

TIBAME GAD245



AI與風險評估

汽車保險理賠數據洞察

Car Insurance Claim Prediction

PRESENTED BY SYLVIA YANG

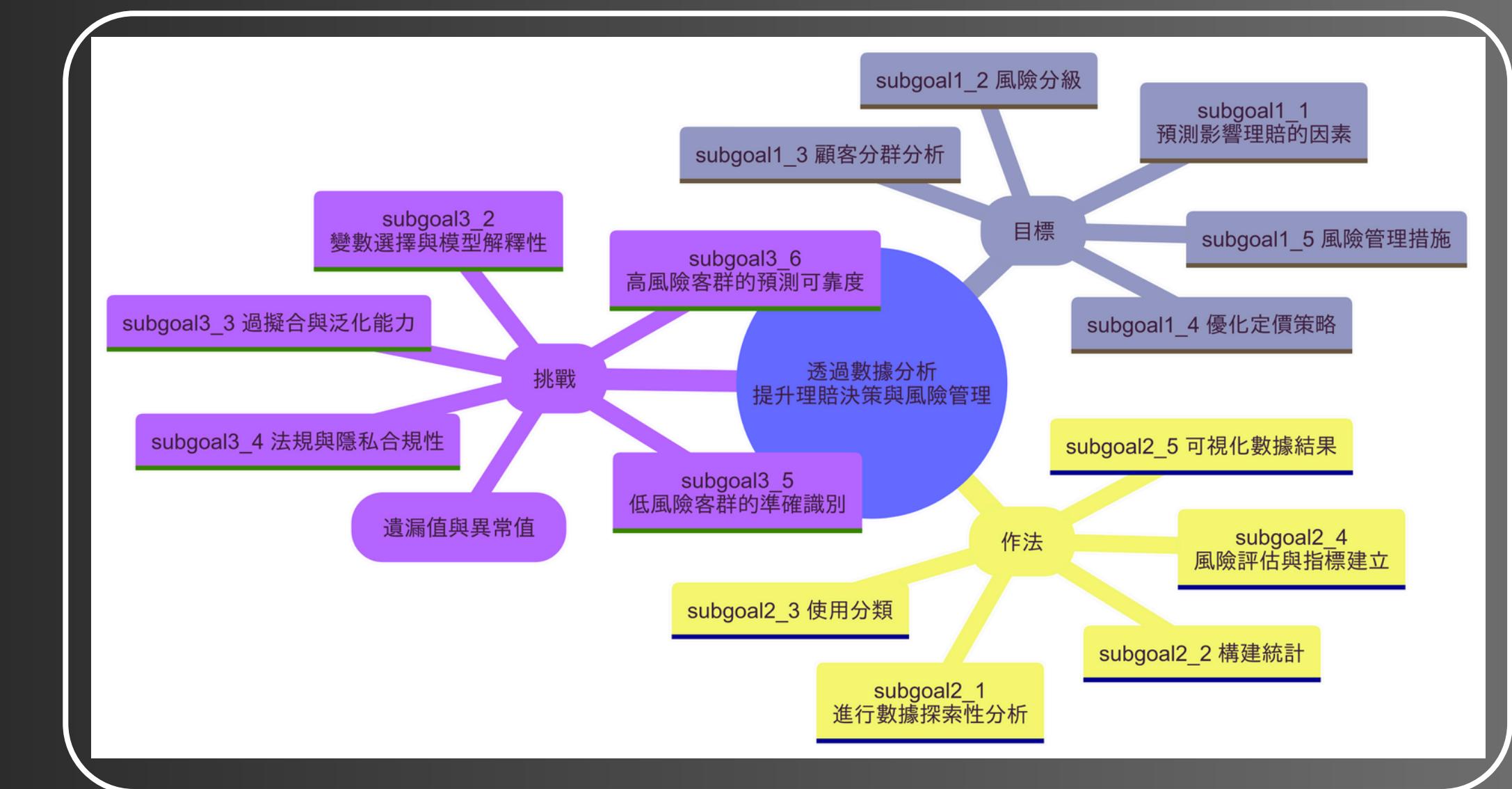


專案目標：

透過數據分析 提升理賠決策與風險管理

主要關注點：

- 影響理賠的關鍵因素
- 風險分級
- 客戶分群
- 個性化風險管理建議
- 決策建議



* * *

- 數據來源

kaggle:

<https://www.kaggle.com/datasets/ifteshanajnin/carinsuranceclaimprediction-classification/data>

- 資料內容

資料集包含保單持有人的信息，包括保單期限、汽車年齡、車主年齡、城市人口密度、汽車品牌和型號、功率、發動機類型等屬性，以及指示的目標變量保單持有人是否在未來6個月內提出索賠。

- 分析方法

1. 風險分級（依據理賠機率將客戶分類）
2. 顧客分群
3. 交叉分析

- 使用工具

1. 數據處理：Python (Pandas, NumPy)
2. 資料視覺化：Matplotlib、Seaborn、Tableau
3. 簡報呈現：canva

1. 問題識別 (Problem Identification)

目前保險公司在理賠預測和風險管理方面面臨多項挑戰，包括：

高理賠成本：由於風險評估機制不夠精確，導致高風險客戶的理賠支出較高。

定價不準確：保費定價與客戶的實際風險不匹配，影響競爭力與盈利能力。

風險分級困難：缺乏有效的數據分析工具，難以準確識別高風險與低風險客戶。

客戶體驗不足：無法提供個性化的保險方案與預防建議，導致客戶滿意度下降。

2. 解決方案 (Solution)

針對上述問題，我們提出了一套基於數據分析與機器學習的解決方案，包括：

數據驅動的風險評估：運用數據探索與機器學習模型，識別影響理賠的關鍵因素，如車齡、駕駛紀錄、地理位置等。

顧客分群分析：透過 Random Forest，細分客戶群體，提供更精準的風險評估與產品推薦。

個性化風險管理建議：針對高風險客戶提供駕駛安全建議，降低潛在理賠風險，提高客戶忠誠度。

客戶體驗，最終達成雙贏的目標。

3. 價值創造 (Value Proposition)

該專案的核心價值體現在以下幾個方面：

提升保險公司盈利能力：降低不必要的理賠支出，提升保單獲利能力。

優化風險管理：提高風險識別精準度，減少保險詐欺與高風險保單的影響。

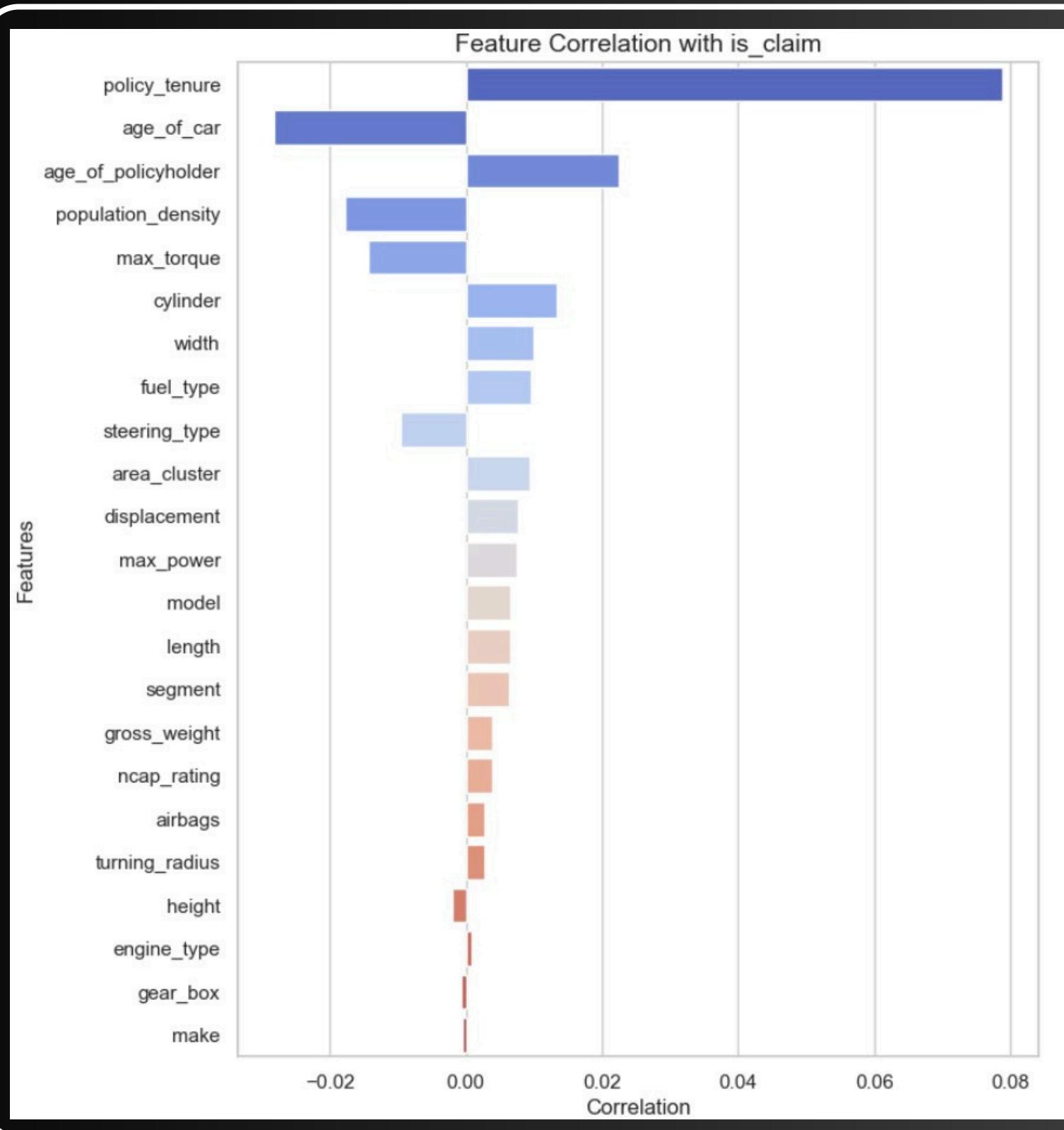
提高客戶滿意度：提供透明、公正的保費定價與個性化建議，提升客戶信任。

市場競爭優勢：透過數據驅動的決策，提高產品定價與風險管理的靈活性，鞏固市場地位。

透過這個專案，保險公司可以更科學地管理風險，提供更具競爭力的產品，並提升客戶體驗，最終達成雙贏的目標。



重要欄位說明-前10個重要的變數：



- POLICY_TENURE (保單持有時間, 相關性 0.078)
持有保單時間越長，理賠的可能性越高，可能與長期持有有關。
- AGE_OF_CAR (車齡, 相關係數 -0.0282)
車齡越高，理賠發生率越低，可能因為舊車投保的車主駕駛經驗較多。
- AGE_OF_POLICYHOLDER (保單持有人年齡, 0.022)
年齡越大，理賠的可能性略高，可能與年長反應速度下降有關。
- POPULATION_DENSITY (人口密度, -0.0178)
人口密度較高的地區，理賠的可能性較低，可能因為都市地區車速較低，事故較少。
- MAX_TORQUE (最大扭力, -0.014)
扭矩較高的車輛可能屬於高性能車型，這些車主可能更熟練駕駛，降低事故風險。
- CYLINDER (氣缸數, 0.013)
氣缸數較多的車輛動力較大，可能影響風險。
- WIDTH (車寬, 0.010)
較寬的車可能在特定情境下較安全或較危險。
- FUEL_TYPE (燃料類型, 0.010)
不同燃料類型可能影響駕駛行為或維護成本。
- STEERING_TYPE (轉向類型, -0.009)
較高級的轉向系統可能搭載在更安全的車輛上，進一步降低事故風險。。
- AREA_CLUSTER (地區群集, 0.009)
不同地區可能有不同的事故風險。

長條圖-地區風險客戶數量:

長條圖顯示：

C8區域的保單數量最高，超過13,000筆，大部分為低風險。

C2區域的保單數量也較高，超過7,000筆，少量中高風險客戶。

多數區域以低風險客戶為主，中高風險客戶所占比例較小。

圓餅圖-客戶風險等級:

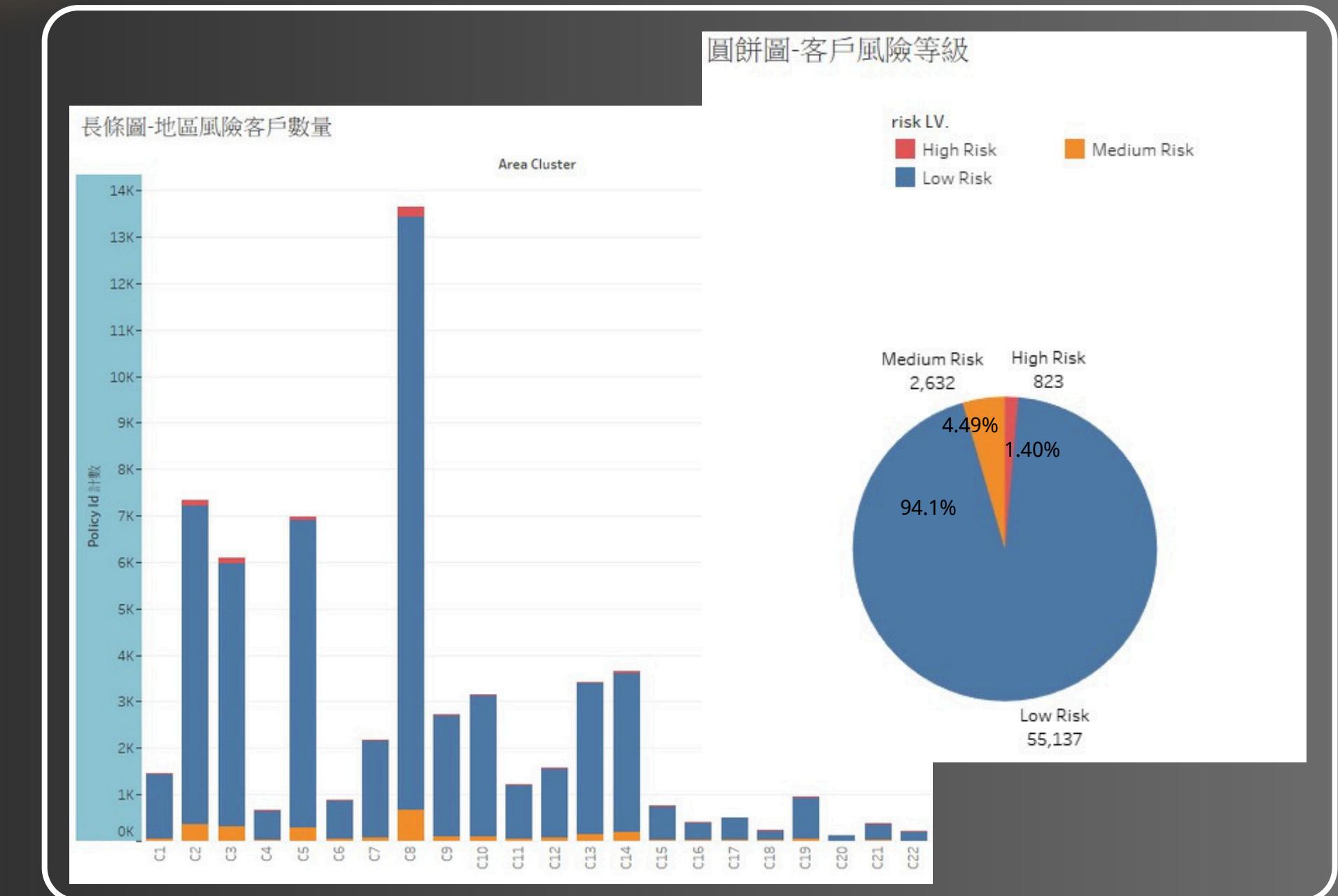
圓餅圖顯示，大部分客戶屬於低風險類別，高風險客戶的占比極小。

分析結論：

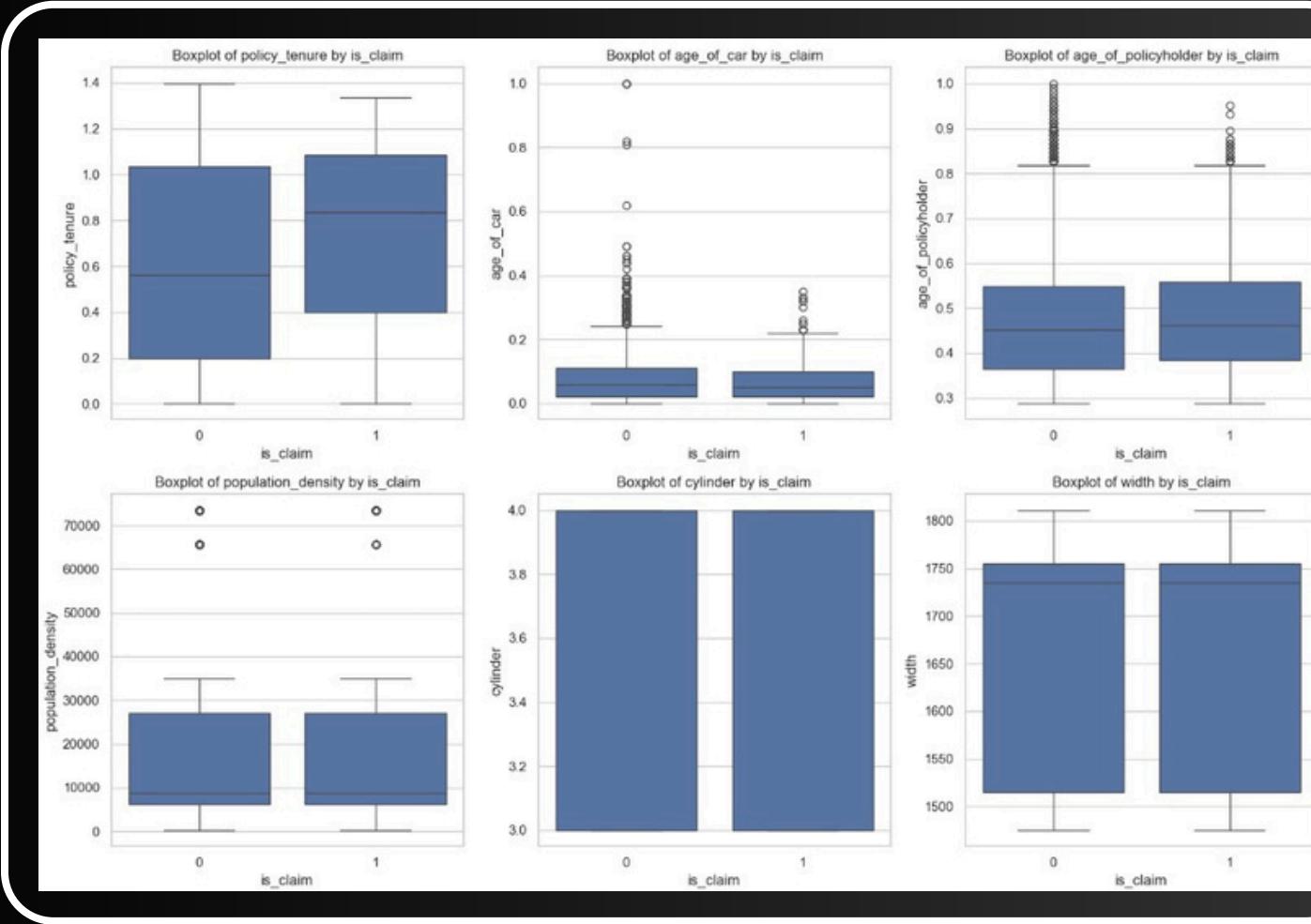
保險客戶中，以低風險客戶居多，超過94%，高風險客戶相對較少。

在某些特定區域（如C8），保單數量極高，可能是主要的市場重點。

可對高風險區域進行額外評估，或在低風險區域擴展業務。



分析圖表與解釋-2



3. AGE_OF_POLICYHOLDER (保單持有者年齡) BY IS CLAIM

兩組的分布幾乎相同，顯示年齡在不同理賠情境下變化不大。
上方有大量異常值 (outliers)，可能是年長投保者，但對理賠影響似乎不大。
下面三張圖分布相似，代表對理賠影響不大。

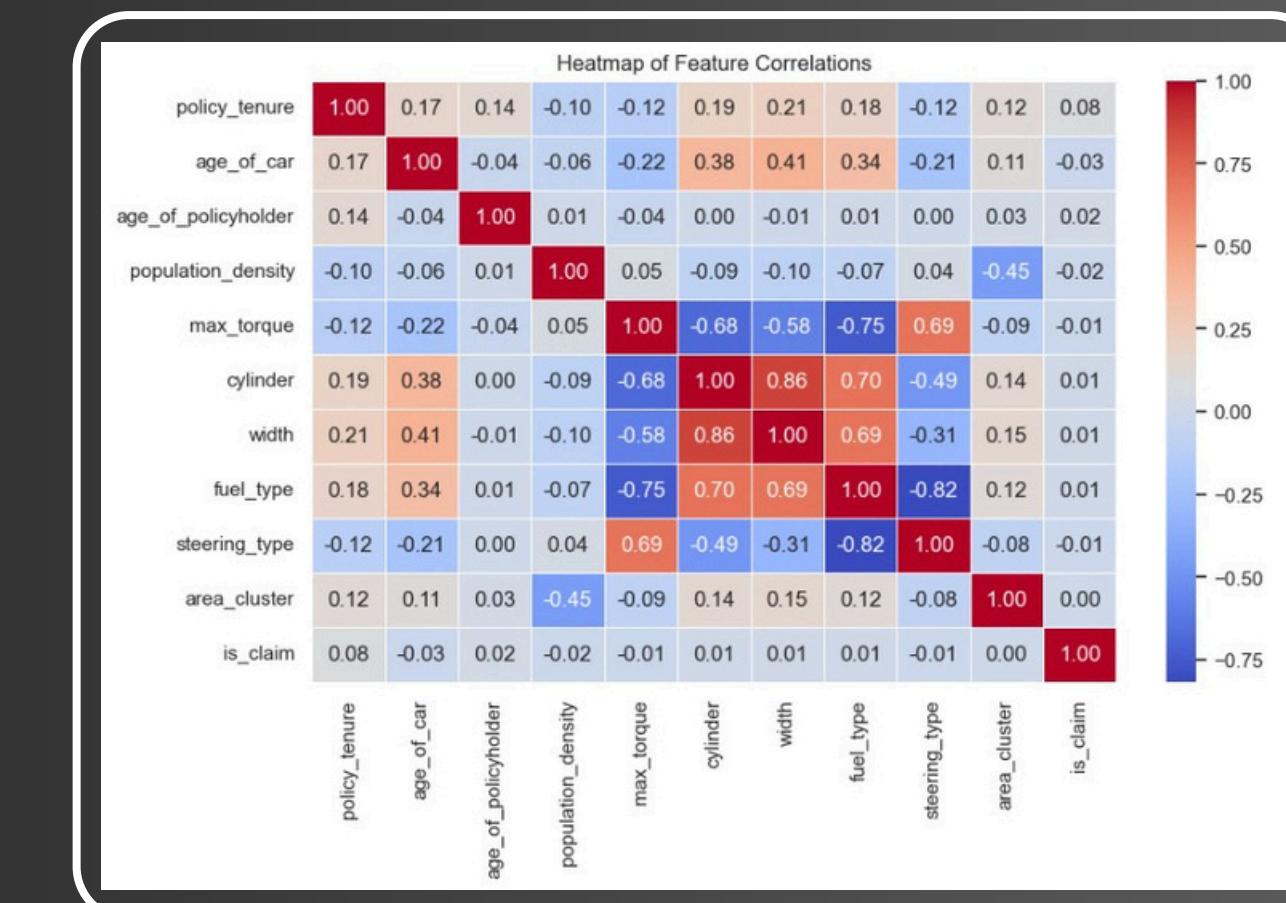
箱型圖：

1. POLICY_TENURE (保單持有時間) BY IS CLAIM

當 $is_claim = 0$ (未理賠) 和 $is_claim = 1$ (有理賠) 時， $policy_tenure$ 的分布相似，但理賠者的中位數略高。
理賠群體 ($is_claim = 1$) 有較多長期持有者，可能表示長期持有的保單更容易出險。

2. AGE_OF_CAR (車齡) BY IS CLAIM

大部分車齡接近 0.1~0.2，表示車輛普遍較新。
但在 $is_claim = 0$ 和 $is_claim = 1$ 的分布幾乎相同，代表車齡對理賠影響不大。
存在很多異常值 (outliers)，特別是在 $is_claim = 0$ 組別中，可能表示一些年齡較大的車輛被納入分析。



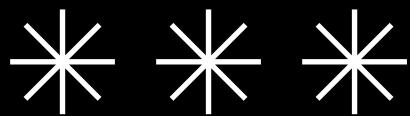
熱圖：

用顏色表示變數之間的相關性

顯示
 $policy_tenure$ 、
 age_of_car 、
 $age_of_policyholder$
對是否理賠具有較高的正或負
相關，

表示這些變數可能影響客戶的
理賠行為。





從DIKW 金字塔 (Data, Information, Knowledge, Wisdom) 的角度，來分析這個專案的價值：

1. 數據 (Data):

專案的基礎在於收集和處理大量的原始數據。這些數據包括：

- 車齡：不同年份車輛的理賠紀錄。
- 保單持有時間：駕駛人持有保單的時長。
- 地理位置：不同地區的理賠發生頻率。
- 客戶基本資料：例如年齡、性別、職業等可能影響風險的因素。
- 保險理賠紀錄：歷史理賠數據、原因等資訊。
- 這些數據是專案的基石，沒有這些原始數據，就無法進行後續的分析和預測。

2. 資訊 (Information):

透過對原始數據的處理和分析，我們將數據轉換成有意義的資訊。這包括：

- 理賠關鍵因素：透過數據探索和機器學習模型，找出哪些因素（例如車齡、保單持有時間）是影響理賠的主要原因。
- 風險評估結果：分析數據後，得出不同客戶的風險等級，例如高風險、中風險、低風險。
- 客戶分群結果：利用 K-means 或層次聚類方法，將客戶分為不同的群體，例如高風險駕駛群、低風險駕駛群等。
- 市場趨勢分析：分析不同地區、不同車型、不同客戶群體的理賠狀況，了解市場趨勢。
- 這些資訊可以幫助保險公司更了解現況，並為後續的決策提供依據。

3. 知識 (Knowledge):

將資訊整合和解釋，形成對保險理賠風險的知識。這包括：

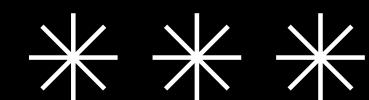
- 風險管理策略：針對不同風險等級的客戶，制定相應的風險管理策略，例如提供安全駕駛建議、調整保險條款等。
- 產品推薦策略：根據客戶分群，推薦更適合其風險狀況的保險產品。
- 客戶行為模式：了解不同客戶群體的行為模式，例如高風險駕駛人可能有哪些特徵，以便更精準地評估風險。

這些知識能夠幫助保險公司制定更有效的營運策略，並提升競爭力。

4. 智慧 (Wisdom):

在知識的基礎上，運用洞察力和判斷力，做出對保險公司整體發展有益的智慧決策。這體現在：

- 提升保險公司盈利能力：降低不必要的理賠支出，提升保單獲利能力，將資源有效配置。
- 優化風險管理：精準識別風險，減少保險詐欺和高風險保單的負面影響，實現風險控管的最大效益。
- 提高客戶滿意度：提供透明、公正的保費定價和個性化服務，提升客戶信任和忠誠度，建立長期的客戶關係。
- 輯固市場競爭優勢：透過數據驅動的決策，提高產品定價和風險管理的靈活性，保持市場領先地位。
- 建立雙贏局面：藉由更精準的定價和風險管理，達到保險公司和客戶的雙贏局面。



開發其他區域市場:

1. C8區域與其他區域保單數量有明顯差異，可開發其他區域提高市佔率。

使用更精細的風險評分模型:

1. 透過大數據與AI技術，提升風險評估的精準度，確保保費能夠合理反映客戶的風險等級。
2. 結合行為數據、信用評分、醫療紀錄、駕駛習慣等多維度資訊，提高風險識別能力。

動態定價:

1. 採取風險基礎定價策略，讓高風險客戶支付更高的保費，補償可能的理賠支出。
2. 可以透過車載設備（Telematics）監測駕駛行為，提供風險適應型的保費方案。



設定更嚴格的承保條件:

1. 提高高風險客戶的自負額（Deductible），讓客戶在小額損失時需自付較高比例，以降低理賠次數。
2. 設置理賠上限，確保單一客戶不會造成過大財務負擔。
3. 對高風險客戶限制特定險種，例如對於曾多次理賠的客戶，限制某些高風險保險產品。

推出行為改變型保單:

1. 提供安全駕駛獎勵，若客戶符合良好駕駛行為，可降低保費或享有回饋。

風險教育與行為管理:

1. 要求高風險駕駛完成安全駕駛課程，以降低事故發生率。
2. 針對高風險區域或職業提供專屬產品
3. 若特定區域的高風險客戶占比高，可推區域性定價方案或限制特定險種。
4. 針對高風險職業（如職業司機、賽車手），設計專門的保險產品，提高風險控制能力。

- 總結

本次分析揭示了影響理賠發生的三大因素：

- 保單持有時間 (Policy Tenure) 是最具影響力的變數，顯示出長期持有保單的客戶更可能申請理賠，可能與長期累積的風險有關。
- 車齡 (Age of Car) 雖有微弱相關性，但整體影響不顯著，代表新舊車輛的理賠機率差異不大，需結合其他變數進一步分析。
- 保單持有人年齡 (Age of Policyholder) 對理賠的影響有限，但異常值偏多，可能與非主要駕駛人的影響有關，值得進一步探討。

這些結果顯示，目前的風險評估模型仍有優化空間，

為了提升理賠決策的準確度與保險公司的競爭力，建議：

1. 採用更精細的風險評分模型 —— 整合更多變數，提升理賠預測能力，降低高風險保單的財務影響。
2. 推動風險基礎定價策略 —— 讓高風險客戶承擔更合理的保費，確保保單的可持續性。
3. 提供安全駕駛獎勵機制 —— 鼓勵良好駕駛行為，減少事故率，降低理賠成本。
4. 調整高風險客戶的自負額 —— 透過適當的自負額設計，讓客戶分擔部分風險，減少小額頻繁理賠的情況。

透過這些策略，不僅能優化風險管理，也能提升保險公司的獲利能力，同時提供更公平且個性化的保障方案，實現客戶與公司的雙贏局面。

END

PAGE 10

THANK YOU AND STAY CONNECTED

We appreciate your time. If you have any questions,
feel free to reach out!

YOYOYA1007@GMAIL.COM *