

## Rapport communiquant toutes les observations et affiche la ou les visualisations produites à partir des données traitées.

### Observation faites sur le jeu de données dfArchive

**dfArchive** contient l'archives des tweet de dog\_rates jusqu'en 2017

- Ce jeu de données ne contenait pas que les tweets de dog\_rates.
- Il contenait aussi les réponses des utilisateurs et les retweets.
- Nous avons donc supprimé toutes ces lignes et avons conservé que les tweets de dog\_rates.
- Nous avons remarqué que dans les colonnes name , doggo , floofer , pupper, puppo que l'expression None était utilisé pour remplacer les valeurs manquantes
- Nous avons donc remplacé le mot None par une chaîne caractère vide
  
- La source des tweets était entre des balises <a> </a>. Nous avons juste gardé que l'expression comprise entre cette balise <a> dans la colonne source.
- Le timestamp fournit ne nous aidait pas pour analyse. Nous avons donc gardé que la date sous forme de yyyy-mm-dd dans la colonne timestamp.
  
- Il y avait beaucoup de colonnes dont, moi je ne jugeais pas important je les ai donc supprimés et je n'ai maintenu que celle que j'ai jugé pertinentes.
- Toutes les colonnes non pertinentes je les ai supprimés.

### Observation sur le jeu de données tweet\_json représenté par dfAnswerApi

**dfAnswerApi** : contient les réponses obtenu par interrogation de Tweepy sur les tweet\_id de dfArchive.

- Ce jeu de données contenait beaucoup de données nous avons juste pris pour chaque tweet\_id de notre dataframeArchive nettoyé le nombre de favoris et le nombre de retweet en utilisant la méthode Merge de pandas

### Observation sur le jeu de données image\_Prediction représenté dfImage

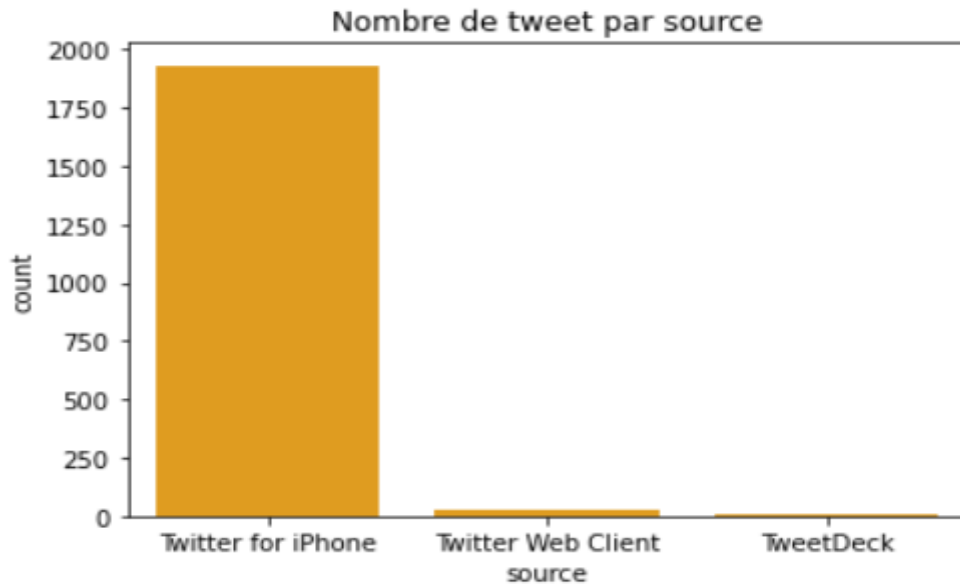
Il comporte surtout l'url vers les images de chaque tweet\_id.

Nous avons donc repérés pour chaque tweet\_id , l'url qui renvoie vers son image.

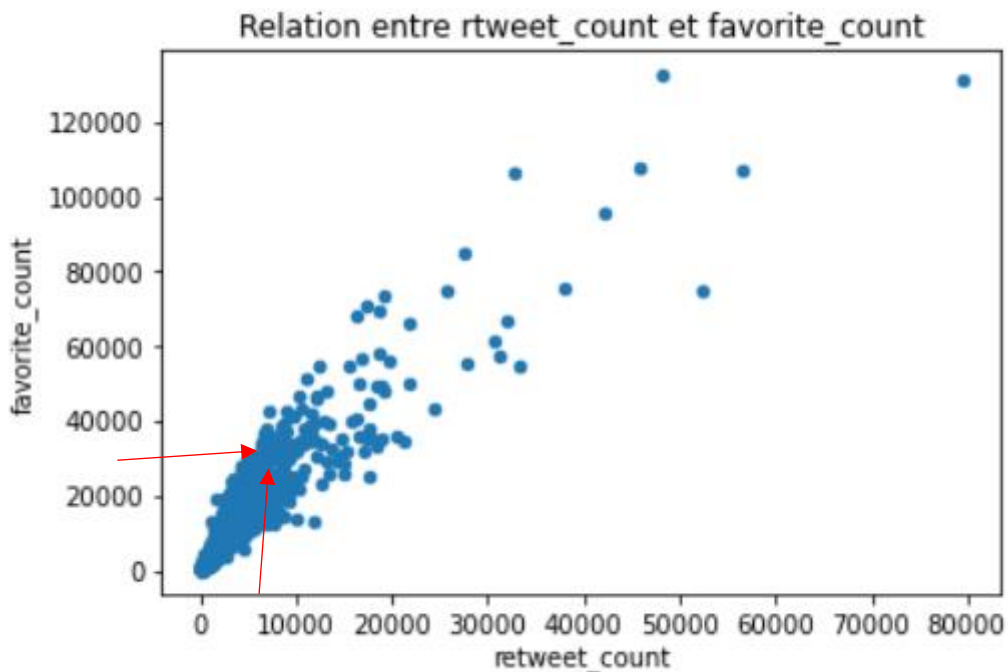
En faisant une jointure en utilisant la méthode **Merge**.

### Quelques analyses et visualisations.

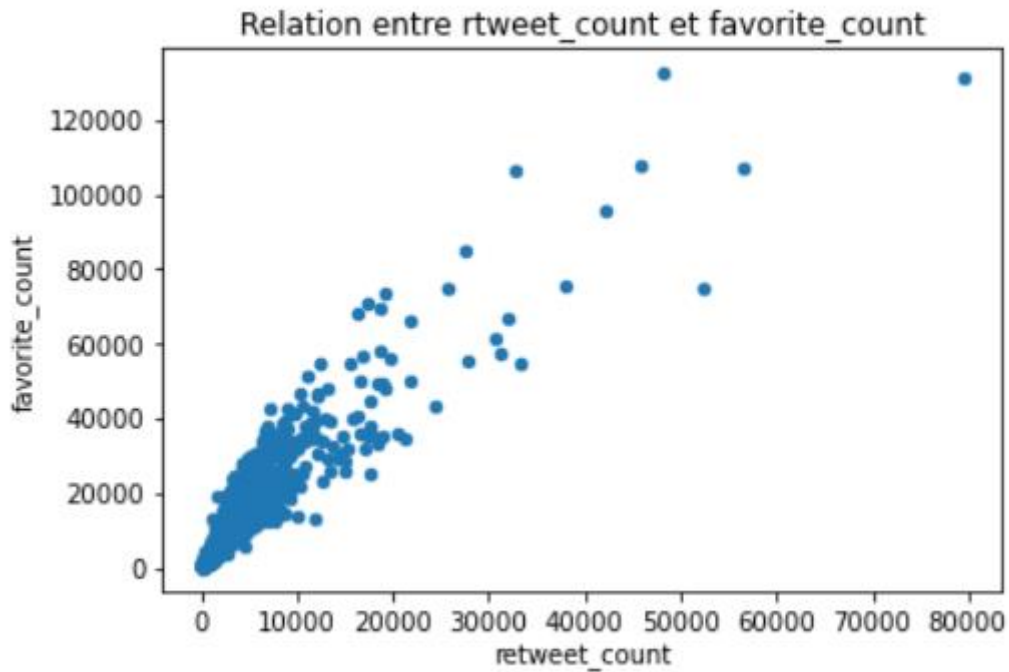
Nous avons remarqué que plus de sources des tweets proviennent de la source **Twitter for iPhone**



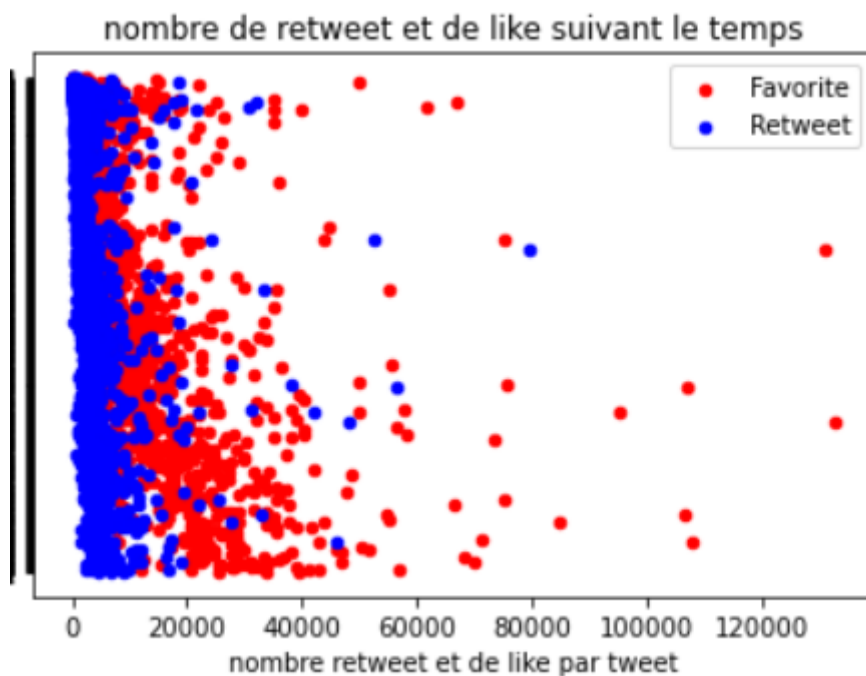
Nous avons remarqué qu'en moyenne un tweet de dog\_rates obtient 12000 retweet et 38 000 favoris.



Nous avons remarqué qu'en général les tweets de dog\_rates obtiennent plus de nombre de favoris que de retweet.



Nous avons remarqué qu'au fil des dates le nombre de favoris et de retweet qu'obtienne un tweet\_id augmentent. Cela peut nous permettre de déduire que soit le nombre d'utilisateur de twitter augmente au fil des ans et ou le nombre d'intéressé ou d'abonnées du compte de dog\_rates augmente.



TOMAVO Clarisse

06/09/2022