



資料分析與商業洞察

1. RFM客群分析
2. 公司治理與ESG分析
3. 決策樹與K-means
4. Google News+NLP

資料分析與商業洞察

RFM客群分析

專案簡介

本專案透過 RFM 模型建立複合指標，評估客戶價值，協助企業進行客戶分群與行銷策略規劃。專案流程涵蓋資料前處理、指標計算與視覺化，最終生成 RFM 立體分布圖，清楚呈現不同客群的價值差異。

主要功能

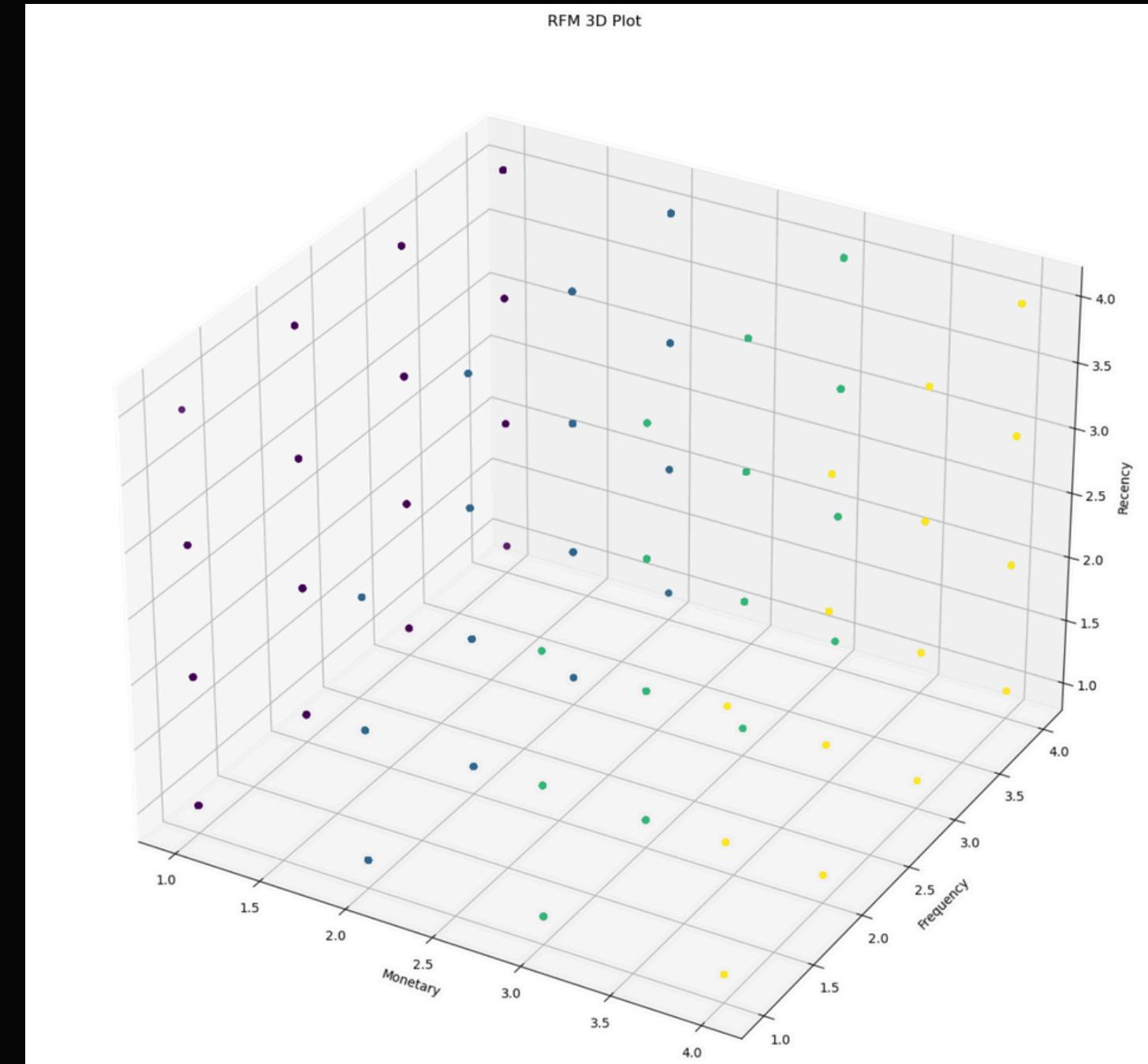
- 讀取並清理 EXCEL 資料集
- 新增總購買金額欄位 TOTALPRICE
- 計算每位客戶的 RECENCY、FREQUENCY 與 MONETARY
- 建立 RFM TABLE 並轉換為字典結構，便於查詢與分群
- 產生 RFM 3D PLOT，呈現各客群分布

使用技術

- PYTHON
- LAMBDA
- APPLY ()
- MATPLOTLIB

成果展示說明

展示圖片為 RFM 3D PLOT，顯示客戶在 RECENCY、FREQUENCY、MONETARY 三項指標的分布，便於辨識高價值、低互動等不同客群，為後續行銷策略提供依據。



資料分析與商業洞察

公司治理與ESG分析

專案簡介

本專案分析企業 ESG (ENVIRONMENTAL、SOCIAL、GOVERNANCE) 分數，探討不同產業的 ESG 表現及透明度對 ESG SCORE 的影響。透過統計檢定與線性回歸，評估各因素對 ESG 分數的重要性，並檢視資料建模的問題，如多重共線性與過擬合。

主要功能

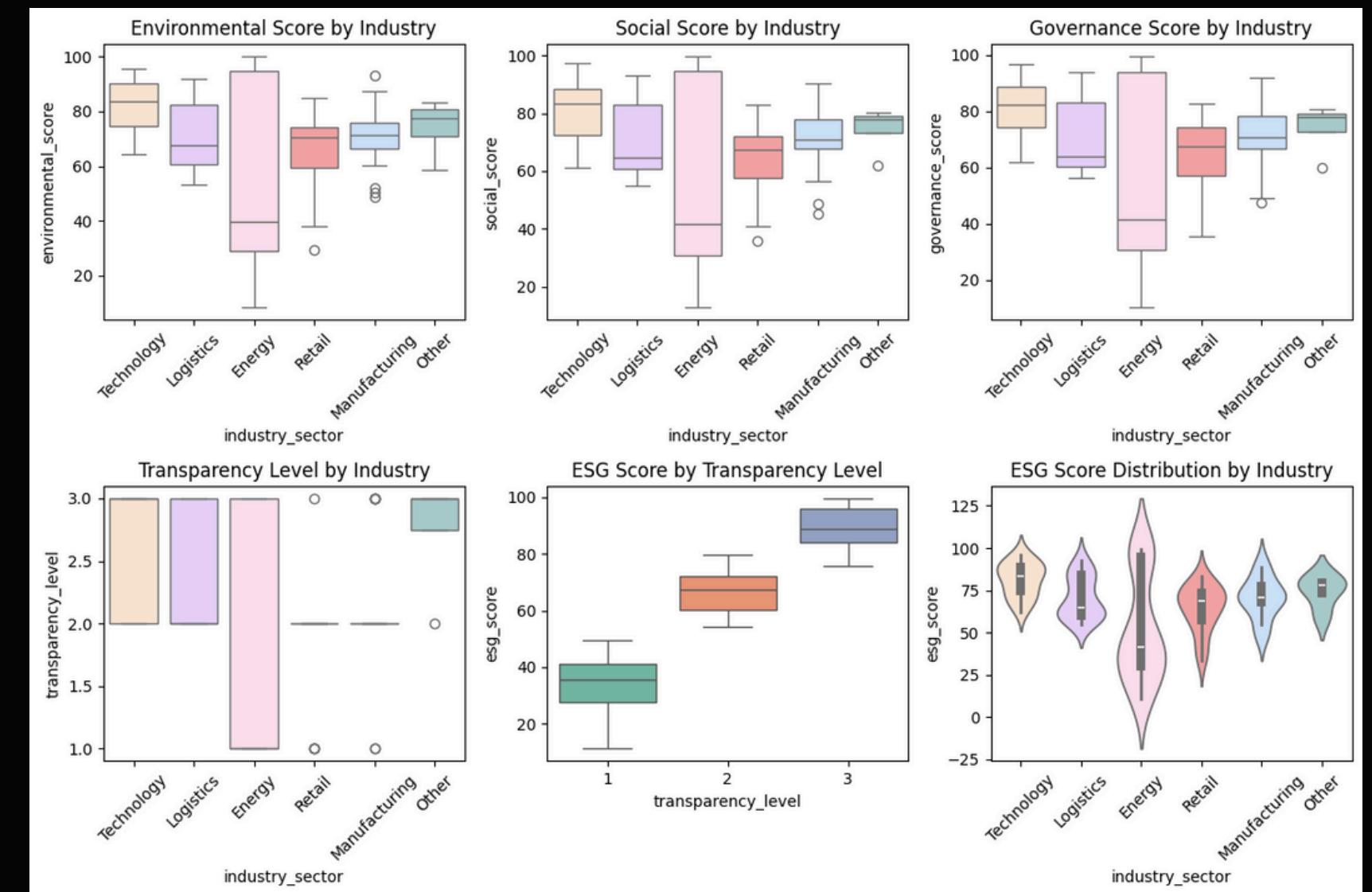
- 生成產業分析的視覺化圖表
- 使用 T 檢定分析不同組別間的差異
- 使用線性回歸模型分析 ESG SCORE 與各分數及透明度的關聯

成果展示說明

六張圖表呈現不同產業的 ESG 組成及透明度差異。回歸分析結果顯示 ESG SCORE 主要由 ENVIRONMENTAL、SOCIAL、GOVERNANCE 三項分數驅動，透明度與產業類別對 ESG SCORE 的影響不顯著。分析結果為企業永續策略提供定量參考。

使用技術

- PYTHON
- PANDAS、NUMPY
- MATPLOTLIB、SEABORN
- 統計檢定與線性回歸分析方法



資料分析與商業洞察

決策樹與K-MEANS

專案簡介

本專案包含兩個分析部分：K-MEANS 群聚分析與決策樹分類分析。透過 K-MEANS 對餐飲業資料進行分群，協助理解不同類型店家的營業特徵；決策樹部分則分析年齡與薪水對購買決策的影響，提供簡單的行為預測模型。

主要功能

- K-MEANS 部分
 - 讀取餐飲業 CSV 資料集
 - 進行資料清理與前處理
 - 使用 K-MEANS 計算群集中心並分配樣本群集
 - 利用 SSE 找出最佳群集數
 - 生成散佈圖展示各群集的分布特徵

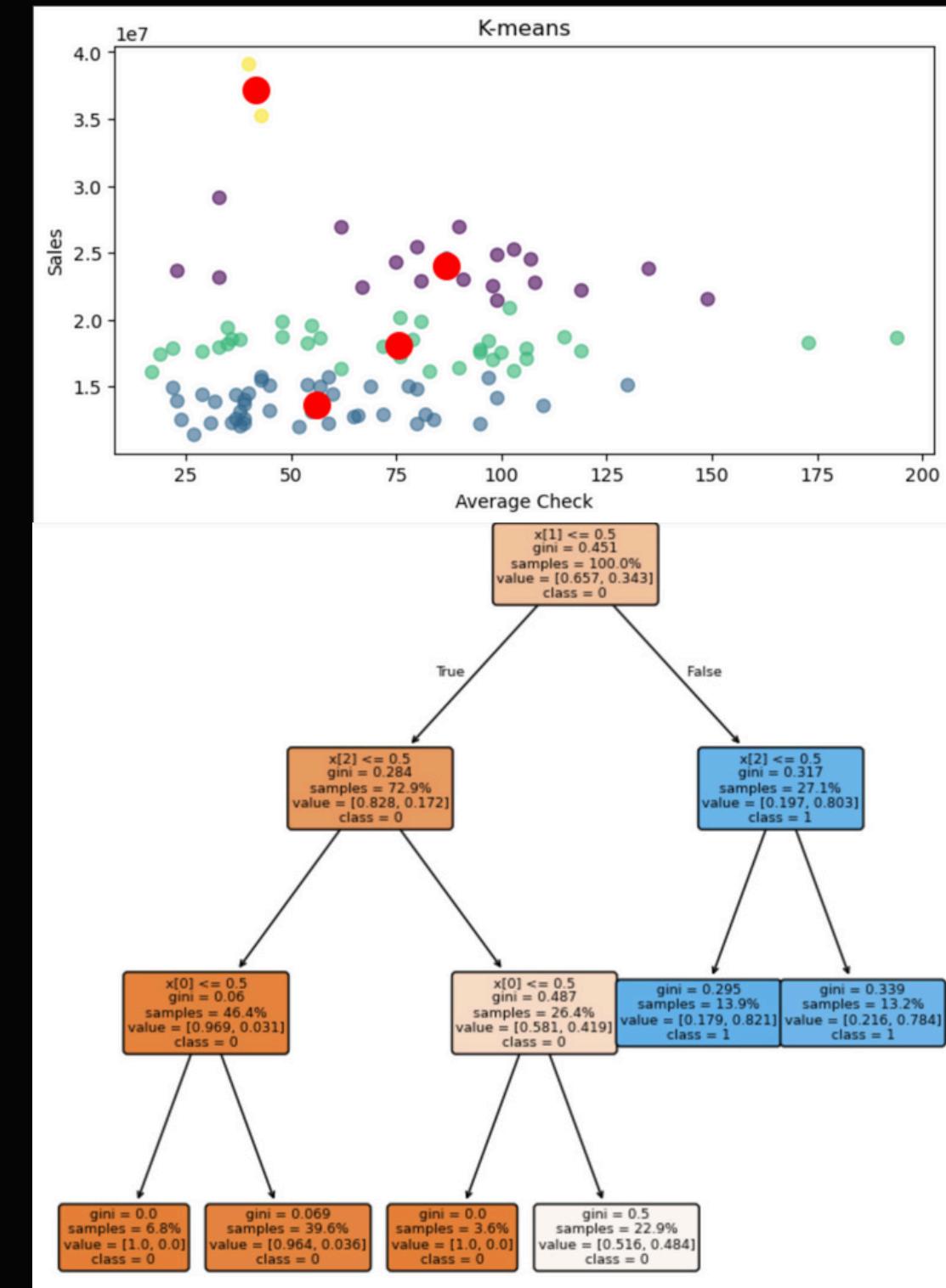
- 決策樹部分
 - 讀取購買行為 CSV 資料集
 - 將類別變數轉換為數值型
 - 將資料分為訓練集與測試集
 - 使用決策樹分類器進行模型訓練與預測
 - 評估模型準確率，將決策樹結構可視化

使用技術

- PYTHON
- PANDAS
- NUMPY
- K-MEANS
- DECISION TREE
- MATPLOTLIB
- SSE

成果展示說明

- K-MEANS 分群圖：四群分布清楚呈現，黃色群集營業額高但規模大，可能為速食店；紫色群集規模次之，平均消費略高，可能為西餐店；綠色群集規模中等，消費範圍廣，可能為咖啡廳或麵包店；藍色群集規模最小，可能為便利商店。
- 決策樹：可視化的樹形結構顯示年齡與薪水對購買決策的影響，模型準確率可作為初步行為預測參考。



資料分析與商業洞察

GOOGLE NEWS + NLP

專案簡介

本專案透過爬取 GOOGLE NEWS 資料，結合自然語言處理技術，對新聞文本進行清理、分詞與文字雲生成，協助快速掌握文章關鍵詞與內容主題，提升資訊整理效率。

主要功能

- 爬取 GOOGLE NEWS 上指定文章並進行資料排序
 - 對文章進行前處理：去除空格與無效符號
 - 使用分詞工具對文本進行切詞，包括精確模式與搜尋引擎模式
 - 生成文字雲，可視化關鍵詞出現頻率
 - 對長詞進行再切分，提高召回率並增強文本分析效果

成果展示說明

展示圖片為文字雲結果，直觀呈現文章中最頻繁出現的關鍵詞，反映文本的主題分布與重點概念。透過精確模式與搜尋引擎模式的切詞對比，可有效捕捉長詞與重要資訊，提高後續分析或搜尋效率。

使用技術

- PYTHON
 - JIEBA
 - WORDCLOUD
 - MATPLOTLIB

