Data Wrangling

udacity

### Project 4

Student:

Muddassar Sohail

Course:

Data Analyst

Data Analysis and Visualization "WeRateDogs" Data

This project aims at practicing data wrangling techniques to process data so as to make it ready for analysis and visualzations. data set being used os of a twitter initiative @dog_rates, aka WeRateDogs who rate dogs, usually as 11/10 :) This report showcases data insights that were assessed from the data set.

Insight 1

Among different types of dogs, golden retriever gets most frequent ratings, it got over 140 ratings overall while cocker spaniel got lowest number of ratings. Remember this doesn't mean that cocker spaniel got bad ratings; that is a different thing, it only means that it got less ratings as compared to others.
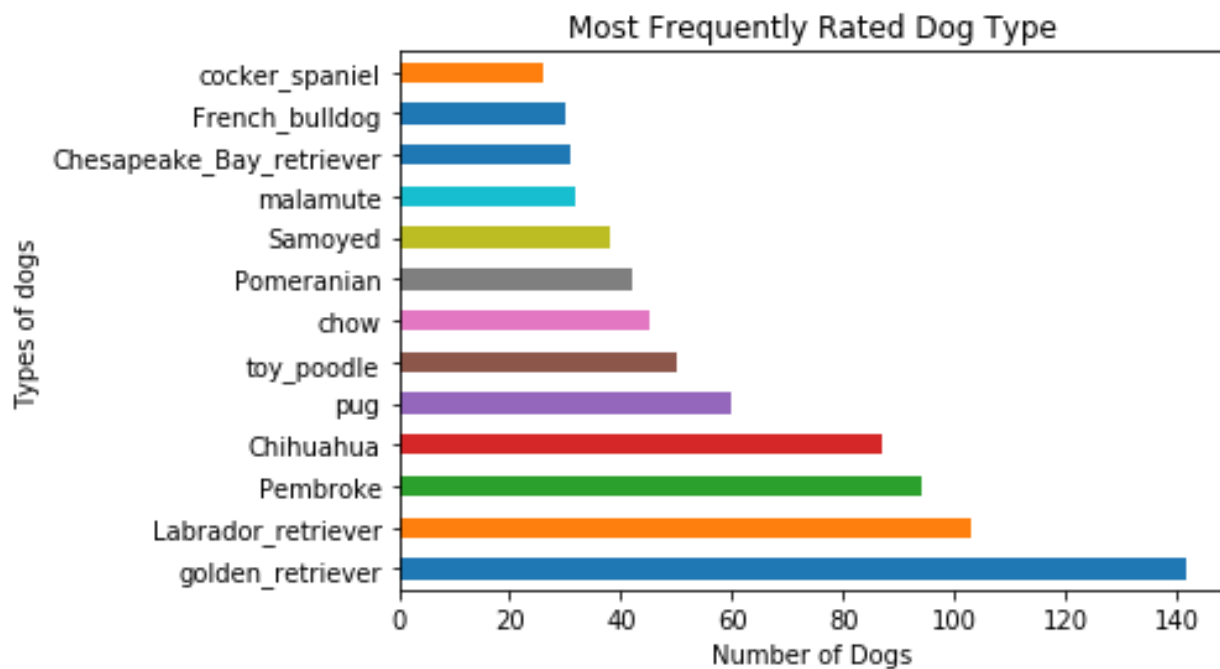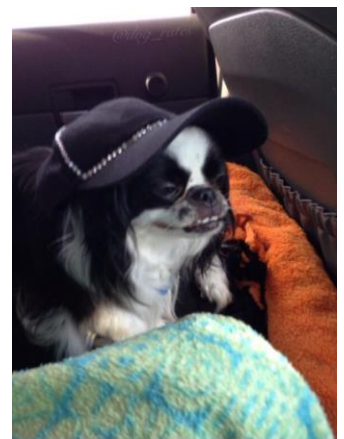
A golden retriever:

Insight 2

Unfortunately, this Japanese spaniel got lowest average rating, a rating of 5/10. Anyways it's still cute. As per this insight, Bouvier_des_Flandres got highest average ratings of 13/10.

Insight 3

We previously saw that golden retriever got most ratings but Bouvier_des_Flandres got highest ratings. So, to conclude a winner we plot this chart, we see that though Bouvier_des_Flandres has highest ratings but it has less tha 10 ratings overall while golden retriever with a slightly less rating got more than 140 ratings. Golden retreiver seems to be the winner in being famous. Other than that we also see that more number of ratings mean a narrow range of average atings, while in case of less ratings there is more diversity in mean ratings.

Average Ratings by Number of Ratings wrt dog type
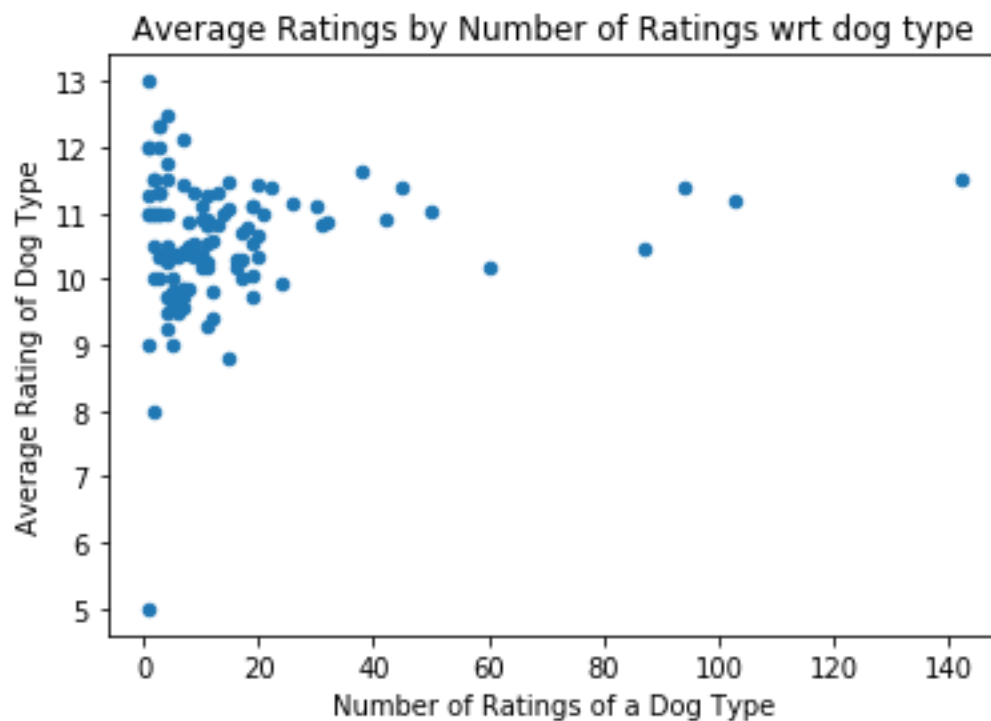
Insight 4

When looking at a glance we see a trend that the higher the ratings the more the number of retweets. But looking closely it was revealed that rating above 14 got lesser retweets as compared to rating of 13. So we can't establish a relationship between these two variables.