作業4: 關聯分析(只能處理文字型資料，在做分析前須先做資料轉換(

Apriori：效能較差，需多次掃描 / support值不宜設太高或太低(無最適當的值，應該視不同的領域調整)

FP-growth：較快

1. 在關聯分析中最為著名的例子不外乎就是啤酒與尿布的例子，您可以簡單說明這是什麼樣的一個故事嗎? (5%)

美國大型超市沃爾瑪(Walmart)，利用數據分析，發現每週五的晚上，啤酒與尿布的銷售量呈現正向關係，也就是每個週五的晚上時段，尿布和啤酒這兩樣東西一起，賣得特別好。

深入探討後發現原因是因為父親會去超市幫嬰兒買尿布，並且順便買啤酒回家，以便週末在家看球賽。

2. 而除了啤酒與尿布的例子外，您還能舉例說明一個關聯分析的例子嗎? (10%)

影音廣告投放媒體OneAD運用了9種不同維度的屬性進行廣告觀看時長的關聯規則分析，包括：廣告投放媒體、投放時間、廣告播放格式、裝置類型、受眾興趣、受眾年齡性別類別、廣告產業類別、廣告產業次類別、廣告投放模式，來探討影音廣告投放與觀看間各項維度的關聯，找出創造高影音廣告觀看長度的各個屬性組合，以家電類舉例，家電類顯著的關鍵規則在訓練資料中出現在觀看長度大於75%時段為下班回家途中(17:31 ~ 19:59)及在家休息時間(20:00 ~ 00:59)，明顯地集中於下班到晚上時間，推測可能為下班路途或在家中比較有時間思考及搜索生活家電的採買或更換，在該時段對於家電類的廣告對受眾有較高的吸引力。

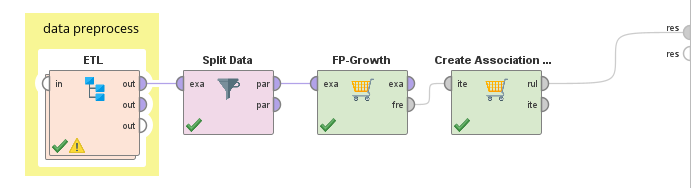
太棒了，這樣您就知道至少2個例子囉.接下來我們要開始來實做!

在開始之前，我們要先了解一下，其實關聯分析是一項　監督式　非監督式　學習 (請圈選)　(5%)

所以.這與您之前所使用的流程不太一樣唷!

3.請開啟gamestudy.csv 這是一個關於遊戲研究的資料集， 如同我們在先前課程所述，在這個作業裡我們希望您可以使用FP-Growth Algorithm 進行關連分析。**只使用90%的資料(只使用90%的資料， 只使用90%的資料)**進行關連分析模型建構，並請將該流程存檔為學號-1.rmp檔　(15%)

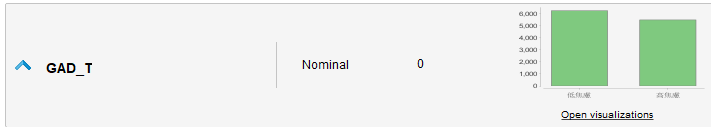
<https://www.kaggle.com/code/ahmed1rabie/classify-gamers-mentality/input>

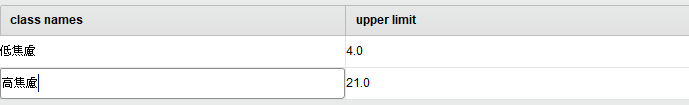


ETL

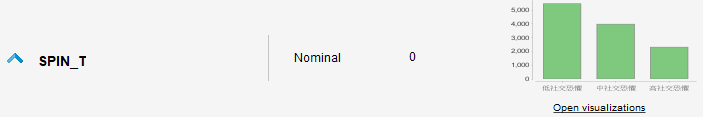
1. 敘述統計
2. 將空值、極端值刪除
3. 資料型別轉換
4. 篩選欄位(將重複性高、資料過於不均、高度相關的變數刪除)
5. 資料離散化與重命名

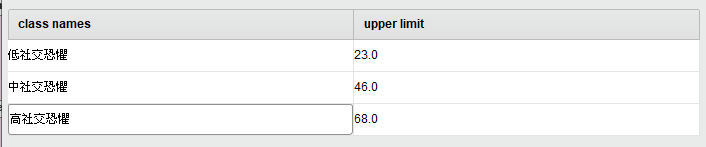
留下的欄位之敘述統計



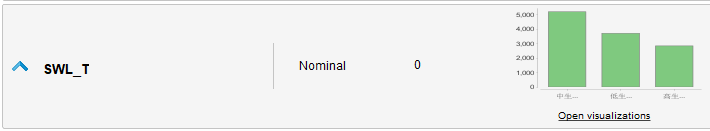


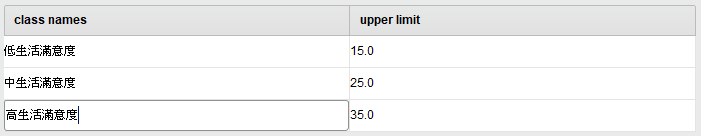
將GAD\_T的資料分為三部分，資料多落在低，此變數與焦慮有關



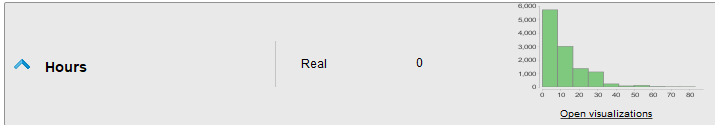


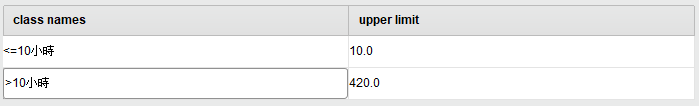
將SPIN\_T的資料分為三部分，資料多落在低，此變數與社交恐懼有關

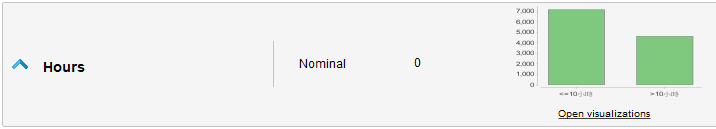


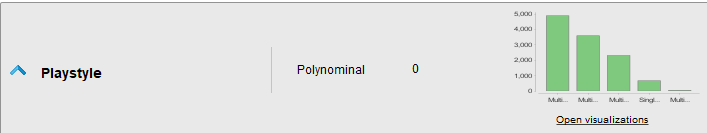


將SWL\_T的資料分為三部分，資料多落在中，此變數與生活滿意度有關

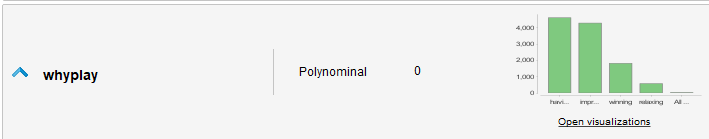




Hours呈現左偏，最多集中在10小時以下，所以後續將此變數分為 <=10小時 / >10小時



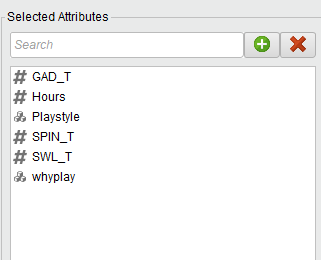
Playstyle多數集中在與自己實際生活中的朋友一起遊玩



whyplay大多集中在having fun

4. 請說明您模型的發現(包括:你留下那些欄位，您期待找到些什麼關連規則;您所尋找到的規則是否符合您的預期，如果可以請在doc檔中貼上您的規則，以方便作業批改(15%)

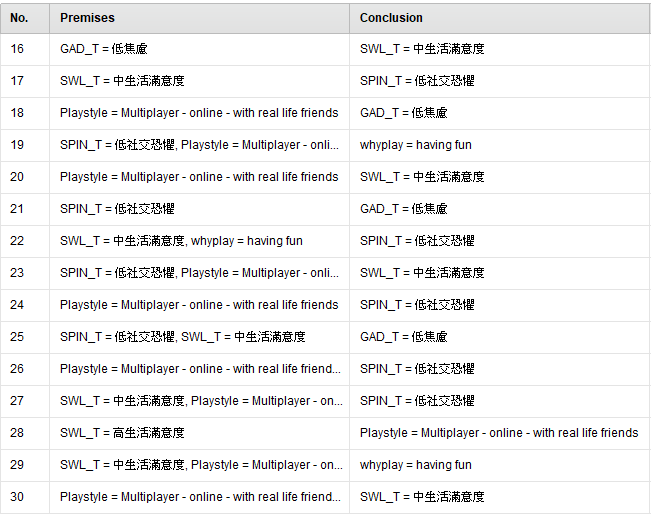
共留下6個欄位



期待的關聯規則

找出心理疾病(GAD、SWL和SPIN)與玩電玩所花費的時間、遊玩原因、遊玩方式之間的關係

實際找出的規則



舉幾個例子來說

no.18可以發現跟好友一起玩電玩，這種人多屬於低焦慮的心理狀態

no.26可以發現跟好友一起玩電玩並且玩電玩的目的是因為有趣，這種人多屬於低社交恐懼的心理狀態

no.28可以發現對生活滿意度高的人，通常會與好友一起玩遊戲

→從這幾個例子可以看出心理疾病與他們玩電玩的原因及方式有關係

5. 最後，我們來進行相關參數調整，請說明support值和confidence調高及調低的影響對於分析結果有何影響(8%)

當support調高時，會在初步就篩選掉出現次數較少的物件，因此這些物件就不可能出現在後續的關聯規則當中，留下的物件較少就導致所產生出來的關聯規則也會減少

當support調低時，會在初步留下較多的物件，所產生出來的關聯規則也會變多

當confidence調高時，所產生的關聯規則會減少，因為符合最小信心度的規則變少了

當confidence調低時，所產生的關聯規則會增加，因為符合最小信心度的規則變多了

6. 最後，請將拓樸圖畫出 (7%)



7.請將您模型存成學號後，用剩下的10%的資料進行驗證，並加以說明您可能的應用結果，並將相關流程存為學號\_2.rmp(15%)

根據這個資料集發現心理疾病與他們玩電玩的原因及方式有關係，通常心理狀態較好的人(低焦慮、低社交恐懼、高生活滿意度)會較常與好友一起玩電玩，並且他們玩電玩的原因是因為覺得電玩很有趣，因此若我是電玩產品銷售商，行銷電玩相關產品時，我會將這些心理狀態較好的人設為目標客群，除了可以增加銷售量外，或許他們也可以將這些產品推薦給他們的朋友，使銷售量更加提升。