**專案標題 (Project Title)**

**摘要 (Abstract)**

疫情期間，遠端工作成為全球企業運營的常態，社會和企業不得不迅速適應這一新的工作模式。隨著疫情逐漸緩解，一些企業開始要求員工回歸現場工作，理由是現場工作能提升生產力。然而，這一轉變引發了員工的強烈反彈，使企業管理層面臨新的挑戰。

本研究旨在探討工作地點（遠端、混合、現場）對企業生產力的影響，深入分析不同工作模式下員工績效的變化。同時，本研究進一步探討員工的壓力水準和心理健康對生產力的影響，並分析這些因素是否會影響員工的流失行為（辭職）。此外，通過綜合分析工作場所、心理健康和壓力水準等變數，本研究將探討這些因素如何預測員工的生產力和流動性，並提出優化工作安排的策略建議，以提高員工績效並降低流失率。

**引言 (Introduction)**

在疫情期間，遠端工作迅速成為全球企業運營的主流模式，社會和企業不得不適應這一工作型態。然而，隨著疫情逐漸緩解，部分企業開始要求員工回到現場工作，認為這樣能提升生產力並促進團隊協作。然而，這一轉變引發了員工的強烈反彈，許多員工認為遠端工作提高了他們的靈活性和生活品質，並不願回到傳統的辦公室工作模式。這引出了關鍵問題：**工作地點是否確實影響員工的績效？** 進一步來說，**工作地點的變化對員工的壓力水準和心理健康有何影響，這些因素是否會進而影響生產力和員工的流失行為？**

理解工作地點、壓力水準與心理健康之間的關聯對企業決策至關重要。隨著工作模式的多樣化，企業需要在提升生產力和保護員工福祉之間找到平衡。如果工作地點的選擇不僅影響生產力，還對員工的心理健康和壓力水準產生負面影響，可能會導致員工士氣下降、績效減弱，甚至帶來更高的離職率。有效解決這一問題，不僅有助於提升生產力，還能改善員工滿意度、降低員工流失率，從而為企業帶來長期的競爭優勢。

然而，現有的解決方案大多僅關注工作地點對生產力的直接影響，忽視了工作地點變化對員工壓力水準和心理健康的間接影響。此外，這些方案通常未考慮到員工流失率對企業生產力和運營成本的長期影響。因此，本研究將進行更全面的評估，整合工作地點、壓力水準、心理健康狀況和員工流失行為，探討這些因素如何共同影響員工的生產力與流失率。這將為企業提供深度的策略建議，幫助優化工作安排，提升員工績效，並有效降低流失率。

**相關研究 (Related Work)**

### 深入討論已知解決方案

1. **遠距工作下職家衝突與工作績效對離職傾向之關聯性研究**

https://www.airitilibrary.com/Article/Detail/U0064-1912202311022661

* + **問題解決方式**：本研究通過深度訪談探討了遠距工作對員工離職傾向及工作績效的影響。研究發現，遠距工作對員工有雙重影響，尤其是對有小孩或高齡家屬的受訪者，工作與家庭的界線模糊可能導致工作效率下降。同時，對於需要面對面溝通的工作，遠距工作對生產力有顯著負面影響。研究認為，遠距工作通過技術工具可以提升部分員工的生產力，但工作績效的提升取決於工作性質及自律性。
  + **局限性**：這項研究依賴質性研究方法，僅基於10名受訪者的訪談結果，樣本量較小，難以推廣到所有遠距工作者。此外，研究並未深入探討遠距工作對不同壓力水準或心理健康狀況的影響，忽略了這些重要因素。

1. **遠距工作模式對工作自主性與工作生產力之影響**

https://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/login?o=dnclcdr&s=id=%22111YUNT0121071%22.&searchmode=basic

* + **問題解決方式**：本研究通過定量問卷調查，探討了工作自主性在遠距工作模式中的作用，並發現工作自主性對工作生產力有顯著正向影響。該研究還引入了「被主管信任」作為調節變數，強調主管信任可以增強員工的自主性，進而提升工作績效。
  + **局限性**：該研究的樣本量較大（287份有效問卷），但依賴自我報告的問卷調查，存在主觀偏差。且僅限於探討自主性對工作績效的影響，未深入分析壓力或心理健康對生產力的影響。

1. **探討遠距工作對工作表現之影響**

https://dba.nkust.edu.tw/uploads/asset/data/623d3a8f2e6356240a2de868/2021.6.8.2.pdf

* + **問題解決方式**：該研究通過定量分析，探討遠距工作對需求滿足、組織承諾與工作表現的影響，並發現遠距工作中的靈活性和頻率對組織承諾有正向影響，進而提升工作表現。此研究證明，當員工感受到內在需求被滿足時，他們的生產力會增加。
  + **局限性**：雖然研究討論了需求滿足對工作表現的影響，但忽略了外部因素如壓力或心理健康對員工績效的影響。此外，該研究對於「靈活性」的定義過於模糊，未具體闡明如何在遠距工作環境中實現靈活性。

**分組討論**

根據專案的主題，可以將這三篇研究分為兩大類：**工作績效與工作條件的關聯性研究** 和 **組織支持與生產力的研究**。這樣的分類可以幫助我們理解這些研究如何為你的專案提供背景參考。

1. **工作績效與工作條件的關聯性研究**
   * **研究1：遠距工作下職家衝突與工作績效對離職傾向之關聯性研究**
     + 此研究關注於**工作場所**與**家庭干擾**的影響，尤其對有家庭責任的員工來說，遠距工作可能帶來工作與生活的界線模糊，進而降低工作績效。這與本專案中探討的「工作地點（遠端、混合、現場）如何影響員工績效」息息相關，提供了遠距工作在不同情境下影響生產力的實際例子。
   * **研究3：探討遠距工作對工作表現之影響**
     + 此研究著重於**需求滿足**和**靈活性**對員工工作表現的正向影響，強調靈活的工作模式能提升員工的忠誠度和生產力。這為本專案提供了有關「工作場所和靈活性」如何影響員工生產力的背景資料。
2. **組織支持與生產力的研究**
   * **研究2：遠距工作模式對工作自主性與工作生產力之影響**
     + 該研究強調了**主管信任**和**工作自主性**在遠距工作環境中的重要性。這可以補充本專案中「心理健康和壓力水準是否影響生產力或流失」的研究，尤其是主管對員工的信任如何幫助減輕員工的壓力，並提升其工作績效。

**區別與優勢**

* **區別**：本專案不僅關注工作地點對生產力的影響，還包括**壓力水準**和**心理健康**這些更廣泛的因素，而現有的研究大多集中於工作場所的靈活性和自主性對績效的影響。這使得本專案能夠在現有文獻的基礎上擴展，探索員工心理因素對工作績效和流失率的關鍵影響，填補現有研究的空白。
* **優勢**：本專案不僅僅是分析單一因素對生產力的影響，而是通過結合多個數據集，研究多個因素（工作場所、心理健康、壓力水準等）對員工生產力和流失的綜合影響。這能提供更全面的視角，有助於為企業管理層提供更具實用價值的策略建議。

**擬定計劃 (Proposed Work)**

* + **資料描述**：

本專案將使用兩組數據集進行分析，分別涵蓋員工的績效、生產力以及遠端工作與心理健康的相關數據。這兩組數據將幫助我們全面了解不同工作模式（遠端、混合、現場）下，員工的生產力、心理健康和工作滿意度之間的關聯。

**一、員工績效和生產力數據**

此數據集包含100,000筆記錄，涵蓋企業環境中員工的績效、生產力及人口統計特徵。數據中的變數詳細記錄了與員工工作習慣、教育背景、工作績效和滿意度等關鍵方面的資訊。這些變數將幫助我們分析哪些因素影響員工生產力和離職率。

* Employee\_ID：每個員工的唯一標識符。
* Department：員工工作的部門（例如，銷售、人力資源、IT）。
* Gender：員工的性別（男性、女性、其他）。
* Age：員工的年齡（22 至 60 歲之間）。
* Job\_Title：員工擔任的角色（例如，經理、分析師、開發人員）。
* Hire\_Date：員工的入職日期。
* Years\_At\_Company：員工為公司工作的年限。
* Education\_Level：最高學歷（高中、學士、碩士、博士）。
* Performance\_Score：員工的績效評級（1 到 5 分）。
* Monthly\_Salary：員工的月薪（美元），與職稱和績效分數相關。
* Work\_Hours\_Per\_Week：每周工作的小時數。
* Projects\_Handled：員工處理的項目總數。
* Overtime\_Hours：去年的總加班時間。
* Sick\_Days：雇員的病假天數。
* Remote\_Work\_Frequency：遠端工作的時間百分比（0%、25%、50%、75%、100%）。
* Team\_Size：員工團隊中的人數。
* Training\_Hours：訓練所花費的小時數。
* Promotions：在任期內獲得的晉升次數。
* Employee\_Satisfaction\_Score：員工滿意度評分（1.0 到 5.0 等級）。
* Resigned：指示員工是否已辭職的布爾值。

**二、遠端工作與心理健康數據**

此數據集包含5,000筆記錄，專注於遠端工作對員工壓力、心理健康及工作滿意度的影響。它收集了員工的工作地點、壓力水準、心理健康狀況和社交孤立感等重要變數，幫助我們深入探討這些因素如何影響員工的生產力與幸福感。

* Employee\_ID：每個員工的唯一標識符。
* 年齡：員工的年齡。
* 性別：員工的性別。
* Job\_Role：員工的當前角色。
* 行業：他們工作的行業。
* Years\_of\_Experience
* Work\_Location：他們是遠端工作、混合工作還是現場工作。
* Hours\_Worked\_Per\_Week：每周工作時數。
* Number\_of\_Virtual\_Meetings：每周參加的虛擬會議次數。
* Work\_Life\_Balance\_Rating：員工對工作與生活平衡的評分。
* Stress\_Level：他們自我報告的壓力水準。
* Mental\_Health\_Condition：報告的任何心理健康情況（焦慮、抑鬱等）。
* Productivity\_Change：疫情期間員工自我報告的生產力變化。
* Social\_Isolation\_Rating：自我報告的關於他們感到孤立的程度的評分 （1-5）。
* Satisfaction\_with\_Remote\_Work：他們對遠端工作安排的滿意度（滿意、中立、不滿意）。
  + **工具與方法**：

為了進行數據分析和建模，本專案將完全使用Python進行開發，並根據需求構建必要的工具以支持數據處理、特徵工程和模型構建。具體技術如下：

* **Python編程語言**：Python將用於數據處理、統計分析和模型訓練。主要使用的庫包括：
  + pandas：用於數據讀取、清洗和操作。
  + numpy：進行數值計算和矩陣運算。
  + scikit-learn：用於資料探勘模型的訓練和評估，包括分類、迴歸和聚類模型的實施。

#### **資料探勘技術**：

1. **分類模型**：使用分類模型，來預測員工的辭職可能性和生產力變化。
2. **聚類分析**：應用聚類技術，將員工分成不同群體，以識別高壓力或低生產力的員工群體。

#### **數據可視化**：

為了更好地展示分析結果，將使用**Altair**進行數據可視化。Altair基於Vega和Vega-Lite，能夠以聲明式的方式創建靈活且互動性強的可視化圖表。Altair將用來展示數據之間的關聯性、模型結果及重要特徵，並通過直觀的圖表幫助解釋模型預測的準確性和表現。

#### **交叉驗證與模型評估**：

為了確保模型的穩健性，將使用**k折交叉驗證**來檢驗模型的性能。

* **Main Task**

本專案的核心目標是將兩個數據集進行串聯，以提供更全面的分析。這兩個數據集雖然獨立，但存在幾個共同的特徵，能夠作為數據串聯的基礎。具體來說，以下4組變數將用於連接這兩個數據集：

1. **年齡**（Age）
2. **性別**（Gender）
3. **職位/部門**：第一個數據集中的「Job\_Role」（職位）與第二個數據集中的「Department」（部門）有邏輯上的關聯，可作為對應變數。
4. **遠端工作頻率/工作地點**：第一個數據集中的「Remote\_Work\_Frequency」（遠端工作頻率）與第二個數據集中的「Work\_Location」（工作地點）反映了員工的工作模式，可用來匹配遠端或現場工作的頻率。

透過這四組共同變數的串聯，本專案將能夠跨數據集進行聯合分析，深入探討工作地點、壓力水準、心理健康、員工績效等多變數之間的相互關聯性。這種整合將使得我們能夠從不同視角進行綜合評估，揭示遠端工作對生產力和員工福祉的全貌，並預測可能的員工流失風險。

**員工績效和生產力數據分析**

一、分析目標

目標是探討影響員工滿意度與生產力和人員流失的關鍵因素，特別關注遠端工作、年資、薪資、工時和團隊規模等變數。將進行相關係數分析，檢查變數之間的線性關係，以便找出這些因素是否對員工滿意度和工作生產力有顯著影響。

**二、進行相關係數分析**

* **目標**：檢查變數之間的線性相關性。
* 數據欄位：
  + Department：員工工作的部門
  + Gender：員工的性別
  + Age：員工的年齡
  + Performance\_Score：員工的績效評級
  + Remote\_Work: 遠端工作類型（遠端、混合、現場）
  + Years\_At\_Company: 員工年資
  + Monthly\_Salary: 月薪
  + Work\_Hours\_Per\_Week: 每週工作時數
  + Projects\_Handled: 處理的項目數
  + Overtime\_Hours: 加班時間
  + Remote\_Work\_Frequency: 遠端工作頻率
  + Team\_Size: 團隊規模
  + Employee\_Satisfaction\_Score: 員工滿意度評分
* 分析項目：
  + **遠端工作**：檢查 Department、Gender、Age、Remote\_Work 是否對員工績效（Performance\_Score）、員工滿意度（Employee\_Satisfaction\_Score）、辭職(Resigned)有顯著影響。
  + **年資與薪資**：檢查 Years\_At\_Company 和 Monthly\_Salary 是否對員工績效（Performance\_Score、員工滿意度（Employee\_Satisfaction\_Score）、辭職(Resigned)有顯著影響。
  + **工作時數、項目數、加班時間**：檢查 Work\_Hours\_Per\_Week、Projects\_Handled 和 Overtime\_Hours 是否對員工績效（Performance\_Score）、員工滿意度（Employee\_Satisfaction\_Score）、辭職(Resigned)有顯著影響。
  + **團隊大小和遠端工作頻率**：探討 Remote\_Work\_Frequency 和 Team\_Size 是否對員工滿意度（Employee\_Satisfaction\_Score）與辭職(Resigned)有顯著影響。

分析結論；只有薪資跟員工績效，Remote\_Work\_Frequency跟 Resigned有相關

**遠端工作與心理健康數據分析**

一、分析目標

目標是探討遠端工作影響員工壓力水準與心理健康影響的關鍵因素，特別關注遠端工作、壓力水準、心理健康狀況、孤立程度等變數。將進行相關係數分析，檢查變數之間的線性關係，以便找出這些因素是否對工作地點的生產力有顯著影響。

**二、進行相關係數分析**

* **目標**：檢查變數之間的線性相關性。
* 數據欄位：
  + Job\_Role：員工工作的部門
  + Gender：員工的性別
  + Age：員工的年齡
  + Work\_Location: 遠端工作類型（遠端、混合、現場）
  + Years\_of\_Experience: 員工年資
  + Hours\_Worked\_Per\_Week: 每週工作時數
  + Number\_of\_Virtual\_Meetings: 虛擬會議次數
  + Work\_Life\_Balance\_Rating：工作與生活平衡評分
  + Stress\_Level：自我報告的壓力水準。
  + Mental\_Health\_Condition：自我報告的心理健康情況
  + Productivity\_Change：工作量的變動
  + Social\_Isolation\_Rating：自我報告感到孤立的程度
  + Satisfaction\_with\_Remote\_Work：對遠端工作的滿意度
* 分析項目：
  + 心理健康狀況：檢查 Job\_Role、Gender、Age、Work\_Location 是否對Work\_Life\_Balance\_Rating、Stress\_Level、Mental\_Health\_Condition、Social\_Isolation\_Rating、Satisfaction\_with\_Remote\_Work有顯著影響。
  + **年資與工作量**：檢查Years\_of\_Experience、Hours\_Worked\_Per\_Week、Number\_of\_Virtual\_Meetings、Productivity\_Change是否對Work\_Life\_Balance\_Rating、Stress\_Level、Mental\_Health\_Condition、Social\_Isolation\_Rating、Satisfaction\_with\_Remote\_Work有顯著影響。

分析結論

只有Social\_Isolation\_Rating跟Gender有相關

**聯合分析**

一、分析目標

目標是探討員工滿意度與員工壓力水準與心理健康對於員工生產力或人員流失的關鍵因素，將比對2個數據集的相關性，找出這些因素是否對員工生產力或人員流失有顯著影響。

#### 二、分析項目

壓力水準與生產力的聯合分析：

* **單獨分析**：
  + **第一個數據集**：檢查 Gender、Age、Remote\_Work、Work\_Hours\_Per\_Week、Projects\_Handled 和 Overtime\_Hours 是否對員工績效（Performance\_Score）有顯著影響。
  + **第二個數據集**：檢查Gender、Age、Stress\_Level是否對工作地點(Work\_Location) 有顯著影響。
* **聯合分析**：
  + 根據兩個數據集的分析結果，透過工作地點的關聯，比對在不同的性別與年齡之間，壓力水準(Stress\_Level)是否影響員工績效（Performance\_Score）

心理健康與辭職行為的聯合分析：

* **單獨分析**：
  + **第一個數據集**：檢查 Gender、Age、Remote\_Work、Work\_Hours\_Per\_Week、Projects\_Handled 和 Overtime\_Hours 是否對員工滿意度（Employee\_Satisfaction\_Score）、辭職(Resigned)有顯著影響。
  + **第二個數據集**：檢查 Gender、Age、Work\_Location 、Work\_Life\_Balance\_Rating、Mental\_Health\_Condition、Social\_Isolation\_Rating 是否對Stress\_Level有顯著影響。
* **聯合分析**：
  + 根據兩個數據集的分析結果，透過高工時帶來高壓力的關聯，比對在不同的性別與年齡之間，工作地點的不同與心理健康狀況是否影響員工滿意度（Employee\_Satisfaction\_Score）與辭職(Resigned)。

數據建模

問題一：提升員工生產力並降低離職率

**I. 模型目標**  
利用員工的性別、年齡、壓力水準與心理健康等特徵，透過分類模型預測適合的工作地點，以提升員工生產力並降低離職率。

**II. 分類模型**

1. **第一個數據集**
   * 基於Gender（性別）、Age（年齡）、Performance\_Score（績效評分）、Work\_Hours\_Per\_Week（每週工時）、Projects\_Handled（處理項目數）和Overtime\_Hours（加班時數），使用分類模型預測不同工作地點（遠端、混合、現場）下員工的高績效可能性。
   * 基於Gender（性別）、Age（年齡）、Remote\_Work（遠端工作頻率）、Work\_Hours\_Per\_Week（每週工時）、Projects\_Handled（處理項目數）和Overtime\_Hours（加班時數），使用分類模型預測員工是否有較高的辭職可能性（Resigned）。
2. **第二個數據集**
   * 基於Stress\_Level（壓力水準）、Work\_Life\_Balance\_Rating（工作與生活平衡評分）、Mental\_Health\_Condition（心理健康狀況）、Social\_Isolation\_Rating（社交孤立評分），使用分類模型預測在不同工作地點下員工的壓力水準或心理健康狀況，並預測哪些工作地點有助於改善心理健康或降低壓力。

問題二：哪些因素影響員工生產力及離職率

**I. 模型目標**  
利用員工的性別、年齡、壓力水準與心理健康等特徵，透過聚類分析，識別出高生產力或低生產力的員工群體，以及辭職率較高的員工群體。

**II. 聚類分析**

1. **第一個數據集**
   * 基於Performance\_Score（績效評分）、Work\_Hours\_Per\_Week（每週工時）、Projects\_Handled（處理項目數）、Overtime\_Hours（加班時數）等變數，對員工進行聚類，識別出不同生產力水平的群體，並進一步分析這些群體中的工作習慣差異。
2. **第二個數據集**
   * 基於Stress\_Level（壓力水準）、Mental\_Health\_Condition（心理健康狀況）、Work\_Location（工作地點）等變數，對員工進行聚類，找出高壓力水平和低心理健康狀況的群體，並進一步分析這些群體與工作地點的關聯性。

**Evaluation**

在評估和驗證模型時，需確保模型在多個方面的表現都符合預期，以保證結果的可靠性和實用性。以下是具體的評估步驟：

* **效能指標**：
  + **分類模型**：使用**準確率**（Accuracy）、**精確度**（Precision）來評估模型的性能。這些指標有助於衡量模型在正確預測辭職或高績效員工方面的表現。
  + **聚類分析**：將使用人工觀察來評估模型性能。
* **效率評估**：
  + 對於較大的數據集（如包含大量員工的數據集），模型的**運行時間**和**計算資源消耗**需進行評估。運行時間過長或資源消耗過高的模型在實際應用中可能不具可行性，因此需要權衡模型的準確度和效率。
* **交叉驗證 (Cross-validation)**：
  + 為確保模型穩健性，應使用**k 折交叉驗證**。這種方法可以將數據集分成k個子集，交替進行訓練和測試，從而更好地評估模型在不同數據集上的性能表現，避免過度擬合。

### 4. ****時間規劃與挑戰 (Timeline and Challenges)****

* **時間表設置**：
  + 可以將專案分為數據清洗、特徵工程、模型訓練、模型評估與驗證、最終報告撰寫等階段。
    - **第1週**：數據整理和清洗，包括處理遺漏值、標準化和特徵轉換。
    - **第2週**：模型選擇與訓練，根據問題選擇分類、迴歸或聚類技術。
    - **第3週**：進行交叉驗證與模型優化，確保模型性能。
    - **第4週**：撰寫專案報告和結果展示。
* **挑戰預估**：
  + **數據獲取與質量**：可能出現數據集中的遺漏值或異常值，需要適當的數據清洗和填補策略。若有無法及時獲取的資料，應準備替代數據源。
  + **模型選擇與調優**：不同問題可能需要嘗試多種模型來找到最優解，這可能需要更多時間進行調參。若運行時間過長，需考慮使用分布式計算或更高效的模型。
  + **數據集大小與運行資源**：在處理大規模數據時，運行時間可能會過長，需考慮雲端資源或採用增量式學習方法來提高效率。

### 5. ****結論與未來工作 (Conclusion and Future Work)****

* **總結**：
  + 專案目標是利用分類、聚類，找出提升員工生產力和降低離職率的關鍵因素。基於兩個數據集的分析，我們通過不同的資料探勘技術探討了工作地點、壓力水準和心理健康等因素對員工績效和流失的影響。該專案展示了如何利用數據來為人力資源決策提供支持，並給出了相關的策略建議。
* **未來工作**：
  + **探索更多特徵**：未來可以引入更多的特徵（如員工培訓效果、工作投入程度等），以進一步提升模型的準確性和可解釋性。
  + **應用深度學習技術**：對於更大的數據集和更複雜的預測問題，可以嘗試使用深度學習技術來進行模型擬合和預測。
  + **跨公司數據比較**：將來可以進行跨行業、跨公司的數據比較，探索遠端工作和壓力管理策略在不同企業中的應用效果，提供更具普適性的建議。
  + **持續監控與優化**：模型在實際應用中，隨著新的數據進入，可以通過增量學習或持續監控進行模型的更新和優化，保證預測結果的準確性和實用性。