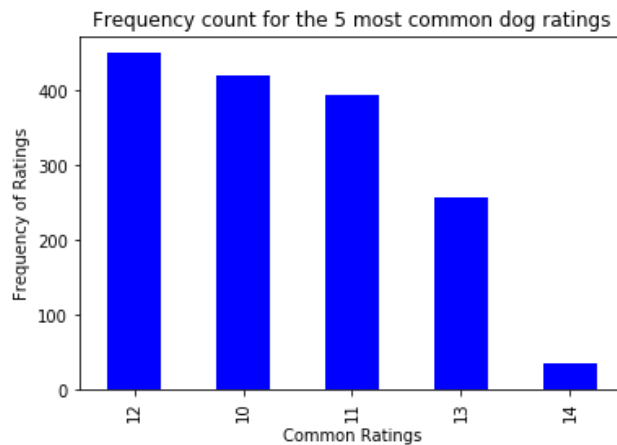


# Analyzing and Visualizing Data

## Insight1: Most popular rating across tweets

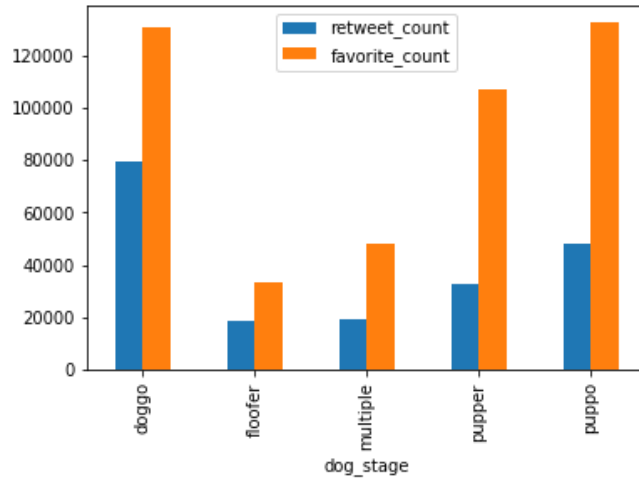
Since WeRateDogs is known for their unique ratings, we decided to view the most common and frequent ratings in the data by looking at the values of `rating_numerator` then plotting the first five values since the rest of the values are not frequent.



From the bar plot above, we can see that the most common rating given by WeRateDogs is **12/10**, followed by **10/10** and **11/10**.

## Insight2: Most retweeted and favorited dog stage

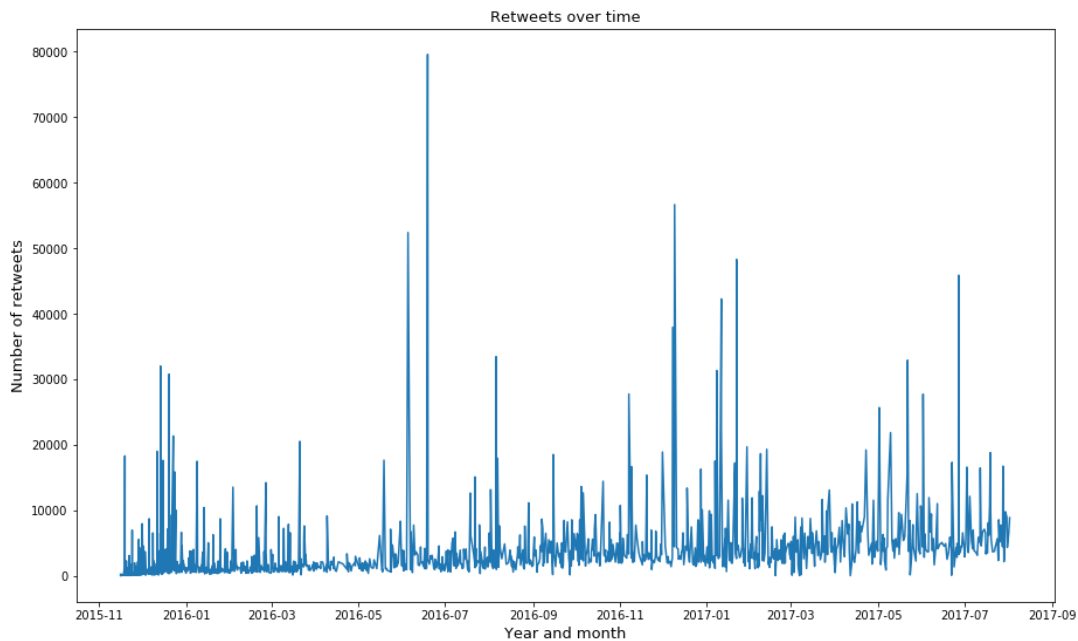
Although most dog stages are null values in our data, we are still curious about the retweet and favorite count of each dog stage. Thus, we created an additional dataframe containing the `dog_stage`, `retweet_count`, and `favorite_count` with no null values. We used `groupby()` and `aggregate()` to group by dog stages and aggregate the dataframe using `max`.



From the bar plot above, we can see that the dog stage with the maximum retweet count is `doggo`, followed by `puppo` then `pupper`. Also, we see that the favorite count is way more than the retweet count. The dog stage `puppo` had the maximum favourite count, followed by `doggo` then `pupper`. The dog stages with the smallest maximum retweet and favorite counts are `floofer` and `multiple`.

### Insight3: Retweet Count Over time

We want to further investigate the retweet count by looking at the timestamp of tweets and the retweet count.



From the plot above, we can see that peak with the most retweet count was in June 2016. Then, the retweet count dropped and reached its peak again in December 2016.