

# IMDB MOVIE ANALYSIS

TRAINITY PROJECT-5





# CONTENT:

- ☐ PROJECT DESCRIPTION
- ☐ APPROACH
- ☐ TECH-STACK USED
- ☐ INSIGHTS
- ☐ RESULT
- ☐ EXCELSHEET HYPERLINK



# PROJECT DESCRIPTION

In this project, I focused on analyzing dataset sourced from IMDB Movie given by team, This project investigate the factors influencing success of movie on IMDB , through the task such as genre analysis, movie duration analysis, language analysis, director analysis and budget analysis, my aim to provide actionable insights .

The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.







# APPROACH:

- Firstly, I imported the IMDB dataset in MS Excel.
- Then thoroughly cleaned the data by manipulating dataset.
- Deleted duplicate rows and removed the blanks.
- Looked into a different aspects according to analytical task and eliminated unwanted columns.
- Performed specific analysis for each task by utilizing Excel's functionalities, different formulas and visualized using pivot tables, charts.
- Moreover, Executed statistical analysis as calculation of mean, median, mode, variance, standard deviation, percentile, profit/loss and correlation, to achieve meaningful insights.

# THE 5 WHYS APPROACH

1. Why do some genres perform better than others?

Viewer preferences and global trends vary

2. Why do certain actors consistently appear in high-rated films?

They choose roles in well-written scripts

5. Why are certain directors' movies always well-received?

They have a unique style and maintain high creative control over their projects.

3. Why do some movies with high budgets still perform poorly?

Budget allocation may prioritize visual effects over plot and character development.

4. Why do some low-budget movies become critically acclaimed?

Why do some low-budget movies become critically acclaimed?

# TECH-STACK USED



❖ **Microsoft Excel:** Microsoft Excel is a powerful spreadsheet application used for data organization, analysis, and visualization. Excel's user-friendly interface and widespread compatibility ensures ease of use and accessibility for stack holders involved in the project.



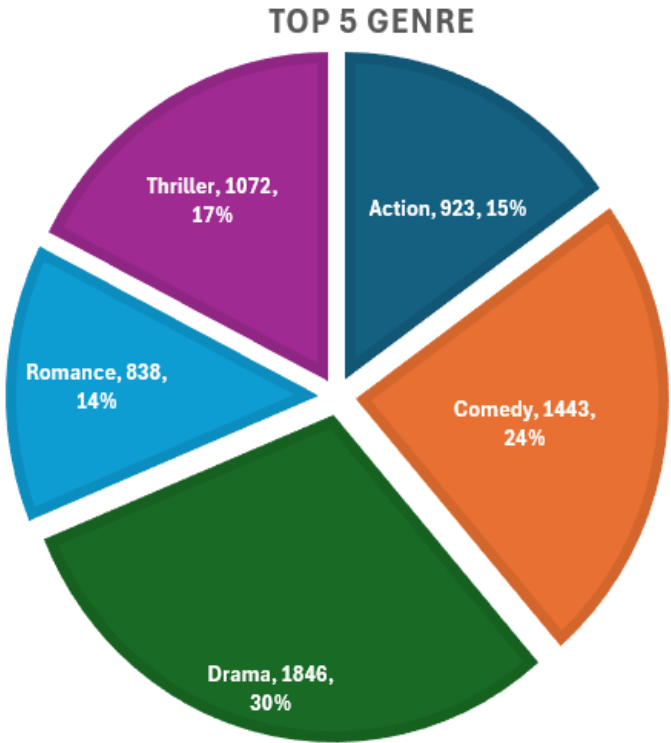
❖ **Microsoft power point:** It is a presentation software used to create slideshows for meetings, lectures, and other events. It allows users to combine text, images, charts, videos, and animations into visually engaging presentations. PowerPoint offers various templates, design tools, and transitions to enhance the visual appeal of slides.

Row Labels	Count of GENRE(Separated)
Action	923
Adventure	751
Animation	194
Biography	237
Comedy	1443
Crime	692
Documentary	45
Drama	1846
Family	431
Fantasy	486
Film-Noir	1
genres	1
History	146
Horror	374
Music	149
Musical	95
Mystery	371
Romance	838
Sci-Fi	478
Sport	143
Thriller	1072
War	149
Western	57
(blank)	
<b>Grand Total</b>	<b>10922</b>

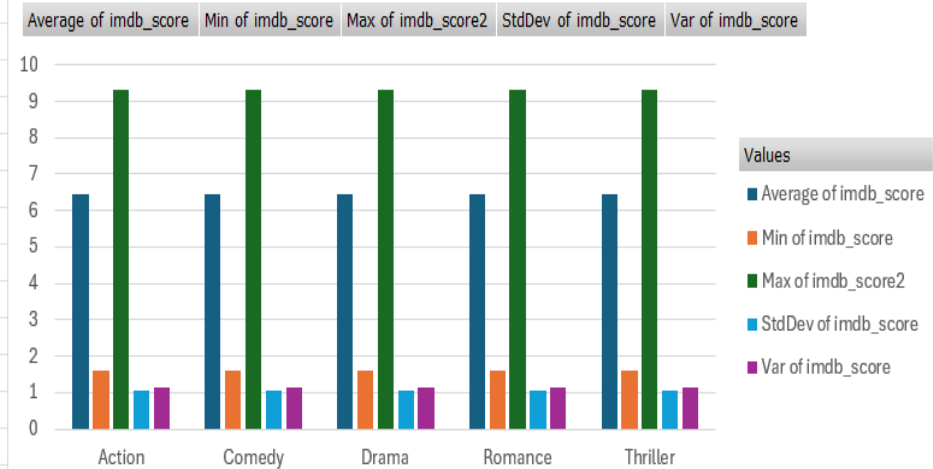
# 1) MOVIE GENRE ANALYSIS

**Task:** Analyze the distribution of movie genres and their impact on the IMDB score.

Count of GENRE(Separated)



genres	Average of imdb_score	Min of imdb_score	Max of imdb_score2	StdDev of imdb_score	Var of imdb_score
Action	6.463474692	1.6	9.3	1.057885897	1.11912257
Comedy	6.463474692	1.6	9.3	1.057885897	1.11912257
Drama	6.463474692	1.6	9.3	1.057885897	1.11912257
Romance	6.463474692	1.6	9.3	1.057885897	1.11912257
Thriller	6.463474692	1.6	9.3	1.057885897	1.11912257
<b>Grand Total</b>	<b>6.463474692</b>	<b>1.6</b>	<b>9.3</b>	<b>1.057885897</b>	<b>1.11912257</b>



GENRE(Separated)



# 1) INSIGHT

In movie genre analysis the main task was separating the genre column from a given dataset ,using excel's Text to Column feature from ribbon I separated the data and combined altogether using formula TOCOL, through this process this task became easy to analyze.

On analysing ,I found the top 5 genre most used in movie industry by filtering and visualized them using pivot table ,those are (Action, Comedy, Drama, Romance, Thriller) afterword I went through the statistical analysis of the top 5 movie genres, after analysing ,I determine that “DRAMA” is the most commonly used genre. Except the ROMANCE , all 4 genres had the maximum imdb ratings .





## 2) MOVIE DURATION ANALYSIS

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

FOR MOVIE DURATION	
MEAN	114.0343
MEADIN	108
SD	28.24518



## 2) INSIGHT

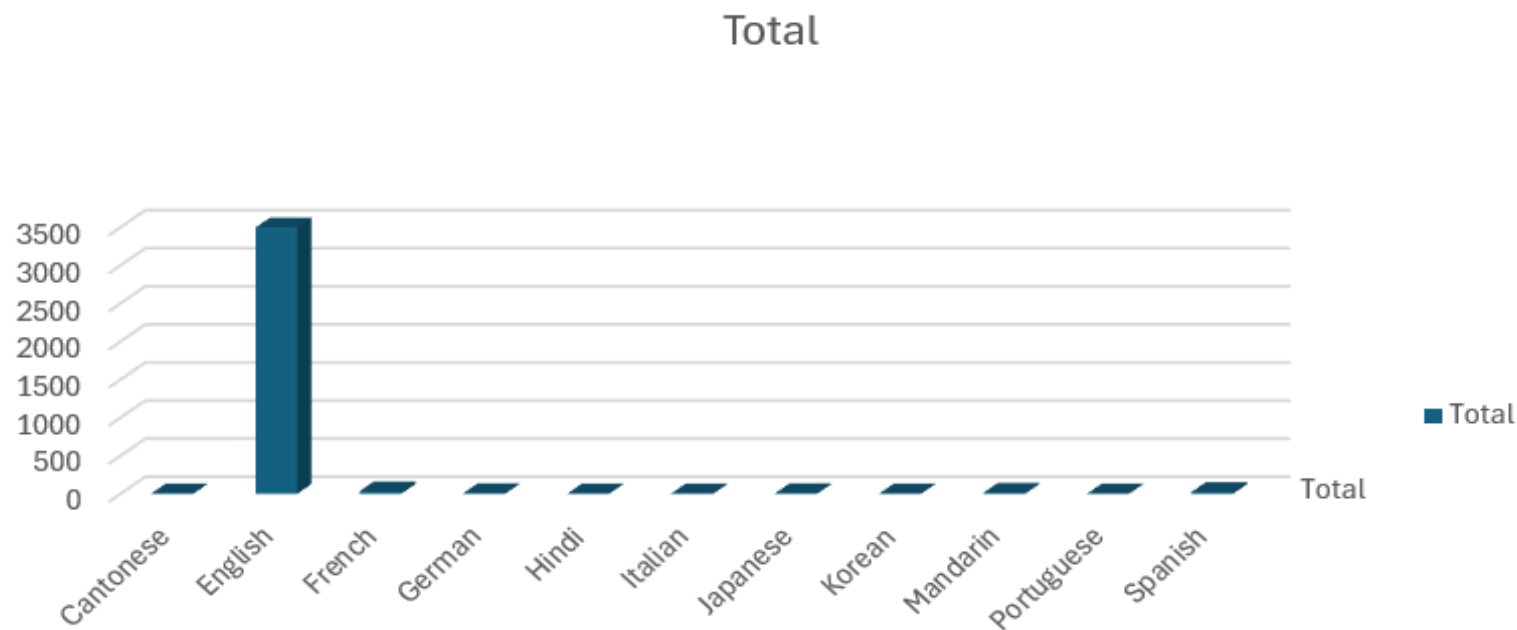
- Through this movie duration analysis, I determine descriptive statistics for imdb scores across various genres. To analyze , I used scatter plot with the duration of movie on x-axis and their respective IMDB score on Y-axis.
- On analysis o scatter plot we understand, There is a slight **positive correlation** between movie duration and IMDB score, as indicated by the upward-sloping trend line. This suggests that longer movies tend to have higher IMDB scores, but the correlation is not very strong.
- Most scores fall between **5 and 8**, with very few exceeding 9 or dropping below 3.



# 3) LANGUAGE ANALYSIS

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

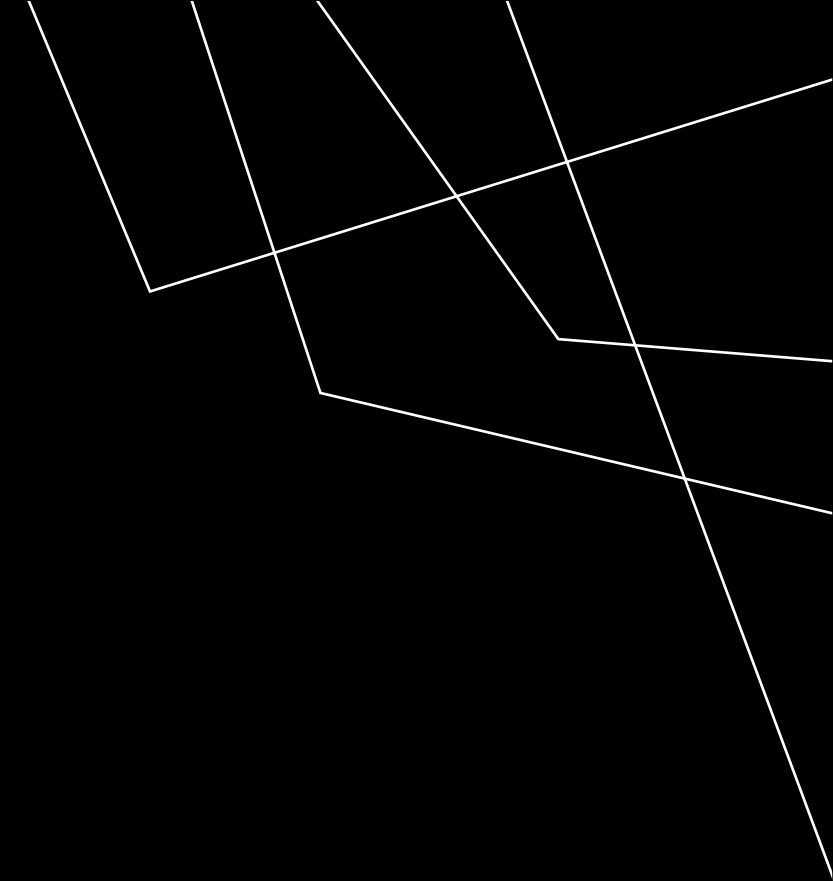
Count of language



Row Labels	Count of language	MEAN	MEDIAN	SD
Aboriginal	2	6.463475	6.6	1.057886
Arabic	1	6.463475	6.6	1.057886
Aramaic	1	6.463475	6.6	1.057886
Bosnian	1	6.463475	6.6	1.057886
Cantonese	7	6.463475	6.6	1.057886
Czech	1	6.463475	6.6	1.057886
Danish	3	6.463475	6.6	1.057886
Dari	2	6.463475	6.6	1.057886
Dutch	3	6.463475	6.6	1.057886
English	3498	6.463475	6.6	1.057886
Filipino	1	6.463475	6.6	1.057886
French	34	6.463475	6.6	1.057886
German	10	6.463475	6.6	1.057886
Hebrew	1	6.463475	6.6	1.057886
Hindi	5	6.463475	6.6	1.057886
Hungarian	1	6.463475	6.6	1.057886
Indonesian	2	6.463475	6.6	1.057886
Italian	7	6.463475	6.6	1.057886
Japanese	10	6.463475	6.6	1.057886
Kazakh	1	6.463475	6.6	1.057886
Korean	5	6.463475	6.6	1.057886
Mandarin	14	6.463475	6.6	1.057886
Maya	1	6.463475	6.6	1.057886
Mongolian	1	6.463475	6.6	1.057886
None	1	6.463475	6.6	1.057886
Norwegian	4	6.463475	6.6	1.057886
Persian	3	6.463475	6.6	1.057886
Portuguese	5	6.463475	6.6	1.057886
Romanian	1	6.463475	6.6	1.057886
Russian	1	6.463475	6.6	1.057886
Spanish	23	6.463475	6.6	1.057886
Thai	3	6.463475	6.6	1.057886
Vietnamese	1	6.463475	6.6	1.057886
Zulu	1	6.463475	6.6	1.057886
(blank)		6.463475	6.6	1.057886
Grand Total	3655	6.463475	6.6	1.057886

# 3) INSIGHT

- To do the language analysis, here I first counted the no. of movies for each language and English is overwhelmingly the most common language, with **3,498** movies, comprising the vast majority of the dataset.
- Apart from English, other languages such as **French (34 movies)**, **Spanish (23 movies)**, **Mandarin (14 movies)**, **Korean (14 movies)**, and **German (10 movies)** have a smaller representation.
- I calculated descriptive statistics such as mean, median, and standard deviation for IMDB scores across different languages and compared statistics to understand the impact of language on movie ratings.
- Visualized languages and its movie counts through pivot table.

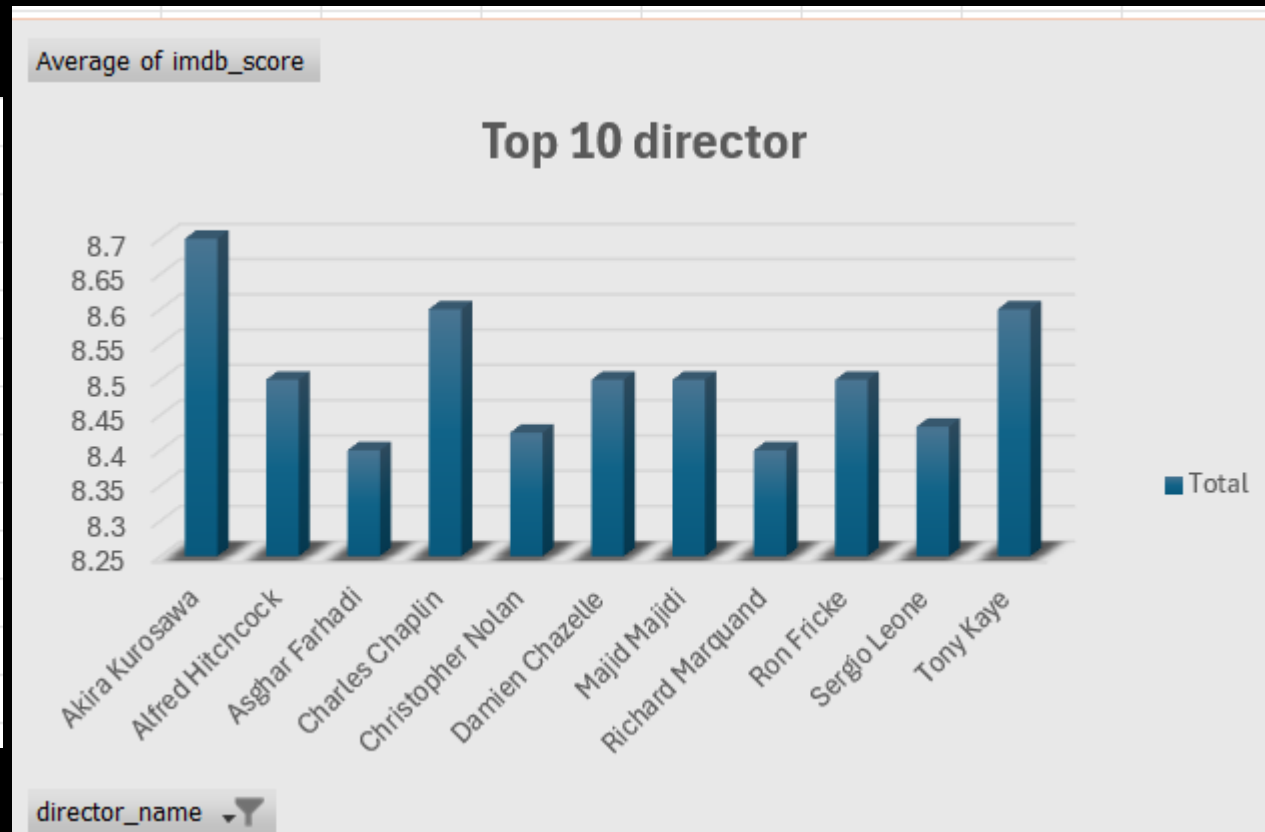




# 4)DIRECTOR ANALYSIS

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

Row Labels	Average of imdb_score
Akira Kurosawa	8.7
Alfred Hitchcock	8.5
Asghar Farhadi	8.4
Charles Chaplin	8.6
Christopher Nolan	8.425
Damien Chazelle	8.5
Majid Majidi	8.5
Richard Marquand	8.4
Ron Fricke	8.5
Sergio Leone	8.43333333
Tony Kaye	8.6
<b>Grand Total</b>	<b>8.47</b>



movie_title	imdb_score	director_name	PERCENTILE
Avatar	7.9	James Cameron	0.929
Pirates of the Caribbean: At World's End	7.1	Gore Verbinski	0.696
Spectre	6.8	Sam Mendes	0.585
The Dark Knight Rises	8.5	Christopher Nolan	0.987
Spider-Man 3	6.2	Sam Raimi	0.342
Tangled	7.8	Nathan Greno	0.908
Avengers: Age of Ultron	7.5	Joss Whedon	0.832
Harry Potter and the Half-Blood Prince	7.5	David Yates	0.832
Batman v Superman: Dawn of Justice	6.9	Zack Snyder	0.623
Superman Returns	6.1	Bryan Singer	0.305
Quantum of Solace	6.7	Marc Forster	0.539
Pirates of the Caribbean: Dead Man's Chest	7.3	Gore Verbinski	0.77
The Lone Ranger	6.5	Gore Verbinski	0.454
Man of Steel	7.2	Zack Snyder	0.733
The Chronicles of Narnia: Prince Caspian	6.6	Andrew Adamson	0.495
The Avengers	8.1	Joss Whedon	0.957
Pirates of the Caribbean: On Stranger Tides	6.7	Rob Marshall	0.539
Men in Black 3	6.8	Barry Sonnenfeld	0.585
The Hobbit: The Battle of the Five Armies	7.5	Peter Jackson	0.832
The Amazing Spider-Man	7	Marc Webb	0.659
Robin Hood	6.7	Ridley Scott	0.54
The Hobbit: The Desolation of Smaug	7.9	Peter Jackson	0.93
The Golden Compass	6.1	Chris Weitz	0.306
King Kong	7.2	Peter Jackson	0.734
Titanic	7.7	James Cameron	0.884
Captain America: Civil War	8.2	Anthony Russo	0.97
...	...	...	...

## 4)INSIGHT

To find the impact of a director on the IMDB Score of a movie:

1. I calculated total number of movies directed by each director.
2. I filtered out top 10 directors who's average IMDB score is great and visualized the relation using pivot table.
3. TO analyze their contribution to the success of movies ,I used PERCENTRANK.INC function, This gave me exact of the statistical values.
4. Percentile function computed values that range between 0 to 1, accordingly I understood , The highest scored movies is Avtar(7.9) movie has a percentile rank of 0.929, meaning it is the 92.9<sup>th</sup> percentile of the data.

# 5)BUDGET ANALYSIS

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

Highest Profit  
52350847

Correlation  
0.093820012

## INSIGHT:

- The correlation between movie budgets and gross earnings is 0.09382.
- The highest profit margin of a movie is 52350847.
- Hence, we get the relation as the budget of movie increase the gross amount of movie also increased.

gross	budget	profit
760505847	237000000	523505847
309404152	300000000	9404152
200074175	245000000	44925825
448130642	250000000	198130642
336530303	258000000	78530303
200807262	260000000	59192738
458991599	250000000	208991599
301956980	250000000	51956980
330249062	250000000	80249062
200069408	209000000	8930592
168368427	200000000	31631573
423032628	225000000	198032628
89289910	215000000	125710090
291021565	225000000	66021565
141614023	225000000	83385977
623279547	220000000	403279547
241063875	250000000	8936125
179020854	225000000	45979146
255108370	250000000	5108370
262030663	230000000	32030663
105219735	200000000	94780265
258355354	225000000	33355354
70083519	180000000	109916481
218051260	207000000	11051260
658672302	200000000	458672302
407197282	250000000	157197282
65173160	209000000	143826840
652177271	150000000	502177271
304360277	200000000	104360277
373377893	200000000	173377893
408992272	200000000	208992272
334185206	200000000	134185206



# ❖ RESULT

Working on real-life like projects always gives so meaningful hands on . By putting my learnings of advance excel into this project helped me to analyze different trends , patterns and understood how a successful project can be done. Now feeling ready to deal With more complex data with more confidence.







# Excel

HERE'S THE LINK TO EXCEL SHEET

An abstract geometric pattern consisting of several overlapping, tilted rectangles and lines, creating a complex, layered effect on the left side of the image.

# THANK YOU

PROJECT MADE BY: GITANJALI PEKAMWAR