IMDB MOVIE ANALYSIS

MIRRA. G

PROJECT DESCRIPTION:

This project entails a detailed analysis of various factors influencing movie success, leveraging a dataset of movie records. The study covers five primary areas: genre, duration, language, director, and budget. Each aspect is meticulously examined using Excel functions to draw out significant insights. The genre analysis explores the prevalence and impact of different movie genres on IMDB scores. Duration analysis looks at how the length of movies correlates with their ratings. The language analysis investigates the distribution of movies across different languages and their associated ratings. Director analysis identifies the top directors based on average IMDB scores and their influence on movie success. Lastly, the budget analysis examines the relationship between movie budgets and gross earnings, highlighting films with the highest profit margins. These analyses collectively provide a comprehensive understanding of the elements that contribute to a movie's performance and success.

TECH STACK USED: Microsoft Excel

DATA CLEANING:

- The columns that are unwanted for the analysis are dropped. In this case we drop the columns 'color', 'director_facebook_likes', 'actor_3_facebook_likes', 'actor_1_facebook_likes', 'actor_2_facebook_likes', 'movie_facebook_likes', 'aspect_ratio', 'movie_imdb_link', 'facenumber_in_posts' and 'cast_total_facebook_likes'.
- The rows with duplicate values are then dropped.
- The rows with null/blank cells are also dropped.

A) *Movie Genre Analysis:* **Analyze the distribution of movie genres and their impact on the IMDB score.**

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.**

Approach:

We first extract the unique genres from each column and find the count of films corresponding to each genre and then calculate the descriptive statistics such as average, median, mode, min, max, variance and standard deviation for the IMDB scores using a combination of the respective functions along with functions like IF, ISNUMBER and SEARCH. We also create a graph between the genres and count to identify the most common genres.

Output:

unique_genre	count	average_IMDB	median_IMDB	mode_IMDB	max_IMDB	min_IMDB	var_IMDB	stddev_IMDB
Action	933	6.27	6.3	6.6	9	2.1	2.1	1.05
Adventure	762	6.45	6.6	6.6	8.9	2.3	2.3	1.12
Drama	1890	6.79	6.9	6.7	9.3	2.1	2.1	0.9
Animation	195	6.7	6.8	7.3	8.6	2.8	2.8	0.99
Comedy	1484	6.17	6.3	6.7	8.8	1.9	1.9	1.05
Mystery	382	6.46	6.5	6.6	8.6	3.1	3.1	1.03
Crime	698	6.55	6.6	6.6	9.3	2.4	2.4	0.99
Biography	239	7.15	7.2	7	8.9	4.5	4.5	0.7
Fantasy	492	6.28	6.4	6.7	8.9	2.2	2.2	1.14
Documentary	54	7.04	7.4	6.6	8.5	1.6	1.6	1.28
Sci-Fi	487	6.31	6.4	7	8.8	1.9	1.9	1.18
Horror	395	5.84	5.9	6.2	8.6	2.2	2.2	1.03
Romance	862	6.43	6.5	6.5	8.6	2.1	2.1	0.97
Family	438	6.2	6.3	6.1	8.6	1.9	1.9	1.17
Western	60	6.7	6.75	6.8	8.9	3.8	3.8	1.04
Musical	98	6.57	6.7	7.1	8.5	2.1	2.1	1.09
Thriller	1095	6.35	6.4	6.5	9	2.7	2.7	0.99
History	149	7.13	7.2	7.7	8.9	4.4	4.4	0.7
Music	241	6.46	6.6	6.2	8.5	1.6	1.6	1.17
War	154	7.05	7.1	7.1	8.6	4.3	4.3	0.81
Sport	147	6.58	6.8	7.2	8.4	2	2	1.07
Film-Noir	2	7.95	7.95	7.7	8.2	7.7	7.7	0.25
News	1	7.1	7.1	7.1	7.1	7.1	7.1	0
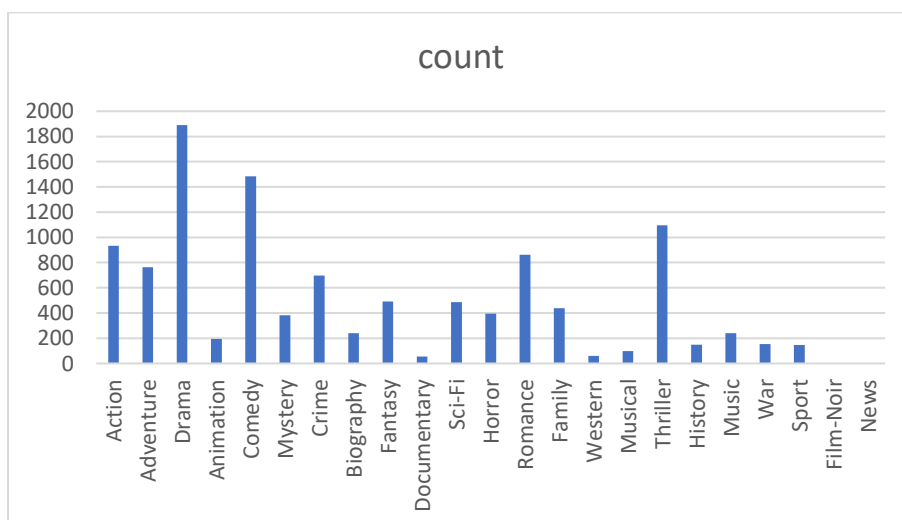
Analysis:

We can see that the genre 'Drama' is the most common one followed by 'Comedy', 'Thriller' and 'Romance'.

B) *Movie Duration Analysis:* **Analyze the distribution of movie durations and its impact on the IMDB score.**
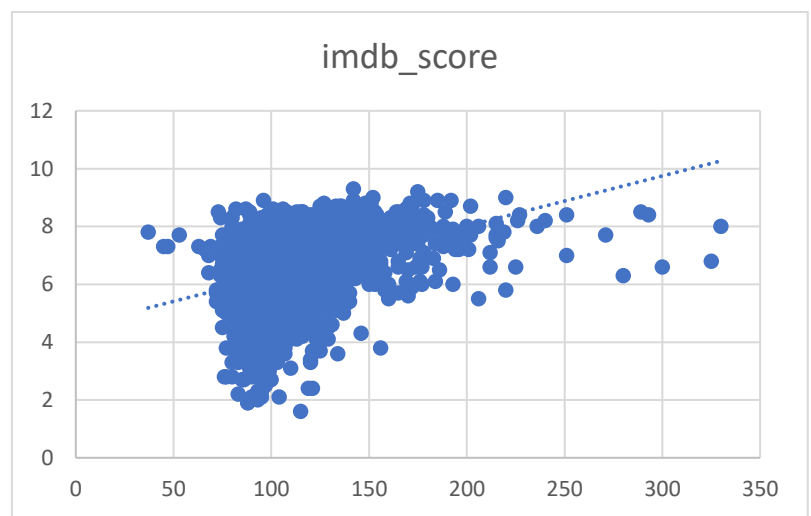
Task: **Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.**

<u>*Approach:*</u>

We analyse the distribution of movie durations by calculating the descriptive statistics of movie duration and identify the relationship between movie duration and IMDB score by plotting a scatter plot between the two variables and derive conclusions.

<u>*Output:*</u>

average_duration	109.7407014
mode_duration	101
median_duration	105
stddev_duration	22.64341264
var_duration	512.7241362

<u>*Analysis:*</u>

The majority of the movies have durations between 50 to 200 minutes. The scatter plot indicates a positive correlation between movie duration and IMDb scores, with longer movies generally receiving higher ratings. However, the correlation is not very strong, and there is considerable variation in scores for movies of similar lengths. This suggests that while duration may influence IMDb scores to some extent, other factors are also crucial in determining a movie's rating.

C) *Language Analysis:* **Situation: Examine the distribution of movies based on their language.**

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

Approach:

We begin by extracting the unique languages into a separate column from the 'languages' column using the UNIQUE function and the count of number of films for each language. We then find the descriptive statistics of IMDB scores with respect to each unique language. We plot the average IMDB along with the unique languages on a bar chart to analyse the impact of languages on IMDB scores.

Output:

unique_languages	count	avg_IMDB	median_IMDB	mode_IMDB	max_IMDB	min_IMDB	var_IMDB	stddev_IMDB
English	3606	6.41	6.5	6.7	9.3	1.6	1.14	1.07
Mandarin	14	7.02	7.25	7.6	7.9	5.6	0.59	0.74
Aboriginal	2	6.95	6.95	6.4	7.5	6.4	0.61	0.55
Spanish	23	7.08	7.2	5.9	8.2	5.2	0.74	0.84
French	34	7.36	7.3	7.2	8.4	5.8	0.27	0.51
Filipino	1	6.7	6.7	6.7	6.7	6.7	0.523	0
Maya	1	7.8	7.8	7.8	7.8	7.8	0.523	0
Kazakh	1	6	6	6	6	6	0.523	0
Cantonese	7	7.34	7.3	7.3	7.8	6.7	0.12	0.32
Japanese	10	7.66	8	7.7	8.7	6	0.98	0.94
Aramaic	1	7.1	7.1	7.1	7.1	7.1	0.523	0
Italian	7	7.19	7	5.3	8.9	5.3	1.33	1.07
Dutch	3	7.57	7.8	7.8	7.8	7.1	0.16	0.33
Dari	16	7.08	7.4	7.6	7.9	5.6	0.54	0.71
German	10	7.77	7.8	7.4	8.5	6.1	0.51	0.68
Mongolian	1	7.3	7.3	7.3	7.3	7.3	0.523	0
Thai	3	6.63	6.6	6.2	7.1	6.2	0.2	0.37
Bosnian	1	4.3	4.3	4.3	4.3	4.3	0.523	0
Korean	5	7.7	7.7	8.1	8.4	7	0.33	0.51
Hungarian	1	7.1	7.1	7.1	7.1	7.1	0.523	0
Hindi	5	7.22	7.4	6	8	6	0.64	0.72
Danish	3	7.9	8.1	7.3	8.3	7.3	0.28	0.43
Portuguese	5	7.76	8	8.1	8.7	6.1	0.96	0.88
Norwegian	4	7.15	7.3	7.6	7.6	6.4	0.33	0.5
Czech	1	7.4	7.4	7.5	7.4	7.4	0.523	0
Russian	1	6.5	6.5	6.5	6.5	6.5	0.523	0
None	1	8.5	8.5	8.5	8.5	8.5	0.523	0
Zulu	1	7.3	7.3	7.3	7.3	7.3	0.523	0
Hebrew	2	7.65	7.65	8	8	7.3	0.25	0.35
Arabic	1	7.2	7.2	7.2	7.2	7.2	0.523	0
Vietnamese	1	7.4	7.4	7.4	7.4	7.4	0.523	0
Indonesian	2	7.9	7.9	7.6	8.2	7.6	0.18	0.3
Romanian	1	7.9	7.9	7.9	7.9	7.9	0.523	0
Persian	3	8.13	8.4	8.4	8.5	7.5	0.3	0.45

avg_IMDB
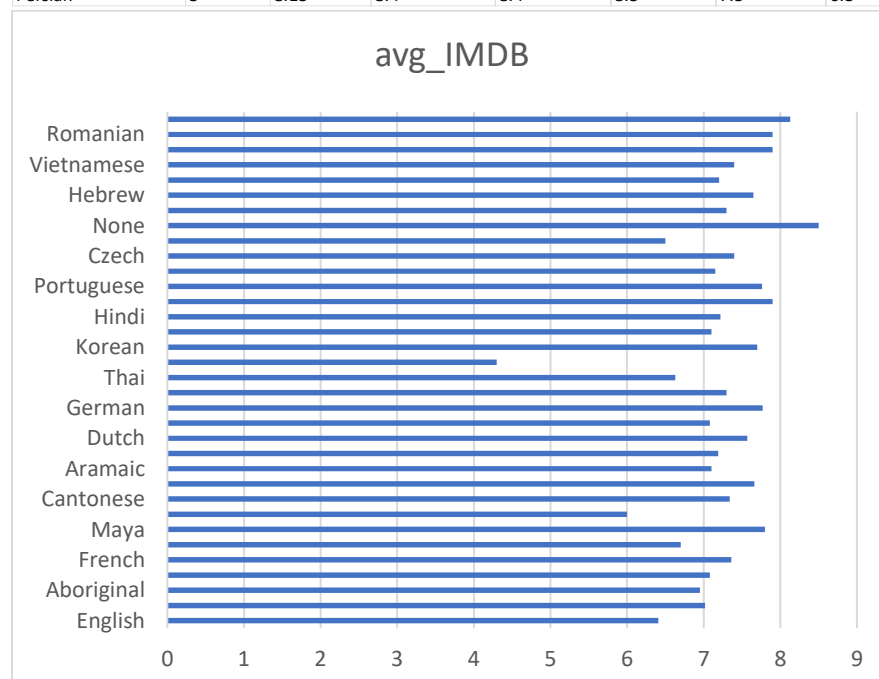
Analysis:

The analysis of the relationship between movie languages and their IMDB scores reveals that languages with a larger number of movies, such as English, French, and Japanese, tend to provide more reliable averages and variances, indicating a broad range of ratings and moderate consistency. English, with the largest dataset, shows a moderate average score and high variability, while Mandarin and Japanese display higher average scores with relatively lower variance, suggesting consistent quality. Conversely, languages with fewer movies, like Bosnian and Thai, exhibit extreme scores and perfect variance, but these results are less statistically significant due to small sample sizes. Overall, while some languages correlate with higher or lower average scores, the reliability of these correlations strongly depends on the number of movies available in each language.

## D) *Director Analysis:* **Influence of directors on movie ratings.**

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

Approach:

We start off by creating a pivot table of the director names and their average IMDB scores. We then sort them based on the average IMDB scores and extract the top 10 directors. A separate column is created and the function 'PERCENTRANK.INC' is used to calculate the percentile for each of the 10 directors. A stacked bar chart with all the three columns is created for visual analysis.

Output:

top 10 directors	avg_imdb	percentile
Akira Kurosawa	8.7	0.994
Tony Kaye	8.6	0.992
Michael Curtiz	8.6	0.992
Charles Chaplin	8.6	0.992
Ron Fricke	8.5	0.987
Majid Majidi	8.5	0.987
Damien Chazelle	8.5	0.987
Sergio Leone	8.4	0.983
Christopher Nolan	8.4	0.983
Richard Marquand	8.4	0.983
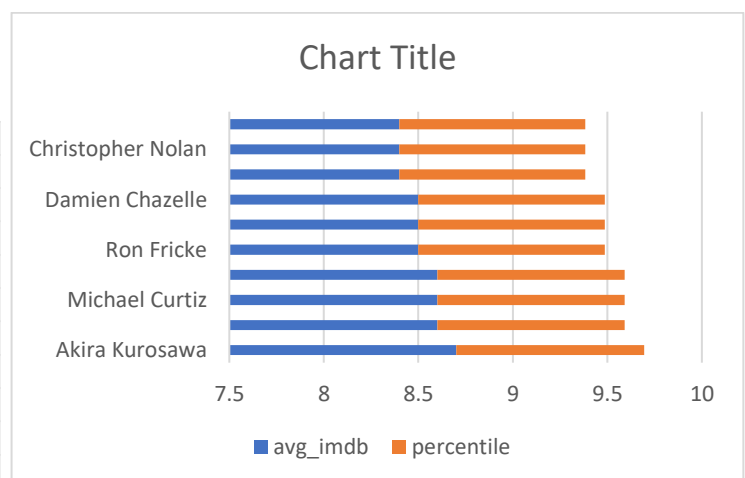
Analysis:

The analysis of the top 10 directors based on their average IMDB scores indicates a high level of consistent excellence in their filmmaking. Akira Kurosawa leads with an average score of 8.7 and is in the 99.4th percentile, showcasing his films' exceptional reception. Directors like Tony Kaye, Michael Curtiz, and Charles Chaplin, all scoring around 8.6, follow closely, reflecting their significant contributions to critically acclaimed cinema. The group includes a mix of directors known for both classic and contemporary films, such as Christopher Nolan and Sergio Leone, each with an average score of 8.4 and a percentile rank of 98.3. This high ranking across the board highlights these directors' consistent ability to produce highly rated films, solidifying their status as some of the most impactful directors in the industry.

E) *Budget Analysis:* **Explore the relationship between movie budgets and their financial success.**

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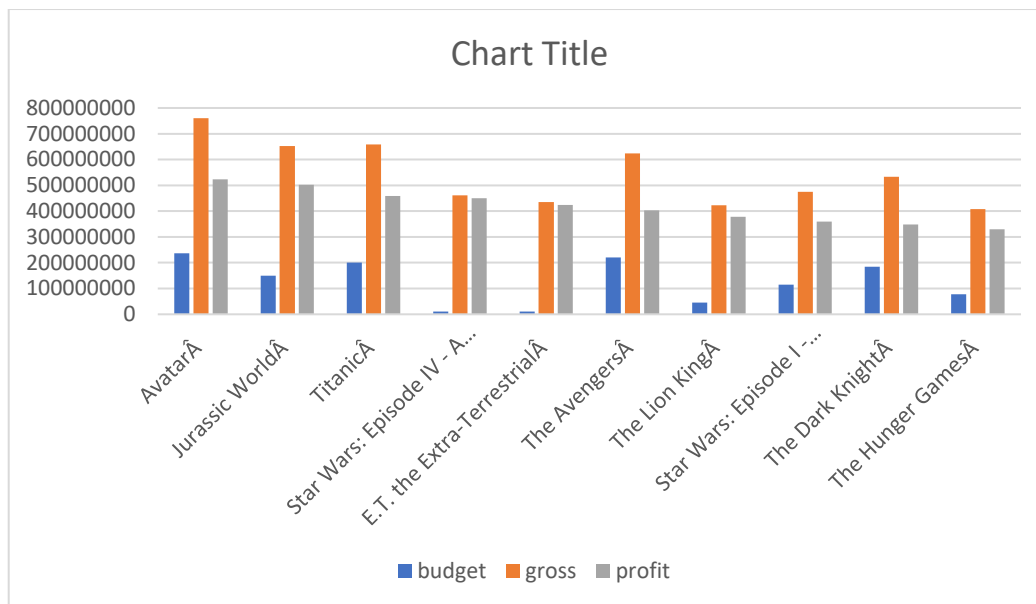
Approach:

We calculate the correlation coefficient between budgets and gross earnings by using the 'CORREL' function to analyse the correlation between them. We create a new column 'Profit' by subtracting gross earnings from the budgets and use the profits to sort the films in descending order to obtain the top 10 highest grossing films. We also create a clustered column chart to visualize the results.

Output:

correlation_coeff	0.095584

movie_title	budget	gross	profit
AvatarÂ	237000000	760505847	523505847
Jurassic WorldÂ	150000000	652177271	502177271
TitanicÂ	200000000	658672302	458672302
Star Wars: Episode IV - A New HopeÂ	11000000	460935665	449935665
E.T. the Extra-TerrestrialÂ	10500000	434949459	424449459
The AvengersÂ	220000000	623279547	403279547
The Lion KingÂ	45000000	422783777	377783777
Star Wars: Episode I - The Phantom MenaceÂ	115000000	474544677	359544677
The Dark KnightÂ	185000000	533316061	348316061
The Hunger GamesÂ	78000000	407999255	329999255

Chart Title

Legend: budget, gross, profit

Categories: AvatarÂ, Jurassic WorldÂ, TitanicÂ, Star Wars: Episode IV - A..., E.T. the Extra-TerrestrialÂ, The AvengersÂ, The Lion KingÂ, Star Wars: Episode I -..., The Dark KnightÂ, The Hunger GamesÂ

Analysis:

The table provides insight into the relationship between movie budgets and their financial success by detailing the budgets, gross earnings, and profits of ten popular films. It reveals that while higher budgets often correlate with higher gross earnings and profits, this is not a strict rule. For instance, "Avatar," with the highest budget of $237 million, achieved the highest gross earnings and profit, indicating a strong positive relationship between budget and financial success. However, movies with relatively lower budgets, such as "Star Wars: Episode IV - A New Hope" and "E.T. the Extra-Terrestrial," also achieved impressive profits, demonstrating that lower-budget films can also yield substantial financial success. Overall, the data suggests that while a higher budget can enhance a movie's potential for financial success, effective storytelling, strong franchises, and audience appeal are crucial factors that can drive profitability regardless of the initial budget.

RESULT:

Hence, we have implemented all the tasks given as a part of the IMDB Movie Analysis project and provided the approach to solve the tasks in MS Excel along with

the outputs generated.