

Act Report

WeRateDogs Data

The **WeRateDogs** Twitter Archive Enhanced comprised of dog ratings and dog names that were extracted from the text of the tweet with dog names but not all text contains this information. Some data were missing from this file which were **retweet count** and **favorite count**. This were downloaded by querying twitter API using my access keys and token as a developer.

I also downloaded image predictions file programmatically using request module.

Wrangling Data

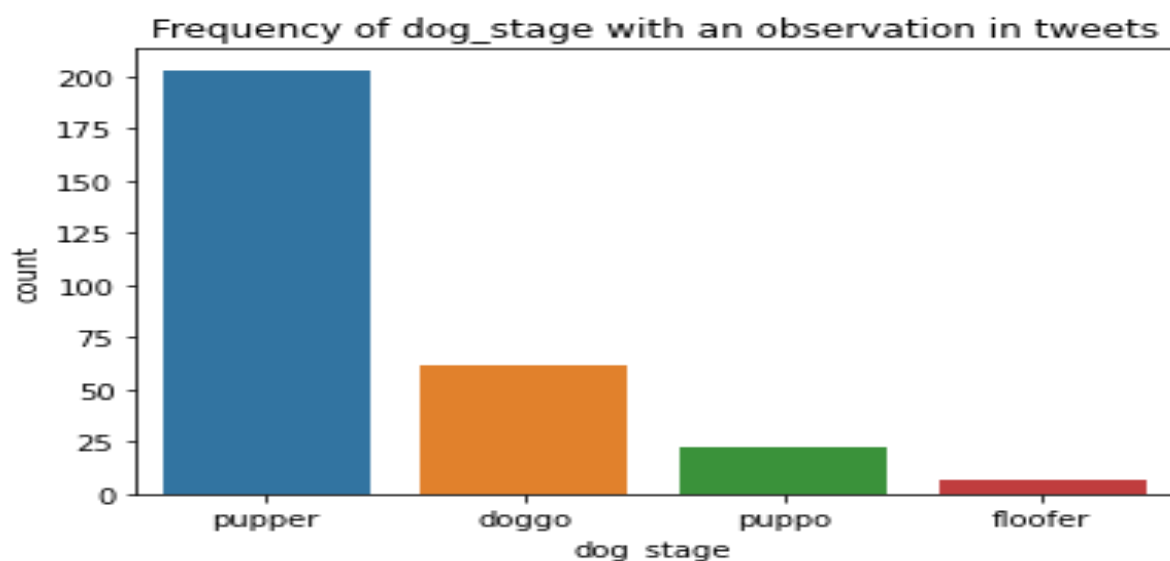
The data were loaded and opened in my jupyter notebook for both visual and programmatical assessment for quality and tidiness, hence reduced the data to **1986 rows** and **13 columns** for analysis and visualisation.

Insights

1. only 305 observations of dog stage
2. The maximum image number in a tweet is 4
3. the maximum first prediction confidence is 100
4. not all tweets have rating, only 1570 tweets have rating
5. maximum numerator rating is 14 and minimum is 0
6. the minimum, 25%, 50% and 75% of image number in a tweet is 1, this implies most of the tweets only have 1 image.

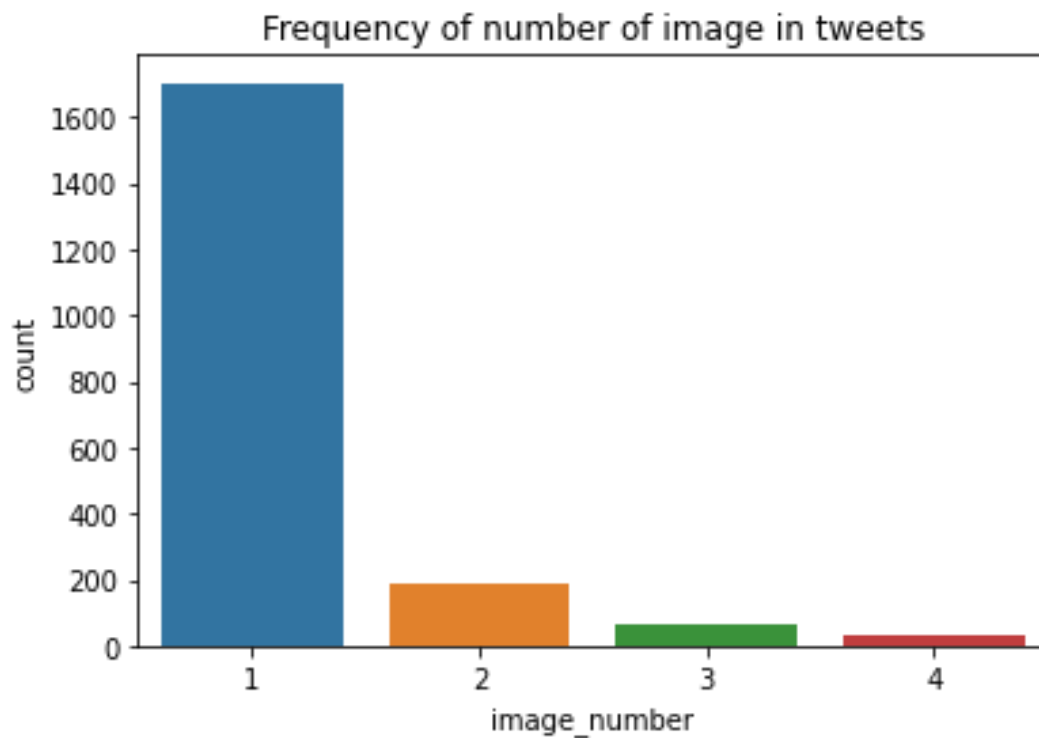
Question 1. What dog stage occur most in the tweets?

From the dataset after cleaning, only 305 observations of dog stage were in the tweets and the highest tweeted dog stage is **PUPPER**



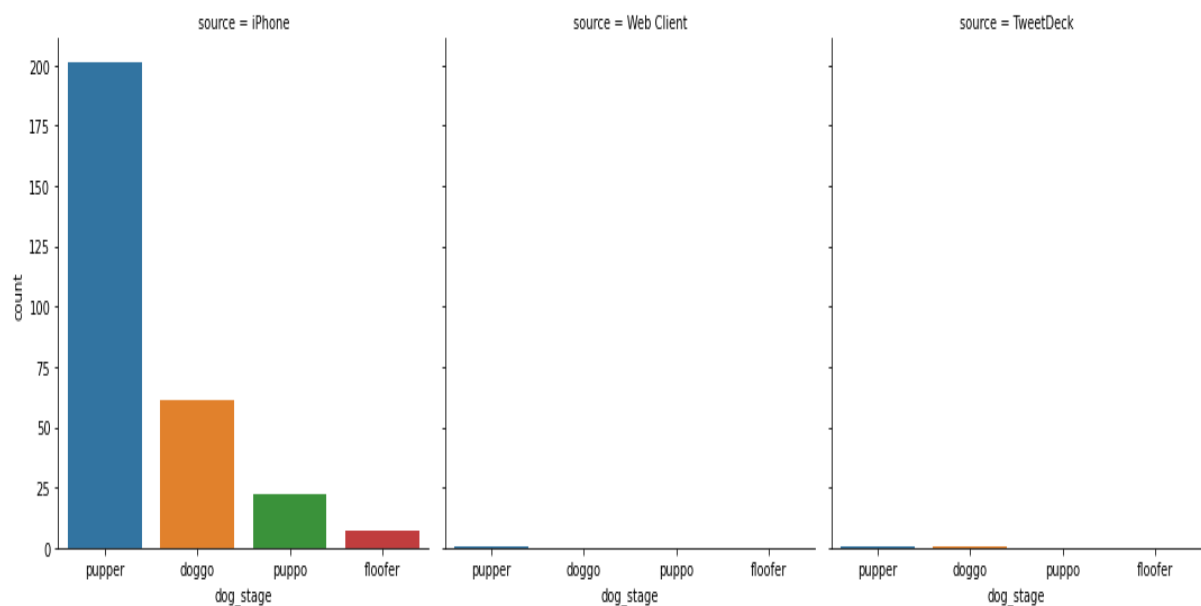
Question 2. what is the frequent number of images in the tweets?

Over 85% of the image that occurs in a tweet was a single image (**1**), this shows that out of the 305 images in the dataset, 85% of it were singly uploaded.



Question 3 What is the highest source of the tweets?

about **98%** of the tweets are from **iPhone source** and **pupper** is **tweeted most**. The might also contribute to the images with each tweets.



Question 4 What is the first 5 breed of dogs with highest favorite count?

The top dog breed is Golden Retriever, followed by Labrador Retriever, Pembroke, Chihuahua, and Samoyed. People love Retrievers according to the tweet analysis.

