

act_report

December 14, 2018

1 ACT REPORT - insights and displays the visualizations

1.0.1 Olive KMG

```
In [1]: from IPython.display import HTML
```

Wrangling WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations is the goal of this report. As I delved into this data some of the questions and thoughts were the following:

- What are the relations
- Was the twitter api - data as worth as the painful wait to get it?
- How good are the neural predictions?
- My hunch was that the prediction was made easier when the dog was a pure breed - so I set out to verify that.

Fig. 1 & 2 The handy pairplot shows the relations in the data. There is a strong positive correlation of 0.86 for favorite_count vs retweets. This is understandable - when its a favorite its shared more. This explains the increase in favorite count as retweet counts increase or viceverser. So the api pain was worth the wait in gold!

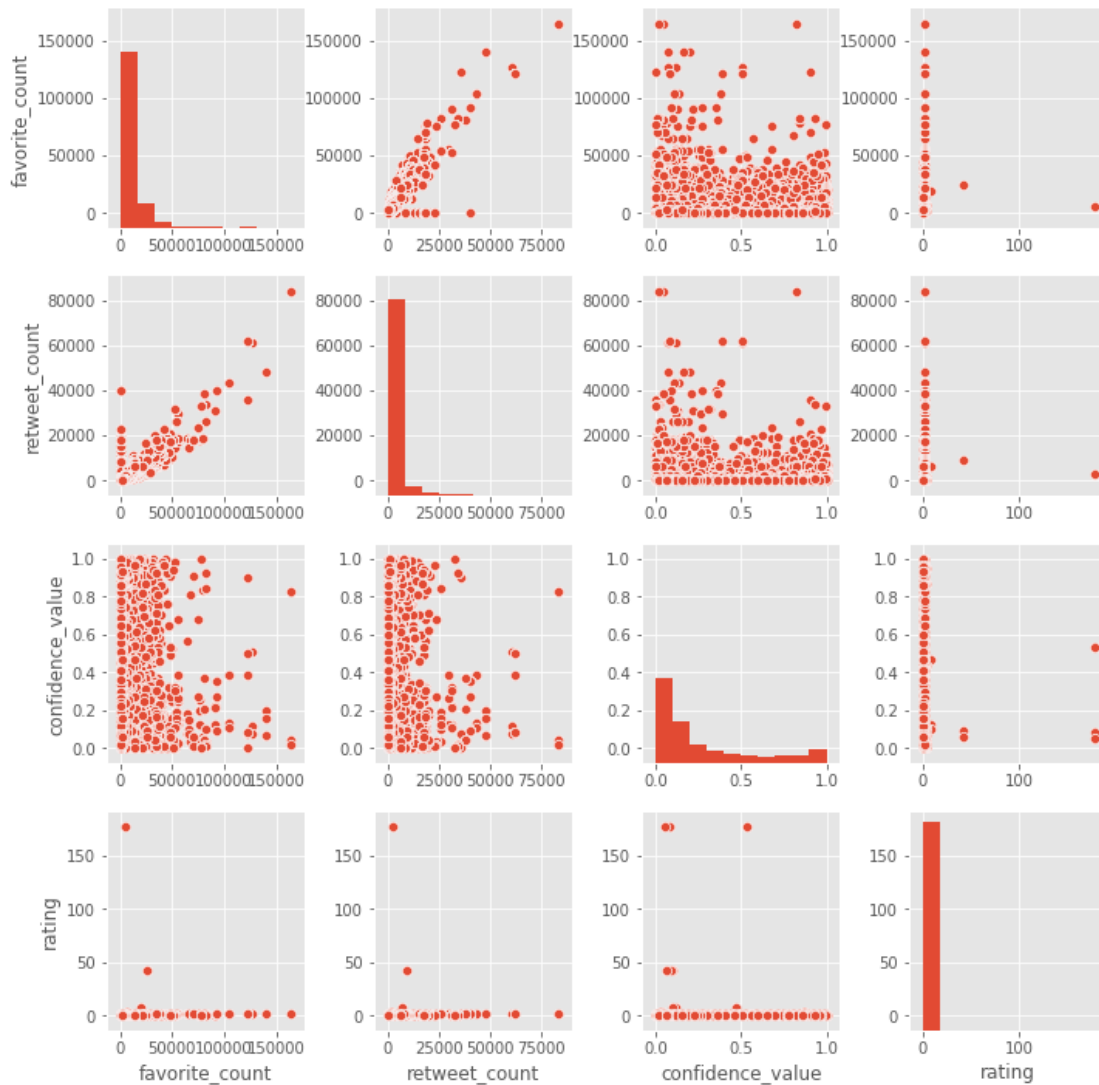
Fig.3 How good was the neural algorithm? Before this - because I love dogs, I was curious to know especially the types of dogs most predicted. The Labrador_retriever and golden_retriever rank highest. See the top 10 below.

Fig.4 How good is the neural algorithm - I wonder! Prediction_values do not have the actual dog names...you find values like zebra! A dog called zebra at P1 level with a P1_conf of 0.997 confidence level. Not sure if its the literal meaning or just for humor. Anyhow - does a pure breed influence the algos' prediction capability or not? - well from the last figure below, the two figures almost seem alike which suggests that this attribute does not in any way skew or influence the predictions.

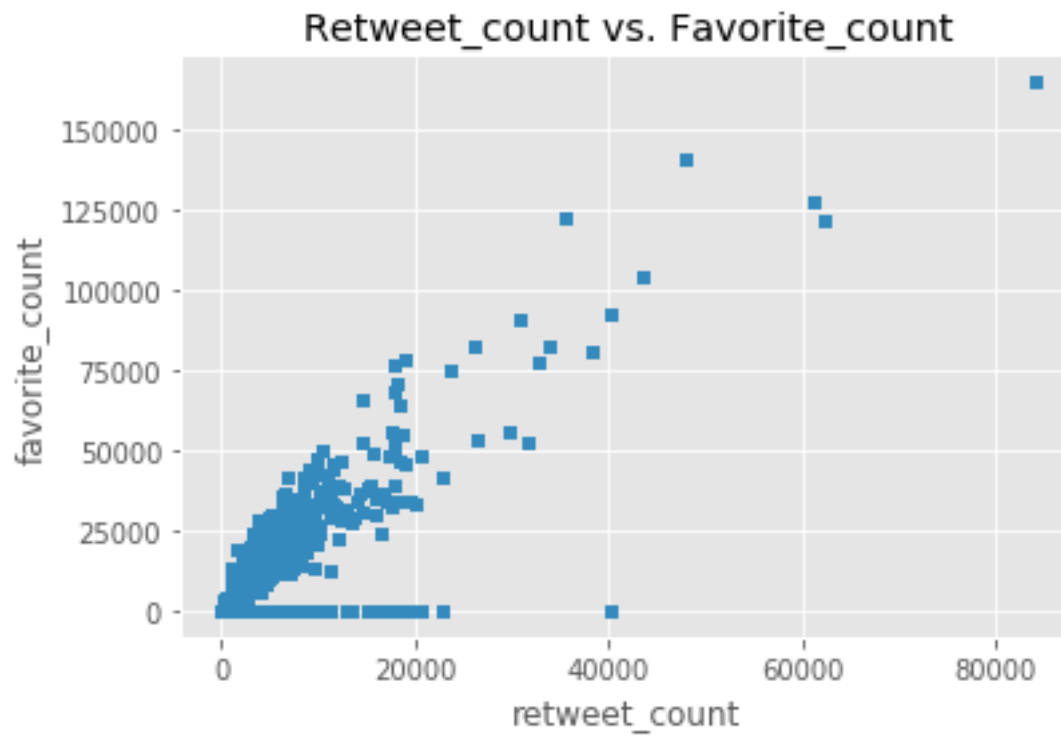
So there you have it. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. Yes its a place to humor you and so does the data!

```
In [22]: #Could there be some relation between favorite count, retweet and rating?
         #Create the default pairplot
         df_num = df_num.drop(['numerator', 'denominator', 'year'], axis = 1)
         sns.pairplot(df_num)
```

Out[22]: <seaborn.axisgrid.PairGrid at 0x7f903d442ba8>

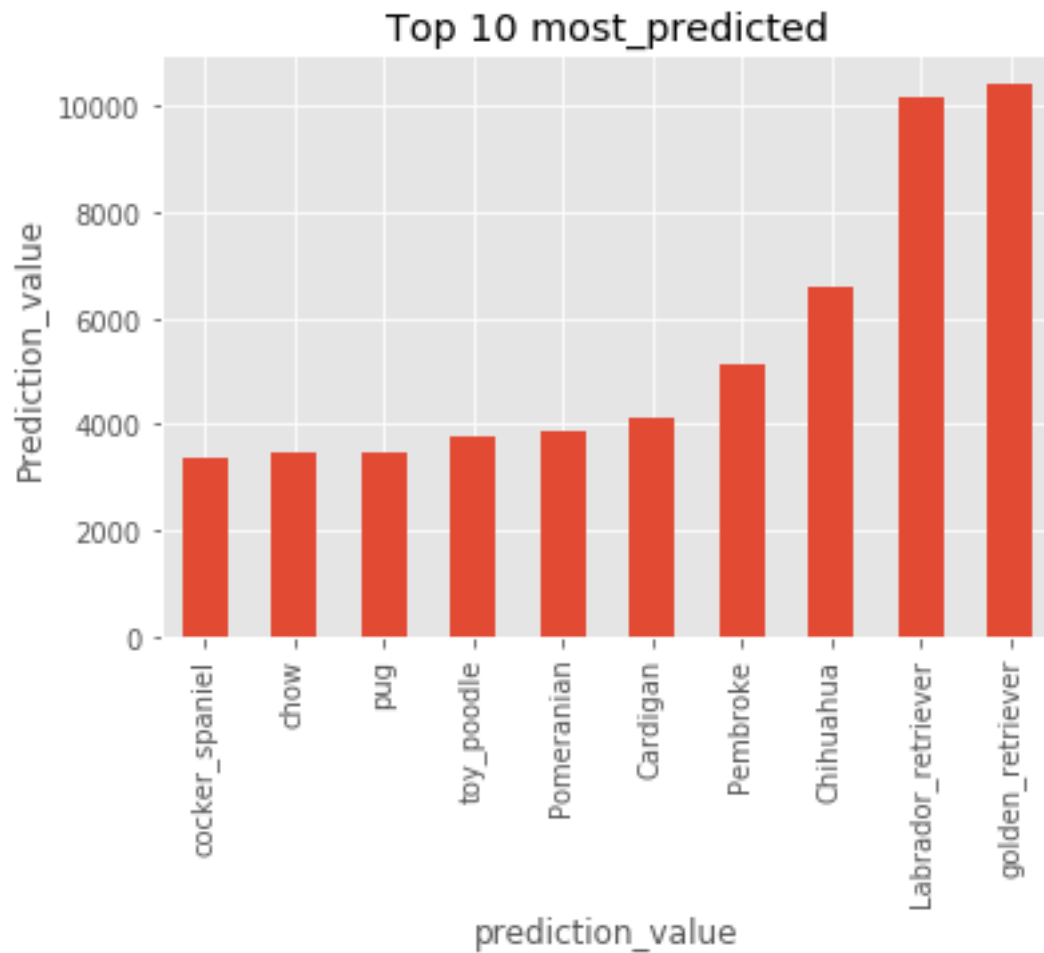


```
In [23]: #positive relation only in retweet and favourite pair - expected as shown above!
#correlation Plot - positive relation
df_master.plot(kind='scatter', x = 'retweet_count', y= 'favorite_count')
plt.title("Retweet_count vs. Favorite_count")
plt.show()
```



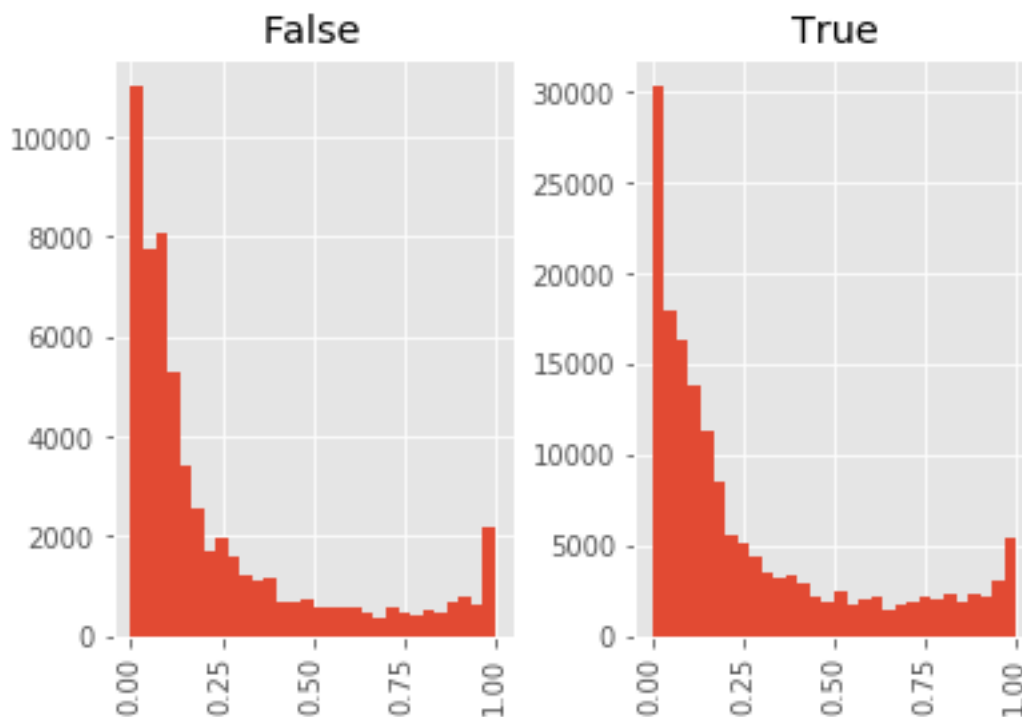
```
In [26]: #plot the 10 most predicted dogs
df_types.sort_values()[-10:].plot(kind='bar')
plt.title('Top 10 most_predicted')
plt.ylabel('Prediction_value')
```

```
Out[26]: Text(0,0.5,'Prediction_value')
```



```
In [28]: #Does being a breed influence the algorithms' prediction - does not seem so!  
df_master.hist(column="confidence_value",by="is_breed",bins=30)
```

```
Out[28]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x7f903fe507b8>,  
                <matplotlib.axes._subplots.AxesSubplot object at 0x7f903d2bcef0>], dtype=object)
```



```
In [2]: HTML('''<script>
code_show=true;
function code_toggle() {
  if (code_show){
    $('div.input').hide();
  } else {
    $('div.input').show();
  }
  code_show = !code_show
}
$( document ).ready(code_toggle);
</script>
The raw code for this IPython notebook is by default hidden for easier reading.
To toggle on/off the raw code, click <a href="javascript:code_toggle()">here</a>.''' )
```

```
Out[2]: <IPython.core.display.HTML object>
```

```
In [ ]:
```