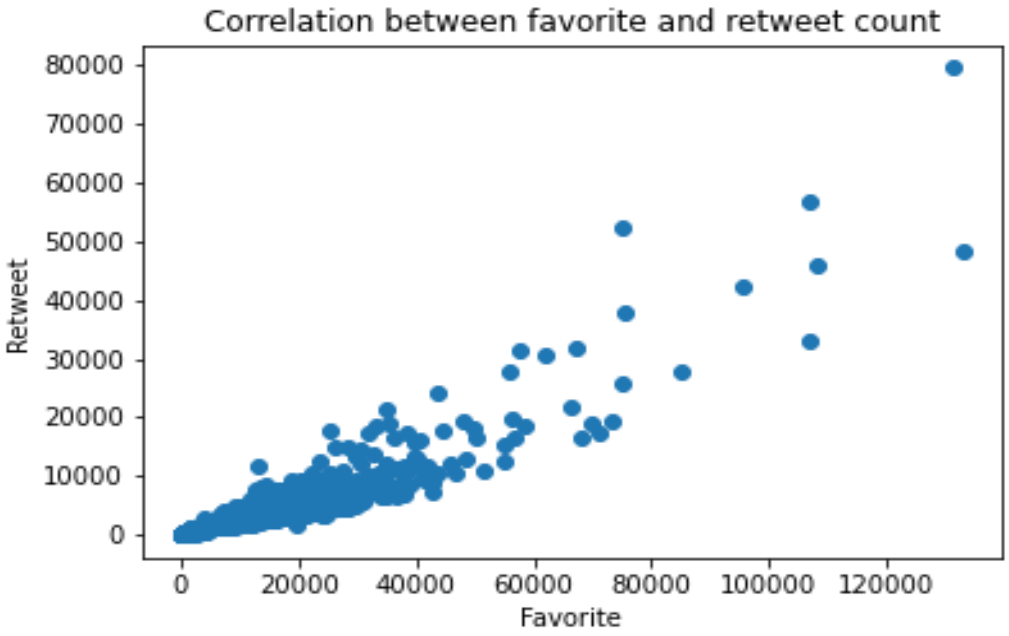
what kind of source is the most used one?

The most used source is Twitter for iPhone.

- Twitter for iPhone 1655
- Twitter Web Client 22
- TweetDeck 9

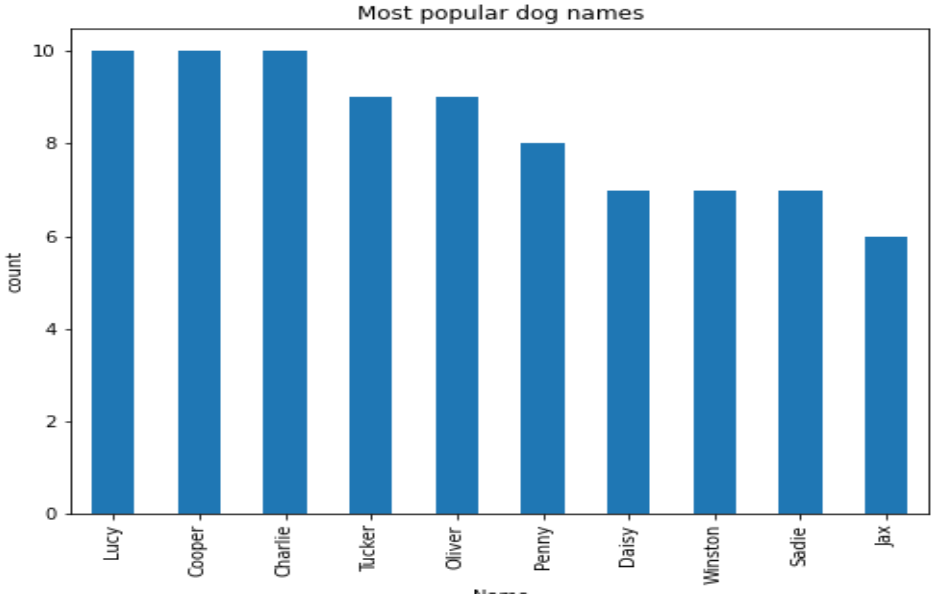
Is there any correlation between favorites and retweet count?

using scatter plot to discover if there is any correlation between favorites and retweets.



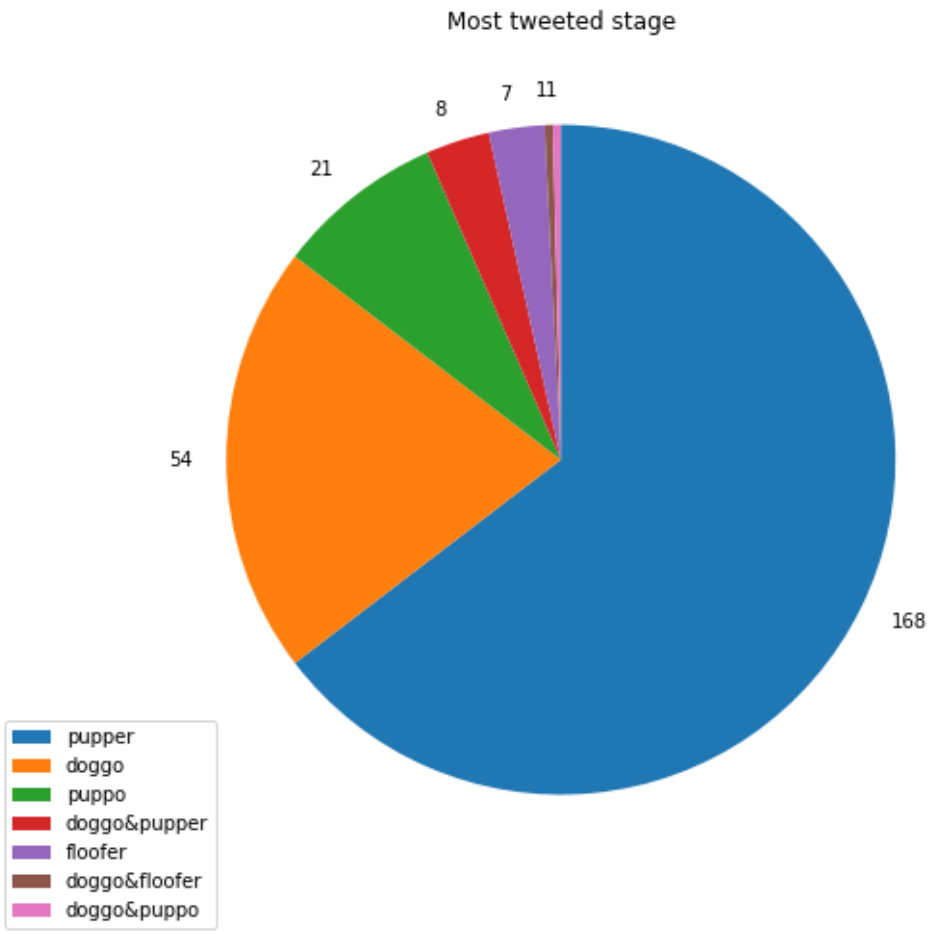
And we can see from the chart that there is positive correlation between favorite and retweet counts, which means when the retweet counts increase the favorites increases too.

what is top 10 most popular dog name?



we can see from the chart above the top 10 most popular dog name in WeRateDogs account twitter archive.

which dog stage have been tweeted about the most?



we can see that the most tweeted dog stage is pupper.