



IMDB Movie Analysis

Final Project-1

Presented by: Ayush Mahanta

INTRODUCTION

In this project, I analyzed a dataset of movies from IMDB. I looked at different aspects of the movies like genres, duration, language, directors, and budget to see how they relate to the movie's IMDB rating and financial success.



Approach

DATA CLEANING

In this step, I preprocess the data to make it ready for analysis. This involves fixing any missing values, removing duplicate entries, changing data types if needed, and possibly creating new features.

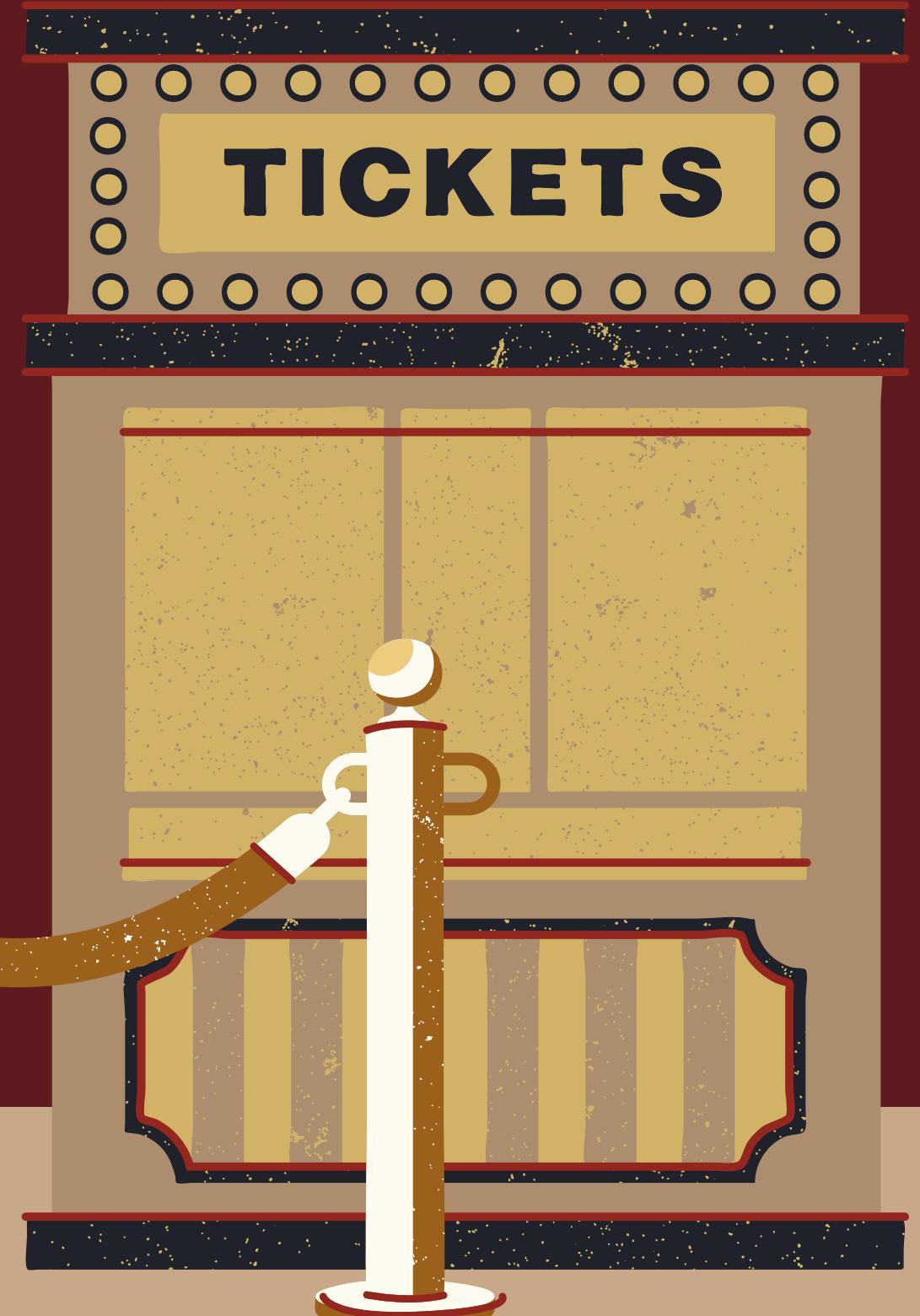
DATA ANALYSIS

Here, I explore the data to understand how different variables are related. I look at things like how movie ratings are influenced by factors such as genre, director, and budget. I also consider the year the movie was released, the actors in it, and other relevant factors.



Report Generation

Finally, I created a report that summarizes my findings. This report includes what I learned from the data analysis and is supported by visualizations and statistics.

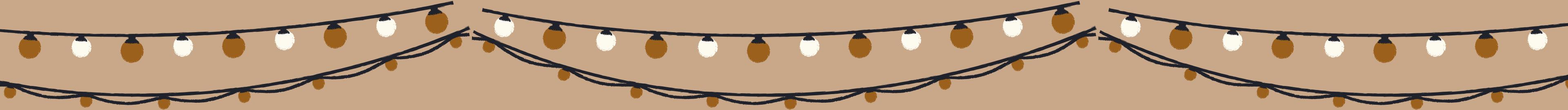


Tech-Stack Used



Insight





TASK



Movie Genre Analysis



Movie Duration Analysis



Language Analysis



Director Analysis



Budget Analysis

Movie Genre Analysis

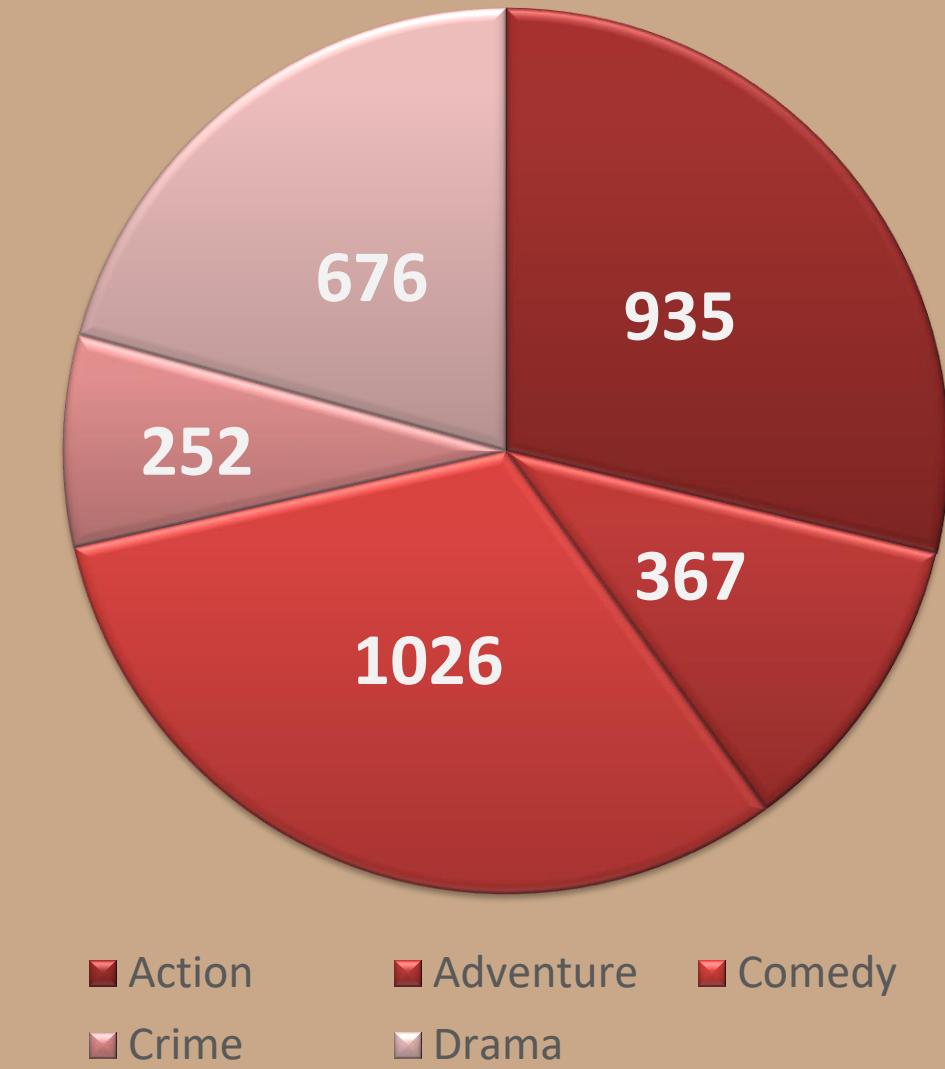
This analysis provides valuable insights for filmmakers and producers regarding audience preferences and industry trends

- **Genre Popularity:** Comedy is the most represented genre, while Family, Musical, Romance, Thriller, and Western are less common.
- **Rating Trends:** Documentaries tend to have the highest average ratings, while Thrillers have the lowest.
- **Score Variability:** Westerns have the most varied ratings, while Biographies are more consistent.
- **Quality and Consistency:** Biographies consistently score well, while Comedies have a more mixed reception.
- **Genre Appeal:** Action, Adventure, Drama, and Crime genres have broad appeal, while Family, Musical, Romance, and Thriller genres are more niche..

Analysis

Row Labels	Count of imdb_score	Average of imdb_score	Max of imdb_score	Min of imdb_score	Var of imdb_score	StdDev of imdb_score
Action	935	6.285989305	9	2.1	1.078186788	1.038357736
Adventure	367	6.561307902	8.6	2.3	1.264345826	1.124431334
Animation	46	6.763043478	8	4.5	0.945937198	0.972593028
cc	206	7.151941748	8.9	4.5	0.489337675	0.69952675
Comedy	1026	6.164424951	8.8	1.9	1.074567328	1.036613393
Crime	252	6.945238095	9.3	3.3	0.75475811	0.868768157
Documentary	43	6.951162791	8.5	1.6	2.005415282	1.41612686
Drama	676	6.821745562	8.8	2.1	0.82072643	0.905939529
Family	3	6.5	7.9	5.7	1.48	1.216552506
Fantasy	35	6.234285714	7.9	4.3	0.799378151	0.894079499
Horror	156	5.813461538	8.5	2.3	1.015624069	1.007781757
Musical	2	6.75	7.2	6.3	0.405	0.636396103
Mystery	23	6.586956522	8.5	3.3	1.23027668	1.109178381
Romance	2	6.65	7.1	6.2	0.405	0.636396103
Sci-Fi	8	6.5875	8.2	5	1.064107143	1.031555691
Thriller	3	5.3	6.3	4.8	0.75	0.866025404
Western	3	6.766666667	8.9	4.1	5.973333333	2.444040371
(blank)						
Grand Total	3786	6.462466984	9.3	1.6	1.118324058	1.05750842

Top 5 Genres



I found the most common genres in the dataset and studied how they affect IMDB scores. I discovered that Action, Biography, Crime, Comedy, and Drama are the most common genres.

Movie Duration Analysis

1. Movie Duration:

- **Mean Duration:** The average duration of movies is approximately 109.81 minutes.
- **Variability:** The duration varies widely, with a range from 34 to 330 minutes, and a standard deviation of 22.76, indicating significant variability in movie lengths.
- **Distribution:** The distribution of movie durations is positively skewed (skewness = 2.35), meaning that there are more shorter-duration movies than longer-duration ones. The kurtosis value of 12.41 indicates a very peaked distribution.

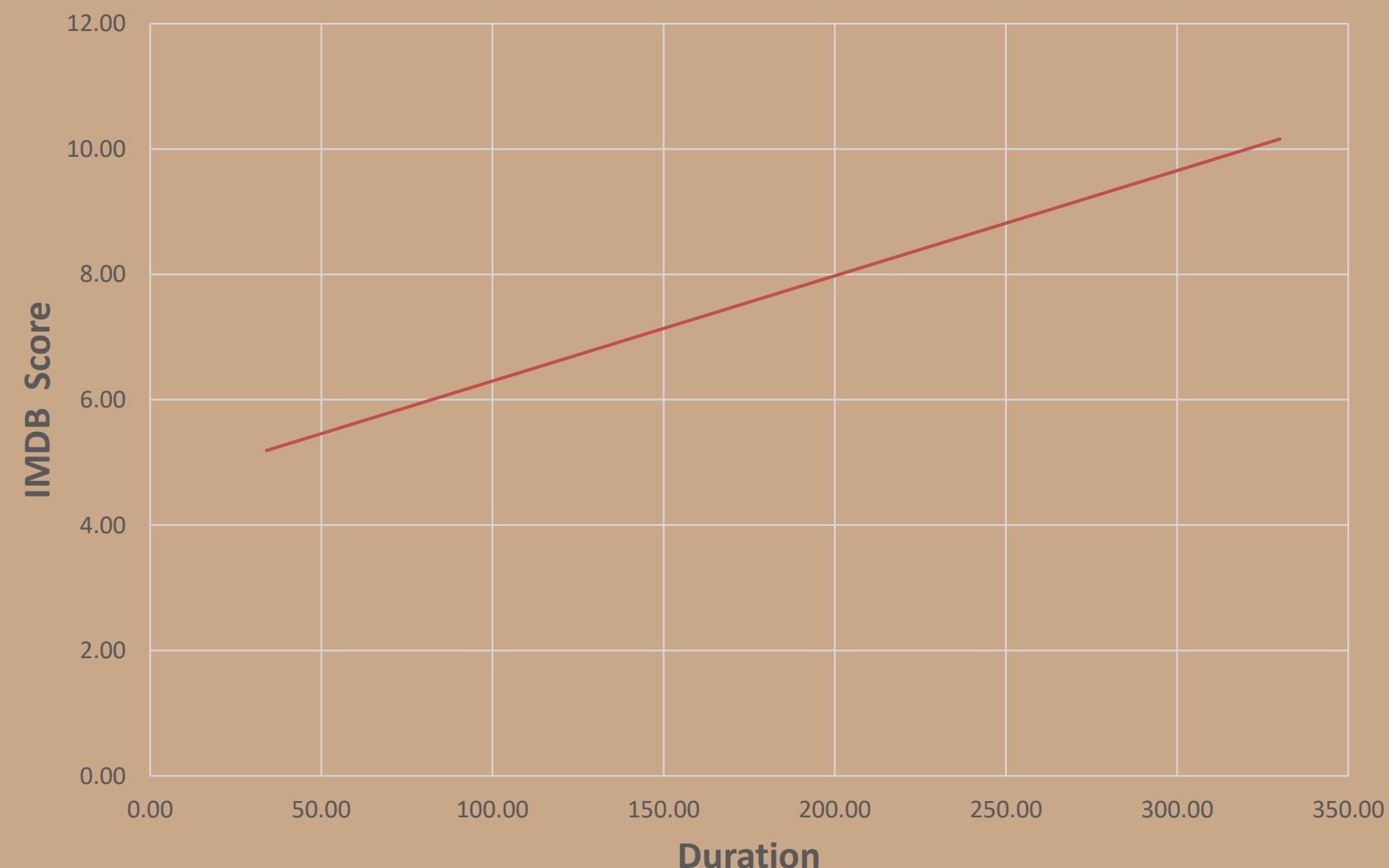
2. IMDb Scores:

- **Average IMDb Score:** The average IMDb score is 6.46 out of 10.
- **Variability:** The IMDb scores range from 1.6 to 9.3, with a standard deviation of 1.06, indicating some variability in the ratings.
- **Distribution:** The distribution of IMDb scores is negatively skewed (skewness = -0.73), meaning that there are more higher-rated movies than lower-rated ones. The kurtosis value of 1.13 indicates a moderately peaked distribution.

Analysis

DURATION		IMDB_SCORE	
Mean	109.808505	Mean	6.462466984
Standard Error	0.369949997	Standard Error	0.017186741
Median	105	Median	6.6
Mode	101	Mode	6.7
Standard Deviation	22.763201	Standard Deviation	1.05750842
Sample Variance	518.16332	Sample Variance	1.118324058
Kurtosis	12.40512587	Kurtosis	1.133330499
Skewness	2.347508256	Skewness	-0.725634682
Range	296	Range	7.7
Minimum	34	Minimum	1.6
Maximum	330	Maximum	9.3
Sum	415735	Sum	24466.9
Count	3786	Count	3786
Largest(1)	330	Largest(1)	9.3
Smallest(1)	34	Smallest(1)	1.6
Confidence Level(95.0%)	0.725320612	Confidence Level(95.0%)	0.033696168

Duration Vs IMDB score



Language Analysis

1. Language Impact on IMDb Scores:

- **Highest Average IMDb Scores:** Languages with the highest average IMDb scores include Portuguese (7.76), German (7.69), and Japanese (7.63), indicating generally well-received movies in these languages.
- **Lowest Average IMDb Scores:** Bosnian (4.3) and Kazakh (6.0) have the lowest average IMDb scores, suggesting lower reception or fewer movies in these languages.

2. Variability in IMDb Scores:

- **Most Consistent Scores:** Arabic, Aramaic, Dzongkha, and Zulu have only one movie each, but with identical IMDb scores, indicating consistent reception for movies in these languages.
- **Most Varied Scores:** Hindi movies have a wide range of IMDb scores, from 4.8 to 8.0, suggesting varied reception among viewers.

3. Impact of Sample Size:

- Languages with larger sample sizes, such as English (3606 movies), may have more stable average IMDb scores compared to languages with smaller sample sizes, like Icelandic (1 movie).

4. Overall Trends:

- The average IMDb score across all languages is 6.46, with a standard deviation of 1.06, indicating a moderate level of variability in movie ratings across languages.
- English movies, with the largest sample size, have an average IMDb score of 6.42, suggesting a relatively consistent reception for English-language movies.

Analysis



Row Labels	Count of Imdb_Score	Average of Imdb_Score	Max of Imdb_Score	Min of Imdb_Score	Varp of Imdb_Score	StdDevp of Imdb_Score
Aboriginal	2	6.95	7.5	6.4	0.3025	0.55
Arabic	1	7.2	7.2	7.2	0	0
Aramaic	1	7.1	7.1	7.1	0	0
Bosnian	1	4.3	4.3	4.3	0	0
Cantonese	8	7.2375	7.8	6.5	0.16984375	0.412121038
Czech	1	7.4	7.4	7.4	0	0
Danish	3	7.9	8.3	7.3	0.186666667	0.43204938
Dari	2	7.5	7.6	7.4	0.01	0.1
Dutch	3	7.566666667	7.8	7.1	0.108888889	0.329983165
Dzongkha	1	7.5	7.5	7.5	0	0
English	3606	6.421436495	9.3	1.6	1.107446744	1.052352956
Filipino	1	6.7	6.7	6.7	0	0
French	37	7.286486486	8.4	5.8	0.306574142	0.553691378
German	13	7.692307692	8.5	6.1	0.379171598	0.615769111
Hebrew	3	7.5	8	7.2	0.126666667	0.355902608
Hindi	10	6.76	8	4.8	1.1124	1.05470375
Hungarian	1	7.1	7.1	7.1	0	0
Icelandic	1	6.9	6.9	6.9	0	0
Indonesian	2	7.9	8.2	7.6	0.09	0.3
Italian	7	7.185714286	8.9	5.3	1.144081633	1.069617517
Japanese	12	7.625	8.7	6	0.741875	0.861321659
Kazakh	1	6	6	6	0	0
Korean	5	7.7	8.4	7	0.26	0.509901951
Mandarin	14	7.021428571	7.9	5.6	0.544540816	0.737930089
Maya	1	7.8	7.8	7.8	0	0
Mongolian	1	7.3	7.3	7.3	0	0
None	1	8.5	8.5	8.5	0	0
Norwegian	4	7.15	7.6	6.4	0.2475	0.497493719
Persian	3	8.133333333	8.5	7.5	0.202222222	0.449691252
Portuguese	5	7.76	8.7	6.1	0.7664	0.875442745
Romanian	1	7.9	7.9	7.9	0	0
Russian	1	6.5	6.5	6.5	0	0
Spanish	26	7.05	8.2	5.2	0.656346154	0.810151933
Swedish	1	7.6	7.6	7.6	0	0
Telugu	1	8.4	8.4	8.4	0	0
Thai	3	6.633333333	7.1	6.2	0.135555556	0.368178701
Vietnamese	1	7.4	7.4	7.4	0	0
Zulu	1	7.3	7.3	7.3	0	0
(blank)						
Grand Total	3786	6.462466984	9.3	1.6	1.118028674	1.05736875

Director Analysis

1. Highly Acclaimed Directors: Directors like Frank Darabont, Francis Ford Coppola, and Christopher Nolan have exceptionally high average IMDb scores of 9.0 or above. This indicates a consistent track record of directing highly acclaimed and well-received movies.

2. Consistency in High Ratings: Directors such as Peter Jackson, Quentin Tarantino, Sergio Leone, and Steven Spielberg also have high average IMDb scores, all around 8.9. This suggests a consistent ability to deliver movies that are highly appreciated by audiences and critics alike.

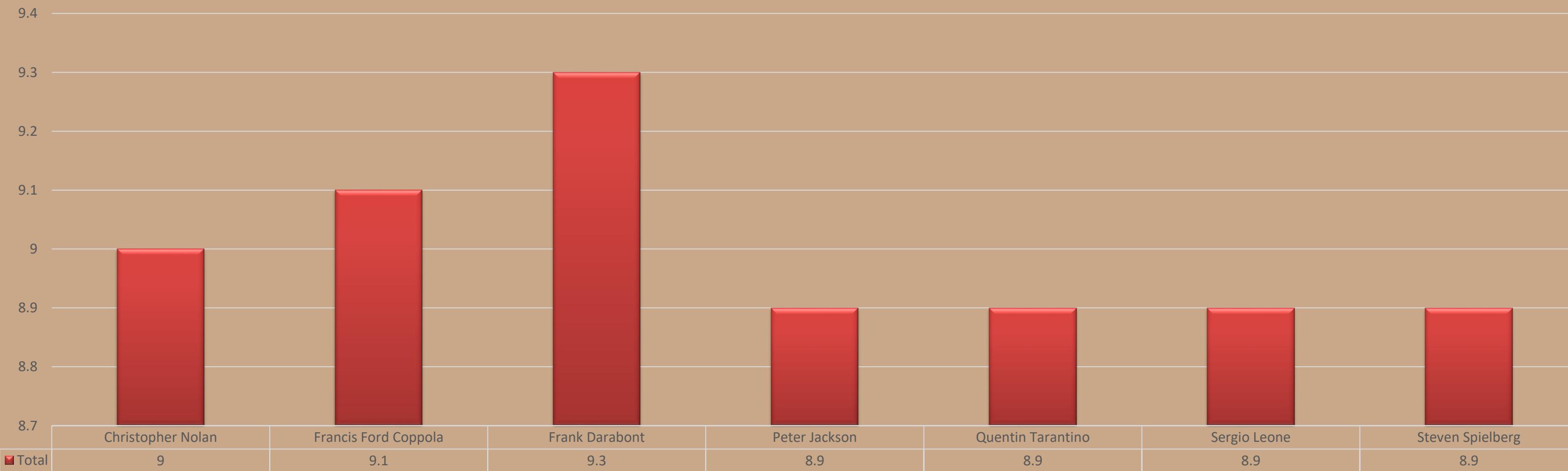
3. Overall Average IMDb Score: The overall average IMDb score across all directors is 9.0125, indicating that the movies directed by these directors are generally considered of high quality and are well-received.



Analysis

Row Labels	Average of imdb_score
Christopher Nolan	8.9
David Fincher	8.8
Francis Ford Coppola	9.1
Frank Darabont	9.3
Irvin Kershner	8.8
Peter Jackson	8.85
Quentin Tarantino	8.9
Robert Zemeckis	8.8
Sergio Leone	8.9
Steven Spielberg	8.9
Grand Total	8.930769231

Top 10 Director



Budget Analysis

I found out how movie budgets relate to financial success by studying the correlation between budgets and gross earnings. I found a positive correlation between the two variables, but the strength of this relationship depended on factors like genre and director.

If we keep the budget over 200,000,000, there's a high chance of earning a greater profit margin.



MAX	523505847
MIN	-12213298588

Analysis



CONCLUSION

This project helped me learn about what makes movies popular, successful and financial performance. By analyzing genres, durations, languages, directors, and budgets, I was able to identify trends and patterns that shed light on audience preferences and industry dynamics. My findings provide actionable insights for filmmakers, producers, and studios seeking to optimize their movie production strategies and maximize audience appeal.

Excel Drive link :

[https://docs.google.com/spreadsheets/d/1rr8cE81SRHX86jjkUBGrdPKZyv9wHLSz/edit?](https://docs.google.com/spreadsheets/d/1rr8cE81SRHX86jjkUBGrdPKZyv9wHLSz/edit?usp=drive_link&ouid=108396890359637253084&rtpof=true&sd=true)



THANK YOU

