WRANGLE AND ANALYZE DATA

BACKGROUND

The objective of this project is to apply the data wrangling process to WeRateDog's twitter archive. The process involves gathering, assessing, and cleaning the data in the accompanying Jupyter notebook.

To analyze the archive data, we merged three data frames created in Jupyter Notebook. The dataframes are the twitter archive, image predictions file, and json data. The archive is a collection of past tweets. The image predictions provides dog breed predictions. The json data allows us to extract more information such as favorites and retweets.

The following are insights of the final cleaned dataframe. Analyzing the data allows us to dig even deeper to reveal interesting or significant findings. These findings can lead to possible recommendations for areas of improvement.

INSIGHTS

MOST POPULAR WEEKDAY FOR TWEETS

```
Monday       339
Wednesday    313
Tuesday      310
Friday       294
Thursday     290
Sunday       261
Saturday     261
```

The most popular day for tweets within WeRateDogs' archive is Monday. Tweets gradually decline as the week progresses.

TOP TWEET BY FAVORITE COUNT

The top tweet by favorite count is tweet_id #822872901745569793. This was tweeted on a Sat. Although Monday is a popular day for tweets, it does not necessarily mean the most engagement.

HIGHEST RATED DOG

The highest rated dog is Atticus with a rating numerator of 1776.

CORRELATION BETWEEN FAVORITES AND RETWEETS

Favorite and retweet count have a positive relationship. As favorite count goes up, retweet count does the same. Most dogs have more favorites than retweets. This finding is not surprising due to the likelihood of people favoriting a tweet over retweeting and sharing it on their personal account.