

DATA SET PREPARATION

Project Name: AGENTBASED Worth2Watch (W2W) DATA ACQUISITION SOFTWARE

- 1- 201805008-Vakkas Karakurt
- 2- 201805020-Begüm İş
- 3- 201805049-Ayşegül Çağlı (MAIN CONTACT)
- 4- 201805070-Emrah Fidan (SCRUM MASTER)
- 5- 221805117-Kürşat Edabalı Yıldırım

- ✓ **Data Acquisition:** We pulled IMDB ids by sending TMDB API requests.
- ✓ **Data Scraping:** Utilizing these IDs, we performed web scraping using BeautifulSoup to extract movie title (MovieTitle), review ID (ReviewID), review text (Review), the number of people who found it helpful (Found Helpful), and the rating (Rating).
- ✓ **Found Helpful Rate Calculation:** By dividing the count of those who found it helpful by the count of individuals who evaluated the review, we obtained the Found Helpful Rate. In other words, we divided the total number of help actions on a review by the number of people who assessed that review.
- ✓ **Rating Adjustment:** We adjusted the rating values from the default range of 1-10 to the range of 0.1–1. This means we compressed the rating scores into a smaller range.

Our dataset consists of 93,450 reviews, and we can fetch more comments if needed

Input 1				Input 2	Input 3	Output 1	
movie_id	Movie_Title	genres	id	review	found_helpful	found_helpful_rate	rating
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8168012	Wow. The first Top Gun is a classic, and as v	2736	0.88	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8177338	If you were a late teen or in your early twe	5330	0.95	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8173679	In one of the more memorable lines in the	1223	0.85	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8178255	This is one sequel that looked like it woul	983	0.85	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8177954	I was reluctantly dragged into the theater,	1885	0.9	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8189765	If there's any movie that deserves to be se	1071	0.86	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8174934	What an excellent sequel - I, in fact, like it	658	0.8	0.9
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8178881	Top Gun (1986) made Tom Cruise a star, an	641	0.77	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8186445	5 out of 5 stars.Top Gun Maverick is an awe	399	0.75	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8184290	See? This is why I LOVE Tom Cruise Post-	406	0.74	0.9
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8205469	There are so many references to the origin	510	0.78	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8190055	It makes more than sense to have seen the	354	0.72	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8173474	This is one of the best theatrical experienc	617	0.81	0.9
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8182413	Having loved Top Gun, I clearly wasn't goin	292	0.72	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8248584	I don't share everyone's unbridled enthusi	577	0.74	0.7
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8172523	Greetings again from the darkness. Heavy c	298	0.71	0.8
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8185617	I appreciate the first film but never been a	372	0.75	0.9
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8197359	Greetings from Lithuania.I wasn't the bigge	212	0.7	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8192822	Tom Cruise is still an ultimate being at 59 y	209	0.69	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8482010	Based on the scores, I was expecting a	381	0.68	0.6
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8680202	Firstly before my small complaint I will poi	68	0.73	0.7
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8183145	I'll preface my review by first stating that I	636	0.85	0.9
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8177919	Wow! As a retired combat veteran who has	250	0.69	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8173363	Just Saw This In IMAX and guys what an exp	275	0.68	1
tt1745960	Top Gun: Maverick	['Action', 'Drama']	rw8253750	Wow!! I have just walked out of the cinema,	116	0.66	1
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8363576	Bad script, bad acting, stupid story, a new c	663	0.85	0.2
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8207383	This movie consists mostly of us watching j	1557	0.85	0.3
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8222042	If I were to list out all the things this movie	977	0.81	0.5
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8265997	Before I went to this film (because my Gra	377	0.78	0.1
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8242296	-Dinosaurs that only survived in warm clir	1029	0.85	0.6
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8229362	When I go to a Jurassic Park movie, I want s	674	0.85	0.3
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8223771	Honestly, I could have adapted any numbe	733	0.84	0.5
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8224696	Colin Trevorrow was so preoccupied with v	455	0.79	0.2
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8215815	This movie attempts to exploit the nostalg	644	0.75	0.6
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8223092	It's just amazing that no dinosaur hasn't bit	497	0.75	0.4
tt8041270	Jurassic World Dom	['Action', 'Adventu	rw8216225	I don't know whether this is a fabled mem	589	0.77	0.2

Movie ID: Each film has a unique identifier. This identifier is used to distinguish the film from others.

Movie Title: It is the full name of the movie.

Genres: Defines the genres of movies. We will use movie genres to classify our dataset.

Review ID: Each review has a unique identifier. This identifier is used to distinguish the review from others.

Review: It is the user's comment about the movie. Comments are in text format and may include opinions on elements such as the content, acting, direction, visuals, etc., of the film.

Found Helpful: It is the count of users who found the comment helpful. This number indicates how much the comment is liked by other users.

Found Helpful Rate: It is the rate of users who found the comment helpful. This rate indicates the overall level of appreciation for the comment. The ratio is calculated by dividing the number of users who found the comment helpful by the total number of users who evaluated the comment. The ratio takes a value between 0 and 1. A value close to 1 signifies that the comment is highly liked.

Rating: It is the rating given by the user who made the comment for the movie. Ratings range from 0.1 to 1, where 0.1 is the lowest and 1 is the highest rating.