

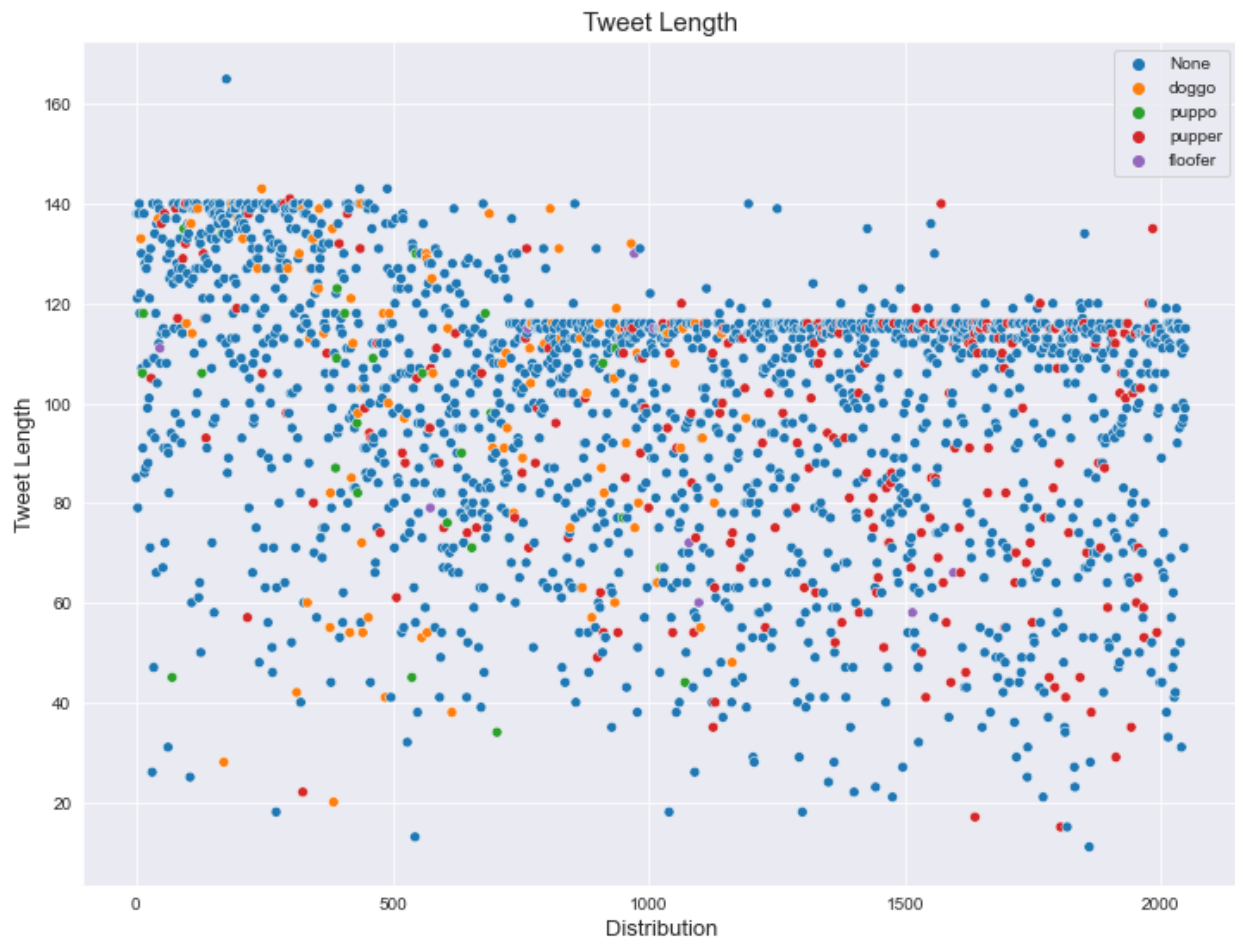
Insights and Visualizations

Introduction

This report discusses the insights derived from a user's [tweet](#) data and visuals explored on these insights. The dataset is being wrangled to solve some posed questions. These visualizations are carried out using special libraries like Matplotlib and Seaborn.

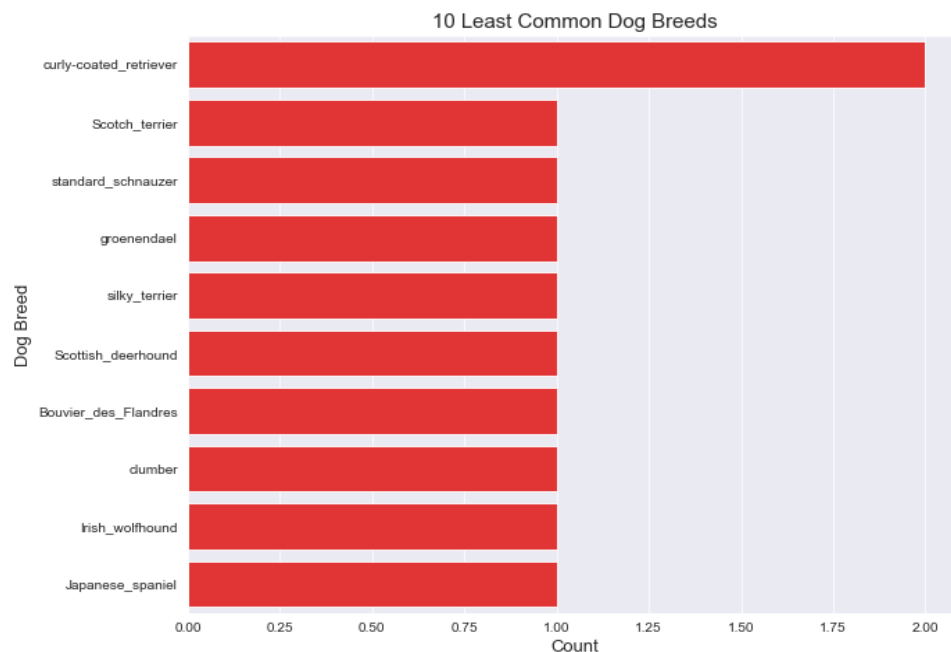
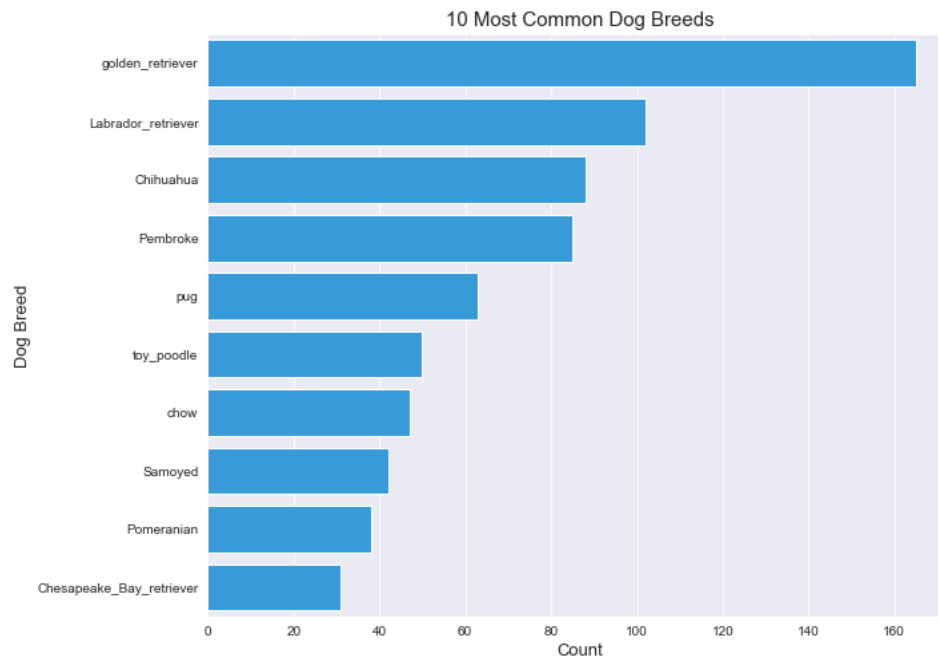
The average tweet length

The average length is the addition of all the characters typed in each [tweet](#) divided by the number of [tweets](#) in the wrangled dataset. After this exploration, the average length is 96.8; this states the fact that in most cases, the length of each [tweet](#) might not all be 96.8 but revolve around this value. The visualization below gives a picture about that.



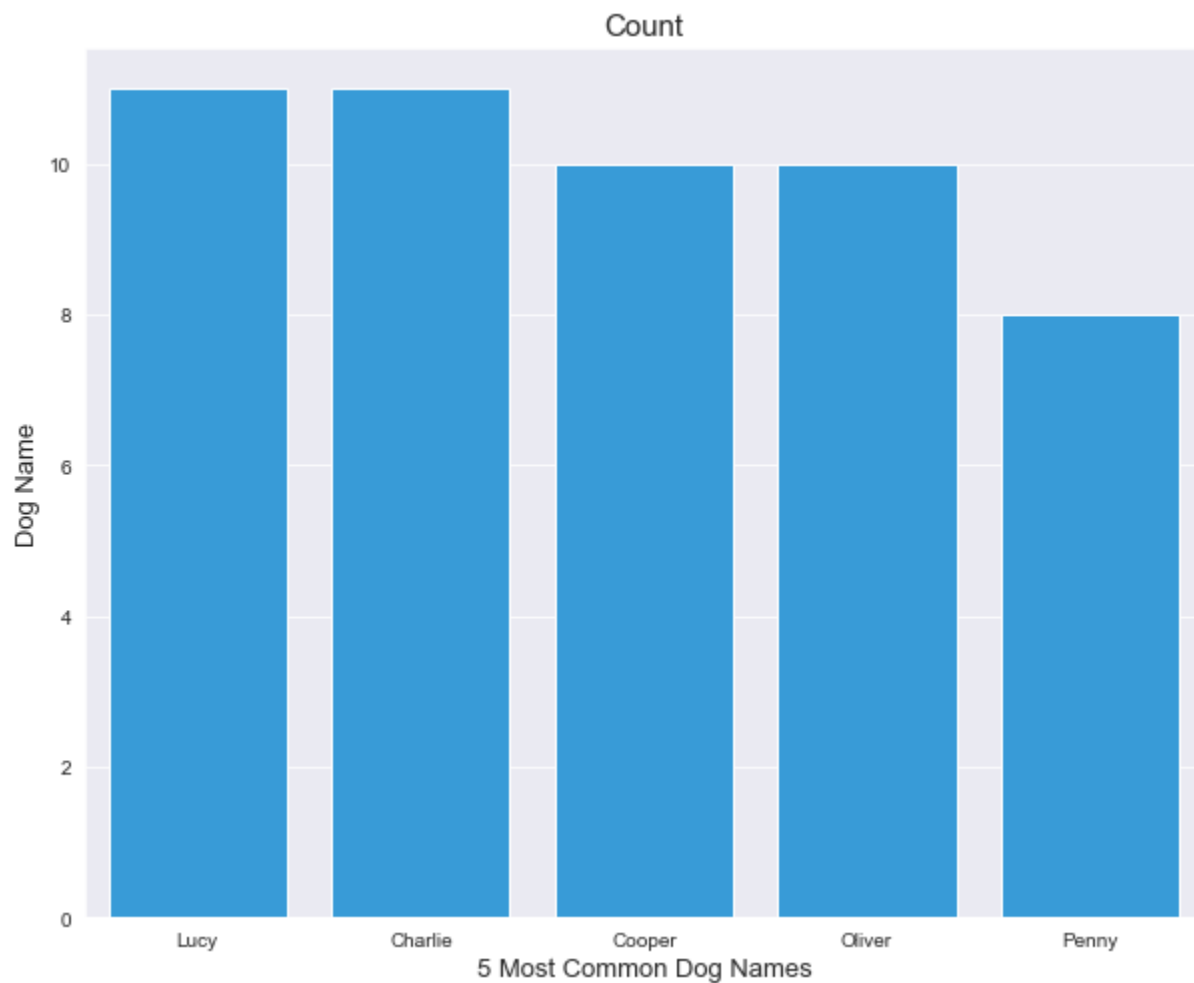
Most and Least common Breeds

There are several ways to grab these features but for this visualization, the count of distinct breeds is derived and the breed with the highest appearance is the most common while the breed with the lowest appearance is the least common breed. The most common breed is **GOLDEN RETRIEVER** followed by **LABRADOR RETRIEVER** while the least are **Scotch_terrier**, **standard_schnauzer**, **groenendael**, **silky_terrier**, **Scottish_deerhound**, **Bouvier_des_Flandres**, **Lumber**, **Irish_wolfhound**, **Japanese_spaniel**. This is visualized using a Barchart to plot this information.



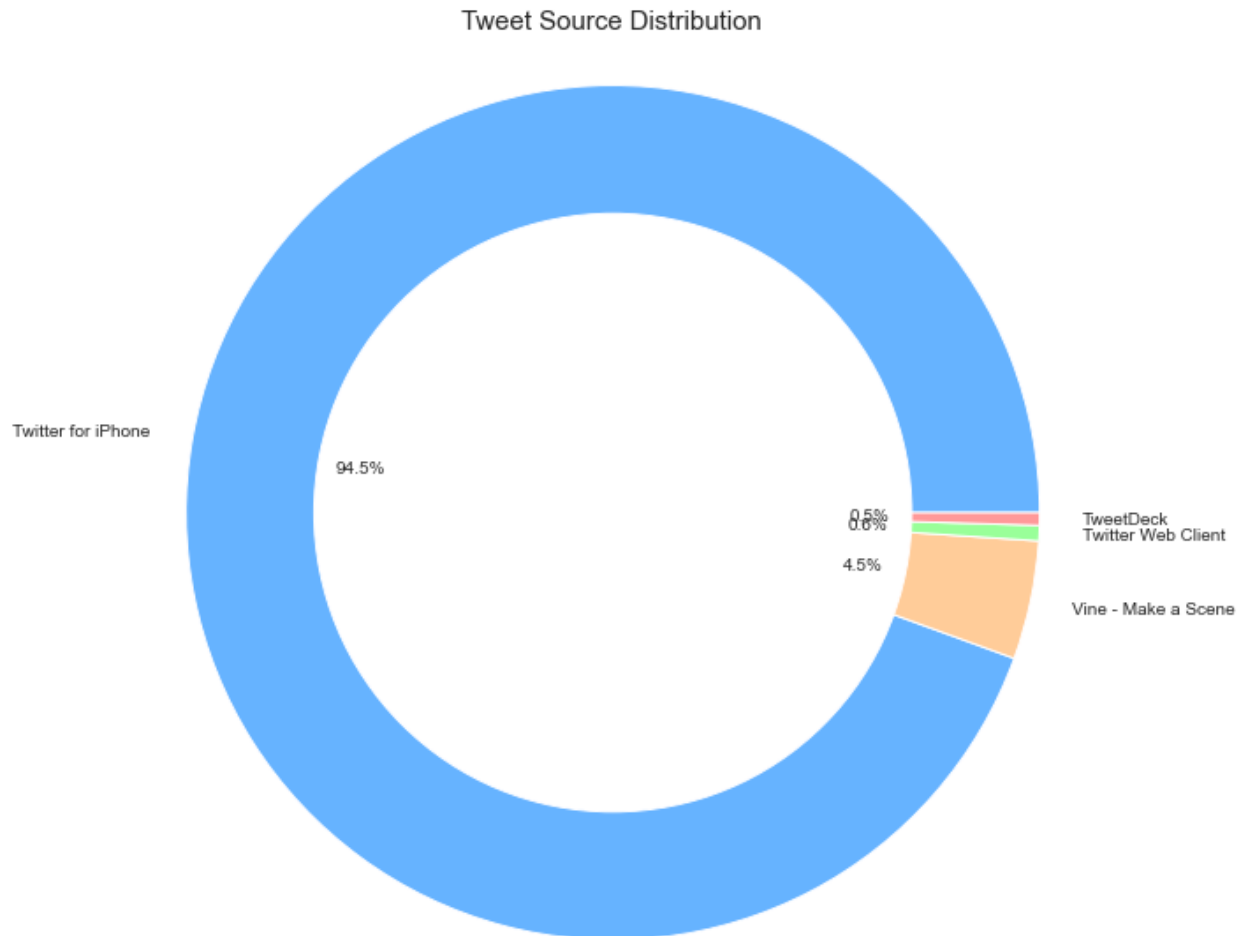
Most common Dog Names

The same method is being used to derive the most common names, the entities with no assigned names are disregarded. **CHARLIE** and **LUCY** is the most common name with 12 Dogs named **Charlie and Lucy**, followed by **COPPER** 11 and **OLIVER** 11. This is also visualized using a Bar chart to plot this information.



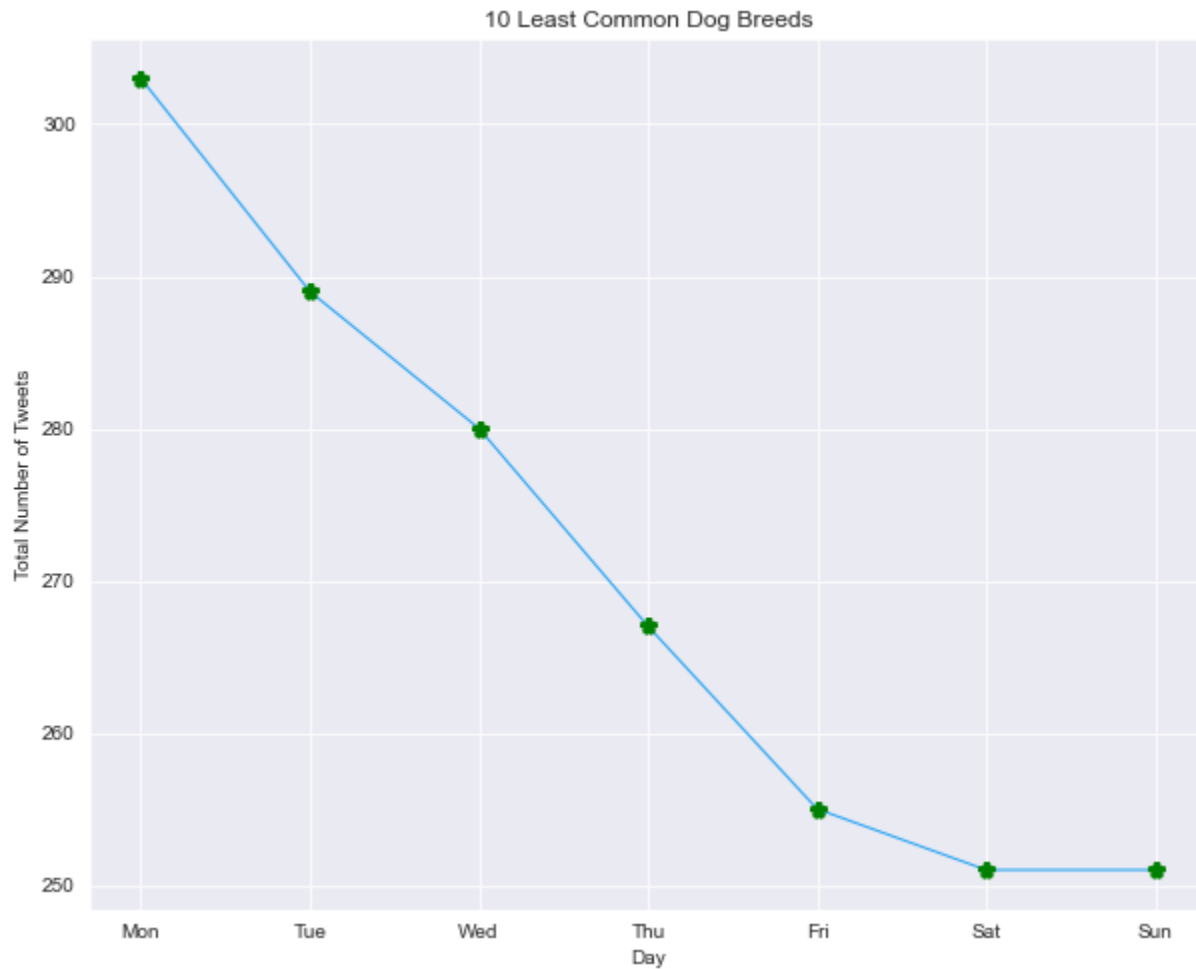
Distribution of Tweet Sources

There are 4 distinct modes of tweeting by the user. **TWITTER for iPhone** is the most common source of the tweets meaning the user tweets most of the information using an iPhone. This source carries 94.5% of the tweets' sources. followed by **VINE - MAKE A SCENE** (4.5%) and **TWITTER WEB CLIENT** (0.6%) and lastly **TWEETDECK** (0.5%) is the least source of tweets. This insight is visualized using Pieplot as shown below.



What day of the week does the user tweet mostly?

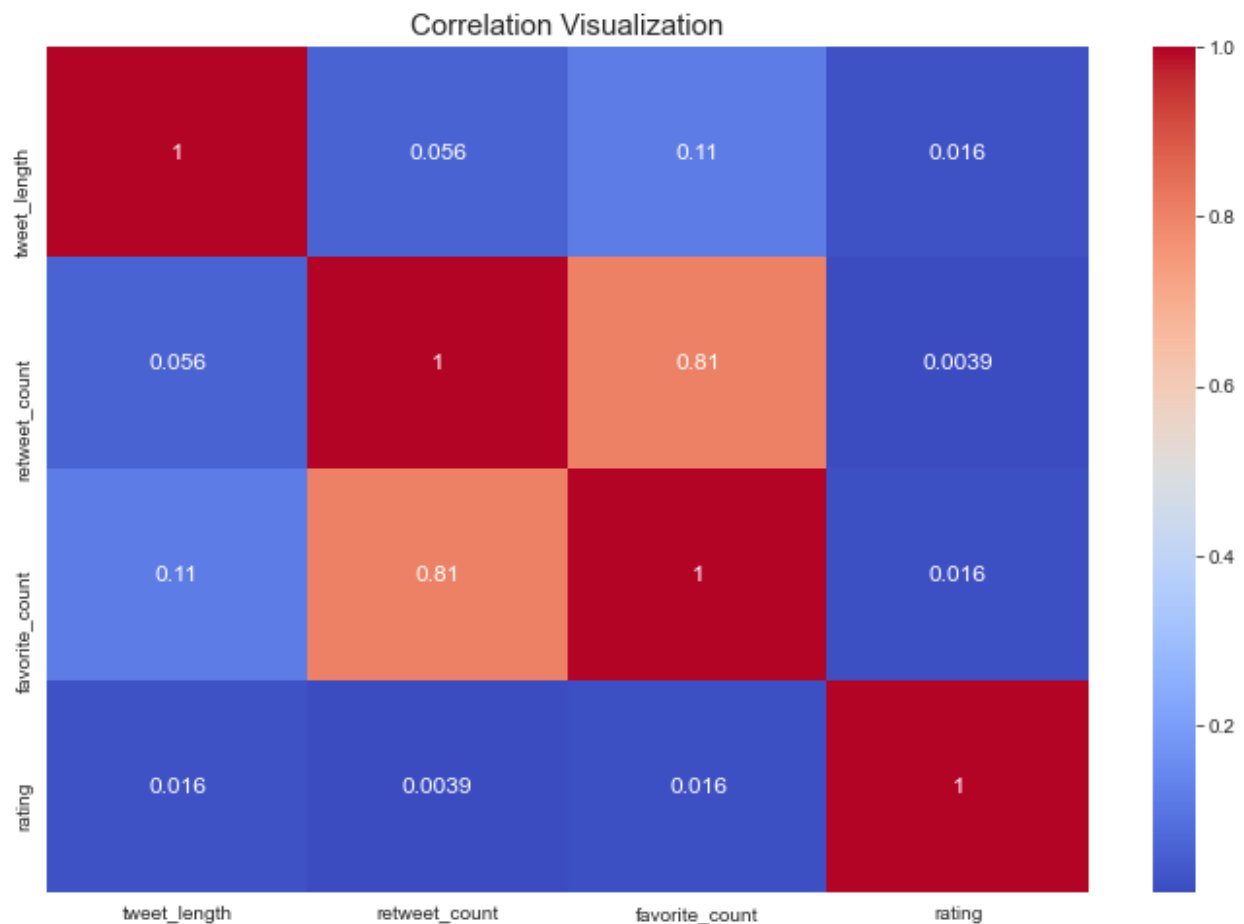
After using the `value_count` method to classify all information in the master dataset to their respective day. It shows, The user [tweets](#) mostly on **Mondays** and rarely on **Saturdays and Sundays**. The visualization is presented below using a line chart.



Relationship between the following underlisted numerical columns

- tweet_length
- retweet_count
- favorite_count
- rating

Correlation is best suited for analyzing relationships across numerical columns and is employed to get how each column relates to one another. After applying the correlation method, the proceeding returns a correlation table which is then visualized using heatmap to give better views and graphics for more understanding. All the columns have positive correlation to one another but some are weaker and some are stronger. Below are the relationship;



- `tweet_length` has weak positive correlation to `retweet_count`
- `tweet_length` has very weak positive correlation to `favorite_count`
- `tweet_length` has very weak positive correlation to `rating`
- `retweet_count` has very high positive correlation to `favorite_count`
- `retweet_count` has very weak positive correlation to `rating`
- `favorite_count` has very weak positive correlation to `rating`