Lại Thị Tâm -21810310492

Netflix Engagement (Jan - Jun 23) +

MÔ TẢ DỮ LIỆU

Bao gồm:

* Title (Tiêu đề)
* Available globally (Có sẵn trên toàn cầu)
* Release Date (Ngày ra mắt)
* Hours Viewed (Số giờ đã xem)
* Numbers of rating (Số lượt xếp hạng)
* Rating (Xếp hạng)
* Genre (Thể loại)
* Key Word (Từ khóa)
* Description (Mô tả)

PHÂN TÍCH DỮ LIỆU

* Title: Có 18214 tiêu đề của phim hoặc series phim
* Available globlely:

+ 4534 phim được chiếu sẵn trên toàn cầu (25%)

+ 13.8K phim không có sẵn (75%)

* Release Date:

+ 4879 phim có lịch ra mắt (27%)

* 01/04/2010 – 28/07/2011: 10 phim
* 23/11/2012 – 22/03/2014: 16 phim
* 22/03/2014 – 19/07/2015: 63 phim
* 19/07/2015 – 14/11/2016: 192 phim
* 14/11/2016 – 12/03/2018: 495 phim
* 12/03/2018 – 09/07/2019: 838 phim
* 09/07/2019 – 04/11/2020: 1023 phim
* 04/11/2020 – 03/03/2022: 1090 phim
* 03/03/2022 – 30/06/2023: 1150 phim

+ 13.5k phim không có lịch ra mắt (73%)

* Hours Viewed:

+ min: 100k giờ xem

+ max: 812m giờ xem

* Numbers of rating:

+ 14.2k phim có đánh giá (78%)

+ 4110 phim thiếu đánh giá (22%)

* Rating:

+ 14.2k phim có đánh giá (78%)

+ 4110 phim thiếu đánh giá (22%)

* + 1.2 (min)
  + 5.8 (25%)
  + 6.6 (50%)
  + 7.3 (75%)
  + 10 (max)
* Genre:

+ 15.8k phim có ghi thể loại (86%)

+ 2571 phim không ghi thể loại (14%)

* Thể loại phổ biến nhất “Comedy”(Hài kịch) 679 (6%)
* Các thể loại khác (80%)
* Key words

+ 12.8k phim có chứa các từ khóa (70%)

+ 5531 phim không có từ khóa (30%)

+ Các từ khóa phổ biến: tv special,stand up comedy,stand up comedy special

* Description:

+ 10.6k phim có phần mô tả (58%)

+ 7713 phim không có phần mô tả (42%)

+ Các mô tả phổ biến: Celebrate the enduring, powerful character of Special Agent Dana Scully, M.D, the scientific phenomenon that she inspired for young women all over the globe, and the importance of female representation both in front and behind the...

XỬ LÝ DỮ LIỆU ĐỂ ĐƯA VỀ DỮ LIỆU CHUẨN

* Xử Lý Thông Tin Về Thời Gian (Release Date):

+ Chuyển Đổi Lịch Ra Mắt:

* + Chuyển đổi định dạng lịch ra mắt thành dạng DateTime để thuận tiện cho việc xử lý và phân tích.
  + Có thể tạo một biểu đồ thời gian để theo dõi xu hướng ra mắt phim qua các năm.
* Xử Lý Dữ Liệu Về Thể Loại (Genre):

+ One-Hot Encoding:

* Sử dụng kỹ thuật one-hot encoding để chuyển đổi thông tin về thể loại thành các biến nhị phân (0 hoặc 1) cho từng thể loại.
* Tạo một dataframe mới với các cột thể loại nhị phân.
* Xử Lý Dữ Liệu Về Đánh Giá và Đánh Giá Rating:

+ Xử Lý Dữ Liệu Thiếu:

* + Đối với những phim thiếu đánh giá, có thể điền giá trị trung bình hoặc giá trị mặc định.

+ Chia Tầng Đánh Giá:

* + Chia tầng đánh giá thành các khoảng để tạo ra các biến nhóm.
* Xử Lý Dữ Liệu Về Key Words và Description:

+ Xử Lý Dữ Liệu Thiếu:

* + Đối với phim không có từ khóa hoặc mô tả, có thể thêm giá trị mặc định hoặc giá trị 'Không có thông tin'.

+ Phân Tích Key Words và Description:

* + Sử dụng kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) để trích xuất thông tin hữu ích từ key words và mô tả, ví dụ như từ vựng phổ biến, tần suất xuất hiện, v.v.
* Xử Lý Dữ Liệu Về Available Globally

+ Chuyển Đổi Dữ Liệu Thành Percentages:

* + Chuyển đổi số lượng phim có sẵn và không có sẵn thành tỷ lệ phần trăm để so sánh dễ dàng hơn.
* Xử Lý Dữ Liệu Về Số Giờ Xem (Hours Viewed):

+ Normalizing:

* + Chuẩn hóa dữ liệu về số giờ xem để đưa về khoảng giá trị thống nhất.