Dựa trên các tên cột hiện có, chúng ta có thể phân loại như sau:

* **Các feature có mặt ở cả 3 bộ dữ liệu (mặc dù có cách đặt tên khác nhau):**
  + **Tuổi:** Có dưới dạng **age** (các file 1 và 3) và **Age** (file 2)
  + **Giới tính:** Có dưới dạng **sex** (các file 1 và 3) và **Sex** (file 2)
  + **Đau thắt ngực:** Có dưới dạng **cp** (file 1 và 3) và **ChestPainType** (file 2)
  + **Huyết áp lúc nghỉ:** **trestbps** (file 1 và 3) và **RestingBP** (file 2)
  + **Cholesterol:** **chol** (file 1 và 3) và **Cholesterol** (file 2)
  + **Đường huyết lúc đói:** **fbs** (file 1 và 3) và **FastingBS** (file 2)
  + **Điện tâm đồ lúc nghỉ:** **restecg** (file 1 và 3) và **RestingECG** (file 2)
  + **Chỉ số oldpeak:** xuất hiện nhất quán ở cả 3 file
  + **Độ dốc ST:** **slope** (file 1 và 3) và **ST\_Slope** (file 2)
* **Các feature chỉ xuất hiện ở 2 trong số 3 bộ dữ liệu:**  
  Do file 2 thường dùng tên khác cho các chỉ số có cùng ý nghĩa, một số feature có tên giống hệt chỉ có mặt ở file 1 và file 3 (và không có bản sao đúng tên đó trong file 2), điển hình như:
  + **cp**: Xuất hiện ở file 1 và file 3
  + **trestbps**: Xuất hiện ở file 1 và file 3
  + **chol**: Xuất hiện ở file 1 và file 3
  + **fbs**: Xuất hiện ở file 1 và file 3
  + **restecg**: Xuất hiện ở file 1 và file 3
  + **slope**: Xuất hiện ở file 1 và file 3
  + **exang**: Xuất hiện ở file 1 và file 3
  + **ca**: Xuất hiện ở file 1 và file 3
  + **thal**: Xuất hiện ở file 1 và file 3

Những biến này đều thể hiện cùng một ý nghĩa lâm sàng trong file 1 và file 3, trong khi file 2 lại sử dụng tên khác (ví dụ, sử dụng **ChestPainType** thay cho **cp**, **RestingBP** thay cho **trestbps**, v.v.) để biểu thị nội dung tương tự.

* **Các feature hoàn toàn độc nhất (chỉ có ở một bộ dữ liệu):**
  + Trong file 1 có: **id**, **dataset**, và **thalch**
  + Trong file 2 có: **MaxHR** và **ExerciseAngina**
  + Trong file 3 có: **target** (trong khi file 1 có **num** và file 2 có **HeartDisease** để biểu thị kết quả)

Tóm lại, có các feature như **cp, trestbps, chol, fbs, restecg, slope, exang, ca, thal** có cùng tên (và cách đánh dấu) chỉ xuất hiện ở 2 file (file 1 và file 3) vì file 2 thay đổi tên cho những chỉ số này.

 **Tuổi**: Age

 **Giới tính**: Sex

 **Đau thắt ngực**: ChestPainType

 **Huyết áp lúc nghỉ**: RestingBP

 **Cholesterol**: Cholesterol

 **Đường huyết lúc đói**: FastingBS

 **Điện tâm đồ lúc nghỉ**: RestingECG

 **Nhịp tim tối đa**: MaxHR

 **Đau thắt ngực do tập luyện**: ExerciseAngina

 **Chỉ số Oldpeak**: Oldpeak

 **Độ dốc ST**: ST\_Slope

 **Số lượng mạch máu chính bị hẹp**: CA

 **Kết quả kiểm tra Thalium**: Thal

 **Biến mục tiêu (kết quả chẩn đoán)**: Target