IMDB MOVIE ANALYSIS

PROJECT DESCRIPTION

The IMDB Movie Analysis project aims to decipher the factors that influence movie success on IMDB by examining relationships between variables like genre, duration, language, directors, and budgets. By leveraging data cleaning and statistical analysis, this project seeks to provide actionable insights for stakeholders in the film industry. Through comprehensive exploration and visualization, the project will unravel key insights into what drives high IMDB ratings, empowering decision-makers to make informed choices in their future movie projects.

Approach

Given the manageable nature of the IMDB Project's dataset, I opted for MS Excel as my tool of choice. The challenges at hand are effectively addressable through Excel's diverse functions and tools. Consequently, I intend to seamlessly import and cleanse the data utilizing Power Query. By leveraging this clean dataset, I'll undertake a transformative journey, converting it into a treasure trove of insightful information through various analytical processes.

Tech-Stack Used

- Excel: Used for data cleaning, analysis, and visualization.
- Power Query: Utilized for data transformation and cleaning.
- Descriptive Statistics Functions: AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, STDEV.
- Scatter Plots: Created scatter plots with trendlines to visualize relationships

Insights

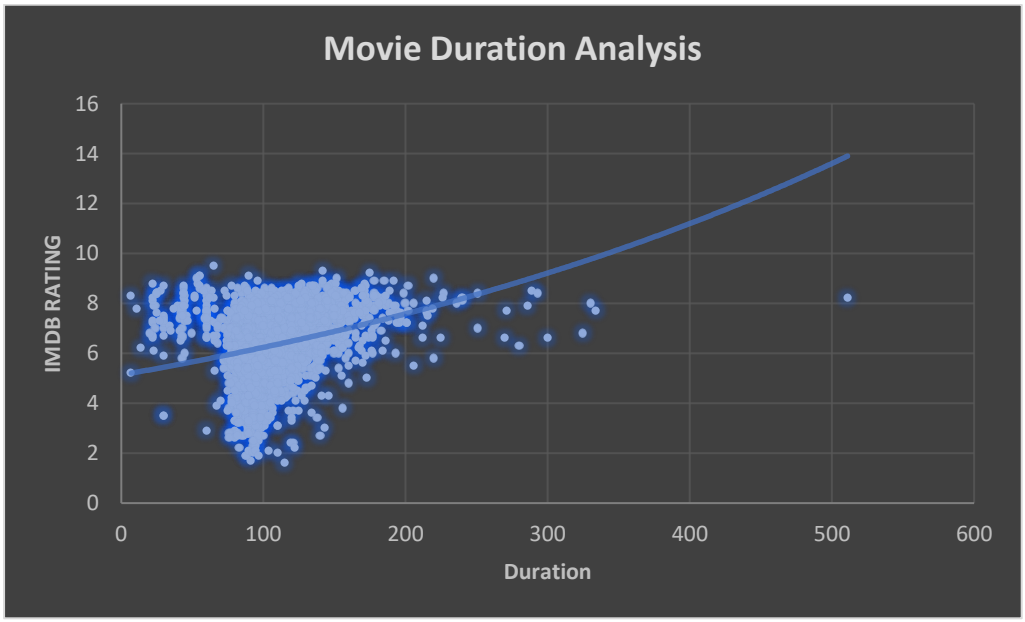
1) **Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.
   Answer:

Genre	No. of Movies	Mean	Median	Mode	Range	Varience	Standard Deviation
Drama	2594	6.763763	6.9	7.2	7.3	0.916173	1.118349975
Comedy	1872	6.195246	6.3	6.7	7.8	1.189021	1.05216187
Thriller	1411	6.314245	6.4	6.1	6.8	1.110832	1.13058319
Action	1153	6.239896	6.3	6.1	7.4	1.250707	0.957169448
Romance	1107	6.450587	6.5	6.5	6.5	0.99119	1.137237831

Adventure	923	6.44117	6.6	6.7	7	1.278218	1.090422487
Crime	889	6.564792	6.6	6.6	6.9	1.052427	1.089667399
Sci-Fi	616	6.281818	6.4	6.7	6.9	1.463695	1.159734063
Fantasy	610	6.307049	6.4	6.7	7.2	1.344983	1.025878858
Horror	565	5.84354	5.9	6.2	6.5	1.275697	0.721281783
Family	546	6.245055	6.4	6.7	7	1.441193	1.209832797
Mystery	500	6.4864	6.6	6.6	6.4	1.187375	1.129467664
Biography	293	7.150171	7.2	7	4.4	0.520247	0.995585159
Animation	242	6.576033	6.7	6.7	6.9	1.29331	1.053960057
Music	214	6.41028	6.6	6.5	6.9	1.383165	0
War	213	7.070423	7.1	7.1	5.9	0.76152	1.200497187
History	207	7.083575	7.2	7.5	6.9	0.784561	1.1760805
Sport	182	6.606044	6.8	7.2	6.7	1.207601	1.037093931
Musical	132	6.507576	6.7	7	6.4	1.491003	1.221066425
Documentary	121	7.180165	7.4	7.5	7.1	1.107045	0.394405319
Western	97	6.689691	6.8	6.5	5.1	1.075564	0.885754556

2) Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

Answer:



Mean	Median	Standard Deviation
107.2011	103	25.19493497

The standard deviation of 25.19493497 indicates the variability or spread of movie durations around the mean. A larger standard deviation suggests that movie durations

vary more widely from the average duration. In the context of movie durations, a higher standard deviation might imply that there's a wider range of durations, including both shorter and longer movies.

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

Answer:

Language	No.of Movies	Mean	Median	Standard Deviation
English	4704	6.398427	6.5	1.121948655
French	73	7.038356	7.2	0.721989287
Spanish	40	6.9375	7.15	0.844300746
Hindi	28	6.632143	6.95	1.37374711
Mandarin	26	6.788462	7.05	1.021810958
German	19	7.342105	7.6	0.928675225
Japanese	18	7.394444	7.6	0.962907762
Russian	11	6.363636	6.5	1.319278541
Cantonese	11	6.954545	7.2	0.671989768
Italian	11	7.227273	7.3	1.186354929
Korean	8	7.3875	7.5	0.772071078
Portuguese	8	7.4875	7.7	0.826797285
Hebrew	5	7.58	7.6	0.299332591
Swedish	5	7.44	7.6	0.677052435
Danish	5	7.5	8.1	0.963327566
Arabic	5	7.38	7.4	0.790948797
Dutch	4	7.425	7.45	0.376662979
Polish	4	8.25	8.25	0.85
Norwegian	4	7.15	7.3	0.497493719
Persian	4	7.575	7.95	1.042532973
Chinese	3	5.666667	5.7	0.449691252
Thai	3	6.633333	6.6	0.368178701
Aboriginal	2	6.95	6.95	0.55
Icelandic	2	7.55	7.55	0.65
Dari	2	7.5	7.5	0.1
None	2	7.95	7.95	0.55
Zulu	2	7.1	7.1	0.2
Indonesian	2	7.9	7.9	0.3
Romanian	2	7.2	7.2	0.7
Filipino	1	6.7	6.7	0
Maya	1	7.8	7.8	0
Kazakh	1	6	6	0
Telugu	1	8.4	8.4	0
Aramaic	1	7.1	7.1	0

Mongolian	1	7.3	7.3	0
Bosnian	1	4.3	4.3	0
Hungarian	1	7.1	7.1	0
Czech	1	7.4	7.4	0
Kannada	1	7.1	7.1	0
Panjabi	1	6.6	6.6	0
Tamil	1	5.1	5.1	0
Dzongkha	1	7.5	7.5	0
Vietnamese	1	7.4	7.4	0
Urdu	1	7	7	0
Slovenian	1	6.4	6.4	0
Greek	1	7.3	7.3	0
Swahili	1	7.4	7.4	0

The language analysis of the dataset reveals intriguing insights into the impact of language on movie ratings. English dominates the dataset with 4704 movies and a mean IMDB score of 6.40, while French, Spanish, and Hindi follow with smaller counts but relatively high mean scores of 7.04, 6.94, and 6.63 respectively. Interestingly, movies in German, Japanese, and Mandarin exhibit higher average ratings of 7.34, 7.39, and 6.79, suggesting that language diversity doesn't necessarily equate to lower scores. Notably, languages with fewer representations like Bosnian and Kazakh have single movies with relatively lower ratings, while languages with a handful of movies like Portuguese and Hebrew garner impressively high mean scores of 7.49 and 7.58. This analysis showcases that language indeed plays a role in movie ratings, where movies in various languages can achieve both high and moderate success, emphasizing the universal appeal of exceptional storytelling transcending linguistic barriers.

4) Director Analysis: Influence of directors on movie ratings.

Answer:

Directors	AVG Rating	Director
John Blanchard	9.5	Top Director
Mitchell Altieri	8.7	Top Director
Sadyk Sher-Niyaz	8.7	Top Director
Cary Bell	8.7	Top Director
Mike Mayhall	8.6	Top Director
Charles Chaplin	8.6	Top Director
Raja Menon	8.5	Top Director
Ron Fricke	8.5	Top Director
Damien Chazelle	8.5	Top Director
Majid Majidi	8.5	Top Director
Sergio Leone	8.475	Top Director
Christopher Nolan	8.425	Top Director
S.S. Rajamouli	8.4	Top Director
Moustapha Akkad	8.4	Top Director
Richard Marquand	8.4	Top Director
Catherine Owens	8.4	Top Director
Rakeysh Omprakash Mehra	8.4	Top Director
Jay Oliva	8.4	Top Director
Robert Mulligan	8.4	Top Director
Asghar Farhadi	8.4	Top Director
Marius A. Markevicius	8.4	Top Director
Bill Melendez	8.4	Top Director

The top one percentage of the best directors is given.

5) Budget Analysis: Explore the relationship between movie budgets and their financial success

Answer:

Profit Margin	movie_title
523505847	Avatar
502177271	Jurassic World
458672302	Titanic
449935665	Star Wars: Episode IV - A New Hope
424449459	E.T. the Extra-Terrestrial
403279547	The Avengers
403279547	The Avengers
377783777	The Lion King
359544677	Star Wars: Episode I - The Phantom Menace
348316061	The Dark Knight

These are the top 10 Movies with highest profit margin.

Since the correlation coefficient is close to -1, you can infer that there is a strong tendency for movies or projects with higher budgets to have lower collections, and vice versa. This could imply that investing more money in a project does not necessarily guarantee higher financial success or returns.

RESULT

This project helped me to recollect all the functions and formulas in Excel and to use the charts and tables for visualising the data.