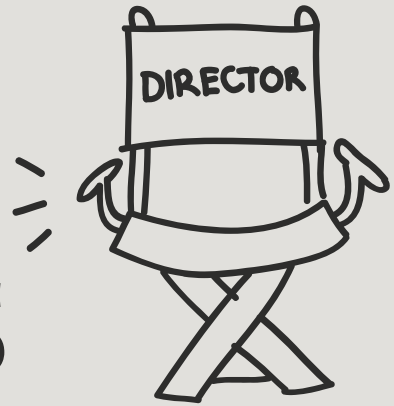
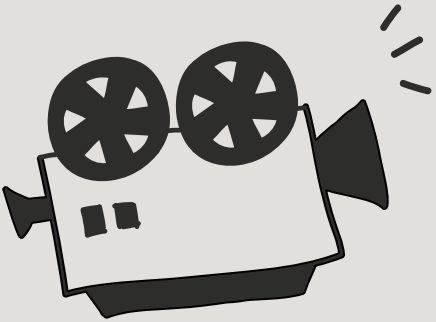
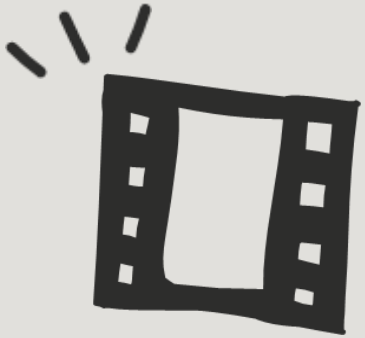
PROJECT-5

IMDB MOVIE ANALYSIS







Name: Bhavya Sri Duggina

PROJECT DESCRIPTION

- Our project aims to delve into the influential factors behind a movie's success on IMDB.
- This project focuses on leveraging Excel and Statistical skills to Conduct IMDB Movie Analysis to make informed decisions in the directors and producers further projects.
- As a Lead Data Analyst ,the tasks involve investigating Films IMDB analysis and understanding trends such as the top movie ratings, profitable movies directors impact on movie can provide valuable insights for the film department for their future films
- It's crucial to emphasize that in the dynamic and competitive movie industry, making informed decisions is paramount. Data analysis serves as a powerful tool for movie producers, directors, and investors, enabling them to navigate the complexities of the industry, mitigate risks, and optimize their chances of creating not just successful but critically acclaimed films.
- The main goal is to analyze dataset to derive valuable insights that can help and improve the producers and directors predictions and draw meaningful conclusions.

APPROACH

1.IMPORTING THE DATASET INTO EXCEL

2.DATA CLEANING AND QUALITY CHECK

3.EXPLORE THE DATASET AND EXTRACT THE INSIGHTS

4.GENERATE EFFICIENT REPORT

TECH STACK USED:

Tech-stack used in this project are Microsoft Excel 2013 and Microsoft PowerPoint

➢ Microsoft Excel2013:

Purpose: Microsoft Excel 2013 is a pivotal tool for this IMDB Movie analytics project. It is utilized for various data-related tasks, including data cleaning, manipulation, and exploratory data analysis (EDA). Excel's features like data validation, pivot tables, and charting capabilities are instrumental in processing and visualizing data.

➢ Microsoft PowerPoint 2013:

Purpose: Microsoft PowerPoint 2013 plays a crucial role in this project by enabling the creation of a compelling and informative presentation. It allows us to present the project's objectives, methodologies, findings, and recommendations in a structured and visually engaging manner. PowerPoint facilitates effective communication of complex data-driven insights to decision-makers.

EXPLORATORY DATA ANALYSIS

- Analyze the dataset and find the columns which are not necessary for the analysis and delete them

- Handling missing values

  Check for the null values and delete them

- Find Duplicates

  Duplicates affects the analysis so check for duplicates and delete them

  101 duplicates are found in the dataset and deleted as deleting duplicates improves accuracy

After successful completion of EDA we can extract insights from the dataset.

A picture is worth a thousand words

INSIGHTS

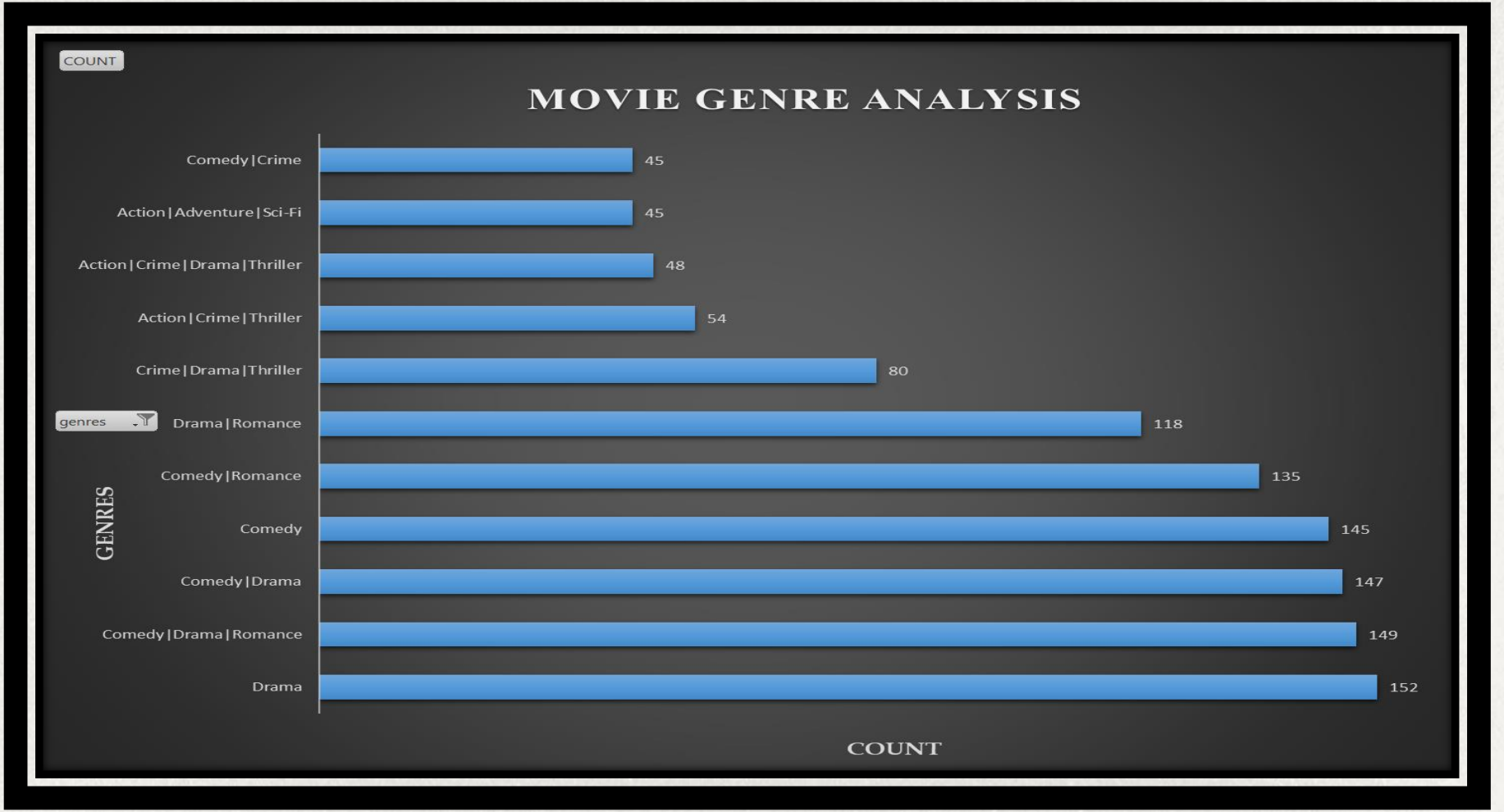
A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics of the IMDB scores.

GENRES	COUNT
Drama	152
Comedy|Drama|Romance	149
Comedy|Drama	147
Comedy	145
Comedy|Romance	135
Drama|Romance	118
Crime|Drama|Thriller	80
Action|Crime|Thriller	54
Action|Crime|Drama|Thriller	48
Action|Adventure|Sci-Fi	45
Comedy|Crime	45
Grand Total	**1118**

Here are the top 11 most common genres

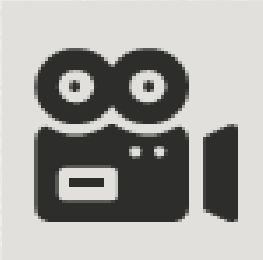
The top three most common movie genres are Drama with 152 films, followed closely by Comedy |Romance |Drama with 149 and Comedy |Drama featuring in 147 movies .


COUNT
MOVIE GENRE ANALYSIS
Comedy|Crime
45
Action|Adventure|Sci-Fi
45
Action|Crime|Drama|Thriller
48
Action|Crime|Thriller
54
Crime|Drama|Thriller
80
genres
Drama|Romance
118
Comedy|Romance
135
Comedy
145
Comedy|Drama
147
Comedy|Drama|Romance
149
Drama
152
GENRES
COUNT

GENRE IMPACT ON IMDB								
GENRES	AVERAGE	MEDIAN	MODE	MIN	MAX	RANGE	STD	VAR
Drama	7.040131579	7.15	7.3	3.4	8.8	5.4	0.828621229	0.686613141
Comedy|Drama|Romance	6.495302013	6.5	6.5	4.3	8	3.7	0.742780516	0.551722895
Comedy|Drama	6.583673469	6.7	6.7	3.3	8.8	5.5	0.854284187	0.729801472
Comedy	5.840689655	6	6.5	1.9	8	6.1	1.213117759	1.471654697
Comedy|Romance	5.896296296	6	6.1	2.7	8.4	5.7	0.873257622	0.762578875
Drama|Romance	6.95	7.1	7.2	4.1	8.1	4	0.746375422	0.557076271
Crime|Drama|Thriller	6.865	7	6.1	5.1	8.5	3.4	0.777994216	0.605275
Action|Crime|Thriller	6.403703704	6.5	6.5	4.4	7.6	3.2	0.630685389	0.39776406
Action|Crime|Drama|Thriller	6.522916667	6.5	6.5	5.1	9	3.9	0.712460403	0.507599826
Action|Adventure|Sci-Fi	6.668888889	6.8	6.6	2.4	8.4	6	1.199967078	1.439920988
Comedy|Crime	6.037777778	6.1	6.7	3.1	8.3	5.2	1.167007358	1.361906173

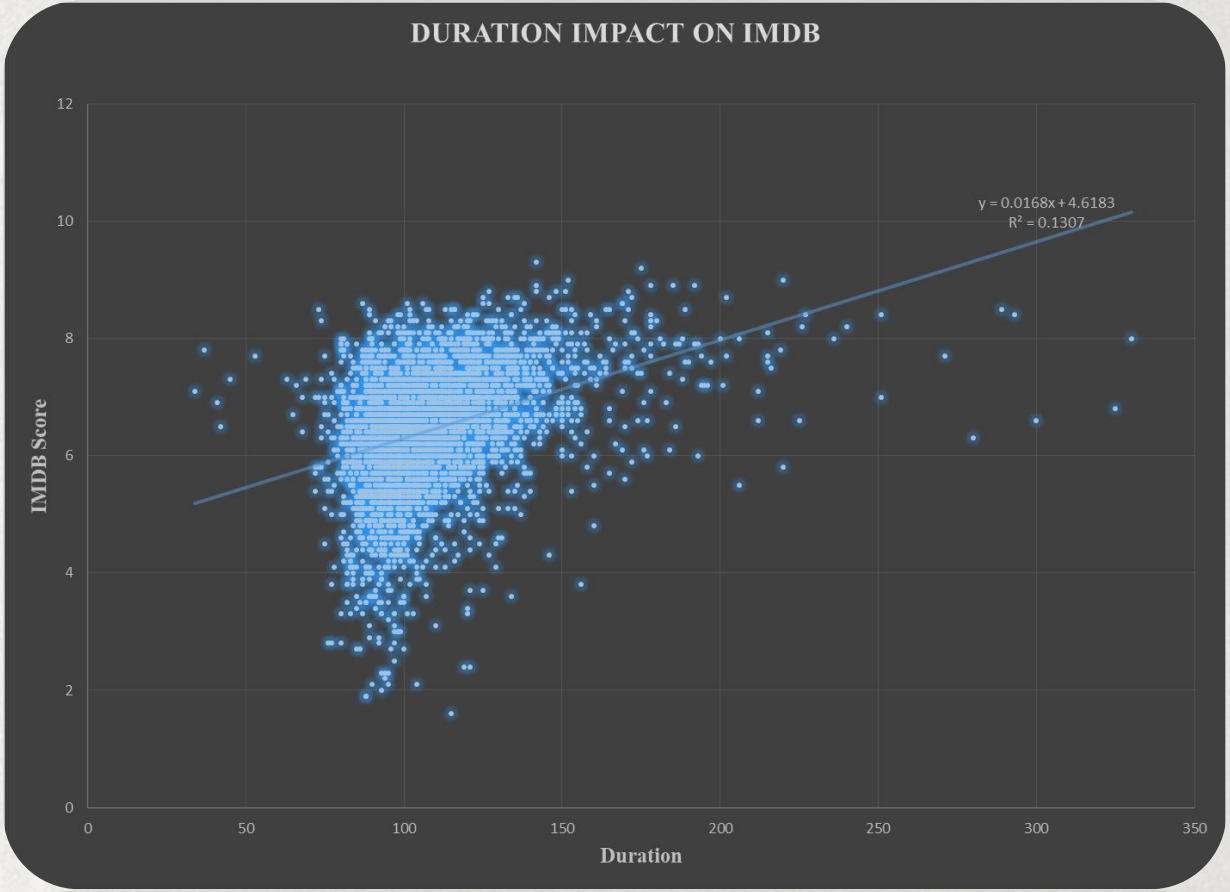
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

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



Descriptive Statistics of Duration

Distribution of movie duration	
MEAN	109.808505
MEDIAN	105
MODE	101
MIN	34
MAX	330
RANGE	296
STD	22.76019457
VAR	518.026457



Scatter plot between Duration and IMDB score

Trendline
Equation y=0.0168x+4.6183
R value=0.1307

Trendline is positive so there is a positive relationship between Duration and IMDB

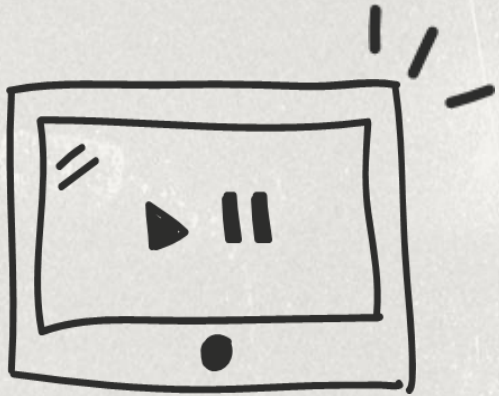
C. Language Analysis: Situation: Examine the distribution of movies based on their language.

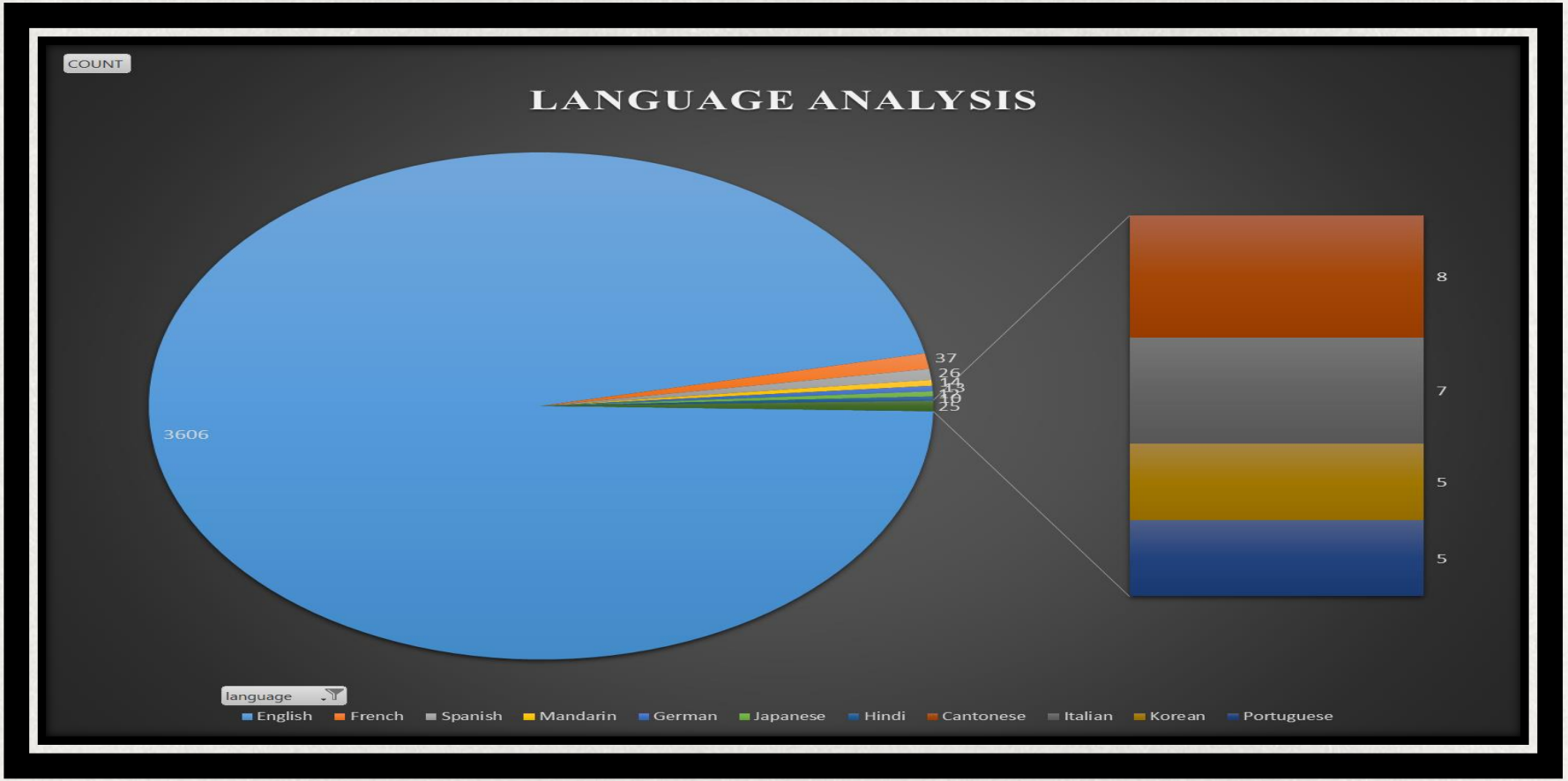
Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

TOP 11 LANGUAGES	COUNT
English	3606
French	37
Spanish	26
Mandarin	14
German	13
Japanese	12
Hindi	10
Cantonese	8
Italian	7
Korean	5
Portuguese	5
Grand Total	**3743**

Here are the Top 11 Languages used in movies

English tops the list by being in 3606 movies followed by French with 37 and Spanish with 26




COUNT
LANGUAGE ANALYSIS
3606
37
26
8
7
5
5
language
English
French
Spanish
Mandarin
German
Japanese
Hindi
Cantonese
Italian
Korean
Portuguese

	IMPACT OF LANGUAGE ON IMDB						
TOP 11 LANGUAGES	MEAN	MEDIAN	MIN	MAX	RANGE	STD	VAR
English	6.421436495	6.5	1.6	9.3	7.7	1.052353	1.107447
French	7.286486486	7.2	5.8	8.4	2.6	0.553691	0.306574
Spanish	7.05	7.15	5.2	8.2	3	0.810152	0.656346
Mandarin	7.021428571	7.25	5.6	7.9	2.3	0.73793	0.544541
German	7.692307692	7.7	6.1	8.5	2.4	0.615769	0.379172
Japanese	7.625	7.8	6	8.7	2.7	0.861322	0.741875
Hindi	6.76	7.05	4.8	8	3.2	1.054704	1.1124
Cantonese	7.2375	7.3	6.5	7.8	1.3	0.412121	0.169844
Italian	7.185714286	7	5.3	8.9	3.6	1.069618	1.144082
Korean	7.7	7.7	7	8.4	1.4	0.509902	0.26
Portuguese	7.76	8	6.1	8.7	2.6	0.875443	0.7664

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.

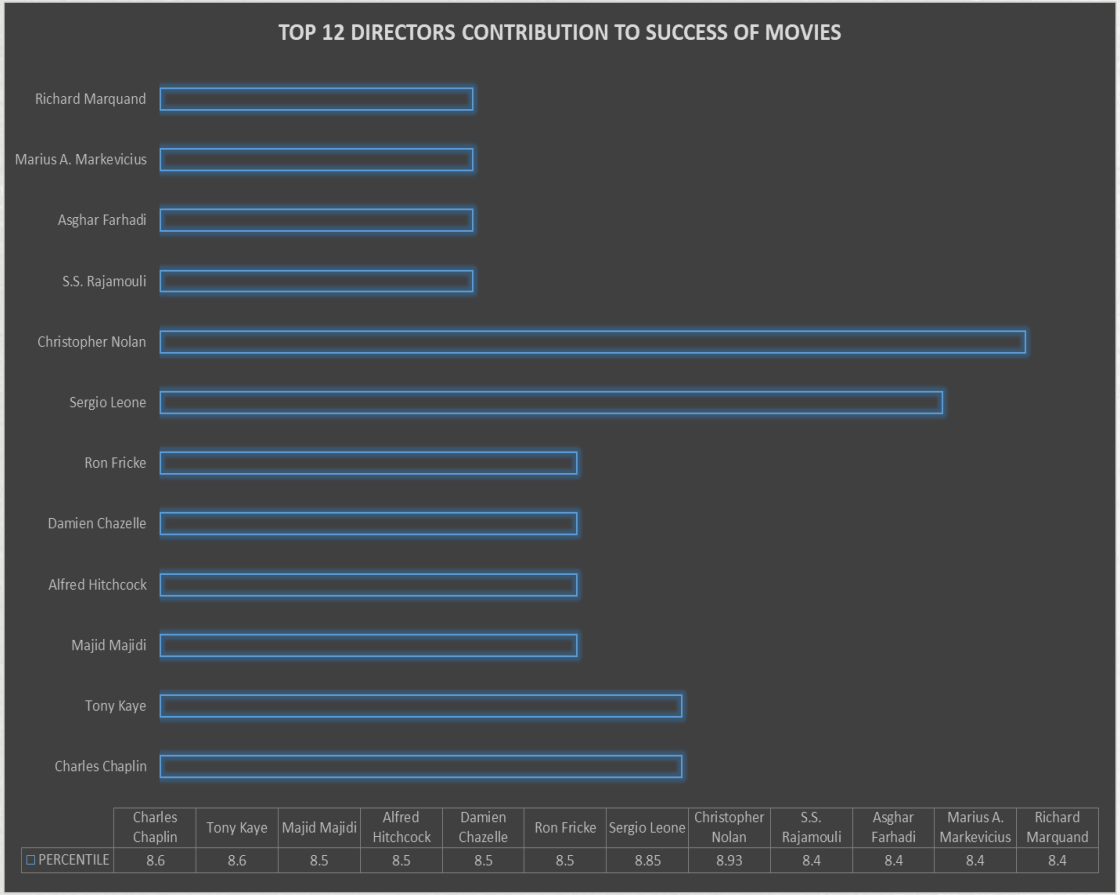
TOP Directors	Average of imdb_score
Charles Chaplin	8.6
Tony Kaye	8.6
Majid Majidi	8.5
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Ron Fricke	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
S.S. Rajamouli	8.4
Asghar Farhadi	8.4
Marius A. Markevic	8.4
Richard Marquand	8.4
Grand Total	**8.452380952**

TOP 12 Directors based on IMDB Average Score

Charles Chaplin and Tony Kaye Tops the list with 8.6 Average IMDB Score

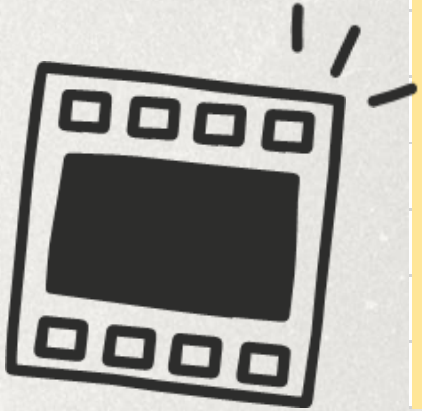
TOP 12 DIRECTORS BASED ON IMDB AVERAGE CONTRIBUTION TO THE SUCCESS OF MOVIES

TOP 12 DIRECTORS	PERCENTILE
Charles Chaplin	8.6
Tony Kaye	8.6
Majid Majidi	8.5
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Ron Fricke	8.5
Sergio Leone	8.85
Christopher Nolan	8.93
S.S. Rajamouli	8.4
Asghar Farhadi	8.4
Marius A. Markevicius	8.4
Richard Marquand	8.4



DIRECTORS WITH HIGHEST PERCENTILE OTHER THAN TOP DIRECTORS WITH HIGHEST AVERAGE IMDB SCORE

DIRECTORS	PERCENTILE
Frank Darabont	9.18
Francis Ford Coppola	9.12
Christopher Nolan	8.93
Peter Jackson	8.86
Sergio Leone	8.85
Quentin Tarantino	8.76
David Fincher	8.71
Irvin Kershner	8.67
Milos Forman	8.66
Akira Kurosawa	8.64



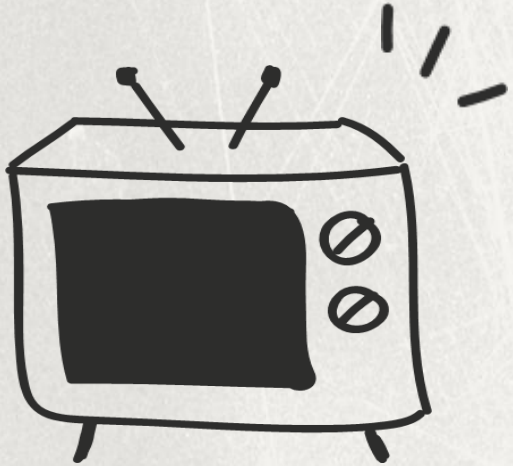
E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

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

Correlation between Movie Budgets and Gross Earnings is 0.0965

There exists a positive correlation between Budget and Movie Earnings

Correlation can be calculated from CORREL( ) function



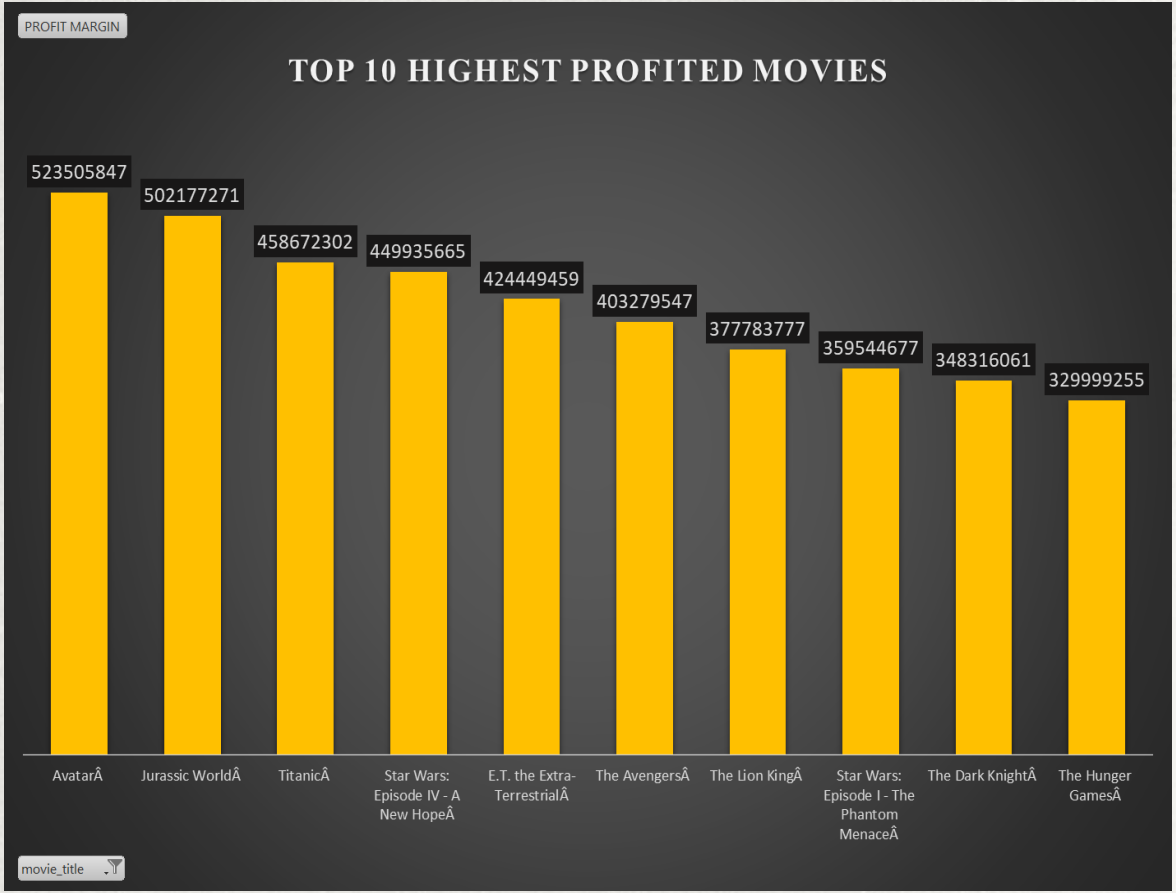
Profit Margin can be calculated as GROSS of movie – BUDGET of Movie

Highest Profited Movie – Avatar
Profit Margin-523505847

Biggest Financial Loss Movie-The HostÂ
Loss- 12213298588

TOP 10 MOVIES WITH HIGHEST PROFIT MARGIN

TOP 10 PROFITED MOVIES	PROFIT MARGIN
AvatarÂ	523505847
Jurassic WorldÂ	502177271
TitanicÂ	458672302
Star Wars: Episode IV - A New HopeÂ	449935665
E.T. the Extra-TerrestrialÂ	424449459
The AvengersÂ	403279547
The Lion KingÂ	377783777
Star Wars: Episode I - The Phantom MenaceÂ	359544677
The Dark KnightÂ	348316061
The Hunger GamesÂ	329999255



RESULT

Through the project, I successfully analyzed a dataset of IMDb movies, gaining valuable insights into various aspects of the film industry. This analysis allowed me to:

- Identify trends: I could discern patterns and trends in movie genres, helping me understand which genres were most commonly used and their combinations.
- Financial insights: By examining box office performance, budget, and profitability, I gained a deeper understanding of the financial aspects of the movie industry.
- Director influence: I could determine the most prolific directors in the dataset, shedding light on the influence of key individuals on film production.
- Language : Analyzing languages used in movies and their impact on movie success helped me grasp the importance of language in the film industry.

Overall, this project significantly contributed to my understanding of IMDB movie analysis by providing practical insights into the factors that influence a movie's success, the popularity of different genres, and the key players in the industry.

THANK YOU

BHAVYA SRI DUGGINA

View the file in Microsoft Excel for better Visualization

EXCEL file