

國立中正大學通識教育課程教學大綱

開課學年度/學期	115 學年度第 1 學期		
課程名稱 (中文)	資料科學分析與應用		
課程名稱 (英文)	Data Science Analysis and Applications		
課 碼	(由通識教育中心填寫)	學分數	2
授 課 方 式	請勾選(可複選)： ☒ 課堂講授 ☐ 網路教學 ☐ 分組討論 ☐ 校外教學 ☒ 其他：電腦教室上機實作		
教學目標及範圍	教學目標及範圍 資料科學是運用程式工具與分析方法，將原始資料轉化為可理解、可驗證並能支持判斷的資訊。資料在進入分析或人工智慧模型之前，必須經過整理、清理、探索、轉換及視覺化；資料的完整性、正確性、代表性與時效性，也會直接影響分析結果及 AI 模型的可信度。 本課程從資料科學的角度出發，適合沒有程式設計背景的學生。前半學期教授 Python 3 基礎，包括變數、資料型態、字串、串列、元組、集合、字典、條件判斷、迴圈、函數及 CSV 資料處理，使學生具備基本的程式閱讀、邏輯思考、資料表示與處理能力。 後半學期介紹 pandas 與 Matplotlib，帶領學生進行表格資料的讀取、檢視、篩選、清理、統計分析及視覺化。課程使用教師加工的 Adult Census Income 公開資料集，讓學生處理缺失值、重複資料、格式不一致、不合理數值、類別不平衡及樣本代表性等常見問題，並觀察資料清理前後對分析結果的影響。 課程也將介紹 AI 訓練資料、特徵、標籤、群體差異及資料漂移等概念。學生將分析不同年齡、性別、職業及教育程度群體的資料分布與結果差異，理解資料問題可能使 AI 模型產生不合理或不公平的結果，並比較原始資料與後續資料的分布變化，判斷資料是否仍適合原有用途。 期中考後，教師將帶領學生使用 Google Colab 的生成式 AI 功能，逐週建立一套「AI 訓練資料品質、偏差與漂移分析工具」。學生將學習描述分析需求、執行與修改 AI 產生的程式、判斷輸出是否合理，並以資料與圖表說明分析結果。期末作業由每週課堂成果逐步累積完成，不要求學生在課後獨立從零開發程式。 在「ISO/IEC 42001:2023 人工智慧管理系統」所關注的資料品質、偏差、公平性、風險監測及持續改善等議題，將作為本課程設計資料分析任務的參考，使學生了解資料科學如何協助發現資料問題、降低不當偏差、監測資料變化，並支		

持可信賴 AI 的建立與持續運作。

完成本課程後，學生應能使用 Python、pandas 與 Matplotlib 處理及呈現資料，辨識常見資料品質問題，分析群體差異與資料漂移，並能運用 Colab AI 協助完成程式，同時對程式與分析結果進行人工檢查及合理解釋。

課程內容比例

Python 3 與程式思維約占 50%；資料科學、資料品質、AI 基礎、偏差與漂移實作約占 50%。

授 課 大 綱
(週次表及每週課程詳細內容說明)

週次	主題
1	資料科學、AI 與 Google Colab 導論
2	Python 變數、資料型態、輸入輸出與運算
3	字串、布林值與條件判斷
4	List、Tuple、索引與切片
5	Dictionary、Set 與巢狀資料
6	For、While、Range 與迴圈
7	函數、參數、回傳值與變數作用域
8	Map、Filter、Lambda、CSV 與期中複習
9	期中考：Python 上機考
10	期末實作啟動：Adult Census Income 與 Colab AI
11	Pandas、DataFrame 與資料概況
12	資料品質檢查：缺失、重複、格式與不合理值
13	資料清理：轉換、補值、類別統一與處理紀錄
14	Matplotlib 與探索性資料分析
15	AI 訓練資料、標籤、偏差與群體差異
16	資料漂移、分布比較與 PSI
17	整合資料健檢工具、驗證與人工監督
18	個人期末成果展示與繳交

每週課程詳細內容說明：
以下為更詳細的週次活動資訊：

每週課程詳細內容說明

第 1 週：資料科學、AI 與 Google Colab 導論

- 說明資料、資訊、知識及資料科學的基本概念，介紹統計分析、人工智慧、機器學習與深度學習的關係。
- 介紹本課程的學習方式、評量方式、Google Colab 操作介面及 Notebook 儲存方式。

- 說明生成式 AI 在本課程中的角色：協助產生程式，但學生仍須閱讀、執行、檢查與說明。

第 2 週：Python 變數、資料型態、輸入輸出與運算

- 介紹變數命名、整數、浮點數、布林值、字串、型態轉換、`print()`、`input()` 及基本運算。
- 透過生活化案例練習把問題拆成輸入、處理與輸出。

第 3 週：字串、布林值與條件判斷

- 介紹字串索引、切片、常用方法、比較運算、`and/or/not` 與 `if/elif/else`。
- 練習資料格式檢查與簡單條件篩選。

第 4 週：List、Tuple、索引與切片

- 介紹 `list` 與 `tuple` 的建立、存取、修改、排序、尋找、加入與刪除。
- 使用串列管理多筆資料，建立後續資料分析所需的基本觀念。

第 5 週：Dictionary、Set 與巢狀資料

- 介紹 `dictionary` 的鍵值結構、存取、更新與巡訪，以及 `set` 的唯一性與集合運算。
- 練習用巢狀 `list/dictionary` 表達結構化資料，並簡介 `JSON`。

第 6 週：For、While、Range 與迴圈

- 介紹 `for`、`while`、`range()`、`break`、`continue` 與巢狀迴圈。
- 結合 `list`、`dictionary` 與 `set`，批次檢查及轉換多筆資料。

第 7 週：函數、參數、回傳值與變數作用域

- 介紹函數定義、參數、預設值、回傳值、區域變數與全域變數。
- 將重複的資料檢查流程封裝為可重複使用的函數。

第 8 週：Map、Filter、Lambda、CSV 與期中複習

- 介紹 `map()`、`filter()`、`lambda`、模組匯入及基本 `CSV` 檔案讀寫。
- 完成 Python 綜合練習並複習期中考範圍。

第 9 週：期中考

- 以電腦上機方式評量 Python 基礎能力。考試原則上不使用生成式 AI。
- 範圍包含變數、資料型態、字串、資料結構、判斷、迴圈、函數及簡單 `CSV` 處理。

第 10 週：期末實作啟動—Adult Census Income 與 Colab AI

- 介紹 `UCI Adult Census Income` 資料的用途、欄位、敏感屬性及使用限制。
- 載入教師加工版資料集，說明期末 `Notebook` 的固定架構與逐週累積方式。
- 練習把需求拆成明確提示詞，請 `Colab AI` 產生程式，再由全班共同執行與修正。

第 11 週: Pandas、DataFrame 與資料概況

- 介紹 pandas、Series、DataFrame、read_csv()、head()、info()、describe()、shape、columns 與資料型態。
- 完成資料筆數、欄位、類型、基本統計及類別分布的概況報告。

第 12 週: 資料品質檢查

- 檢查缺失值、特殊缺失符號、重複資料、前後空白、大小寫不一致、型態錯誤及不合理值。
- 使用函數與表格彙整每一欄的問題數量，建立資料品質檢查清單。

第 13 週: 資料清理與處理紀錄

- 使用 dropna()、fillna()、replace()、astype()、字串方法及條件篩選進行資料清理。
- 比較清理前後筆數與分布，記錄每一項清理規則，避免無法追溯的任意修改。

第 14 週: Matplotlib 與探索性資料分析

- 介紹長條圖、直方圖、圓餅圖、箱型圖及標題、座標軸與圖例。
- 繪製年齡、教育程度、職業、工時及收入標籤的分布，練習從圖表提出合理問題。

第 15 週: AI 訓練資料、偏差與群體差異

- 介紹特徵、標籤、訓練資料、測試資料、類別不平衡、敏感屬性與代理變數。
- 比較不同年齡群、性別、職業與教育程度的樣本數及高收入比例。
- 強調群體差異不等於歧視結論，必須結合資料來源、情境與人工審查。

第 16 週: 資料漂移、分布比較與 PSI

- 介紹資料漂移、概念漂移及模型部署後持續監測的必要性。
- 比較基準資料與教師提供的新資料，計算類別比例變化、平均值變化與簡化 PSI。
- 建立漂移警示表並討論何時需要重新檢查資料或模型。

第 17 週: 整合工具、驗證與人工監督

- 整合資料概況、品質檢查、清理、視覺化、偏差及漂移分析功能。
- 加入錯誤處理、欄位存在性檢查、分析摘要及人工審查提醒。
- 驗證 Google Colab Notebook 從頭執行是否成功，檢查 AI 生成程式是否與需求一致。

第 18 週: 個人期末成果展示與繳交

- 學生於課堂完成最後修正並展示個人 Notebook。
- 說明資料問題、主要圖表、偏差與漂移發現、人工審查需求及改善建議。
- 繳交可執行 Google Colab Notebook 及簡短的 AI 使用與驗證說明。

教科書及 延伸閱讀	主要教材 • 教師自編講義：Python 3 基礎、pandas、Matplotlib、資料品質、偏差與漂移。 • 教師製作之 Adult Census Income 加工版資料集、Google Colab Notebook 骨架與每週實作範例。 • Python 官方文件：https://docs.python.org/3/ • pandas 官方文件：https://pandas.pydata.org/docs/ • Matplotlib 官方文件：https://matplotlib.org/stable/ • UCI Machine Learning Repository: Adult Census Income 資料集。 • ISO/IEC 42001:2023 相關概念摘要：資料品質、偏差、公平性、漂移監測、人類監督、可追溯性與持續改善。 延伸閱讀 • 洪錦魁，《Python 入門邁向高手之路：王者歸來》，深石。 • 陳允傑，《Python 資料科學與人工智慧應用實務》，旗標。
評 量 方 式	☒ 課堂參與 D 類 ☒ 期中考 B 類 ☐ 期末考 C 類 ☐ 小組報告 ☐ 小組討論 ☐ 書面報告 ☒ 課後作業 A 類 ☐ 平時測驗 ☐ 心得分享 ☒ 學習紀錄 C 類 ☒ 專題創作 C 類 ☒ 其他：出席 D 類 A 類（程式練習）40%；B 類（期中 Python 上機考）20%；C 類（期末個人作業）20%；D 類（出席及課堂參與）20%。 評量說明 • 出席及課堂參與 20%：本課程採課堂逐週實作，出席、操作參與及當週成果均納入評量。 • 程式練習 40%：包含期中前 Python 基礎練習，以及期中後 pandas、Matplotlib、資料品質、偏差與漂移的分段成果。 • 期中考 20%：Python 上機考，檢核學生閱讀、撰寫及修改基礎程式的能力；原則上不使用生成式 AI。 • 期末個人作業 20%：第 10 至 18 週在課堂中逐步完成，不要求學生課外獨立從零開發。評量重點為 Notebook 可執行性、資料處理與圖表、偏差及漂移分析、結果解釋、人工審查與改善建議。
與聯合國永續發展 目標(SDGs)及 細項之對應 (請參閱 SDGs 對照表)	目標 4：優質教育 細項 4.4、4.7 培養學生資訊與資料素養，並理解科技、倫理與永續社會之關係。 目標 10：減少不平等 細項 10.2、10.3 透過資料偏差與公平性案例，理解資料與 AI 決策可能擴大或減少群體不平等。

目標 16：和平、正義及健全制度 細項 16.6、16.7 強調透明、可追溯、負責任的資料處理，以及人工監督與可問責的 AI 應用。			
核心能力指標設定	通識課程 核心能力指標 (請勾選主要的 3-5 項)	說明	課程能培養學生此項核心能力者請打✓
	(1) 思考與創新	能夠進行獨立性、批判性、系統性或整合性等面向的思考，或能以創意的角度來思考新事物。	V
	(2) 道德思辨與實踐	能夠對於社會、文化中相關的倫理或道德議題，進行明辨、慎思與反省，或能實踐在日常生活中。	V
	(3) 生命探索與生涯規劃	能夠主動探索自我的價值或生命的真諦，或能具體實踐在自我生涯的規劃或發展。	
	(4) 公民素養與社會參與	能夠尊重民主與法治的精神、關心公共事務及議題，或能參與社會事務及議題的討論與決策。	V
	(5) 人文關懷與環境保育	能夠具備同理、關懷、尊重、惜福等人文素養，或能擴及到更為廣泛的環境及生態議題。	
	(6) 溝通表達與團隊合作	能夠善用各種不同的表達方式進行有效的人際溝通，或能理解組織運作，與他人完成共同的事物或目標。	
	(7) 國際視野與多元文化	能夠了解國際的情勢與脈動，具備廣博的世界觀，或能尊重或包容不同文化間的差異。	
	(8) 美感與藝術欣賞	能夠領略各種知識、事物或領域中的美感內涵，或能據此促成具美感內涵之實踐力。	
	(9) 問題分析與解決	能夠透過各種不同的方式發現問題，解析問題，或能進一步透過思考以有效解決問題。	V
授課教師資料	姓名：王俊堯 ☐ 專任教師 學系(所，中心)： 職稱： ☒ 兼任教師 服務單位：中正通識中心 職稱：助理教授 國立中正大學 資訊工程系 博士、碩士。 國立中正大學 資訊工程系 兼任助理教授， 虎尾科技大學、南華大學 資訊工程系兼任助理教授。 專業領域：分散式計算、資訊安全、信任架構。		

備 註	• 本課程全程於電腦教室上課，學生須具備可登入 <code>alum.ccu.edu.tw</code> 帳號及使用 Google Colab 的能力。 • 本課程不要求程式設計先備經驗，但學生須依進度完成課堂練習。 • 期中考後允許且要求使用 Google Colab AI 協助產生程式；學生仍須閱讀、執行、驗證、修改並說明程式。 • 不得將真實個人資料、機敏資料、未公開研究資料或受限制資料上傳至生成式 AI 服務。 • 期末作業使用教師提供之 Adult Census Income 加工版資料，於課堂中逐週完成。
--------	---