

## **Key Friend Profile:**

### Data to collect:

UTC Offset  
Number of followers  
Number of friends  
Date of creation  
Number of Tweets

### Preparation of data:

For each data collected, put it in a distribution array that contains an unique element of each different element collected.

For example, if we have five times the data “UTC Offset” equal to UTC -03, we put this data only once in the array

### List of subindex:

- Offset
- Age
- Friends
- Followers
- Tweets number

### Calcul of the probability:

- Length of the offset array divided by 8
- Length of the age array divided by 50
- Length of the friends array divided by 50
- Length of the followers array divided by 50
- Length of the tweets number array divided by 50
- If any of the score above is more than 1, we fix it to 1.

Then, for the score of everything above, we multiply by two the offset score and add it to all others score. Then, everything is divided by 10.

All the calculation above are made once for the friends of the profile analyzed, and once for the followers of the profile.

Once we got the two scores. We calculate the final friend index score by multiplied the score of followers by 1,5x and add it to the friends score. Finally, we divide it by 3.

### More explanations:

The offset score is the most important, this is the reason why it count twice in the final score.

This is not very common for an user to have friends/followers from a lot of different timezone.

We consider others data like more common to be different and care less about their value in the final score.

The score of friends is most important than the score of followers since friends are chosen by the user, so this score count 1,5x more in the final score.