

# IMDB Movie Analysis

## 1. Project Description

- **Objective:** The primary objective of this project is to analyze the factors that influence the success of a movie on IMDB. Success is defined by high IMDB ratings, which can be valuable information for producers, directors, and investors looking to make informed decisions.
- **Scope:** The analysis focused on exploring various dimensions of movie data, including genre, duration, language, director influence, and budget-profit relationships, to identify which factors most impact movie ratings and profitability.
- **Approach Summary:** A systematic approach was used, starting with data cleaning and preprocessing to handle missing values, duplicates, and necessary transformations. This was followed by detailed analysis using pivot tables and descriptive statistics to reveal patterns within the data.

## 2. Approach

- **Data Cleaning:** The dataset was reviewed for missing values and duplicates, and unnecessary columns were removed to streamline analysis. Basic preprocessing included standardizing column values (such as language names) and separating multiple genres when necessary.
- **Data Analysis:** Each task was approached with specific techniques:
  - **Genre Analysis:** Conditional averages, median, mode, and other descriptive statistics were used to analyze IMDB ratings across different genres, identifying genre-specific trends.
  - **Duration and Rating:** Scatter plots and trendlines were used to visualize and assess the relationship between movie duration and ratings.
  - **Language Impact:** By grouping and summarizing ratings based on language, insights were drawn about language-specific success trends.
  - **Director Influence:** Average ratings were calculated for each director, highlighting the impact of directorial influence on movie ratings.
  - **Budget and Profit Margin:** A calculated profit margin column helped examine financial success and the correlation between budget and gross earnings.

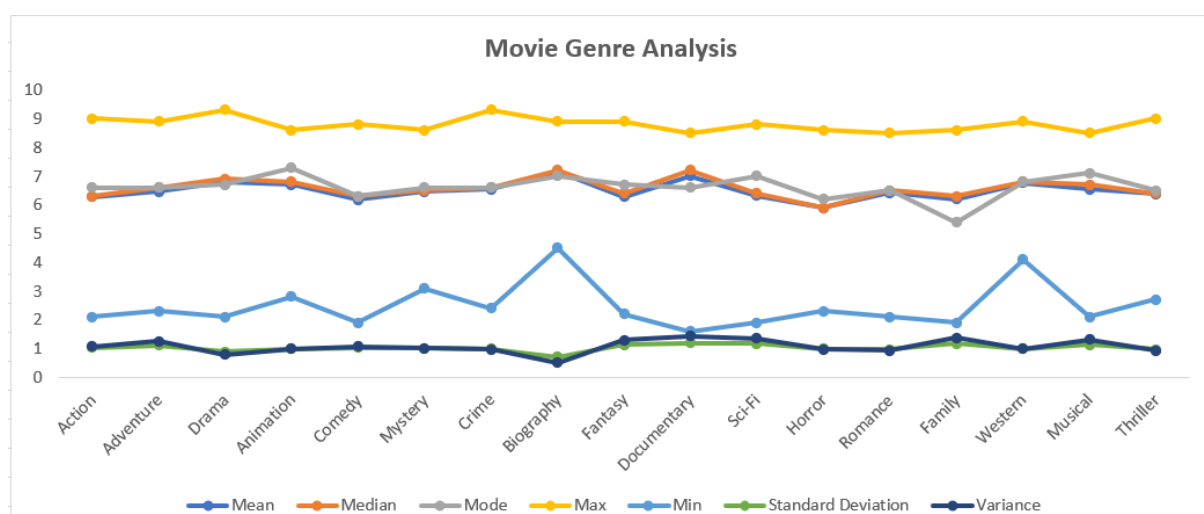
### 3. Tech-Stack Used

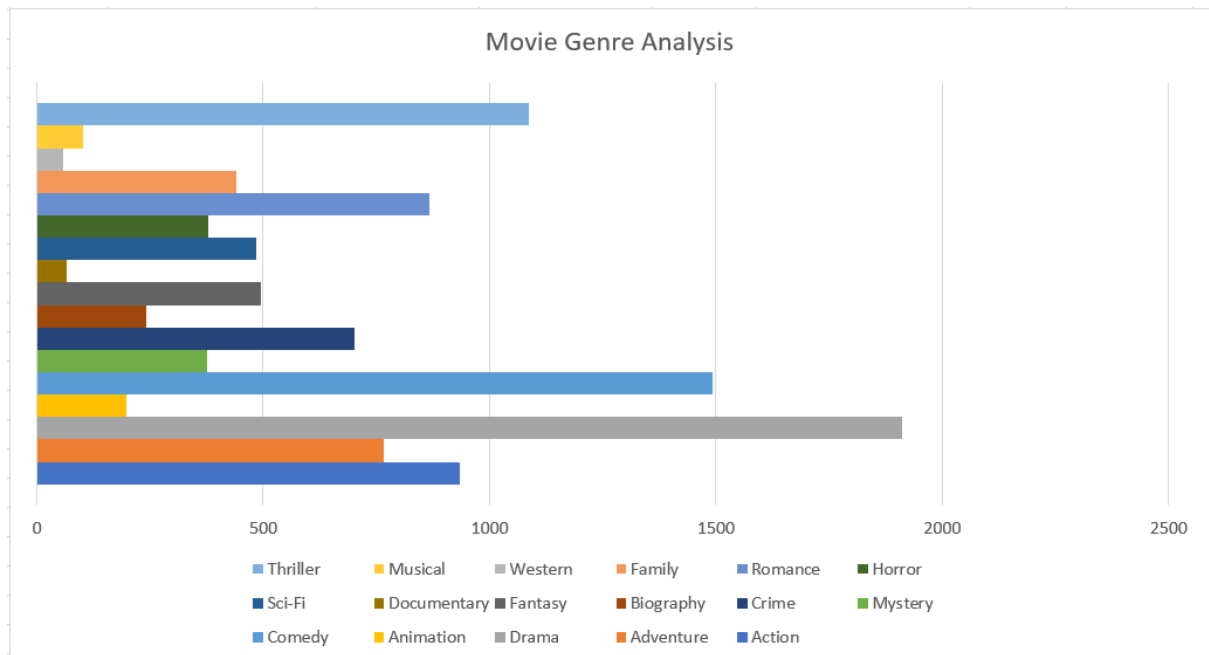
- **Microsoft Excel 2019 & Excel for the Web:** Both versions of Excel were used. Microsoft Excel 2019 provided primary data analysis capabilities, but Excel for the web was essential for functions unavailable in the 2019 desktop version, such as calculated fields in pivot tables.
- **Data Visualization:** Scatter plots, bar charts, and pivot tables were used to create visual representations of the data for better interpretability and trend analysis.

### 4. Insights

- **Movie Genre Analysis:** Certain genres, such as drama and action, had consistently higher IMDB ratings. Descriptive statistics highlighted that these genres typically appeal to a wider audience, contributing to better ratings.

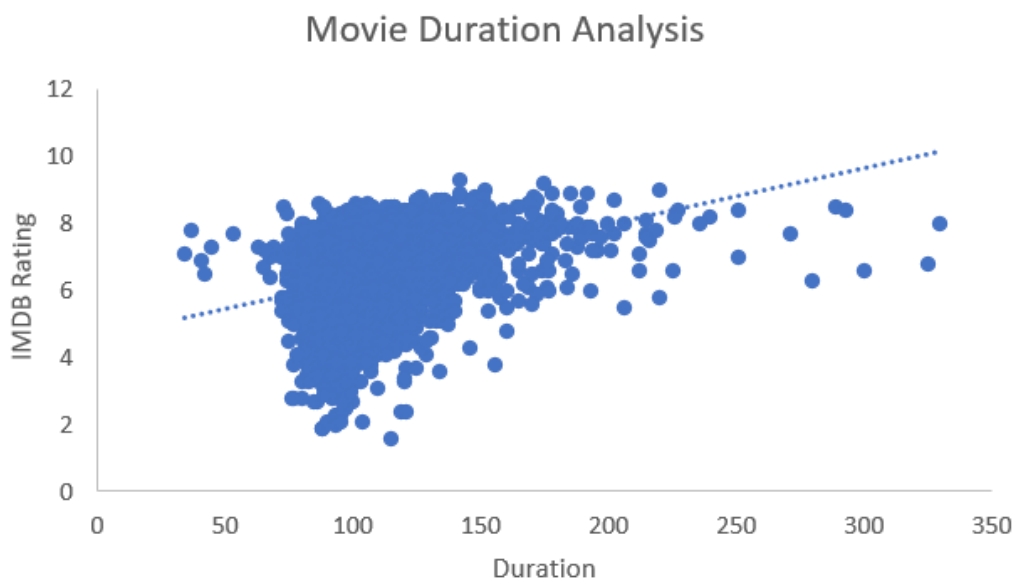
Genre	Mean	Median	Mode	Max	Min	Standard Deviation	Variance
Action	6.2859893	6.3	6.6	9	2.1	1.038357736	1.078186788
Adventure	6.4549608	6.6	6.6	8.9	2.3	1.116926308	1.247524378
Drama	6.7891156	6.9	6.7	9.3	2.1	0.891064898	0.793996652
Animation	6.7005076	6.8	7.3	8.6	2.8	0.993627525	0.987295659
Comedy	6.183311	6.3	6.3	8.8	1.9	1.039919012	1.081431552
Mystery	6.469496	6.5	6.6	8.6	3.1	1.007391835	1.014838309
Crime	6.5481481	6.6	6.6	9.3	2.4	0.984105199	0.968463042
Biography	7.1400826	7.2	7	8.9	4.5	0.71009671	0.504237338
Fantasy	6.2850806	6.4	6.7	8.9	2.2	1.140414241	1.30054464
Documentary	7.0119403	7.2	6.6	8.5	1.6	1.199939694	1.439855269
Sci-Fi	6.3272727	6.4	7	8.8	1.9	1.16718415	1.362318841
Horror	5.9039578	5.9	6.2	8.6	2.3	0.991023285	0.982127152
Romance	6.4262125	6.5	6.5	8.5	2.1	0.968996249	0.938953731
Family	6.2	6.3	5.4	8.6	1.9	1.169576458	1.367909091
Western	6.7655172	6.8	6.8	8.9	4.1	0.998516746	0.997035693
Musical	6.5509804	6.7	7.1	8.5	2.1	1.143535	1.307672297
Thriller	6.3723091	6.4	6.5	9	2.7	0.969078327	0.939112803





- Movie Duration Analysis:** Movies within a moderate duration range (90-120 minutes) showed a tendency to have higher ratings. Extremely short or long movies often received mixed reviews, possibly indicating a viewer preference for standard-length movies.

Operations	Values
Mean	109.80851
Median	105
Mode	101
Standard Deviation	22.763201
Variance	518.16332



- **Language Analysis:** English-language films made up the majority and often received higher ratings, although several foreign-language films also achieved significant success, indicating a growing diversity in audience preferences.

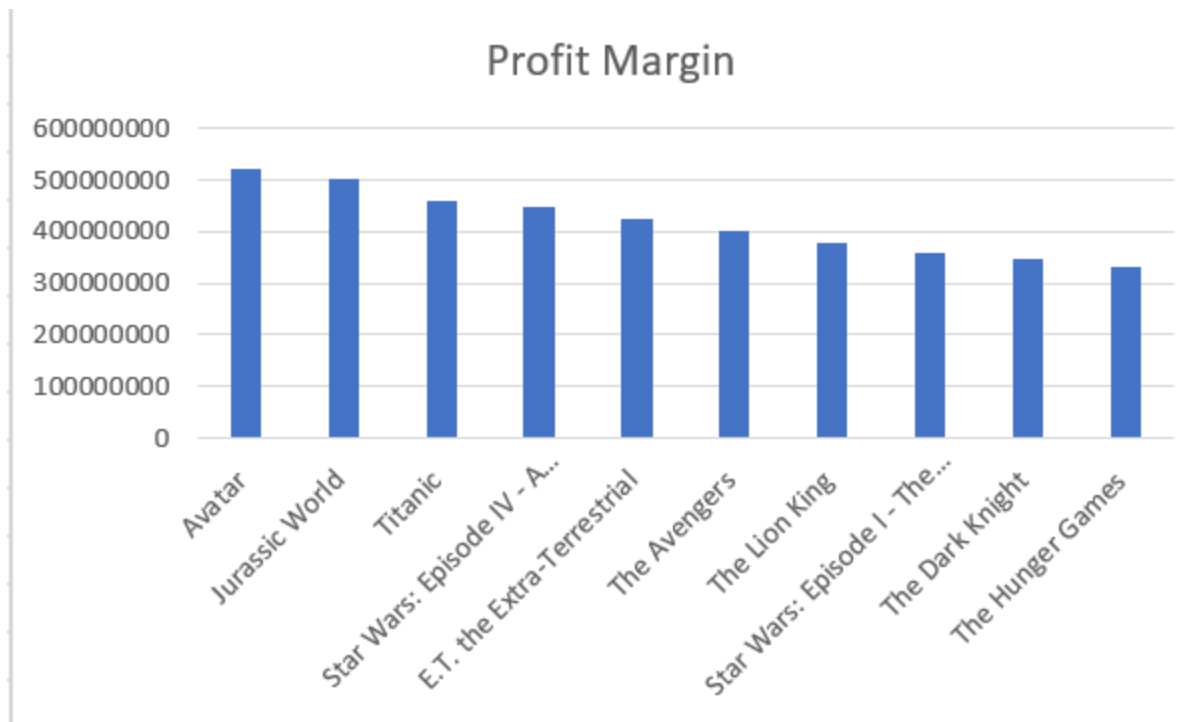
Languages	No. of Movies	Average rating	Median	Standard Deviation
English	3606	6.421436495	6.5	1.052498903
Mandarin	14	7.021428571	7.25	0.765786244
Aboriginal	2	6.95	6.95	0.777817459
Spanish	26	7.05	7.15	0.826196103
French	37	7.286486486	7.2	0.561328861
Filipino	1	6.7	6.7	N/A
Maya	1	7.8	7.8	N/A
Kazakh	1	6	6	N/A
Telugu	1	8.4	8.4	N/A
Cantonese	8	7.2375	7.3	0.440575922
Japanese	12	7.625	7.8	0.899621132
Aramaic	1	7.1	7.1	N/A
Italian	7	7.185714286	7	1.155318962
Dutch	3	7.566666667	7.8	0.404145188
Dari	2	7.5	7.5	0.141421356
German	13	7.692307692	7.7	0.640912811
Mongolian	1	7.3	7.3	N/A
Thai	3	6.633333333	6.6	0.450924975
Bosnian	1	4.3	4.3	N/A
Korean	5	7.7	7.7	0.570087713
Hungarian	1	7.1	7.1	N/A
Hindi	10	6.76	7.05	1.111755369
Icelandic	1	6.9	6.9	N/A
Danish	3	7.9	8.1	0.529150262
Portuguese	5	7.76	8	0.978774744
Norwegian	4	7.15	7.3	0.574456265
Czech	1	7.4	7.4	N/A
Russian	1	6.5	6.5	N/A
None	1	8.5	8.5	N/A
Zulu	1	7.3	7.3	N/A
Hebrew	3	7.5	7.3	0.435889894
Dzongkha	1	7.5	7.5	N/A
Arabic	1	7.2	7.2	N/A
Vietnamese	1	7.4	7.4	N/A
Indonesian	2	7.9	7.9	0.424264069
Romanian	1	7.9	7.9	N/A
Persian	3	8.133333333	8.4	0.550757055
Swedish	1	7.6	7.6	N/A

- **Director Analysis:** A few directors consistently achieved high average ratings, suggesting that their brand and storytelling style resonate well with audiences. These findings could serve as a benchmark for directors aiming to produce high-rated films.

Director	Average Rating	Percentile
Tony Kaye	8.6	0.996
Charles Chaplin	8.6	0.994
Alfred Hitchcock	8.5	0.993
Ron Fricke	8.5	0.987
Damien Chazelle	8.5	0.989
Majid Majidi	8.5	0.961
Sergio Leone	8.433333333	0.993
Christopher Nolan	8.425	0.995
S.S. Rajamouli	8.4	0.99
Richard Marquand	8.4	0.991
Asghar Farhadi	8.4	0.969
Marius A. Markevicius	8.4	0.971
Lee Unkrich	8.3	0.991
Fritz Lang	8.3	0.985
Lenny Abrahamson	8.3	0.986

- **Budget Analysis:** A positive correlation was found between budget and gross earnings, with higher-budget movies generally achieving greater profitability. However, high profit margins were sometimes achieved by mid-budget films, highlighting the potential for financial success without a large budget.

Movie Title	Budget	Gross	Profit Margin	Correlation
Avatar	237000000	760505847	523505847	0.85881224
Jurassic World	150000000	652177271	502177271	0.85881224
Titanic	200000000	658672302	458672302	0.85881224
Star Wars: Episode IV - A New Hope	11000000	460935665	449935665	0.85881224
E.T. the Extra-Terrestrial	10500000	434949459	424449459	0.85881224
The Avengers	220000000	623279547	403279547	0.85881224
The Lion King	45000000	422783777	377783777	0.85881224
Star Wars: Episode I - The Phantom Menace	115000000	474544677	359544677	0.85881224
The Dark Knight	185000000	533316061	348316061	0.85881224
The Hunger Games	78000000	407999255	329999255	0.85881224



## 5. Result

- **Achievements:** This project allowed for the successful identification of key factors affecting movie success on IMDB. By analyzing genres, language, duration, directorial impact, and budget, a comprehensive view of what drives higher ratings and profitability was achieved.
- **Contribution to Understanding:** This analysis deepened the understanding of how different elements contribute to a movie's success. The ability to derive meaningful insights from raw data emphasized the importance of data analytics in the entertainment industry, highlighting how stakeholders can make data-driven decisions for future projects.

For the excel file: [Click here](#). Note that Only task E is inside a separate worksheet and all other tasks are in the cleaned dataset worksheet(Look beside the cleaned dataset one by one everything is done)