

# Analysis of WeRateDogs Twitter Data

Hagop Bozawglanian

## Intro

As a student of the Udacity Data Analyst Nano degree, we were given a prompt to wrangle, clean, and analyze the data from the Twitter account WeRateDogs.

We gathered data from three sources:

1. An Archive of tweets from the WeRateDogs twitter, given as a CSV File.
2. A TSV of image predictions that were given to us via link and download (programmatically).
3. A live feed from Twitter using the Tweepy python library.

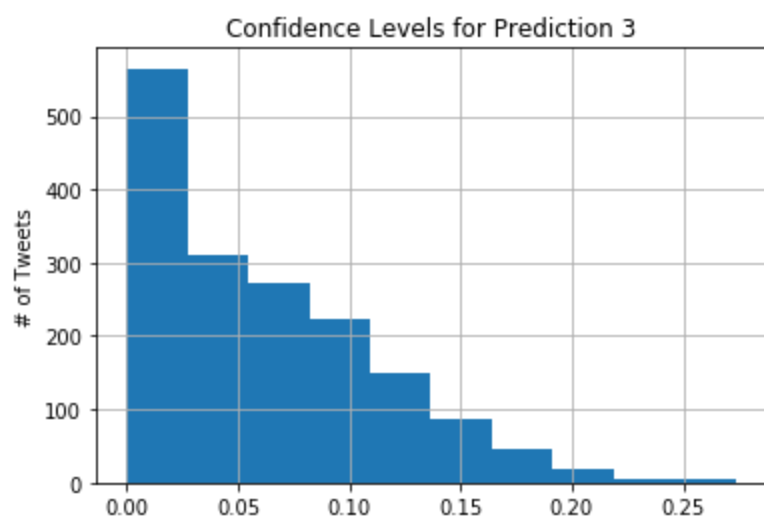
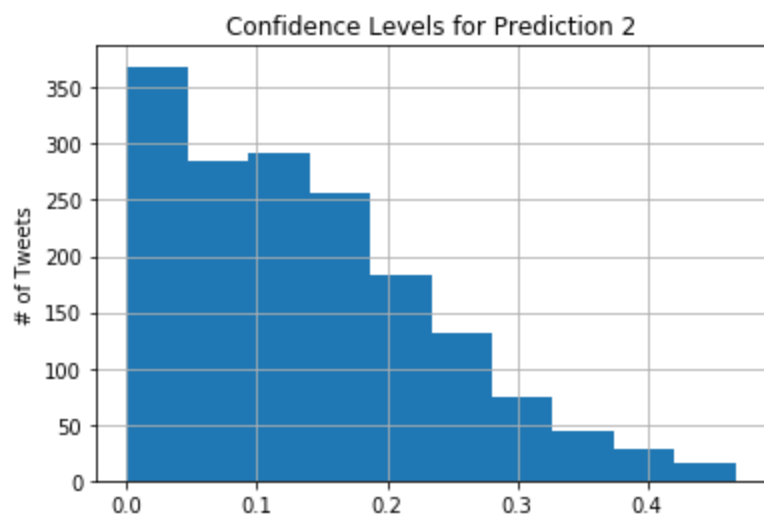
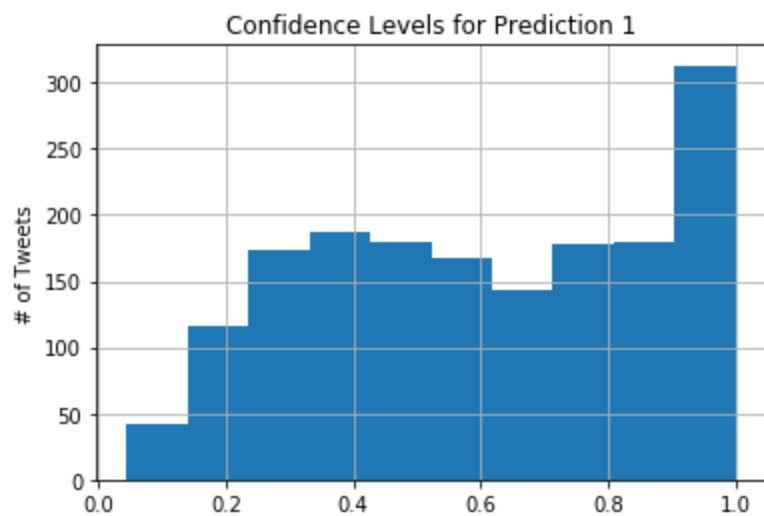
At the end of the wrangling and cleaning my data set consisted of BLANK original and unique tweets (retweets and replies were removed).

I was able to analyze the data and analyzed three observations from my findings.

## Observation 1

With this we can see that 1679 out of the 2060 rows had at least one dog in the three predictions and we can see the distribution shifting from left trailing to right trailing as the predictions continue, which is what we would expect in this situation. This is due to the image recognition model recognizing non dog objects.

I was able to create some visualization to showcase the confidence levels of the predictions made through the 3 iterations:



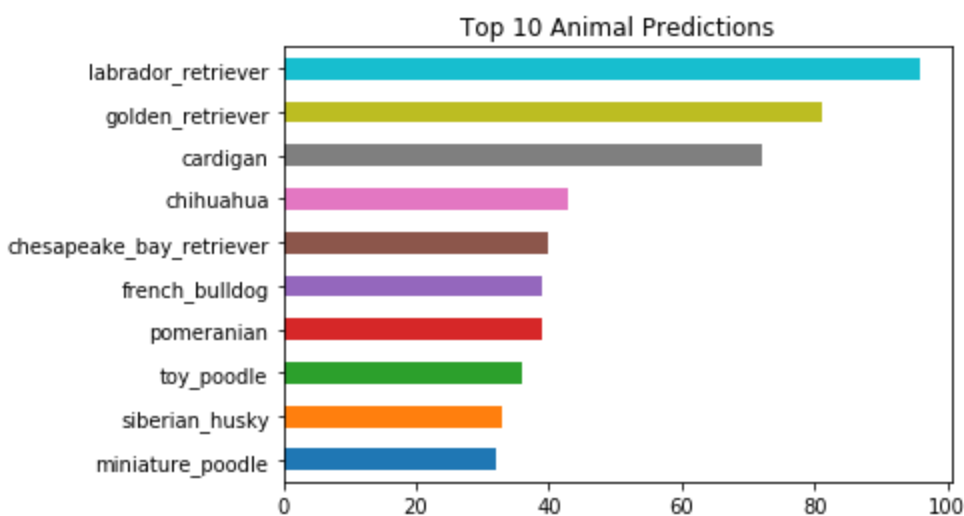
Observations made about the distributions:

- Prediction 1: Confidence ranges from 0 to 100%
- Prediction 2: Confidence ranges from 0 to 50%
- Prediction 3: Confidence ranges from 0 to 30%

I would say the prediction AI model needs some more training to more accurately produce correct information.

## Observation 2

I was able to take the predictions and map them to the top animals seen within the data set.



I choose to show the top 10 predictions based off the first prediction made since it has the highest chances of being correct. We can see the most popular dogs and use this data in marketing or other areas.

## Observation 3

I also used the data to categorize the types of dogs into the dog stages.

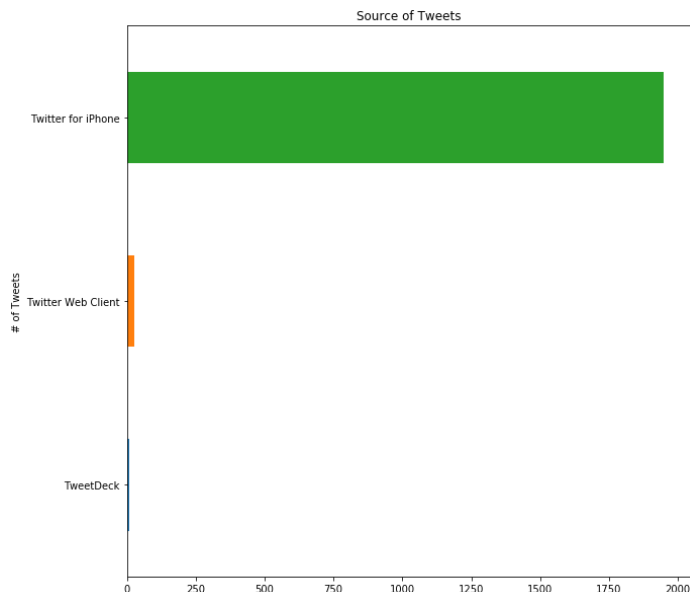
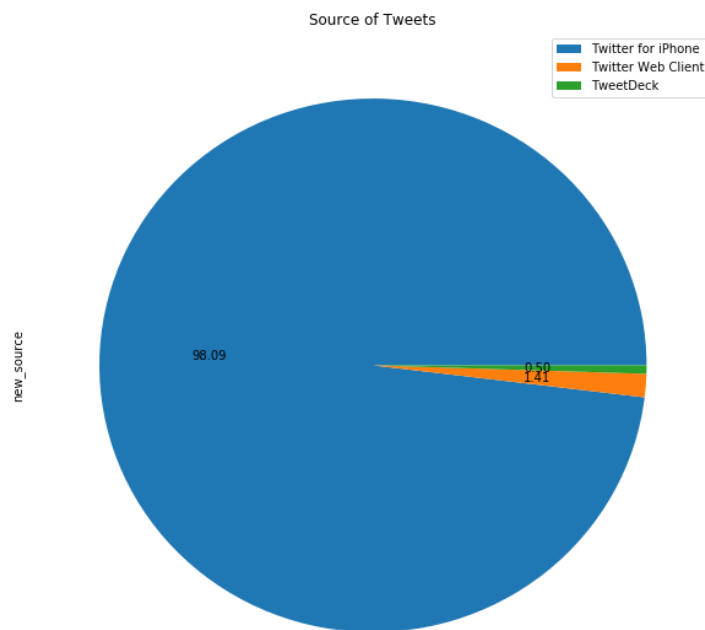
- doggo: 73 of 1987 (3.67%) tweets
- floofer: 8 of 1987 (0.40%) tweets
- pupper: 212 of 1987 (10.67%) tweets
- puppo: 23 of 1987 (1.16%) tweets
- total dog: 316 of 1987 (15.9%) tweets

We can also use this data along the data above with the definitions of the stages to see the most popular dogs.

## Observation 4

The final observation made was the source of the tweets that were made. We can see that there are three different sources we received tweets from:

1. Twitter for iPhone
2. Twitter Web Client
3. TweetDeck



To no surprise we can see that 98% of the tweets that came in are from the iPhone.