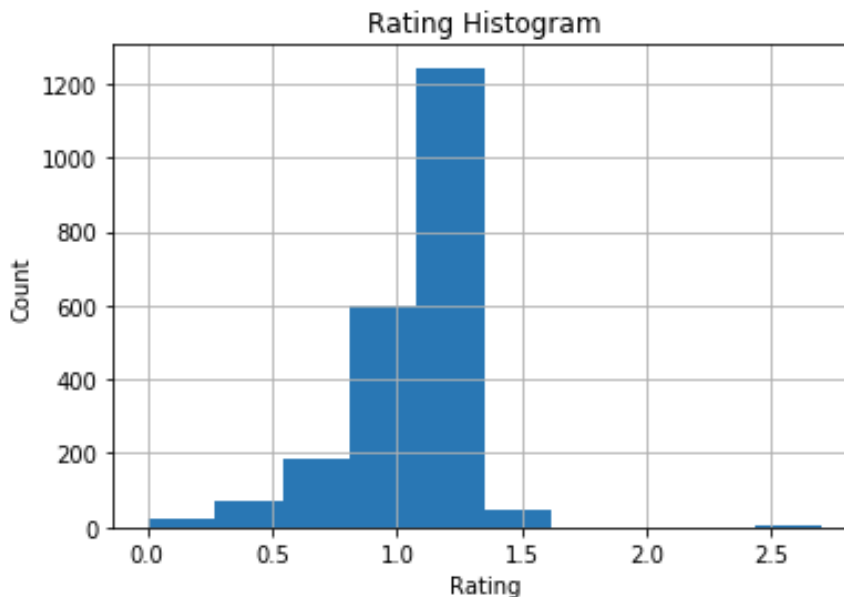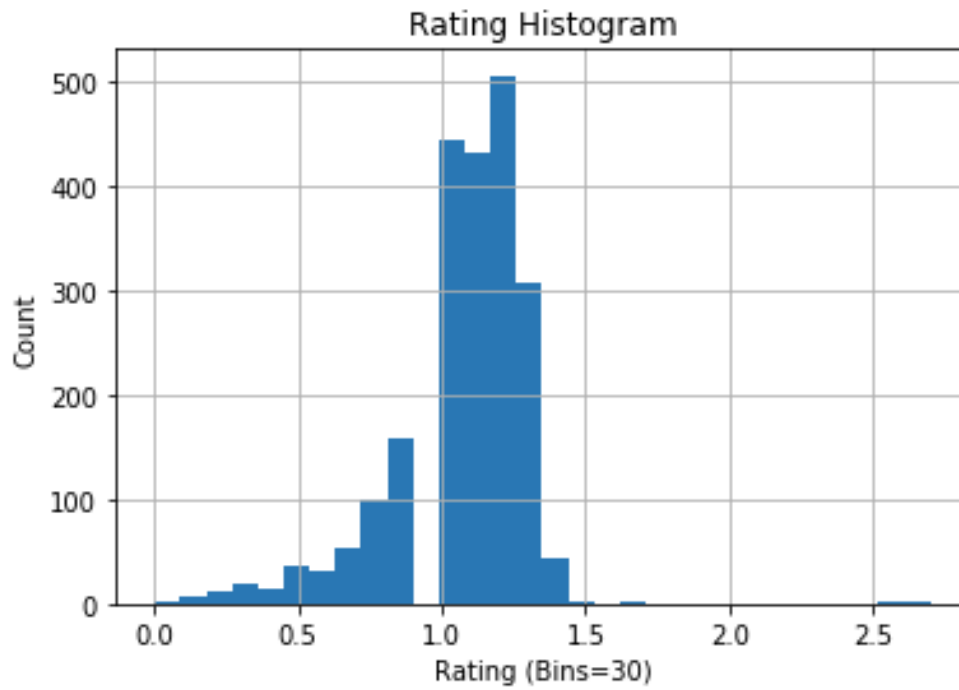Analysis and Visualization

- I first got the statistics for the numerical variables using the describe function.
- Initially, my merged data had all the ratings. But on using the describe function, I noticed the maximum as inifinite values in ratings. So, I then made a decision to remove the records with ratings that were extremely high. I ket the threshold as 3.0.
- Running the statistics on the new dataset gave me more reasonable statistics. However, this is also a limitation since the outliers are effectively removed. So, any significance of the outliers is not showcased in this analysis.
- Next, I made a visualization of the ratings as a histogram. On adding number of bins as 30, I could actually get the range of popular rating scores: values are 12-13.
- I also made a histogram of the favorites (count). Here too, increasing the number of bins gave me the popular range as 0-2000.
- My other visualizations were bar plots to see the stage and rating as adjacent bar plots. Due to the number of values, the visual is not very insightful. I also realized that my clubbed 'stage' needs additional wrangling to remove the appended names.
- Further, I plotted the p2 values as a bar plot and found that it was similarly not very insightful owing to the large number of values.
- Next, I computed the correlations for the numerical values.
- I found that based on the correlations, user_followers_cnt and user_favourites_cnt showed a correlation. It is expected as number of favorites increases with number of followers.
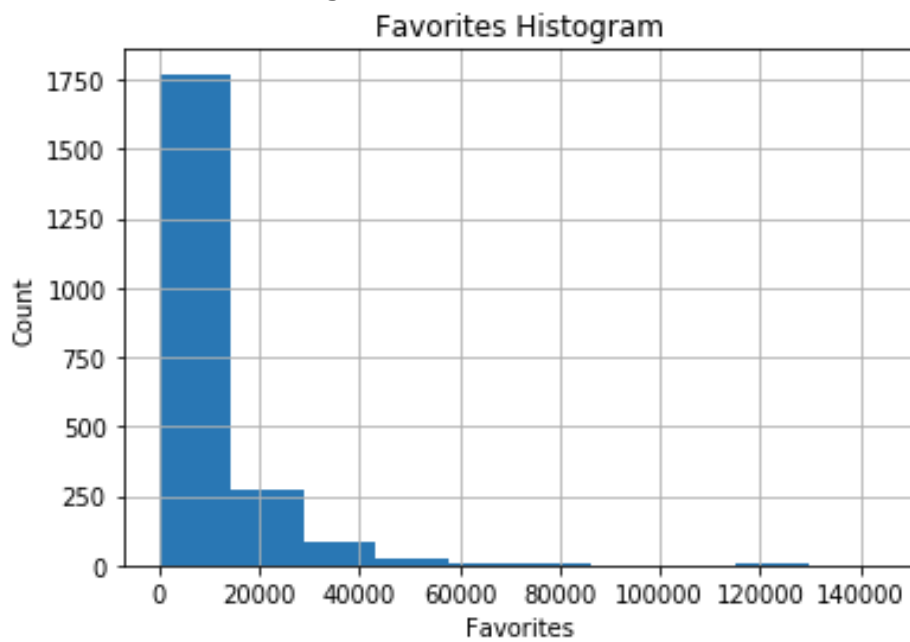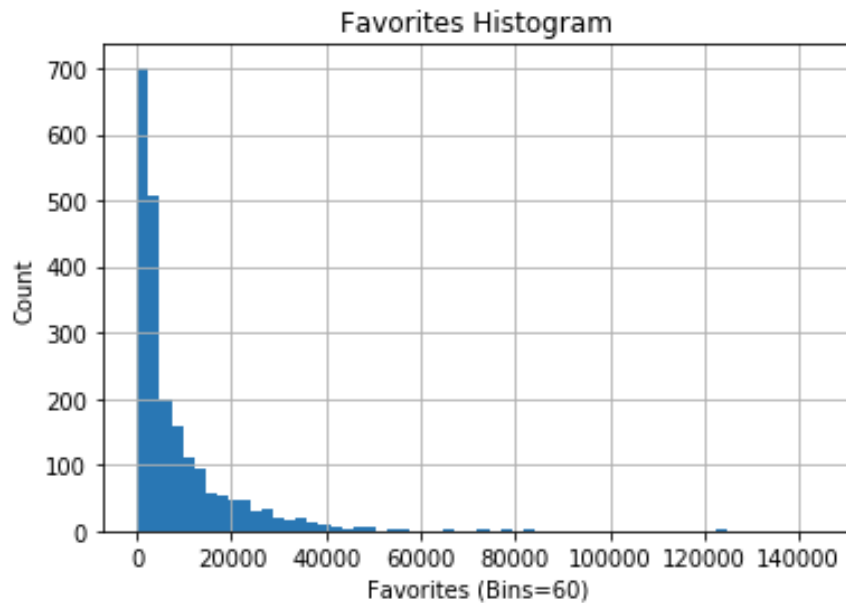
Visualizations are as follows:
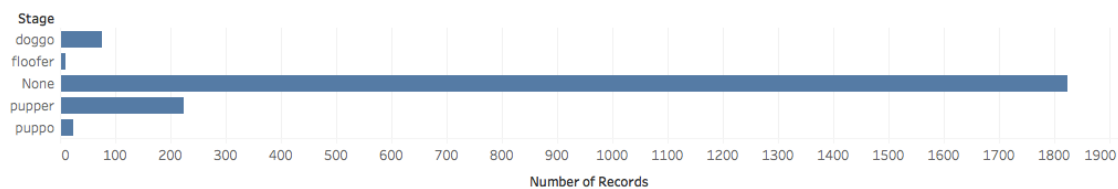Ratings as a Histogram is:

With bins=30, we get:

Rating Histogram

Favorites variable as a Histogram is as follows:

Favorites Histogram

With bins=60 is:

Favorites Histogram

With Python, my visualization for count of records was not very insightful. But with Tableau, the plots for count of records in Stage is:

Final Insights
- Ratings range from 0 to 2.7 -barring the extreme value records that we ignored
- Mean is 1.06 which shows that most dogs are rated 10/10 on average.
- Popular rating is 12-13.
- The user_followers_cnt on average is 6.3314e6 with minimum being 6.3310e6 and maximum being 6.3317e6. This highlights that the user community is popular on the whole.
- Based on the correlations, user_followers_cnt and user_favourites_cnt showed a correlation. It is expected as number of favorites increases with number of followers.
- Popular range of number of favorites is 0-2000.