問題導向之資料科學與機器學習應用\_Final Project

組員：N07041334許嘉倪、Na6084030曾子秦、H34064024 劉育瑄

**資料來源：**<https://www.kaggle.com/c/titanic>

**一、資料前處理：**1.將較用不到的欄位刪除(如Cabin等)

2.將空白欄位填入中位數

3.將Female轉換為1、male轉換為0

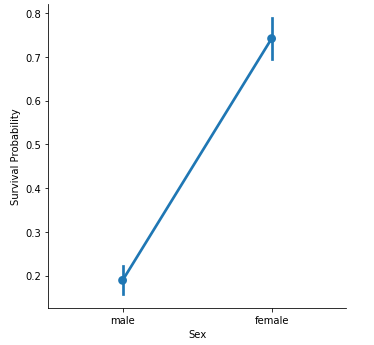
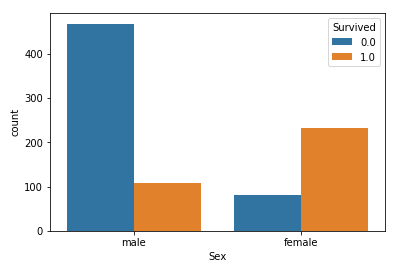
4.將年齡分為5個區隔:16歲(含)以下

16歲以上至32歲以下(含)

32歲以上至48歲以下(含)

48歲以上至64歲以下(含)

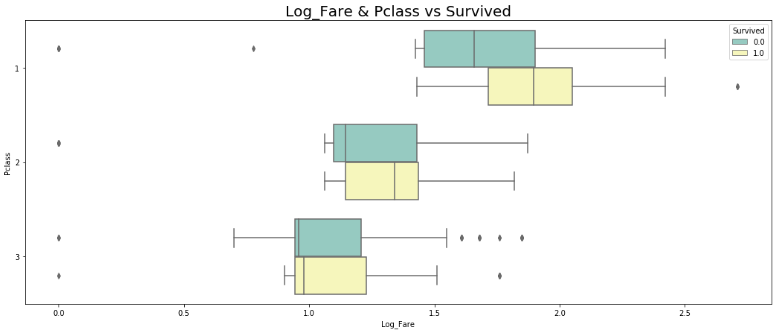
64歲以上

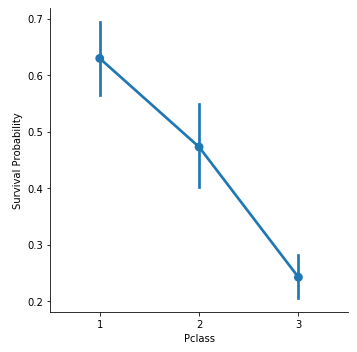
二、**資料分析(生還性別分析)：**

1. 由以上圖表中可知，女生生還者比男性多

2. 女性生還者機率高達75%，而男性僅有20%

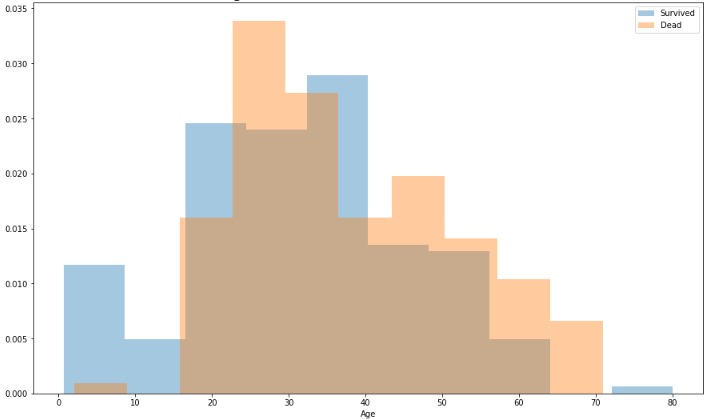
**三、資料分析(票價.艙等分析)：**

****

****

票價如同等於艙等等級，票價越高者，住的艙等越好，

能早點接收緊急的消息，也享有較優先逃生的權利，所以生還者較多。

**四、資料分析(年齡分析)：**

1. 由上圖中可知，生還者分佈在40歲以下，死亡者則在16歲以上至70歲以下。
2. 有0~8歲及72~80歲的生還者，且以20~40歲的生還者佔多數，

由此推測，小孩、老人可優先搭船逃生，而20~40歲的生還者有可能是帶著0~8歲的孩童，所以佔最多數。

**五、遇到的問題：**1.沒有準確率

2.決策樹後的圖無法輸出

**六、對這堂課的建議與希望學到的東西：**

1.希望可以在開課前先公告修課學生應具備之知識，或者調查上課學生的背景以製作教材。

2.希望老師的講義與助教操作實作的內容可以相符，降低落差。