Data Analysis of WeRateDogs twitter Account

Data analysis on the twitter account WeRateDogs was performed on their original tweets that contain an image and were tweeted before 2 Aug. 2017

Data was gathered from 3 different sources and combined and cleaned

1. WeRateDogs tweet archive in a .CSV file
   This file contained the following attributes
2. Neural Network Dog breed predictions for the tweets in a .tsv file
3. Tweet details gathere through twitter API by tweet ID

After gathering the data visual and programmatic assessment were performed.
The assessment resulted in 8 data quality issues and 4 data tidiness issues

The 8 data quality issues were

1. Some tweets are retweets and not originals ('retweeted_status' == True and/or there's a value for in_reply_to_status_id, in_reply_to_user_id, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp
2. Some tweets are replies (reply_status == True)
3. Some tweets do not have images or do not exist anymore (values in "fails_dict")
4. Some image prediciton doesn't have any dog breed in the result (p1_dog & p2_dog & p3_dog are all False)
5. Some tweets do not have the correct dog rating ta_df.query('rating_denominator != 10')
6. Some tweets have vine videos instead of photos (td_df['tweet_source'] == "Vine - Make a Scene")
7. Only a small fraction of tweets contain dog stage (less than 400 tweets) so this data is incomplete
8. Some entries have a wrong dog name (all entries under name starting with a small letter are wrong, other entries starting with capital letters are mostly correct)

The 4 data tidiness issues were

1. Some columns have the wrong data type
2. Tables are separated
3. Some columns are redundant
4. Some columns are not well named

After cleaning the data the results was saved in a csv file for easy access in the future.

Basic analysis of the data was performed

The findings of the analysis are as follow

- The most common dog name in the tweets we have is Cooper with 10 occurences
- The most tweeted tweet for WeRateDogs is https://t.co/DkBYaCAg2d it was retweeted 54647 times it rates Stephan with 13/10
- The most favorited tweet for WeRateDogs is https://t.co/DkBYaCAg2d it was favorited 116922 times it rates Stephan with Stephan/10
- The most retweeted tweet is the most favorited tweet

The relationship between Rating and retweet counts s as follows