

## Titre : Analyse des publications Twitter de WeRateDogs

### Question 1

Partant de la question qui était celle de savoir quel était la race de chien la moins bien notée et par extension la moins appréciée, nous avons premièrement récupéré les entrées (lignes) qui correspondait à un chien selon l'algorithme de classement. Et il se trouvait que nous nous retrouvons avec 550 enregistrements correctement identifiés comme celles des chiens et ayant des notes correctes. Sur ce, nous avons exécuté le code ci-dessous

```
liste = ['rating']
```

```
df_tmp.groupby(['race'])[liste].mean().reset_index().sort_values(by='rating')
```

et avons découvert que la race '**Afghan\_hound**' est la moins appréciée de toutes les races.

### Question 2

Notre question étant : Quelle race est la mieux reconnue par l'algorithme de classement d'image ?

Nous nous sommes basés sur les résultats obtenus à la première question dont le Dataframe. Nous y avons appliqué la même procédure et nous avons trouvé que la race '**komodor**' est celle qui a obtenu le plus de succès avec l'algorithme. Elle est suivie de près par la race '**keeshond**' et '**weimaraner**'

Le premier a été diagnostiqué avec 99.9956% de réussite.

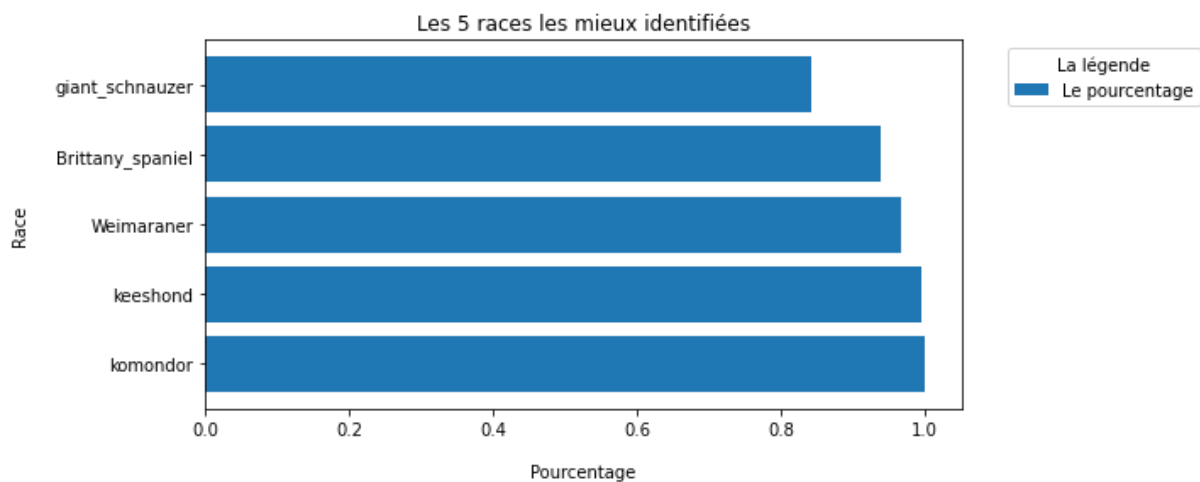


Figure 1: les 5 races les mieux identifiées par l'algorithme de classement d'images

### Question 3

La question 3 est : Quels sont les tweets qui ont été les plus partagés ?

Etant donné qu'on ne fait pas de différence entre les races dans cette question, nous revenons sur le DataFrame que nous avons stocké avant la résolution de la question 1. Par elle, nous exécutons la commande

```
liste = ['retweet_count']
```

```
df_full.groupby('tweet_id')[liste].sum().reset_index().sort_values(by='retweet_count',  
ascending=False)
```

et nous nous retrouvons avec le tweet dont l'id vaut **761672994376806400** en tête de liste.

Et il ne s'agirait même pas de la photo d'un chien selon l'algorithme.