

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



BÁO CÁO ĐỒ ÁN
TRỰC QUAN HÓA DỮ LIỆU

GV : Thầy Nguyễn Thanh Bình

Lớp : 18TTH

Họ và tên

MSSV :

Nguyễn Minh Hoàng

18110095

Nguyễn Quốc Bảo

18110053

Mục lục :

1. Mở đầu	4
2. Mô tả thuộc tính	5
3. Trực quan thuộc tính	6
4. Giải pháp	27



1 Mở đầu :

Bệnh tim mạch là bệnh về những rối loạn ảnh hưởng đến tim và các mạch máu bao gồm bệnh mạch vành, mạch máu não, tăng huyết áp, nhồi máu cơ tim, đột quỵ...

Hiện nay 25% dân số mắc bệnh về tim mạch và tăng huyết áp, đặc biệt, trong những năm gần đây, bệnh lý tăng huyết áp đang trẻ hóa.

Theo Tổ chức Y tế thế giới, mỗi năm, toàn cầu có 17,5 triệu người chết về các bệnh tim mạch, nhiều hơn gấp 4 lần tổng số người tử vong của 3 bệnh lý HIV/AIDS, sốt rét và lao phổi.



Nhìn vào các thông số trên , ta đã thấy được sự đáng sợ của bệnh tim gây cho mọi người , hiểu được các tính chất về bệnh tim sẽ giúp chúng ta trang bị các kiến thức cần thiết để đề phòng bệnh tim trong tương lai.

2 Mô tả các thuộc tính của dữ liệu :

Dữ liệu này bắt đầu từ năm 1988 và bao gồm 4 databases : Cleveland, Hungary, Switzerland, and Long Beach V. Nó có 76 thuộc tính bao gồm thuộc tính được dự đoán, nhưng bản thí nghiệm được công bố chỉ sử dụng 14 trong số đó. Cột "target" dùng để chỉ khả năng bệnh tim của bệnh nhân đó (0 nếu bệnh và 1 nếu ngược lại.)

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

age: tuổi của bệnh nhân

sex: Giới tính của bệnh nhân (1 = nam, 0 = nữ)

cp: Kiểu đau ngực

- 0: không có triệu chứng
- 1: đau thắt ngực không điển hình
- 2: không đau thắt ngực
- 3: đau thắt ngực điển hình

trestbps: Huyết áp khi nghỉ ngơi (mm Hg)

chol: Đo cholesterol tính bằng mg / dl

fbs: Đường huyết lúc đói của người đó (> 120 mg / dl, 1 = true; 0 = false)

restecg: Đo điện tâm đồ khi nghỉ ngơi

- 0: bình thường
- 1: có bất thường sóng ST-T
- 2: hiển thị phì đại thất trái có thể xảy ra hoặc xác định theo tiêu chuẩn của Estes

thalach: Nhịp tim tối đa của một người đạt được

exang: Đau thắt ngực do tập thể dục (1 = có; 0 = không)

oldpeak: ST trầm cảm gây ra do tập thể dục liên quan đến nghỉ ngơi ('ST' liên quan đến các vị trí trên đồ thị ECG)

slope: độ dốc của đoạn ST tập thể dục đỉnh cao

- 0: dốc xuống
- 1: phẳng
- 2: lên cao

ca: số lượng các mạch chính liên quan đến bệnh tim (0-3)

thal: Một chứng rối loạn máu được gọi là thalassemia

- 0: **NULL** (bị loại bỏ từ tập dữ liệu trước đó)
- 1: khuyết tật cố định (không có lưu lượng máu ở một số phần của tim)
- 2: lưu lượng máu bình thường
- 3: khuyết tật có thể hồi phục (quan sát thấy dòng máu chảy nhưng không bình thường)

target: Chuẩn đoán bệnh (0 = có, 1 = không)

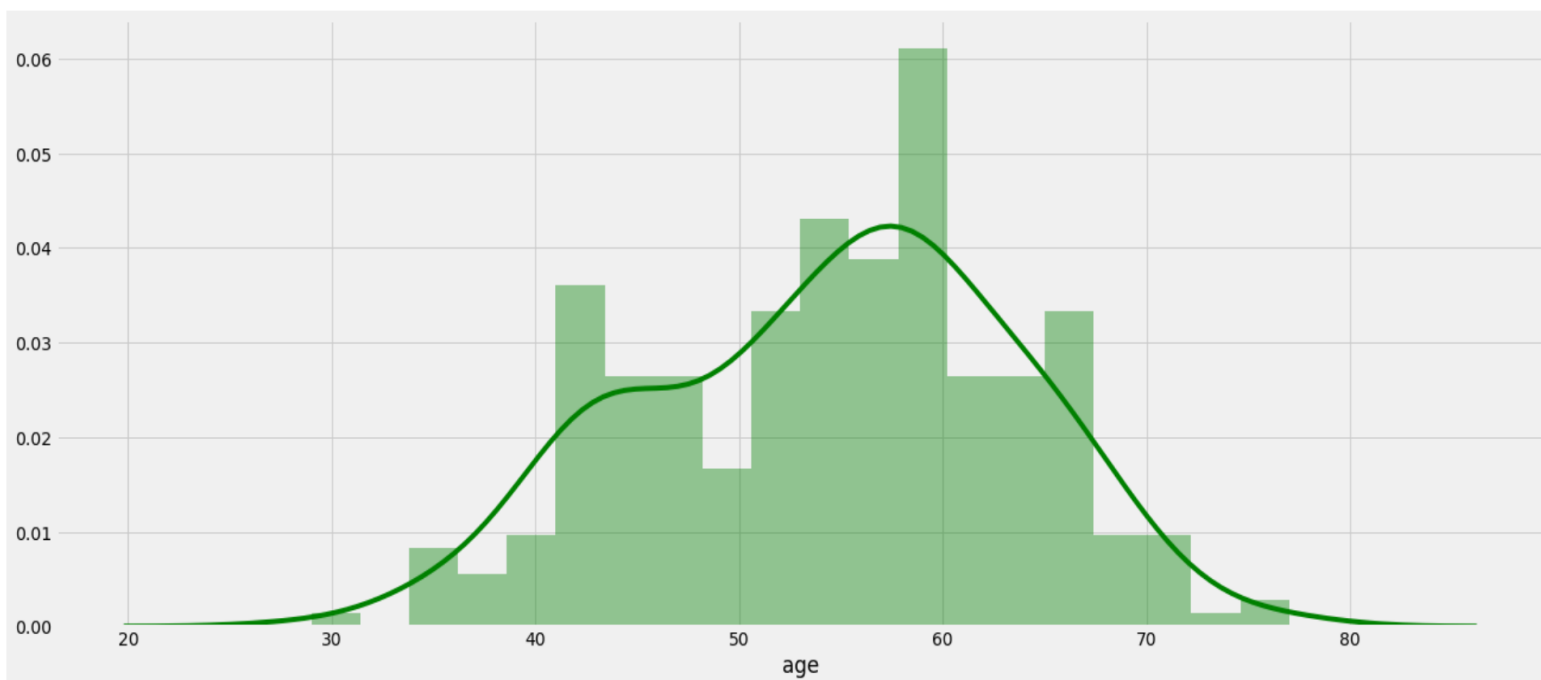
3 Trục quan hóa các thuộc tính :

Bệnh tim mạch có rất nhiều nguyên nhân gây bệnh, chính vì vậy cũng có nhiều yếu tố nguy cơ dẫn đến các bệnh lý về tim mạch. Hiểu rõ các yếu tố nguy cơ này sẽ giúp ta phòng tránh bệnh tim mạch hiệu quả nhất.

a) Trục quan 1 thuộc tính :

Tuổi tác

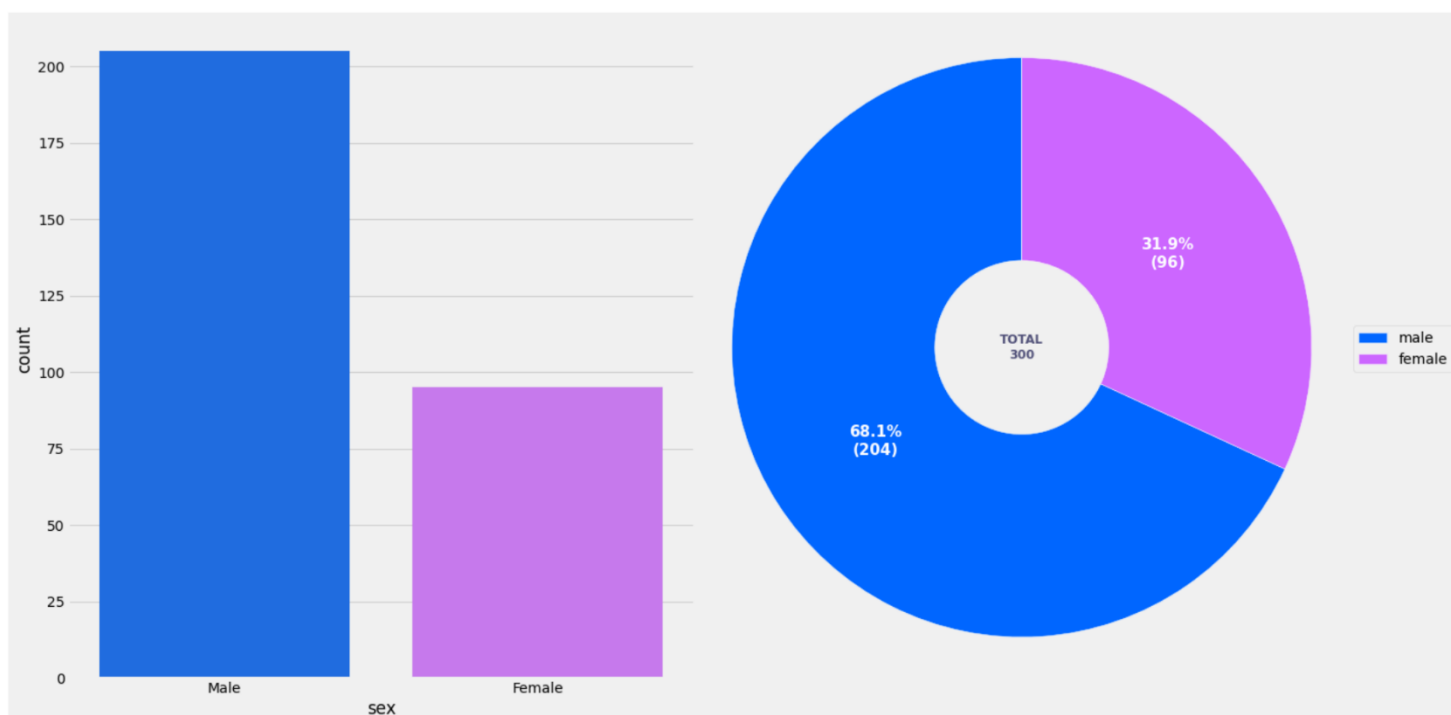
Tuổi già là một trong những yếu tố nguy cơ của bệnh tim mạch. Càng lớn tuổi, hoạt động của tim càng kém hiệu quả. Thành tim dày lên, các động mạch cứng lại khiến cho quá trình bơm máu cũng trở nên khó khăn, đó là lý do vì sao nguy cơ mắc bệnh tim mạch gia tăng theo tuổi.



Dựa vào hình trên ta thấy rằng , tập dữ liệu này sẽ tập trung khảo sát ở các nhóm tuổi từ 50- 65 là nhiều nhất , vì đó là nhóm tuổi lớn , có khả năng gây bệnh tim cao.

Giới tính:

Kết quả nghiên cứu bệnh tim mạch cho thấy, trong độ tuổi từ nhỏ đến trung niên, tỷ lệ nam giới bị bệnh tim mạch nhiều hơn nữ giới, nhưng đến tuổi mãn kinh thì tỷ lệ bệnh tim mạch ở nữ giới xấp xỉ nam giới.



Vì vậy xét theo mặt bằng chung thì nam giới có khả năng bệnh tim nhiều hơn nữ giới, nên tập dữ liệu của chúng ta sẽ tập trung khảo sát nam nhiều hơn với số lượng nam chiếm gần 70% trong tập khảo sát , gấp đôi số lượng nữ .

Đau ngực :

Đau ngực là cảm giác đau ở vùng ngực từ mức ngang vai đến trên cơ hoành. Triệu chứng có thể biểu hiện với nhiều mức độ, xuất hiện đột ngột hay tái diễn, có thể đi kèm với các triệu chứng khác như khó thở, vã mồ hôi ... Có nhiều **nguyên nhân gây đau ngực**, trong đó có các nguyên nhân là nghiêm trọng có thể ảnh hưởng đến tính mạng như do bệnh lý phổi, **tim mạch** .

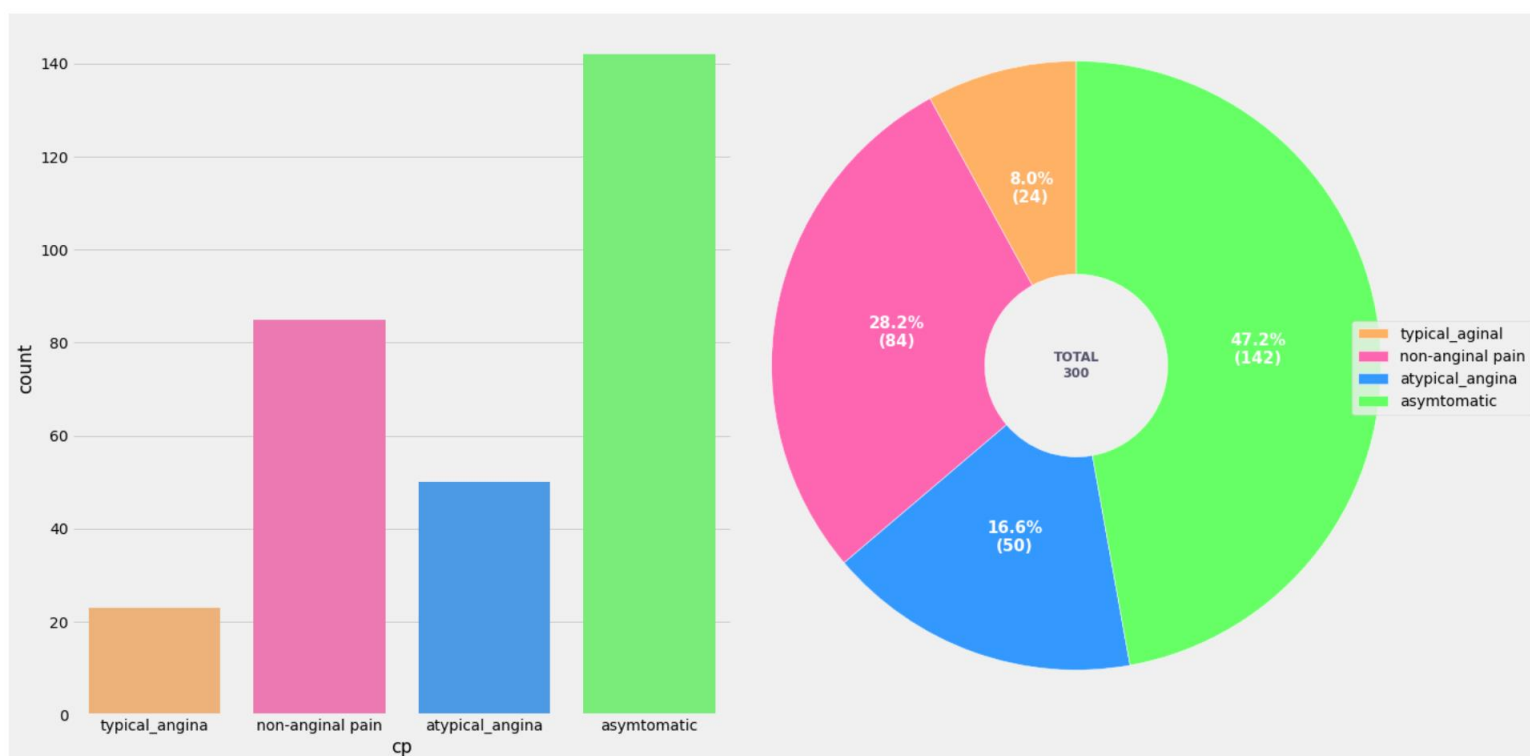


Các loại đau ngực :

Đau thắt ngực điển hình là cảm giác khó chịu được ghi nhận khi tim không nhận đủ máu hoặc oxy. Thông thường, điều này là do sự tắc nghẽn của mảng bám tích tụ trong động mạch vành.

Đau thắt ngực không điển hình được định nghĩa là cơn đau thắt ngực thay đổi hoặc trầm trọng hơn. Nó có ít nhất một trong ba tính năng sau: nó xảy ra khi nghỉ ngơi (hoặc khi gắng sức tối thiểu), thường kéo dài hơn 10 phút nó nghiêm trọng và mới khởi phát (tức là trong vòng 4-6 tuần trước) nó xảy ra với một mô hình siêu việt (tức là, rõ ràng là nghiêm trọng hơn, kéo dài hoặc thường xuyên hơn trước)

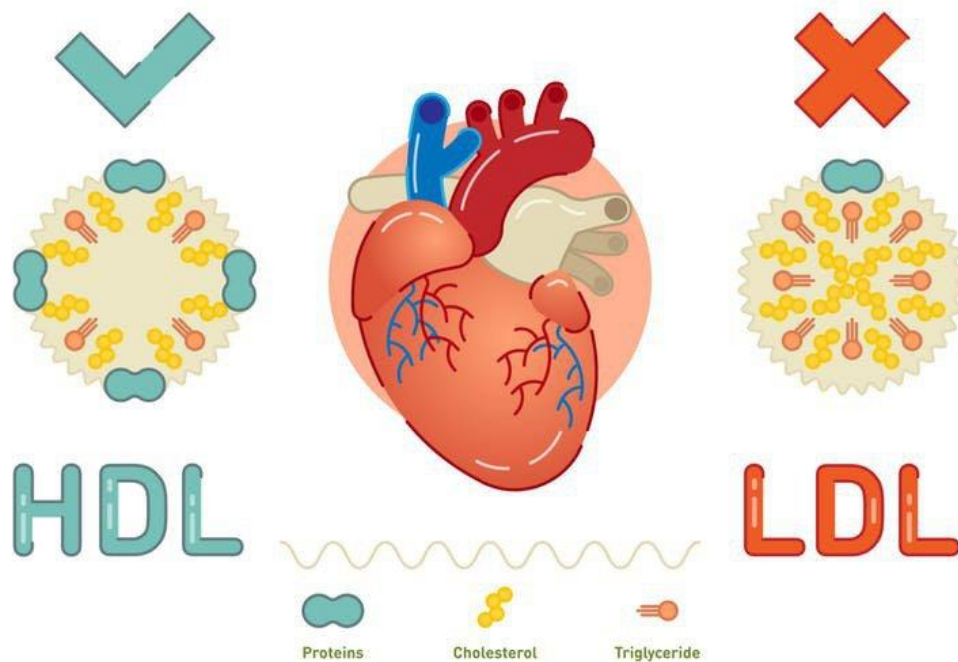
Đau thắt ngực không triệu chứng gây ra bởi các vấn đề với các động mạch nhỏ bên trong cơ tim, đóng vai trò quan trọng trong việc điều chỉnh cung cấp máu cho tim. Trước đây nó được gọi là hội chứng tim X



Dựa vào đồ thị trên ta thấy rằng trong các loại đau ngực thì số bệnh nhân đau ngực không triệu chứng chiếm nhiều nhất với gần 50% , thứ hai là không đau ngực (28.3%) , còn lại là đau ngực điển hình và không điển hình .

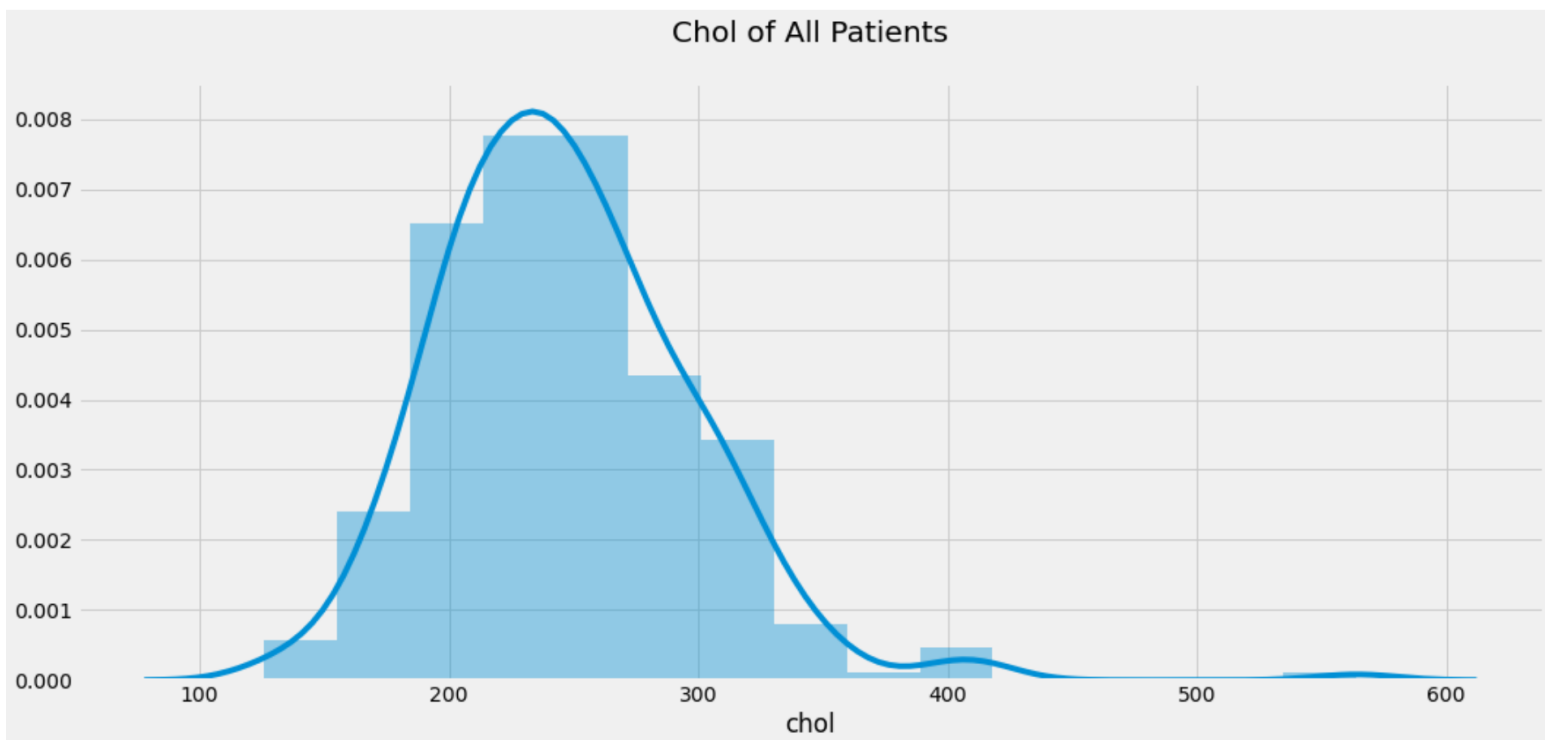
Cholesterol :

Là một chất dạng sáp, mềm có thể được tìm thấy trong máu và trong toàn bộ các tế bào của cơ thể. 75% cholesterol được tạo ra ở gan và 25% còn lại đến từ thức ăn. Trong cơ thể người Cholesterol có một vai trò quan trọng vì được sử dụng để hình thành màng tế bào, một số loại hormon và cần thiết cho nhiều chức năng khác của cơ thể. Tuy nhiên, nồng độ cholesterol cao quá mức cho phép sẽ ảnh hưởng xấu tới sức khỏe, đặc biệt là gây ra các **bệnh tim mạch** điển hình như: bệnh mạch vành và **nhồi máu cơ tim**



Cholesterol không hòa tan trong máu. Do đó khi cholesterol lưu thông trong máu nó phải được bao quanh bằng một lớp áo protein hay còn gọi là lipoprotein. Có 2 loại lipoprotein quan trọng là lipoprotein tỷ trọng thấp (**LDL**) và lipoprotein tỷ trọng cao (**HDL**).

Cholesterol loại LDL chuyên chở hầu hết cholesterol của cơ thể. Khi có nhiều LDL trong máu, thành động mạch sẽ lắng đọng mỡ gây ra xơ vữa động mạch, do đó LDL được gọi là **cholesterol xấu**. Ngược lại, HDL lại lấy cholesterol ra khỏi máu và ngăn chúng không xâm nhập vào thành động mạch, do đó **HDL** được gọi là **cholesterol tốt**.



Ở biểu đồ này , ta nhận thấy rằng lượng chol của các bệnh nhân tập trung ở mức 200-300. Đây được đánh giá là lượng cholesterol khá cao so với người bình thường . Khi lượng cholesterol vượt mức cho phép sẽ là nguyên nhân dẫn tới bệnh tim và đột quỵ, hai nguyên nhân gây tử vong hàng đầu ở Mỹ.

Huyết áp :

Huyết áp là áp lực của mạch máu lên thành động mạch. Huyết áp thì bằng cung lượng tim nhân với sức cản ngoại vi từ công thức này ta có thể biết được những yếu tố làm tăng giảm huyết áp.

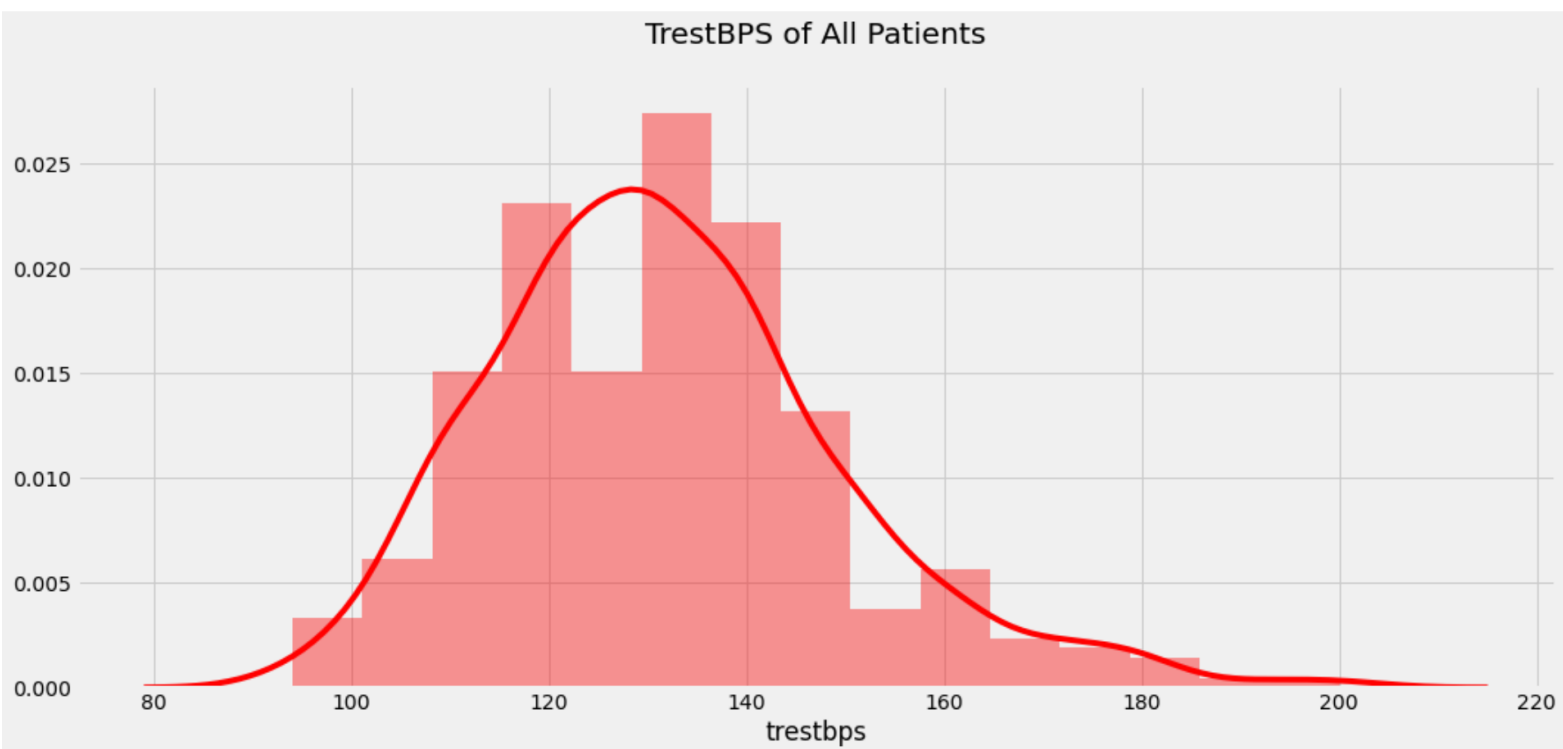
Chứng cao huyết áp sẽ dẫn đến nguy cơ đau tim và đột quỵ

Huyết áp có thể thay đổi tùy theo điều kiện vận động và tuổi tác, nhưng về cơ bản, chỉ số huyết áp ở người lớn khi đang nghỉ ngơi nên ở mức 120/80.

Tăng huyết áp là một trong những căn bệnh không lây nhiễm phổ biến nhất thế giới hiện nay nhưng về lâu dài, căn bệnh này lại là điều đáng lo ngại với rất nhiều những biến chứng trầm trọng trong đó có suy tim.

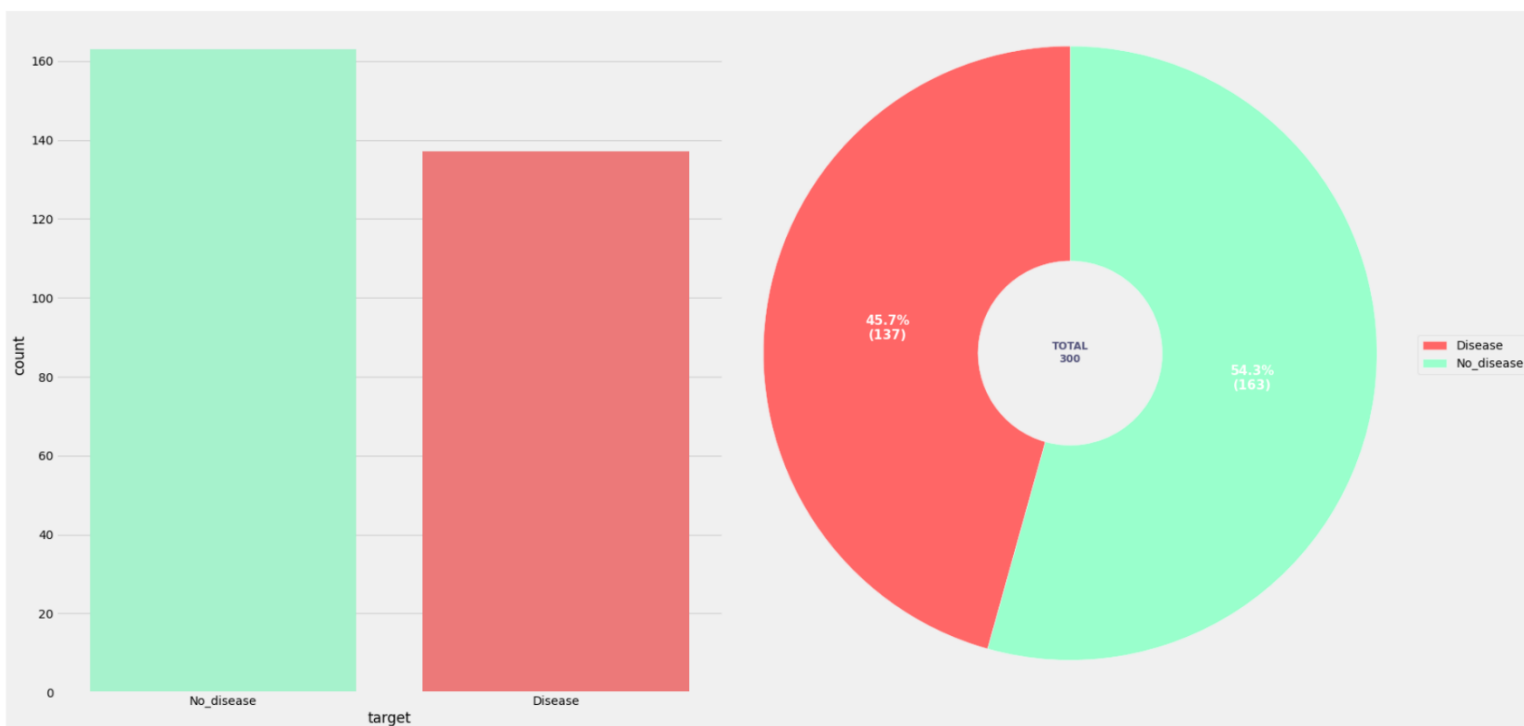


Được xác định nếu con số huyết áp luôn vượt quá **140/90mmHg**. Tăng huyết áp được coi là **“kẻ giết người thầm lặng”** vì huyết áp tăng thường không có triệu chứng gì và gây ra một loạt các biến chứng nguy hiểm đến tim mạch như bệnh lý động mạch vành, bệnh động mạch chủ, bệnh động mạch ngoại biên...



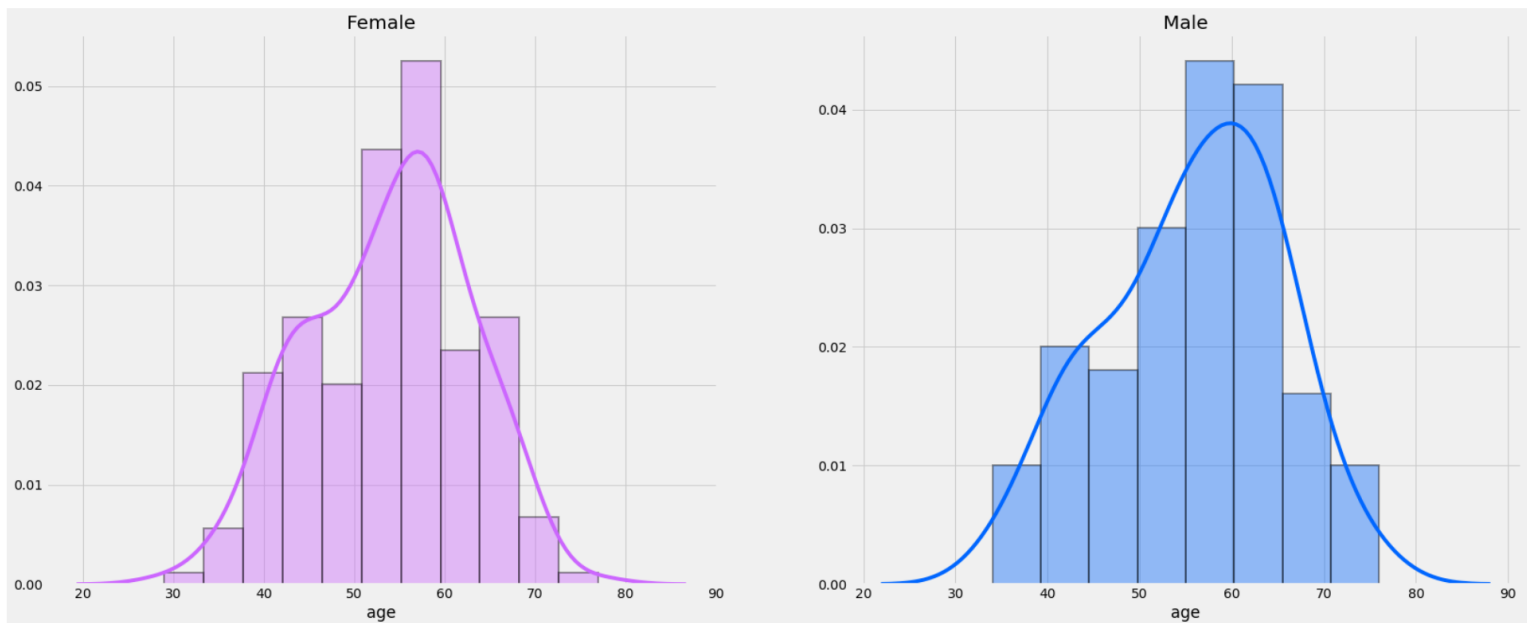
Đồ thị trên thể hiện lượng huyết áp của các bệnh nhân khảo sát tập trung ở khoảng 120-140 mmHg. Chỉ số này thuộc loại tăng huyết áp độ 1 và tăng huyết áp độ 2, thuộc 2 loại huyết áp dẫn đến khả năng bệnh tim cao .

Kết quả chuẩn đoán :



Dựa vào hình trên , ta nhận thấy rằng trong số 300 bệnh nhân , thì số bệnh nhân không mắc bệnh chiếm gần 55% nhiều hơn 10% so với số bệnh nhân mắc bệnh. Chỉ số này tuy lệch nhưng không đáng kể .

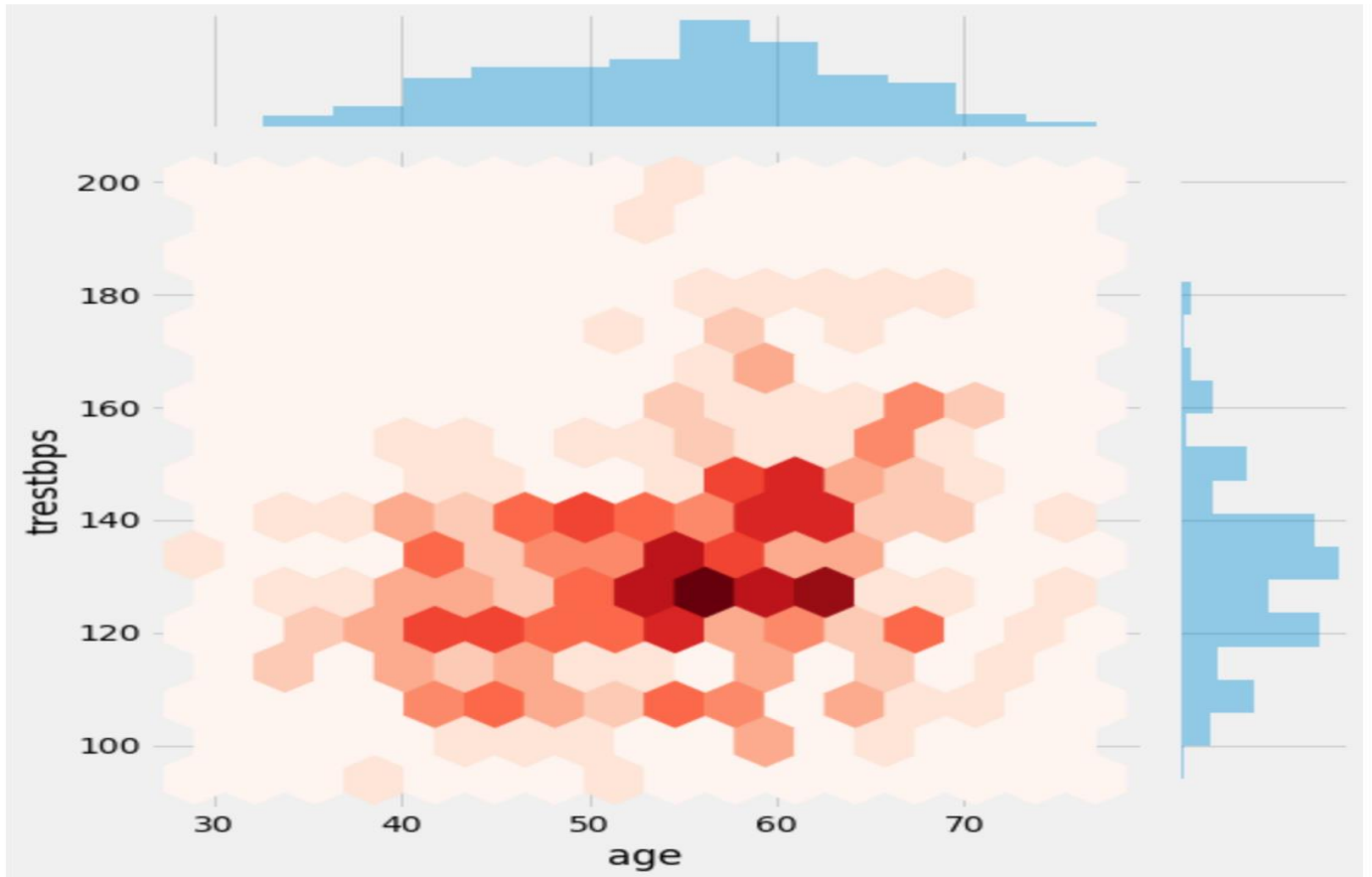
b) Trục quan nhiều thuộc tính :



Biểu đồ này cho thấy phân bố về lượng tuổi dựa trên giới tính của những người tham gia khảo sát.

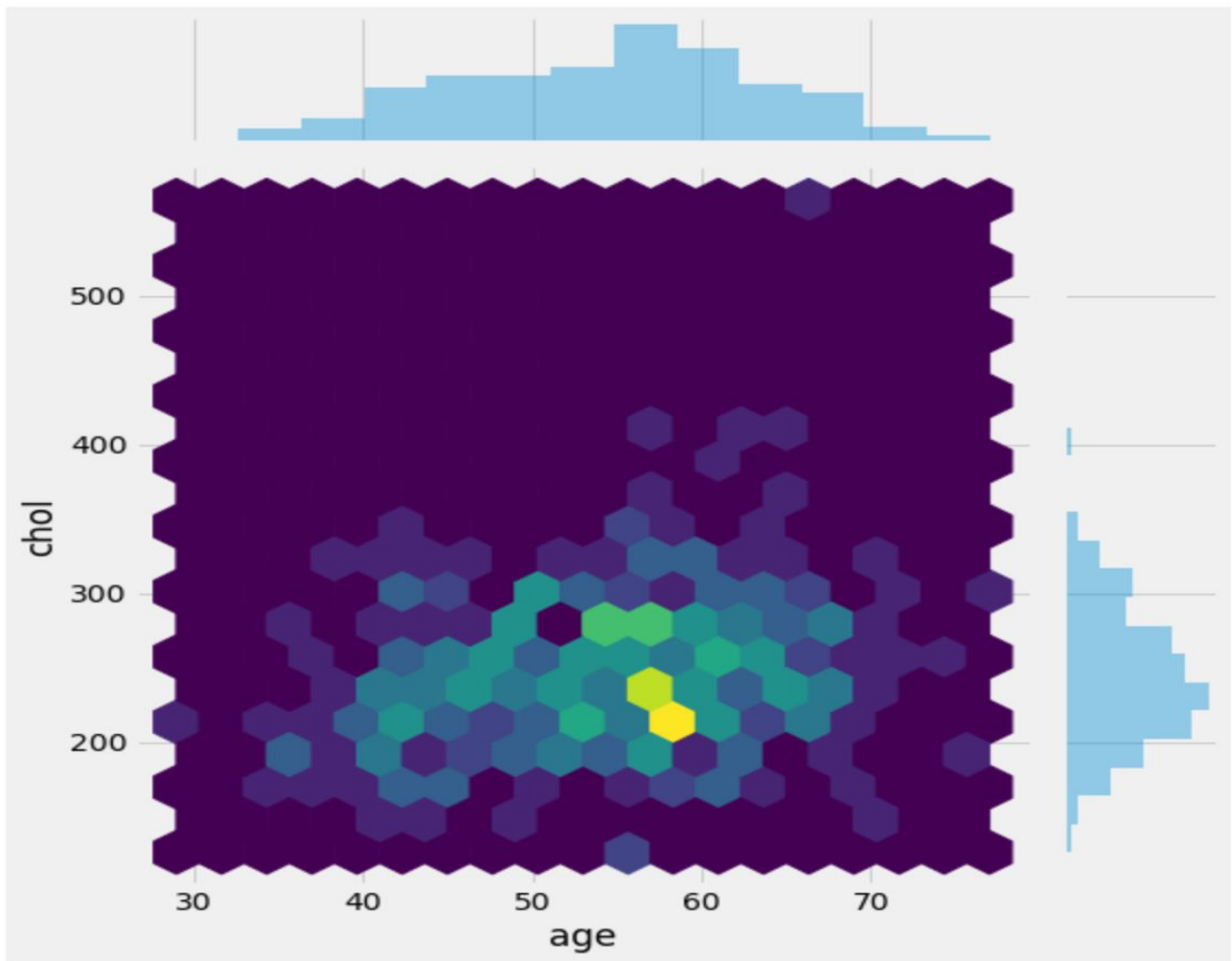
Ta nhận thấy rằng ở nam sẽ tập trung ở khoảng tuổi từ 50-65 là khoảng tuổi dễ gây bệnh tim .

Trong khi đó , đối với nữ , mặc dù nữ chỉ chiếm 30% tập khảo sát nhưng không có nghĩa là nữ sẽ ít khả năng bệnh tim hơn nam vì một báo cáo của các nhà khoa học đưa ra rằng phụ nữ sau tuổi mãn kinh sẽ có khả năng bị bệnh tim tương đương nam. Vì thế đối với nữ giới , họ sẽ tập trung vào khoảng tuổi 50-60 (tuổi sau mãn kinh).



Đồ thị này cho ta thấy mối liên hệ giữa tuổi và huyết áp

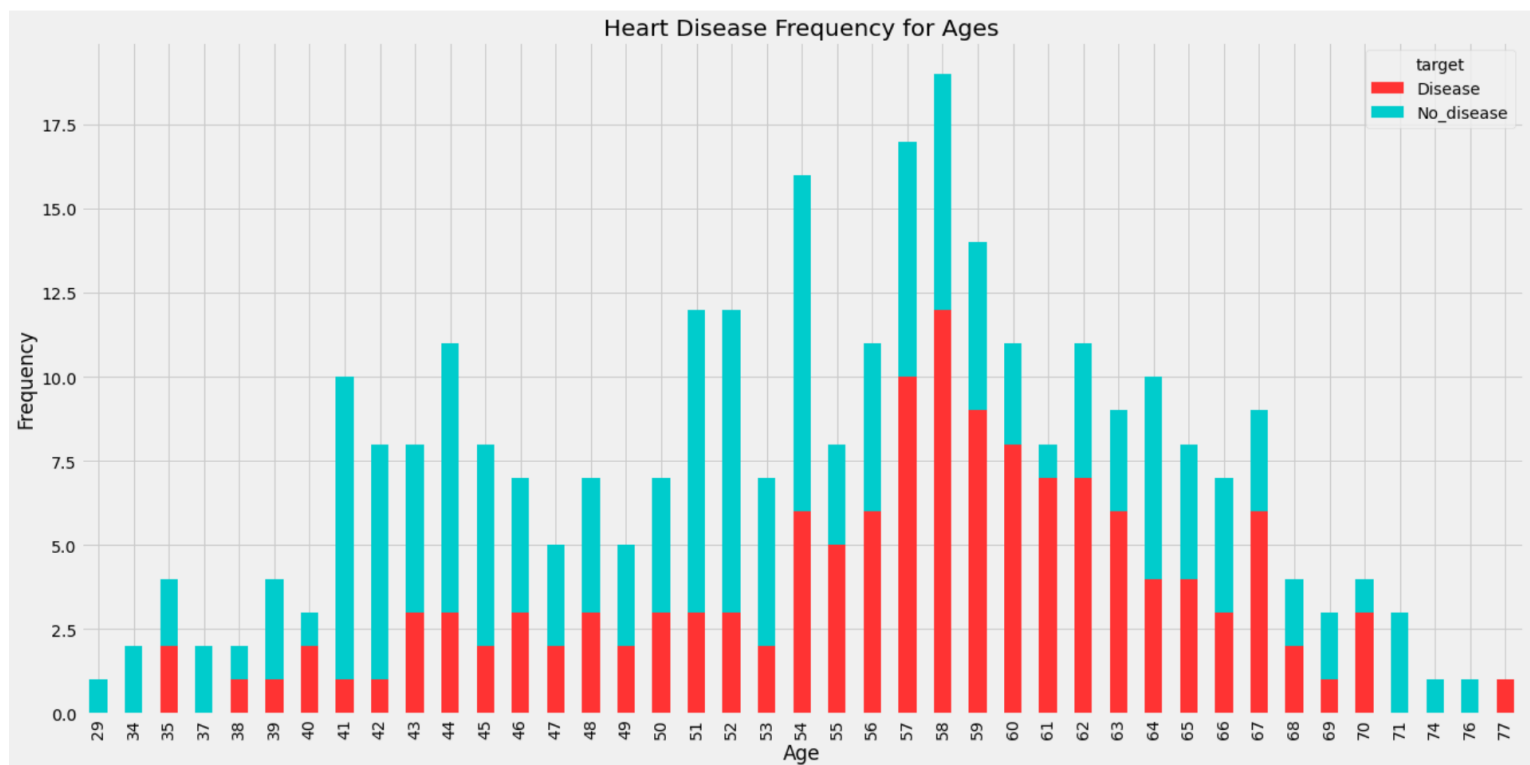
Như đã thấy trên đồ thị, các điểm càng đậm chỉ ra rằng huyết áp và tuổi sẽ xảy ra ở điểm đó nhiều nhất. Như khoảng tuổi từ 50 - 60 , ta sẽ thấy rằng có khá nhiều người ở độ tuổi này có huyết áp trong khoảng 120-140 tức là huyết áp cao. Lí do là vì Người lớn tuổi ,Hệ thống thành mạch máu không còn duy trì được độ đàn hồi như trước, dẫn đến cao huyết áp



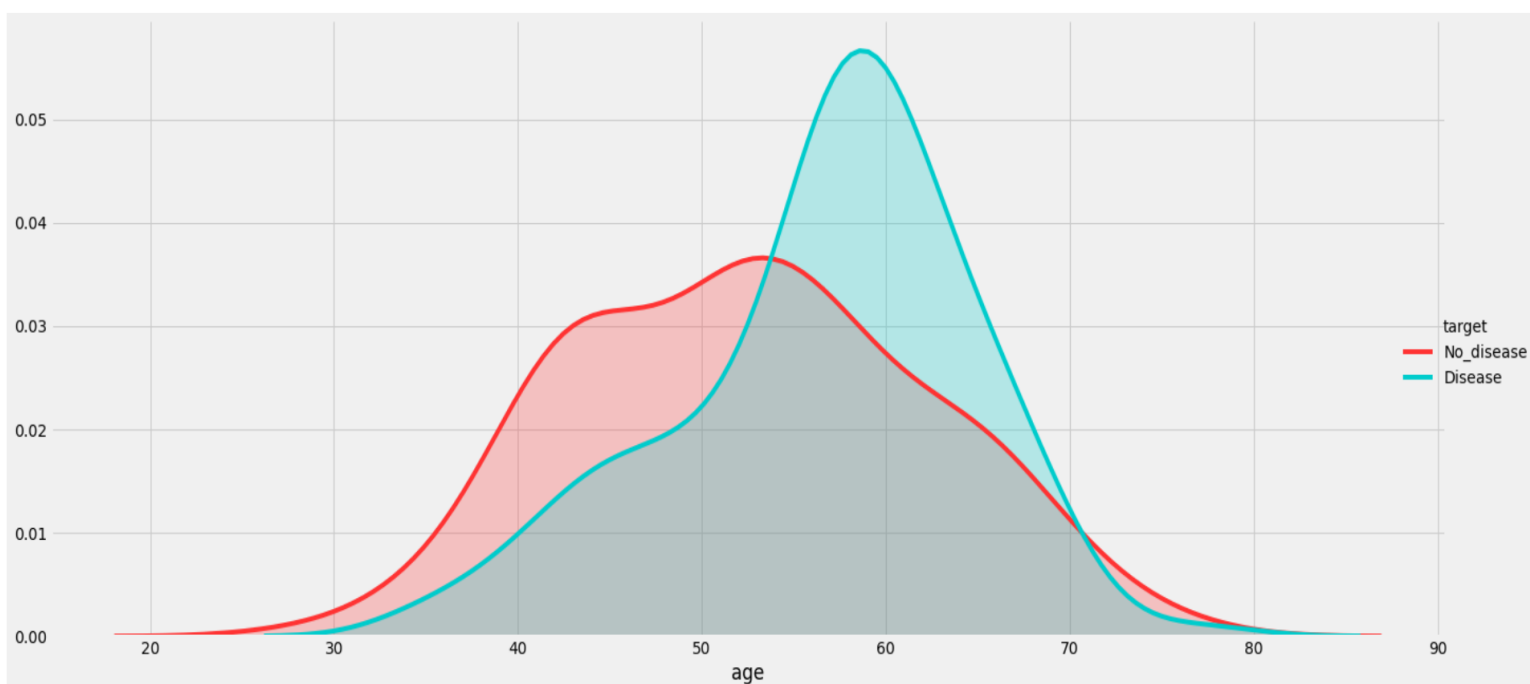
Đồ thị này cho ta thấy mối liên hệ giữa tuổi và Cholesterol

Tương tự như huyết áp thì lượng cholesterol cũng tăng theo độ tuổi. Ta ghi nhận bệnh nhân trong khoảng 50-60 tuổi đa số có cholesterol là từ 240 mg/dL trở đi tức là cholesterol cao

Vì Người càng lớn tuổi thì nguy cơ mỡ máu cao càng tăng, khả năng xơ vữa động mạch và biến chứng cũng cao hơn.



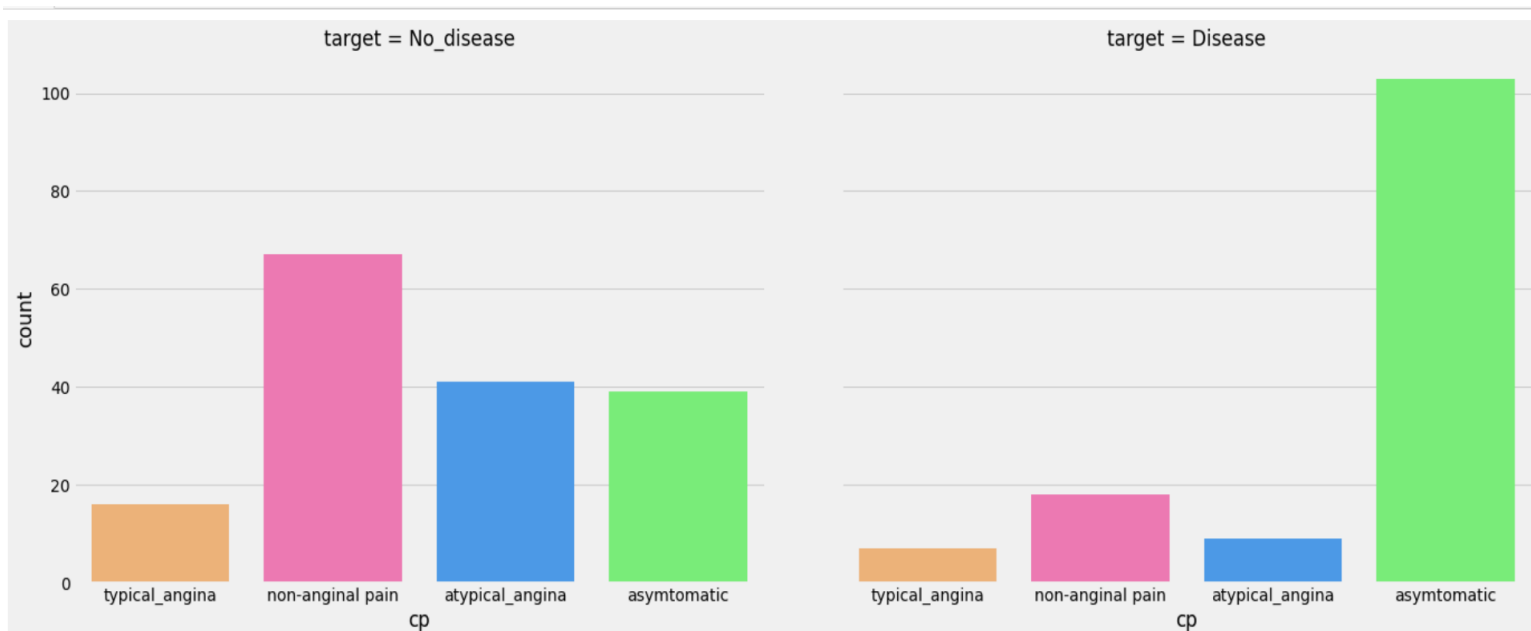
Biểu đồ này cho thấy mức độ phân bố của bệnh tim trên từng loại tuổi. Cột nào có màu xanh nhiều hơn sẽ cho thấy ở độ tuổi đó có ít bệnh nhân có khả năng bệnh tim hơn. Cột nào có màu đỏ chiếm nhiều hơn sẽ cho thấy bệnh nhân đó mắc bệnh tim nhiều hơn



Biểu đồ này sẽ cho thấy được rõ hơn mức độ phân bố của người bệnh tim so với nhóm tuổi.

Người bệnh tim sẽ tập trung ở nhóm tuổi từ 55-60 nhiều nhất

Trong khi đó những người không có bệnh tim sẽ phân bố nhiều ở khoảng tuổi trẻ hơn.



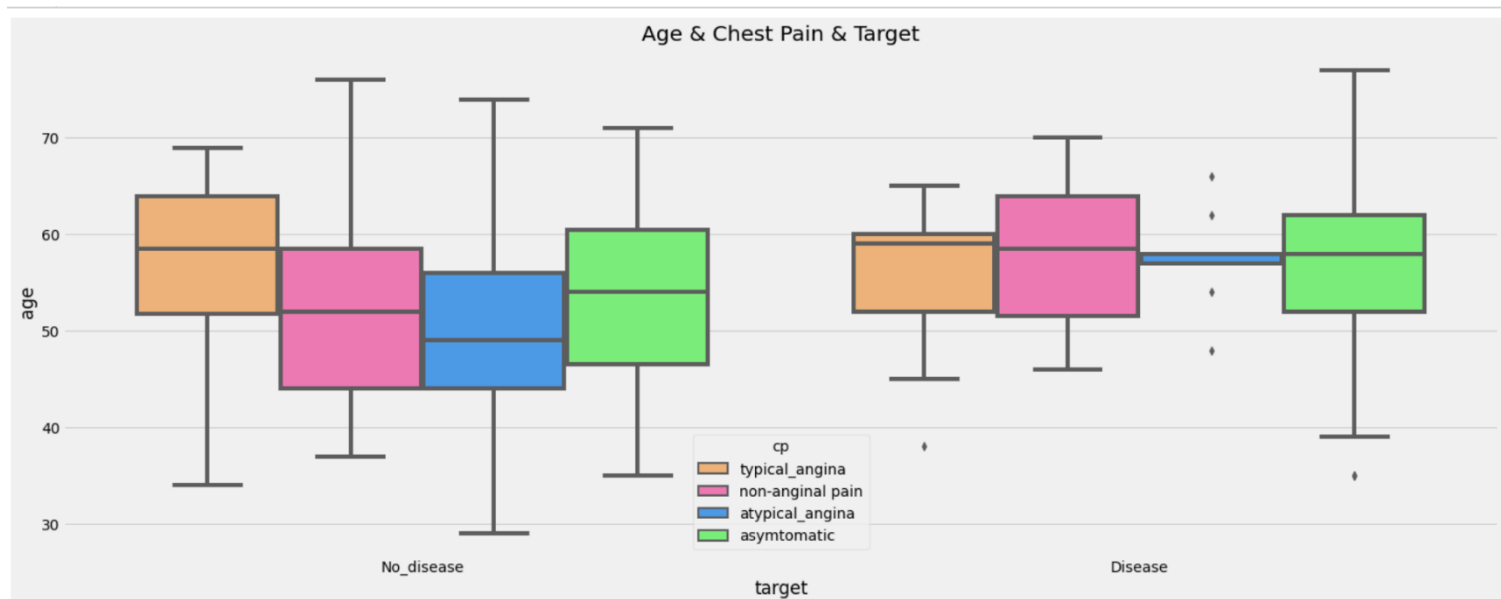
Đồ thị này cho ta thấy được 2 điểm đặc biệt là :

Trong 2 loại đau ngực phổ biến nhất trong tập dữ liệu này : đau ngực không triệu chứng và không đau ngực

Thì đau ngực không triệu chứng lại ghi nhận số bệnh nhân bị bệnh tim với trên 100 bệnh nhân chiếm cao nhất so với 3 loại đau ngực còn lại.

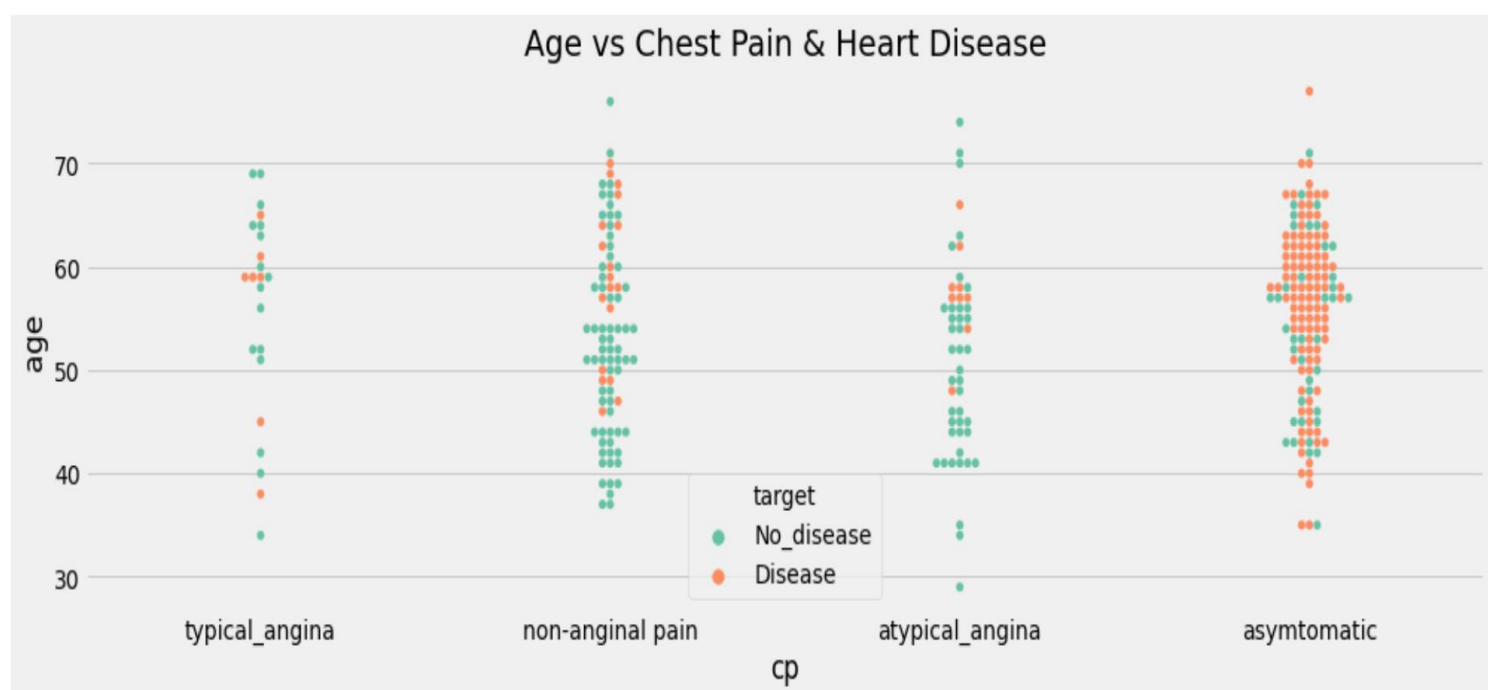
Trong khi đó người không đau ngực lại ghi nhận số bệnh nhân không mắc bệnh nhiều nhất .

Đồng thời 2 loại đau ngực còn lại là đau ngực không điển hình và điển hình đều có số người không bệnh ít hơn người bệnh .



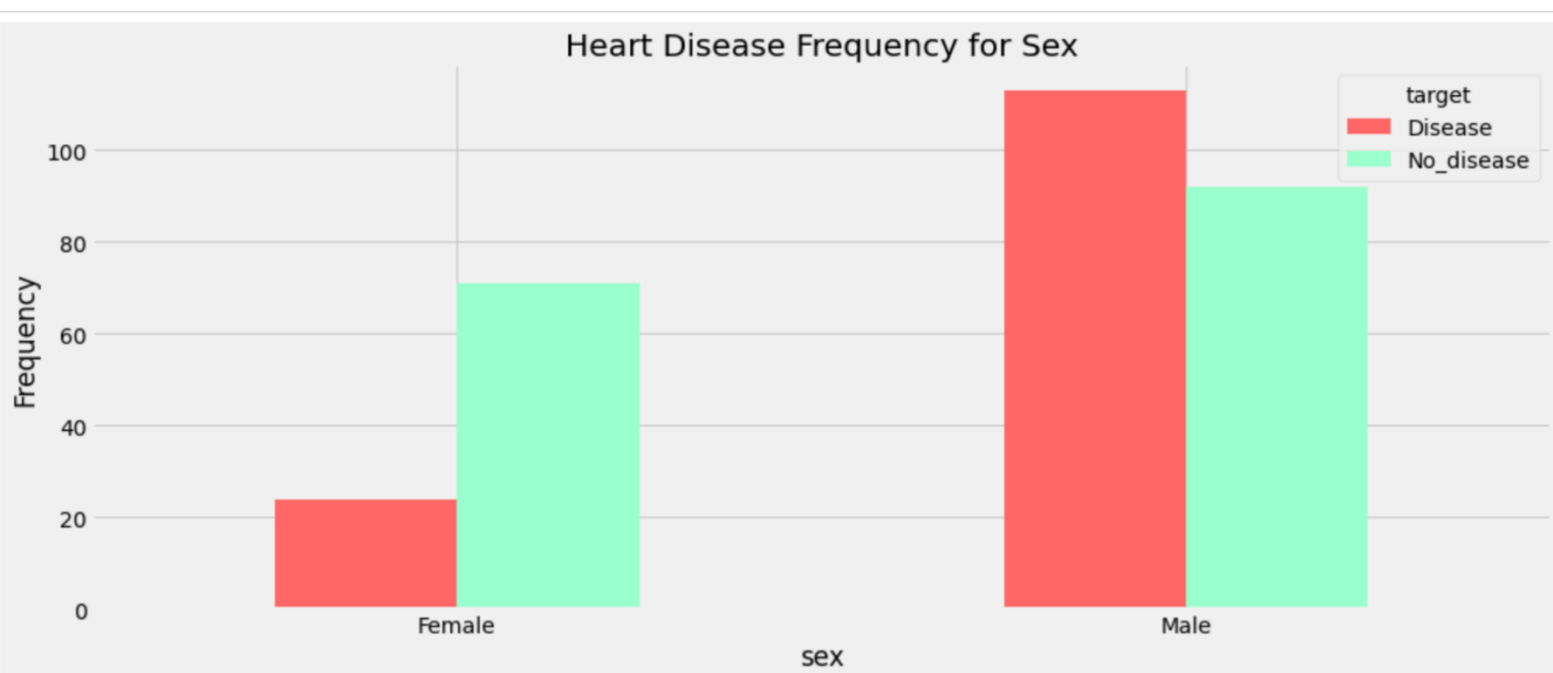
Độ tuổi và đau ngực ảnh hưởng tới bệnh tim như thế nào ?

Ở biểu đồ này ta thấy đa số các loại đau ngực đều được ghi nhận bệnh tim khi số tuổi của họ cao hơn . Đồng thời có 1 điều đặc biệt ở đây là những người đau thắt ngực không điển hình lại tập trung lượng lớn ở độ tuổi từ 55-57.



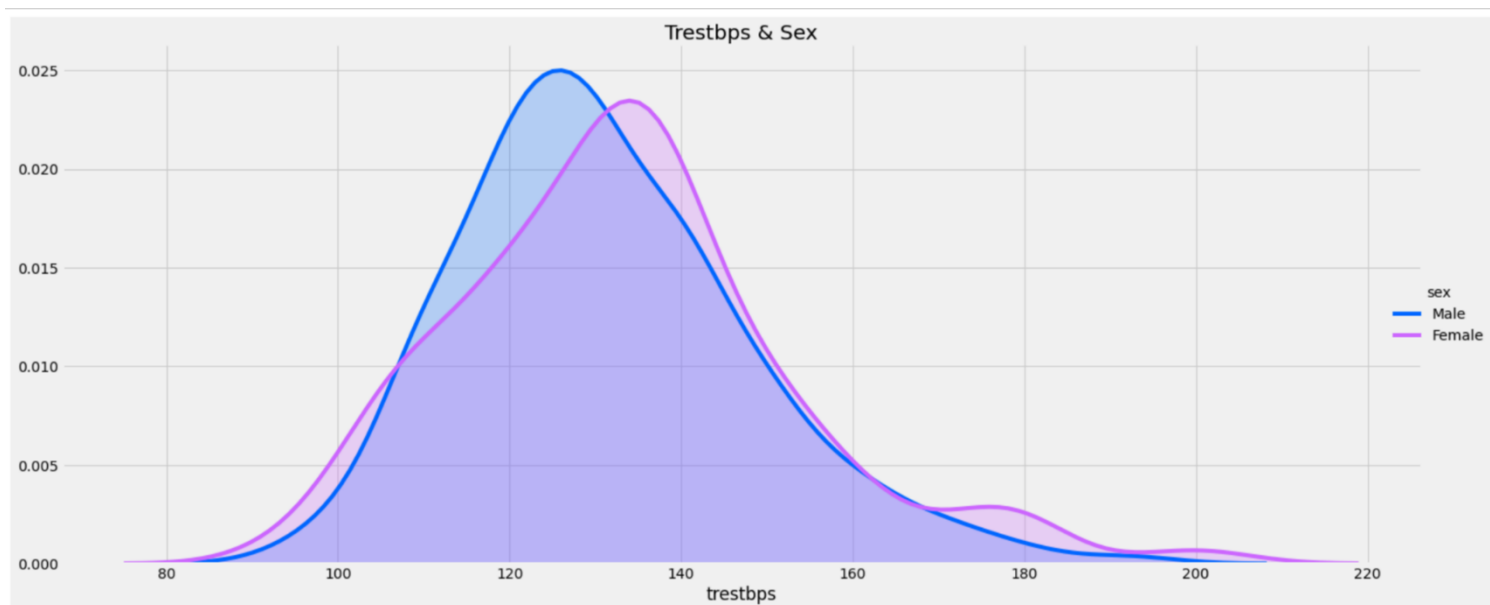
Ở đồ thị này , ta thấy được rằng đa số các loại đau ngực đều ghi nhận các bệnh nhân bị bệnh tim khi tuổi đã lớn. Nhưng đối với đau ngực không triệu chứng thì ngoài ghi nhận người lớn tuổi bị bệnh tim , ta còn thấy rằng bệnh tim có thể xảy ra ở người già đối với loại đau ngực này .

Đau tức ngực là tình trạng rất thường gặp đối với tất cả mọi người. Không chỉ phổ biến ở người cao tuổi, đau tức ngực còn có xu hướng trẻ hóa dần và đi kèm với nhiều dấu hiệu khác nhau như: tức ngực khó thở, đau ngực buồn nôn, đau thắt ngực bên trái, bên phải, ở giữa kèm đau nhói ở tim.

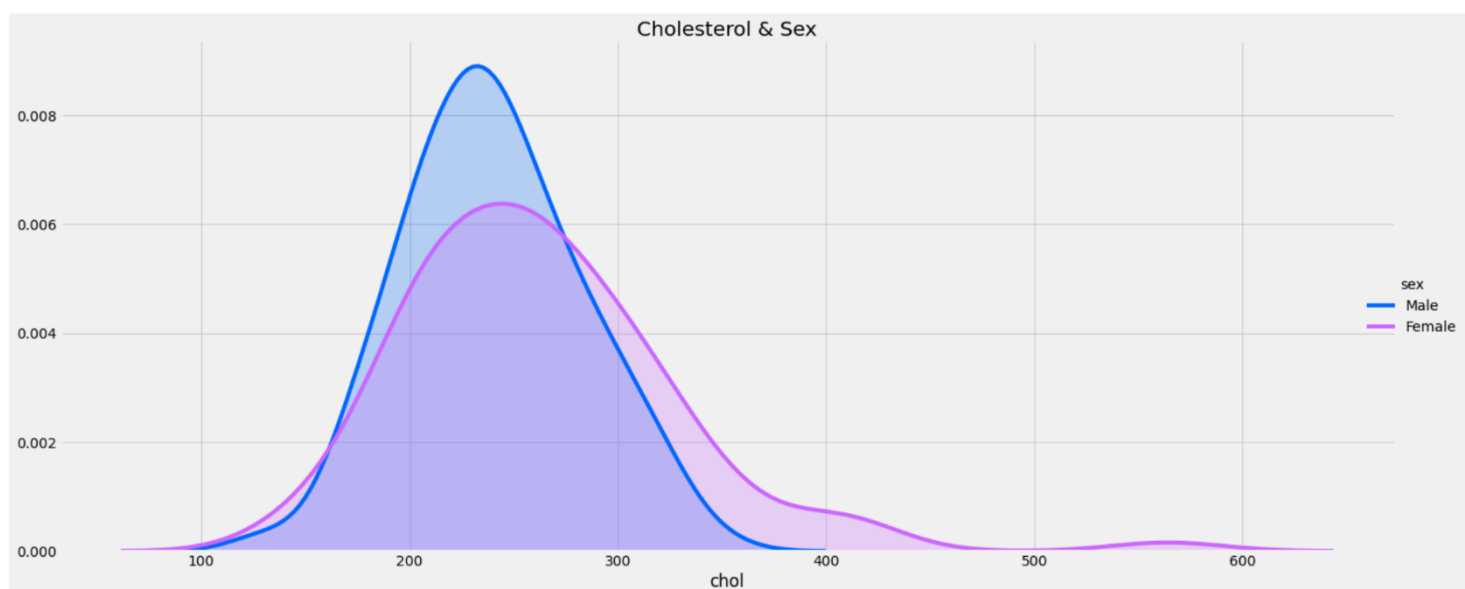


Đồ thị này cho ta thấy được rằng khả năng bệnh tim của nam luôn cao hơn nữ.

Điều này là dễ hiểu vì vào thời điểm đó (năm 1988) vẫn còn nạn kì thị giới tính , tức là Nữ giới sẽ không thể đảm nhiệm chức vụ quan trọng trong xã hội . Và chỉ Nam giới mới làm việc đó vì vậy họ bị chịu rất nhiều căng thẳng, những áp lực trong công việc, trách nhiệm gia đình ở mức độ rất cao. Bên cạnh đó, nam giới cũng ít khi bày tỏ cảm xúc so với nữ giới nên thường tích tụ lại, vấn đề này chính là lý do khiến nam giới dễ bị phát bệnh về tim mạch nhiều hơn so với phụ nữ.



Biểu đồ này cho ta thấy được lượng huyết áp của nam tập trung ở khoảng từ 120-140 là nhiều nhất, trong khi đối với nữ thì huyết áp của họ có phần nhỉnh hơn khi tập trung ở khoảng từ 130 trở đi .

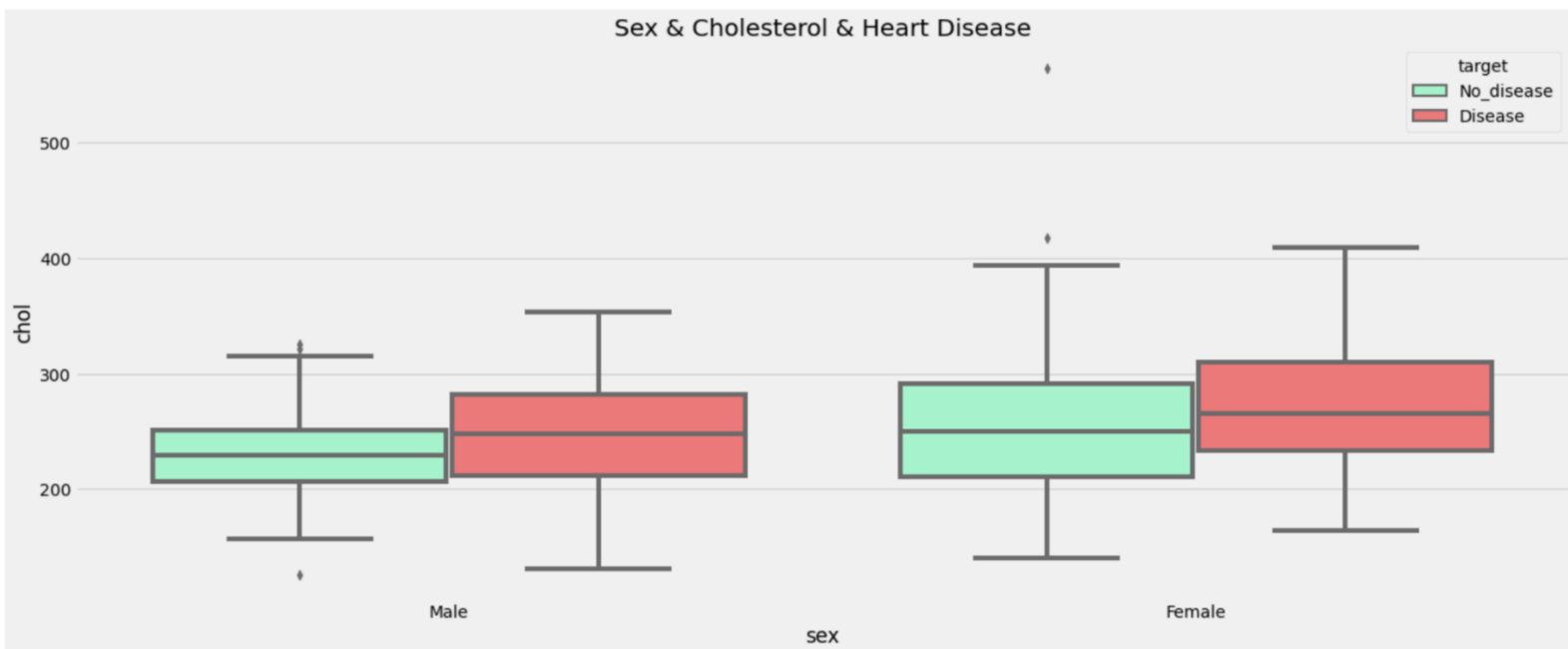


Còn đối với lượng Cholesterol thì ta thấy được rằng nam nằm ở khoảng 200-300 là nhiều nhất trong khi đối với nữ thì họ tập trung trải dài từ 200-300 mmHg

Sau khi tìm thấy các mối liên hệ giữa giới tính và lượng cholesterol với huyết áp, Ta sẽ tìm hiểu xem chúng ảnh hưởng như thế nào đối với kết quả chuẩn đoán bệnh tim

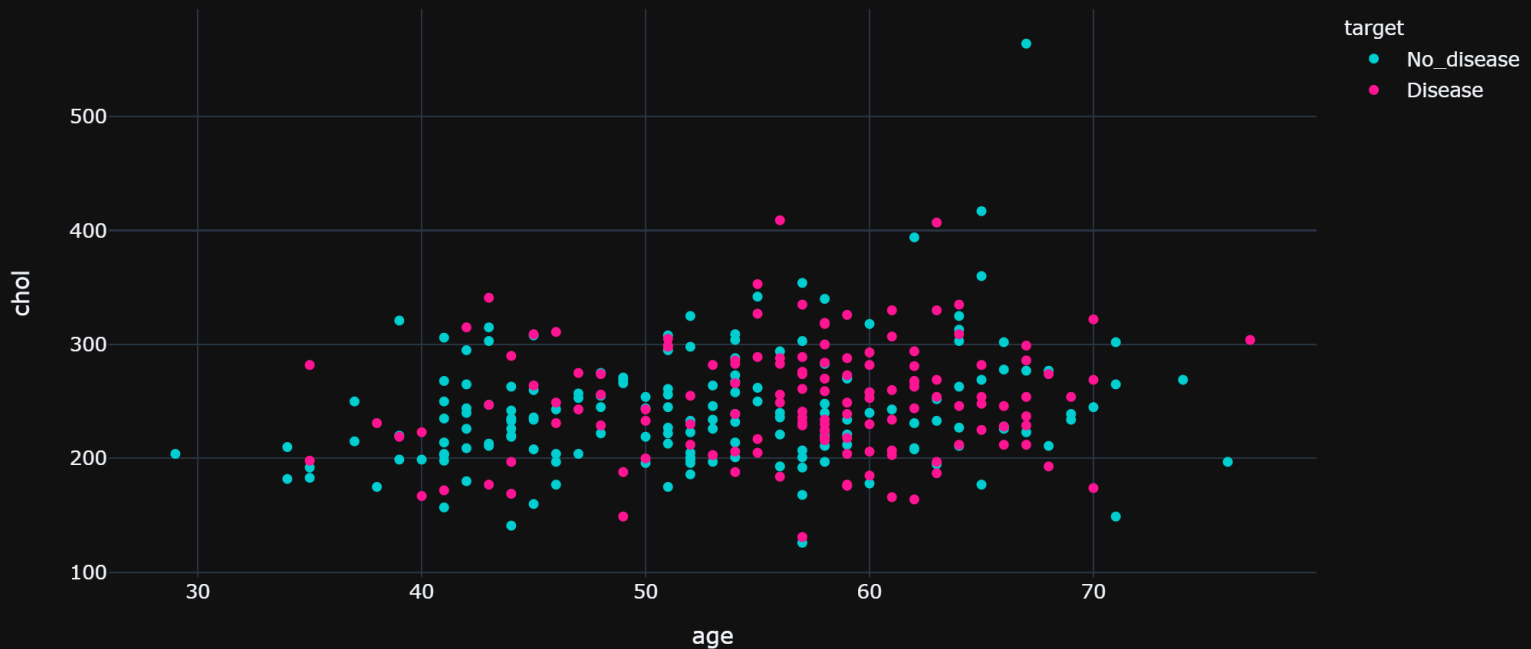


Đồ thị này cho thấy ở cả 2 giới tính, thì lượng huyết áp cao hơn sẽ ghi nhận ở người bị bệnh tim.



Tương tự với lượng huyết áp, lượng cholesterol cũng ghi nhận ở tình trạng tương tự.

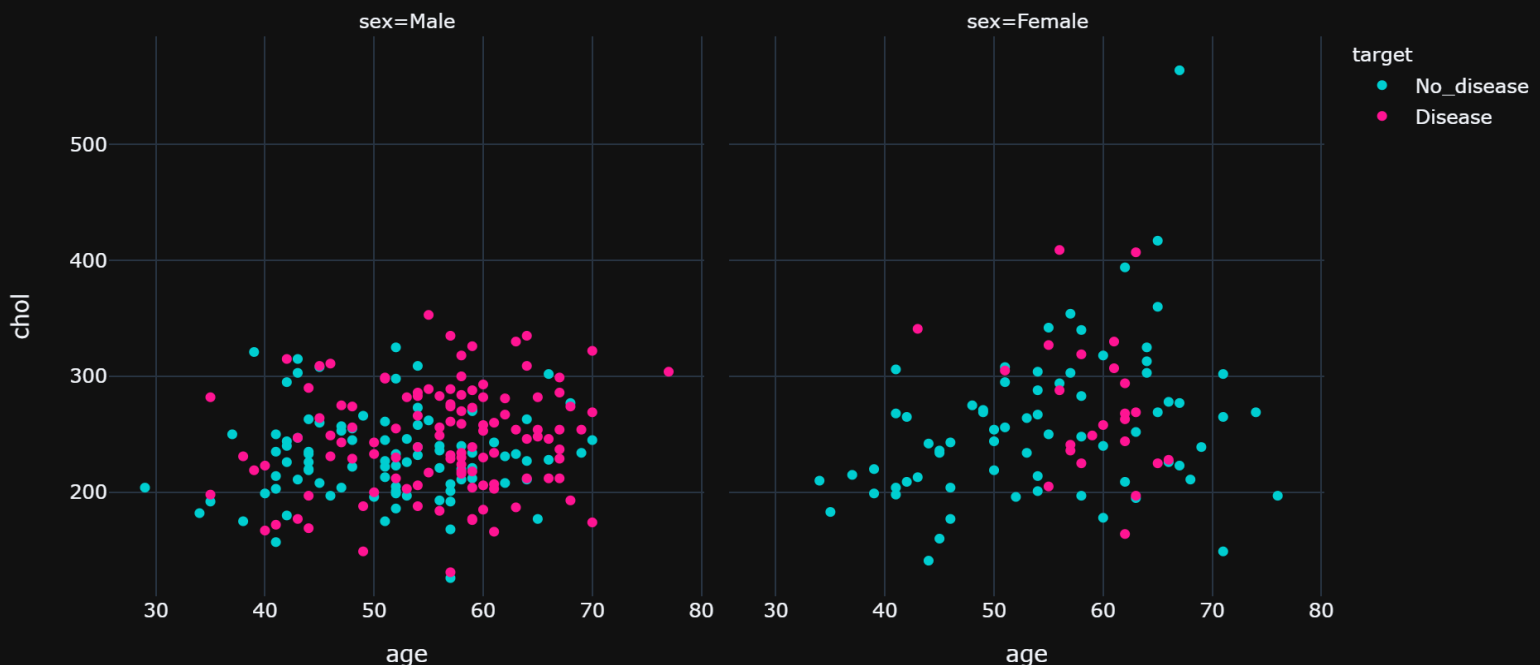
Cholesterol vs Age vs Heart Disease



Từ các đồ thị bên trên , ta biết được rằng độ tuổi càng cao sẽ dẫn đến lượng cholesterol càng cao. Nhưng lượng cholesterol cao có ảnh hưởng tới bệnh tim không ?

Câu trả lời là có . Như ta đã thấy bên hình trên , các chấm đỏ chỉ những người bị bệnh tim và chấm xanh ngược lại . Khi đó ta thấy được các chấm đỏ tập trung đa số ở những độ tuổi cao và lượng chol của vị trí đó cũng ghi nhận ở mức lớn (≥ 200).

Cholesterol vs Age vs Heart Disease vs Sex



Một cách rõ ràng hơn, khi ta phân biệt ra giới tính của những bệnh nhân này ta thấy được rằng :

Đối với Nam lượng cholesterol của họ tập trung nhiều ở khoảng 200-300 nhưng những người bệnh tim thường có lượng cholesterol nhỉnh hơn. Bên cạnh đó , bệnh tim cũng xuất hiện ở những độ tuổi khá trẻ ở Nam. Đối với nữ thì lượng cholesterol có phần thấp hơn hoặc bằng nam nhưng từ khoảng tuổi 50 trở đi lượng cholesterol của họ bắt đầu tăng cao.

Điều này xảy ra là do :

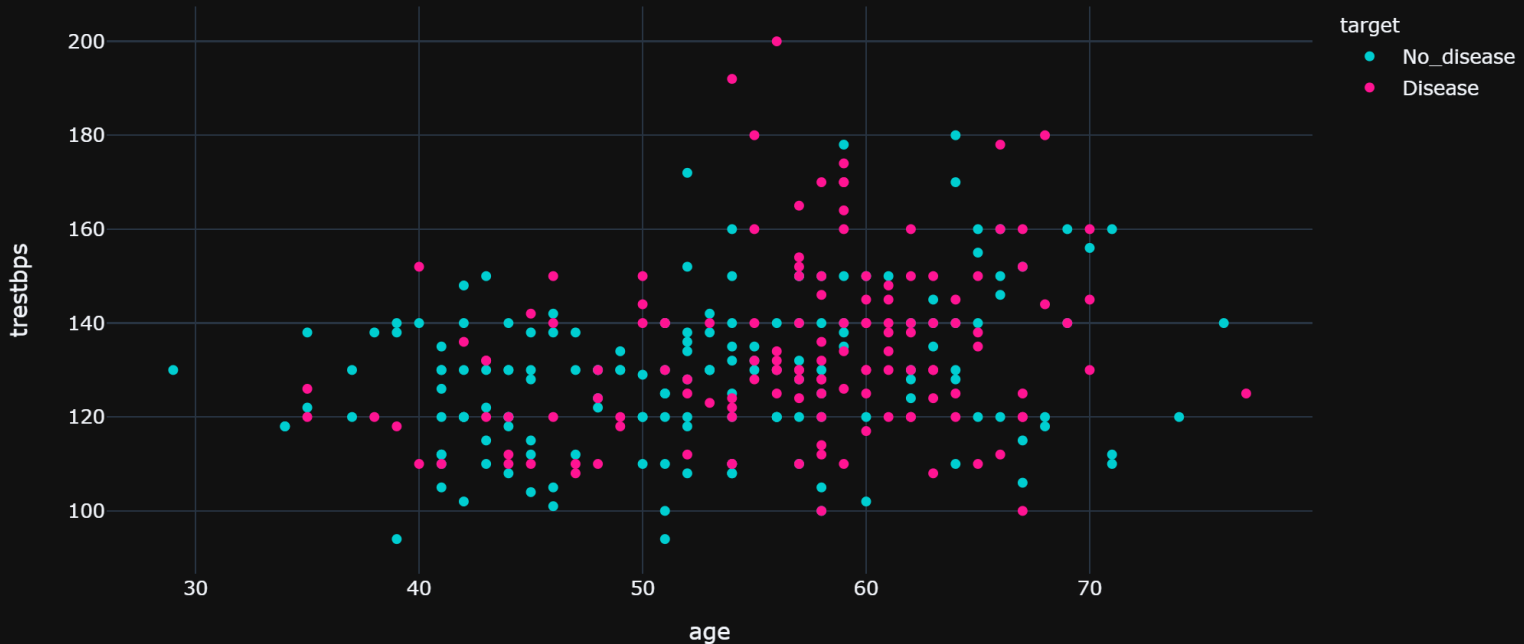
Etrogen ở phụ nữ giúp làm giảm lượng cholesterol có hại và làm tăng cholesterol có lợi, từ đó làm giảm nguy cơ tim mạch. Ngược lại, testosterone ở nam giới lại làm tăng lượng cholesterol có hại trong máu và giảm lượng cholesterol có lợi, khiến cho nam giới có lượng cholesterol cao hơn ở khoảng tuổi 50.

Khi phụ nữ mãn kinh (sau tuổi 50) cholesterol và triglycerides của họ có xu hướng tăng lên.

Nhìn vào đồ thị ta không thể nói rằng lượng cholesterol cao gây ra bệnh tim , nhưng lại là tác nhân trực tiếp gây ra bệnh tim :

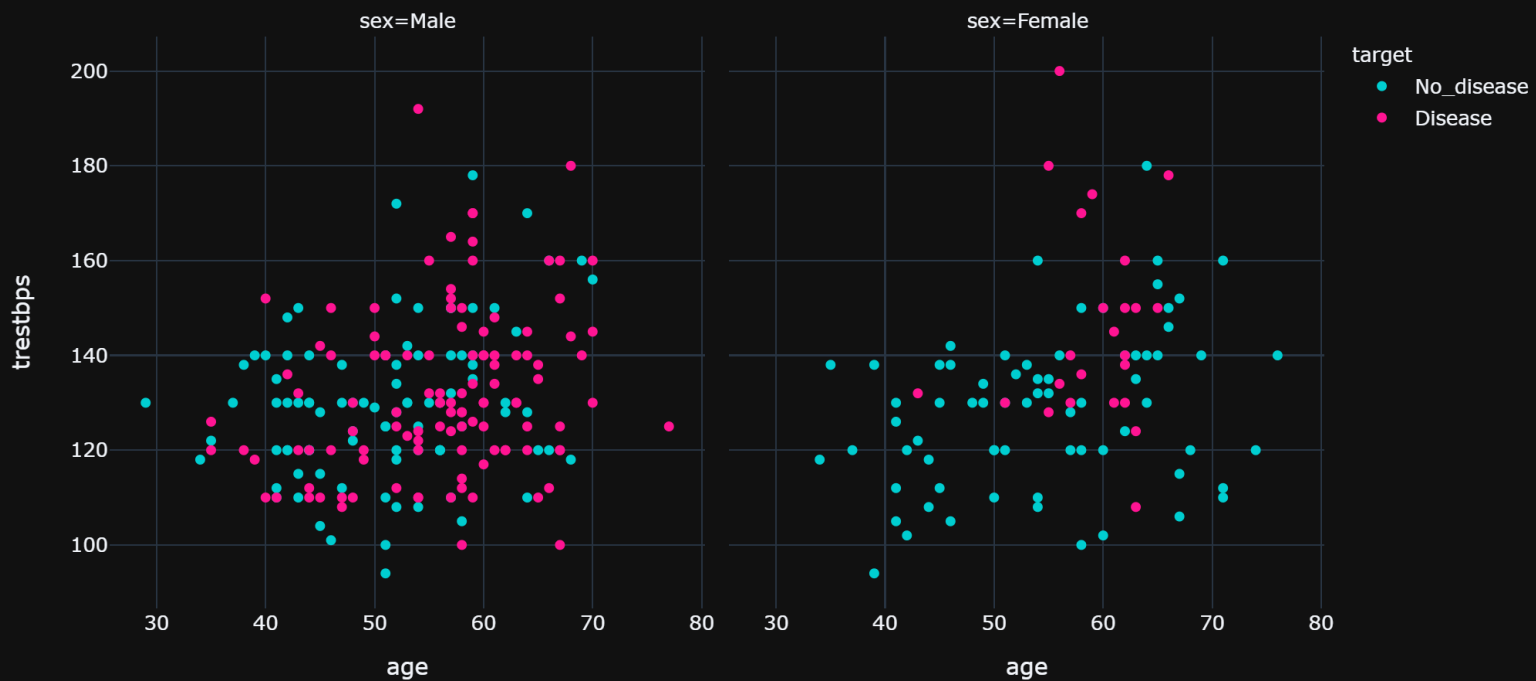
Khi chỉ số cholesterol trong máu cao hơn bình thường, ta có thể bị lắng đọng chất béo trong các mạch máu. Các chất lắng đọng này có thể làm cản trở sự vận chuyển của máu qua động mạch. Do đó, tim không nhận đủ máu giàu oxy và bạn có thể bị [đau tim](#).

Trestbps vs Age vs Heart Disease



Tương tự với cholesterol, huyết áp cùng độ tuổi cao cũng chính là nguyên nhân gây ra bệnh tim. Như đồ thị trên , các chấm đỏ tập trung nhiều nhất ở khoảng tuổi 50-60 và lượng cholesterol từ 120 trở đi .

Trestbps vs Age vs Sex vs Heart Disease



Tiếp tục phân tích rõ hơn khi tách nhóm trên ra thành 2 giới tính :

Ta thấy rằng ở nam bên cạnh những người có độ tuổi cao cùng với lượng huyết áp lớn dễ mắc bệnh hơn thì ta thấy rằng ở nhóm tuổi trẻ hơn (≤ 50) cũng có khả năng mắc bệnh cao. Còn đối với nữ thì những người trẻ tuổi sẽ có khả năng bị ít hơn nhưng đối với người già họ còn mắc bệnh với lượng huyết áp cao hơn cả nam.

Việc thường xuyên phải đối mặt với căng thẳng công việc, chế độ ăn uống, sinh hoạt thiếu lành mạnh là nguyên nhân khiến đàn ông dưới 50 tuổi dễ mắc huyết áp cao hơn phụ nữ. Bạn có thể dễ dàng nhận thấy, đàn ông thường xuyên tiếp xúc với bia rượu và các chất kích thích như cà phê nhiều hơn phụ nữ. Vì vậy nam giới dưới 50 tuổi dễ mắc các bệnh như huyết áp cao, tim mạch, tiểu đường.

Trong giai đoạn từ 20- 40 tuổi, phụ nữ được bảo vệ bởi estrogen, sau 55 tuổi cũng là độ tuổi bắt đầu mãn kinh hoặc sau mãn kinh ở phụ nữ. Tại thời điểm này, các nội tiết tố của phụ nữ thay đổi rất lớn nên đây là nguyên nhân lớn dẫn đến huyết áp cao ở phụ nữ.

4 GIẢI PHÁP :

Nhận biết được độ nguy hiểm của bệnh tim và các tác nhân trực tiếp gây ra nó. Ta có thể làm những việc đơn giản sau đây để giảm nguy cơ mắc bệnh tim

1. Biết rõ tình trạng sức khỏe của mình :

Biết rõ tình trạng sức khỏe của mình sẽ giúp bạn khỏe mạnh hơn. Hãy tới cơ sở y tế để kiểm tra lượng cholesterol, lượng đường trong máu và các dấu hiệu của bệnh tiểu đường, bệnh tăng huyết áp để biết chắc chắn về tình trạng sức khỏe hiện tại của mình.



2. Chăm tập luyện thể dục

Tập thể dục có thể làm tăng cholesterol "tốt", và giảm cholesterol "xấu".



3. Hạn chế uống rượu, bia.

Vì uống nhiều rượu, bia làm huyết áp tăng và làm tăng trọng lượng cơ thể. Khi trọng lượng cơ thể tăng lại là yếu tố nguy cơ gây tăng huyết áp.



-----HẾT-----