**組員:113598065王丞頤 113598078 廖宥霖**

**題目:** NBA 的球員表現跟薪水和上場時間的關聯性分析 以及 NBA球員得分效率分析 並透過分群方式去評斷此球員是否為超值球員

**流程步驟:**

步驟1: 將球員得分跟薪水的資料做資料清洗 清除含空值row 並做資料的正規化

**薪水資料集** : (紅色為本次會用到的column)

1. playerName
2. seasonStartYear
3. salary
4. inflationAdjSalary

**球員得分資料集 :**

1. game\_id
2. teamName
3. playerName
4. MP
5. FG
6. FGA
7. 3P
8. 3PA
9. FT
10. FTA

步驟2:

1. 計算每個球員間的平均上場時間以及生涯中領到的薪水平均值
2. 透過FGA和3PA去計算二分球嘗試次數 (2PA)
3. 透過FG和3P計算二分球進球次數 (2P)
4. 透過FT 和 2P 和 3P去計算場均得分
5. 計算各別得分項的效率 (命中率)
6. 用關聯性套件如:Apriori 或 隨機森林等方法去計算各別命中率對上場時間和薪水的重要性特徵影響
7. 透過命中率跟上場時間做得分效率分析
8. 將資料切割成訓練跟測試集 去訓練分群 ex:給定測試集的球員上面的得分特徵去預測他的薪水區間 並與真實標籤去比較是否超值

**資料集:**[https://www.kaggle.com/datasets/patrickhallila1994/nba-data-from-basketball](https://www.kaggle.com/datasets/patrickhallila1994/nba-data-from-basketball-reference?select=salaries.csv)

**prefered\_time\_slot : 1/8**