THE ACT_REPORT.

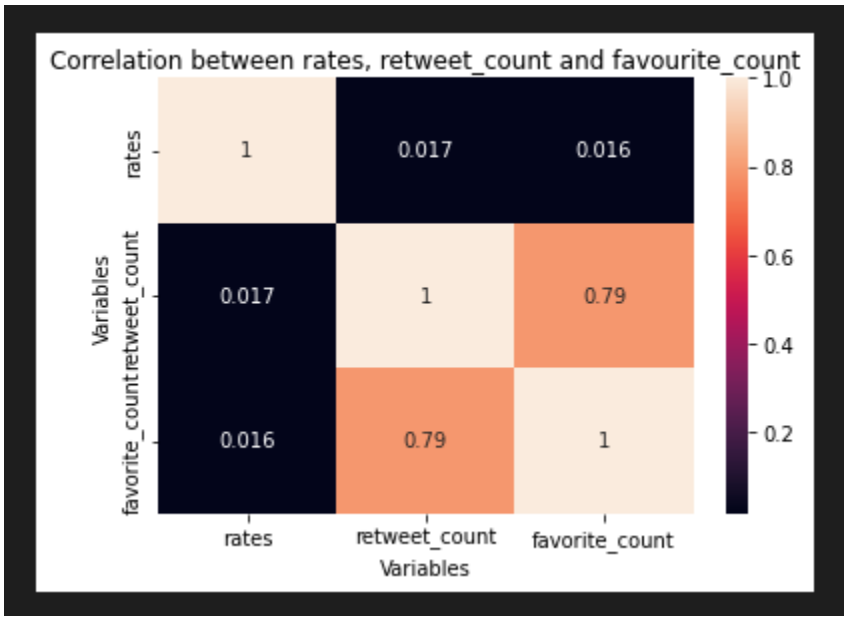
After gathering, assessing and cleaning all the three dataframes, they were merged ready to be analyzed and visualized.

Insights from df_4.

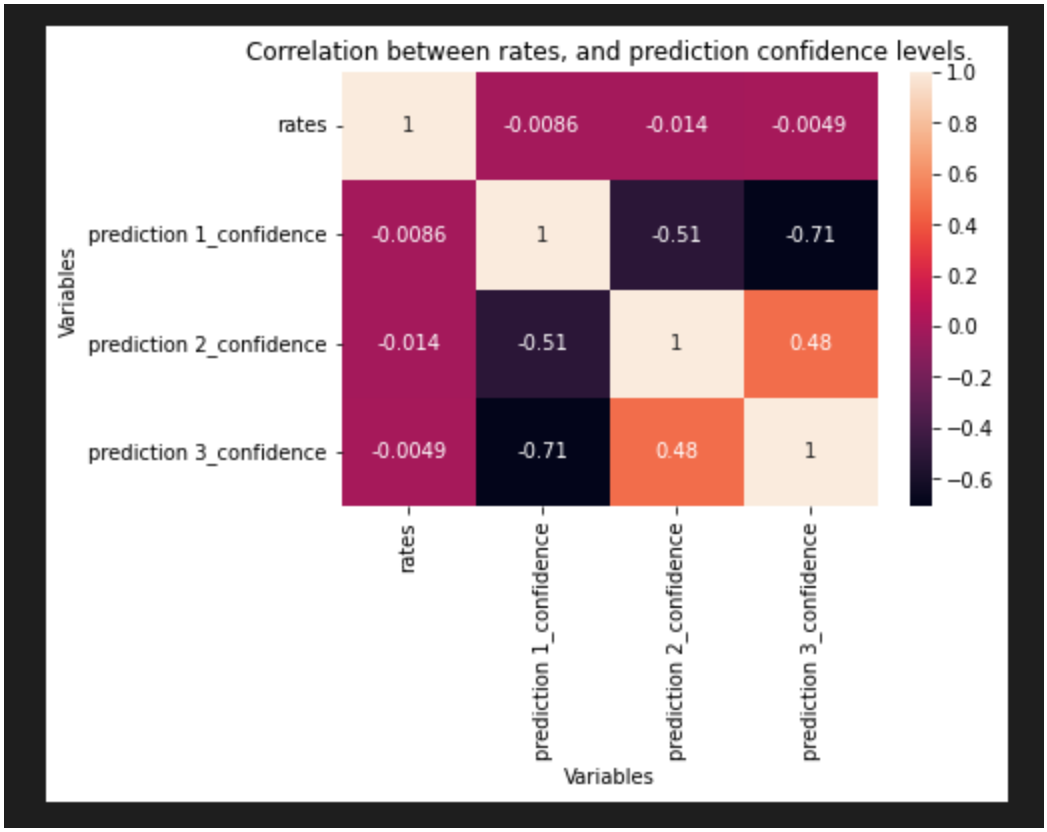
1. The mean of the ratings of the dogs is 1.17 and this gives us a clue as to where the dataset's center value is. Because it contains data from each observation in a dataframe.
2. 1.2 rate occurred the most with 1908 value counts, followed by 1.0, 1.1 and 1.3. The rates that had the least were 42, 2.6, 0.6, 0.8, 177.6 and 1.5 with a 4.
3. The sources for the tweets were, Twitter for iPhone (8128), Twitter Web Client (120) and TweetDeck (44). This indicates that most users had iPhones.
4. Most dogs were rated 12 out of 10 because the rating_numerator 12 had the most value counts which was 1892, followed by 10 with 1716, the 11 with 1652, 13 with 1132. 420 was the highest numerator value that was given and it had a value count of 4.

Insights from the plotted visuals from the wrangled data.

1. When investigating the correlationship between the rate, retweet_count and favorite_count variables, there was a very weak correlationship between rate and the other two variables. There was a positive strong relationship between retweet_count and favorite_count.



2. In the second visual, there's a negative correlationship between the variables (rate, prediction 2_confidence, prediction 3_confidence) because the values are negative. The only positive correlationship that exists is the one between prediction 2_confiedence and prediction 3_confidence by 0.48.


Correlation between rates, and prediction confidence levels.
rates
prediction 1_confidence
prediction 2_confidence
prediction 3_confidence
Variables
1
-0.0086
-0.014
-0.0049
-0.0086
1
-0.51
-0.71
-0.014
-0.51
1
0.48
-0.0049
-0.71
0.48
1
rates
prediction 1_confidence
prediction 2_confidence
prediction 3_confidence
Variables
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6