

IMDb MOVIE ANALYSIS USING EXCEL

PRESENTED BY –
SOUVIK KARMAKAR

IMDb

Project Description:-

Title: "IMDb Movie Analysis: Decoding the Secrets of Cinematic Success"

- In this captivating project, we delve into the vast world of IMDb movies to unravel the factors that contribute to cinematic success, as defined by high IMDb ratings.
- Our exploration aims to provide valuable insights for movie producers, directors, and investors, empowering them to make informed decisions in their future projects.



Approach:-

- **Get the Data in Excel:**

To get started with our analysis, we first downloaded and opened the dataset in Excel.

- **Clean Up the Data:**

We then removed superfluous columns and duplicate values.

- **Find Answers with Formulas and Filters:**

To solve problems and find answers, I used Excel's formula and filter functions.

- **Show Data with Charts:**

To make the data easy to see and comprehend, Excel was used to create charts and graphs.

- **Write a Short Report:**

In our project report, we have summarized our findings.

Techstack Used:-

Excel has been utilized to comprehend the data and derive some insights from it.



Data Analytics Task:-

required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

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 (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

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

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

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

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

D. Director Analysis: Influence of directors on movie ratings.

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

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

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

INSIGHTS

- Movie Genre Analysis

Determined the most popular genres and how they affect ratings on IMDB.

- Movie Duration Analysis:

Patterns were discovered that suggest how long a film is affects how audiences perceive it.

- Language Analysis:

Analyzed languages to determine how language affects IMDB ratings.

- Director Analysis:

The importance of directors in influencing the success of films was uncovered through director analysis.

- Budget Analysis:

Correlation between film budgets and gross receipts was investigated in the budget analysis, also apart from that We examined the movie that has made the most money.



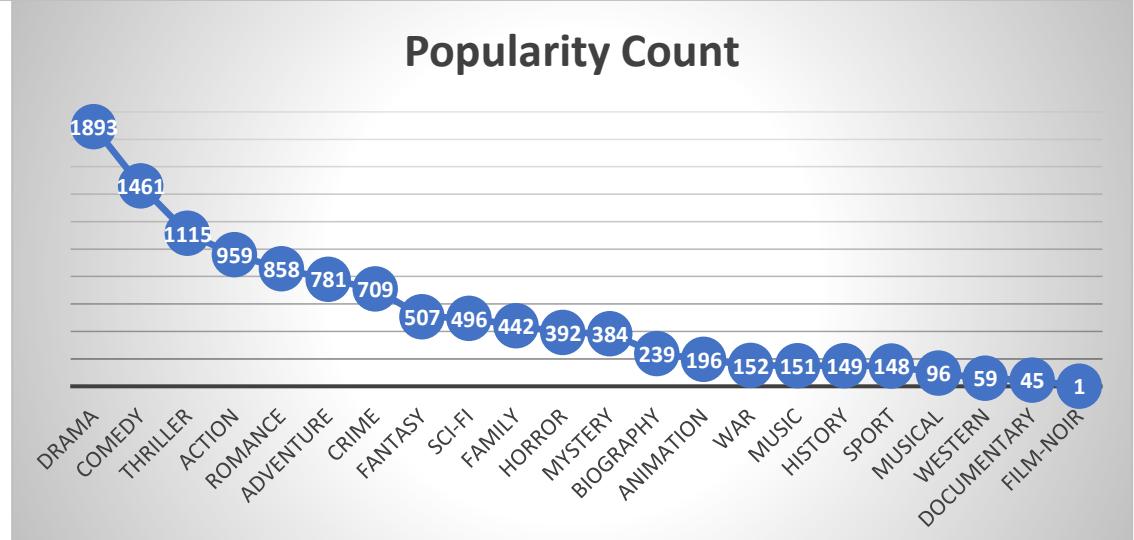
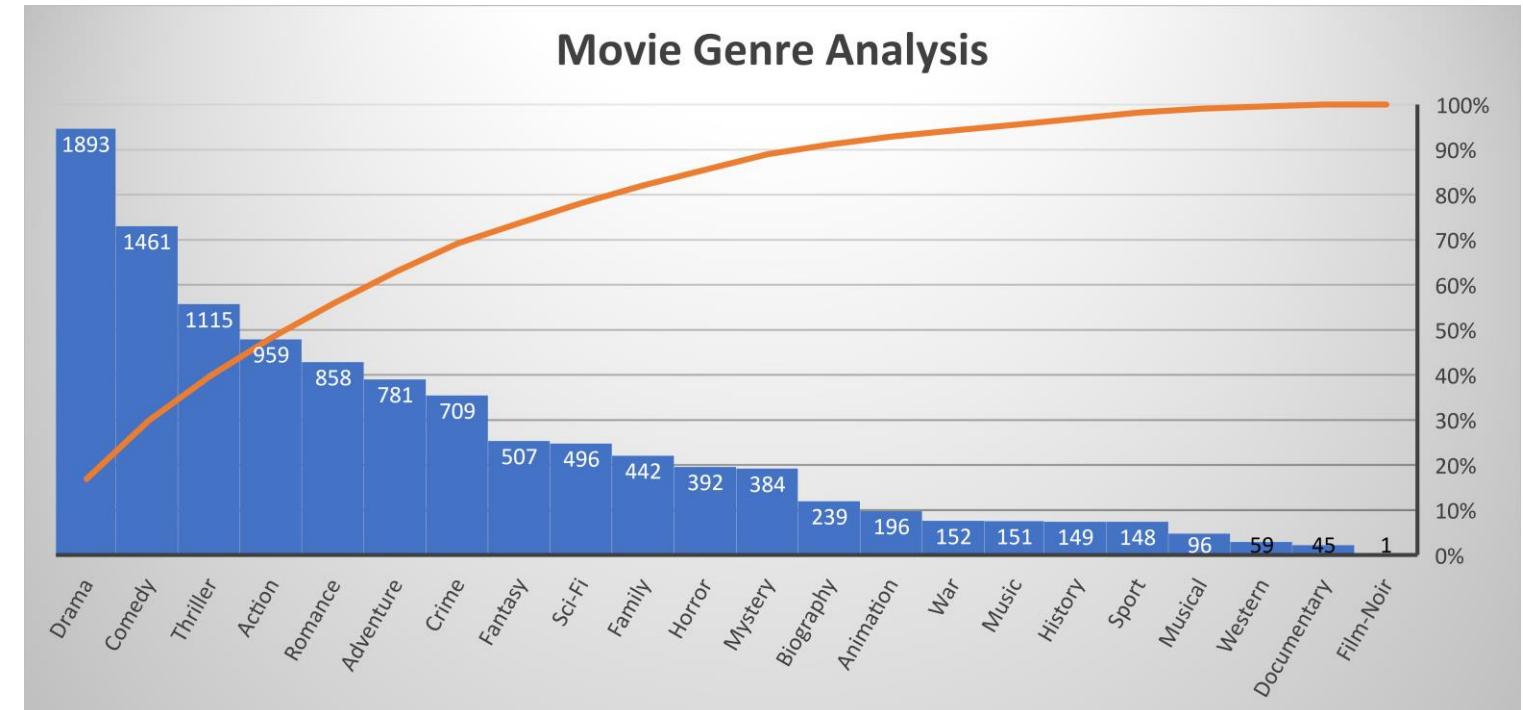
Data Preprocessing:-

- Removed unnecessary columns
- In order to clean, we first eliminated 45 duplicate values, leaving 4998 unique values.
- Since the color column is unimportant to our study, we have omitted it.



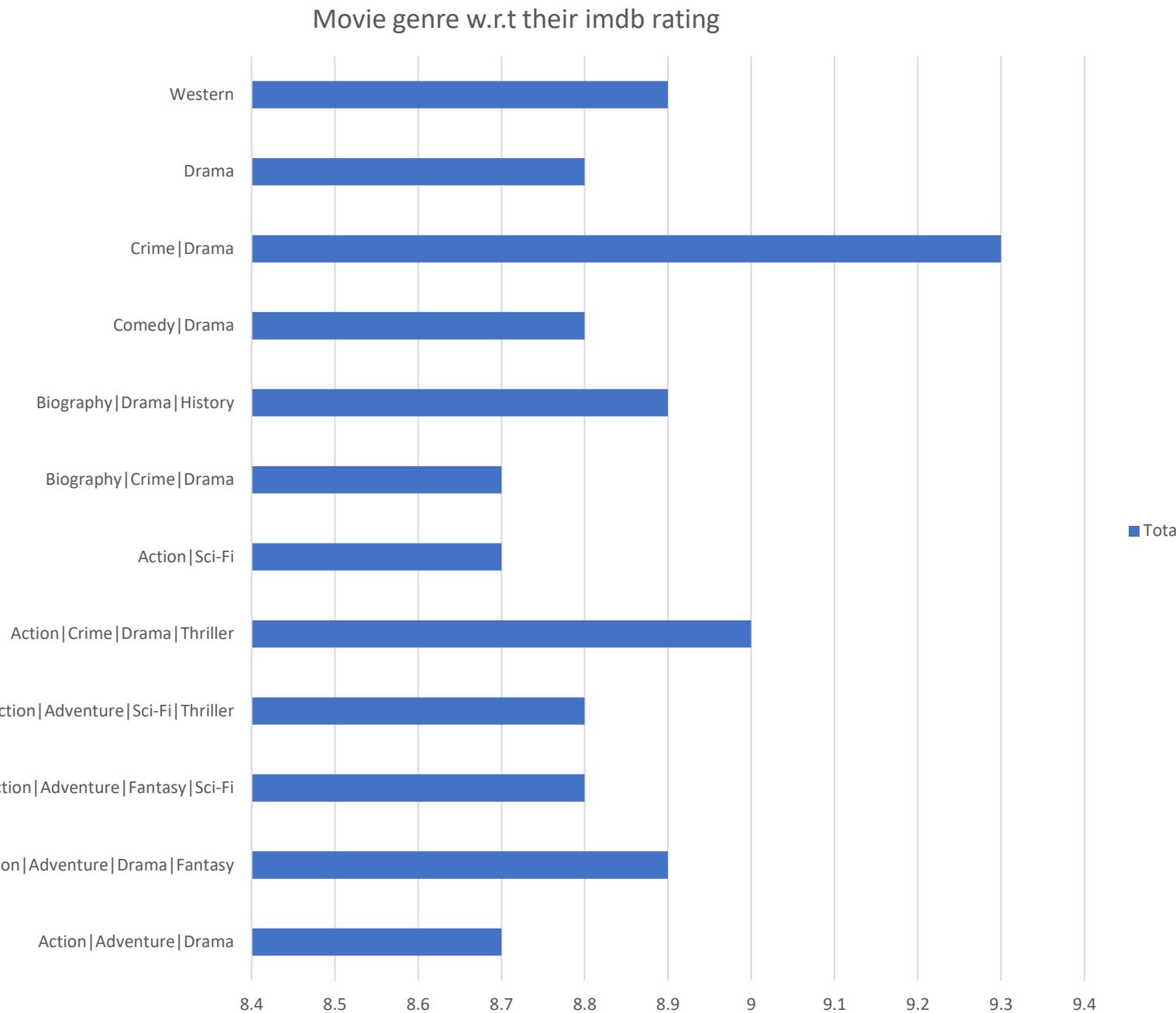
Task A:-(Movie Genre Analysis)

Movie Genre	Popularity Count	Percentage
Drama	1893	16.85213211
Comedy	1461	13.00632066
Thriller	1115	9.926110567
Action	959	8.537345322
Romance	858	7.638208849
Adventure	781	6.952728568
Crime	709	6.311759993
Fantasy	507	4.513487047
Sci-Fi	496	4.415561293
Family	442	3.934834862
Horror	392	3.489717796
Mystery	384	3.418499065
Biography	239	2.127659574
Animation	196	1.744858898
War	152	1.35315588
Music	151	1.344253539
History	149	1.326448856
Sport	148	1.317546515
Musical	96	0.854624766
Western	59	0.525238138
Documentary	45	0.400605359
Film-Noir	1	0.008902341



So from the data we can say that most popular genre Is 'Drama', 2nd most popular genre is 'comedy'

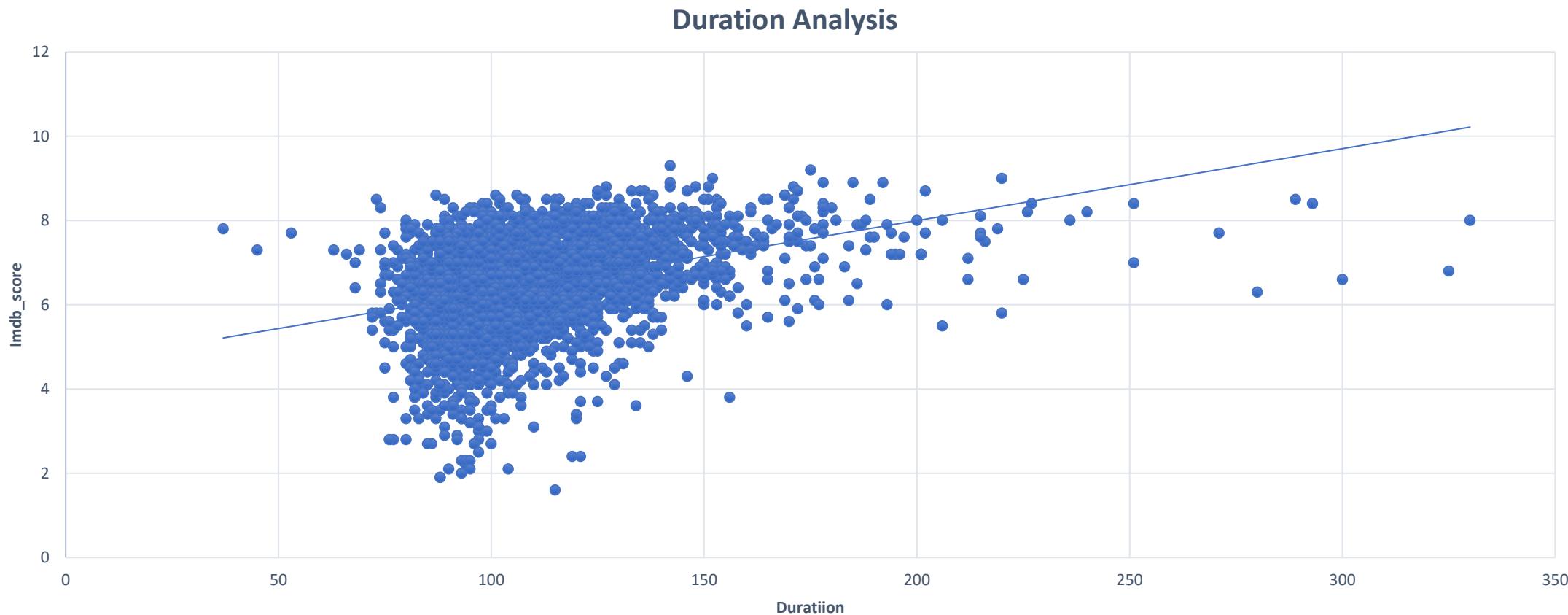
- Among all genres, Crime | Drama has the highest IMDb Score.
- Crime, Drama, Action, and Thriller ranked second on the list



Descriptive Statistics for Movie Genre Analysis:-

Row Labels	StdDev of imdb_score5	Var of imdb_score2	Sum of imdb_score	Count of imdb_score4	Average of imdb_score2	Max of imdb_score3	Min of imdb_score
Action Adventure Drama	1.2	1.39	33.6	5	6.7	8.7	5.8
Action Adventure Drama Fantasy	1.4	2.09	44.9	6	7.5	8.9	6.1
Action Adventure Fantasy Sci-Fi	1.0	1.04	74.3	10	7.4	8.8	5.8
Action Adventure Sci-Fi Thriller	0.9	0.77	216.5	34	6.4	8.8	5
Action Crime Drama Thriller	0.7	0.53	306.6	47	6.5	9	5.1
Action Sci-Fi	1.2	1.40	109.8	16	6.9	8.7	4.1
Biography Crime Drama	0.6	0.34	91.1	12	7.6	8.7	6.6
Biography Drama History	0.7	0.44	102.9	14	7.4	8.9	6.4
Comedy Drama	0.8	0.70	549.6	81	6.8	8.8	3.3
Crime Drama	0.9	0.73	260.1	34	7.7	9.3	5.3
Drama	0.6	0.39	611.1	82	7.5	8.8	4.6
Western	1.1	1.28	16.2	2	8.1	8.9	7.3
Grand Total	0.9	0.85	2416.7	343	7.0	9.3	3.3

Task B:-Movie Duration Analysis:-



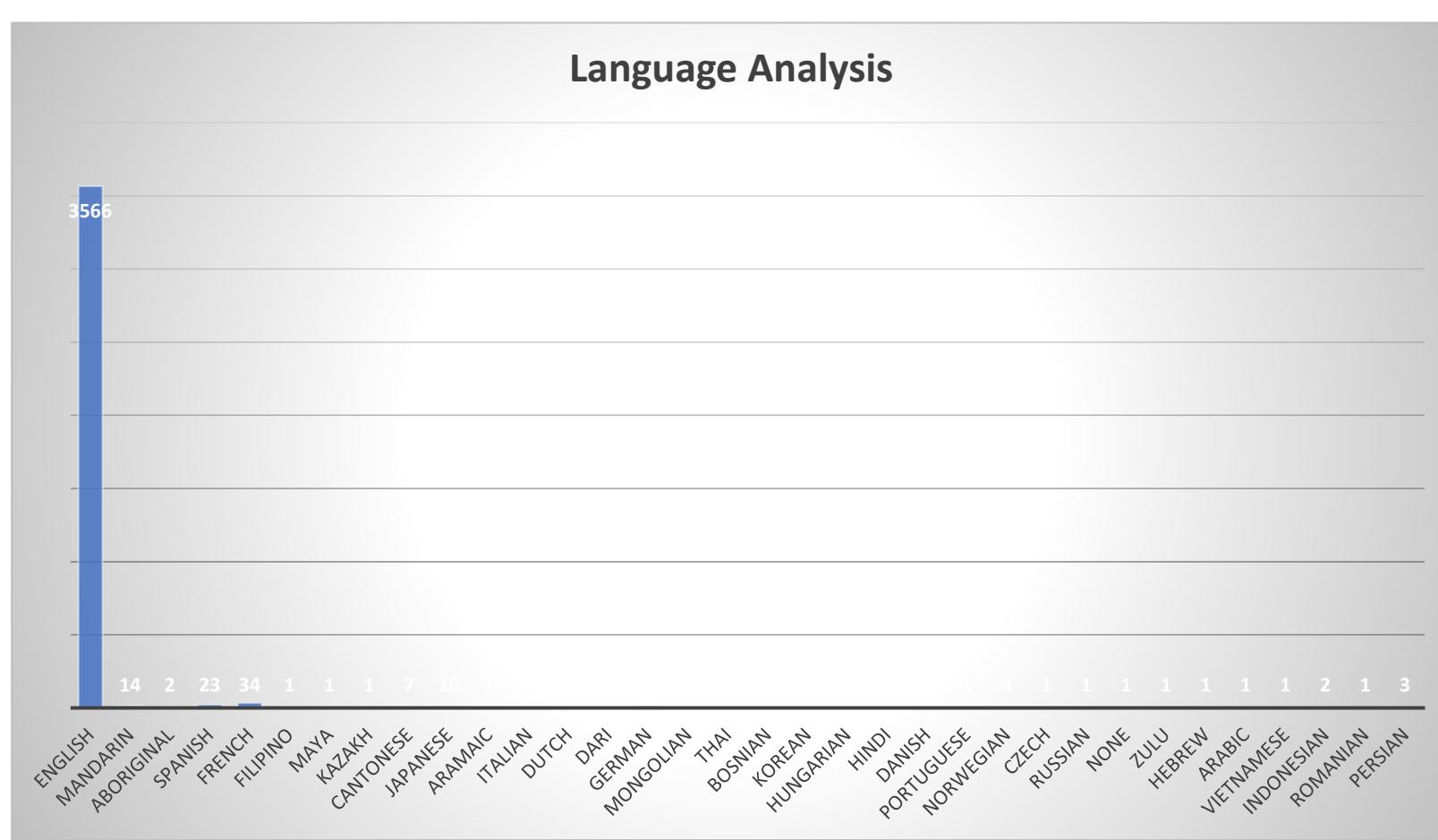
- We created a scatter plot to compare the duration of the movie and the imdb rating; the graph indicates positive correlation between the two.
- The maximum duration for movies with an average IMDb rating is 95–120 minutes.

Descriptive statistics for movie duration analysis:-

Movie_Title	Duration	imdb_score	mean	median	STD
Avatarâ	178	7.9	110.2635	106	22.67832
Pirates Of The Caribbean: At World'S Endâ	169	7.1	110.2453	106	22.65417
Spectreâ	148	6.8	110.2295	106	22.63672
The Dark Knight Risesâ	164	8.5	110.2194	106	22.63128
John Carterâ	132	6.6	110.2049	106	22.61713
Spider-Man 3â	156	6.2	110.199	106	22.61735
Tangledâ	100	7.8	110.1867	106	22.60791
Avengers: Age Of Ultronâ	141	7.5	110.1895	106	22.61033
Harry Potter And The Half-Blood Princeâ	153	7.5	110.1812	106	22.60772
Batman V Superman: Dawn Of Justiceâ	183	6.9	110.1696	106	22.59984
Superman Returnsâ	169	6.1	110.15	106	22.57124
Quantum Of Solaceâ	106	6.7	110.1342	106	22.5536
Pirates Of The Caribbean: Dead Man'S Chestâ	151	7.3	110.1353	106	22.55654
The Lone Rangerâ	150	6.5	110.1243	106	22.54959
Man Of Steelâ	143	7.2	110.1135	106	22.54312
The Chronicles Of Narnia: Prince Caspianâ	150	6.6	110.1046	106	22.53969
The Avengersâ	173	8.1	110.0939	106	22.5332
Pirates Of The Caribbean: On Stranger Tidesâ	136	6.7	110.0769	106	22.51253
Men In Black 3â	106	6.8	110.0699	106	22.51153

Task C:- (Language Analysis):-

<u>language</u>	<u>Count</u>
English	3566
Mandarin	14
Aboriginal	2
Spanish	23
French	34
Filipino	1
Maya	1
Kazakh	1
Cantonese	7
Japanese	10
Aramaic	1
Italian	7
Dutch	3
Dari	2
German	10
Mongolian	1
Thai	3
Bosnian	1
Korean	5
Hungarian	1
Hindi	5
Danish	3
Portuguese	5
Norwegian	4
Czech	1
Russian	1
None	1
Zulu	1
Hebrew	1
Arabic	1
Vietnamese	1
Indonesian	2
Romanian	1
Persian	3



Out of all the movies, English language films are the ones with the greatest distribution.

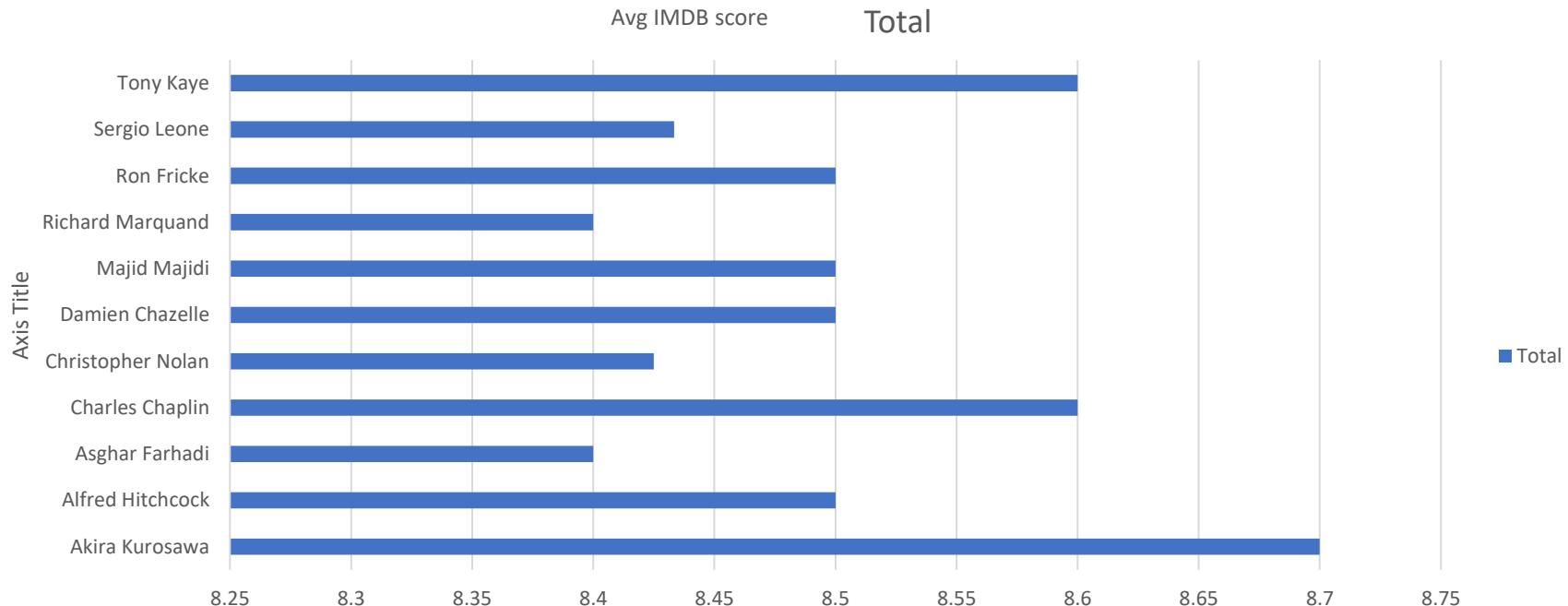
Approximately 97% of the entire movie database was in English.

Descriptive statistics for Language Analysis-

Row Labels	Var of imdb_score2	Sum of imdb_score2	StdDev of imdb_score5	Min of imdb_score3	Max of imdb_score	Count of imdb_score4	Average of imdb_score4
Aboriginal	0.6	13.9	0.8	6.4	7.5	2	7.0
Aramaic	0.0	7.1	0.0	7.1	7.1	1	7.1
Bosnian	0.0	4.3	0.0	4.3	4.3	1	4.3
Cantonese	0.1	44.1	0.4	6.7	7.8	6	7.4
Danish	0.0	16.4	0.1	8.1	8.3	2	8.2
Dari	0.0	7.6	0.0	7.6	7.6	1	7.6
Dutch	0.0	7.8	0.0	7.8	7.8	1	7.8
English	1.0	18475.5	1.0	1.6	9.3	2807	6.6
Filipino	0.0	6.7	0.0	6.7	6.7	1	6.7
French	0.2	157	0.5	6.6	8.4	21	7.5
German	0.2	71.5	0.4	7.3	8.5	9	7.9
Hebrew	0.0	8	0.0	8	8	1	8.0
Hindi	0.2	38	0.5	6.9	8	5	7.6
Indonesian	0.2	15.8	0.4	7.6	8.2	2	7.9
Italian	0.4	24.6	0.6	7.7	8.9	3	8.2
Japanese	0.1	57.6	0.4	7.7	8.7	7	8.2
Korean	0.3	38.5	0.6	7	8.4	5	7.7
Mandarin	0.2	74.4	0.5	6.5	7.9	10	7.4
Maya	0.0	7.8	0.0	7.8	7.8	1	7.8
Mongolian	0.0	7.3	0.0	7.3	7.3	1	7.3
Norwegian	0.4	21	0.6	6.4	7.6	3	7.0
Persian	0.0	16.9	0.1	8.4	8.5	2	8.5
Portuguese	0.1	24.8	0.4	8	8.7	3	8.3
Romanian	0.0	7.9	0.0	7.9	7.9	1	7.9
Russian	0.0	6.5	0.0	6.5	6.5	1	6.5
Spanish	0.5	97.4	0.7	5.9	8.2	13	7.5
Telugu	0.0	8.4	0.0	8.4	8.4	1	8.4
Thai	0.4	13.3	0.6	6.2	7.1	2	6.7
Zulu	0.0	7.3	0.0	7.3	7.3	1	7.3
Grand Total	1.0	19287.4	1.0	1.6	9.3	2914	6.6

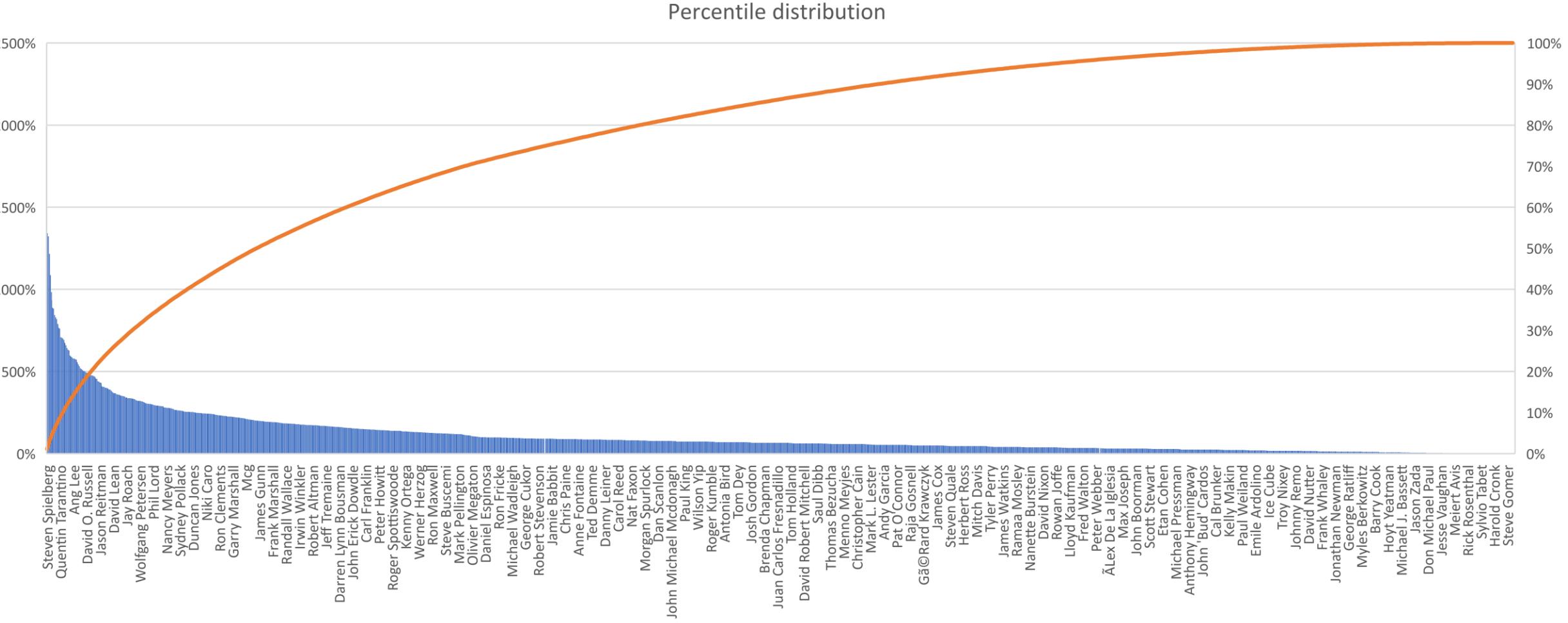
Task D:-(Director Analysis: Influence of directors on movie ratings)

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	8.433333333
Tony Kaye	8.6
Grand Total	8.47



- According to the graph we can easily conclude that Best director is **Akira Kurosawa** with mean IMDB score of **8.7**.

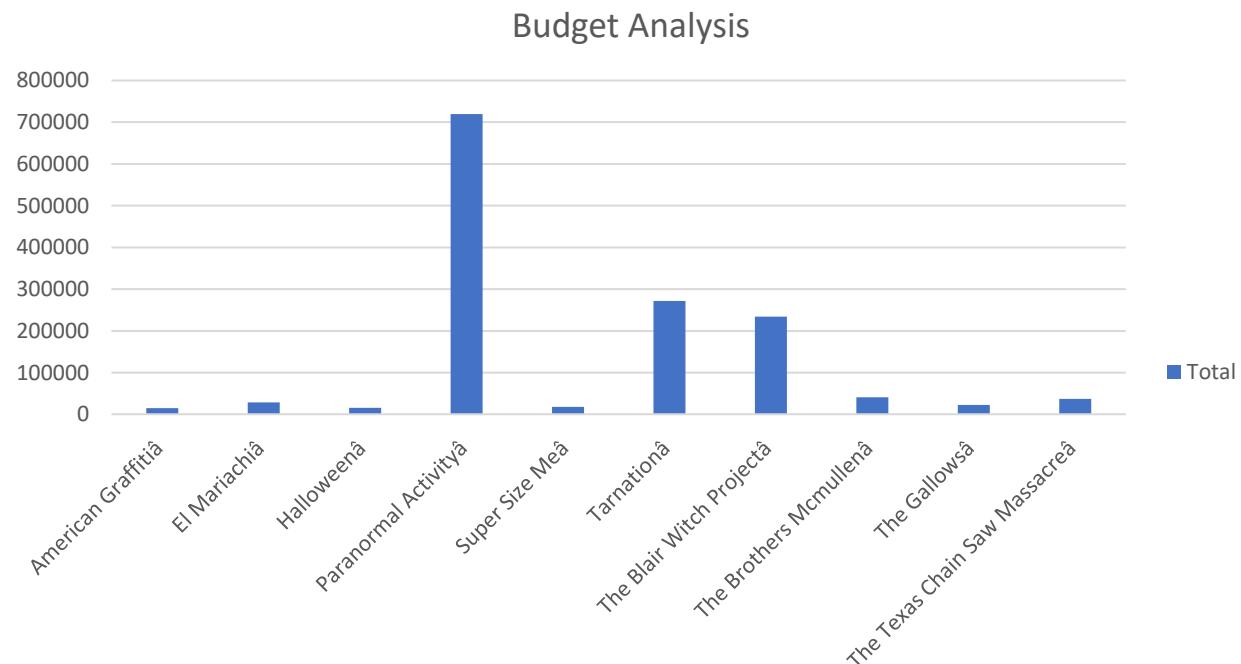
Percentile Distribution:-



Task E:-(Budget Analysis:-Top Highest profit given Movies)

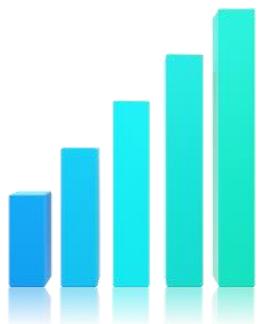
Movie_Title	budget	Gross	Profit Margin	Corr
Paranormal Activityâ	15000	107917283	719348.5533	0.098318
Tarnationâ	218	592014	271466.055	0.09837
The Blair Witch Projectâ	60000	140530114	234116.8567	0.098338
The Brothers McMullenâ	25000	10246600	40886.4	0.098428
The Texas Chain Saw Massacreâ	83532	30859000	36842.72854	0.098385
El Mariachiâ	7000	2040920	29056	0.09837
The Gallowsâ	100000	22757819	22657.819	0.098338
Super Size Meâ	65000	11529368	17637.48923	0.098318
Halloweenâ	300000	47000000	15566.66667	0.098287
American Graffitiâ	777000	115000000	14700.5148	0.098283

- “Paranormal Activityâ ”took the 1st place among all the movies with profit margin of 719348.5533
- Most of the movies are with Low and Medium budget and got Low and Medium Gross
- Risk is high as the budget is getting high



Results:-

- IMDB success factors for movies have been successfully identified.
- Analytical tools in Excel were used for a thorough understanding.
- based on the data set, provided useful insights.
- Enhanced decision-making for future projects that will pay off financially and increase viewer engagement and ratings.



Important Links:-

- Excel File Link:-

https://1drv.ms/f/s!Ap4mXYr_YuhTulaZjXvtEBVwM5is?e=w2oWHA

- Presentation Video link:-

https://1drv.ms/v/s!Ap4mXYr_YuhTulokhwnEZxL9gdzf?e=4SI4V1

