**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

A blue circle with text

Description automatically generated****

**BÀI TẬP LAB 3**

**ĐỀ TÀI: PHÂN TÍCH KHÁM PHÁ VỀ**

**BỆNH ĐÁI THÁO ĐƯỜNG**

**Danh sách sinh viên thực hiện:**

**3122410138 – Bùi Minh Huy**

**3122410128 – Phan Huy Hoàng**

**3112410336 – Hoàng Minh Quang**

**3122410106 – Nguyễn Trần Công Hậu**

**Giảng viên hướng dẫn: Đỗ Như Tài**

**Thành phố Hồ Chí Minh – Tháng 09/2025**

**MỤC LỤC**

[**1.Định nghĩa vấn đề 4**](#_Toc209736546)

[**2.Chuẩn bị vấn đề 5**](#_Toc209736547)

[**2.1.Khai báo thư viện 5**](#_Toc209736548)

[**2.2.Nạp dữ liệu 5**](#_Toc209736549)

[**2.3.2. Xử lí giá trị "0" không hợp lệ 6**](#_Toc209736550)

[**3.Phân tích khám phá dữ liệu 8**](#_Toc209736551)

[**3.1.Phân tích đơn biến (Univariable Analysis) 8**](#_Toc209736552)

[**3.1.1. Hiển thị một số thông tin về dữ liệu 8**](#_Toc209736553)

[**3.1.2. Các tính chất thống kê trên dữ liệu số 10**](#_Toc209736554)

[**3.1.3. Tần số xuất hiện trên dữ liệu phân loại 11**](#_Toc209736555)

[**3.1.4. Phân loại tình trạng dinh dưỡng theo BMI 12**](#_Toc209736556)

[**3.1.5. Phân loại tình trạng dinh dưỡng theo nồng độ Glucose huyết tương sau 2 giờ 14**](#_Toc209736557)

[**3.2.Phân tích đa biến (Multivariable Analysis) 16**](#_Toc209736558)

[**3.2.1. So sánh trung bình các đặc điểm theo nhóm Outcome 16**](#_Toc209736559)

[**3.2.2. Phân tích độ dày da theo độ tuổi 45**](#_Toc209736560)

[**3.3.Phân tích mối liên hệ giữa các yếu tố nguy cơ và bệnh tiểu đường 49**](#_Toc209736561)

[**4. Kết luận 53**](#_Toc209736562)

**PHÂN CÔNG CÔNG VIỆC**

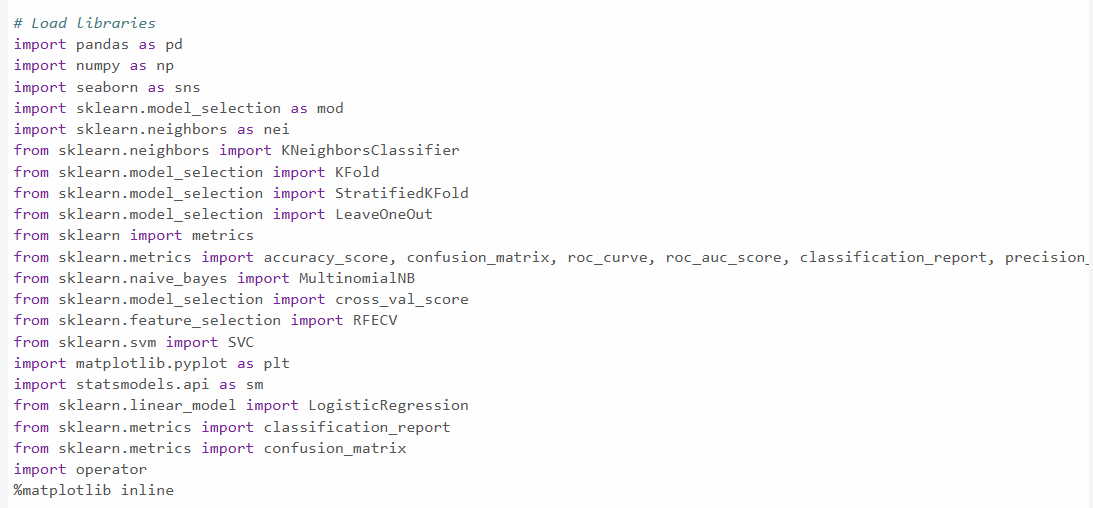
|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Công việc** |
| 1 | Bùi Minh Huy | 3122410138 | - Thực hiện EDA  - Viết báo cáo cuối Word |
| 2 | Phan Huy Hoàng | 3122410128 | - Đọc 3 paper (NDDG 1979, ADAP NN 1988, WHO 1999)  - Xác định ý nghĩa bài toán: Input / Output / Và mục tiêu  - Chuẩn bị slide thuyết trình |
| 3 | Hoàng Minh Quang | 3112410336 | - Thực hiện EDA  - Viết báo cáo cuối Word |
| 4 | Nguyễn Trần Công Hậu | 3122410106 | - Thực hiện EDA  - Chuẩn bị slide thuyết trình |

# 1.Định nghĩa vấn đề

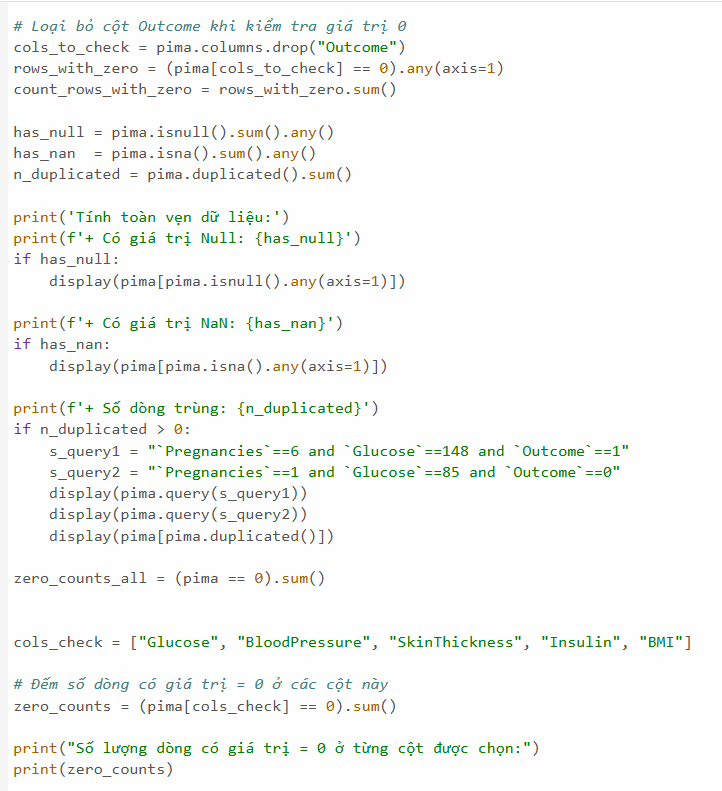
* **Mô tả**:
  + Bộ dữ liệu gồm 768 mẫu từ bệnh nhân đều là nữ giới, ít nhất 21 tuổi, và có nguồn gốc người da đỏ Pima, sống gần Phoenix, Arizona.
  + 8 đặc điểm được đo từ mỗi mẫu + 1 biến phân loại (có/không mắc tiểu đường)
* **Dữ liệu vào**:
  + Số lần mang thai (lần)
  + Nồng độ glucose huyết tương sau 2 giờ trong xét nghiệm dung nạp glucose đường uống (mg/dL)
  + Huyết áp tâm trương (mm Hg)
  + Độ dày nếp gấp da cơ tam đầu (mm)
  + Nồng độ insulin huyết thanh sau 2 giờ (mu U/ml)
  + Chỉ số khối cơ thể BMI (kg/m²)
  + Hàm phả hệ tiểu đường ()
  + Tuổi (năm)
  + Biến phân loại (nhị phân)
* **Dữ liệu ra**:
  + Dự đoán bệnh nhân có tiểu đường không (Biến phân loại 0 hoặc 1)
* **Mục tiêu**:
  + Phân tích bộ dữ liệu bệnh đái tháo đường nhằm hiểu rõ các yếu tố liên quan đến nguy cơ mắc bệnh.
  + Vận dụng kiến thức về dinh dưỡng để đánh giá ảnh hưởng của chế độ ăn uống đến tình trạng bệnh.

# 2.Chuẩn bị vấn đề

## 2.1.Khai báo thư viện



## 2.2.Nạp dữ liệu



Tính toàn vẹn dữ liệu:

+ Có giá trị Null: False

+ Có giá trị NaN: False

+ Số dòng trùng: 0

Số lượng dòng có giá trị = 0 ở từng cột được chọn:

Glucose 5

BloodPressure 35

SkinThickness 227

Insulin 374

BMI 11

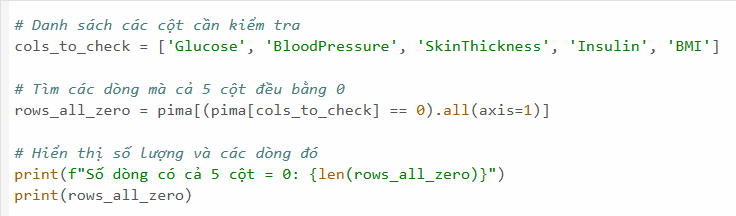
dtype: int64

**Nhận xét**:

* Không có dữ liệu nào trùng nhau
* Dữ liệu không có giá trị rỗng (NaN, Null).
* Tuy nhiên, khi kiểm tra chi tiết, phát hiện một số cột có giá trị = 0 không hợp lý về mặt y học, được coi là dữ liệu thiếu:
  + Glucose: không thể bằng 0 vì cơ thể luôn có đường huyết lưu hành.
  + BloodPressure: không thể bằng 0, vì huyết áp 0 chỉ xảy ra khi đã tử vong.
  + SkinThickness: không thể bằng 0, lớp mỡ dưới da luôn có độ dày nhất định.
  + Insulin: không thể bằng 0, vì tuyến tụy luôn tiết insulin nền.
  + BMI: không thể bằng 0, vì chỉ số khối cơ thể = 0 nghĩa là cân nặng = 0, điều này không thực tế.

### 2.3.2. Xử lí giá trị "0" không hợp lệ

* **Hướng xử lí**:
  + **Bước 1:** Loại bỏ các dòng có cả 5 cột trên đều bằng 0.
  + **Bước 2:** Thay thế các giá trị bằng 0 bằng trung vị của từng cột.



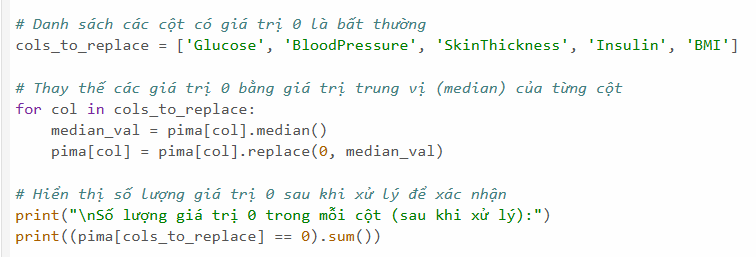
Số dòng có cả 5 cột = 0: 0

Empty DataFrame

Columns: [Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome]

Index: []

**Nhận xét**: Không có dòng nào mà 5 cột có dữ liệu bằng 0. Chuyển sang bước 2



Số lượng giá trị 0 trong mỗi cột (sau khi xử lý):

Glucose 0

BloodPressure 0

SkinThickness 0

Insulin 0

BMI 0

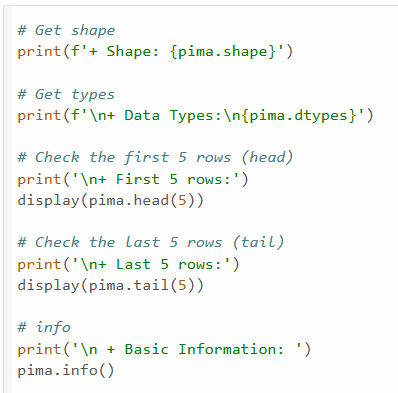
dtype: int64

# 3.Phân tích khám phá dữ liệu

## 3.1.Phân tích đơn biến (Univariable Analysis)

### 3.1.1. Hiển thị một số thông tin về dữ liệu

* Số dòng, số cột của dữ liệu
* Kiểu dữ liệu của từng cột
* 5 dòng đầu và 5 dòng cuối của bảng dữ liệu
* Thông tin chung về dữ liệu



+ Shape: (768, 9)

+ Data Types:

Pregnancies int64

Glucose int64

BloodPressure int64

SkinThickness int64

Insulin float64

BMI float64

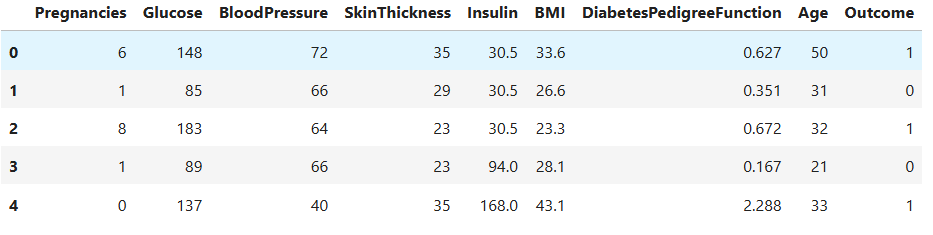
DiabetesPedigreeFunction float64

Age int64

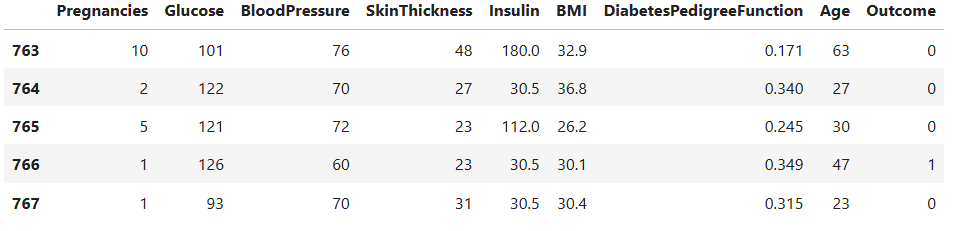
Outcome int64

dtype: object

+ First 5 rows:



+ Last 5 rows:



+ Basic Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 768 entries, 0 to 767

Data columns (total 9 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Pregnancies 768 non-null int64

1 Glucose 768 non-null int64

2 BloodPressure 768 non-null int64

3 SkinThickness 768 non-null int64

4 Insulin 768 non-null float64

5 BMI 768 non-null float64

6 DiabetesPedigreeFunction 768 non-null float64

7 Age 768 non-null int64

8 Outcome 768 non-null int64

dtypes: float64(3), int64(6)

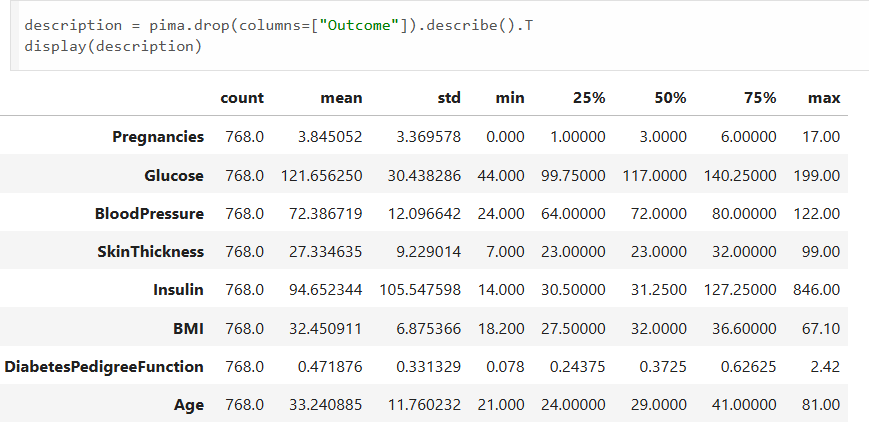
memory usage: 54.1 KB

**Nhận xét**:

* Dữ liệu có 8 tính chất để phân lớp: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age
* Các giá trị đều là số học, thuộc kiểu int64 và float64
* Tổng số dòng dữ liệu là 768 dòng
* Dữ liệu phân lớp nằm ở cột Outcome, gồm 2 giá trị: 0 (không mắc bệnh) và 1 (mắc bệnh)

### 3.1.2. Các tính chất thống kê trên dữ liệu số

* Count, Mean, Standard Deviation, Minimum Value
* 25th Percentile, 50th Percentile (Median), 75th Percentile, Maximum Value

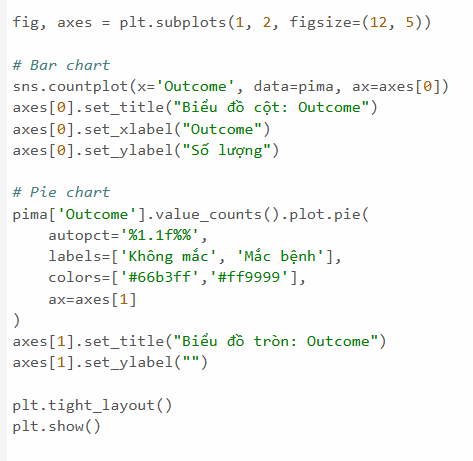


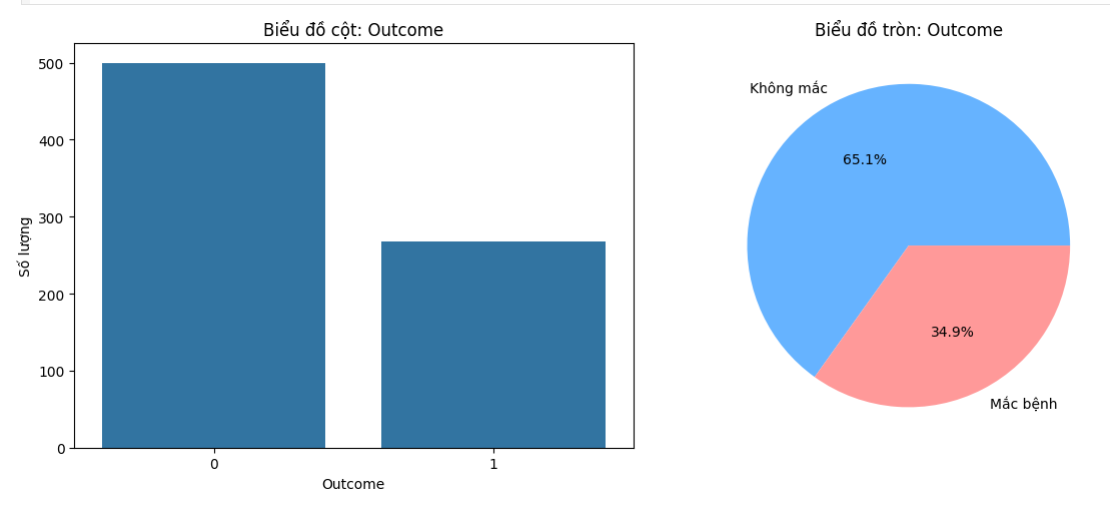
**Nhận xét**:

* **Pregnancies:** Trung bình 3.84, dao động 0–17 → hợp lý, vì có thể có phụ nữ chưa mang thai.
* **Glucose:** Giá trị trung bình là 120.89, với mức cao nhất là 199. Phân phối cho thấy có sự đa dạng về mức đường huyết trong nhóm nghiên cứu, phản ánh rõ nguy cơ tiểu đường.
* **BloodPressure:** Huyết áp trung bình là 69.1, nằm trong ngưỡng bình thường. Giá trị cao nhất là 122, cho thấy có một số trường hợp tăng huyết áp.
* **SkinThickness:** Độ dày lớp mỡ dưới da trung bình là 20.5, với giá trị cao nhất là 99. Phân phối khá rộng, thể hiện sự khác biệt về thể trạng giữa các cá nhân.
* **Insulin:** Mức insulin trung bình là 79.8, nhưng giá trị cao nhất lên tới 846, cho thấy có một số trường hợp rối loạn chuyển hóa rõ rệt.
* **BMI:** Chỉ số khối cơ thể trung bình là 31.99, nằm trong ngưỡng thừa cân. Một số cá nhân có BMI rất cao (tối đa 67.1), cho thấy nguy cơ béo phì.
* **DiabetesPedigreeFunction:** Hệ số di truyền trung bình là 0.47, dao động từ 0.078 đến 2.42. Phân phối này phản ánh mức độ ảnh hưởng của yếu tố di truyền đến nguy cơ mắc bệnh.
* **Age:** Tuổi trung bình là 33.24, dao động từ 21 đến 81. Phân phối tuổi phù hợp với nhóm dân số trưởng thành, có nguy cơ tiểu đường tăng theo tuổi.

### 3.1.3. Tần số xuất hiện trên dữ liệu phân loại

Câu hỏi: Tỷ lệ người mắc và không mắc bệnh tiểu đường trong tập dữ liệu là bao nhiêu?



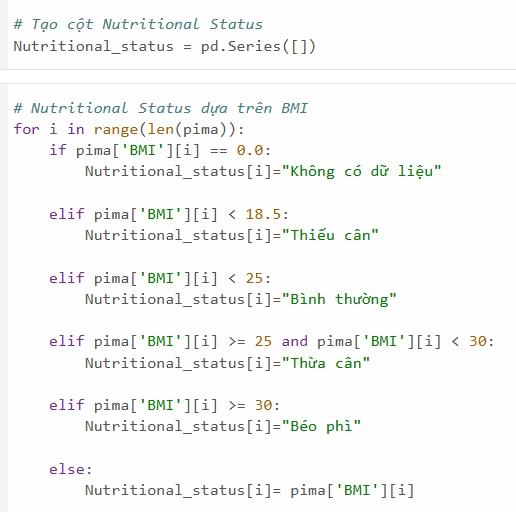


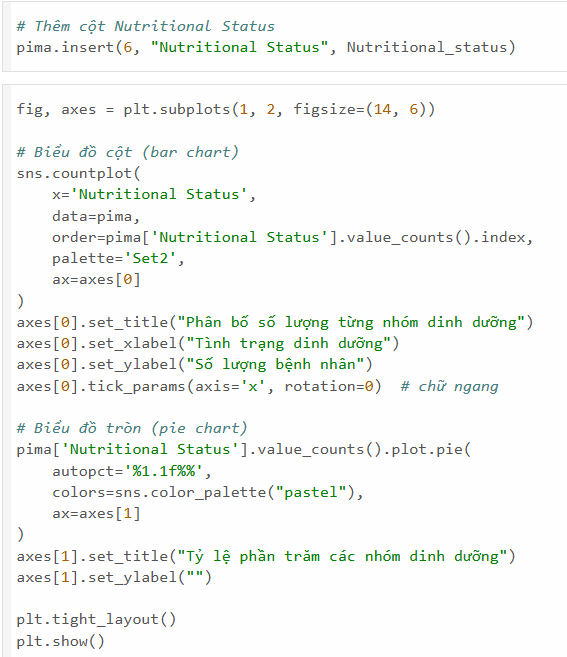
**Nhận xét**:

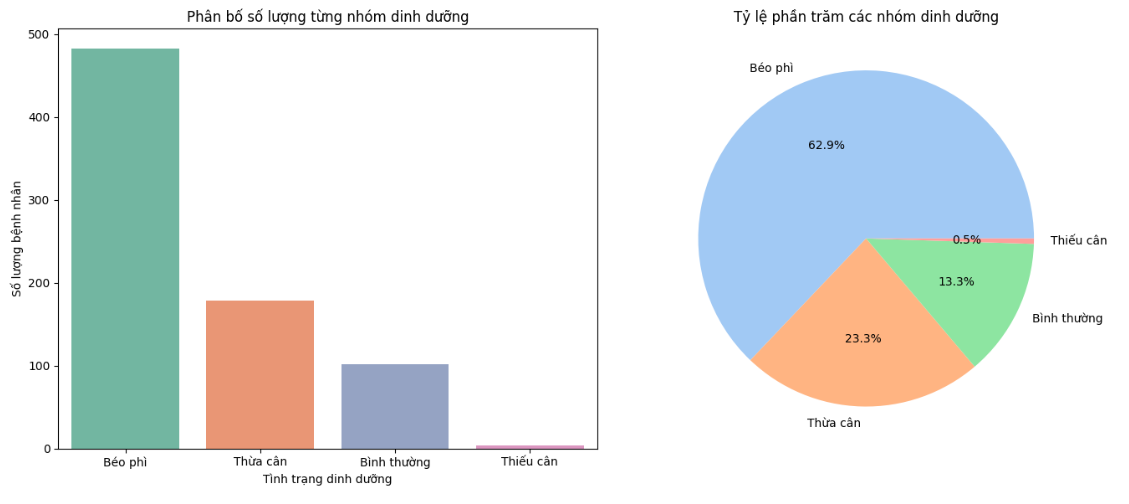
* Dữ liệu không cân bằng: lớp 0 chiếm khoảng 65%, lớp 1 chiếm khoảng 35%, điều này có thể làm mô hình thiên lệch về lớp đa số.
* **Giải pháp**: có thể áp dụng oversampling (SMOTE) để tăng số mẫu lớp 1, hoặc undersampling để giảm số mẫu lớp 0

### 3.1.4. Phân loại tình trạng dinh dưỡng theo BMI

* Câu hỏi: Tình trạng dinh dưỡng của bệnh nhân được phân bố như thế nào theo chỉ số BMI?







**Nhận xét**:

* Nhóm "Béo phì" chiếm tỷ lệ áp đảo (62.89%) trong dân số Pima, trong khi nhóm "Thiếu cân" chỉ chiếm 0.52%, phản ánh xu hướng thừa cân rõ rệt trong cộng đồng này.
* Trong tổng số 768 phụ nữ, có 102 người có cân nặng bình thường và 4 người thuộc nhóm thiếu cân; phần lớn còn lại thuộc nhóm thừa cân hoặc béo phì, phản ánh xu hướng dinh dưỡng mất cân đối và cho thấy mối liên hệ tiềm năng giữa tình trạng dinh dưỡng và nguy cơ mắc tiểu đường.

### 3.1.5. Phân loại tình trạng dinh dưỡng theo nồng độ Glucose huyết tương sau 2 giờ

* Câu hỏi: Tình trạng dung nạp glucose của bệnh nhân được phân bố như thế nào theo kết quả OGTT?

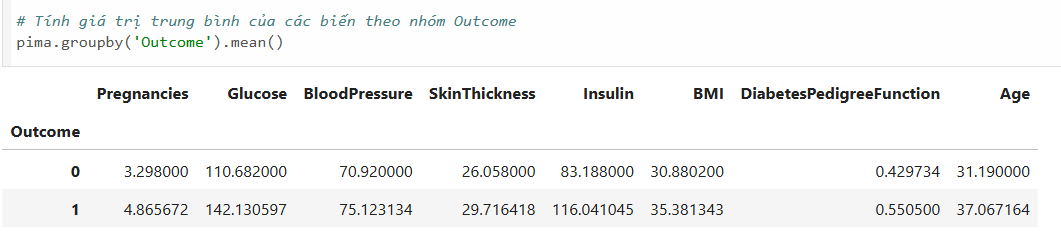


**Nhận xét**:

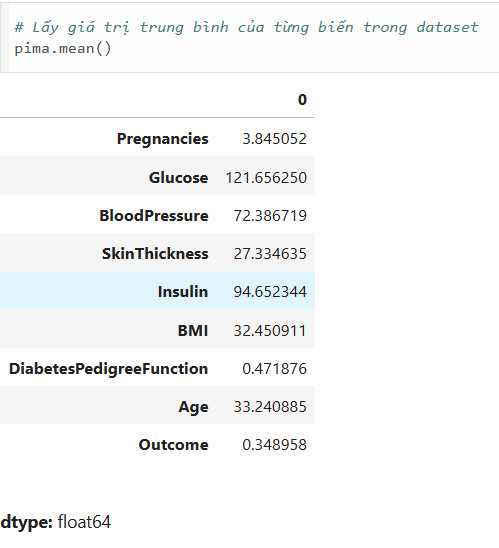
* Không có cá nhân nào trong tập dữ liệu có mức glucose sau uống đạt ngưỡng tiểu đường (≥ 198 mg/dL).
* Có 131 người thuộc nhóm rối loạn dung nạp glucose (141–198 mg/dL) và đã được chẩn đoán mắc tiểu đường.
* Có 136 người có mức glucose bình thường (≤ 140 mg/dL) nhưng vẫn mắc tiểu đường.
* Điều này cho thấy: mức glucose cao không phải là yếu tố duy nhất gây bệnh, và tiểu đường có thể xuất hiện ngay cả khi glucose bình thường.
* Khả năng tồn tại kháng insulin ở nhóm glucose bình thường là điều đáng chú ý.
* Việc tính toán chỉ số HOMA-IR sẽ giúp đánh giá chính xác hơn mức độ kháng insulin trong mẫu nghiên cứu.

## 3.2.Phân tích đa biến (Multivariable Analysis)

### 3.2.1. So sánh trung bình các đặc điểm theo nhóm Outcome



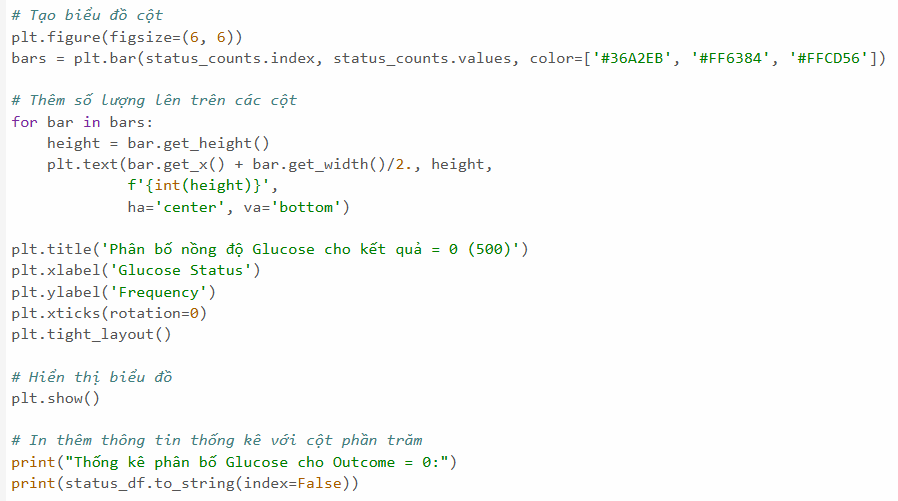
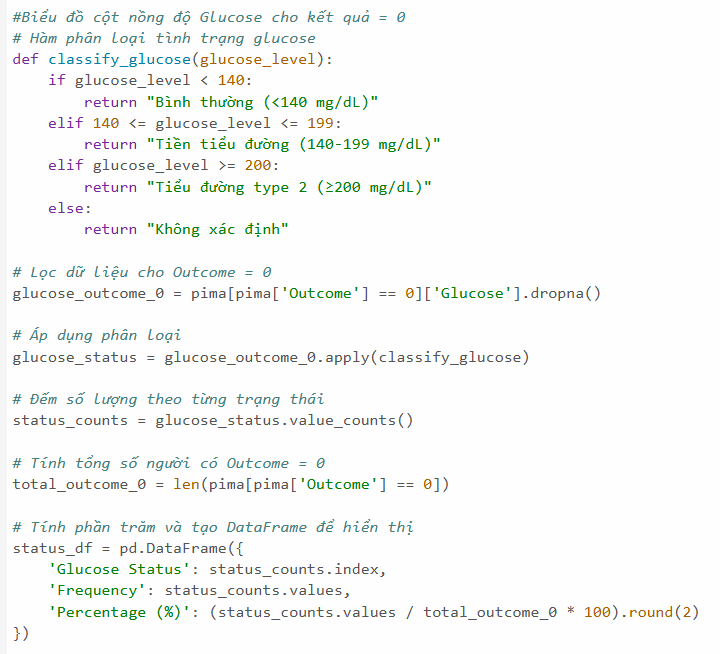
**Nhận xét**: Kết quả cho thấy nhóm phụ nữ mắc tiểu đường có xu hướng sở hữu các đặc điểm sinh lý với giá trị trung bình cao hơn so với nhóm không mắc bệnh.

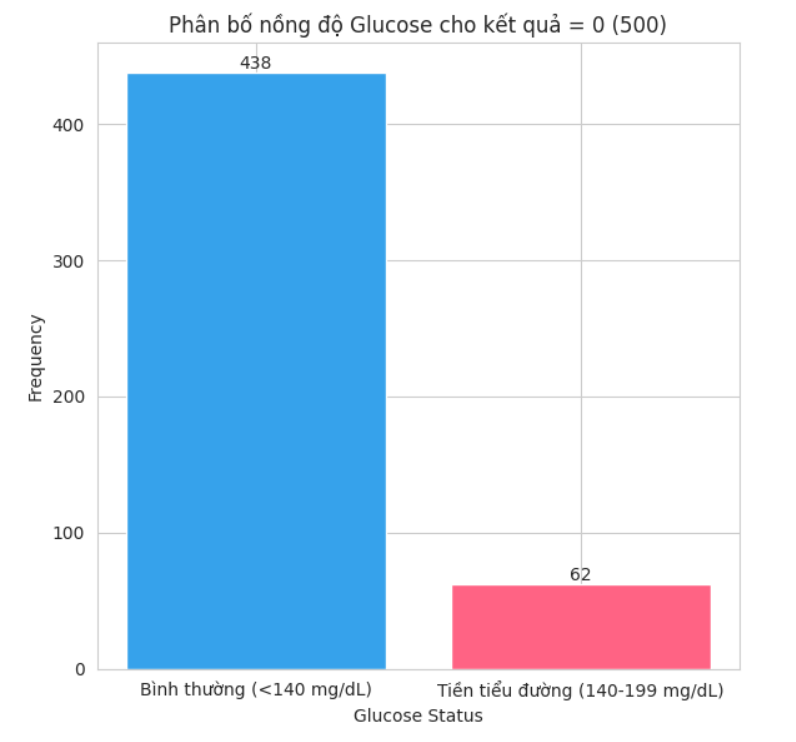


**Nhận xét**:

* **Pregnancies:** Trung bình 3.85 lần mang thai, phản ánh nhóm phụ nữ đã từng sinh nở, phù hợp với đối tượng nghiên cứu.
* **Glucose:** Mức đường huyết trung bình là 121.66 mg/dL, tiệm cận ngưỡng chẩn đoán tiểu đường (126 mg/dL), cho thấy nguy cơ bệnh trong tập dữ liệu là đáng chú ý.
* **BloodPressure:** Huyết áp trung bình là 72.39 mmHg, nằm trong ngưỡng bình thường. Đây là mức ổn định, tuy nhiên vẫn cần kiểm tra sự khác biệt giữa các nhóm bệnh.
* **SkinThickness:** Độ dày lớp mỡ dưới da trung bình là 27.33 mm, cho thấy thể trạng tương đối ổn định trong nhóm nghiên cứu. Đây là một đặc điểm có thể liên quan đến tình trạng béo phì và nguy cơ rối loạn chuyển hóa.
* **Insulin:** Mức insulin trung bình là 94.65 µU/mL, cho thấy chức năng chuyển hóa ở mức phổ thông. Tuy nhiên, đây vẫn là biến cần theo dõi vì có thể liên quan đến tình trạng kháng insulin.
* **BMI:** Chỉ số khối cơ thể trung bình là 32.45, vượt ngưỡng 30 – cho thấy tình trạng thừa cân hoặc béo phì là phổ biến trong tập dữ liệu.
* **DiabetesPedigreeFunction:** Hệ số di truyền trung bình là 0.47, cho thấy mức độ ảnh hưởng di truyền ở mức vừa phải.
* **Age:** Tuổi trung bình là 33.24, phù hợp với nhóm dân số trưởng thành, độ tuổi có nguy cơ tiểu đường bắt đầu tăng.
* **Outcome:** Tỷ lệ trung bình là 0.34, tức khoảng 35% người trong tập dữ liệu có kết quả dương tính với tiểu đường — đủ để phân tích sự khác biệt giữa hai nhóm.

#### 3.2.1.1. So sánh trung bình Glucose theo nhóm Outcome





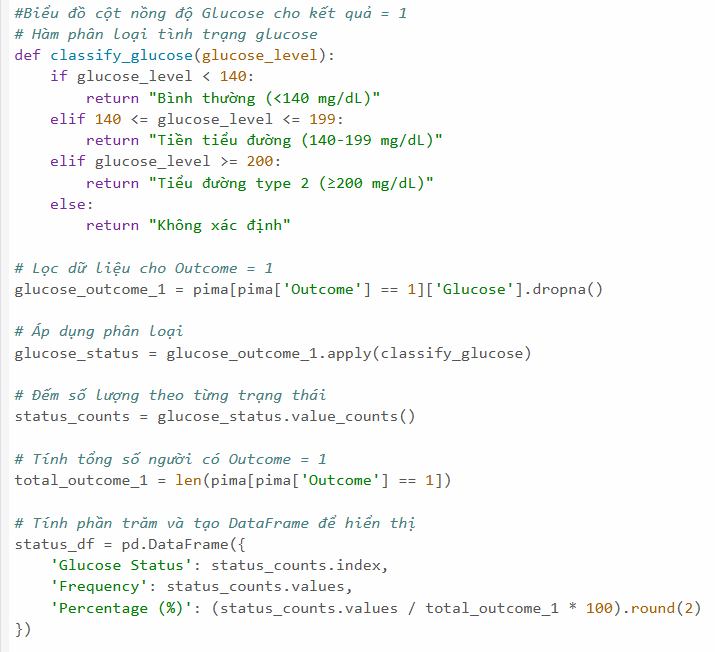
Thống kê phân bố Glucose cho Outcome = 0:

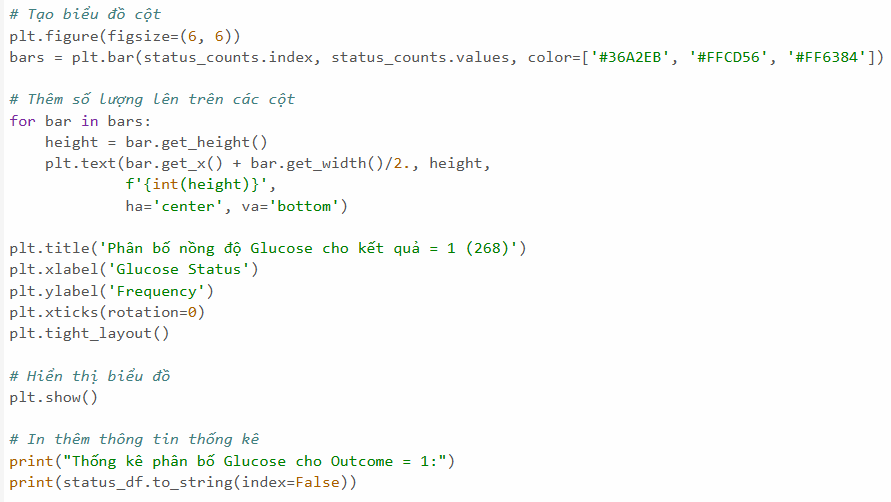
Glucose Status Frequency Percentage (%)

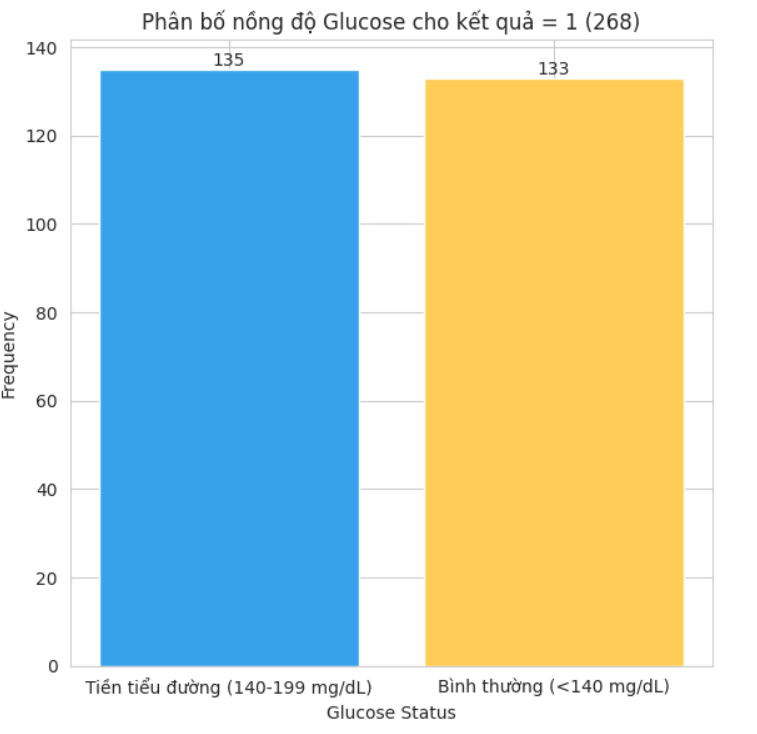
Bình thường (<140 mg/dL) 438 87.6

Tiền tiểu đường (140-199 mg/dL) 62 12.4

**Nhận xét:** Phân bố dữ liệu: Phần lớn (87.6% với 438 người) có mức glucose bình thường (< 140 mg/dL), trong khi chỉ một phần nhỏ (12.4% với 62 người) rơi vào mức tiền tiểu đường (140-199 mg/dL). Điều này cho thấy xu hướng chung là glucose ở mức bình thường trong nhóm không mắc bệnh.







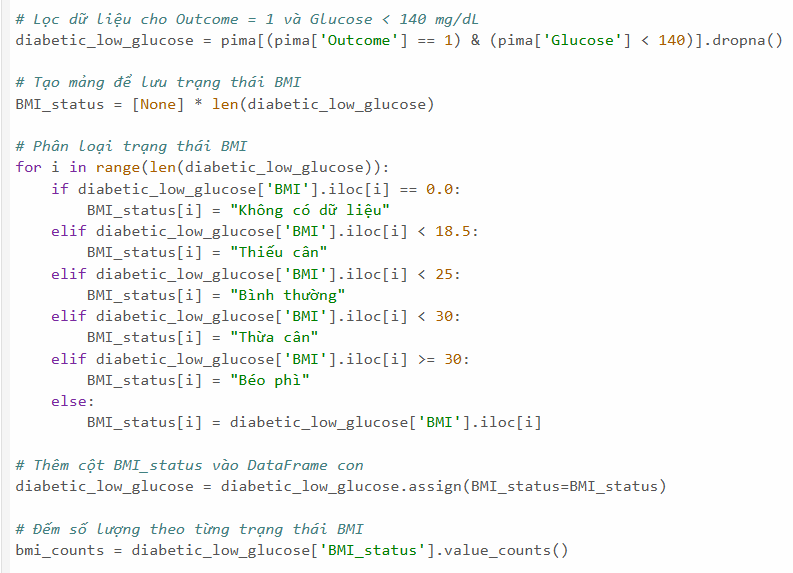
Thống kê phân bố Glucose cho Outcome = 1:

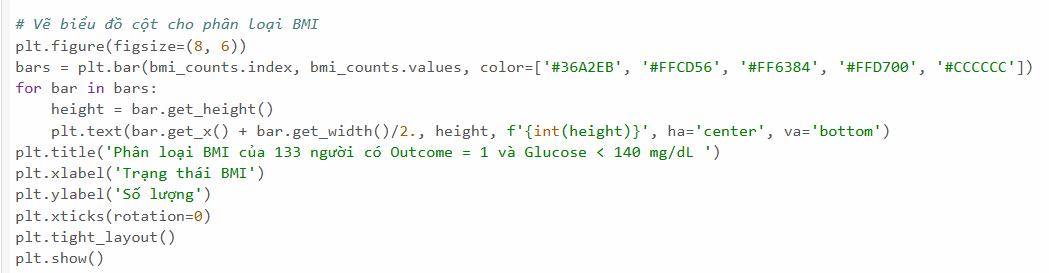
Glucose Status Frequency Percentage (%)

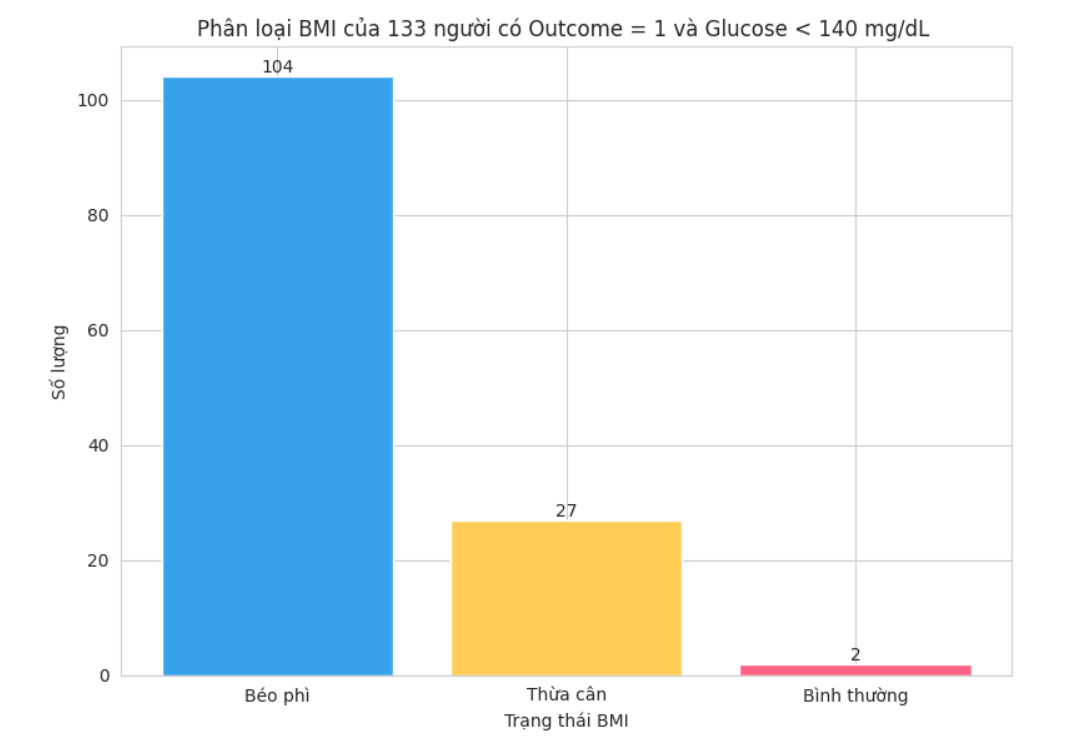
Tiền tiểu đường (140-199 mg/dL) 135 50.37

Bình thường (<140 mg/dL) 133 49.63

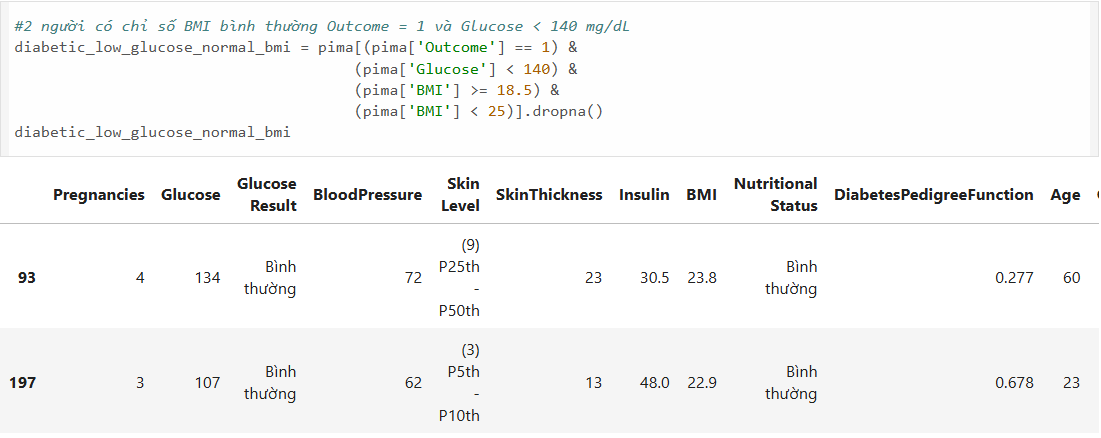
**Nhận xét:** Hai nhóm có số lượng và phần trăm rất gần nhau (135 người, 50.37% cho "Tiền tiểu đường" và 133 người, 49.63% cho "Bình thường"), cho thấy sự phân bố khá cân bằng giữa các mức glucose trong nhóm mắc đái tháo đường. Điều này có thể ngụ ý rằng một số người mắc bệnh có mức glucose bình thường, có thể do chẩn đoán dựa trên các yếu tố khác







**Nhận xét:** BMI của 133 người có Outcome = 1 và Glucose < 140 mg/dL thì có 102 người bị béo phì và 27 người thừa cân



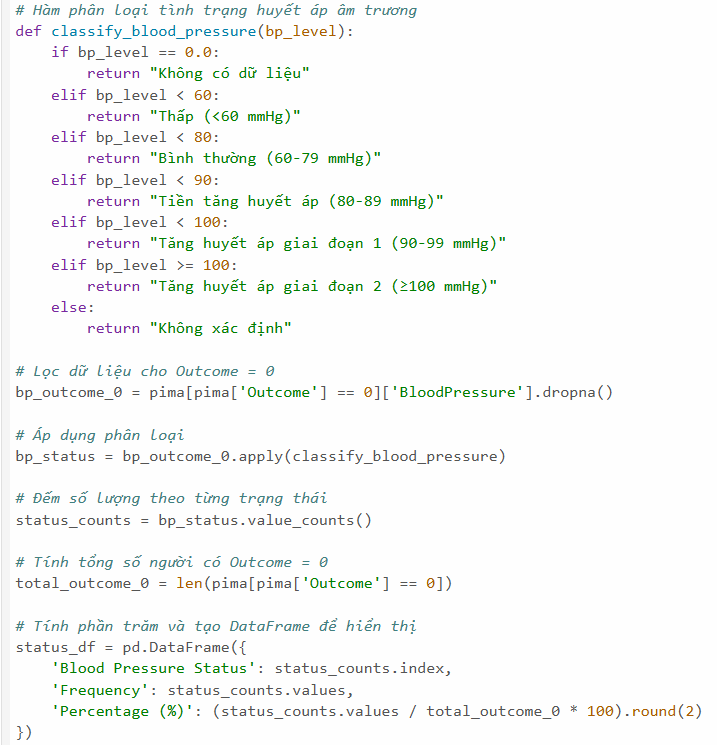
**Sự phân bố nồng độ glucose ở nhóm không mắc bệnh và mắc bệnh khác nhau như thế nào?**

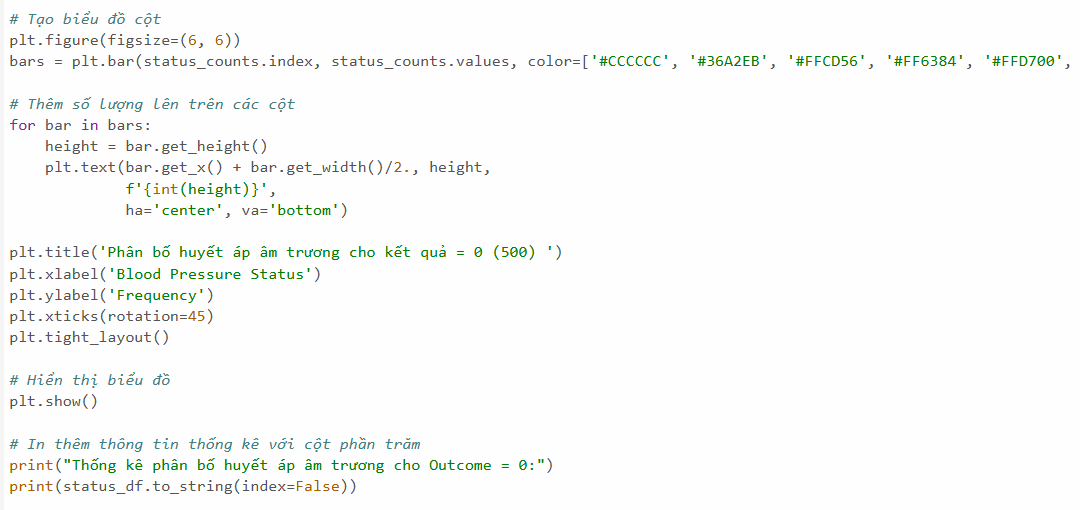
* Sự phân bố nồng độ glucose ở nhóm không mắc bệnh và mắc bệnh khác nhau rõ rệt. Nhóm không mắc bệnh chủ yếu có nồng độ glucose thấp (<140 mg/dL), trong khi nhóm mắc bệnh có xu hướng phân bố đồng đều hơn giữa nồng độ thấp và trung bình (140-199 mg/dL), với tỷ lệ nồng độ cao hơn đáng kể. Điều này hỗ trợ giả thuyết rằng nồng độ glucose cao có thể liên quan đến nguy cơ mắc bệnh tiểu đường.

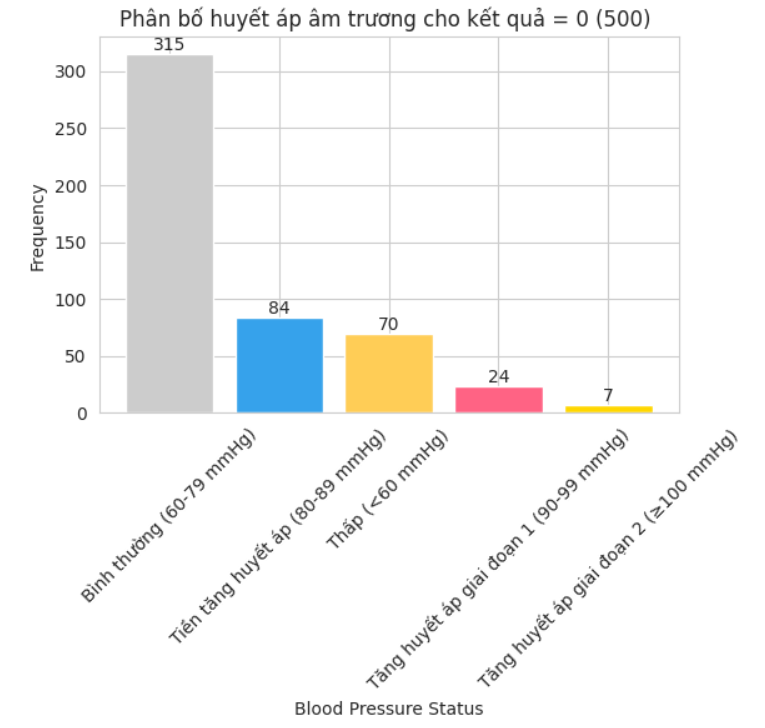
**Tại sao có những người có nồng độ glucose bình thường nhưng vẫn mắc bệnh?**

* Việc có nồng độ glucose bình thường nhưng vẫn mắc bệnh tiểu đường phản ánh tính phức tạp của bệnh, vượt ra ngoài chỉ số glucose lúc đói. Các yếu tố như xét nghiệm bổ sung (HbA1c, OGTT), cơ chế tự miễn, biến thiên sinh học, và yếu tố di truyền đều đóng vai trò quan trọng. Để hiểu rõ hơn, cần phân tích thêm các biến số khác trong tập dữ liệu (như BMI, huyết áp, hoặc insulin) hoặc tham khảo hồ sơ y tế chi tiết.

#### 3.2.1.2. So sánh trung bình BloodPressure theo nhóm Outcome







Thống kê phân bố huyết áp âm trương cho Outcome = 0:

Blood Pressure Status Frequency Percentage (%)

Bình thường (60-79 mmHg) 315 63.0

Tiền tăng huyết áp (80-89 mmHg) 84 16.8

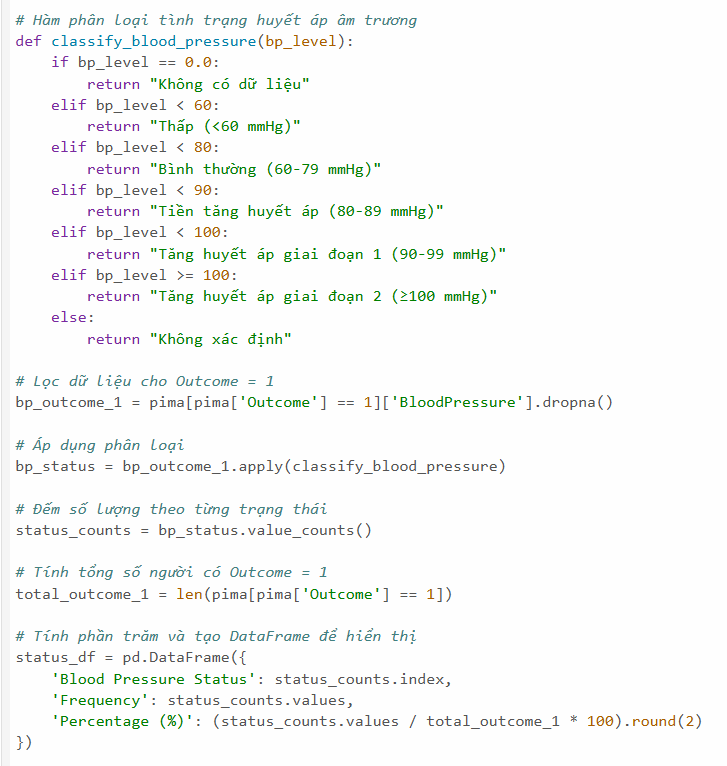
Thấp (<60 mmHg) 70 14.0

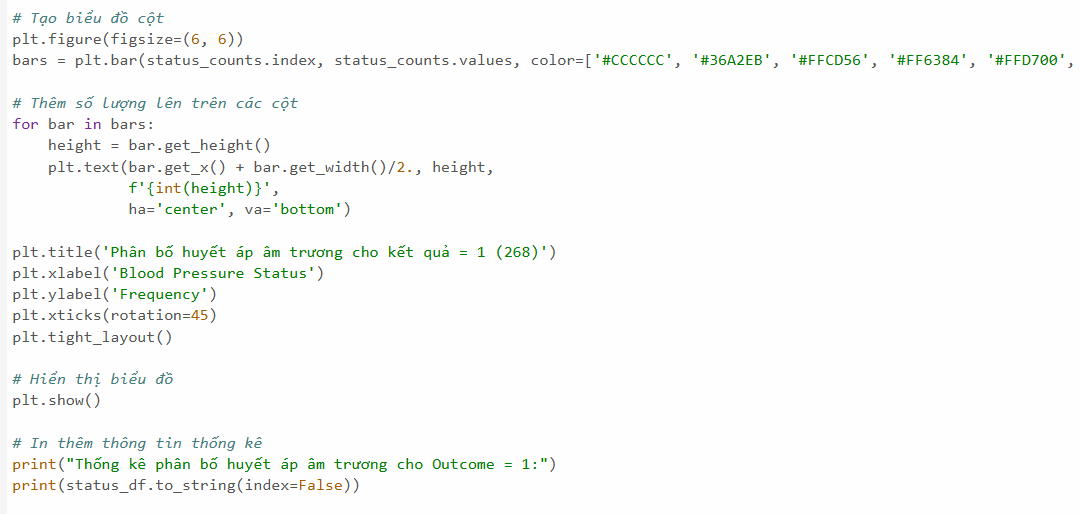
Tăng huyết áp giai đoạn 1 (90-99 mmHg) 24 4.8

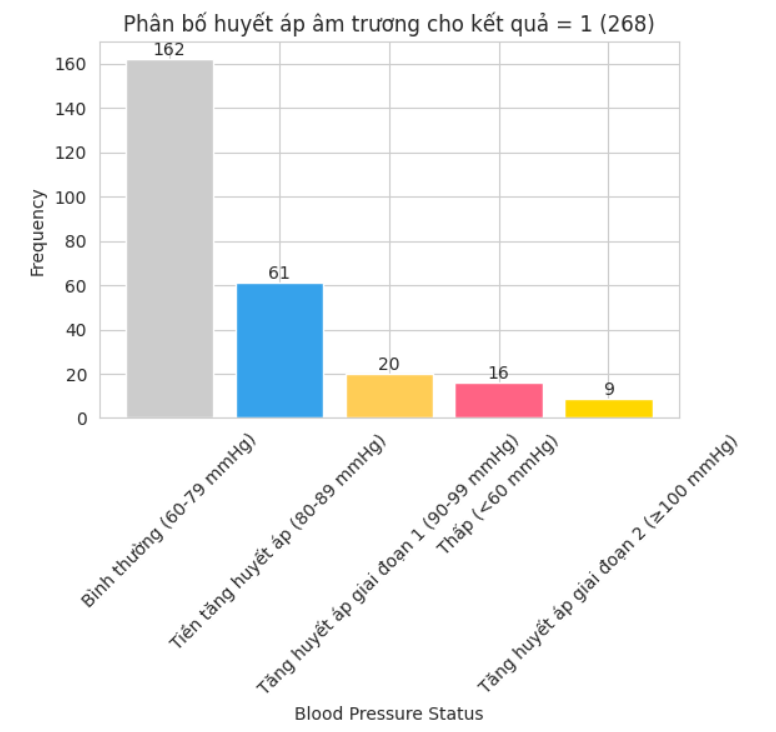
Tăng huyết áp giai đoạn 2 (≥100 mmHg) 7 1.4

**Nhận xét:**

* Đa số không tiểu đường có huyết áp tâm trương bình thường 63% nằm trong khoảng 60–79 mmHg, tức là mức khỏe mạnh.Đây là tín hiệu tích cực
* Đa số người không bị tiểu đường cũng có huyết áp ổn định. Một phần nhỏ có dấu hiệu tiền tăng huyết áp (16.8%) Huyết áp 80–89 mmHg chưa phải bệnh nhưng là cảnh báo sớm.
* Có nhóm huyết áp thấp (14%)<60 mmHg thường ít gặp, có thể do yếu tố sinh lý, thiếu dữ liệu chính xác, hoặc tình trạng sức khỏe đặc thù.







Thống kê phân bố huyết áp âm trương cho Outcome = 1:

Blood Pressure Status Frequency Percentage (%)

Bình thường (60-79 mmHg) 162 60.45

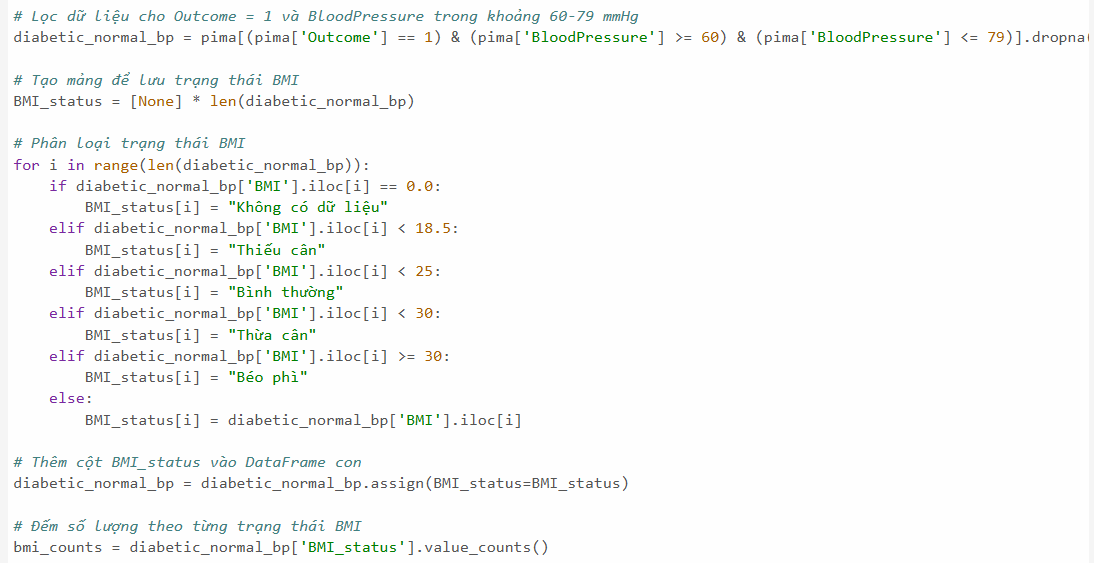
Tiền tăng huyết áp (80-89 mmHg) 61 22.76

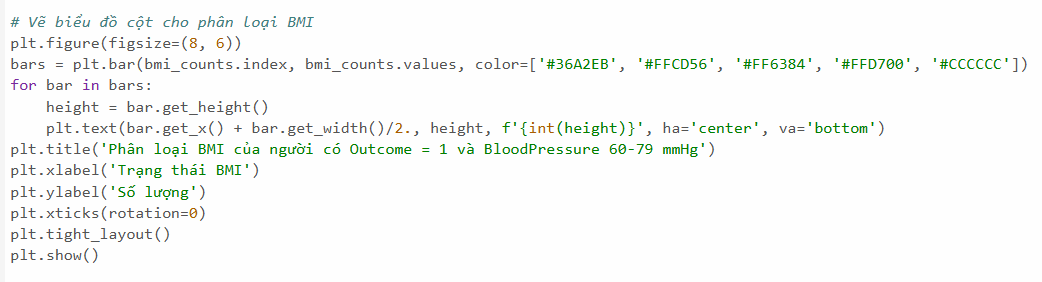
Tăng huyết áp giai đoạn 1 (90-99 mmHg) 20 7.46

Thấp (<60 mmHg) 16 5.97

Tăng huyết áp giai đoạn 2 (≥100 mmHg) 9 3.36

**Nhận xét:** Nhóm tiểu đường có tỷ lệ huyết áp cao hơn Tiền tăng huyết áp + Tăng huyết áp (gđ 1 + gđ 2) = 33.6%. Trong khi ở nhóm không tiểu đường chỉ 23%. Nghĩa là người tiểu đường có xu hướng huyết áp tâm trương cao hơn.





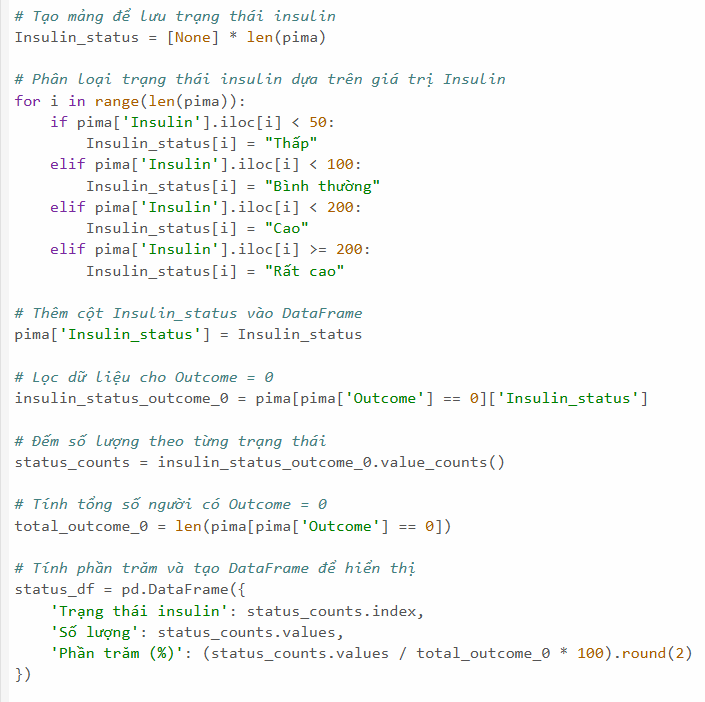


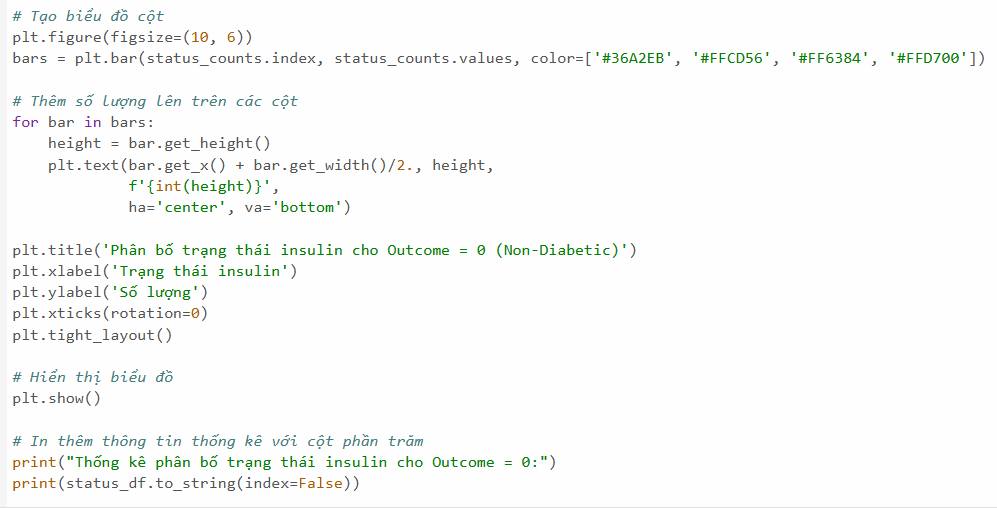
**Nhận xét**: Béo phì chiếm đa số có tới 126 người trong tổng số 162 béo phì trong nhóm này, áp đảo so với các trạng thái khác.Điều này cho thấy béo phì là yếu tố nguy cơ rất lớn với bệnh tiểu đường.

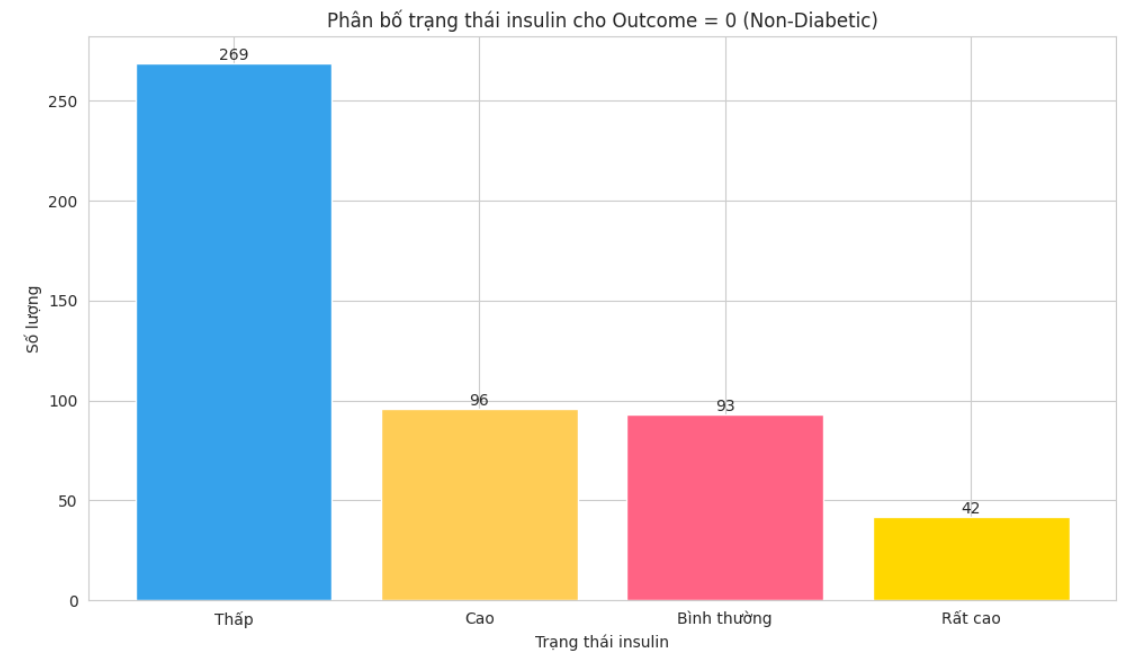
**Có mối liên hệ nào giữa huyết áp tâm trương và nguy cơ mắc bệnh tiểu đường không?**

* Có mối liên hệ giữa huyết áp tâm trương và nguy cơ mắc bệnh tiểu đường trong tập dữ liệu này, nhưng mối liên hệ này không hoàn toàn tuyến tính: Huyết áp cao (≥ 100 mmHg) có liên quan mạnh mẽ nhất với nguy cơ mắc bệnh (56.25%), gợi ý rằng tăng huyết áp nghiêm trọng có thể là yếu tố nguy cơ.

#### 3.2.1.3. So sánh trung bình Insulin theo nhóm Outcome







Thống kê phân bố trạng thái insulin cho Outcome = 0:

Trạng thái insulin Số lượng Phần trăm (%)

Thấp 269 53.8

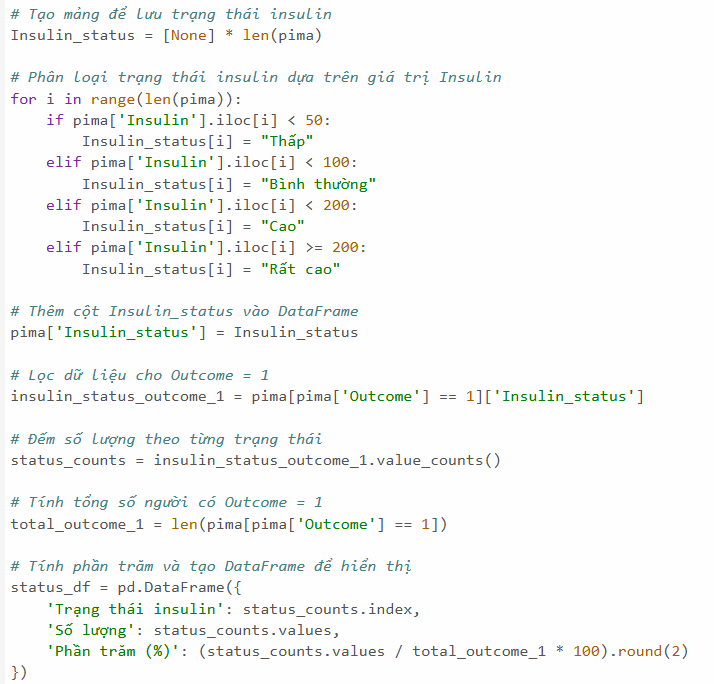
Cao 96 19.2

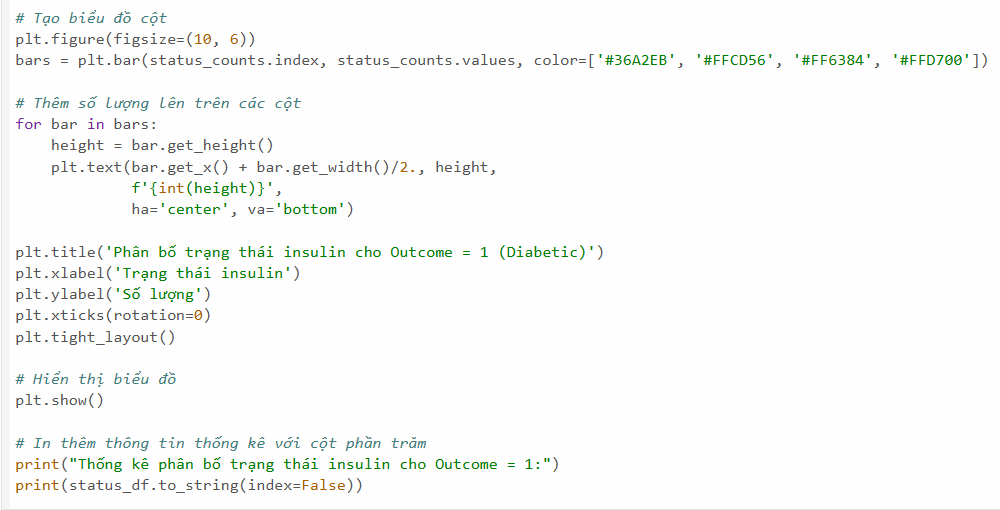
Bình thường 93 18.6

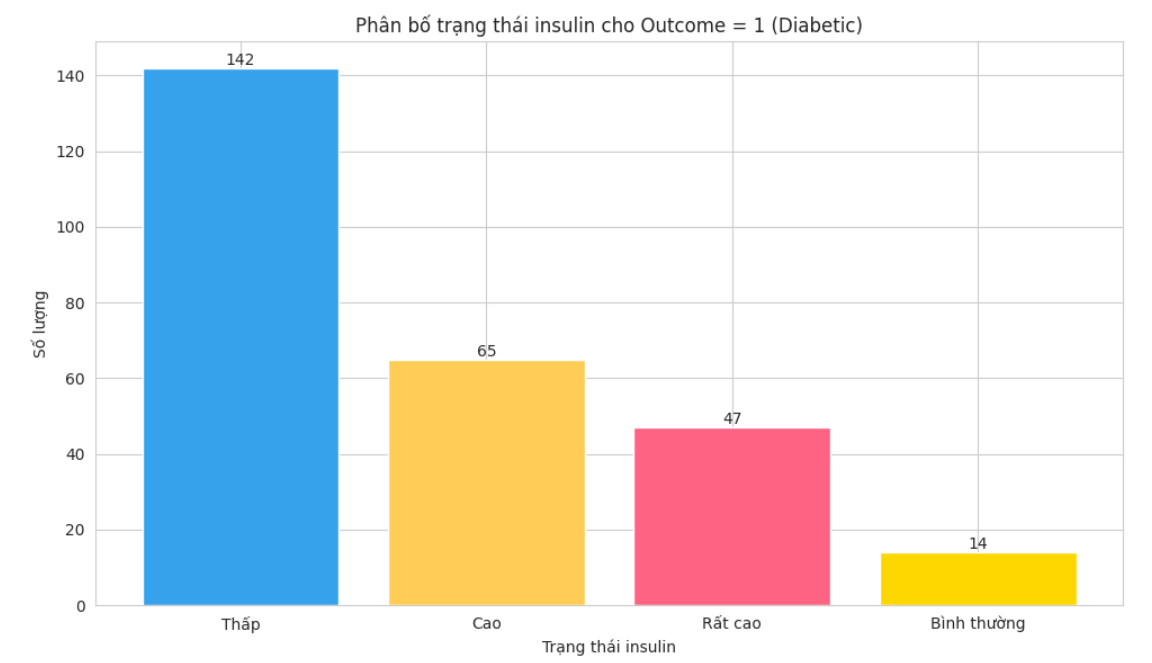
Rất cao 42 8.4

**Nhận xét:**

* Hơn nửa số người (53.8% với 269 người) có mức insulin Thấp, cho thấy xu hướng đáng chú ý trong nhóm không mắc đái tháo đường.
* Các nhóm còn lại phân bố khá đồng đều, với mức Rất cao (8.4% với 42 người) là nhóm nhỏ nhất.







Thống kê phân bố trạng thái insulin cho Outcome = 1:

Trạng thái insulin Số lượng Phần trăm (%)

Thấp 142 52.99

Cao 65 24.25

Rất cao 47 17.54

Bình thường 14 5.22

**Nhận Xét:**

* Hơn nửa số người (52.99% với 142 người) có mức insulin Thấp, cho thấy xu hướng đáng chú ý trong nhóm mắc đái tháo đường.
* Nhóm Bình thường chỉ chiếm tỷ lệ nhỏ (5.22% với 14 người)
* Các nhóm Cao và Rất cao chiếm tỷ lệ đáng kể (24.25% và 17.54% tương ứng). Điều này có thể gợi ý mối liên hệ giữa mức insulin thấp hoặc cao bất thường với bệnh đái tháo đường.

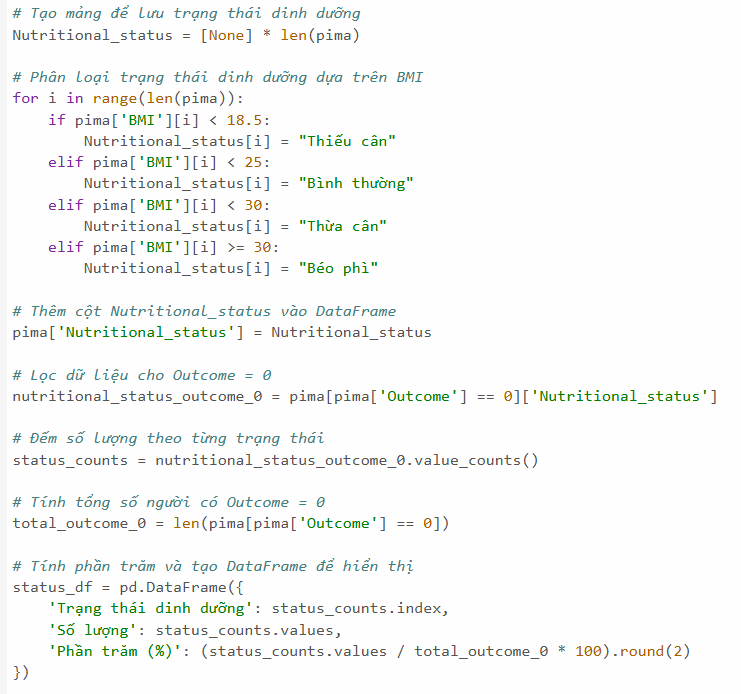
**Sự phân bố nồng độ insulin ở nhóm mắc bệnh và không mắc bệnh có khác biệt không?**

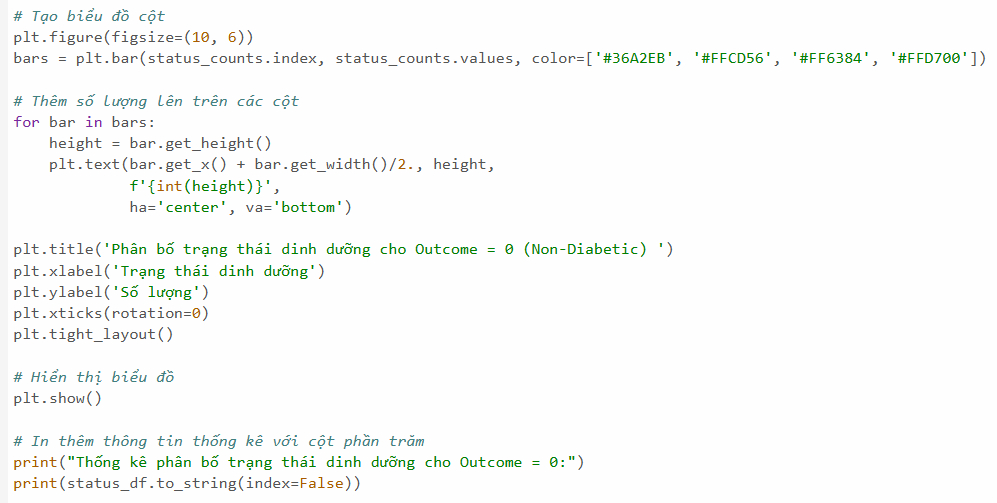
* Có, sự phân bố nồng độ insulin giữa nhóm mắc bệnh và không mắc bệnh có khác biệt. Nhóm Diabetic có tỷ lệ cao hơn ở các mức insulin cao và rất cao (24.25% và 17.54%) so với Non-Diabetic (19.20% và 8.40%), trong khi tỷ lệ ở mức bình thường thấp hơn đáng kể (5.22% so với 18.60%). Sự khác biệt này cho thấy mức insulin bất thường (đặc biệt là rất cao) có thể là yếu tố nguy cơ liên quan đến đái tháo đường, trong khi mức insulin bình thường có xu hướng bảo vệ.

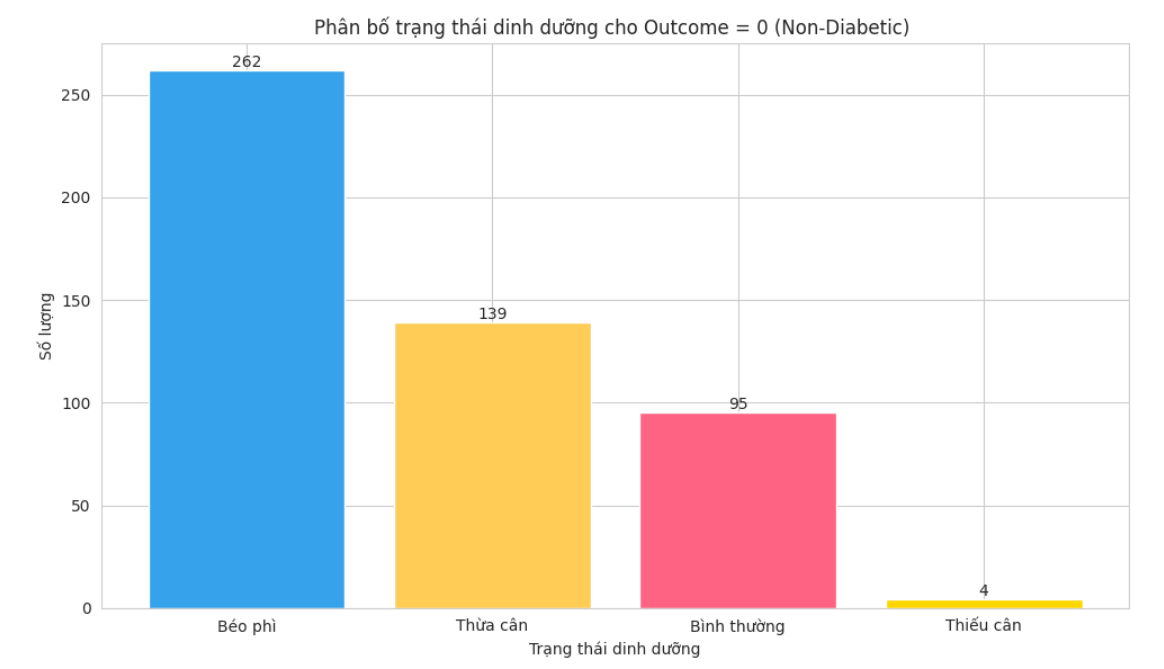
**Nồng độ insulin có mối liên hệ như thế nào với nồng độ glucose? (Có dấu hiệu kháng insulin không?)**

* Nồng độ insulin và glucose có mối quan hệ phức tạp trong tập dữ liệu này. Ở nhóm Non-Diabetic, insulin thấp hoặc bình thường tương ứng với glucose bình thường, phù hợp với cơ chế điều hòa khỏe mạnh. Ở nhóm Diabetic, mức insulin cao hoặc rất cao xuất hiện cùng với glucose tăng (tiền tiểu đường), cho thấy sự mất cân bằng trong điều hòa glucose.

#### 3.2.1.4. So sánh trung bình BMI theo nhóm Outcome







Thống kê phân bố trạng thái dinh dưỡng cho Outcome = 0:

Trạng thái dinh dưỡng Số lượng Phần trăm (%)

Béo phì 262 52.4

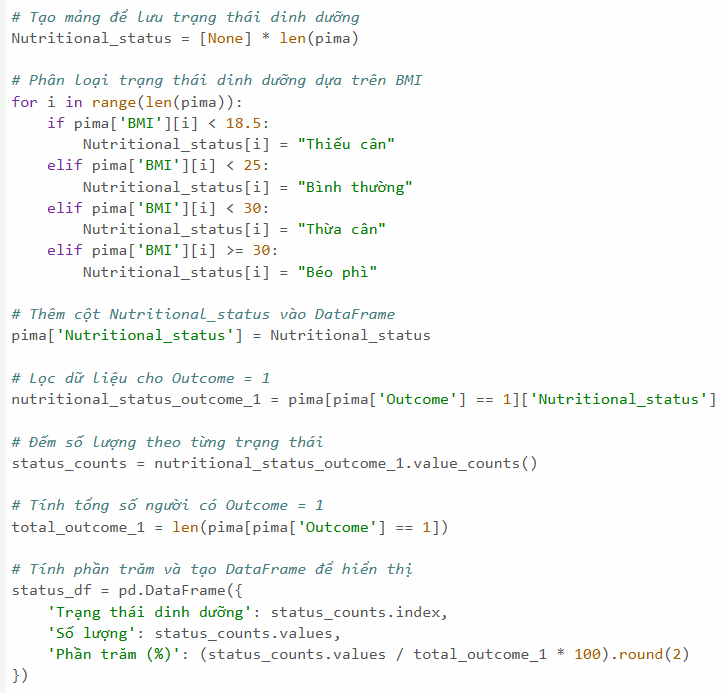
Thừa cân 139 27.8

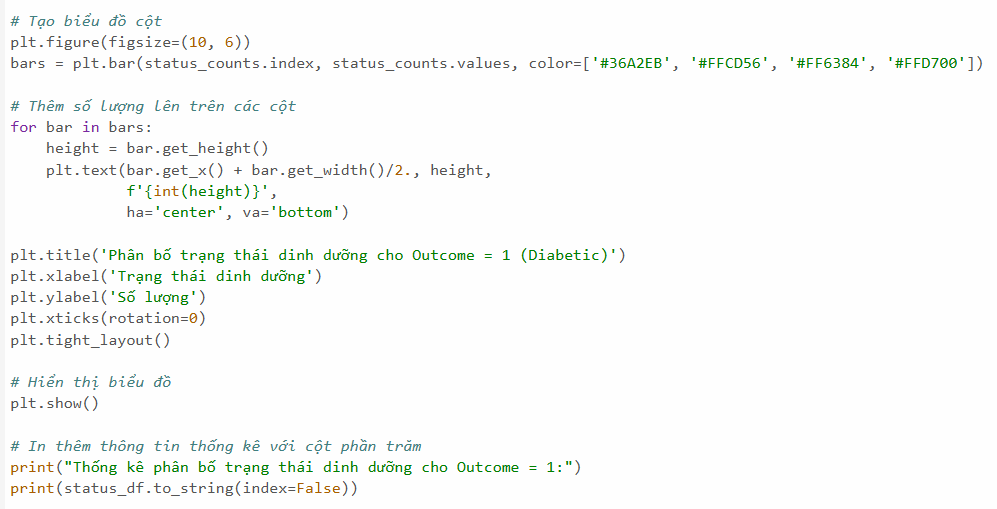
Bình thường 95 19.0

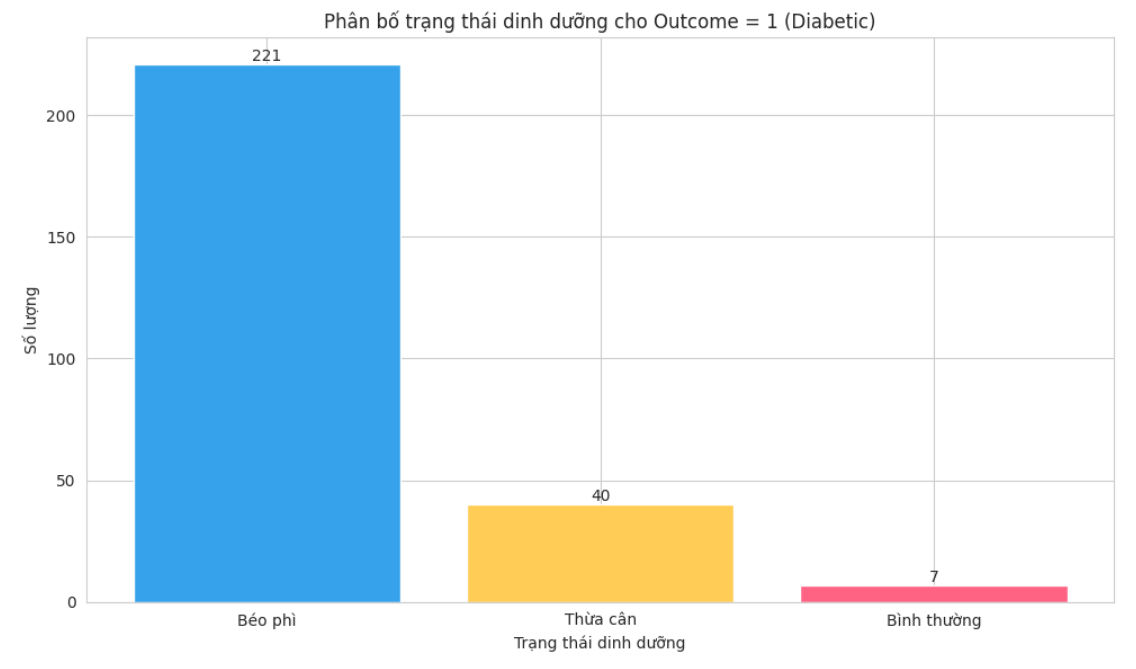
Thiếu cân 4 0.8

**Nhận xét**

* Béo phì vẫn chiếm tỷ lệ cao ở nhóm không tiểu đường (52.4%) → chứng tỏ béo phì không đồng nghĩa chắc chắn sẽ mắc tiểu đường, nhưng là một yếu tố nguy cơ.
* Thừa cân + Bình thường chiếm 46.8% → cao hơn hẳn so với nhóm tiểu đường (chỉ khoảng 18%).







**Thống kê phân bố trạng thái dinh dưỡng cho Outcome = 1:**

**Trạng thái dinh dưỡng Số lượng Phần trăm (%)**

**Béo phì 221 82.46**

**Thừa cân 40 14.93**

**Bình thường 7 2.61**

**Nhận xét**

* **Từ biểu đồ tỉ lệ béo phì rất cao 219 người (81.72%) bệnh nhân tiểu đường thuộc nhóm béo phì.Điều này cho thấy béo phì là yếu tố nguy cơ chính gây tiểu đường trong tập dữ liệu.**
* **Thừa cân cũng đáng kể 40 người (14.93%). Nhóm này cũng có nguy cơ cao so với người bình thường.**

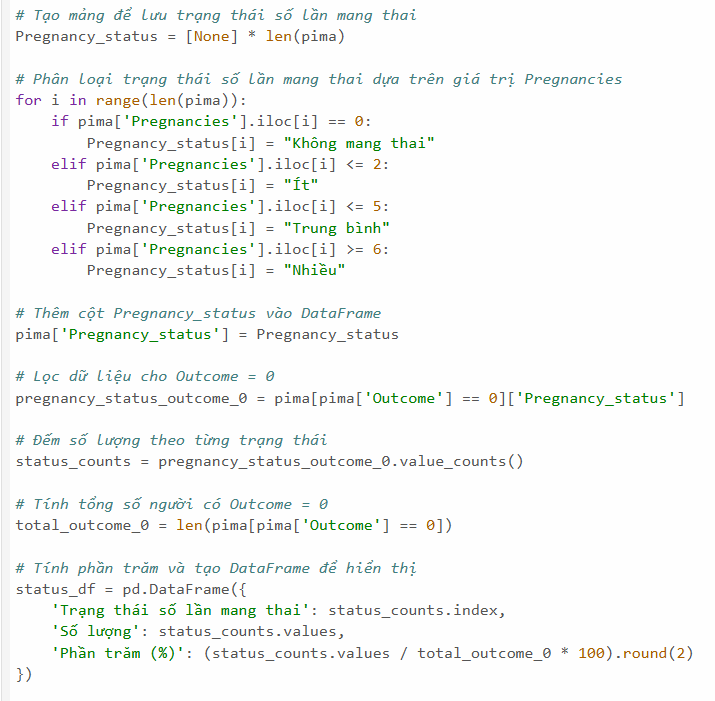
**BMI có phải là yếu tố dự báo mạnh mẽ cho bệnh tiểu đường trong tập dữ liệu này không?**

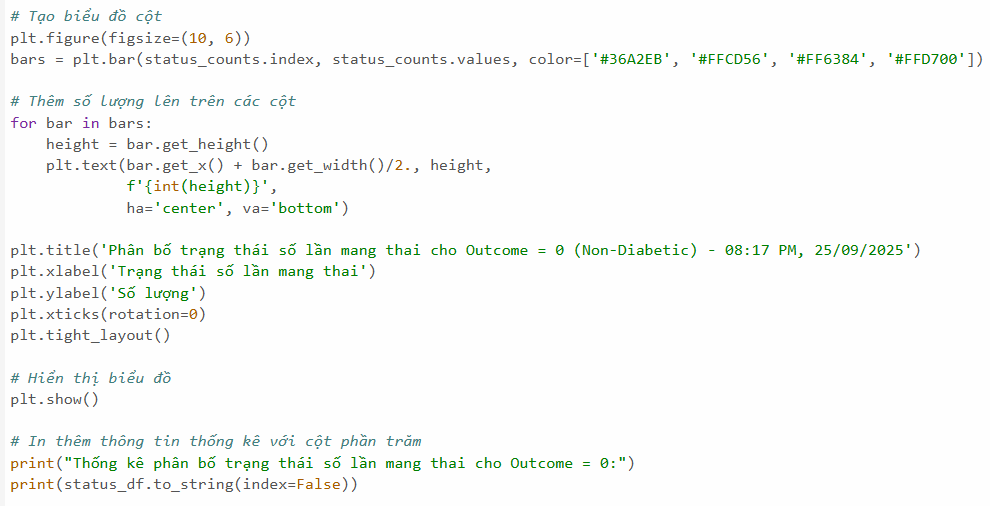
* **Có, BMI dường như là một yếu tố dự báo mạnh mẽ cho bệnh tiểu đường trong tập dữ liệu này, đặc biệt khi xem xét mối liên hệ tích cực và rõ rệt giữa béo phì (BMI ≥ 30) và nguy cơ mắc bệnh. Tỷ lệ béo phì cao trong nhóm Diabetic (82.46%) so với Non-Diabetic (52.40%) và tỷ lệ mắc bệnh tăng dần theo mức BMI cao hơn (45.76% ở béo phì so với 6.86% ở bình thường) củng cố điều này.**

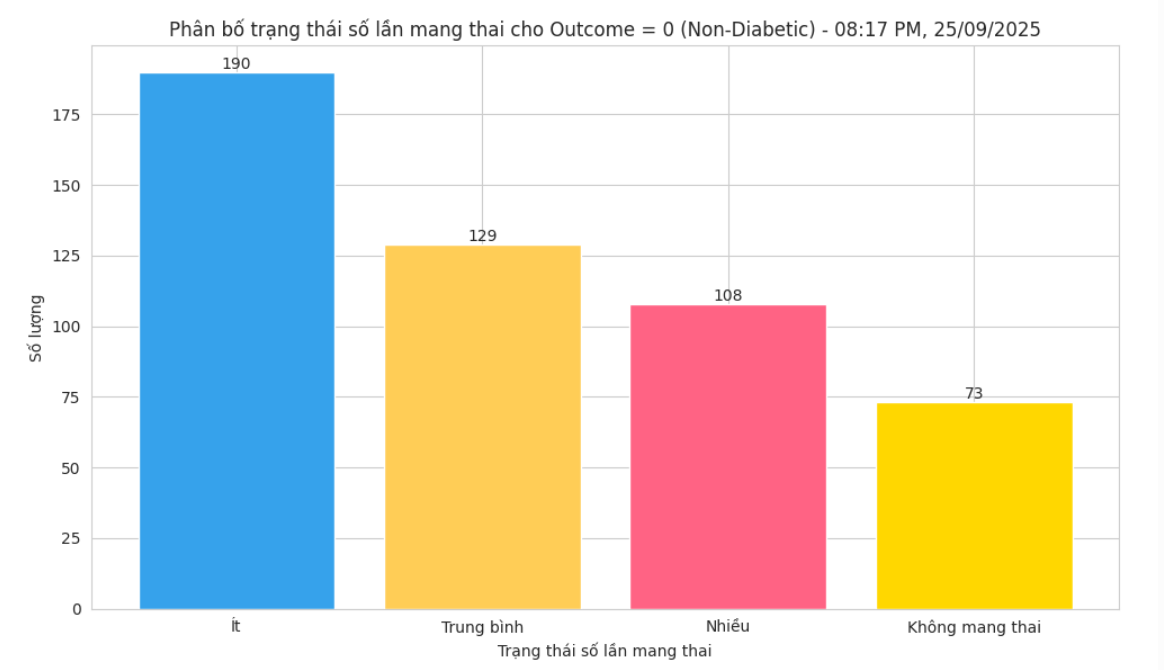
**Tỷ lệ người béo phì, thừa cân, bình thường, thiếu cân ở nhóm mắc bệnh và không mắc bệnh khác nhau như thế nào?**

* **Tỷ lệ phân loại BMI khác biệt rõ rệt giữa hai nhóm, với nhóm Diabetic có tỷ lệ béo phì cao gấp hơn 1.5 lần so với Non-Diabetic, trong khi các nhóm thừa cân, bình thường và thiếu cân đều thấp hơn.**

#### 3.2.1.5. So sánh số lượng mang thai theo nhóm Outcome

****

****

****

**Thống kê phân bố trạng thái số lần mang thai cho Outcome = 0:**

**Trạng thái số lần mang thai Số lượng Phần trăm (%)**

**Ít 190 38.0**

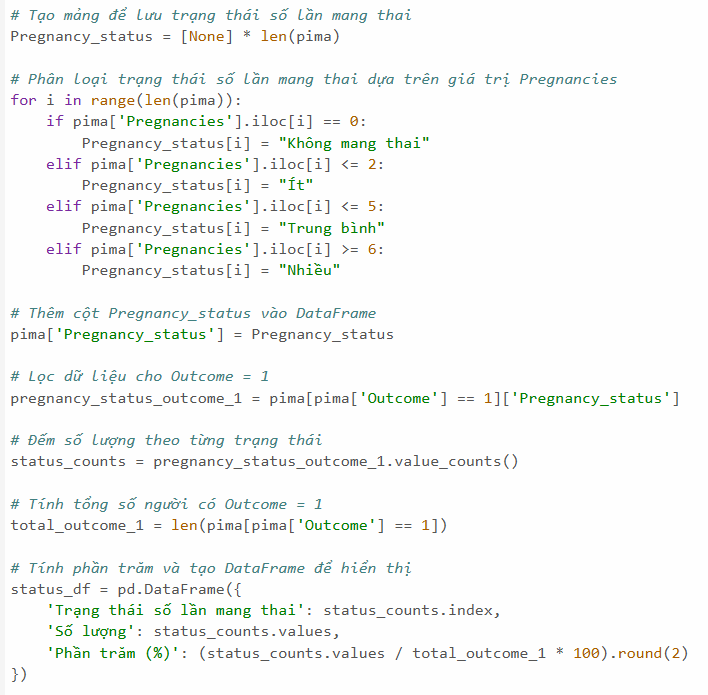
**Trung bình 129 25.8**

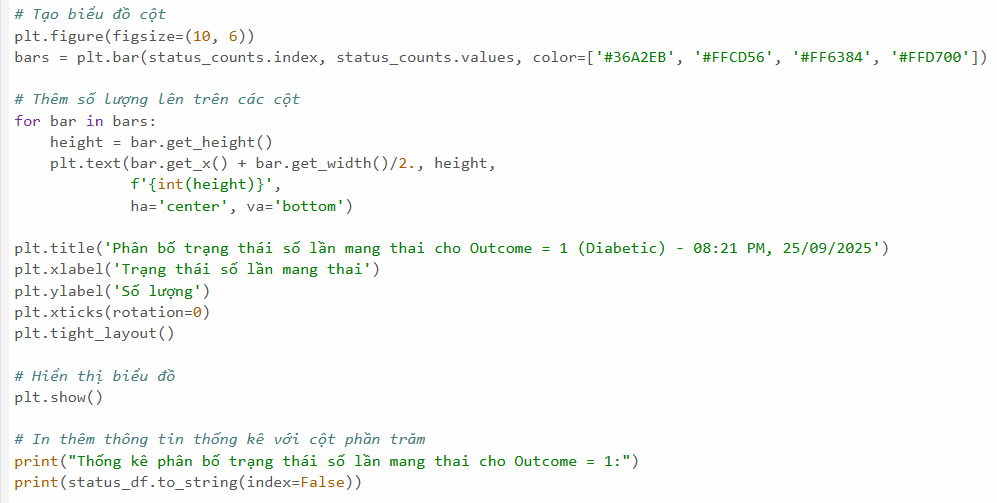
**Nhiều 108 21.6**

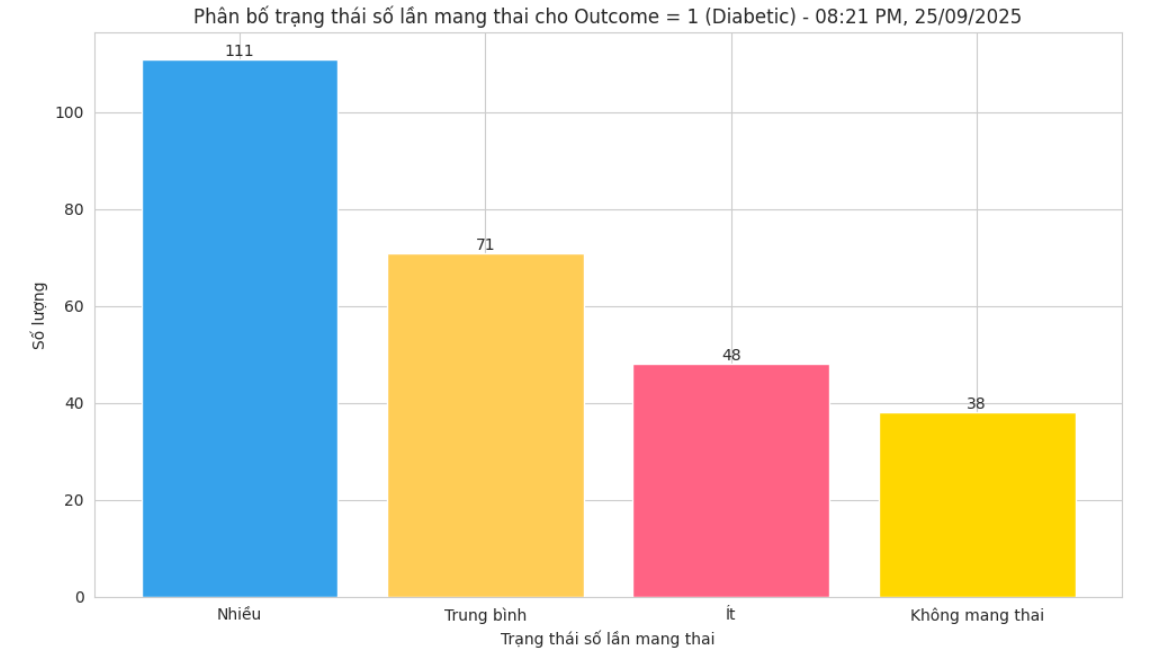
**Không mang thai 73 14.6**

**Nhận xét**

* **Nhóm "Ít" (38.0% với 190 người) chiếm tỷ lệ cao nhất, cho thấy xu hướng phổ biến trong nhóm không mắc bệnh.**
* **Nhóm "Không mang thai" (14.6% với 73 người) là nhóm nhỏ nhất, có thể phản ánh đặc điểm dân số hoặc cách phân loại.**
* **Các nhóm còn lại phân bố khá đồng đều, với sự khác biệt không quá lớn.**

****

****

****

**Thống kê phân bố trạng thái số lần mang thai cho Outcome = 1:**

**Trạng thái số lần mang thai Số lượng Phần trăm (%)**

**Nhiều 111 41.42**

**Trung bình 71 26.49**

**Ít 48 17.91**

**Không mang thai 38 14.18**

**Nhận xét:**

* Nhóm "Nhiều" (41.42% với 111 người) chiếm tỷ lệ cao nhất, cho thấy xu hướng đáng chú ý trong nhóm mắc bệnh.
* Nhóm "Không mang thai" (14.18% với 38 người) là nhóm nhỏ nhất, có thể phản ánh đặc điểm dân số hoặc mối liên hệ với bệnh lý.
* Các nhóm còn lại phân bố khá đồng đều, với sự khác biệt không quá lớn giữa "Trung bình" và "Ít".

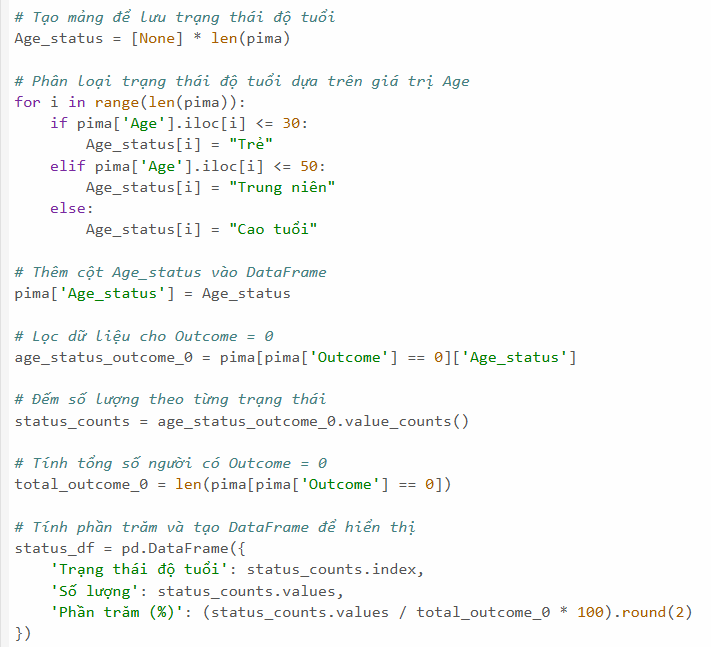
Số lần mang thai trung bình ở nhóm mắc bệnh và không mắc bệnh có khác biệt không?

* Có, số lần mang thai trung bình ở nhóm mắc bệnh (Outcome = 1) cao hơn nhóm không mắc bệnh (Outcome = 0) khoảng 0.91 lần. Sự khác biệt này cho thấy có thể tồn tại mối liên hệ giữa số lần mang thai và nguy cơ mắc đái tháo đường, với nhóm có số lần mang thai cao hơn (trung bình ~3.81) có xu hướng mắc bệnh nhiều hơn.

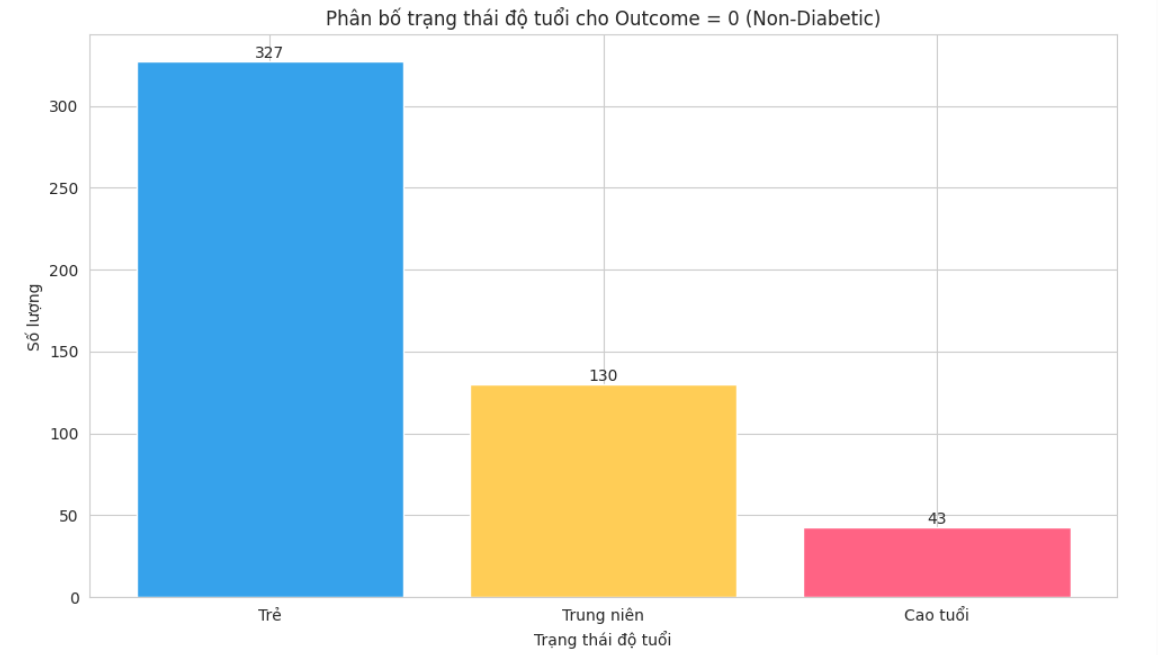
Có nhóm số lần mang thai cụ thể nào có nguy cơ mắc bệnh tiểu đường cao hơn không?

* Có, nhóm số lần mang thai "Nhiều" (≥ 6 lần) có nguy cơ mắc bệnh tiểu đường cao hơn so với các nhóm khác. Điều này có thể liên quan đến tác động lâu dài của thai kỳ lên cơ thể, chẳng hạn như tăng nguy cơ kháng insulin hoặc rối loạn chuyển hóa glucose. Tuy nhiên, để xác định chính xác, cần phân tích thêm các yếu tố khác (như tuổi, BMI, hoặc glucose) và dữ liệu y tế chuyên sâu.

#### 3.2.1.6. So sánh độ tuổi theo nhóm Outcome

****

****

****

**Thống kê phân bố trạng thái độ tuổi cho Outcome = 0:**

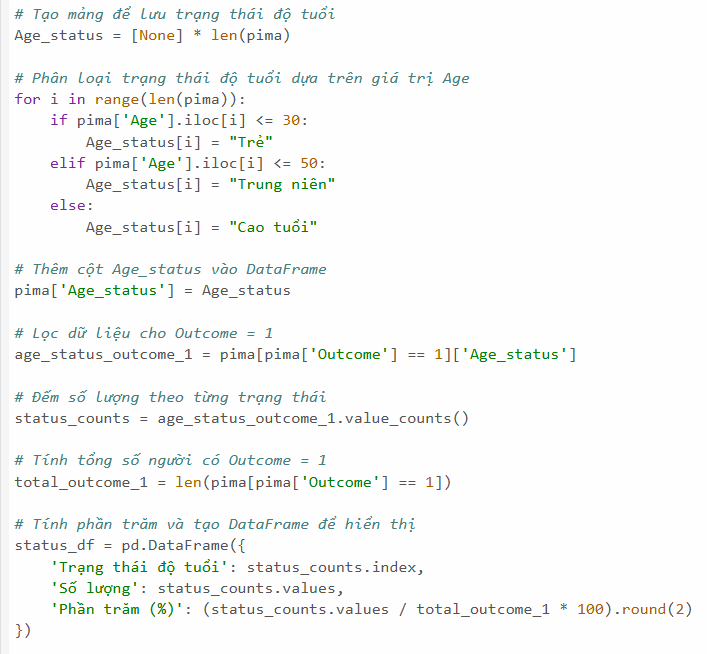
**Trạng thái độ tuổi Số lượng Phần trăm (%)**

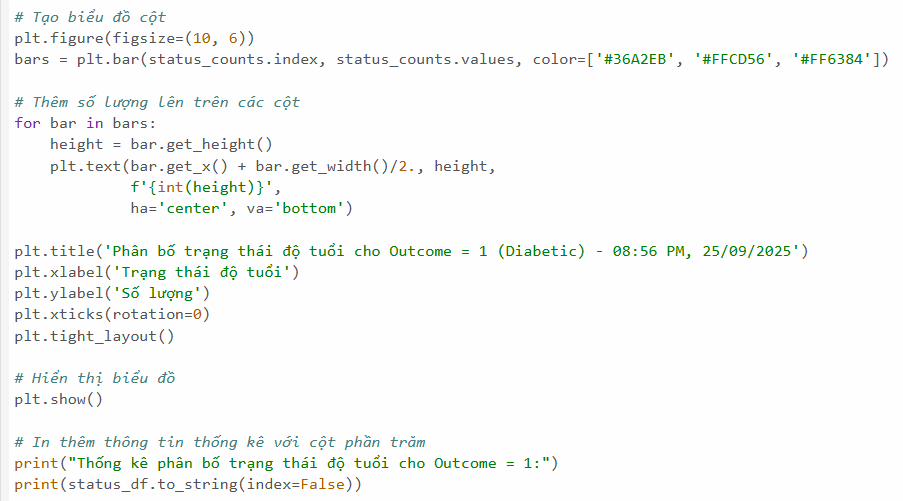
**Trẻ 327 65.4**

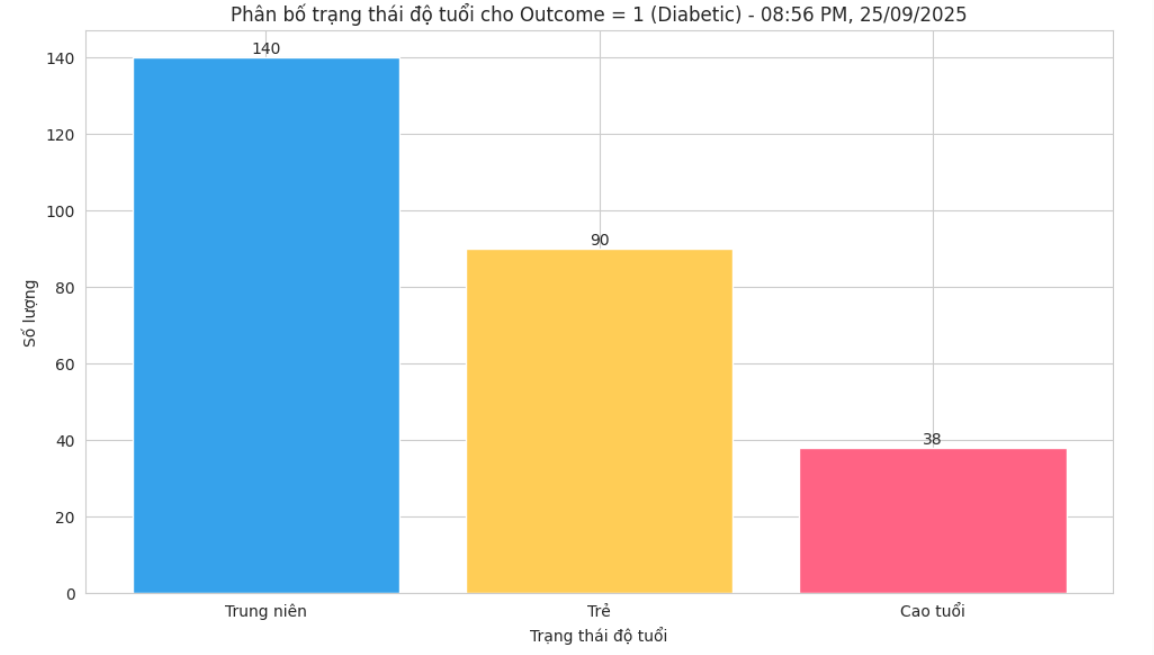
**Trung niên 130 26.0**

**Cao tuổi 43 8.6**

**Nhận xét:** Kết quả cho thấy phần lớn những người không mắc tiểu đường nằm ở nhóm tuổi "Trẻ" (≤ 30 tuổi), điều này có thể gợi ý rằng nguy cơ tiểu đường tăng lên khi tuổi cao hơn

****

****

****

**Thống kê phân bố trạng thái độ tuổi cho Outcome = 1:**

**Trạng thái độ tuổi Số lượng Phần trăm (%)**

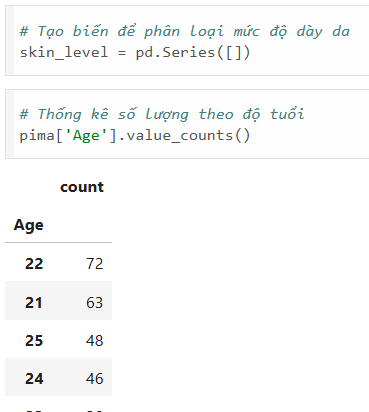
**Trung niên 140 52.24**

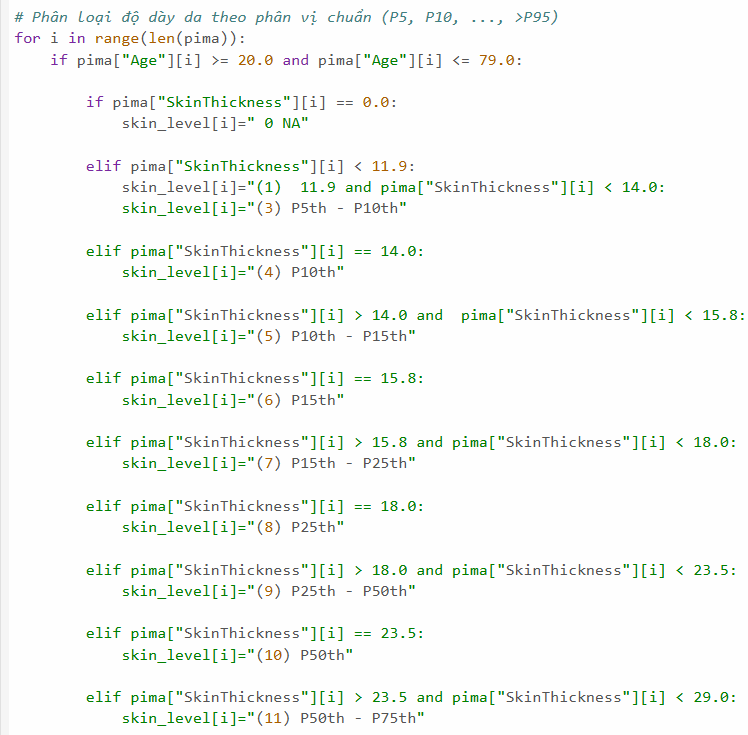
**Trẻ 90 33.58**

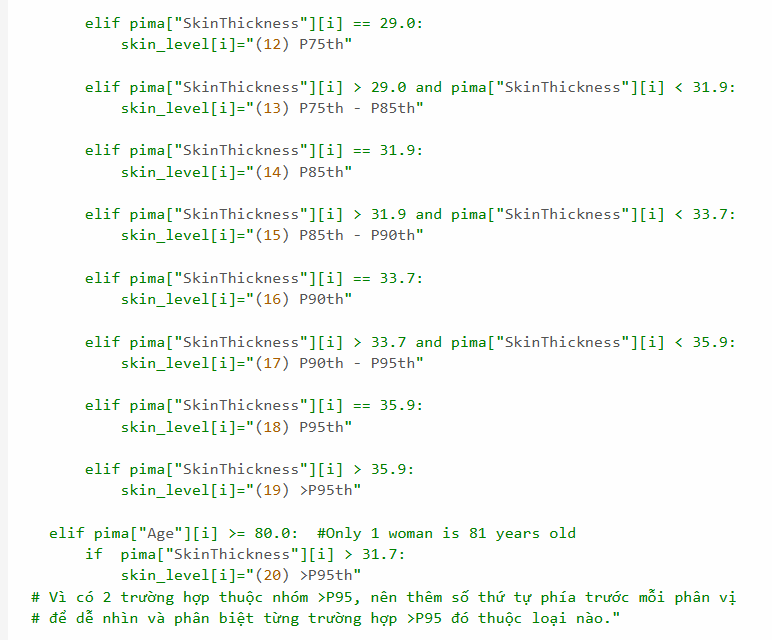
**Cao tuổi 38 14.18**

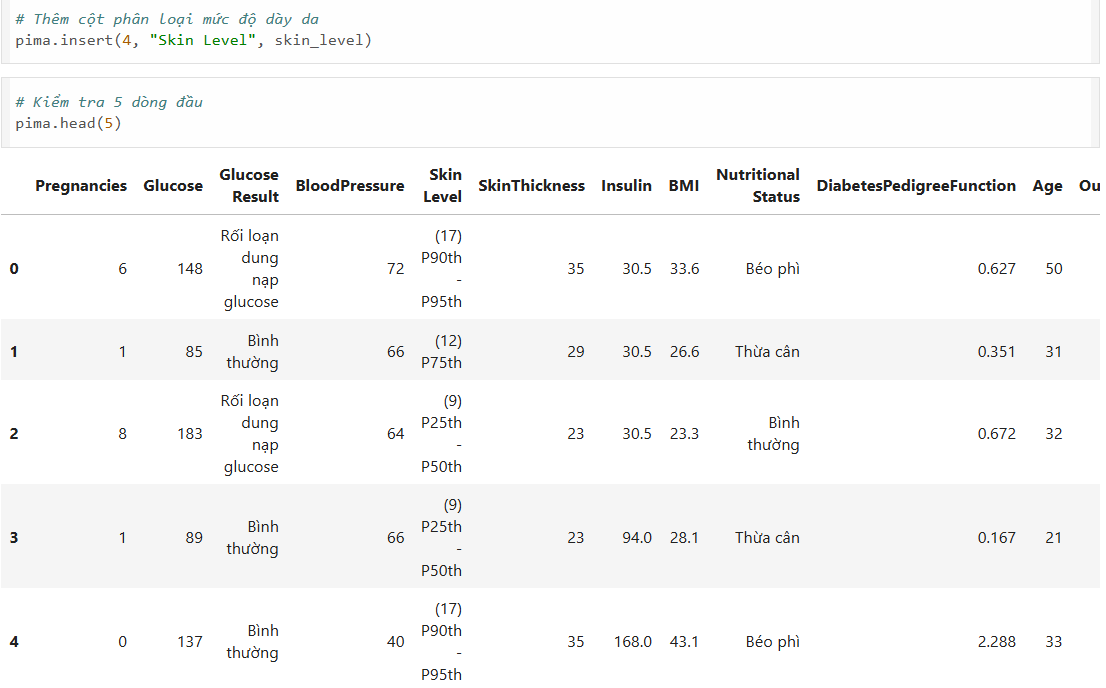
**Nhận xét:** Kết quả cho thấy nhóm "Trung niên" (52.24%) chiếm tỷ lệ cao nhất trong số những người mắc tiểu đường, tiếp theo là "Trẻ" (33.58%) và "Cao tuổi" (14.18%). Điều này gợi ý rằng nguy cơ mắc tiểu đường có thể cao hơn ở độ tuổi trung niên

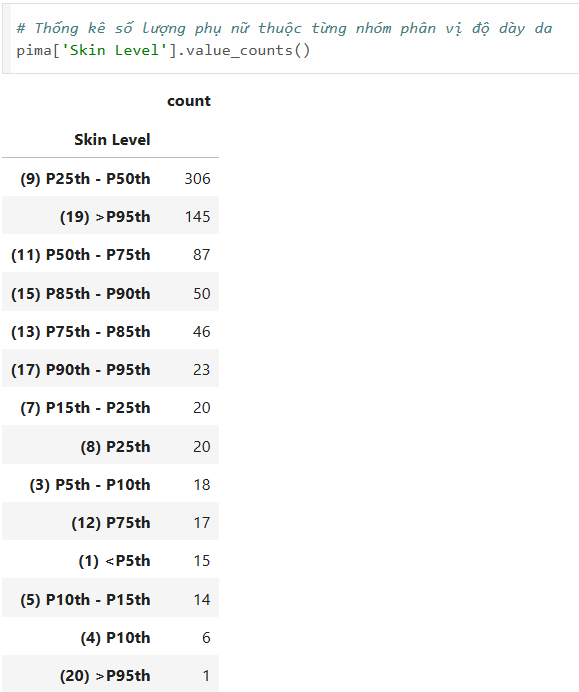
### 3.2.2. Phân tích độ dày da theo độ tuổi

****

****

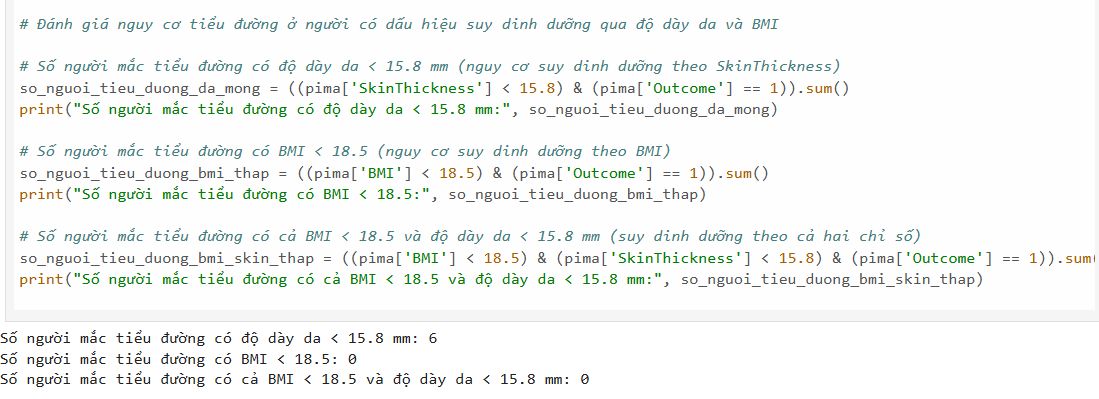
****

****

****

**Nhận xét:**

* Phân vị P25–P50 chiếm phần lớn mẫu nghiên cứu với 306 người, cho thấy đa số phụ nữ trong tập dữ liệu có độ dày da ở mức trung bình. Đây là nhóm thể trạng phổ biến, không quá gầy cũng không quá béo.
* Có 53 người nằm dưới phân vị thứ 15 (P15), phản ánh nhóm thể trạng gầy hoặc có khả năng thiếu dinh dưỡng. Nhóm này có thể cần được phân tích riêng để đánh giá nguy cơ chuyển hóa thấp.
* Chỉ có 1 người nằm trên phân vị thứ 95 (P95), cho thấy mức độ mỡ dưới da cao là rất hiếm gặp trong tập dữ liệu. Đáng chú ý, người này cũng là cá nhân duy nhất từ 80 tuổi trở lên.

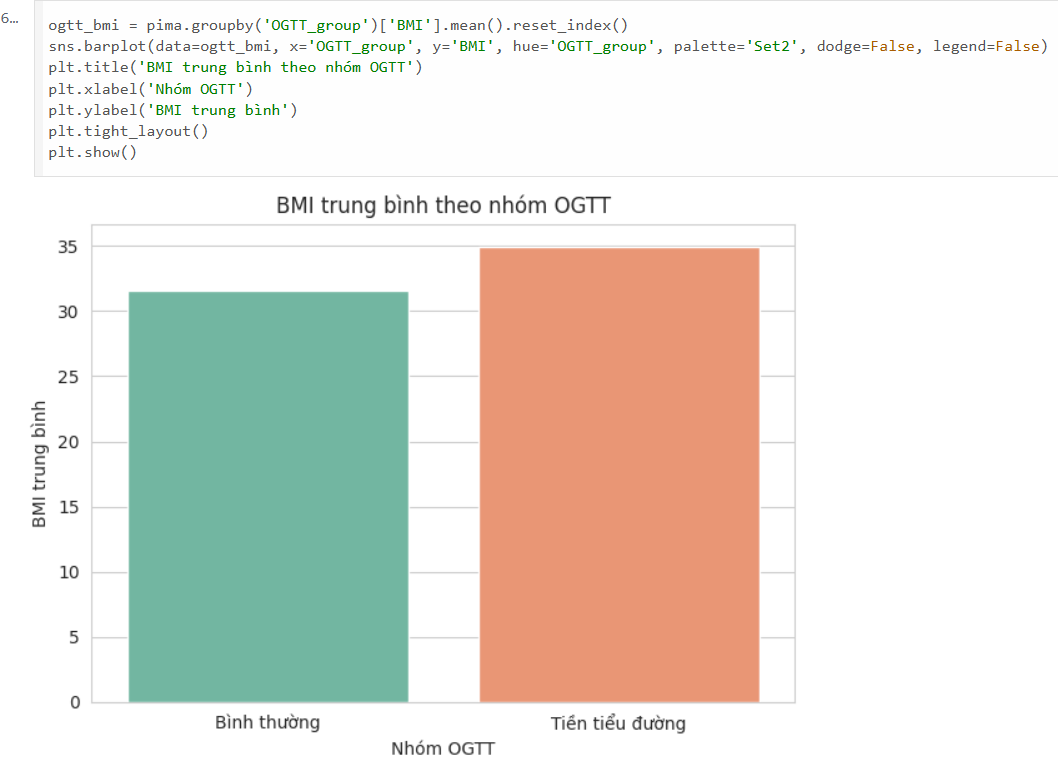
****

**Nhận xét":**

* Có 6 người mắc tiểu đường có độ dày da thấp (< 15.8 mm), cho thấy một nhóm nhỏ bệnh nhân có nguy cơ suy dinh dưỡng theo lớp mỡ dưới da. Đây là dấu hiệu thể trạng gầy, có thể liên quan đến chuyển hóa bất thường.
* Không có người mắc tiểu đường nào có chỉ số BMI thấp (< 18.5), cho thấy suy dinh dưỡng toàn thân là rất hiếm hoặc không xuất hiện trong nhóm bệnh nhân tiểu đường của tập dữ liệu này.
* Không có trường hợp nào đồng thời có cả độ dày da thấp và BMI thấp, cho thấy sự chồng chéo giữa hai nguy cơ suy dinh dưỡng là không xảy ra trong nhóm bệnh nhân tiểu đường.
* Kết quả này cho thấy rằng tuy có một số bệnh nhân tiểu đường có dấu hiệu thể trạng gầy theo độ dày da, nhưng nhìn chung suy dinh dưỡng không phải là đặc điểm phổ biến trong nhóm người mắc bệnh tiểu đường trong tập dữ liệu hiện tại.

## 3.3.Phân tích mối liên hệ giữa các yếu tố nguy cơ và bệnh tiểu đường

**Câu 1: Người có OGTT bất thường có BMI cao hơn trung bình không?**

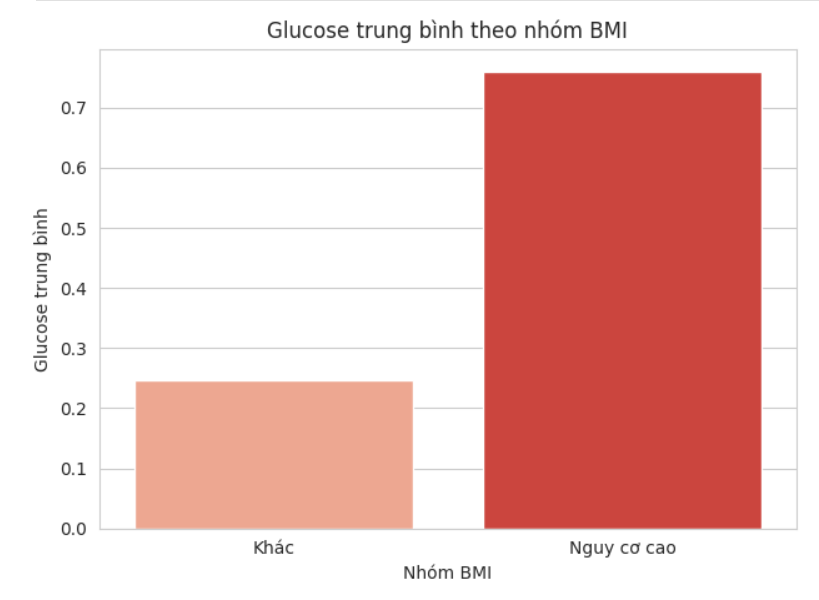
****

**Nhận xét:** Nhìn vào biểu đồ: Nhóm Bình thường có BMI trung bình ≈ 31.5. Và Nhóm Tiền tiểu đường (OGTT bất thường) có BMI trung bình ≈ 35, cao hơn rõ rệt.

* Kết luận: Người có OGTT bất thường (tiền tiểu đường) có BMI trung bình cao hơn so với nhóm bình thường, cho thấy thừa cân/béo phì có thể liên quan đến nguy cơ rối loạn đường huyết.

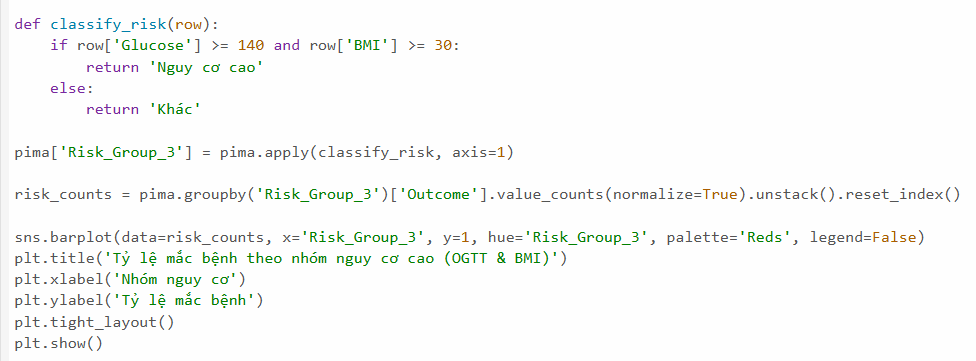
**Câu 2: BMI càng cao thì Glucose càng tăng?**

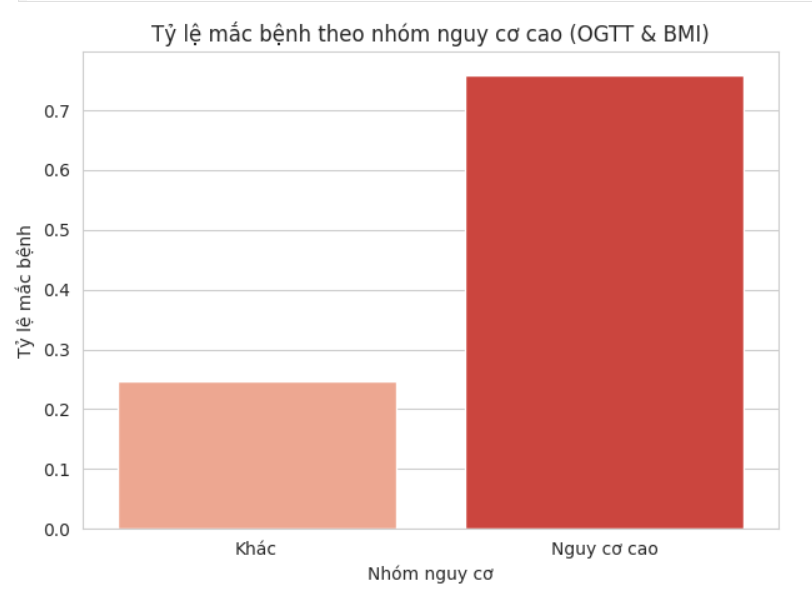
****

****

**Nhận xét:** Dựa trên biểu đồ, có thể thấy glucose trung bình tăng dần theo mức BMI: nhóm gầy khoảng 95, nhóm bình thường khoảng 105, nhóm thừa cân khoảng 112 và nhóm béo phì cao nhất với khoảng 123. Như vậy, BMI càng cao thì glucose trung bình càng tăng, cho thấy thừa cân/béo phì có liên quan đến nguy cơ rối loạn đường huyết cao hơn.

Câu 3: Người có cả OGTT bất thường và BMI ≥ 30 có tỷ lệ mắc bệnh cao nhất không?

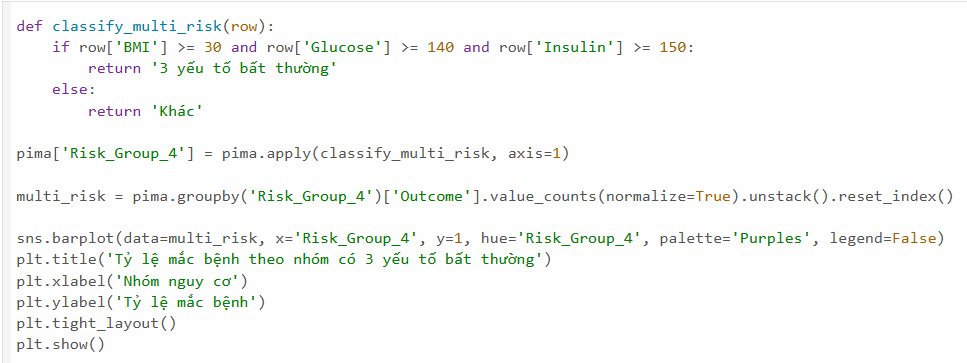
****

****

**Nhận xét:**

* Nhóm “Nguy cơ cao” có tỷ lệ mắc bệnh tiểu đường cao hơn rõ rệt so với nhóm “Khác”.
* Điều này cho thấy sự kết hợp giữa rối loạn dung nạp glucose và béo phì làm tăng nguy cơ mắc bệnh.
* Nhóm “Khác” gồm những người không đồng thời có cả hai yếu tố, và có tỷ lệ mắc bệnh thấp hơn.

**Câu 4: Có sự cộng hưởng giữa nhiều yếu tố bất thường (BMI cao, Glucose cao, Insulin cao) trong nhóm mắc bệnh không?**

****

****

**Nhận xét:**

* Nhóm “3 yếu tố bất thường” có tỷ lệ mắc bệnh lên đến 75%, cao hơn rõ rệt so với nhóm “Khác” (30%).
* Điều này cho thấy sự cộng hưởng giữa các yếu tố nguy cơ làm tăng xác suất mắc bệnh.
* Nhóm “Khác” gồm những người không đồng thời có cả 3 yếu tố, và có tỷ lệ mắc bệnh thấp hơn đáng kể.

# 4. Kết luận

* Nồng độ Glucose huyết tương sau 2 giờ là chỉ số chẩn đoán tiểu đường quan trọng nhất; mức cao phản ánh rõ tình trạng rối loạn chuyển hóa đường.
* Dù phần lớn cả hai nhóm đều có huyết áp tâm trương trong ngưỡng bình thường, người mắc tiểu đường có xu hướng huyết áp cao hơn và tỷ lệ tăng huyết áp cũng vượt trội.
* Mặc dù nhiều người có mức Insulin thấp, nhưng Insulin cao và rất cao có liên hệ rõ với nguy cơ tiểu đường; nhóm mắc bệnh có tỷ lệ Insulin cao cao hơn đáng kể.
* BMI và Glucose có mối liên hệ nhất định: Người có BMI cao thường có Glucose cao hơn, dù mức tương quan không mạnh, nhưng xu hướng cùng tăng là rõ rệt.
* OGTT bất thường đi kèm với BMI cao: Nhóm có kết quả OGTT bất thường (Glucose ≥ 140) có BMI trung bình cao hơn nhóm bình thường, cho thấy tình trạng thừa cân có thể liên quan đến rối loạn dung nạp glucose.
* Cộng hưởng yếu tố nguy cơ làm tăng tỷ lệ mắc bệnh:
  + Nhóm có cả OGTT bất thường và BMI ≥ 30 có tỷ lệ mắc bệnh cao hơn rõ rệt.
  + Khi kết hợp thêm Insulin cao (≥ 150), nhóm có cả 3 yếu tố bất thường có tỷ lệ mắc bệnh lên đến ~75%, cao nhất trong toàn bộ phân tích.
* Phân tích đa yếu tố là cần thiết: Việc đánh giá nguy cơ tiểu đường không nên dựa vào từng chỉ số riêng lẻ, mà cần xem xét tổng thể các yếu tố chuyển hóa như BMI, Glucose, Insulin và tuổi.