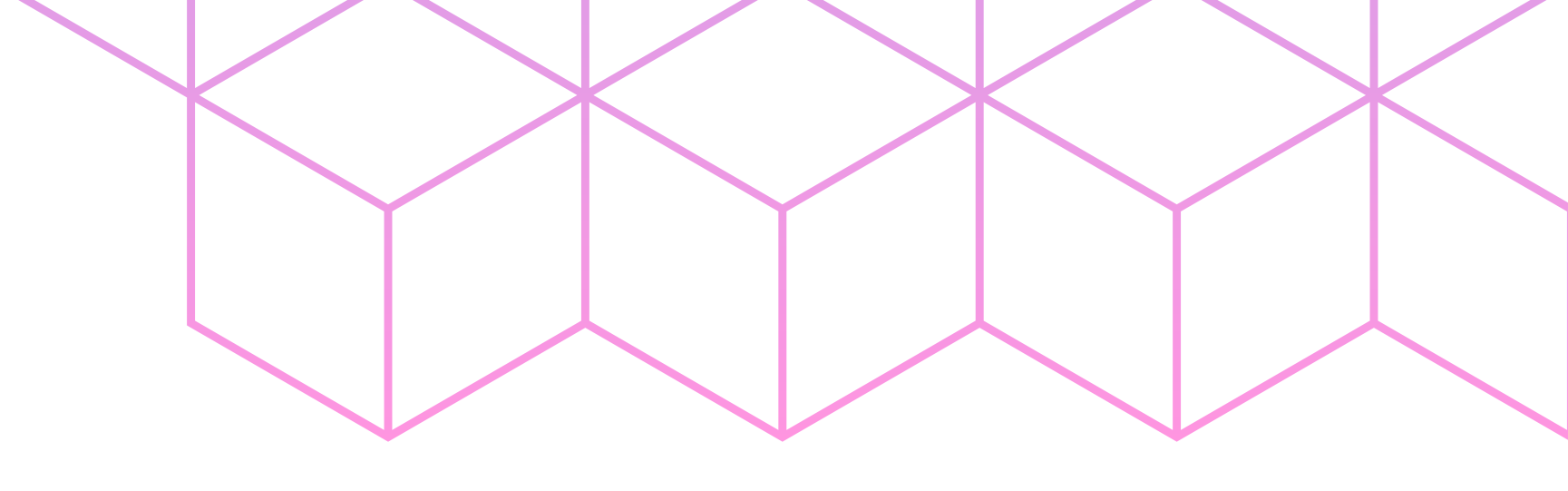


人工智慧小組報告

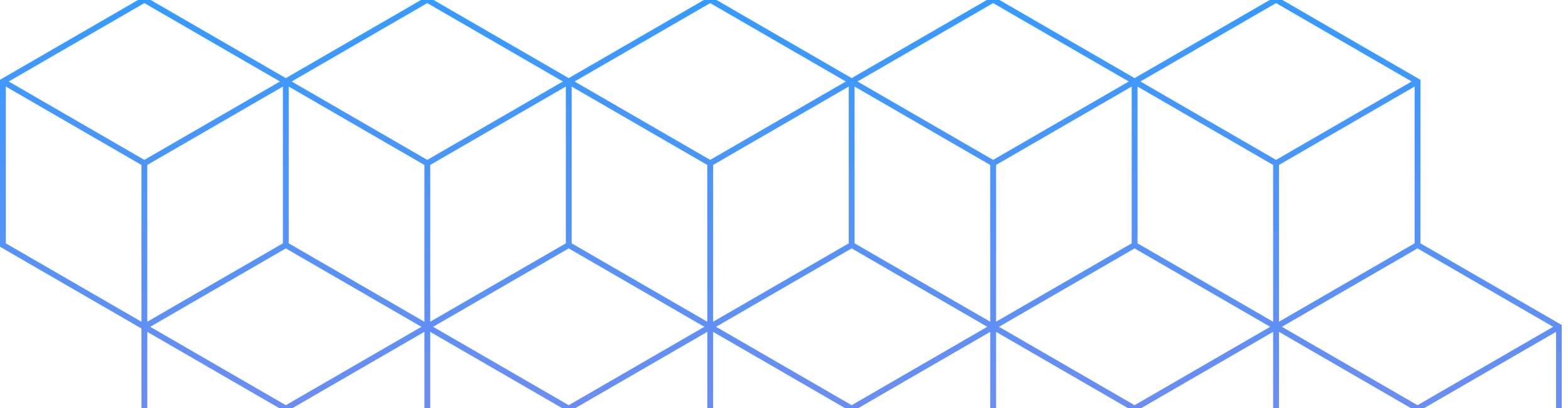
腎臟病患者分群分析

組長:陳鈺昕

組員:賴兆信, 黃子晏, 宋苡瑄, 黃天芸

- 
- 分析步驟
 - 離群值檢查與數值特徵處理
 - 分群前數值特徵標準化與 K-Means 分群

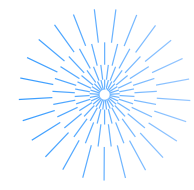
- 分群結果群組命名與醫療建議
- 時事結合
- 衛教方式
- 推薦健康講座



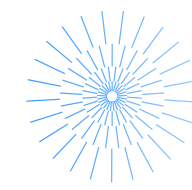
目錄

分工表

姓名	分工內容
陳鈺昕	根據分群結果圖表處理 上台報告
賴兆信	資料分群(包含數據說明) 上台報告
黃子晏	簡報製作 上台報告
宋苡瑄	簡報製作
黃天芸	簡報製作



分析步驟



01

讀取原始資料與前處理

02

數值特徵處理與離群值檢查

03

分群前數值特徵標準化與 K-Means 分群

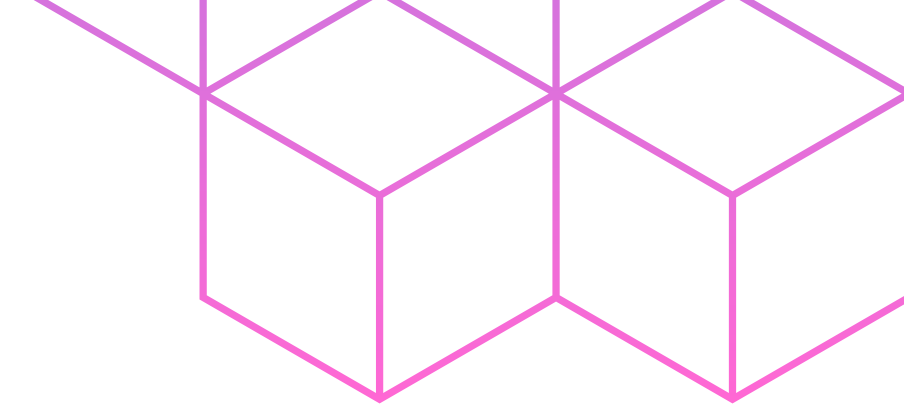
04

分群結果視覺化與描述統計

05

根據分群結果進行群組命名與醫療建議

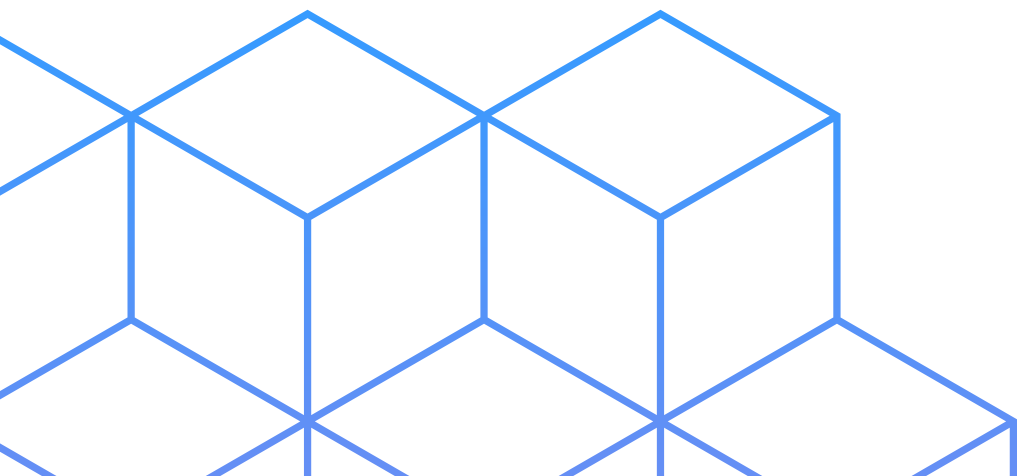
程式碼COLAB



- <https://drive.google.com/file/d/121g9FnufVDYUS8IGxv-avby1hKRYs3J4/view?usp=sharing>

分群敘述word

- <https://drive.google.com/file/d/121g9FnufVDYUS8IGxv-avby1hKRYs3J4/view?usp=sharing>



缺失值檢查與欄位型別

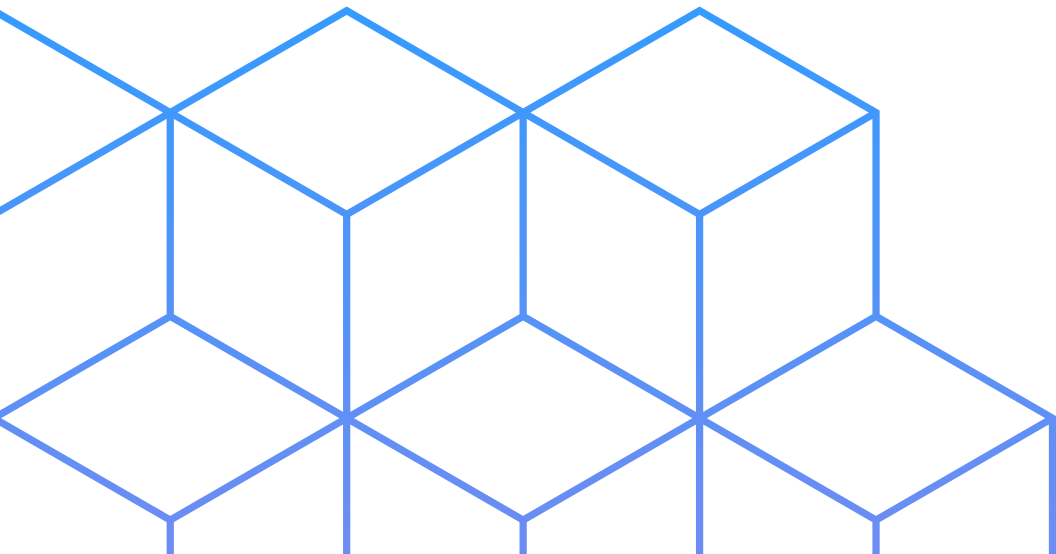
- 全部沒有缺失值
- 欄位資料型別都是數值

#	Column	Non-Null Count		Dtype
0	Age (yrs)	400	non-null	float64
1	Blood Pressure (mm/Hg)	400	non-null	float64
2	Specific Gravity	400	non-null	float64
3	Albumin	400	non-null	float64
4	Sugar	400	non-null	float64
5	Blood Glucose Random (mgs/dL)	400	non-null	float64
6	Blood Urea (mgs/dL)	400	non-null	float64
7	Serum Creatinine (mgs/dL)	400	non-null	float64
8	Sodium (mEq/L)	400	non-null	float64
9	Potassium (mEq/L)	400	non-null	float64
10	Hemoglobin (gms)	400	non-null	float64
11	Packed Cell Volume	400	non-null	float64
12	White Blood Cells (cells/cmm)	400	non-null	float64
13	Red Blood Cells (millions/cmm)	400	non-null	float64
14	Red Blood Cells: normal	400	non-null	float64
15	Pus Cells: normal	400	non-null	float64
16	Pus Cell Clumps: present	400	non-null	float64
17	Bacteria: present	400	non-null	float64
18	Hypertension: yes	400	non-null	float64
19	Diabetes Mellitus: yes	400	non-null	float64
20	Coronary Artery Disease: yes	400	non-null	float64
21	Appetite: poor	400	non-null	float64
22	Pedal Edema: yes	400	non-null	float64
23	Anemia: yes	400	non-null	float64
24	Chronic Kidney Disease: yes	400	non-null	float64

離群值檢查與數值特徵處理



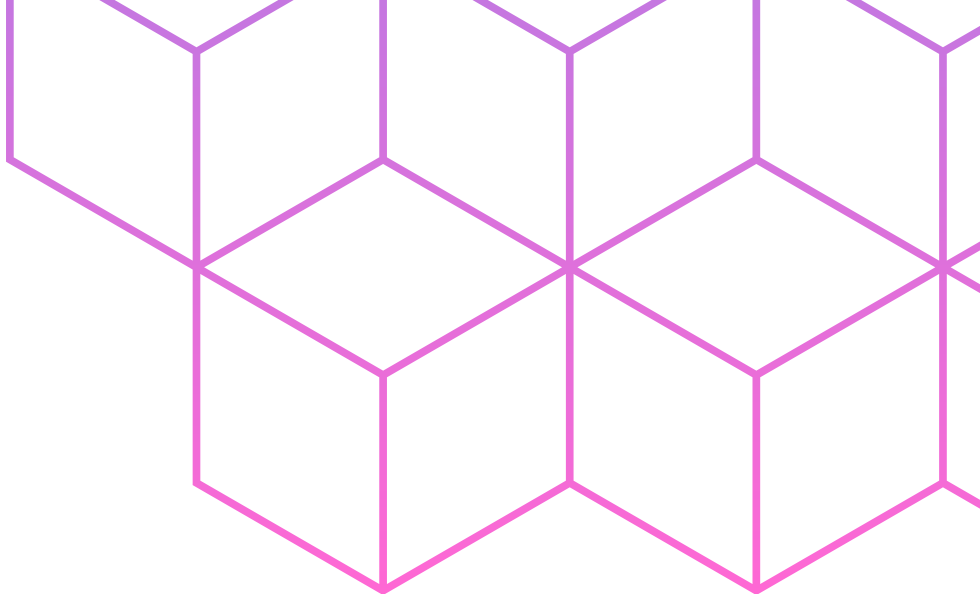
- 右圖為資料異常值圖表
- 在資料分群前找出異常的數值
- 利用超出設定的四分位距，將的資料寫入異常欄位




	Column	Outlier_Count	Outlier_Percent	Lower_Bound	Upper_Bound
21	Appetite: poor	82	0.2050	0.000000	0.000000
15	Pus Cells: normal	80	0.2000	1.000000	1.000000
22	Pedal Edema: yes	76	0.1900	0.000000	0.000000
14	Red Blood Cells: normal	68	0.1700	1.000000	1.000000
4	Sugar	61	0.1525	0.000000	0.000000
23	Anemia: yes	60	0.1500	0.000000	0.000000
7	Serum Creatinine (mgs/dL)	52	0.1300	-1.950000	5.650000
5	Blood Glucose Random (mgs/dL)	42	0.1050	11.125000	248.125000
16	Pus Cell Clumps: present	42	0.1050	0.000000	0.000000
6	Blood Urea (mgs/dL)	39	0.0975	-30.375000	122.625000
1	Blood Pressure (mm/Hg)	36	0.0900	55.000000	95.000000
20	Coronary Artery Disease: yes	34	0.0850	0.000000	0.000000
17	Bacteria: present	22	0.0550	0.000000	0.000000
8	Sodium (mEq/L)	18	0.0450	126.000000	150.000000
12	White Blood Cells (cells/cmm)	13	0.0325	2384.289679	13892.850536
0	Age (yrs)	7	0.0175	7.500000	99.500000
9	Potassium (mEq/L)	7	0.0175	2.400000	6.400000
2	Specific Gravity	7	0.0175	1.007500	1.027500
13	Red Blood Cells (millions/cmm)	1	0.0025	1.950000	7.150000
10	Hemoglobin (gms)	1	0.0025	4.071225	21.237265
11	Packed Cell Volume	1	0.0025	14.000000	62.000000
3	Albumin	0	0.0000	-3.000000	5.000000
19	Diabetes Mellitus: yes	0	0.0000	-1.500000	2.500000
18	Hypertension: yes	0	0.0000	-1.500000	2.500000

分群前標準化與K-Means 各群分數

- 先標準化數據（避免數值範圍影響分群）
- 右圖為將K=2~10帶入得出的WCSS數值

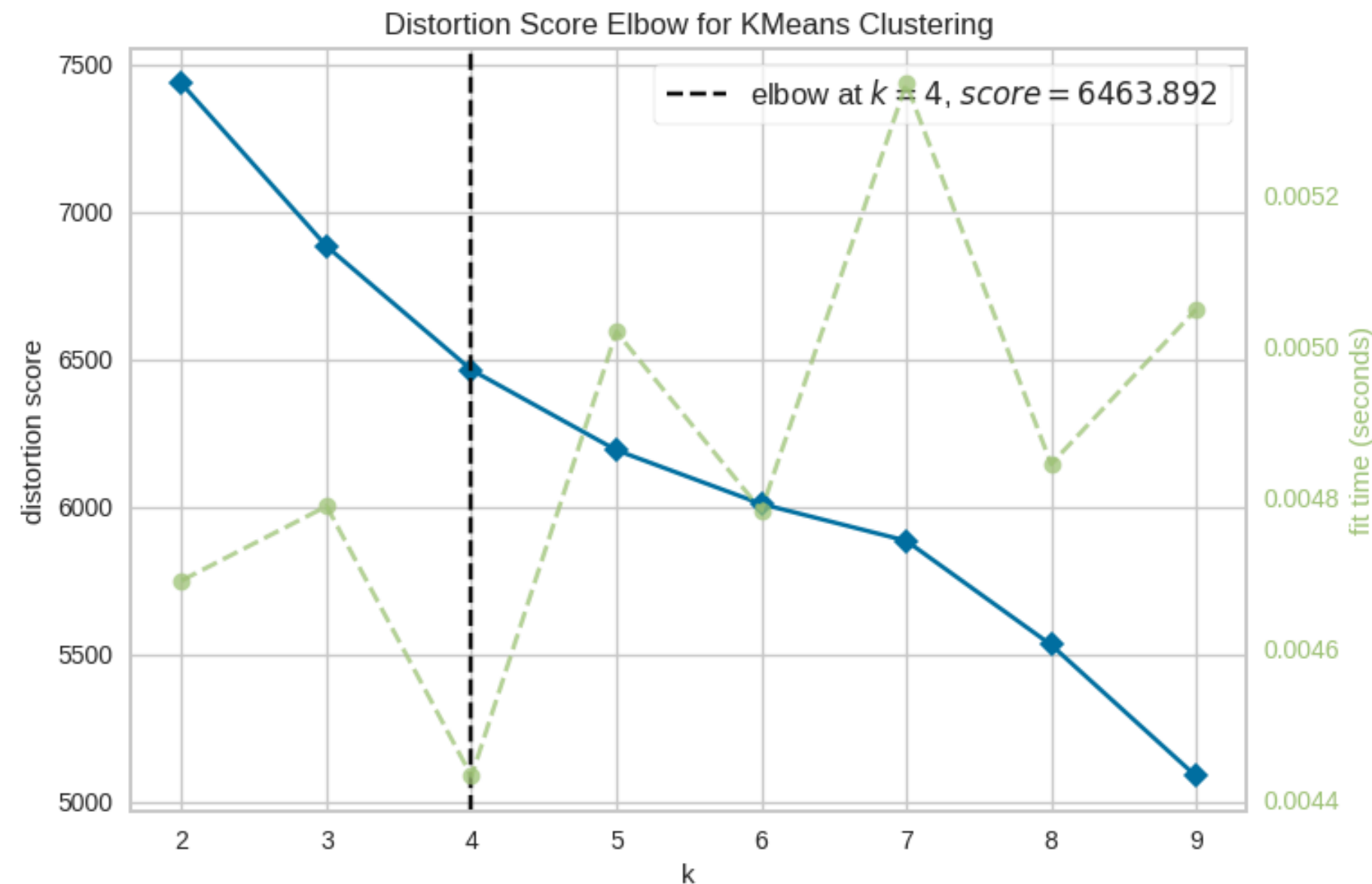


	K	WCSS
0	2	7435.220662
1	3	6860.894465
2	4	6468.437944
3	5	6075.639036
4	6	5757.877553
5	7	5423.967001
6	8	5139.660465
7	9	4925.395272
8	10	4760.639529


$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

決定分群K值

- 圖：左 Y 軸 distortion score
(失真數分)
- 圖：右 Y 軸 fit time (seconds)
(訓練耗時)
- 使用 Elbow Method 選擇最佳 k
- 執行 K-means 分群



分群結果視覺化與描述統計

各群平均數值與中位數

目的：計算並展示各群的平均數值與中位數，了解群組基本特徵。

- Cluster 0 為中老年且腎功能有中度異常的患者群，風險較高。
- Cluster 1 為低風險年輕健康群，腎功能正常，預防為主。
- Cluster 2 為極高風險群，患者腎功能極差，代謝指標嚴重異常。
- Cluster 3 為中度風險群，雖全數 CKD，但損害程度介於 Cluster 0 與 Cluster 2 之間。

各群平均數據：

Cluster	Age (yrs)	Blood Pressure (mm/Hg)	Specific Gravity	Albumin \
0	57.477273	78.409091	1.012841	2.590909
1	44.315633	72.105263	1.020921	0.178947
2	55.638298	83.404255	1.015846	1.815290
3	59.432596	79.411765	1.014244	1.184874

Cluster	Sugar	Blood Glucose Random (mgs/dL)	Blood Urea (mgs/dL) \
0	0.818182	186.053659	75.036364
1	0.031579	109.318210	32.298397
2	0.446809	159.777648	141.042553
3	0.798319	184.336124	55.928180

Cluster	Serum Creatinine (mgs/dL)	Sodium (mEq/L)	Potassium (mEq/L)	...	\
0	3.547727	133.638916	4.424750
1	1.084773	141.052088	4.265196
2	11.132979	130.074468	6.416130
3	2.720269	135.812058	4.378848

Cluster	Red Blood Cells: normal	Pus Cells: normal	Pus Cell Clumps: present	\
0	0.636364	0.090909	0.840909	
1	0.973684	0.989474	0.000000	
2	0.595745	0.553191	0.042553	
3	0.764706	0.857143	0.025210	

分群結果視覺化與描述統計

用統計檢定 p 對各組檢查

目的：檢查各群在不同變數上的差異是否顯著。

- 依群別收集每個變數的數據，若 $p < 0.05$ ，輸出顯著性結果。

Age (yrs): 差異顯著 ($p=0.0000$)
Blood Pressure (mm/Hg): 差異顯著 ($p=0.0000$)
Specific Gravity: 差異顯著 ($p=0.0000$)
Albumin: 差異顯著 ($p=0.0000$)
Sugar: 差異顯著 ($p=0.0000$)
Blood Glucose Random (mgs/dL): 差異顯著 ($p=0.0000$)
Blood Urea (mgs/dL): 差異顯著 ($p=0.0000$)
Serum Creatinine (mgs/dL): 差異顯著 ($p=0.0000$)
Sodium (mEq/L): 差異顯著 ($p=0.0000$)
Potassium (mEq/L): 差異顯著 ($p=0.0000$)
Hemoglobin (gms): 差異顯著 ($p=0.0000$)
Packed Cell Volume: 差異顯著 ($p=0.0000$)
White Blood Cells (cells/cmm): 差異顯著 ($p=0.0000$)
Red Blood Cells (millions/cmm): 差異顯著 ($p=0.0000$)
Red Blood Cells: normal: 差異顯著 ($p=0.0000$)
Pus Cells: normal: 差異顯著 ($p=0.0000$)
Pus Cell Clumps: present: 差異顯著 ($p=0.0000$)
Bacteria: present: 差異顯著 ($p=0.0000$)
Hypertension: yes: 差異顯著 ($p=0.0000$)
Diabetes Mellitus: yes: 差異顯著 ($p=0.0000$)
Coronary Artery Disease: yes: 差異顯著 ($p=0.0000$)
Appetite: poor: 差異顯著 ($p=0.0000$)
Pedal Edema: yes: 差異顯著 ($p=0.0000$)
Anemia: yes: 差異顯著 ($p=0.0000$)

分群結果視覺化與描述統計

對比個群的差異

目的：針對每個群別，找出變異性最大的 top_k 特徵並生成文字描述。

總結說明

- 各群在主要臨床與實驗室指標上均存在顯著差異，這說明分群方法能夠有效區分不同健康狀態的患者：
- Cluster 0 顯示中高風險特徵，提示需要進一步的監控與干預。
- Cluster 1 表現出相對理想的健康狀態，屬於低風險群。
- Cluster 2 則顯示極高風險，尤其在腎功能方面可能存在嚴重問題。
- Cluster 3 屬於中等風險，需要進一步評估與持續追蹤。

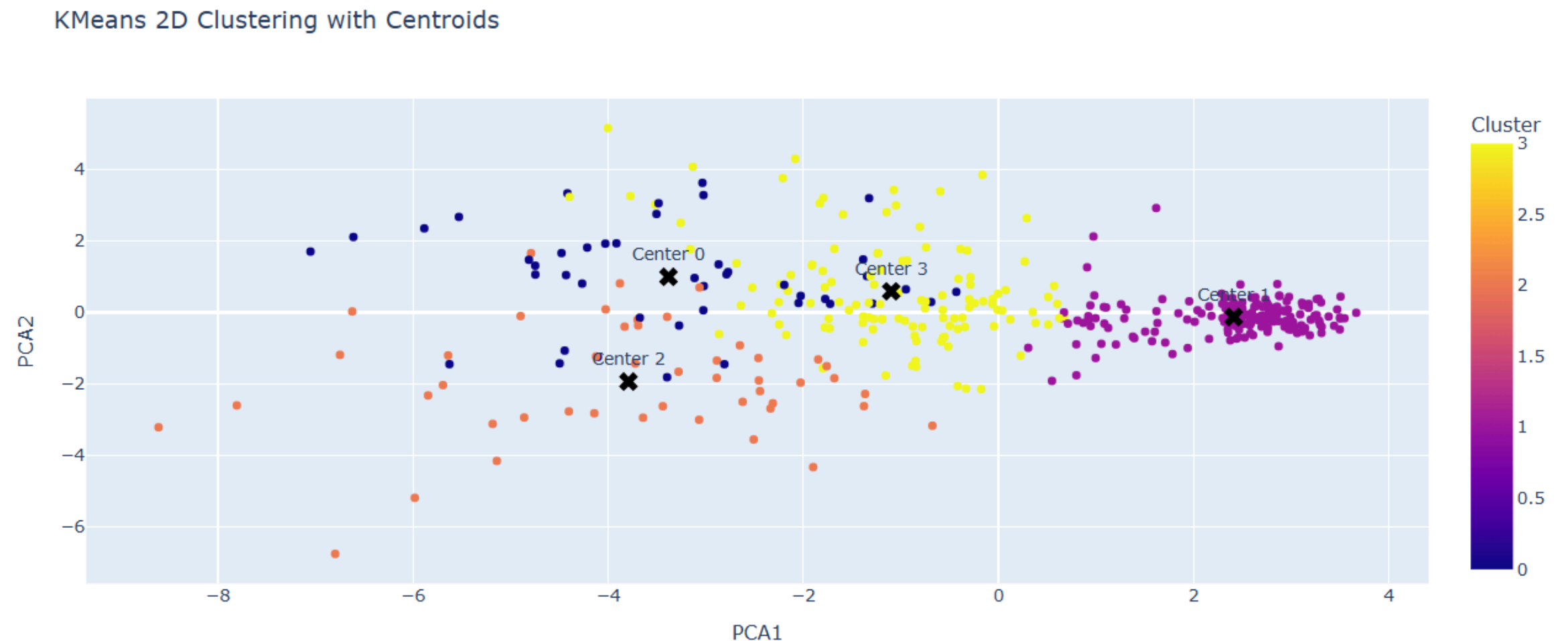
Cluster 0	特徵明顯變化:	White Blood Cells (cells/cmm)	均值 9910.70,	Blood Urea (mgs/dL)	均值 75.04,	Blood Glucose Random (mgs/dL)	均值 186.05,	Packed Cell Volume	均值 31.41,	Age (yrs)	均值 57.48
Cluster 1	特徵明顯變化:	White Blood Cells (cells/cmm)	均值 7770.14,	Blood Urea (mgs/dL)	均值 32.30,	Blood Glucose Random (mgs/dL)	均值 109.32,	Packed Cell Volume	均值 44.98,	Age (yrs)	均值 44.32
Cluster 2	特徵明顯變化:	White Blood Cells (cells/cmm)	均值 7695.01,	Blood Urea (mgs/dL)	均值 141.04,	Blood Glucose Random (mgs/dL)	均值 159.78,	Packed Cell Volume	均值 25.89,	Age (yrs)	均值 55.64
Cluster 3	特徵明顯變化:	White Blood Cells (cells/cmm)	均值 8979.57,	Blood Urea (mgs/dL)	均值 55.93,	Blood Glucose Random (mgs/dL)	均值 184.34,	Packed Cell Volume	均值 35.32,	Age (yrs)	均值 59.43

分群結果視覺化與描述統計

PCA 二維視覺化

目的：計算並展示各群的平均數值與中位數，了解群組基本特徵。

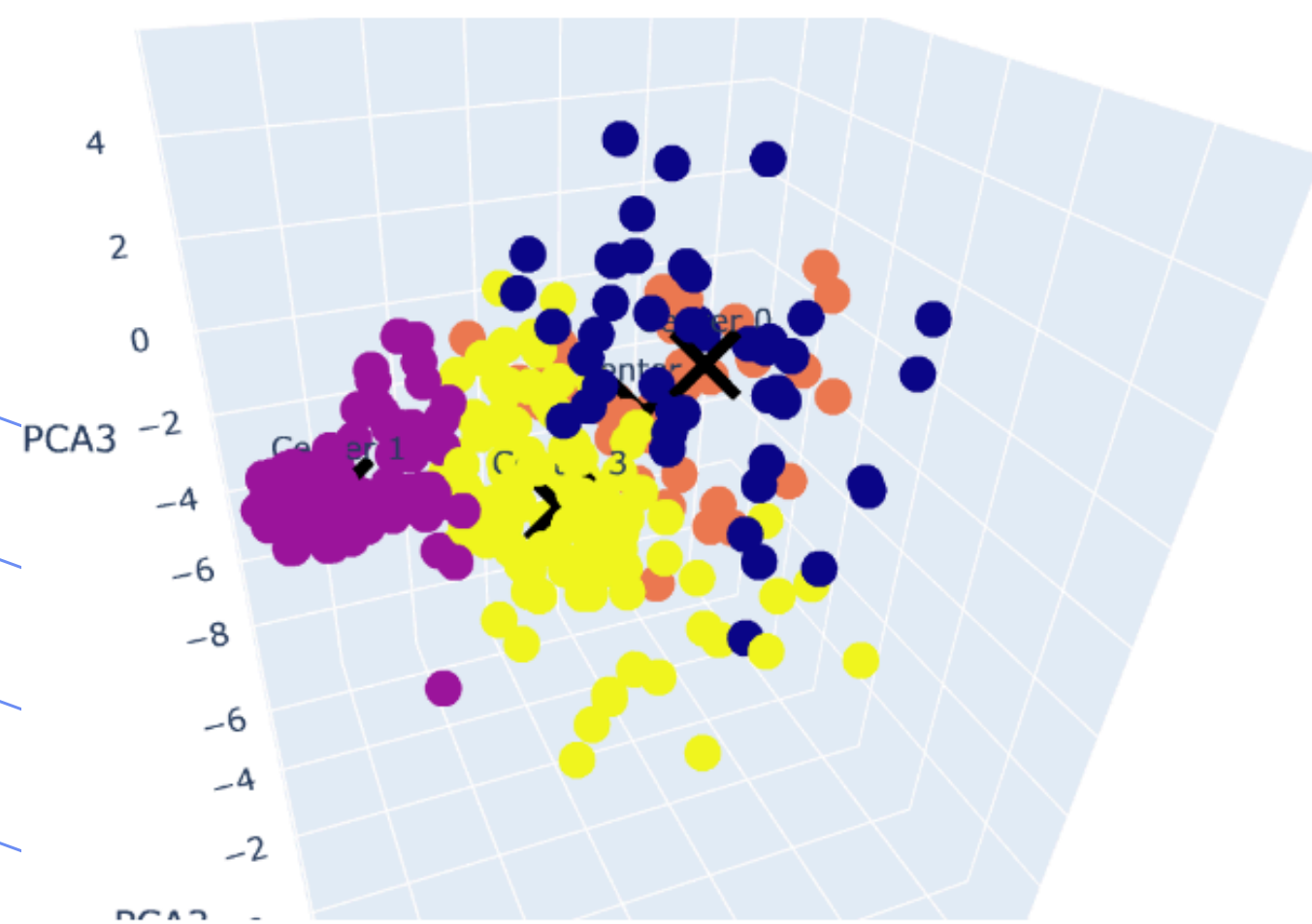
- 各群之間沒有明顯重疊，說明 K-means 分群結果能夠有效區分不同的患者群體。
- 以黑色“x”標記顯示了各群的群中心，證實了分群結果的合理性



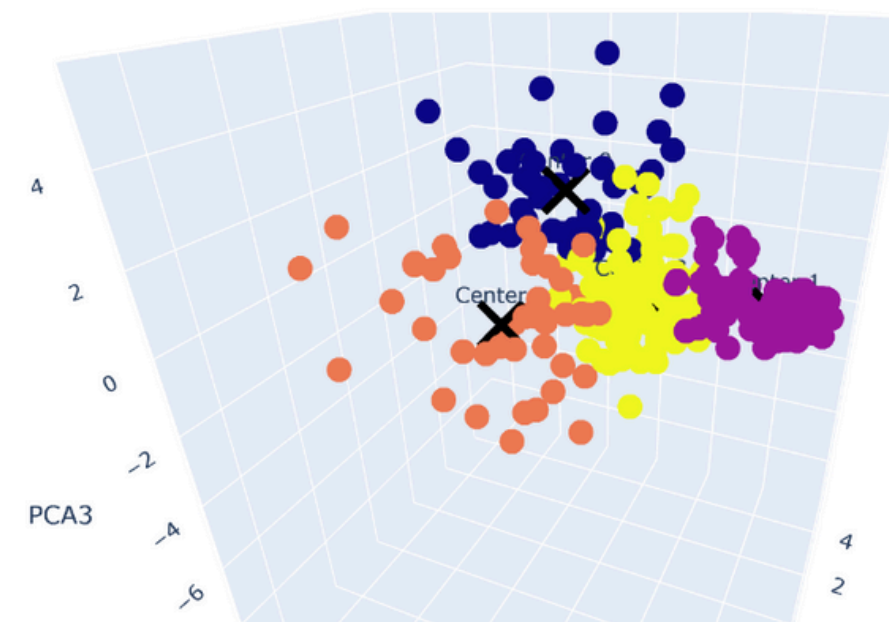
分群結果視覺化與描述統計

PCA 三維視覺化

目的：進一步將數據降維到 3 維，生成三維散點圖，方便觀察群組分佈及群中心位置。



Means 3D Clustering with Centroids



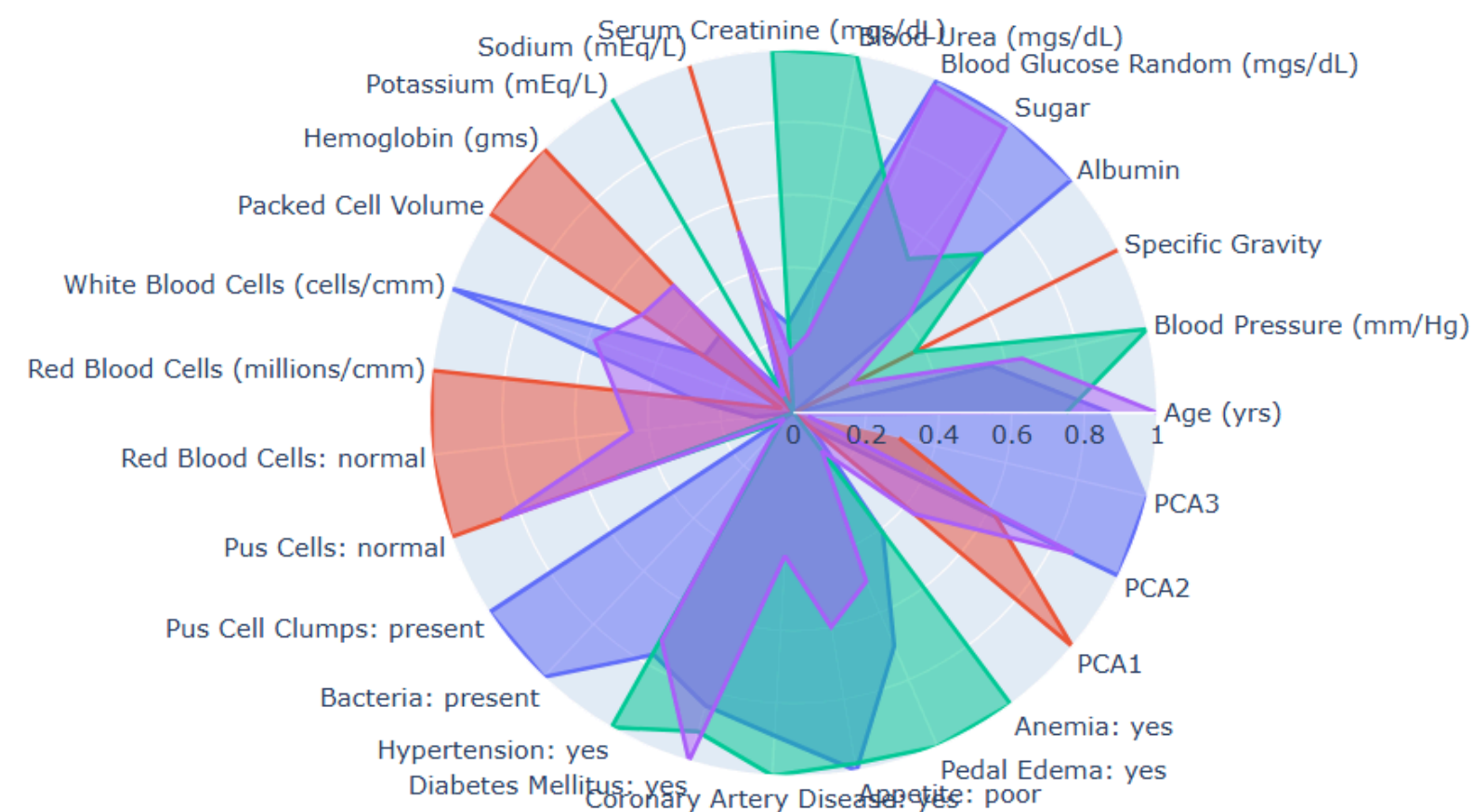
分群結果視覺化與描述統計

各群核心特徵雷達圖（標準化後）

目的：以雷達圖展示各群在多維特徵上的表現，便於直觀比較。

各群核心特徵雷達圖（標準化後）

- 各群標準化後雷達表現，視覺化 代謝、血液、感染、慢病負荷 四大系統強度
- 雷達圖 高度可視化了疾病進程的「能量場」與死亡熱區
- Cluster 2 的全面爆表，是系統代謝的崩潰實像
- 不同群體在不同生理系統中呈現典型分層，這是數據對應生理病理的直接證明



分群結果視覺化與描述統計

分群後引入 CKD 欄位並進行腎臟病比例分析

目的：在分群結果中引入 Chronic Kidney Disease: yes 欄位，並計算各群中腎臟病患者的比例。

各群腎臟病患者比例 (Yes/No):

Chronic Kidney Disease: yes	0.0	1.0
Cluster		
0	NaN	1.000000
1	0.789474	0.210526
2	NaN	1.000000
3	NaN	1.000000

分群結果視覺化與描述統計

群體輪廓表

目的：針對關鍵醫療變數生成各群的平均數，並將部分百分比數據格式化顯示，方便後續比較。

	Cluster	Creatinine	Blood Urea	Potassium	Anemia	Diabetes Mellitus (%)	\
0	0	3.547727	75.036364	4.424750	0.295455	56.82%	
1	1	1.084773	32.298397	4.265196	0.010526	1.58%	
2	2	11.132979	141.042553	6.416130	0.702128	61.7%	
3	3	2.720269	55.928180	4.378848	0.100840	67.23%	
		Hypertension (%)					
0		65.91%					
1		2.11%					
2		85.11%					
3		62.18%					

分群結果視覺化與描述統計

群體輪廓表

目的：針對關鍵醫療變數生成各群的平均數，並將部分百分比數據格式化顯示，方便後續比較。

Cluster 0

Creatinine：平均值約 3.55 mgs/dL，顯示腎功能有中度受損的可能性。

Blood Urea：平均約 75.04 mgs/dL，提示腎臟排毒功能可能異常。

Potassium：平均約 4.42 mEq/L，屬於正常範圍，但仍需注意電解質平衡。

Anemia：平均值約 0.30，表示約 30% 的患者存在貧血現象。

Diabetes Mellitus (%)：約 56.82%，說明該群超過一半的患者有糖尿病病史或糖尿病風險。

Hypertension (%)：約 65.91%，顯示高血壓的比例也相當高。

解讀：Cluster 0 的患者群中，腎功能指標（Creatinine 與 Blood Urea）異常，且糖尿病與高血壓比例較高，提示此群患者可能存在多重代謝及心血管風險，需針對腎功能與慢性疾病加強管理。

分群結果視覺化與描述統計

群體輪廓表

目的：針對關鍵醫療變數生成各群的平均數，並將部分百分比數據格式化顯示，方便後續比較。

Cluster 1

Creatinine：平均約 1.08 mgs/dL，屬於正常範圍，顯示腎功能較好。

Blood Urea：平均約 32.30 mgs/dL，較低，支持腎功能正常的判斷。

Potassium：平均約 4.27 mEq/L，正常範圍內。

Anemia：極低約 0.01，幾乎無貧血現象。

Diabetes Mellitus (%)：僅 1.58%，幾乎可以排除糖尿病風險。

Hypertension (%)：僅 2.11%，高血壓比例極低。

解讀：Cluster 1 的各項指標均顯示正常或接近理想狀態，無論是腎功能、電解質、貧血，或是慢性疾病（糖尿病、高血壓）的比例都非常低，表明該群屬於低風險、健康狀態良好的患者群。

分群結果視覺化與描述統計

群體輪廓表

目的：針對關鍵醫療變數生成各群的平均數，並將部分百分比數據格式化顯示，方便後續比較。

Cluster 2

Creatinine：平均約 11.13 mgs/dL，顯著偏高，顯示腎功能嚴重受損。

Blood Urea：平均約 141.04 mgs/dL，也明顯偏高，與腎功能損害相符。

Potassium：平均約 6.42 mEq/L，顯著高於正常值，可能導致心律不整等併發症。

Anemia：平均約 0.70，貧血情形明顯，約 70% 患者有貧血問題。

Diabetes Mellitus (%)：約 61.7%，糖尿病比例偏高。

Hypertension (%)：約 85.11%，顯示高血壓極為普遍。

解讀：Cluster 2 患者的腎功能極差（Creatinine 與 Blood Urea 明顯升高），伴有明顯的電解質失衡（高鉀）、高比例貧血、糖尿病與高血壓，該群屬於極高風險群，需密切監控並積極進行治療干預。

分群結果視覺化與描述統計

群體輪廓表

目的：針對關鍵醫療變數生成各群的平均數，並將部分百分比數據格式化顯示，方便後續比較。

Cluster 3

Creatinine：平均約 2.72 mgs/dL，略高於正常但低於 Cluster 0 與 2。

Blood Urea：平均約 55.93 mgs/dL，較 Cluster 0 與 Cluster 2 低。

Potassium：平均約 4.38 mEq/L，接近正常範圍。

Anemia：平均約 0.10，貧血情形輕微。

Diabetes Mellitus (%)：約 67.23%，糖尿病比例偏高。

Hypertension (%)：約 62.18%，高血壓比例中等。

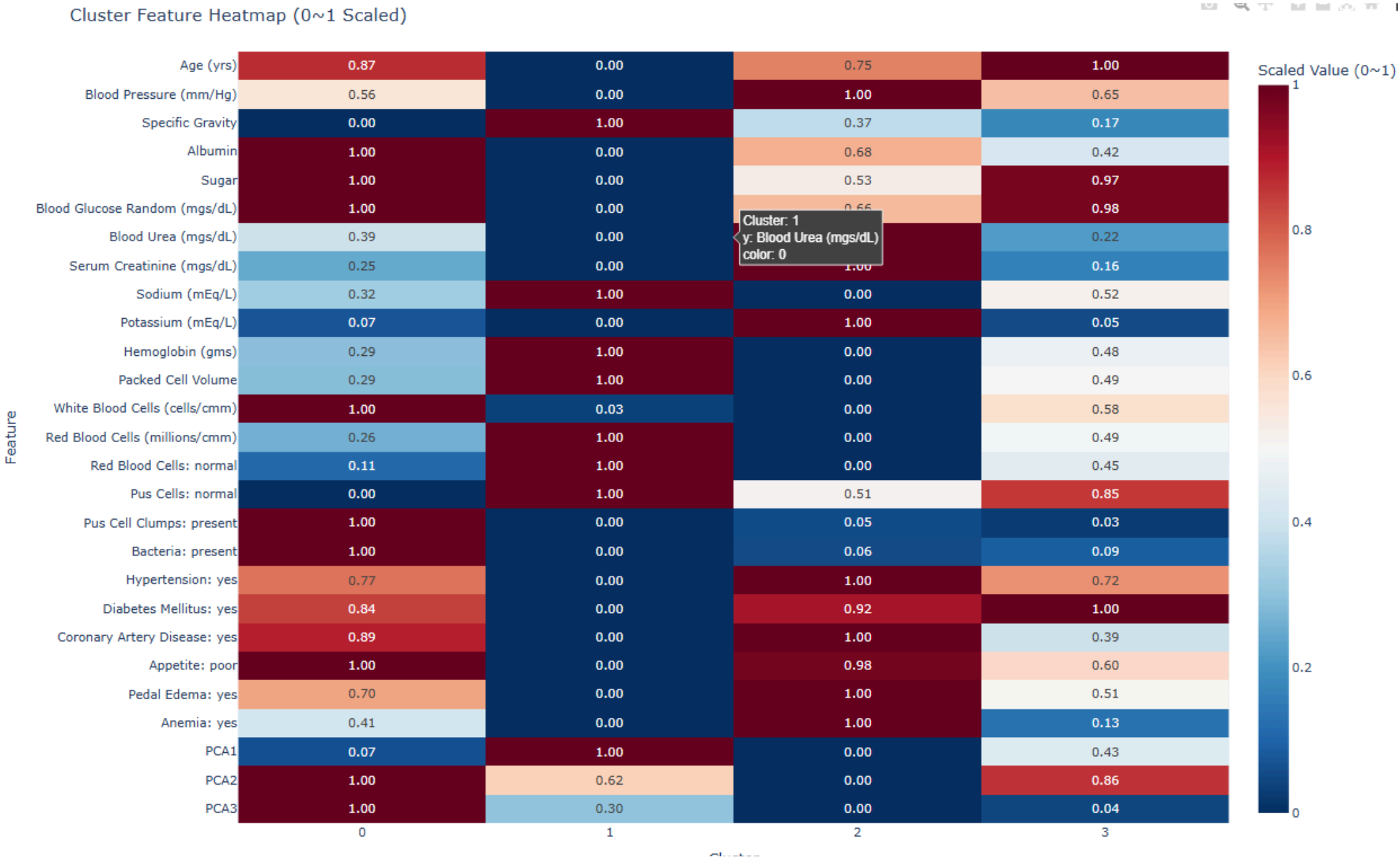
解讀：Cluster 3 的腎功能指標輕度異常（Creatinine 略高），但 Blood Urea 較低，電解質與貧血狀況較好；然而糖尿病與高血壓的比例依然偏高，提示該群可能存在代謝與心血管風險，但相較於 Cluster 2，風險程度較低，屬於中等風險群。

分群結果視覺化與描述統計

特徵差異熱圖

目的：將各群特徵數據進行標準化後以熱圖展示，便於比較各群差異。

- Cluster 2 血紅熱區爆發，全面指向代謝崩潰
- BUN、Creatinine、Potassium、Anemia 全面暴紅
- 明確呈現 CKD 病人死亡的主路徑：高鉀 + 高尿毒 + 貧血



分群結果視覺化與描述統計

風險分數計算與風險排名

目的：根據多個關鍵變數計算每個患者的風險分數，並生成風險排名與各群風險統計。

- 公式：

$$\text{Risk_Score} = 0.4 * \text{Creatinine} + 0.3 * \text{BUN} + 0.2 * \text{Potassium} + 0.1 * \text{Anemia}$$

- 解讀：

Cluster 2 的風險分數 48.1 (極高)

Cluster 1 僅 10.97 (低風險)

風險分數直接對應 生理死亡風險累積過程

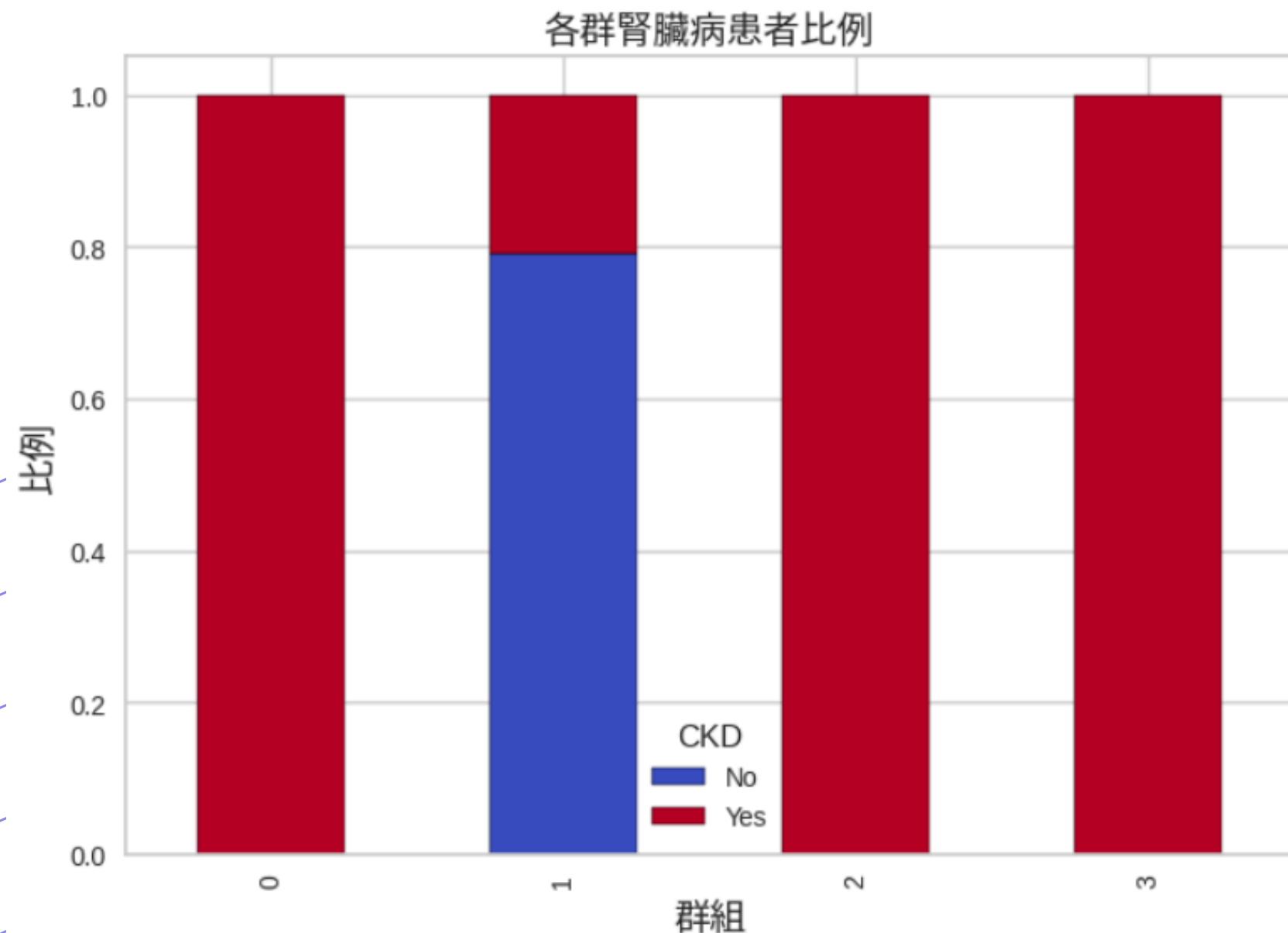
【各 Cluster 風險輪廓】					
	cluster	mean	max	min	count
0	0	24.844495	69.994376	6.64	44
1	1	10.977520	35.260000	3.93	190
2	2	48.119396	137.900000	6.82	47
3	3	18.752415	57.640000	5.54	119

【風險排名前10】			
	Risk_Score	Serum Creatinine (mgs/dL)	Blood Urea (mgs/dL) \
0	137.900000	32.0	391.0
1	102.860000	13.0	322.0
2	99.420000	13.3	309.0
3	85.380000	76.0	180.0
4	78.720000	13.4	241.0
5	76.960000	14.2	235.0
6	75.540000	18.1	223.0
7	71.820000	15.2	215.0
8	71.440000	12.2	219.0
9	69.994376	10.2	217.0

分群結果進行群組命名與醫療建議

群組命名

目的：依據各群統計特徵（如 CKD 比例、年齡、血壓等）給出適當的群組命名。



- 0: 重度腎臟病患者群
- 1: 低風險年輕健康群
- 2: 高血壓重度腎病群
- 3: 老年慢性腎病群

分群結果進行群組命名與醫療建議

醫療建議與衛教方式

- 依照命名後的分組提出醫療建議與衛教方式。

重度腎臟病患者群（群組0）：

1. 嚴格限制蛋白質攝入，採用低鹽、低蛋白飲食，以減輕腎臟負擔。
2. 定期監測腎功能指標（如肌酐、尿素及電解質），以便早期發現病情惡化。
3. 與腎臟科醫師密切合作，評估是否需要進一步調整治療方案，並考慮使用腎保護藥物。
4. 加強健康衛教，提醒患者戒除不良嗜好，控制血糖與血壓，維持良好的生活習慣。

低風險年輕健康群（群組1）：

1. 維持均衡飲食和規律運動，保持理想體重，促進心血管及腎功能健康。
2. 定期進行健康檢查，包括血壓、血糖和腎功能評估，以預防未來慢性疾病發生。
3. 加強衛教，建立健康生活模式，避免熬夜和吸菸等不良習慣。

分群結果進行群組命名與醫療建議

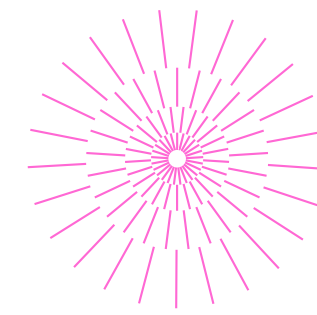
醫療建議與衛教方式

高血壓重度腎病群（群組2）：

1. 必須嚴格控制血壓，可能需要使用 ACEI 或 ARB 類藥物來保護腎臟功能。
2. 採用低鹽、低蛋白飲食，並密切監控血尿素、肌酐及鉀含量，防止電解質失衡。
3. 定期接受腎功能檢查，根據檢查結果及時調整治療方案，並積極進行營養諮詢。
4. 衛教上需特別強調藥物服用的重要性及日常生活管理，避免過度勞累。

老年慢性腎病群（群組3）：

1. 定期進行全面健康檢查，特別是腎功能、血壓與血糖的監測，及早發現及治療共病。
2. 採用個別化的低鹽、低蛋白飲食，並在營養師指導下改善飲食結構，確保營養均衡。
3. 鼓勵適度運動，如太極、散步等，以促進血液循環和增強體能，但避免過度負荷。
4. 衛教上要提醒患者注意正確服藥，避免藥物間的不良互動，並保持良好的生活作息。



時事結合

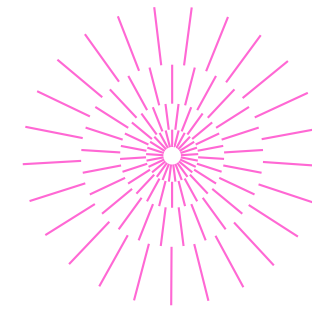
[健身族GFR數值起伏大，數值遠超過正常值的5倍，進一步切片檢查，才確診為慢性腎臟病]

- 健身族群若長期高蛋白飲食，可能增加腎臟負擔，導致慢性腎病風險。
- 科學家正在研究「硫酸吡哆酚（IS）」檢測，可能成為早期發現腎病的重要指標。



出處：信傳媒(<https://www.cmmedia.com.tw/home/articles/53293>).

[重度吸菸者 慢性腎病風險增三成]

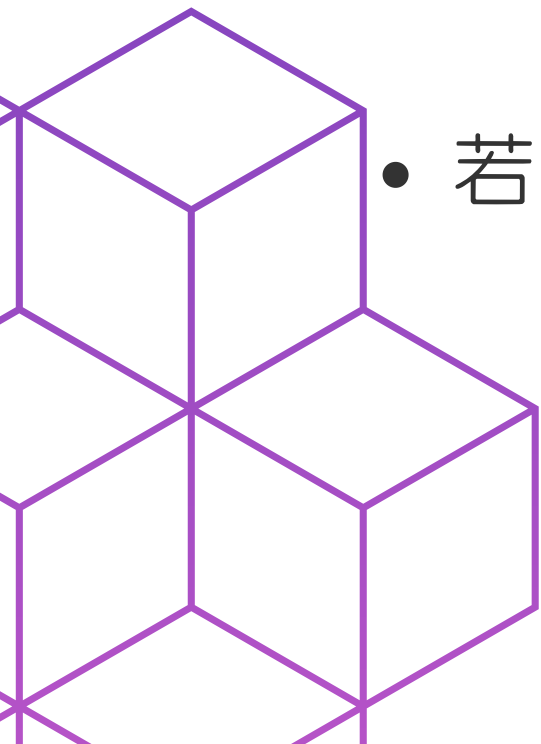


時事結合

- 香菸中的尼古丁會導致腎絲球過濾率增加，長期可能引發腎絲球硬化
- 菸草中的鎘會在體內累積，對腎小管造成毒性，最終導致腎小管及腎間質纖維化
- 若在起床後5分鐘內吸菸，風險更提高至70%

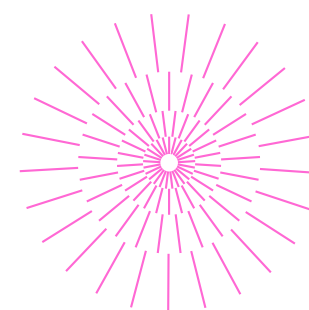


出處：<https://health.ltn.com.tw/article/paper/1676326>



慢性腎臟病「無聲」且不可逆 三高為高危族群

今年起腎功能篩檢下修至30歲



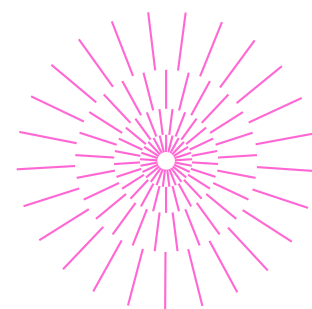
時事結合

- 慢性腎臟病（CKD）隱匿性高：早期無明顯症狀，晚期才易被察覺，且不可逆。
- 高風險族群：糖尿病、高血壓、肥胖患者，其中三高族群約 20-30% 有腎功能異常。
- 政策調整：台灣自 2025 年起將腎功能篩檢年齡下修至 30 歲，以提高早期發現率。
- 提升腎病防治意識：國健署加強宣導，鼓勵民衆主動篩檢。

護腎8招

- 均衡飲食
- 控制體重
- 飲水充足
- 戒菸保健康
- 規律運動
- 謹慎用藥
- 定期量血壓 722
- 控制三高

出處：<https://health.ltn.com.tw/article/paper/1676326>



推薦健康講座

財團法人腎臟病防治基金會



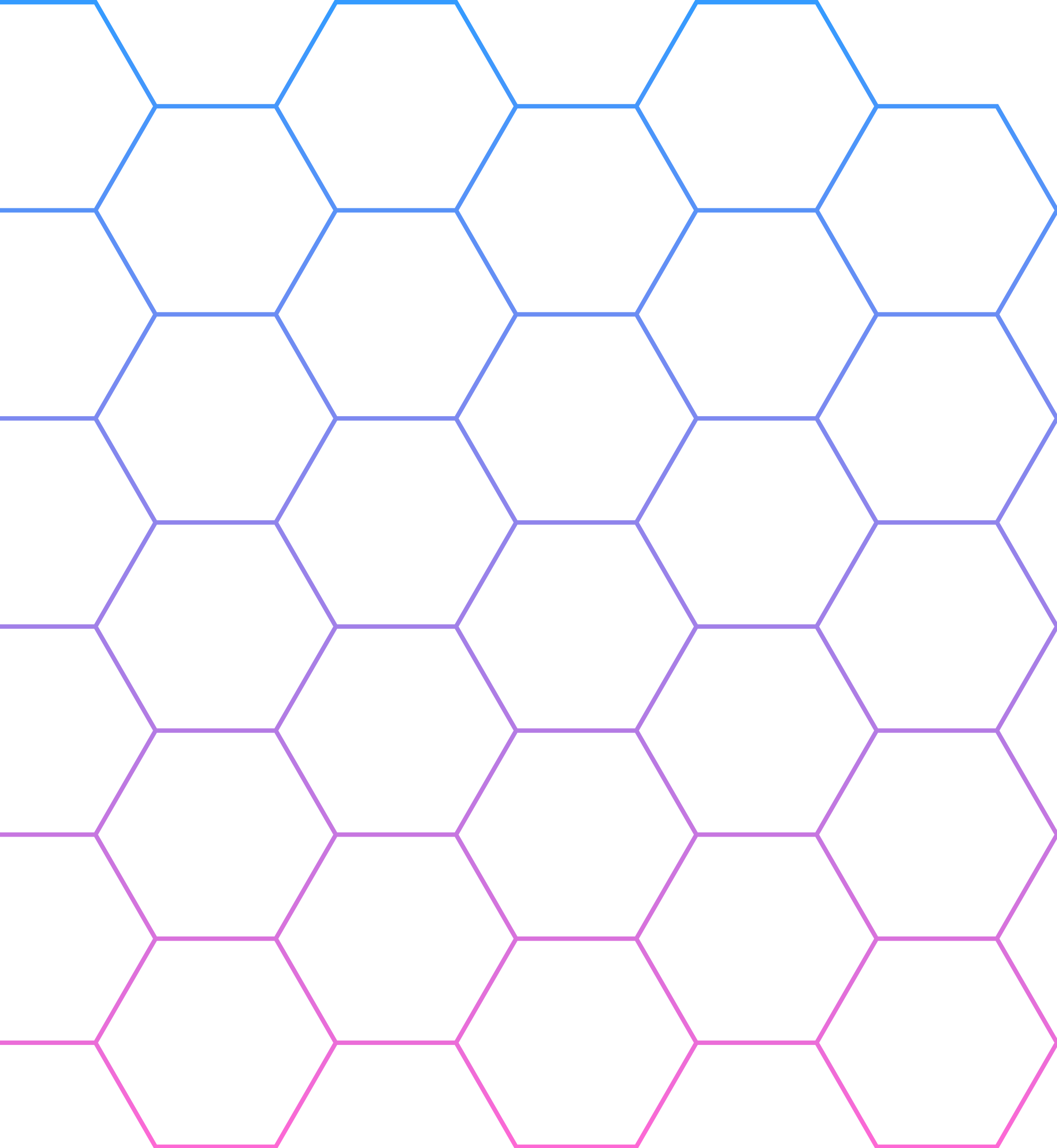
2025/3/23「一起聊腎事」衛教講座活動

2025年的第一場講座來囉 這次我們要前往台中了
歡迎住在附近的朋友一起團來參加 趕快點擊連結...

tckdf.org.tw

日期	時間	主題	講師
3/1(五)	PM 2:00 ~ 3:00	破解腎臟病的迷思	毛方聖醫師
3/1(五)	PM 3:00 ~ 4:00	預防腎病的飲食原則	蔡書玫營養師
3/29(五)	PM 2:00 ~ 3:00	淺談憂鬱症	陳君豪醫師
3/29(五)	PM 3:00 ~ 4:00	動出好心情	賴聖文治療師

臺大醫院民眾健康講座



感謝聆聽