期中企劃書：

《遊戲行為是否影響學生成績？》

班級：資工系4B 姓名：江采菁 學號：110021061

# 1. 資料來源：Kaggle 公開資料平台

本研究使用的資料集來源為 Kaggle，為全球最大的開放數據與機器學習競賽平台，擁有來自世界各地的高品質數據集。該資料集包含學生的遊戲行為（如遊戲時長、頻率、年資）、家庭背景（收入、父母教育程度）、以及學業表現（分數、等第）。

# 2. 選擇該資料的理由與應用

豐富變數：該資料集涵蓋行為面（遊戲）、學業表現面（成績）、背景面（家庭與教育）。

應用價值：

　- 探討學生娛樂與學習時間的平衡。

　- 分析特定變數是否具有干擾或緩衝作用（例如家庭背景）。

　- 作為教育政策或學習輔導的參考依據。

# 3. 預計使用的數據清理手法

　- 異常值處理：使用 IQR 或 Z-score 檢測遊戲時數是否有極端不合理值。

　- 缺失值處理：針對缺漏欄位做刪除或填補。

　- 資料型別轉換：將不合適的字串欄位轉為數值、日期或布林值。

# 4. 預計使用的資料視覺化工具與原因

|  |  |
| --- | --- |
| 視覺化類型 | 用途與理由 |
| 散佈圖 (Scatter Plot) | 觀察遊戲時長與成績是否有線性或非線性關聯 |
| 箱型圖 (Box Plot) | 比較不同遊戲頻率下的成績分布，尋找群體差異 |
| 熱力圖 (Heatmap) | 顯示各變數間的相關係數，幫助篩選顯著變數 |
| 直條圖/長條圖 (Bar Chart) | 呈現分類變數（例如不同收入層級）對成績的影響 |

# 5. 時間規劃甘特圖

|  |  |
| --- | --- |
| 週次 | 項目 |
| 第1週 | 資料集整理與欄位標註 |
| 第2週 | 進行數據清理與預處理 |
| 第3週 | 初步視覺化與關聯分析 |
| 第4週 | 建立簡單迴歸模型 |
| 第5週 | 敘述性分析與故事撰寫 |
| 第6週 | 統整報告、修訂與提交 |

# 6. 預期數據故事內容

分析完成後預期可產出以下敘事主軸：

　- 學生平均每日遊戲時數與成績呈現什麼關係？

　- 遊戲頻率過高是否真的對學業表現有害？

　- 家庭背景是否是遊戲行為與成績之間的調節因子？

　- 是否存在某些「高遊戲、高成績」的例外族群？

故事性內容將強調「不是所有玩遊戲的學生成績都差」，希望透過數據找出背後的真相與關聯。