

analysing **IMBD** **Movies**

- ANJALI SAHNI

Project Description



This project aims to **analyze** a dataset containing information about **IMDB-listed movies** to **determine** which **factors** are most **strongly associated** with **higher ratings**.



The **goal** is to **identify patterns** that can **help** industry professionals—like **producers, directors, and investors**—understand what contributes to a movie's **positive reception** and better **decision-making**.

Approach and Tech Stack used



For this project, **Microsoft Excel 2022** was selected as the primary tool for **data analysis** due to its **versatility** and **robust capabilities** in handling tabular data.

Specific techniques such as **pivot tables**, **charts**, and **formulas** were employed to analyze the data.

I have used **canva** for making this presentation as it contains required Elements, Graphs, Images which made this project more **attractive**.

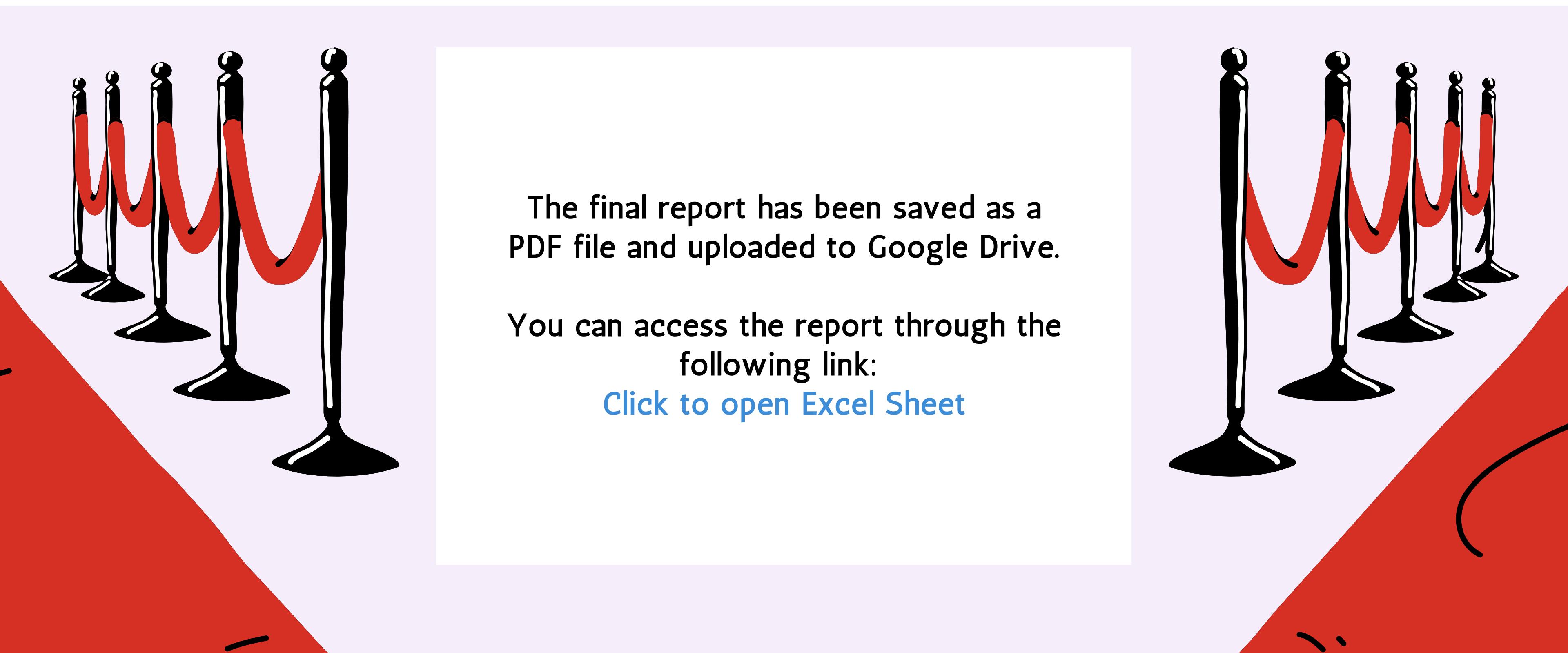


Insights and Results

- Movies of certain **genres** and **directors** often receive **higher IMDB ratings**.
- **Well-known actors** tend to **boost** audience ratings and movie **popularity**.
- Films with a **runtime** between 90 to 120 minutes are generally more **well-received**.
- A large budget doesn't guarantee high ratings—**content quality is key**.



Drive Link

A decorative graphic at the bottom of the slide features a red carpet (represented by a wavy red line) leading up a set of black stairs. The stairs have black railings and black spherical tops. The background behind the carpet and stairs is a light purple color.

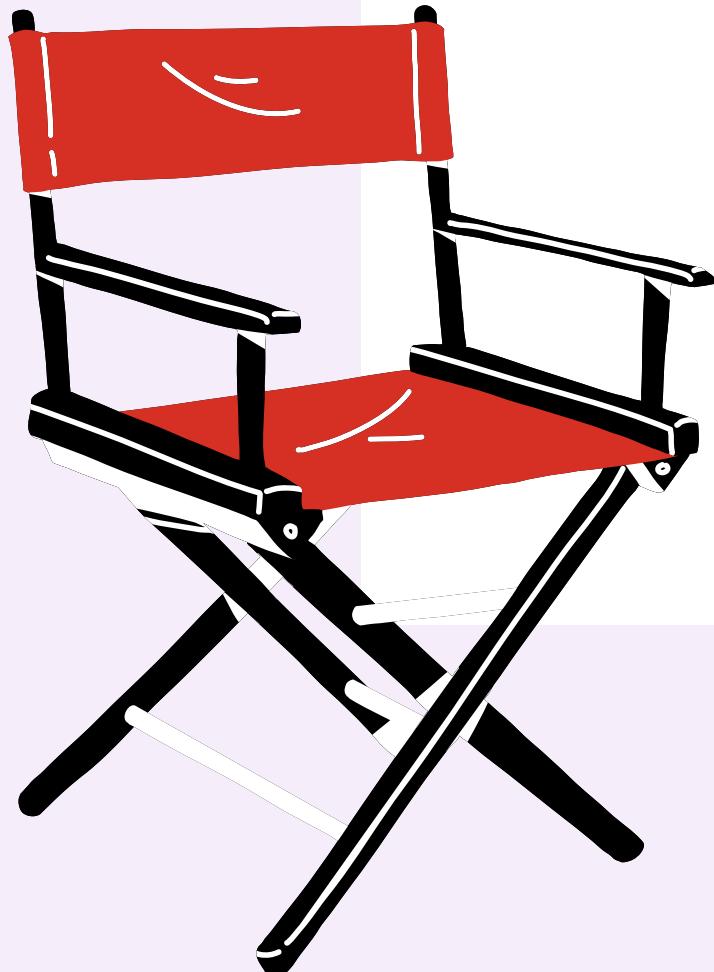
The final report has been saved as a PDF file and uploaded to Google Drive.

You can access the report through the following link:

[Click to open Excel Sheet](#)

Key Approach

to solve the problem



Steps to Solve the problem

1

Know the Problem statement i.e.
"What factors influence the success
of a movie on IMDB?"

2

Pre-process the dataset to extract
useful insights. It includes handling
missing values, removing duplicates,
converting data types

3

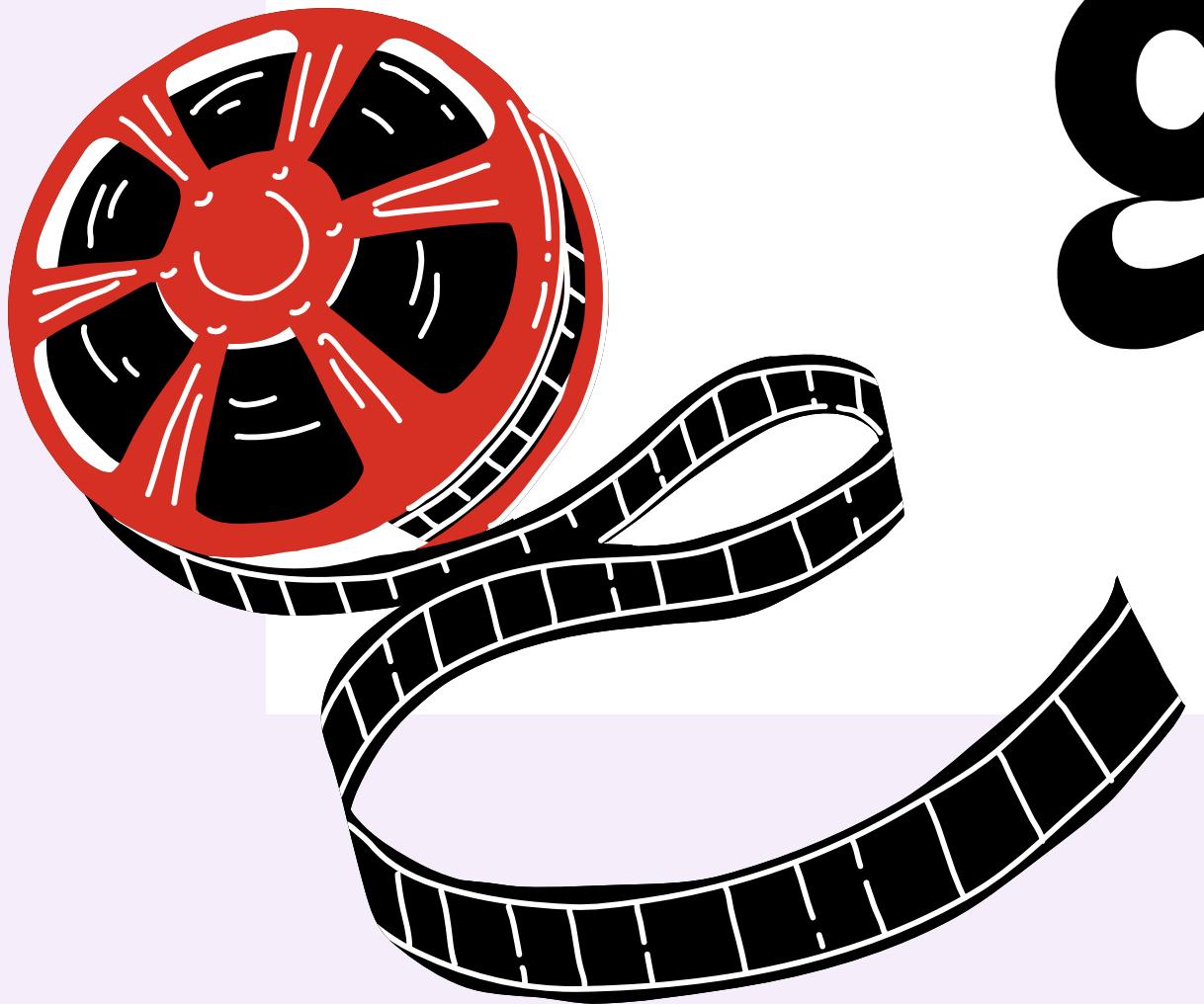
Delete the columns which are not
required to do analysis such as
facebook likes, color etc.

4

Apply data analysis knowledge to
understand the relationships between
different variables.



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.
calculate descriptive statistics of the IMDB scores

The most common movie genres are :- Comedy, Action, Drama, Adventure, Crime
This can be concluded by using the formula of =COUNTIF(\$B:\$H,J2)

Also, we manipulated the 'genres' column to separate multiple genres for a single movie by using Delimiter function and removing the duplicates from each column to extract the unique list of genres.

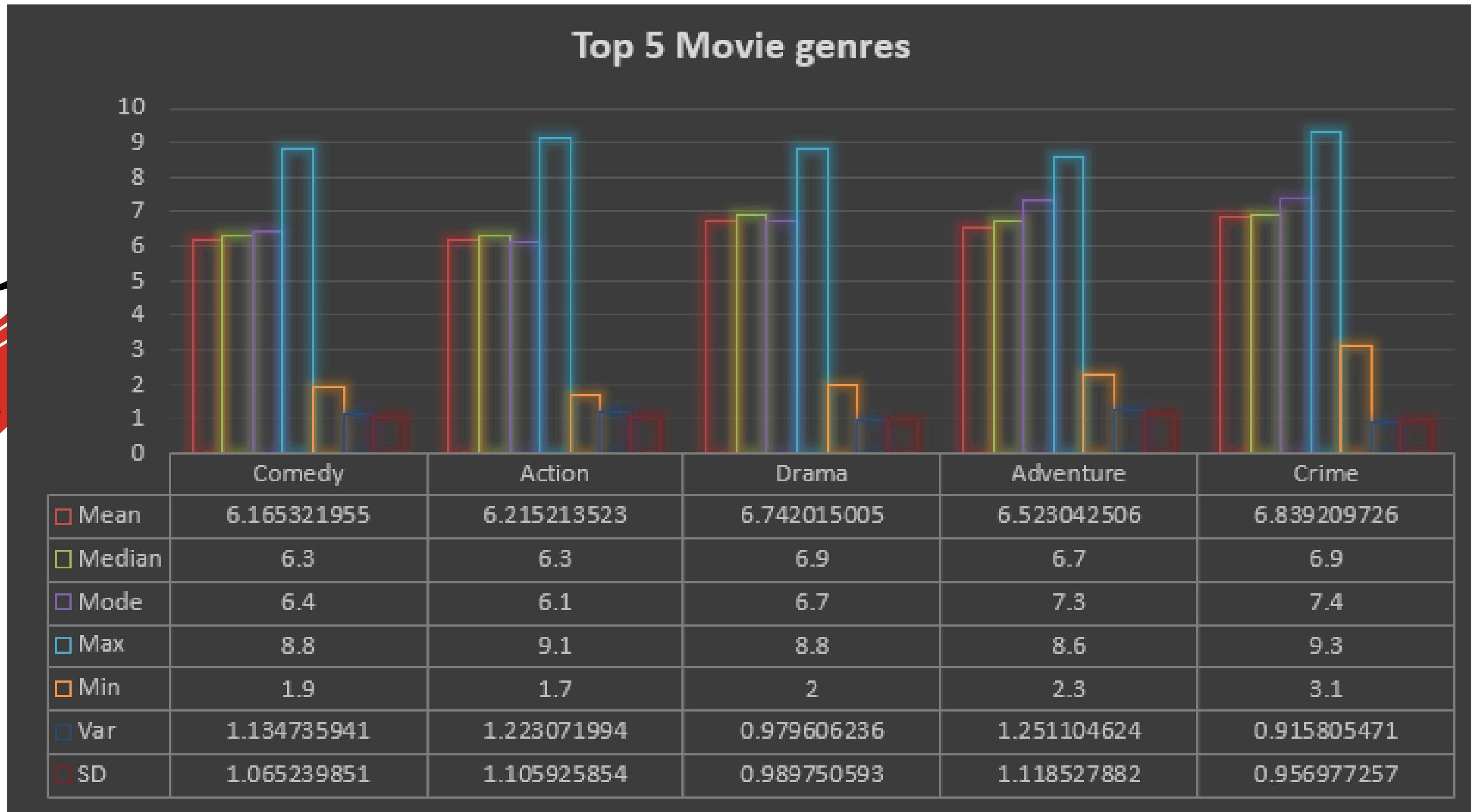
Formula used to calculate Descriptive analysis :-
=AVERAGEIF(\$B\$2:\$H\$4892,J2,\$I\$2:\$I\$4892)
=MEDIAN(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))
=MODE.SNGL(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))
=MAX(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))
=MIN(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))
=VAR.S(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))
=STDEV.S(IF(\$B\$2:\$H\$4892=J2,\$I\$2:\$I\$4892))





D E S C R I P T I V E A N A L Y S I S

Graphical Representation of Top 5 Genres



Key Analysis

1

Documentary is the most consistently high-rated genre (mean 7.18, median 7.5, mode 7.5).

2

Biography is both popular and stable with a strong mean (7.16) and lowest variance (0.48).

3

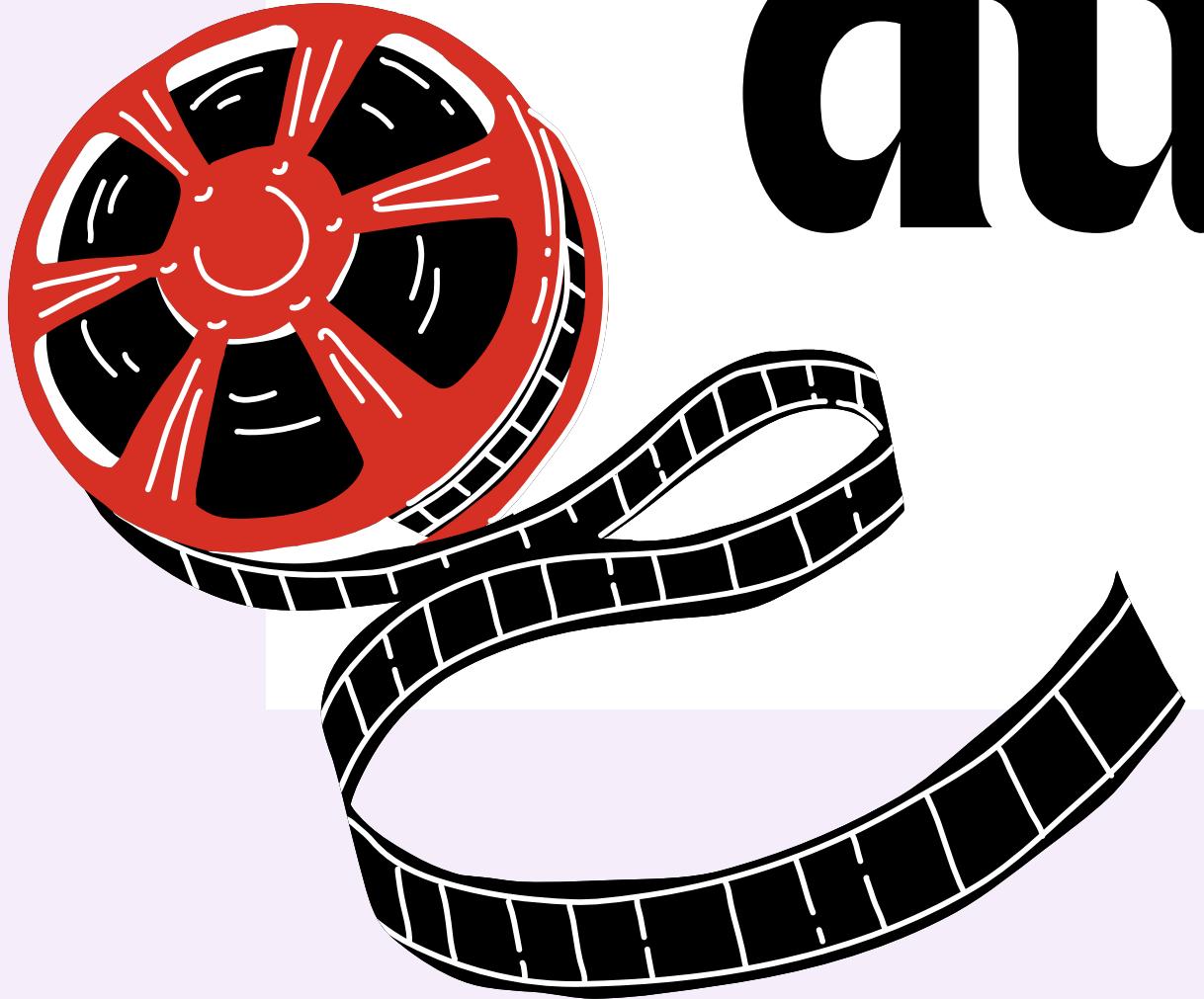
Family has the most unpredictable ratings – lowest mean among top genres (5.71) and highest variance (3.87).

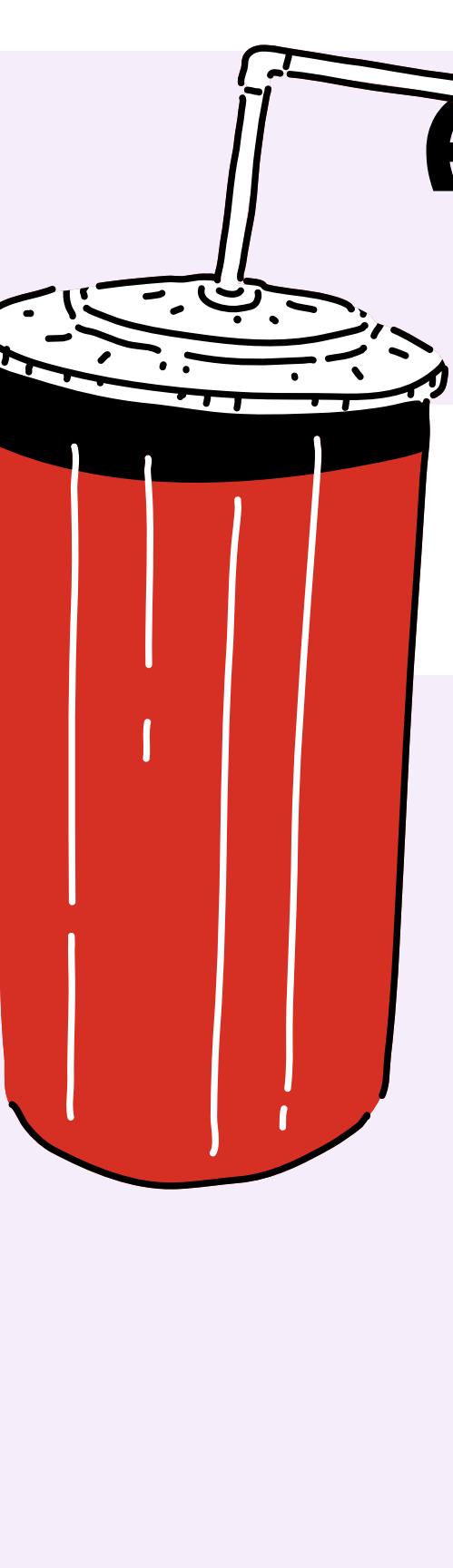
4

Thriller shows widely mixed reception, with the lowest mean (5.58) and high rating spread (variance 1.77).



Movie duration Analysis





Analyze the distribution of movie durations and its impact on the IMDB score.

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

The average duration of movies are 108 minutes

A standard deviation of 22.5 minutes indicates a wide range of movie durations, from short films to long ones.

Formula used to calculate Descriptive analysis :-

=AVERAGE(A:A)

=MEDIAN(A:A)

=MODE.SNGL(A:A)

=STDEV.S(A:A)



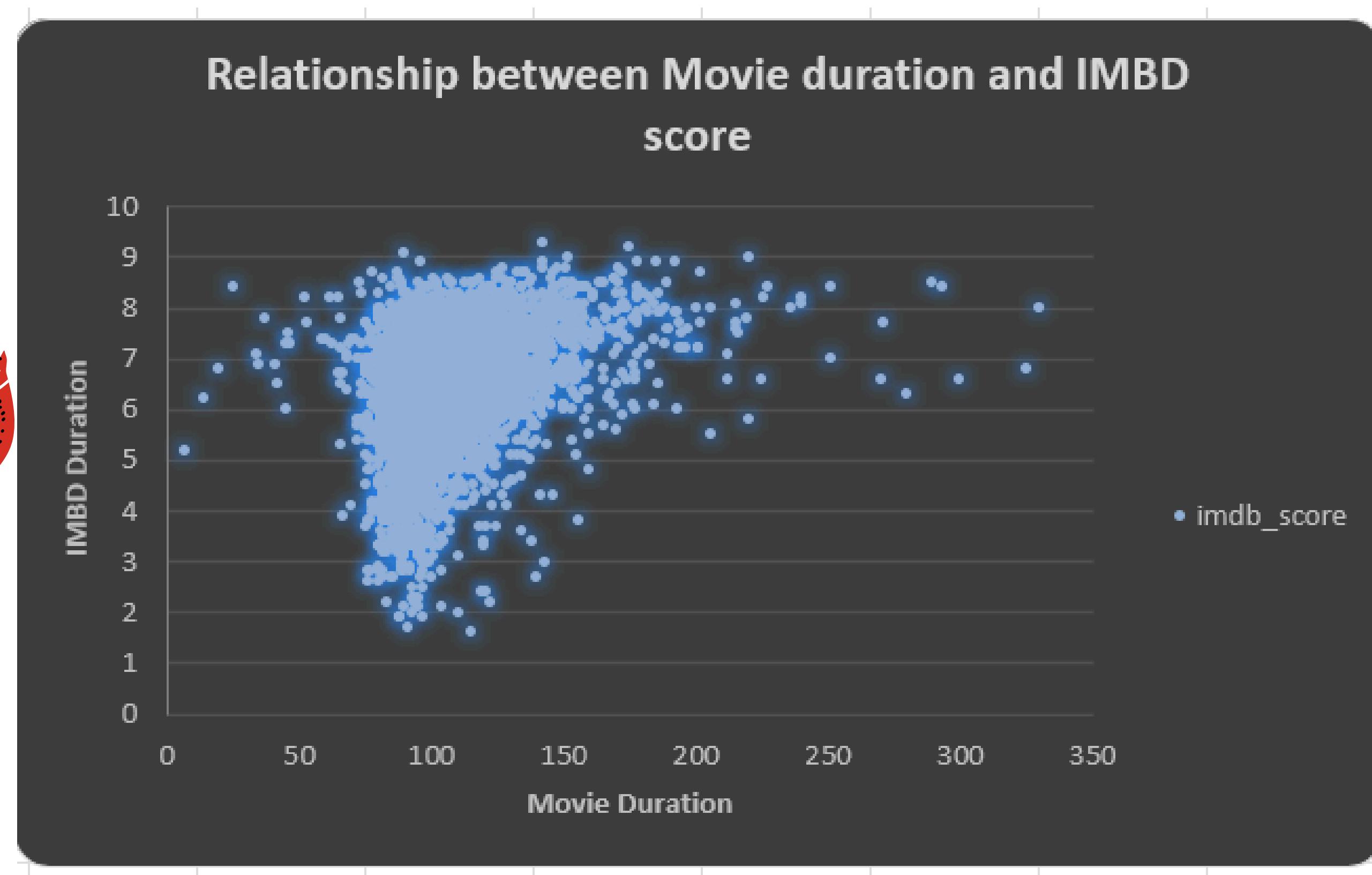


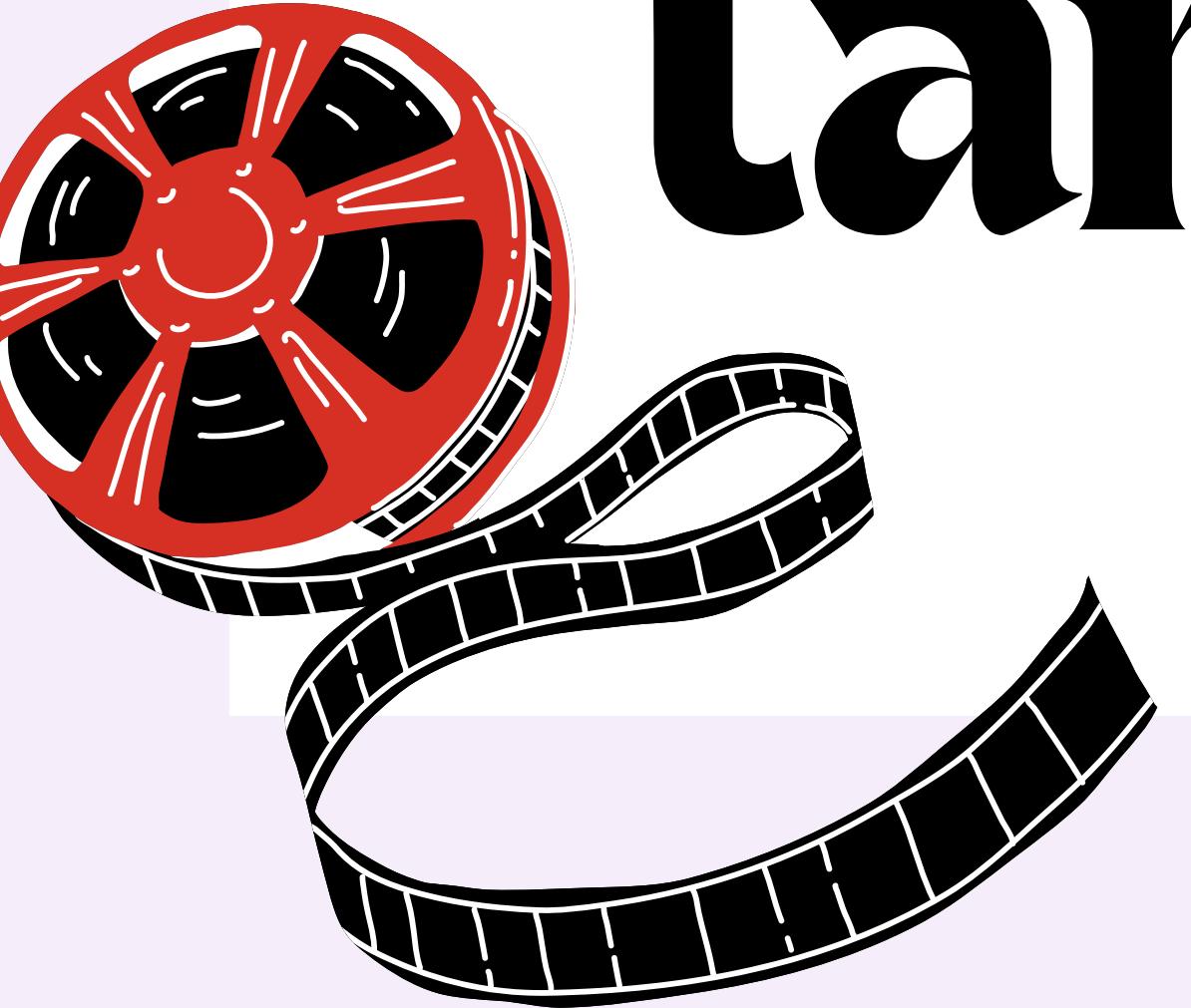
| duration | imdb_score |
|----------|------------|
| 178 | 7.9 |
| 169 | 7.1 |
| 148 | 6.8 |
| 164 | 8.5 |
| 132 | 6.6 |
| 156 | 6.2 |
| 100 | 7.8 |
| 141 | 7.5 |
| 153 | 7.5 |
| 183 | 6.9 |
| 169 | 6.1 |
| 106 | 6.7 |
| 151 | 7.3 |
| 150 | 6.5 |
| 143 | 7.2 |
| 150 | 6.6 |
| 173 | 8.1 |
| 136 | 6.7 |
| 106 | 6.8 |
| 164 | 7.5 |
| 153 | 7 |
| 156 | 6.7 |
| 186 | 7.9 |
| 113 | 6.1 |
| 201 | 7.2 |

| Mean | Median | Mode | STDEV |
|----------|--------|------|----------|
| 108.1687 | 104 | 90 | 22.55464 |



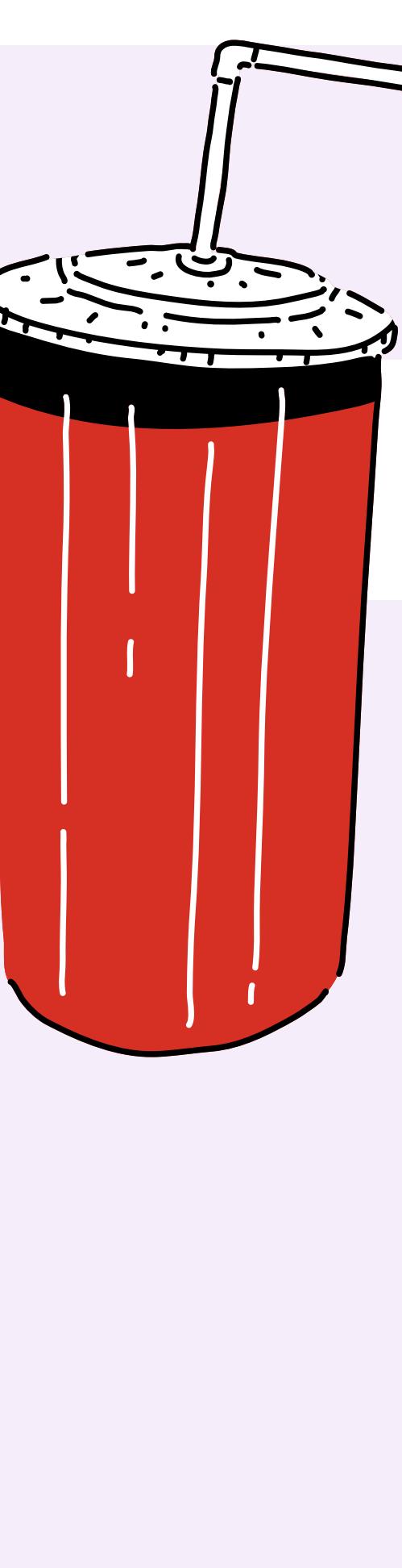
Graphical Representation with scatter graph





Movie language Analysis





Examine the distribution of movies based on their language.

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

The most common languages of movies are English, French and Spanish.

Formula used =COUNTIF(\$A:\$A,C2)

Hebrew content has the highest mean rating (7.58) and lowest standard deviation (0.35), making it the most consistently well-rated.

Chinese is lowest rated (3.67) with little variation.

English has the highest count 4,566 but mixed ratings (mean 6.37, high SD).

Formula used to calculate Descriptive analysis :-

=AVERAGEIF(\$A\$2:\$A\$4892,[@language],\$B\$2:\$B\$4892)

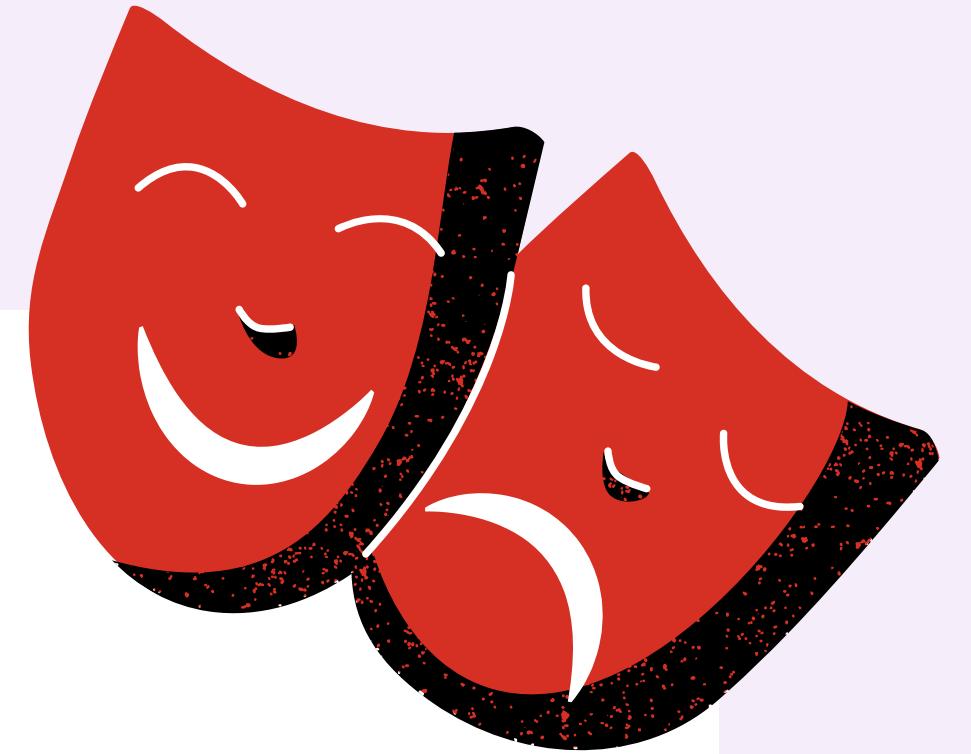
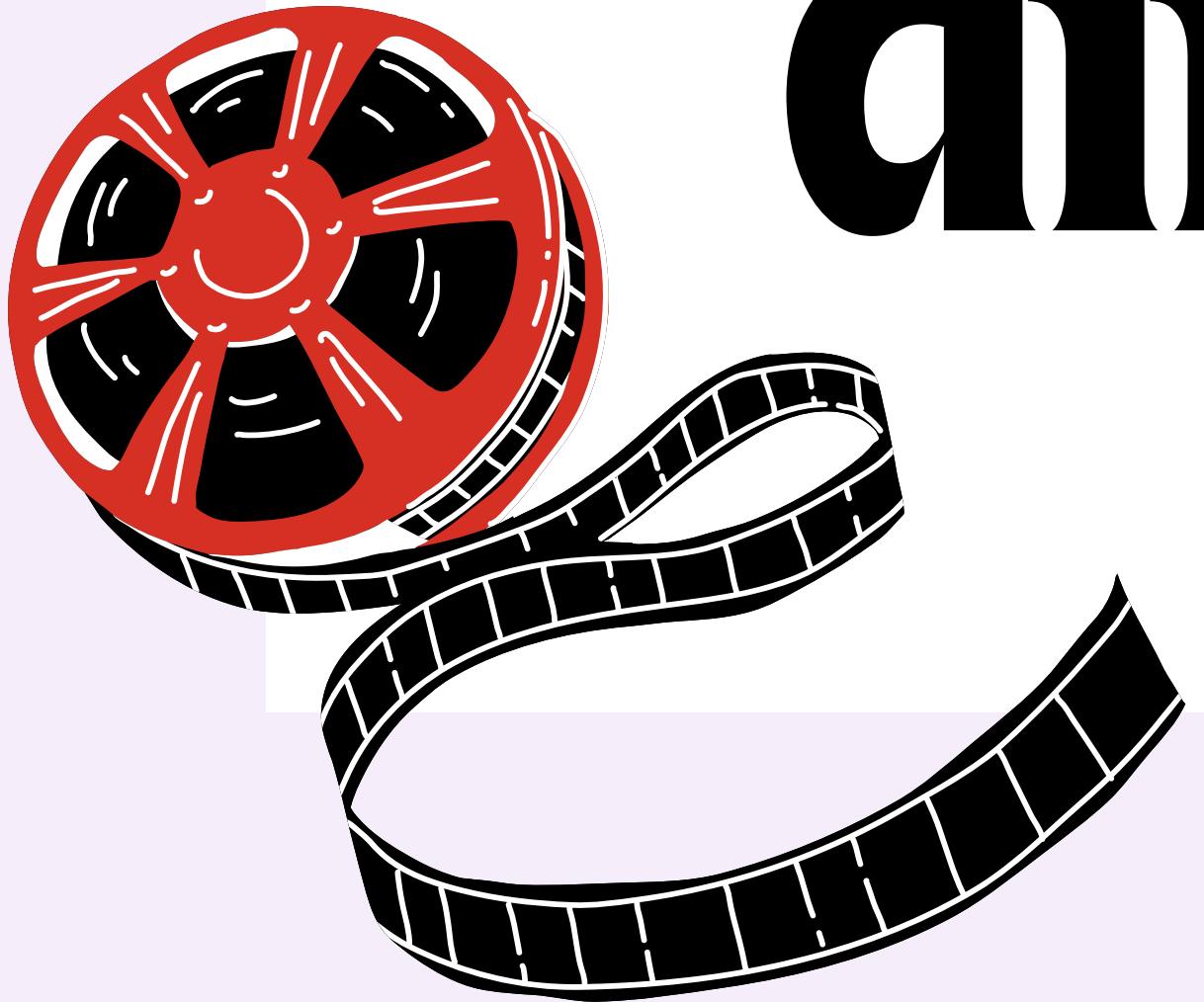
=MEDIAN(IF(\$A\$2:\$A\$4892=[@language],\$B\$2:\$B\$4892))

=STDEV.S(IF(\$A\$2:\$A\$4892=[@language],\$B\$2:\$B\$4892))

| language | Count | Mean | Median | STDEV | |
|------------|-------|----------|--------|----------|--|
| English | 4566 | 6.373675 | 6.5 | 1.110371 | |
| French | 72 | 7.020833 | 7.2 | 0.716395 | |
| Spanish | 40 | 6.9375 | 7.15 | 0.855057 | |
| Hindi | 28 | 6.632143 | 6.95 | 1.398956 | |
| Mandarin | 24 | 6.7875 | 7.05 | 1.036848 | |
| German | 19 | 7.342105 | 7.6 | 0.954123 | |
| Japanese | 16 | 7.36875 | 7.6 | 1.028733 | |
| Russian | 11 | 6.363636 | 6.5 | 1.383671 | |
| Cantonese | 11 | 6.954545 | 7.2 | 0.704789 | |
| Italian | 10 | 7.08 | 7.15 | 1.20628 | |
| Unknown | 9 | 6.8 | 6.7 | 1.459452 | |
| Korean | 8 | 7.3875 | 7.5 | 0.825379 | |
| Portuguese | 8 | 7.4875 | 7.7 | 0.883883 | |
| Hebrew | 5 | 7.58 | 7.6 | 0.334664 | |
| Danish | 5 | 7.5 | 8.1 | 1.077033 | |
| Arabic | 5 | 7.38 | 7.4 | 0.884308 | |
| Dutch | 4 | 7.425 | 7.45 | 0.434933 | |
| Swedish | 4 | 7.275 | 7.15 | 0.763217 | |
| Norwegian | 4 | 7.15 | 7.3 | 0.574456 | |
| Persian | 4 | 7.575 | 7.95 | 1.203813 | |
| Chinese | 3 | 5.666667 | 5.7 | 0.550757 | |
| Thai | 3 | 6.633333 | 6.6 | 0.450925 | |
| Aboriginal | 2 | 6.95 | 6.95 | 0.777817 | |
| Dari | 2 | 7.5 | 7.5 | 0.141421 | |
| None | 2 | 7.95 | 7.95 | 0.777817 | |
| Zulu | 2 | 7.1 | 7.1 | 0.282843 | |
| Indonesian | 2 | 7.9 | 7.9 | 0.424264 | |
| Romanian | 2 | 7.2 | 7.2 | 0.989949 | |
| 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 | |
| Icelandic | 1 | 6.9 | 6.9 | 0 | |
| Czech | 1 | 7.4 | 7.4 | 0 | |
| Kannada | 1 | 7.1 | 7.1 | 0 | |
| Punjabi | 1 | 6.6 | 6.6 | 0 | |
| Polish | 1 | 7.4 | 7.4 | 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 | |



Movie director Analysis





Influence of directors on movie ratings.

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

The average IMBD score of each director can be calculated with
Formula =AVERAGEIF(\$C\$2:\$C\$4892,C2,\$B\$2:B\$4892)

90th Percentile can be calculated by =PERCENTILE.INC(D2:D2396, 0.9)

Director with average IMBD greater than the 90th percentile can be
extracted with sort function.

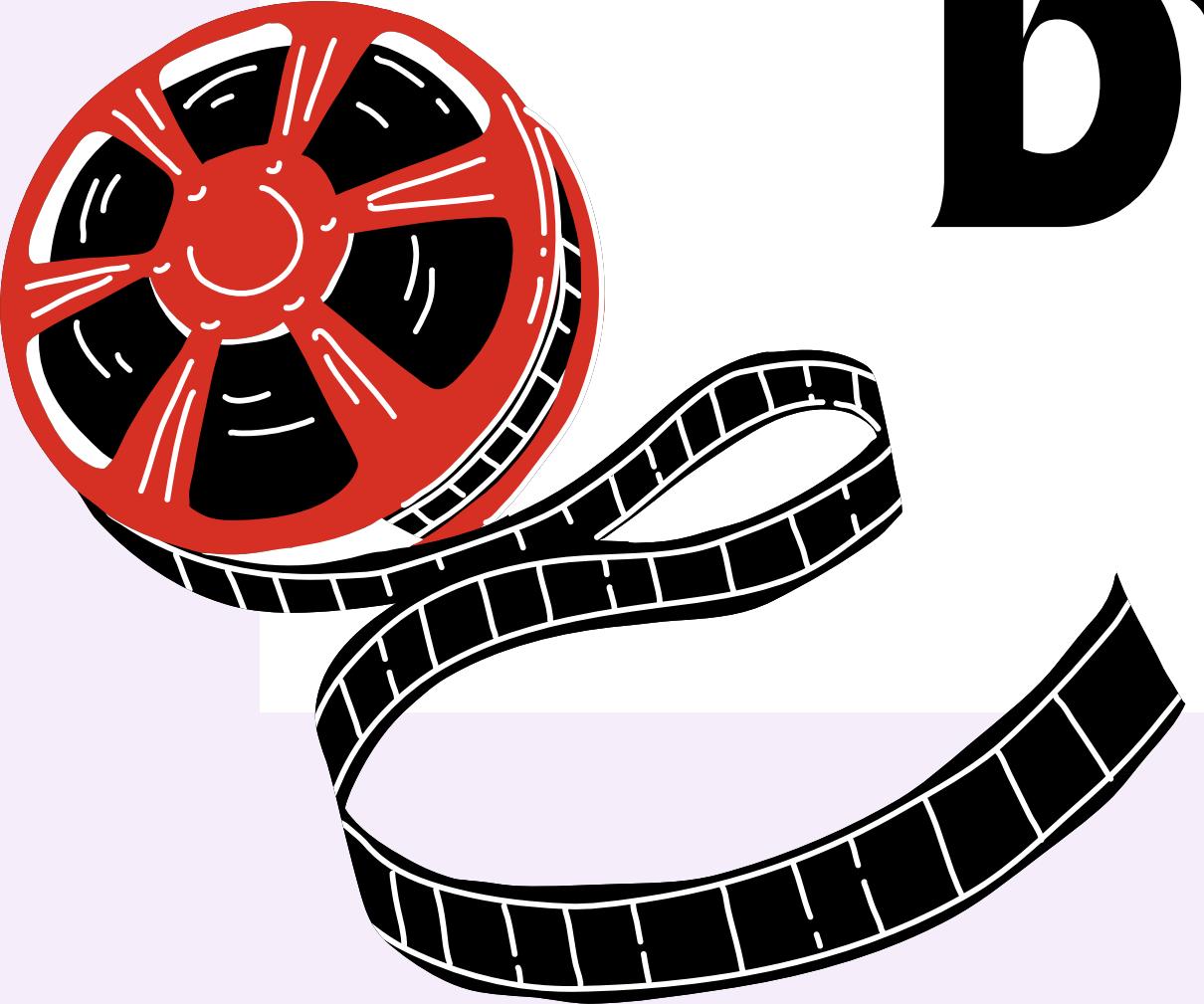


| A | director_name | average IMBD | Lee Tamahori | 8.2 |
|---|-----------------------|--------------|---------------------|-----|
| V | James Cameron | 7.9 | Paul Feig | 7.9 |
| E | Christopher Nolan | 8.5 | Peter Ramsey | 7.8 |
| R | Nathan Greno | 7.8 | Dean Parisot | 7.8 |
| A | Marc Webb | 8.1 | Phillip Noyce | 8 |
| G | Shane Black | 7.9 | Darren Aronofsky | 8.3 |
| E | Dan Scanlon | 7.7 | Guy Ritchie | 8.1 |
| I | Michael Bay | 8.2 | Frank Coraci | 7.7 |
| M | Martin Campbell | 7.8 | Bo Welch | 8 |
| B | Robert Zemeckis | 8.3 | Peter Chelsom | 7.9 |
| D | Jon Favreau | 8 | Paul Weitz | 7.7 |
| | Martin Scorsese | 7.8 | Pitof | 8.1 |
| | Roland Joffé | 8.4 | Lawrence Guterman | 8.1 |
| | James Gunn | 9 | Sergey Bodrov | 7.8 |
| | Hideaki Anno | 8.3 | Tony Bancroft | 8.1 |
| | Matthew Vaughn | 7.9 | Roger Allers | 8 |
| | Duncan Jones | 7.9 | Gary Winick | 8.1 |
| | Oliver Stone | 8.3 | Joe Pytka | 7.8 |
| | Eric Darnell | 7.8 | Des Mcanuff | 8.8 |
| | Byron Howard | 7.8 | Robert Luketic | 8.1 |
| | Christopher Mcquarrie | 7.9 | Andy Tennant | 8.5 |
| | Joe Johnston | 7.8 | Florent-Emilio Siri | 7.7 |
| | M. Night Shyamalan | 8.2 | Ron Shelton | 8 |
| | David Bowers | 8.1 | Henry Selick | 8.5 |
| | Joe Wright | 8.6 | Clay Kaytis | 7.8 |
| | Rob Minkoff | 8.8 | Betty Thomas | 8.5 |
| | | | POP CORN | |



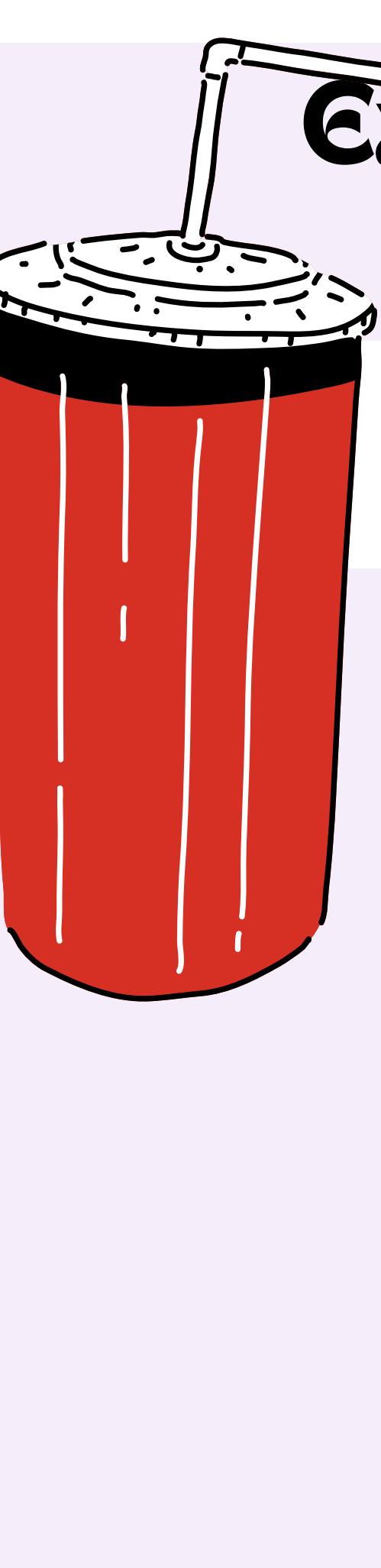
Directors with highest score :-

| director_name | average IMBD |
|-----------------------|--------------|
| James Cameron | 7.9 |
| Christopher Nolan | 8.5 |
| Nathan Greno | 7.8 |
| Marc Webb | 8.1 |
| Shane Black | 7.9 |
| Dan Scanlon | 7.7 |
| Michael Bay | 8.2 |
| Martin Campbell | 7.8 |
| Robert Zemeckis | 8.3 |
| Jon Favreau | 8 |
| Martin Scorsese | 7.8 |
| Roland Joffé | 8.4 |
| James Gunn | 9 |
| Hideaki Anno | 8.3 |
| Matthew Vaughn | 7.9 |
| Duncan Jones | 7.9 |
| Oliver Stone | 8.3 |
| Eric Darnell | 7.8 |
| Byron Howard | 7.8 |
| Christopher McQuarrie | 7.9 |
| Joe Johnston | 7.8 |
| M. Night Shyamalan | 8.2 |
| David Bowers | 8.1 |
| Joe Wright | 8.6 |
| Rob Minkoff | 8.8 |



Movie budget Analysis





Explore the relationship between movie budgets and their financial success.

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

The profit of each movie can be calculated with
Formula Profit margin = (gross earnings - budget)

Correlation between movie budgets and gross earning can be
calculated by =CORREL(A:A,B:B)

Correlation is 0.1107967 means there is a very weak positive correlation, suggesting that higher budgets are slightly associated with higher earnings.

Movie with highest profit is Avatar with net profit of 523505847
calculated by =MAX(C:C)





| budget | gross | Profit |
|---------------|--------------|---------------|
| 237000000 | 760505847 | 523505847 |
| 300000000 | 309404152 | 9404152 |
| 245000000 | 200074175 | -44925825 |
| 250000000 | 448130642 | 198130642 |
| 263700000 | 73058679 | -190641321 |
| 258000000 | 336530303 | 78530303 |
| 260000000 | 200807262 | -59192738 |
| 250000000 | 458991599 | 208991599 |
| 250000000 | 301956980 | 51956980 |
| 250000000 | 330249062 | 80249062 |
| 209000000 | 200069408 | -8930592 |
| 200000000 | 168368427 | -31631573 |
| 225000000 | 423032628 | 198032628 |
| 215000000 | 89289910 | -125710090 |
| 225000000 | 291021565 | 66021565 |
| 225000000 | 141614023 | -83385977 |
| 220000000 | 623279547 | 403279547 |
| 250000000 | 241063875 | -8936125 |
| 225000000 | 179020854 | -45979146 |
| 250000000 | 255108370 | 5108370 |
| 230000000 | 262030663 | 32030663 |
| 200000000 | 105219735 | -94780265 |
| 225000000 | 258355354 | 33355354 |
| 180000000 | 70083519 | -109916481 |
| 207000000 | 218051260 | 11051260 |

Correlation

0.1107967



Movie with Highest Profit

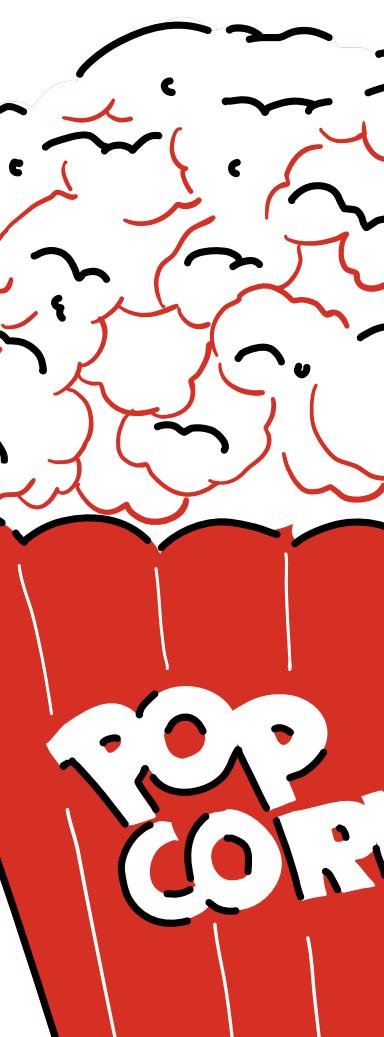


| movie_title | Movie with highest profit |
|-------------|---------------------------|
| Avatar | 523505847 |

Top 10 Movies with Highest Profit



| movie_title | Profit |
|---|-----------|
| Avatar | 523505847 |
| The Avengers | 403279547 |
| Titanic | 458672302 |
| Jurassic World | 502177271 |
| The Dark Knight | 348316061 |
| Star Wars: Episode I - The Phantom Menace | 359544677 |
| The Hunger Games | 329999255 |
| The Lion King | 377783777 |
| Star Wars: Episode IV - A New Hope | 449935665 |
| E.T. the Extra-Terrestrial | 424449459 |



Thankyou

