Act Report

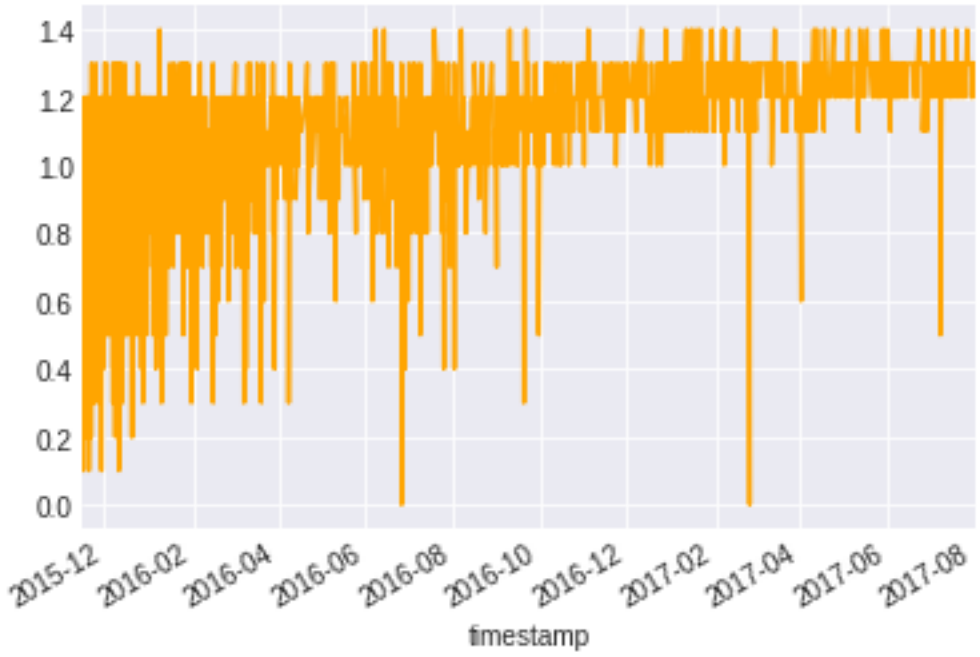
Dataset

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1991 entries, 0 to 2068
Data columns (total 23 columns):
tweet_id                 1991 non-null object
in_reply_to_status_id    1991 non-null object
in_reply_to_user_id      1991 non-null object
timestamp                1991 non-null datetime64[ns]
source                   1991 non-null category
text                     1991 non-null object
expanded_urls            1991 non-null object
name                     1991 non-null object
meme                     1991 non-null category
retweet_count            1991 non-null int64
favorite_count           1991 non-null int64
jpg_url                  1991 non-null object
img_num                  1991 non-null int64
p1                       1991 non-null object
p1_conf                  1991 non-null float64
p1_dog                   1991 non-null bool
p2                       1991 non-null object
p2_conf                  1991 non-null float64
p2_dog                   1991 non-null bool
p3                       1991 non-null object
p3_conf                  1991 non-null float64
p3_dog                   1991 non-null bool
rating                   1991 non-null float64
dtypes: bool(3), category(2), datetime64[ns](1), float64(4), int64
(3), object(10)
```

This clean dataset has 1991 observations and 23 features. The key features are rating, retweet_count, favorite_count.

Visualization

Fig. 1 Ratings over time



As the time goes by, the average of image rating is increasing. The reason is that tweeters have more experience in how to photo. There are 2 times rating 0 around July 2016 and March 2013 showing in this figure. The related images are downloaded as followed. The first picture has nothing related to dog, and the second one is a proving of someone's plagiarism. The reason of 0 rating is obvious.

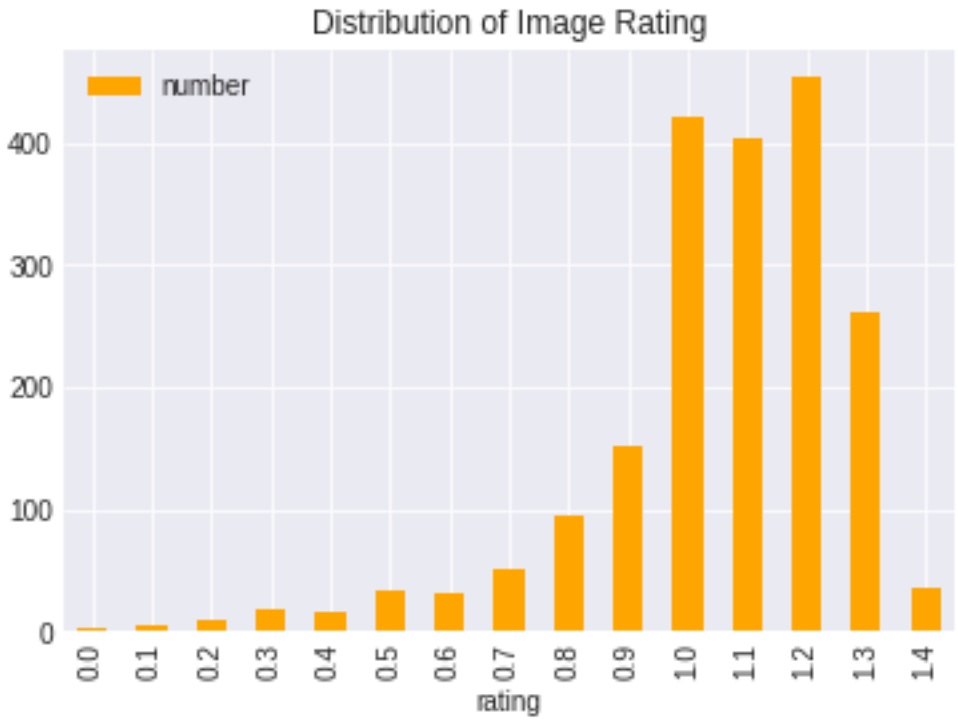
WeRateDogs™

@dog_rates

Follow

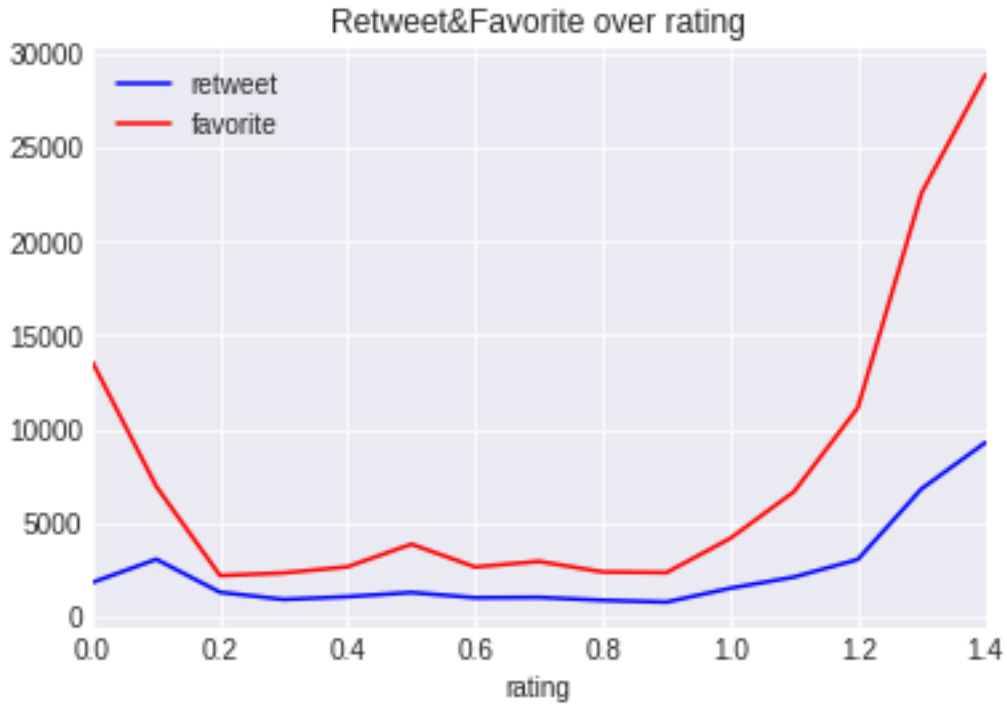
PUPDATE: can't see any. Even if I could, I couldn't reach them to pet. 0/10 much disappointment

Fig 2. Distribution of Image Rating



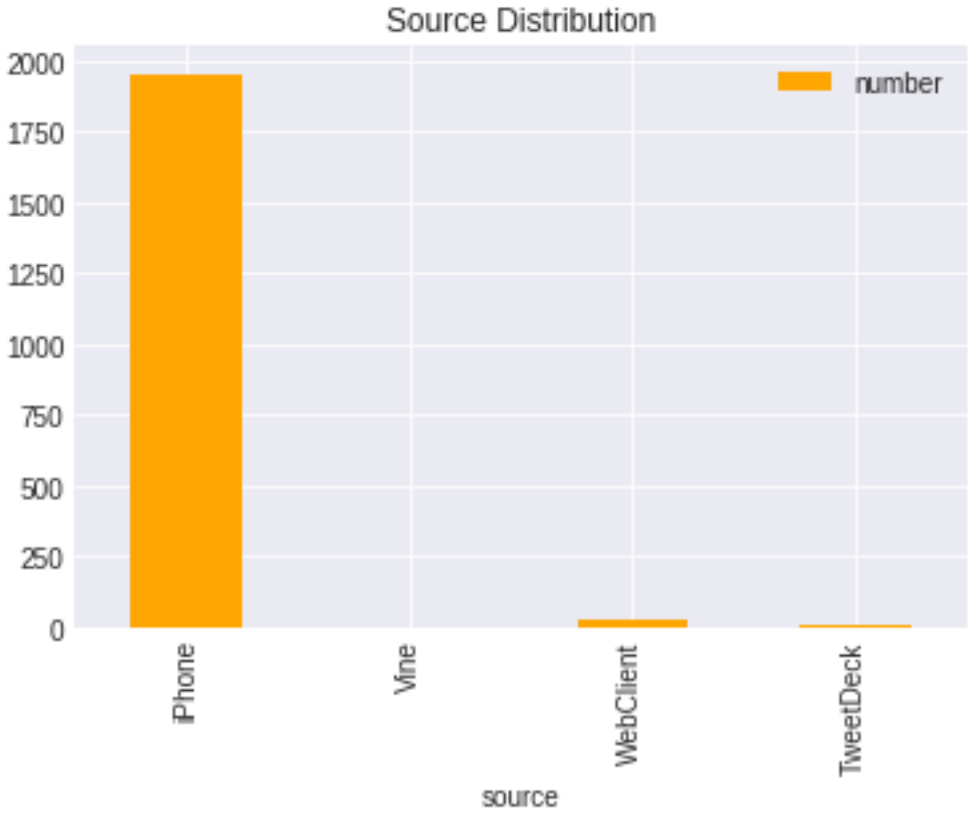
The distribution of image rating is left skewed. Ratings 1.0, 1.1, 1.2 are the most common ratings. Only a few images have very high or very low rating.

Fig 3. Retweet&Favorite over rating



Generally, tweeters prefer 'like' to 'share' interesting since the action of liking a post is easier. And people like tweets with extreme rating which means more attractive.

Fig 4. Source Distribution



Most of Tweeters use cellphone to tweet since using mobile to surf the internet is much easier than computer.