

# act\_report

November 15, 2022

## 1 Data Wrangling Insights Report from WeRateDogs twitter archive

### 1.0.1 By Arikanawasi Udoka

The @WeRateDogs twitter page is a popular and well-beloved page on the twitter website where people send pictures of their dogs to be rated by the handler. The handler usually gives them a rating out of 10 and usually the ratings are greater than 10 because, “They’re good dogs, Brent.”

Each tweet usually contains, the dog’s name, a funny sentence, their rating, their dog stage (pupper, puppo, floofer, doggo), coined by the handler himself and finally the dog’s rating.

In this article, we’re going to explore which ratings, image numbers and dog stage is posted the most on the account.

```
In [20]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [21]: master = pd.read_csv('twitter_archive_master_clean.csv')
```

We will use this dataset where null/none rows have been dropped

```
In [22]: master.info()
```

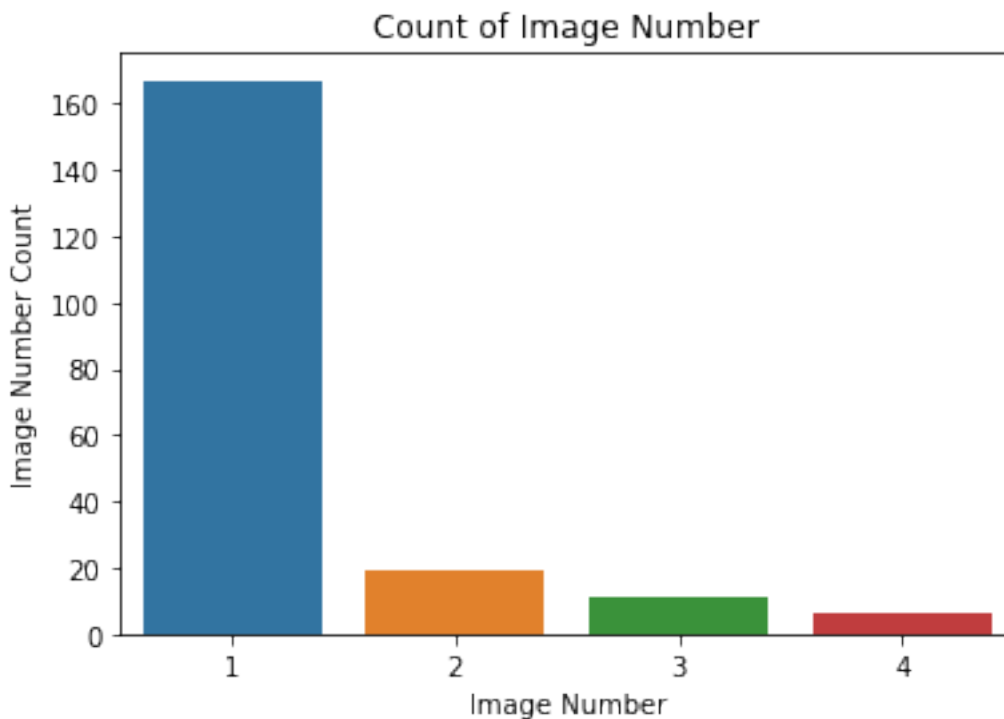
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 203 entries, 0 to 202
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   created_at            203 non-null   int64
1   tweet_id              203 non-null   int64
2   retweet_count         203 non-null   int64
3   favorite_count        203 non-null   int64
4   rating_numerator      203 non-null   int64
5   rating_denominator    203 non-null   int64
6   name                  203 non-null   object
7   timestamp             203 non-null   object
8   dog_stage             203 non-null   object
9   img_num              203 non-null   int64
10  dog_breed             203 non-null   object
```

```
11 p1_dog          203 non-null    bool
dtypes: bool(1), int64(7), object(4)
memory usage: 17.8+ KB
```

From the dataset above we will extract the following insights.

### 1.0.2 What is the most and least used "number" of images?

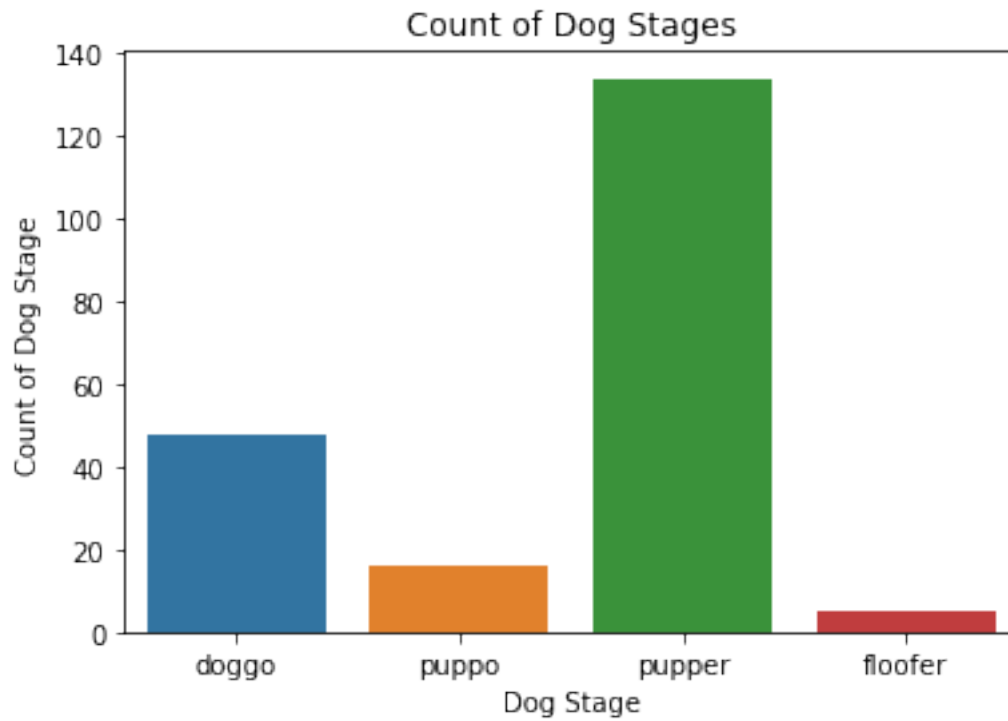
```
In [23]: sns.countplot(data=master, x='img_num').set(title='Count of Image Number');
plt.xlabel('Image Number');
plt.ylabel('Image Number Count');
```



A very high number of tweets contained only 1 image, with 4 being the least used image layout.

### 1.0.3 What is the most popular dog stage?

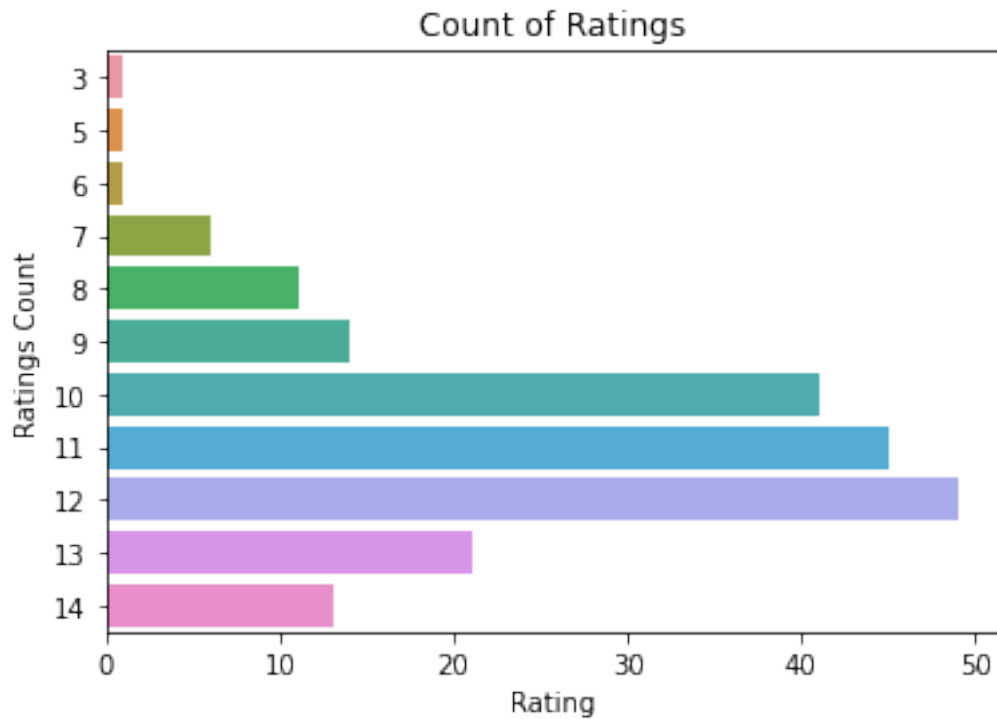
```
In [24]: sns.countplot(data=master, x='dog_stage').set(title='Count of Dog Stages');
plt.xlabel('Dog Stage');
plt.ylabel('Count of Dog Stage');
```



Pupper beat all the other dog stages hands down with floofer coming in last.

#### 1.0.4 What is the most popular dog rating?

```
In [25]: sns.countplot(data=master, y='rating_numerator').set(title='Count of Ratings');  
plt.xlabel('Rating');  
plt.ylabel('Ratings Count');
```



12 is the most popular dog rating with 11 and 12 following closely behind. We can see they're really good dogs as lower ratings have way smaller counts.

**Conclusion** This report is an overview on what the most popular ratings, dog stages and image layouts are. We can see that most dogs that get ratings are rated very high with most of their ratings falling between 10 - 12 and being puppies.