

IMDB MOVIE ANALYSIS

Project Report

by
Shubhangi Chaudhary

Content

01 Project Description

02 Approach

03 Tech stack

04 Insights

05 Results

06 Thank you

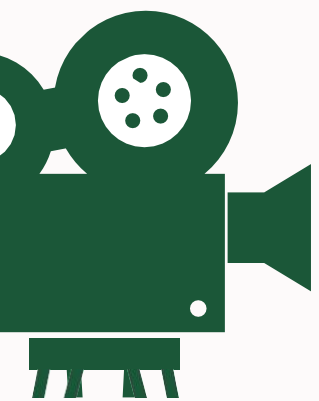




Project Description

The dataset related to IMDB Movies is provided. Several problems are to be investigated. The impact of these 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. It will involve Data Cleaning, Data Analysis, Five 'Whys' Approach and more.

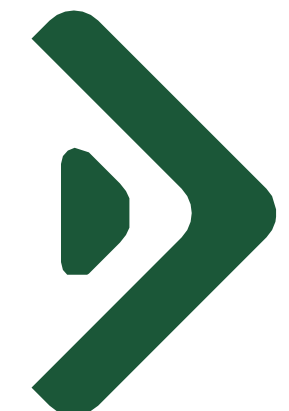
The aim is to use visualizations to help tell a story and make your findings more understandable. The goal is not just to answer questions but to provide insights that can drive decision-making. The analysis is aimed to provide actionable insights that can help stakeholders make informed decisions.





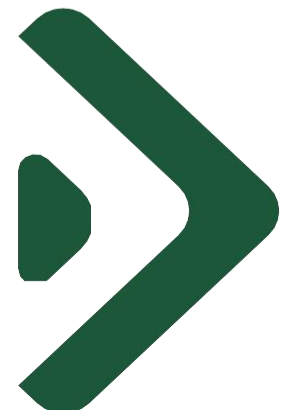
Approach

01



Data Cleaning: It involves preprocessing the data to make it suitable for analysis.

02



Data Analysis: Exploring the data to understand the relationships between different variables.

03



Inference: To better interpret the results from the collected data.





TECH STACK

The project required me to extensively work with MS-Excel allowing me to gain a better understanding related to its various features and how I could seek relevant insights with the same when dealing with a huge quantity of data.

Movie Genre Analysis

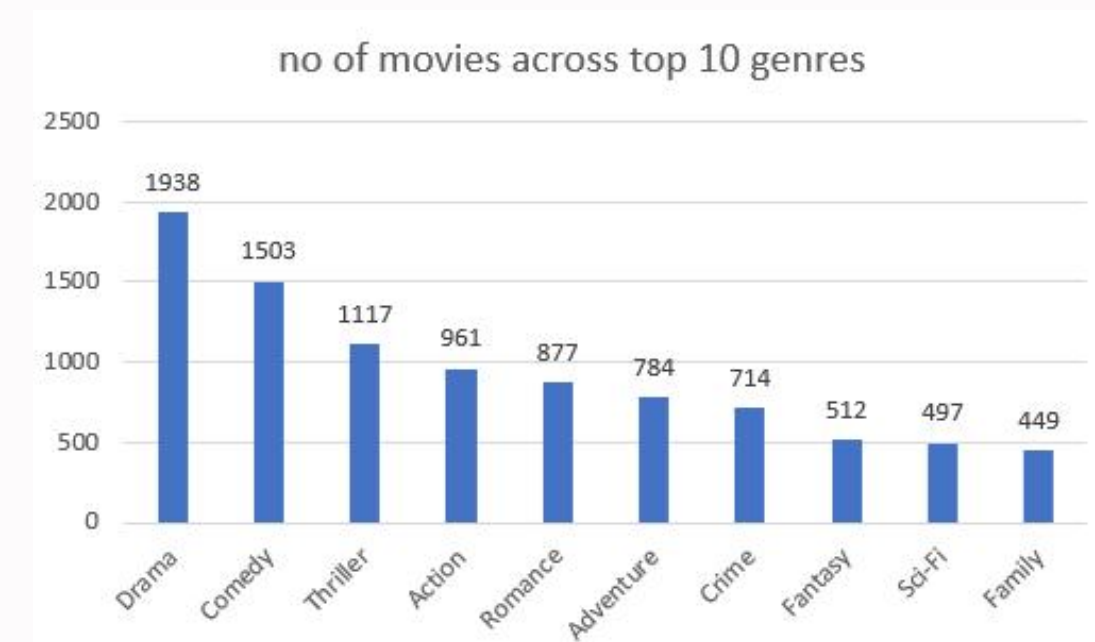
Insights

The aim was to analyze the distribution of movie genres and their impact on the IMDB score.

genres2	movies in the genre
Drama	1938
Comedy	1503
Thriller	1117
Action	961
Romance	877
Adventure	784
Crime	714
Fantasy	512
Sci-Fi	497
Family	449
Horror	391
Mystery	382
Biography	243
Animation	198
War	160
Music	158
History	152
Sport	151
Musical	103
Documentary	61
Western	58
Short	1
Film-Noir	1

The “**Drama**” genre had the highest number of movies with a decent Average IMDB score, the highest mean IMDB rating was for the genre ‘**Biography**’.

It can be so because audiences in general enjoy more of movies in the categories of drama followed closely by comedies and thrillers, while biographies tend to get better ratings since they often tend to be more acclaimed



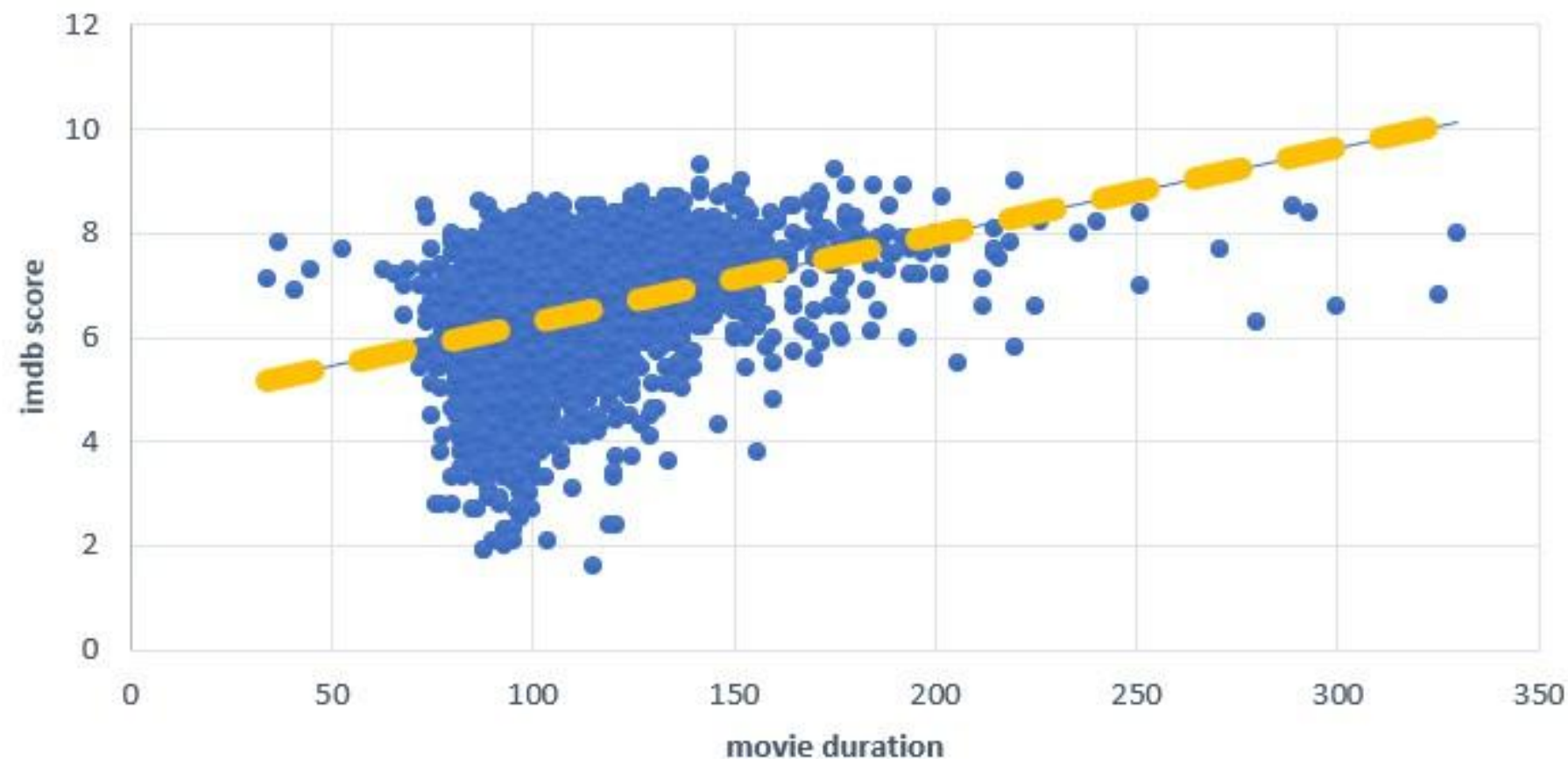
genres2	movies in the genre	mean	median	mode	var	stdv	min	max
Drama	1938	6.81	6.9	6.7	0.7951	0.8917	2.1	9.3
Comedy	1503	6.17	6.3	6.7	1.0774	1.0380	1.9	8.8
Thriller	1117	5.30	6.4	6.5	0.9335	0.9662	2.7	9
Action	961	6.29	6.3	6.1	1.0645	1.0317	2.1	9
Romance	877	6.65	6.5	6.5	0.9356	0.9673	2.1	8.5
Adventure	784	6.56	6.6	6.7	1.2323	1.1101	2.3	8.9
Crime	714	6.94	6.6	6.6	0.9607	0.9802	2.4	9.3
Fantasy	512	6.28	6.4	6.7	1.2768	1.1300	2.2	8.9
Sci-Fi	497	6.59	6.4	6.7	1.3366	1.1561	1.9	8.8
Family	449	6.50	6.3	6.7	1.3489	1.1614	1.9	8.6
Horror	391	5.84	6	5.9	0.9942	0.9971	2.3	8.6
Mystery	382	6.65	6.5	6.6	1.0144	1.0072	3.1	8.6
Biography	243	7.15	7.2	7	0.5027	0.7090	4.5	8.9
Animation	198	6.74	6.8	6.7	0.9761	0.9880	2.8	8.6

Movie Duration Analysis

Insights

The aim was to analyze the distribution of movie durations and its impact on the IMDB score..

determining relationship between duration and
imdb score



Analyzing the distribution of movie durations and its impact on the IMDB score, we notice a positive trend. It implies that a lot of movies have a runtime of about 110 minutes. Many of which gain good IMDB ratings between 6-10.

This is possibly because it allows the audience more time to indulge in the story, and the storyline can be better expressed at a stable pace adding value to the progress of the story. However, the trends also show that a runtime exceeding 200 minutes shows a decline in ratings.

descriptive analysis for movie duration	
mean	109.97
median	106
standard deviation	22.72680286

Language Analysis.

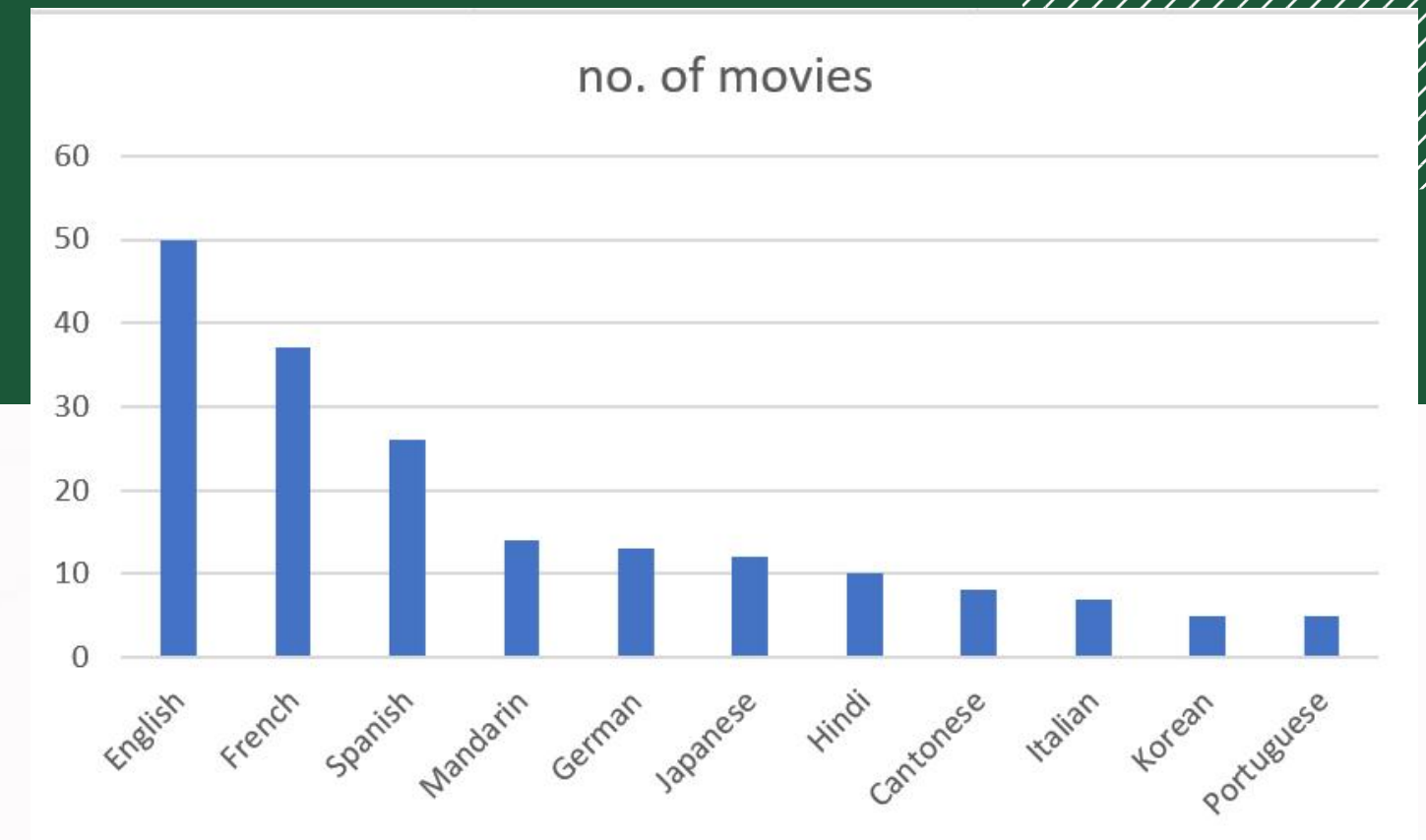
Insights

The aim was to examine the distribution of movies based on their language.

English as a language dominated the number of movies made in the language, followed by French.

The significant difference in the data could be attributed to the fact that the audience tends to better receive media in English due to the global use of the language and the ease of its understanding.

languages	Count of movie_title	avg	median	stdv
Cantonese	8	7.2	7.3	0.44058
English	3662	6.4	6.5	1.04899
French	37	7.3	7.2	0.56133
German	13	7.7	7.7	0.64091
Hindi	10	6.8	7.05	1.11176
Italian	7	7.2	7	1.15532
Japanese	12	7.6	7.8	0.89962
Korean	5	7.7	7.7	0.57009
Mandarin	14	7.0	7.25	0.76579
Portuguese	5	7.8	8	0.97877
Spanish	26	7.1	7.15	0.82620



language	no. of movies
English	3662
French	37
Spanish	26
Mandarin	14
German	13
Japanese	12
Hindi	10
Cantonese	8
Italian	7
Korean	5
Portuguese	5

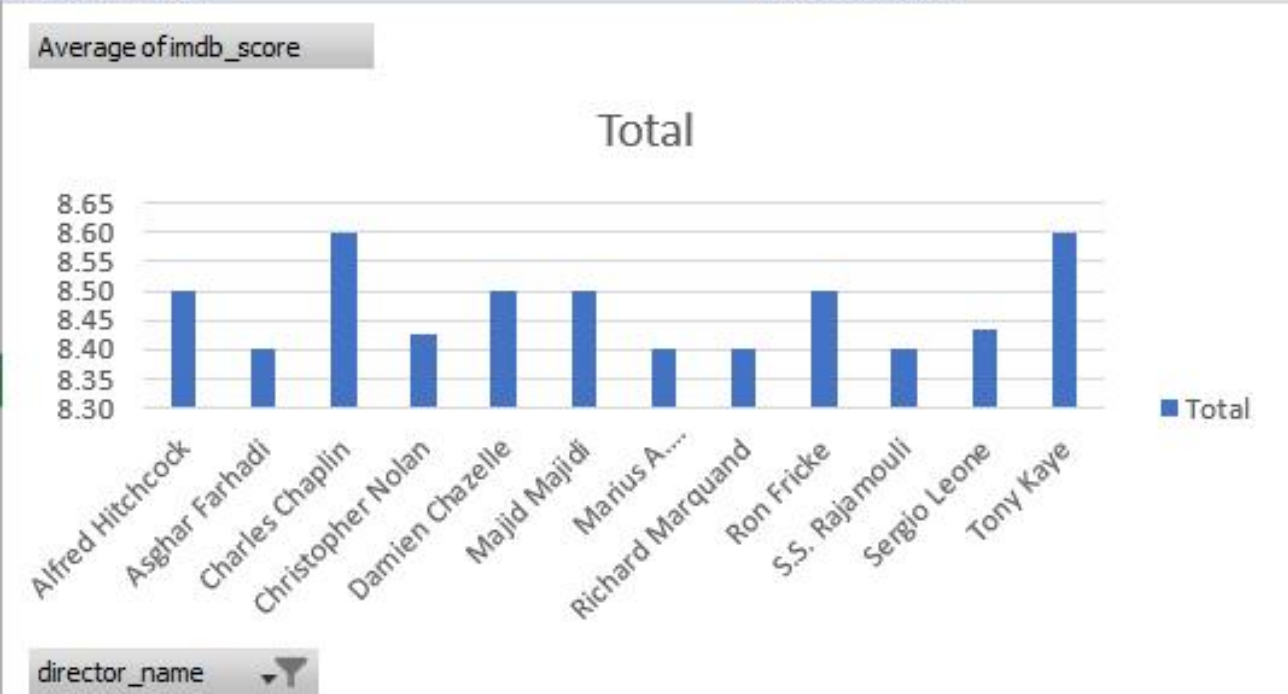
Director Analysis

Insights

The aim was to analyze the influence of directors on movie ratings..

Analyzing the average IMDB rating of various movie directors, we noticed that Charles Chaplin had the highest average IMDB rating. He tied with Tony Kaye, both having an average IMDB score of 8.60, with several other directors standing at 8.50.

Row Labels	Average of imdb_score	percentile
Alfred Hitchcock	8.50	8.59
Asghar Farhadi	8.40	8.59
Charles Chaplin	8.60	8.60
Christopher Nolan	8.43	8.51
Damien Chazelle	8.50	8.52
Majid Majidi	8.50	8.53
Marius A. Markevicius	8.40	8.54
Richard Marquand	8.40	8.55
Ron Fricke	8.50	8.56
S.S. Rajamouli	8.40	8.56
Sergio Leone	8.43	8.57
Tony Kaye	8.60	8.59
Grand Total	8.452380952	



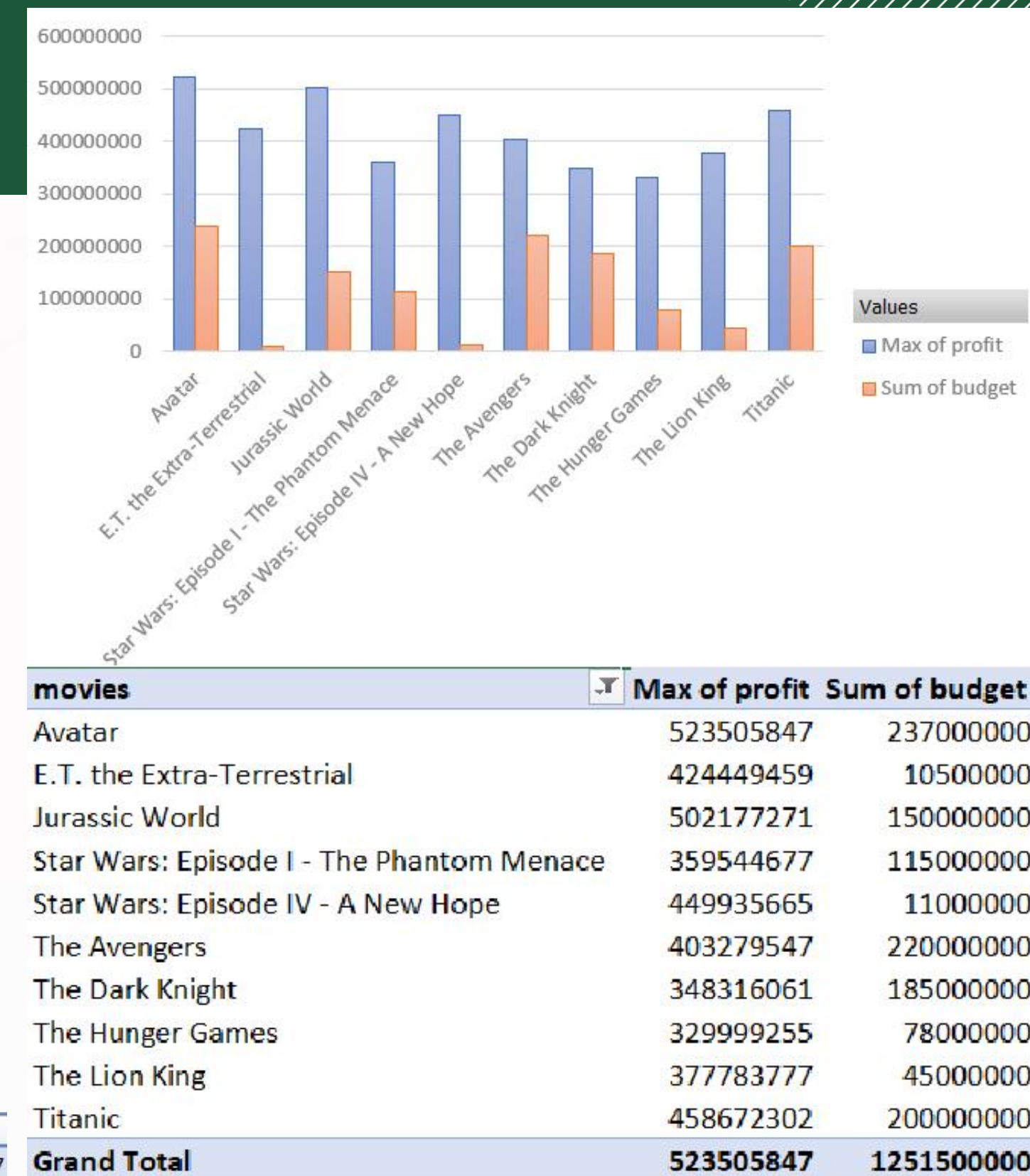
Budget Analysis

Insights

The aim was to analyze the relationship between movie budgets and their financial success.

Analyzing the data we see that “**Avatar**” was the movie that gained the most profits, followed by “**Jurassic World**” and “**Titanic**”.

Observing the data, it becomes apparent that movies with a high budget also earned great profits. This can be attributed to the fact that a bigger budget allows for better production and quality, which intrigues more people.



movie name1	language	budget	imdb_score	profit	correlation	max profit
Avatar	English	237000000	7.9	523505847	0.10082156	523505847

INSIGHTS SUMMARISED

- The “**Drama**” genre had the highest number of movies with a decent Average IMDB score, the highest mean IMDB rating was for the genre “**Biography**”.
- Analyzing the distribution of movie durations and its impact on the IMDB score, we notice a positive trend. It implies that a lot of movies have a runtime of about 110 minutes. Many of which gain good IMDB ratings between 6-10.
- English as a language dominated the number of movies made in the language, followed by French.
- Analyzing the average IMDB rating of various movie directors, we noticed that Charles Chaplin had the highest average IMDB rating. He tied with Tony Kaye, both having an average IMDB score of 8.60.
- Analyzing the data we see that “**Avatar**” was the movie that gained the most profits, followed by “**Jurassic World**” and “**Titanic**”.

Results

Through this project, I got to better understand the many ways in which we can gather meaningful insights from a large amount of data. MS Excel can indeed be quite a powerful tool that can be helpful in answering a number of questions extensively to better address various doubts and make informed decisions by observing trends over time.

Link to excel workbook:

<https://docs.google.com/spreadsheets/d/1B5JG4AkUyT16EzNLTfcL6OQqpVernNOZ/edit?usp=sharing&ouid=106859028556133454851&rtpof=true&sd=true>

[Link to project video](#)

THANK YOU

