Act Report

By: Ahmed Elsayed
December 2020

Introduction

@WeRateDogs is the username of a Twitter account that rates dogs' images and giving them a nice comment on each image. This account founded in 2015 by Matt Nelson, who received international interest and attention for this account.

They have about 8.9 million followers now.

Data Gathering

I used three sources while the data gathering process, and these sources differ from manually and programmatically sources.

The first source was the archive of the Twitter account (@WeRateDogs) that contains all their tweets, retweets, and replies from November 2015 to August 2017. It is a CSV format file that was downloaded directly using pd.read_csv import into a pandas data frame.

The second source of data is the (Image Predictions) file, which was downloaded programmatically from Udacity's server through this [URL](URL) into the workspace, using the requests library.

The third and final source is Tweets Json File, which is a query from Twitter API that using Python's Tweepy library and stored the entire data in a file named tweet-json.txt.
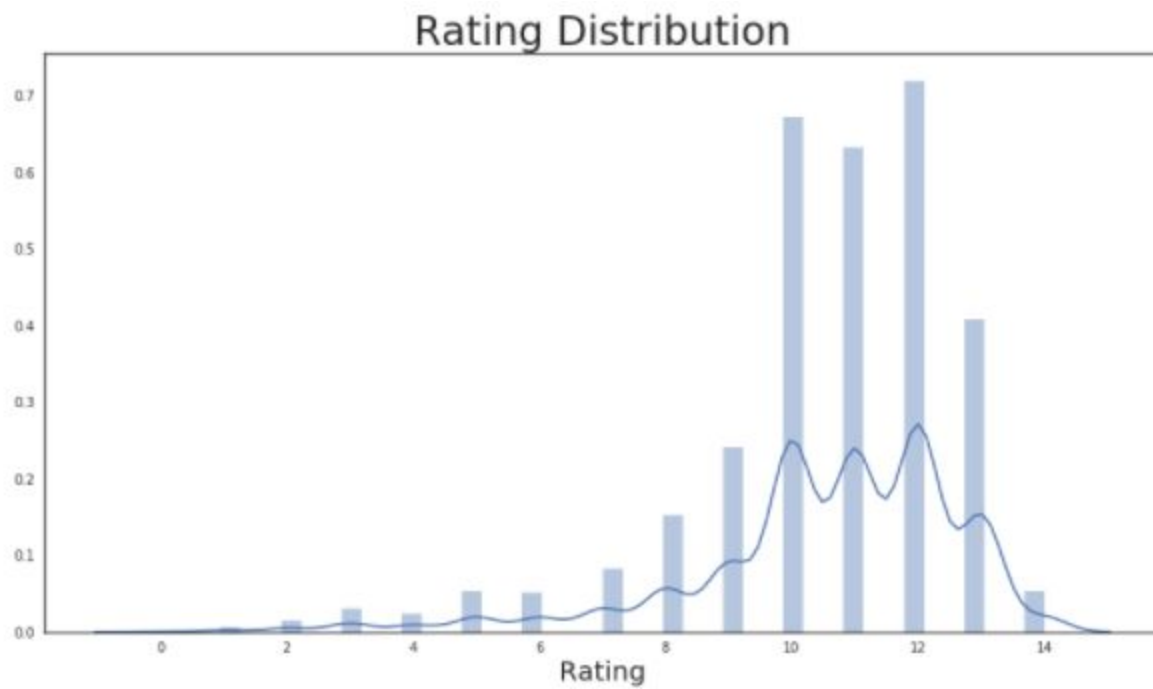
Data Assessing

After gathering the data as explained previously, I copied the three dataframes in order to assess them programmatically using Jupyter Notebook with pandas library, and then visually using Google Sheets and Microsoft Excel.

This process was so important so I can get all the quality and tidiness issues, and then clean all these issues found to be able to start storing and visualizing data.
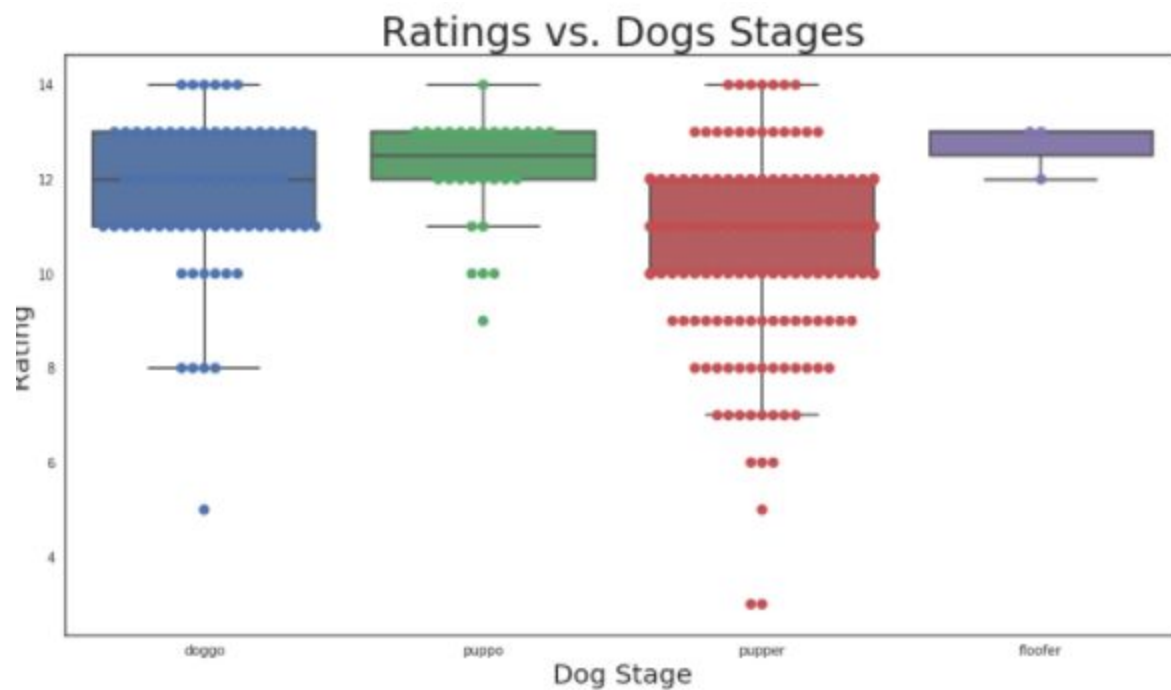
Data Analysis

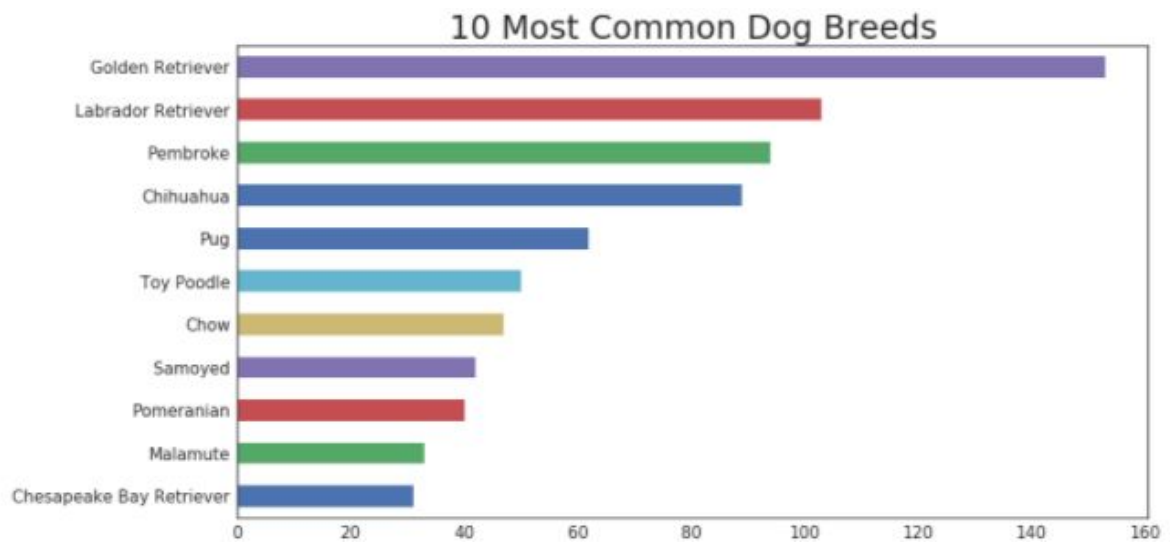My analysis focused on the different factors that affect dogs' ratings such as:
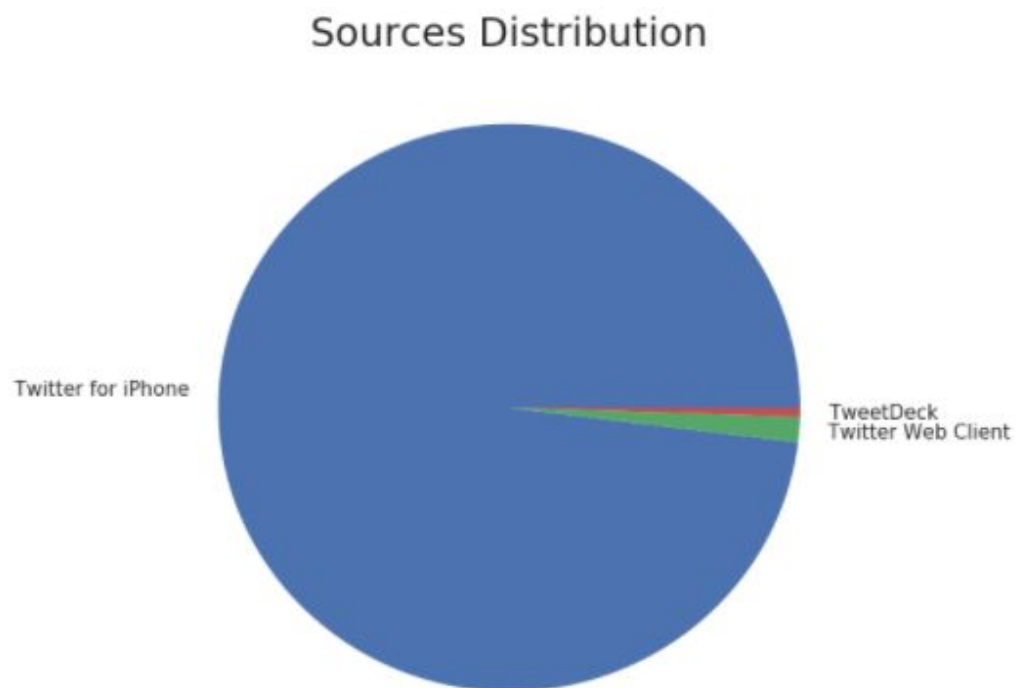
Rating Distribution:

Ratings vs. Dog Stage

Most Common Dog Breeds

10 Most Common Dog Breeds

Source Distribution

Sources Distribution

Average Rating for the 3 sources