act_report

December 7, 2018

1 Analyzing and Visualizing

In this report consists of visual display based on the dataset that I cleanned in following project called Data Wrangling - WeRateDogs.

Here we want to observe the rating, favorite_count and retweet_count variables.

1.0.1 Dataset Overview

- The dataset consists of 1971 rows and 22 columns.
- Highest rating is 1776.
- Highest favorite_count is 132810.
- Highest retweet_count is 79515.

1. Rating histogram plot:

From the plot we can see most of the dogs are rated at 12 and the distribution are left skewed.

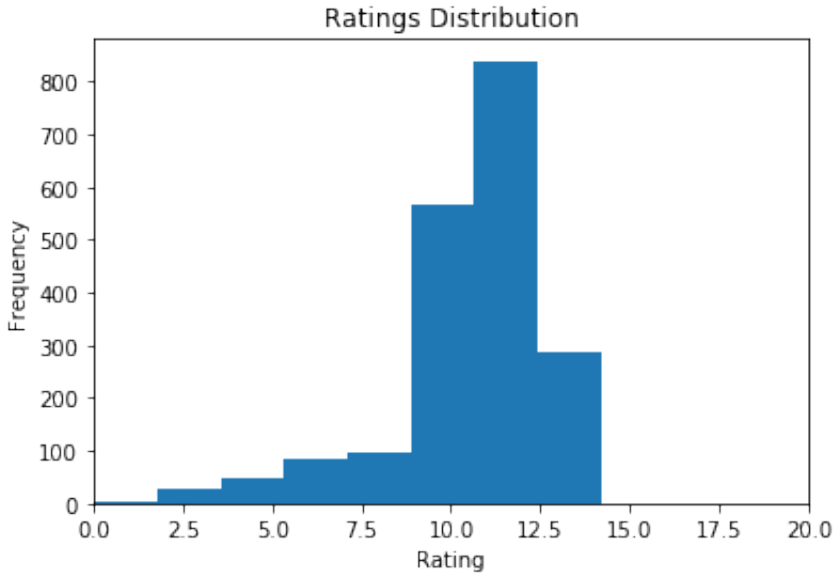
2. Rating favorite_count plot:

3. Rating retweet_count plot:

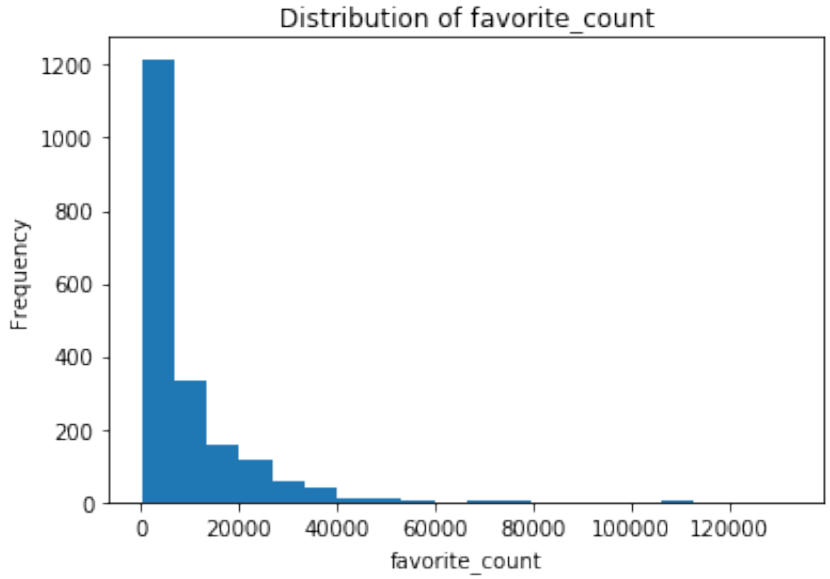
Both distribution of favorite_count and retweet_count are similar. Let plot scatter between this two variables to see the relationship.

4. favorite_count vs. retweet_count scatter plot:

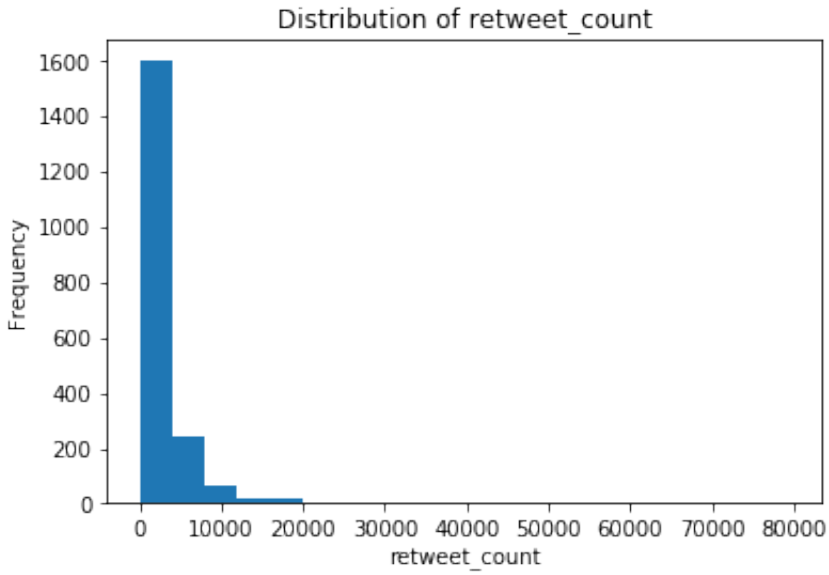
It shows strong positive correlation between favorite count and retweet count with the correlation 0.91


Ratings Distribution
Frequency
Rating
0
100
200
300
400
500
600
700
800
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

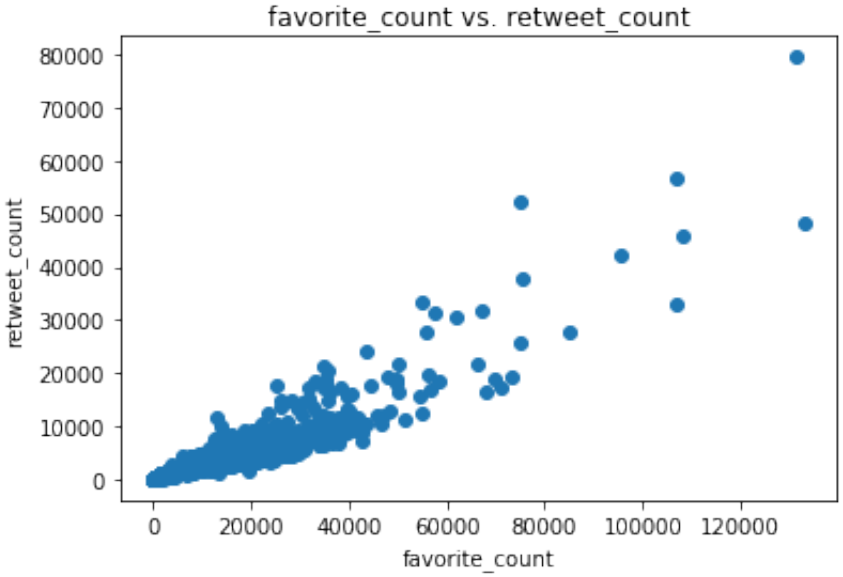
title


Distribution of favorite_count
Frequency
favorite_count
0
200
400
600
800
1000
1200
0
20000
40000
60000
80000
100000
120000

title


Distribution of retweet_count
Frequency
retweet_count
0
200
400
600
800
1000
1200
1400
1600
0
10000
20000
30000
40000
50000
60000
70000
80000

title


favorite_count vs. retweet_count
retweet_count
favorite_count
0
10000
20000
30000
40000
50000
60000
70000
80000
0
20000
40000
60000
80000
100000
120000

title