IMDB Movie Analysis

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:-

Functions used are average, mode, median, variance, Standard Deviation and range.

Genre	Mean	Median	Mode	Variance	Stdev	Range	Count of Movies
Action	6.239896	6.6	6.7	1.210583	1.100265	7.9	1153
Adventure	6.441278	6.6	6.7	1.27952	1.131159	7	923
Fantasy	6.307049	6.4	6.7	1.347192	1.160686	7.2	610
Sci-Fi	6.280032	6.4	6.7	1.465471	1.210566	6.9	616
Thriller	6.314245	6.4	6.1	1.1116196	1.054334	6.8	1411
Documentary	7.180165	7.4	7.5	1.11627	1.056537	7.1	121
Romance	6.450587	6.5	6.5	0.992086	0.996035	6.5	1107
Animation	6.576033	6.7	6.7	1.2986763	1.139595	6.9	242
Comedy	6.194658	6.3	6.7	1.1893568	1.090576	7.8	1872
Family	6.245055	6.4	6.7	1.4438379	1.201598	7	546

Musical	6.507576	6.7	7	1.5023849	1.225718	6.4	132
Mystery	6.4864	6.6	6.6	1.1897545	1.090759	6.4	500
Western	6.689691	6.8	6.5	1.0867676	1.042481	5.1	97
Drama	6.763338	6.9	7.2	0.9167002	0.957445	7.3	2594
History	7.084058	7.2	7.5	0.787366	0.887336	6.9	207
Sport	6.606044	6.8	7.2	1.2142727	1.10194	6.7	182
Crime	6.564679	6.6	6.6	1.0534583	1.026381	6.9	889
Horror	5.841593	5.9	6.2	1.2755897	1.12942	6.5	565
War	7.070423	7.1	7.1	0.7651116	0.874707	5.9	213
Biography	7.150171	7.2	7	0.5220291	0.722516	4.4	293
Music	6.41028	6.6	6.5	1.3896591	1.178838	6.9	214
Game-Show	2.9	2.9	0	0	0	0	1
Reality-TV	4.75	4.75	0	6.845	2.616295	3.7	2
News	7.533333	7.4	0	0.2633333	0.51316	1	3
Short	6.38	6.5	0	0.557	0.746324	1.9	5
Film-Noir	7.633333	7.65	0	0.1866667	0.432049	1.1	6

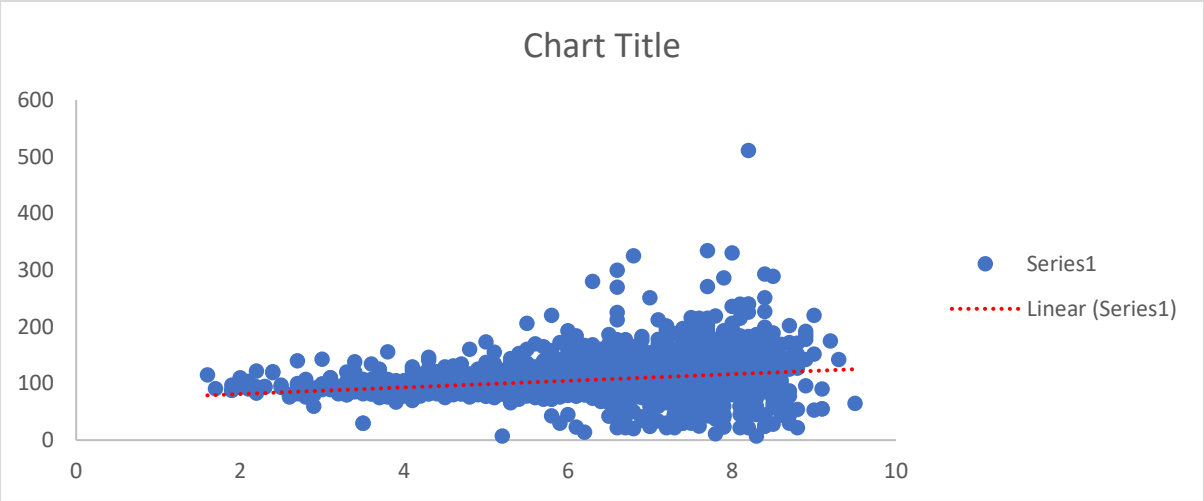
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

- Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Output:-

Functions used are average, median and Standard Deviation.

Quartile, Inter-Quartile Range, Lower Limit, Higher Limit



Scatter plot showing a correlation between duration of movie and it's IMDB Rating.

As correlation coefficient is 0.260886 and scatterplot is also showing there is week correlation between duration of movie and it's imdb rating.

Mean	Median	Mode	Standard Deviation	Variance
107.2011	103	90	25.19744	634.911

Quartile 1	93
Quartile 2	103
Quartile 3	118
Inter Quartile Range	25
Lower limit	55.5

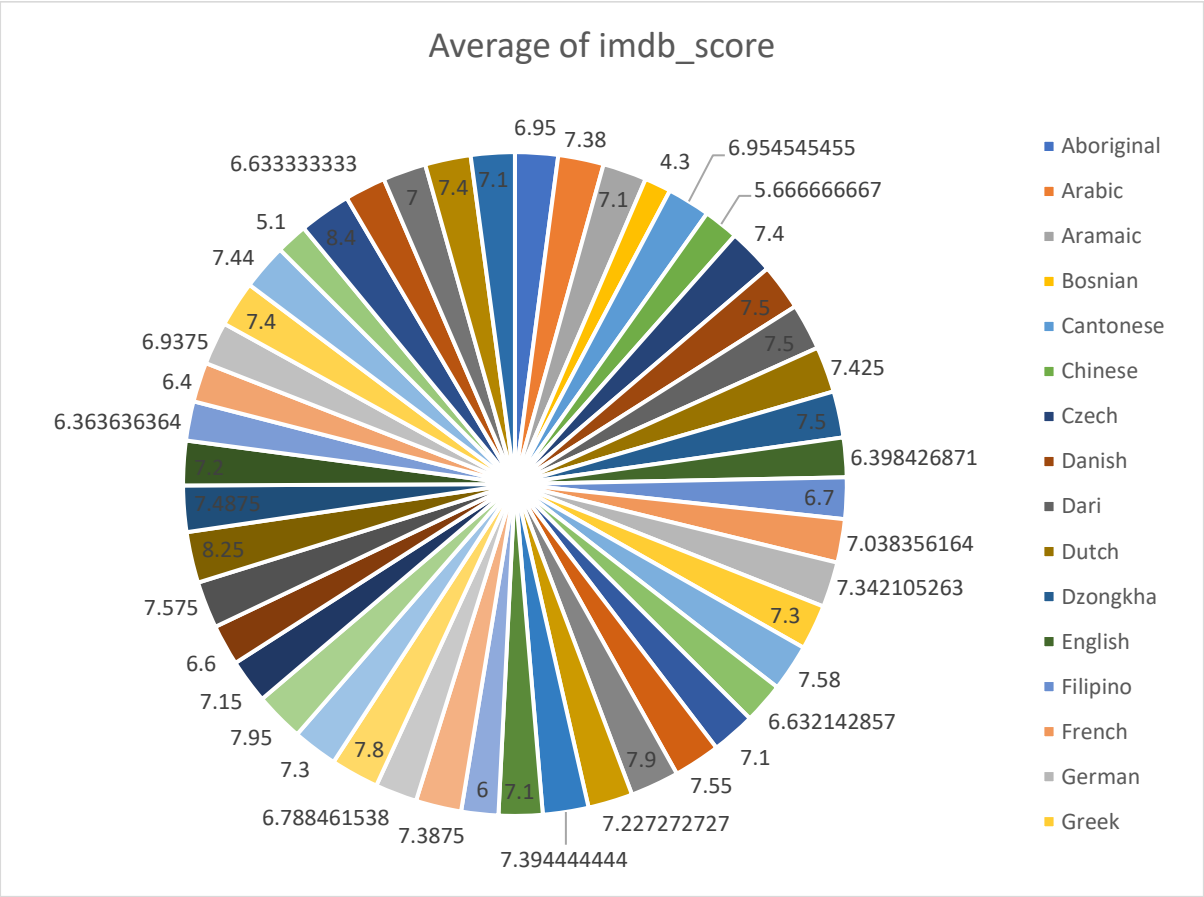
Higher Limit	155.5

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.

Output:-

Functions used Countif, Averageif



Pie Chart showing Language correlation with imdb rating.

Row Labels	Average of imdb_score	Count of movie_title
Aboriginal	6.95	2
Arabic	7.38	5
Aramaic	7.1	1

Bosnian	4.3	1
Cantonese	6.954545455	11
Chinese	5.666666667	3
Czech	7.4	1
Danish	7.5	5
Dari	7.5	2
Dutch	7.425	4
Dzongkha	7.5	1
English	6.398426871	4704
Filipino	6.7	1
French	7.038356164	73
German	7.342105263	19
Greek	7.3	1
Hebrew	7.58	5
Hindi	6.632142857	28
Hungarian	7.1	1
Icelandic	7.55	2
Indonesian	7.9	2
Italian	7.227272727	11
Japanese	7.394444444	18
Kannada	7.1	1
Kazakh	6	1
Korean	7.3875	8
Mandarin	6.788461538	26
Maya	7.8	1
Mongolian	7.3	1
None	7.95	2
Norwegian	7.15	4
Panjabi	6.6	1
Persian	7.575	4
Polish	8.25	4
Portuguese	7.4875	8
Romanian	7.2	2
Russian	6.363636364	11
Slovenian	6.4	1
Spanish	6.9375	40
Swahili	7.4	1
Swedish	7.44	5
Tamil	5.1	1
Telugu	8.4	1
Thai	6.633333333	3
Urdu	7	1
Vietnamese	7.4	1
Zulu	7.1	2

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.

Output:-

Function used are average and Percentrank.exc

It shows top 20 directors with highest average movie imdb ratings in percentile and percentage.

director_name	Average imdb percentile	Percentile	Percentage
John Blanchard	9.5	0.999	100%
Mitchell Altieri	8.7	0.998	100%
Sadyk Sher-Niyaz	8.7	0.998	100%
Cary Bell	8.7	0.998	100%
Mike Mayhall	8.6	0.997	100%
Charles Chaplin	8.6	0.997	100%
Raja Menon	8.5	0.995	100%
Ron Fricke	8.5	0.995	100%
Damien Chazelle	8.5	0.995	100%
Majid Majidi	8.5	0.995	100%
Sergio Leone	8.475	0.995	100%
Christopher Nolan	8.425	0.994	99%
S.S. Rajamouli	8.4	0.99	99%
Moustapha Akkad	8.4	0.99	99%
Richard Marquand	8.4	0.99	99%
Catherine Owens	8.4	0.99	99%
Rakeysh Omprakash Mehra	8.4	0.99	99%
Jay Oliva	8.4	0.99	99%
Robert Mulligan	8.4	0.99	99%
Asghar Farhadi	8.4	0.99	99%
Marius A. Markevicius	8.4	0.99	99%
Bill Melendez	8.4	0.99	99%

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.

Function used are correl, max().

Correlation coeffecient	0.102179		
Movie with highest profit margin		523505847	AvatarÂ

As high budgets give movies with high profit margin but it also has exceptions such as The Host.

"To summarize, our analysis revealed:

1. **Film-Noir and News are high-performing genres, however movie with highest rating is Towering Inferno which is a comedy movie with rating of 9.5.**
2. **The ideal movie duration is 103-118 minutes.**
3. **Language diversity can boost ratings.**
4. **Top directors consistently produce successful movies.**
5. **Higher budgets often lead to higher earnings but aren't the sole factor."**

Related Excel File