

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	通過數據品質評估進行和審查數據的預處理和轉換
編號	111140L6
應用範圍	探索數據分析和使用一系列數據審計技術和方法，對數據進行預處理和轉換，並審查數據品質，以確保其採用適合專案目的的最佳格式、佈局或形狀
級別	6
學分	6（僅供參考）
能力	<p>表現要求</p> <ol style="list-style-type: none"> 1. 具備預處理和探索性數據分析知識（見注1） 2. 評估和管理數據的預處理和轉換，以確保其採用適合專案目的（即特徵工程）的最佳格式、佈局或形狀 3. 進行探索性數據分析，對數據進行初步調查，以發現模式/異常，測試假設，並藉助匯總統計和圖形表示檢查假設 4. 根據對數據的廣泛瞭解，對缺失值，重複和格式不一致等潛在問題以及對數據科學/分析過程的影響進行數據質量評估
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"> • 具備相關知識，並進行適當的預處理、探索性數據分析及數據質素評估，以確保資料以最佳格式進行，並對資料質素進行初步調查（見注2）
備註	<ul style="list-style-type: none"> • 1. 資料預處理將包括: 通過分析分類或數字數據、可視化和一些統計決策來執行。探索性數據分析（EDA）是一個術語，通常在早期的分析過程用於描述某些類型的初始分析。 • 2. 數據應有效、準確解釋並適用於所需情況