Report

Wrangle and analyze data project

The steps that I have followed in this project is:

Gathering

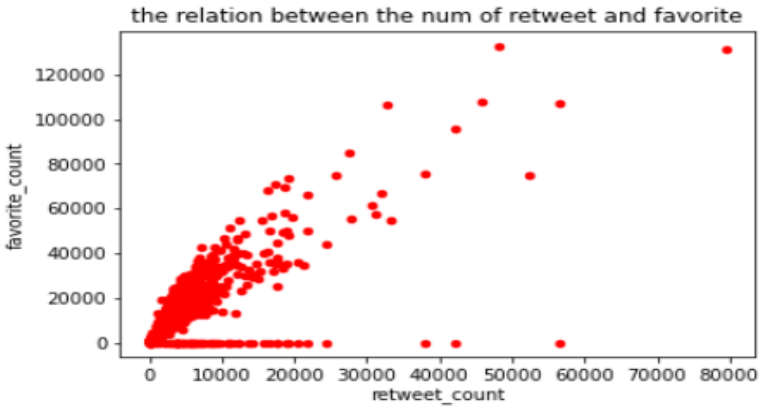
Assessing

Cleaning

Sorting and Visualizing

In this file I will talk briefly about the visuals and insights that I have made.

[505]:
```
df.plot('retweet_count','favorite_count',color='red',kind='scatter')
plt.xlabel('retweet_count')
plt.ylabel('favorite_count')
plt.title('the relation between the num of retweet and favorite')
plt.show()
```

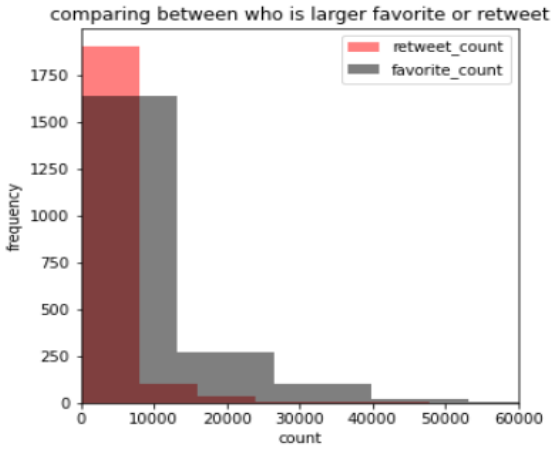


Insight 1

This gives the relation between the number of the favorite count and the retweet count , its seem that there is a positive strong relation between them so as retweet increase the favorite increase, and the larger favorite and retweet from the pic above is approximately 138000 and 47000 respectively.

]:
```
df['retweet_count'].plot(kind='hist',alpha=0.5,color='red',label='retweet_count',figsize=(5,5))
df['favorite_count'].plot(kind='hist',alpha=0.5,color='black',label='favorite_count',figsize=(5,5))
plt.legend()
# to change the default range of the x axis
plt.xlim(xmin=0,xmax=60000)
plt.xlabel('count')
plt.ylabel('frequency')
plt.title('comparing between who is larger favorite or retweet')
```

]:
```
Text(0.5, 1.0, 'comparing between who is larger favorite or retweet')
```

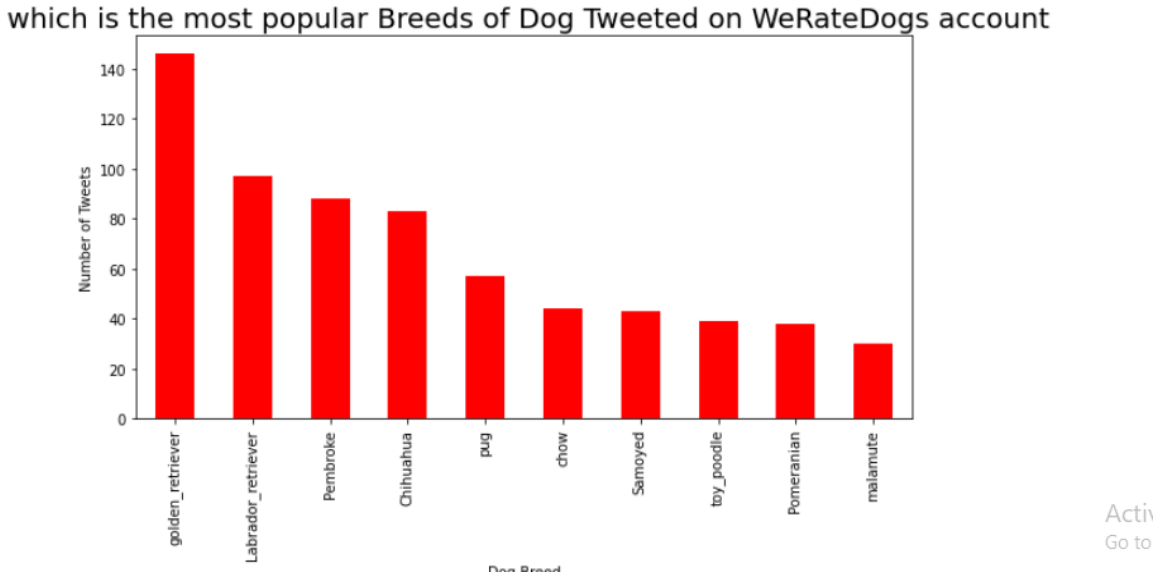


Insight 2

This pic tells us which range has more frequent in the data.

We can say that the favorite is more than the retweet in many ranges ,that is the pictures in WeRateDogs account has more favorite than the retweet.

```
[533]: fig = plt.figure(figsize=(10,5))
df.groupby('breed')['breed'].count().sort_values(ascending=False).head(10).plot(kind='bar',color='red')
plt.title(" which is the most popular Breeds of Dog Tweeted on WeRateDogs account",fontsize=20)
plt.xlabel("Dog Breed")
plt.ylabel("Number of Tweets");
```



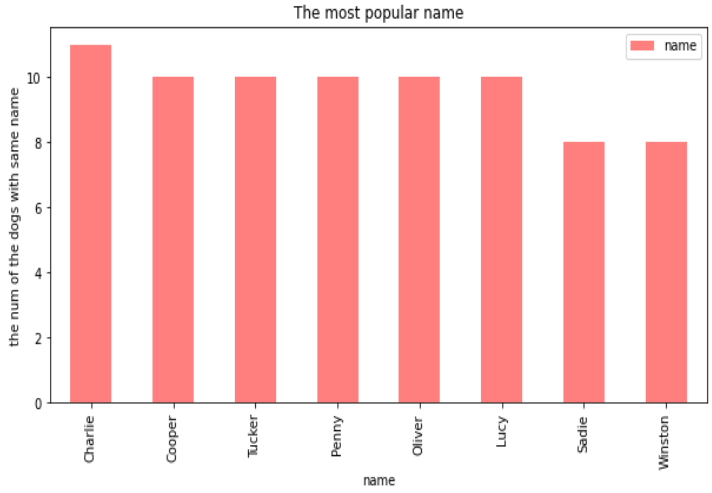
Insight 3

From the pic above we are trying to compare between the dogs breed to see which breed is more popular in WeRateDogs account.

And we figure out that the golden_retriever breed is the most popular in this project.

```
In [532]: df['name'].value_counts().sort_values(ascending=False).head(8).plot(kind='bar',alpha=0.5,color='red',label='name',figsize=(10,
          plt.legend()
          plt.xlabel('name')
          plt.ylabel('the num of the dogs with same name')
          plt.title('The most popular name')
```

```
Out[532]: Text(0.5, 1.0, 'The most popular name')
```



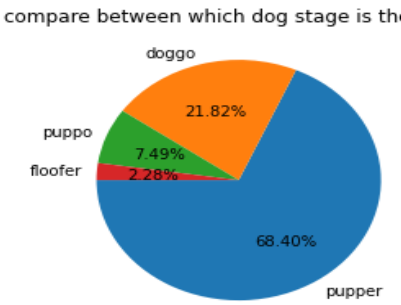
Insight 4

Here we are looking for the most popular dog name in WeRateDogs account and it seems that Charlie is the most popular name in this project with 11 dogs and

then followed by Lucy ,Copper ,Oliver ,Tucker and Penny with 10 dogs for each dog name.

```
: plt.title('compare between which dog stage is the larger')
  plt.axis('equal')
  plt.pie(perc,labels=['pupper','doggo','puppo','floofer'],autopct='%1.2f%%',startangle=180)
```

```
: ([<matplotlib.patches.Wedge at 0x29a93238040>,
    <matplotlib.patches.Wedge at 0x29a9323f640>,
    <matplotlib.patches.Wedge at 0x29a9323fbe0>,
    <matplotlib.patches.Wedge at 0x29a93244280>],
   [Text(0.601146989610214, -0.9212069783075776, 'pupper'),
    Text(-0.2946553264986708, 1.059801037254617, 'doggo'),
    Text(-1.022089812397166, 0.40661088941877377, 'puppo'),
    Text(-1.0971790568876423, 0.07872812157764277, 'floofer')],
   [Text(0.3278983579692076, -0.5024765336223149, '68.40%'),
    Text(-0.16072108718109315, 0.5780732930479727, '21.82%'),
    Text(-0.5575035340348177, 0.22178775786478566, '7.49%'),
    Text(-0.5984613037568958, 0.04294261176962332, '2.28%')])
```



Insight 5

In the last visual we are trying to compare between the four dog stage and see which dog stage is the most in this data .

From the pie-chart above we can see that the Pupper dog stage has the mosr percentage compared to the other three with 68.4%