

WeRateDogs Data Wrangling Insight

Twitter provides lots of nice features and even makes it possible to collaborate to build a topics specific group where lots of likeminded people and sometimes animals show off themselves. So as WeRateDogs do, a Twitter account that rates people's dogs in a most creative fashion. On outstanding thing is the unusual rating of overestimating or like shooting to the star rating by mostly giving ratings like 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." (Meme) WeRateDogs has over 4 million followers and has received international media coverage.

Udacity provided all necessary data from twitter and also the possibility to enrich the dataset by using the tweepy API to get more updated data from twitter. Before it was possible to give some insights or answer question about it, data had to be gathered, assessed and finally cleaned. After all the tedious and exhausting part; ending up with one clean dataset was a relieve and now coming up with some basic question about the dataset and the group participants.

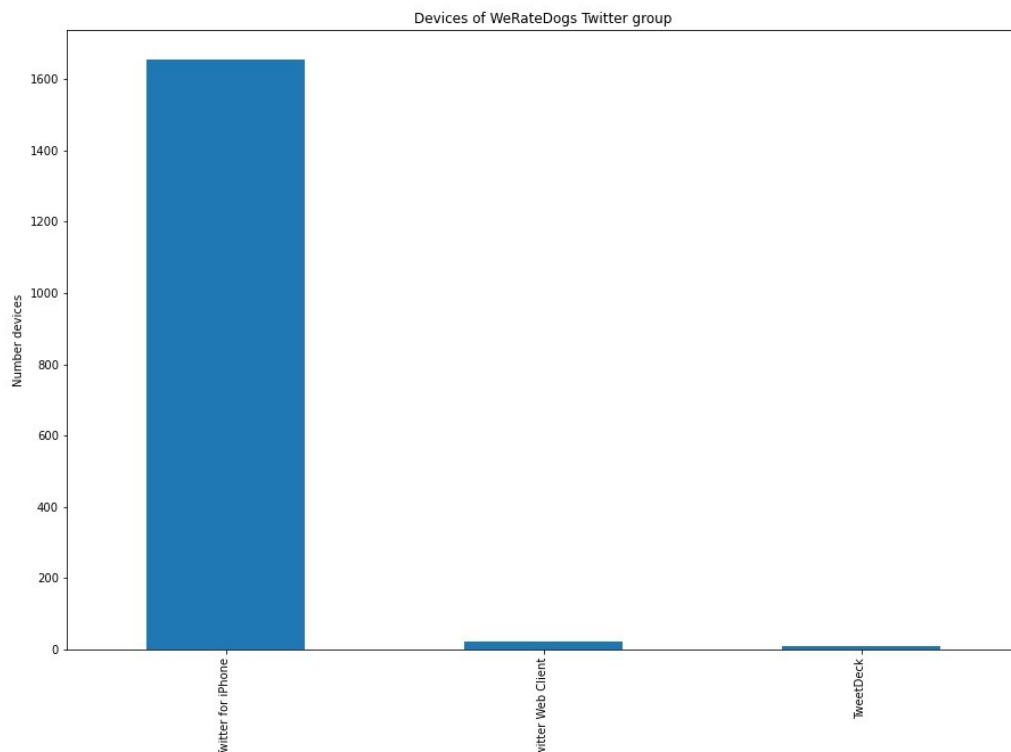
Info about the dataset

Initially there were more than 2500 datasamples in more than 50 different features. After cleaning there are 23 features with 1685 samples available.

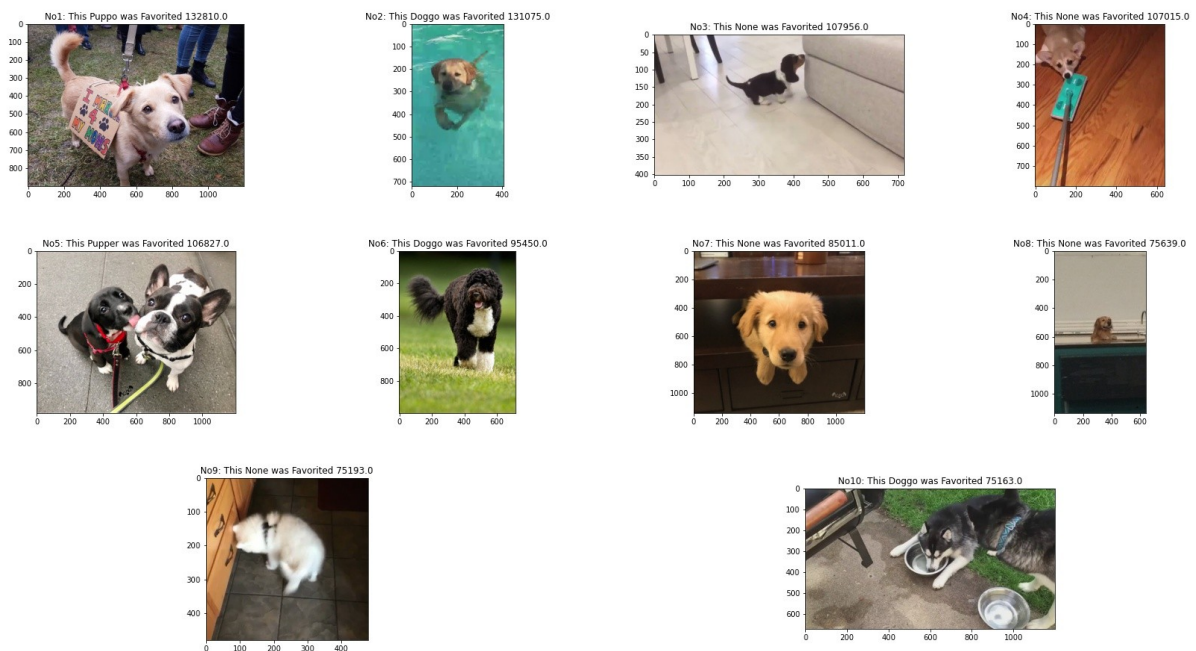
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1685 entries, 0 to 1684
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  -
0   tweet_id              1685 non-null   int64
1   timeStamp             1685 non-null   object
2   text                  1685 non-null   object
3   expanded_urls         1685 non-null   object
4   rating_numerator      1685 non-null   int64
5   rating_denominator    1685 non-null   int64
6   name                  1685 non-null   object
7   jpg_url               1685 non-null   object
8   img_num               1685 non-null   float64
9   created_at            1685 non-null   object
10  id_str                 1685 non-null   float64
11  full_text              1685 non-null   object
12  truncated              1685 non-null   bool
13  retweet_count          1685 non-null   float64
14  favorite_count         1685 non-null   float64
15  possibly_sensitive     1685 non-null   float64
16  possibly_sensitive_appealable  1685 non-null   float64
17  lang                   1685 non-null   object
18  dog_race               1685 non-null   object
19  device                 1685 non-null   object
20  device_url             1685 non-null   object
21  dogtionary             1685 non-null   object
22  text_length            1685 non-null   int64
dtypes: bool(1), float64(6), int64(4), object(12)
memory usage: 291.4+ KB
```

What devices do users tweet from?

From the following plot we can see, that WeRateDogs twitter users mostly use their iPhone for posting their content.

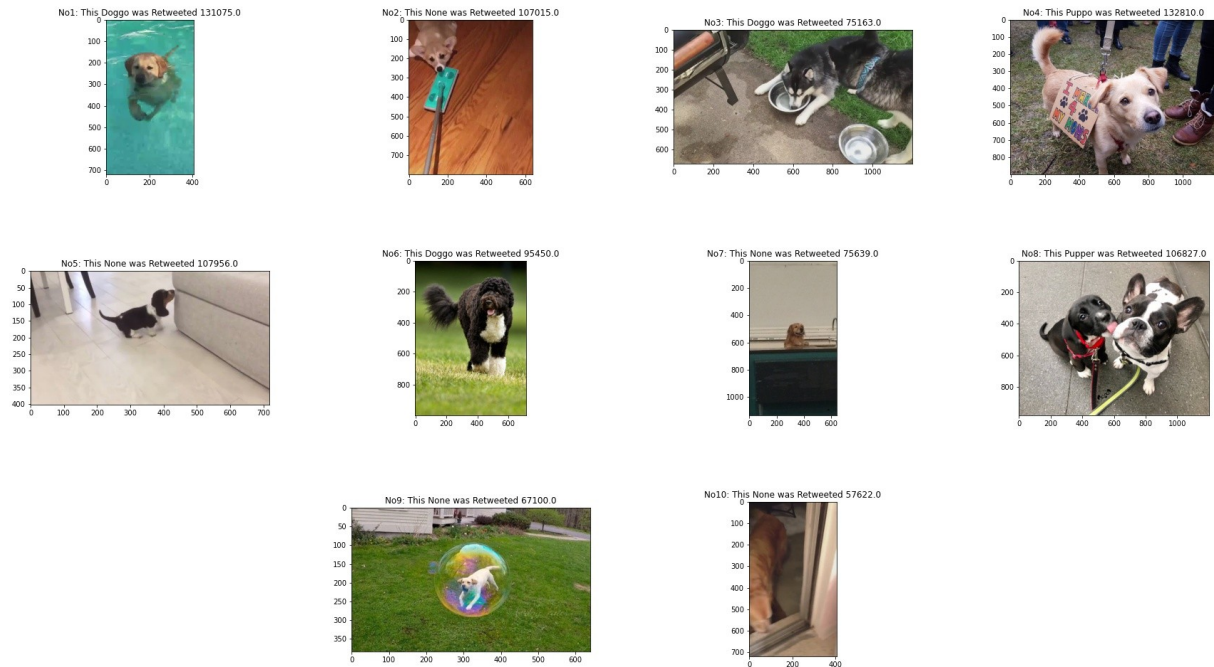


What are the top 10 favorites?



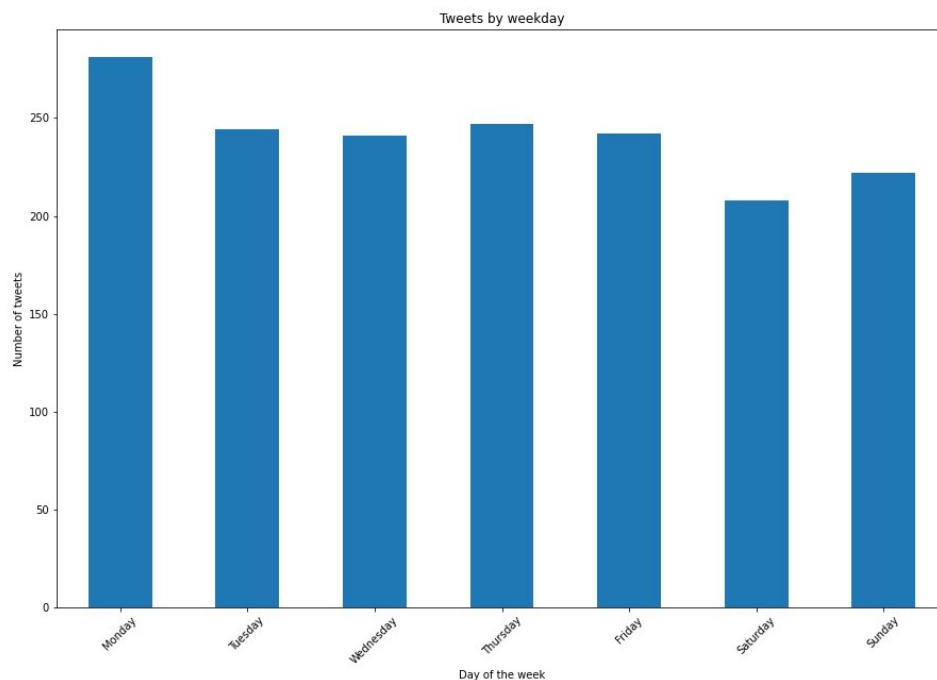
For my unprofessional eye, is mostly about the situation or pose or toys what the dogs are at the moment. So, it's less about the dog at itself, but more about the setup...

What are the top 10 retweeted tweets?

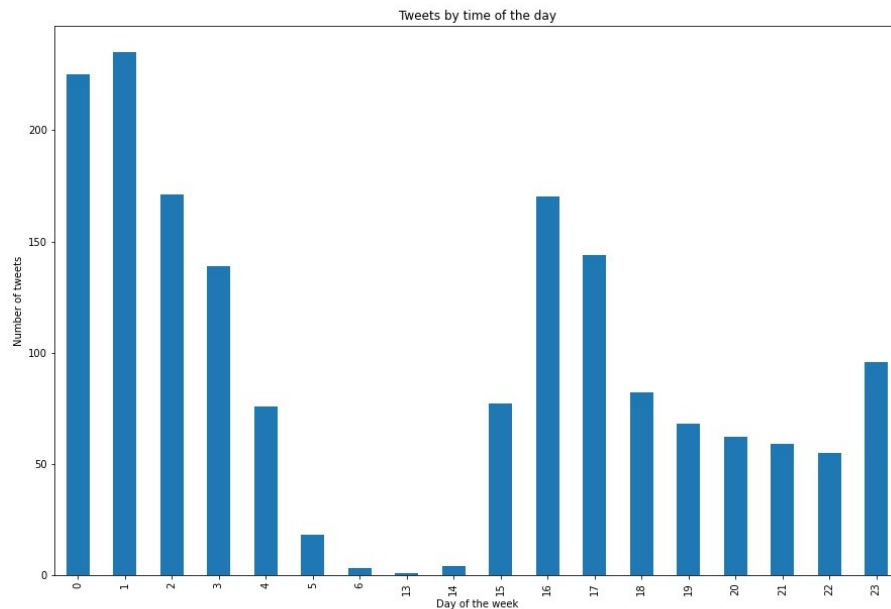


Situation here is similar as above mentioned. They almost the same, in a different order and some exchanges. So, retweeting is not the same but similar in it's information content about being „famous“ or „important“. That said, a most favorite dog might not be the most retweeted one but is likely in the top 10.

What day of the week and time of the day to people tweet the most?

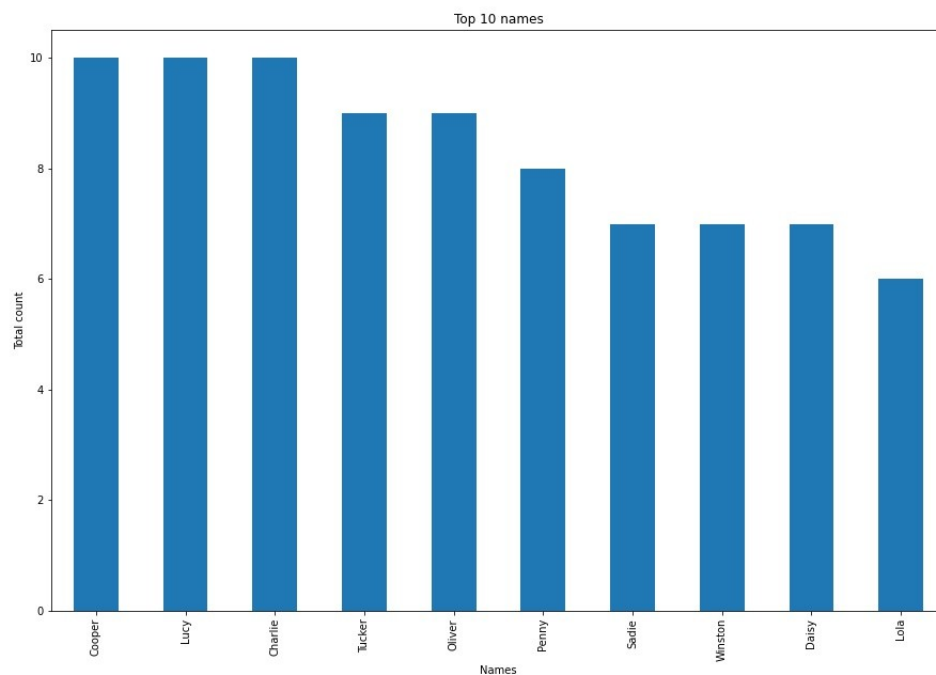


It is common sense that people normally have more time for twitter stuff during the free time... Surprisingly peak tweet day is monday and little less on weekends compared to the average on during the week.



The peak tweet time is at 0 and 1 in the morning... Which is GMT and for the US that might be around 4 and 5 pm. Besides that during brunch time there are some smaller peaks as well. What is striking as well is, that there are 7 hours of no tweeting at all.

The top 10 dog names (after no name provided)



The dognames are less like classic „spot“ dognames but more names as someone would give their children. So, they must really love their dogs and it might even be a children or partner replacement. And as we can see everyday, people spend a lot of cash on their beloved dogs.