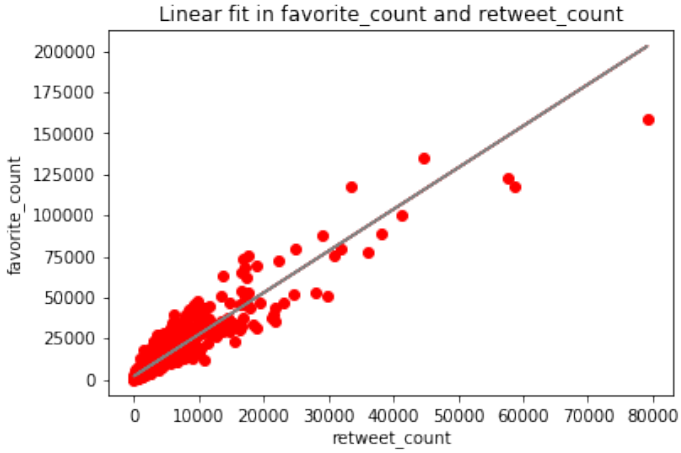
After gathering and cleaning the data, the next step is naturally to be assessing the data. In this part, multiple questions are asked, and detailed analysis are performed through different tools provided in python.

Firstly, it can be noticed that the dataset provides ‘favorite_count’ and ‘retweet_count’. What's the relationship between the number of favorites a tweet receives and its retweet number? Is it more likely to get retweeted if the tweet receives more favorites? To answer this question, firstly, create a new dataframe with only 'favorite_count' and 'retweet_count'. Then, fit all the data points with a linear line. The following plot is obtained.



The x-axis is retweet_count and y-axis is the favorite_count. The line equation shown in the plot is retweets = retweets = (2.541137111928492)*favorites + 1919.2141354580426. The r-value is 0.927.

Although the relationship between favorite_count and retweet_count might be complicated and can't be simply described by linear line, it still gives us a good indication. The r_value is 0.927, which shows that the two variable are highly correlated.

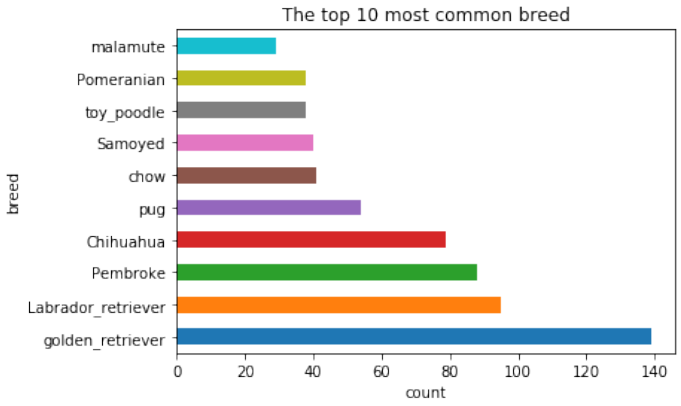
To further study the relationship between favorite_count and retweet_count, we want to see does the dog which has the most favorite also have the most retweets? What does it look like?

By outputting the row in the dataframe with the most favorites, it’s found that indeed, the the dog which has the most favorite also have the most retweets.

Which are the top 10 breeds? Compare the top 3 breeds' favorite_count. It’s found the top 10 are the following:

1. golden_retriever 139
2. Labrador_retriever 95
3. Pembroke 88
4. Chihuahua 79
5. pug 54
6. chow 41
7. Samoyed 40
8. toy_poodle 38

9. Pomeranian 38
10. malamute 29



total favorites of golden_retriever are 1601728.0
total favorites of Labrador_retriever are 1006600.0
total favorites of Pembroke are 941574.0