IMDB
Movie Analysis

SPREADSHEET LINK

HTTPS://DOCS.GOOGLE.COM/SPREADSHEETS/D/1AEBL_J8Q9AG_4HKS9UJXJNPLWZFRAOAD/EDIT?USP=DRIVE_LINK&OUID=104966616771126772005&RTPOF=TRUE&SD=TRUE

VIDEO LINK

Project Description

My major responsibility as a data analyst in a multinational corporation is to investigate "What factors influence the success of a movie on IMDB?"

One of my responsibilities is to draw important conclusions from the dataset. These observations will provide useful information to different business teams. For filmmakers, financiers, and producers of motion pictures who wish to know what makes a picture successful so they may make wise choices for their upcoming productions, this issue is a significant turn.

The dataset that is offered has been used with IMDB Movies. One area of possible interest for further research would be: "What factors influence a movie's success on IMDB?" In this case, having high IMDB ratings is a sign of success. For filmmakers, producers, and financiers who wish to know what makes a film successful so they may make wise choices for their upcoming productions, this issue has a big impact.

Approach

- Downloading the dataset: Open Excel on your device and download the file (.csv). Verify that the file you downloaded has the (.xlsx) extension.

- Understanding the worksheet: Examining the table's structure that contains the data in the Excel sheet is the next step(movies.csv from IMDB).

- Finding the important tables: Determining the main key from the IMDB_movies dataset.

- The preparation stage known as "data cleaning" prepares the data for analysis. It covers handling null values and eliminating duplicates.

Data Analysis Tasks

A. Movie Genre Analysis

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.

- The red cells show the maximum values in the table and the green highlighted cells show the minimum value in the table. The most common genre as **"Drama"**.

MOVIE GENRE ANALYSIS

| Genres	Occurrence	Mean	Median	Mode	Maximum	Minimum	Variance	Std.Deviation
Action	959	6.29		6.6	9	2.1	1.08	1.04
Adventure	781	6.45	6.6	6.7	8.9	2.3	1.24	1.11
Animation	0	6.70	6.8	6.7	8.6	2.8	0.97	0.99
Biography	239	7.16	7.2	7	8.9	4.5	0.48	0.69
Crime	709	6.55	6.6	6.6	9.3	2.4	0.97	0.98
Comedy	1461	6.19	6.3	6.7	8.8	1.9	1.07	1.04
Documentary	45	6.99	7.4	7.6	8.5	1.6	1.92	1.38
Drama	1893	6.79	6.9	6.7	9.3	2.1	0.80	0.90
Family	442	6.21	6.3	6.7	8.6	1.9	1.35	1.16
Fantasy	507	6.28	6.4	6.7	8.9	2.2	1.29	1.13
History	149	7.16	7.2	7.7	8.9	5.5	0.45	0.67
Horror	392	5.92	6	5.9	8.6	2.3	1.00	1.00
Musical	96	6.60	6.75	7.1	8.5	2.1	1.21	1.10
Music	233	6.46	6.6	6.2	8.5	1.6	1.40	1.18
Mystery	384	6.47	6.5	6.6	8.6	3.1	1.03	1.02
Romance	859	6.44	6.5	6.5	8.5	2.1	0.91	0.95
Sci-Fi	496	6.33	6.4	6.7	8.8	1.9	1.35	1.16
Sport	148	6.59	6.8	7.2	8.3	2	1.09	1.04
Thriller	1117	6.38	6.4	6.5	9	2.7	0.94	0.97
War	152	7.06	7.1	7.1	8.6	4.3	0.64	0.80
Western	59	6.79	6.8	6.8	8.9	4.7	0.87	0.93

Genre vs Occurrence
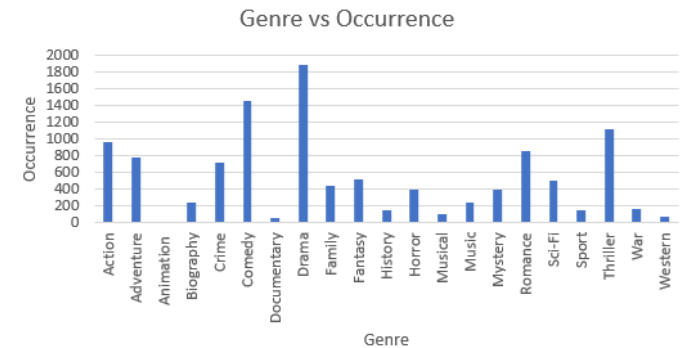
B. Movie Duration Analysis

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

MOVIE DURATION ANALYSIS	
Mean	110.26
Median	106.00
Std. Dev	22.65

- Mostly the movies are of mean duration **110 min**.(Approx 2 hrs.)

Duration vs IMDB Score
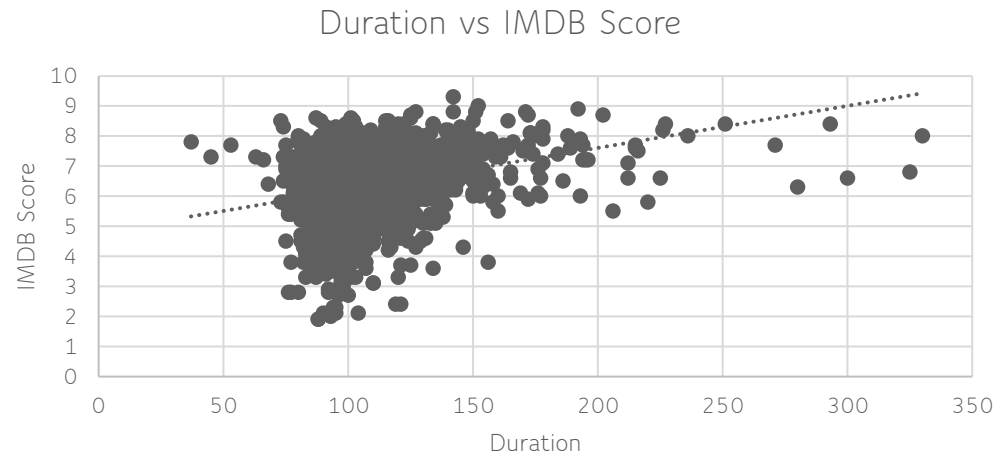
C. Language Analysis

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

LANGUAGE ANALYSIS			
Language	Mean	Median	Std. Dev
English	6.43	6.5	1.05
Mandarin	7.08	7.4	0.77
Aboriginal	6.95	6.95	0.78
Spanish	7.08	7.2	0.86
French	7.36	7.3	0.52
Filipino	6.70	6.7	#DIV/0!
Maya	7.80	7.8	#DIV/0!
Kazakh	6.00	6	#DIV/0!
Cantonese	7.34	7.3	0.35
Japanese	7.66	8	0.99
Aramaic	7.10	7.1	#DIV/0!
Italian	7.19	7	1.16
Dutch	7.57	7.8	0.40
Dari	7.50	7.4	0.74
German	7.77	7.8	0.71
Mongolian	7.30	7.3	#DIV/0!
Thai	6.63	6.6	0.45
Bosnian	4.30	4.3	#DIV/0!
Korean	7.70	7.7	0.57
Hungarian	7.10	7.1	#DIV/0!
Hindi	7.22	7.4	0.80
Danish	7.90	8.1	0.53
Portuguese	7.76	8	0.98
Norwegian	7.15	7.3	0.57
Czech	7.40	7.4	#DIV/0!
Russian	6.50	6.5	#DIV/0!
None	8.50	8.5	#DIV/0!
Zulu	7.30	7.3	#DIV/0!
Hebrew	8.00	8	#DIV/0!
Arabic	7.20	7.2	#DIV/0!
Vietnamese	7.40	7.4	#DIV/0!
Indonesian	7.90	7.9	0.42
Romanian	7.90	7.9	#DIV/0!
Persian	8.13	8.4	0.55
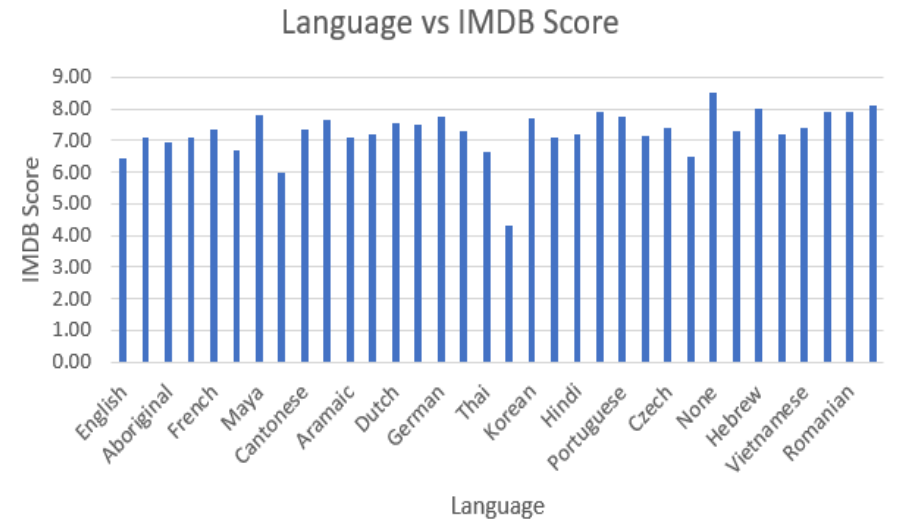
- **English** is the most occurred language.

#DIV/0! Error occurs due to STDEV function. This function can only perform with more than two available values.

D. Director Analysis

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

DIRECTOR ANALYSIS				Percentile	Percentile Share
Director Name ▼	**No. of Movies** ▼	**Average IMDB Score** ▼			
Akira Kurosawa	1	8.70		100.00%	8.70
Tony Kaye	1	8.60		90.00%	7.50
Charles Chaplin	1	8.60		80.00%	7.10
Alfred Hitchcock	1	8.50		70.00%	6.90
Ron Fricke	1	8.50		60.00%	6.68
Damien Chazelle	1	8.50		50.00%	6.45
Majid Majidi	1	8.50		40.00%	6.20
Sergio Leone	3	8.43		30.00%	5.90
Christopher Nolan	8	8.43		20.00%	5.60
Richard Marquand	1	8.40		10.00%	5.10
Asghar Farhadi	1	8.40		0.00%	2.10

- The most popular director based on their average movie rating is **Akira Kurosawa.**
- 100 percentile score among average movie rating is **8.70**

E. Budget Analysis

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

BUDGET ANALYSIS
Movie with heighest profit margi Avatar	
Profit Made	523505847
Corelation Coefficient	0.099

The movie with the highest profit is **"Avatar"** with a profit of **523505847** dollars.

The correlation coefficient between the gross and budget came out to be **0.099**

Tech Stack Used

- Microsoft Excel: It is a spreadsheet program from Microsoft and a component of its Office product for business applications. This enables users to format, calculate and organize data in a spreadsheet.

- MS Excel Functions: They are predefined formulas that perform calculations by using specific values, called arguments, in a particular order or structure. Some of the functions are:

- Text functions: clean(), substitute(), replace(), concatenate(), trim(), search(), find(), etc.

- Mathematical and Statistical functions: sum(), sumif(), count(), countif(), round(), avg(), min(), max(), subtotal(),averageif(),median(),mode(), etc.

- Data Visualization in Excel: Bar, Column, Line, Histogram, Pie, Scatter, Boxpot, Treemap, etc.

Insights

- We were able to determine the most common genres of movies dataset. We also added the descriptive statistics for each genre.

- To identify the relationship between the movie duration and the IMDB score. The most mean movie duration came out to be 110 minutes.

- Determining the most common languages used in movies and how language impacts the IMDB score.

- Checking the influence of the movie directors on the IMDB ratings. The most popular director came out to be Akira Kurosawa.

- Analyzing the correlation between the movie budgets and their financial success(gross income).

Results

- These learned insights helped me understand specific business questions which were addressed by MS Excel.

- Learning about Excel Text and Statistical functions. The importance of max(), average(), min(), text() functions.

- We were able to build different charts for visualization for answering the business questions. Some of the charts used were scatter plot and line graph.

- Achieving the ability to learn and write MS Excel functions to execute different business questions.

- Solving Company related problems using different visualization charts offered by MS Excel.

Thankyou.