ACT REPORT

PROJECT 4:WRANGLE AND ANALYZE DATA

Introduction

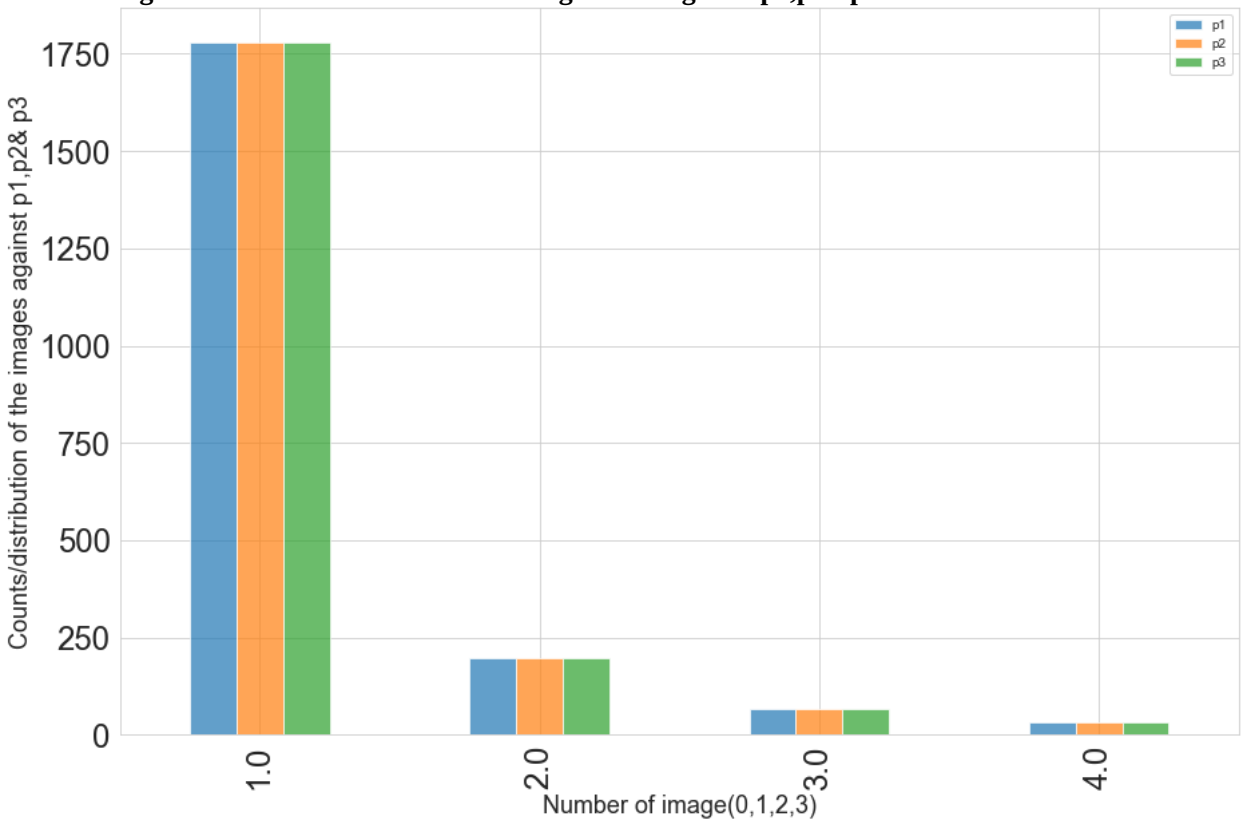
This report contains the findings following the data wrangling and analysis processes on the tweet archive of Twitter user [@dog_rates], also known as [WeRateDogs]. [WeRateDogs] is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings have a denominator of 10. The numerators always greater than 10. 11/10, 12/10, 13/10. The WeRateDogs has over 4 million followers and has received international media coverage.

A SAMPLE OF TWEETS

In the wrangle and analyze data project data was gathered from three datasets file formats, it was assessed cleaned and then the master dataset was analyzed.

VISUALIZATIONS
The following visualizations were created using python seaborns and matplotlyb libraries

Plot showing the counts/distribution of the image labels against p1,p2&p3

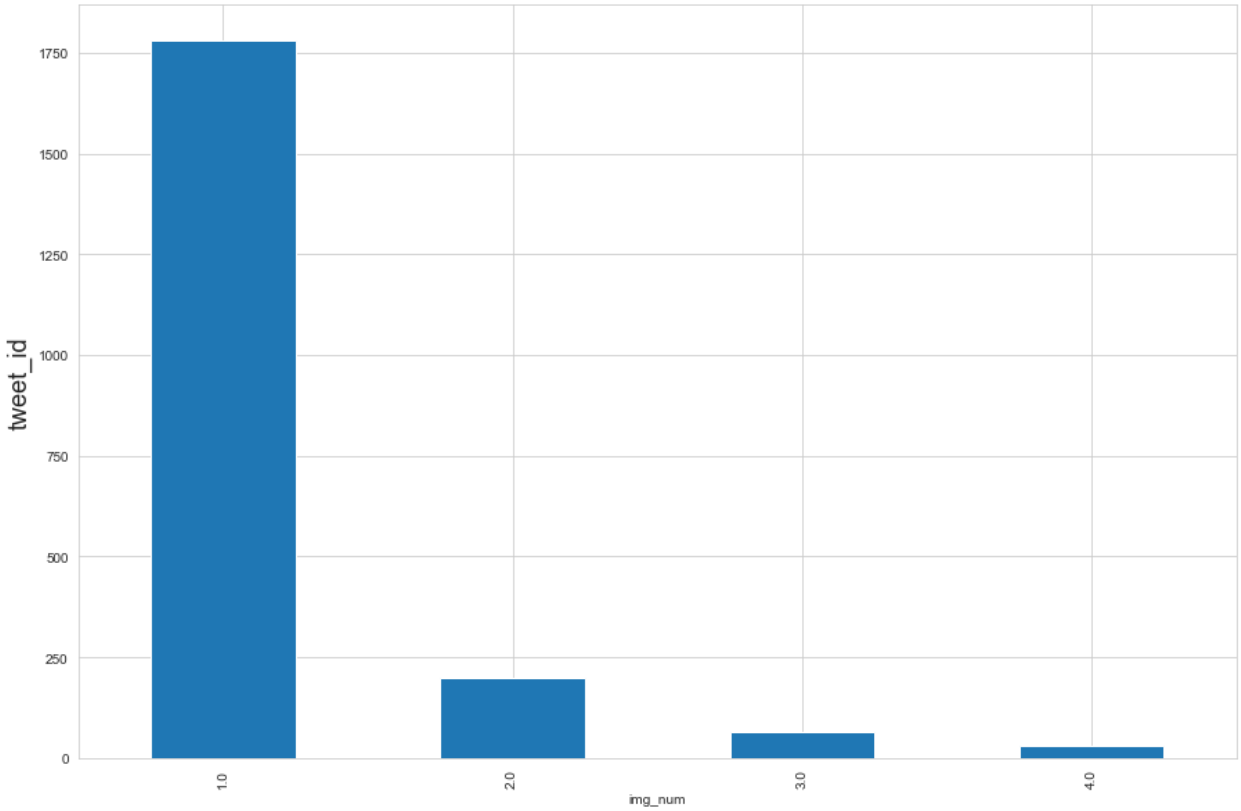


FINDING NO:1

labels 1 had the highest number of counts compared to other labels when plotted against 'p1','p2' and 'p3' as shown above.This could have been due to the strategy used for the sampling of the images used in the neural network. In all instances the proportion of images was the same.

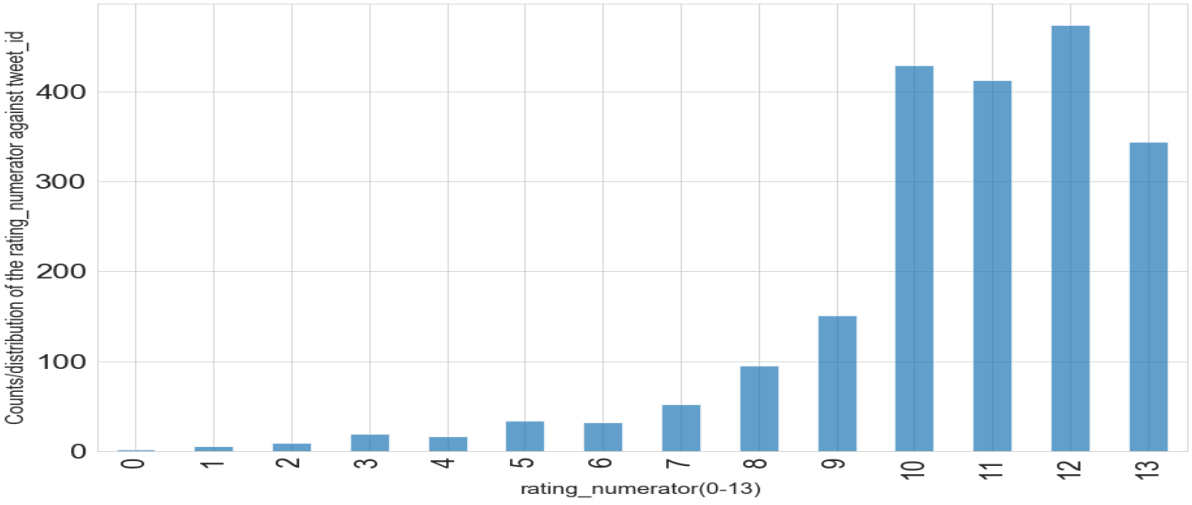
Similar findings are shown below when comparing the number of tweets id against the labels where label one has the highest proportion.

Plot showing the image number/label that had the highest number of tweets as per the tweet_id



From image classifications labels 1 had the highest number of counts compared to other labels when plotted against the tweet_id

Plot showing the counts/distribution of the rating numerator against tweet_id



FINDING NO:2
The rating_numerator contains multiple labels ranginging from 0-13, despite the specifications of the API ratings being greater than 10, several tweet_ids had rating below 10 which were detected as outliers.Majority of the ratings were between 10-13 with mean being 13.

FINDING NO.3
The tweets were gathered across three years spanning from November 2015 to August 2017.The number of tweets were highest in every months of November 2015,December 2015 and January 2016 with December 2015 recording the highest of over 350 tweets. This could have been caused by the excitement created by the new features of the API.

Wow!That's the END of my findings