
Heart Disease Prediction

第三組：410410236 魏品華．410410863 黃裕權．410410749 陳南天
410410111 洪碩廷．410410988 郭恩均．410411580 吳姿妤

資料集名稱: heart_2020_cleaned.csv

下載網址：

[HTTPS://WWW.KAGGLE.COM/CO
DE/GEORGYZUBKOV/HEART-
DISEASE-EXPLORATORY-DATA-
ANALYSIS/NOTEBOOK#LIST-TAB](https://www.kaggle.com/code/georgyzubkov/heart-disease-exploratory-data-analysis/notebook#list-tab)

資料集介紹:

PERSONAL KEY INDICATORS OF HEART 是 2020 年美國疾病控制與預防中心 (CDC) 對 40 萬成年人的年度調查數據的精簡版本。

針對每個患者 (列) 包含該人的健康狀況。數據是以電話調查的形式收集的。每年，疾病預防控制中心都會打電話給大約 40 萬美國居民，詢問他們的健康狀況，絕大多數問題都是“是”或“否”。

#	Column	Non-Null Count	Dtype
0	HeartDisease	319795 non-null	object
1	BMI	319795 non-null	float64
2	Smoking	319795 non-null	object
3	AlcoholDrinking	319795 non-null	object
4	Stroke	319795 non-null	object
5	PhysicalHealth	319795 non-null	float64
6	MentalHealth	319795 non-null	float64
7	DiffWalking	319795 non-null	object
8	Sex	319795 non-null	object
9	AgeCategory	319795 non-null	object
10	Race	319795 non-null	object
11	Diabetic	319795 non-null	object
12	PhysicalActivity	319795 non-null	object
13	GenHealth	319795 non-null	object
14	SleepTime	319795 non-null	float64

資料數量: 319795

欄位數量: 18

(資料並無遺漏值)

關於欄位名稱及解釋：

- HeartDisease：受訪者是否有心臟疾病。(Yes or No)
- BMI
- Smoking：是否在您的整個生活中至少吸過100支香煙？(Yes or No)
- AlcoholDrinking：是否為重度飲酒者（成年男性每週超過14杯飲酒，成年女性每週超過7杯飲酒）。(Yes or No)
- Stroke：是否被告知您有中風？(Yes or No)
- PhysicalHealth：想一想您的身體健康，包括身體疾病和受傷，過去30天您的身體健康不好有多少天？（0-30天）。
- MentalHealth：您的心理健康，過去30天您的心理健康不好有多少天？（0-30天）。
- DiffWalking：您是否有嚴重行走困難或爬樓梯困難？(Yes or No)
- Sex：您是男性還是女性？(Female or Male)
- AgeCategory：年齡區段 (ex:55-59、75-79、80 or older)
- Race：推斷的種族/族裔 (ex: white、black)

```
HeartDisease
['No' 'Yes']

BMI
[16.6 20.34 26.58 ... 62.42 51.46 46.56]

Smoking
['Yes' 'No']

AlcoholDrinking
['No' 'Yes']

Stroke
['No' 'Yes']

PhysicalHealth
[ 3.  0. 20. 28.  6. 15.  5. 30.  7.  1.  2. 21.  4. 10. 14. 18.  8. 25.
 16. 29. 27. 17. 24. 12. 23. 26. 22. 19.  9. 13. 11.]

MentalHealth
[30.  0.  2.  5. 15.  8.  4.  3. 10. 14. 20.  1.  7. 24.  9. 28. 16. 12.
  6. 25. 17. 18. 21. 29. 22. 13. 23. 27. 26. 11. 19.]

DiffWalking
['No' 'Yes']

Sex
['Female' 'Male']

AgeCategory
['55-59' '80 or older' '65-69' '75-79' '40-44' '70-74' '60-64' '50-54'
 '45-49' '18-24' '35-39' '30-34' '25-29']

Race
['White' 'Black' 'Asian' 'American Indian/Alaskan Native' 'Other'
 'Hispanic']
```

關於欄位名稱及解釋:

- Diabetic : 是否被告知您有糖尿病? (Yes or No)
- PhysicalActivity : 過去30天, 除了您的常規工作之外, 您是否報告過進行身體活動或運動? (Yes or No)
- GenHealth : 您認為整體而言您的健康狀況是? (ex: poor、fair、good)
- SleepTime : 平均每天睡眠多少小時?
- Asthma : 是否被告知您有哮喘? (Yes or No)
- KidneyDisease : 是否被告知您有腎臟疾病? (Yes or No)
- SkinCancer : 是否被告知您有皮膚癌? (Yes or No)

```
Diabetic
['Yes' 'No' 'No, borderline diabetes' 'Yes (during pregnancy)']

PhysicalActivity
['Yes' 'No']

GenHealth
['Very good' 'Fair' 'Good' 'Poor' 'Excellent']

SleepTime
[ 5.  7.  8.  6. 12.  4.  9. 10. 15.  3.  2.  1. 16. 18. 14. 20. 11. 13.
 17. 24. 19. 21. 22. 23.]

Asthma
['Yes' 'No']

KidneyDisease
['No' 'Yes']

SkinCancer
['Yes' 'No']
```

-
- 目標：透過機器學習讓我們能快速找出可能存在心臟病風險的人並指出哪些變量對患心臟病的可能性有顯著影響。
 - label column : HeartDisease欄位
 - 問題類型：分類 (有病 or 沒病)

T H E E N D .