===== 原始資料（未降維）分析 =====

Silhouette = 0.4540

===== PCA 降維分析 =====

PCA 1D → silhouette = 0.6118

PCA 2D → silhouette = 0.4540

===== 最佳 PCA 維度 =====

最佳維度：1D(silhouette = 0.6118)

你最後會得到：

複製

```
原始資料 silhouette score = S_raw
```

## 2️⃣ PCA 降維分析

你要測試多個降維版本（例如 PCA 1 維 / 2 維 / 3 維）。

但你的資料只有兩個欄位，所以最多只能做到 2 維：

- PCA 1 維
- PCA 2 維（其實 = 原始資料）

**PCA 過程（概念）：**

PCA 會做兩件事：

1. 找出資料最有變化方向（主成分）
2. 投影到較低維空間

## 3️⃣ PCA 之後重新做 K-means

每一個 PCA 維度，你都要重做一次 K-means 分群。

例如：

**PCA → 1 維**

- 原本（chinese, english）變成一個數字
- 用這個 1 維資料做 K-means
- 算 silhouette score → S1

- 算 silhouette score → S1

**PCA → 2 維**

- 其實跟原始一樣（因為本來就兩欄）
- 得到 silhouette score → S2

如果你有 3 維資料（例如三科成績），就會測試到 PCA=3。

---

## 4️⃣ 比較所有 silhouette score

你將得到：

| 降維維度 | silhouette score |
| --- | --- |
| 原始資料 | S_raw |
| PCA 1 維 | S1 |
| PCA 2 維 | S2 |

**選擇分數最高的那一個，那個就是最佳 PCA 維度。**

---

## ✔ 你的最後分析要回答三件事：

### (A) 原始資料（未降維）結果如何？

例如：

- K-means 後分群合理嗎？
- silhouette score 是多少？

### (B) 降維後效果有改善嗎？


從原本的 kmean 讀資料的程式下去改
先從原始資料（2d）且 k=3 去算 Silhouette

之後投影 2d 和 1d 去算降維後的資料分界度