

IMDB MOVIE ANALYSIS

ADVANCE EXCEL



PROJECT DESCRIPTION

The goal of this project is to analyze a dataset of IMDB movies and draw insights from the data. The dataset includes various columns such as movie names, budgets, gross revenue, and IMDB ratings. To complete the project, you will need to use a combination of Excel formulas and SQL commands to clean and manipulate the data. You will be asked to complete specific tasks, such as identifying the movie with the highest profit or the top IMDB movies, as well as share your own insights by identifying any problems or trends in the data. You may also be asked to use charts and visualizations to present your findings. The overall objective of the project is to gain a better understanding of the movie industry by analyzing the data and drawing meaningful conclusions

APPROACH

In order to find the insight, Excel was used. Using Excel, we performed Exploratory Analysis first like understanding data columns, checking for missing data, checking and removing outliers, etc. After that, we performed various operations like selecting, sorting, applying different formulas, etc. to get the insights we needed.

TECH STACK USED

MS EXCEL to explore , analyse and visualize my data

PROJECT INSIGHTS

1 . Cleaning the data::

This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

Your task: Clean the data.

Ans: To clean the data, I arranged the columns in the correct format and increased the column width to improve readability. I also removed null values and duplicates by selecting them using the "Find & Select" option and deleting the entire row. This helped to ensure that the data was accurate and ready for analysis.

Before Cleaning: 5044 rows and 28 columns

After Cleaning: 3850 rows and 13 column

2. Movies with the highest profit: Find the movies with the highest profit.

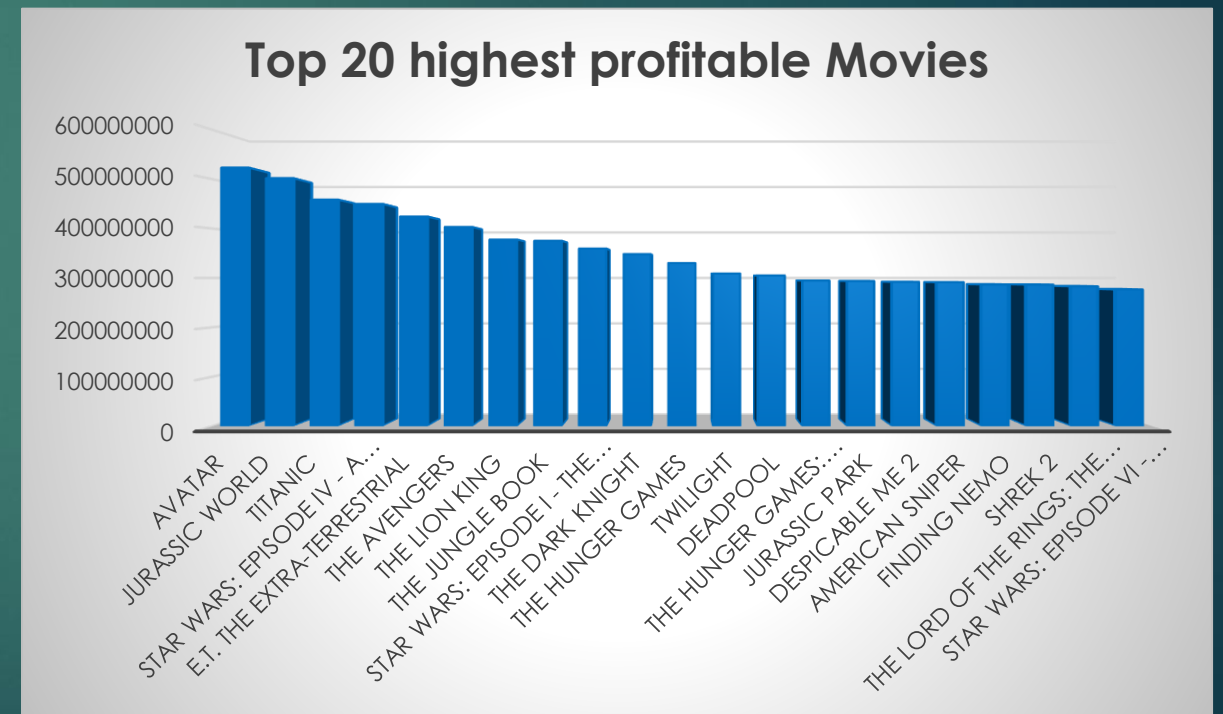
Excel Formula -

Created Profit column by subtracting gross with budget using formula-
=I2-W2

Created Pivot Table and found out the top 20 movies. Created Chart for the

Result-

Row Labels	Sum of PROFIT
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459
The Avengers	403279547
The Lion King	377783777
The Jungle Book	375290282
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Hunger Games	329999255
Twilight	308898950
Deadpool	305024263
The Hunger Games: Catching Fire	294645577
Jurassic Park	293784000
Despicable Me 2	292049635
American Sniper	291323553
Finding Nemo	286838870
Shrek 2	286471036
The Lord of the Rings: The Return of the King	283019252
Star Wars: Episode VI - Return of the Jedi	276625409



3. Top 250: Find IMDB top 250.

Excel Formula-

Applied filters based on the requirement and extracted columns which needed in a different sheet.

Result-

A new sheet has been created to show IMDB's top 250 movies with votes>25000 and IMDB's Top Foreign language film.

Rank	movie_title	num_voted_users	language	imdb_score
1	The Shawshank Redemption	1689764	English	9.3
2	The Godfather	1155770	English	9.2
3	The Dark Knight	1676169	English	9
4	The Godfather: Part II	790926	English	9
5	The Lord of the Rings: The Return of the King	1215718	English	8.9
6	Schindler's List	865020	English	8.9
7	Pulp Fiction	1324680	English	8.9
8	The Good, the Bad and the Ugly	503509	Italian	8.9
9	Inception	1468200	English	8.8
10	The Lord of the Rings: The Fellowship of the Ring	1238746	English	8.8
11	Fight Club	1347461	English	8.8
12	Forrest Gump	1251222	English	8.8
13	Star Wars: Episode V - The Empire Strikes Back	837759	English	8.8
14	The Lord of the Rings: The Two Towers	1100446	English	8.7
15	The Matrix	1217752	English	8.7
16	Goodfellas	728685	English	8.7
17	Star Wars: Episode IV - A New Hope	911097	English	8.7
18	One Flew Over the Cuckoo's Nest	680041	English	8.7
19	City of God	533200	Portuguese	8.7

Rank	movie_title	num_voted_users	imdb_score	language
1	The Good, the Bad and the Ugly	503509	8.9	Italian
2	City of God	533200	8.7	Portuguese
3	Seven Samurai	229012	8.7	Japanese
4	Spirited Away	417971	8.6	Japanese
5	Samsara	22457	8.5	None
6	The Lives of Others	259379	8.5	German
7	Children of Heaven	27882	8.5	Persian
8	Princess Mononoke	221552	8.4	Japanese
9	Das Boot	168203	8.4	German
10	Oldboy	356181	8.4	Korean
11	A Separation	151812	8.4	Persian
12	Metropolis	111841	8.3	German
13	Downfall	248354	8.3	German
14	The Hunt	170155	8.3	Danish
15	Howl's Moving Castle	214091	8.2	Japanese
16	The Act of Killing	23836	8.2	Indonesian
17	Pan's Labyrinth	467234	8.2	Spanish
18	The Secret in Their Eyes	131831	8.2	Spanish
19	Tae Guk Gi: The Brotherhood of War	31943	8.1	Korean

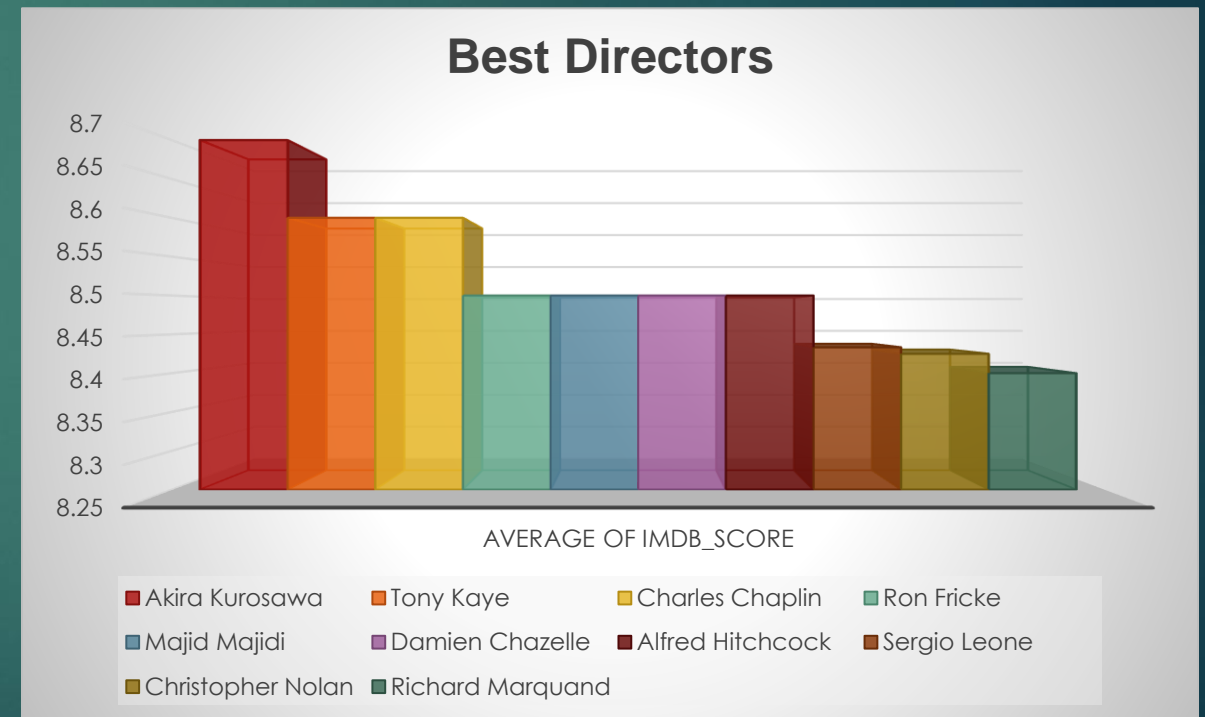
4. Best Directors: The top 10 directors for whom the mean of imdb_score is the highest. The best directors list is shown.

Excel Formula

Done using Pivot table. The grouping feature was used to find the mean of imdb score.

Result:-

Row Labels	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.43333333
Christopher Nolan	8.425
Richard Marquand	8.4



5. Popular Genres: Find the most popular genres.

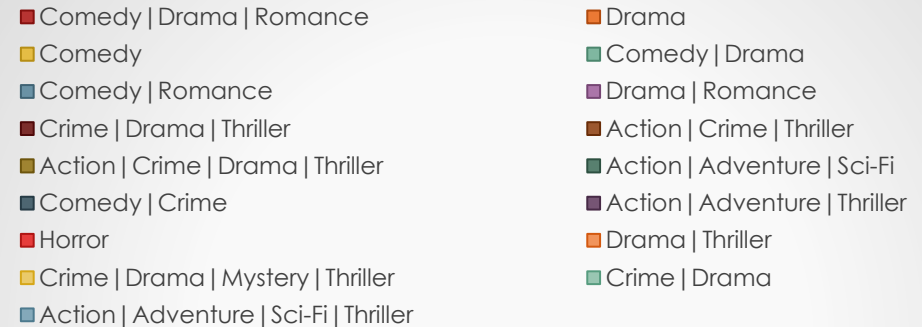
Excel Formula-

Done using Pivot table. The grouping feature was used to find the count of grouped genres.

Result:-

Row Labels	Count of genres
Comedy Drama Romance	147
Drama	140
Comedy	137
Comedy Drama	137
Comedy Romance	131
Drama Romance	114
Crime Drama Thriller	81
Action Crime Thriller	55
Action Crime Drama Thriller	50
Action Adventure Sci-Fi	46
Comedy Crime	45
Action Adventure Thriller	45
Horror	43
Drama Thriller	42
Crime Drama Mystery Thriller	41
Crime Drama	40
Action Adventure Sci-Fi Thriller	34

Most popular Grouped Genres



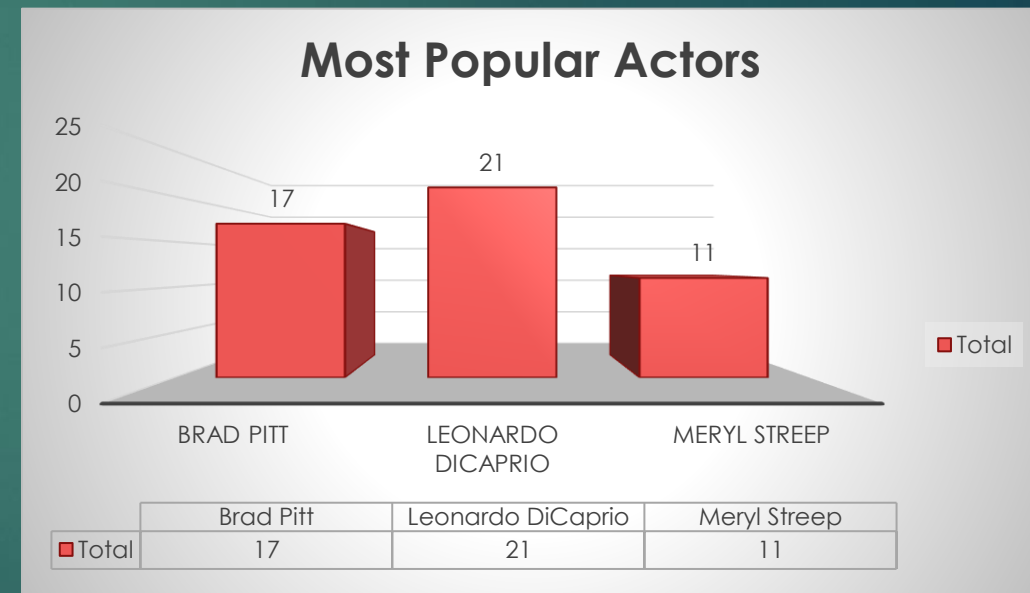
6. Charts: Used charts and graphs to find different-different insights.

Excel Formula-

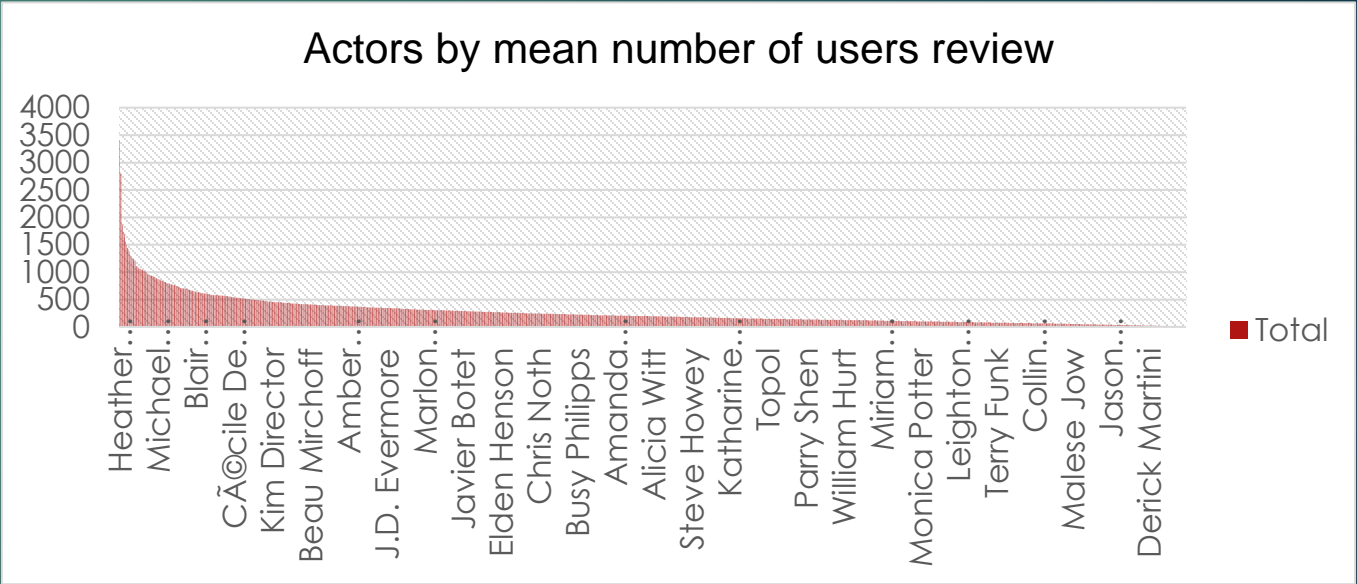
Pivot Tables and Graphs were used to find these insights. For finding the decade, I used the formula:

=LEFT(Y2,3)&"0s"

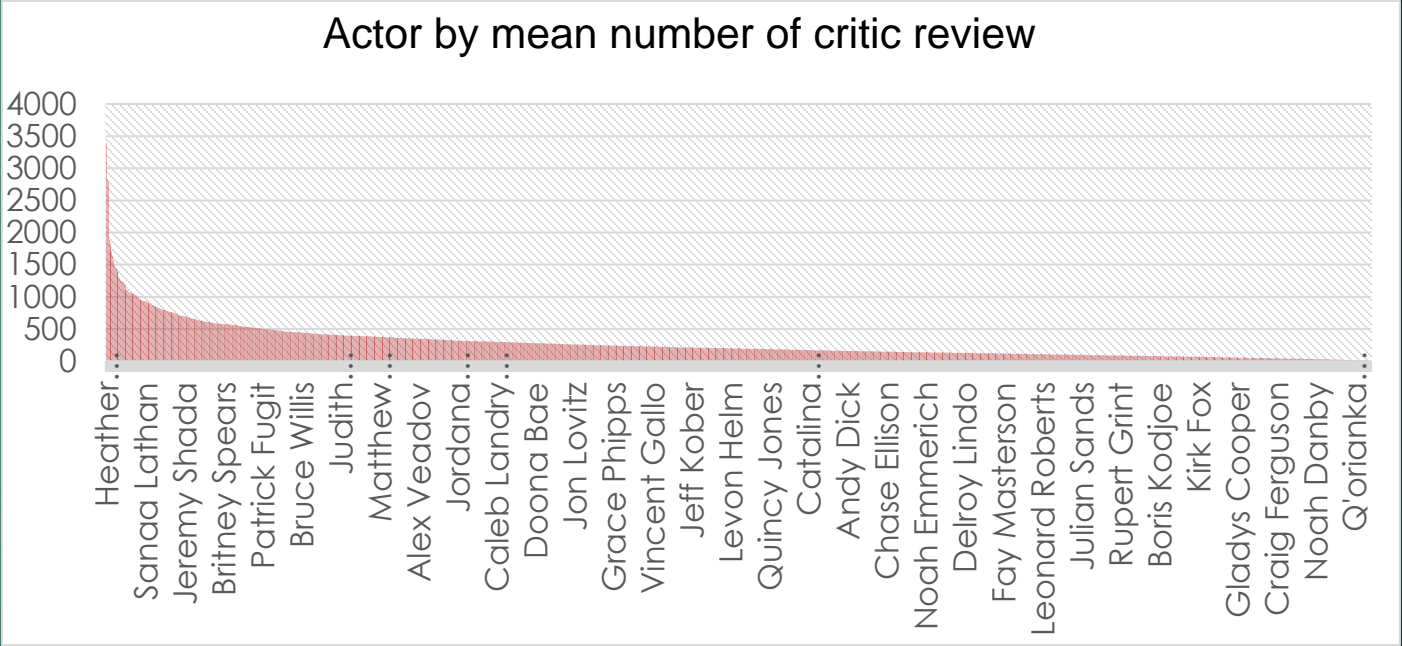
Row Labels	Count of actor_1_name
Brad Pitt	17
Leonardo DiCaprio	21
Meryl Streep	11
Grand Total	49



Row Labels	Average of num_user_for_reviews
Heather Donahue	3400
Christo Jivkov	2814
Steve Bastoni	2789
Phaldut Sharma	1885
Orlando Bloom	1842
Keir Dullea	1736
Eva Green	1708.333333
Chen Chang	1641
Nick Stahl	1562
Albert Finney	1498
Kevin Rankin	1445
Noah Huntley	1441
Osama bin Laden	1416
Seychelle Gabriel	1382
Mathieu Kassovitz	1314
Essie Davis	1285.5
Sharlto Copley	1262

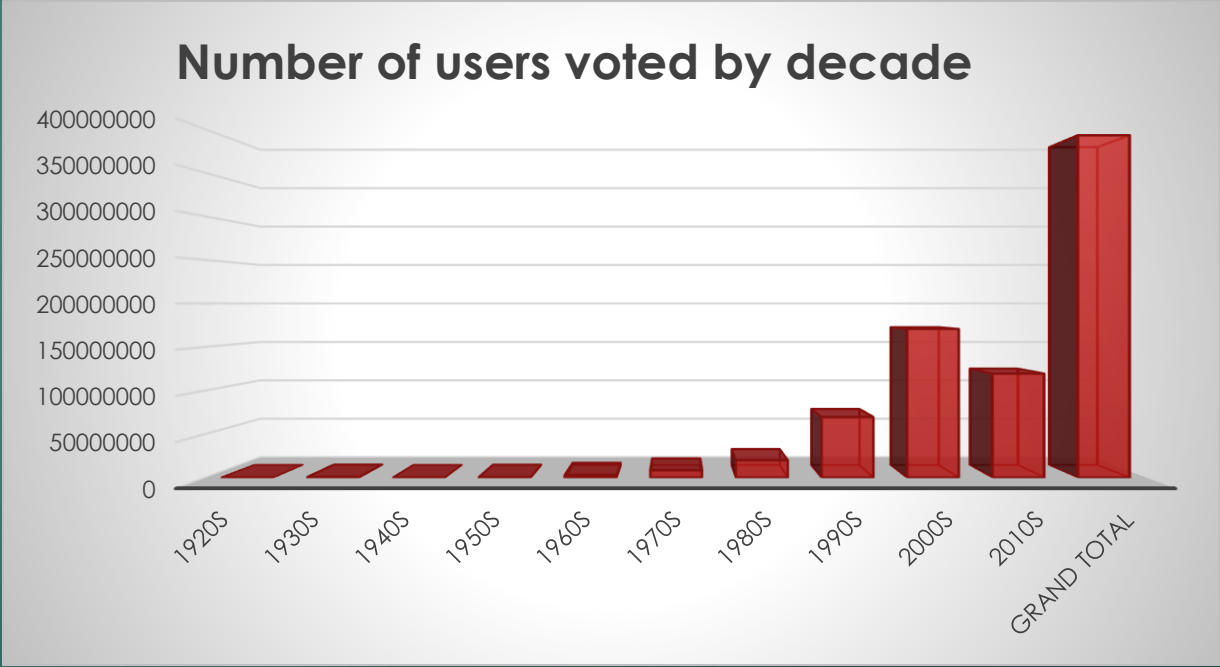


Row Labels	Average of num_critic_for_reviews
Albert Finney	750
Phaldut Sharma	738
Peter Capaldi	654
Craig Stark	596
BÃ©rÃ©nice Bejo	576
Suraj Sharma	552
Ellar Coltrane	548
Mike Howard	546
Lou Taylor Pucci	543
Joel Courtney	539
Maika Monroe	533
Tim Holmes	525
Elina Alminas	489
Kurt Fuller	487
Iko Uwais	481
QuvenzhanÃ© Wallis	478.6666667
Edgar Arreola	478
Sharlto Copley	472
Cory Hardrict	452
Matt Frewer	451
Isabelle Huppert	447
Elizabeth McGovern	447
Michael Fassbender	434.3333333
Wood Harris	432
Jennifer Lawrence	418.9285714
Anil Kapoor	418





Row Labels	sum of num_voted_users
1920s	116387
1930s	804839
1940s	159517
1950s	678336
1960s	2982551
1970s	8523299
1980s	19987476
1990s	69581866
2000s	170711435
2010s	119432961
Grand Total	392978667



PROJECT INSIGHTS

- ▶ It appears that movies like "Avatar" and "Jurassic Park" have the potential to earn high profits. If the goal is to maximize profit, it may be advisable to consider making movies with similar themes or characteristics.
- ▶ It appears that the movie "The Shawshank Redemption" has the highest IMDB score among those with a minimum of 25,000 voted users.
- ▶ From the top 250 IMDB movies, we can conclude that only 37 of them are not in the English language. This suggests that English is a more preferable language for these films.
- ▶ Consider working with Tony Kaye or Charles Chaplin as a director on future projects, as their past work has received high ratings from audiences and critics.
- ▶ It appears that the Crime|Drama|Fantasy|Mystery genre has the highest average IMDB score, indicating that it is a more preferable genre.
- ▶ It appears that Johnny Depp is the audience favorite and critic favorite actor

THANK YOU