

Final Project – Nhóm 4

I.	Xử lý dữ liệu	2
1.	Bộ dataset sử dụng và công cụ xử lý dữ liệu	2
2.	Các bước xử lý dữ liệu	2
II.	Phân tích và nhận xét dữ liệu	2
1.	Các câu hỏi đặt ra	2
	Câu 1: Liệu phim có rank nhỏ (thứ hạng cao) thì sẽ nhận được nhiều lượt yêu thích (Favorites) và điểm số (Score)?	2
	Câu 2: Top 5 thể loại phổ biến nhất trong tập dataset này là gì?	2
	Câu 3: Top 10 nhà sản xuất phim ảnh hưởng như thế nào đến dataset? (top 10 nói về số lượng phim sản xuất)	2
	Câu 4: Nguồn phim chuyển thể ở đâu là nhiều nhất trong dataset?	2
	Câu 5: Lượng rating ảnh hưởng thế nào tới lượng favorites?	2
	Câu 6: Top 5 những studio có lượng phim nhiều nhất.	2
	Câu 7: Phim một tập hay phim dài tập được yêu thích hơn?	2
2.	Vẽ và phân tích dữ liệu	2
	Câu 1: Liệu phim có rank nhỏ (thứ hạng cao) thì sẽ nhận được nhiều lượt yêu thích (Favorites) và điểm số (Score)?	3
	Câu 2: Top 5 thể loại phổ biến nhất trong tập dataset này là gì?	4
	Câu 3: Top các nhà sản xuất phim ảnh hưởng như thế nào đến dataset? (top 10 nói về số lượng phim sản xuất)	5
	Câu 4: Nguồn phim chuyển thể ở đâu là được yêu thích nhiều nhất trong dataset?	6
	Câu 5: Lượng rating ảnh hưởng thế nào tới lượng favorites?	7
	Câu 6: Top 5 những studio có lượng phim nhiều nhất.	7
	Câu 7: Phim một tập hay phim dài tập được yêu thích hơn?	8
III.	Kết luận	8

I. Xử lý dữ liệu

1. Bộ dataset sử dụng và công cụ xử lý dữ liệu

+ Dataset: Anime.csv

+ Công cụ xử lý dữ liệu: Python

2. Các bước xử lý dữ liệu

+ Bước 1: Import thư viện Pandas và import dataset

+ Bước 2: Liệt kê các cột trong bộ dataset và tìm các cột có dòng mang giá trị Null.

+ Bước 3: Loại bỏ các cột không cần thiết trong việc phân tích dữ liệu

+ Bước 4: Thực hiện loại bỏ các dòng có giá trị Null hoặc các dòng có Favorites = 0.

+ Bước 5: Xuất file mới "anime_new.csv" và tiến hành phân tích dữ liệu.

II. Phân tích và nhận xét dữ liệu

1. Các câu hỏi đặt ra

Câu 1: Liệu phim có rank nhỏ (thứ hạng cao) thì sẽ nhận được nhiều lượt yêu thích (Favorites) và điểm số (Score)?

Câu 2: Top 5 thể loại phổ biến nhất trong tập dataset này là gì?

Câu 3: Top 10 nhà sản xuất phim ảnh hưởng như thế nào đến dataset? (top 10 nói về số lượng phim sản xuất)

Câu 4: Nguồn phim chuyển thể ở đâu là nhiều nhất trong dataset?

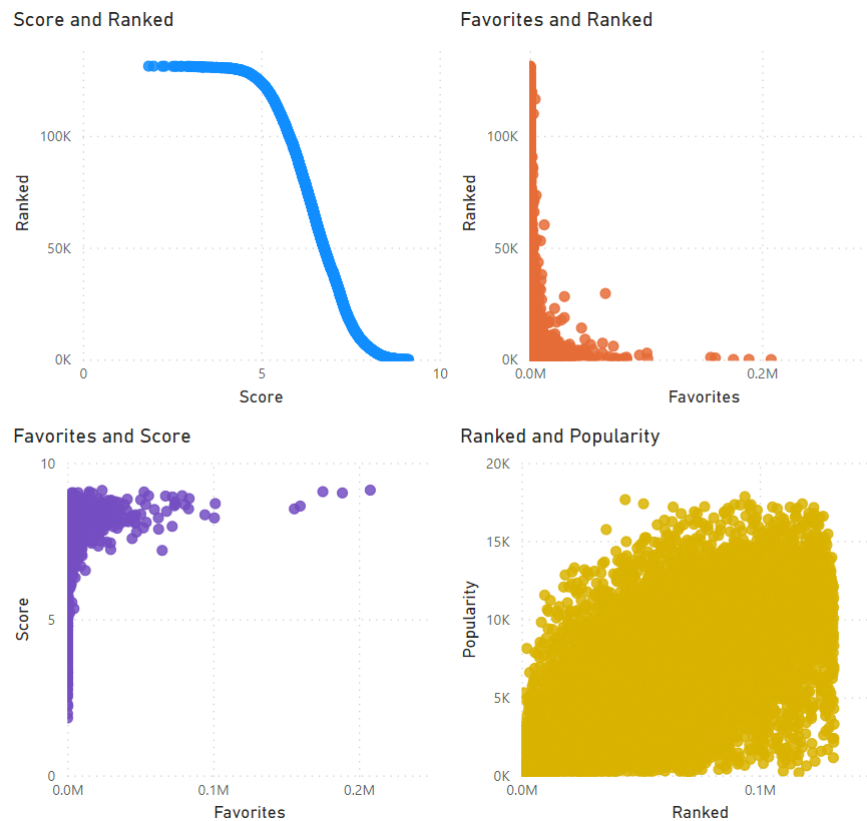
Câu 5: Lượng rating ảnh hưởng thế nào tới lượng favorites?

Câu 6: Top 5 những studio có lượng phim nhiều nhất.

Câu 7: Phim một tập hay phim dài tập được yêu thích hơn?

2. Vẽ và phân tích dữ liệu

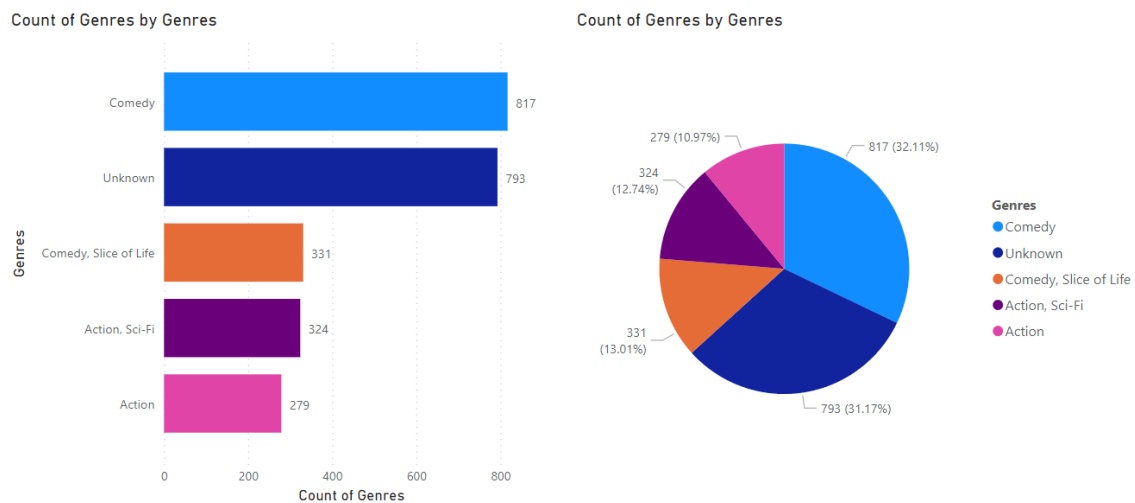
Câu 1: Liệu phim có rank nhỏ (thứ hạng cao) thì sẽ nhận được nhiều lượt yêu thích (Favorites) và điểm số (Score)?



Theo biểu đồ ta thấy, do rank được lấy dựa trên score người dùng chấm nên mối quan hệ của 2 chỉ tiêu này tỷ lệ thuận với nhau, khác với chỉ tiêu rank và popularity, trên biểu đồ ta thấy 2 chỉ tiêu này không có mối tương quan tỷ lệ thuận. Do đó ta có thể kết luận rằng các bộ phim có rank nhỏ (thứ hạng cao) chưa chắc có là bộ phim được các người dùng cho vào list của họ nhiều nhất hay có độ phổ biến (popularity) cao nhất nhưng chắc chắn có điểm score cao. Chỉ tiêu Favorites theo biểu đồ ta thấy có mối tương quan gần trùng với Score và rank (điểm cao, rank nhỏ => nhiều favorites).

Ta có thể kết luận những bộ phim được chấm điểm cao và thứ hạng cao cũng gần trùng với các bộ phim người dùng cho vào list yêu thích của họ.

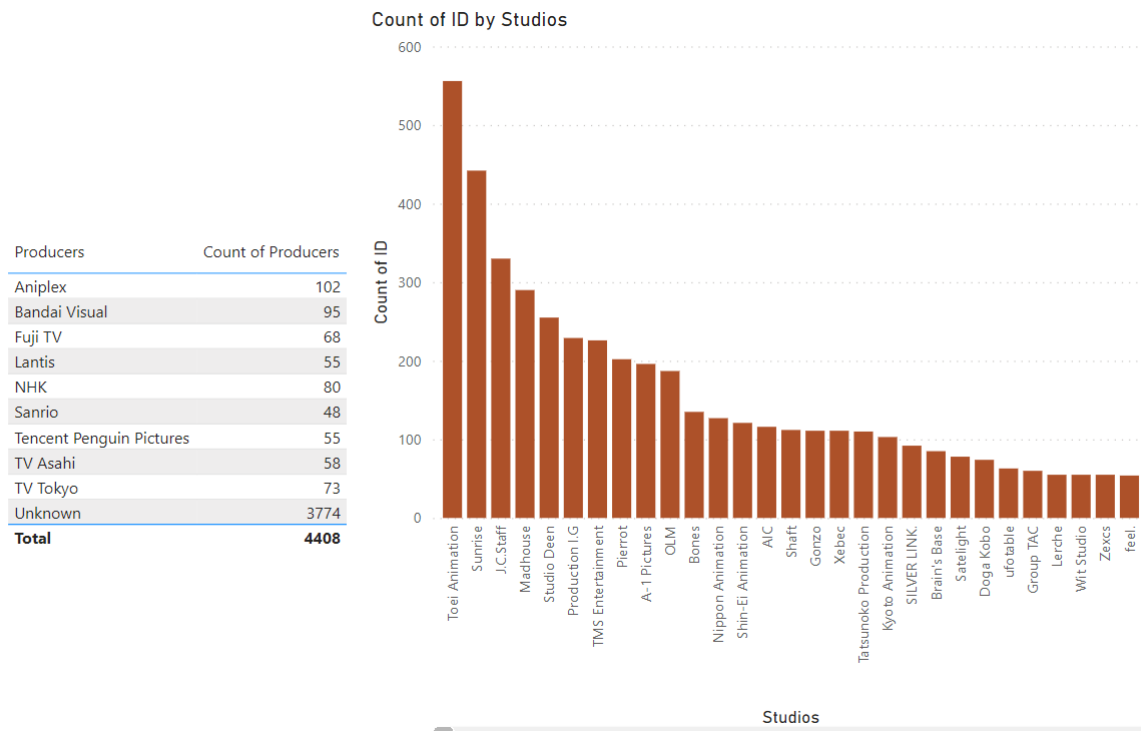
Câu 2: Top 5 thể loại phổ biến nhất trong tập dataset này là gì?



Theo biểu đồ cho thấy, nếu không xét đến những bộ phim không xác định được thể loại (unknown), thì thể loại phổ biến nhất là những bộ phim có yếu tố hài (comedy), số lượng cao nhất là comedy và comedy, slice of life, sau đó đến yếu tố hành động (action) (comedy nhiều hơn hành động 1%).

Ta có thể kết luận đây là 2 thể loại được các nhà làm phim sản xuất nhiều nhất. Tuy nhiên, theo độ yêu thích của người dùng thể loại action là thể loại được yêu thích nhất, rồi mới đến comedy.

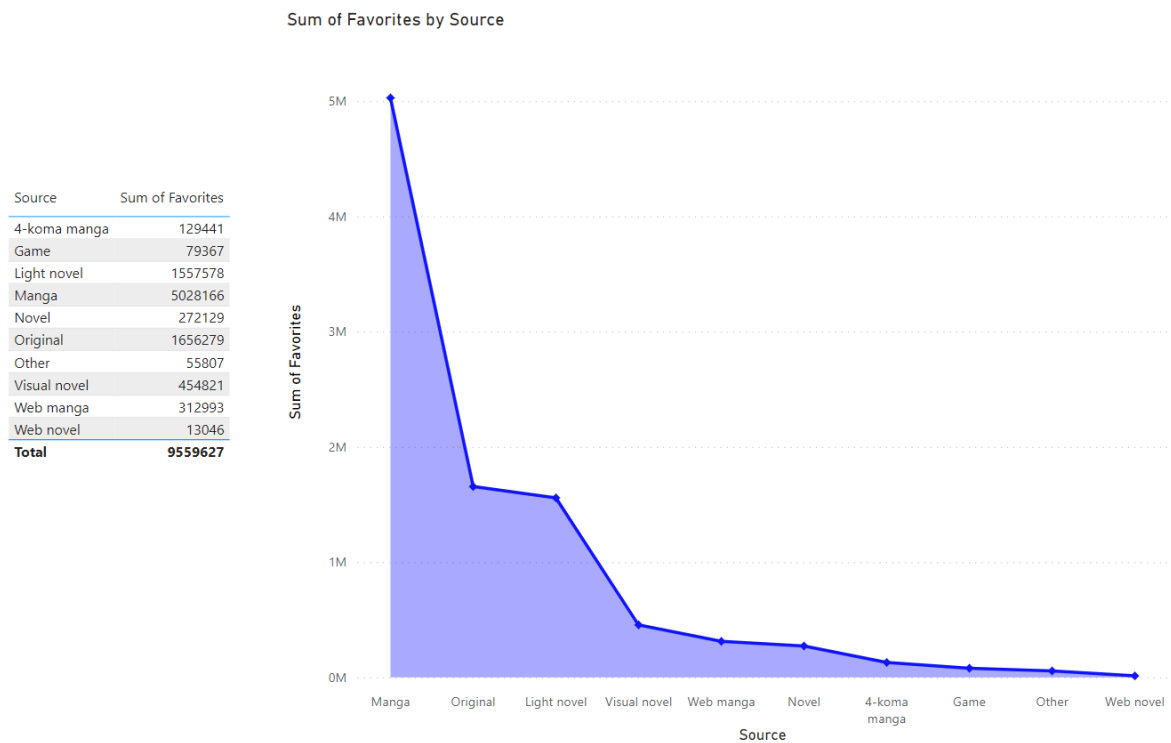
Câu 3: Top các nhà sản xuất phim ảnh hưởng như thế nào đến dataset? (top 10 nói về số lượng phim sản xuất)



Top các có nhà sản xuất Aniplex sản xuất nhiều phim nhất tuy nhiên trong top 10 nhà sản xuất có phim đạt favorite cao nhất, Aniplex chỉ đứng thứ 2 nhưng cũng không phải nhà sản xuất có điểm rank cao nhất. Nhà sản xuất có lượng Favorite cao nhất là Vap, Nippon Television Network và điểm rank cao nhất là các bộ phim không xác định được nhà sản xuất và các bộ phim của Bandai Visual, nhà sản xuất này cũng đứng thứ 2 trong top 10 sản xuất nhiều phim nhất.

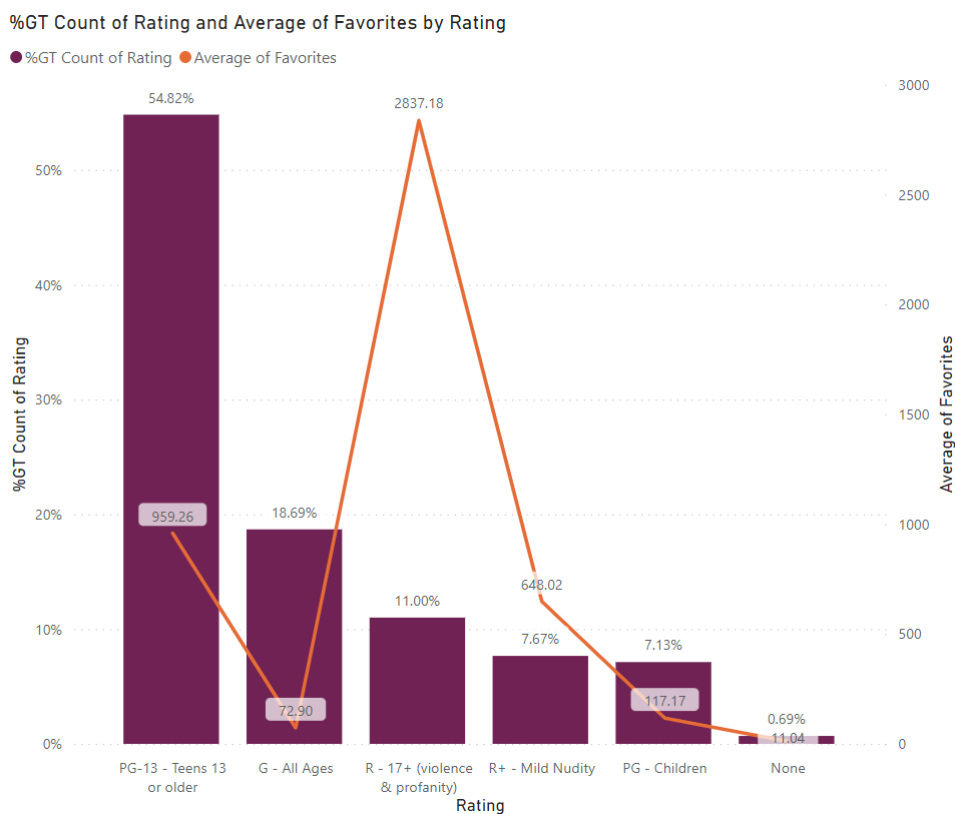
Do đó ta có thể kết luận khi chọn nhà sản xuất ta có thể chọn Aniplex (nhờ được yêu thích gần nhất) và Bandai Visual.

Câu 4: Nguồn phim chuyên thể ở đâu là được yêu thích nhiều nhất trong dataset?



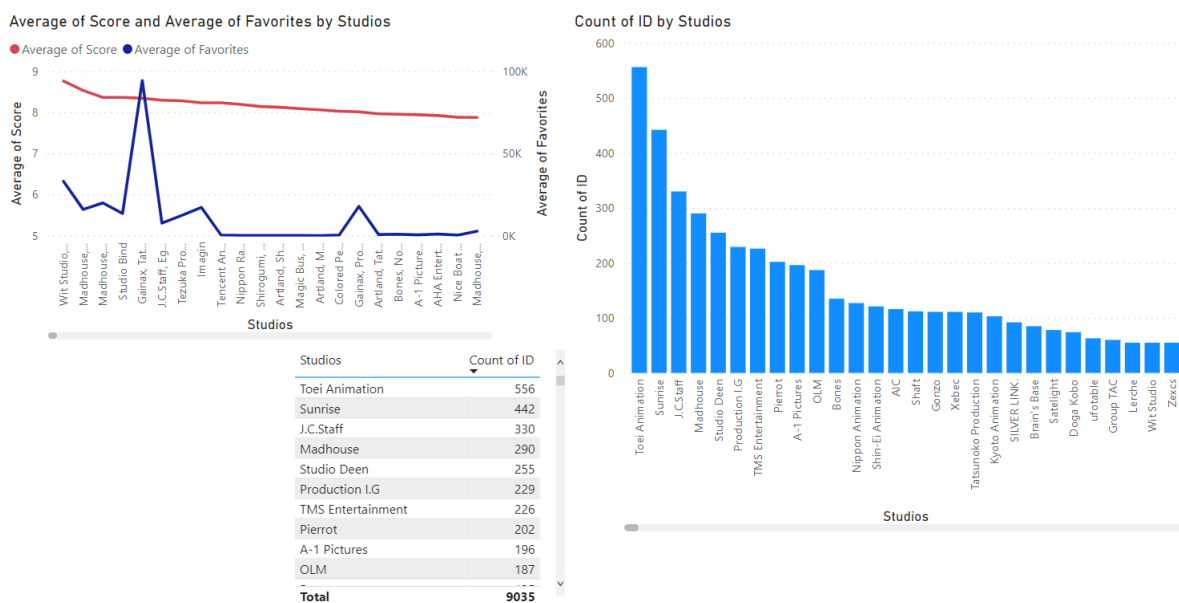
Nguồn từ manga được yêu thích nhất và lượng yêu thích cao gần gấp đôi lượng yêu thích của phim từ nguồn được yêu thích thứ 2 là original và thứ 3 là light novel.

Câu 5: Lượng rating ảnh hưởng thế nào tới lượng favorites?



PG-13 and older là rating có lượng phim nhiều nhất và nhưng được yêu thích nhất lại là dòng phim có rating R – 17.

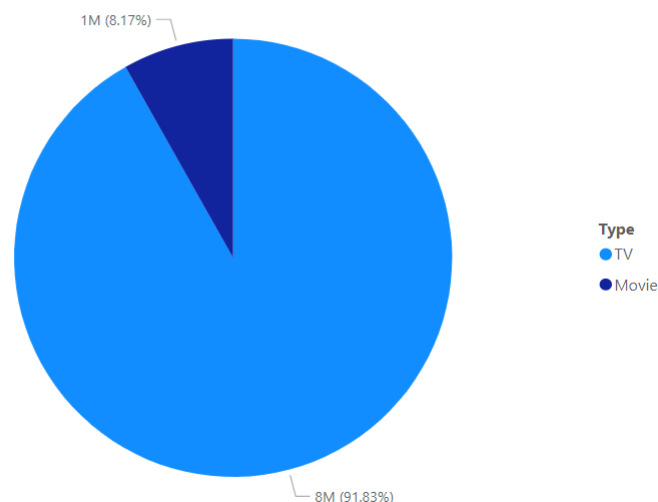
Câu 6: Top 5 những studio có lượng phim nhiều nhất.



Studio có lượng phim nhiều nhất là Tobi animation và đứng thứ 2 là Sunshine tuy nhiên nhận được nhiều yêu thích và score cao là những bộ phim có studio gồm ít lượng phim.

Câu 7: Phim một tập hay phim dài tập được yêu thích hơn?

Sum of Favorites by Type



Giữa hai dòng phim thì phim dài tập (chiếu trên TV) được yêu thích hơn (tổng Favorites là 8 triệu, chiếm hơn 91%) và phim chiếu rạp chỉ chiếm phần nhỏ.

Vậy phim dài tập (TV) được người xem yêu thích hơn.

III. Kết luận

Để sản xuất bộ phim đạt doanh thu cao, nên tập trung vào thể loại action hoặc comedy vì 2 thể loại này có lượng yêu thích cao nhất. Nguồn có thể từ manga hoặc original và chọn producer là Aniplex hoặc Bandai Visual, Studio là Tobi Animation hoặc Sunshine, rating nên là PG-13, độ tuổi có lượng phim được yêu thích nhiều nhất và lượng ID người dùng cũng nhiều nhất.

Title	Popularity	Favorites	Score	Ranked	Producers	Type	Source	Genres	Rating
Fullmetal Alchemist: Brotherhood	3	207772	9.13	12	Aniplex, Square Enix, Mainichi Broadcasting System, Studio Moriken	TV	Manga	Action, Adventure, Drama, Fantasy	R - 17+ (violence & profanity)
Kaguya-sama wa Kokurasetai: Ultra Romantic	261	23770	9.12	22	Aniplex, Mainichi Broadcasting System, Magic Capsule, Shueisha, JR East Marketing & Communications	TV	Manga	Comedy, Suspense	PG-13 - Teens 13 or older
Gintama*	337	14967	9.08	32	TV Tokyo, Aniplex, Dentsu	TV	Manga	Action, Comedy, Sci-Fi	PG-13 - Teens 13 or older
Steins;Gate	13	175248	9.08	42	Frontier Works, Media Factory, Movie, AT-X, Kadokawa Pictures Japan, Nitroplus	TV	Visual novel	Drama, Sci-Fi, Suspense	PG-13 - Teens 13 or older
Shingeki no Kyojin Season 3 Part 2	28	52595	9.07	52	Production I.G, Dentsu, Mainichi Broadcasting System, Pony Canyon, Kodansha, Techno Sound,	TV	Manga	Action, Drama	R - 17+ (violence & profanity)

Dựa vào visualize và phân tích, phim có lượng Popularity và Ranked cao, lượng người yêu thích lớn dựa theo dòng phim, thể loại, và rating thì phim **Fullmetal Alchemist: Brotherhood** là phù hợp nhất. Bộ phim có lượng yêu thích nhất và có độ phổ biến thứ 3 và thuộc nguồn manga, thể loại action và đến từ nhà sản xuất Aniplex.