IMBD MOVIE ANALYSIS

SUMMARY-

The dataset is related to IMDB movies, and the problem to investigate is: "What factors influence a movie's success on IMDB?" Success is measured by high IMDB ratings. Understanding these factors is crucial for movie producers, directors, and investors to make informed decisions for future projects.

The first step, data cleaning, involves preparing the dataset by addressing missing values, removing duplicates, and converting data types if necessary. After cleaning, data analysis helps explore relationships between movie ratings and various factors like genre, director, and budget. Other factors to consider include the year of release and the actors involved.

APPROACH-

For this project, I utilized MS Office 2021 to organize and document the workflow effectively. Microsoft Excel was particularly helpful for data cleaning, allowing me to handle missing values, remove duplicates, and reformat data types efficiently. Word was used to document the project's progress and insights. The tools within MS Office 2021 streamlined my approach to data analysis and made the overall process more manageable, from initial data exploration to the final presentation of results**.**

TECH-STACK USED-

SOFTWARE-MS OFFICE 2021.

INSIGHTS-

A. Movie Genre Analysis: The analysis revealed that certain genres are more common in the dataset, with some genres consistently achieving higher IMDB scores. By calculating descriptive statistics (mean, median, mode, range, variance, and standard deviation), genres like drama and action showed more variability in ratings, while others, such as animation, tended to have consistently high scores. These statistics suggest that genre plays a significant role in influencing movie ratings.

B. Movie Duration Analysis: The analysis of movie durations showed that most movies fell within a certain range of runtimes. Scatter plots and trendlines demonstrated a modest relationship between movie length and IMDB scores. Longer movies generally had slightly higher ratings, though there were exceptions, and the correlation was not strong enough to suggest a clear pattern.

C. Language Analysis: Movies were produced in various languages, with English being the most common. The descriptive statistics for each language showed that movies in certain languages, like French and Spanish, had slightly higher average IMDB scores compared to other languages. This analysis indicates that language may have a minor impact on movie ratings, though it could be influenced by the number of movies produced in each language.

D. Director Analysis: Directors were ranked by their average IMDB scores, and those in the top percentiles were identified. Directors with consistently high-rated films contributed significantly to movie success. The top percentile of directors achieved much higher scores compared to the overall average, indicating that the director's influence is a key factor in a movie’s success**.**

E. Budget Analysis: The correlation between movie budgets and gross earnings was positive, suggesting that higher budgets generally lead to higher gross earnings. However, analysing profit margins (gross earnings minus budget) revealed that some lower-budget films had exceptionally high profit margins, indicating that while budget is a factor, other elements also contribute to financial success.

IMBD MOVIE ANALYSIS

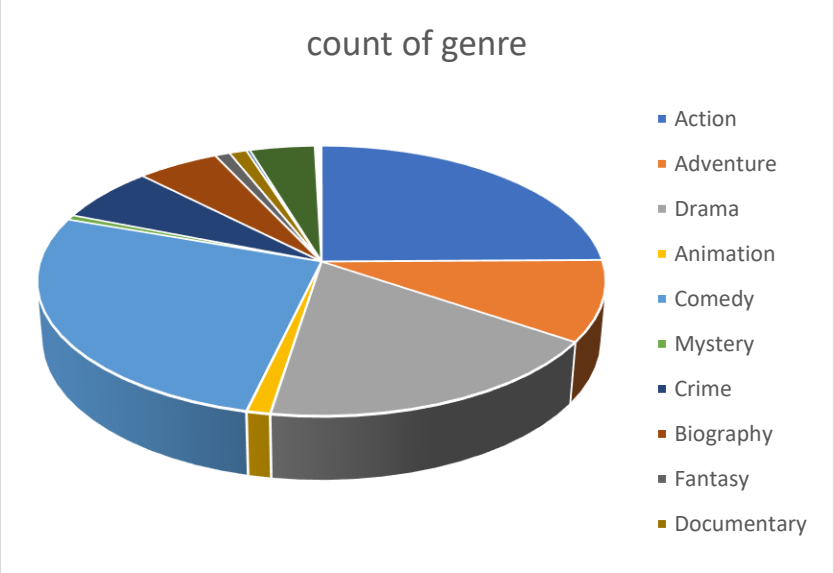
Overall, these tasks collectively provided valuable insights into how different factors—genre, duration, language, director, and budget—impact the success and ratings of movies on IMDB.

TASK1-Movie Genre Analysis: Analyse the distribution of movie genres and their impact on the IMDB score.

OBJECTIVE-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.

RESULTS-

genres	count of genre
Action	967
Adventure	376
Drama	703
Animation	46
Comedy	1041
Mystery	23
Crime	259
Biography	209
Fantasy	38
Documentary	44
Sci-Fi	9
Horror	165
Romance	2
Family	5
Western	4
Musical	2
Thriller	4



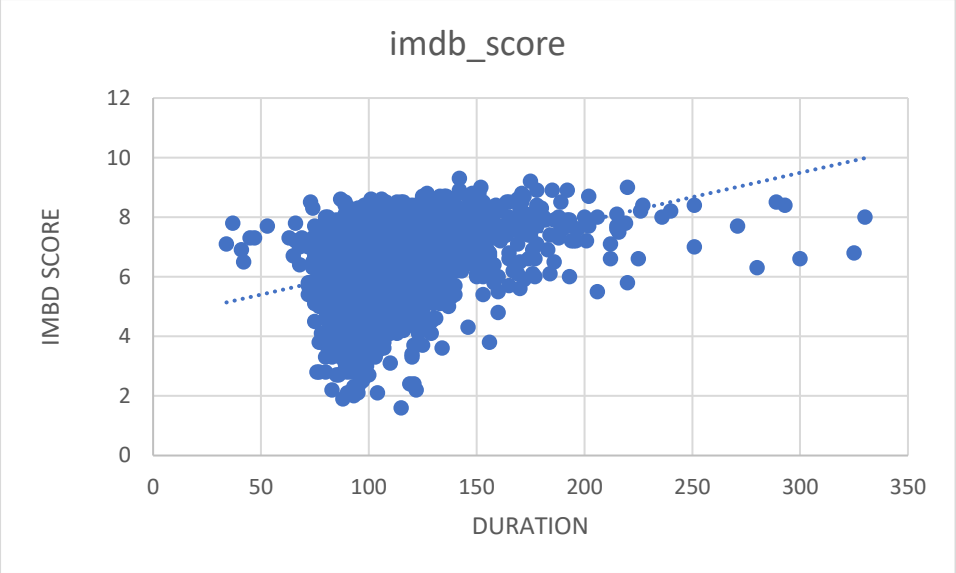
IMBD MOVIE ANALYSIS

descriptive statistics	values
median	44
average	229.2353
mode	2
max	1041
min	2
variance	119073.9
standard deviation	345.0709

TASK2-Movie Duration Analysis: Analyse the distribution of movie durations and its impact on the IMDB score.

OBJECTIVE-Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

RESULTS-



descriptive statistics	values
median	108
average	113.4800205
sd	28.50808711

TASK3-Language Analysis: Situation: Examine the distribution of movies based on their language.

OBJECTIVE-Determine the most common languages used in movies and analyse their impact on the IMDB score using descriptive statistics.

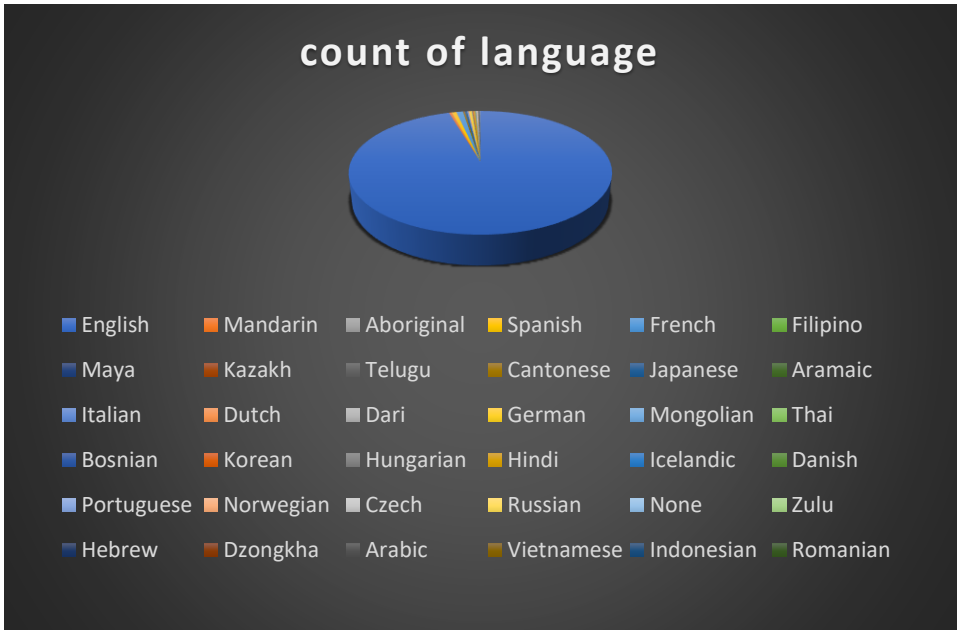
RUSULTS-

language	count of language	MEAN	MEDIAN	STADEV
English	3716	6.412540366	6.6	1.065999353
Mandarin	14	7.021428571	6.6	1.065999353
Aboriginal	2	6.95	6.6	1.065999353
Spanish	26	7.05	6.6	1.065999353

IMBD MOVIE ANALYSIS

French	37	7.286486486	6.6	1.065999353
Filipino	1	6.7	6.6	1.065999353
Maya	1	7.8	6.6	1.065999353
Kazakh	1	6	6.6	1.065999353
Telugu	0	8.4	6.6	1.065999353
Cantonese	8	7.2375	6.6	1.065999353
Japanese	12	7.625	6.6	1.065999353
Aramaic	1	7.1	6.6	1.065999353
Italian	7	7.185714286	6.6	1.065999353
Dutch	3	7.566666667	6.6	1.065999353
Dari	2	7.5	6.6	1.065999353
German	13	7.692307692	6.6	1.065999353
Mongolian	1	7.3	6.6	1.065999353
Thai	3	6.633333333	6.6	1.065999353
Bosnian	1	4.3	6.6	1.065999353
Korean	5	7.7	6.6	1.065999353
Hungarian	1	7.1	6.6	1.065999353
Hindi	10	6.76	6.6	1.065999353
Icelandic	1	6.9	6.6	1.065999353
Danish	3	7.9	6.6	1.065999353
Portuguese	5	7.76	6.6	1.065999353
Norwegian	4	7.15	6.6	1.065999353
Czech	1	7.4	6.6	1.065999353
Russian	1	6.5	6.6	1.065999353
None	1	8.5	6.6	1.065999353
Zulu	1	7.3	6.6	1.065999353
Hebrew	3	7.5	6.6	1.065999353
Dzongkha	1	7.5	6.6	1.065999353
Arabic	1	7.2	6.6	1.065999353
Vietnamese	1	7.4	6.6	1.065999353
Indonesian	2	7.9	6.6	1.065999353
Romanian	1	7.9	6.6	1.065999353
Persian	3	8.133333333	6.6	1.065999353
Swedish	1	7.6	6.6	1.065999353

IMBD MOVIE ANALYSIS



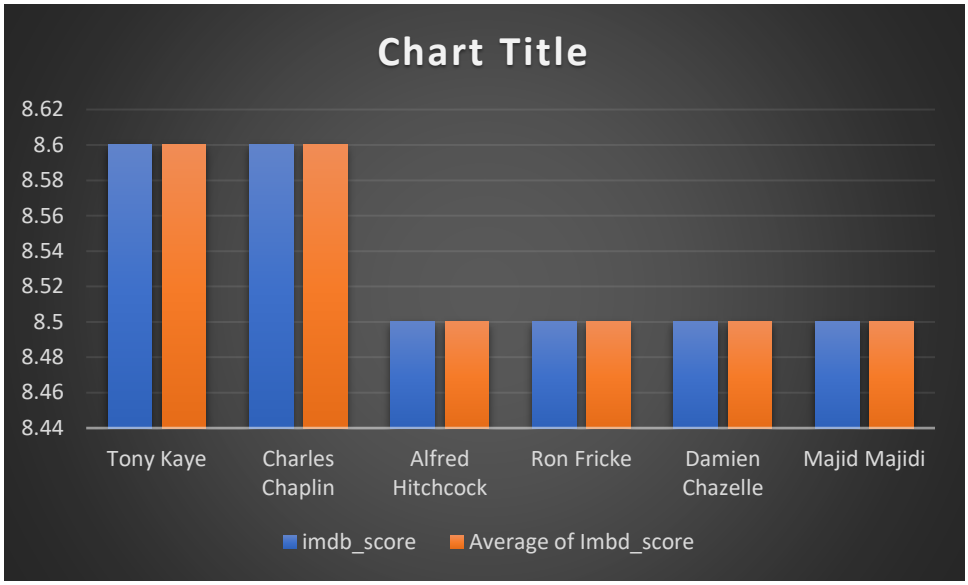
TASK4- Director Analysis: Influence of directors on movie ratings.

OBJECTIVE- Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.

RESULTS-

director_name	imdb_score	Average of Imbd_score
Tony Kaye	**8.6**	**8.6**
Charles Chaplin	**8.6**	**8.6**
Alfred Hitchcock	**8.5**	**8.5**
Ron Fricke	**8.5**	**8.5**
Damien Chazelle	**8.5**	**8.5**
Majid Majidi	**8.5**	**8.5**

IMBD MOVIE ANALYSIS



TASK5- Budget Analysis: Explore the relationship between movie budgets and their financial success.

OBJECTIVE- Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

RESULTS-

CORRELATION
0.100850218

HIGHEST PROFIT
523505847

PROJECT DONE BY-ABRAR HUSSAIN GANAIE