# Documentation of data analysis, insights and visualization

After collecting data of WeRateDog Twitter account focusing on retweets and favorites count for each tweet and it's relation with the rest of data mostly relation between I created correlation assumptions

Assumption 1

- retweets count and favorite count

Assumption 2

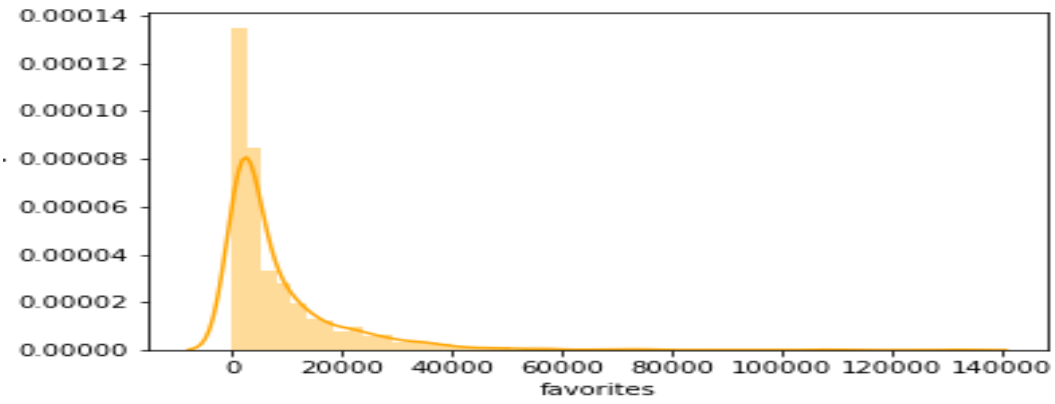
- img_num and retweet count
- img_num and favorites count

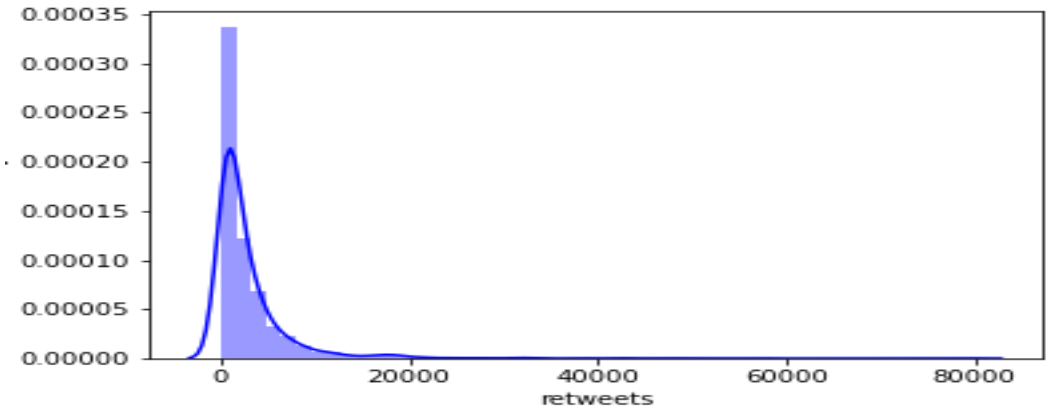
Assumption 3

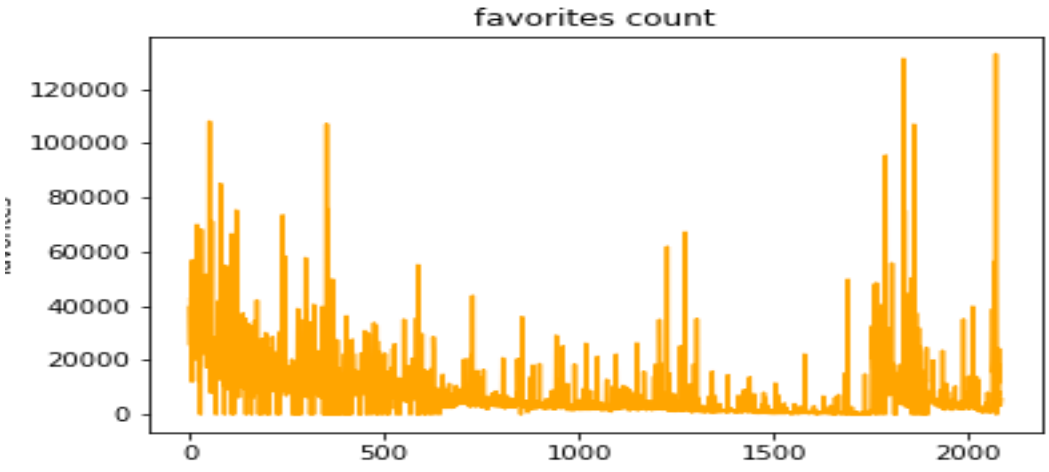
- rating and retweet counts
- rating and favorites count

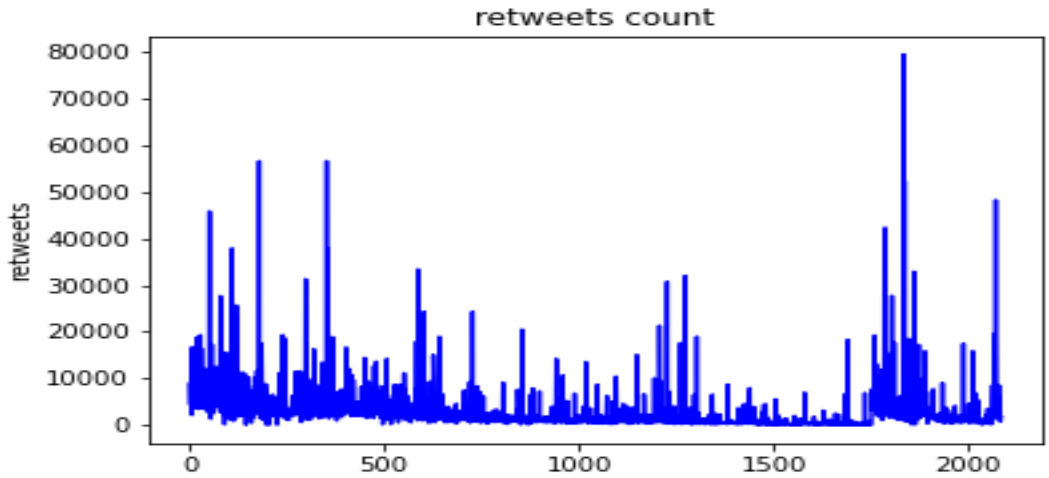
Calculated median of retweets =1423 , median of favorites =3897 which refers it's popular account have posts with many retweets and favorites.

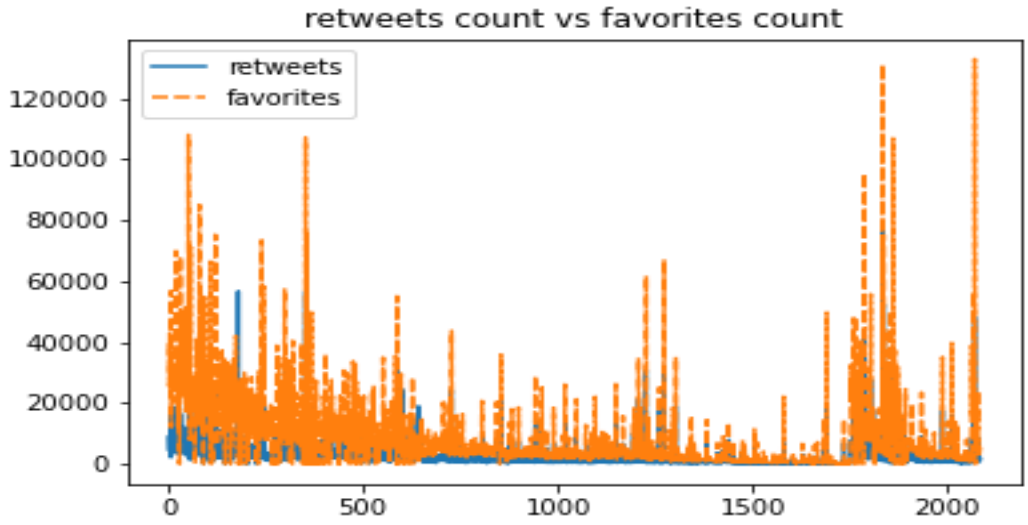
the below figures represent density of retweets counts , density of favorites counts , retweet data distribution , favorites data distribution separated and combined.




0.00035
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0
20000
40000
60000
80000
retweets


favorites count
120000
100000
80000
60000
40000
20000
0
0
500
1000
1500
2000

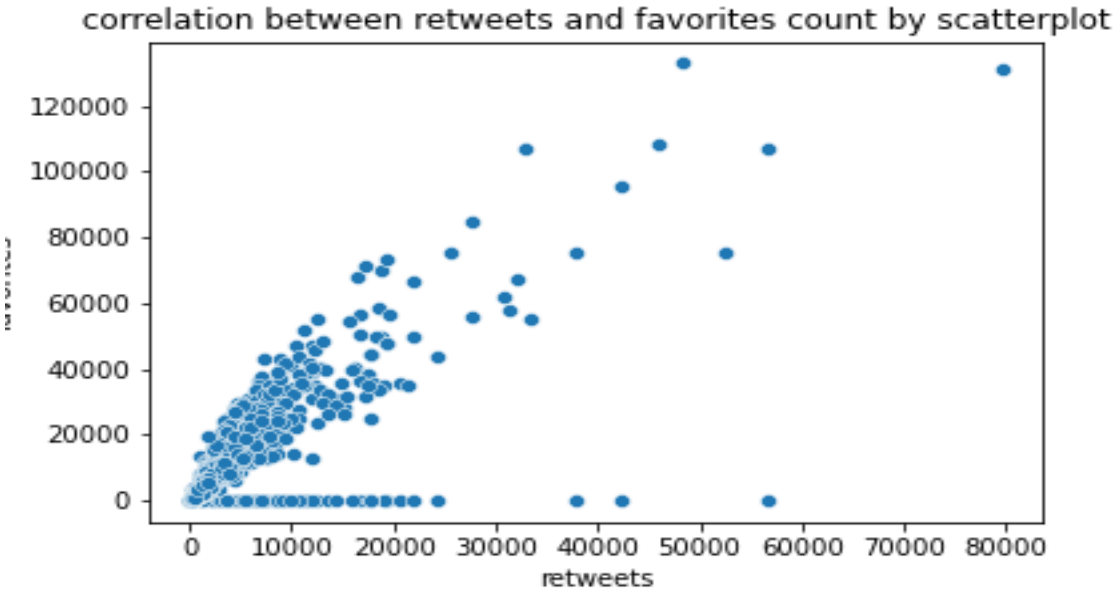

retweets count
retweets
80000
70000
60000
50000
40000
30000
20000
10000
0
0
500
1000
1500
2000

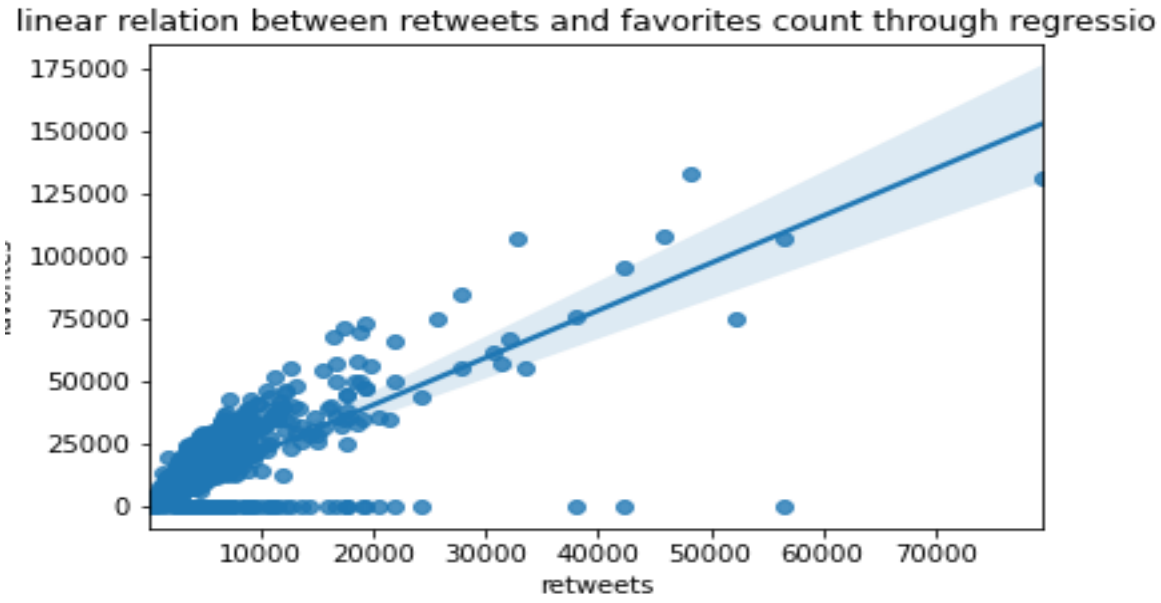



Assumption 1

Assumed relation between retweets counts and favorites count then tested by appling seaborn.corr() and seaborn.corrwith() to result value **R=0.78** which refers to **strong positive correlation** between retweet and favorites.

Correlation between retweets and favorites as shown in the below graphs using scatterplot ,and regplot seaborn functions.






Trying to find more  correlations between data and testing it like

1. img_num and retweet count
2. img_num and favorites count
3. rating and retweet counts
4. rating and favorites count

Assumption 2

Assumed the more the post have uploaded images expecting more retweets and favorites counts tested by appling seaborn.corr() and seaborn.corrwith() to result **very weak correlation** with both retweets and favorites **R=0.10, 0.13**

Assumption 3

Assumed the more the post have numenator value expecting more retweets and favorites counts tested to find **very weak correlation** with both retweets and favorites **R=0.30, 0.37**