IMDB Movie Analysis

Today's agenda

- Project recap
- Data Analytics Tasks:
- Process
- Insights
- Summary

Project Description

In the scope of this project, we are tasked with conducting a comprehensive analysis of IMDb's Movies dataset spanning the years 1927 to 2016. Our objective is to extract valuable insights from this dataset through a series of structured tasks and analyses.

In addition to the analytical tasks, we will leverage various visualization techniques to effectively communicate our findings in a clear and easily understandable manner.

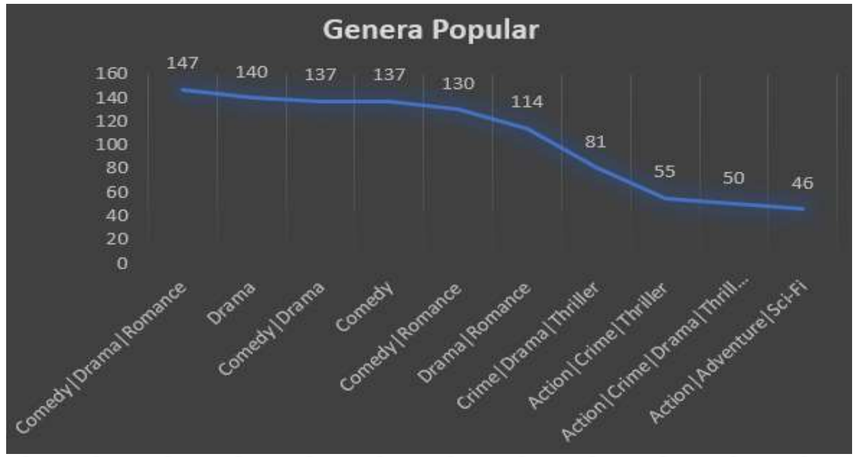
Data Analytics Tasks:

1. Movie Genre Analysis
2. Movie Duration Analysis
3. Language Analysis
4. Director Analysis
5. Budget Analysis

➢ Movie Genre Analysis:

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.

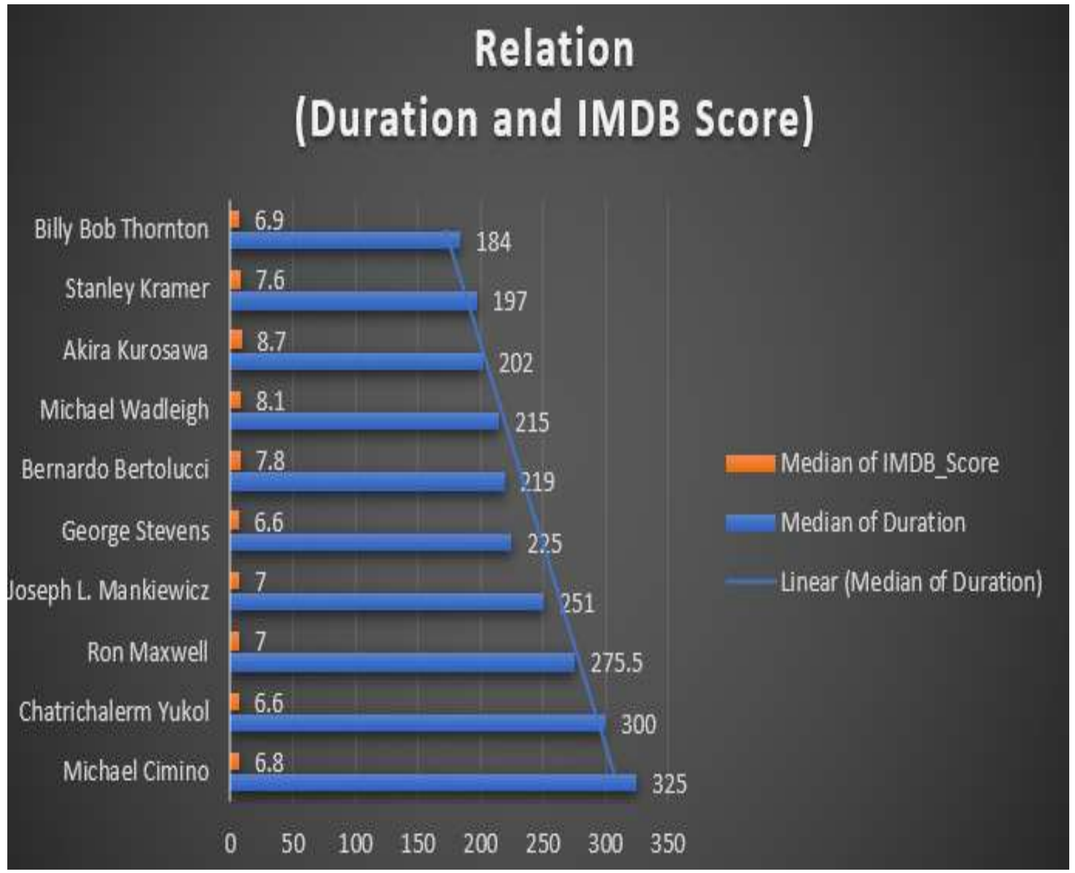
1	Genera	Count	Sum	MEDIAN	Max	Min	Mode	Var	Standered Deviation	Average
2	Comedy|Drama|Romance	147	953.5	6.5	8	4.3	6.5	0.57	0.76	6.486394558
3	Drama	141	998.9	7.2	8.8	3.4	7.3	0.69	0.83	7.084397163
4	Comedy|Drama	138	906.5	6.7	8.8	3.3	6.7	0.76	0.87	6.56884058
5	Comedy	138	807.2	6	8	1.9	6.2	1.50	1.23	5.849275362
6	Comedy|Romance	131	777.9	6	8.4	2.7	6.1	0.70	0.84	5.938167939
7	Drama|Romance	115	802	7.1	8.1	4.1	7.2	0.54	0.74	6.973913043
8	Crime|Drama|Thriller	82	562.5	7	8.5	5.1	6.1	0.61	0.78	6.859756098
9	Action|Crime|Thriller	56	359.2	6.5	7.6	4.4	6.5	0.39	0.63	6.414285714
10	Action|Crime|Drama|Thriller	50	324.9	6.5	9	5.1	6.5	0.51	0.72	6.498
11	Action|Adventure|Sci-Fi	48	319.3	6.8	8.4	2.4	6.6	1.54	1.24	6.652083333



➢ Movie Duration Analysis:

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

S	T	U	V
Unique Directors	Sum Of Duration	Median	Standeredv Deviation
James Cameron	1098	154	28.25
Colin Trevorrow	210	105	26.87
George Lucas	655	136	12.49
Steven Spielberg	3429	135	18.90
Joss Whedon	433	141	27.15
Roger Allers	156	78	7.07
Christopher Nolan	1122	139	20.98
Gary Ross	421	140	1.53
Tim Miller	108	108	#DIV/0!
Francis Lawrence	630	123	16.20
Pierre Coffin	185	92.5	7.78
Clint Eastwood	2503	134	10.43
Andrew Stanton	330	100	19.08
Andrew Adamson	483	121.5	33.80
Peter Jackson	1866	172	31.55
Richard Marquand	134	134	#DIV/0!
Robert Zemeckis	1583	118	17.62
Irvin Kershner	248	124	4.24
Chris Columbus	1398	124	22.60
Sam Raimi	1417	125.5	32.73
Kyle Balda	91	91	#DIV/0!
M. Night Shyamalan	831	106	6.03
Chris Buck	102	102	#DIV/0!
Yarrow Cheney	87	87	#DIV/0!
Chris Weitz	529	101	16.39
Todd Phillips	696	101	7.48
Joel Zwick	188	94	1.41
David Slade	341	113	10.02



➢ Language Analysis:

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

Language	Count	Medean	Standard Deviation
English	3565	6.5	1.05
French	34	7.3	0.52
Spanish	23	7.2	0.86
Mandarin	14	7.25	0.77
German	10	7.8	0.71
Japanese	10	8	0.99
Cantonese	7	7.3	0.35
Italian	7	7	1.16
Portuguese	5	8	0.98
Korean	5	7.7	0.57
Hindi	5	7.4	0.80
Norwegian	4	7.3	0.57
Persian	3	8.4	0.55
Danish	3	8.1	0.53
Dutch	3	7.8	0.40
Thai	3	6.6	0.45
Indonesian	2	7.9	0.42
Dari	2	7.5	0.14

Language	Column1
Persian	8.13
Hebrew	8.00
Romanian	7.90
Indonesian	7.90
Danish	7.90

(Languages with **Highest** Average IMDB Score.)

Language	Mean
Bosnian	4.30
Kazakh	6.00
English	6.43
Russian	6.50
Thai	6.63

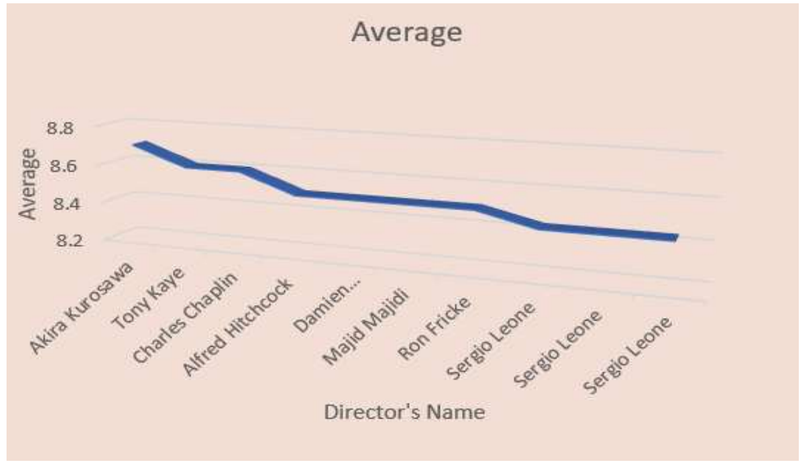
(Languages with **least** Average IMDB Score.)

➢ Director Analysis:

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

Unique Directors	Average	Percent
Akira Kurosawa	8.7	Higher
Tony Kaye	8.6	Higher
Charles Chaplin	8.6	Higher
Alfred Hitchcock	8.5	Higher
Damien Chazelle	8.5	Higher
Majid Majidi	8.5	Higher
Ron Fricke	8.5	Higher
Sergio Leone	8.4333333	Higher
Sergio Leone	8.4333333	Higher
Sergio Leone	8.4333333	Higher

Percentile
7.5



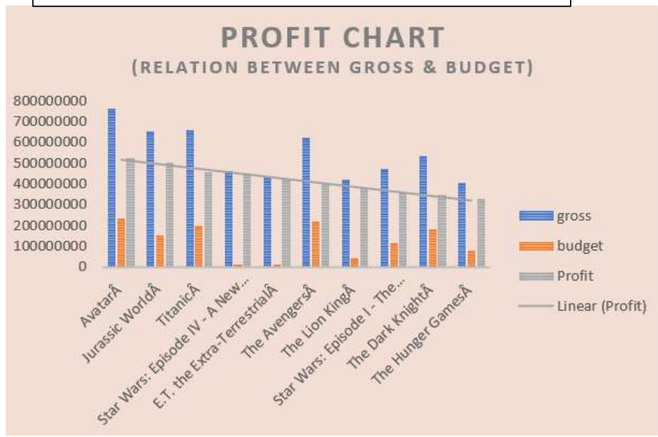
(Top 10 Director’s with Highest Average IMDB Score.)

➢ Budget Analysis:

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

movie_title	gross	budget	Profit
AvatarÂ	760505847	237000000	523505847
Jurassic WorldÂ	652177271	150000000	502177271
TitanicÂ	658672302	200000000	458672302
Star Wars: Episode IV - A New	460935665	11000000	449935665
E.T. the Extra-TerrestrialÂ	434949459	10500000	424449459
The AvengersÂ	623279547	220000000	403279547
The Lion KingÂ	422783777	45000000	377783777
Star Wars: Episode I - The Pha	474544677	115000000	359544677
The Dark KnightÂ	533316061	185000000	348316061
The Hunger GamesÂ	407999255	78000000	329999255

Co-relation between Gross and Budget
0.098306764



(Top 10 Most Profited Movies.)

➢ Insights

- After conducting our analysis, we have determined that the movie **'Avatar'** stands out as the most profitable film in the period spanning 1927 to 2016, with an impressive profit of approximately 523,505,847 dollars. It is followed by **'Jurassic World'** in the second position and **'Titanic'** in the third position in terms of profitability.

- Upon thorough data analysis, it has been determined that the most popular movie genre is **'Comedy'** specifically **'Comedy | Drama | Romance,'** which boasts a popularity score of 147. This is followed closely by the genre **‘Crime | Drama | Thriller’**.

Data Sheet Link: -

https://docs.google.com/spreadsheets/d/1llp7oGFqo8BwJUTCvJ7cQ6AwV3e3RV3/edit?usp=sharing&ouid=107328674588948887778&rtpof=true&sd=true

523505847

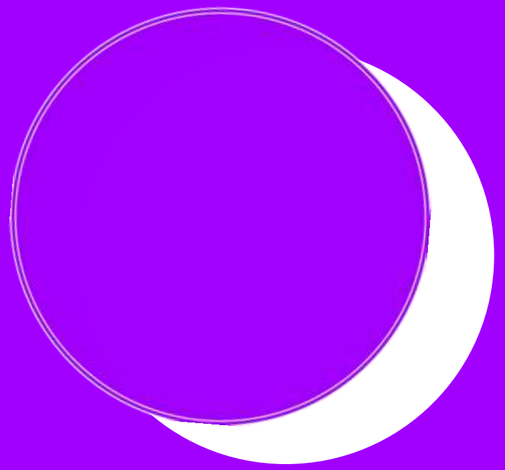
147

➢ Tech Used:-

Excel by Microsoft Corporation – For extracting & manipulating data
PowerPoint by Microsoft Corporation – For creating the project report.

➢ Approach:-

Our initial approach involves comprehensively understanding the dataset, including its structure, data types, and available information. Following this, we will undertake a data cleaning process to rectify any blank spaces resulting from human or system errors, as well as address inconsistent values. Subsequently, we will proceed to analyze the refined data, extracting the necessary insights. Finally, we will compile these insights into a professional and presentable report, ensuring effective communication of our findings.



Thank you!

ANY QUESTIONS?