Introduction

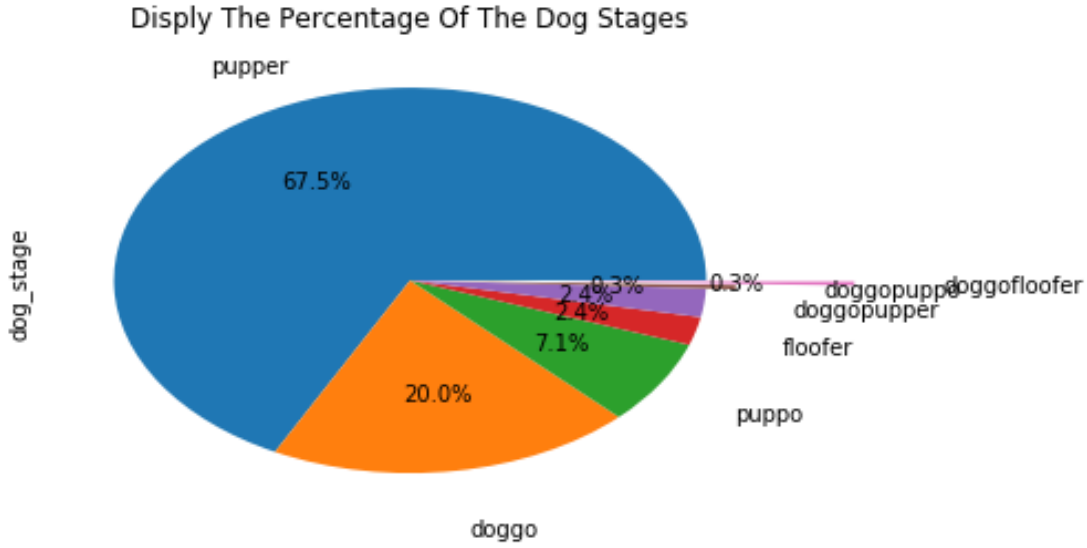
The objective of this project is to put in practice what I learned in data wrangling section from Udacity Data Analysis Nanodegree program. I use WeRateDogs dataset from user @dog_rates archive in tweeter.

WeRateDogs is a popular Twitter account, and it has a 9 million flowers, tweeters rate dogs to show how lovely the dog is.

After I cleaned data and wrangling this data, now I can make some data analysis and visualizations, and summarize some results.

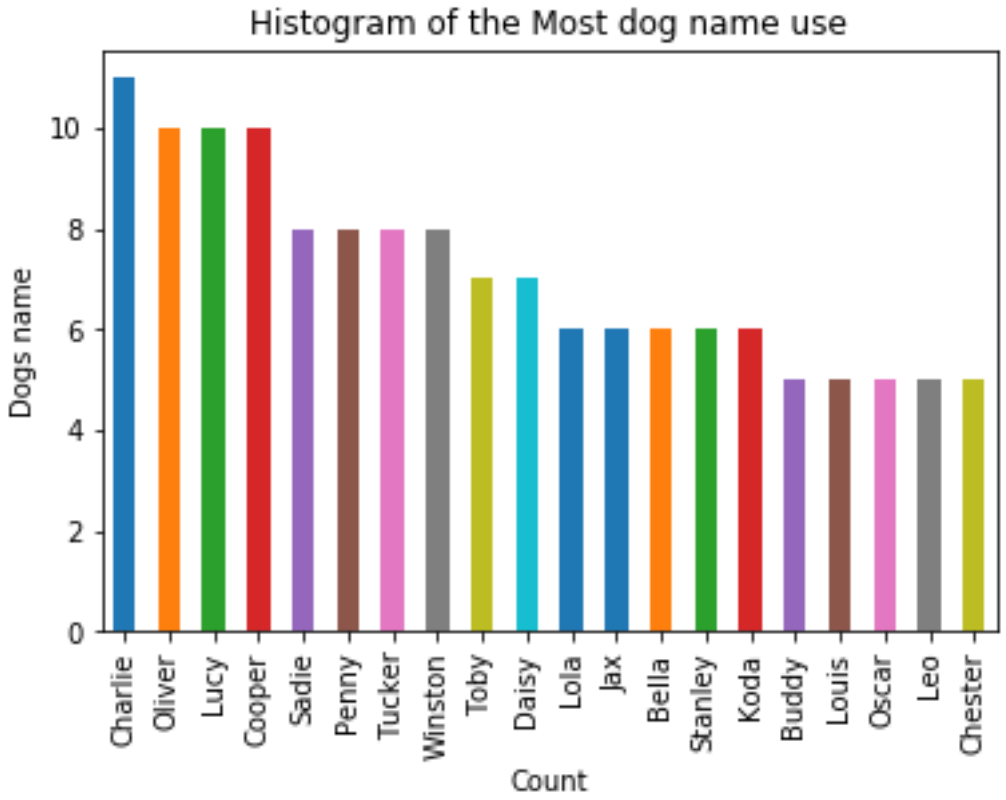
Q: There are many stages of dogs: pupper, doggo, puppo, floofer, DoggoPupper, DoggoFloofer and DoggoPuppo. So which of this stages have the most frequency?

we can see the answer to this question in pie chart below:



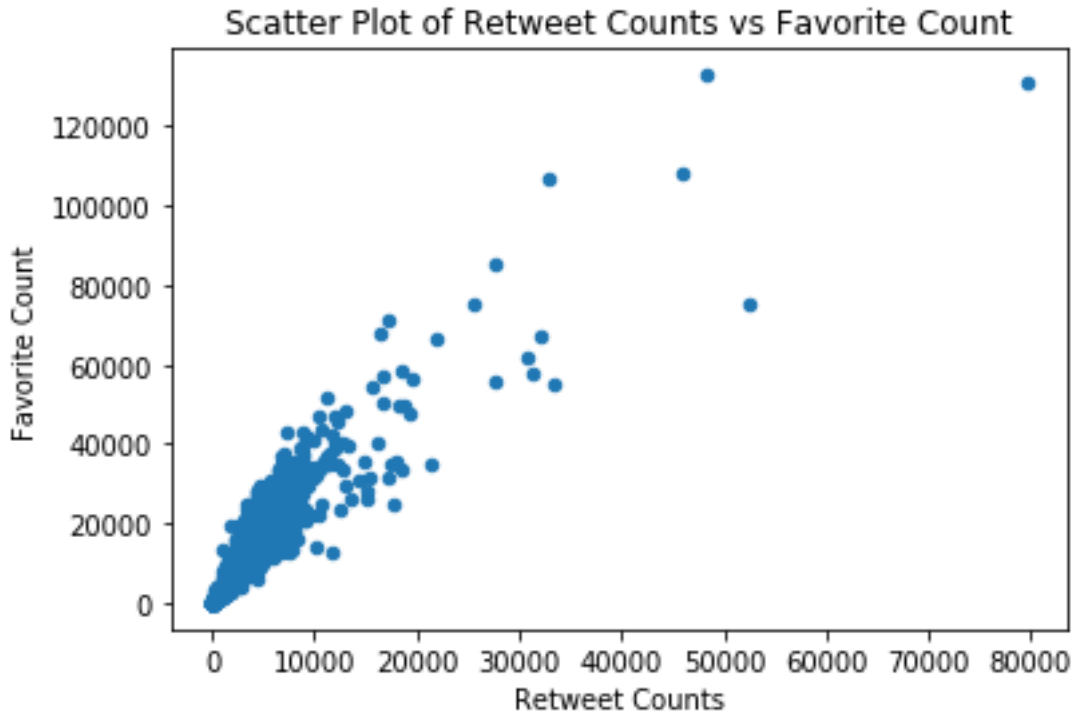
Pupper has the most frequency.

Q: In this dataset there is a lot of dog's names. What is the most names people like to call their dogs?



From the above picture the most name people like to call their dogs is Charlie then Oliver, Lucy and Cooper.

Q: Is there a relationship between favorite count and retweet count? and how this relation affects in both retweet and favorite counts?

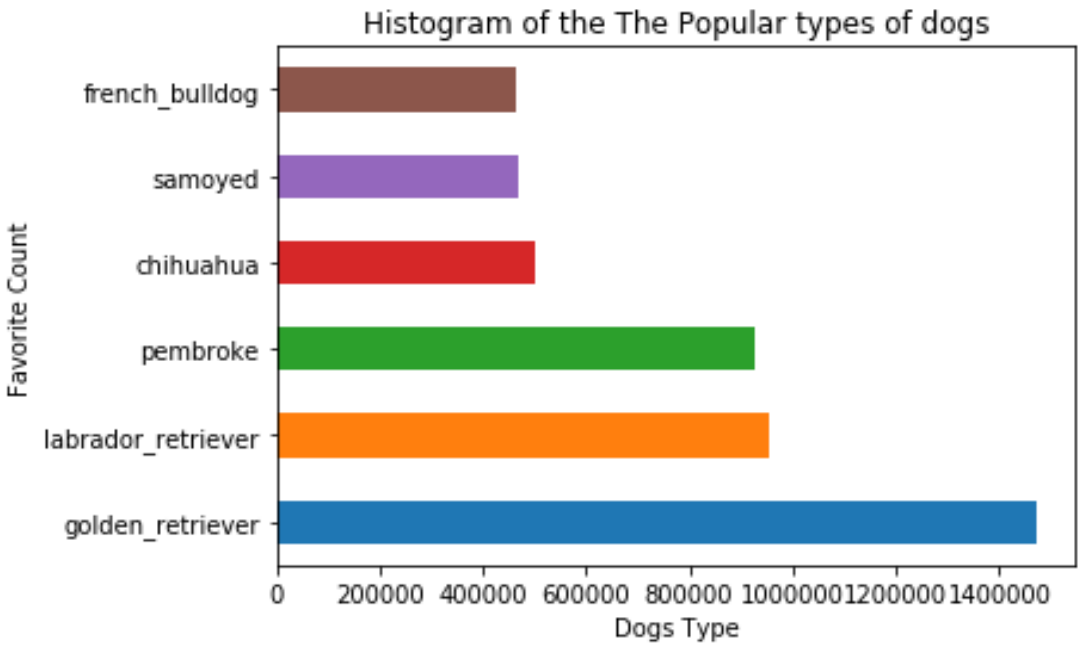


From The scatter above we can say favorite count and retweet count have a Positive relationship because they are increase together and they are affected each other positively.

Insight

WeRateDogs dataset has more than 5000 tweet. After wrangling this data, I analyzed more than 1000 tweet and I found that, the most dog type that people favorite is golden retriever with more than 140k.

You can see that clearly in the horizontal bar chart below:



Here are some pictures to golden retriever dog.