Act_Report

By Maria George

The act report includes the basic dataanalysis of WeRateDogs twitter account data from the final dataset "twitter_archive_master" it provide four insights from the analysis and visualization results
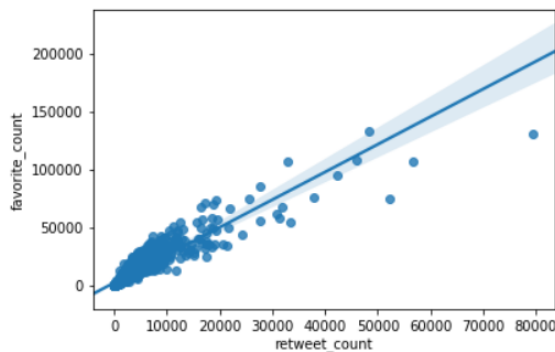
1- Retweet_count and favorite_count
I test the correlation between "retweet_count" and favorite_count . The pearson is 0.912 that is a very high value showing the strong positive relation between them and the plot below confirm the hypothsis

```
In [57]:    # Preason number is 0.912947 a strong relation between retweet_count and favorite_count

            sns.regplot(x="retweet_count", y="favorite_count", data=twitter_archive_master)

Out[57]:    <matplotlib.axes._subplots.AxesSubplot at 0x2372d989c88>
```


2- The classification of dog analysis result
These is the top 10 dog breed this model predicated , Golden retriever and Labrador retriever are top 10 , may be because the are most common in US
The plot show how confident the algorithm is in it's first predication as 100% is the most cases , the amount from 0.1 to 0.8 dominate the entire distribution that also suggest model is not good enough

Out[58]: golden_retriever   139
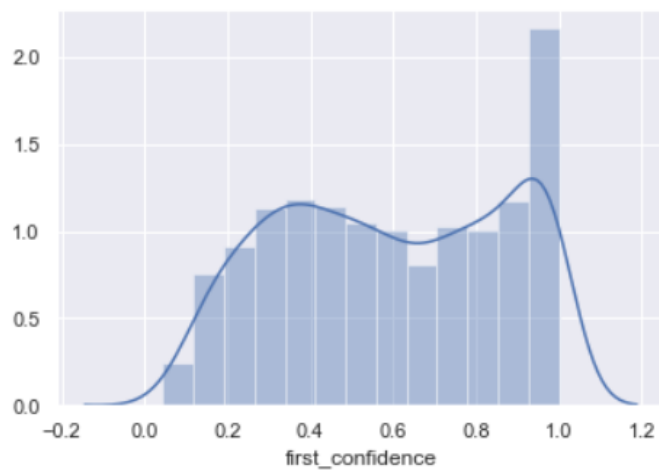   Labrador_retriever   95
         Pembroke   88
         Chihuahua   79
         pug   54
         chow   41
         Samoyed   40
         Pomeranian   38
         toy_poodle   38
         malamute   29
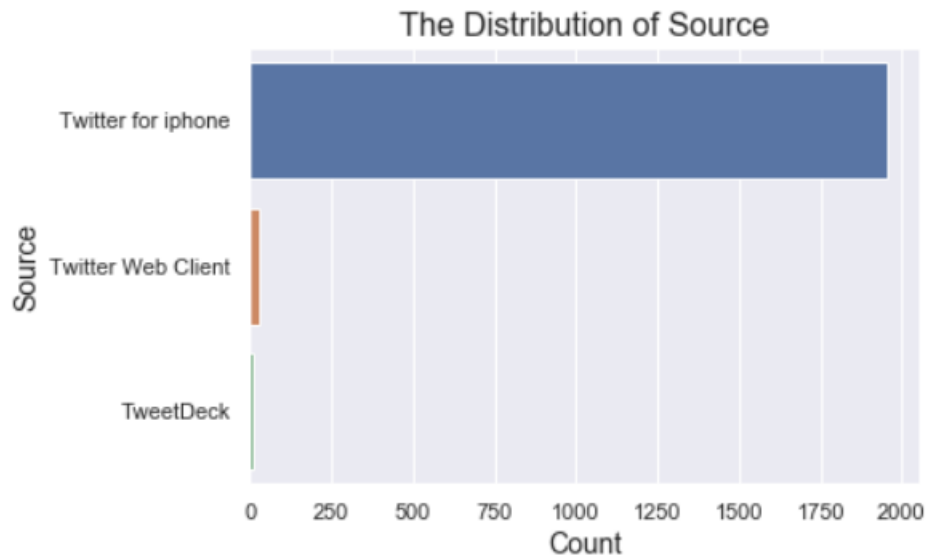   Name: first_prediction, dtype: int64

```
In [59]: sns.set(style='darkgrid')
         ax = sns.distplot(twitter_archive_master['first_confidence'])
```


3- The distribution of the source
   The plot show the distribution of the source , the dominante source of tweet is from iphone twitter app wich is 94% approx. , that mean the twitter app is the main channel for people using tweet , retweet , post and others

The Distribution of Source

4- Dog stage distribution with avg_retweet and avg_ favorite
Visually analysis show the avg_retweet and avg_ favorite for each dog_stage comparing the avg_retweet and avg_ favorite to each stage of dog we notice that there is a huge amount of missing data

```
t[61]:  <matplotlib.legend.Legend at 0x2372dd9ec88>
```