DECI project (`we rate dogs`): act_report

After cleaning the data given it was stored as 'twitter_archive_master' and the following questions were answered:

1.top 5 timestamps with the highest favorite_count

2.top 5 timestamps with the highest retweet_count

3.Finding the relation between favorite count and rating_denominator andrating_numerator the relation between favorite count and rating_numerator and rating_denominator

Here are the insights:

- Favorite count and retweet count has been found to reach their peaks in June. This can be rationally attributed to the fact that dog festival normally occur during this period. Followed by this month is January and December for favorite count and retweet count respectively. Third on the list is also December and January (respectively). This may be due to increased festive activities during the perionds

- Also, as expected, the correlation between favorite count and retweet count is, positively, very strong (0.79) Hence, favorite tweets are more likey to be retweeted.

- On the other hand, the correlation between the each feature (favorite count and retweet count) and numerator rating is and denominator rating is very weak, positive for the former and negative for the latter.

Recommendations

- Dog events should be hosted around June, December or January.
- Another factor should be used in predicting probability of retweeing as the numerator and denominator ratings are not effective.

1. top 5 timestamps with the highest favorite_count

2. top 5 timestamps with the highest retweet_count

Favorite count and retweet count has been found to reach their peaks in June. This can be rationally attributed to the fact that dog festival normally occur during this period. Followed by this month is January and December for favorite count and retweet count

respectively. Third on the list is also December and January (respectively). This may be due to increased festive activities during the perionds
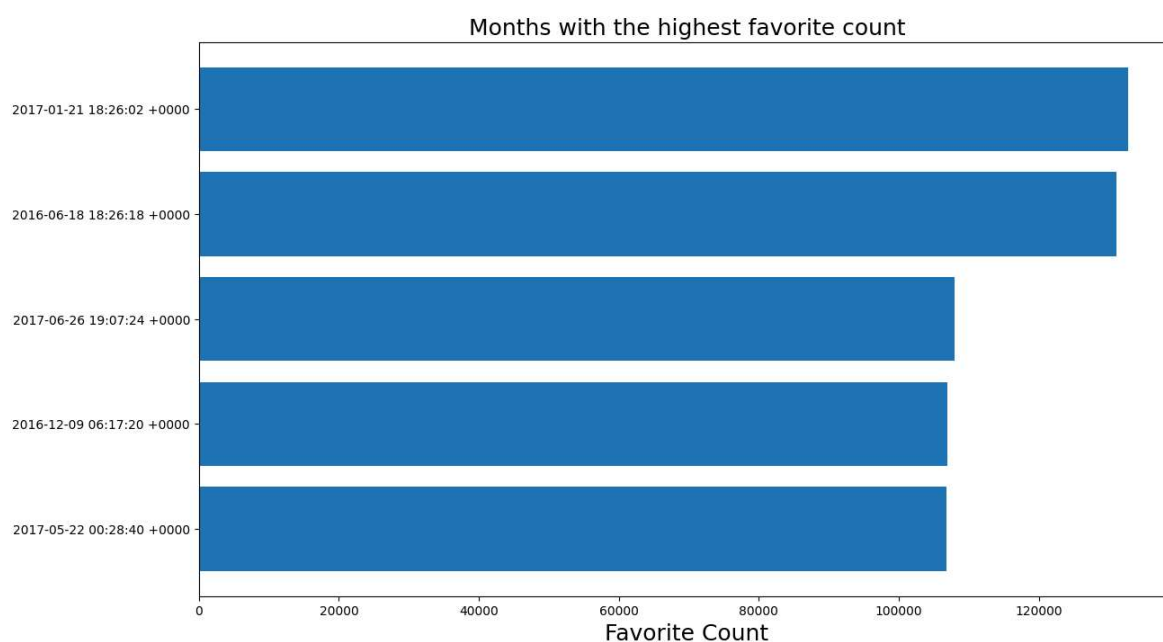
3. Finding the relation between favorite count and rating_denominator andrating_numerator the relation between favorite count and rating_numerator and rating_denominator

the correlation between favorite count and retweet count is, positively, very strong (0.79) Hence, favorite tweets are more likey to be retweeted

Correlation between the each feature (favorite count and retweet count) and numerator rating is and denominator rating is very weak, positive for the former and negative for the latter.

1. top 5 timestamps with the highest favorite_count

In [ ]:

Months with the highest favorite count

2. top 5 timestamps with the highest retweet_count

In [ ]:

Months with the highest retweet count