

# IMDB MOVIE ANALYSIS

- Iswariya S

The IMDb logo is displayed in a bold, black, sans-serif font. The letters are thick and closely spaced. The 'i' in 'IMDb' is lowercase and features a distinctive dot. The logo is centered within a bright yellow rectangular background.

[illegible]

- 
- The image is a dark-themed collage representing IMDb Live events. In the center is a large yellow rounded square containing the "IMDb" logo in bold black letters. Surrounding this central element are several smaller rectangular panels showing different scenes: 
- Top left: A woman in a pink dress standing outdoors near water.
- Top middle-left: A man wearing sunglasses and a white shirt speaking into a microphone with the IMDb logo.
- Top middle-right: A close-up of actor Brad Pitt smiling.
- Top right: A person holding a bottle at an event with a backdrop.
- Middle left: A woman in a patterned dress smiling next to a snowman prop.
- Middle top-center: A panel titled "LIVE VIEWING PARTY" with a grid of small video thumbnails.
- Middle top-right: A man in a suit looking at a wall covered with framed photos or awards.
- Middle right: A person holding a bottle against a checkered background.
- Bottom left: Two men wearing festive hats, one with "IMDb" written on it.
- Bottom center: A view of a stage or event space with equipment.
- Bottom right: An outdoor night scene of a building decorated for Christmas with lights and snow.
- Bottom far left: Three people sitting together, two holding white pillows with the IMDb logo.
- Bottom center-bottom: A woman in a red dress walking on a stage.
- Bottom far right: Two men standing side-by-side, smiling for a photo.

# INTRODUCTION

The dataset pertains to IMDB Movies, with a focus on understanding the factors influencing a movie's success, defined by high IMDB ratings. This research is valuable for movie industry stakeholders to make informed decisions. The project involves analyzing a dataset containing 28 columns and 5044 rows of IMDB Movies, including movie names, release year, actors, budget, gross, and genre. The process includes defining the problem, cleaning the data, exploring insights, and using root cause analysis. Visual representations like charts and graphs are employed to answer specific questions and create reports for data-driven decision-making.



## APPROACH

The "IMDB Movie Analysis" project involves a comprehensive dataset examination to extract insights into factors influencing movie success on IMDB.

---

Our analysis was conducted using Microsoft Excel, utilizing its built-in formulas to systematically address the questions posed regarding the dataset. Following this, we will identify and eliminate any duplicates in the dataset using Excel's built-in "remove duplicate" function

Additionally, we will examine the dataset for missing values. To identify outliers, we will employ scatter plots, and if any are found, we will eliminate them and subsequently provide a summary and draw conclusions based on the analysis.



PowerPoint



Microsoft  
Excel 2016

## TECH-STACK USED

---

- **Microsoft PowerPoint**

For presentation of the project

- **Microsoft Excel**

For graphical representation of the results and to understand the results set better

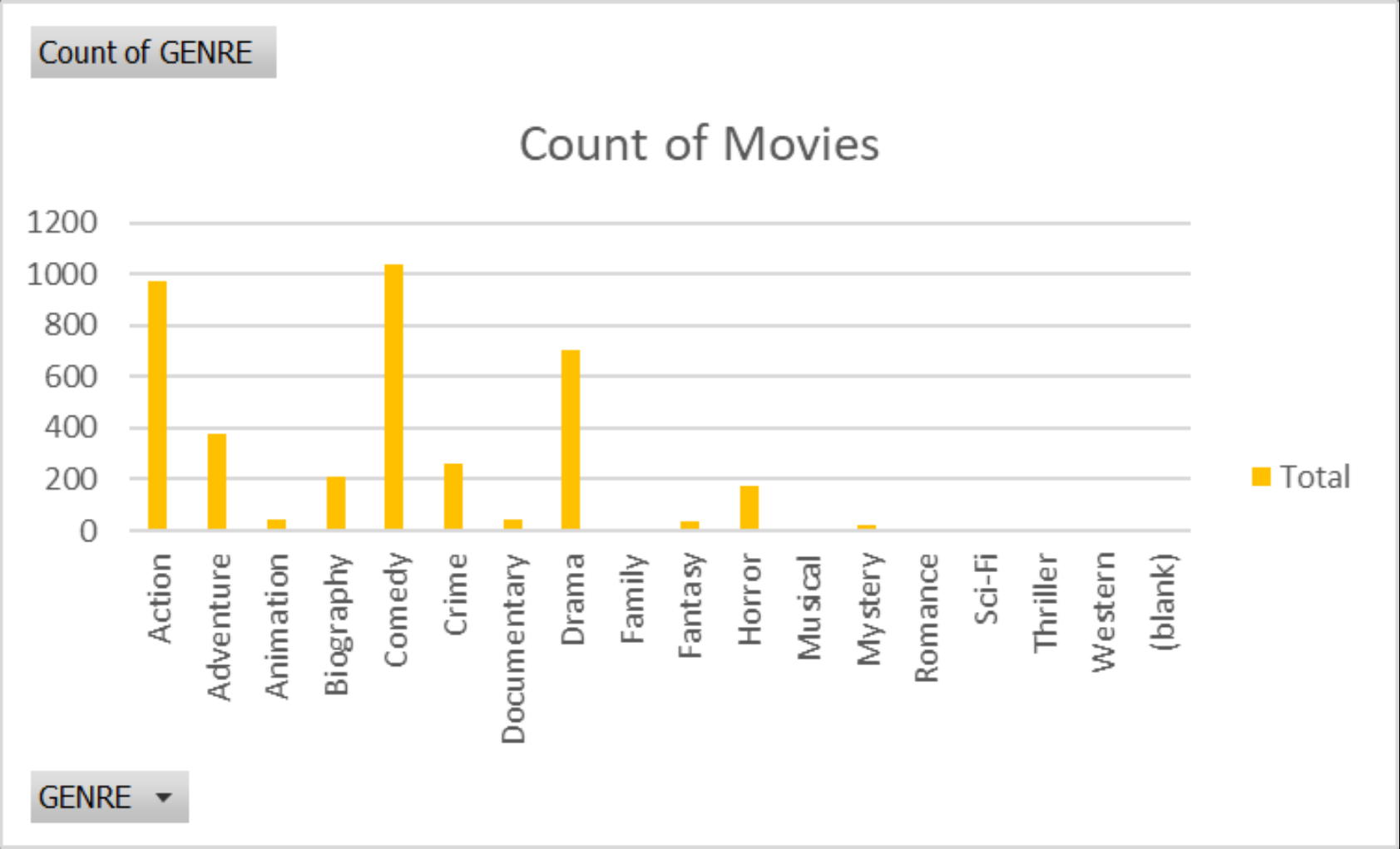


# INSIGHTS

## A. Movie Genre Analysis:

**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.

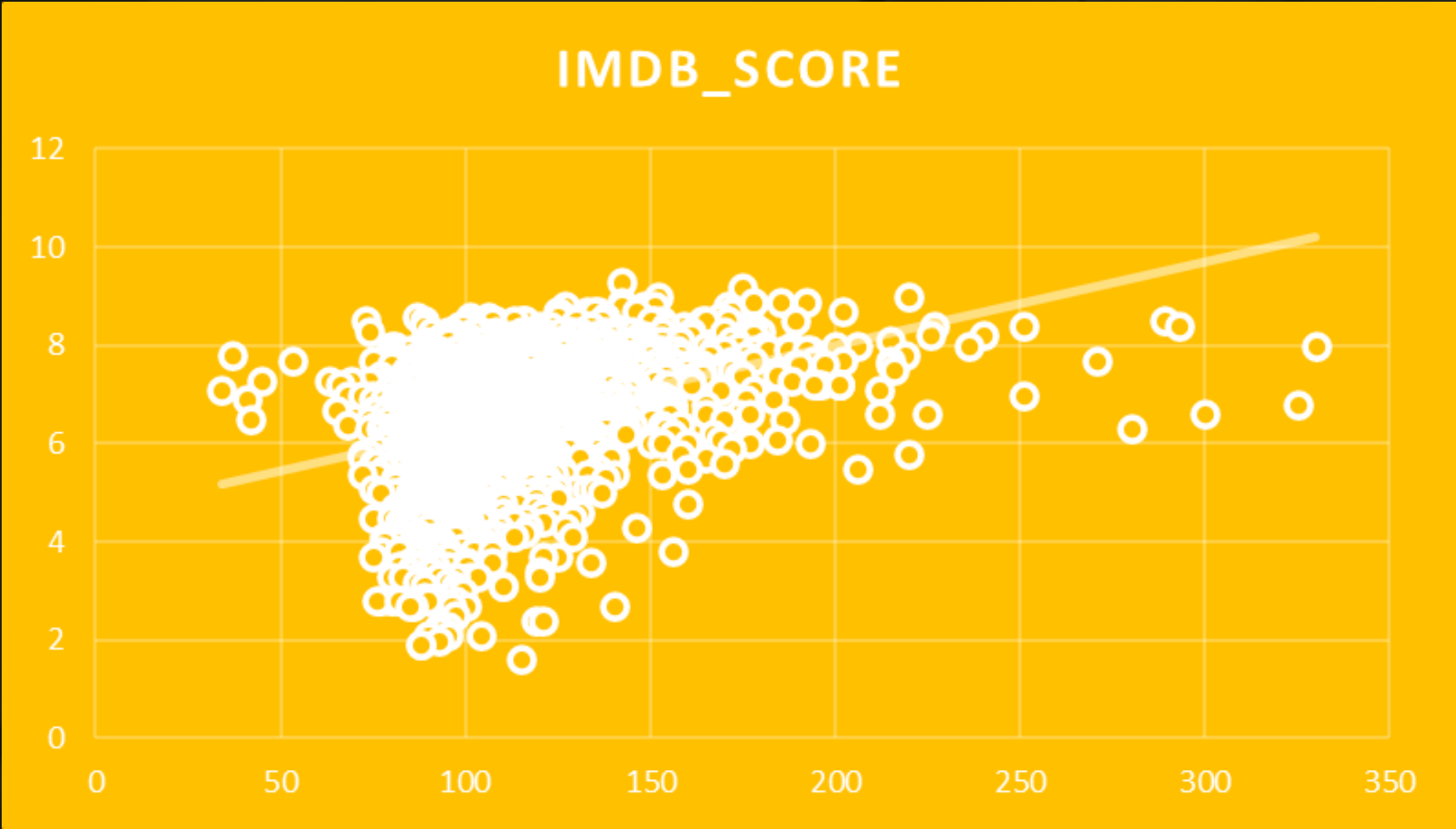
GENRE	Count of GENRE	Average	Median	Mode	Variance	SD	Max	Min
Action	976	6.3	6.3	6.6	1.1	1.04	9	2.1
Adventure	380	6.5	6.7	7.3	1.3	1.1	8.6	2.3
Animation	47	6.7	7	7.1	1	1	8	4.5
Biography	209	7.1	7.2	7	0.5	0.7	8.9	4.5
Comedy	1041	6.2	6.3	6.4	1.1	1	8.8	1.9
Crime	262	6.9	7	7.4	0.8	0.9	9.3	3.1
Documentary	43	7.0	7.3	7.3	4	2	8.5	1.6
Drama	704	6.8	6.9	7.3	0.8	0.9	8.8	2.1
Family	3	6.5	5.9	N/A	1.5	1.2	7.9	5.7
Fantasy	37	6.3	6.5	6.8	0.8	0.9	7.9	4.3
Horror	173	5.8	5.8	5.9	1.1	1.1	8.5	2.2
Musical	2	6.8	6.75	N/A	N/A	N/A	7.2	7.2
Mystery	23	6.7	6.7	7.1	1.2	1.1	8.5	3.3
Romance	2	6.7	6.65	N/A	N/A	N/A	7.1	7.1
Sci-Fi	10	6.0	6.2	N/A	1.4	1.2	8.2	5.0
Thriller	3	5.3	4.8	N/A	0.6	0.8	4.5	3.4
Western	4	6.5	6.4	N/A	1.4	2.5	7.3	3.8
(blank)								
Grand Total	3919							



The most common movie genre is Comedy - 1041

## B. Movie Duration Analysis:

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

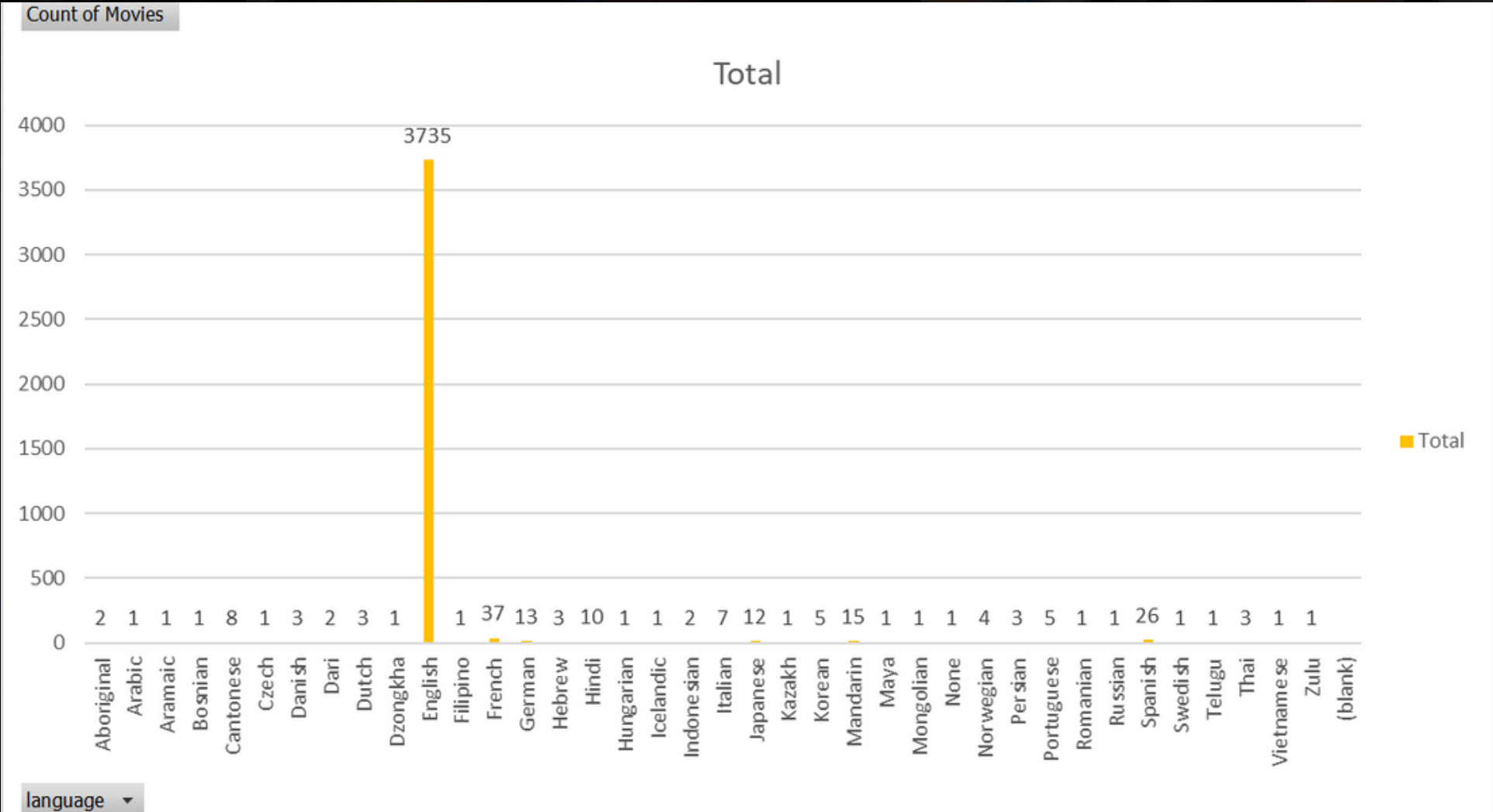


## Descriptive Analysis

mean	6.449859658
Standard Error	0.017050391
Median	6.6
Mode	6.7
Standard Deviation	1.067387146
Sample Variance	1.13931532
Range	7.7
Minimum	1.6
Maximum	9.3
Sum	25277
Count	3919

# C. Language Analysis:

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



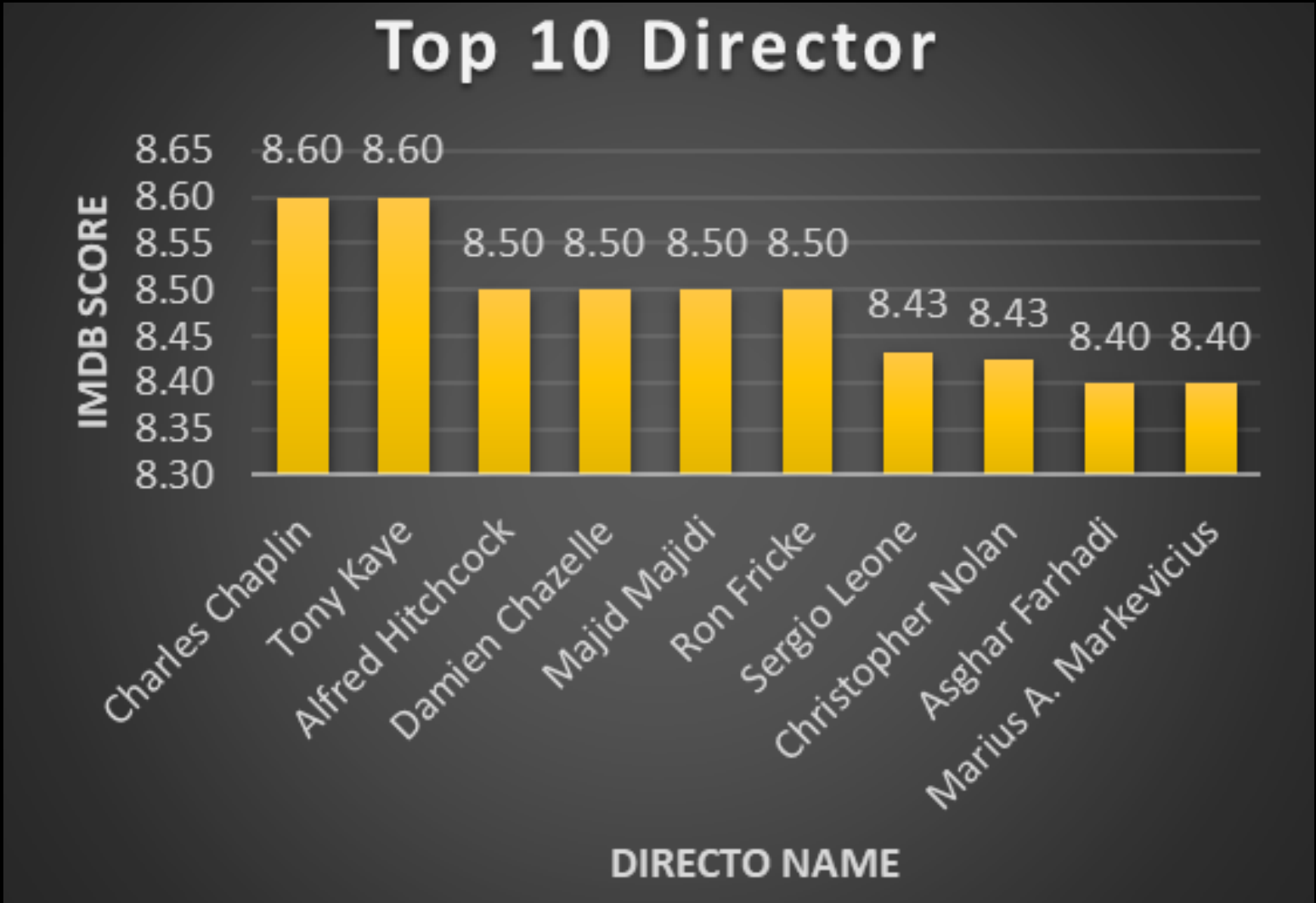
The most common language is - English with 3735 number of movies )

language	Count of Movies	Average of imdb	Median	S.d
Aboriginal	2	7.2	7.0	0.8
Arabic	1	7.1	7.2	0.0000
Aramaic	1	4.3	7.1	0.0000
Bosnian	1	7.2	4.3	0.0000
Cantonese	8	7.4	7.3	0.44
Czech	1	7.9	7.4	0.0000
Danish	3	7.5	8.1	0.53
Dari	2	7.6	7.5	0.1414
Dutch	3	7.5	7.8	0.4041
Dzongkha	1	6.4	7.5	0.0000
English	3735	6.7	6.5	1.0624
Filipino	1	7.3	6.7	0.0000
French	37	7.7	7.2	0.5613
German	13	7.5	7.7	0.6409
Hebrew	3	6.8	7.3	0.4359
Hindi	10	7.1	7.1	1.1118
Hungarian	1	6.9	7.1	0.0000
Icelandic	1	7.9	6.9	0.0000
Indonesian	2	7.2	7.9	0.4243
Italian	7	7.6	7.0	1.1553
Japanese	12	6.0	7.8	0.8996
Kazakh	1	7.7	6.0	0.0000
Korean	5	7.1	7.7	0.5701
Mandarin	15	7.8	7.4	0.7720
Maya	1	7.3	7.8	0.0000
Mongolian	1	8.5	7.3	0.0000
None	1	7.2	8.5	0.0000
Norwegian	4	8.1	7.3	0.5745
Persian	3	7.8	8.4	0.5508
Portuguese	5	7.9	8.0	0.9788
Romanian	1	6.5	7.9	0.0000
Russian	1	7.1	6.5	0.0000
Spanish	26	7.6	7.2	0.8262
Swedish	1	8.4	7.6	0.0000
Telugu	1	6.6	8.4	0.0000
Thai	3	7.4	6.6	0.4509
Vietnamese	1	7.3	7.4	0.0000
Zulu	1	7.1	7.3	1.0000
(blank)				
Grand Total	3916			



D. Director Analysis:

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



LARGE - 8.60  
PERCENTILE RANK - 0.89  
PERCENTILE - 8.60

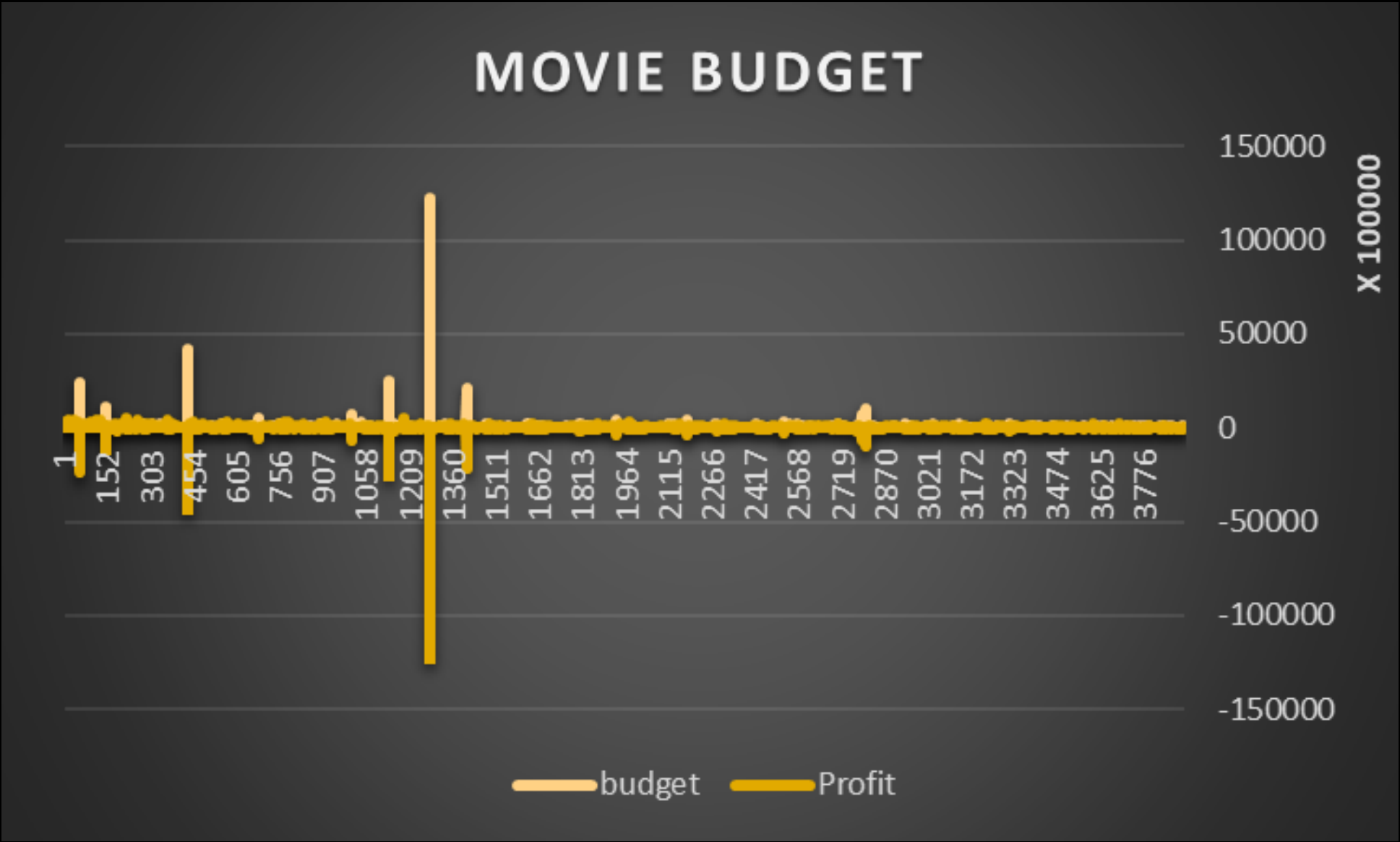
Top 10 Directors	Average of imdb_score
Charles Chaplin	8.60
Tony Kaye	8.60
Alfred Hitchcock	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
Asghar Farhadi	8.40
Marius A. Markevicius	8.40

E. Budget Analysis:

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

HIGHT PROFIT MARGIN MOVIE	HIGHT PROFIT MARGIN
The Hobbit: The Desolation of Smaug	523505847

CORRELATION - 0.102177361



# RESULT

Data cleaning ensured the dataset's quality by addressing missing values, duplicates, and data type issues, establishing a strong foundation for subsequent unbiased analysis.

- The dataset highlighted comedy as the most common genre (1041 instances) and English as the predominant language (3735 instances).
- Descriptive statistics revealed variations in IMDB scores across genres.
- Analysis showed a mild positive correlation between movie durations and IMDB scores, indicating slightly higher ratings for longer movies.

- Top directors with consistently high IMDB scores significantly influenced movie success compared to the overall distribution.
- Correlation analysis indicated a moderate positive relationship (0.102177361) between movie budgets and gross earnings.
- Movies with the highest profit margins totaled 523,505,847.





# CONCLUSION

In conclusion, this project has offered a profound insight into the factors impacting the success of movies on IMDB. From the initial data cleaning to the 'Five Whys' analysis and individual data analytics tasks, the acquired insights can empower stakeholders in the movie industry to make informed, data-driven decisions, elevate the quality of movies, and increase audience satisfaction. The comprehensive report generated by the project encapsulates these valuable insights, offering a convenient reference for decision-making purposes.





**THANK YOU!**