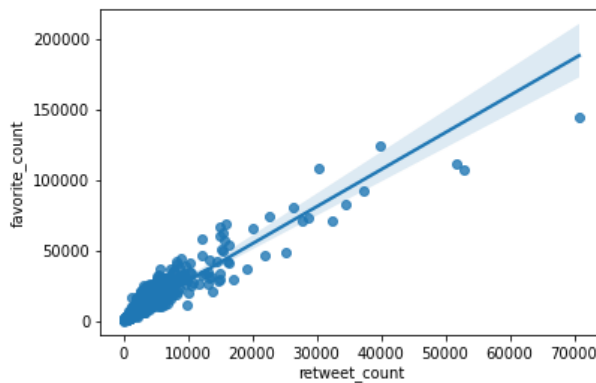After the merging of the now-clean three datasets on the twet_id column, some insights/conclusions were got from it.

The correlation between number of retweets and number of favorite counts was positive and there were no surprises there. The more people like a picture, the more likely they would retweet it also.

```
In [48]:  ▶ sns.regplot(x=combined_df['retweet_count'],y=combined_df['favorite_count'])
   Out[48]: <AxesSubplot:xlabel='retweet_count', ylabel='favorite_count'>
```
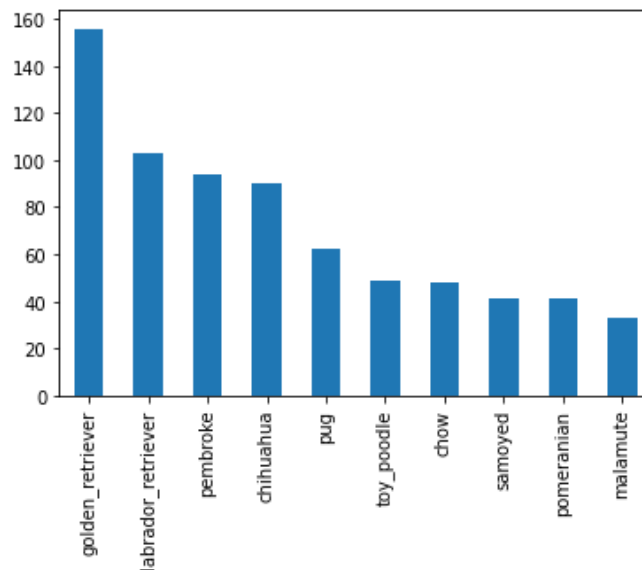

The little surprise came when plotting for the correlation for rating numerator vs retweet_count and as noticeable from the plot, there are a lot of outliers and it is not positively strongly correlated.

The bar plot for the 10 most common dog breeds were visualized using value_counts, index slicing. Golden_retriever was the most common dog (and by some margin) followed by lambrador_retriever, pembroke, chihuaha and pug.

```
In [50]:  ▶ combined_df.predicted_breed.value_counts()[:10].plot(kind='bar')
```

Out[50]: <AxesSubplot:>

That you have the most common dog doesnot guarantee you likes and retweets from other tweeter users. The most likely dog to get a retweet or like was gotten using the groupby pandas function, grouping the dataframe by dog breeds and applying the mean of favorite count and/or retweet_counts. The results from this 2 should be similar considering the fact that favorite count and retweet counts have strong positive correlation.

The 5 most likely pictures of breeds of dogs to be retweeted are:

1. bedlington_terrier
2. afghan_hound
3. standard_poodle
4. french_bulldog
5. english_springer.

The 5 most likely pictures of breeds of dogs to be liked are:

1. bedlington_terrier
2. saluki
3. French_bulldog
4. bouvier_des_flandres

5. afghan_hound,

Doing this has made known that one of the keys of getting likes/retweets is not by having the most popular dog (golden_retriever) but by having a bedlington_terrier.