Data Analysis Report

By Hani Jubail

Overview

This project is about analysing data from a Twitter account called WeRateDogs.

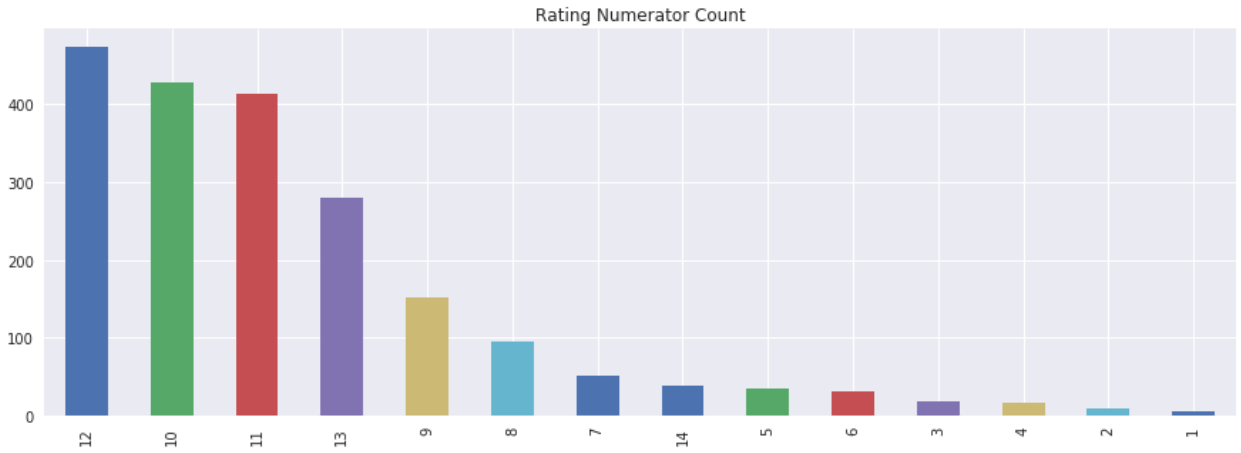
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

The data sets for this project contains 22 columns and 2075 entries between 11-2015 to 09-2017 .

With the twitter data we also used and image prediction data that was developed as a part of neural network project in Udacity's Data Scientist Nanodegree, Machine Learning Engineer Nanodegree and Artificial Intelligence Nanodegree programs.

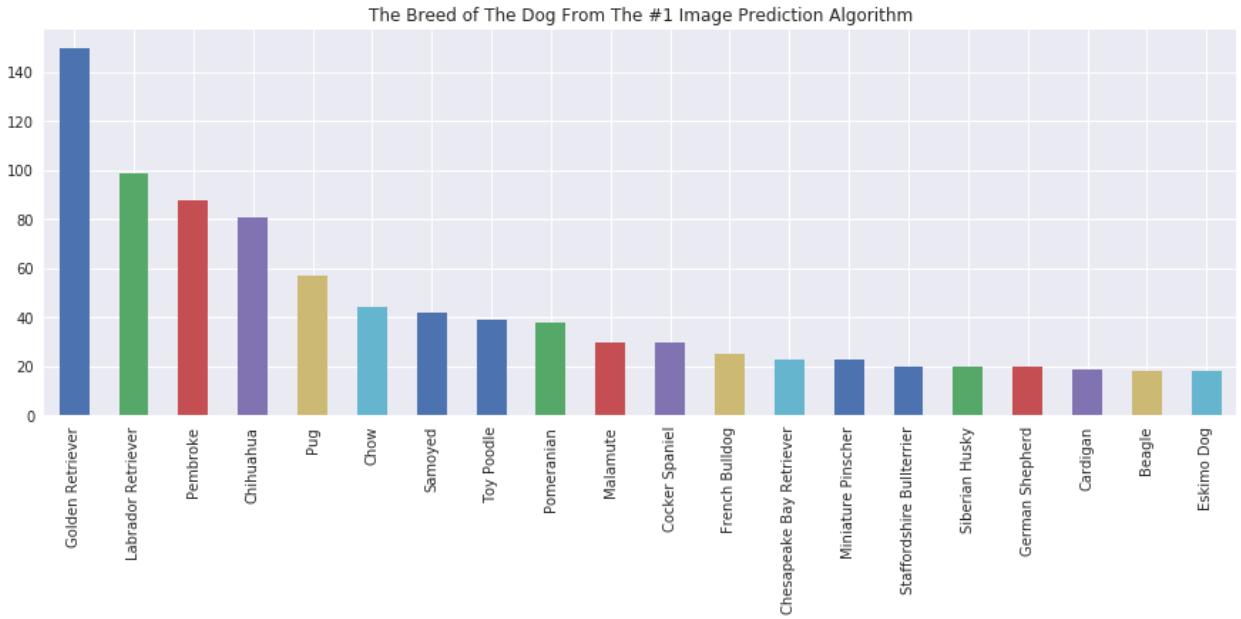
Visualization

Rating count



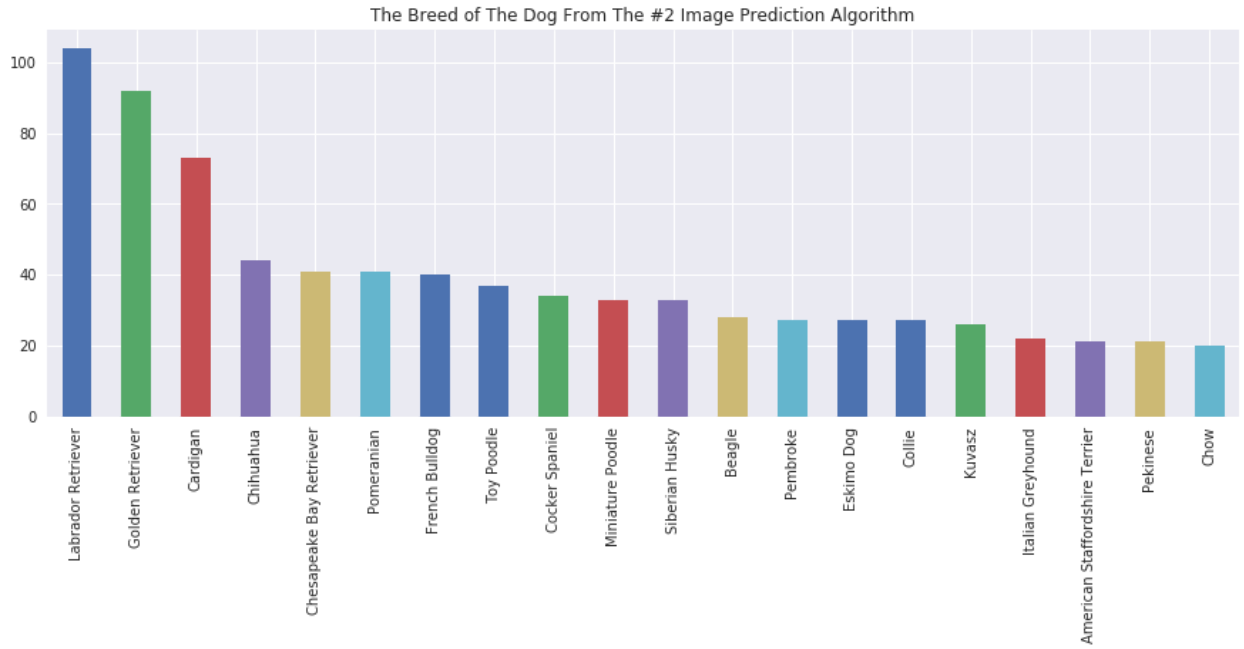
From the bar plot above we can see that the most common rate is 12,10,11,13 and there is a small amount of a lower or higher rates for the dogs .

The Breed of The Dog From The First Image Prediction Algorithm



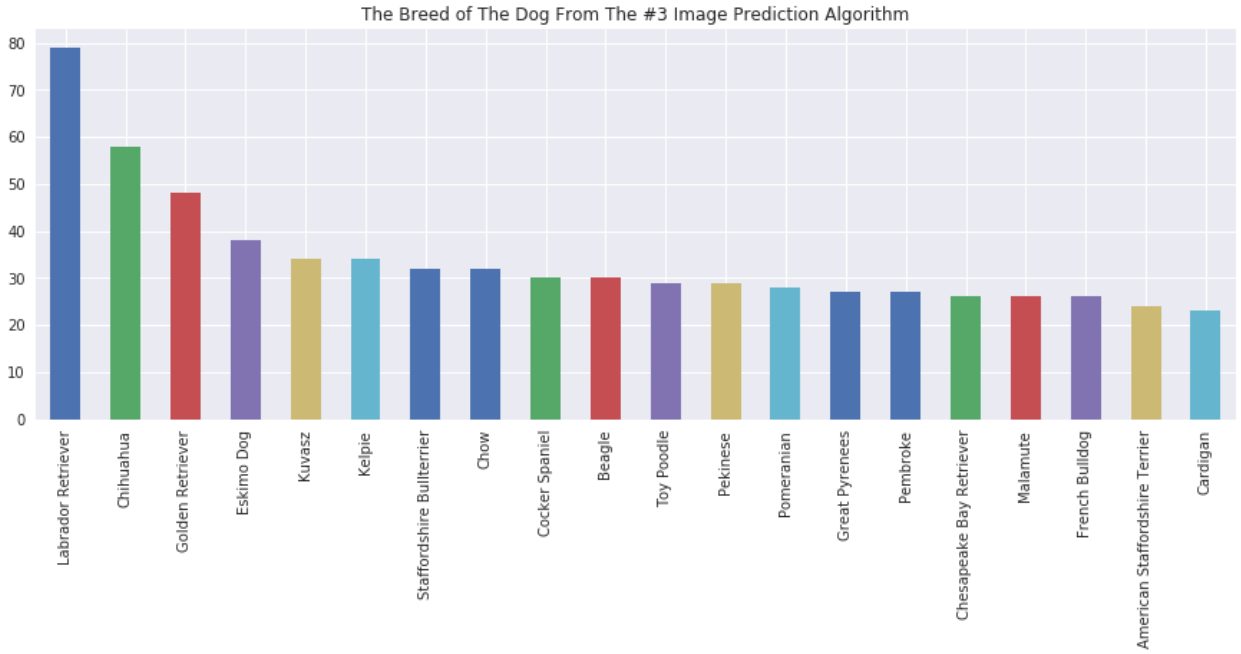
From the bar plot above we can see that the first image prediction algorithm gave us the most popular breed of dogs from the tweets is Golden Retriever over 140 dogs , Labrador Retriever at 100 dogs , Pembroke and Chihuahua around 90 dogs .

The Breed of The Dog From The Second Image Prediction Algorithm



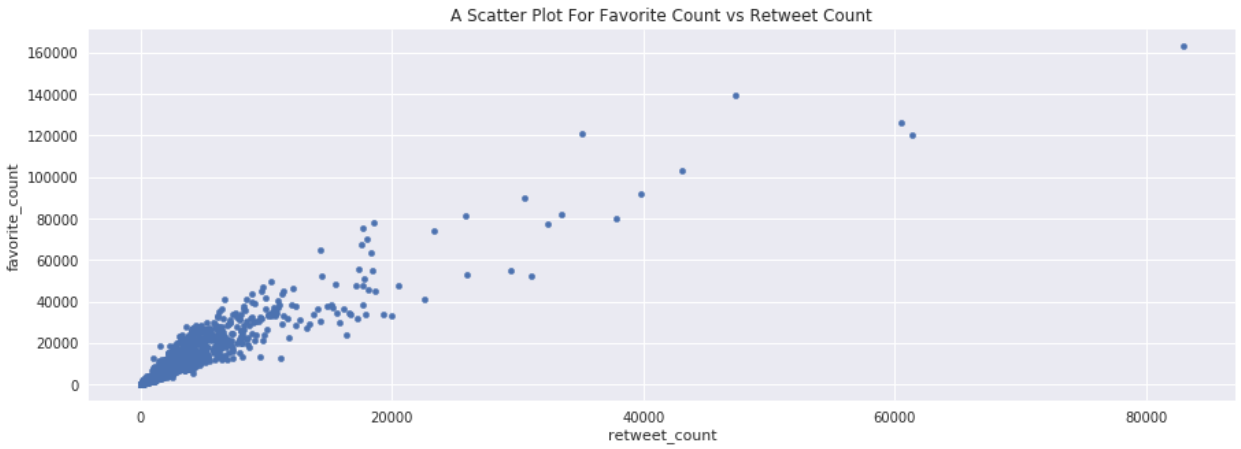
The 2nd algorithm most popular breed of dogs from the tweets is Labrador Retriever ,Golden Retriever,Cardigan and Chihuahua.

The Breed of The Dog From The Third Image Prediction Algorithm



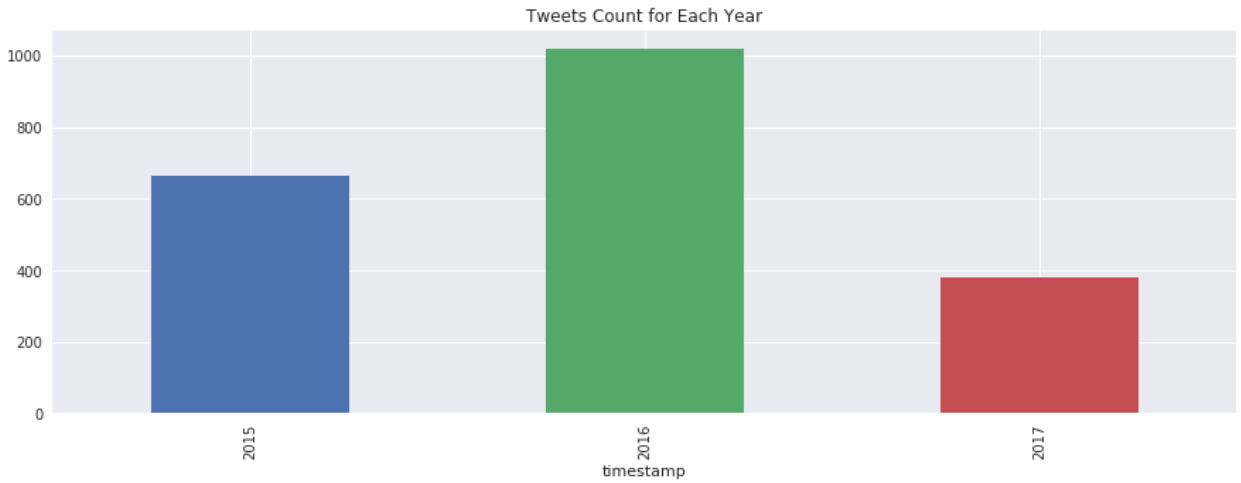
For the 3nd algorithm, the most popular breed of dogs from the tweets is Labrador Retriever,Chihuahua,Golden Retriever and Eskimo Dog.

Correlation Between Retweets and Likes



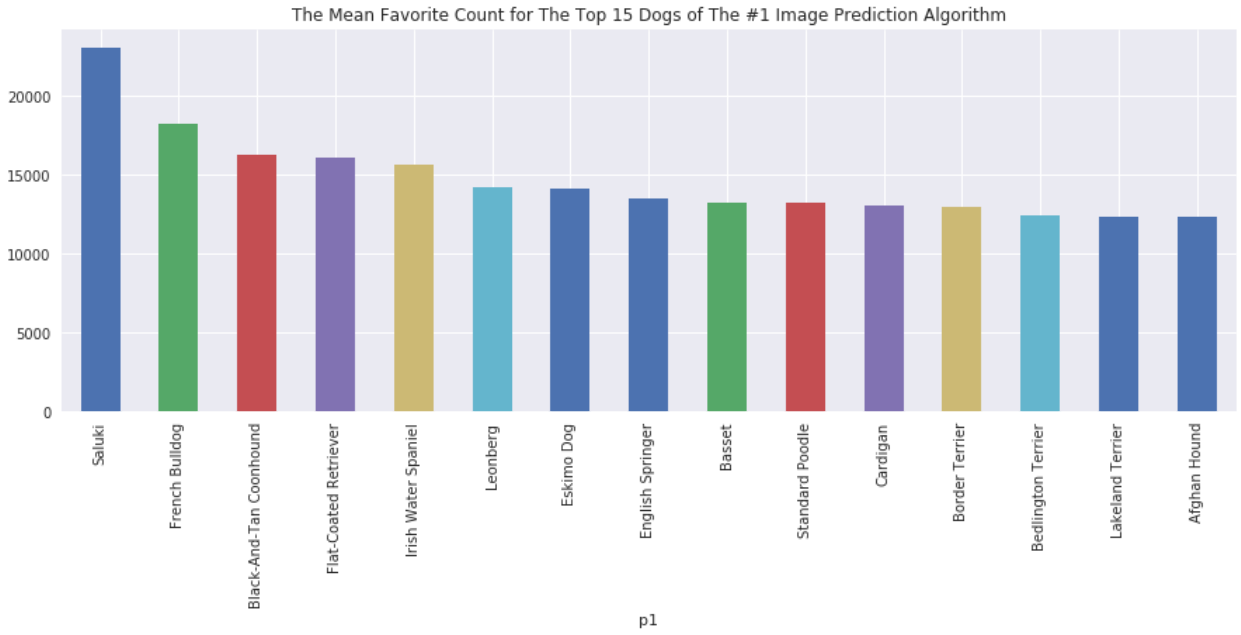
The scatterplot above shows a very strong correlation between the number of retweets and the favorite count which makes a lot of sense .

Tweets Per Year



In 2015 WeRateDogs account tweeted around 700 tweets and the most active period and population was in 2016 with over a 1000 tweets then it reached all time low in 2017 with almost a 400 tweets to 09-2017 .

The Most Liked Dog From The Tweets



So Because The First Image Prediction algorithm have The highest Confidence level mean we can see that "Saluki" is The most liked Dog from WeRateDogs Twitter account .

Conclusion

- The WeRateDogs Twitter account shows us how user react to the photos and the tweets that contains pictures of the dogs .
- The image prediction algorithm gave us an idea about the breeds of dogs in those tweets and how popular some dogs are .
- When it comes to dogs rating the owner of the twitter account “WeRateDogs” is the only one who rates so of course we can’t take those ratings as an accurate measure of how people would like the dog in the tweets .
- instead of taking the rates by the owner of the account as the measure we can use the retweets count and favorites counts to understand how the users react the a tweet .
- The owner of the account used an iphone device for tweeting for the vast majority of the tweets .
- If i have enough time i would go deeper into the data set to find other interesting relationships between the variables .