Shekhah Alhuwaydi

Analyzing Wrangled WeRateDogs Twitter Data

Introduction

The dataset that I will be wrangling (and analyzing and visualizing) is WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.
It was started in 2015 by college student Matt Nelson and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter.

These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs, Brent." WeRateDogs has over 4 million followers and has received international media coverage.

In this report, I provide the analysis of the clean WeRateDogs data and show the insight and the Visualization to answer my questions for the data :
How many Tweets and favorites?
What is the most repeated name?
What is the top prediction?
What is Top Dog Stage?

The insight:

1-Descriptive Statistics :

	rating_numerator	rating_denominator	rating	id_str	retweet_count	favorite_count	pre_conf
count	2151.000000	2151.0	2151.000000	2.151000e+03	2151.000000	2151.000000	1975.000000
mean	12.251976	10.0	1.225198	7.372893e+17	2762.140865	8781.050209	0.464325
std	42.699237	0.0	4.269924	6.762896e+16	4655.203638	12094.697050	0.339481
min	0.000000	10.0	0.000000	6.660209e+17	0.000000	52.000000	0.000000
25%	10.000000	10.0	1.000000	6.767941e+17	608.000000	1902.500000	0.140041
50%	11.000000	10.0	1.100000	7.099013e+17	1343.000000	4077.000000	0.458746
75%	12.000000	10.0	1.200000	7.896140e+17	3222.000000	11117.000000	0.773964
max	1776.000000	10.0	177.600000	8.924206e+17	79515.000000	132810.000000	0.999956

2- How many Tweets and favorites?
- we can see that the highest number of retweets of one tweet is 79515. and the favorite is 132810.

3- I wanted to see the count and, the top doge Name :

```
1  #count the top Name

Lucy          11
Charlie       11
Cooper        10
Oliver        10
Penny          9
              ..
Staniel        1
Chef           1
Fwed           1
Jeffri         1
Zooey          1
Name: name, Length: 928, dtype: int64
```
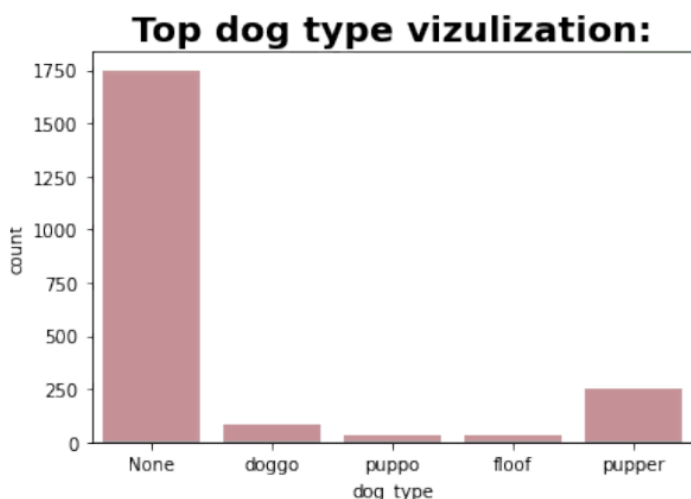
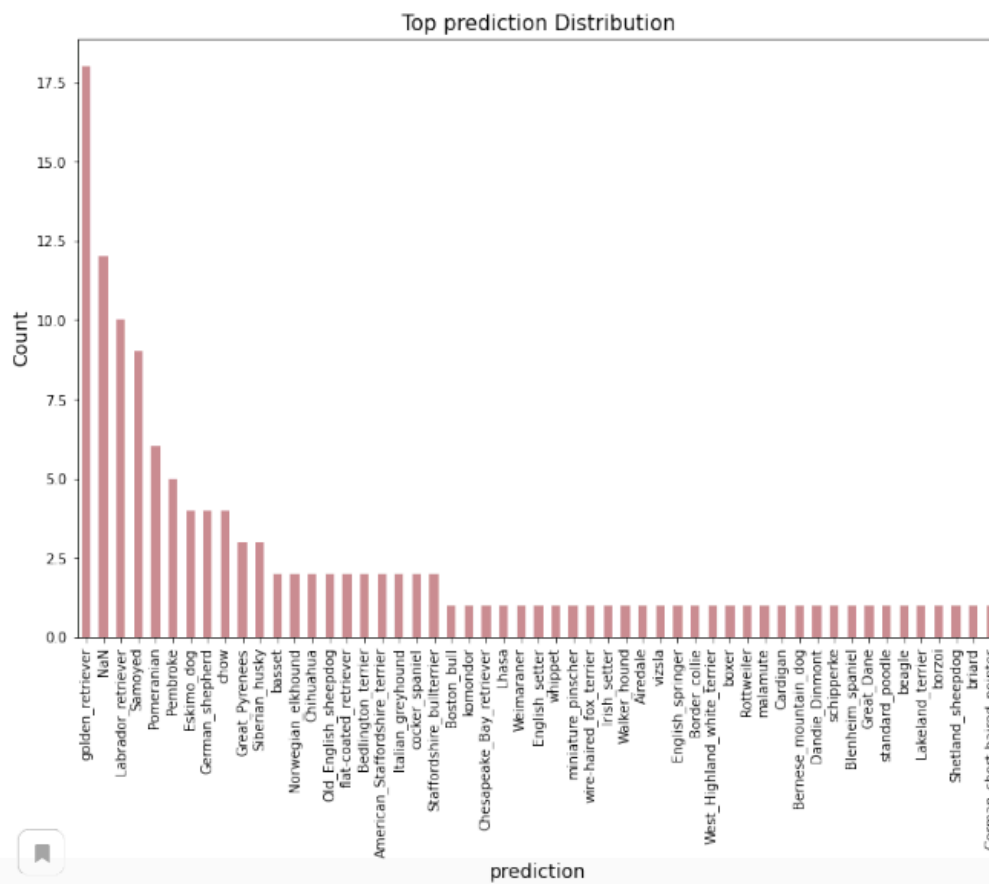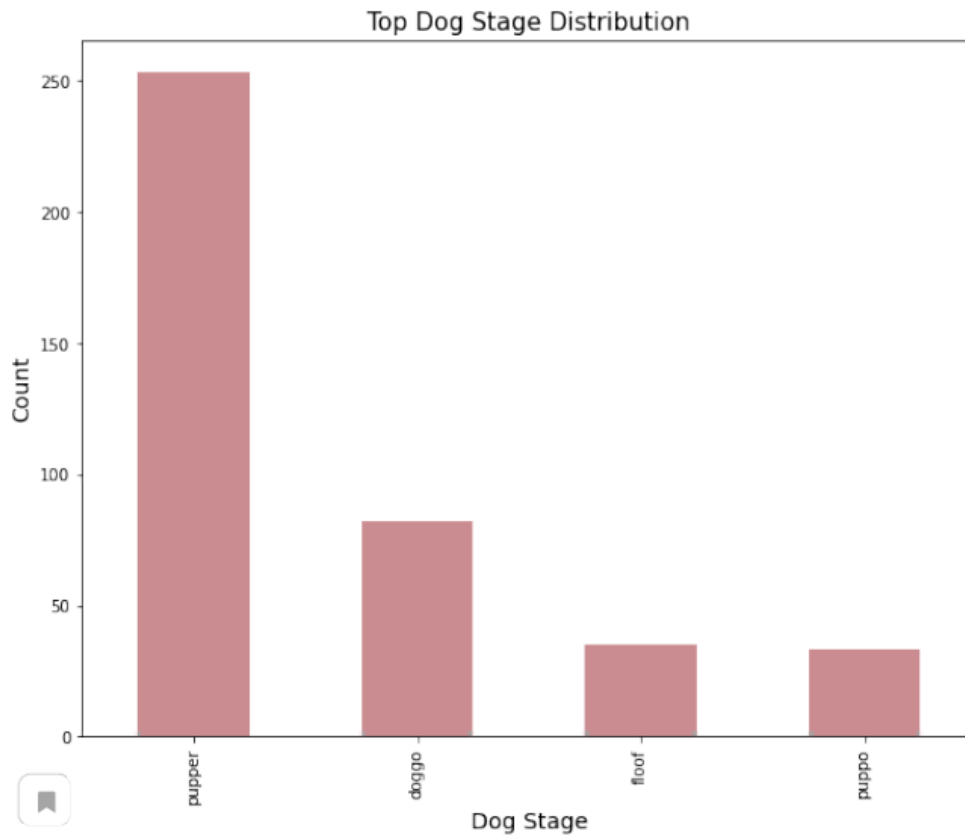we see that The top Names are Lucy and Charlie

4-What is the top prediction?

```
1  ## Top prediction:
2  Master2.prediction.value_counts()
```

```
NaN                      307
golden_retriever         154
Labrador_retriever       105
Pembroke                  95
Chihuahua                 90
                        ...
Scotch_terrier             1
Bouvier_des_Flandres       1
standard_schnauzer         1
silky_terrier              1
Japanese_spaniel           1
Name: prediction, Length: 114, dtype: int64
```

5-What is Top Dog Stage?

We can conclude from the previous analysis that There are dogs with undiscovered dog stages and also "NAN" predictions It is a very large number and will consume more time to clean. For now, I will exclude the category 'None' and continue to analyze it.

Top Dog Stage Distribution

Top prediction Distribution

So the result shows that - We can see that the top prediction is the golden_retrievevr and the Top Dog Stage is pupper after it is doggo