ACT REPORT

I am in course of Data Analytics at Udacity, we have an interesting assignment to analyze some twitter activity.

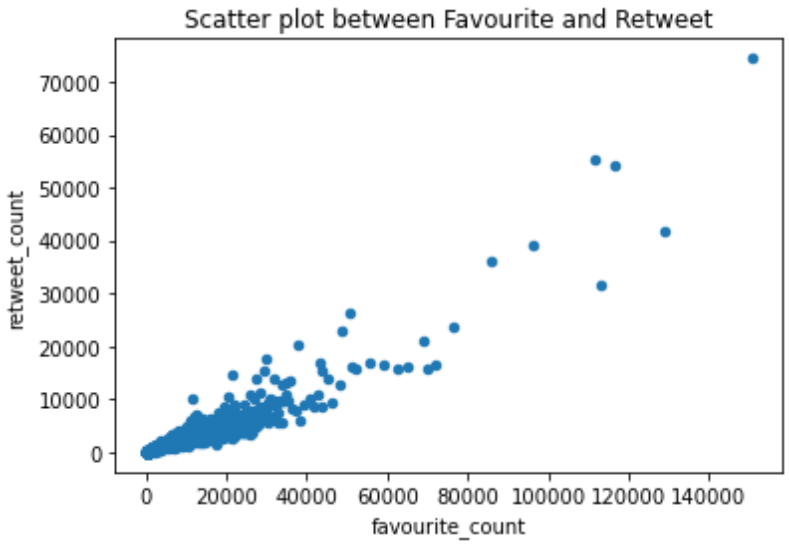
This is not a simple task, first we must gather all the information that we need from various sources (included a connection to Twitter API).

Then assess the data and find quality and tidiness issues in the databases so we can clean them and finally make meaningful insights.

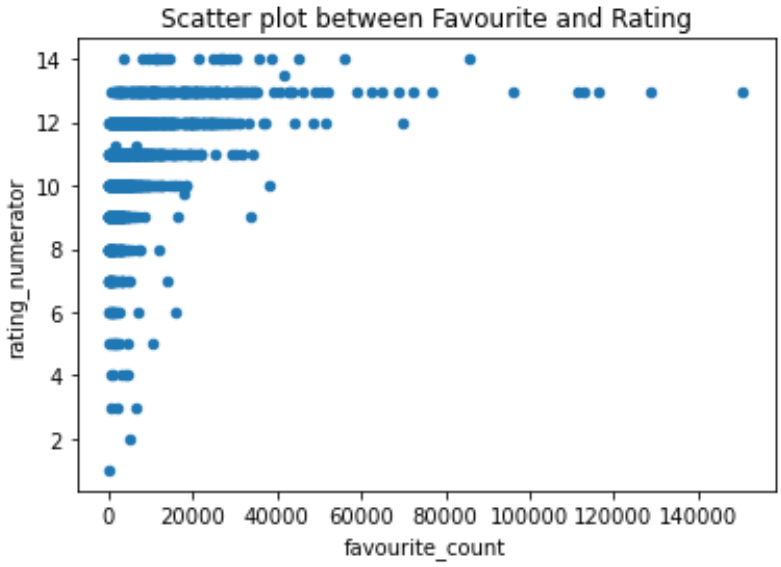
This process is especially important, there is a lot of work previously to get some findings. For example, I found:

1) After merging the twitter archive with the prediction of breed, there are 14 breeds that make the 50% of the total tweets. The top 5 most common dogs are Golden retriever, Labrador retriever, Pembroke, Chihuahua and Pug.
   But common dogs doesn´t mean that there are the favorites, neither of these breeds are in the top 18 of most favorites' tweets.

2) There is a positive relation between favorite and retweet counts, it makes sense that these two indicators move along. It is a good way to track the popularity of a tweet and what makes something "viral".
   It´s important to say that we have "popular" tweets that are way above the average, for example the top 1 it is a Labrador swimming in the pool.



3) There is also a relation between the rating that the owners give and the favorite counts, this means that higher numerators also have higher popularity. You know that your dog is doing something special and people notice too.



In conclusion we have a lot of work getting the data, the first step could be the harder but it will make all the next analysis valuable and accurate.