**Data Cleaning:** In the dataset given we can see that the GENDER column has 19953 non-null values, which we are considering as a target feature. Cleaning of dataset is very important as we will have a better understanding and analysis of the data.

I have dropped the following features which were not impacting a lot on the target variable and also deleted some irrelevant values where gender is not male nor female nor brand.

**Features which were not relevant:**

* Profileimage
* tweet\_id\_golden
* gender\_gold
* profile\_yn\_gold
* profile\_yn:confidence
* profile\_yn
* user\_timezone
* fav\_number
* \_unit\_state
* \_trusted\_judgments
* \_last\_judgment\_at
* \_unit\_id
* Name
* tweet\_created
* tweet\_coord
* link\_color
* sidebar\_color
* tweet\_location

**Dependent Features for finding out Gender are**

* Created
* Description
* Retweet\_count
* Tweet\_count
* Text

**EDA Analysis**

Performed the below analysis and understood the correlation between the features and analysed the data :

1.Plot of gender

2. Plot of tweet Popularity

3. Plot number of tweets per hour

4. Plot number of words used per year

5. Categorical plotting of text and tweet by gender

6. Categorical plotting of gender per hour tweet

7. Categorical plotting of words of emotion used by male and female

# Classifier method which I used is Ensemble Classification for predicting gender and also applied neural network methodologies to predict the gender