

# Analyzing and Visualizing Data

## INTRODUCTION

---

[WeRateDogs](#) is a popular twitter account that rates people's dogs with a humorous comment about the dog. We combined the WeRateDogs data with neural network image prediction from Udacity, and analyzed the dogs.

## MOST POPULAR DOG BREEDS

---

According to the image prediction these are the most popular dog breeds in the account:

golden_retriever	139
Labrador_retriever	95
Pembroke	88
Chihuahua	79
pug	54
chow	41
Samoyed	40
Pomeranian	38
toy_poodle	38
malamute	29

15 percent of the dogs belongs to the three most popular breeds (if we accept the most probable image prediction breeds).

## AND THE WINNER IS . . .

---

The name and the image of the most popular dog by retweets is unknown. Its retweet count is 84230.

Fortunately, the score winner is known. He is Atticus, and has a rating of 1776/10. Atticus is a real patriot.



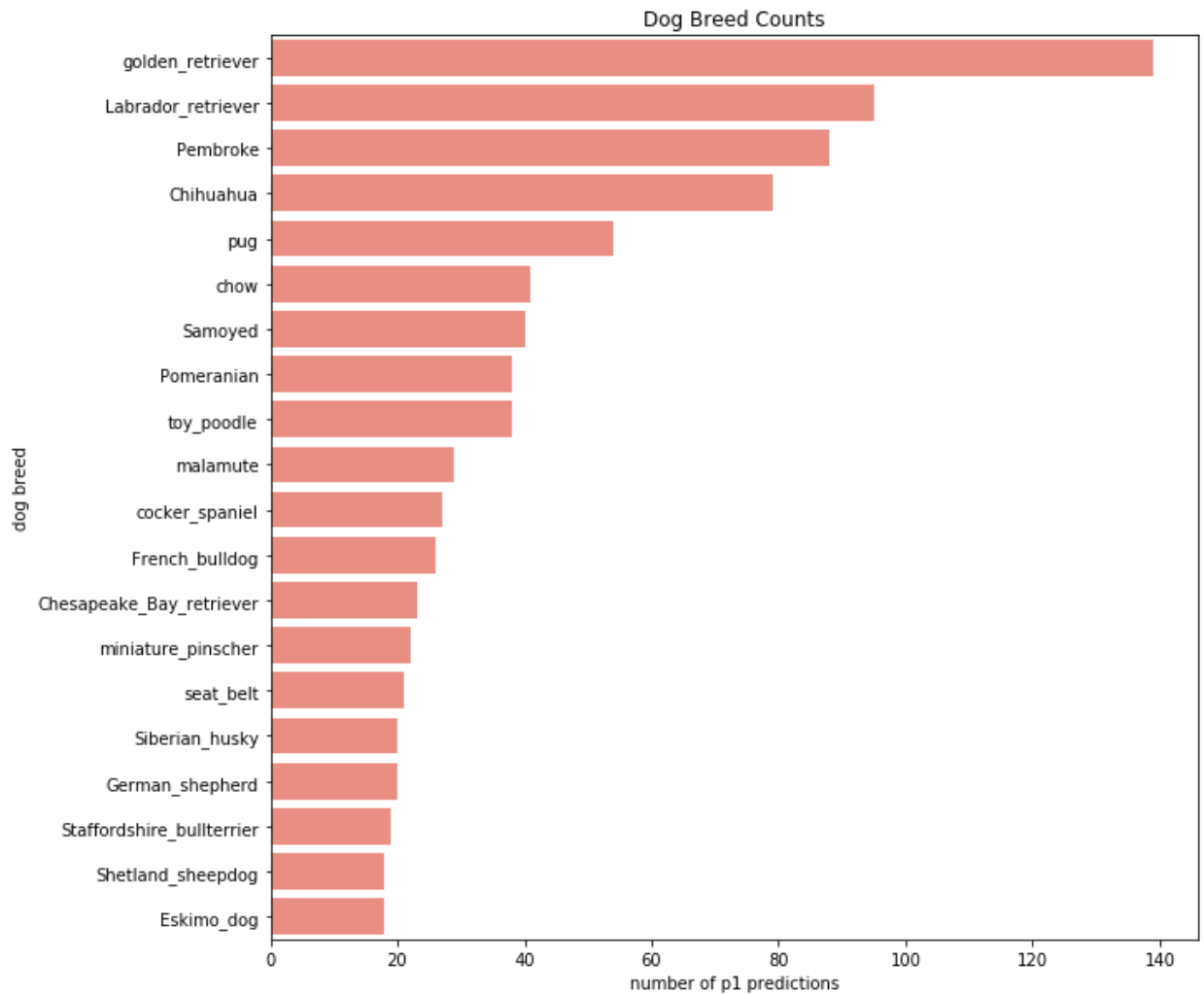
## MOST POPULAR WORDS

---

I wanted to see the most popular words with real information. So I didn't consider personal pronouns, articles, conjunctions and other words needed to build a sentence. I wanted only words with contextual meaning. The most used words with contextual meaning are: meet (216), pet (174), pupper (161) and dog (138). 3 from the first 4 is a kind of "dog word".

## DOG BREED COUNTS

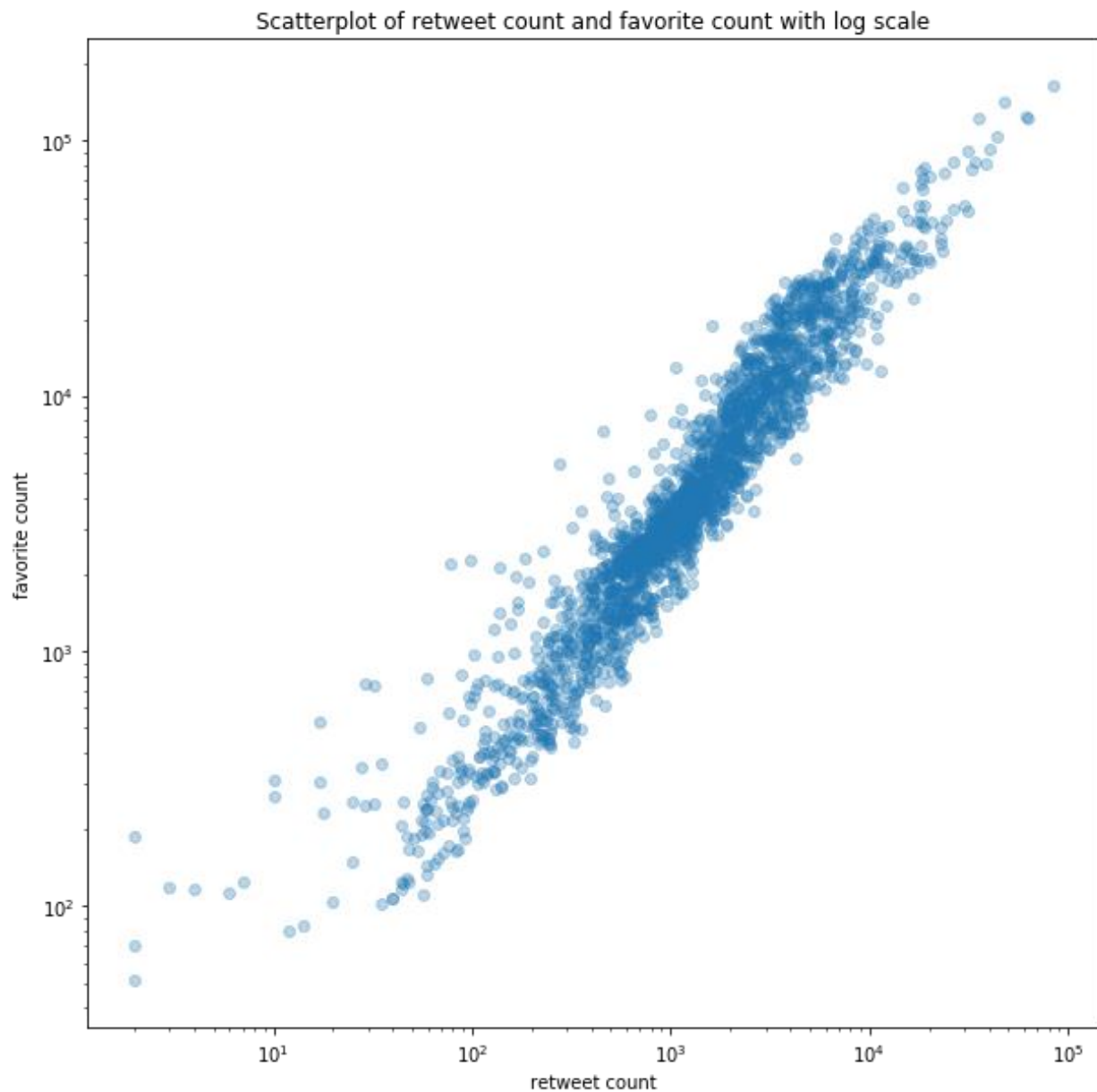
---



This plot shows the most popular dog breeds by occurrence according to the image predictions p1 (most probable) guess. It seems that the retrievers are the stars.

## RETWEET COUNT — FAVORITE COUNT SCATTERPLOT

---



In the plot above we can see that as the larger the retweet count of a dog the larger the favorite count most of the time. This is what we would assume. The scatterplot uses log scales for the visualization.