

# Act Report

## Introduction

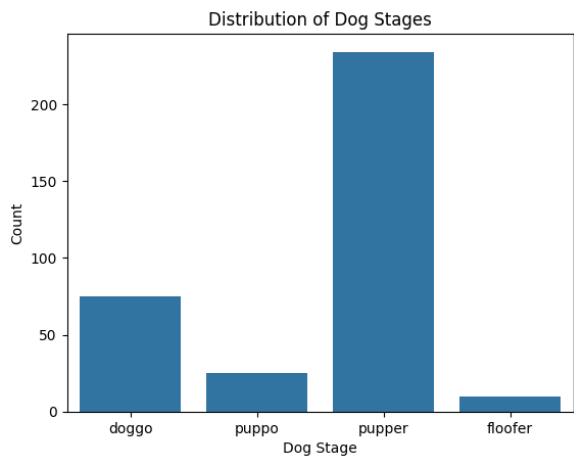
The WeRateDogs Twitter account is known for its humorous dog ratings. In this project, we wrangled and analyzed the WeRateDogs Twitter data to uncover interesting insights about the ratings and the dogs featured.

## Insights

- Distribution of Dog Stages:** We found that the most common dog stage is "pupper," followed by "doggo." This suggests that WeRateDogs tends to feature younger dogs more frequently.
- Ratings Analysis:** The ratings given by WeRateDogs are almost always greater than 10/10, with the most common rating being 12/10. This unique rating system contributes to the account's popularity.
- Engagement Analysis:** There is a positive correlation between retweet counts and favorite counts. Tweets with higher ratings tend to receive more engagement, indicating that followers appreciate the higher-rated dogs.

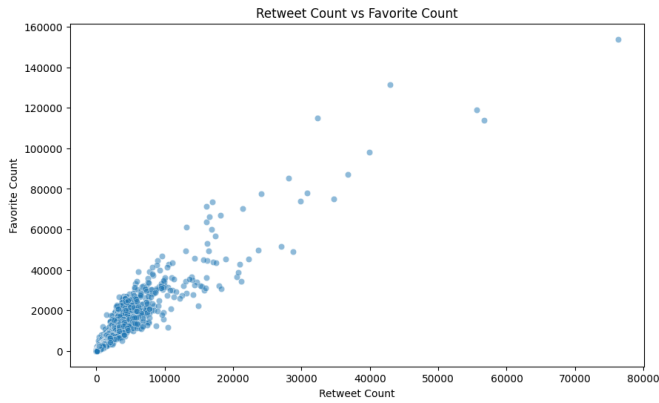
## Visualizations

### Distribution of Dog Stages



The bar plot above shows the count of each dog stage. "Pupper" is the most common stage, followed by "doggo."

### Engagement Analysis



The scatter plot above shows the relationship between retweet counts and favorite counts. There is a clear positive correlation, indicating that tweets with more retweets also tend to have more favorites.

# Wrangle Report

## Introduction

The goal of this project was to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The data sources included the WeRateDogs Twitter archive, image predictions from a neural network, and additional tlet data gathered via the Twitter API.

## Data Gathering

I gathered data from three sources:

1. **Twitter Archive:** The WeRateDogs Twitter archive was provided as a CSV file containing basic tlet information.
2. **Image Predictions:** I downloaded the image predictions file programmatically using the Requests library.
3. **Twitter API:** I used the Tlepy library to query the Twitter API and gather additional data such as retlet counts and favorite counts.

## Data Assessing

I assessed the data both visually and programmatically:

- **Visual Assessment:** I displayed the data in the Jupyter Notebook to identify obvious issues.
- **Programmatic Assessment:** I used pandas functions to detect data quality and tidiness issues. For example, I checked for missing values, duplicated rows, and incorrect data types.

I identified several quality and tidiness issues, including missing values, incorrect data types, and inconsistencies in dog names and stages.

## Data Cleaning

I addressed the identified issues through the following steps:

1. **Define:** For each issue, I defined the cleaning task.
2. **Code:** I wrote code to clean the data, such as filling missing values, correcting data types, and standardizing dog names and stages.
3. **Test:** I tested the cleaned data to ensure the issues were resolved.

After cleaning, I merged the datasets to create a tidy master DataFrame, which I stored in a CSV file for further analysis.