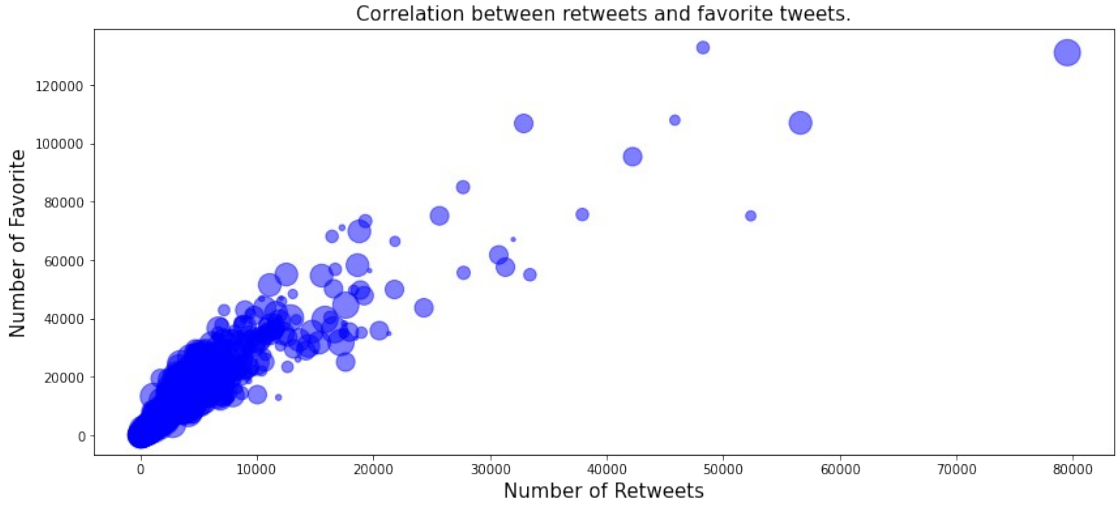
Data Act Report from "We Rate Dogs" dataset

After gathering, assessing and cleaning the datasets. I moved on to analyze and visualize in other to derive insights from the cleaned data. The first step was to create viz dataset by selecting specific columns relevant to my analysis. After importing all necessary libraries for visualization, My first question was I wanted to see if there was a correlation between **retweet_counts (retweets)** and **favorite_count (likes)**. To answer this question, I plotted a scattered plot as shown below

```
plt.figure(figsize= [14,6])
plt.scatter(x = viz_df['retweet_count'],
            y = viz_df['favorite_count'],alpha=0.5,color= 'blue',
sizes = np.array([20,50,100,200,200,300,60,90,10,300,200,400,75]))
plt.xlabel('Number of Retweets', fontsize = 15)
plt.ylabel('Number of Favorite', fontsize = 15)
plt.title('Correlation between retweets and favorite tweets.',
          fontsize = 15)
plt.show()
```



the correlation between retweets and likes is somewhat positive

Q2: Which is the most popular stage among dogs

```
viz_df.stage_name.value_counts()
```

```
                1650
pupper           201
doggo             63
puppo             22
doggo, pupper      8
floofer            7
doggo, puppo       1
doggo, floofer     1
Name: stage_name, dtype: int64
```

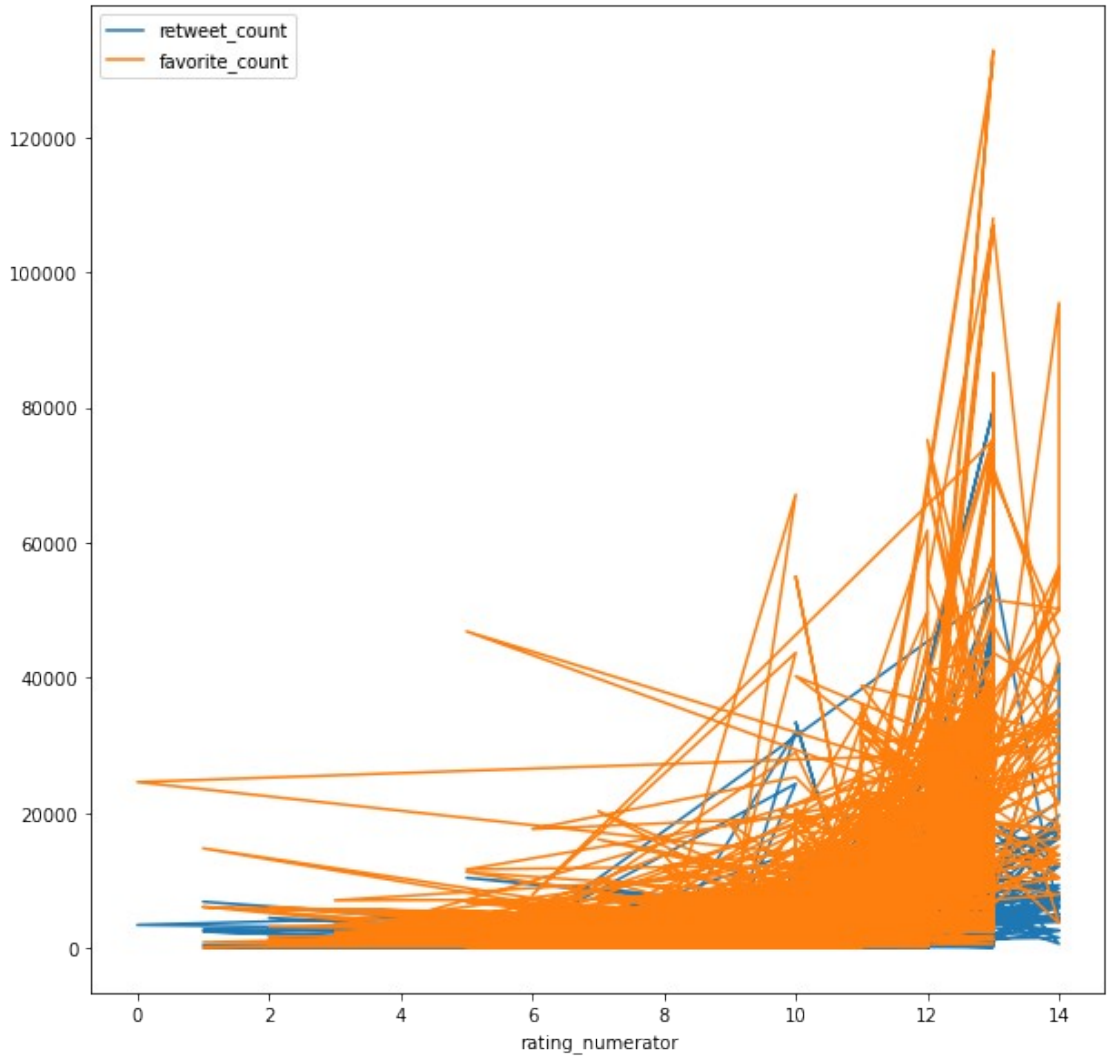
We can see that Pupper is very popular amongst dog stages

Q3: Does a high numerator rating correlates to high retweets and favorite count?

```
viz_df.plot([x ="retweet_count", "favorite_count"],
y="numerator_rating", kind="line",bins=7)
```

```
  Input In [482]
    viz_df.plot([x ="retweet_count", "favorite_count"],
y="numerator_rating", kind="line",bins=7)
                  ^
SyntaxError: invalid syntax
```

```
viz_df.plot(x="rating_numerator",
y=["retweet_count","favorite_count"], kind="line", figsize=(10,10));
```



Though the viz isnt that great but we can see that likes and retweets increases as rating numerator increases

Insights:

1. There's a probability that the rating numrator is actually a factor of wether the photo gets more likes and retweets
2. Pupper is the most popular stage dogs on the dataset
3. The is a strong correlation between favorite_count and retweet_count