* Ngữ cảnh: Nhu cầu giải trí của con người hiện nay ngày càng phát triển và đa dạng các loại hình. Tuy nhiên, phim ảnh vẫn đóng vai trò quan trọng trong đời sống tinh thần của con người. Mỗi năm có hàng ngàn bộ phim mới được sản xuất ra thị trường với nội dung lôi cuốn, kỹ xảo hấp dẫn. Nắm bắt được nhu cầu đó, nhóm bắt đầu thực hiện tìm kiếm dữ liệu về các bộ phim, sau đó thực hiện trực quan hóa dữ liệu để tìm hiểu các mối liên hệ giữa các trường dữ liệu với nhau.
* Chủ đề: Dữ liệu mà nhóm quyết định tìm hiểu là Bảng xếp hạng 100 bộ phim nổi tiếng nhất năm, từ năm 2003 đến năm 2022, dữ liệu được thu thập từ trang iMDB (Internet Movie Database) - một trang cơ sở dữ liệu trực tuyến về điện ảnh thế giới.
* Dữ liệu được cấp phép công khai “CC0: Public Domain” cho phép phân phối miễn phí, người khác có thể sử dụng nội dung mà không gặp bất kỳ ràng buộc nào.
* Dữ liệu được thu thập bằng web scraping, là quá trình sử dụng các phần mềm máy tính để trích xuất dữ liệu và nội dung từ các website, dữ liệu thu thập được sau đó được lưu vào định dạng hữu ích, cụ thể là bảng tính.