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 anayses will focus on popularity and distrybution of different features.

In [ ]:
```
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')
```

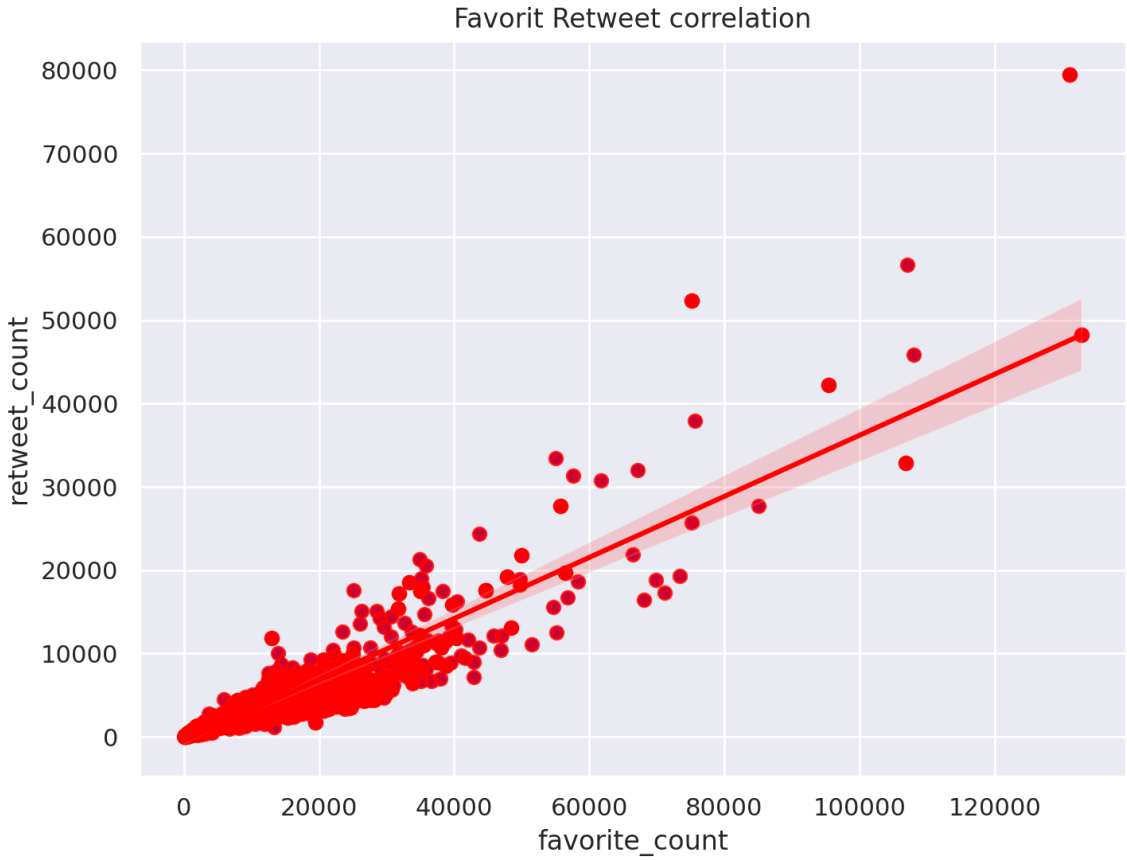
Question 1: How is the favorit count correlating to the retweets?

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


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

# 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

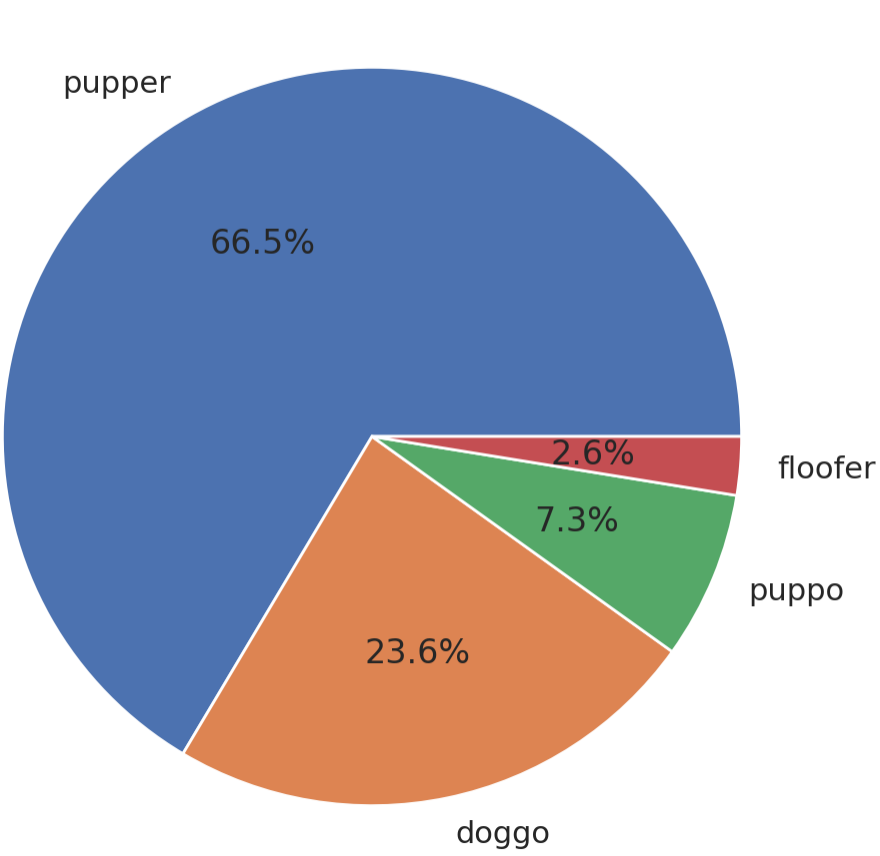
Question 2: How many Pupper, Floofer, puppo and Doggo's do we have in our Dataset?

In [ ]:
```
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')
```

Out[ ]:
```
Text(0.5, 1.0, '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 realy sad

Question 3: The top 5 Names for Dogs

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

count
```

Out[ ]:
```
None       669
Cooper      13
Oliver      12
Charlie     12
Lucy        11
the          9
Winston      9
Tucker       9
Penny        9
Sadie        8
Name: name, dtype: int64
```

Looking at all real names we can see clearly what Cooper is the most popular Name followed by Oliver, Charlie, Lucy, Winston