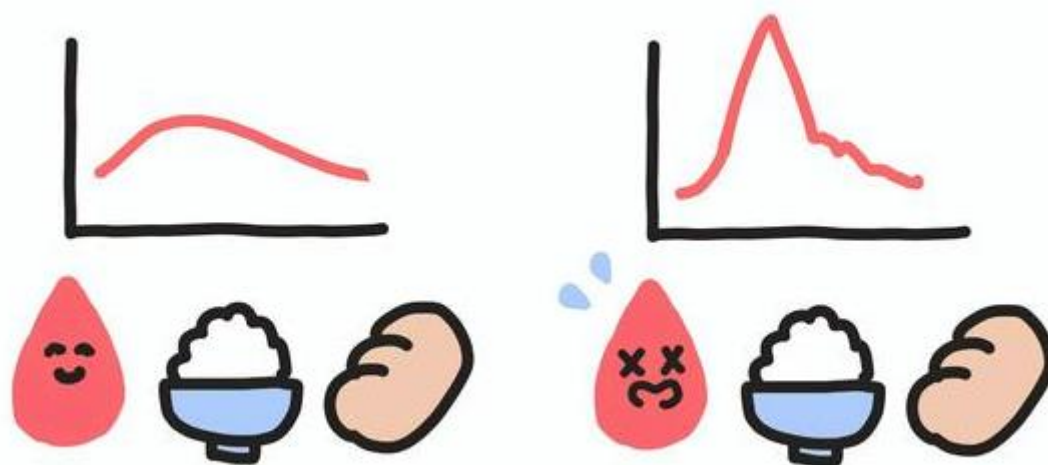


AI-糖尿病預測模型



摘要

- 探討各項危險因子與**DM**的關聯性
 - 利用Domain knowledge判斷與預測數據分析結果
- 藉由個案的資料來預測是否患有**Diabetes**
 - 藉由撰寫Python程式進行數據分析
 - 建立演算法模型呈現結果與對應的關聯性
- 以分析預測數據為依據協助醫療人員執行最適合的醫療處置

Background

- 中山醫學大學 營養學系
- [iSpan資展國際股份有限公司\(原資策會\)](#)
 - AIoT智慧機器人就業養成班
 - 資料分析: Numpy, pandas, matplotlib, seaborn
 - AI相關技術: Scikit learn, Tensorflow, OpenCV
 - 專題:AI臉部辨識化妝品推薦系統



陳冠宇

目錄

- [摘要](#)
- [背景](#)
- [前言](#)
 - [專業能力](#)
 - [領域知識](#)
- [糖尿病診斷標準](#)
- [Skills](#)
- [工作流程](#)
- [Data Analysis](#)
 - [資料型態與來源](#)
 - [資料前處理與分析](#)
- [建立預測模型](#)
 - [LogisticRegression](#)
 - [Confusion Matrix](#)
- [Disscusion](#)
 - [研究限制](#)
 - [未來發展](#)
 - [女性vsDM](#)
- [總結](#)
- [Renference](#)

糖尿病診斷標準

	項目	正常	糖尿病前期	糖尿病
1	糖化血色素%	<5.7	5.7-6.4	≥6.5
2	空腹血糖mg/dl	<100	100-125	≥126
3	口服75公克葡萄糖第2小時 血糖mg/dl	<140	140-199	≥200
4	高血糖症狀且隨機血糖 mg/dl	-	-	≥200

除明顯高血糖症狀，須至少2次異常才可診斷

Skills:

1. 資料清整
2. 資料分析EDA
 - 資料視覺化
 - 數據代表性與關聯
3. 建立演算法模型與優化
4. 探討分析結果與學理的差異



WorkFlow



A light blue rectangular graphic with rounded corners, designed to look like a scroll. It features a vertical strip on the left side and a small circular element at the top right corner, suggesting a rolled-up document.

數據分析

Data Analysis

資料型態與來源

<https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data>

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0
...
99995	Female	80.0	0	0	No Info	27.32	6.2	90	0
99996	Female	2.0	0	0	No Info	17.37	6.5	100	0
99997	Male	66.0	0	0	former	27.83	5.7	155	0
99998	Female	24.0	0	0	never	35.42	4.0	100	0
99999	Female	57.0	0	0	current	22.43	6.6	90	0

100000 rows × 9 columns

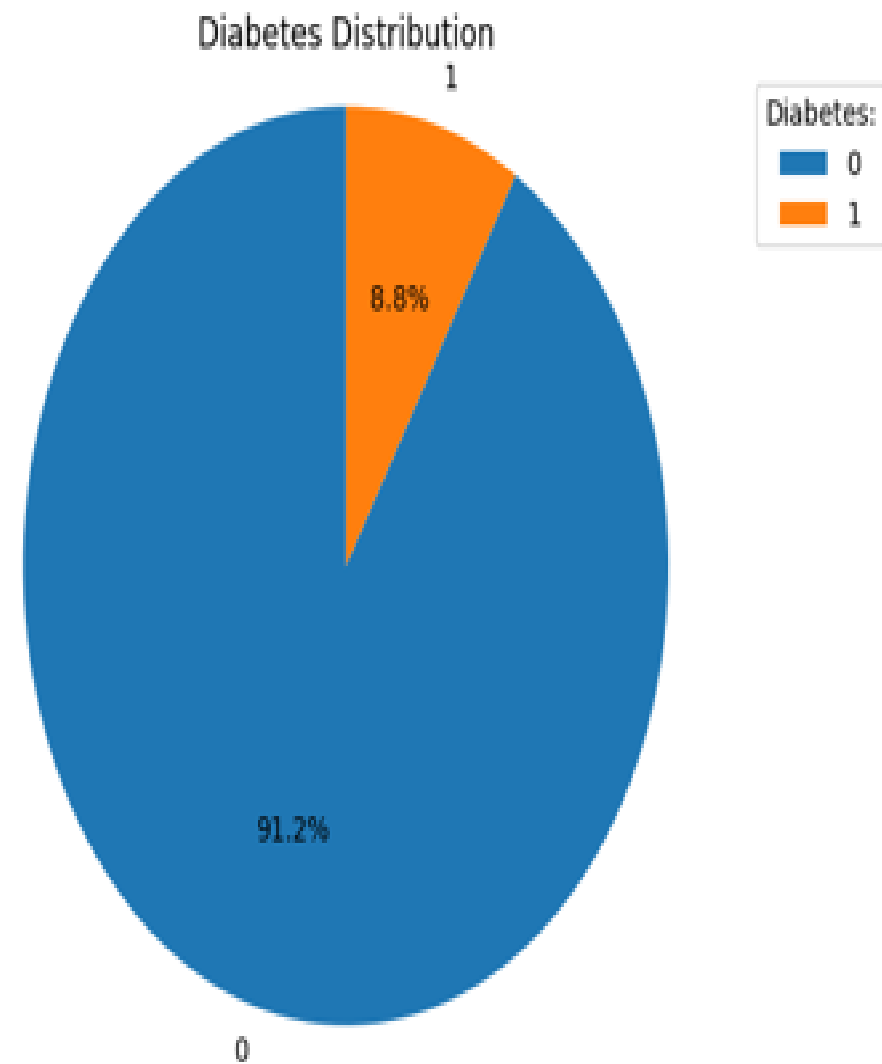
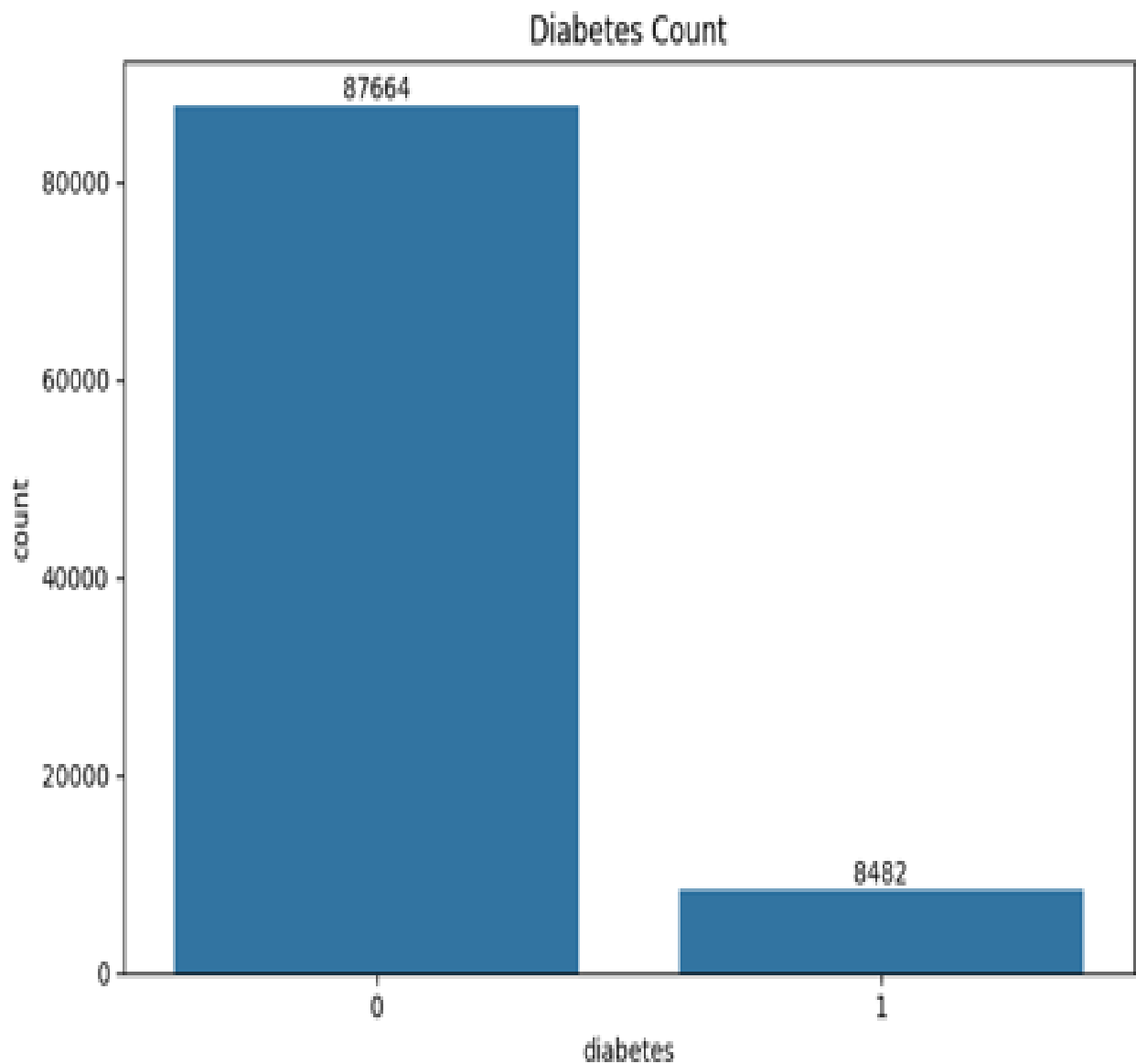
➤ Predict Key Factor: age & HbA1c level

資料前處理與分析

	age	hypertension	heart_disease	bmi	HbA1c_level	blood_glucose_level	diabetes
count	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000
mean	41.885856	0.07485	0.039420	27.320767	5.52750	138.058060	0.085000
std	22.516840	0.26315	0.194593	6.636783	1.07067	40.708136	0.278883
min	0.080000	0.00000	0.000000	10.010000	3.500000	80.000000	0.000000
25%	24.000000	0.00000	0.000000	23.630000	4.800000	100.000000	0.000000
50%	43.000000	0.00000	0.000000	27.320000	5.800000	140.000000	0.000000
75%	60.000000	0.00000	0.000000	29.580000	6.200000	159.000000	0.000000
max	80.000000	1.00000	1.000000	95.690000	9.000000	300.000000	1.000000

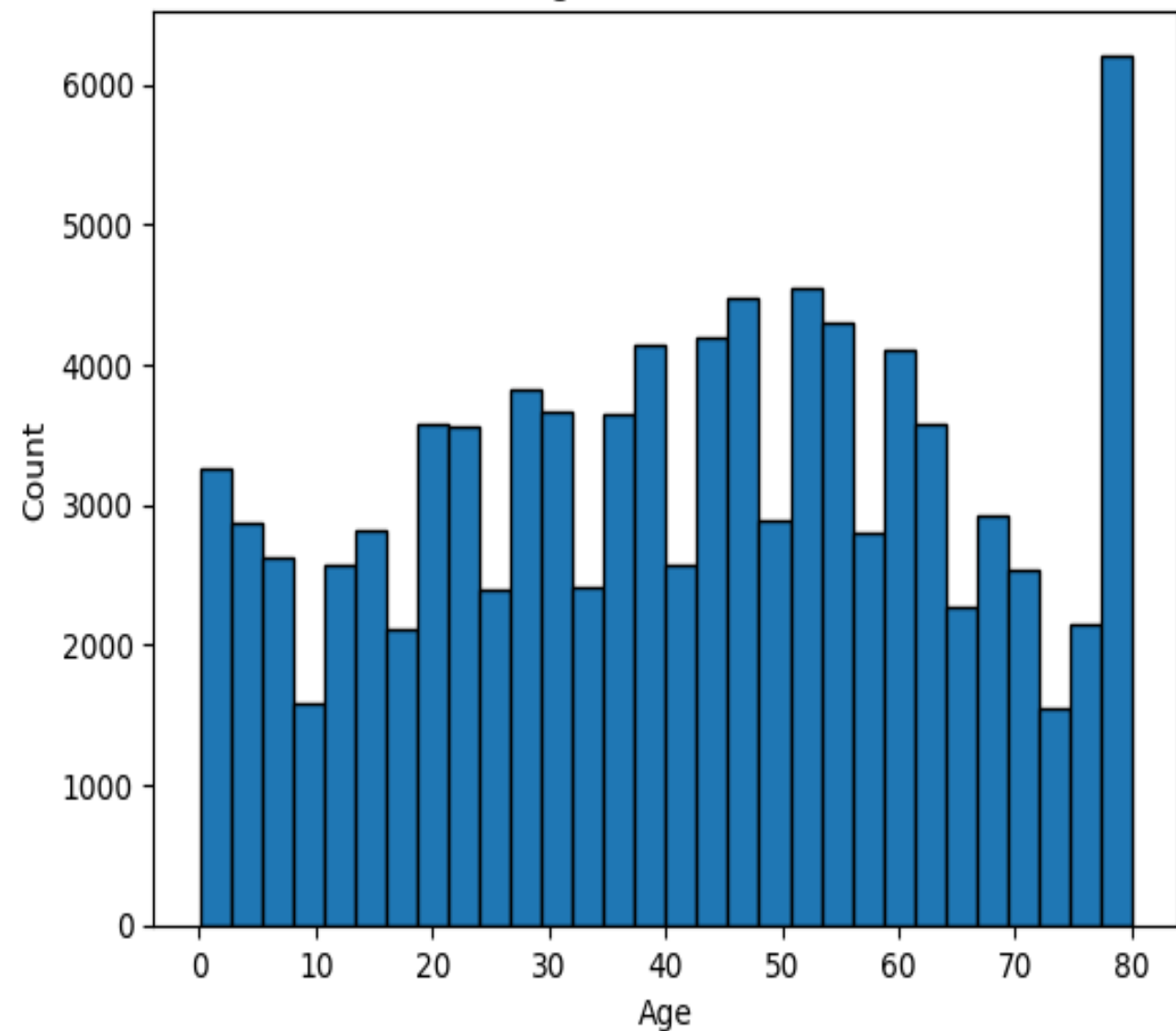
Domain knowledge: 分析前就能找出資料問題點

資料前處理與分析

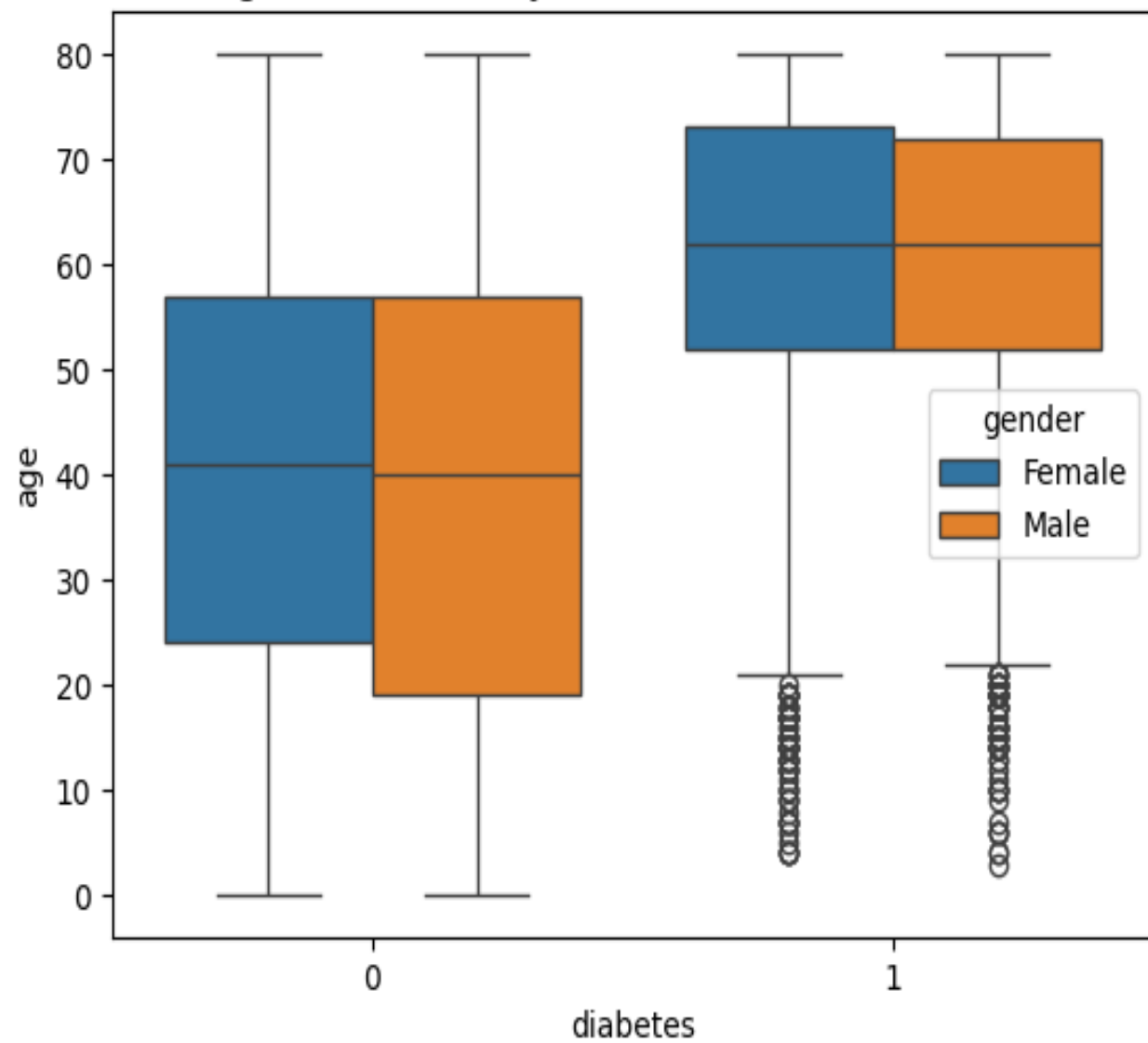


資料前處理與分析

Age Distribution

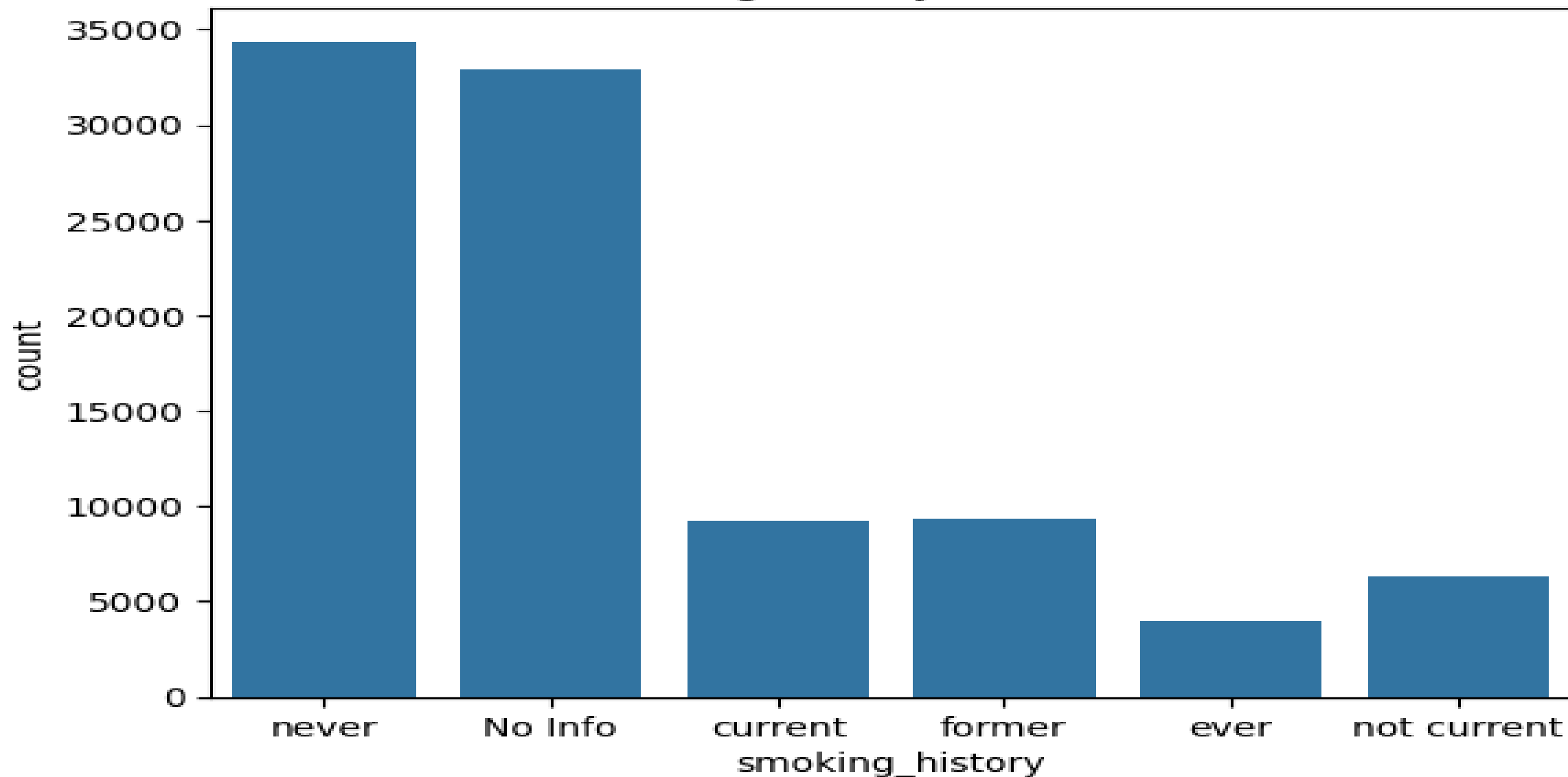


Age Distribution by Diabetes Status and Gender



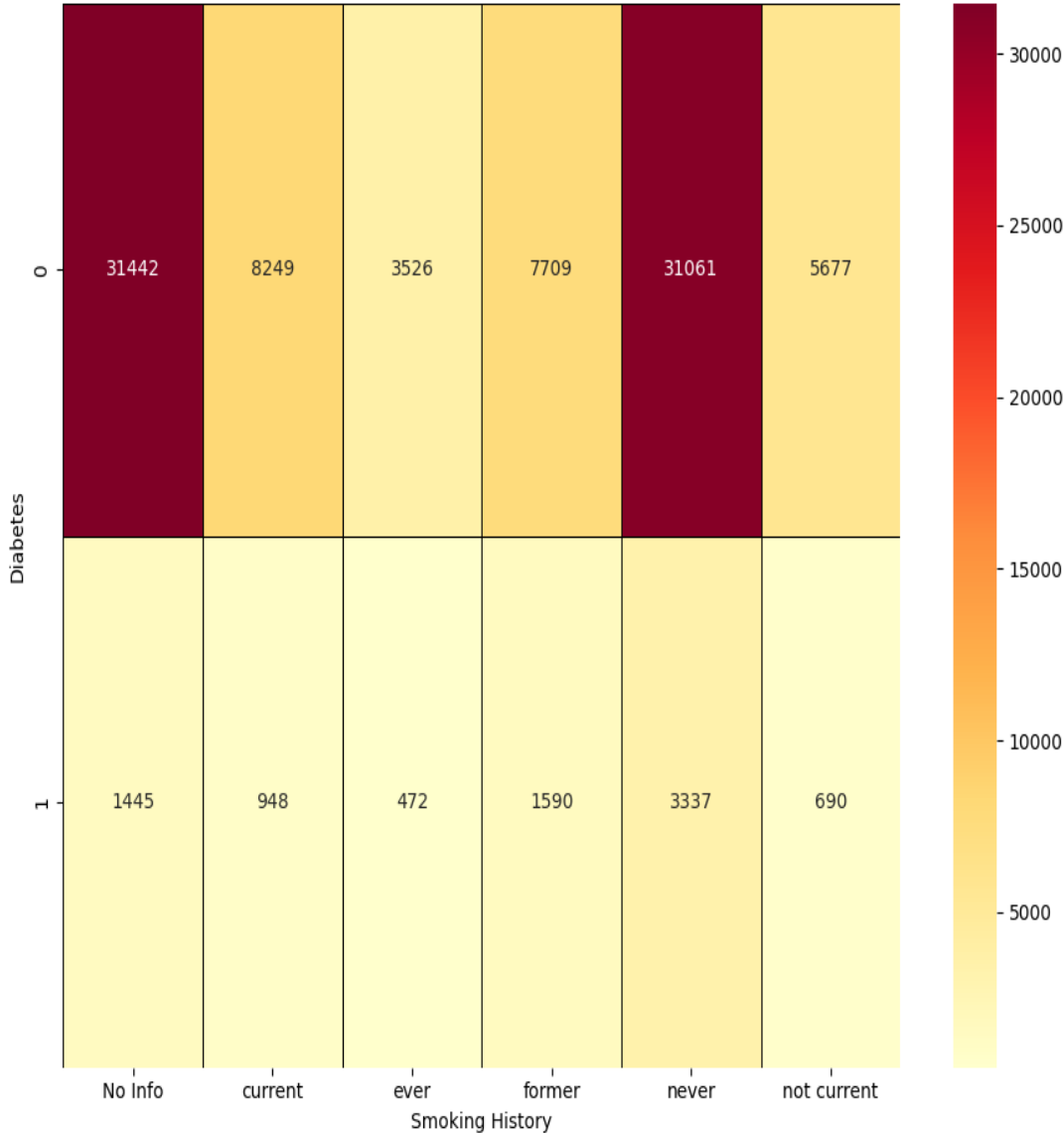
資料前處理與分析

Smoking History Distribution

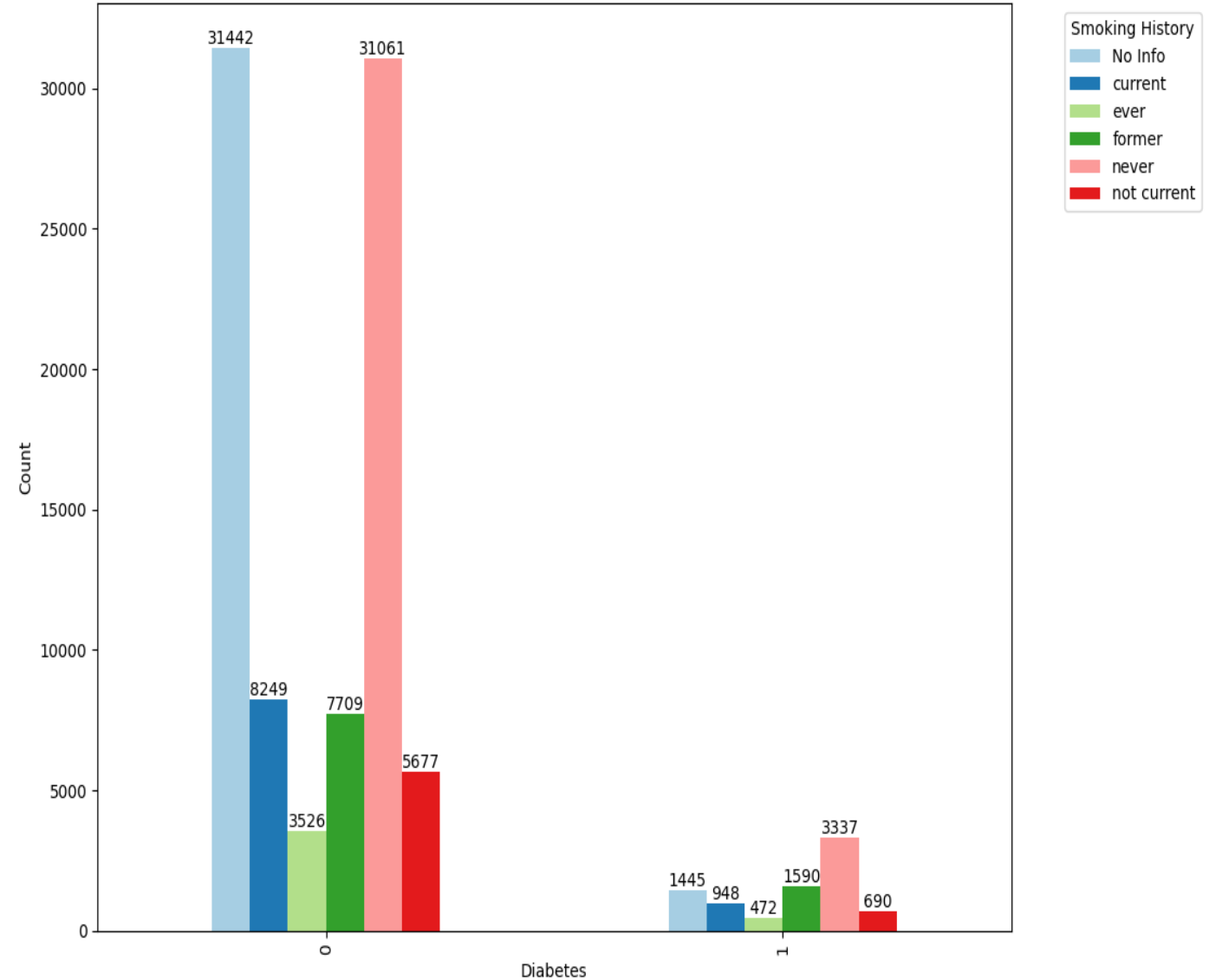


資料前處理與分析

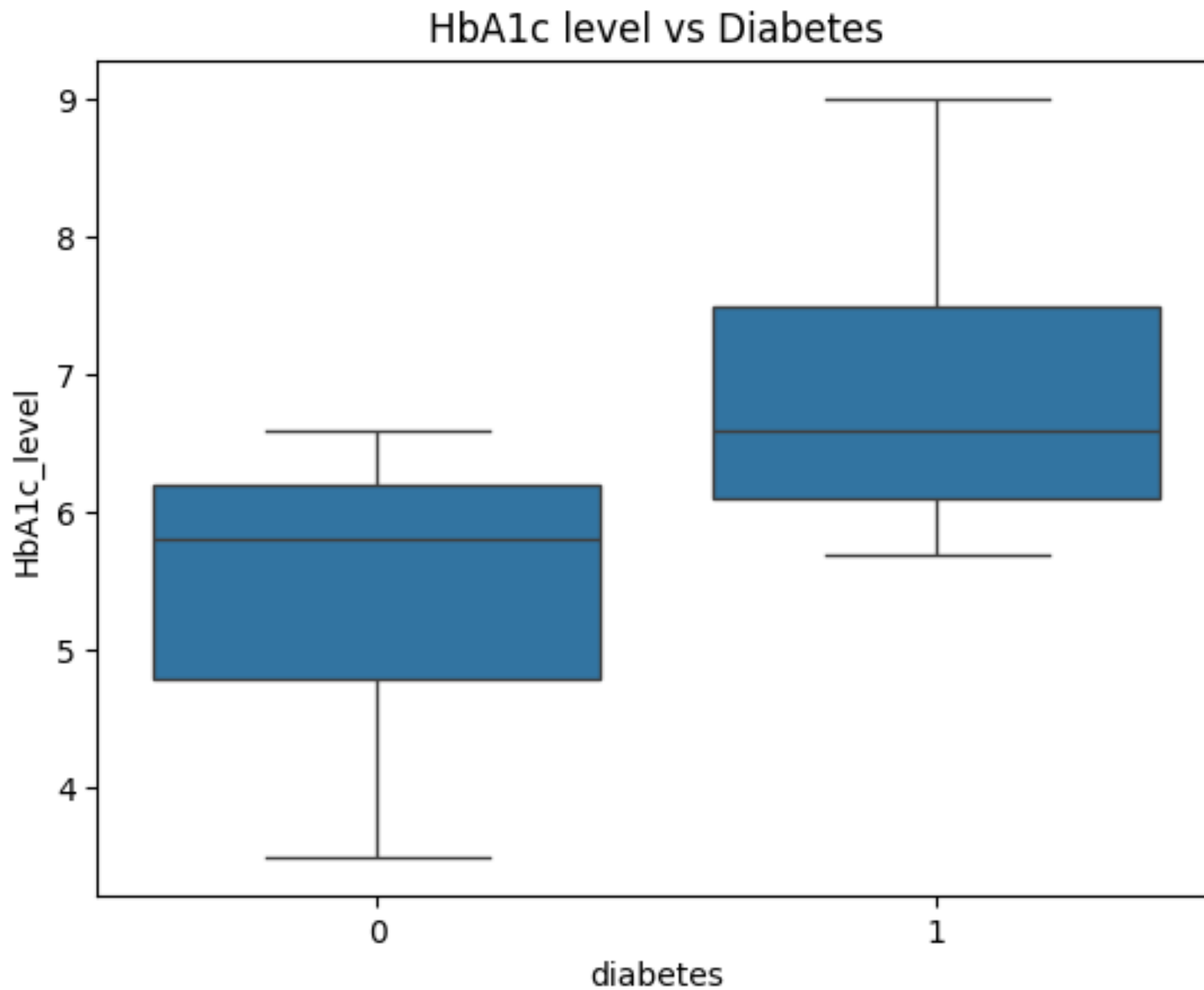
Diabetes and Smoking History (Heatmap)



Diabetes and Smoking History (Bar Plot)

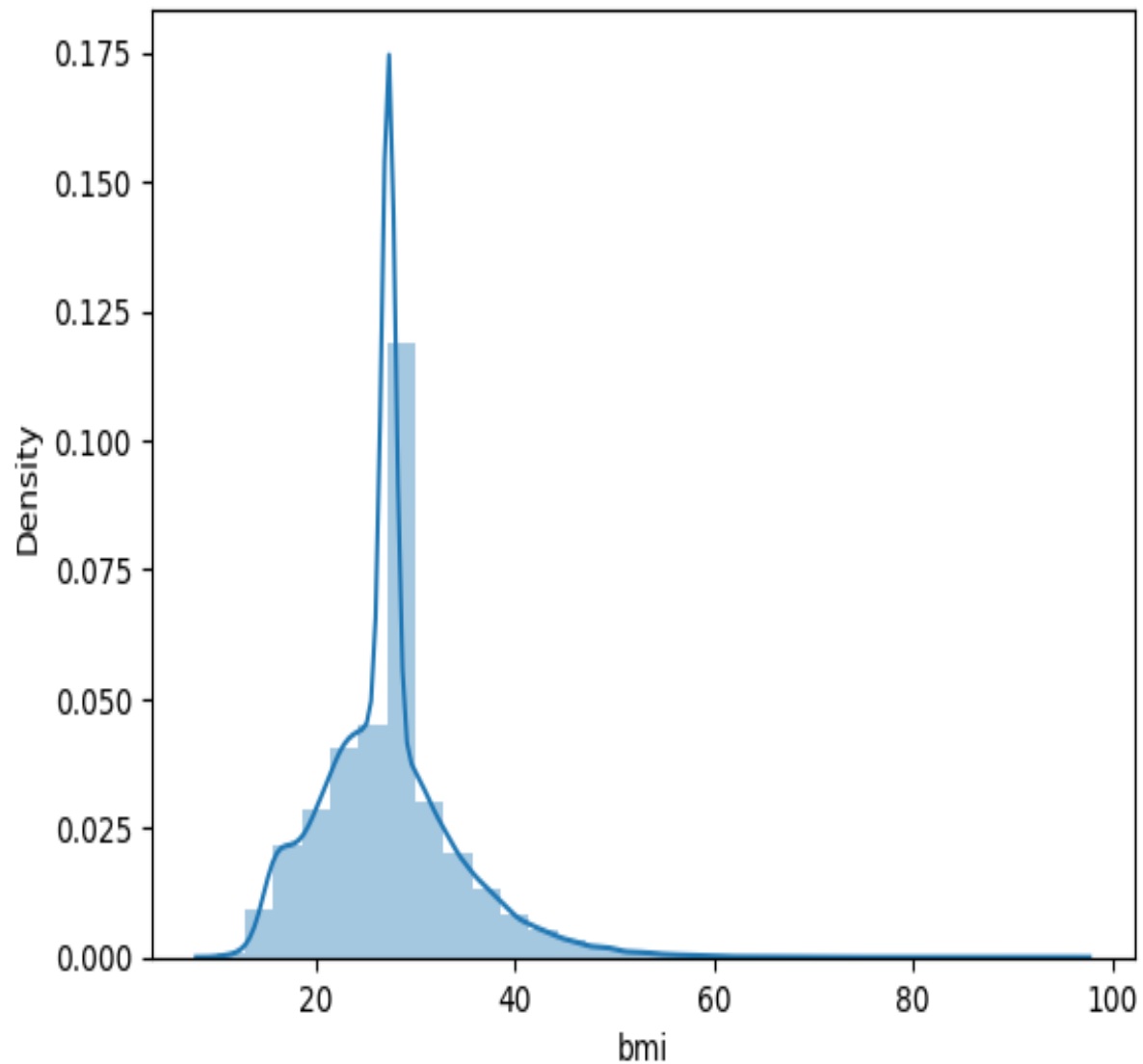


資料前處理與分析

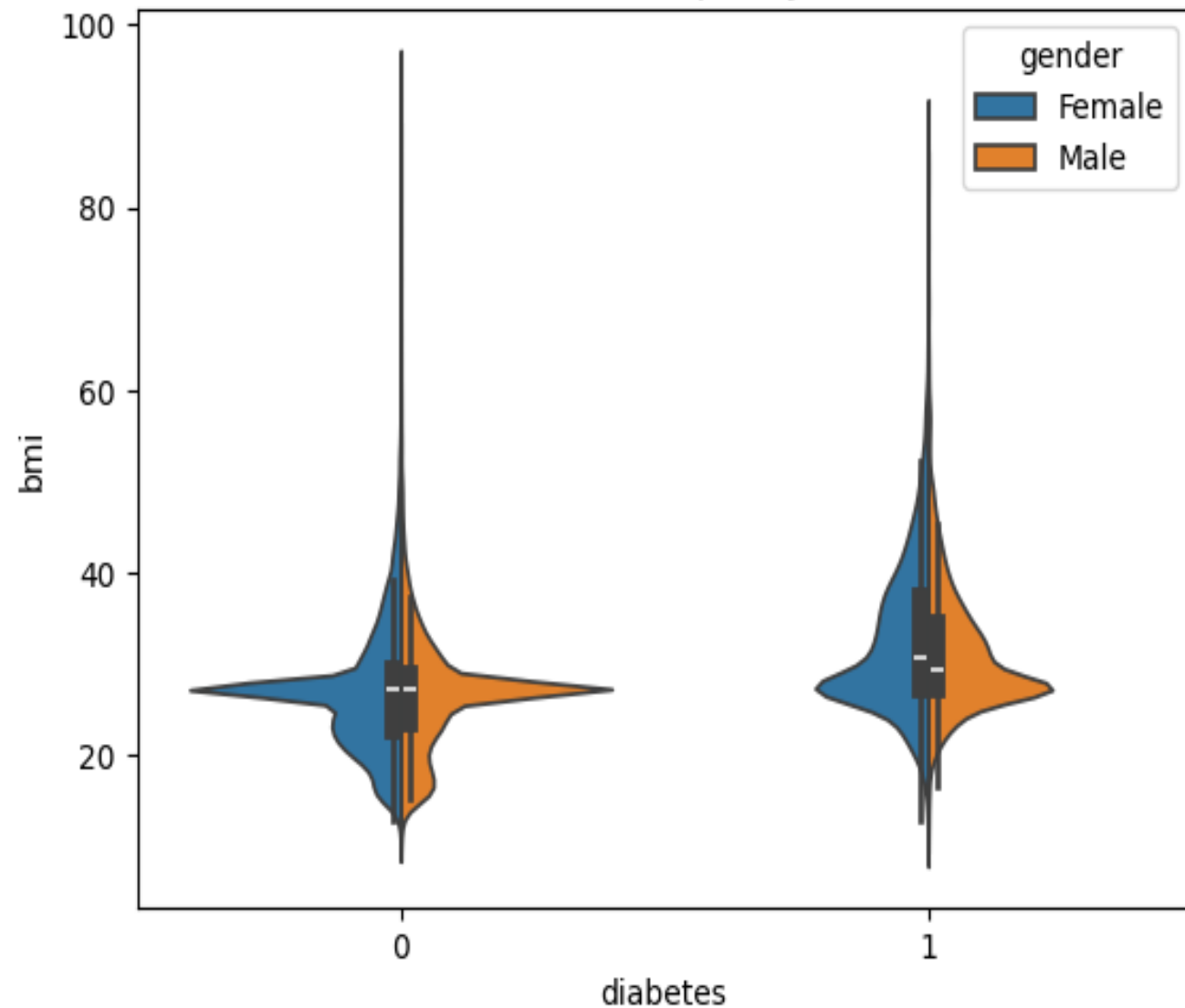


資料前處理與分析

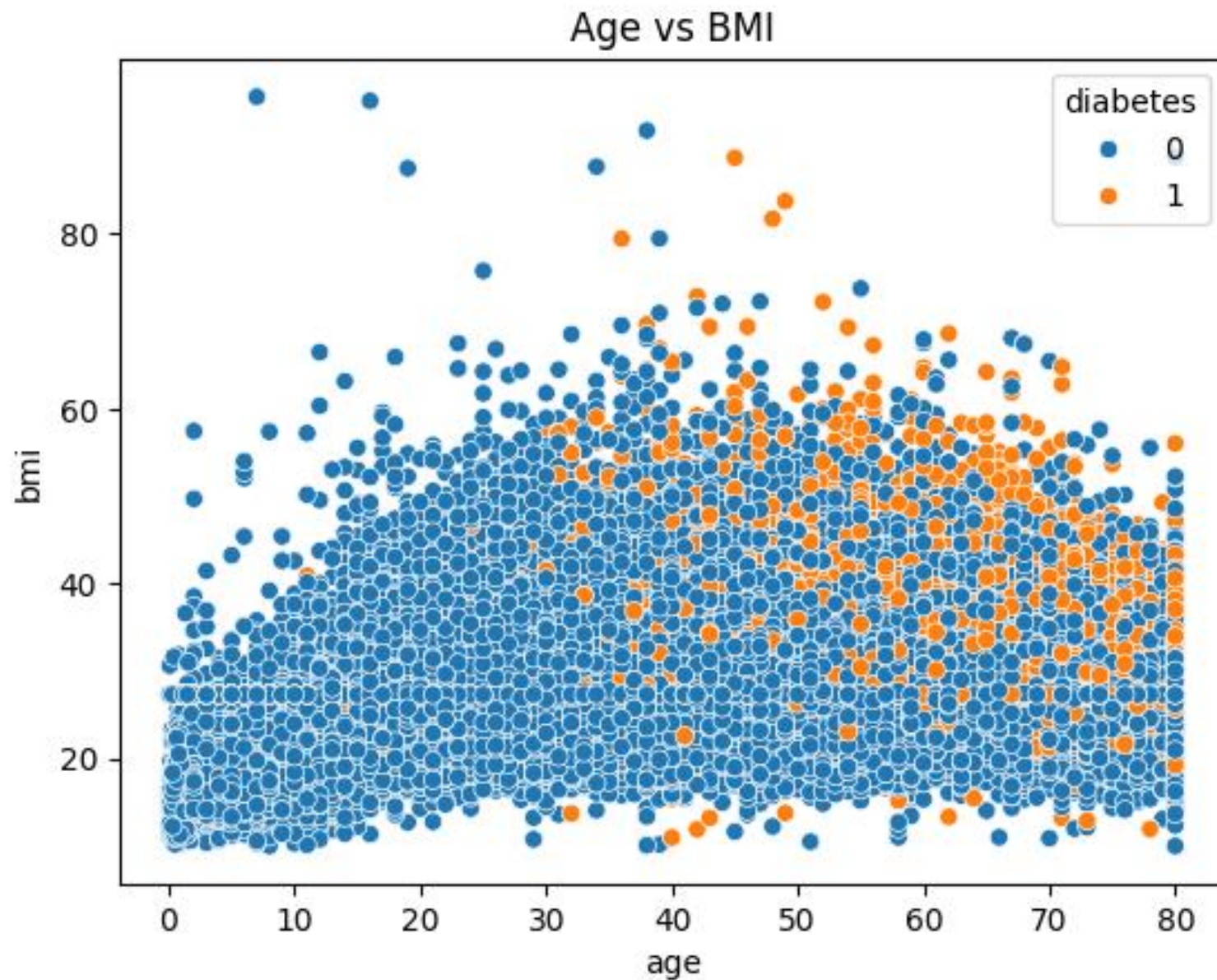
BMI Distribution



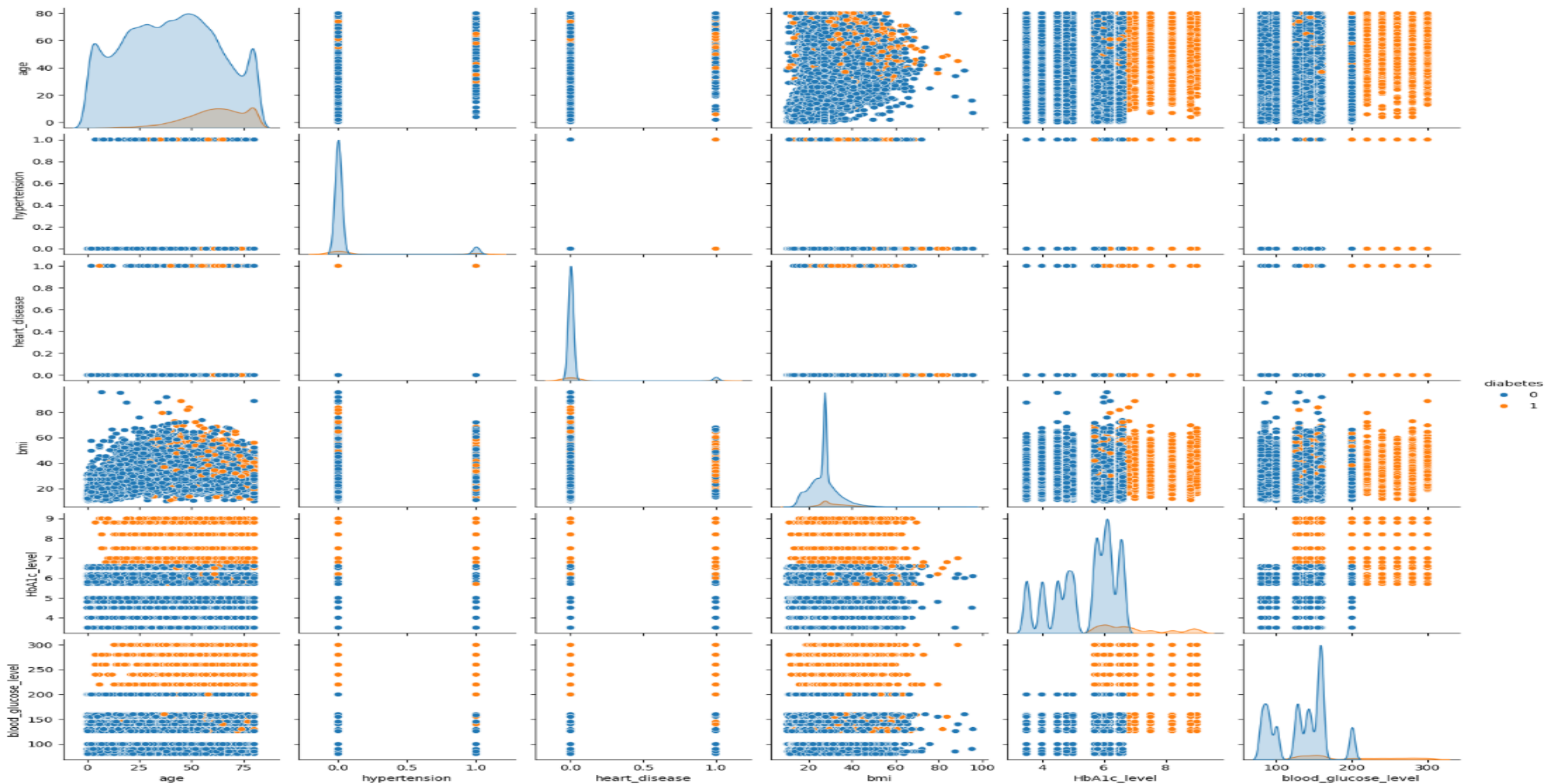
BMI vs Diabetes split by Gender



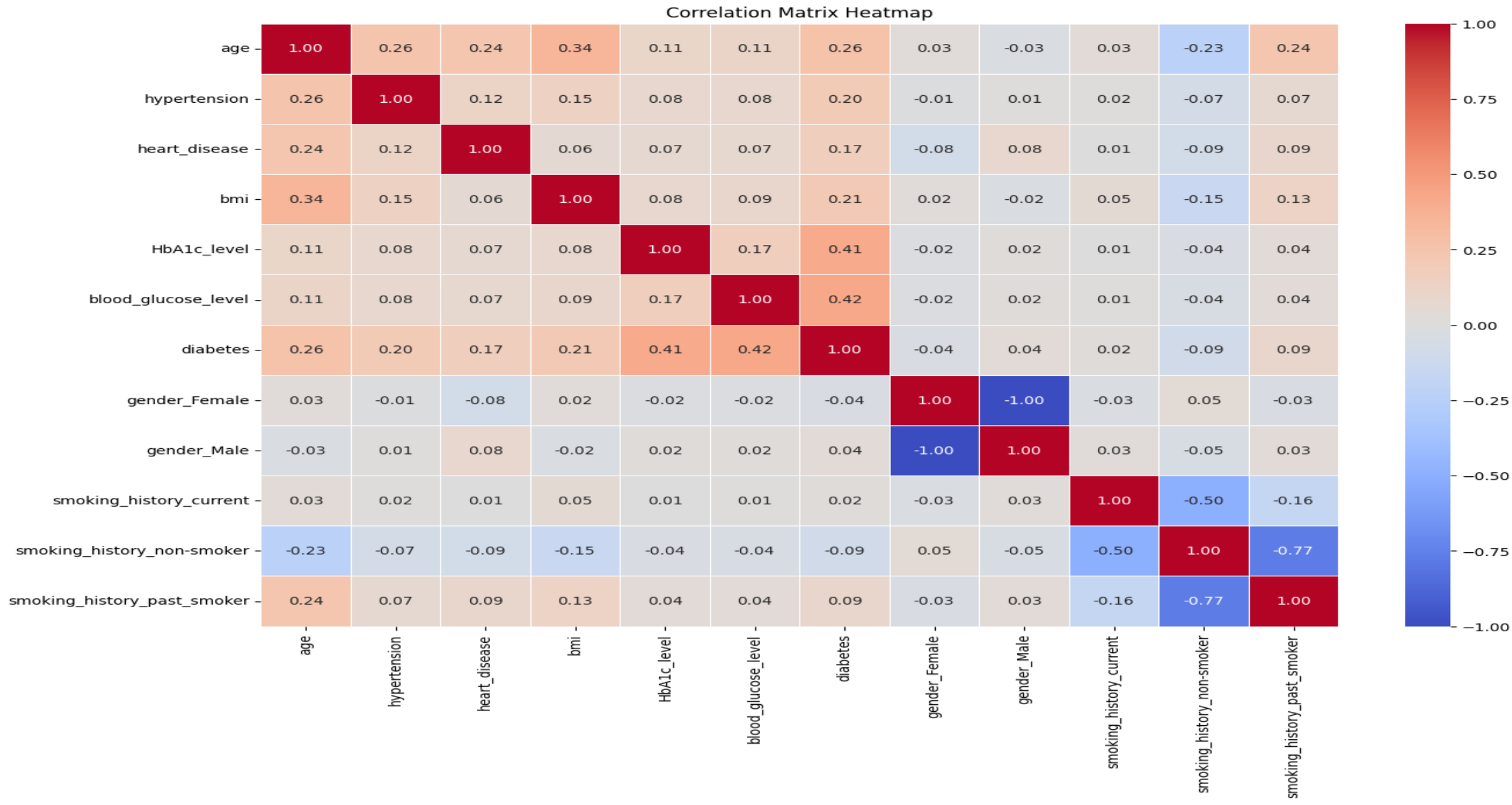
資料前處理與分析



資料前處理與分析



資料前處理與分析



建立預測模型

--各特徵對結果的相關係數

Model: LogisticRegression

	gender	age	Hyper tension	heart_ disease	smoking _history	BMI	HbA1c_ level	blood_ glucose _level
Model Coefficients:	-0.14853434	0.91901222	0.23589913	0.14149198	0.04910917	0.56882591	2.55423499	1.40671821

Accuracy : 0.9464828230895069

Mean Squared Error (MSE): 0.053517176910493104

R-squared (R^2): 0.4593542426715038

Model: LogisticRegression

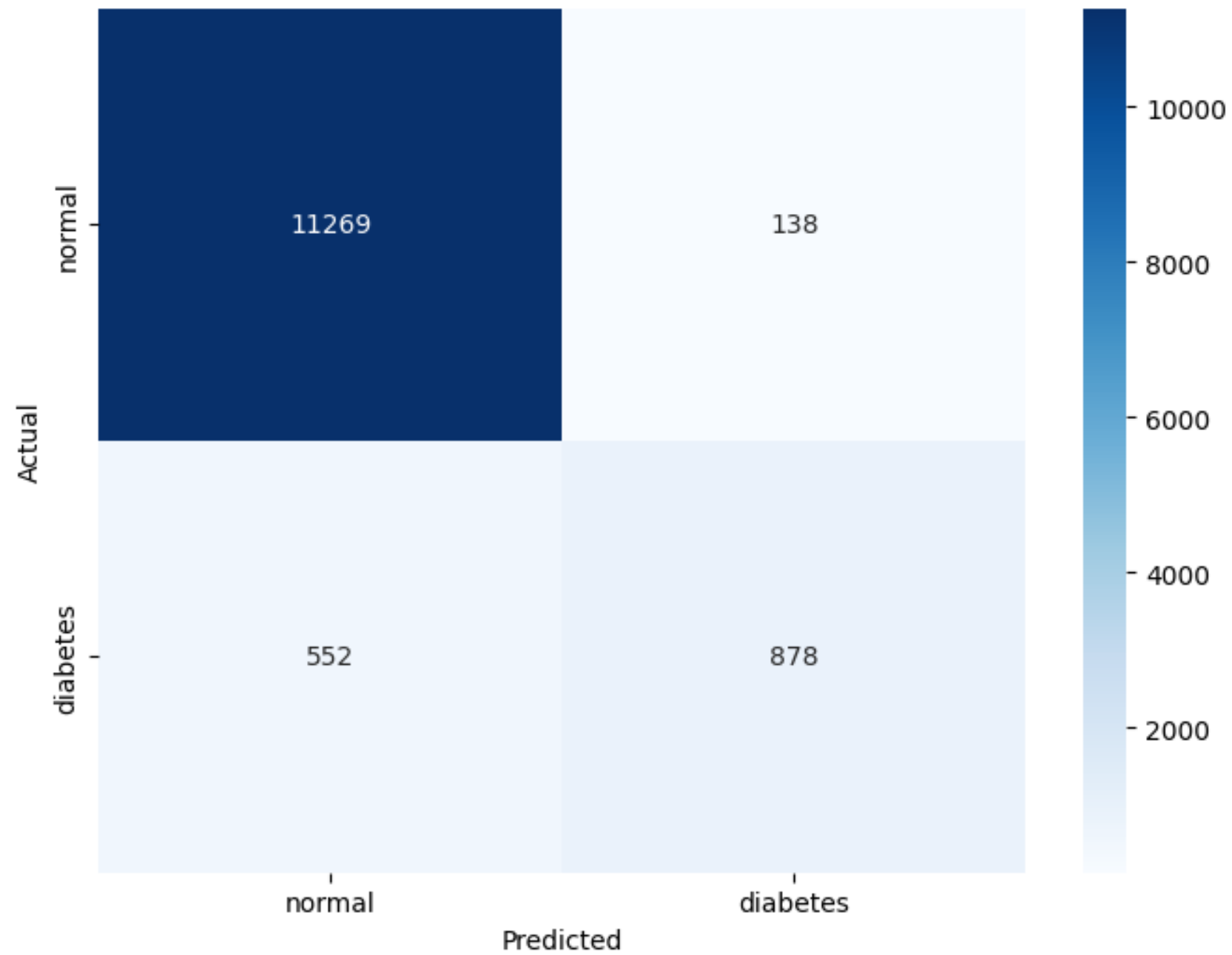
	age	BMI	HbA1c_ level	blood_ glucose_ level
Model Coefficients:	0.05294228	0.09039665	2.33620439	0.03332892

Accuracy : 0.9462491236270156

Mean Squared Error (MSE): 0.05375087637298434

R-squared (R^2): 0.45699334416788595

Confusion Matrix



Discussion

--研究限制與未來方向

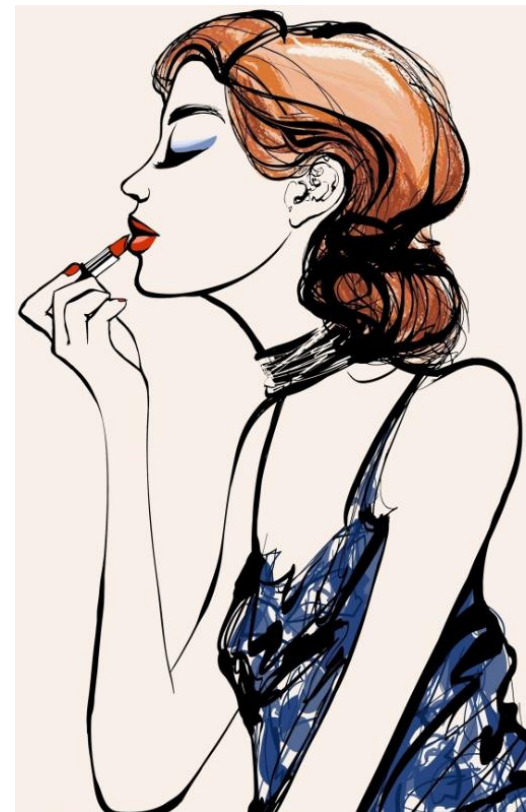
研究限制

- 更加精確的量化抽菸的數量與頻率
- **BS**的取樣方式難以透過模型呈現相關性



未來展望

- 使用TA種族的dataset為標準建立分析與模型
- 列入家族史、女生是否有GDM為預測模型的Factor
- 更嚴謹的方式檢測血糖數據
- BMI數值



女性vsDM

- **Estrogen**

- 血糖:增加細胞對於胰島素的反應
- 脂質代謝：調節膽固醇代謝、降低血液中LDL
- 血壓：增加動脈彈性，讓血管壁不易黏附血脂，保護心血管的功能

- **停經後血糖控制困難:**

1. Estrogen與progesterone濃度改變
2. 熱潮紅與盜汗
3. 體重增加
4. 陰道乾燥

改善方法！



健康生活習慣

均衡飲食、避免吸菸、減少酒精攝取、增加身體活動。



找醫師 改善停經期症狀

賀爾蒙補充治療是可以考慮的治療方式，陰道潤滑劑則可以改善陰道乾燥的症狀。



定時規律測量血糖

比平時更為頻繁的測量血糖，並在夜間測量，幫助理解停經症狀對血糖的影響。



補充維生素D/鈣質

停經後雌激素下降會造成骨質疏鬆和骨折風險增加，須藉由飲食補充鈣質以及維生素D，並定期測量骨密度改善骨質的強度。

更年期停經 與 糖尿病

總結

- **Action**-透過AI技術分析數據中影響DM的關鍵因子
- **Benefit**-協助精準診斷和治療，針對治療方法提供優化建議
- **Conclusion**-準確度94.6%，未來納入種族與家族史並詳細規範抽菸與血糖取樣方式可讓預測模型更加精準且符合台灣患者真實的情形

參考資料

- 數據庫來源:
- <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data>
- 吸菸vs血糖控制:
- <https://blog.health2sync.com/relations-between-diabetes-and-smoking/> (from衛生福利部與2012、2014paper)
- 糖尿病診斷標準:
- <https://www.jah.org.tw/news/index-1.asp?m=9&m1=7&m2=71&id=3186>
- 糖尿病衛教之飲食原則(奇美醫院內分泌新陳代謝科):
- https://www.chimei.org.tw/main/cmh_department/59012/info/7360/A7360106.html
- Estrogen:
- <https://www.commonhealth.com.tw/article/64239> (康健)
- <https://www.edh.tw/article/15809>
- <https://blog.health2sync.com/menopause-and-blood-sugar/>

Thank you for attention!

