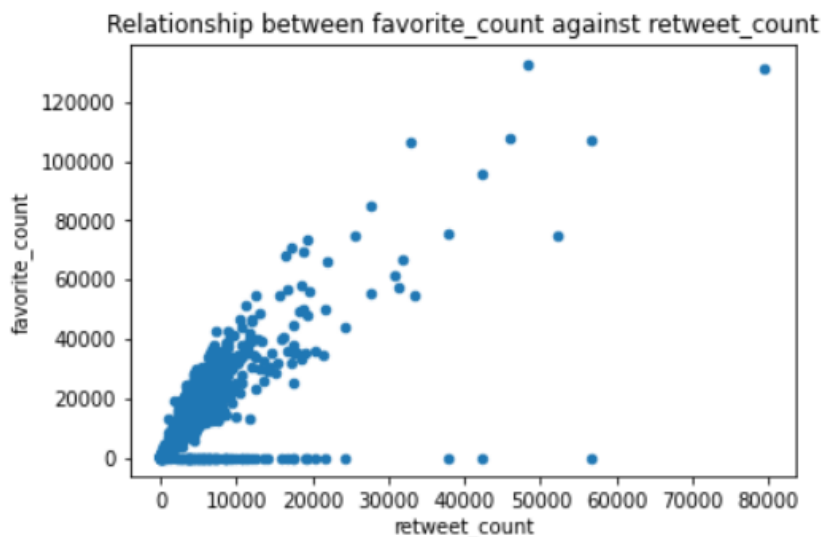WeRateDogs Twitter Dataset wrangling and analyzing project

Visualizations and Insights

The graph **below** shows the relationship between the favorite count and the retweet count using a scatterplot, and it can be seen that there is a positive relationship between the two.

This is likely expected as the most favored are also likely to get the most retweets.
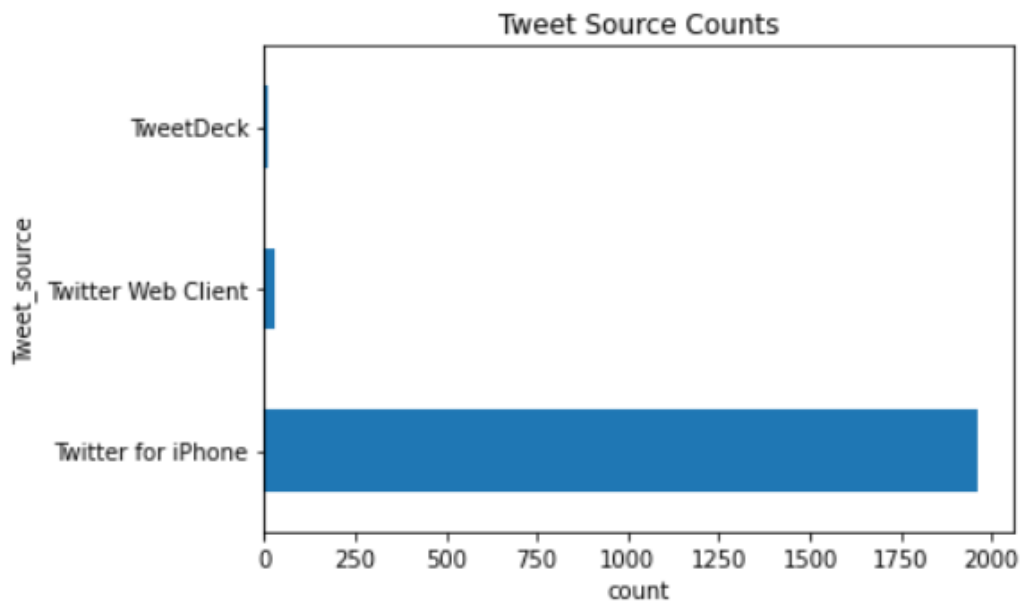
```
df.plot(x='retweet_count', y='favorite_count', kind='scatter')
plt.title('Relationship between favorite_count against retweet_count')
plt.show()
plt.savefig('Scatterplots of favorite_count against retweet_count.jpg')
files.download('Scatterplots of favorite_count against retweet_count.jpg')
```


Relationship between favorite_count against retweet_count

```
<Figure size 432x288 with 0 Axes>
```

The graph **below** shows the plot of the count of the tweet sources, and the source with the highest tweet is the iPhone.

```
[98]  df.source.value_counts().plot(kind='barh')
      plt.xlabel('count')
      plt.ylabel('Tweet_source')
      plt.title('Tweet Source Counts')
      plt.show()
      plt.savefig('Tweet Source Counts.jpg')
```



```
<Figure size 432x288 with 0 Axes>
```
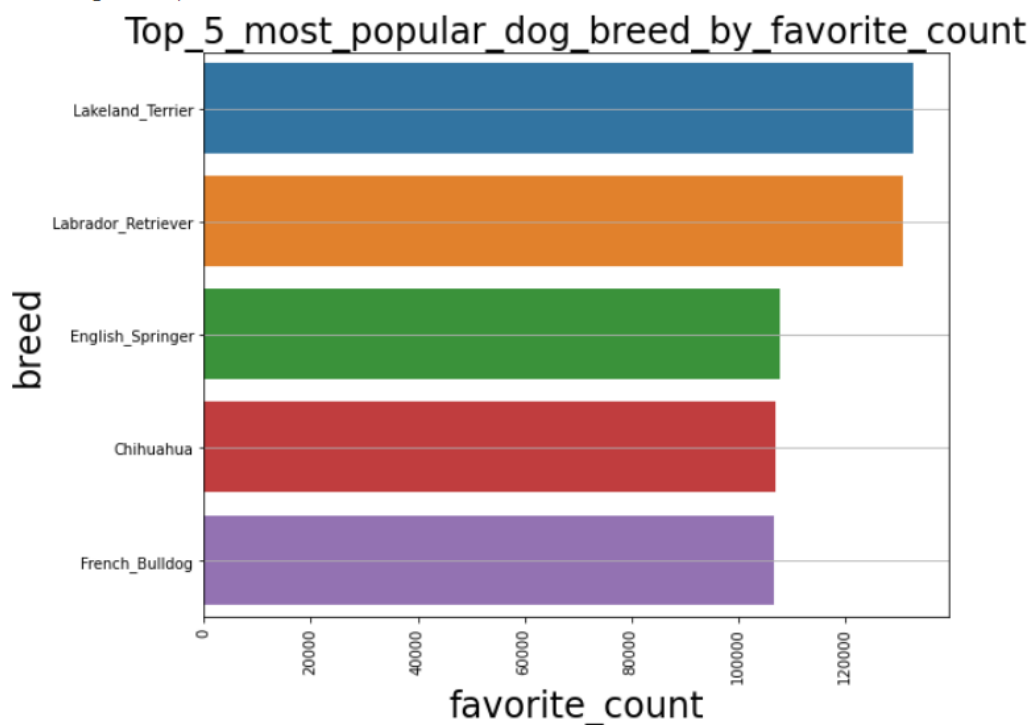
The graph **below** shows the top 5 most popular dogs categorized by favorite count, and it showed that the most popular dog by favorite count is the Lakeland_terrier, closely followed by the Labrador_Retriever.

Top 5 most popular dog breed by favorite count?

```
[99] Top_5_most_popular_dog_breed_by_favorite_count = df[['breed', 'favorite_count
     data = Top_5_most_popular_dog_breed_by_favorite_count
     data
```

	breed	favorite_count
1672	Lakeland_Terrier	132810
1158	Labrador_Retriever	131075
1942	English_Springer	107956
1571	Chihuahua	107015
1889	French_Bulldog	106827

Top_5_most_popular_dog_breed_by_favorite_count

```
<Figure size 432x288 with 0 Axes>
```

The graph **below** shows the top 10 most common dog names; the most common name is Charlie, followed by Cooper. Charlie is likely a male name,

and Cooper is also likely a male name, but the dataset doesn't tell us the gender, which is a limitation.

Top 10 most common dog names?

```
[107] df.name.value_counts().head(10)

     Charlie     11
     Cooper      10
     Oliver      10
     Lucy        10
     Penny        9
     Tucker       9
     Sadie        8
     Winston      8
     Daisy        7
     Lola         7
     Name: name, dtype: int64
```

The breed with the highest retweet count was assessed, and it was seen that Labrador_Retriever is the breed with the highest retweet count. It is also worthy of note that it ranked the second highest when the favorite count grouped it.

Which breed has the highest retweet count?

```
[102] most_popular_dog_breed_by_retweet_count = df[['breed', 'retweet_count']].sort_values(
     data = most_popular_dog_breed_by_retweet_count
     data
```

	breed	retweet_count
1158	Labrador_Retriever	79515

Summary of key findings

1. There is a positive relationship between retweet_count and favorite_count

2. The most common tweet source is twitter_iphone
3. The most common breed by favorite count is the Lakeland Terrier
4. The most common breed by retweet count is also the Labrador retriever
5. The most common dog name is Charlie

Limitations

1. The dataset could not tell the gender of the dogs, which could have been useful for further analysis