Họ tên: Nguyễn Hoàng Hải Mã sinh viên: B21DCCN319

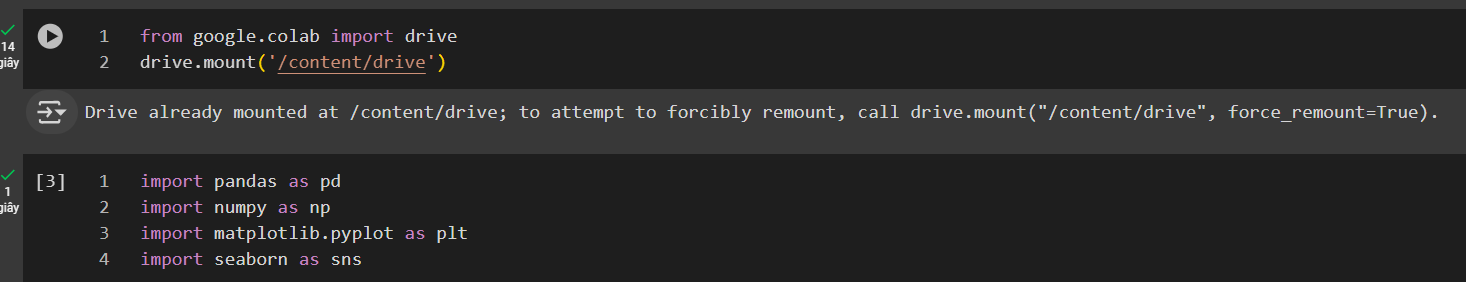
**BÀI TẬP TRỰC QUAN HOÁ DỮ LIỆU**

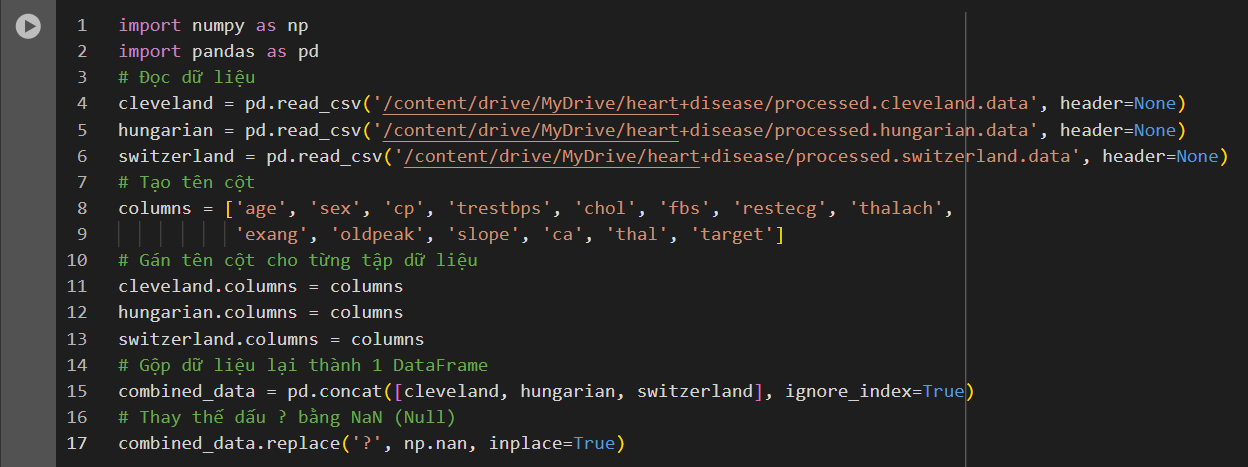
1. **Mô tả dữ liệu:** Gồm các thuộc tính sau:

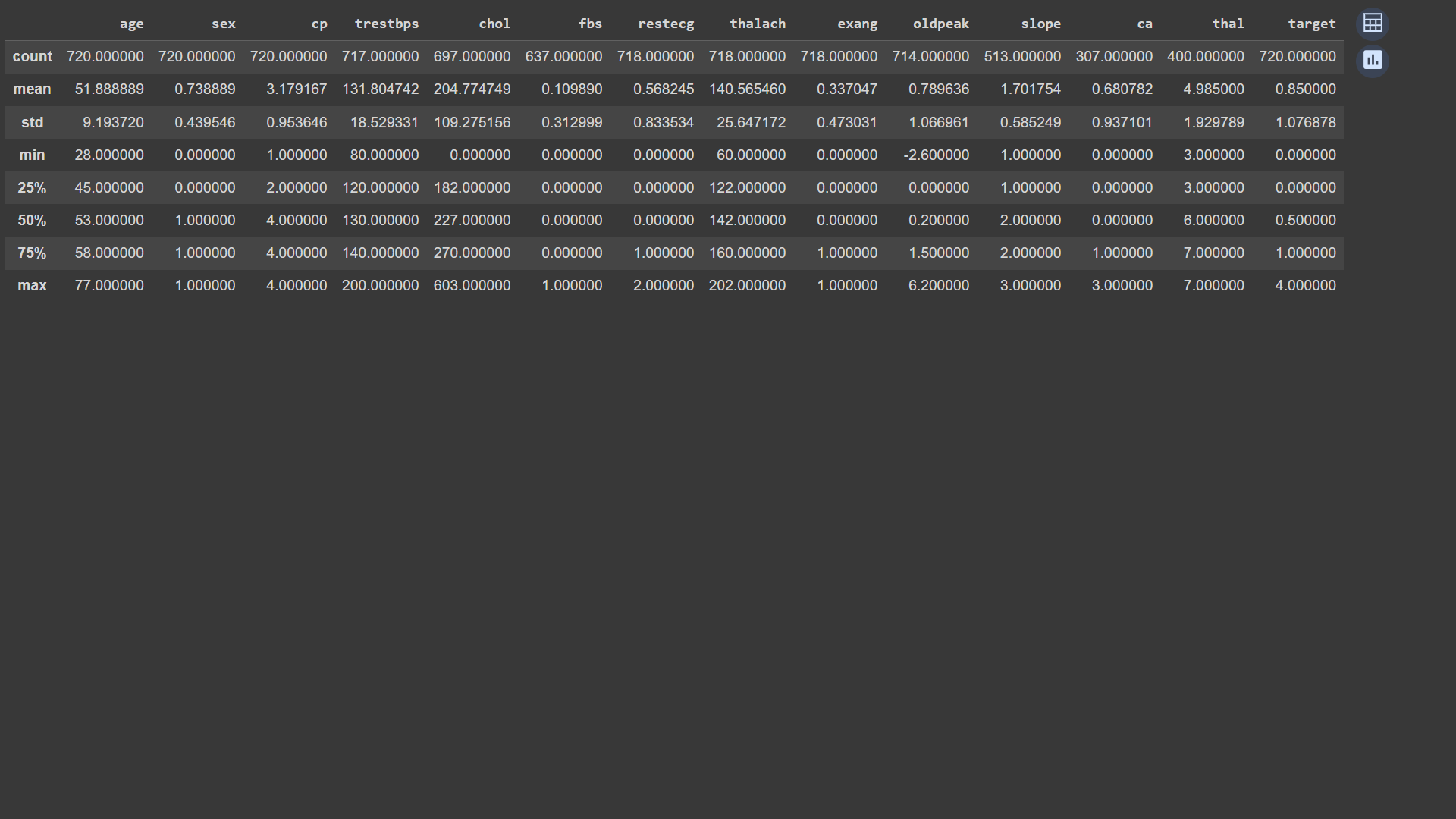
* Tuổi
* Giới tính: Nam – 1, Nữ: 0
* Loại đau ngực: (1 - Đau thắt ngực điển hình, 2: Đau thắt ngực không điển hình, 3: Đau không do ngực, 4: Không triệu chứng
* Huyết áp khi nghỉ (trestbps)
* Cholesterol
* Đường huyết khi đói (< 120, với giá trị là True hoặc False): Cho biết đường huyết khi đói của bệnh nhân có dưới 120mmHg không
* Điện tâm đồ khi nghỉ: 3 trạng thái 0, 1, 2 tương ứng là bình thường, bất thường và tăng huyết áp
* Nhịp tim tối đa
* Đau thắt ngực do vận động (True hoặc False)
* Oldpeak (số đo khác, không rõ mô tả chi tiết)
* Độ dốc của đoạn ST (1: lên, 2: đi ngang, 3: đi xuống)
* Thal: Tình trạng tim mạch khi xét nghiệm Thalium (3: Bình thường, 6: Tổn thương không hồi phục, 7: Tổn thương có hồi phục)
* **Nhãn phân loại**: Buff là khoả mạnh, disease là bị bệnh tim
* **Target**: 0 là khoẻ mạnh, 1 2 3 4 mức độ bệnh tăng dần

1. **Trực quan hoá dữ liệu và nhận xét**

* Gom nhóm dữ liệu 3 khu vực lại thành 1:

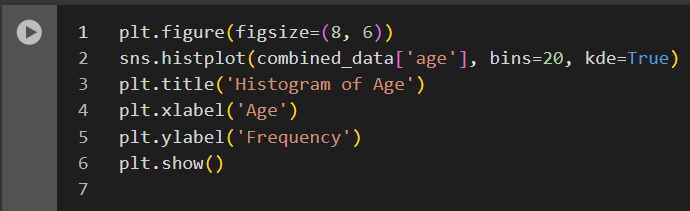


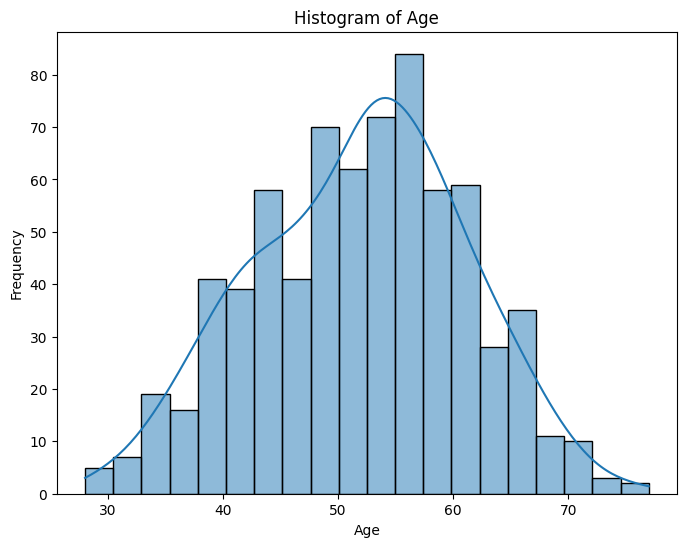




**Vẽ các biểu đồ và nhận xét**

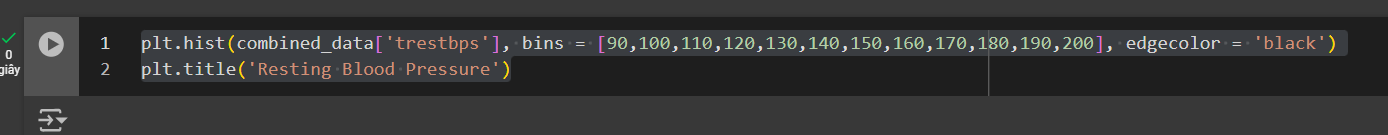
* **Biểu đồ theo tuổi:**

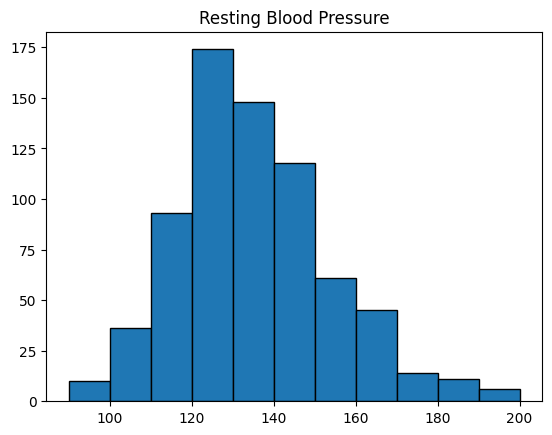
****



Dữ liệu tuổi có dạng phân phối gần với phân phối, tập trung chủ yếu ở nhóm tuổi từ 40 đến 70, với phần lớn người ở độ tuổi 50-60. Điều này có thể là do bệnh tim thường phổ biến hơn ở nhóm tuổi trung niên và cao tuổi, như được thấy qua sự phân bố này.

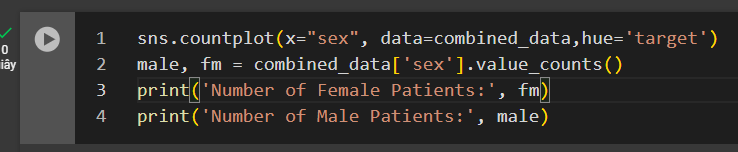
* Biểu đồ huyết áp khi nghỉ ngơi:



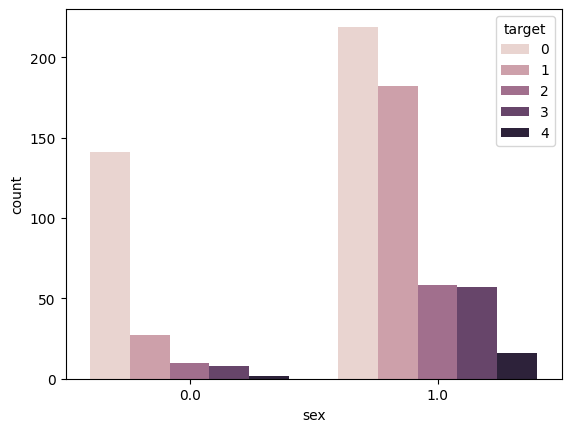


Huyết áp tâm thu có xu hướng phân bố không đối xứng, với phần lớn dữ liệu tập trung ở khoảng 120 đến 140 mmHg. Có một đỉnh cao nhất (mode) ở khoảng 120 mmHg, cho thấy rằng nhiều người có huyết áp ở mức này. Biểu đồ cho thấy tầm trải rộng của huyết áp từ khoảng 100 đến 200 mmHg, với một số ít trường hợp nằm ở các mức huyết áp cực thấp (dưới 110 mmHg) và cực cao (trên 180 mmHg).

* Biểu đồ thể hiện số lượng bệnh nhân theo mức độ và giới tính:

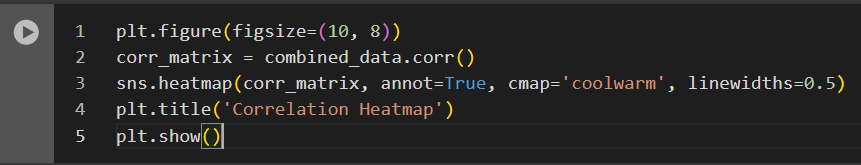


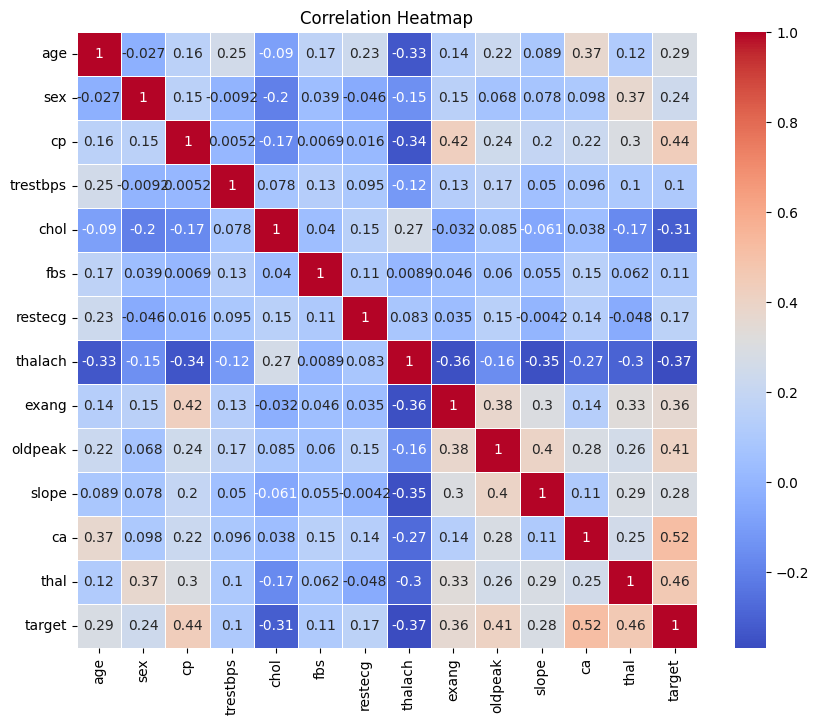
Number of Female Patients: 188; Number of Male Patients: 532



Tỷ lệ mắc bệnh tim ở nam giới cao hơn,chiếm đa số trong các mức độ bệnh nặng hơn (các mức 1, 2, 3, và 4). Ở nữ giới (giá trị 0 trên trục x), số người không mắc bệnh (mức 0) cao hơn rất nhiều so với những người mắc bệnh tim. Điều này cho thấy tỷ lệ nữ giới không mắc bệnh tim là rất lớn, và nữ giới mắc bệnh tim ở các mức nặng (mức 3 và 4) có số lượng khá nhỏ.

* Biểu đồ tương quan giữa các đặc trưng:



**Tương quan với target (mục tiêu về tình trạng bệnh)**:

* ca (số lượng mạch máu bị ảnh hưởng) có tương quan dương mạnh nhất với target (0.52), tức là khi số lượng mạch máu bị ảnh hưởng tăng, khả năng mắc bệnh tim tăng lên.
* cp (loại đau ngực) có tương quan dương 0.44, chỉ ra rằng loại đau ngực càng nghiêm trọng thì khả năng mắc bệnh tim càng cao.
* exang (đau ngực do gắng sức) có tương quan dương 0.36, nghĩa là bệnh nhân có dấu hiệu đau ngực do gắng sức có khả năng mắc bệnh tim cao hơn.
* Ngược lại, thalach (nhịp tim tối đa) có tương quan âm (-0.37) với target, chỉ ra rằng những người có nhịp tim tối đa thấp thường có khả năng mắc bệnh tim cao hơn.

**Tương quan giữa các thuộc tính**:

* cp và exang có tương quan dương (0.42), cho thấy mối liên hệ giữa loại đau ngực và khả năng đau ngực do gắng sức.
* oldpeak và exang cũng có tương quan dương (0.38), cho thấy khi ST depression tăng, khả năng đau ngực do gắng sức cũng tăng.
* age và thalach có tương quan âm (-0.33), có nghĩa là người lớn tuổi thường có nhịp tim tối đa thấp hơn.