

# act\_report

December 15, 2018

## 1 Report after analyse

### 1.1 Question

In this project, we explore following problems after wrangle the data.

1. Which stage of dogs is most tweeted?
2. Which stage get most highest score?
3. Is there any relation between retweet\_count and favorite\_count?
4. Is there any relation between time and favorite\_counts?
5. How many source of tweets? What platform is most popular?
6. Will follower become more with time past?

### 1.2 Explore

At first two questions, we focus on dog stage and related data, try to figure out which type of dog is most popular among people.

#### **Q1. Which stage of dogs is most tweeted?**

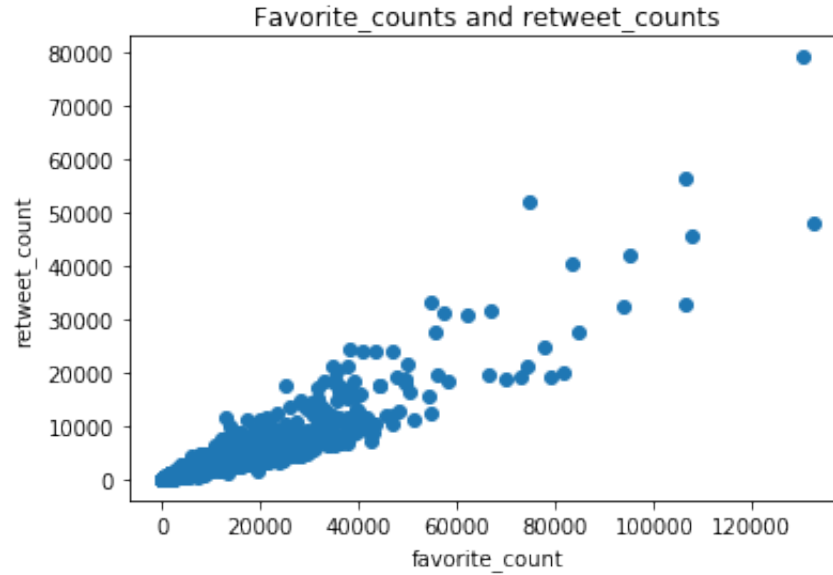
We use value\_counts function to explore how many tweets are there about each stage of dogs, and it's shown as below:

Stage	count
pupper	232
doggo	84
puppo	25
floofer	9

We could easily tell from the table that pupper is the most tweeted stage of dogs.(We note there are large amount of data not labeled, which could try to use neural network label the data and use them in this analyse.)

#### **Q2. Which stage get most highest score?**

For this question, we take a look at statistic information of score each stage of dogs get, we have following table:



favorite-retweet

stage	count	mean	std	min	25%	50%	75%	max
doggo	84.0	1.173810	0.156096	0.5	1.1	1.2	1.3	1.4
floofer	9.0	1.188889	0.105409	1.0	1.1	1.2	1.3	1.3
pupper	232.0	1.084914	0.200402	0.3	1.0	1.1	1.2	2.7
puppo	25.0	1.208000	0.125565	0.9	1.2	1.2	1.3	1.4

Since there's a unbalance between example number of four stage, we use basic statistic data, mean value, std value and qartile to describe the score. We can see that pupper have lowest score and highest score at same time, which isn't surprising since it's most tweeted. naturally, it has highest std value and lowest mean value. doggo is almost same situation as pupper. From the table, we could point out puppo may have highest score among four stages.

From data we have, people maybe willing to tweet their pupper on the internet, and more people love puppo, but consider great number of pupper examples, we could say it's also welcome on the Internet.

Next, we'll explore some data relates to tweet itself, other than it's content.

### Q3. Is there any relation between retweet\_count and favorite\_count?

Like many people, we may curious about is there any connection between retweet\_count and favorite\_count? Normally we may think tweets get more favorite than retweet, since it's not convenience to retweet.

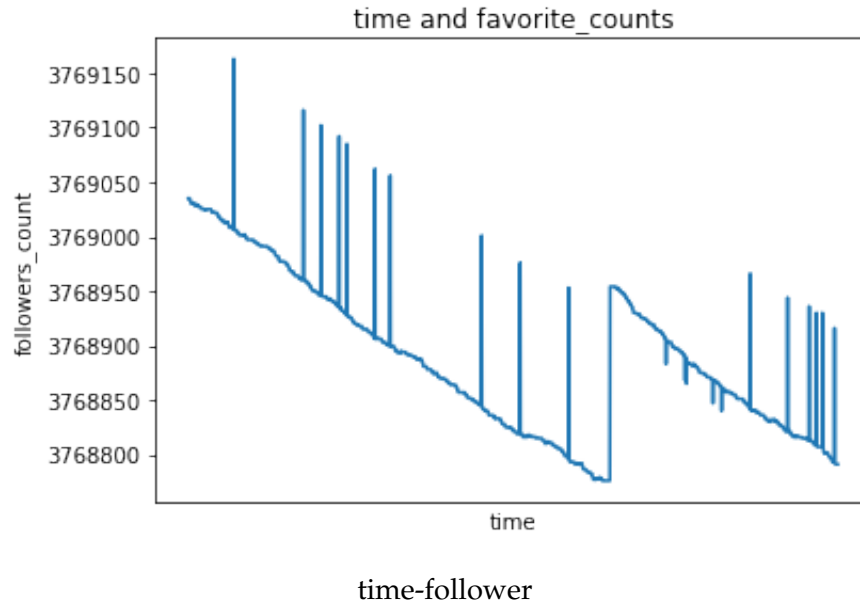
Basic on the data we have, I plot a scatter image:

From this image, we could tell that retweet count usually less than favorite count, which certify our guess. Also, this two variables have a positive correlation, more people favorite the tweet, more people retweet, obviously.

### Q4. Will follower become more as time past?

Everyone wants more followers, so, will follower increase as time past? To understand this question, we plot this chart:

Sadly, followers don't increase as time past, on contrary, people even unfollow as time past.



What could be behind it? Maybe a user quit tweeter, maybe they just clean their follow list... Nevertheless, we could see some peaks on the plot, seems after a period of time, this account get more followers. At the same time, follower number fluctuates around 3,768,950 and the difference between max and min is less than 2,000, which makes sense.

**Q5. Is there any relation between time and favorite\_counts?**

Now, what relation between time and favorite count? We plot two charts to answer this question. First, we plot a chart with whole time:

There's a positive correlation between these two variables, which is not very rigid, but this graphic inspires me with another question, will time you tweet affect favorite it gets? So, we have the following graph:

We split a day into four periods, 23:00:00 - 6:00:00 as night, 6:00:00 - 13:00:00 as morning, 13:00:00 - 18:00:00 as afternoon and 18:00:00 - 23:00:00 as evening. We could observe that tweets tweeted in day get more favorites than those tweeted in night. So, if you'd like to get more favorites, choose a good time, 6:00 to 13:00 maybe a nice choice since it gets the highest favorite in this case.

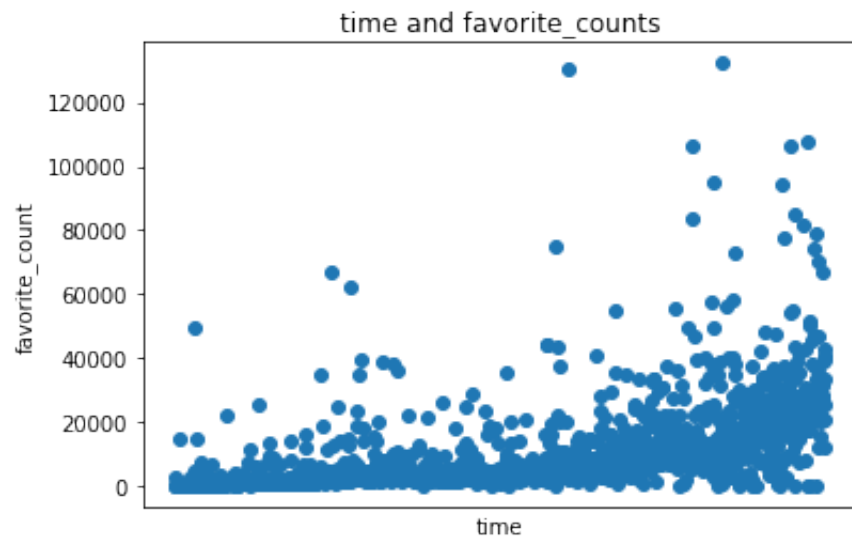
**How many sources of tweets? What platform is most popular?**

Last but not least, we focus on the source of tweets. Which platform do they use often? Mobilephones, PCs, or some other kind of equipment? So we use value\_count again and get the table below:

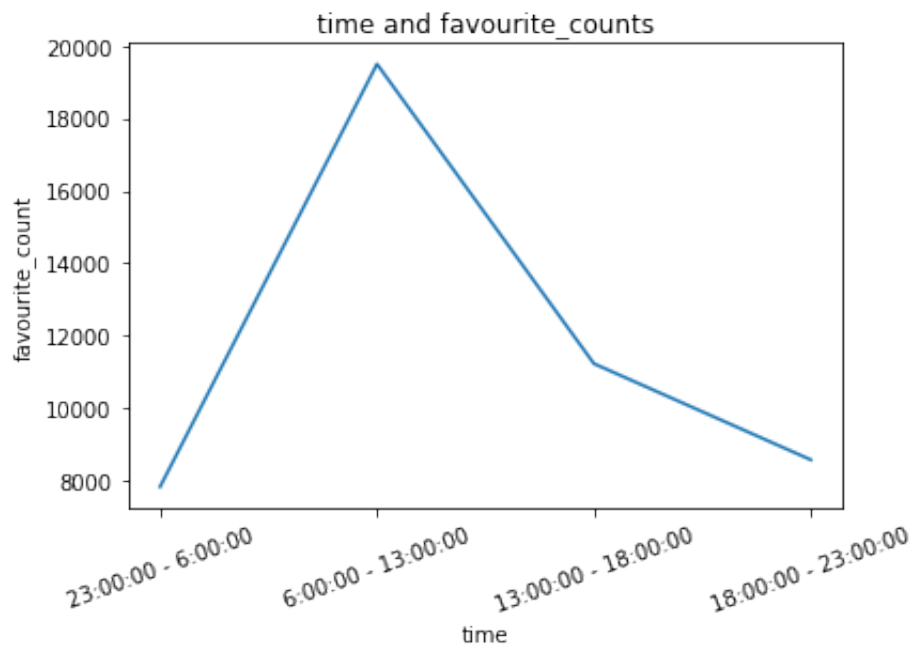
Source	count
Twitter for iPhone	2047
Vine - Make a Scene	92
Twitter Web Client	31
TweetDeck	11

No wonder that iPhone is the most used platform. Vine, which seems a website came as second position, following is twitter's web application, then their deck software, which only used 11 times.

Maybe twitter will pay more effort on iPhone or other mobile apps since it's most used to tweet, then they could improve their web application, as for deck, it seems not that important in this product system.



time-favorite



time-favorite-2