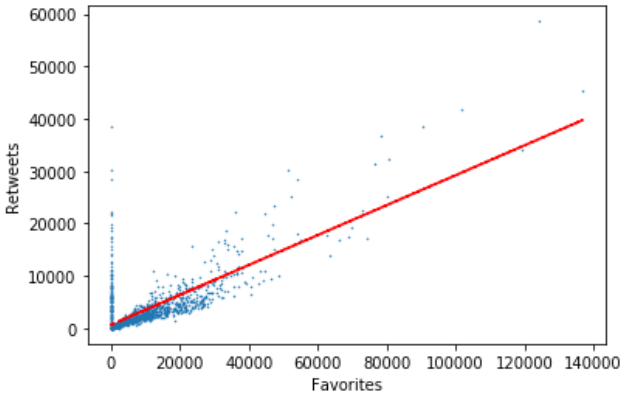
Data Exploration Report

After exploring the clean dataset, here are some insights:

- ## Insight 1

  The more “Favorites”, the more re-tweets. But the plot indicates people tend to favorite it instead of re-tweet it. From the retweets-favorites scatter plot, seems there is a clear pattern: a tweet with 20000 favorites usually got around 5000 retweets, much less than favorites counts. Guessing people like it first before retweeting it. We can also observe that low favorite counts can also be along with high retweet counts, suggesting people are talking about them but they are not popular.

  Additionally, Pearson correlation analysis is performed. We found that favorites are highly correlated to retweets. However, rating is not correlated to favorites and retweets as the correlation coefficient is quite low.



	favorites	rating_numerator	retweets
favorites	1.000000	0.257892	0.777238
rating_numerator	0.257892	1.000000	0.203083
retweets	0.777238	0.203083	1.000000

- ## Insight 2

  Generally speaking, user tend to like higher rating as time goes by. There is an upward high rating trend. For example, in 2017, there is less and less tweet with rating less than 10. It may due to weratedogs had changed the way they rate a dog.

- **Insight 3**

  Most of tweet are for pupper, but this dog type usually got low rating, both mean and median value. From the boxplot, it indicates that pupper dog stage usually got a bad rating with intermediate variance. We can also observe some extremely low-rating dog in pupper dog type. However, the dog type categories are not as various as I expect. Besides, the records of dog types are not sufficient to get a concrete conclusion.