

# Act Report

Mahmoud Khayrallah

## Introduction :

In this steps I used the merged dataframe master\_df (saved to 'twitter\_archive\_master.csv') to perform some analytics on the data which mainly included :

- descriptive statistics of different variables like the mode of names and breeds and the average ratings
- visual exploration or EDA through visualizations and more coding to draw insights and find relations between variables

## Descriptive statistics :

It helped me understand the data more and get some insights like the most used dog names , the most common breeds , the highest rated breeds and so on

## EDA :

Actually this is what the whole process about (exploring data), so we can include the above step in it too. I used visualizations and plots of one and more than one variables to draw insights and get more understanding of the data

## Insights :

### insights

- most common dog names in the data-set :
  1. Charlie (12)
  2. Oliver (11)
  3. Laila, Cooper, Lucy, Bob (10)
- the most common rating is 12
- the average rating is 10.85 and the median is 11
- most common dog breeds in the data-set :
  1. Golden retriever
  2. Labrador retriever
  3. Pembroke
- highest rated breeds are Samoyed, Golden retriever and Great pyrenees  
lowest rated breed is Soft-coated wheaten terrier
- the higher the rating is, the more retweets and favorites it grabs

■

**Some visualizations (the rest are in the notebook) :**

