ACT REPORT: WeRateDogs

Table of contents

1. Analysis

I first analyzed the cleaned data using the methods describe, info and shape to get a brief understanding of what I was working with. Here are the results of the first analysis.

	tweet_id	rating_numerator	rating_denominator	p1_conf	p2_conf	p3_conf	retweet_count	favorite_count
count	1.968000e+03	1968.000000	1968.0	1968.000000	1.968000e+03	1.968000e+03	1968.000000	1968.000000
mean	7.357836e+17	11.697154	10.0	0.593152	1.348269e-01	6.029205e-02	2230.750000	7689.712398
std	6.757143e+16	40.934157	0.0	0.272093	1.007142e-01	5.080034e-02	3998.776866	11344.813889
min	6.660209e+17	0.000000	10.0	0.044333	1.011300e-08	1.740170e-10	11.000000	66.000000
25%	6.757330e+17	10.000000	10.0	0.360456	5.419487e-02	1.616932e-02	491.750000	1623.750000
50%	7.081401e+17	11.000000	10.0	0.587222	1.184015e-01	4.959335e-02	1073.000000	3456.000000
75%	7.878678e+17	12.000000	10.0	0.843827	1.955617e-01	9.174633e-02	2550.000000	9525.500000
max	8.924206e+17	1776.000000	10.0	1.000000	4.880140e-01	2.710420e-01	70332.000000	144247.000000

The questions I was looking to answer include:

a) Is there a relationship between ratings and retweet and favorite counts?
b) What are the top 10 most retweeted dogs?
c) What are the top 10 most favorite dogs?

- **Is there a relationship between ratings and retweet and favorite counts?**
  I grouped the data by retweet count and displayed the Top 10 mean ratings.

```
(retweet_count
 2270    1776.000000
 7555     420.000000
 5705      75.000000
 1490      27.000000
 438       15.333333
 5104      14.000000
 3992      14.000000
 7475      14.000000
 3978      14.000000
 5710      14.000000
 Name: rating_numerator, dtype: float64,
 None)
```

I also grouped the data by favorite count and displayed the top 10 mean ratings.

```
(favorite_count
 4766     1776.0
 22015     420.0
 17319      75.0
 6175       27.0
 1571       26.0
 6730       14.0
 82406      14.0
 18305      14.0
 42881      14.0
 20470      14.0
 Name: rating_numerator, dtype: float64,
 None)
```

- **What are the top 10 most retweeted dogs?**
  I grouped the data by the name of the dogs and displayed the top ten retweeted dogs.

```
(name
 Stephan    51423.0
 Duddles    37263.0
 Jamesy     30057.0
 Hurley     16933.0
 Aja        15734.0
 Lilly      15364.0
 Canela     14952.0
 Pipsy      14563.0
 Kenneth    13858.5
 Ken        13527.0
 Name: retweet_count, dtype: float64,
 None)
```

- **What are the top 10 most favorite dogs?**
  I grouped the data by the name of the dogs and display the top ten favorite dogs.

```
(name
 Stephan    111195.0
 Jamesy     108495.0
 Duddles     92500.0
 Aja         69042.0
 Lilly       62249.0
 Canela      59979.0
 Pipsy       42135.0
 Barney      41086.0
 such        40589.0
 Alfy        39041.0
 Name: favorite_count, dtype: float64,
 None)
```

Insights

- There is no specific relationship between ratings and retweet and favorite counts.
- Stephan is the most retweeted dog.
- Stephan is the most favorite dog (Such a celebrity !)

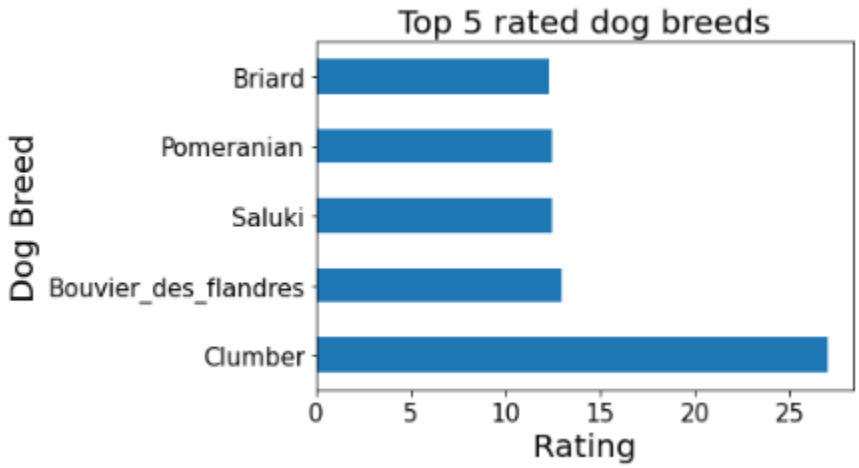
2. Visualization

What are the top 5 highest rated dog breeds?

I first grouped the data by dog breed and displayed their main ratings.

```
11.000000    9
10.000000    7
11.333333    5
10.333333    4
10.500000    4
             ..
11.307692    1
11.611111    1
11.176471    1
10.444444    1
10.200000    1
Name: rating_numerator, Length: 74, dtype: int64
```

I plotted the bar graph of the top five dog breeds against their mean ratings.



Clumber is the highest rated dog breed. Here is a photo.