Udacity Data Analytics Nanodegree

Project 04 – Data Wrangling

Act Report

Introduction

This report main focus is to make some analysis on the WeRateDogs' tweets. With more than 8.5 million followers, this account is famous for rating dogs in a unusual way, plus, showing pictures of cute puppies.

Their unusual way of rating dogs is what's above the denominator: almost all numbers are higher than 10, so the dog's rates are slightly hightoned. But, according to the page's followers, there are no problems with that, because all they want is to see fluffy doggos.

The dataset provide by Udacity has 2354 rows, but after cleaning this data, there are 2105 rows. Each row has a tweet id, range of the tweets, date, the complete text that was tweeted, number of retweets and how many times was checked as favorite, and a few more interesting data. The date record range of this dataset is from November 15th, 2015 to August 1st 2017, which gives us a good time range for analysis, and some time consistency.

After all wrangling part, that included the gathering, assessing and cleaning, it was possible to create a quick data analysis, a great opportunity to practice everything that have been learned this far.

Analyzing the Data

Looking forward to better understand how the WeRateDogs tweets dataset is displayed, a few questions were made:

1. How's the distribution of tweets length?

   In a quick analysis we can see that WeRateDogs always uses the previous Twitter's maximum capacity of tweet length. Looks like they are always trying to give as much adjectives to the puppies as they can.

   This is distribution was expected since in Twitter users tend to use the whole character space to write their messages.
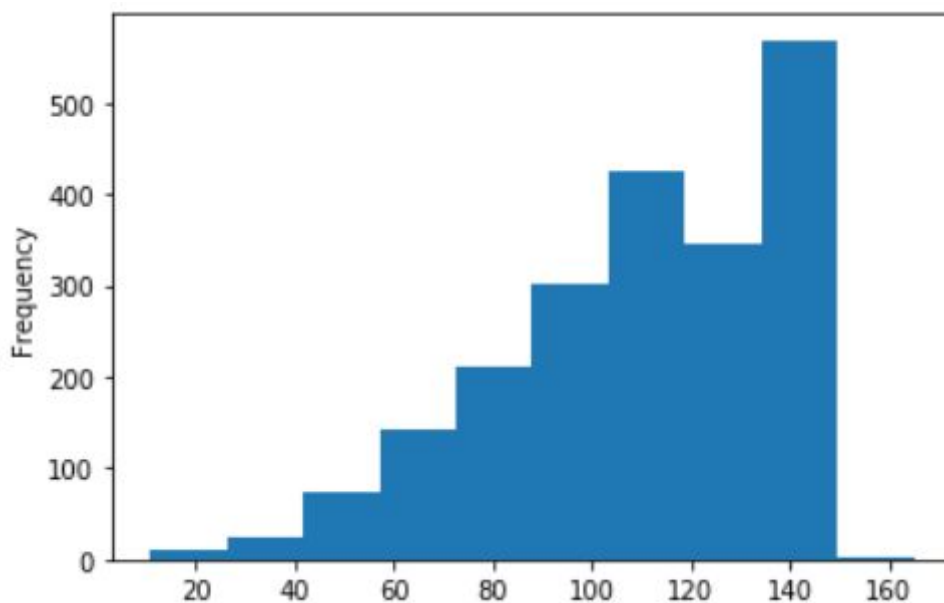
2. What's the distribution of dog's stage?

In this chart, we can see that the most rated dogs are classified as pupper. This is the biggest dog stage, with about 70% of all dogs.
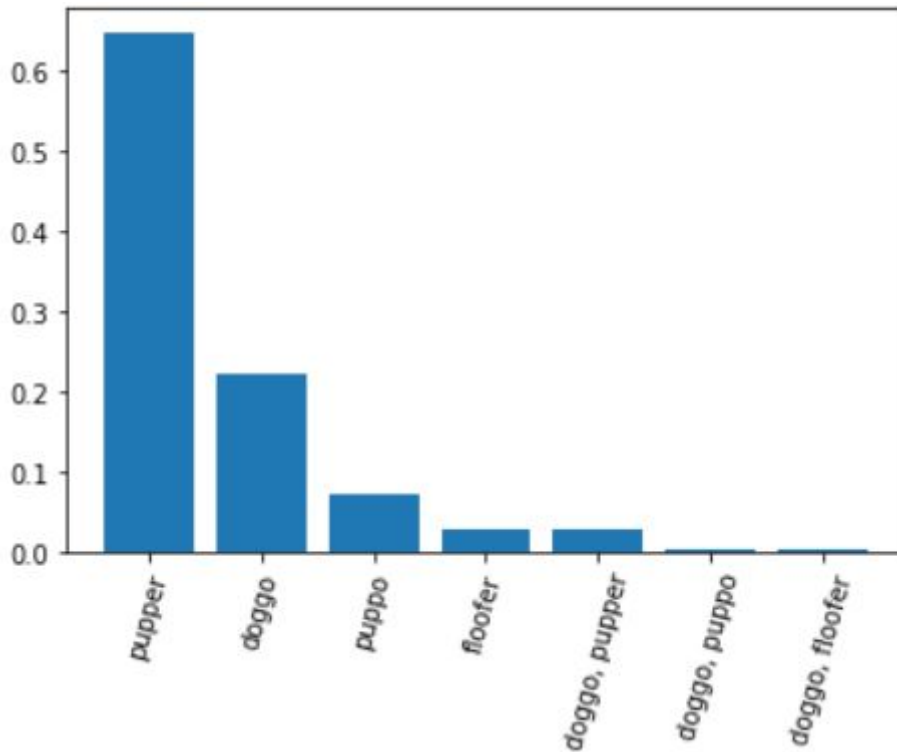
These percentages show that the young dogs are the ones with most appearance. It's very understandable, since every person that gets a new dog want to share cute pictures of them right away.

3. What's the average of dogs rate?

Dogs are cute, that's why they deserve rates above the limit. That's why we can see that the mode of rating is about 12.5. Not a problem, they deserve it!
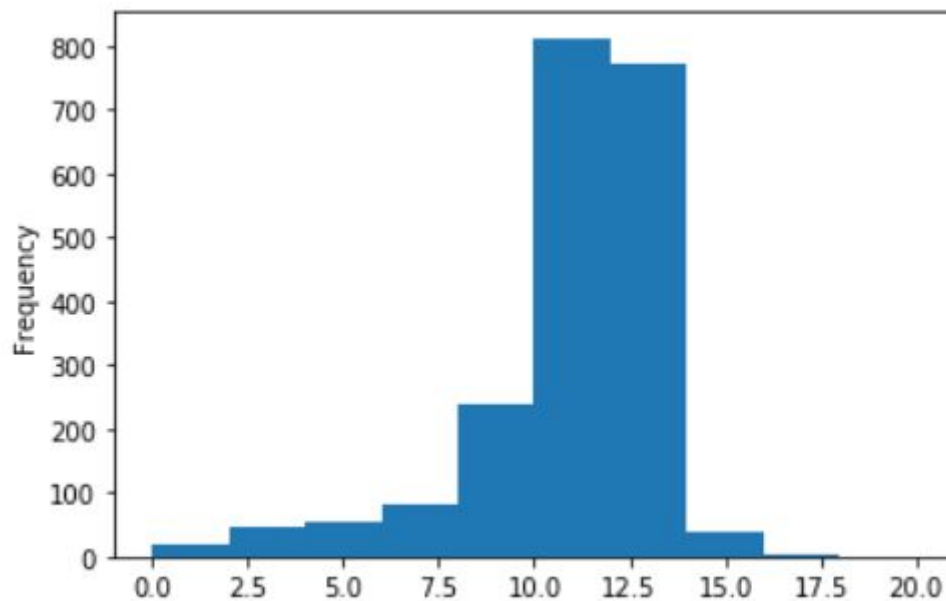
For this quick analysis, I was also expecting this kind of histogram shape. The rates are usually above 10 and not much higher than that. This means that the average rate is a little bit bigger 10, so does the mode.