act_report

May 28, 2018

```
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt

        %matplotlib inline
```

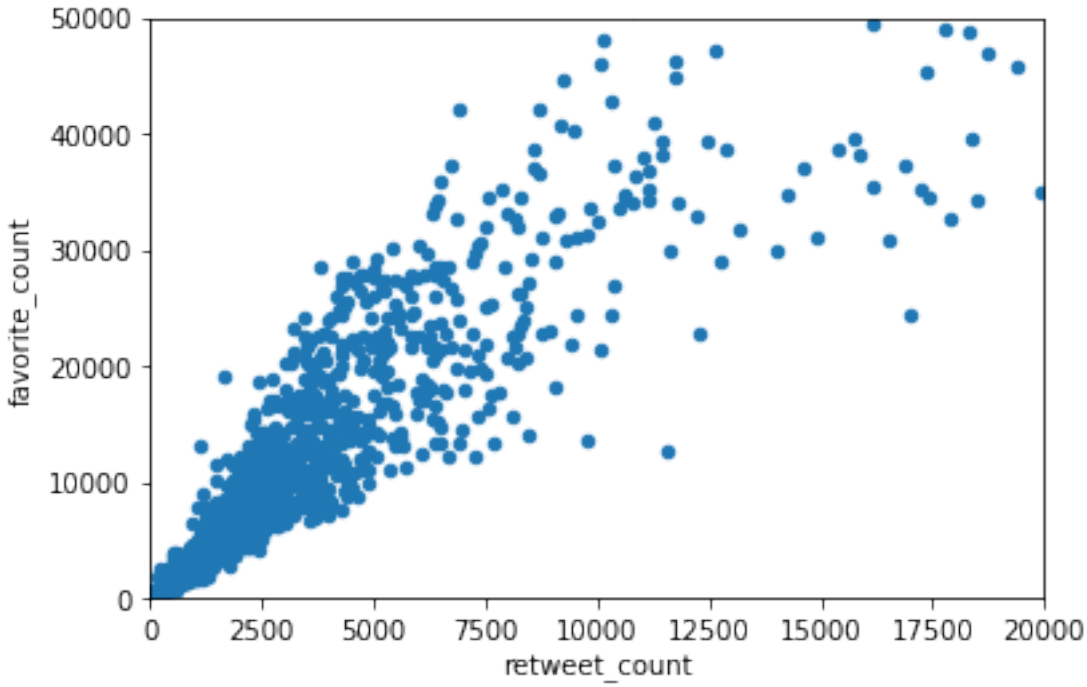
```
In [8]: df = pd.read_csv('twitter_archive_master.csv')
```

0.1 Visualizations

0.1.1 Retweet Count vs. Favorite Count

```
In [5]: df.plot.scatter(x='retweet_count', y='favorite_count', xlim=[0, 20000], ylim=[0, 50000])
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd68cd619b0>
```

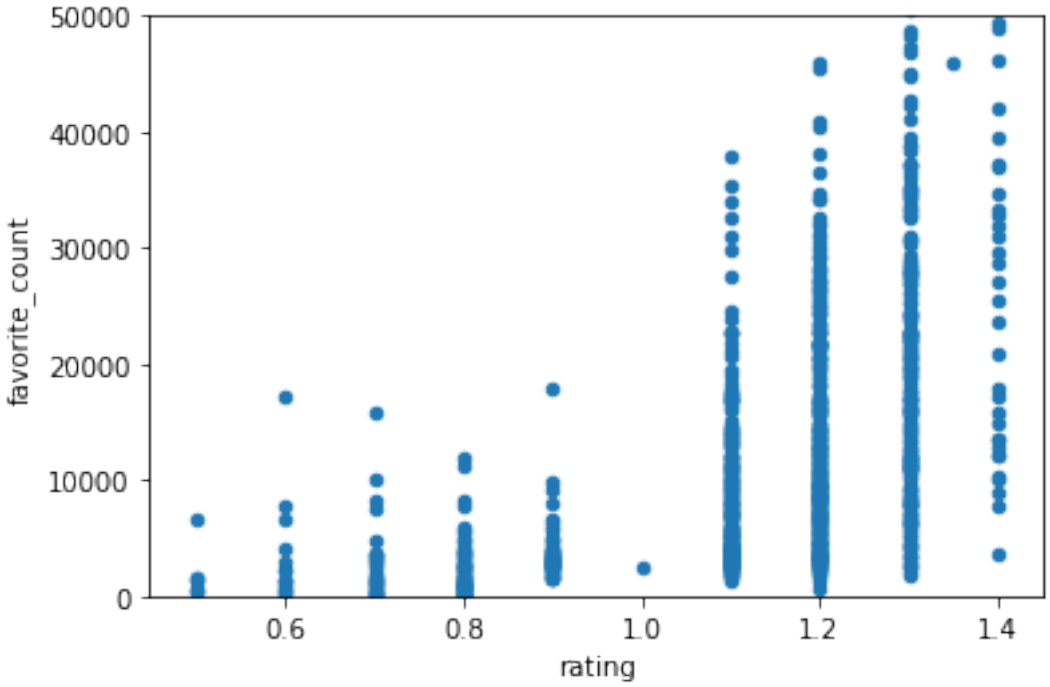


The first relationship I wanted to take a look at was the one between the favorite count and the retweet count. My expectation was that there would be a strong positive correlation between the two variables, meaning as one went up, the other would go up as well. In order to confirm my suspicions, I created a scatter plot of the two variables. Originally, I couldn't really tell much from the graph, as the range on the x and y axes were too wide. In order to fix this, I used xlim and ylim, making it so the scatter plot was now focused enough to see relationships. By looking at the scatter plot, (shown above), we can see that as the retweet count increases, so does the favorite count, confirming my suspicion.

0.1.2 Retweet Count and Favorite Count vs. Rating

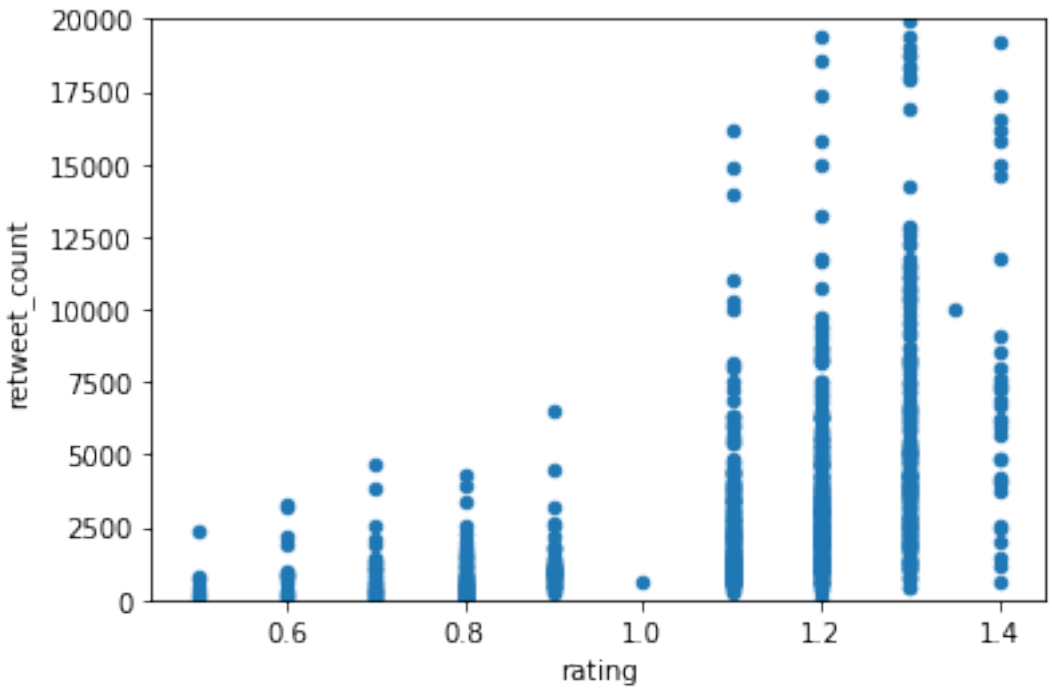
```
In [7]: df.plot.scatter(x='rating', y='favorite_count', ylim=[0, 50000])
```

```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd68d5b7c50>
```




```
In [9]: df.plot.scatter(x='rating', y='retweet_count', ylim=[0, 20000])
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fd68cd6d6d8>
```



I also took a look at the relationships between favorite count and rating and the one between retweet count and rating. My suspicion was that those tweets with higher ratings would have higher retweet counts, but I wasn't sure. Once again, a scatter plot proved to be the most effective in viewing these relationships to confirm my suspicions. The relationships between favorite count, retweet count, and rating were different from the one between favorite count and retweet count as there was an increase in the number of tweets along with the increase of rating. While the graphs do seem to show a positive correlation between favorite count, retweet count, and rating, these relationships are not nearly as obvious as the one between favorite and retweet count. The graph however did show me that the majority of dogs in the data received ratings of about 13/10.