

UDACITY
DATA ANALYTICS FUNDAMENTALS NANODEGREE

Project 04: Wrangle and Analyze Data
Act Report: Analysis and
Visualizations

Hernán Adasme

1 Analysis and Visualization

Three basic analytical procedures have been performed on the data. Each of those analysis presents a research question to be answered. These questions can be addressed only because the data wrangle process has been successfully completed.

1.1 What are the 10 most popular dog names?

The following bar graph shows the ten most popular dog names. The Two most popular dog names extracted from the data are Charlie and Lucy, two classic names, with eleven counts each, out of 931 different names. The next two are Oliver and Cooper, both with 10 preferences. Holding a really honorable fifth place, the graph shows Penny and Tucker. Special mention to the name "Hobbes" that was found two times, probably in honor of the great philosopher Thomas Hobbes.

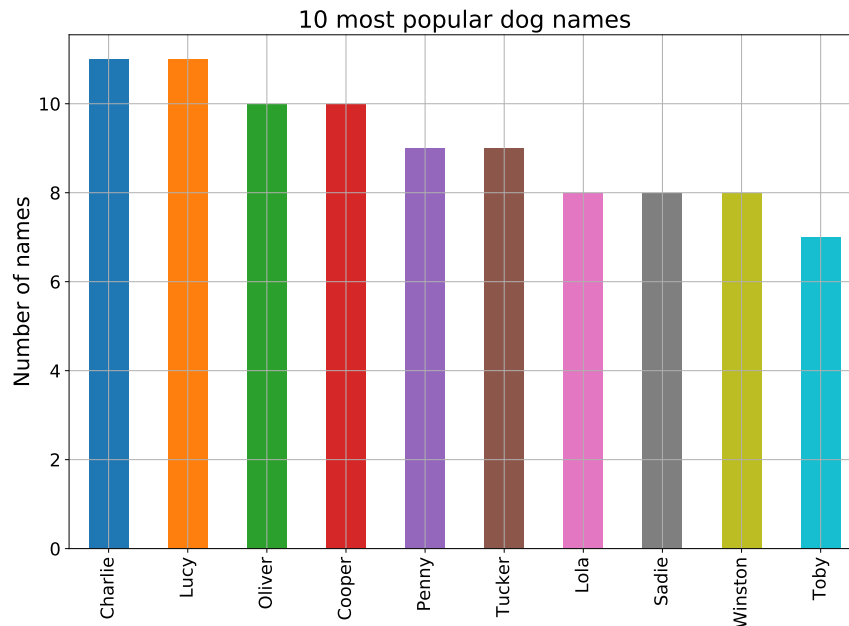


Figure 1.1: Bar plot that shows 10 most common dog names.

1.2 What are the 10 most popular dog breeds?

The following bar graph shows the ten most popular dog breeds predicted by the algorithm. With 139 appearances out of approximately 370 different breeds, Golden Retriever is the sole winner. In a second place, Labrador Retriever shows up on 95

instances. However, we do not know how many of these 95 Labrador Retrievers are Black, Yellow or Brown. Pembroke is placed in a third place, with 88 puppies, and in a fourth place, Chihuahua, with 75 specimens. Even though big breeds like Labrador Retriever and Golden Retriever seem to be Twitter user's favorites, smaller breeds dominate the top ten with 7 breeds on the top ten position. Malamute, the other big breed, is placed on the tenth position.

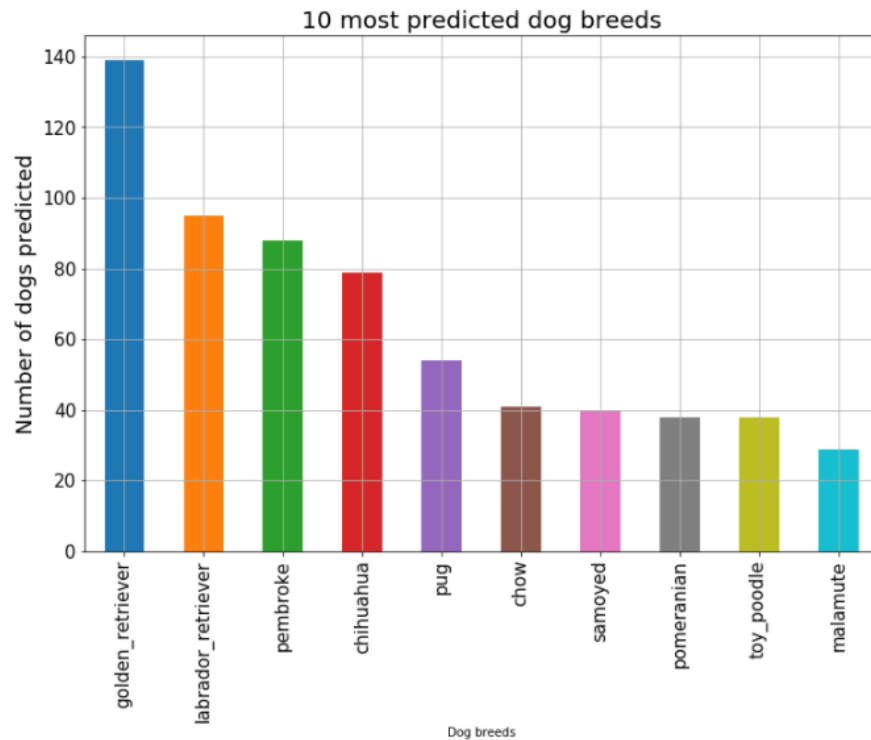


Figure 1.2: Bar plot that shows 10 most common dog breeds predicted by the algorithm .

1.3 What are the 5 fake dog breeds predicted with the highest normalized ratings?

The predictions performed by the algorithms are truly great but in several cases their predictions were not correct. Some of the predictions made by the algorithm were taken from tweets with high normalized ratings. The following graph shows the top five fake dog breeds with the highest rating scores. Bow-Tie, and actually awesome name for a dog, takes up the first place. Interesting other cases are "military uniform" and "microphone".

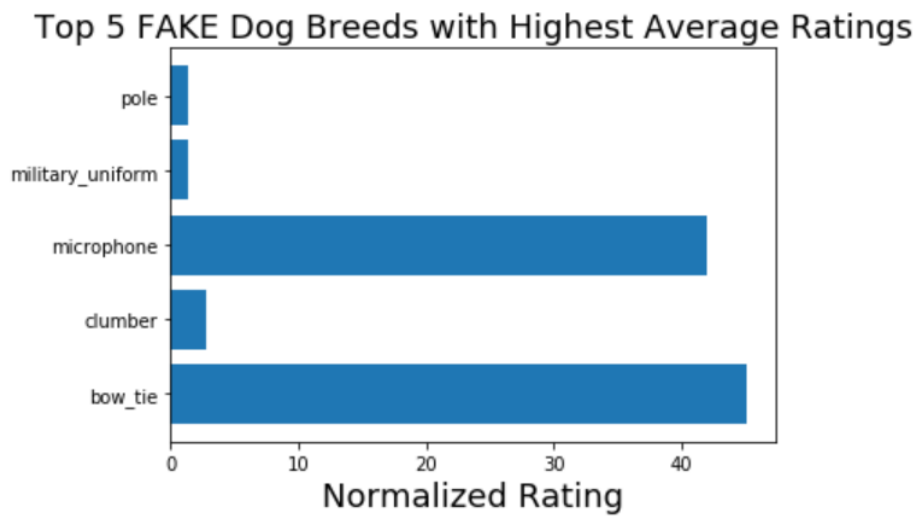


Figure 1.3: Bar plot shows the fake predictions with the highest scores.