

國立中正大學課程大綱

課程名稱(中文)	資料科學導論	開課單位：	數學系
課程名稱(英文)	Introduction to Data Sciences	課程代碼	2104556
授課教師	賴振耀	學分數	3
必/選修	選修	開課年級	四
先修科目或先備能力	微積分、線性代數、程式語言		
課程概述	此課程在介紹資料分析中使用的基本方法，搭配 python 進行練習，並利用實際案例進行資料分析。		
學習目標	● 介紹資料科學相關子領域其應用，內容將涵蓋資料收集與預處理、相關數學基礎回顧、Python 套件工具實作。 ● 培養自主學習的能力。		
參考書籍	1. J. Vander Plas (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media. 2. R. Mitchell (2018). Web Scraping with Python, 2nd ed., O'Reilly Media. 3. A. Anderson and D. Semmelroth (2015). Statistics for Big Data for Dummies, Wiley. 4. J. Grus (2019). Data Science from Scratch, ed., O'Reilly Media.		

課程大綱

單元主題	內容綱要	上課週數
基礎與環境	● 課程簡介。 ● Python for Data Science : Jupyter Notebook / Google Colab、Numpy 陣列操作、Pandas DataFrame 操作、基本資料探索 (EDA)。	3-4
資料獲取與前處理	● 資料來源與開放資料:公開資料平台 (政府資料、Kaggle)、API 介紹 (JSON, RESTful API)。 ● 網路爬蟲 (I): Requests, BeautifulSoup、HTML 結構與 DOM 解析。 ● 網路爬蟲 (II): Selenium 自動化爬蟲、爬蟲倫理與 robots.txt。 ● 資料清理 (Data Wrangling):缺失值、異常值處理、資料型態轉換、合	5

	併與切割。	
探索與分析	● 敘述統計與機率基礎回顧：平均數、變異數、相關係數、常見分布（常態、二項、Poisson 等）。 ● 資料視覺化 (I): Matplotlib / Seaborn 基本繪圖探索與分析、折線圖、長條圖、散佈圖。 ● 資料視覺化(II): 高階視覺化 (heatmap, boxplot) 探索與分析分群視覺化。 ● 非監督式學習 (聚類分析): K-means, Hierarchical clustering、應用案例。 ● 降維 (Dimensionality Reduction): PCA 原理（主成分分析）、t-SNE / UMAP 直觀概念。	5
應用與進階	● 文字資料處理 (NLP 入門): 文本清理、詞頻分析、WordCloud。 ● 時間序列分析入門: 移動平均、季節性趨勢。 ● 資料倫理與隱私: 資料偏差 (Bias) 、AI 公平性、資料隱私。	4-5

評量方式

- (1) 出席率：10%
- (2) 作業、小考、隨堂測驗：60%
- (3) 期中/期末報告：30%