Twitter report

The Problems

- Null Values are alot in this data there is might be for different reasons. “In_reply_status_id” and “in_reply_to_user_id” has the same null values that might be for the same reason.
- “Timestamp” column as object
- There is alot of null values in other columns like Name,doggo,floofer,pupper,puppo,doggo,expanded urls.
- Entities column is a dic. And have empty data that does count as missing values
- display_test_rang column has a list of 2 values that i have to make each value in a separtated column

Quality :

- **Inconsistent dog stage labels in twitter_archive_enhanced.csv.**
- **Missing values in name, doggo, floofer, pupper, puppo columns in twitter_archive_enhanced.csv.**
- **Inconsistent date/time format in timestamp column in twitter_archive_enhanced.csv.**
- **Missing values in image predictions for some tweets in image-predictions.tsv.**
- **Unclear meaning of some columns in tweet-json.txt like "entities"**
- **Inconsistent capitalization in dog breed predictions in image-predictions.tsv.**

Tidiness issues :

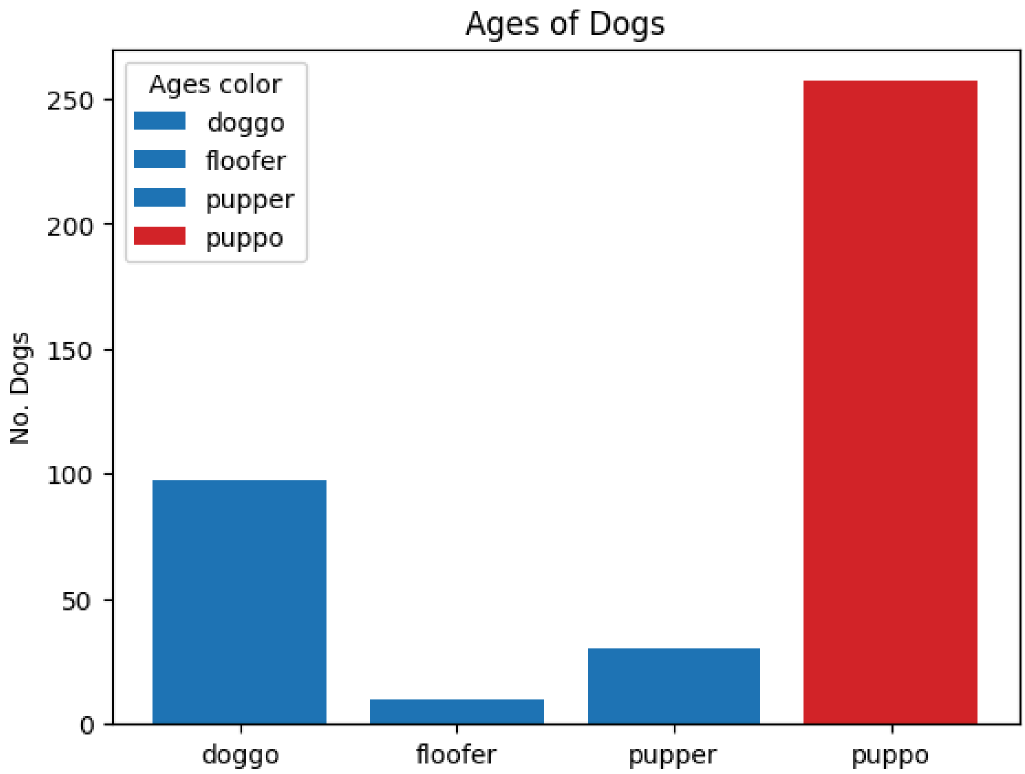
- **Dog stage information spread across multiple columns (doggo, floofer, pupper, puppo) in twitter_archive_enhanced.csv.**
- **Display text range is in a list**

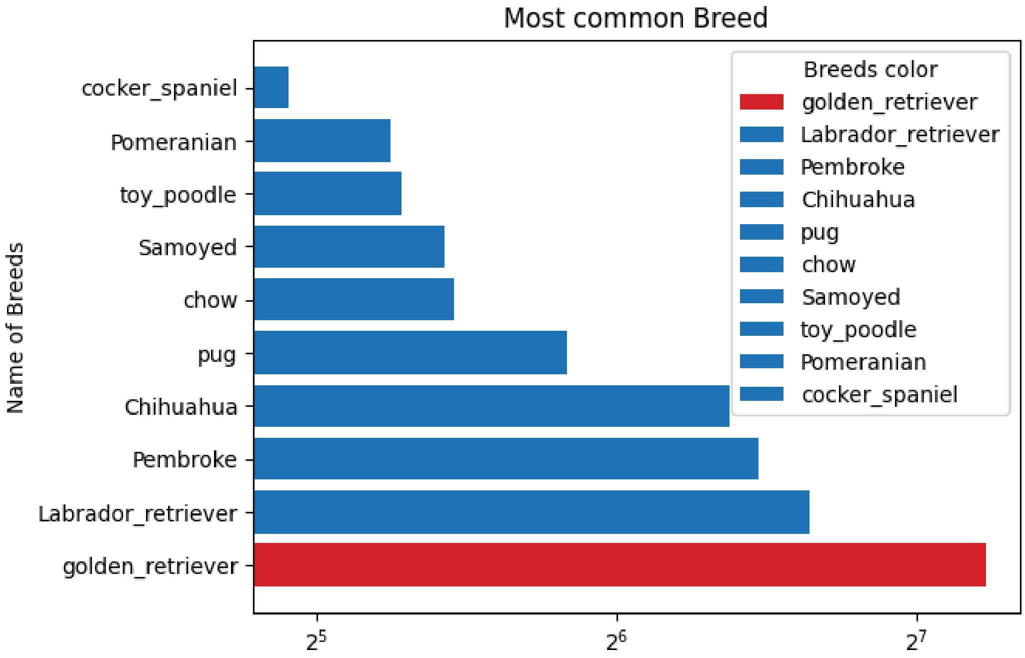
- **user Column have the same value in all rows and that does not make sense**

How i will treat it

will Drop the null values in some columns that i have to drop from it and i will unpack the dic. to create more usefull columns to be able to view them in a better way but not all values only the important ones ,change the time from object to datetime, there is 2 columns i will fill the nan value of them with zero cause they only has two values 0 and nan, i will do some visualization with colorful view,drop some useless

columns,fill dog names with other column





From the above we can say that the most common lang is English and

the most common dogbreed is Golden and the most common age is puppo