**Python assignment**

* Files Used:

1. slurs\_file\_.xlsx: This file contains few slurs words, its representation and reasons.
2. tweets\_.csv: This file contains 31,961 tweets collected from kaggle.

* Mounted google drive with google colab.
* I separated each input text into sentences by [‘ . ’, ‘ , ’ ,’ ! ‘ ,” ; “].
* By listing a few conjunctions, I separated each sentence into parts to find the degree of profanity.
* Example:

A and B. C

Where,

A- contains racial slurs words

B- doesn’t contain racial slurs words

C- doesn’t contain racial slurs words

1. In the first sentence, the degree of profanity is ½.

2. In the second sentence, the degree of profanity is 0.

3. On total, as no.of sentences=2

=> degree of profanity = (½ +0)/2=¼.

* I defined a function for the above function “find\_dop\_score”.
* I imported the tweets\_.csv file and created a new column as “DOP” to store the degree of profanity for each tweet accordingly.
* I showed the tweets where dop>0.0
* I implemented the function with a sample text:

“ABC is a kind , Do you know why. Because I hate them”

1. As the same format in above example, (ABC is a racial slur word because it means, American-Born Chinese. An Americanized Chinese person who does not understand Chinese culture.
2. So, the score will be 0.25 (same, as explained in example)