

Data Wrangling of Twitter Data

Qi Zhao

Lehigh University

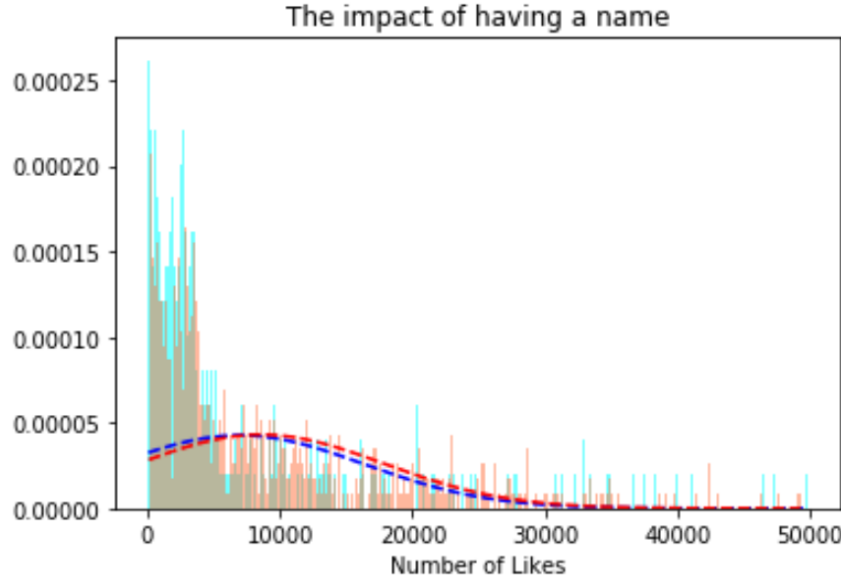


Fig. 1. 2 hyperplanes

We gathered the tweets from the twitter account "WeRateDogs", including the contents, number of likes, re-tweets, and the time of the tweets. We tried to get some insights from the data.

After wrangling the data from the data, we have some interesting findings that worth further research.

First of all, the 'likes' number of the tweet is increasing dramatically over the time. From the visualization we can see that it is going up with the 'J' shape.

The number of re-tweets is also growing, however, not as fast as the number of 'likes'. This is probably because more and more people start to click 'likes' without re-tweeting.

Besides, although we can see a little difference between the likes number of the named dogs and unnamed dogs, it is not significant at all. Thus probably giving your dog a name is not necessary to get more likes.

Lastly, the score and the number of likes have a positive relation, which means the person who give the score has a similar tastes as his or her readers. However, it may also because the score has an impact on the readers.