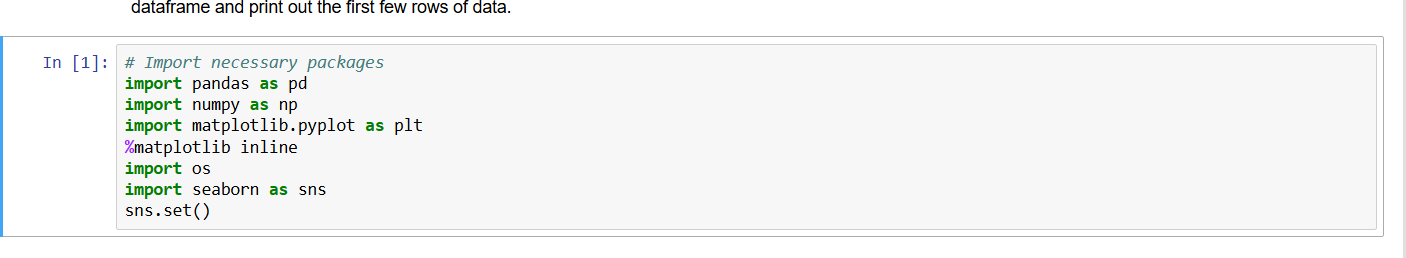
## **Họ và tên: Dương Xuân Bách**

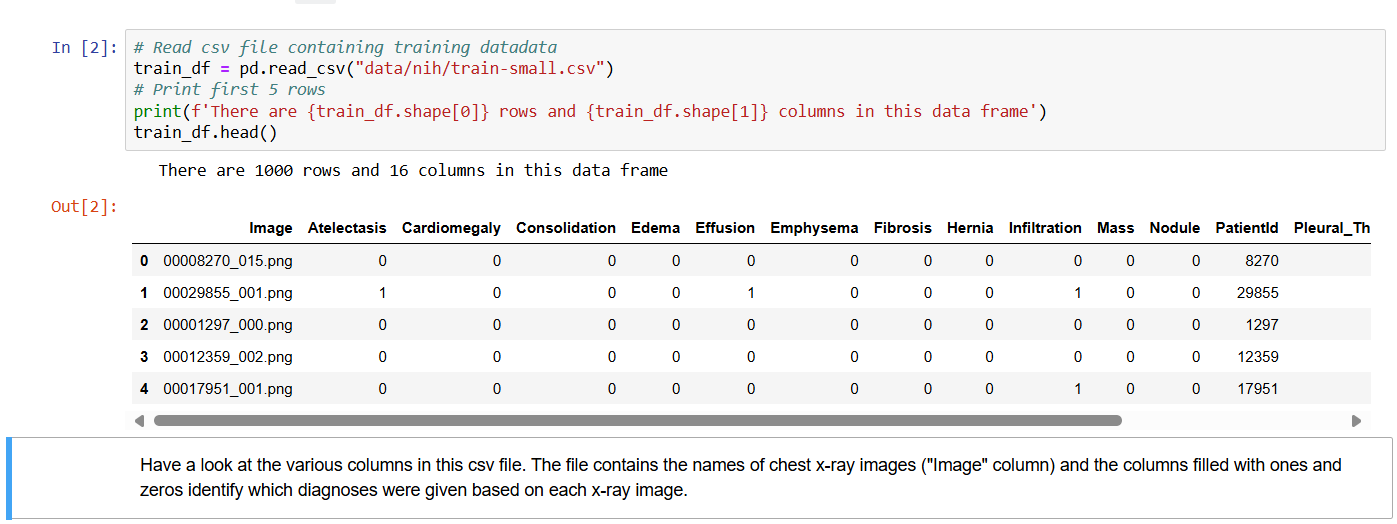
## **MSSV: HE176027**

### **1. Import library**

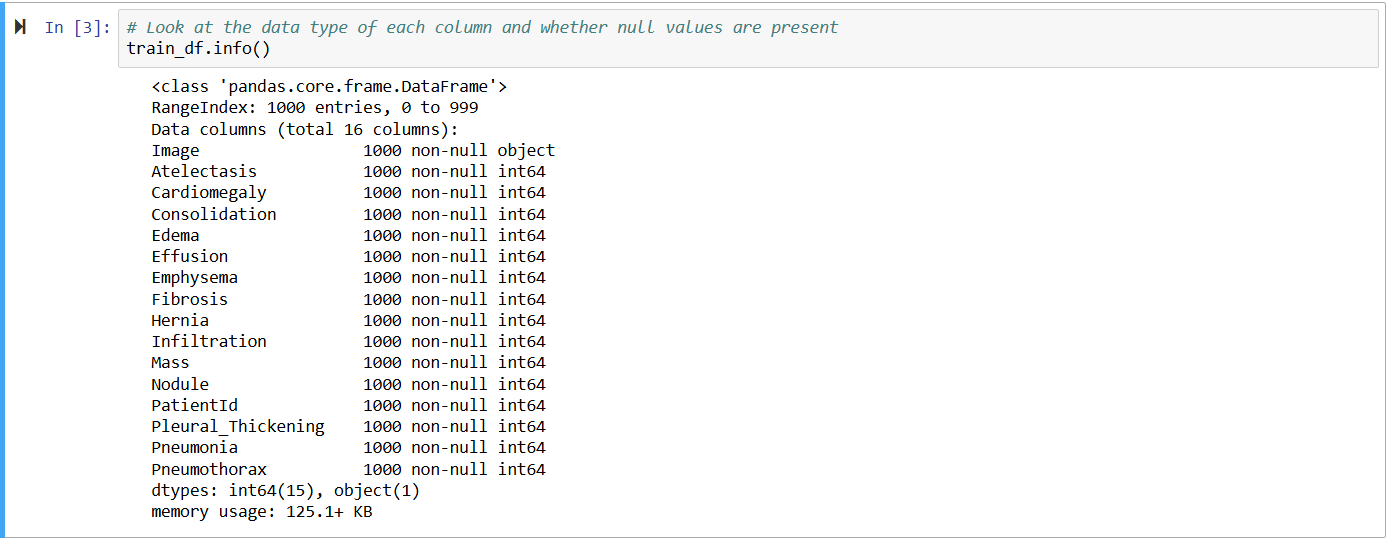


Cài đặt những thư viện cần thiết cho bài toán.

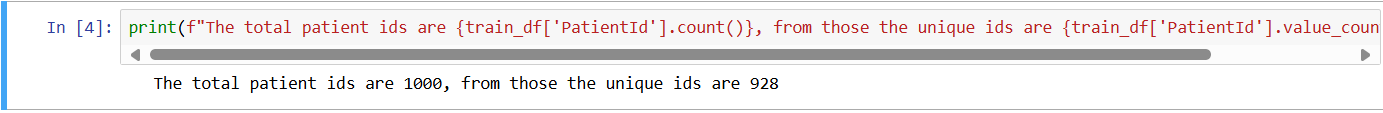
### **2. Exploration**



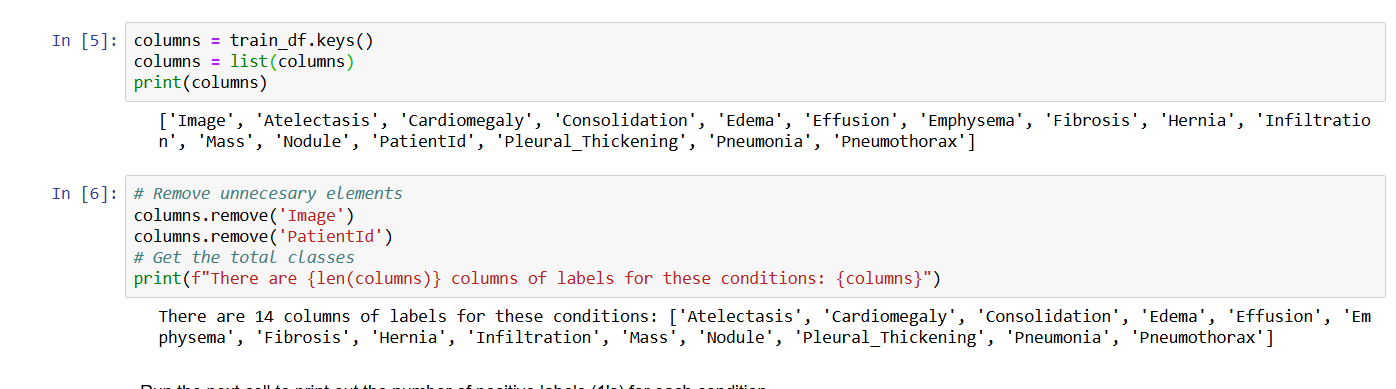
* Đọc dữ liệu từ file csv, tổng quan dữ liệu: 1000 hàng (điểm dữ liệu) và 16 cột (thuộc tính)



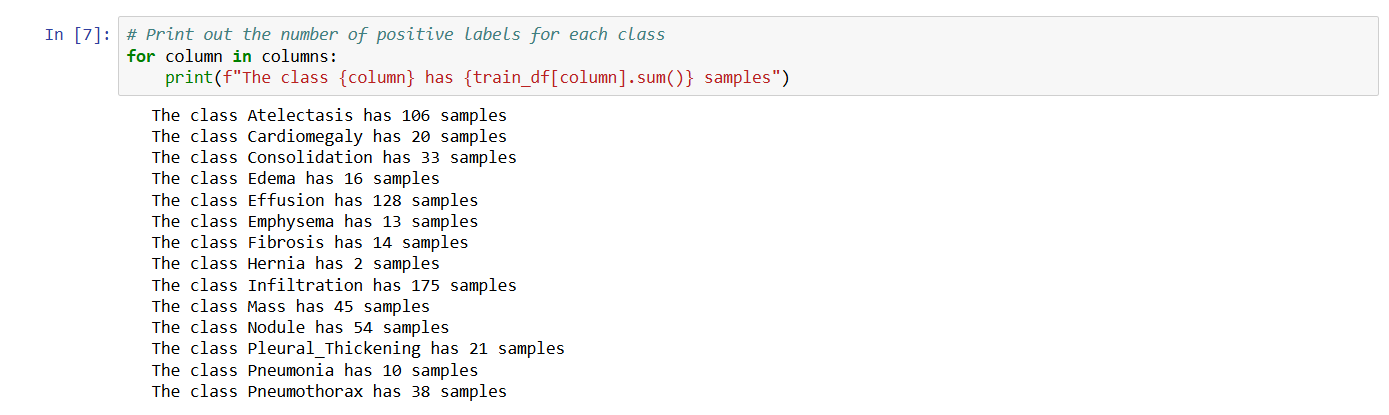
* Kiểm tra kiểu dữ liệu cùng và kiểm tra xem có điểm dữ liệu nào Null không:
  + Chủ yếu kiểu dữ liệu là int64, ngoài dữ liệu ảnh
  + Không có dữ liệu bị khuyết



* Kiểm tra số ID độc nhất, ở đây có ghi lại 1000 IDs, trong đó có 928 IDs độc nhất, nghĩa là có nhiều ID bệnh nhân bị trùng lặp

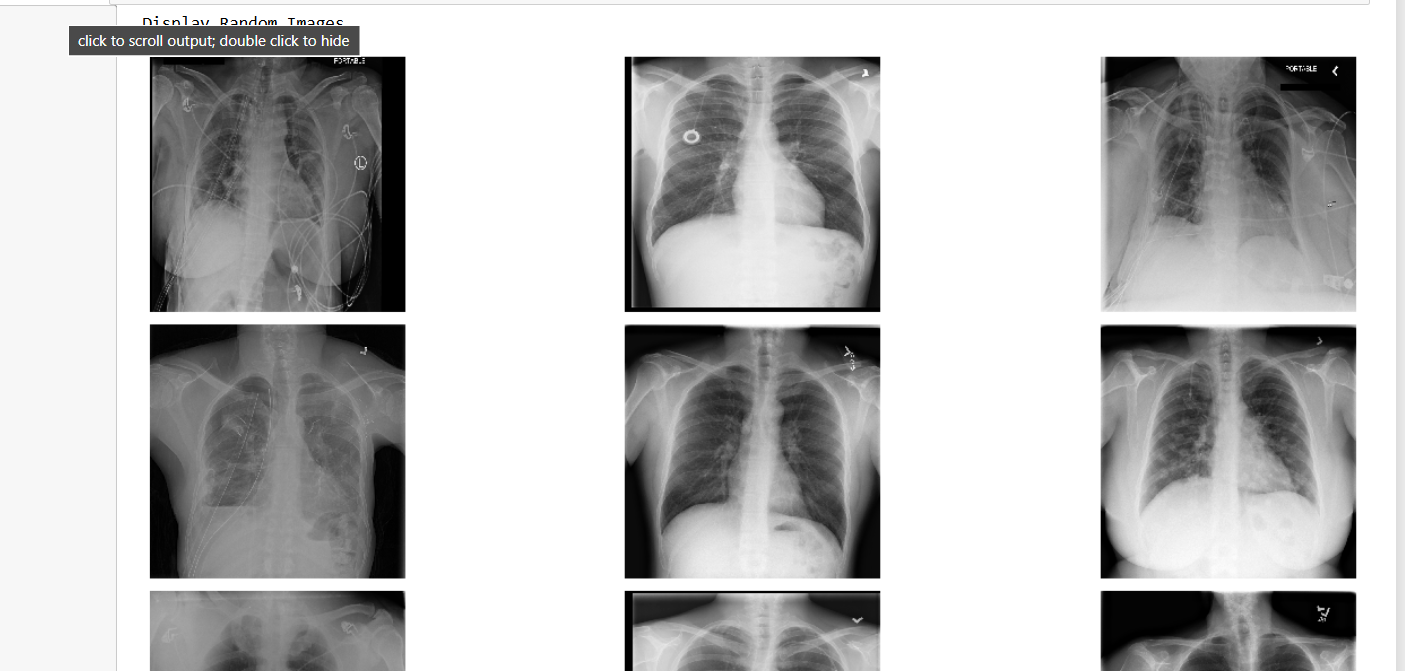


* Tạo danh sách kiểm tra các loại bệnh và triệu chứng mà bệnh nhân gặp phải

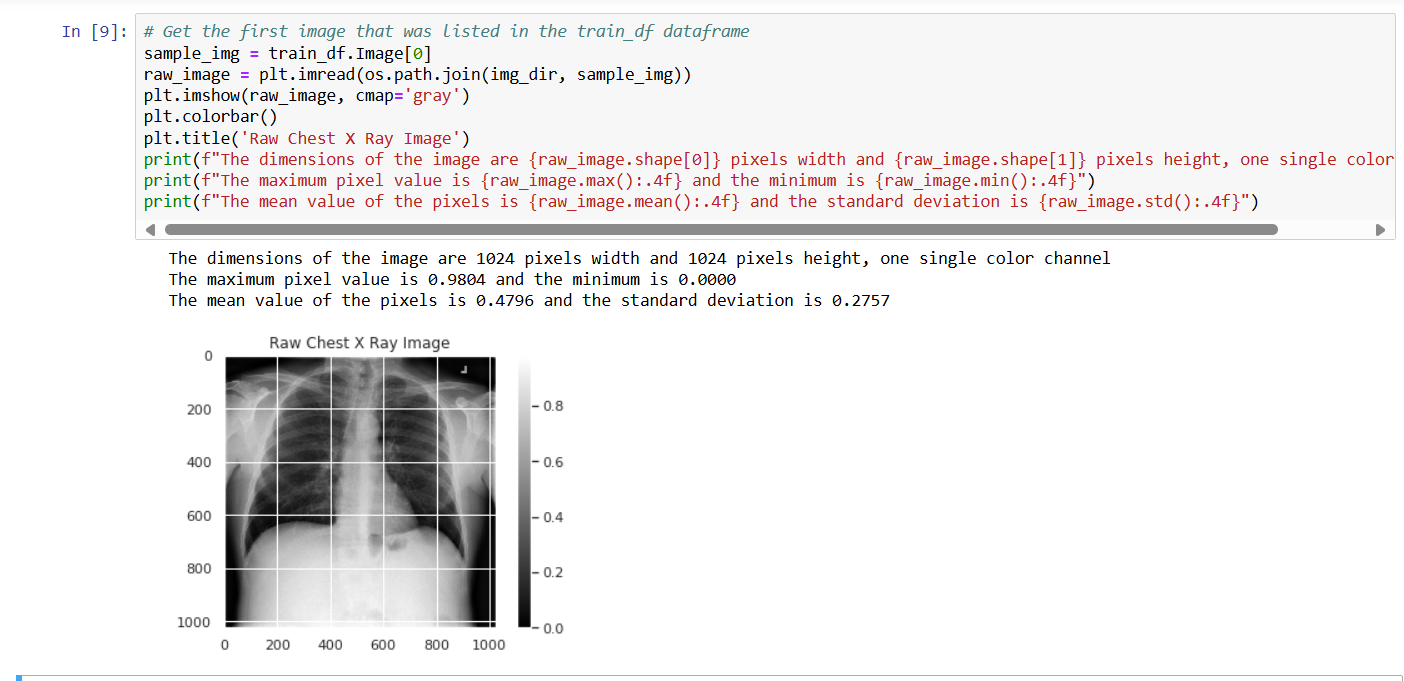


* Kiểm tra số lượng bệnh nhân có trong trong từng bệnh: Dữ liệu bị mất cân bằng

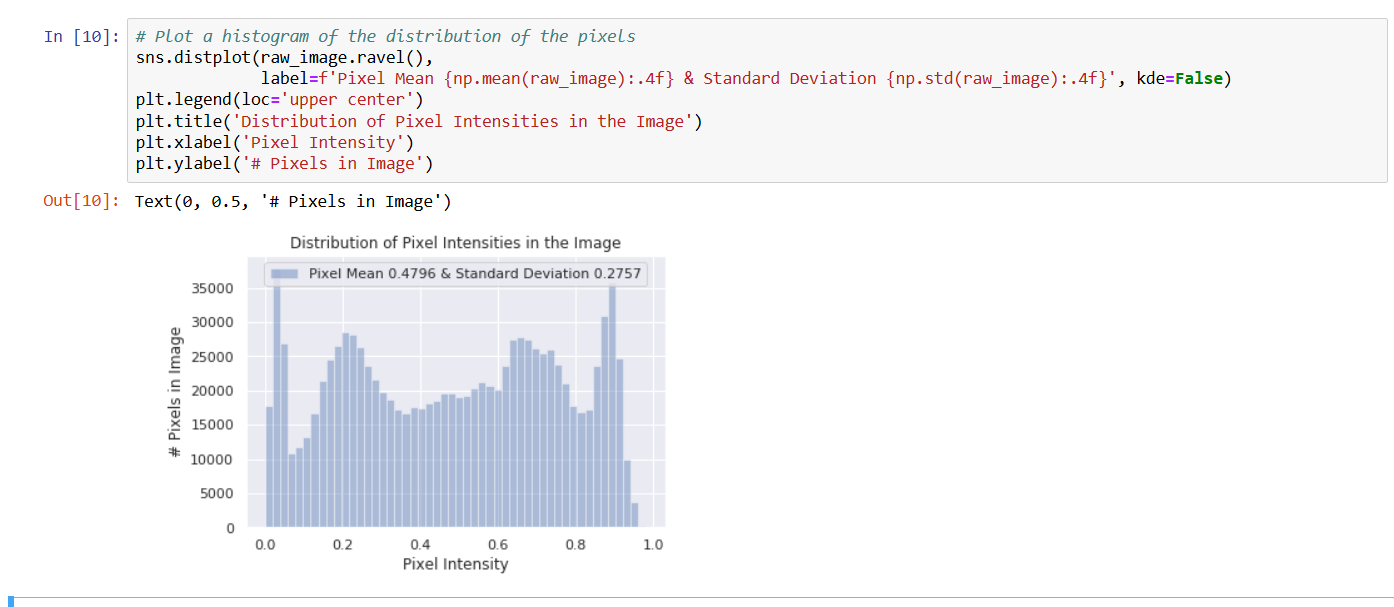




* In ra một số ảnh ngẫu nhiên trong bộ dữ liệu



* In ra ảnh đầu tiên và một số thông số về ảnh:
  + Chiều dài và chiều rộng của ảnh
  + Giá trị min max của pixel trong ảnh
  + Mean (Giá trị trung bình) và Standard Deviation (Độ lệch chuẩn)



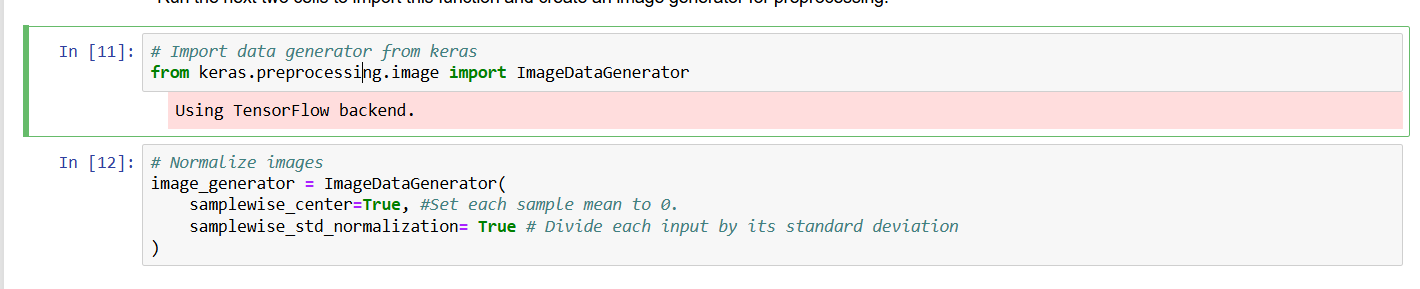
* In ra biểu đồ cột biểu diễn phân phối của các giá trị pixels trong ảnh (kèm theo thông tin về Mean và Standard Deviation của chúng

### 

### 

### 

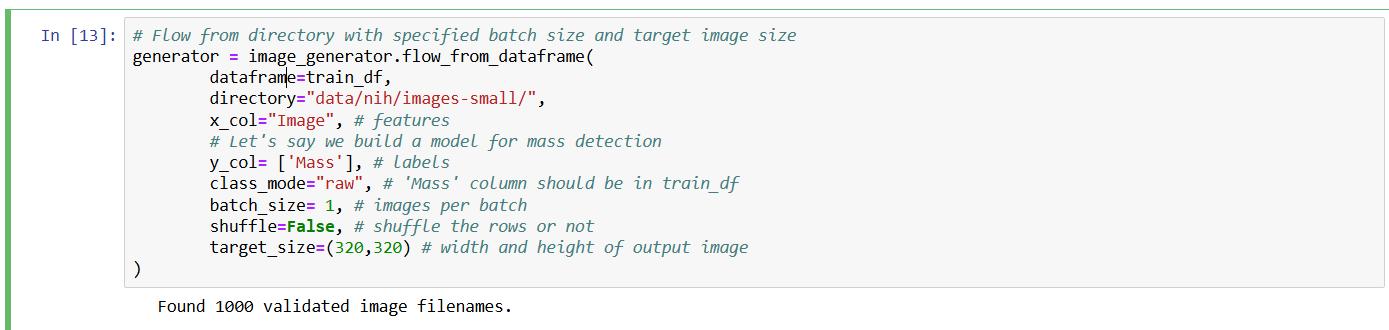
### **3. Image Processing in Keras**



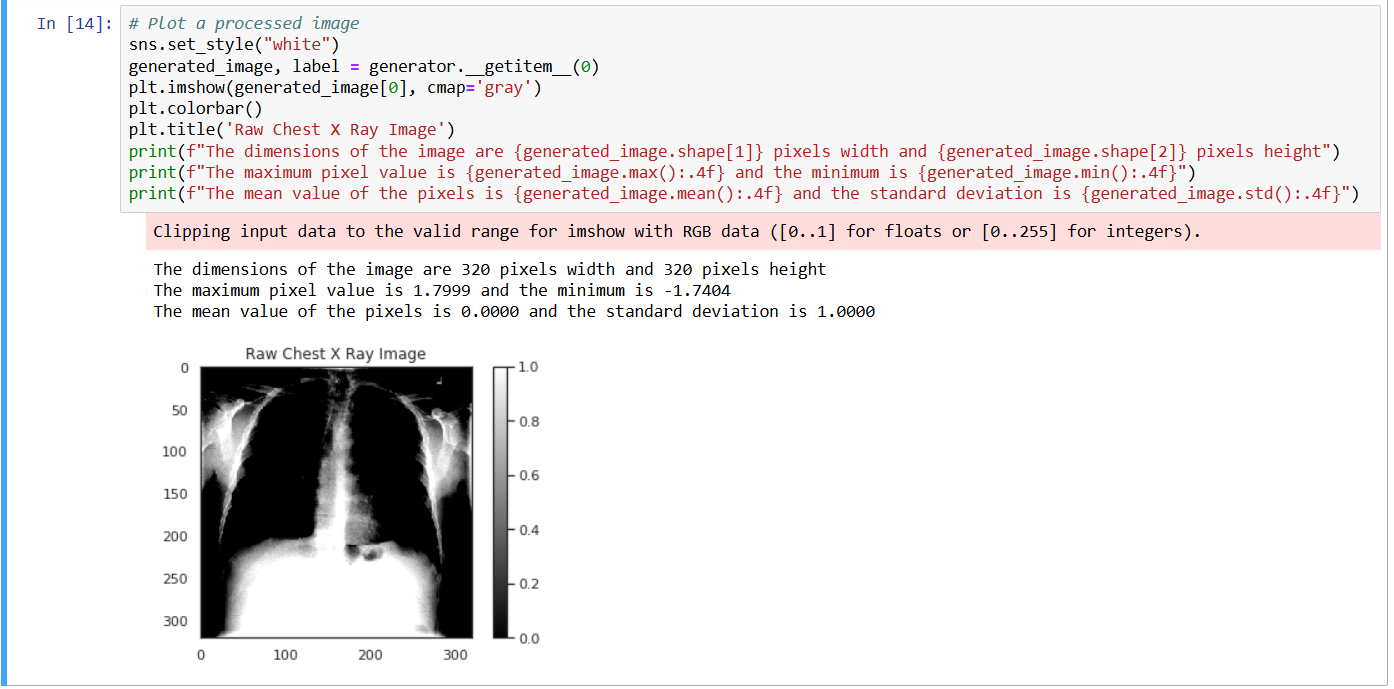
* Tạo hàm image\_generator từ Keras để chỉnh lại ảnh (Image Preprocessing Tiền xử lý ảnh) cho phù hợp trước khi đưa vào mạng CNN:
  + samplewise\_center=True: Đặt giá trị trung bình của ảnh xuống 0
  + samplewise\_std\_normalization= True: Chuẩn hóa lại độ lệch chuẩn về 1

-> Đưa phân phối giá trị pixels của ảnh về phân phối chuẩn

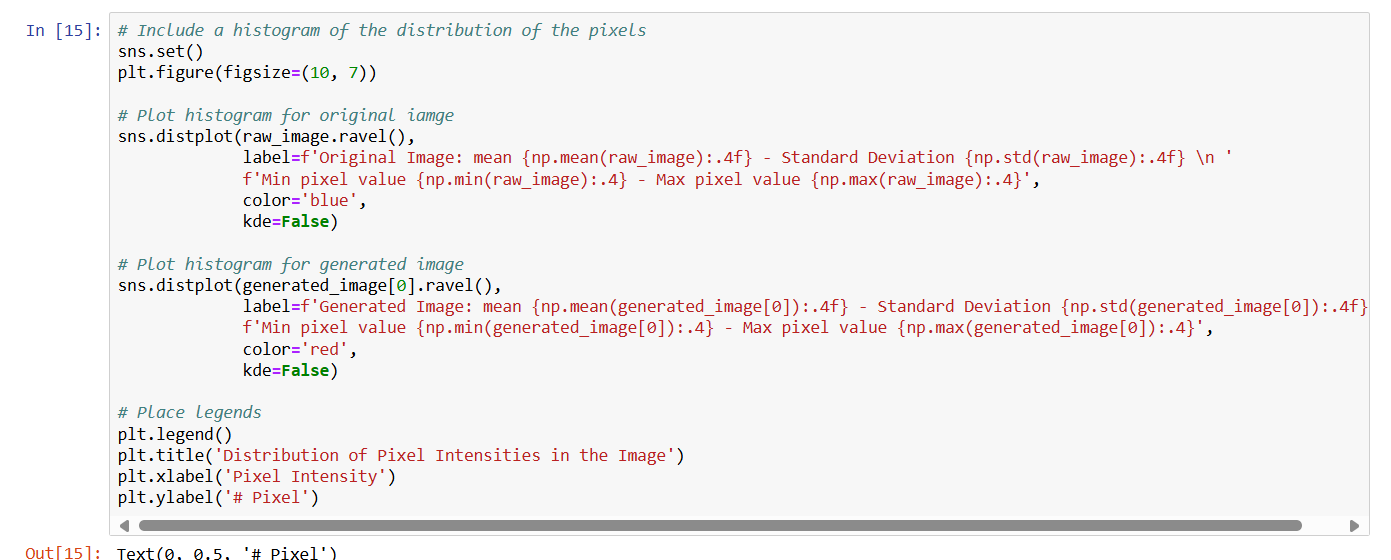
-> Giúp mô hình học nhanh dễ hội tụ hơn, đặc biệt là ảnh có độ lệch về giá trị pixels cao

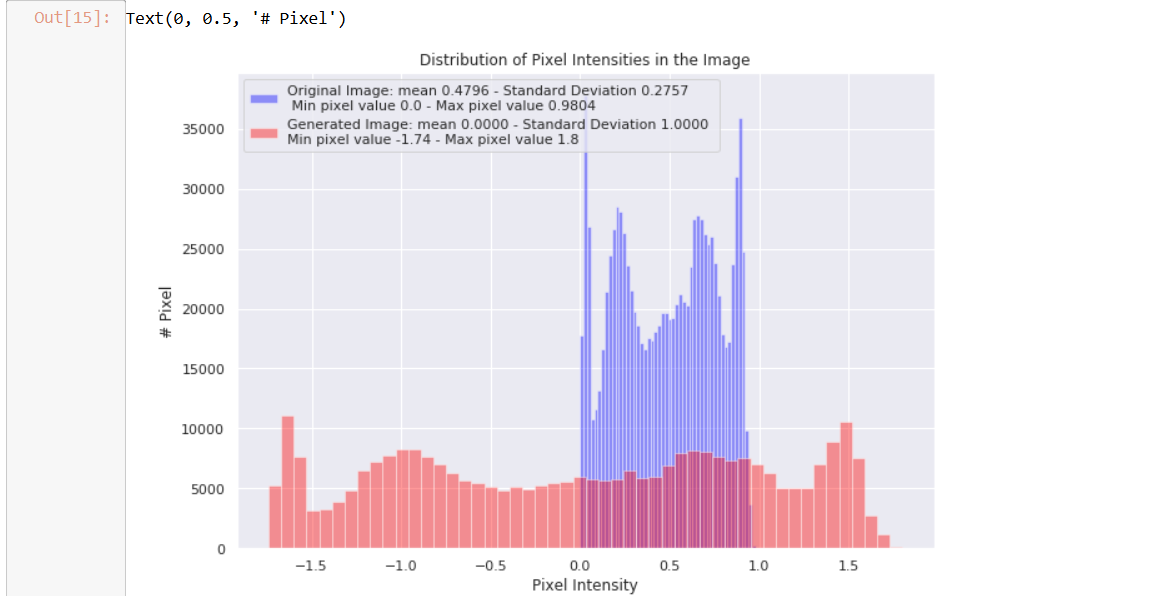


* Bắt đầu normalization các ảnh: Ngoài những thông tin trên, một số thông tin đi kèm là resize ảnh về (320, 320) với batch size là 1



* Kiểm tra kết quả sau khi tinh chỉnh





* In và so sánh phân phối giá trị pixels của ảnh sau khi thực hiện normalization và trước khi thực hiện normalization

-> Phân phối ảnh được đưa về dạng phân phối chuẩn N(0, 1)

-> Phân phối đều hơn và giảm được độ lệch giữa giá trị pixels thấp nhất và giá trị pixels cao nhất của ảnh