

IMDB Movie Analysis

PROJECT DESCRIPTION

We are a production company, and we analyse data from our films to track our progress over time. We have information on the top-rated films from the preceding 100 years, as well as various other data, in this project. We are calculating profit using our gross revenue and the budget of the film.

We also want to know which of our top films had positive viewer feedback and performed well commercially. We'd also like to know who our top favourite directors are, as they're among our most popular and promising. We want to know which genres are popular among our target demographic. In order to uncover answers, we are looking at our top films, directors, genres, and financial results.

APPROACH

We are pursuing techniques such as what are our top movies that have received strong ratings from audiences, directors with good imdb ratings, and emphasizing the top ten. We are also discovering popular genres, favourite performers, and favourite films among the public. We are answering questions such as what are the top movies, how much profit we are generating from these movies, which movie had the highest rating, and which filmmaker may be named best director based on their imdb rating. To gather deeper insights, we are conducting a 5 whys study on our project data.

TECH-STACK USED

Microsoft Excel is a Microsoft software tool that uses spreadsheets to arrange numbers and data using formulas and functions. Microsoft Excel is a robust data visualization and analysis program that employs spreadsheets to store, organize, and track data sets using formulae and functions.

INSIGHTS

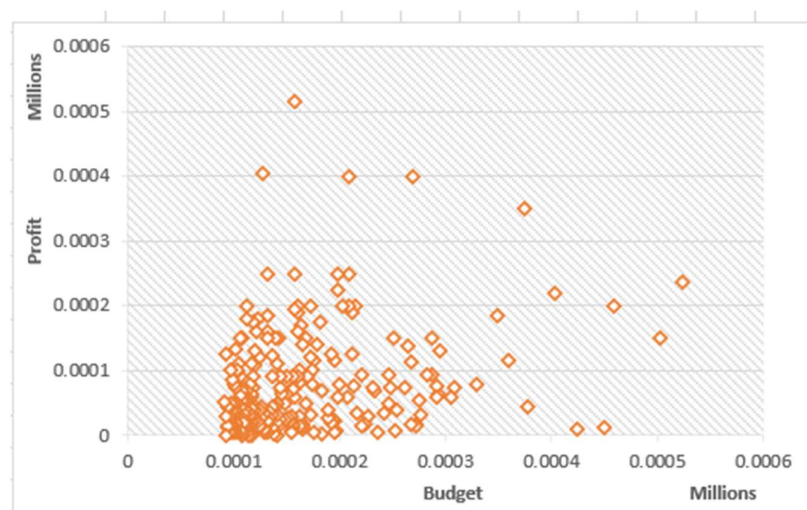
We must first clean up the noisy data we have and only maintain the columns that are necessary for our research. Therefore, we begin by looking for duplicate values, and if any are found, we must get rid of them. Then, we'll look for any null values and eliminate them as well. After reducing part of the data, we must now determine which columns we actually need, and after choosing the final columns, we are deleting any columns that are no longer necessary or only minimally necessary for analysis. Additionally, we are including more calculated custom columns in our data.

Once we have the necessary, clean data, we analyse it as follows:

1. Movies with Highest Profit

We computed our profit for each film by deducting the film's budget from its gross revenue. After assessing profit, we select films with a large profit margin for our production company. The movies with the highest profit are shown below, and we are analysing profit vs budget using a scatter plot to see the difference and detect outliers.

Movie	TotalProfit	TotalBudget
Avatar	523.51	237
Jurassic World	502.18	150
Titanic	458.67	200
Star Wars: Episode IV - A New Hope	449.94	11
E.T. the Extra-Terrestrial	424.45	10.5
The Avengers	403.28	220
The Lion King	377.78	45
The Jungle Book	375.3	350
Star Wars: Episode I - The Phantom Menace	359.54	115
The Dark Knight	348.32	185
The Hunger Games	330	78
Twilight	308.9	74
Deadpool	305.02	58
The Hunger Games: Catching Fire	294.65	130
Jurassic Park	293.78	63
Despicable Me 2	292.05	76
American Sniper	291.32	58.8
Finding Nemo	286.84	94
Shrek 2	286.47	150
The Lord of the Rings: The Return of the King	283.02	94
Star Wars: Episode VI - Return of the Jedi	276.63	32.5
Forrest Gump	274.69	55
Star Wars: Episode V - The Empire Strikes Back	272.16	18
Juno	271.98	15
Alice in Wonderland	268.38	400
Home Alone	267.76	18
Star Wars: Episode III - Revenge of the Sith	267.26	113
Spider-Man	264.71	139
Minions	262.03	74



2. Top 250 Movies with High IMDB Score

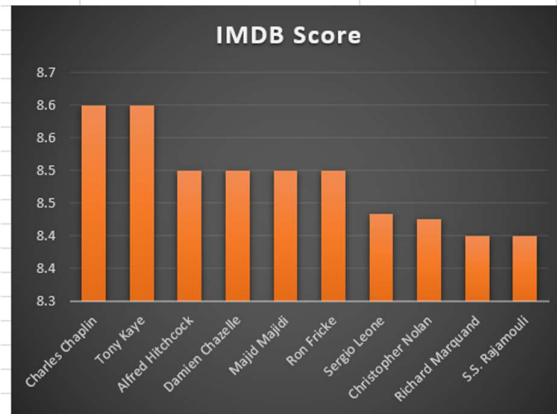
We are listing the top 250 films with high IMDb ratings as well as films with more than 25000 user votes. The top films according to user ratings and votes are shown below, with their ranking determined by the total number of votes.

Movie	NumberOfVotes	Ratings	Rank
The Shawshank Redemption	1689764	9.3	1
The Dark Knight	1676169	9	2
Inception	1468200	8.8	3
Fight Club	1347461	8.8	4
Pulp Fiction	1324680	8.9	5
Forrest Gump	1251222	8.8	6
The Lord of the Rings: The Fellowship of the Ring	1238746	8.8	7
The Matrix	1217752	8.7	8
The Lord of the Rings: The Return of the King	1215718	8.9	9
The Godfather	1155770	9.2	10
The Dark Knight Rises	1144337	8.5	11
The Lord of the Rings: The Two Towers	1100446	8.7	12
Skyfall	1044078	15.6	13
Se7en	1023511	8.6	14
The Avengers	995415	8.1	15
Gladiator	982637	8.5	16
Batman Begins	980946	8.3	17
Django Unchained	955174	8.5	18
King Kong	948065	21.6	19
Casino Royale	940984	16	20

3. Best Director

Based on their imdb ratings, we have chosen the top director out of the top 10 filmmakers.

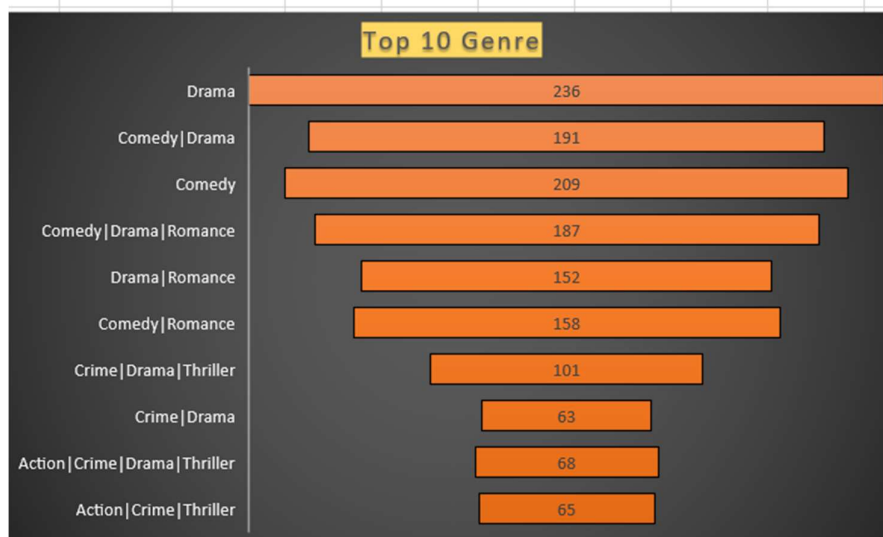
Director Name	IMDB Score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.4
Christopher Nolan	8.4
Richard Marquand	8.4
S.S. Rajamouli	8.4



4. Popular Genres

There are several genres available, but what our audience prefers is very important. Based on their IMDB scores and the number of films in that genre, we have chosen the top genres.

Genre	TotalMovieCount	IMDB_Rating
Drama	236	1634.6
Comedy Drama	191	1254.6
Comedy	209	1224
Comedy Drama Romance	187	1220.5
Drama Romance	152	1054.9
Comedy Romance	158	936.4
Crime Drama Thriller	101	683.4
Crime Drama	63	462.7
Action Crime Drama Thriller	68	439.8
Action Crime Thriller	65	412.8



5. Popular Actor

Audiences adore their leading men, and some films' box office success is sometimes solely attributable to them. In order to determine which of these performers has movies that have garnered great praise from reviewers and audience members, we have chosen our three favourite lead actors: Brad Pitt,

Meryl Streep, and Leonardo DiCaprio. And using this figure, we may determine who is our best actor in the eyes of the critics.

First, we compiled a list of the movies in which each actor portrayed the lead part.

Meryl Streep	Brad Pitt	Leonardo DiCaprio
A Prairie Home Companion	Babel	Blood Diamond
Hope Springs	By the Sea	Body of Lies
It's Complicated	Fight Club	Catch Me If You Can
Julie & Julia	Fury	Django Unchained
Lions for Lambs	Interview with the Vampire: The Vampire Chronicles	Gangs of New York
One True Thing	Killing Them Softly	Inception
Out of Africa	Mr. & Mrs. Smith	J. Edgar
The Devil Wears Prada	Ocean's Eleven	Marvin's Room
The Hours	Ocean's Twelve	Revolutionary Road
The Iron Lady	Seven Years in Tibet	Romeo + Juliet
The River Wild	Sinbad: Legend of the Seven Seas	Shutter Island
	Spy Game	The Aviator
	The Assassination of Jesse James by the Coward Robert Ford	The Beach
	The Curious Case of Benjamin Button	The Departed
	The Tree of Life	The Great Gatsby
	Troy	The Man in the Iron Mask
	True Romance	The Quick and the Dead
		The Revenant
		The Wolf of Wall Street
		Titanic

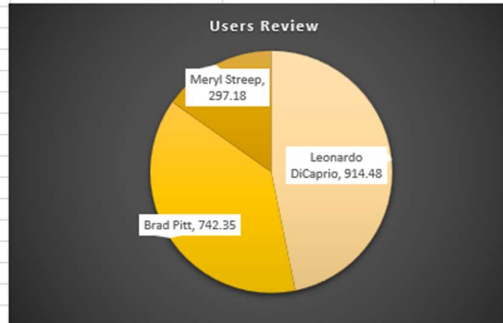
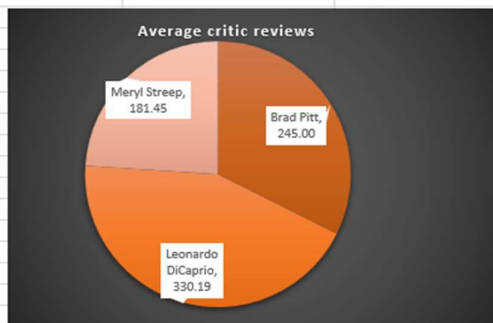
The names of all the movies are then combined, and each actor's name is used to classify them.

Combined	Actor1_Name
A Prairie Home Companion	Meryl Streep
Hope Springs	Meryl Streep
It's Complicated	Meryl Streep
Julie & Julia	Meryl Streep
Lions for Lambs	Meryl Streep
One True Thing	Meryl Streep
Out of Africa	Meryl Streep
The Devil Wears Prada	Meryl Streep
The Hours	Meryl Streep
The Iron Lady	Meryl Streep
The River Wild	Meryl Streep
Babel	Brad Pitt
By the Sea	Brad Pitt
Fight Club	Brad Pitt
Fury	Brad Pitt
Interview with the Vampire: The Vampire Chronicles	Brad Pitt
Killing Them Softly	Brad Pitt
Mr. & Mrs. Smith	Brad Pitt
Ocean's Eleven	Brad Pitt
Ocean's Twelve	Brad Pitt
Seven Years in Tibet	Brad Pitt
Sinbad: Legend of the Seven Seas	Brad Pitt
Spy Game	Brad Pitt
The Assassination of Jesse James by the Coward Robert Ford	Brad Pitt
The Curious Case of Benjamin Button	Brad Pitt
The Tree of Life	Brad Pitt
Troy	Brad Pitt
True Romance	Brad Pitt
Blood Diamond	Leonardo DiCaprio
Body of Lies	Leonardo DiCaprio
Catch Me If You Can	Leonardo DiCaprio
Django Unchained	Leonardo DiCaprio
Gangs of New York	Leonardo DiCaprio
Inception	Leonardo DiCaprio
J. Edgar	Leonardo DiCaprio
Marvin's Room	Leonardo DiCaprio
Revolutionary Road	Leonardo DiCaprio

We then determined the average user and critic reviews for each actor, as well as the number of lead roles each actor has played, using the data from above.

Critics Review		
Actor Name	Average of num_critic_for_review	Count of movie_title
Brad Pitt	245.00	17
Leonardo DiCaprio	330.19	21
Meryl Streep	181.45	11
Grand Total	267.24	49

Users Review		
Actor Name	Average of num_user_for_reviews	Count of movie_title
Leonardo DiCaprio	914.48	21
Brad Pitt	742.35	17
Meryl Streep	297.18	11
Grand Total	716.18	49



6. Movies by Decade

When it comes to film or fashion, each decade has its own preferences. We are categorizing our films by decade and tallying user votes for each decade's films. This will assist us understand how much our audience has increased throughout the years, where we have interacted with our audience the least, and which decades have seen the most support for us.

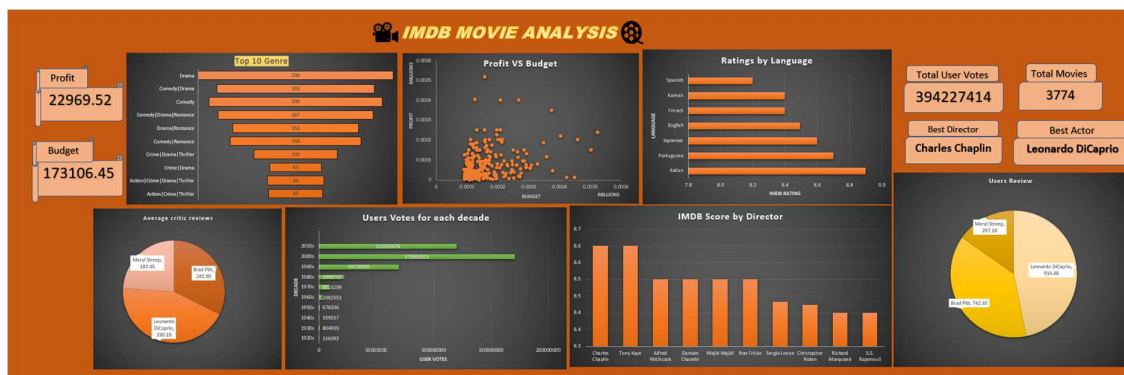
Decade	Year	Users Votes
1920s	1920-1929	116392
1930s	1930-1939	804839
1940s	1940-1949	159517
1950s	1950-1959	678336
1960s	1960-1969	2982551
1970s	1970-1979	8523299
1980s	1980-1989	19987476
1990s	1990-1999	69720305
2000s	2000-2009	170859021
2010s	2010-2019	120395678

Additionally, we determined which decade had the most reviews for our top favourite actors.

Critics Review				
Actor	1980s	1990s	2000s	2010s
Brad Pitt		633	1997	1535
Leonardo DiCaprio		612	1891	4431
Meryl Streep	66	106	1259	565

Users Review				
Actor	1980s	1990s	2000s	2010s
Brad Pitt		3953	6561	2106
Leonardo DiCaprio		3565	6568	9071
Meryl Streep	200	181	2360	528

RESULTS



Following a successful examination of our movie data, we were able to determine the following findings, including the 5 Whys of production, the characteristics of our audience, the most popular genres, and the finest actor performances.

1. The first question we ask is, "What is our profit for each movie?" As we can see from our findings, **"Avatar"** has the most **profit (523 million)**, whereas **"The Host"** hasn't generated any.
2. **'The Shawshank Redemption'** has an overall **9.3** imdb rating with **16,89,764 votes**. We are determining which movies received high votes and their imdb scores, and we have chosen our top 250 movies that have received high ratings. We can infer that this film has been a crowd favourite.
3. But when it comes to only good ratings, the movie **"King Kong"** comes out on top with an overall IMDB score of **21.6** with **profit of 33.15 million**, ranking **19th** overall based on the votes this movie received.
4. Our Total profit is **22,963 million** and our total budget is **1,73,106 million**.
5. Additional to English-language movies **Japanese-language** films appear to be popular with viewers as evidenced by their **14,37,135 votes** and **91.5 IMDB score**.
6. The **drama with comedy genre** appears to be our audience's favourite, receiving the highest IMDB rating.
7. We chose three actors—Leonardo DiCaprio, Meryl Streep, and Brad Pitt—and counted all of the evaluations each actor has received from critics over the years. It appears that Leonardo DiCaprio is the best actor among them all because he received the most reviews from both critics and users, making him our top choice.
8. Based on the findings, we can conclude that not all films with high imdb ratings have made more money than films with average scores.
9. Profit, it is reasonable to state, is not based on IMDB rating.
10. After the 1990s, movie shows started to make money and have continued ever since. Leonardo is regarded as the best actor by critics and audiences because audiences adore their lead performers, which is why a movie's financial success depends on who plays the lead role.
11. Because drama movies have a high imdb score and action movies have a low imdb score, we might think about making both to see how much money we can make.