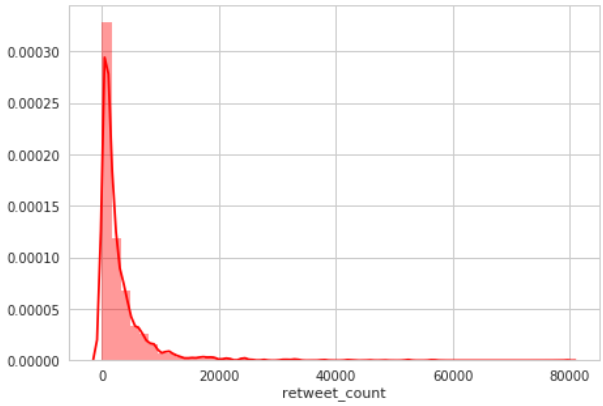
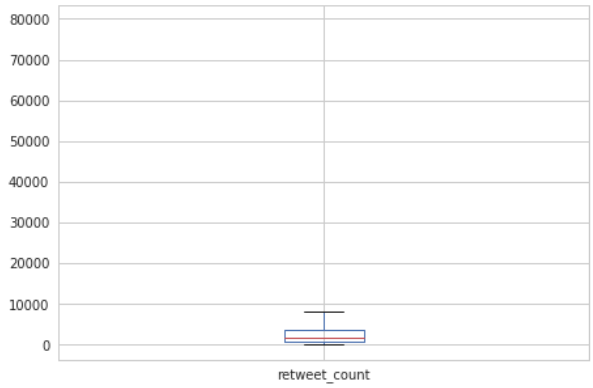
Twitter users activity on WeRateDogs Portal

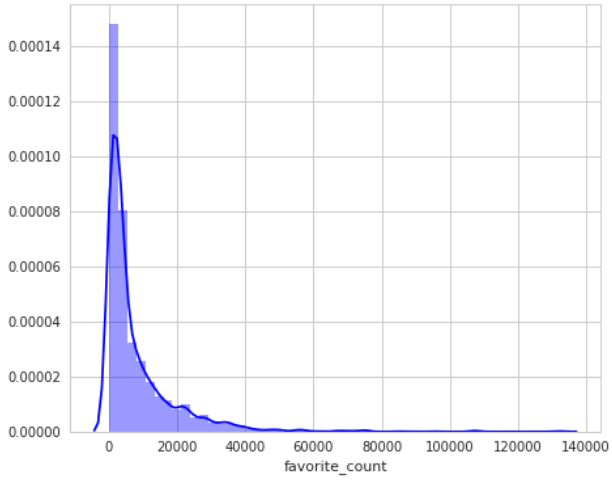
I started to visualize the most important variables (*retweet_count , favorite_count*)

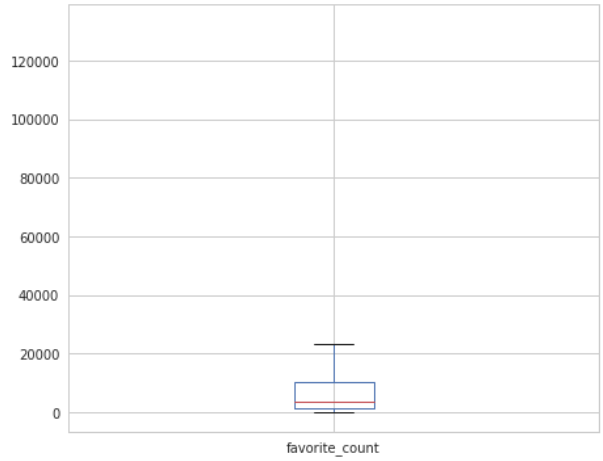
1. **Retweet counts**



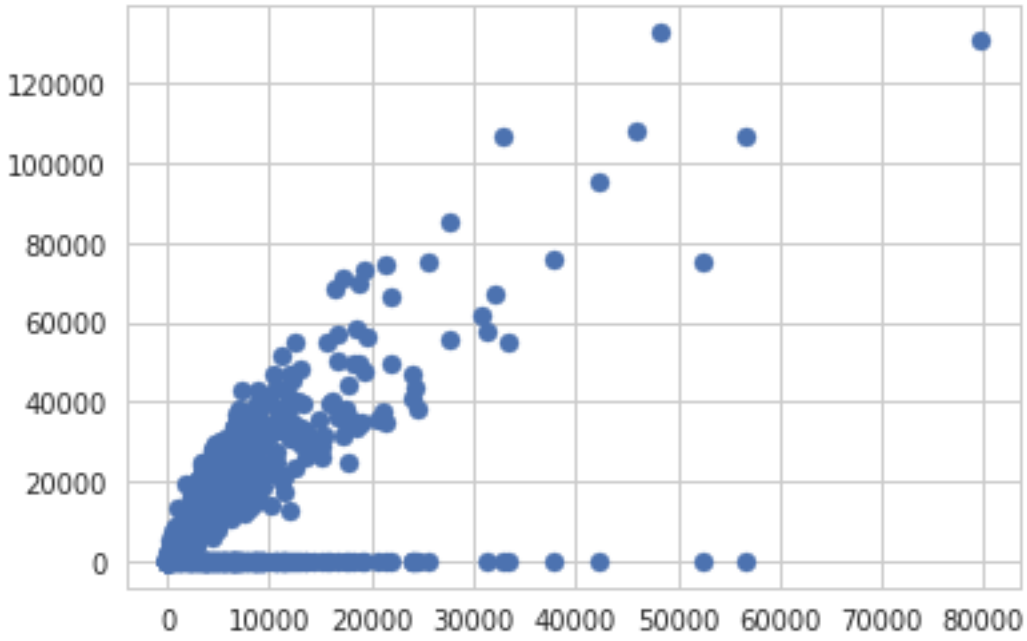


2. **Favorite counts**





3. **Correlation between the favorite and retweet counts**

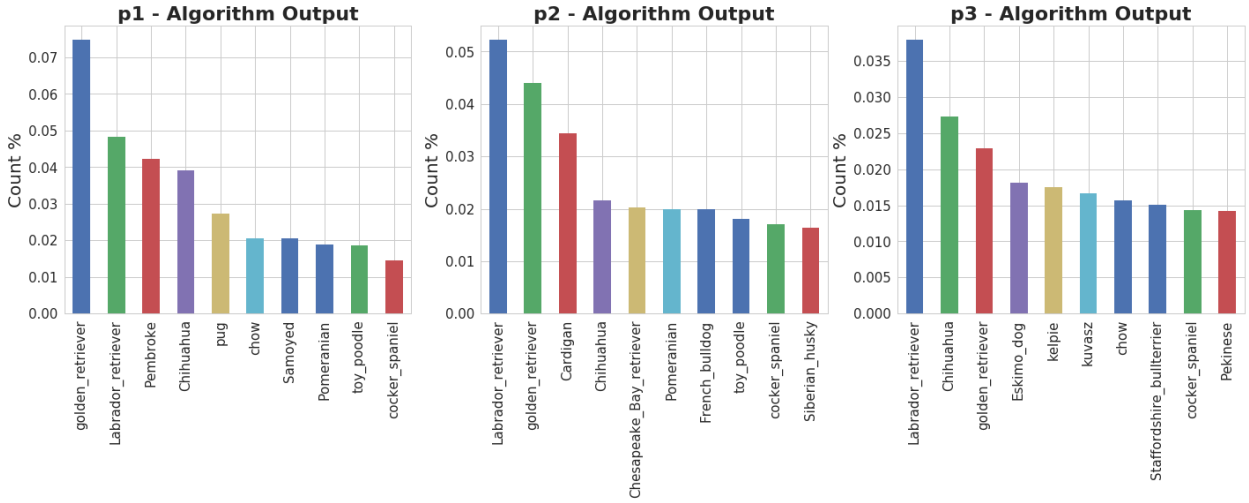


It seems like Retweets and Favorites are highly correlated. This shows that people who have retweeted have also done with favorites tag. Hence both the distributions shows similar behavior

To confirm this correlation, I checked for a scatter plot between Retweets and Favorites. It clearly shows that they both follow a similar trend and go in sync together.

There is a very strong relationship between retweet and favorite counts. As the user shows interest on Favorites, one can expect to see Retweets to increase and vice versa.

look at the image predictions output which shows the output produced from different algorithms



Majority of the predictions by 3 algorithms is done for golden_retriever and Labrador_retriever

Resources :

https://video.udacity-data.com/topher/2018/November/5be5fb4c_twitter-api/twitter-api.py

https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html

https://github.com/shravankoninti/Udacity_DataAnalyst/tree/master/Project-4%20-%20Data_Wrangling

https://www.geeksforgeeks.org/python-pandas-dataframe-drop_duplicates/

https://cmdlinetips.com/2018/03/how-to-change-column-names-and-row-indexes-in-pandas/#:~:text=One%20way%20to%20rename%20columns,list%20to%20column%20names%20directly.&text=This%20will%20assign%20the%20names,the%20data%20frame%20%E2%80%9Cgapminder%E2%80%9D.

https://pandas.pydata.org/docs/getting_started/intro_tutorials/03_subset_data.html