

# INSIGHTS AND VISUALIZATION OF TWITTER DATASET

By Opeyemi Ogundile

## Introduction

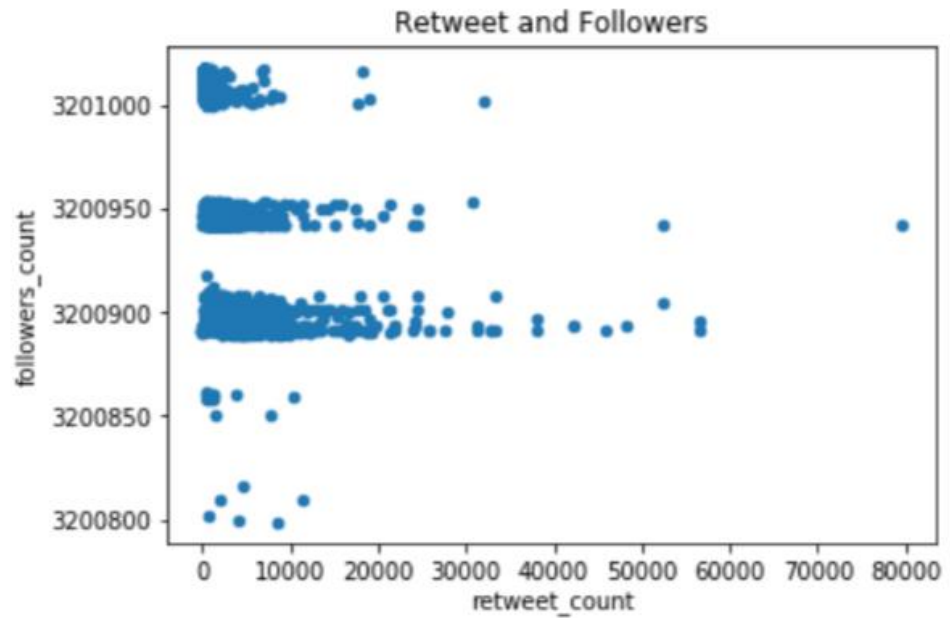
This project is based on the "WeRateDogs" Twitter page, a page which is used to rate pictures and videos of dogs. My task here is to go through the wrangling processes: by gathering, accessing, and cleaning data. Create Insights and visualizations. I will be gathering data from three different sources. It is also mandatory for me to assess the data manually and programmatically in order to identify some qualities and tidiness issues associated with the datasets.



I did a little research by visiting the account and discovered that they already have 9.3M followers and over 18.7K tweets. The user does weekly dog rating.

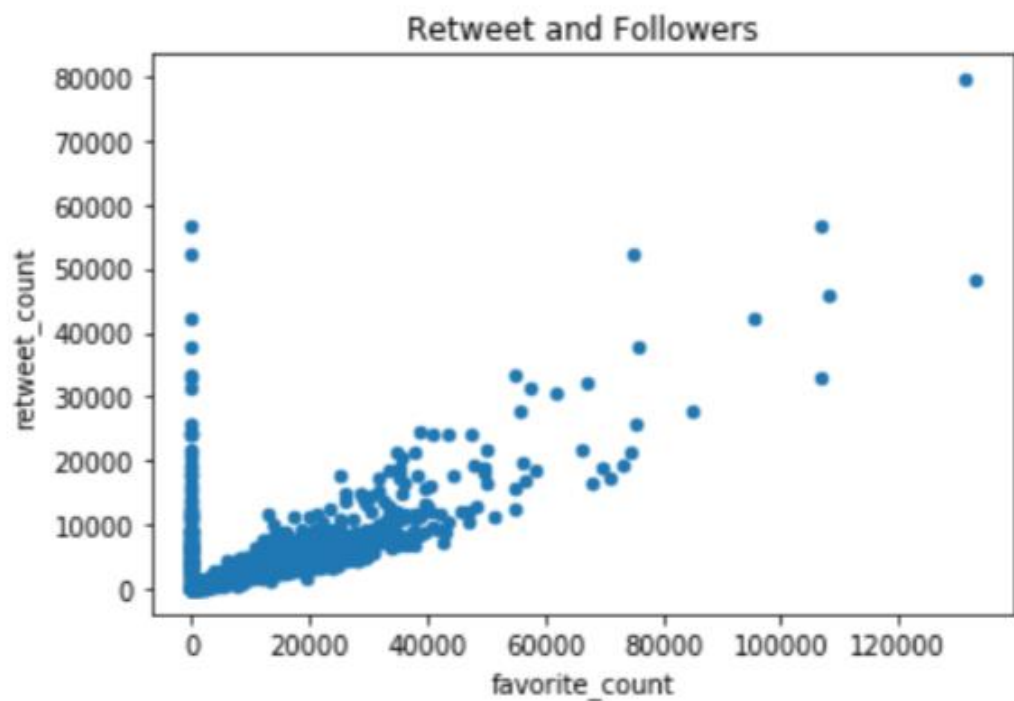
## Insights and visualization

1. Relationship between retweet and follower between retweet and followers



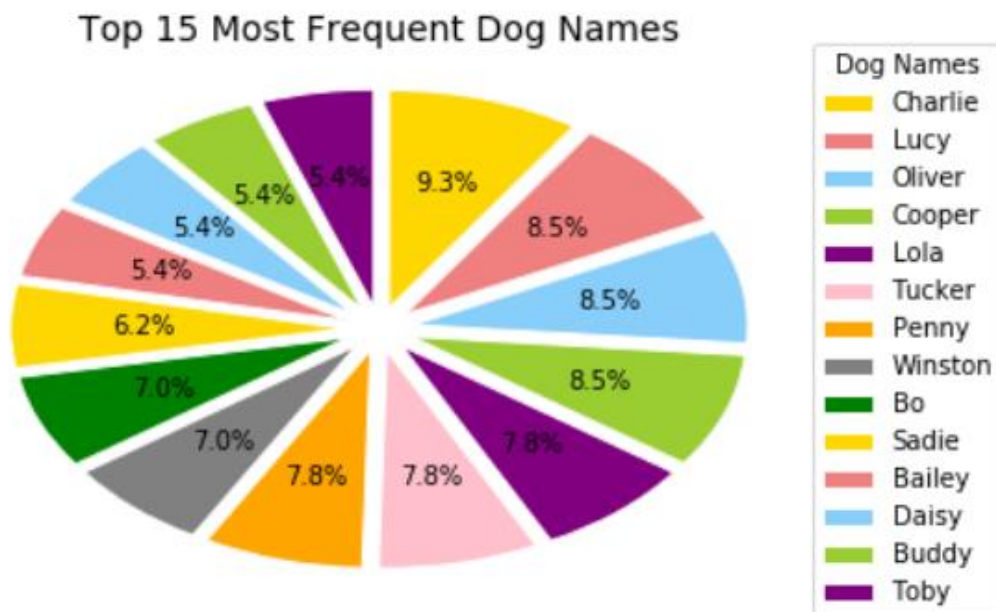
The result produced a correlation of  $-0.35787803731403972$  which indicates weak correlation between the two factors.

2. Relationship between favorite and retweet



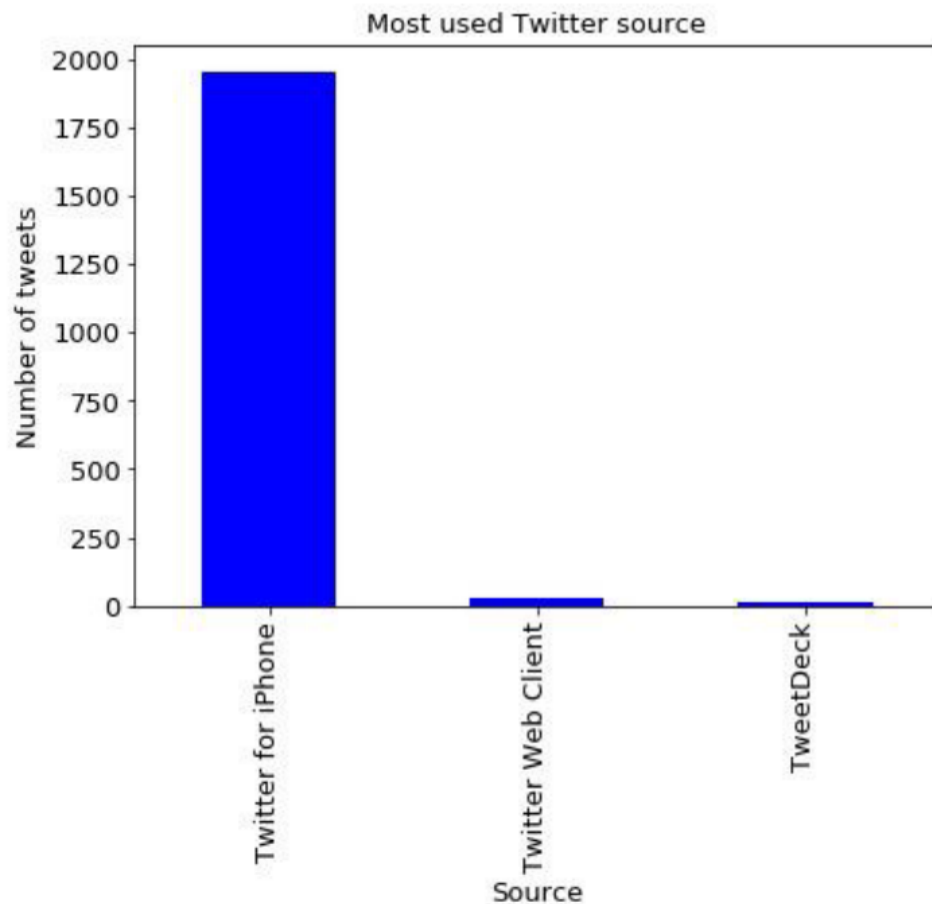
With correlation of 0.70288135961450338. The result suggests positive correlation between favourite and retweet. I.e as favorite tweet increases there is an increase in retweet as well.

3. What is the most frequent dog names?



From the pie chart, Charlie happens to be the top the list of Dog names with 9.1%, and Lucy, Oliver, and cooper as joint second with 8.5%.

#### 4. Most Used twitter sources



From the result obtained it is observed that Twitter for iPhone have the highest number as the most used platform for tweeting.

## Conclusion

In this project, I have gone through the process of gathering, assessing, and cleaning data which is called wrangling data. I have also done some analysis and visualisation of data.

#### References:

[https://twitter.com/dog\\_rates?ref\\_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor](https://twitter.com/dog_rates?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor)

<https://stackoverflow.com/questions/45035929/creating-new-pandas-dataframe-from-certain-columns-of-existing-dataframe>

<https://stackoverflow.com/questions/25351968/how-to-display-full-non-truncated-dataframe-information-in-html-when-converting>