

IMDB Movie Analysis



Agenda

- Project Description
- Approach
- Tech-Stack Used
- Insights
- Result



PROJECT DESCRIPTION

- The project is all about IMDB movie analysis, the success of a movie is defined by the rating of the IMDB rating so it is very crucial for the movie directors ,producers and entire team for the upcoming movie what kind of genre they should select and the actors for the upcoming movies .The IMDB rating will depend on the different aspects like genre, actors involved, release date , reviews etc .
- The team has given a dataset in excel sheet of different movies and their data like directors, gross, buget,imdb score, movie budget etc for it should be cleaned for analyzing.



TECH STACK USED

- **Software: Microsoft Excel**

A spreadsheet application by Microsoft used for data analysis, visualization, and automation with formulas, charts, and VBA.

- **Operating System: Windows**

A widely used OS by Microsoft, known for its user-friendly interface, multitasking, and software compatibility.



APPROACH

- To analyze any data, we should first clean the data like empty blanks, duplicates , null values ,design ,Outlier Detection etc.
- So I started cleaning the given data
- I followed the below steps for cleaning the data
 - 1)Firstly I have converted the raw data into table format.
 - 2)I analysed the given columns , there are some unwanted columns like



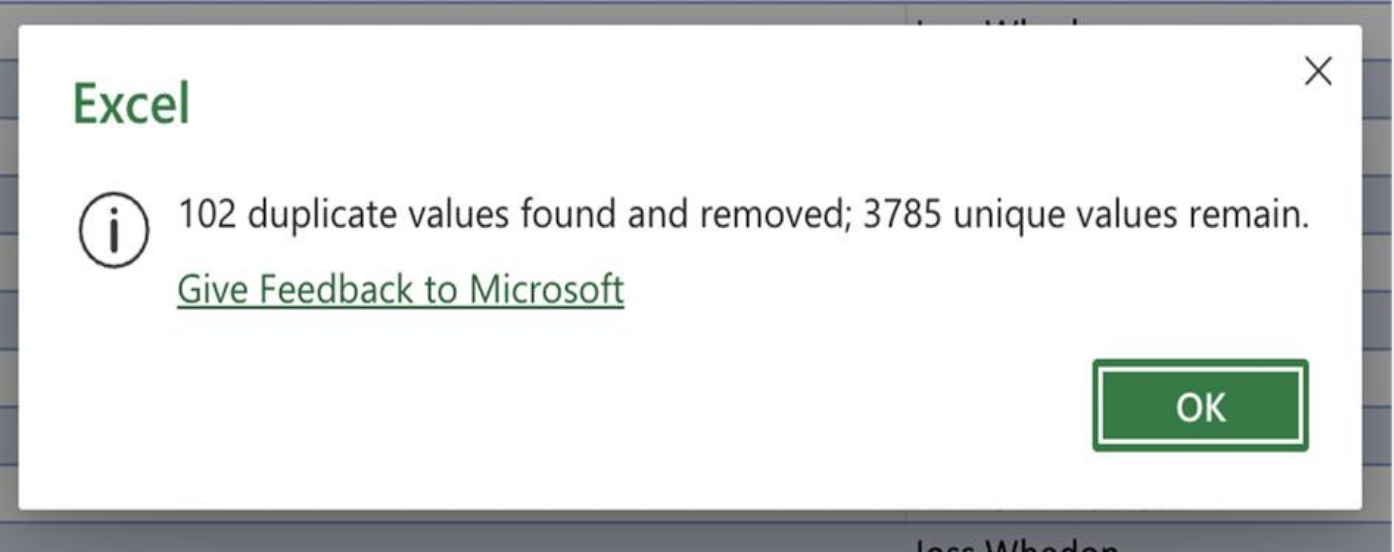
- | | |
|-----------------------------|--------------------------|
| • color | • facenumber_in_poster |
| • Num_critic_for_review | • plot_keywords |
| • Director_facebook_likes | • movie_imdb_link |
| • actor_2_name | • num_user_for_reviews |
| • actor_3_facebook_likes | • content_rating |
| • actor_1_facebook_likes | • title_year |
| • actor_1_name | • actor_2_facebook_likes |
| • Num_voted_users | • aspect_ratio |
| • cast_total_facebook_likes | • movie_facebook_likes |
| • actor_3_name | |



After deleting all the unnecessary columns the data looks like, and also I deleted rows which have blanks

movie_title	director_name	duration	gross	genres	language	budget	imdb_score
Avatar	James Cameron	178	760505847	Action Adventure Fantasy Sci-Fi	English	237000000	7.9
Pirates of the Caribbean: At World's End	Gore Verbinski	169	309404152	Action Adventure Fantasy	English	300000000	7.1
Spectre	Sam Mendes	148	200074175	Action Adventure Thriller	English	245000000	6.8
The Dark Knight Rises	Christopher Nolan	164	448130642	Action Thriller	English	250000000	8.5
John Carter	Andrew Stanton	132	73058679	Action Adventure Sci-Fi	English	263700000	6.6
Spider-Man 3	Sam Raimi	156	336530303	Action Adventure Romance	English	258000000	6.2
Tangled	Nathan Greno	100	200807262	Adventure Animation Comedy Family Fantasy Musical Romance	English	260000000	7.8
Avengers: Age of Ultron	Joss Whedon	141	458991599	Action Adventure Sci-Fi	English	250000000	7.5
Harry Potter and the Half-Blood Prince	David Yates	153	301956980	Adventure Family Fantasy Mystery	English	250000000	7.5
Batman v Superman: Dawn of Justice	Zack Snyder	183	330249062	Action Adventure Sci-Fi	English	250000000	6.9

3.) I checked for duplicates



4) Data Analytics Tasks

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.

Ans: Here the column genre have mixed type of genre to analyze each genre so I should split into individual genres so I went to the data in that text to columns → delimited to make individual

genres
Action Adventure Fantasy Sci-Fi
Action Adventure Fantasy
Action Adventure Thriller
Action Thriller
Action Adventure Sci-Fi
Action Adventure Romance
Adventure Animation Comedy Family Fantasy Musical Romance
Action Adventure Sci-Fi
Adventure Family Fantasy Mystery
Action Adventure Sci-Fi
Action Adventure Sci-Fi
Action Adventure
Action Adventure Fantasy
Action Adventure Western
Action Adventure Fantasy Sci-Fi

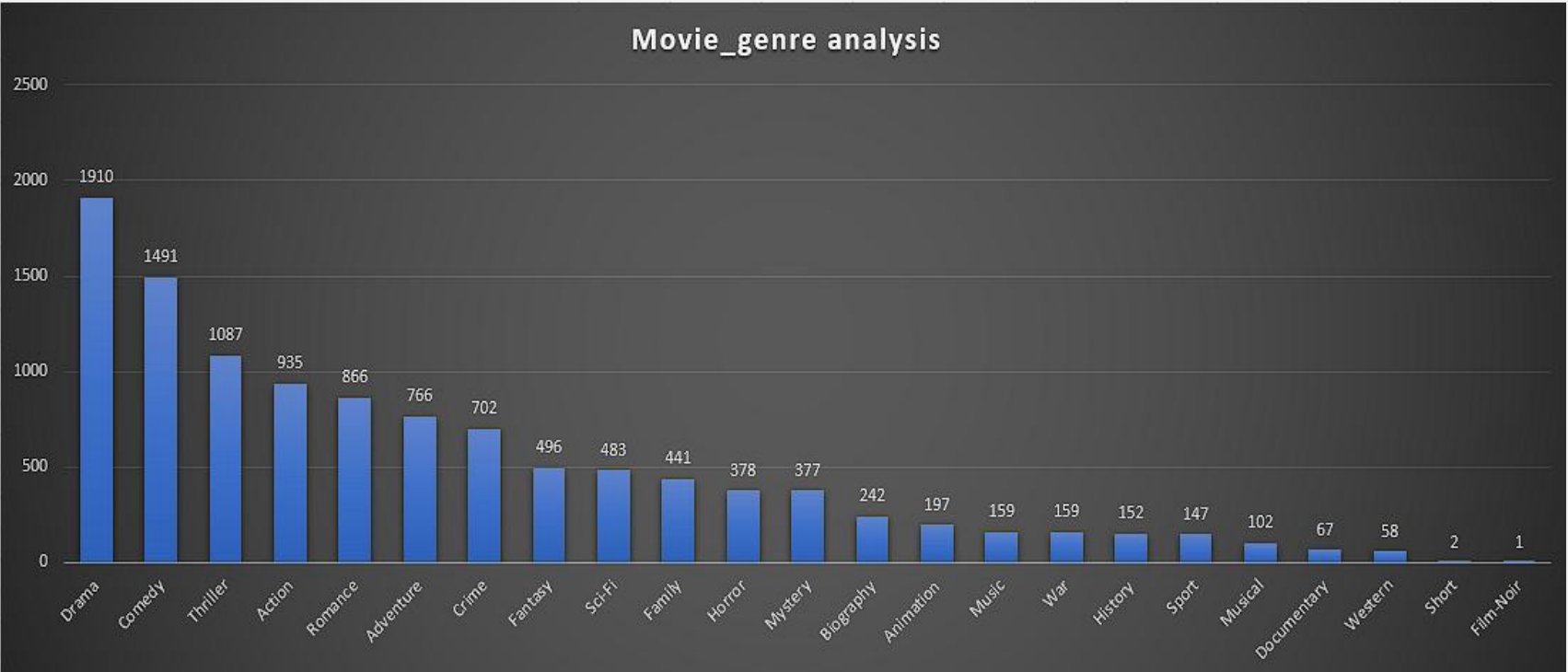


Genres
Action
Action
Action
Action
Action
Action
Adventure
Action
Adventure
Action
Action
Action
Action
Action
Action
Action



Formulas used:	
count	=COUNTIF(B:B,[@[Unq_Genres]])
mean	=AVERAGEIF(Table1[genres], "*" & [@[Unq_Genres]] & "*", Table1[imdb_score])
median	=MEDIAN(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score]))
mode	=IFNA(MODE(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score])), "")
range	=MAX(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score]))-MIN(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score]))
variance	=IFERROR(VAR(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score])), "")
standard deviation	=IFERROR(STDEV(IF(ISNUMBER(SEARCH("'" & [@[Unq_Genres]] & "'", Table1[genres])), Table1[imdb_score])), "")

Unq_Genres	count	Mean	Median	Mode	Range	Variance	Std.Dev
Drama	1910	6.78901	6.9	6.7	7.2	0.794389267	0.8912852
Comedy	1491	6.18276	6.3	6.3	6.9	1.081709406	1.0400526
Thriller	1087	6.37231	6.4	6.5	6.3	0.939112803	0.9690783
Action	935	6.28599	6.3	6.6	6.9	1.078186788	1.0383577
Romance	866	6.42621	6.5	6.5	6.4	0.938953731	0.9689962
Adventure	766	6.45496	6.6	6.6	6.6	1.247524378	1.1169263
Crime	702	6.54815	6.6	6.6	6.9	0.968463042	0.9841052
Fantasy	496	6.28508	6.4	6.7	6.7	1.30054464	1.1404142
Sci-Fi	483	6.32588	6.4	7	6.9	1.364204359	1.1679916
Family	441	6.2	6.3	5.4	6.7	1.367909091	1.1695765
Horror	378	5.90106	5.9	6.2	6.3	0.981537339	0.9907257
Mystery	377	6.4695	6.5	6.6	5.5	1.014838309	1.0073918
Biography	242	7.14008	7.2	7	4.4	0.504237338	0.7100967
Animation	197	6.70051	6.8	7.3	5.8	0.987295659	0.9936275
Music	159	6.45668	6.7	6.2	6.9	1.413359666	1.188848
War	159	7.04843	7.1	7.1	4.3	0.652386753	0.8077046
History	152	7.13158	7.2	7.7	3.4	0.451578947	0.6719962
Sport	147	6.60136	6.8	7.2	6.4	1.09876526	1.04822
Musical	102	6.55098	6.7	7.1	6.4	1.307672297	1.143535
Documentary	67	7.01194	7.2	6.6	6.9	1.439855269	1.1999397
Western	58	6.76552	6.8	6.8	4.8	0.997035693	0.9985167
Short	2	6.8	6.8		0.6	0.18	0.4242641
Film-Noir	1	7.7	7.7		0		
Total	11218						



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

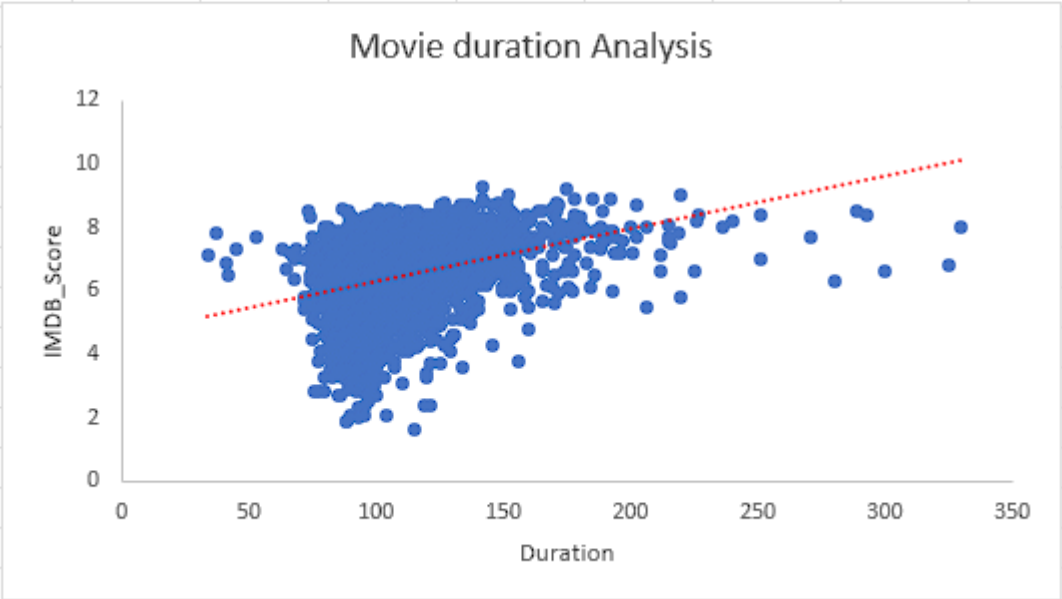
- Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

ANS:

Movie_Title	Duration	IMDB_Score
Avatar	178	7.9
Pirates of the Caribbean: At World's End	169	7.1
Spectre	148	6.8
The Dark Knight Rises	164	8.5
John Carter	132	6.6
Spider-Man 3	156	6.2
Tangled	100	7.8
Avengers: Age of Ultron	141	7.5
Harry Potter and the Half-Blood Prince	153	7.5
Batman v Superman: Dawn of Justice	183	6.9
Superman Returns	169	6.1
Quantum of Solace	106	6.7
Pirates of the Caribbean: Dead Man's Chest	151	7.3
The Lone Ranger	150	6.5
Man of Steel	143	7.2
The Chronicles of Narnia: Prince Caspian	150	6.6
The Avengers	173	8.1
Pirates of the Caribbean: On Stranger Tides	136	6.7
Men in Black 3	106	6.8
The Hobbit: The Battle of the Five Armies	164	7.5
The Amazing Spider-Man	153	7
Robin Hood	156	6.7
The Hobbit: The Desolation of Smaug	186	7.9

2)

Mean	109.8085
Median	105
Std.Dev	22.76621



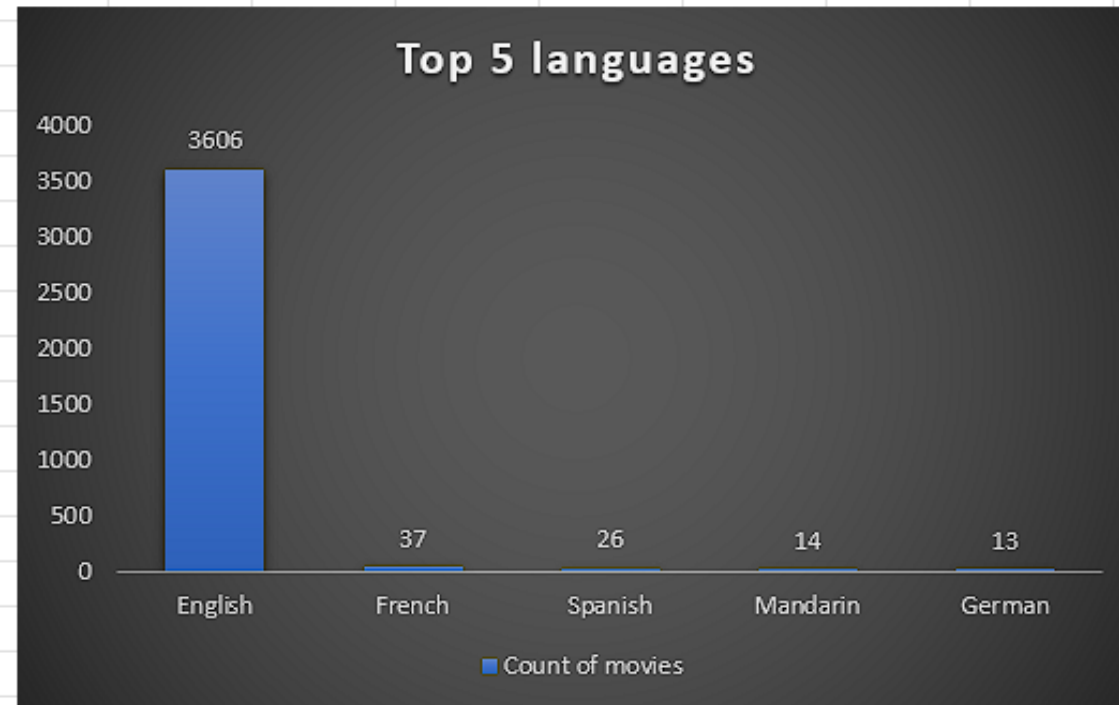
C. Language Analysis: Situation: Examine the distribution of movies based on their language.

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

	Language ▾	count ↕	mean ▾	median ▾	Std.dev ▾
3)	English	3606	6.45	6.5	1.0524989
	French	37	6.35	7.2	0.5613289
	Spanish	26	6.53	7.15	0.8261961
	Mandarin	14	6.24	7.25	0.7657862
	German	13	6.9	7.7	0.6409128
	Japanese	12	6.95	7.8	0.8996211
	Hindi	10	7.08	7.05	1.1117554
	Cantonese	8	6.34	7.3	0.4405759
	Italian	7	6.79	7	1.155319
	Portuguese	5	6.04	8	0.9787747
	Norwegian	4	7.03	7.3	0.5744563
	Korean	4	6.63	7.9	0.4787136
	Thai	3	6.3	6.6	0.450925
	Hebrew	3	6.1	7.3	0.4358899
	Dutch	3	7.23	7.8	0.4041452
	Danish	3	6.27	8.1	0.5291503
	Persian	3	7.53	8.4	0.5507571
	Aboriginal	2	3.95	6.95	0.7778175
	Dari	2	7.25	7.5	0.1414214
	Indonesian	2	7.9	7.9	0.4242641
	Bosnian	1	6.7	4.3	
	Kazakh	1	7	6	
	Russian	1	7.3	6.5	
	Filipino	1	8.1	6.7	
	Icelandic	1	7.3	6.9	

Top 5 languages

English	3606
French	37
Spanish	26
Mandarin	14
German	13



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

ANS:

PivotTable Fields

Choose fields to add to report:

Search

- ☒ director_name
- ☐ duration
- ☐ gross
- ☐ genres
- ☐ language
- ☐ budget
- ☒ imdb_score

More Tables...

Drag fields between areas below:

Filters	Columns
Rows	Values
director_name	Average of imdb_score

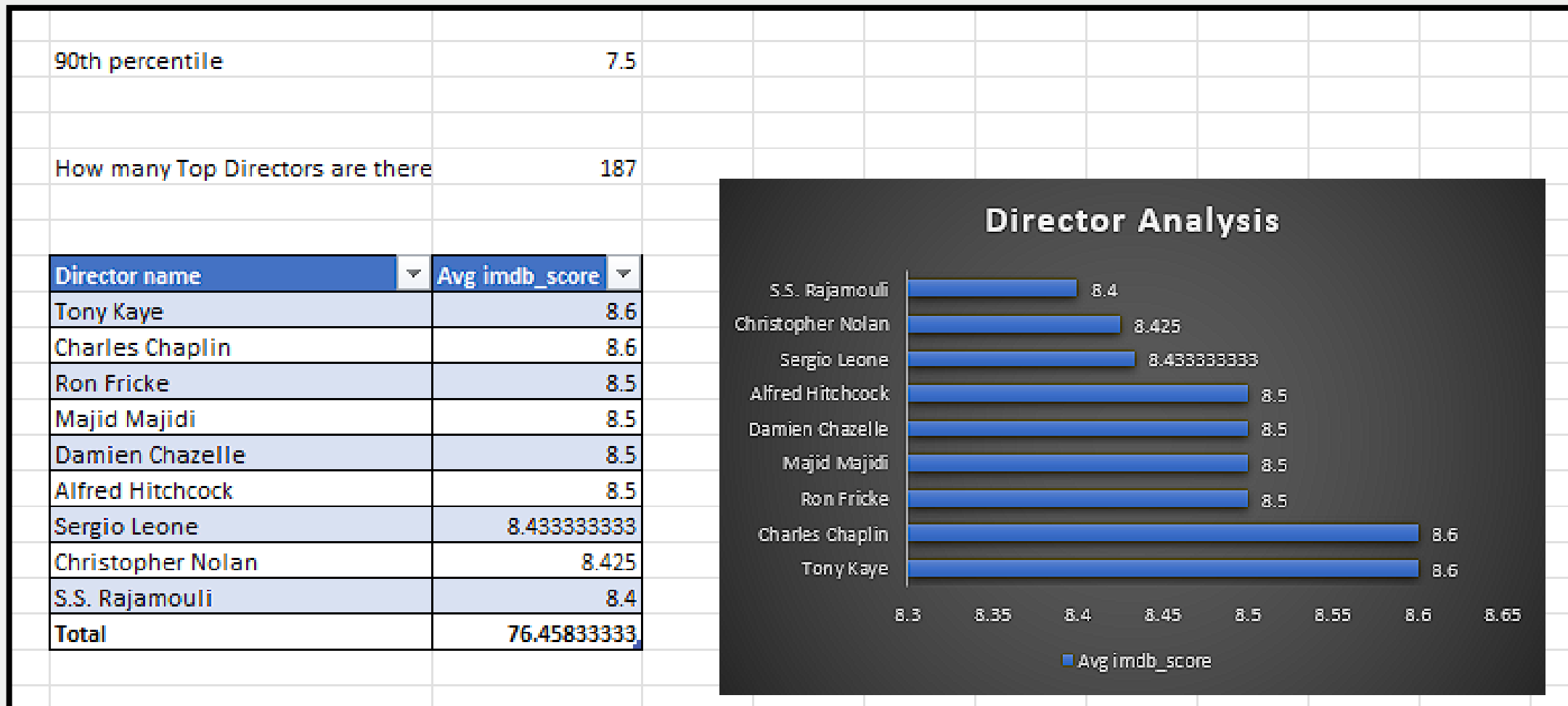
Director Name	Average of imdb_score
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
S.S. Rajamouli	8.4
Richard Marquand	8.4
Marius A. Markevicius	8.4
Asghar Farhadi	8.4
Lee Unkrich	8.3
Lenny Abrahamson	8.3
Fritz Lang	8.3
Billy Wilder	8.3
Pete Docter	8.233333333
Hayao Miyazaki	8.225
Quentin Tarantino	8.2
Juan Jos�� Campanella	8.2
Joshua Oppenheimer	8.2
George Roy Hill	8.2
Elia Kazan	8.2



FORMULAS USED:

=PERCENTILE(C4:C1754,0.9)

=COUNTIF(C4:C1754,">=7.5")



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

Movie_title	Budget	Gross	Profit Margin
Avatar	237000000	760505847	₹ 52,35,05,847.00
Jurassic World	150000000	652177271	₹ 50,21,77,271.00
Titanic	200000000	658672302	₹ 45,86,72,302.00
Star Wars: Episode IV - A New Hope	11000000	460935665	₹ 44,99,35,665.00
E.T. the Extra-Terrestrial	10500000	434949459	₹ 42,44,49,459.00
The Avengers	220000000	623279547	₹ 40,32,79,547.00
The Lion King	45000000	422783777	₹ 37,77,83,777.00
Star Wars: Episode I - The Phantom Menace	115000000	474544677	₹ 35,95,44,677.00
The Dark Knight	185000000	533316061	₹ 34,83,16,061.00
The Hunger Games	78000000	407999255	₹ 32,99,99,255.00
Deadpool	58000000	363024263	₹ 30,50,24,263.00
The Hunger Games: Catching Fire	130000000	424645577	₹ 29,46,45,577.00
Jurassic Park	63000000	356784000	₹ 29,37,84,000.00
Despicable Me 2	76000000	368049635	₹ 29,20,49,635.00
American Sniper	58800000	350123553	₹ 29,13,23,553.00
Finding Nemo	94000000	380838870	₹ 28,68,38,870.00
Shrek 2	150000000	436471036	₹ 28,64,71,036.00
The Lord of the Rings: The Return of the King	94000000	377019252	₹ 28,30,19,252.00
Star Wars: Episode VI - Return of the Jedi	32500000	309125409	₹ 27,66,25,409.00

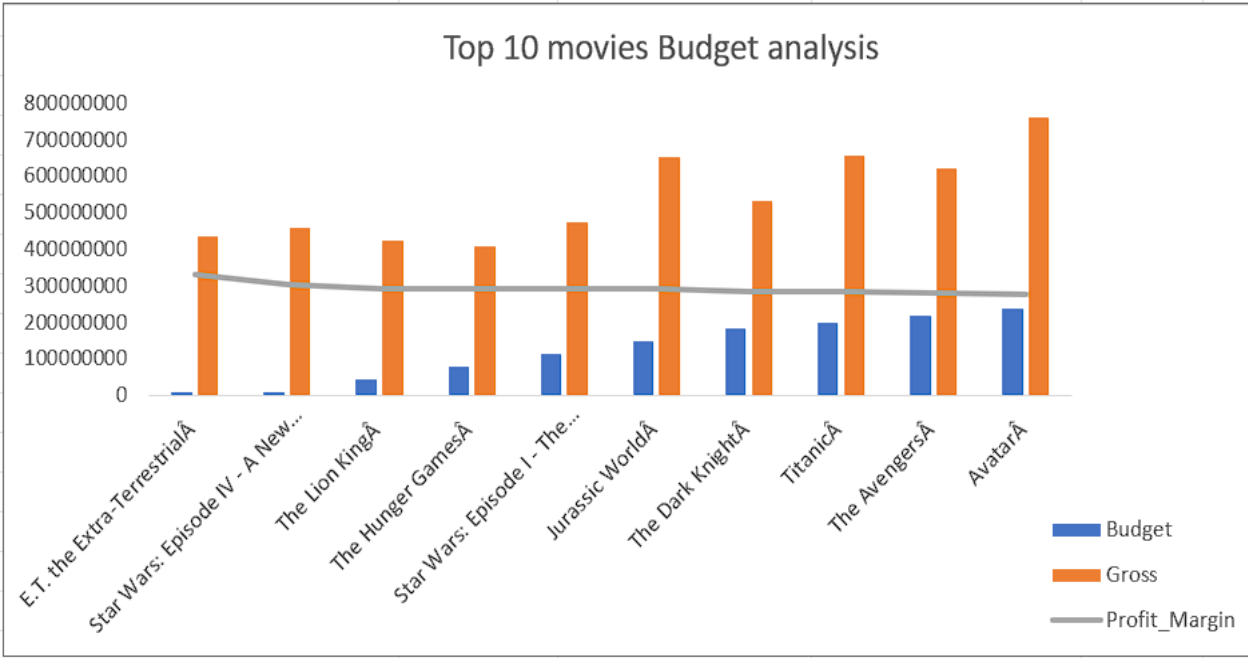


Formulas Used:

=MAX(Table13[Profit Margin])

=CORREL(Table13[Budget],Table13[Gross])

The movie with highest profit margin	Avatar	₹ 52,35,05,847.00	
Correlation coefficient between Gross & Budget	0.223115594		
Movie_Title	Budget	Gross	Profit_Margin
E.T. the Extra-Terrestrial	10500000	434949459	₹ 32,99,99,255.00
Star Wars: Episode IV - A New Hope	11000000	460935665	₹ 30,50,24,263.00
The Lion King	45000000	422783777	₹ 29,46,45,577.00
The Hunger Games	78000000	407999255	₹ 29,37,84,000.00
Star Wars: Episode I - The Phantom Menace	115000000	474544677	₹ 29,20,49,635.00
Jurassic World	150000000	652177271	₹ 29,13,23,553.00
The Dark Knight	185000000	533316061	₹ 28,68,38,870.00
Titanic	200000000	658672302	₹ 28,64,71,036.00
The Avengers	220000000	623279547	₹ 28,30,19,252.00
Avatar	237000000	760505847	₹ 27,66,25,409.00



INSIGHTS

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

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

C. Language Analysis: Examined the distribution of movies based on their language and analyzed their impact on the IMDB score.

D. Director Analysis: Identified the top directors based on their average IMDB score and analyzed their contribution to the success of movies.

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



RESULT

I analyzed movie data, including variables like title, director name, language, IMDb score, budget, and gross earnings. The analysis revealed key factors impacting IMDb scores and profitability margins, such as the director's reputation, language, budget, and the balance between production costs and revenue. These insights can help guide future movie productions and marketing strategies for better success.



REFERENCE LINK:

The link for excel sheet file



[IMDB Excel Sheet](#)



