Data Insights Report

By Ahmed Abdelhamed Elgayar

The Udacity Data Professional Track Project 2 : this is a report for insights of “WeRateDogs” project.

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”.

When i assess the data , i found that there are incorrect and weird values of denominators and numerators

If we open the expanded_urls in twitter-archive-enhanced.csv file we will find that there are multiplied dogs in the photos and one of them for “Snoop Dogg” so we can’t used this data until we clean it , so i think we should open every weird denominators and numerators and check what is the issue and decide how we deal with it .

Data from image_predictions.tsv and teet_json:

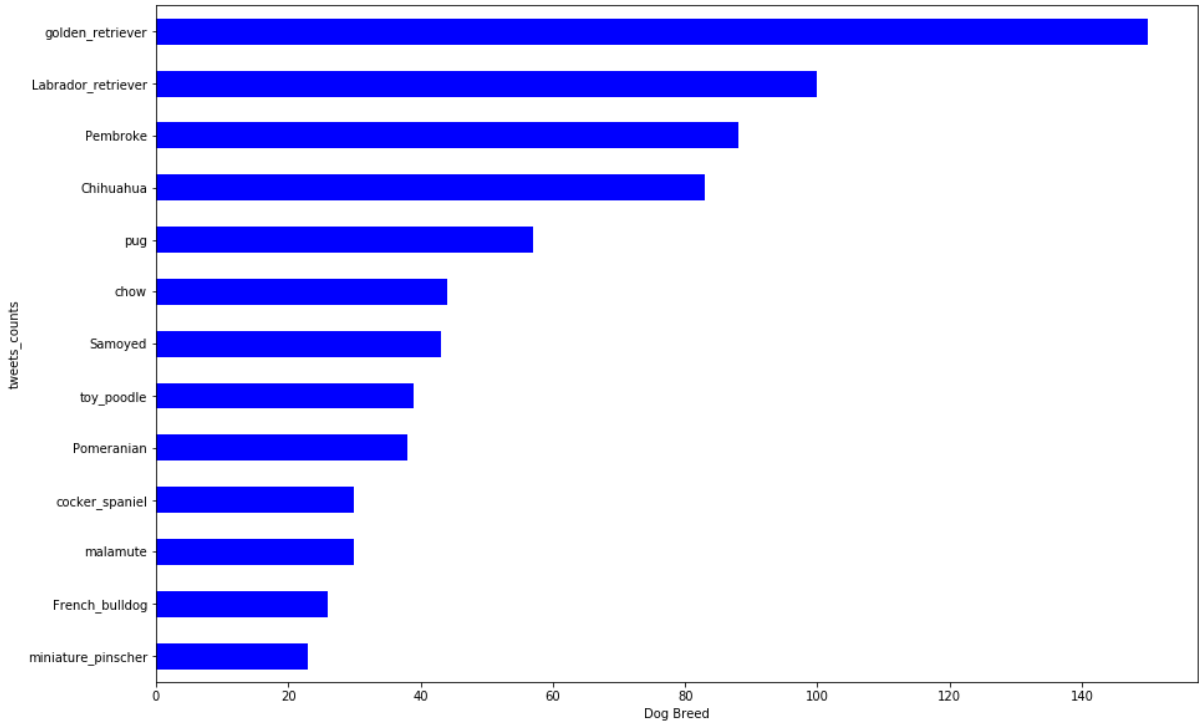
After cleaning and merging the data we found that there are (prediction_level & prediction)

When we check some of them we will find that there are some of photos for dogs and some of them for other animals.

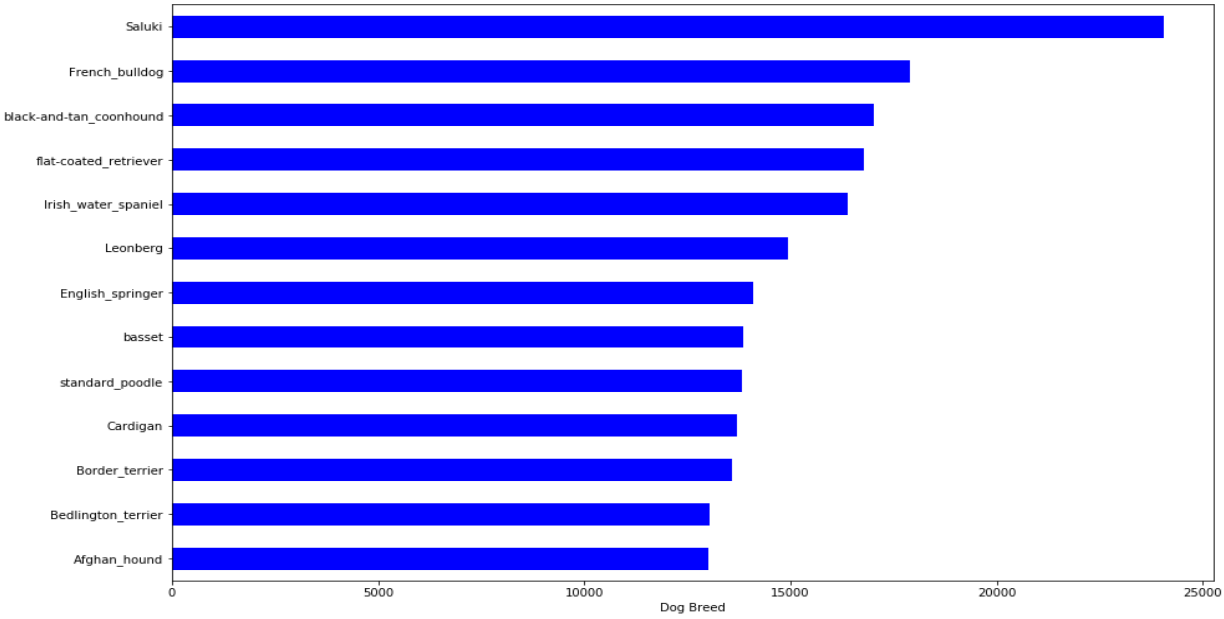
So I prefer to analysis and visual the data for that we are sure it is a dog so I chose the prediction_level “3” and prediction “True”.

Insights from data analysis and visualization:

1) The top dogs breed frequent in the tweets.



2) The top dogs breed ranking in terms of generating favorites.



3) The top dogs breed ranking as per the retweets they had .