act_report

December 18, 2018

1 ACT REPORT - Insights and visualizations

This project was about wrangling WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. of this report. AS I delved into this data I wondered as to whether the data had any serious relations given the surprising numerator vs denominator ratings. How good was the neural algorithm? Was the prediction made easier when the dog was a pure breed? So I ran a pairplot to get answers.

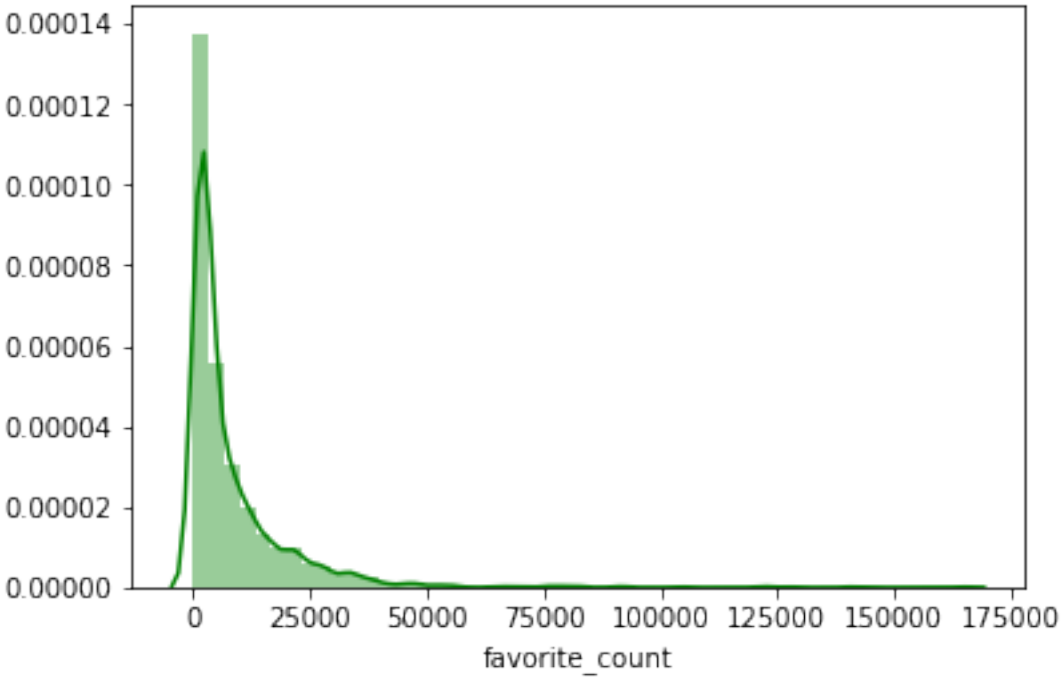
Fig. 1 The density distribution plots of favorite_count and the retweet_count both have quite high sharp peaks about zero and long right tails. Meaning they are occassional or rare occurances and so different from the norm.

Fig. 2 The only significant relationship according to the pairplot is that between the favorite_count and retweets. There is a strong positive correlation of 0.86. Its expected a good tweet will easily increase its count as its shared around; so - as its count increases the number its retweeted also increases.

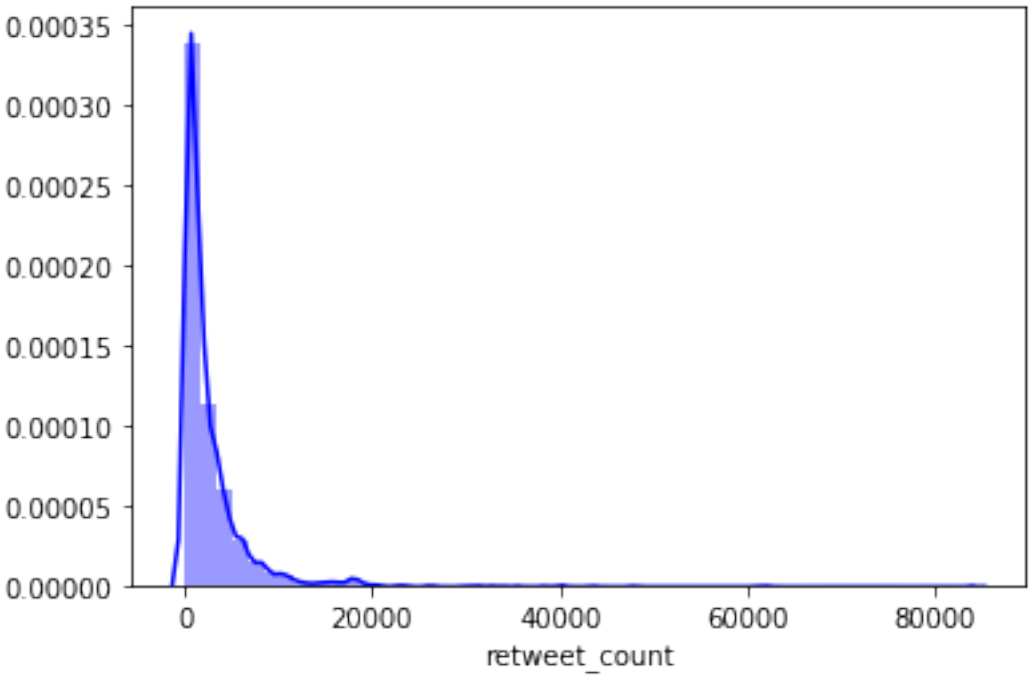
Fig. 3 How good was the neural algorithm? By the way - I love dogs. Therefore I was curious to know especially which types of dogs are most predicted and why. Fig 2 below ranks the Labrador_retriever and golden_retriever rank highes in the Top 10. Now I know why - Labrador Retriever's are America's most popular dog for the 27th. However - I think the major reason the labrador and malamute, Pomeranian, samoyed, chow, toy_poodle, pug, Chihuahua,Pembroke, Labrador_retriever top the list is simply because they are adorable creatures!

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. Yes its a place to get a laugh. findout

```
In [70]: #Fig. 1 - code for density plot of favourite count
         sns.distplot(df_final.favorite_count, color = 'green', label = 'Favorite_count')
         plt.show()
```

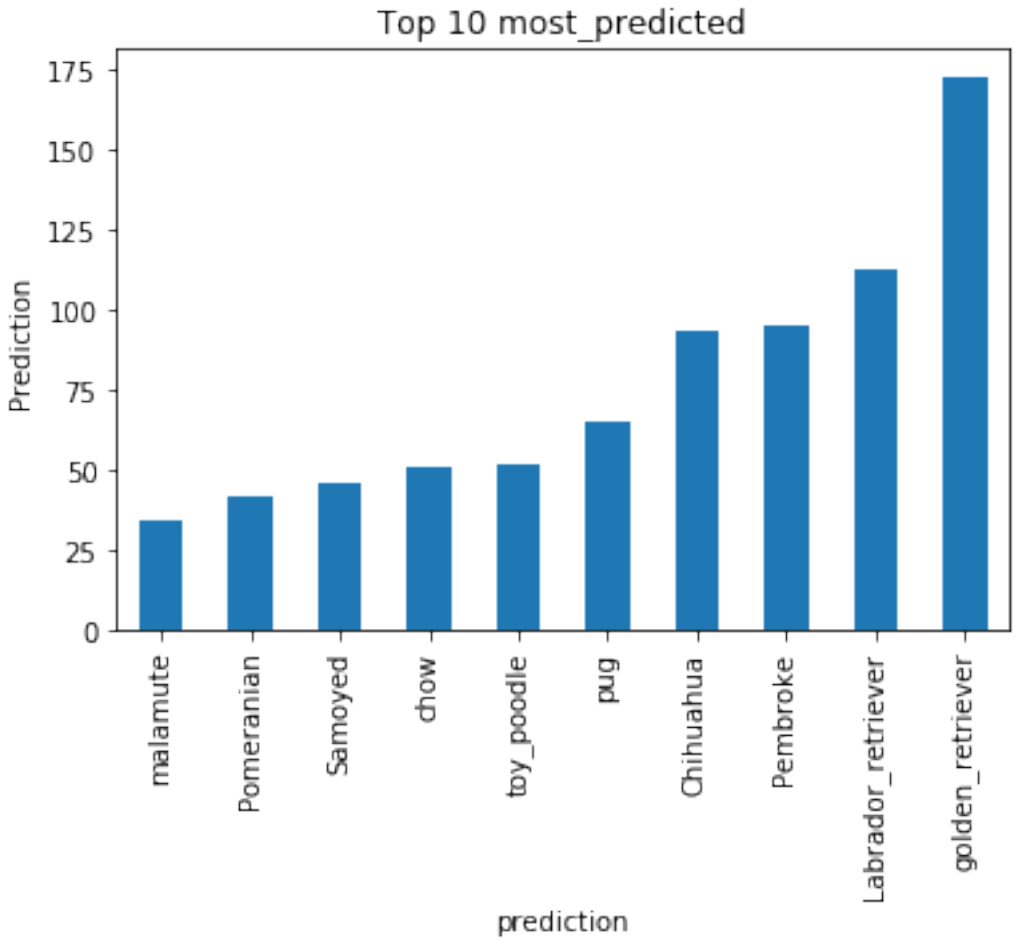


In [75]:
```
# code for density plot of favourite count
sns.distplot(df_final.retweet_count, color = 'blue', label = 'retweet_count')
plt.show()
```

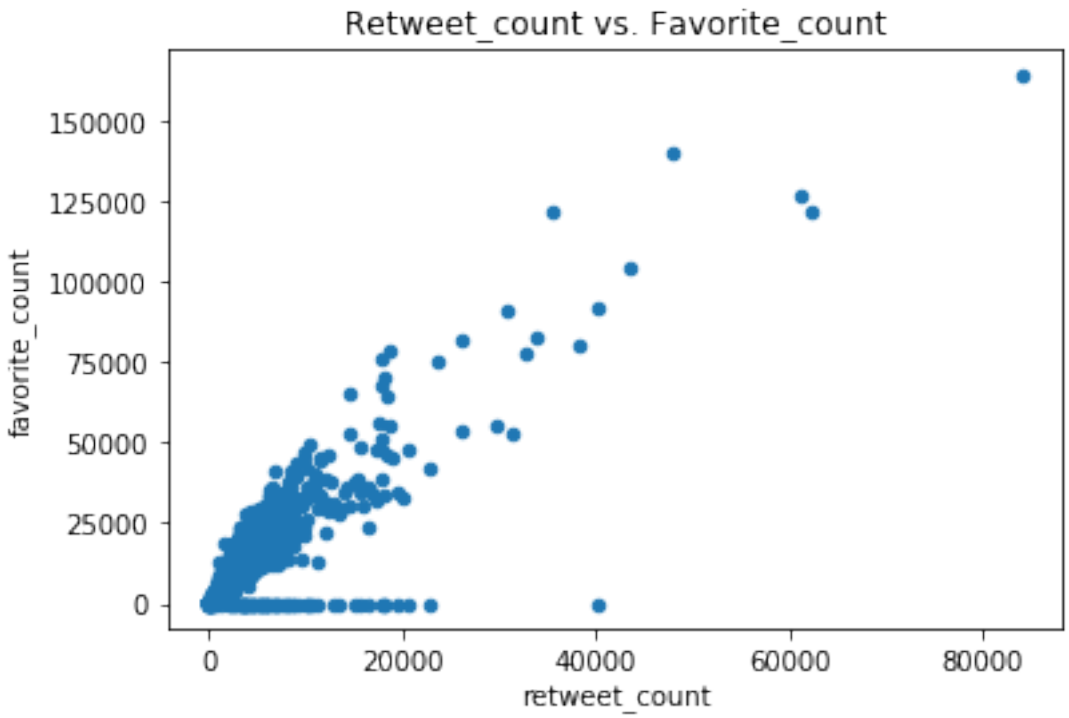


In [95]:
```
#plot the 10 most predicted dogs
df_types.sort_values()[-10:].plot(kind='bar')
plt.title('Top 10 most_predicted')
plt.ylabel('Prediction')
```

Out[95]: Text(0,0.5,'Prediction')



In [96]:
```
#postive relation only in retweet and favourite pair - expected as shown above!
#correlation Plot - positive relation
df_final.plot(kind='scatter', x = 'retweet_count', y= 'favorite_count')
plt.title("Retweet_count vs. Favorite_count")
plt.show()
```




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

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 [ ]:
```