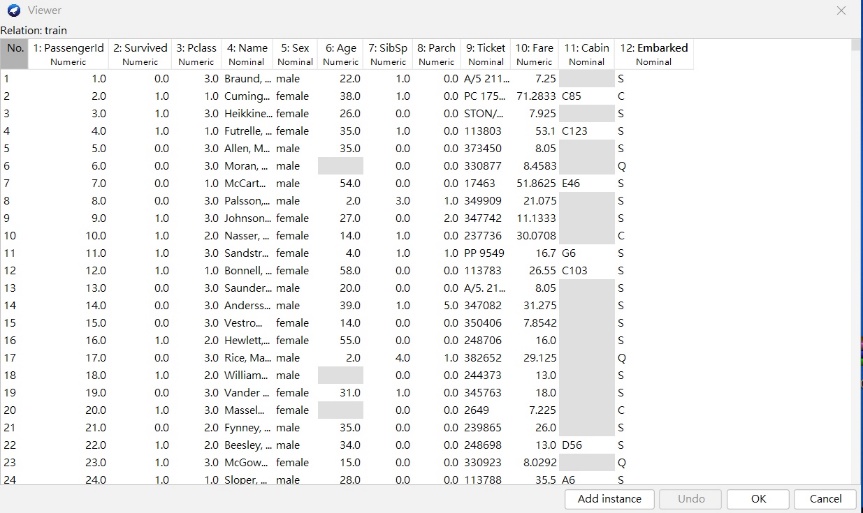
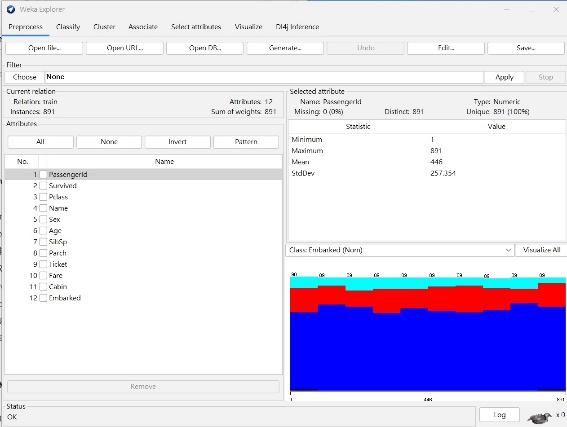
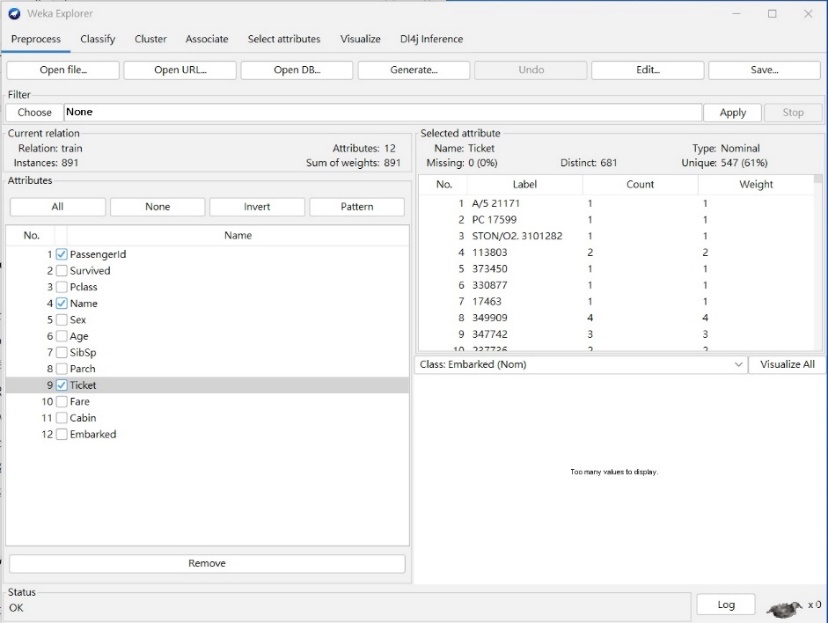
電子商務技術\_HW2

108403201 資管三A 黃名揚

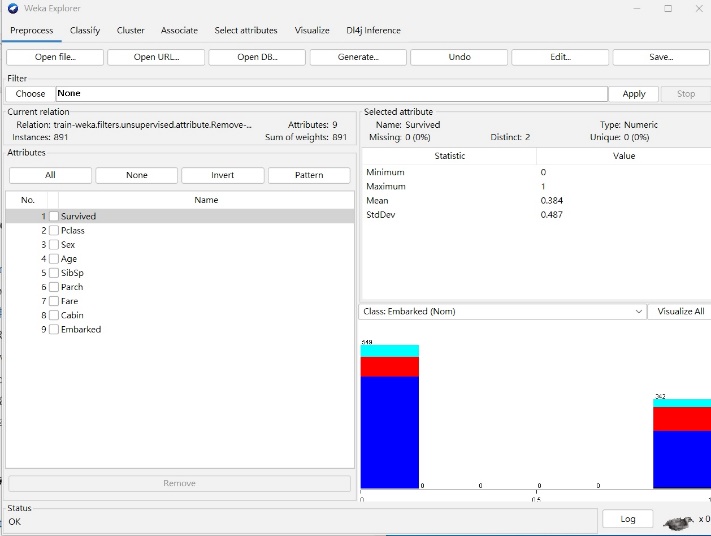
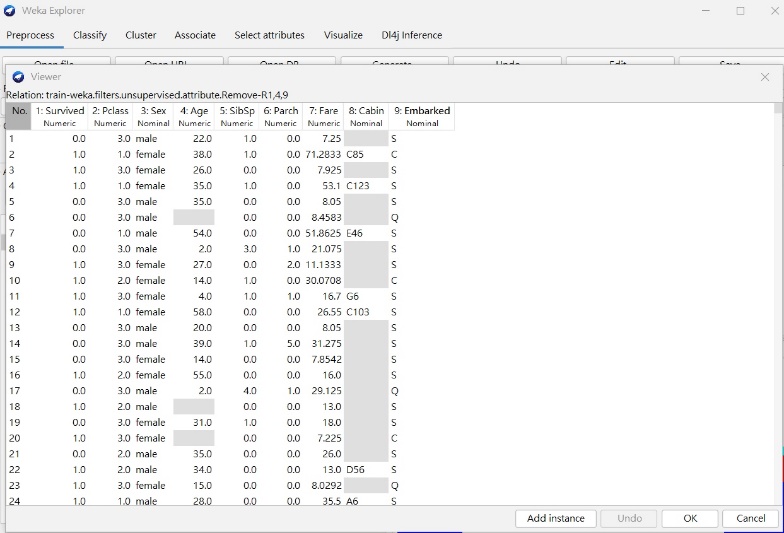
壹、請使用 weka 完成以下題目，並截圖結果附上適當說明

1.1 載入 train.arff，將PassengerId、Name、Ticket 欄位刪除

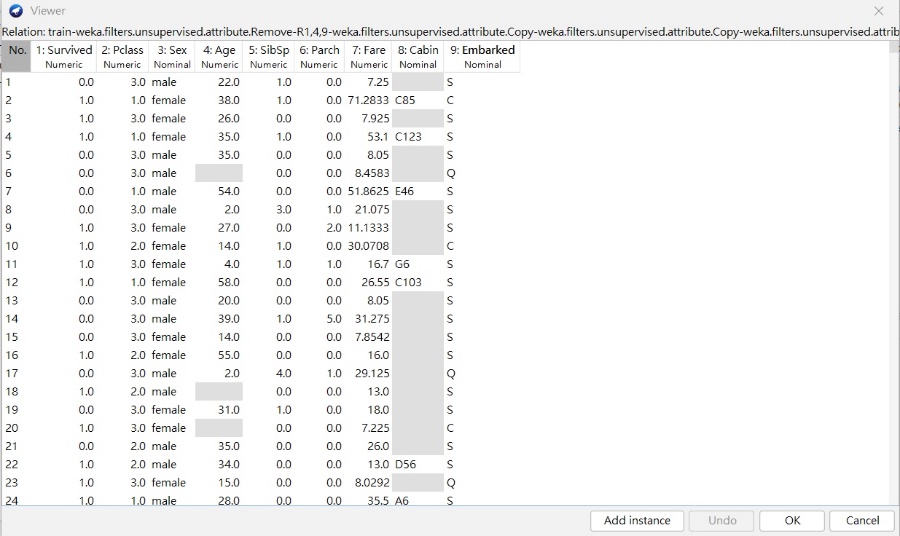
Step(1)載入train.arff



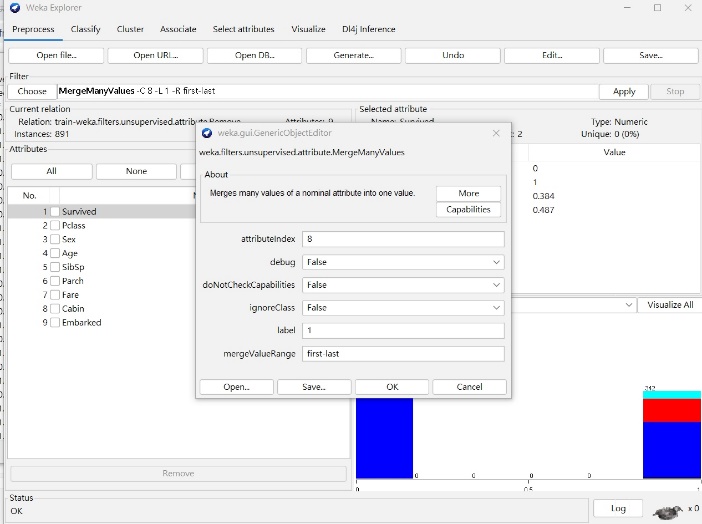
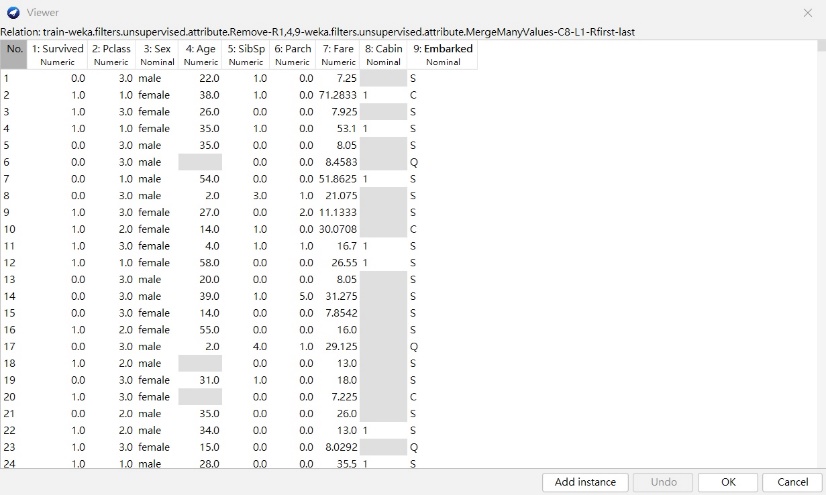
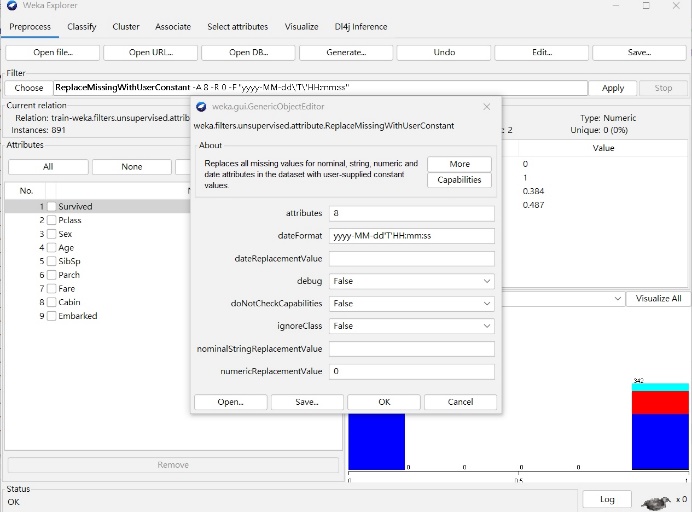
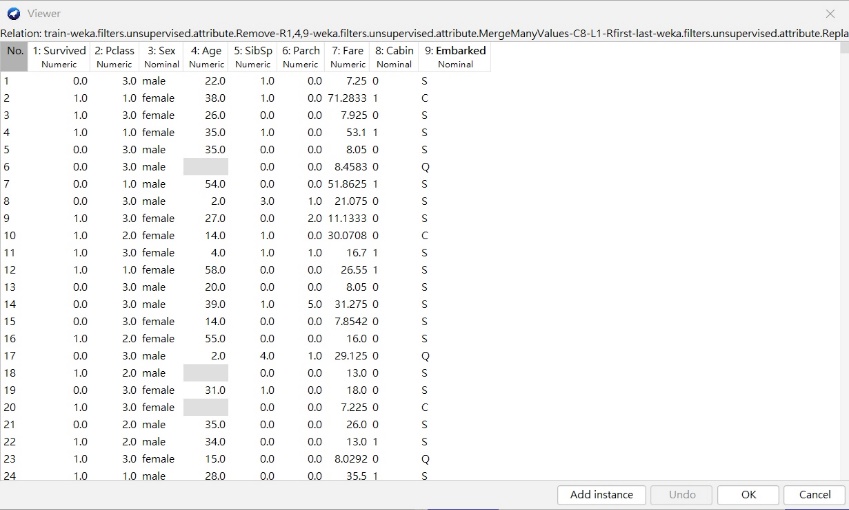
Step(2) 選取PassengerId、Name、Ticket 欄位並且刪除

Step(3)刪除完成後資料

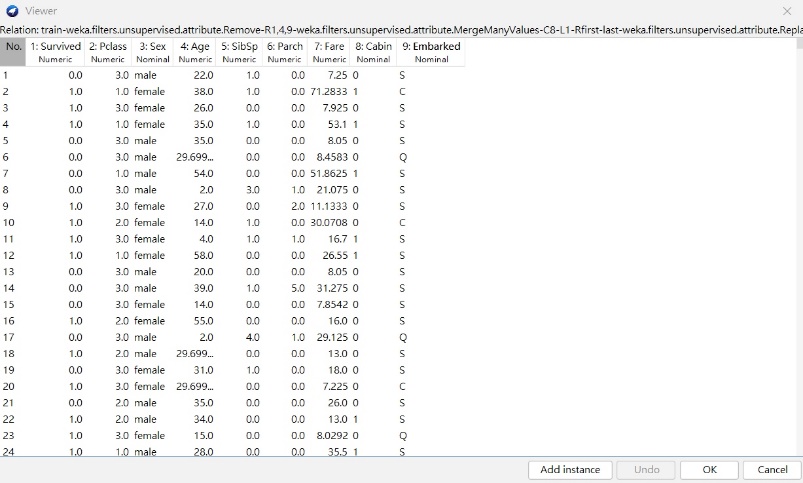
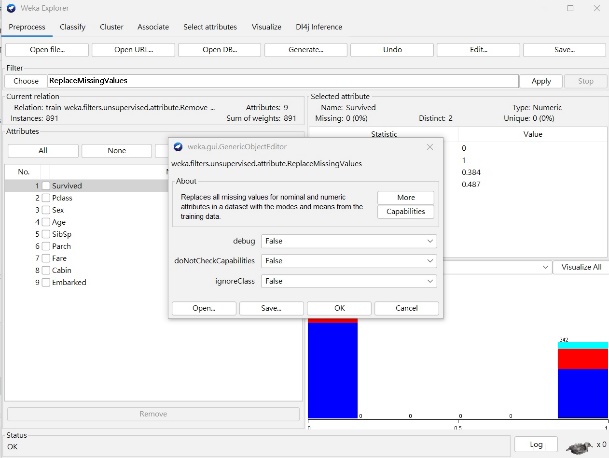
1.2 將 Cabin 的非空值以 1 替代，空值以 0填入（先用MergeManyValues 取代非空值，再用ReplaceMissingWithUserConstant 取代空值）



Step(1)處理前資料

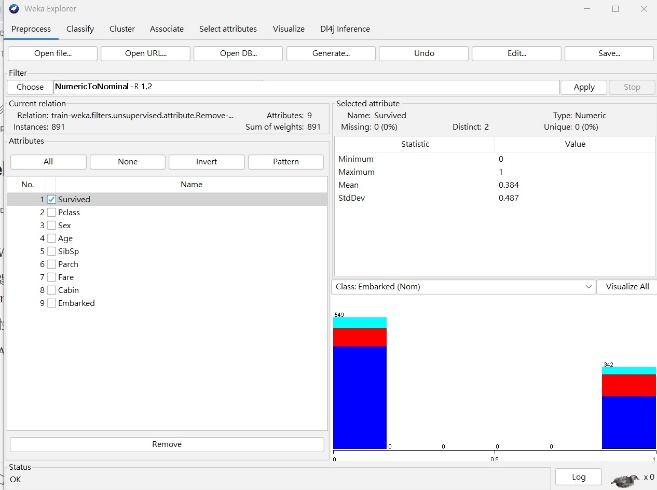
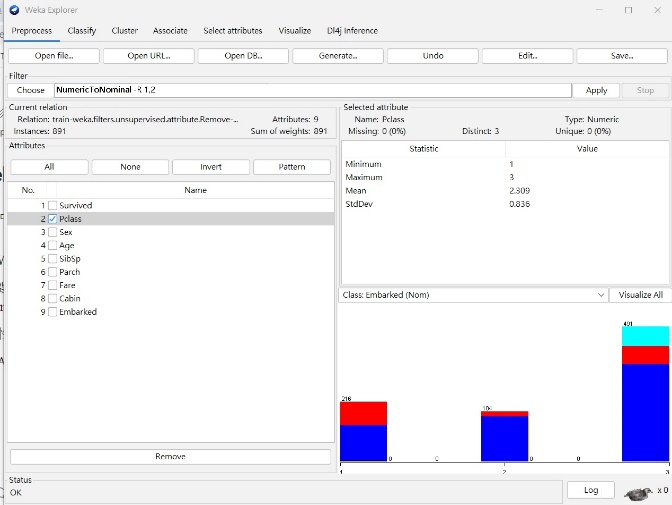
Step(2) 用MergeManyValues 取代非空值

Step(3) 再用ReplaceMissingWithUserConstant 取代空值

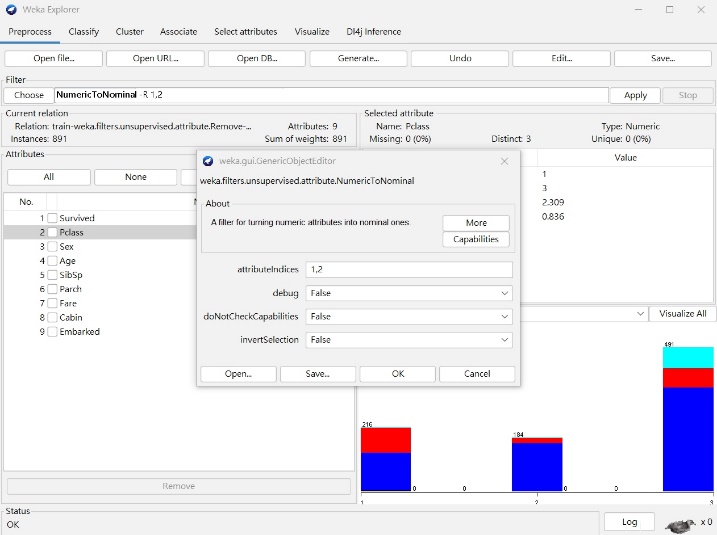
1.3 使用 ReplaceMissingValues 將 Age 的空值以 Age 平均數填入

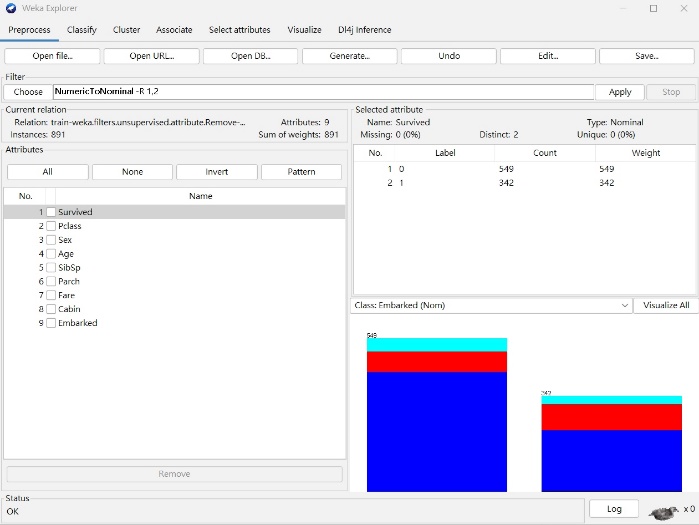
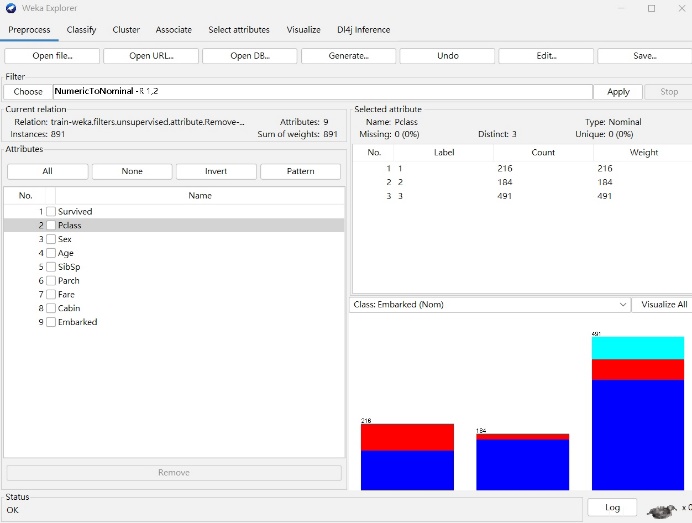
Step(1) 使用ReplaceMissingValues 將 Age 的空值以 Age 平均數填入

1.4 將 Survived 與 Pclass 轉為 Nominal，並說明為何 Numeric 無法使用在 Decision tree

Answer: Numeric為數字型連續型的變數，無法進行分類，如果要配合使用決策樹，需要進行離散化。

Step(1)處理前資料

Step(2) 將 Survived 與 Pclass 轉為 Nominal

step(3)處理後資料

1.5 以 70% 切割訓練資料，使用 J48 對 Survived 進行分類，並截圖分類 準確率、混淆矩陣及視覺化的Decision tree