## Reporting: wragle_report

I Wrangled data to better understand the relationship between different variables and to get insights from it. the data wrangling possess was fairly straightforward, after Gathering the data from different sources into three data frames I started by Listing the **quality issues** and beginning with the **visual assessment** to spot any unusual values or inconsistencies.

First I remove the retweets, as a user can retweet their tweet, to do that I filter texts starting with **RT** @ which means retweet. so I found there were Unusual variables in the name column which are Unidentified observations mixed with the name like *a such, his, old, etc.* I used the **Numpy** function np.nan to convert it into Nan values. After that, I drop unnecessary columns and changed column Name id to tweet_id in the tweet_data to match with other data frames when merging. Returning to the name column I found that the representation of the null values was wrong, it used the string None instead of Nan. Then came the **programmatic assessment** and I found many columns that had the wrong data type so I change them to the most convenient datatype. Then I dealt with the outlier values in the rating_numerator column and set the value for the rating_denumerator to 10 as it should. Finally, I list the **tidiness issues** and found the four columns *doggo, Floofer,pupper, and puppo* are the representation of the dog stage, So I combined them into one single column. The final step is to merge all three data frames and used the pandas merge function to put all of them into one data frame.