

# WeRate Dogs Tweet's Analyses Report

WeRate Dogs is a Twitter Account and the goto place for sharing and rating Dogs.

As this Twitter account focuses on Dog rating my analyses will focus on popularity and distyrbution of different features.

```
In [4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
%config InlineBackend.figure_format = 'retina'

df_master = pd.read_csv('./twitter_archive_master.csv')
```

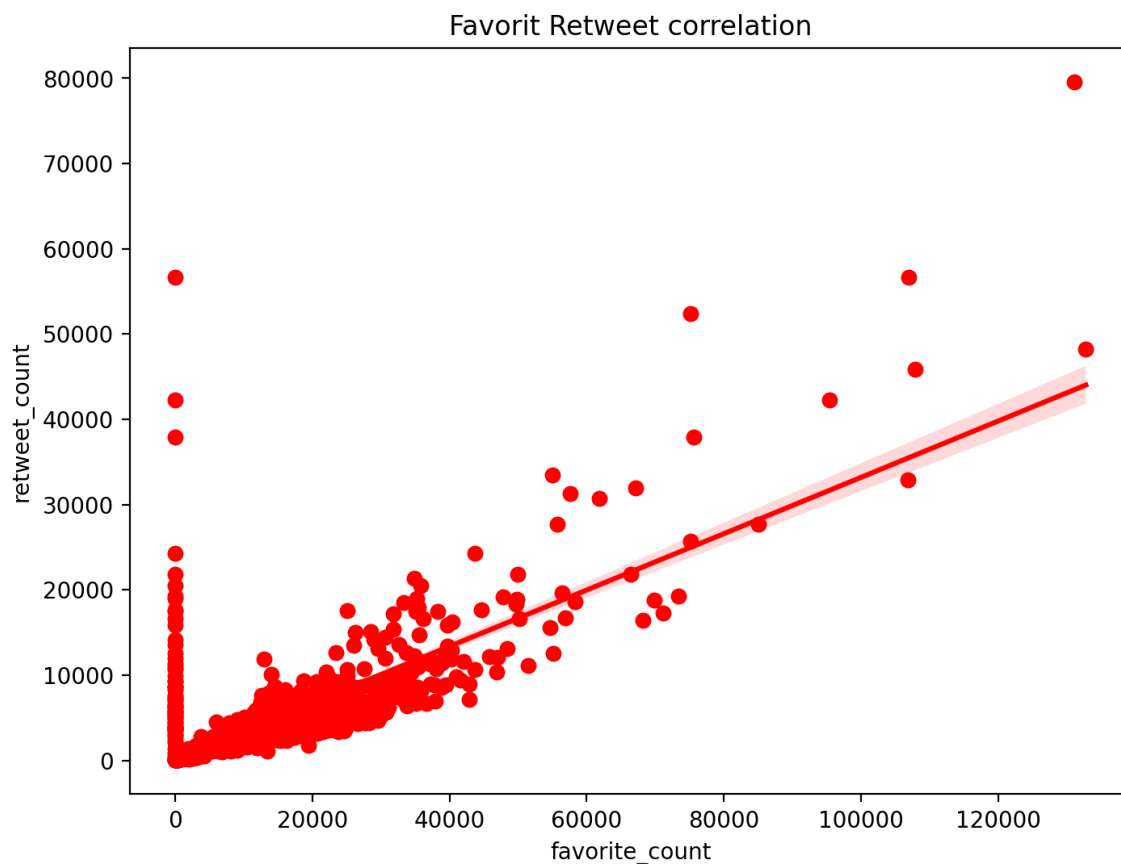
Question 1: How is the favorit count correlating to the retweets?

```
In [5]: plt.figure(figsize=(8, 6))
plt.title("Favorit Retweet correlation")

# Create scatter plot
sns.scatterplot(x="favorite_count", y="retweet_count", data=df_master, color=

# Add trendline
sns.regplot(x="favorite_count", y="retweet_count", data=df_master, color=

plt.show()
```



Interesting to see we we have a lot of tweets which are retweetetd but not marked as a favorite

**Question 2: How many Pupper, Floofer, puppo and Doggo's do we have in our Dataset?**

```
In [6]: counts = df_master[df_master['life_stage'] != 'None']['life_stage'].value
        colors = sns.color_palette('pastel')[0:4]

sns.set()

plt.figure(figsize=(6, 6))
plt.pie(counts.values, labels=counts.index, autopct='%1.1f%%')
plt.title('Dog Life stage distribution')

plt.pie(df_master.life_stage, colors=colors, autopct='%1.1f%%')
```

```

-----
ValueError                                Traceback (most recent call last)
/tmp/ipykernel_31301/4141980010.py in <module>
      8 plt.title('Dog Life stage distribution')
      9
--> 10 plt.pie(df_master.life_stage, colors=colors, autopct="%.0f%%")

/opt/conda/lib/python3.9/site-packages/matplotlib/pyplot.py in pie(x, explode, labels, colors, autopct, pctdistance, shadow, labeldistance, startangle, radius, counterclock, wedgeprops, textprops, center, frame, rotatelabels, normalize, data)
    2754         textprops=None, center=(0, 0), frame=False,
    2755         rotatelabels=False, *, normalize=True, data=None):
-> 2756     return gca().pie(
    2757         x, explode=explode, labels=labels, colors=colors,
    2758         autopct=autopct, pctdistance=pctdistance, shadow=shadow,

/opt/conda/lib/python3.9/site-packages/matplotlib/__init__.py in inner(ax, data, *args, **kwargs)
    1410     def inner(ax, *args, data=None, **kwargs):
    1411         if data is None:
-> 1412             return func(ax, *map(sanitize_sequence, args), **kwargs)
    1413
    1414         bound = new_sig.bind(ax, *args, **kwargs)

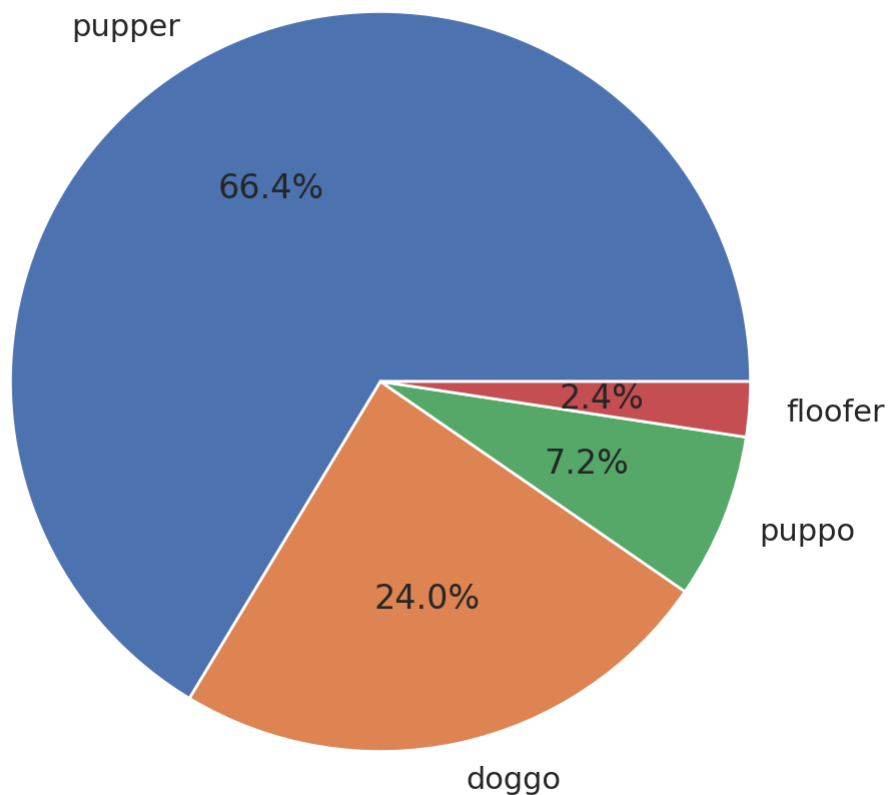
/opt/conda/lib/python3.9/site-packages/matplotlib/axes/_axes.py in pie(self, x, explode, labels, colors, autopct, pctdistance, shadow, labeldistance, startangle, radius, counterclock, wedgeprops, textprops, center, frame, rotatelabels, normalize)
    3038         # The use of float32 is "historical", but can't be changed without
    3039         # regenerating the test baselines.
-> 3040         x = np.asarray(x, np.float32)
    3041         if x.ndim > 1:
    3042             raise ValueError("x must be 1D")

/opt/conda/lib/python3.9/site-packages/pandas/core/series.py in __array__(self, dtype)
    870         dtype='datetime64[ns]')
    871         """
-> 872         return np.asarray(self._values, dtype)
    873
    874         # -----
-----

ValueError: could not convert string to float: 'None'

```

## Dog Life stage distribution



The conclusion is that most of our Dogs are pupper! The smallest amount of dogs the dataset includes are floofer which is really sad

### Question 3: The top 5 Names for Dogs

```
In [7]: df_master.name.value_counts()[0:10]
```

```
Out[7]: None      2308  
a           220  
Charlie     44  
Penny       40  
Oliver      40  
Tucker      40  
Cooper      40  
Lucy        40  
Bo          32  
Lola        32  
Name: name, dtype: int64
```

We can see clearly what Charlie is the most popular Name followed by Penny, Oliver, Tucker, Cooper and Lucy!

So clearly it is not possible to give you only 5 as they have the same count. Now we have the top 6