act_report

June 21, 2022

0.1 Report: act_report

WeRateDogs is a humoristic Twitter account. They publish tweets about dogs (mostly) with the name of the dog, a picture or more, a rating and a funny text explaining the context of the picture.

I have gathered three datasets regarding the tweets published by WeRateDogs. I assessed the data, cleaned it and merged it into a master dataframe. Then I proceeded to analyze the data briefily and visualize it.

0.1.1 Analysis and visualisation

A very interesting part of the recovered data included image recognition. The pictures from the tweets had been put through a neural image recognition program that made three guesses (p1, p2 and p3) on the content of the pictures. Next to each guess, there was a confidence interval for the guess produced by the neural program i.e. how likely it is that the guess is right.

Are there breeds easier to recognise? Some breeds seemed easier to identify than others. So I asked myself which three breeds got -in average- a higher confidence interval?

The three breeds are the following :

```
1st Komondor        -  97.25%
```

```
In [3]: from IPython import display
        display.Image("./Komondor.png")

Out[3]:
```

```
    2nd Clumber          -   94.67%
```

```
In [5]: from IPython import display
        display.Image("./Clumber.png")

Out[5]:
```

3rd Brittany Spaniel - 87.45%

```
In [6]: from IPython import display
        display.Image("./Brittany-spaniel.png")

Out[6]:
```

Image credits: all three pictures above are from WeRatedDogs. These are the pictures that got the highest confidence interval for each of the breeds in the image recognition podium.

Which are the more common dog names in the dataset? One of the limitations of the final dataframe is that some information is partially missing. That's the case for 526 tweets out of the 1657 we kept in the cleaned dataframe. That accounts for roughly 32% of the tweets. Nevertheless, I was curious about the most common or popular dog names in the dataset. Here are the top five:
- Cooper(9 entries) - Tucker(9 entries) - Charlie(8) - Oliver(8) - Lucy(8)

```
In [2]: from IPython import display
        display.Image("./pie-names.png")

Out[2]:
```

Most popular dog names
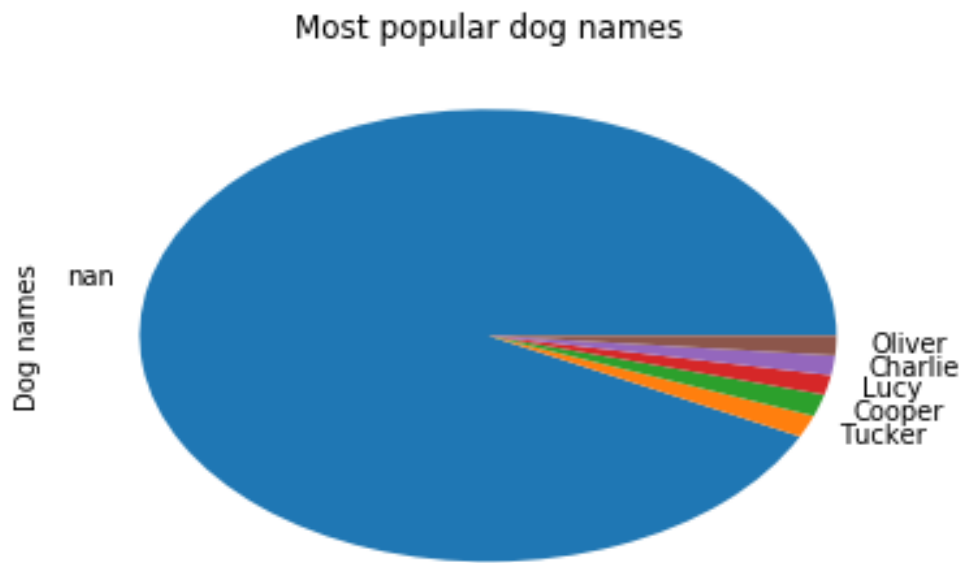
The pie-chart displays the 5 most popular names in our dataset: Tucker, Cooper, Lucy, Charlie and Oliver. The blue slice corresponds to the 526 tweets without a dog name.

Name of the best rated dog(s)   Not all tweets' text includes the name of the dog in it (when it's a dog). So I wondered if the best rated dog's tweet did.

After a quick code research I found out that there was not one, but 22 tweets and dogs (at least, because sometimes there is more than one dog in the picture) that got the best score.

The best rating given by WeRateDogs in the tweets from our dataset was 1.4 or 14/10.

Some of these dogs' names in the wall of fame are : Cassie, Puffy, Emmy, Walter, Cermet, Kuy, Doobert, Gabe, Bo, Gary, Ollie, Loki or Sundance. For the rest of the 22 dogs with the highest rating the name wasn't mentioned in the picture of the tweet.

Is there any correlation between likes and ratings?   Finally, I asked myself whether there was a correlation between the rating given to the dogs by the account WeRateDog and the likes the tweet -as well as the photos and the dog(s) in it- received from Twitter users.

In order to answer to this question more easily I plotted the ratings and the like_count in a scatter chart.

```
In [3]: from IPython import display
        display.Image("./scatter-rating-likes.png")

Out[3]:
```

Correlation between Like_count and Rating
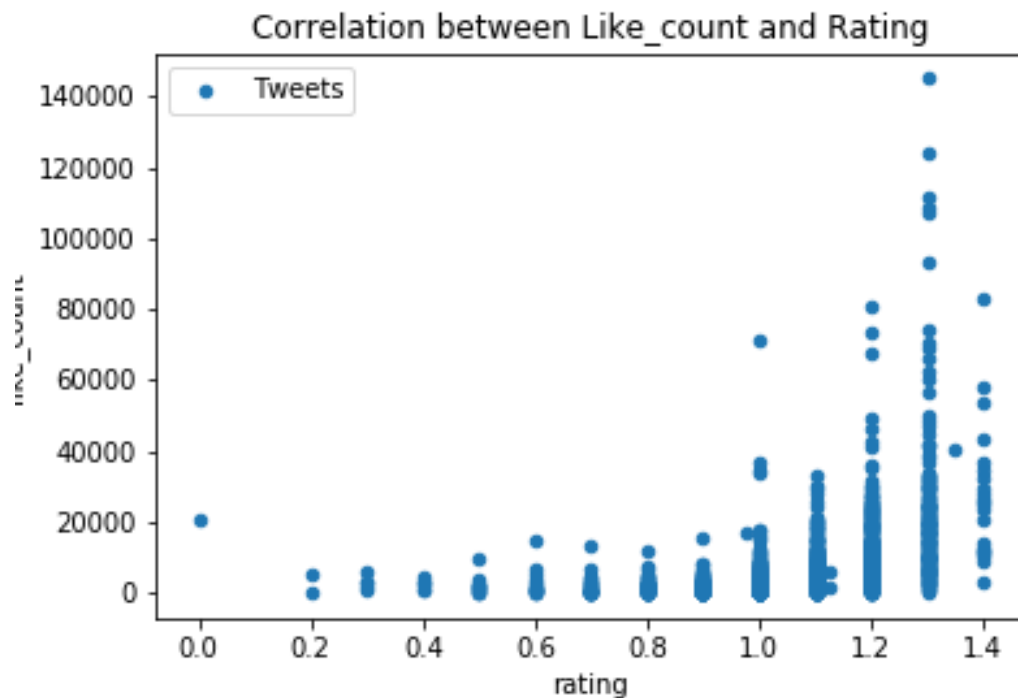
In the scatter plot we can see the rating given by WeRateDogs to the dog(s) in the X axis. In the Y axis, we can see the like counts from Twitter users. Once we are over the rating of 1 on the horizontal, there are more points (representing tweets) that have a higher score (over 60K). On the other hand, tweets that got a rating below 1 or 10/10 got mostly less than 20K likes.

Therefore, we can conclude that there is a quite strong positive correlation between likes and rating. That is, that best rated tweets got generally more likes from Twitter users.