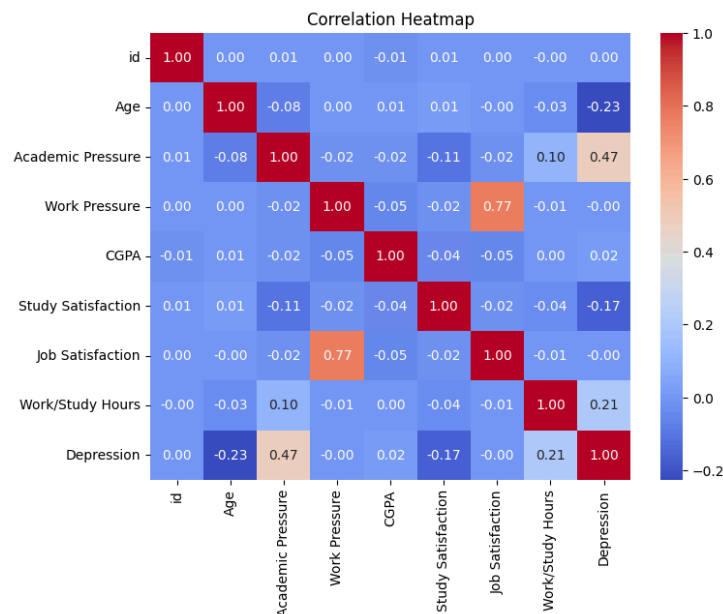


Group 6

Slide 7

Q:所提到的內容(清資料、理解資料)應該不算是狹義上的Challenge

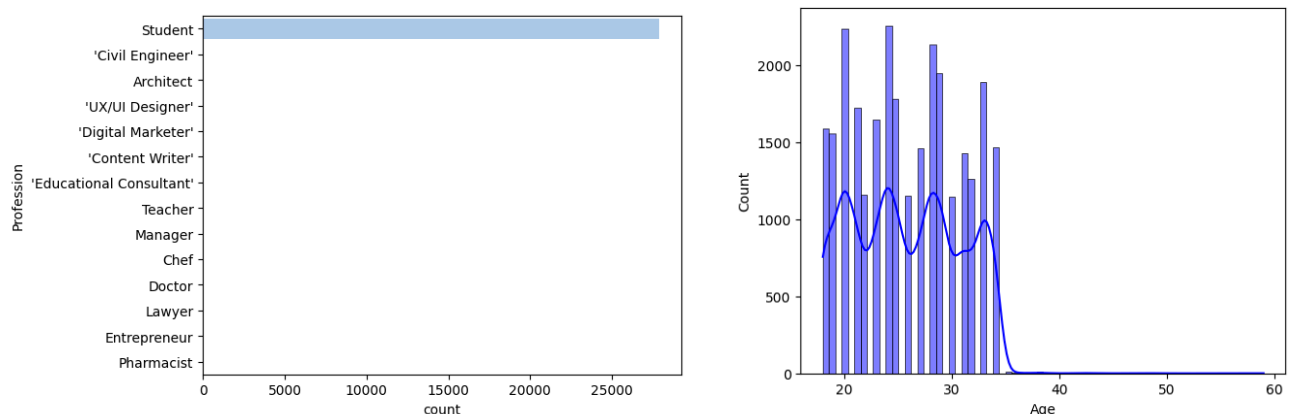
A:將原先的challenge改為資料的特徵數量不夠多，篩選過後給模型會有侷限性，而我們先對 int type feature先進行correlation，發現除了academic pressure外，其他相關度都不太高



Slide 10

Q:題目是做學生壓力，不過Dataset中年齡資料分布要注意

A:我們資料年齡層雖有18-59，但大多數年齡層落在於18-35之間，而且資料數在總共27901裡面中，是學生的資料就有高達27870，基本上可說是裡面只有學生資料。



Slide 14

Q:問題定義不明確

A:用字要注意將” high risk ”替換掉，已將” Predict whether an individual is at high risk for depression.” 改成 ”Predicting whether a person has depression.”

Q: Prediction的window size

A: 資料蒐集過程沒有包含時間資訊，故預測的結果為根據輸入的當前情況做預測是否有憂鬱情形，而非預測某個時間點是否有憂鬱情形。

Slide 15

Q: accuracy跟平常的不太一樣

A: 這是我們自己想的，我們只是換一個比較好懂的方式去表示。P : Predict、G : Ground truth，因為我們預測是否有憂鬱只有兩種表示，1與0，因此若預測結果正確， $P - G = 0$ ，反之相減後結果可能為 1 或 -1，因此取平方確保結果皆為正，N為Test數量，機率因此也就是 $1 - \text{失敗比率} = \text{預測正確比率}$ 。

Slide 17

Q: 怎麼預期transformer會比現在的版本都來的高

A: 我們想說目前看到Kaggle上其他人都是直接使用Random Forest等模型去做預測，而transformer有Attention 機制，我們認為可以先透過此篩選出高度相關的特徵，再送進預測模型的話效果會來的更好。