Act Report

Insights and display of visualization

Having figured out the problem with the datasets through both visual and programmatic assessments we launched into the cleaning stage.
The cleaning stage is divided into three namely:

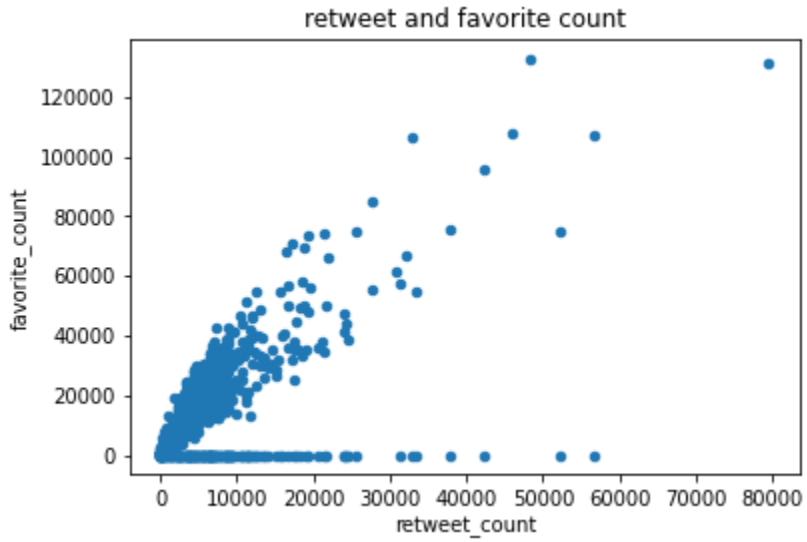
- Gathering
- Assessing
- Cleaning

The cleaning stages can also be segmented into three namely:

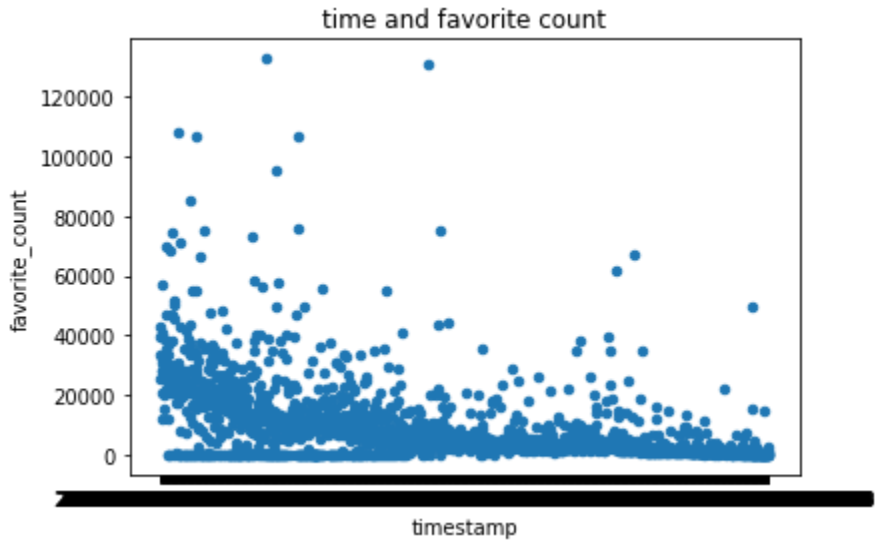
- Issue
- Define
- Code

After the cleaning stage, we launch into the analysis and visualization stage. Even though the three stages are a cycle that can always be repeated as we see fit. Visualization involves the charting of certain data sets or the display of the relationship between two or more datasets.

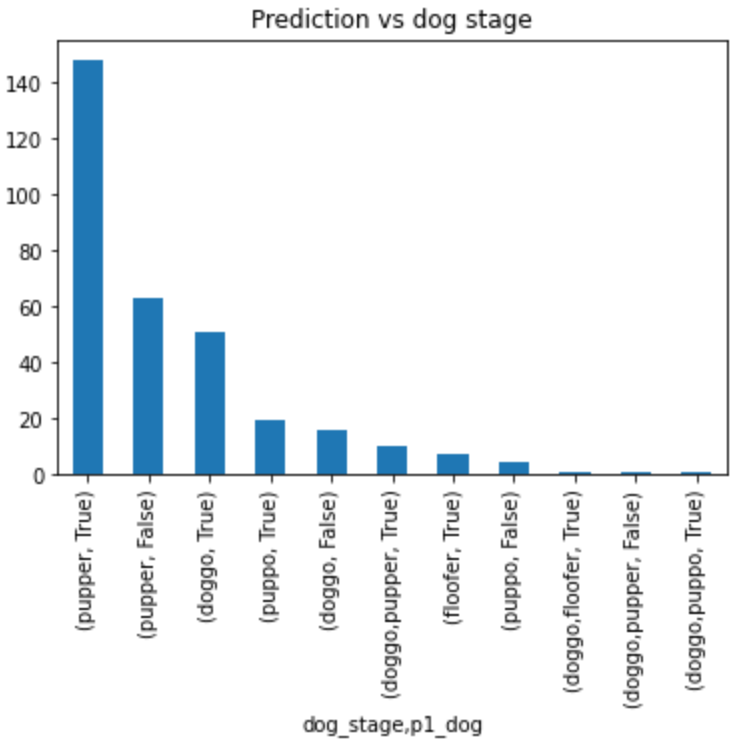
First, we examined the relationship between retweet count and favorite count as shown below:
From the visual below, we can see a perfect positive correlation between the two datasets.

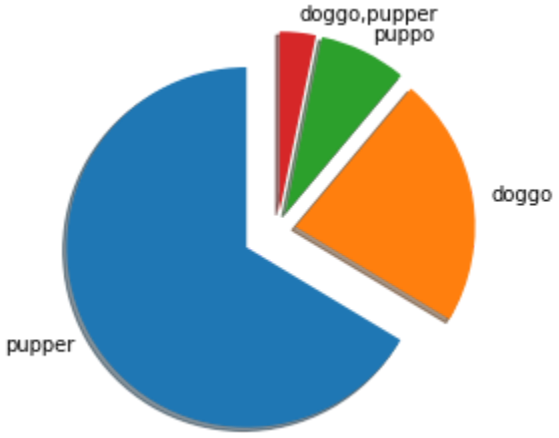


From the visual below, we can see that there is no significant relationship between the time of the day and people's interaction with the post. This implies that retweet and favorite count are not significantly influenced by the time of post.

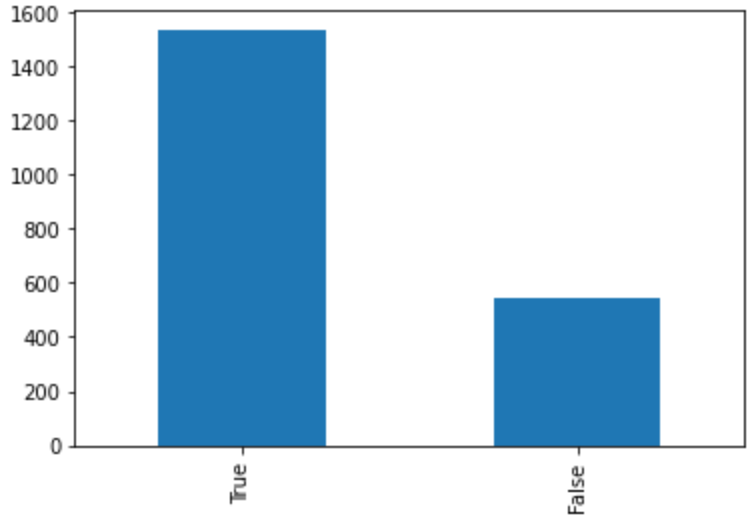


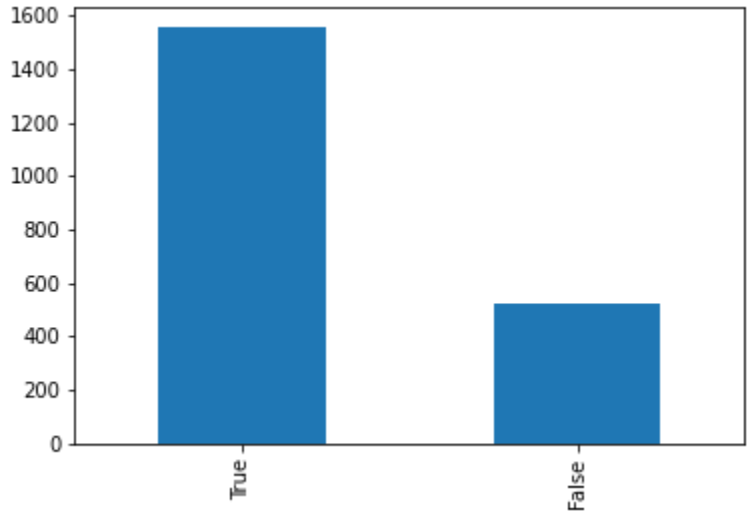
The distribution of the dog stage was also examined as shown below:
From the visuals below, we can see that the majority of the dogs rated are in the 'pupper stage', followed weakly by the 'doggo' and 'puppo'. Also, worthy of note is the fact that most of the "p1" predictions for the stage "pupper" came out to be true.

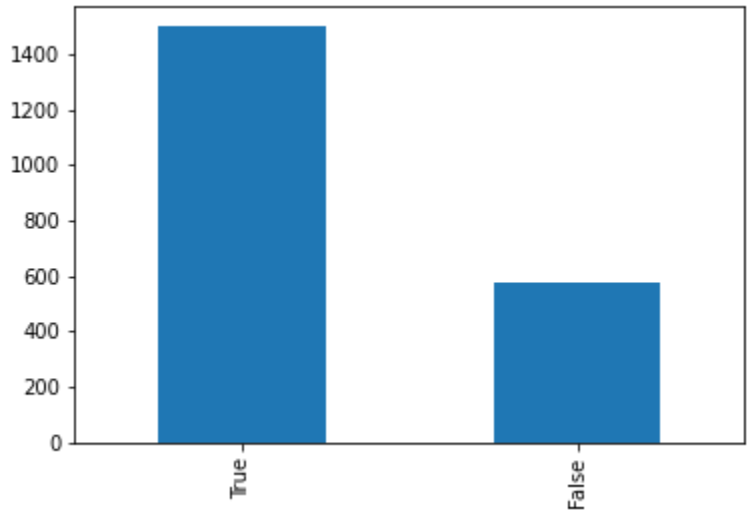




The bar charts below show each prediction (p1, p2, and p3) and they all show a visually consistent distribution of Trues and Falses, meaning that each prediction on its own merit has significant prediction power.


1600
1400
1200
1000
800
600
400
200
0
True
False


1600
1400
1200
1000
800
600
400
200
0
True
False


1400
1200
1000
800
600
400
200
0
True
False

Insights

1. The majority of the dogs rated are in the pupper stage.
2. Retweet and favorite count are not significantly influenced by the time of post.
3. Most of the “p1” predictions for the “pupper stage” are true.
4. Other predictions have similar predictions.

Limitations

1. Data Accuracy: The dataset used was collated over a long period of time.
2. Data Representativeness: Missing data dropped could lead to data representative issues.