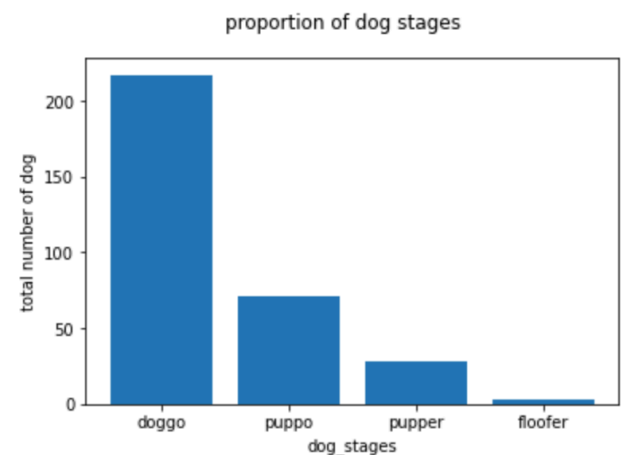# Insights

proportion of dog stages

From the first plot, we can see the total number of dogs per each stage. The majority of the dogs have been successfully identified in their stage are in doggo stage. We can see the proportion, the doggo stage is significantly higher than the other stages. I personally think that dogs in doggo stage are a lot more active than the other stages. People do a lot more activities, tweeting and tagging the dogs a lot more during this stage.

```
breed
golden_retriever       33
Labrador_retriever     16
Pembroke               13
Pomeranian              8
pug                     8
```

The second analysis was to find top five breed from the data. By grouping by the breed column and taking the aggregate value (the count of each breed, we are able to conclude that golden retriever is number one top pick dog breed. What factor affects this analysis? External factors may probably be the reason.

For the third analysis, I had to create a new column of "Time of Day" which is a categorical value showing the number of tweets per parts of day. By creating new variables (morning, afternoon, evening, night), and grouping it by and getting the count, we are able to conclude that majority of the people tweet during night time. People seem to be more active on the phones at night, when the day is over and they have a lot more free time before bed.

Number of tweets per parts of day
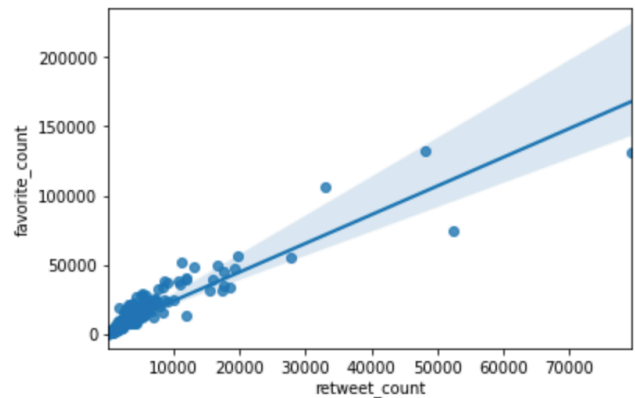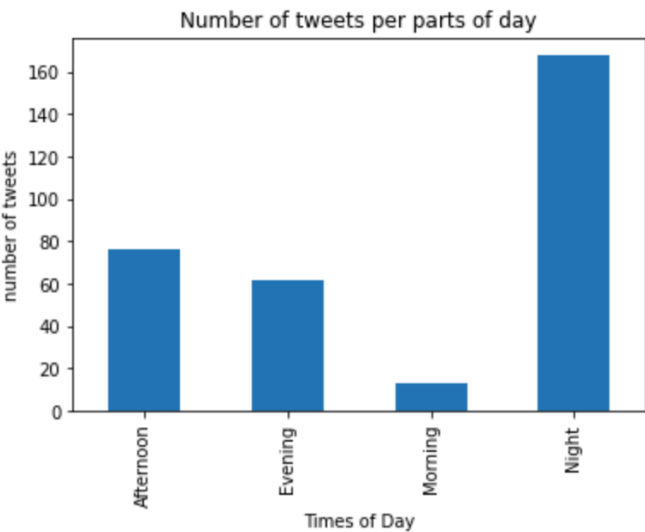
For the last plot, we can see the somewhat positive linear relationship between retweet count and favorite count. As one variable increases, the other also increases. The higher the favorite count, the higher the retweet. We can see that tweets with a lot of favorite count tend to get retweeted.