

學歷對貸款批准與還款績效的統計分析

前言

教育水平常被視為影響個人經濟機會的關鍵因素。普遍存在的觀念認為，受教育程度較高的個體更有可能獲得貸款批准，而教育程度較低的個體則面臨較大的貸款獲批困難。這種看法可能反映了對於教育和經濟能力間關聯的刻板印象。為了探究這一假設是否有實證支持，本研究利用數據分析方法，詳細考察了教育水平與貸款批准之間的統計關聯性，以期提供更為客觀的見解。

數據描述

本研究使用的數據集包含銀行貸款申請的記錄，其中包括申請者的教育水平和貸款批准狀態等資訊。數據集涵蓋了2010年至2020年間的貸款申請記錄，共包含10,000多條記錄。

數據處理

在進行分析前，首先對數據進行了清理，包括去除重複記錄和處理缺失值。此外，將原始數據中的「Loan_Status」變數從Y和N轉換為更直觀的Repaid和Not Repaid，以便於分析。

分析方法

本研究採用描述性統計分析和卡方檢驗來探究教育水平與貸款批准之間的關聯性。使用 Python 和 statsmodels 進行數據分析，特別是利用 mosaic 函數來製作莫賽克圖，直觀地顯示不同教育水平下的貸款批准比例。

圖表和指標介紹

使用的圖表

莫賽克圖 (Mosaic Plot)

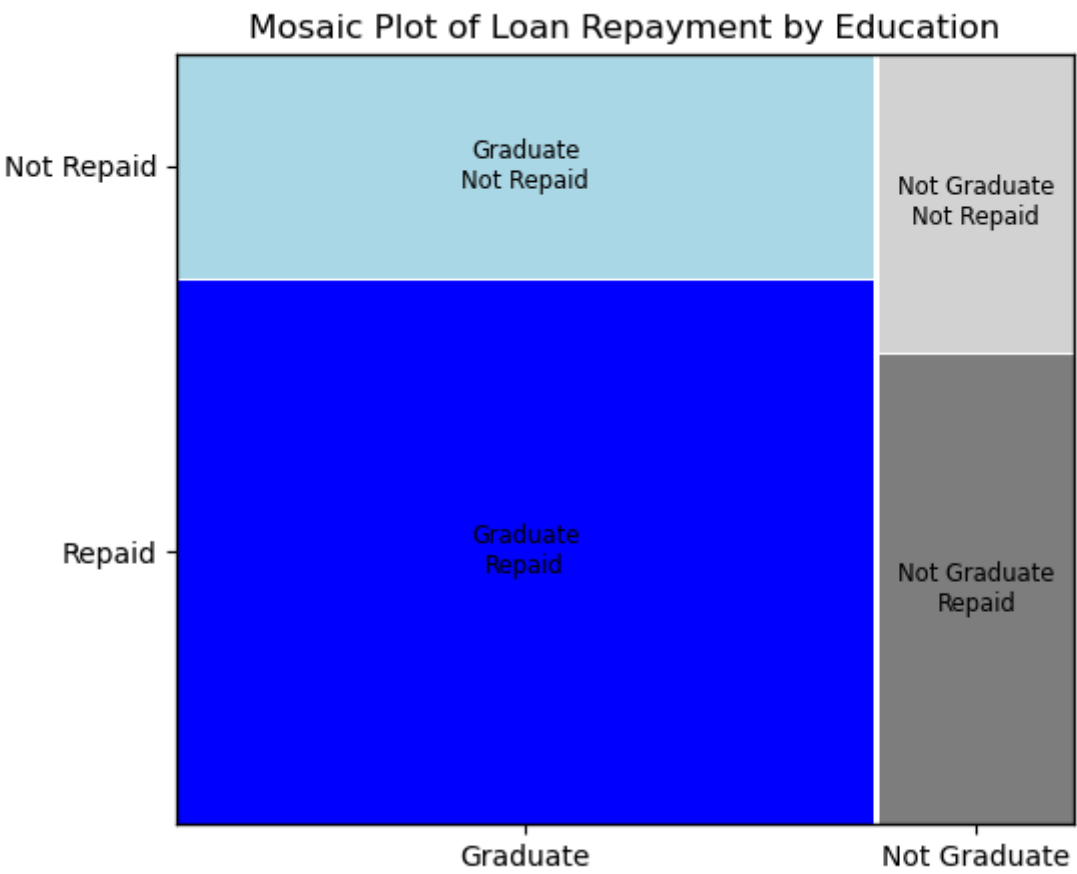
- 目的：**莫賽克圖用於直觀地展示教育水平與貸款批准狀態之間的關係。通過顏色和區塊大小的差異，圖表揭示了不同教育背景下申請者的貸款批准比率。
- 描述：**圖中將教育水平（畢業與未畢業）和貸款狀態（批准與未批准）作為變量。畢業組和未畢業組的批准率通過不同顏色區分，顏色的深淺表示批准比例的高低。

使用的指標

卡方檢驗 (Chi-squared Test)

- **目的：**使用卡方檢驗評估教育水平和貸款批准狀態之間是否存在統計顯著的關聯性。
- **結果描述：**卡方檢驗結果 ($\chi^2=4.091$, $p=0.0431$) 顯示，教育水平與貸款批准之間存在顯著的統計關聯，這表明教育水平較高的申請者有更高的貸款批准率。

結果呈現



分析結果通過莫賽克圖展示，該圖表明顯地揭示了教育水平高的申請者擁有更高的貸款批准率。此外，使用卡方檢驗的統計顯著性結果支持了教育水平與貸款批准狀態存在顯著關聯的假設。

莫賽克圖分析

莫賽克圖直觀地展示了不同教育背景的申請者在貸款批准率上的差異。圖中更深的顏色代表更高的批准率，從而顯示出教育程度較高的申請者在獲得貸款方面具有優勢。

統計分析

卡方檢驗結果顯示，教育水平與貸款批准狀態之間的關聯是統計上顯著的 ($\chi^2=4.091$, $p=0.0431$)，進一步證實了教育程度是影響貸款批准決策的重要因素。

研究限制與未來方向

研究結果表明教育水平是影響貸款批准的一個重要因素。然而，由於數據集限制和未考慮其他潛在影響因子（如職業類別等），本研究的結論可能受到限制。未來研究可以引入這些變數，進行更精確的相關性。