# 數據分析文件-學歷與職場發展之關聯

## 目錄

[數據分析文件 1](#_Toc178017955)

[目錄](#_Toc178017956) 1

[前言 2](#_Toc178017957)

[資料來源 (Kaggle) 2](#_Toc178017958)

[匯入之套件 3](#_Toc178017959)

數據分析項目 [4](#_Toc178017960)

[1. 【晉升機會(Years\_to\_Promotion)之相關分析】(許瑞宏) 4](#_Toc178017961)

[2. 【分析項目】(Name) 8](#_Toc178017962)

[3. 【分析項目】(Name) 13](#_Toc178017963)

[分析之結果總結 14](#_Toc178017963)

## 前言

本技術文件詳細介紹了對 Kaggle 數據集之 **"Education Career Success"** 數據分析過程。我們將使用 **Pandas , Numpy**等套件進行數據處理，並透過 **Plotly, Matplotlib, Seaborn** 等套件進行視覺化分析，以探索學歷、技能與職場表現（如起薪、工作滿意度）之間的關係。

(Kaggle數據資料來源:<https://www.kaggle.com/datasets/adilshamim8/education-and-career-success>)

1. **數據集簡介**

**此份Kaggle中"Education Career Success"的數據集包含 5000 筆數據(X軸)和 20 個特徵(Y軸)，以下是數據集的主要欄位：**

|  |  |  |
| --- | --- | --- |
| **類別** | **欄位名稱** | **說明** |
| 基本資訊 | Student\_ID | 學生編號 |
| 基本資訊 | Age | 年齡 |
| 基本資訊 | Gender | 性別（Male / Female / Other） |
| 學術背景 | High\_School\_GPA | 高中 GPA |
| 學術背景 | SAT\_Score | SAT 考試分數 |
| 學術背景 | University\_Ranking | 大學排名（數值越小代表排名越高） |
| 學術背景 | University\_GPA | 大學 GPA |
| 學術背景 | Field\_of\_Study | 學習領域（如 Computer Science、Law 等） |
| 技能與經歷 | Internships\_Completed | 完成的實習數量 |
| 技能與經歷 | Projects\_Completed | 完成的專案數量 |
| 技能與經歷 | Certifications | 取得的證書數量 |
| 技能與經歷 | Soft\_Skills\_Score | 軟技能評分（1-10） |
| 技能與經歷 | Networking\_Score | 人脈網絡評分（1-10） |
| 職業發展 | Job\_Offers | 獲得的工作機會數 |
| 職業發展 | Starting\_Salary | 起薪（美元） |
| 職業發展 | Career\_Satisfaction | 職業滿意度（1-10） |
| 職業發展 | Years\_to\_Promotion | 晉升所需年數 |
| 職業發展 | Current\_Job\_Level | 當前職位層級（Entry, Mid, Senior） |
| 職業發展 | Work\_Life\_Balance | 工作與生活平衡（1-10） |
| 職業發展 | Entrepreneurship | 是否創業（Yes/No） |

## 

1. **實作步驟**
2. **數據檢視與清理**

(1)

|  |  |
| --- | --- |
| 程式碼實作 | 程式碼解析 |
|  | **匯入套件並進行數據預覽(**education\_career\_success.csv**)**使用 pandas.read\_csv() 讀取 CSV 檔案後的結果。 |

(2)

|  |  |
| --- | --- |
| 程式碼實作 | 程式碼解析 |
|  | ****Method1:******使用**df.info()** 檢視個特徵類別的筆數與類型。**  * 從Non-Null 得知並無缺失值   **Method 2:**   * 使用**df.insull().sum()** 亦顯示無缺失值 |

(3)

|  |  |
| --- | --- |
| 程式碼實作 | 程式碼解析 |
|  | **使**df.describe()**得到統計資料** |

1. **欄位分析與可視化圖表**
2. 【學歷】與【晉升機會】分析

|  |  |
| --- | --- |
| 程式碼實作與可視化圖表 | 程式碼解析 |
|  | 轉換分類大學排名(間距排名333(含)以下，333-667(含)，以及667-1000(含)University\_Tier 欄位把學校排名轉換為類別。(Top Tier, Mid Tier, and Lower Tier)計算 **不同學校層級的畢業生，在職場晉升所需年數**。 |
|  | 1. **Top Tier（頂級學校）** 的畢業生通常晉升較快（較少的 Years\_to\_Promotion）。 2. **Lower Tier（較低排名學校）**   的畢業生晉升較慢，但仍然有一定比例在短時間內晉升。 Mid Tier（中等學校） 則處於中間地位。 |

**小結**

* **學歷影響職場晉升速度**，頂級學校的畢業生通常能夠更快獲得晉升機會。
* **中等學校畢業生晉升節奏較為平均**，沒有明顯的優勢或劣勢，屬於職場的中堅份子。
* **較低排名學校的畢業生晉升較慢，但仍有一部分能夠在短時間內晉升**，這可能取決於個人能力、職業選擇或產業需求。

1. **【性別】與【晉升機會】分析**

|  |  |
| --- | --- |
| 程式碼實作與可視化圖表 | 程式碼解析 |
|  | 此程式碼使用Plotly 來繪製 不同性別在各晉升年數上的人數分佈，以直方圖顯示，並區分性別。histfunc="count" 讓 Y 軸顯示該晉升年數的性別總數。  * **barmode="group"：**讓不同性別的數據分開顯示（不堆疊），方便比較。 |
|  | * 男性與女性的晉升年數分佈較為相似，但男性員工人數稍多。 * Other（其他性別）的人數較少，對晉升年數的影響較小。 * 晉升時間大致分佈均勻，沒有明顯的集中於特定年數。 |

**小結**

* 男性在前期晉升可能因人數原因上升，並到中間年數有男性較趨向於轉換跑道或離職，並在往後繼續招募男性新鮮人比例較高。
* 可能要再計算不同性別的平均晉升年數 來確定是否存在性別差異。
* 並進一步分析各性別在不同職位層級的分佈，以確保晉升公平性。

1. **【性別】與【晉升機會】分析**

|  |  |
| --- | --- |
| 程式碼實作與可視化圖表 | 程式碼解析 |
|  | * **數據範圍較寬**:說明某組的 Networking\_Score 變異較大。 * **形狀較厚的部分**:代表該區間人數較多（更常見）。 |
|  | * **第 1、2、4 年晉升者：**Networking Score 分佈較廣，部分低分者仍能晉升，高分者可能更快獲得晉升。 * **第 3 年晉升者：**分佈較集中，顯示晉升標準較一致。 * **第 5 年晉升者：**變異較小，影響力較低，可能由資歷或績效決定晉升時機。 |

**小結**

* Networking Score 與晉升有一定關聯，但並不是唯一因素。
* 高 Networking Score 的員工，可能更早獲得晉升（1~2 年晉升者中，有部分高分者）。
* 晉升時間較長（5 年）的員工，Networking Score 影響力較小，可能由其他因素決定晉升時機。