# Act report

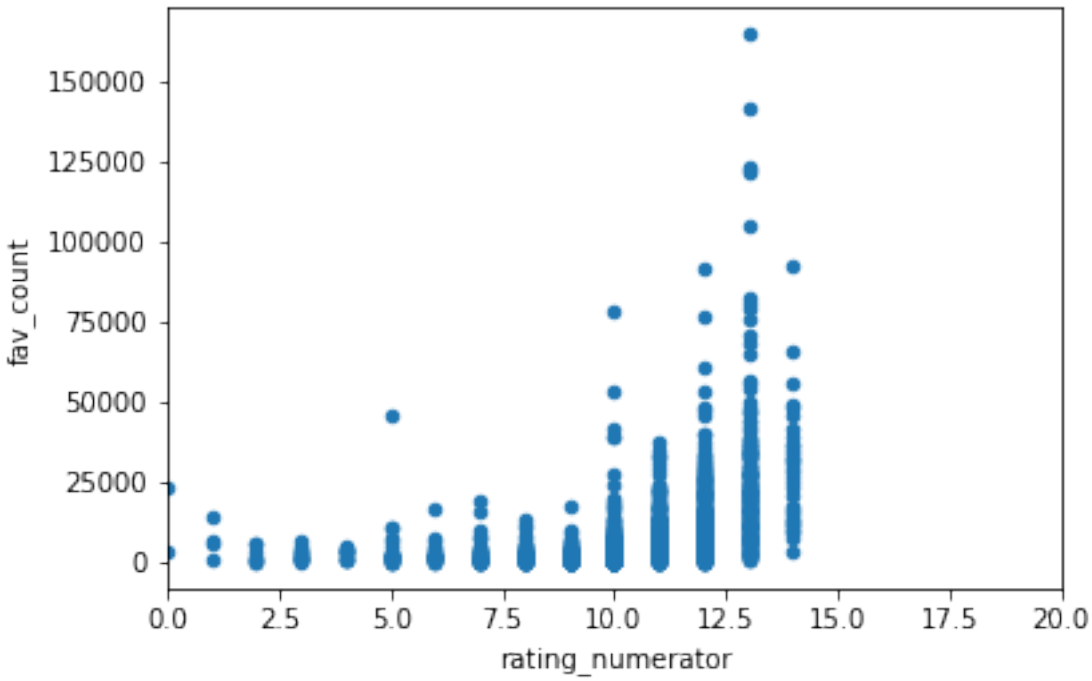
September 18, 2018

## 1 Insights

### 1.1 Favorite count and rating numerator

In [42]:
```
df.plot('rating_numerator','fav_count',kind='scatter')
plt.xlim(0,20)
```

Out[42]: (0, 20)




In [36]:
```
df.fav_count.describe()
```

Out[36]:
```
count      1992.000000
mean       8804.153614
std       12828.611725
```

```
min          80.000000
25%        1895.750000
50%        3995.000000
75%       11010.500000
max      164706.000000
Name: fav_count, dtype: float64
```

```
In [37]: df.rating_numerator.describe()
```

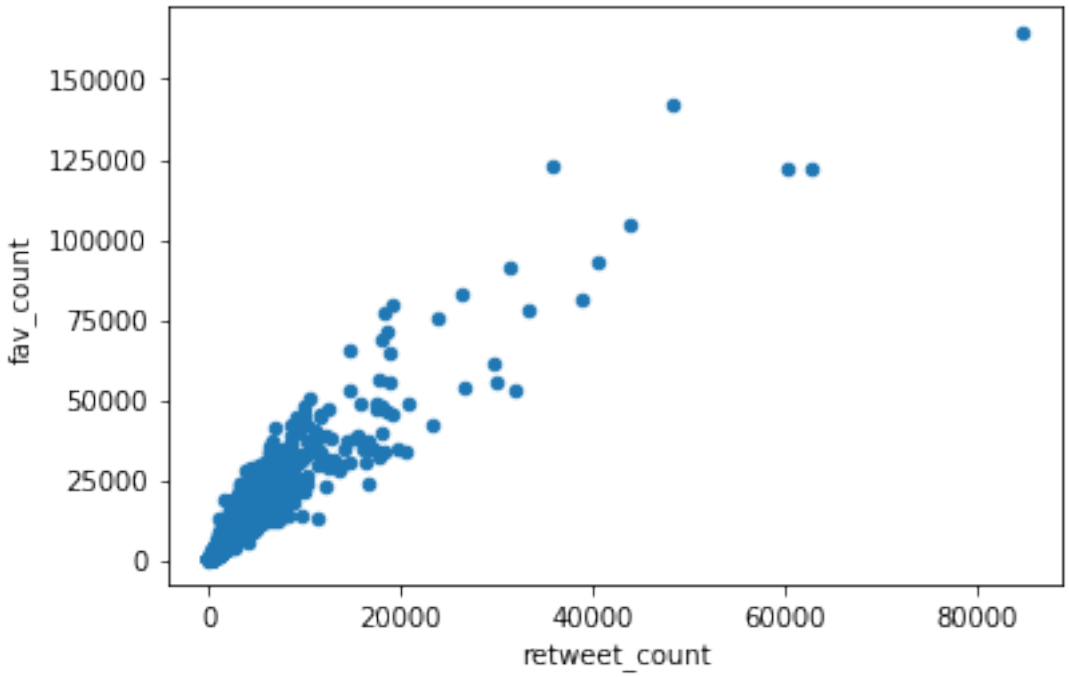
```
Out[37]: count    1992.000000
         mean       12.283133
         std        41.516835
         min         0.000000
         25%        10.000000
         50%        11.000000
         75%        12.000000
         max      1776.000000
         Name: rating_numerator, dtype: float64
```

There seem to be a direct relation between rating numerator and favorite counts. High ratings seem to have high favorite counts. Median ratings is 11 where as median to 3rd quartile show rating of 12. The maximum rating of 1776 is the highest rating given to any dog by We Rate Dogs. The plot show rating numerator under 20 since that is where majority of the ratings fall.

### 1.2 Retweet count and favorite count

```
In [45]: df.plot('retweet_count','fav_count',kind='scatter')
```

```
Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2fc3257860>
```

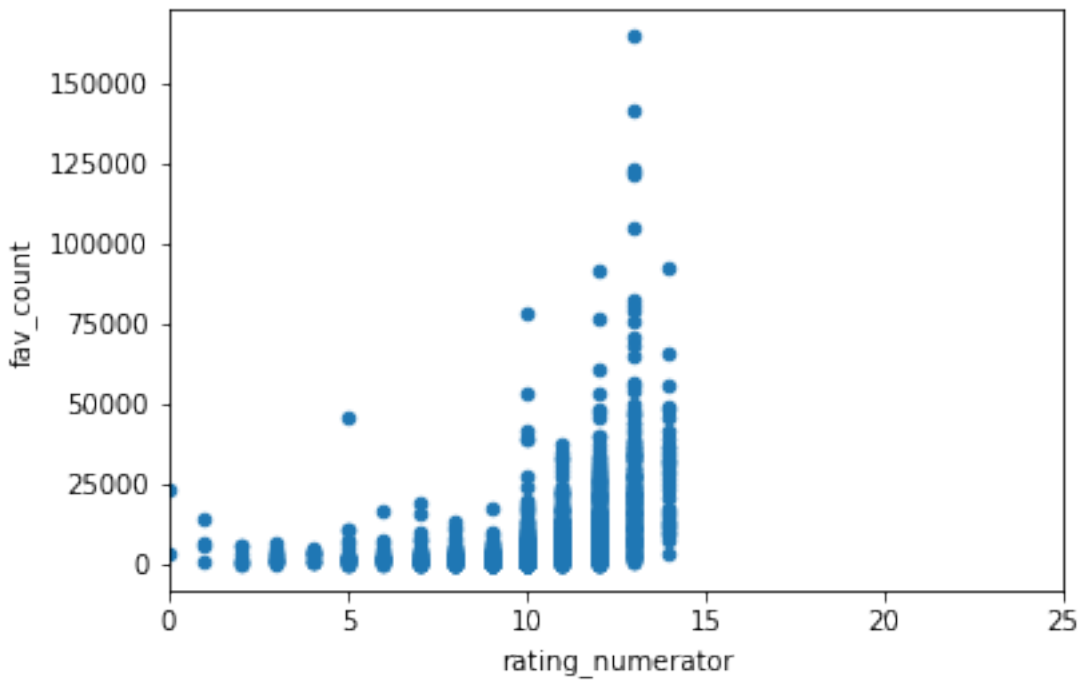


```
In [34]: df.retweet_count.describe()

Out[34]: count     1992.000000
         mean      2689.287149
         std       4768.275811
         min         12.000000
         25%        598.000000
         50%       1293.000000
         75%       3085.250000
         max      84566.000000
         Name: retweet_count, dtype: float64
```

Tweets that have high rating counts have high favorites as well. Median retweet count for tweets above 1000 which shows these ratings are highly followed by the twitter users. The favorite count for the ratings is close to 4000 per tweet.

### 1.3 Rating numerator and favorite count

```
In [49]: df.plot('rating_numerator','fav_count',kind='scatter')
         plt.xlim(0,25)

Out[49]: (0, 25)
```



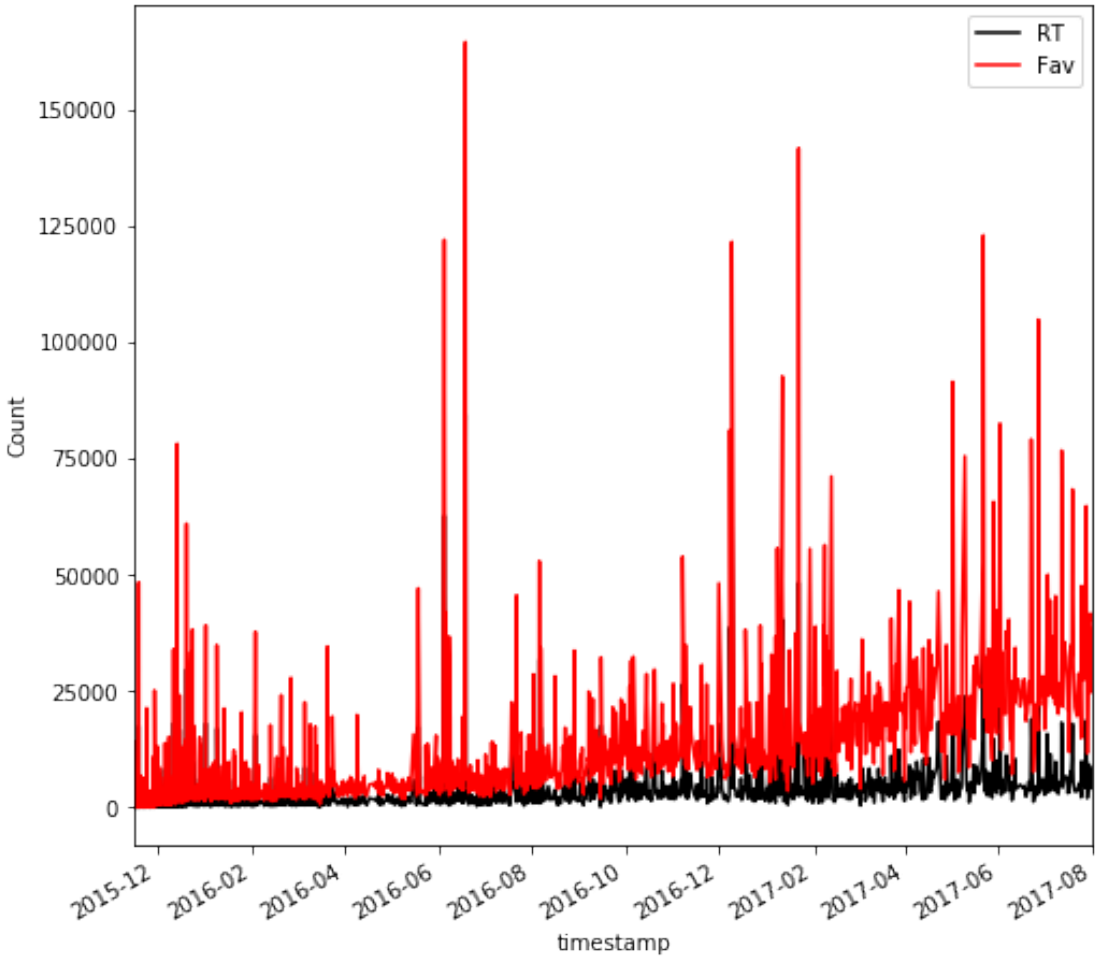
Favorite count also has a direct relation with the rating numerator. Tweets that received ratings above 10 show more favorites and that might be the reason of the fame of we rate dogs as more people like and agree with them based on the ratings they give to dogs.

### 1.4 Retweet and favorite count with respect to the timestamp of tweet

```
In [ ]: df.set_index('timestamp', inplace=True) ## Converting timestamp to index so that it is s
```

```
In [74]: df['retweet_count'].plot(color='black',label='RT',figsize=(8,8))
         df['fav_count'].plot(color='red',label='Fav')
         plt.legend(loc='upper right')
         plt.ylabel('Count')
```

```
Out[74]: Text(0,0.5,'Count')
```




The plot show retweet count and favorite count as a function of timestamp. This shows that for all tweets, the number of favorite count is always greater than the retweet count. There is a spike in favorite count in June 2016 and February 2017. Further analysis would be required to investigate these spikes.