1. Introduction

After the Data Wrangling step we had the final Dataframe in file called “twitter_archive_master.csv” this dataframe shape is 2114 x 23 and it is ready to be analyzed.

2. Insights

2.1. Dog Stages

Dogs are divided into 4 stages which is:

1. Doggo
2. Floofer
3. Pupper
4. Puppo

The final dataframe contained 1976 “None” values that we will not try in this project but the rest of data shows the most common stages as following:

Dog Stage	Number of Occurrence
Pupper	245
Doggo	83
Puppo	29
Multiple	14
Floofer	9

There was tweets that classify the dog into two stages such as “doggo&pupper”, “doggo&puppo” so these stages was replaced with “Multiple” stage classification.

So the most common dog stage is Pupper

2.2. Average rating based on dog stage

A new column in this dataframe was calculated by dividing rating numerators over rating denominator to find the rating in %, also should notes the fact that the rating numerators are greater than the denominators does not need to be cleaned. This unique rating system is a big part of the popularity of WeRateDogs, and the relationship is as follows:

Dog Stage	Avg. Rating
Puppo	121.1%
Doggo	119.2%
Floofer	118.8%
Multiple	113.1%
Pupper	107.6%

So the Puppo dog stage is the most favorited dog in our data.

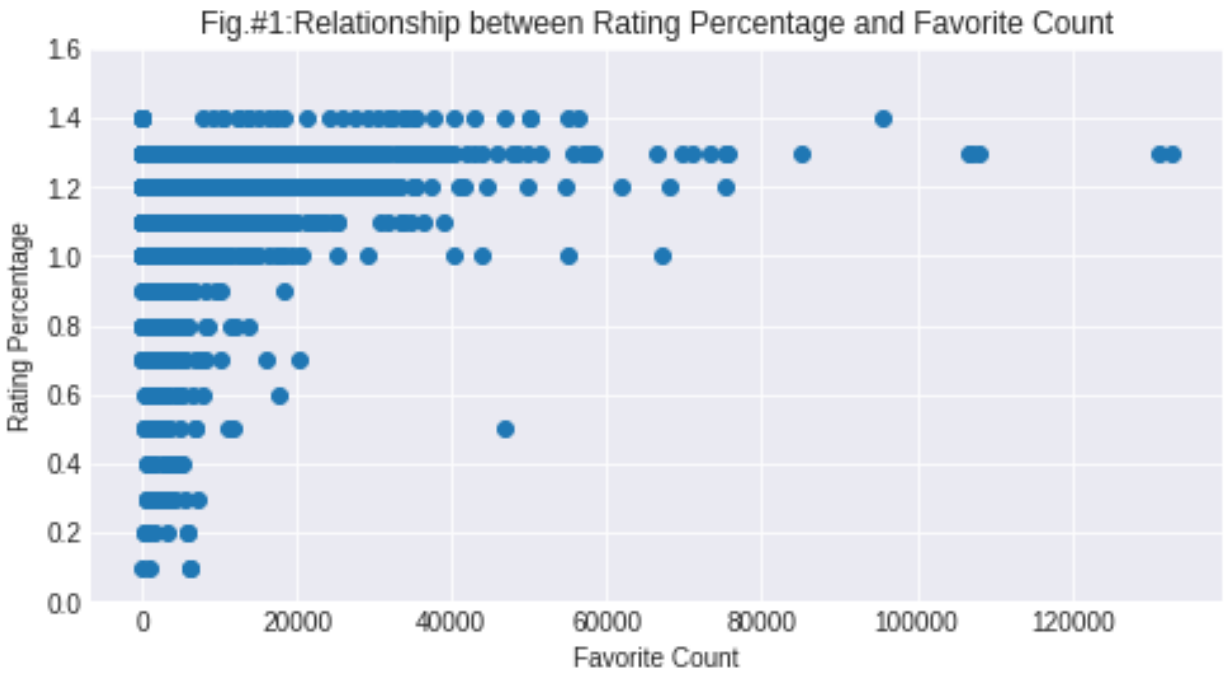
2.3. Most Common 10 Dog Names

The most common 10 do names are the following:

Dog Name	Number of Occurrence
Charlie	12
Cooper	11
Oliver	11
Tucker	10
Penny	10
Lucy	10
Lola	9
Bo	9
Winston	9
Sadie	8

3. Visualization

In this part I have done a visual that explains the relationship between rating percentage and favorite count.



This figure excluded the outliers of the rating percentage (every data above 140% was considered as outlier).

This figure shows that the higher the rating is for a dog in a tweet the more favorite this tweet will get.