

IMDB MOVIE ANALYSIS

By
R Anand Bhaskar



AGENDA



PROJECT DESCRIPTION



APPROACH



TECH-STACK USED



INSIGHTS



RESULTS

PROJECT DESCRIPTION

- The IMDb Movie Analysis project aims to explore and analyze a dataset of movies from IMDb database to understand various trends, patterns and insights.
- By applying Data Analysis techniques, this project helps to understand how different factors like genre, duration, language, director and profit margin impact IMDb scores.
- The goal of this project is to analyse the IMDb movie dataset and provide actionable insights that can help stakeholders to make informed decision.

APPROACH

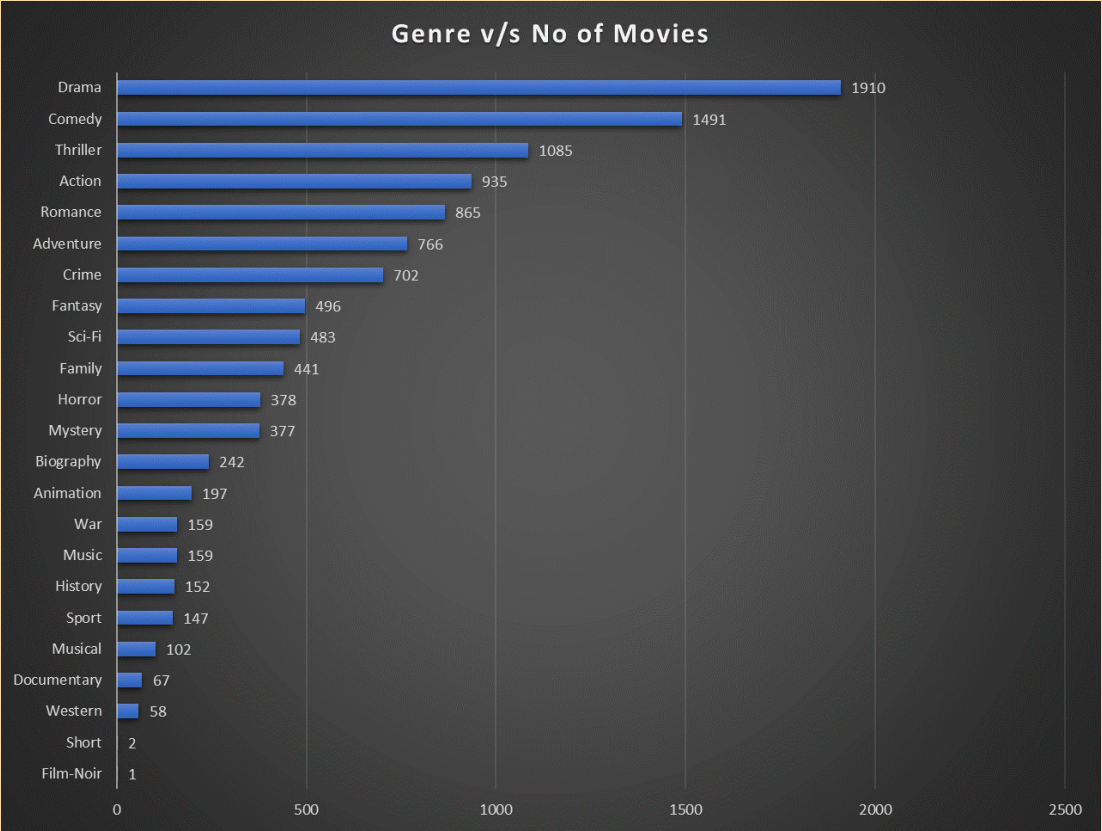
- Downloaded the dataset provided and used excel for data cleaning and analysis.
- Data cleaning process was done by removing duplicate columns and rows with missing values. Only columns that were required for the analysis process was taken.
- The Data Analysis process involves exploring the data to understand relationships and analysing impact of factors like genre, director, language, budget etc on IMDb scores to provide reasonable insights.

TECH-STACK USED

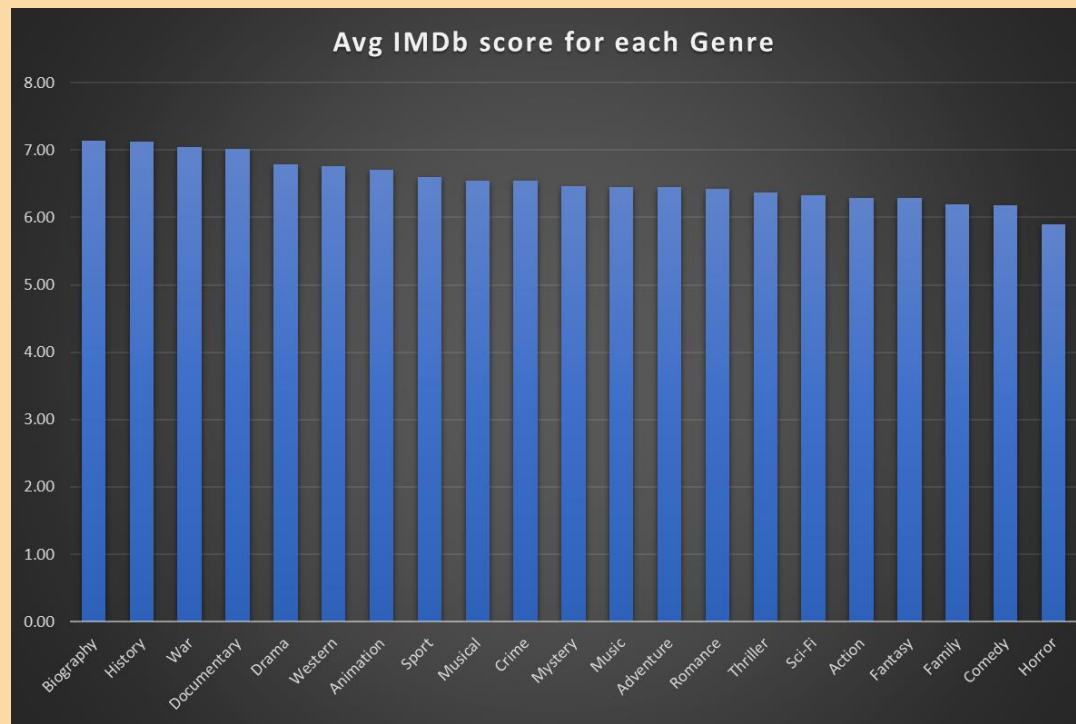
- Microsoft Excel 2022 – For data cleaning, analysing and visualisation

A. MOVIE GENRE ANALYSIS

Task : Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.



Genres	No. of Movies	Mean	Median	Mode	Range	Variance	Standard Deviation
Biography	242	7.14		7.2	7	4.4	0.502154
History	152	7.13		7.2	7.7	3.4	0.448608
War	159	7.05		7.1	7.1	4.3	0.648284
Documentary	67	7.01		7.2	6.6	6.9	1.418365
Drama	1910	6.79		6.9	6.7	7.2	0.793973
Western	58	6.77		6.8	6.8	4.8	0.979845
Animation	197	6.70		6.8	7.3	5.8	0.982284
Sport	147	6.60		6.8	7.2	6.4	1.091291
Musical	102	6.55		6.7	7.1	6.4	1.294852
Crime	702	6.55		6.6	6.6	6.9	0.967083
Mystery	377	6.47		6.5	6.6	5.5	1.012146
Music	159	6.46		6.5	6.5	6.9	1.464671
Adventure	766	6.45		6.6	6.6	6.6	1.245896
Romance	865	6.43		6.5	6.5	6.4	0.938428
Thriller	1085	6.37		6.4	6.5	6.3	0.938067
Sci-Fi	483	6.33		6.4	6.7	6.9	1.36138
Action	935	6.29		6.3	6.6	6.9	1.077034
Fantasy	496	6.29		6.4	6.7	6.7	1.297923
Family	441	6.20		6.3	5.4	6.7	1.364807
Comedy	1491	6.18		6.3	6.7	6.9	1.080984
Horror	378	5.90		5.9	5.9	6.3	0.978941



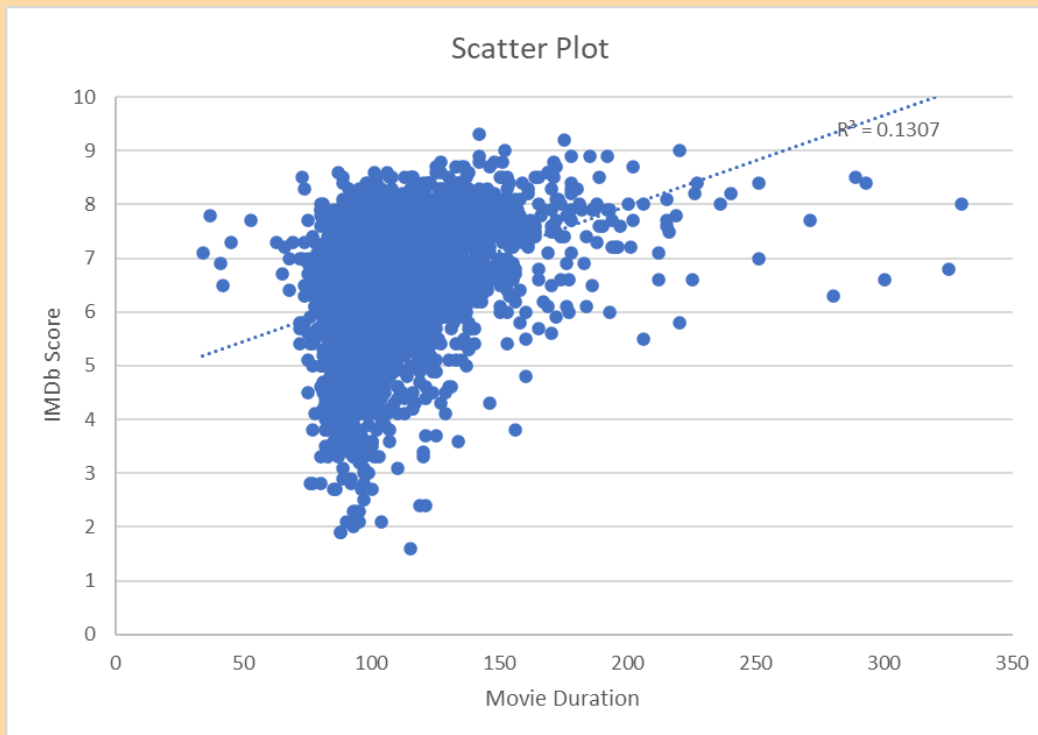
- Biography, History and War genres have relatively higher average IMDb scores and have lower variability suggesting they are well-received and have consistent ratings among audiences.
- The Drama genre is a widely represented genre with 1910 movies and have a moderate rating of 6.79.
- Music, Documentary, Family and Sci-Fi genres have higher standard deviations indicating varied audience responses.
- Genres like Adventure, Romance, Action, and Fantasy have moderate mean and median IMDb scores, suggesting consistent but not extremely high ratings.
- Genres like Documentary, Drama, Western, and Animation have relatively higher ranges, indicating more variation in IMDb scores. This suggests that while some movies in these genres do very well, others might not.

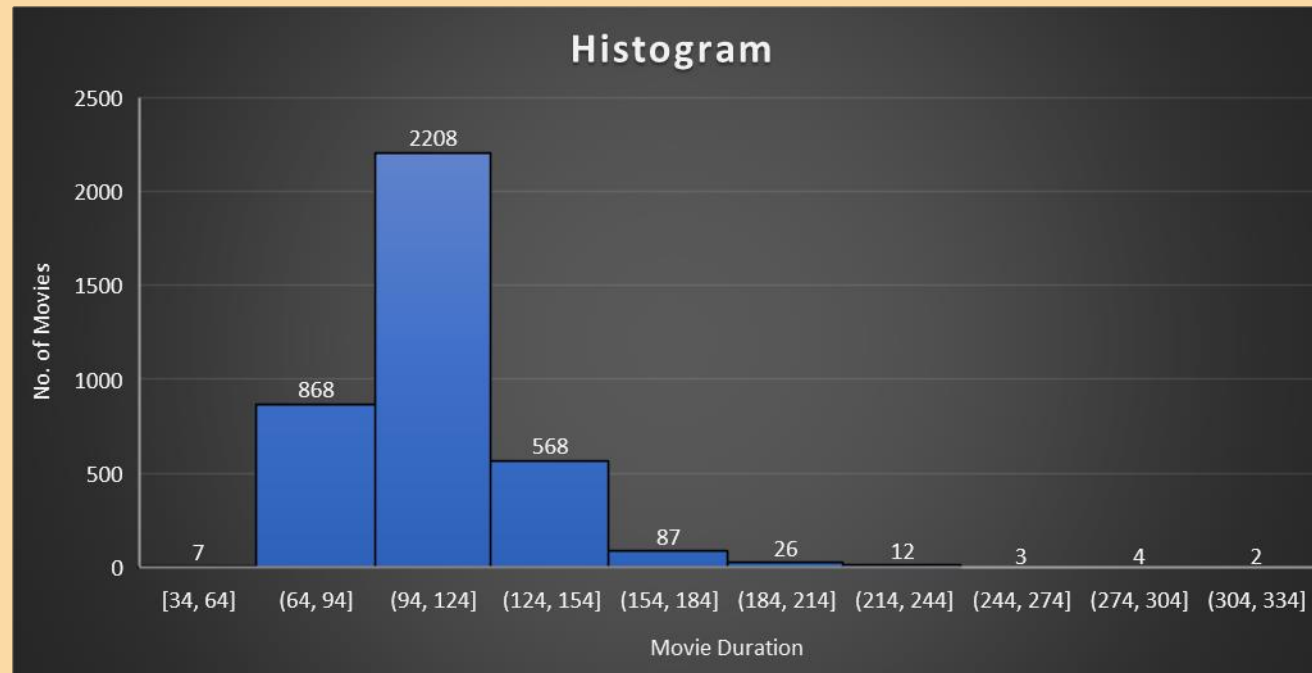
B. MOVIE DURATION ANALYSIS

Task : Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

The mean, median and standard deviation for movie durations are -

Mean	Median	Standard Deviation
109.8084544	105	22.76320079





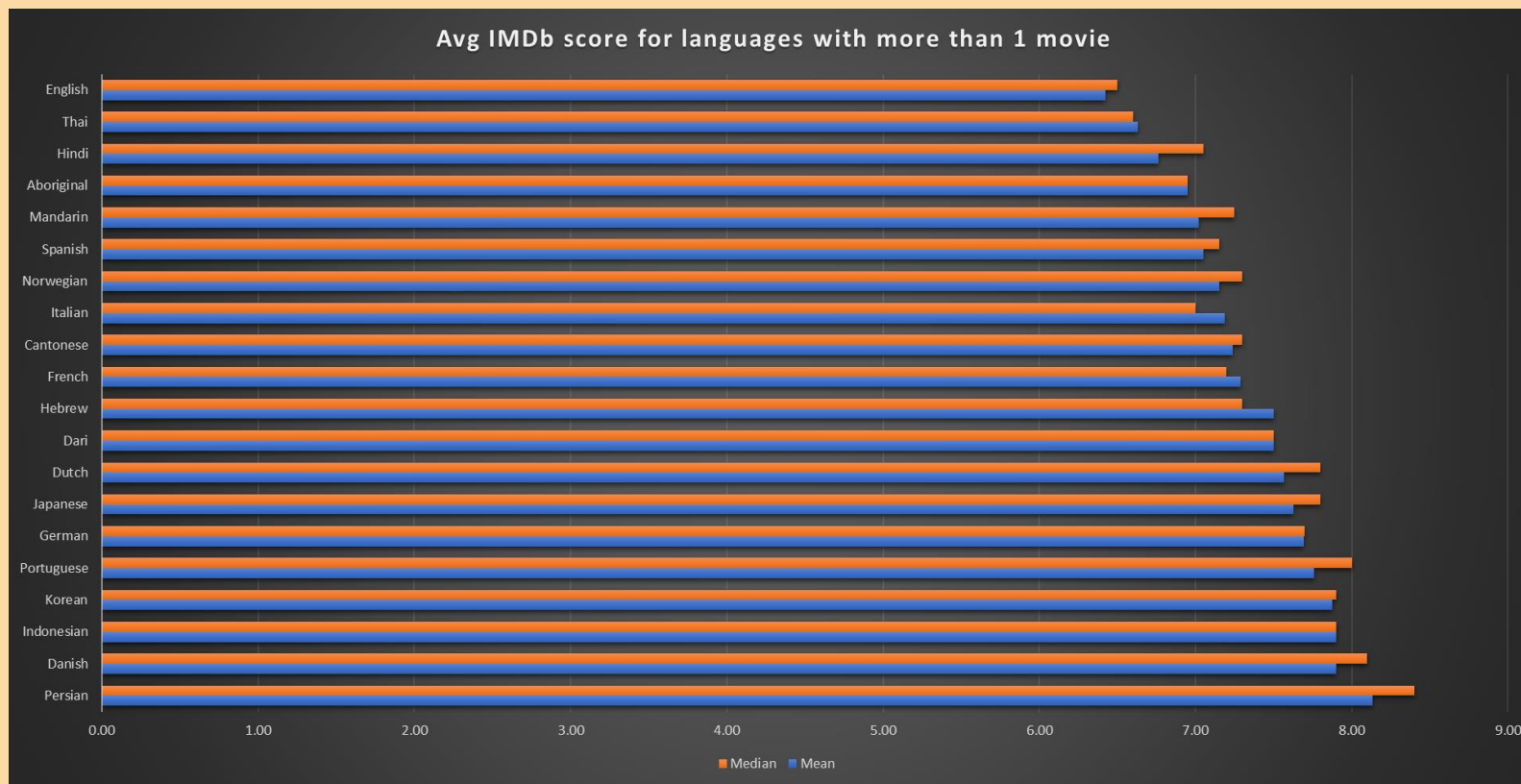
- The average duration for movies are 109.8 minutes and the median value is also close, which is 105 minutes. Most movies have duration in the range of 94-124 minutes.
- The positive slop in the trendline indicates a tendency for longer movies to have higher IMDb scores. However, the strength of this relationship is relatively low, as indicated by the R-squared value of 0.13.
- A weak relationship exists between movie duration and imdb scores.

C. LANGUAGE ANALYSIS

Task : Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Language	No. of Movies	Mean	Median	Standard Deviation
Persian	3	8.13	8.4	0.449691252
Danish	3	7.90	8.1	0.43204938
Indonesian	2	7.90	7.9	0.3
Korean	4	7.88	7.9	0.414578099
Portuguese	5	7.76	8	0.875442745
German	13	7.69	7.7	0.615769111
Japanese	12	7.63	7.8	0.861321659
Dutch	3	7.57	7.8	0.329983165
Dari	2	7.50	7.5	0.1
Hebrew	3	7.50	7.3	0.355902608
French	37	7.29	7.2	0.553691378
Cantonese	8	7.24	7.3	0.412121038
Italian	7	7.19	7	1.069617517
Norwegian	4	7.15	7.3	0.497493719
Spanish	26	7.05	7.15	0.810151933
Mandarin	14	7.02	7.25	0.737930089
Aboriginal	2	6.95	6.95	0.55
Hindi	10	6.76	7.05	1.05470375
Thai	3	6.63	6.6	0.368178701
English	3606	6.42	6.5	1.052352956

Language	No. of Movies	Mean
English	3606	6.42
Mandarin	14	7.02
Aboriginal	2	6.95
Spanish	26	7.05
French	37	7.29
Filipino	1	6.70
Maya	1	7.80
Kazakh	1	6.00
Telugu	1	8.40
Cantonese	8	7.24
Japanese	12	7.63
Aramaic	1	7.10
Italian	7	7.19
Dutch	3	7.57
Dari	2	7.50
German	13	7.69
Mongolian	1	7.30
Thai	3	6.63
Bosnian	1	4.30
Korean	4	7.88
Hungarian	1	7.10
Hindi	10	6.76
Icelandic	1	6.90
Danish	3	7.90
Portuguese	5	7.76
Norwegian	4	7.15
Czech	1	7.40
Russian	1	6.50
None	1	8.50
Zulu	1	7.30
Hebrew	3	7.50
Dzongkha	1	7.50
Arabic	1	7.20
Vietnamese	1	7.40
Indonesian	2	7.90
Romanian	1	7.90
Persian	3	8.13
Swedish	1	7.60



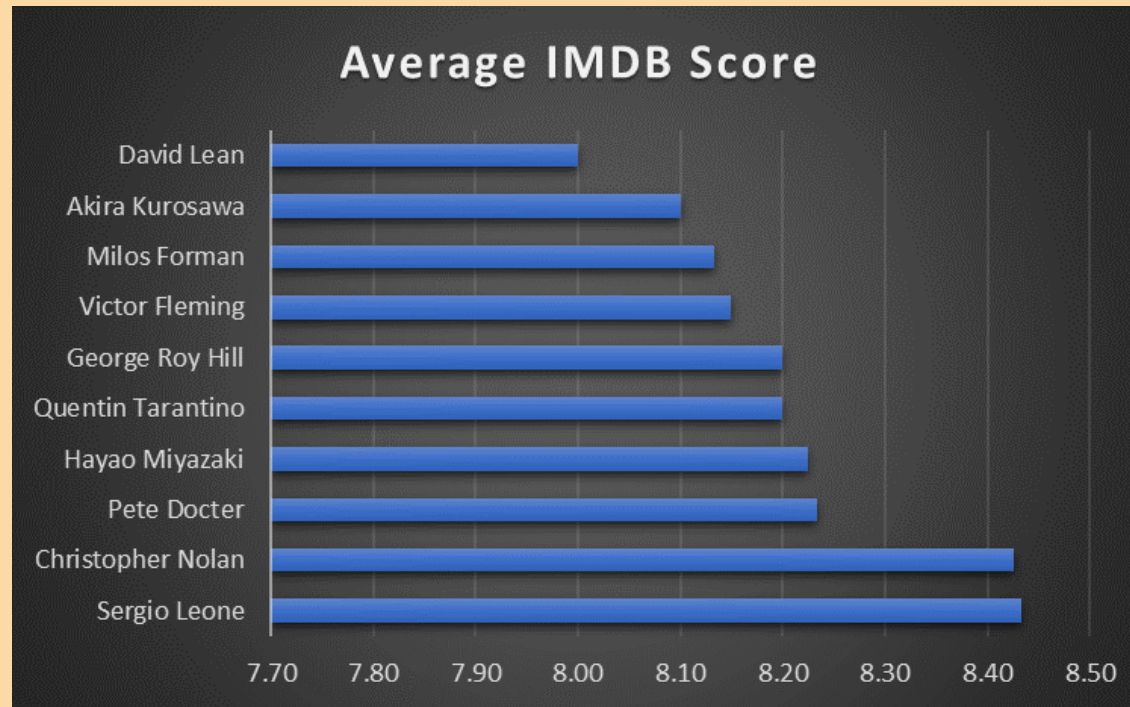
- The most common language is English with a count of 3606 movies, but the average IMDb score is only 6.42 due to varied reception.
- Among the languages with more than one movie count, persian language has an average score of 8.13.
- Persian, Danish, Indonesian and Korean languages have good IMDb scores with less standard deviation indicating that movies in these languages have consistent ratings and less variability among audiences
- Languages like Italian, Hindi, English have higher standard deviation indicating that they have a diverse audience response and wide range of ratings.

D. DIRECTOR ANALYSIS

Task : Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Top 10 directors based on their average IMDb score (more than one movie directed) -

Director	No of Movies	Average IMDB Score
Sergio Leone	3	8.43
Christopher Nolan	8	8.43
Pete Docter	3	8.23
Hayao Miyazaki	4	8.23
Quentin Tarantino	8	8.20
George Roy Hill	2	8.20
Victor Fleming	2	8.15
Milos Forman	3	8.13
Akira Kurosawa	2	8.10
David Lean	4	8.00



- Sergio Leone and Christopher Nolan are the top directors with an average IMDb score of 8.43.
- The 90th percentile for overall IMDb scores is 7.7 and that of top directors is 8.42.
- The average IMDb scores of top directors range from 8-8.43, which is well above 7.7. This indicates that these director's movies consistently achieve higher IMDb scores.

E. BUDGET ANALYSIS

Task : Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

- The movie with highest profit margin is Avatar.

Movie Title	Gross	Budget	IMDb score	Profit margin
Avatar	760505847	237000000	7.9	523505847.

- The correlation coefficient between movie budgets and gross earnings is 0.223. It indicates a relatively weak positive correlation.
- As the correlation coefficient is closer to 0, it suggests that changes in gross earnings are not strongly dependent on movie budgets.

movie_title	gross	budget	Profit Margin
Lady Vengeance	211667	4200000000	-419978833
Fateless	195888	2500000000	-249980411
Princess Mononoke	2298191	2400000000	-239770180
Steamboy	410388	2127519898	-212710951
Akira	439162	1100000000	-109956083
Godzilla 2000	10037390	1000000000	-98996261

- But it can be found that huge loss movies are mostly high budget movies.

movie_title	Profit Margin	imdb_score
Avatar	523505847	7.9
Jurassic World	502177271	7
Titanic	458672302	7.7
Star Wars: Episode IV - A New Hope	449935665	8.7
E.T. the Extra-Terrestrial	424449459	7.9
The Avengers	403279547	8.1
The Lion King	377783777	8.5
Star Wars: Episode I - The Phantom Menace	359544677	6.5
The Dark Knight	348316061	9
The Hunger Games	329999255	7.3



- The Dark Knight is the highest rated among top profited movies with an IMDb score of 9.0.

INSIGHTS

The key insights derived from the analysis are:

- Drama, Comedy, Action and Thriller are the most common movie genres and among all the genres Drama and documentaries have higher IMDb scores. Genres like Action, Adventure and Fantasy have higher variability in scores.
- Movies with duration of 100 – 150 minutes have decent IMDb scores making it an optimal range of duration, and extremely short or long movies might receive lower ratings.
- Languages like Persian and Danish have an average IMDb score of above 8 and English despite having most movies has a relatively lower IMDb score.
- Directors mentioned in the top 10 like Sergio Leone and Christopher Nolan have high average IMDb ratings and they maintain IMDb scores above the 90th percentile, having consistent success.
- There's a weak positive correlation between Budget and gross earnings suggesting that high budget does not always lead to high gross earnings.
- Profitability can be accounted beyond budget and depends on various other factors like marketing, storytelling and wide audience reception.

RESULTS

- The Insights derived from the project can be useful for stakeholders to make informed decisions and prepare best strategies to make a film with higher IMDb scores as well as profit margins.
- The analysis was done on various factors like genre, language, duration, director and budget which are key to the success of a movie.
- In summary, this project helped to extract meaningful information's from raw data to provide insights and better understanding of various factors affecting movie ratings. It helped to understand how data analysis can be applied to any industry.
- File link - <https://docs.google.com/spreadsheets/d/1kimPR3z-r74DmB35rsnifRImzv27mDf5/edit?usp=sharing&oid=113891162099459296915&rtpof=true&sd=true>