A New Age for Data Analysis

Arturo Guerrero Enterría (Udacity DAND)

Since centuries we are analysing data. And humans tend to analyse each step we take in every field in our life's.

But a new age for the data analysis has come due to the irruption (disruption) of new technologies and the Big Data Analysis. Every day millions of users load tons and tons of data on Internet and this massive amount of data is available for it´s analysis through new sophisticated data analysis methods and tools.

Every minute 4.3 million of videos are visualized in Youtube, 13 million of text messages are sent, 73.000 are online transaction are made through Amazon and 750.000 songs are heard in Spotify (Abderasoul 2018). Could you imagine that you can get this information and analyse it? Now we can.

Let´s look on how Data Analyst can do it:

We have proceeded to get information concerning Twitter:
Three different files with more than 5000 tweets related to only one Twitter group "WeRateDogs".

What we can get through these files is enormous. We can know which is the best rated dog in this Twitter group. We can see what type of dog it is, the name, how many times it has been retweeted on Twitter.

We also can get information concerning the dog stages:

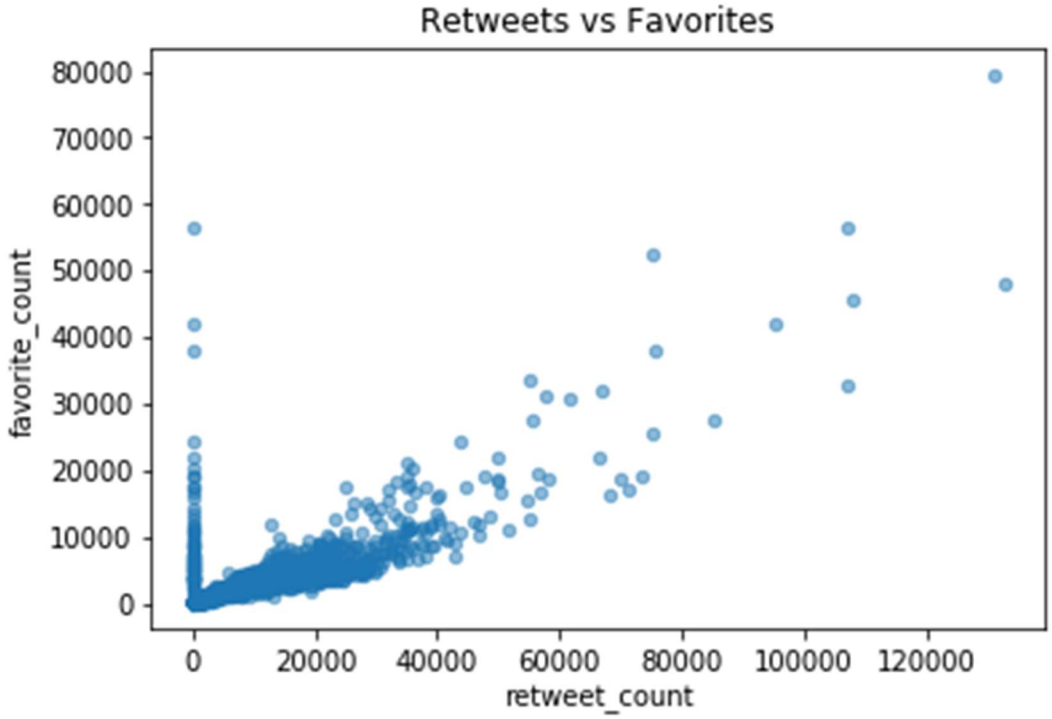
```
dog_stage
doggo      23
floofer     5
pupper     72
puppo      10
```

Or link the retweets to the dog stages:

dog_stage	retweet_count	favorite_count
doggo	143481	466068
floofer	11236	47795
pupper	178307	611874
puppo	35978	130694

All this information is gather through specific digital tools and programming (In this case API´s and Python, but we could do it through different software and programming languages available as opensource for everyone). But we can get even more information if we get closer.

We can change the point of view and have a deeper understanding of how retweets and the favourites are linked and visualizing this relation we will find that increasing the number of retweets we will get higher number of favourites.



But we could go further and analysed this massive number of tweets, retweets and favourites and check if the visualization is giving us a right insight. For doing this we can check the results programmatically. We can run the Pearson correlation test and we will get this result: 0.924631686087.

We could go on and on through this analysis and we would get very useful information. And now is your turn! Imagine what could you know if you are able to analyse all this information. The amount of data available today is bigger than ever in human history. Is only waiting to be analysed.