

資訊及通訊科技業 《能力標準說明》 能力單元

「數據科學」職能範疇

名稱	制定並執行數據採集和收集程序，進行預處理和探索性數據分析
編號	111131L5
應用範圍	進行數據採集過程，包括破解/整理複雜數據，選擇與問題相關的適當技術。利用一系列的數據審計技術和方法進行數據質量評估和數據質量驗證。對有噪音、不完整的數據或有數據質量問題的數據進行數據清理
級別	5
學分	3 (僅供參考)
能力	<p>表現要求</p> <ol style="list-style-type: none"> 1. 掌握數據獲取和收集過程以及預處理和探索性數據分析的知識 <ul style="list-style-type: none"> • 有能力 <ul style="list-style-type: none"> ◦ 具有數據獲取和收集過程的知識 (見備註1) ◦ 具有預處理和探索性數據分析的知識 (見備註2) 2. 評估和管理不同的數據採集方法 <ul style="list-style-type: none"> • 能夠使用適當的方法，從內部或外部來源獲取數據 3. 對數據進行預處理和轉換 <ul style="list-style-type: none"> • 能夠對數據進行預處理和轉換，以確保使用特徵工程或探索性數據分析，使其達到項目的最佳格式、佈局或形狀 4. 進行和審查數據質量過程 <ul style="list-style-type: none"> • 能夠進行和審查數據質量評估，以發現缺失值、重複值、多個來源和不一致的格式
評核指引	<p>此能力單元的綜合成效要求為：</p> <ul style="list-style-type: none"> • 具有數據獲取和收集過程以及預處理和探索性數據分析的知識 • 開發和執行數據獲取和收集過程，進行預處理和探索性數據分析，以獲得有效的、準確解釋的和適用於所需環境的數據
備註	<ul style="list-style-type: none"> • 1. 各種數據採集方法是收集新的數據；轉換/改造遺留數據；共享/交換數據；以及購買數據。這包括自動收集（如傳感器派生的數據），人工記錄經驗觀察，以及從其他來源獲得現有數據。 • 2. 探索性數據分析（EDA）是指對數據集進行的某些類型的初步分析和發現，通常是在分析過程的早期。對數據的預處理將通過分析數據的分類或數字，使其可視化和一些統計決策來進行。