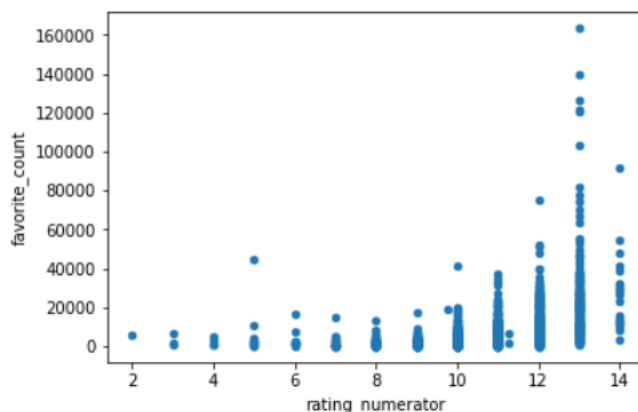


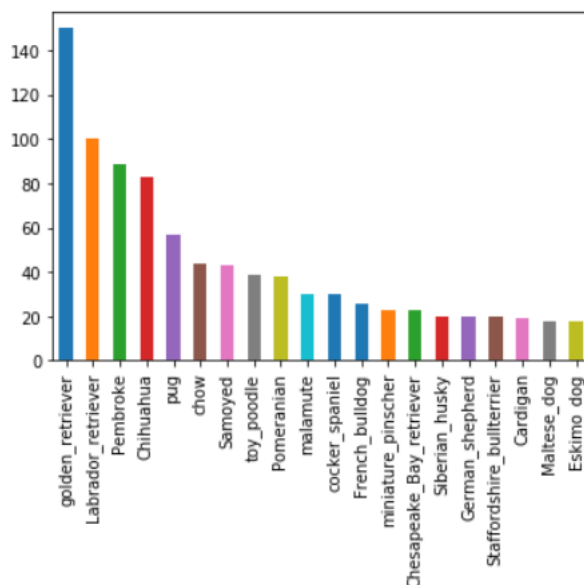
Insights and Visualization of Data

I had three questions about this WeRateDogs data.

My first question was “Does favorite count go up, as ratings go up?”. So I visualized the correlation between these two factors. I used scatter plot to visualize the ‘rating_numerator’ column and ‘favorite_count’ column. And the result showed that as ratings go up, favorite counts also tend to increase. But in case that rating was 14, the favorite counts actually decreased a little bit, compared to the case that rating was 12.



My second question was “Which breed are most popular in WeRateDogs?” I found out top 20 breeds that posted most frequently according to this data set. It turns out that golden retriever was posted most frequently(150 times), followed by Labrador retriever(100 times), Pembroke(89 times) and so on. Golden retriever and Labrador retriever were the only breeds that are posted more than 100 times. I made bar plots to visualize these top 20 breeds.



“Does posting multiple images influence on ratings?” was my last question. I used the ‘img_num’ column to find out the answer to this question. Using ‘value_counts’ function, I found out most people post only one image on WeRateDogs. And by groupby function, I calculated the mean of ratings of each number of images. It turns out that the mean ratings of four images was the highest, followed by three images, two images and one image. Therefore, it is possible to say that the more images you post, the better ratings you are likely to get.