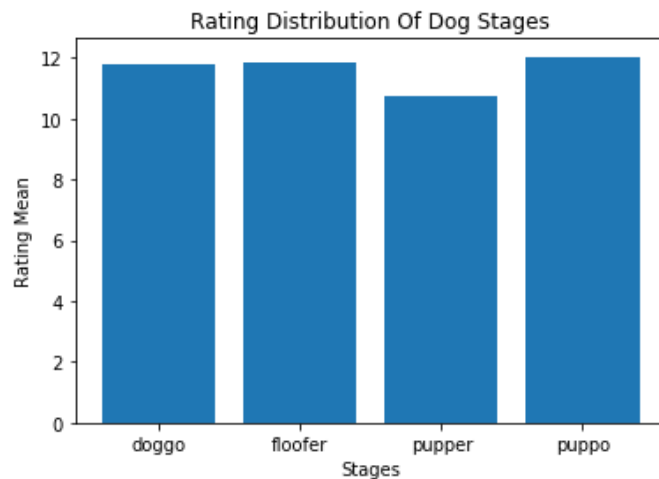


Data Insights and Visualization

The dataset created after data wrangling process of the 'WeRateDogs' Twitter dataset comprises of two dataframes: 'archived_clean' and 'images_clean'.

On analysis of this dataset, three key characteristics of the data is identified.

- 1) In the tweets, if the dog stage: 'doggo', 'floofer', 'pupper' or 'puppo' is named, then the rating is seen to be in a standard format i.e., the denominator rating will always be 10. Also, the numerator rating is seen to be closer to the denominator. The mean of the numerator rating is seen to be greater than ten, and close to each other values.



- 2) In this dataset, the numerator rating ranges from 0 to 1000+ and denominator rating ranges from 2 to 170. However, the overall rating_denominator mean is 10.5, which is very close to the WeRateDogs twitter group standard, 10. The rating_numerator mean is 11.9 which is greater than denominator mean, this follows the Unique rating system of WeRateDogs.
- 3) In the dataset, the algorithm is able to predict the dog in picture in many instances. However, when the dog is in the background or when is being partially seen or when placed in the side of the picture instead of being placed at the center of the picture, then the algorithm is not able to identify the dog and instead identifies the other objects in the picture. The prediction visualization is shown below.

