WE RATE DOGS DATA ANALYSIS PROJECT

Reporting: wrangle_report

Introduction

This Data Analysis and Visualization project was carried out on the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. 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.

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for us to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. More on this soon.

The software used and Requirements

I completed this project inside the Udacity classroom on the Project Workspace, using the Jupyter Notebook provided there.

For a person who wants to work outside of the Udacity classroom, the following software requirements apply:

- You need to be able to work in a Jupyter Notebook on your computer.
- The following packages (libraries) need to be installed. You can install these packages via conda or pip.
  a) `pandas`
  b) `NumPy`
  c) `requests`
  d) `tweepy`
  e) `json`

- You need to be able to create written documents that contain images and you need to be able to export these documents as PDF files. This task can be done in a Jupyter Notebook, but you might prefer to use a word processor like Google Docs, which is free, or Microsoft Word.
- A text editor, like Sublime, which is free, will be useful but is not required.

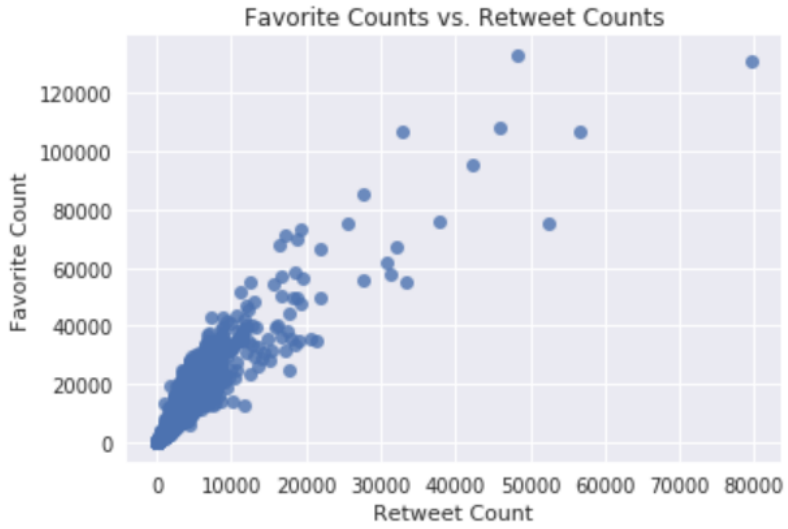
Below are the insights generated:

1. *There exists a highly positive correlation between Favorite couts and retweet counts*
2. *Over 90% of the Tweets are from Twitter for iPhone*
3. *The Most common dog stage is pupper, followed by doggo and the third is puppo.*

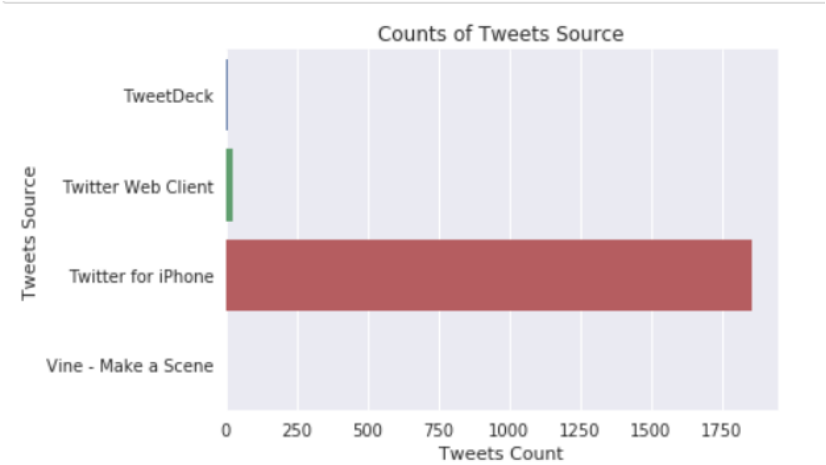
The comparisons through visualizations below support the insights generated:

Visualizations

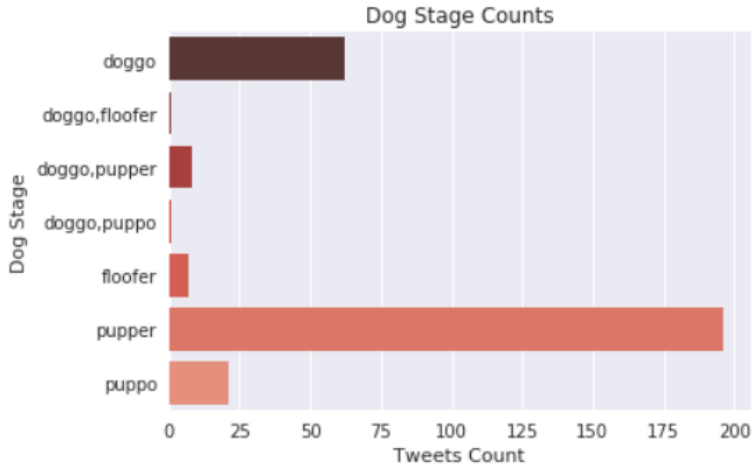
1. *Relationship between 'retweet_count' and 'favoite_count' using a Reg plot*



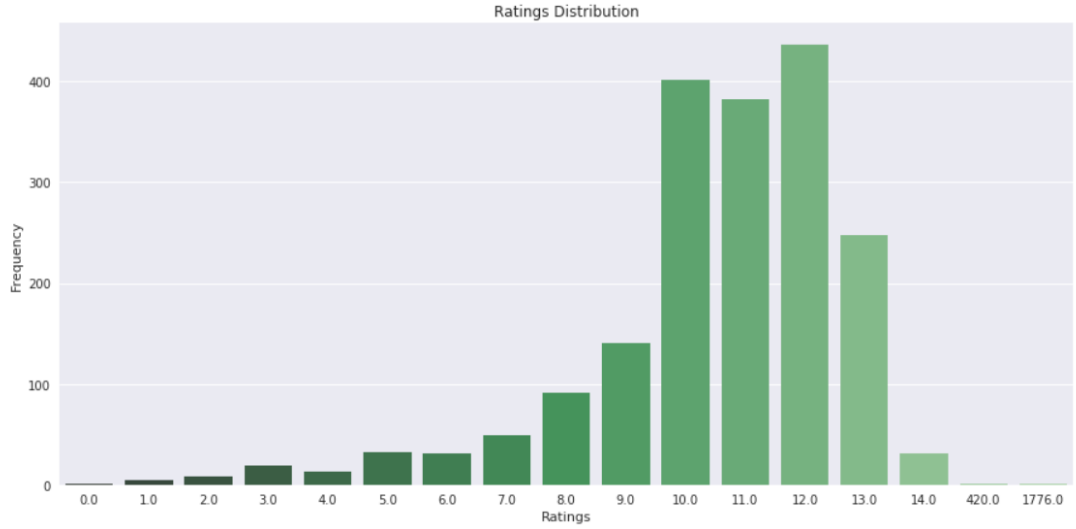
2. *Compare different sources of tweets using seaborn's barplot*



3. *Compare the dog stages using the barplot from seaborn*



4. *Ratings distribution using a bar chart*



5. *The most popular dog names*