Act Report

Author: Nguyen Gia Bao Le

October 17, 2022

This is the analysis report for Udacity Course - Wrangle and Analyze Data Project

Introduction:

The project is about wrangling and analyzing Dog Rate dataset. The dataset contains the tweets archive of Twitter user @dog_rates, also known as WeRateDogs. Detail about the dataset, you can see in "wrangle_report.pdf" file, I describe all in it.
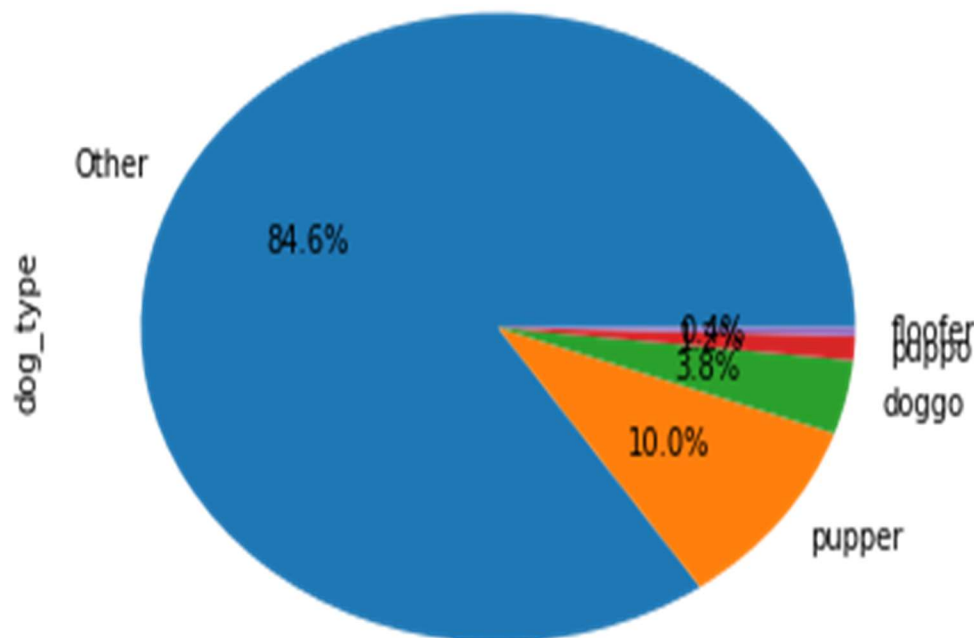
After gathering, accessing, cleaning, the final dataset is stored in "twitter_archive_master.csv" file.

Report:

1. The distribution of the breed of dog:

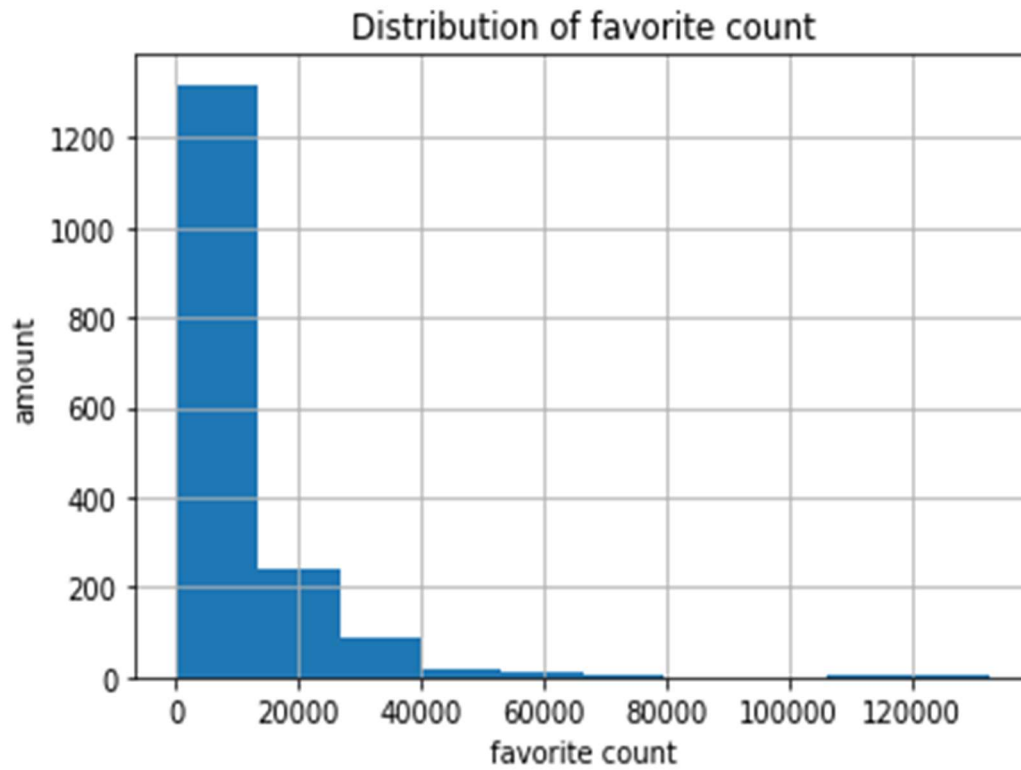| Breed of Dog	Amount	Ratio (percentage)
Other	1426	84.5788849
Pupper	168	9.96441281
Doggo	64	3.79596679
Puppo	21	1.2455516
Floofer	7	0.41518387

Distribution of breeds of dog

2. The distribution of favorite count in the tweets

   Max favorite count: 132810

   Max of frequent favorite count: 0-1000.

   Max of frequent 1200 favorite

Distribution of favorite count

3. The relationship between favorite_count and rating ratio of a tweet

The chart shows that, the more favorite count, the more rating ratio. With tweets have favorite count less than 1000, the average rating ratio is 0.95. In another hand, the tweets have favorite count greater than 30000, the averate rating ratio is approximately 1.25.
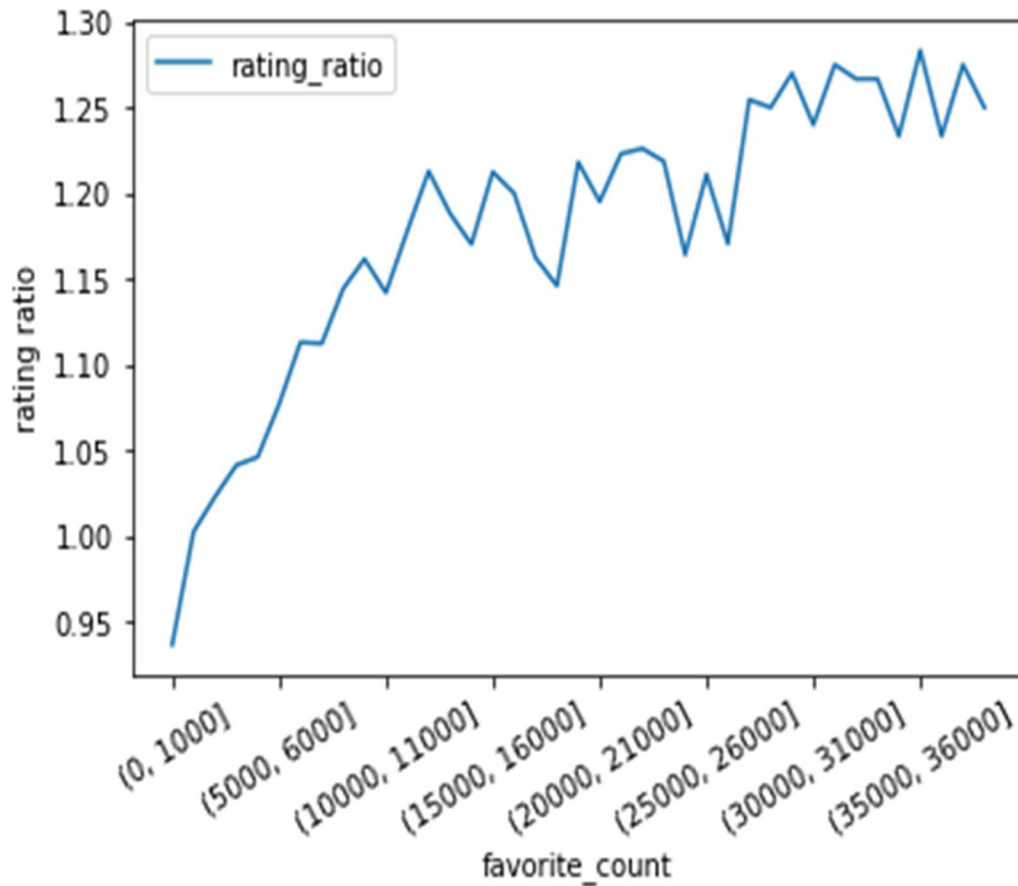
4. The relationship between tweet post time and rating ratio

The statistics shows that the time you should post the tweet to has more rating ratio is 5am to 6am.

timestamp	rating_ratio
(5, 6]	1.2
(13, 14]	1.15
(15, 16]	1.140802
(16, 17]	1.136111
(21, 22]	1.129091
(22, 23]	1.108371
(18, 19]	1.082353
(0, 1]	1.077119
(17, 18]	1.073171
(20, 21]	1.069932
(14, 15]	1.068831
(19, 20]	1.05
(1, 2]	1.034024
(3, 4]	1.022368

(2, 3]	1.010791
(12, 13]	1
(4, 5]	0.994444

relationship between tweet post time and rating ratio