act_report

July 24, 2019

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sb

        %matplotlib inline
```

```
In [5]: master_clean_data = pd.read_csv('twitter_archive_master.csv', sep=',')
```

0.0.1 Analysis

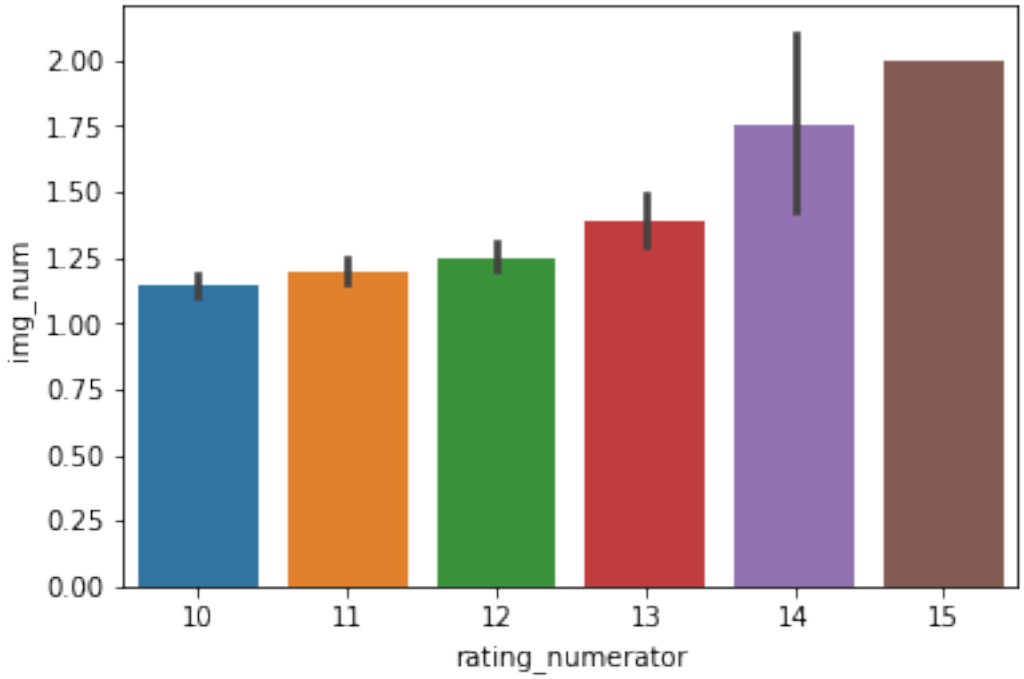
Now since we have cleaned and tidy data. So, now we can start with our analysis.

```
In [7]: #Getting an Idea About our data.
        master_clean_data.head(2)
```

```
Out[7]:              tweet_id  in_reply_to_status_id_x  in_reply_to_user_id_x  \
        0  892420643555336193                        0                      0
        1  892177421306343426                        0                      0

                                                   text  retweeted_status_id  \
        0  This is Phineas. He's a mystical boy. Only eve...                  0
        1  This is Tilly. She's just checking pup on you...                   0

           retweeted_status_user_id  rating_numerator  rating_denominator     name  \
        0                         0                13                  10  Phineas
        1                         0                13                  10    Tilly

          floofer   ...   img_num         p1   p1_conf p1_dog        p2   p2_conf  \
        0    None   ...         1     Orange  0.097049  False     Bagel  0.085851
        1    None   ...         1  Chihuahua  0.323581   True  Pekinese  0.090647

          p2_dog        p3   p3_conf  p3_dog
        0  False    Banana  0.076110   False
        1   True  Papillon  0.068957    True

        [2 rows x 42 columns]
```

Insight #1.

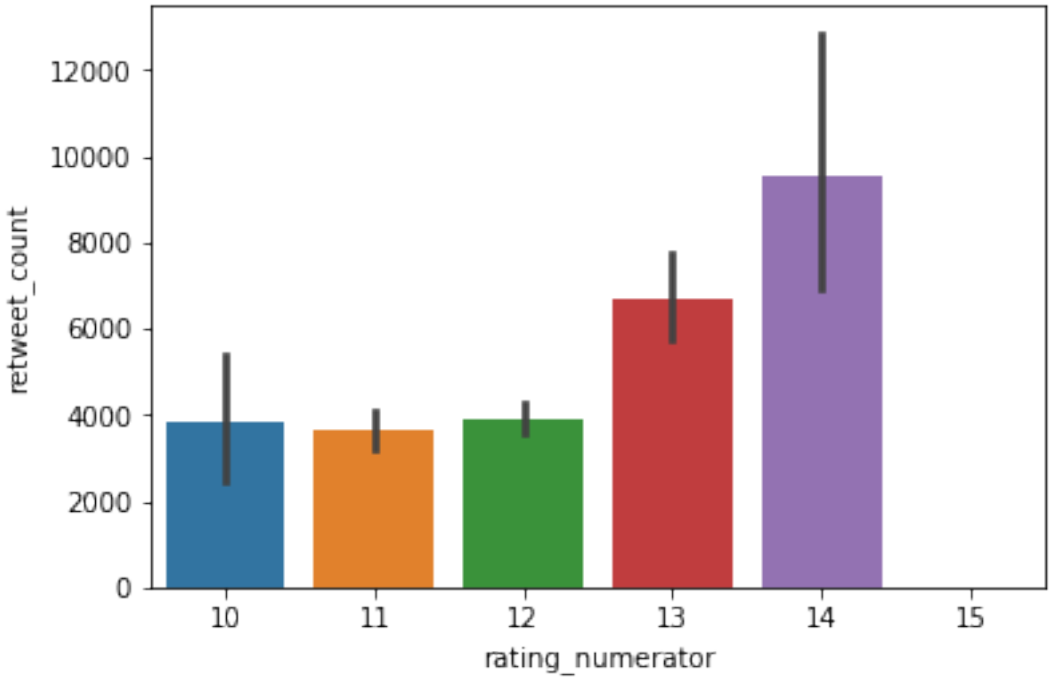
```
In [8]: # Insight #1.
        sb.barplot(data = master_clean_data, y = 'img_num', x = 'rating_numerator');
```



It is indicated that the more the number of images of Dogs can result into more rating numerator. This is just an identication and not confirmatory.

Insight #2.

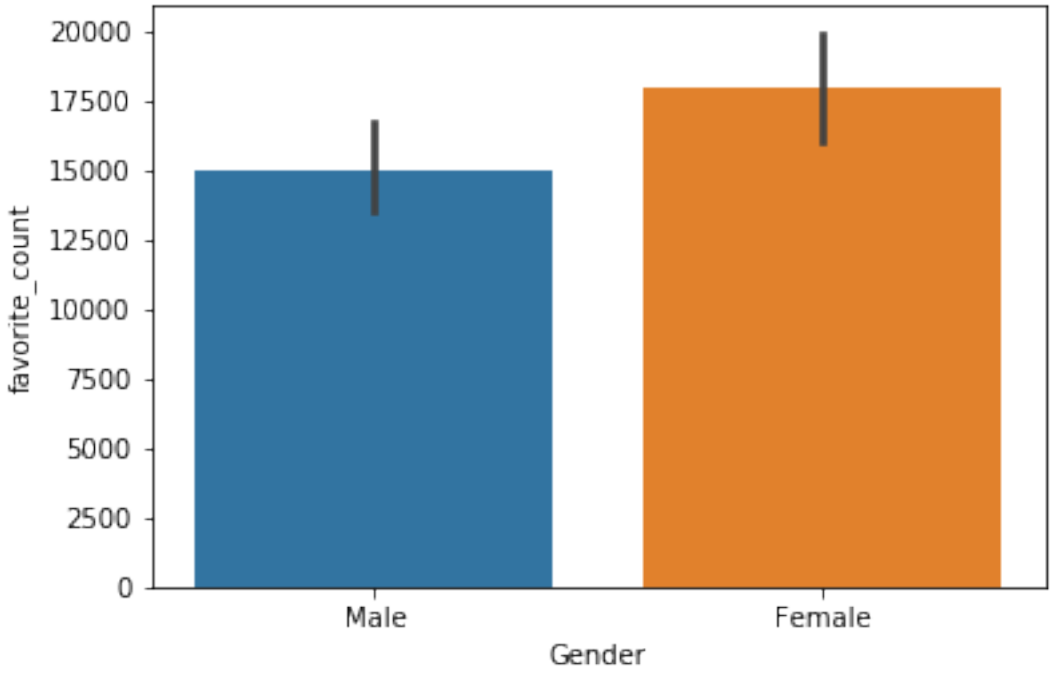
```
In [9]: # Insight No. 2
        sb.barplot(data = master_clean_data, y = 'retweet_count', x = 'rating_numerator');
```



It can be seen that with higher Rating Numerator the Retweet Count also increases with slight exception being at 10 Rating Numerator. Although this is just an insight.

Insight #3.

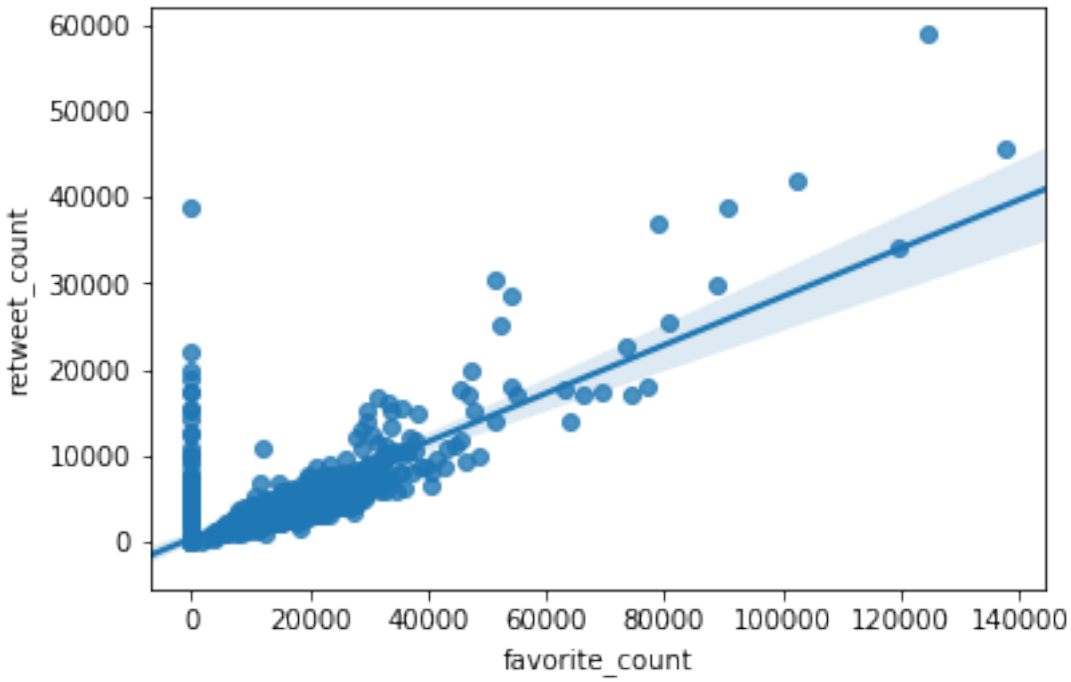
```
In [10]: #Insight No.3
         sb.barplot(data = master_clean_data, x = 'Gender', y = 'favorite_count');
```



It is likely that Female Dogs are having more Favorite Counts than Male Dogs.

Insight #4.

```
In [11]: #Insight No. 4
         sb.regplot(data = master_clean_data, x = 'favorite_count', y = 'retweet_count');
```

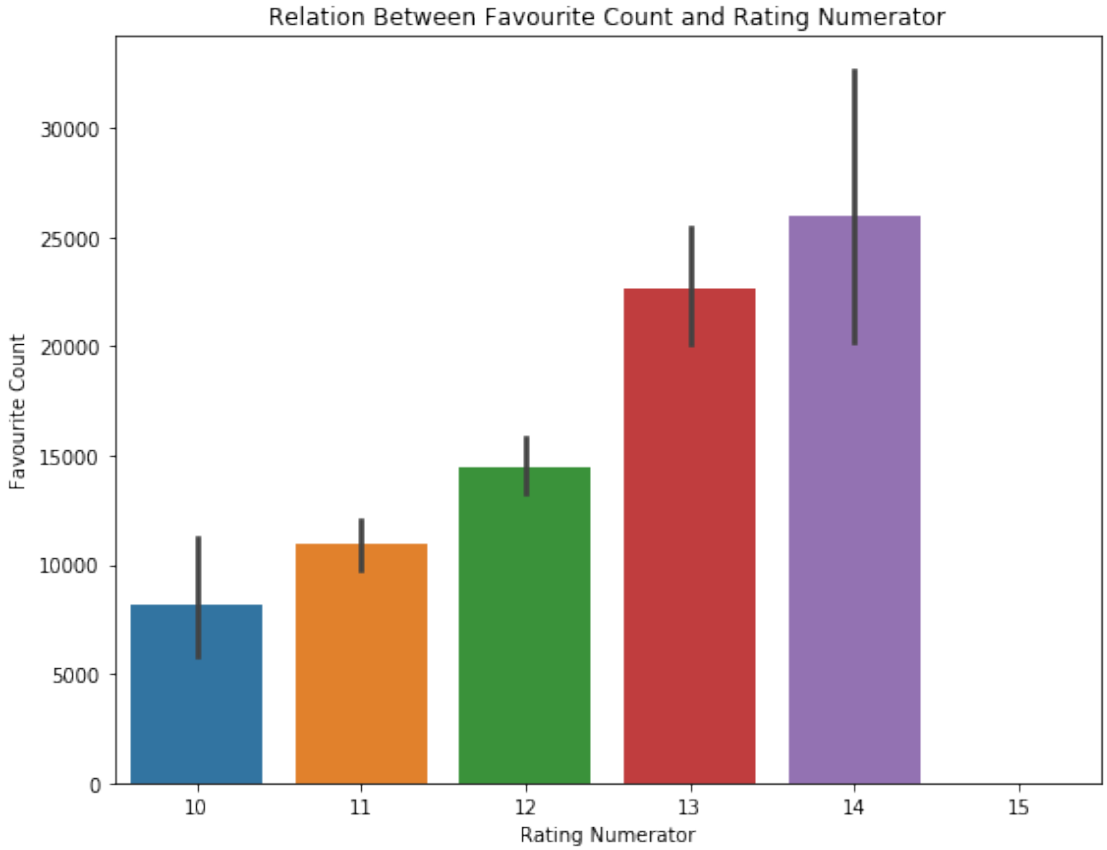


There exists a relationship between Retweet Count and Favorite Count. i.e. The more the retween counts, the more are the favorite counts. It jjust a preliminary insight and it needs further investigation.

0.0.2 Visualization

0.0.3 Q. Is there a relationship between Favorite Count and Rating Numerator ?

In [12]:
```
# Visualisation Presentation bivariate
plt.figure(figsize=(9,7))
sb.barplot(data = master_clean_data, y = 'favorite_count', x = 'rating_numerator')


plt.title("Relation Between Favourite Count and Rating Numerator")
plt.ylabel("Favourite Count")
plt.xlabel("Rating Numerator");
```



Yes there seems to be a relationship between Favorite Counts for each dog and its Rating Numerator. With Higher Rating Numerator, the Favorite Counts also increases. But, this is just an insight and needs further statistical investigation.

In [ ]: