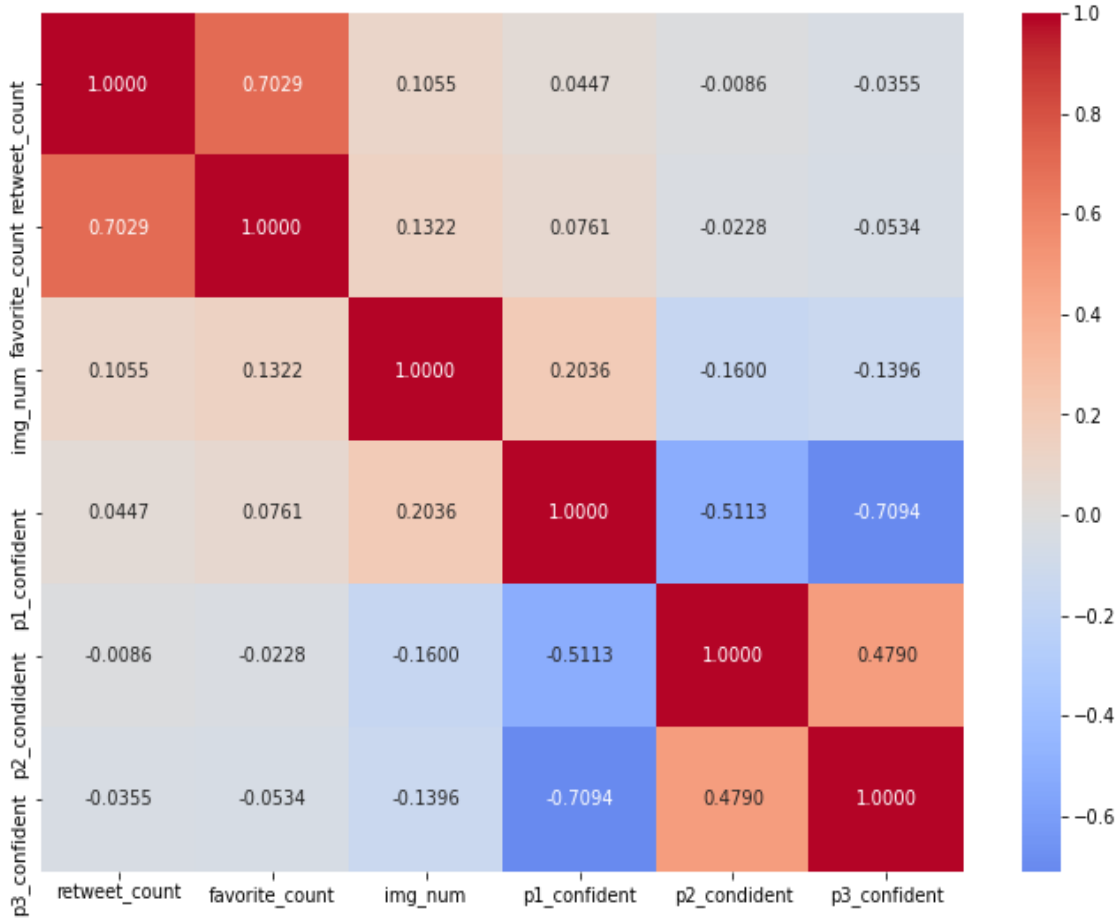
Conclusion

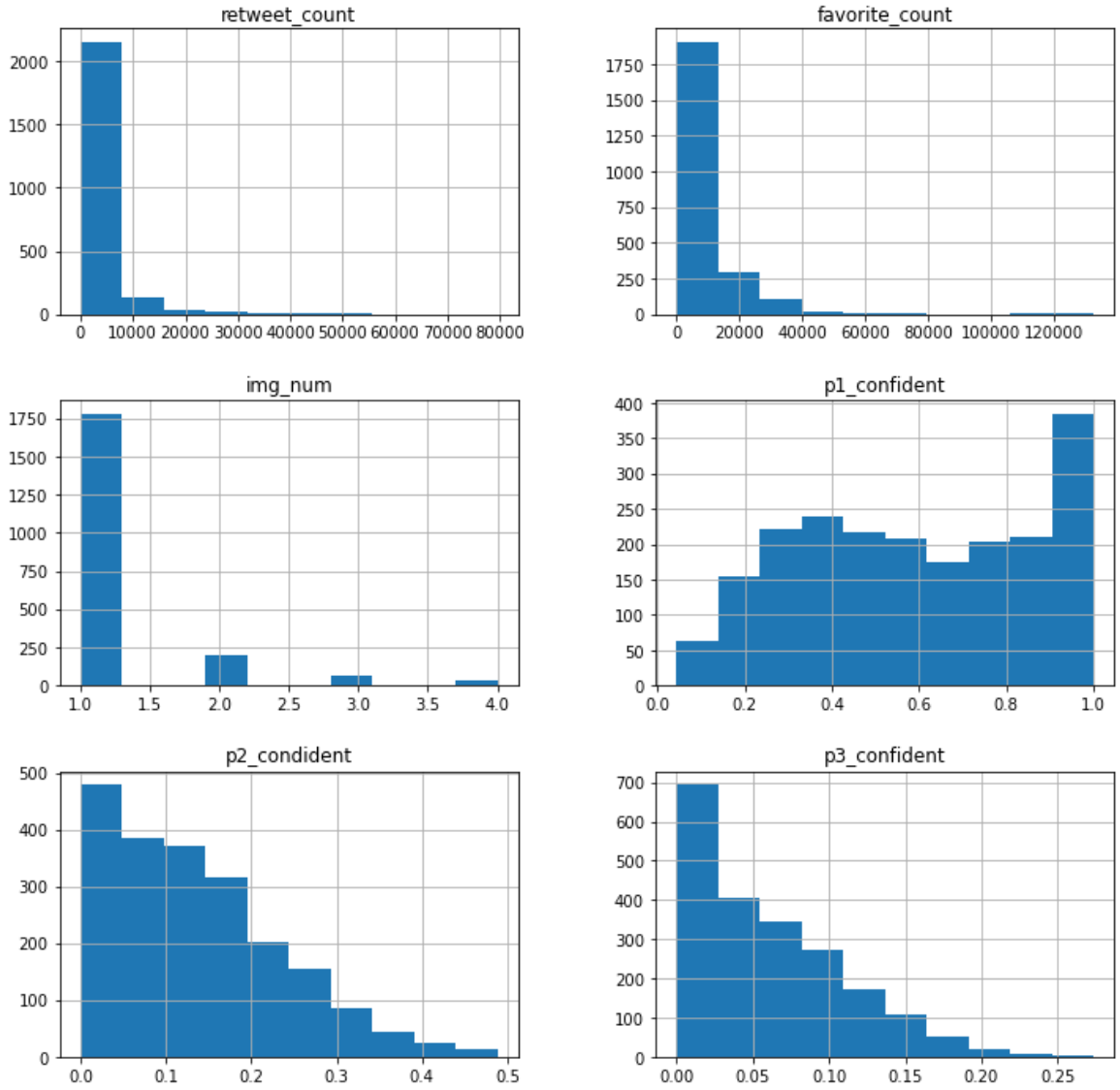
Showing the relationship between numeric columns in the heatmap, histogram and Scatter plot



Insights

- The only relationship we can mention here, is the strong positive correlation between the retweets count & favorite count. and that makes sense.

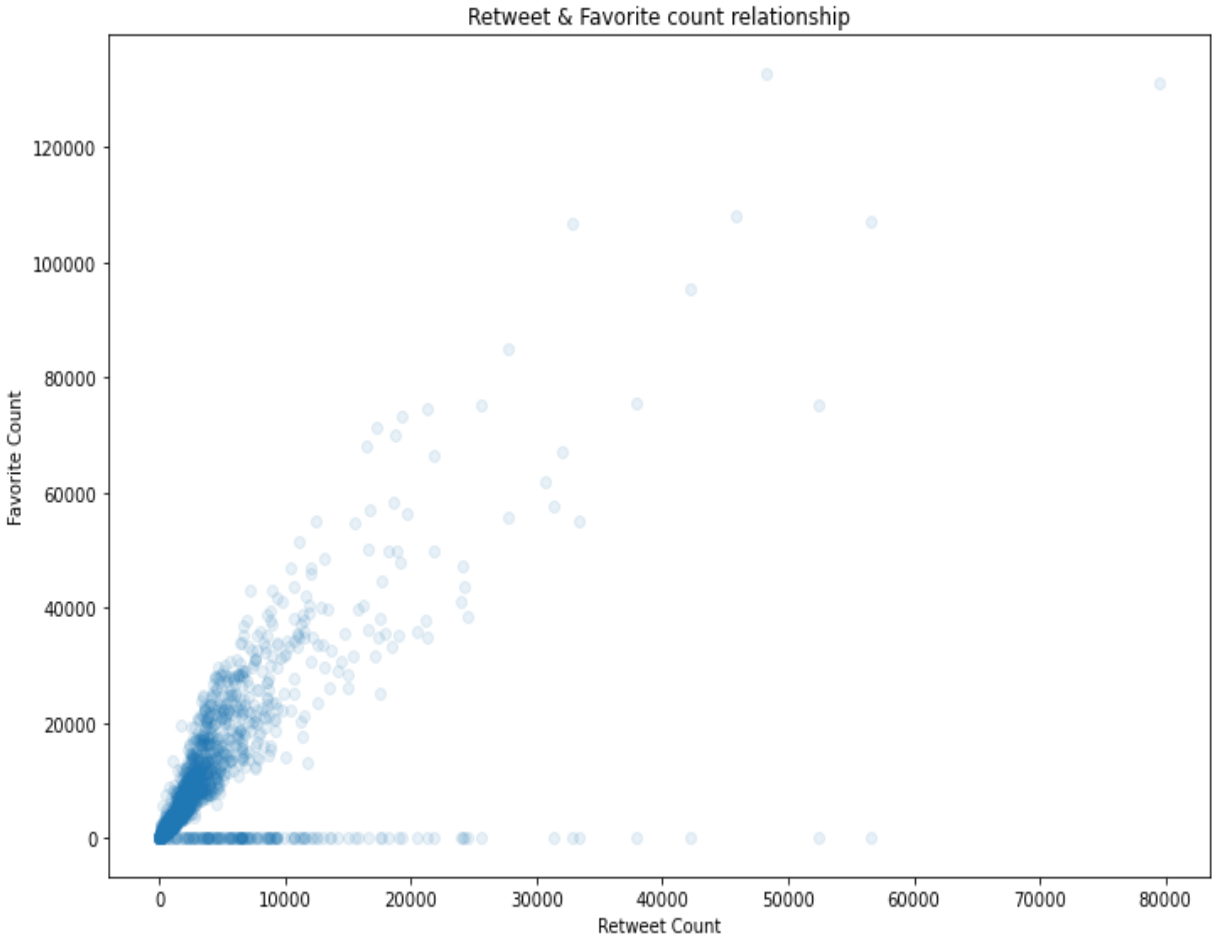
- And we can say that the more images the tweet has, the more likely to get retweets & favorites



Insights

- Only the p1_confident is left skewed, that means it is where the most values occurred at.

- The Rest of numeric columns are right skewed

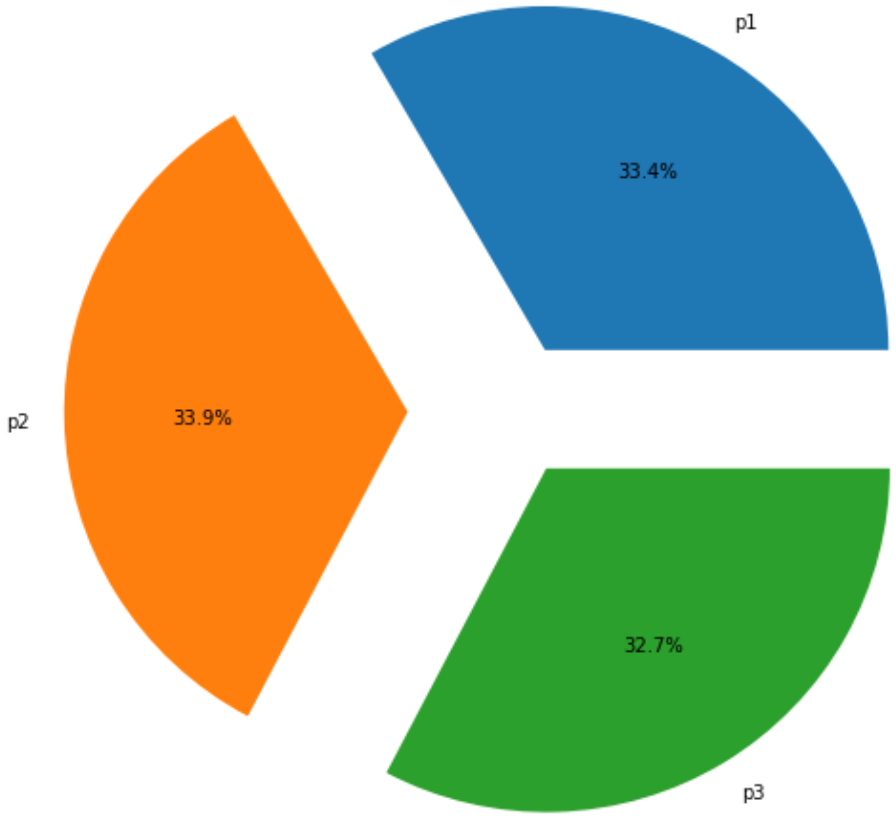


Insights

- Most of tweets had retweets less than 10,000 and favorites up to 32,000.

- We have a few tweets that got much more favorites & retweets.

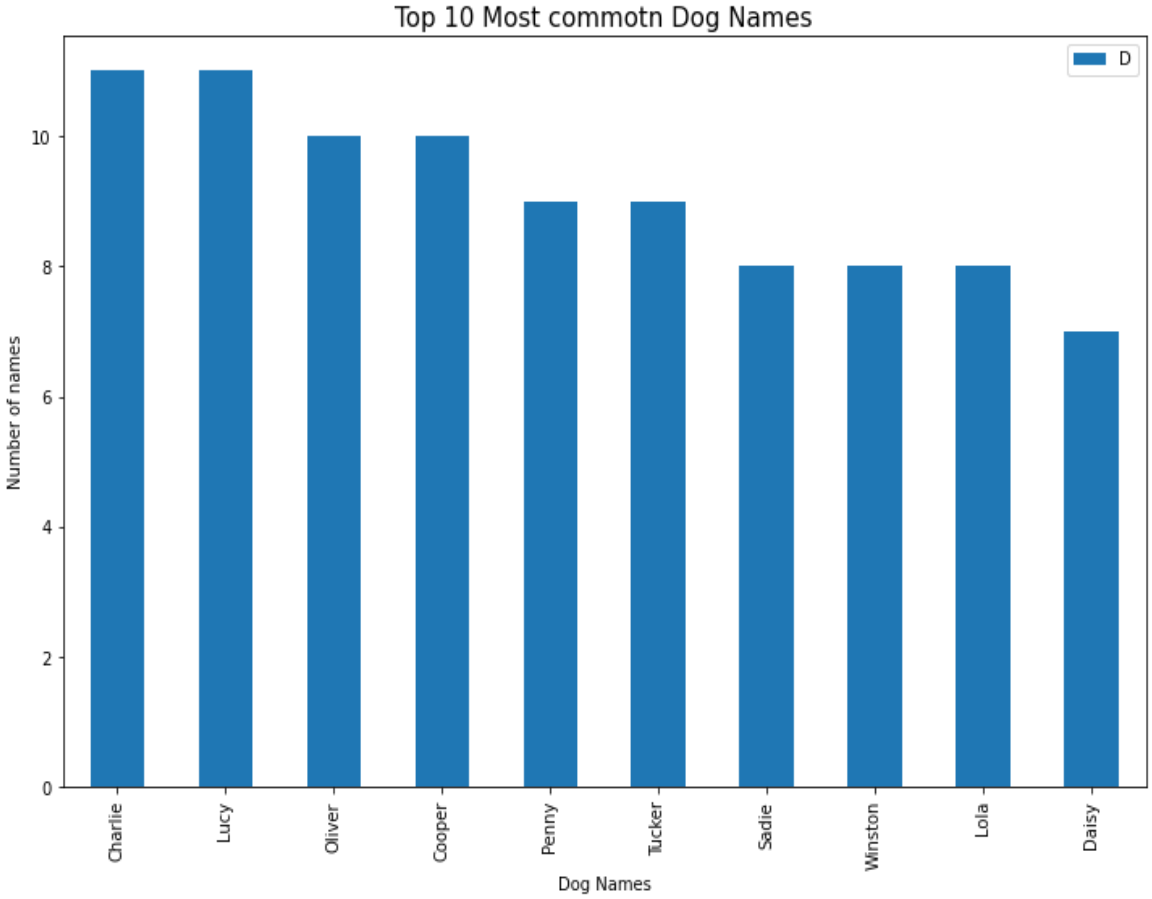
The portion of success algorithm in dogs



Insights

- p2_algorithm is the most successful algorithm
- p3_algorithm is the least successful algorithm
- all algorithms tend to roll around the similar portion (only a little difference).

Most Common dog names



Insights

- The top 10 most common names are Lucy, Charlie, Oliver, Cooper, Penny, Tucker, Sadie, Lola, Winston and Daisy

- Lucy and Charly have the same number of common names

- Lucy, charlie Oliver and Cooper are the top 10 names to be common in more than 9 times

Conclusion

Based on the whole insights we had here in the analysis and visualization step, we can finally say that the data we have seem to be accurate and effective for analysis because it shows a positive correlation between the retweent_count and retweet favorite. Also, most numeric values show a strong positive colour in the heatmap and the histograms of numeric values are nearly all right skewed except for the p1_cofident one. But that's okay because there are always some exceptions that can occur within the data.

The algorism in the dogs is almost equal in proportions, just a little bit difference in 1%, 0.3% 0. 7% around. But the most successful algorithm in the data is p2_algorithm.