Udacity

17/07/2019

@dog_rates Account

Insights and Observations

In this report, I'll study a Twitter account name (WeRateDogs https://twitter.com/dog_rates). The account that I will be studing (and analyzing and visualizing) is Twitter user @dog_rates, also known as 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. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 8 million followers, over 10 thousand tweet and has received international media coverage.

Image – WeRateDogs Profile

@WeRateDogs Account

Insights and Observations

My goal here is to wrangle WeRateDogs Twitter data to create interesting and trustworthy insights and visualizations. And to do that, I'll have to wrangle the data (asses and clean) to have somehow a clean data with higher quality.

Insights

The insights represent some of the observations and conclusions I found during my assessing and cleaning the data. Here I'll represent six insights with some visualizations to give you a better understanding.

Insight-1:

The tweet prediction with 100% confidence was wrong. Also, the algorithm predicts that some images doesn't have a dog in it although it was a dog picture. Thus, the training algorithm for this neural network needs to be modified to avoid any similar issue in the future. As an example, for the inaccurate date, The following picture was linked to tweet_id= 667866724293877760. The following picture shows a dog sits on a sofa, but the algorithm predicts that this image isn't for a dog and its confidence is 100%.

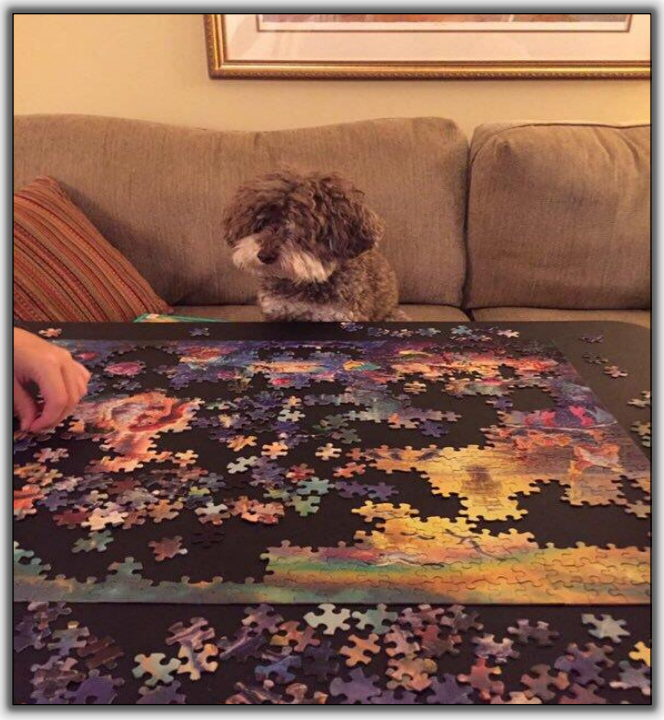
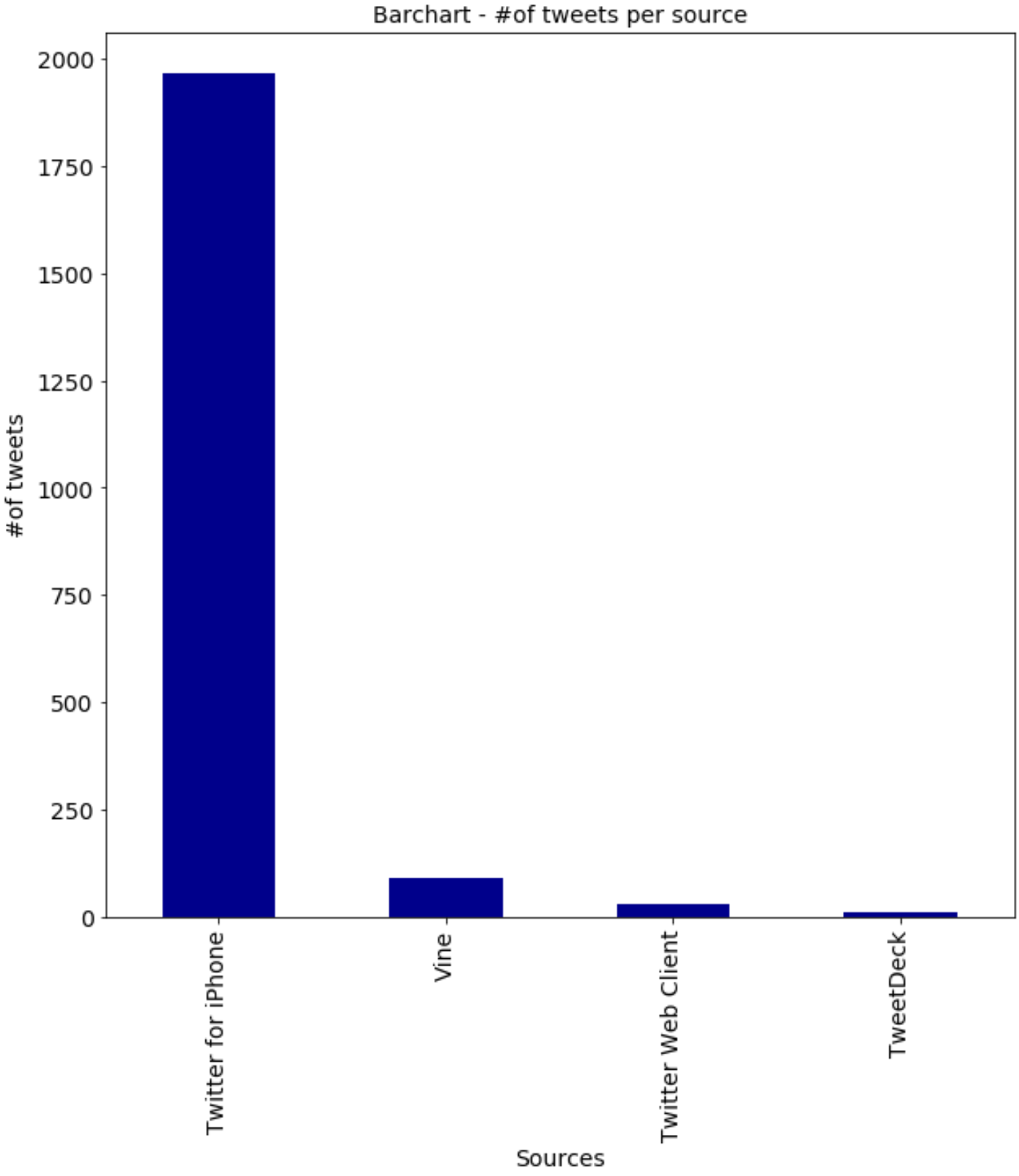


Image – a dog on a sofa

Also, the algorithm predicts the dog stage for some of the images in tweets but it still predicts that the image doesn't have

Insight-2:

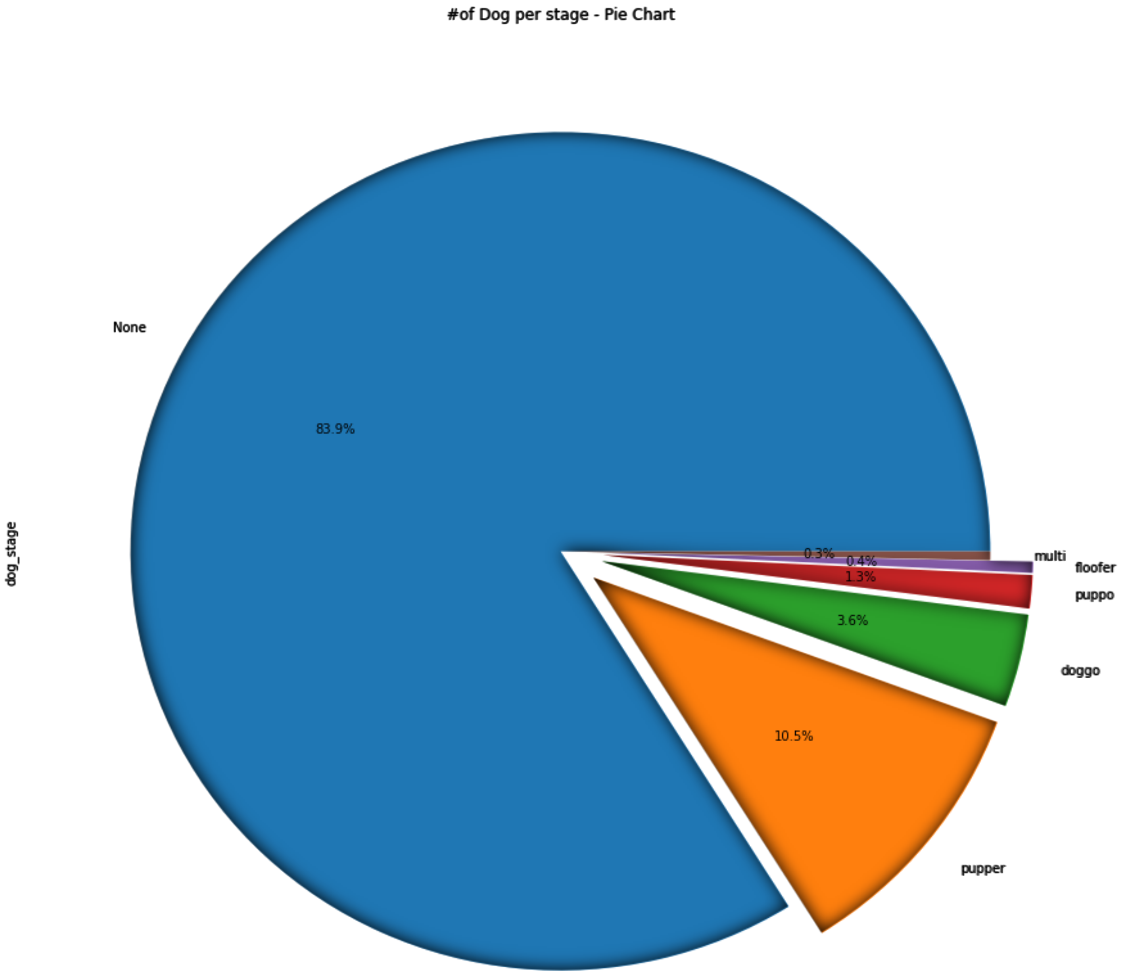
The following bar chart represents the sources for tweets' links. The source of most of the tweets was from IPhone.



Bar chart – Tweets' Source

Insight-3:

Most Images have unknown names. And this because that most of the tweets don't have name for the dogs. Also, the algorithm couldn't identify the dog stage for most of the images



Pie chart – Dog Stage for the image

We can see that most tweets' entry has value of 'None' for the dog stage. In second place, comes pupper stage.

Conclusion

This twitter account has a rich and valuable contents that, if the account owner cleaned it and assessed it carefully, can lead to creating one of the best datasets to train neural networks. But for now, the dataset has many quality and tidiness issues that need to be rectified. Also, the account sometimes tweets images of other animals which can give inaccurate predictions.