Project Description

This Project involves analyzing what factors influence the success of a movie on IMDB. Here, success can be defined by high IMDB ratings. This analysis is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

As a data analyst, first I will understand the project requirements and purpose and then I will understand the data which is provided in the project attachment and will perform analysis to get the meaningful insights.

Analysis: involves understanding he relationships between different variables, and statistical analysis to look for impact of factors like factors like movie duration, budget, direction on IMDB scores, and correlation between movie ratings and other factors like genre, director, budget, etc
For example, low-budget films that have achieved tremendous financial success. This suggest that factors other than budget, such as story quality, directorial skill, popularity of actors and market timing, can also significantly impact a movie's financial success

Approach

1. **Five 'Whys' Approach**: This technique will help you dig deeper into the problem. For instance, if you find that movies with higher budgets tend to have higher ratings, you can ask "Why?" repeatedly to uncover the root cause.

2. **Identifying the Analysis**: Based on the questions, I identified the types of analysis required to perform, which included descriptive statistics, correlation analysis, trend analysis, comparative analysis, and percentile calculations.

3. **Descriptive Statistics**: Calculated measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation) to find the impact of movie Genre on movie ratings.

4. **Correlation Analysis**: I calculated to find correlation between the 'Budget' and 'Gross Income' of movies.

5. **Trend Analysis**: Created Scatter Plot chart to analyze the impact and trend of movie duration on IMDB scores

6. **Comparative Analysis**: From this analysis, we can see that the language of a movie can have an impact on its IMDB score. French and Spanish movies, for example, tend to have higher ratings compared to English, Hindi, and Mandarin movies.

Throughout this process, my goal was to provide clear, accurate, and helpful insights and answers to questions.

Tech-Stack Used

The software and versions you used for the project:

Microsoft Excel
Pivot Tables to classify the data and to calculate Mean, StdDev etc.
Power Query to filter and sort the data
Data analysis tool pack for descriptive statistics

Insights:

I will provide my insights in the form of answers for the questions posed by management team

1. 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.

Output:

Stats	Drama	Comedy	Thriller	Action	Romance
Average	4.890343309	3.593018796	6.1975361	0.8914137	3.5930188
Average b	6.76718107	6.19819413	5.590909	6.239896	5.883333
Median	6.9	6.3	6.4	6.3	6.5
Mode	7.2	6.7	6.1	6.1	6.5
Max	9.3	9.5	9	9.1	8.6
Min	0	0	2.2	0	2.1
Range	9.3	9.5	6.8	9.1	6.5
Variance	0.916173353	1.1890212	1.111811	1.250707	0.991704
StdDev	0.957169448	1.09042249	1.054425	1.11835	0.995844

1. **Mean:** If we categorize each movie under single genre and calculate average IMDB score for each genre, then Drama movies have high score comparatively. But the genre column has multiple categories for each movie. So, if we consider all categories for each movie then Thriller movies has highest score.

   So, the movies with Thriller as primary genre and with combination of other genres have highest IMDB Scores.

2. **Median:** The median rating is highest for the Drama and Romance genres and lowest for the Comedy and Action genres. This suggests that the “typical”

rating is higher for Drama and Romance movies than for Comedy and Action movies.

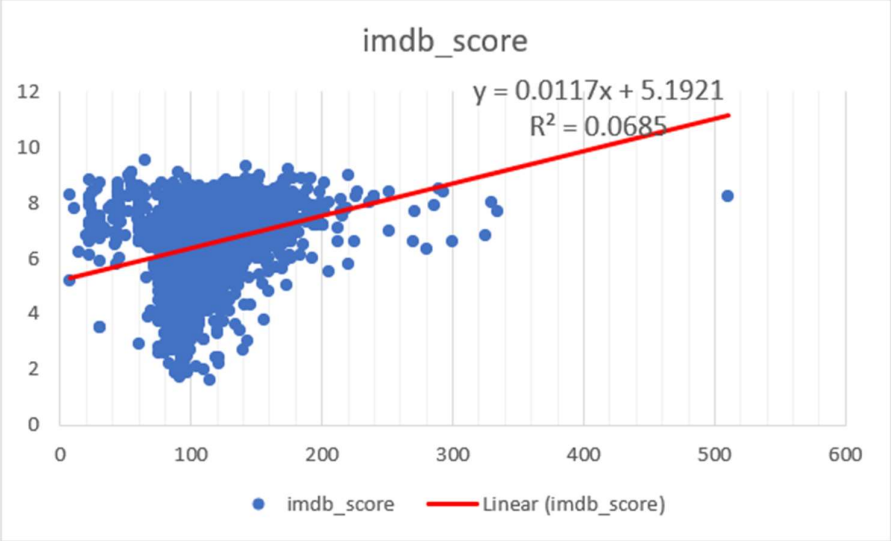
3. **Mode**: The mode (most common rating) is highest for the Drama genre and lowest for the Thriller and Action genres. This suggests that the most common rating is higher for Drama movies than for Thriller and Action movies.
4. **Max and Min**: The maximum rating is highest for the Comedy genre and lowest for the Thriller genre. The minimum rating is lowest (indicating a poor rating) for the Drama, Comedy, and Action genres. This suggests that Comedy movies have the potential to receive very high ratings, but Drama, Comedy, and Action movies also have the potential to receive very low ratings.
5. **Range**: The range (difference between the maximum and minimum ratings) is highest for the Drama and Comedy genres and lowest for the Romance genre. This suggests that the ratings for Drama and Comedy movies can vary widely, while the ratings for Romance movies are more consistent.
6. **Variance and Standard Deviation**: The variance and standard deviation are measures of how spread out the ratings are. They are highest for the Action genre and lowest for the Drama genre. This suggests that the ratings for Action movies can vary widely, while the ratings for Drama movies are more consistent.
7. Also, other factors such as the director, actors, budget, etc., can also impact movie ratings.

2. **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.

Output:

Distribution of movie durations and its impact on the IMDB score.



As per above scatter plot movies with duration between 90 minutes and 120 minutes have maximum number of ratings and also have high IMDB Score.

3. 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.

Output:

Movie Language	Count of language	Average of imdb_score	StdDevp of imdb_score	Varp of imdb_score	Median
English	4712	6.40	1.12	1.26	6.5
French	73	7.04	0.72	0.52	7.2
Spanish	40	6.94	0.84	0.71	7.15
Hindi	29	6.60	1.36	1.85	6.9
Mandarin	27	6.87	1.09	1.20	7.1
German	19	7.34	0.93	0.86	7.6
Japanese	18	7.39	0.96	0.93	7.6
Russian	12	6.33	1.27	1.61	6.25
Italian	11	7.23	1.19	1.41	7.3
Cantonese	11	6.95	0.67	0.45	7.2
Portuguese	8	7.49	0.83	0.68	7.7
Korean	8	7.39	0.77	0.60	7.5
Swedish	6	7.53	0.65	0.43	7.8
Hebrew	5	7.58	0.30	0.09	7.6
Danish	5	7.50	0.96	0.93	8.1
Arabic	5	7.28	0.70	0.63	7.4

1. **Average**: The average IMDB score is highest for movies in the Polish language and lowest for movies in the Chinese language. This suggests that, on average, Polish movies receive higher ratings than Chinese movies.
2. **Median**: The median IMDB score is highest for movies in the Danish and Persian languages and lowest for movies in the Russian language. This suggests that the “typical” rating is higher for Danish and Persian movies than for Russian movies.
3. **Mode**: The mode (most common rating) is highest for movies in the Polish language and lowest for movies in the Thai and Chinese languages. This suggests that the most common rating is higher for Polish movies than for Thai and Chinese movies.
4. **Max and Min**: The maximum IMDB score is highest for movies in the Polish language and lowest for movies in the Thai and Chinese languages. This suggests that Polish movies have the potential to receive very high ratings, but Thai and Chinese movies also have the potential to receive very low ratings.
5. **Range**: The range (difference between the maximum and minimum ratings) is highest for movies in the Hindi language and lowest for movies in the Zulu, Icelandic, Romanian, Aboriginal, Indonesian, and None languages. This suggests that the ratings for Hindi movies can vary widely, while the ratings for Zulu, Icelandic, Romanian, Aboriginal, Indonesian, and None movies are more consistent.

6. **Variance and Standard Deviation**: The variance and standard deviation are measures of how spread out the ratings are. They are highest for movies in the Hindi language and lowest for movies in the Hebrew language. This suggests that the ratings for Hindi movies can vary widely, while the ratings for Hebrew movies are more consistent.

Please note that these are general trends and there may be exceptions. Also, other factors such as the director, actors, budget, etc., can also impact movie ratings.

4. 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.

Output:

Row Labels	Average of imdb_score	Count of movie_title	90th Percenti	25th Percent	50th Percent	75th Percent
Steven Spielberg	7.48	26	8.4	7.1	7.6	7.975
Woody Allen	7.01	22	7.7	6.6	7.0	7.375
Martin Scorsese	7.66	20	8.3	7.2	7.5	8.2
Clint Eastwood	7.23	20	8.1	6.6	7.3	7.825
Ridley Scott	7.07	17	8.3	6.6	7.0	7.8
Steven Soderbergh	6.71	16	7.5	6.3	7.0	7.125
Tim Burton	6.93	16	7.9	6.5	7.0	7.525
Spike Lee	6.57	16	7.7	6.4	6.6	7.075
Renny Harlin	5.75	15	6.6	5.5	5.7	6.35
Oliver Stone	6.95	14	7.9	6.4	7.2	7.375
Ron Howard	6.93	13	8.1	6.2	6.7	7.7
Sam Raimi	6.91	13	7.6	6.4	6.6	7.5
Robert Rodriguez	5.69	13	7.3	4.9	5.6	6.9
Robert Zemeckis	7.31	13	8.4	6.6	7.4	7.7
Michael Bay	6.64	13	7.3	6.3	6.6	6.9
Joel Schumacher	6.41	13	7.4	6.4	6.6	7.1
John Carpenter	6.92	13	7.9	6.1	7.2	7.9
Barry Levinson	6.58	13	7.5	6.0	6.6	7.2

1. **Most Prolific Directors**: Steven Spielberg and Woody Allen have directed the most movies, with 26 and 22 respectively.
2. **Highest Average IMDb Score**: David Fincher has the highest average IMDb score of 7.75.
3. **90th Percentile Performance**: Francis Ford Coppola's movies have the highest score at the 90th percentile, with a score of 9.0.
4. **Consistent High Performers**: Directors like Steven Spielberg, Martin Scorsese, and Peter Jackson consistently have high scores across all percentiles.

5. 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

Output:

Correlation b\n Budget and Gross	0.102179454
AvatarÂ	₹ 52,35,05,847.00

1. **Highest Grossing Movies**: 'Avatar', 'Jurassic World', and 'Titanic' are the top three movies in terms of gross revenue.
2. **Most Profitable Movies**: 'Avatar', 'Jurassic World', and 'Titanic' also have the highest profit margins.
3. **Budget vs Profit**: Movies like 'Star Wars: Episode IV - A New Hope' and 'E.T. the Extra-Terrestrial' have relatively low budgets but high profit margins, indicating a good return on investment.

Result

1. **Director Influence:** Directors like Steven Spielberg, Martin Scorsese, and David Fincher have high average IMDb scores, indicating that their movies are generally well-received.
2. **Movie Count:** Directors like Steven Spielberg and Woody Allen have a high count of movies, suggesting they are prolific in the industry.
3. **Score Distribution:** The 25th, 50th, and 75th percentiles give us an idea about the distribution of IMDb scores for each director's movies. For instance, Martin Scorsese's movies have a median score of 7.5, with most scores falling between 7.2 (25th percentile) and 8.2 (75th percentile).
4. **Top Performers:** The 90th percentile shows the score below which 90% of a director's movie ratings fall. Directors like Steven Spielberg, Martin Scorsese, and Francis Ford Coppola have high 90th percentile scores, indicating that they have made several highly-rated movies.

.

Drive Link

https://docs.google.com/spreadsheets/d/1ktUHMcOet-Q6R2JZI1f3Bvrd2DBAkCFv/edit?usp=sharing&ouid=106678072727235537296&rtpof=true&sd=true