**鐵達尼號分析**

1. **數據來源：**

Kaggle Titanic - Machine Learning from Disaster

<https://www.kaggle.com/competitions/titanic/data>

1. **Data資料夾：**

train.csv → MySQL、Python、R需要

test.csv() → Python、R需要

gender\_submission.csv → 用R預測存活率時需要

train\_forModeler.csv → SPSS modeler需要

1. **MySQL資料儲存：**

「Train.sql」儲存資料庫為titanic、資料表為train的資料，請以Data Import匯入。

1. **Python資料分析&圖表視覺化：**

以google colab環境編寫

<https://colab.research.google.com/drive/1vXjSgma-Ruw9uJAJAgxNAgqdnYxdgeoL?usp=sharing>

參考：<https://chtseng.wordpress.com/2017/12/24/kaggle-titanic%E5%80%96%E5%AD%98%E9%A0%90%E6%B8%AC-1/>

1. **SPSS modeler機器學習分析：**

「SPSS Modeler分析titanic.docx」有分類、關聯、分群分析的截圖

「SPSS Modeler」資料夾內有所有.str

1. **R語言資料分析&機器學習(使用Random Forest演算法)**

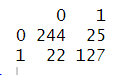
\*檔案在「R」資料夾內

「titanic.R」按照此檔案內的程式碼執行，可獲得執行結果。

「R分析titanic.docx」有分析結果的截圖。

→預測test(測試集)的存活率

成功率(每次執行結果皆會不同)：可高達(244+127)/(244+127+25+22)×100% = 88.76%

(Survived=0為死亡; Survived=0為存活)

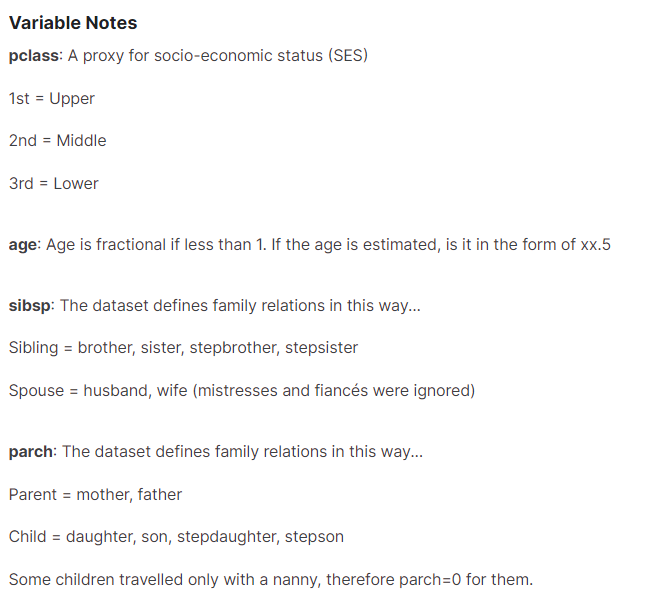
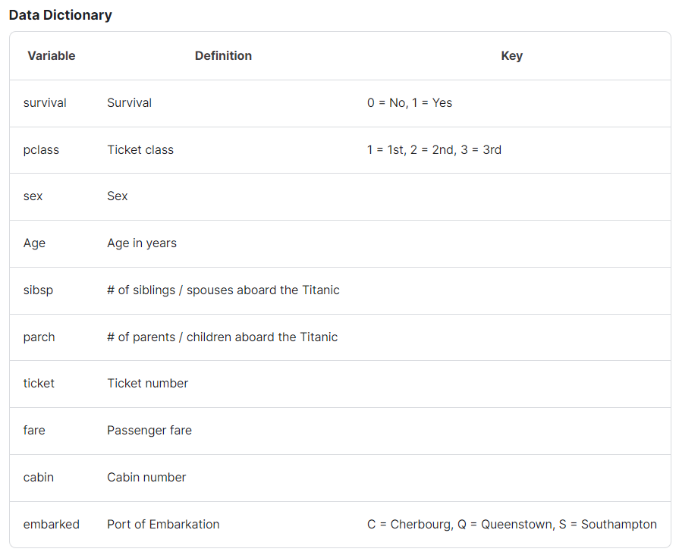
參考：

上集<https://ithelp.ithome.com.tw/articles/10232841>

中集<https://ithelp.ithome.com.tw/articles/10232911>

下集<https://ithelp.ithome.com.tw/articles/10232972>

1. **Kaggle Titanic資料集說明：**



鐵達尼號的沉沒是歷史上最著名的船隻失事之一。

1912 年 4 月 15 日，在處女航期間，被廣泛認為“永不沉沒”的皇家郵輪鐵達尼號在與冰山相撞後沉沒。 不幸的是，船上的每個人都沒有足夠的救生艇，導致 2224 名乘客和船員中有 1502 人死亡。

雖然倖存有一些運氣因素，但似乎有些人比其他人更有可能活下來。

在這個挑戰中，我們要求您建立一個預測模型來回答這個問題：“什麼樣的人更有可能生存？(哪些因素可能導致某人倖存或死亡？)” 使用乘客數據（即姓名、年齡、性別、社會經濟階層等）。

train.csv 將包含機上部分乘客的詳細信息（準確地說是 891 人），重要的是，將揭示他們是否倖存，也稱為“基本事實”。

test.csv 數據集包含類似的信息，但沒有透露每位乘客的“基本事實”。 預測這些結果是你的工作。

使用您在 train.csv 數據中找到的模式，預測機上其他 418 名乘客(在 test.csv 中找到)是否倖存。