act_report

June 15, 2020

0.1 Act_report

Some of the insights that found in the analysis of the dataset:

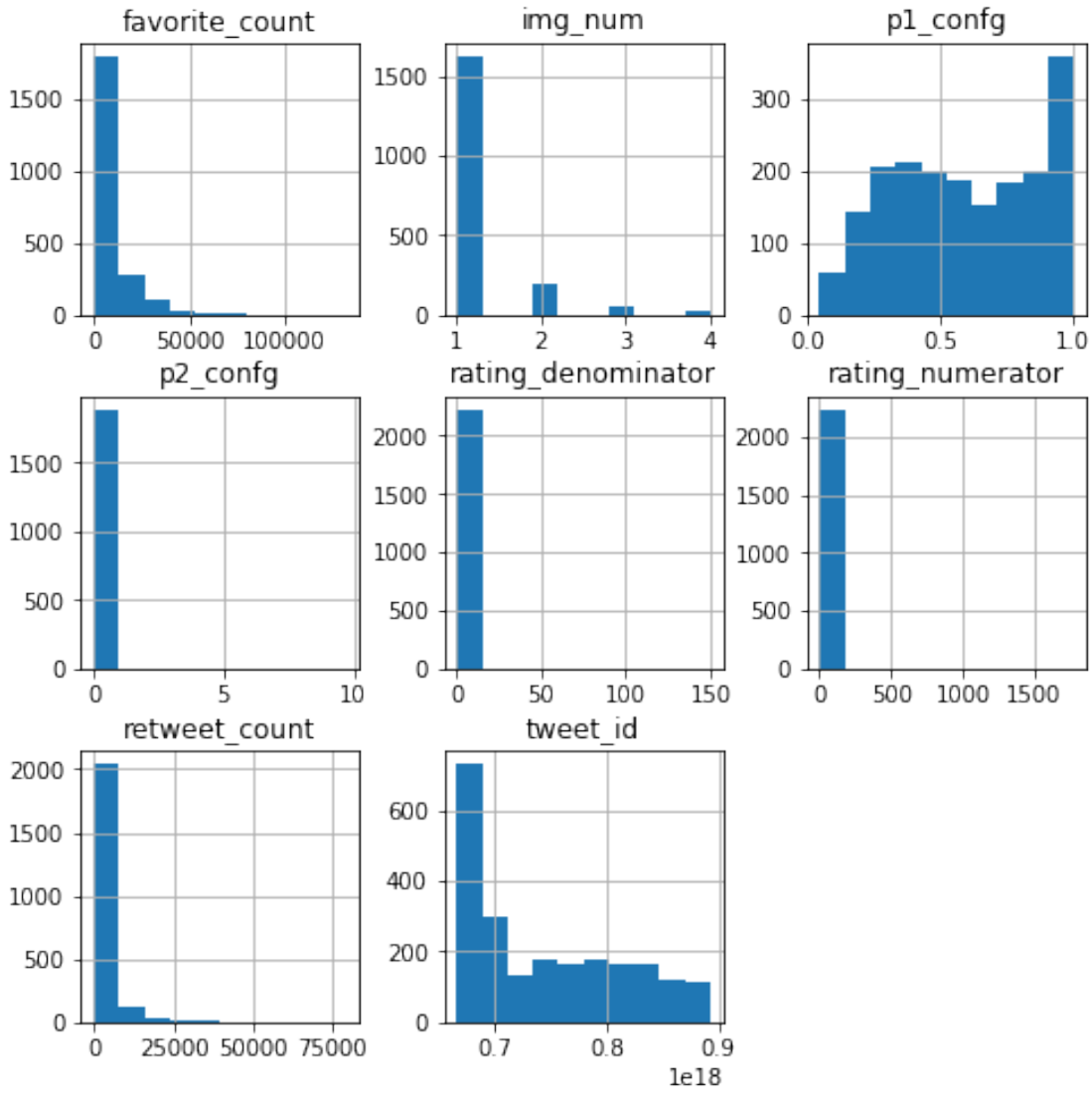
- No. of retweet_count is almost directly propotional to no. of favorite_count i.e more more favorite_count will result in more retweet.
- The algorithms after seeing the images of animals predicted that approximately 75% of them are dogs.This means that if there are 100 images then 75 of them were of dogs.
- Around 88% of the dogs are rated between 8-12.This means that weratedogs has a strict policies of rating dogs and they take it very seriously.

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
        df=pd.read_csv('weratedogmaster.csv')
        df.head()
```

```
Out[1]:              tweet_id                  timestamp  \
        0  892420643555336193  2017-08-01 16:23:56 +0000
        1  892177421306343426  2017-08-01 00:17:27 +0000
        2  891815181378084864  2017-07-31 00:18:03 +0000
        3  891689557279858688  2017-07-30 15:58:51 +0000
        4  891327558926688256  2017-07-29 16:00:24 +0000

                                                source  \
        0  <a href="http://twitter.com/download/iphone" r...
        1  <a href="http://twitter.com/download/iphone" r...
        2  <a href="http://twitter.com/download/iphone" r...
        3  <a href="http://twitter.com/download/iphone" r...
        4  <a href="http://twitter.com/download/iphone" r...

                                                  text  \
        0  This is Phineas. He's a mystical boy. Only eve...
        1  This is Tilly. She's just checking pup on you...
        2  This is Archie. He is a rare Norwegian Pouncin...
        3  This is Darla. She commenced a snooze mid meal...
        4  This is Franklin. He would like you to stop ca...
```

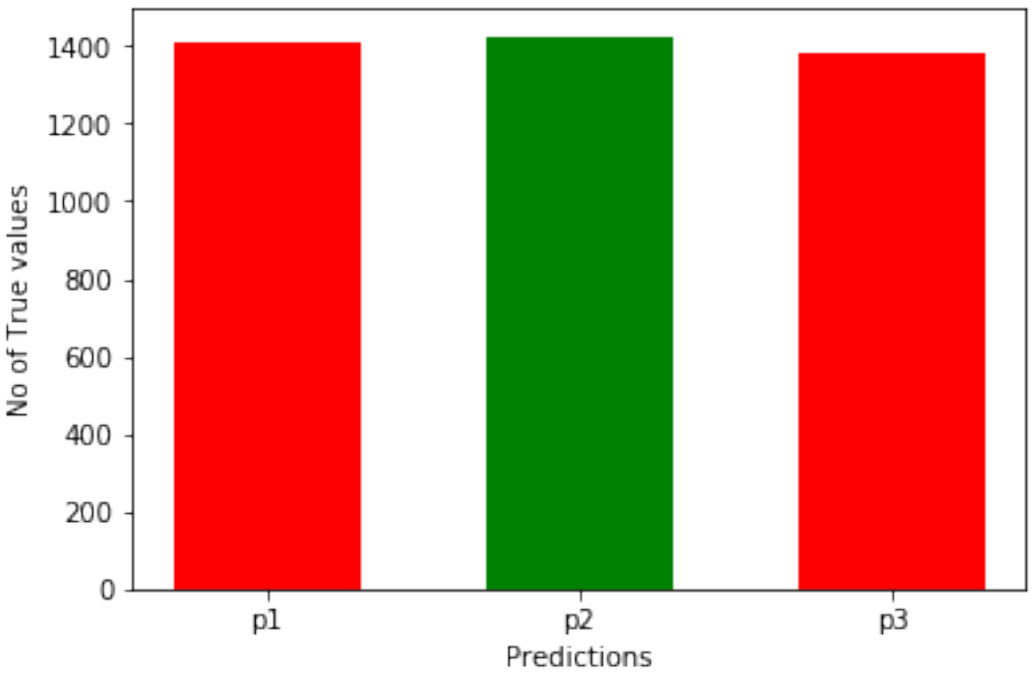
```
                                          expanded_urls  rating_numerator  \
0  https://twitter.com/dog_rates/status/892420643...              13.0
1  https://twitter.com/dog_rates/status/892177421...              13.0
2  https://twitter.com/dog_rates/status/891815181...              12.0
3  https://twitter.com/dog_rates/status/891689557...              13.0
4  https://twitter.com/dog_rates/status/891327558...              12.0

   rating_denominator      name  retweet_count  favorite_count  ...   img_num  \
0                10.0   Phineas         8853.0         39467.0  ...       1.0
1                10.0     Tilly         6514.0         33819.0  ...       1.0
2                10.0    Archie         4328.0         25461.0  ...       1.0
3                10.0     Darla         8964.0         42908.0  ...       1.0
4                10.0  Franklin         9774.0         41048.0  ...       2.0

            p1  p1_confg p1_dog                 p2 p2_confg p2_dog  \
0       orange   0.09704  False              bagel  0.08585  False
1    Chihuahua   0.32358   True           Pekinese  0.09064   True
2    Chihuahua   0.71601   True           malamute  0.07825   True
3  paper_towel   0.17027  False  Labrador_retriever  0.16808   True
4       basset   0.55571   True   English_springer  0.22576   True

                           p3 p3_confg p3_dog
0                      banana  0.07611  False
1                    papillon  0.06895   True
2                      kelpie  0.03137   True
3                     spatula  0.04083  False
4  German_short-haired_pointer  0.17521   True

[5 rows x 22 columns]
```

In [2]:
```
df.hist(figsize=(8,8));
```

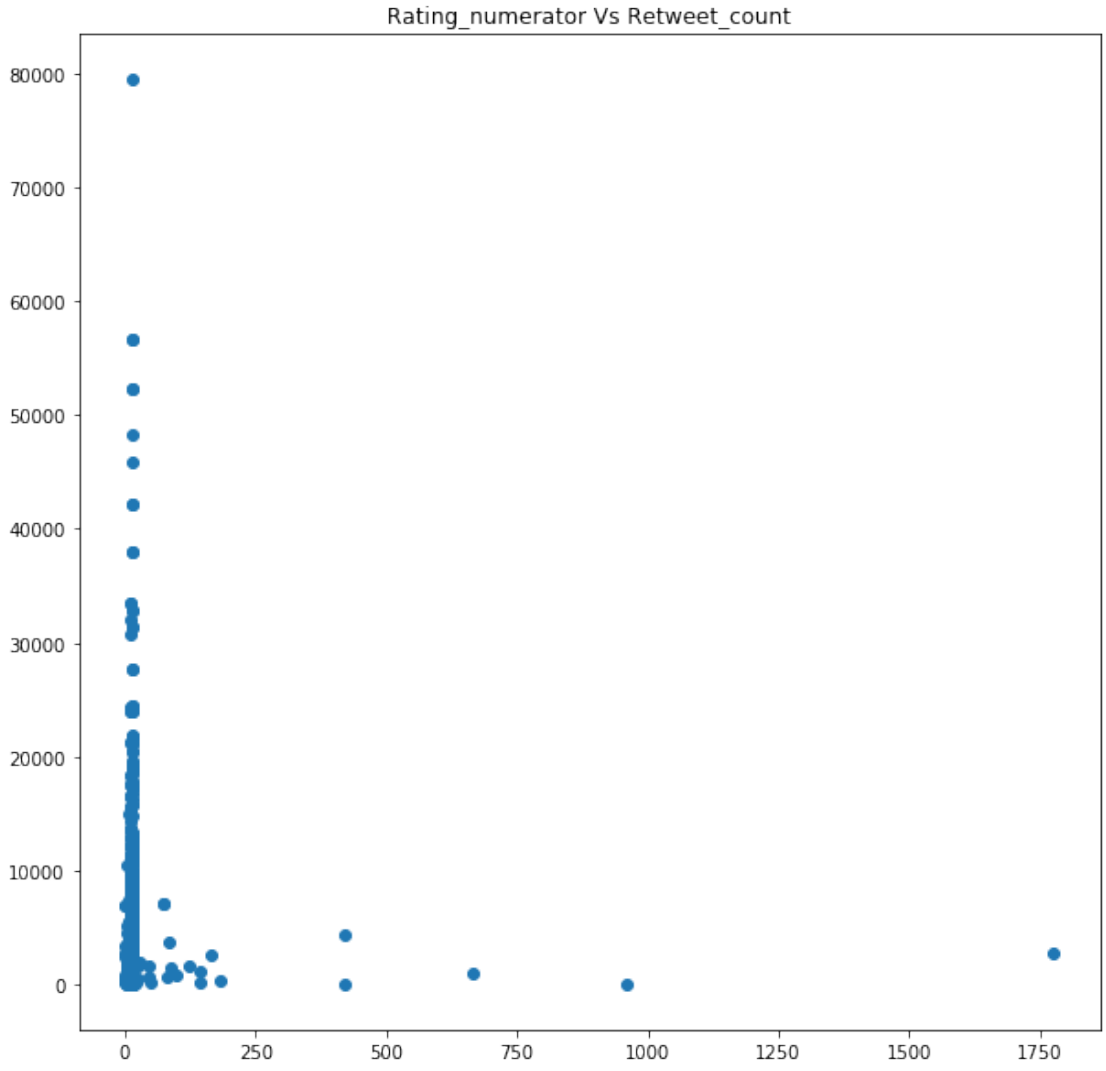



```
In [3]: p1=df.groupby('p1_dog').count()['p1']
        p1

Out[3]: p1_dog
        False     499
        True     1406
        Name: p1, dtype: int64

In [4]: p2=df.groupby('p2_dog').count()['p1']
        p2

Out[4]: p2_dog
        False     480
        True     1425
        Name: p1, dtype: int64
```

```
In [5]: p3=df.groupby('p3_dog').count()['p1']
        p3

Out[5]: p3_dog
        False     522
        True     1383
        Name: p1, dtype: int64

In [6]: location=[1,2,3]
        heights=[p1[1],p2[1],p3[1]]
        labels=['p1','p2','p3']
        plt.bar(location,heights,color=['red','green'],width=0.6,tick_label=labels)
        plt.ylabel('No of True values')
        plt.xlabel('Predictions')

Out[6]: Text(0.5,0,'Predictions')
```

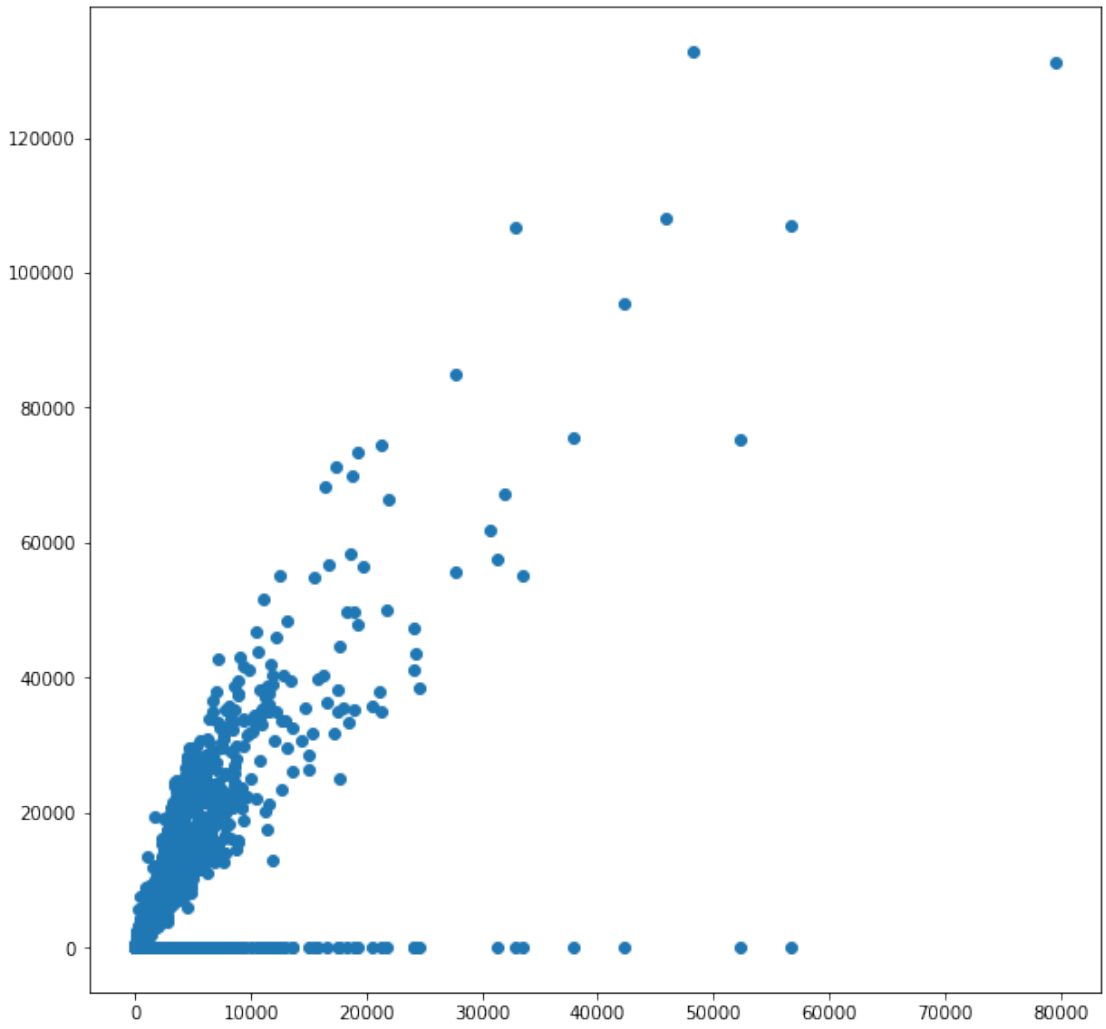



```
In [7]: plt.figure(figsize=(10,10))
        ax=plt.scatter(df.rating_numerator,df.retweet_count)
        plt.title('Rating_numerator Vs Retweet_count')

Out[7]: Text(0.5,1,'Rating_numerator Vs Retweet_count')
```




```
In [8]: plt.figure(figsize=(10,10))
        plt.scatter(df.retweet_count,df.favorite_count)
```

```
Out[8]: <matplotlib.collections.PathCollection at 0x7f2ff16b30f0>
```


120000
100000
80000
60000
40000
20000
0
0
10000
20000
30000
40000
50000
60000
70000
80000

In [ ]: