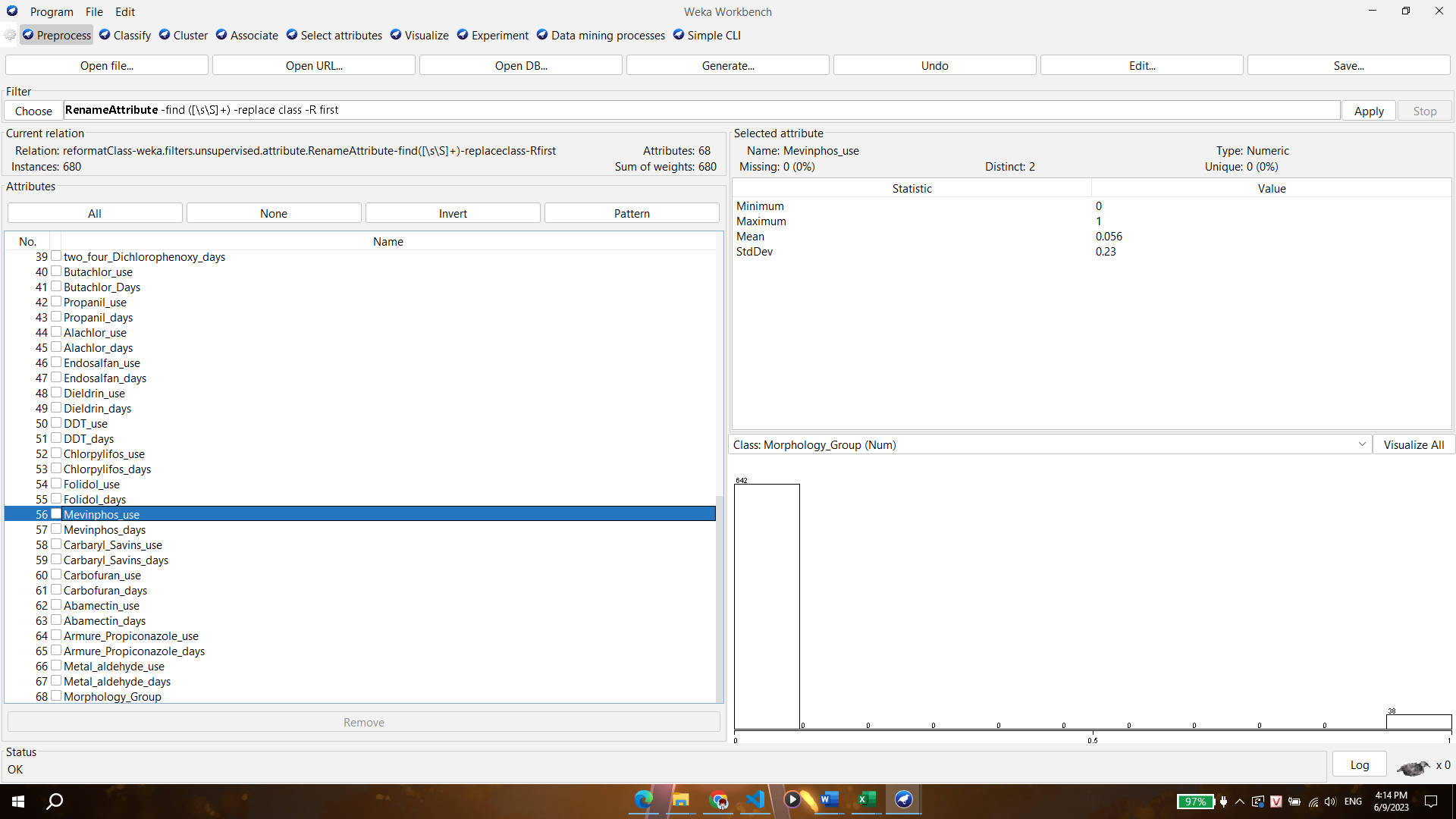
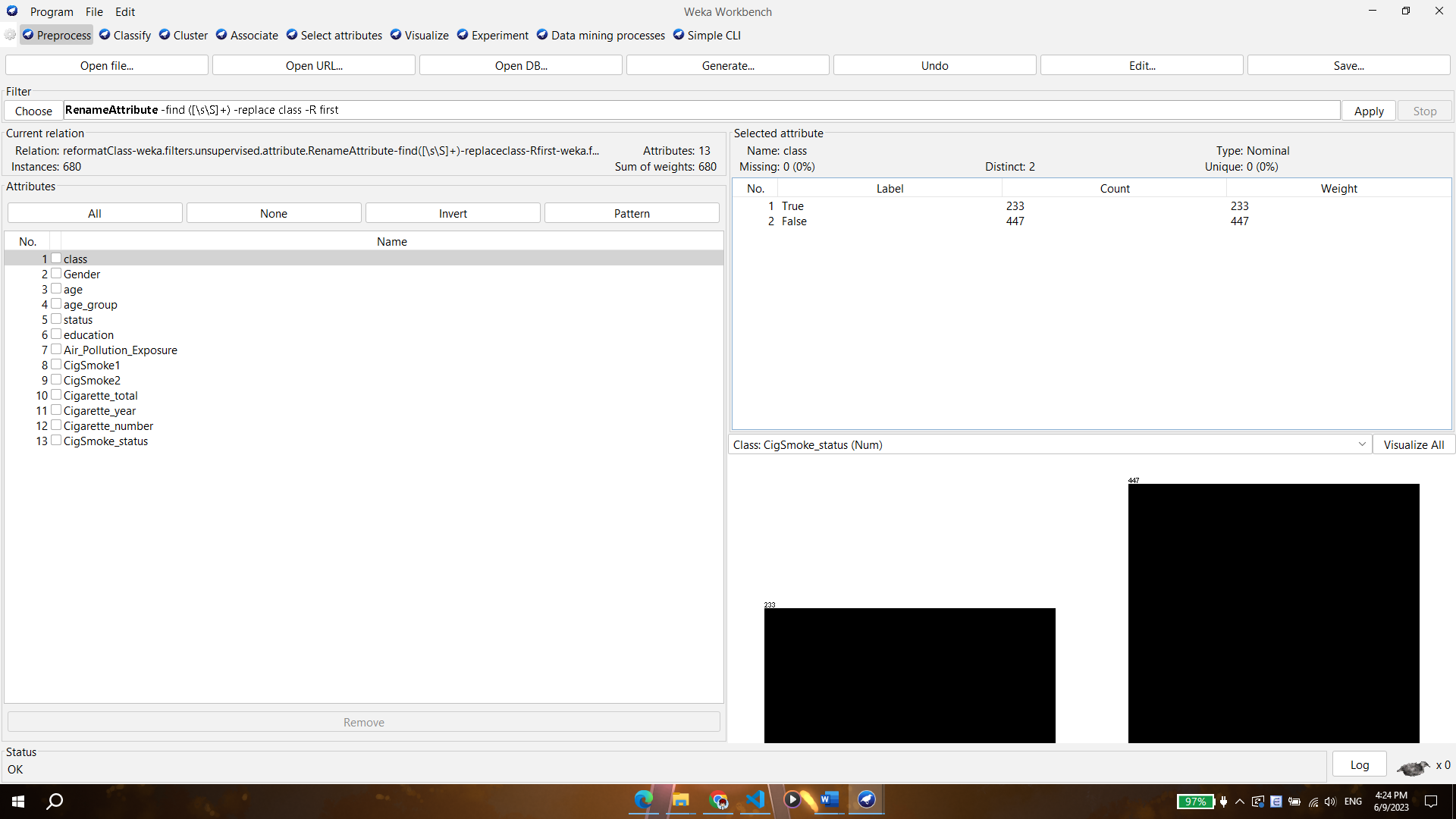
Sử dụng Weka

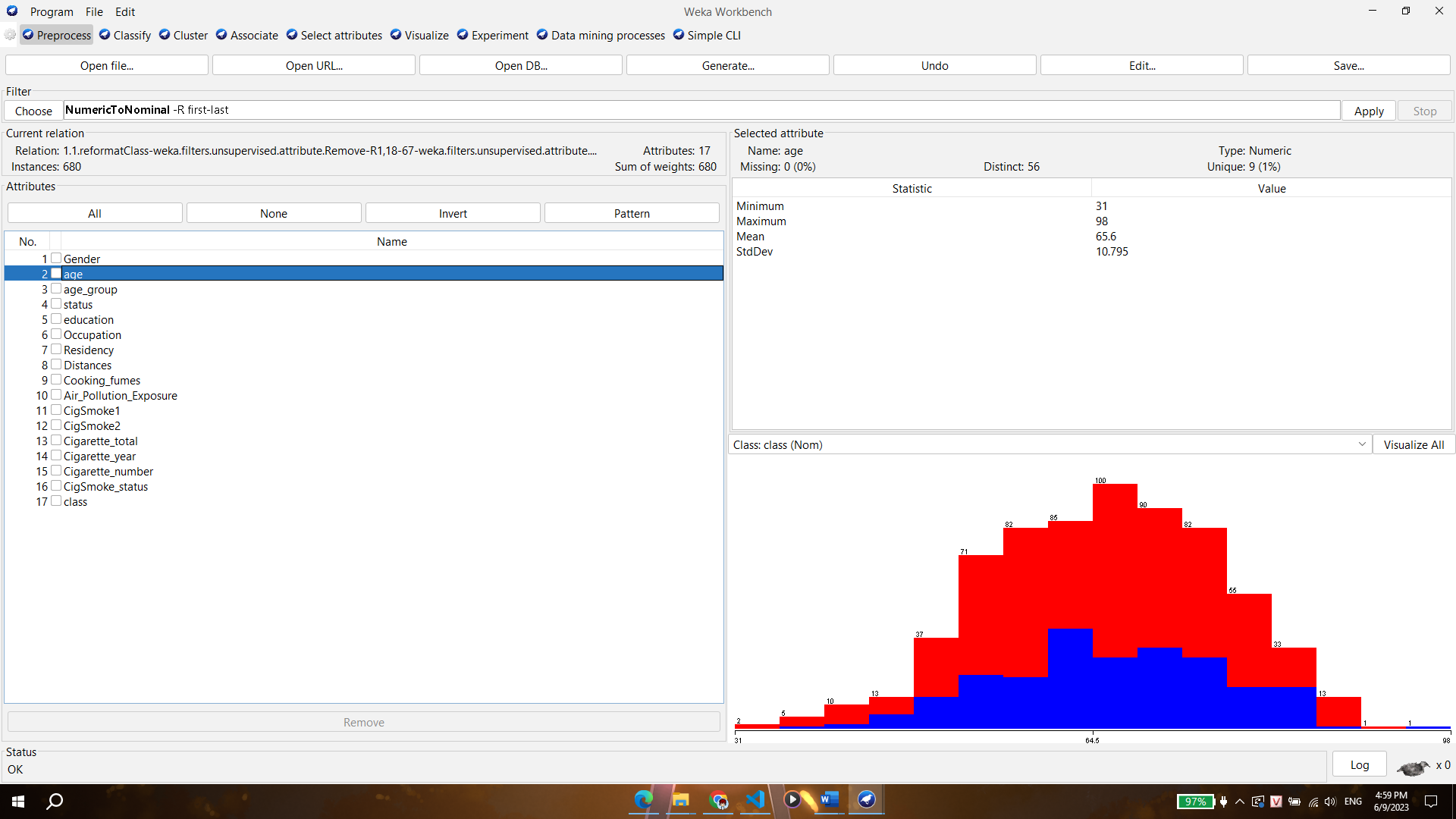
Dữ liệu gốc:



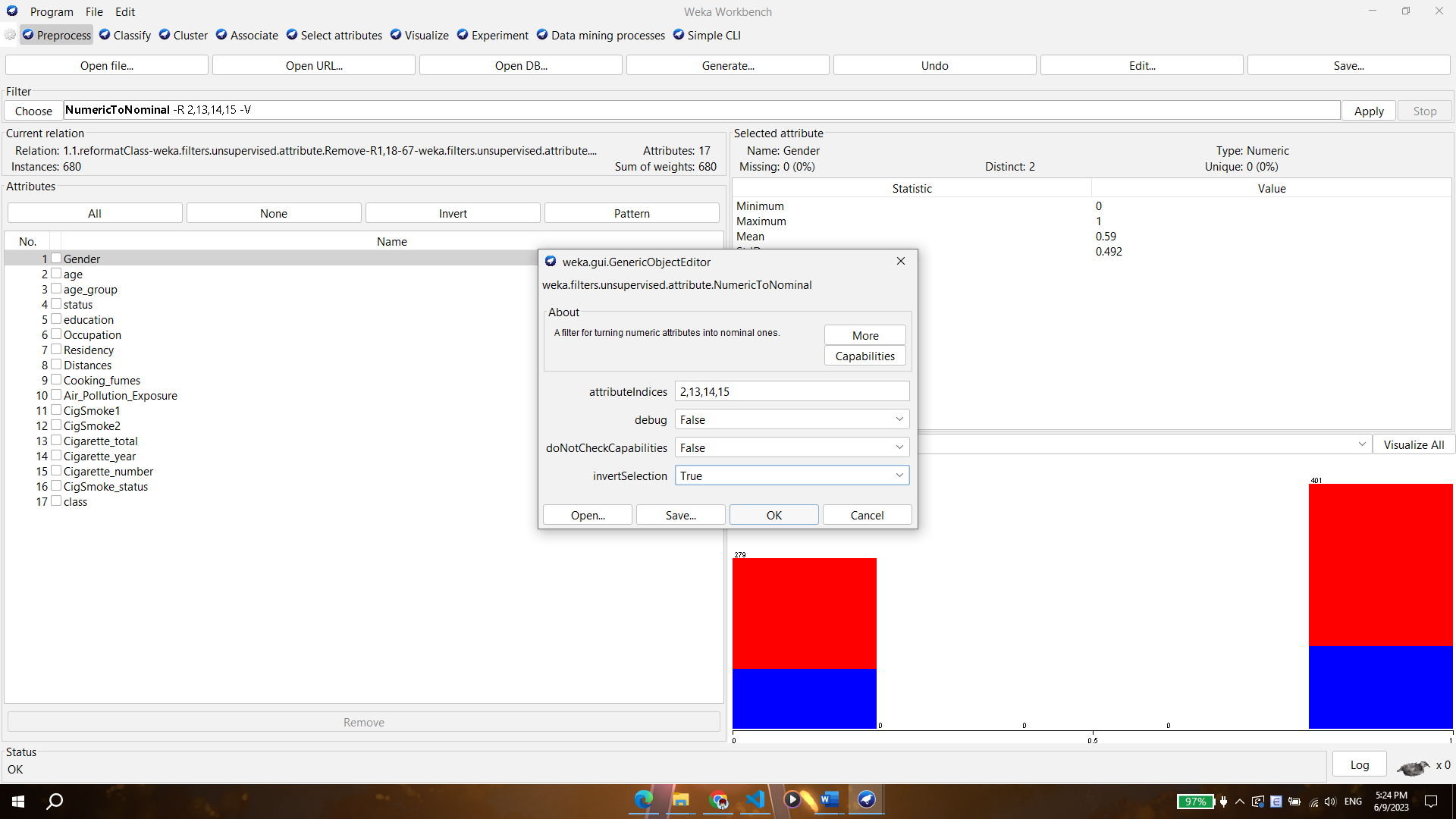
Thu giảm dữ liệu: loại bỏ bớt các thuộc tính về tính trạng ảnh hưởng từ thuốc trừ sâu, hay yếu tố nông nghiệp



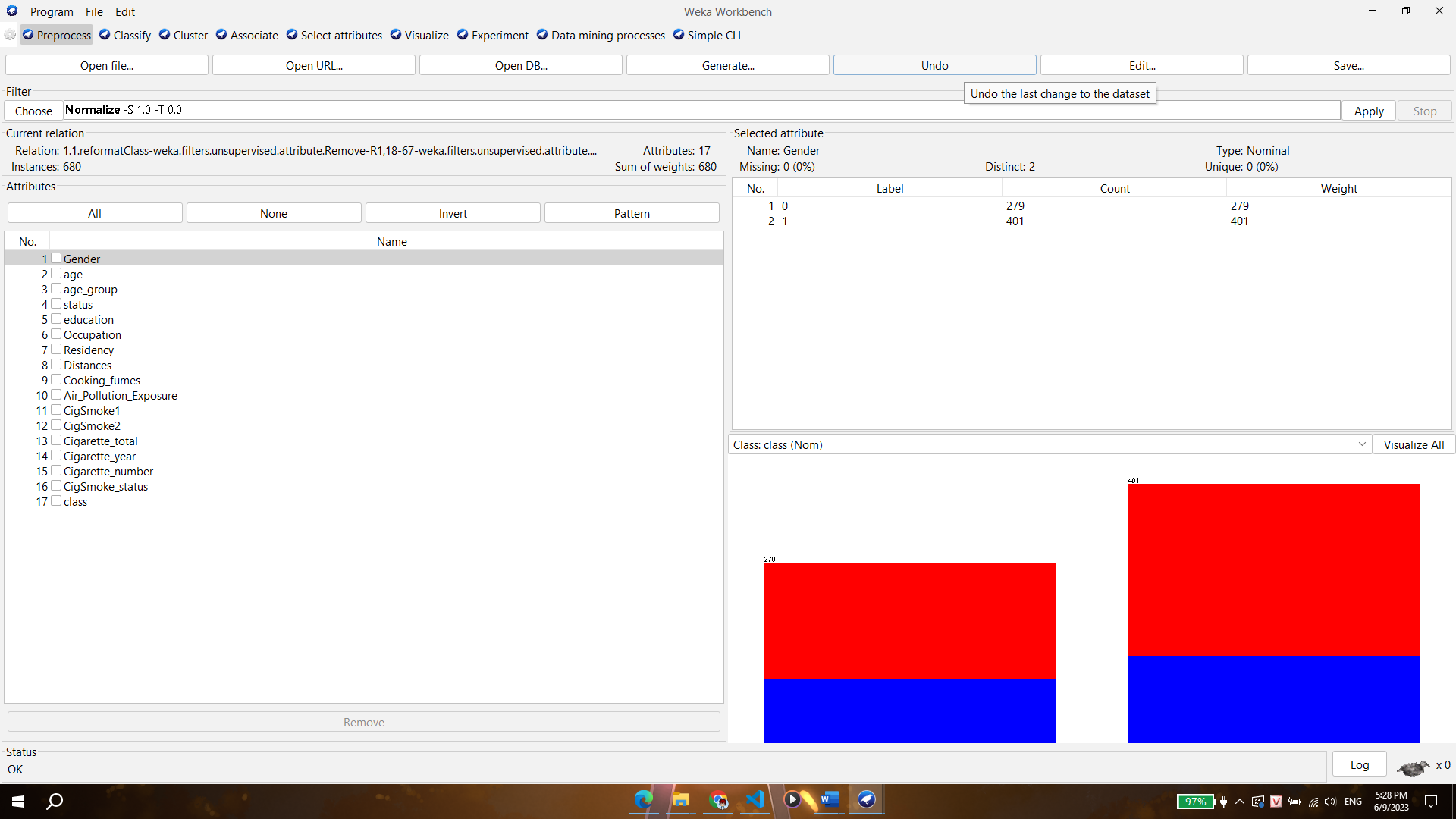
1. Chuyển đổi dữ liệu
   1. Tiêu chuẩn hóa dữ liệu (Stardardlize)
      1. Chuyển dữ liệu sang dạng Norminal: trừ các thuộc tính age, cigaratte\_total, cigaratte\_year, cigaratte\_number



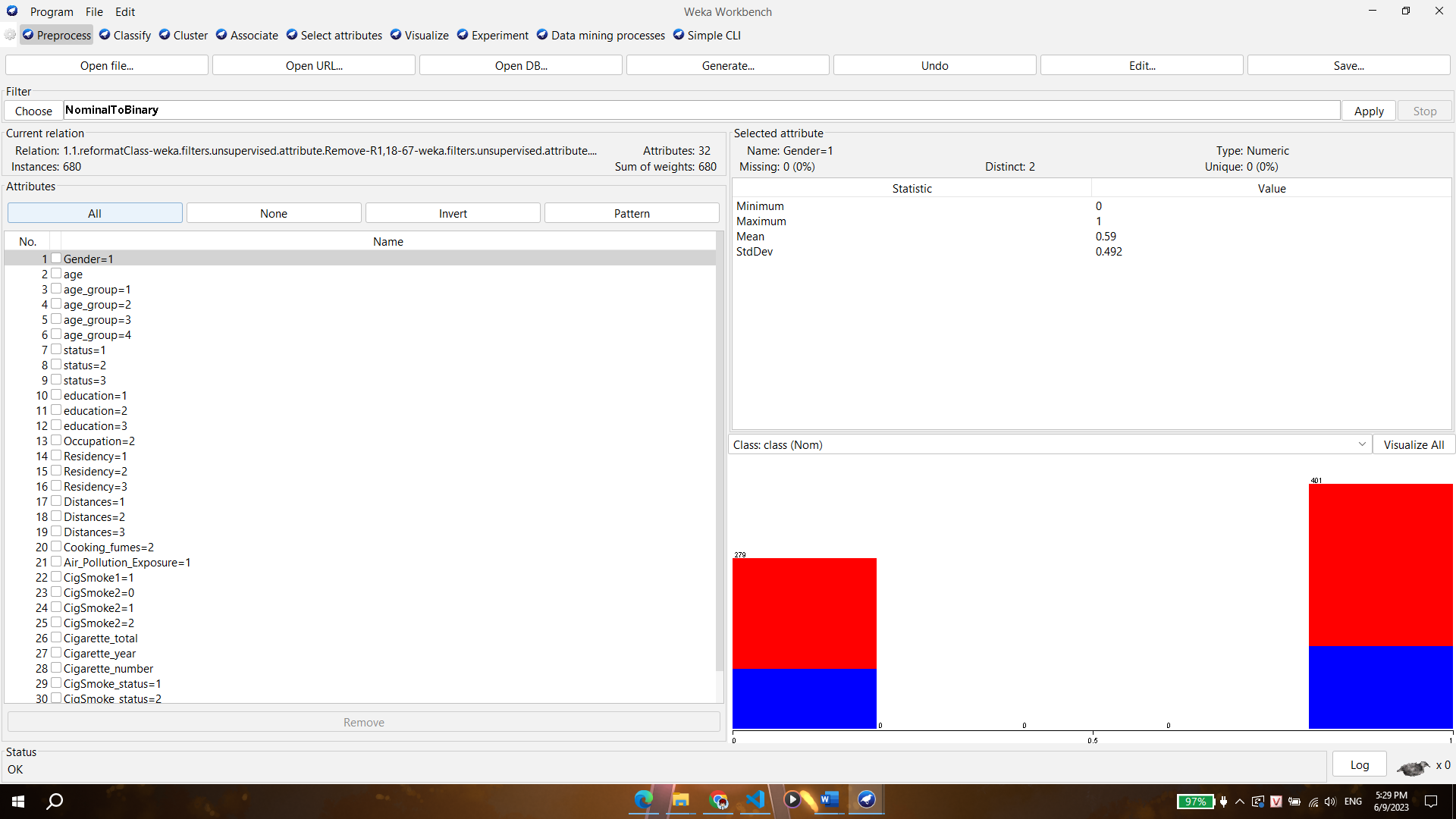
Sau khi chuyển đổi



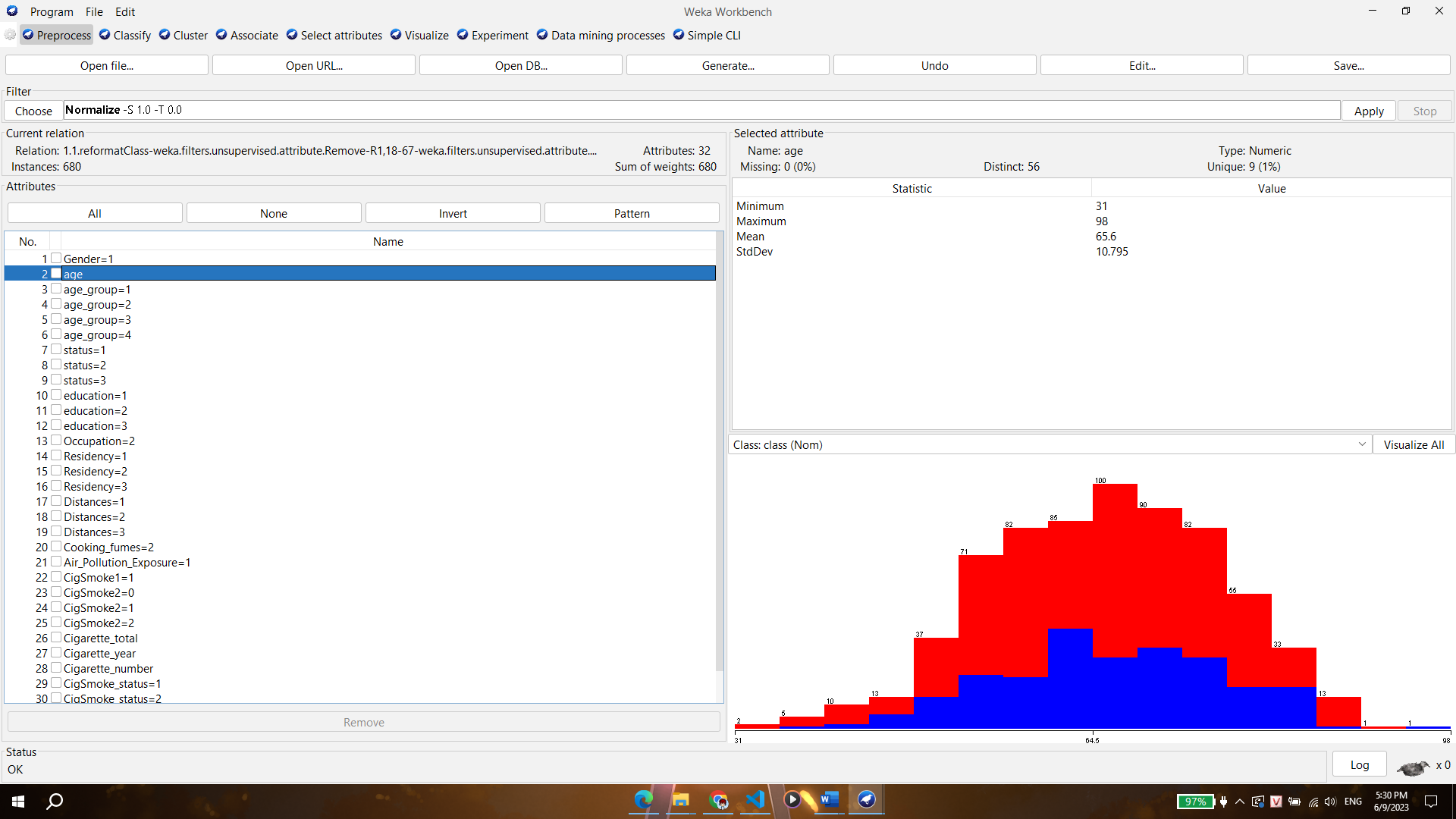
* + 1. Tạo các thuộc tính giả: Chuyển các dữ liệu có vị trí 3 -> 12 và vị trí 16, qua filter NorminalToBinary



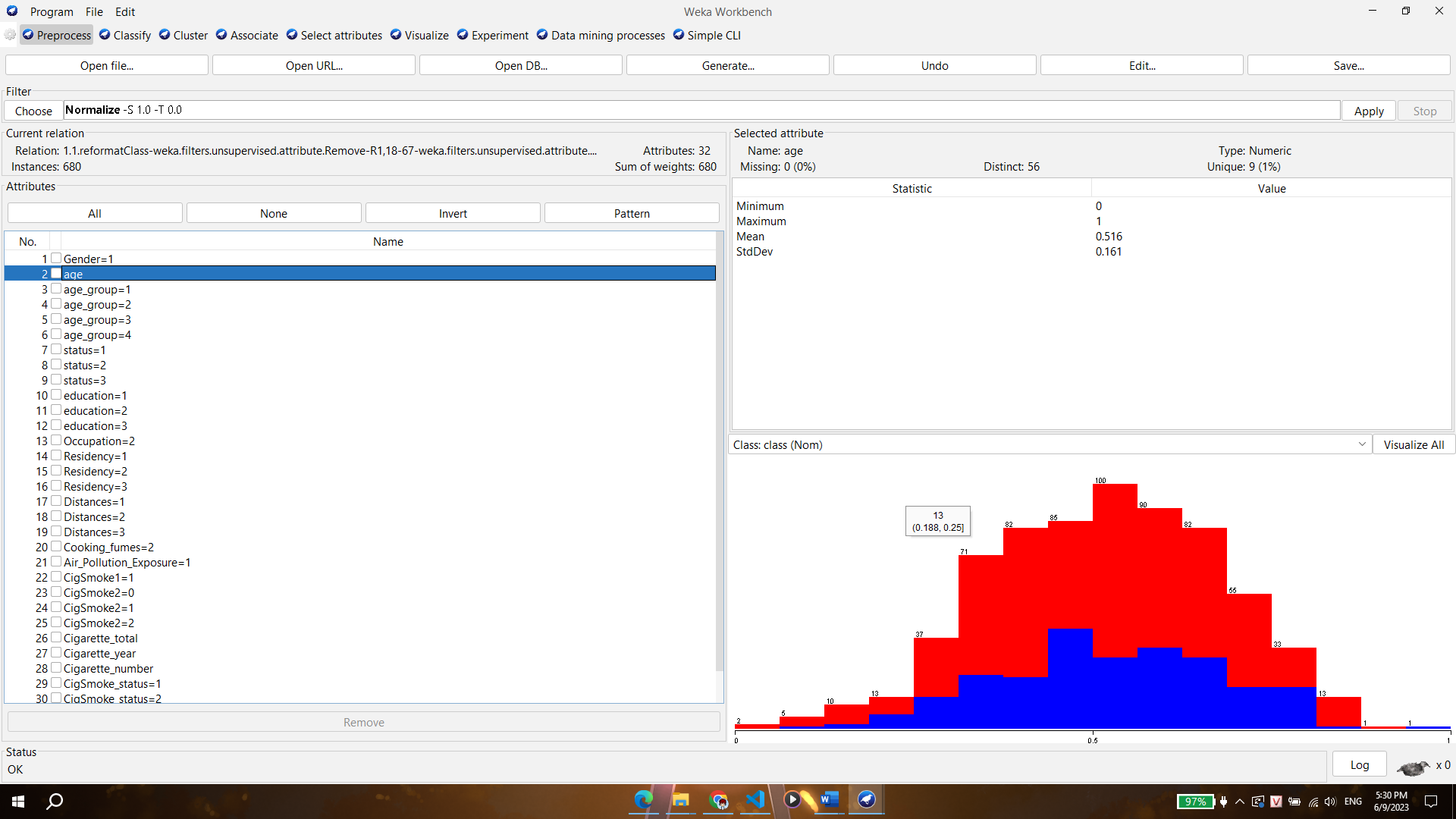
Kết quả thu được



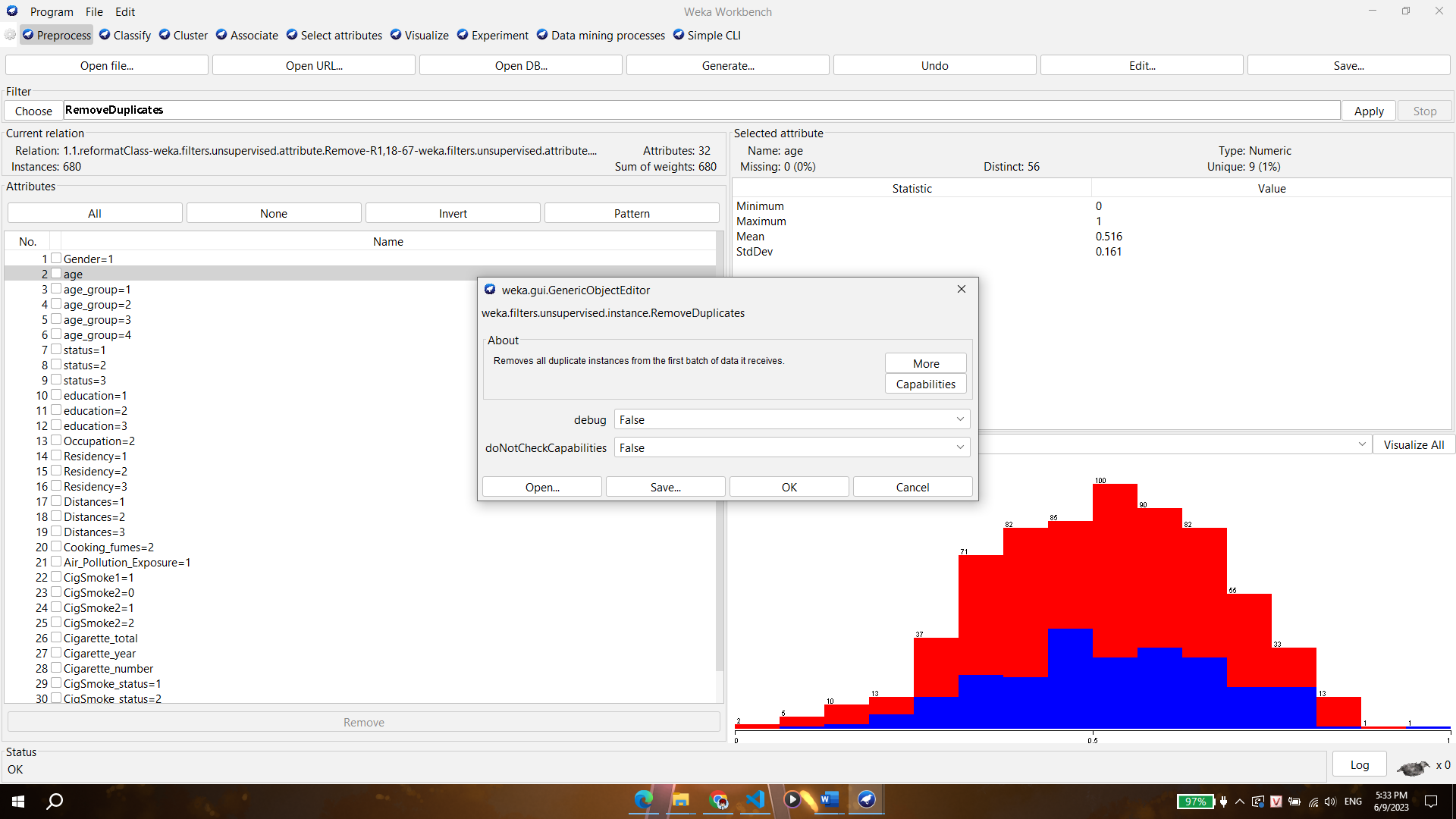
* 1. Chuẩn hóa dữ liệu (Normalize)



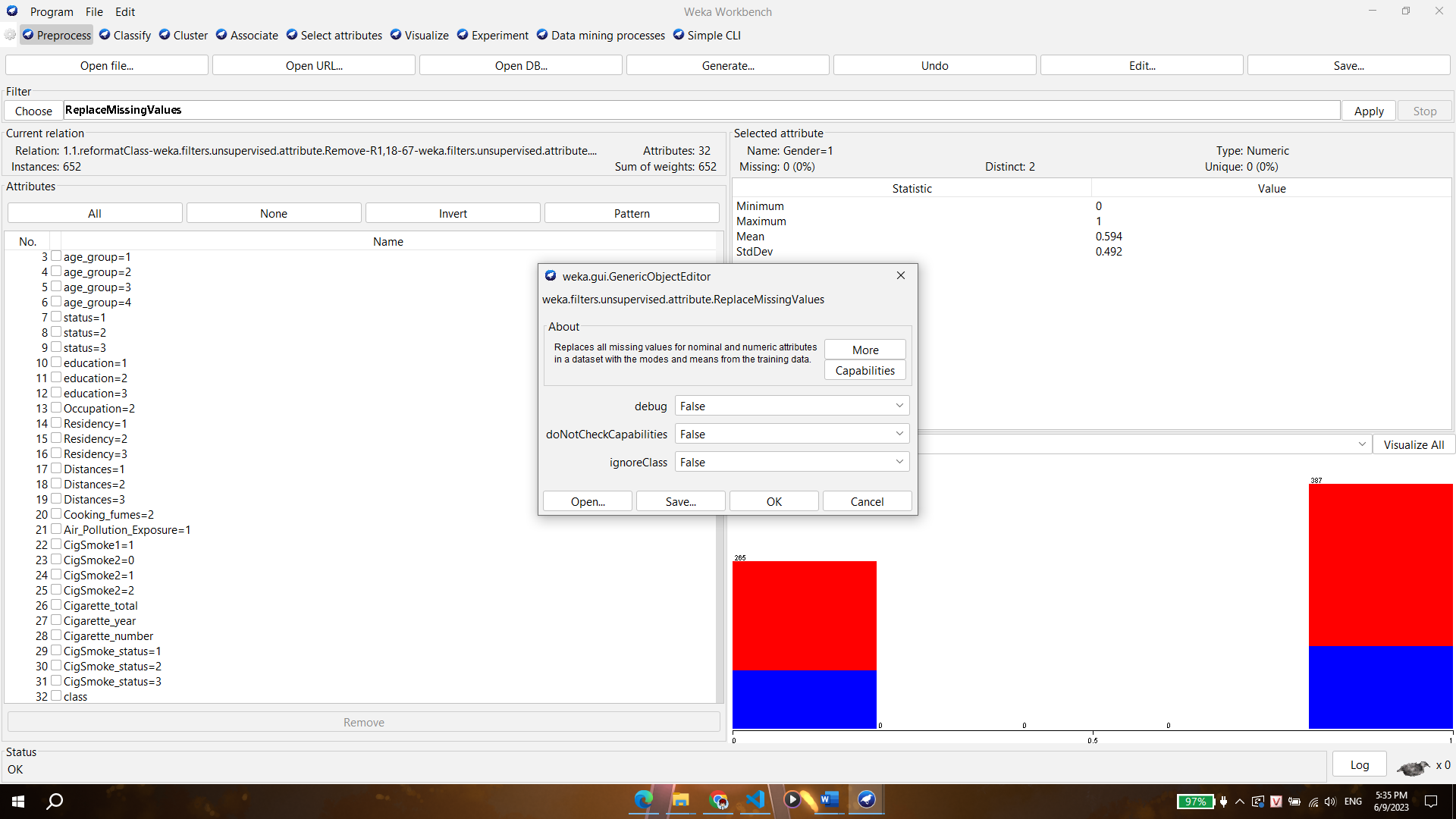
Sau khi chuẩn hóa



1. Thu giảm dữ liệu
2. Làm sạch dữ liệu
   1. Xóa dữ liệu trùng lặp



* 1. Tạo giá trị none cho các thuộc tính bị thiếu giá trị



Test split train

