Analysing & Visualizing Data

Introduction

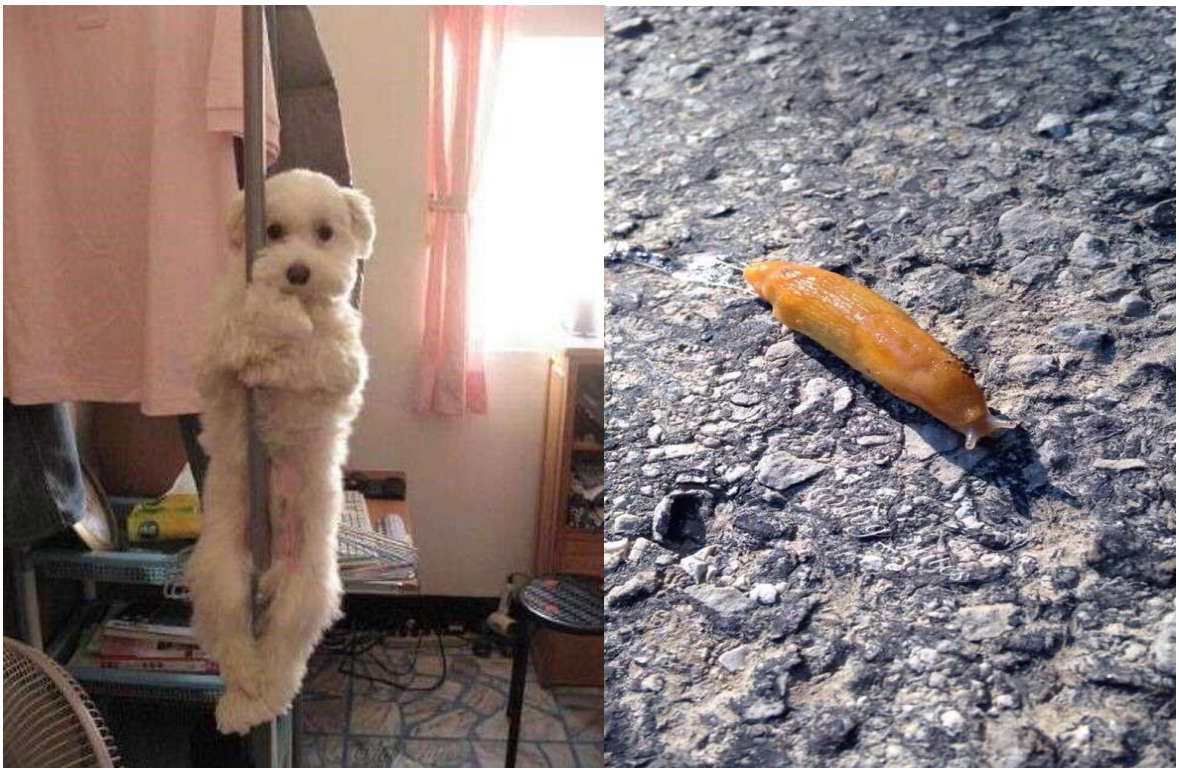
This report aims to provide insights and visual displays based on the analysis made in the project called Data Wrangling Project – Twitter Account WeRateDogs.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. WeRateDogs asks people to send photos of their dogs which are rated on a scale of one to ten, but are invariably given ratings in excess of the maximum, such as "13/10". WeRateDogs has over 4 million followers and has received international media coverage.

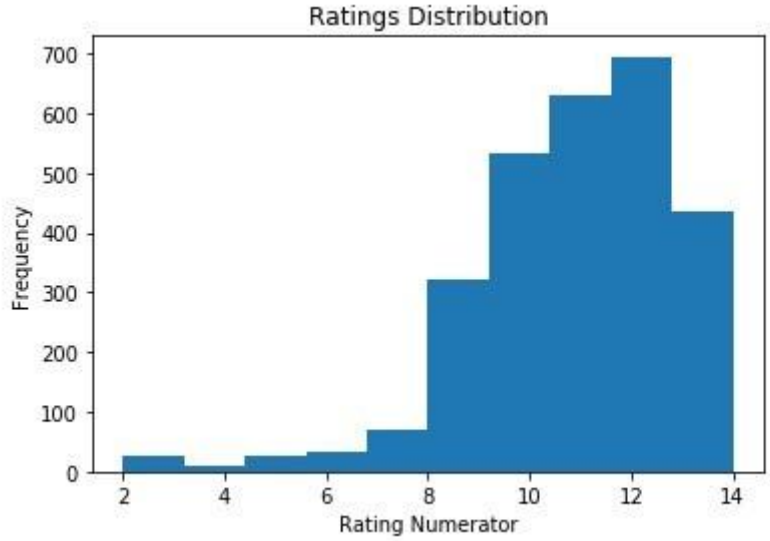
Analysis & Visualizing Data

1. The dataset contains 2786 observations and 21 variables.
2. The highest rating numerator is 1776 and is awarded to a dog named Atticus. The number of likes is around 5500. And it has been retweeted 2700 times.

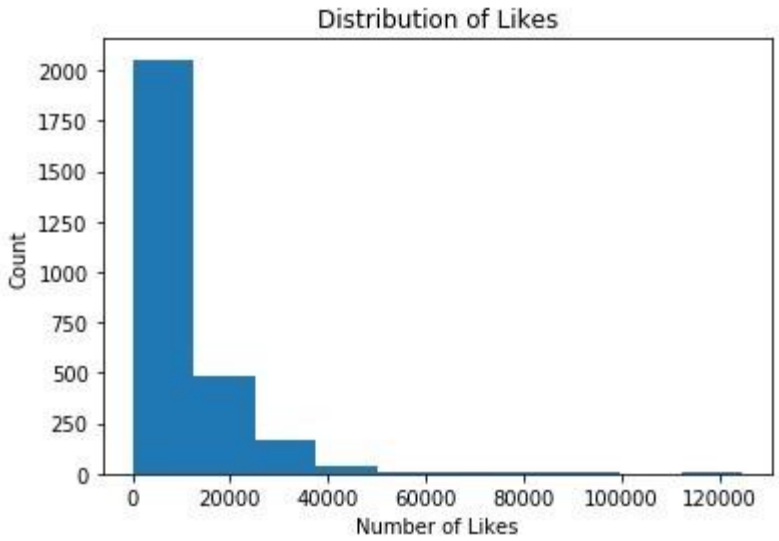
3. The lowest rating numerator is 2 and is awarded to Crystal and Henry (who is not a dog at all). Paradoxically, Crystal has around 5800 likes and 2800 retweets, which means that the ratings are not given based on likes and retweets.



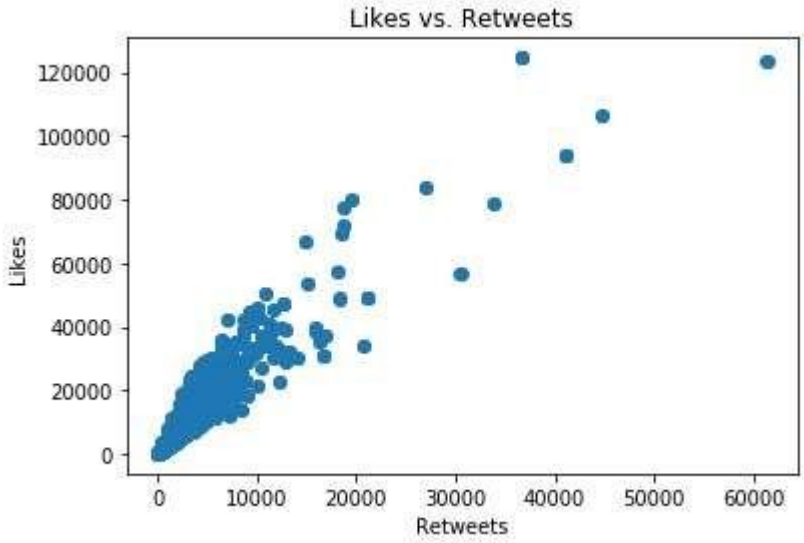
4. Most of the dogs receive a rating of 12. We can also see in the distribution below. We have a left-skewed distribution that peaks at 12.



5. The most common dog stage is pupper, which is a small dog, usually younger, as presented by the @dog rates account. We also have many dogs that are not classified.
6. The highest number of likes is 124644. The median number of likes is 5539 and the mean is 10071. Here we have a distribution of likes which clearly states that we also have outliers, thus, a higher mean.



7. We have a strong positive relationship between likes and retweets. As the number of likes increases the retweets increase also.



8. The dog with the highest number of likes is named Jamesy. He has a rating numerator of 13.

Conclusion

An interesting insight that I found is the fact that the ratings are awarded upfront, so they are not based on likes and retweets. The ratings are subjective, based on the picture of the dog.

From 2786 observations I noticed that the average numerator is 12, so is still near 10.

Surprisingly I discovered that all the pictures are not only of dogs, as we just saw.