act_report

December 4, 2023

```
In [1]: from IPython import display
```

0.1 Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

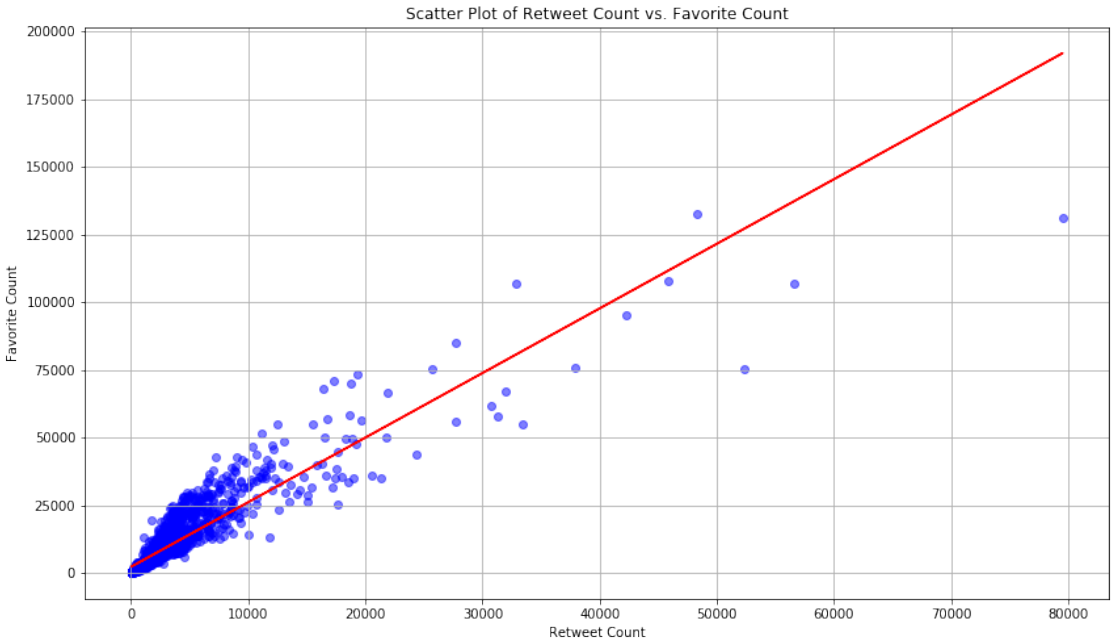
0.1.1 Insight 1: Rates and Confidence Intervals

- The rating distribution shows an average of 1.08, with the range from 0 to 42.
- The majority align with a denominator of 10, emphasizing the centrality of this scale.
- Confidence in predictions follows a hierarchical pattern, with the first prediction boasting the highest confidence (average of 59.39%), followed by the second (average of 13.44%), and the third prediction exhibiting the least confidence (average of 6.025%).

0.1.2 Insight 2: The correlation between retweet count and favorite count

```
In [2]: display.Image("correlation_retweet_favorite.png")
```

```
Out[2]:
```

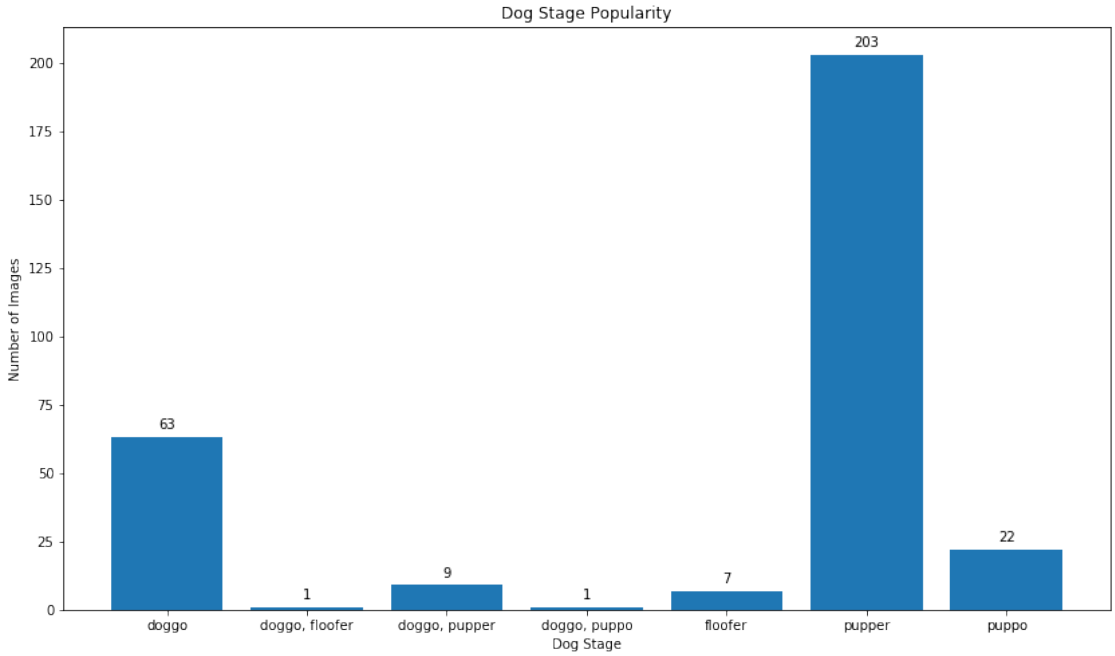


- The correlation analysis between retweet count and favorite count shows a correlation coefficient of 0.91.
- This strong positive correlation suggests that as a tweet receives more likes, it is likely to be retweeted more often, indicating a strong pattern of engagement in the Twitter.

0.1.3 Insight 3: What is the most popular dog stage?

```
In [3]: display.Image("distribution_of_dog_stages.png")
```

```
Out[3]:
```



- "Pupper" takes the crown as the most frequently shared dog stage, with a count of 203. In contrast, "doggo, floofer" and "doggo, puppo" appear less frequently, each with its own count is 1.
- This bar plot visually showcases the preferred dog stages among Twitter users.

0.1.4 Analysis Decision 1: Rate's Outlier Removal

- Recognizing the impact of extreme outliers, I decided to remove outliers in the rating variable. This decision was motivated by the persistence of an extreme value even after the removal of the iconic 1776 rating. The subsequent analysis revealed a notable improvement in the correlation between rate and retweet count, reinforcing the importance of outlier management in data exploration.

- In conclusion, the project has uncovered fascinating insights into the dynamics of ratings, social media engagement, and user preferences. As we navigate this rich tapestry of information, these findings serve as a testament to the power of data analysis in unraveling the stories embedded in each Twitter pup's post.