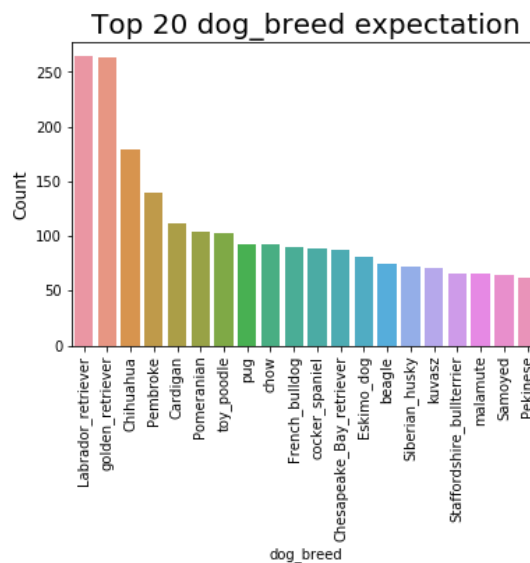Analysis report

Jinsoo Kim

9 June 2018

Insight 1

1. Labrador_retriever, 2. golden_retriever, 3. Chihuahua, 4. Pembroke, and 5. Cardigan are expected the most. According to below picture, We can see top 20 expected dog breeds.
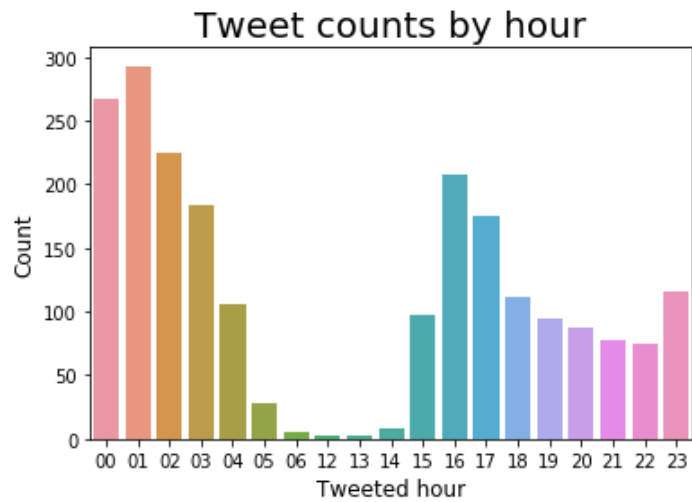

According to '20 Most Popular Dog Breeds in America' by Nancy Barber& Michelle Lanz(https://mom.me/pets/18028-10-most-popular-dog-breeds-america/), labrador_retriever got the first place and golden_retriever got the second. This can be reliable in some ways.
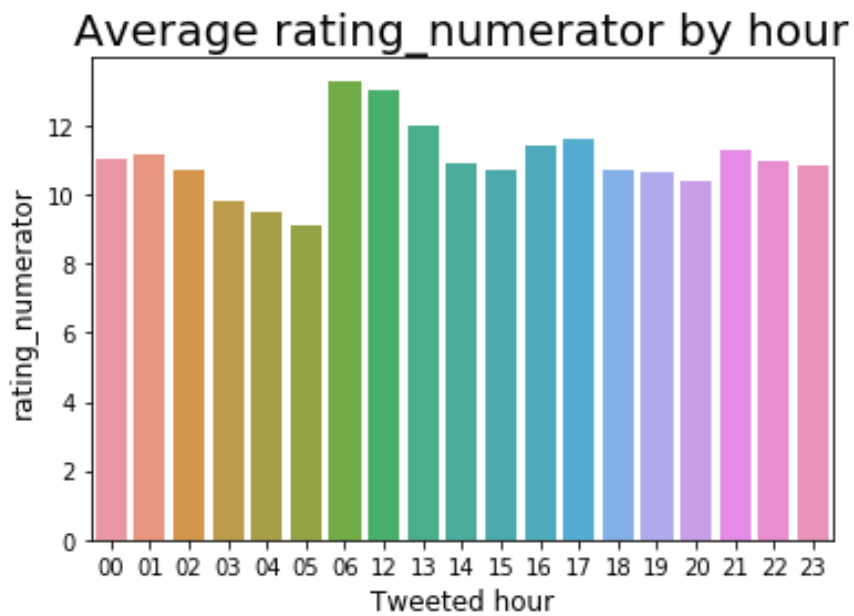
Since expectation modeling is not perfect for every case, we do not know which dog breed truly be rated the most.

Insight 2

There are certain time zone that WeRateDog works hard.

Tweet counts by hour
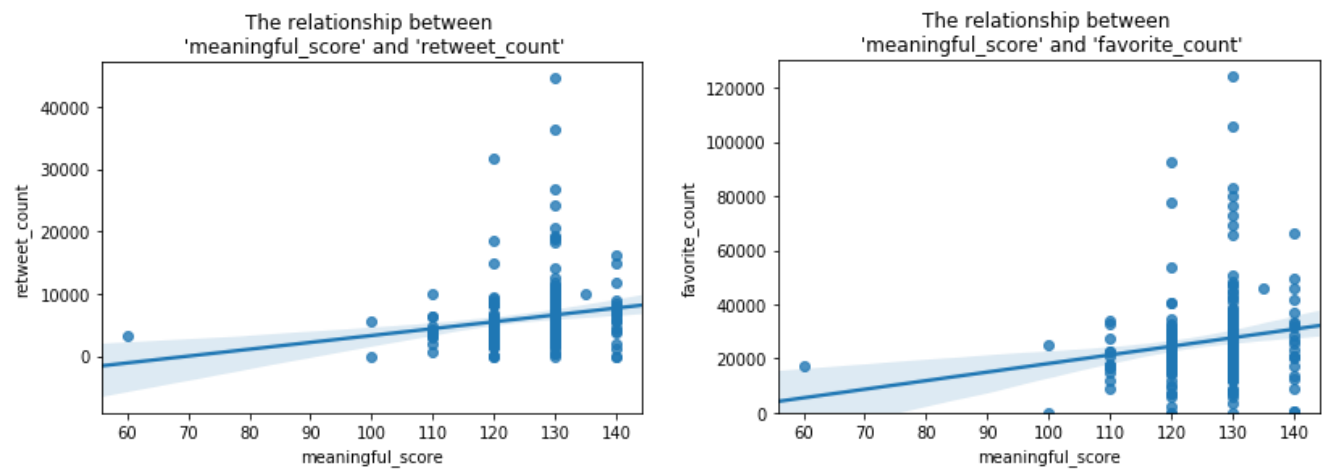
As we can see above, from November 2015 to August 2018 for 07-12 o'clock not even one tweets are made. WeRateDog mainly normaly works hard 00-04 o'clock and 16-18 o'clock.

Average rating_numerator by hour

Except some anomalies, rating scores are quite similarly distributed.

Insight 3

I thought high rating guarantees many favorite count and retweet count. But it was not.

Correlation coefficient between 'meaningful_score' and 'favorite_count' is 0.170056.

Correlation coefficient between 'meaningful_score' and 'retweet_count' is 0.178414.

Both two graphs and two correlation coefficient suggest they are not related.

On the other side, Tweets' 'favorite_count' and 'retweet_count' are strongly related. Because their correlation coefficient is 0.948006 and below picture prove it.