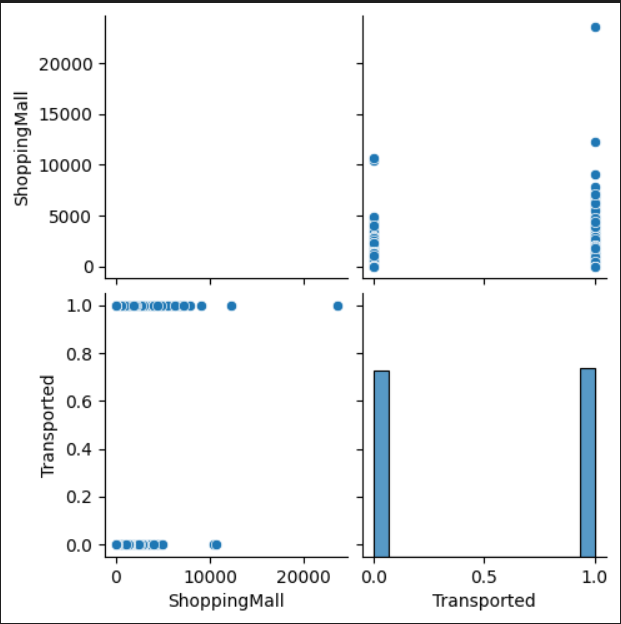
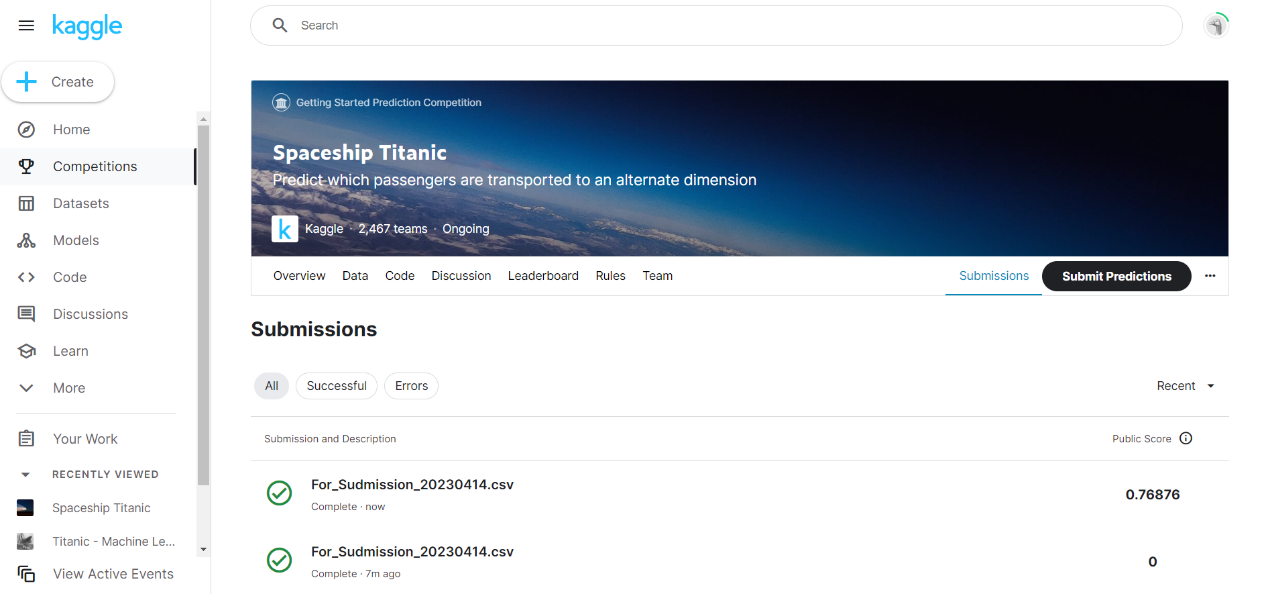
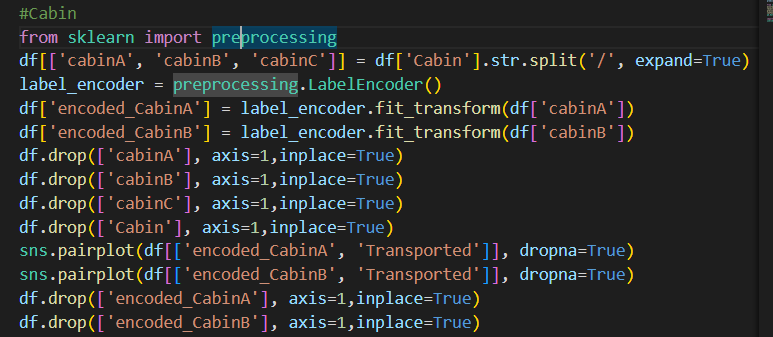
機器學習概論—期中作業心得

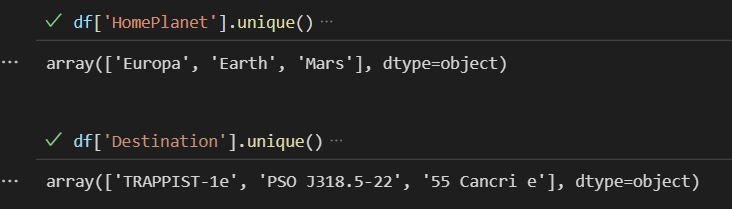
【108304021化材系–陳彥宇】

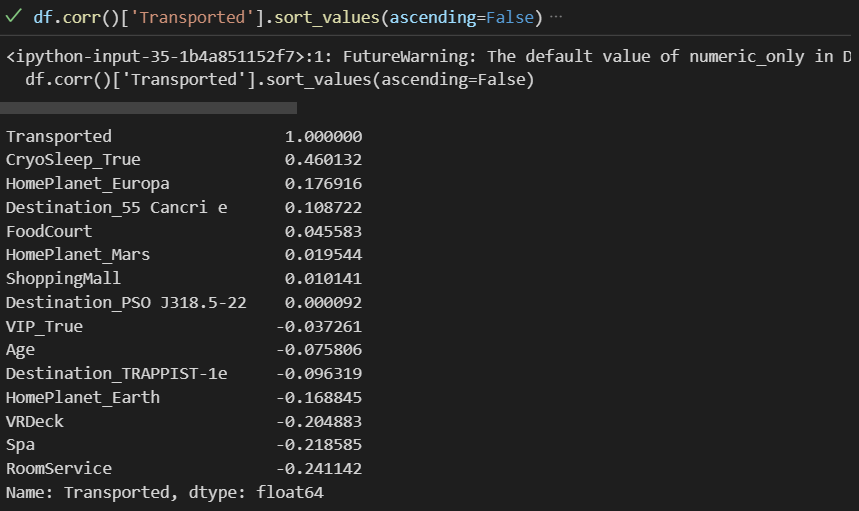
我是程式設計的新手，因為本身是化材(化工與材料)系學生，只學過一點C++，懂一些程式運作的必要條件之類的。機器學習的部分，是沒接觸過的領域，前陣子一個神奇的東西：「CHAT GPT」的出現，讓我對機器學習有初步的認識。但是在老師介紹機器學習的機制後，我了解到電腦其實和人類很像，都需要學習、練習，考試還不會全對XD，人類還是很厲害的！一堆考試滿分的卷哥卷姊阿。

這裡先放上在Kaggle上參加競賽的證明：這次我是從「Get Started」中選擇與老師上課已經做過的鐵達尼號生存預測類似的題目，只是這次要判斷哪些人會被傳送到異空間—當然是在遙遠的以後，不知道是否真的能實現？首先在資料整理的時候出現的瓶頸，不知道該從哪個資料開始整理。後來開始使用seaborn畫出關係圖之後才有了突破。以Shopping Mall為例，與結果「是否被傳送」沒有直接關係，因此可以刪除。

接著是補足缺失的資料，物件部分例如：Destination以及Cabin皆是補足數量最多的資料；數字部分例如Spa及Food Court皆是利用中位數補足；比較特殊的Age則是利用平均數補足。

然後是把Cabin拆成數部分分析，我求助同堂課的朋友，利用「string split」的方式把Cabin拆成三部分再利用seaborn畫出關係圖，發現其沒有特別關聯之後刪除此項。

 再來就是老師上課教過的「get dummies」部分，將物件(object)轉為數字(int)，在這裡我學會先用「unique」先找出資料有幾種分類，因為當資料太多就不能使用「get dummies」了。

最後是利用「correlation」檢查資料與結果的關聯性，為後把絕對值小於0.05的選項刪除。接著就是利用機器學習「sklearn」還有別人的模組練習、考試，初步完成第一階段。

第二階段皆是使用老師上課教授的方法做出，不知道老師有沒有發現我在Kaggle上有一次零分的紀錄XD，原來是因為上課範例鐵達尼號生存預測要求的是Survived得預測結果訊號「0/1」；這次作業要求出是否被傳送的「True/False」因為格式不對變成0分。

總體來說在自己的實作演練下我已經充分掌握老師上課教授的相關知識，相信近年來科技進步，多了解一些關於機器學習的觀念能夠對我的學習成長有很大幫助