all\_label\_final.csv

simple\_label :

在 bureau, pos\_cash, installments, credit\_card 項目內, 任一合約有欠交紀錄者

sum\_avg\_late:

bureau違約數/bureau總合約數 + pos\_cash違約數/pos\_cash總合約數 +

installments違約數/installments總合約數 + credit\_card違約數/credit\_card總合約數

total\_score:

依據各合約計算其違約狀況係數(以後詳述), 並加總

注意:

bureau違約係數: 0-6 整數

pos\_cash違約係數: 0-3 整數

installments違約係數: 0-3 整數

credit\_card違約係數: 0~2 精確到小數下一位浮點數

違約係數表示該合約中違約狀況的嚴重程度

大致上欠款頻率越高, 最大欠款越多, 欠款發生在一年內或是最近一次合約, 就越嚴重

各係數range不同, 因為考量過weighting:

bureau各合約權重為其他種類合約的兩倍

credit\_card各合約的計算比較草率, 加上標準比較寬鬆, 因此weighting較低

sum\_avg\_score:

bureau違約係數/bureau總合約數 + pos\_cash違約係數/pos\_cash總合約數 +

installments違約係數/installments總合約數 + credit\_card違約係數/credit\_card總合約數

建議用 simple\_label 或 total\_score 來建模

程式碼請見

<https://github.com/JenCHuang/PyStudy/tree/master/HomeCredit_labeling>