**BÁO CÁO TIẾN ĐỘ THỰC HIỆN ĐỒ ÁN CUỐI KỲ**

Nhóm 19:

Ngô Phi Lít 18133024

Nguyễn Phạm Duy Khiêm 19133027

**1/ Nguồn Data:** Tập dữ liệu Phim và chương trình truyền hình trên Netflix.

* Link nguồn data: <https://www.kaggle.com/datasets/shivamb/netflix-shows>
* Giới thiệu về nguồn data: Netflix là một trong những nền tảng phát trực tuyến video và phương tiện phổ biến nhất. Họ có hơn 8000 bộ phim hoặc chương trình truyền hình trên nền tảng của mình, tính đến giữa năm 2021, họ có hơn 200 triệu Người đăng ký trên toàn cầu. Tập dữ liệu dạng bảng này bao gồm danh sách tất cả các bộ phim và chương trình truyền hình có sẵn trên Netflix, cùng với các thông tin chi tiết như - dàn diễn viên, đạo diễn, xếp hạng, năm phát hành, thời lượng,…

**2/ Tạo data source:**

* Là một tệp csv với 12 thuộc tính như sau:

+ show\_id: mã id.

+ type: loại (Movie or TV show)

+ title: tiêu đề

+ director:

+ cast: diễn viên

+ country: quốc gia

+ data\_added: ngày thêm vào Netflix

+ release\_year: năm phát hành

+ rating: phân loại phim dựa trên độ tuổi ( hệ thống phân loại MPAA trên lãnh thổ của Hoa Kỳ)

+ duration: thời lượng

+ listed\_in: thể loại phim

+ description: mô tả

**3/ Định nghĩa các câu hỏi cần trả lời:**

* Tổng quát về số lượng phim và chương trình truyền hình của mỗi quốc gia.
* Tình hình sản xuất phim và chương trình truyền hình của mỗi quốc gia qua các năm.
* Số lượng của từng loại phân loại phim trên độ tuổi.