

Act Report with WeRateDog Data

July 26, 2019

1 Act Report with WeRateDog Data

1.1 Data Set Intro

- A Text Sample :
- This example is a good representative of the data set with a lot of issues to be fixed.
- For example :
- smile 24/7 , 24/7 is taken as a rating score
- Another problem is that it is not the original tweet but a retweet from gofundme. The user post it to ask for help but not for rate the dog.

In []:

In []:

1.2 Data Wrangle :

12 quality issues and 4 tidiness issues are fixed before proceeding to data visulization and analysis

1.3 Data Visulization and Data Analysis:

Q 1) What is the relationship between favorite_count and retweet_count ?

Q 2) What are the most popular breeds?

Q 3) What is the relationship between dog stages and retweet counts and favorite counts ?

2 Evaluation of Answer:

Q 1) What is the relationship between favorite_count and retweet_count ?

Answer:

According to the plot, we can see there are two activity peaks 0-3 and 16-17. Users like to post tweets either at the time before off work or late night(even 3 am belongs to early morning but It is more likely they stay late until 3 am rather than they get up at 3 am to post tweets).

It is shown that tweets posted at evening time or late night gains more retweets and favorites. It is generally consistent with the two activity peaks. More users are active at the peorids, which generate more retweets and favorite counts.

Q 2) What are the most popular breeds?

Answer:

*** P1 column ***

The most popular breeds of dogs are Golden Retriever and Labrador retriever. In both 'retweet_count' and 'favorite_count' the most popular breeds are the same, only the order changes.

The four most popular breeds are:

Golden retriever

Labrador retriever

Chihuahua

Pembroke

An image depicting a Golden retriever on average gets 11,310.85 favorites and 3,874.47 retweets.

*** P2 column ***

The most popular breeds of dogs are Labrador retriever and Golden Retriever. In both 'retweet_count' and 'favorite_count' the most popular breeds are the same, only the order changes.

The four most popular breeds are:

Labrador retriever

Golden retriever

pomeranian

cardigan

An image depicting a Labrador retriever on average gets 1400000 favorites and 450000 retweets.

*** P3 column ***

The most popular breeds of dogs are Labrador retriever and chihuahua. In both 'retweet_count' and 'favorite_count' the most popular breeds are the same, only the order changes.

The four most popular breeds are:

Labrador retriever

chihuahua

kelpie

golden retriever

An image depicting a Labrador retriever on average gets 700000 favorites and 200000 retweets.

Q 3) What is the relationship between dog stages and retweet counts and favorite counts ?

Answer:

We can observe that in our plot that most of the tweets have less than 40,000 favorites and less than 20,000 retweets. We also notice that from the equation for the line of best fit that there tends to be 3 times as many favorites as retweets. However we notice that as the values get larger, the ratio of retweets to favorites get larger (further above the line of best fit).

When we highlight the dog stages (ignoring those that were not classified), we see the smaller values bunch below the original line of best fit.

In []:

In []: