

醫療數據可視化工具

你將使用我們在 Replit 的初始化項目來完成這個項目。

- 首先在 Replit 中導入項目。
- 接着，你將看到一個 `.replit` 窗口。
- 選擇 `Use run command` 並點擊 `Done` 按鈕。

我們仍在開發 Python 課程的交互式教學部分。目前，你可以在 YouTube 上通過 freeCodeCamp.org 上傳的一些視頻學習這個項目相關的知識。

- [Python for Everybody Video Course \(14 hours\)](#)
- [如何使用 Python Pandas 分析數據 \(10 小時\)](#)

在本項目中，您將使用 matplotlib、seaborn 和 pandas 來對體檢數據進行可視化和計算。數據集的數值是從體檢中收集的。

數據說明

數據集中的行代表患者，列代表身體測量、各種血液檢查的結果和生活方式等信息。您將使用該數據集來探索心臟病、身體測量數據、血液標誌物和對生活方式的選擇之間的關係。

文件名：medical_examination.csv

項目	變量類型	變量名	變量值類型
年齡	客觀特徵	<code>age</code>	int (days)
身高	客觀特徵	<code>height</code>	int (cm)
體重	客觀特徵	<code>weight</code>	float (kg)
性別	客觀特徵	<code>gender</code>	分類編碼
收縮壓	檢測特徵	<code>ap_hi</code>	int
舒張壓	檢測特徵	<code>ap_lo</code>	int
膽固醇	檢測特徵	<code>cholesterol</code>	1：正常，2：高於正常，3：遠遠高於正常值

血糖值	檢測特徵	gluc	1：正常，2：高於正常，3：遠遠高於正常值
吸菸問題	主觀特徵	smoke	binary
飲酒量	主觀特徵	alco	binary
體育活動	主觀特徵	active	binary
是否有心血管疾病	目標變量	cardio	binary

任務

創建一個類似於 `examples/Figure_1.png` 的圖表，其中我們顯示 `cholesterol`、`gluc`、`alco`、`active` 和 `smoke` 變量，用於不同面板中 `heart=1` 和 `heart=0` 的患者。

在 `medical_data_visualizer.py` 中使用數據完成以下任務：

- 給數據添加一列 `overweight`。要確定一個人是否超重，首先通過將他們的體重（公斤）除以他們的身高（米）的平方來計算他們的 BMI。如果該值是 > 25 ，則此人超重。使用值 0 表示不超重，使用值 1 表示超重。
- 使用 0 表示好的和 1 表示壞，來規範化數據。如果 `cholesterol` 或 `gluc` 的值為 1，則將值設為 0。如果值大於 1，則將值設為 1。
- 將數據轉換為長格式並使用 seaborn 的 `catplot()` 創建一個顯示分類特徵值計數的圖表。數據集應按“Cardio”拆分，因此每個 `cardio` 值都有一個圖表。該圖表應該看起來像 `examples/Figure_1.png`。
- 清理數據。過濾掉以下代表不正確數據的患者段：
 - 舒張壓高於收縮壓（使用 `(df['ap_lo'] <= df['ap_hi'])` 保留正確的數據）
 - 高度小於第 2.5 個百分位數（使用 `(df['height'] >= df['height'].quantile(0.025))` 保留正確的數據）
 - 身高超過第 97.5 個百分位
 - 體重小於第 2.5 個百分位
 - 體重超過第 97.5 個百分位
- 使用數據集創建相關矩陣。使用 seaborn 的 `heatmap()` 繪製相關矩陣。遮罩上三角。該圖表應類似於 `examples/Figure_2.png`。

每當變量設置為 `None` 時，請確保將其設置為正確的代碼。

單元測試是在 `test_module.py` 下為你編寫的。

開發

對於開發，你可以使用 `main.py` 來測試你的函數。單擊“運行”按鈕，`main.py` 將運行。

測試

爲了你的方便，我們將測試從 `test_module.py` 導入到 `main.py`。只要你點擊“運行”按鈕，測試就會自動運行。

提交

複製項目的 URL 並將其提交給 freeCodeCamp。

解決方案鏈接

ex: <https://replit.com/@camperbot/hello>

我已經完成這個挑戰

獲得提示

請求幫助