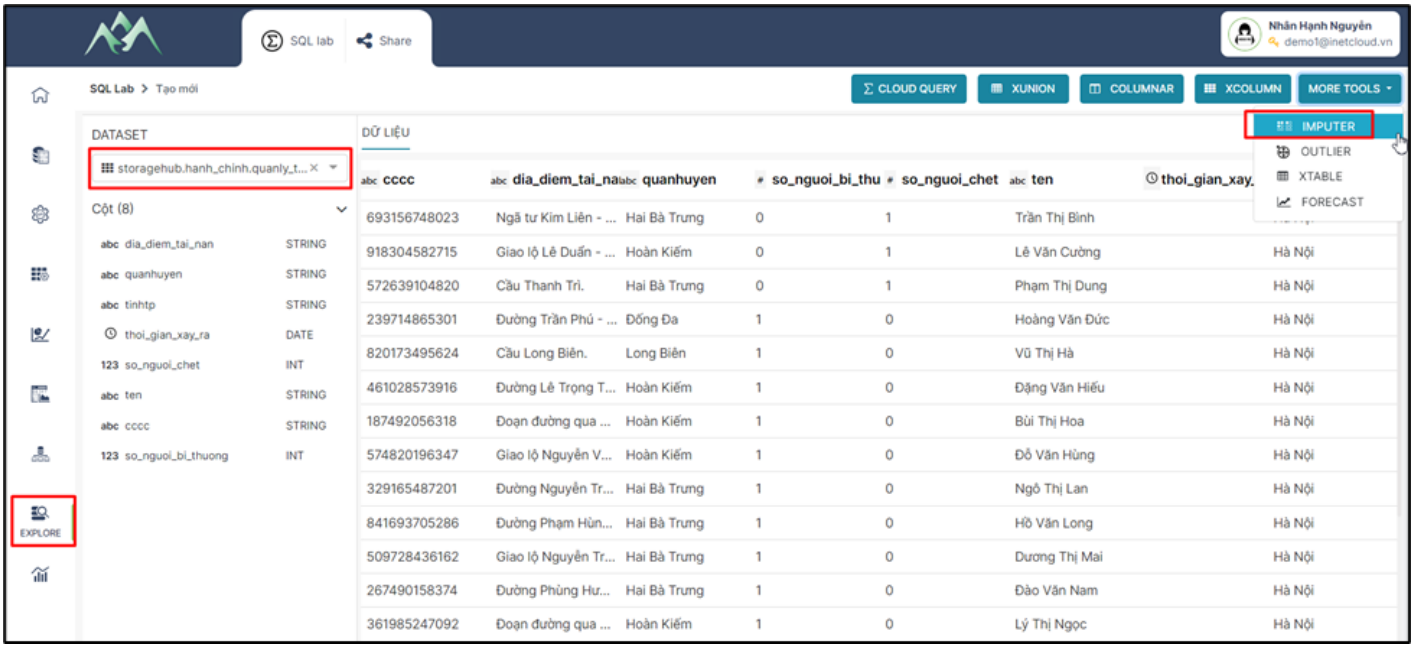
3. IMPUTER

Mục tiêu IMPUTER: được sử dụng để xử lý các giá trị thiếu trong dữ liệu, tức là các ô dữ liệu mà không có giá trị (có thể là NaN, null, hoặc các giá trị không có ý nghĩa khác). Khi dữ liệu bị thiếu, việc sử dụng Imputer giúp điền các giá trị thay thế vào những vị trí thiếu này để có thể tiếp tục phân tích dữ liệu một cách chính xác.

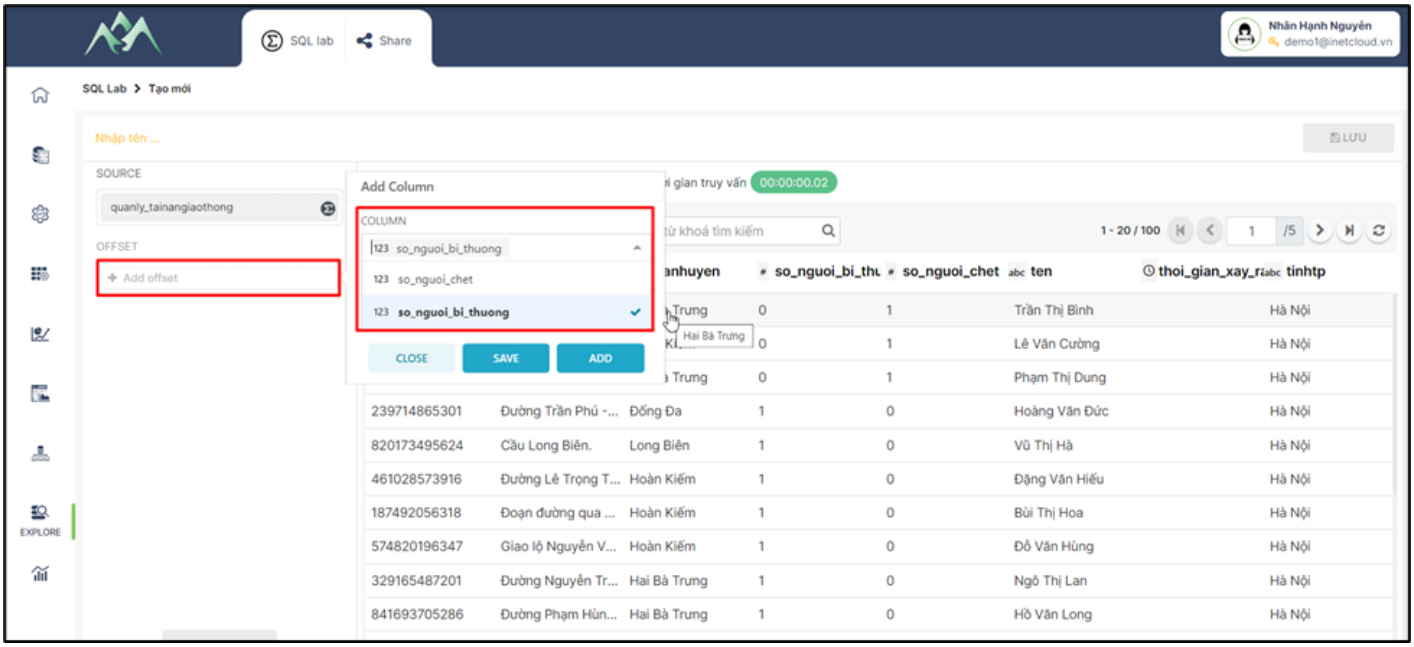
- Thực hiện trên hệ thống theo các bước hướng dẫn sau:

Thực hiện trên hệ thống theo các bước hướng dẫn sau:

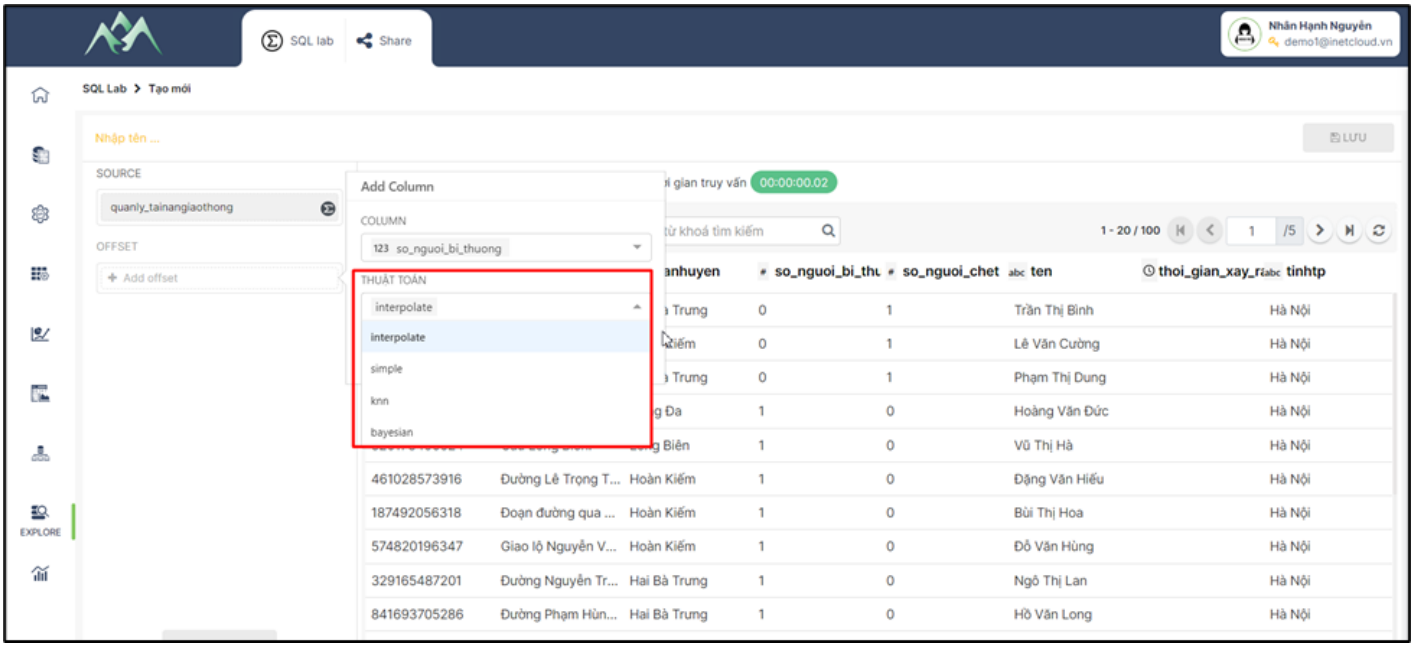
Bước 1: Vào hệ thống chọn module Explore, chọn tập dữ liệu cần tổng hợp và tạo mới một tập dữ liệu tổng hợp từ IMPUTER



Bước 2: Chọn cột dữ liệu để thuật toán Imputer giúp điền các giá trị thay thế vào những vị trí thiếu.



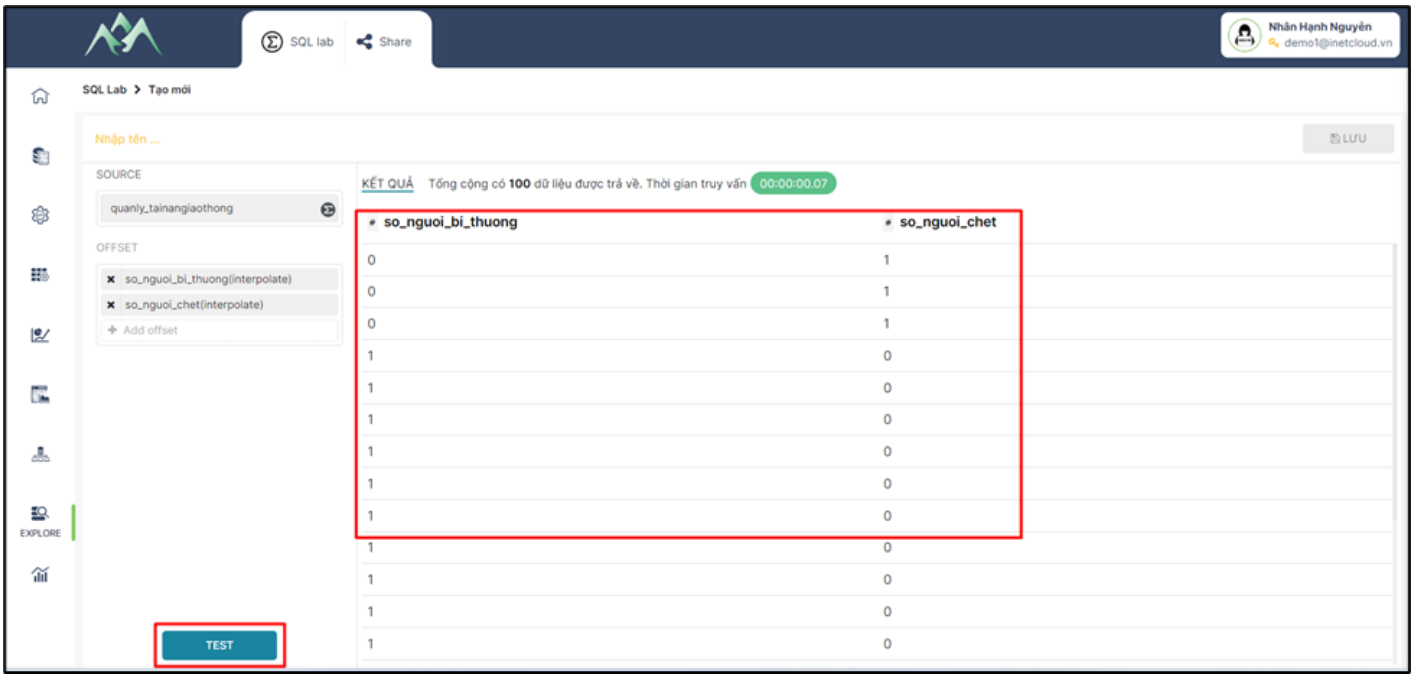
Bước 3: Chọn thuật toán xử lý cho cột dữ liệu đã chọn.



Trong đó:

- Thuật toán nội suy (interpolation) được sử dụng để ước lượng giá trị tại các điểm trung gian dựa trên các giá trị đã biết của một tập dữ liệu.
- Thuật toán Simple được dùng để xử lý các bài toán như tìm kiếm tuyến tính, sắp xếp nổi bọt, hoặc tính giai thừa.
- Thuật toán KNN là một thuật toán dựa trên khoảng cách để phân loại hoặc dự đoán giá trị của một điểm dữ liệu mới dựa trên các điểm dữ liệu đã biết trong không gian đa chiều.
- Thuật toán Payesian là một trong những thuật toán phân loại phổ biến nhất trong học máy, đặc biệt là đối với các bài toán phân loại văn bản và lọc thư rác.

Bước 4: “RUN” chạy test dữ liệu được tổng hợp.



Bước 5: Nhập tên cho tập dữ liệu mới được tổng hợp, chọn Lưu thông tin tập dữ liệu được tổng hợp bằng công cụ IMPUTER