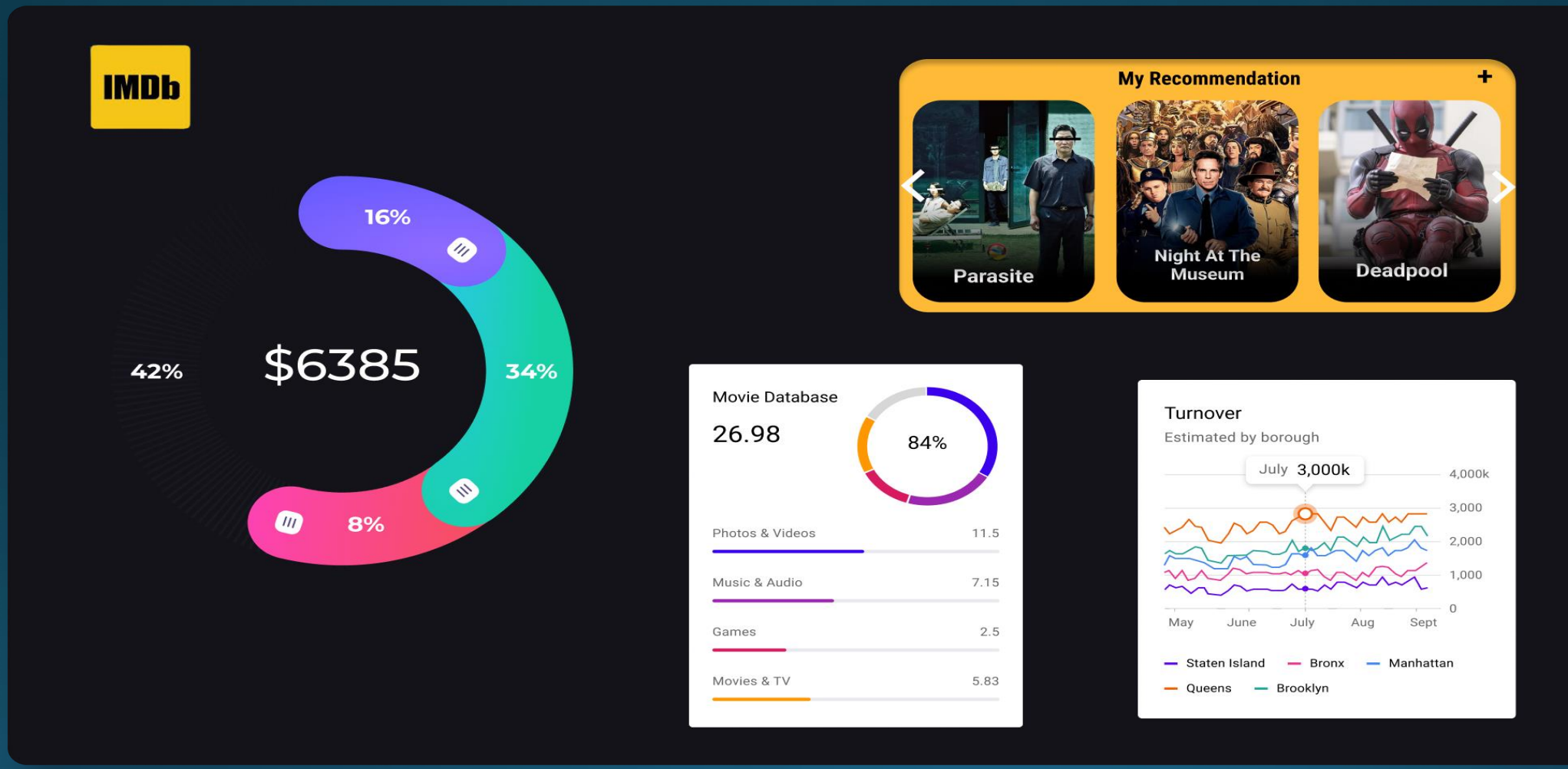
IMDb Movie Analysis

By - Raj Rathod



Project Description

- IMDb, short for Internet Movie Database, stands as a comprehensive website housing an extensive collection of movies, web-series, TV shows, and more. It serves as a platform where users rate their watched content on a scale of 1 to 10 stars, facilitating others in gauging the quality of a movie.
- The primary goal of this project is to uncover correlations between various attributes of movies and their corresponding IMDb ratings. By doing so, we aim to provide actionable insights for stakeholders to optimize future movie projects.
- Understanding the factors influencing IMDb ratings can guide stakeholders in making strategic decisions related to movie production, marketing, and distribution.
- By optimizing various aspects of movie projects based on data-driven insights, stakeholders can increase the likelihood of producing successful and well-received content.

Project Approach

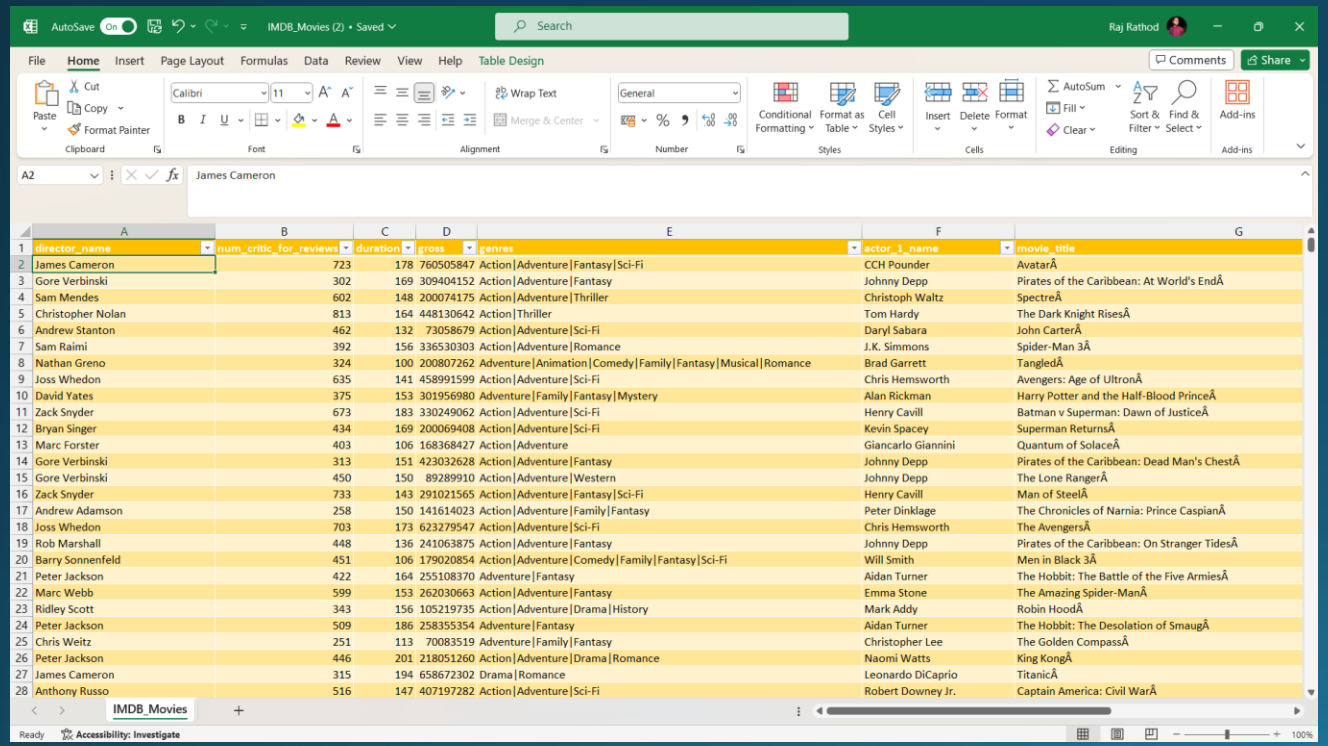
- **Assessment:** Careful examination of dataset to outline tasks.
- **Data Cleansing:** Excel-based cleaning, removal of null values, and unnecessary columns.
- **Genre Identification:** Popular genres identified through analysis.
- **Movie Duration Analysis** : Calculated the Top Movies w.r.t Time and IMDb Ratings.
- **Language Analysis** : Sorted the Top 5 Language according to No of Movies.
- **Profit Calculation:** Computation of profit for each movie, sorting by profit.
- **Director Analysis:** Grouping data by director to determine top 10 with highest mean IMDb scores.

Tech-Stack Used

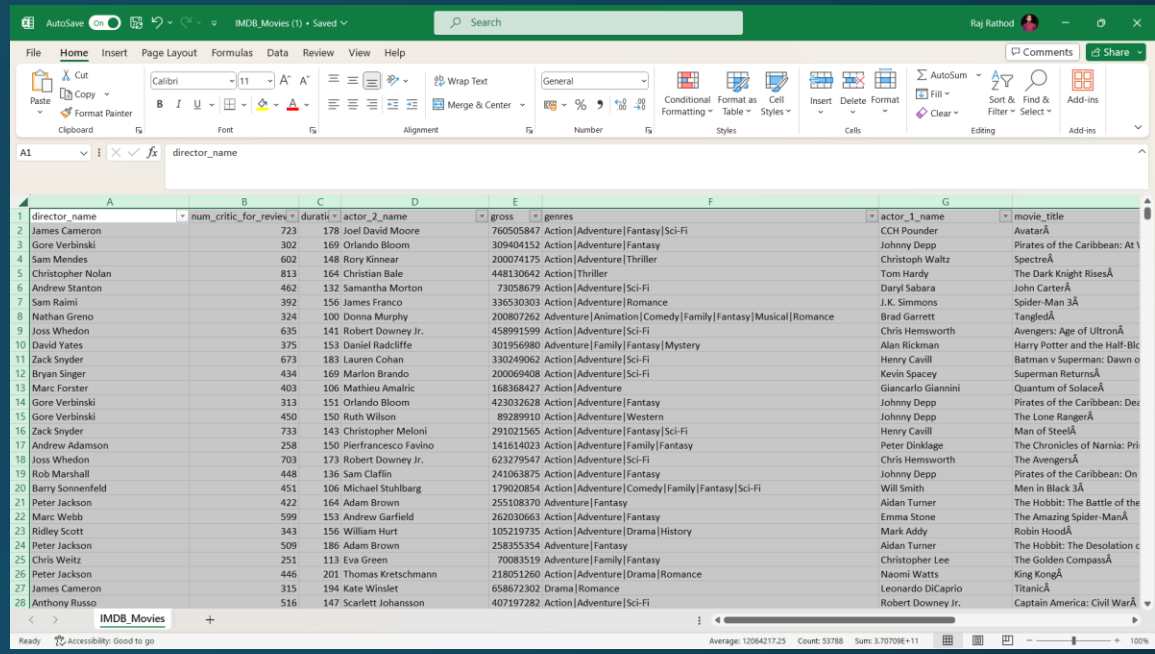
- **Microsoft Excel 2021:**Essential tool for data cleaning, analysis, and visualization.
- Sorting, filtering, and formula capabilities used for data cleansing and profit calculation.
- Functions employed for extracting top movies and grouping data.
- Charting features utilized for effective data visualization.
- User-friendly interface and diverse functionalities made it the ideal choice for the project.

Insights

- Task 1 – Data Cleaning



director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title
James Cameron	723	178	760505847	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder	AvatarÂ
Gore Verbinski	302	169	309404152	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: At World's EndÂ
Sam Mendes	602	148	200074175	Action|Adventure|Thriller	Christoph Waltz	SpectreÂ
Christopher Nolan	813	164	448130642	Action|Thriller	Tom Hardy	The Dark Knight RisesÂ
Andrew Stanton	462	132	73058679	Action|Adventure|Sci-Fi	Daryl Sabara	John CarterÂ
Sam Raimi	392	156	336530303	Action|Adventure|Romance	J.K. Simmons	Spider-Man 3Â
Nathan Greno	324	100	200807262	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romance	Brad Garrett	TangledÂ
Joss Whedon	635	141	458991599	Action|Adventure|Sci-Fi	Chris Hemsworth	Avengers: Age of UltronÂ
David Yates	375	153	301956980	Adventure|Family|Fantasy|Mystery	Alan Rickman	Harry Potter and the Half-Blood PrinceÂ
Zack Snyder	673	183	330249062	Action|Adventure|Sci-Fi	Henry Cavill	Batman v Superman: Dawn of JusticeÂ
Bryan Singer	434	169	200069408	Action|Adventure|Sci-Fi	Kevin Spacey	Superman ReturnsÂ
Marc Forster	403	106	168368427	Action|Adventure	Giancarlo Giannini	Quantum of SolaceÂ
Gore Verbinski	313	151	423032628	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's ChestÂ
Gore Verbinski	450	150	89289910	Action|Adventure|Western	Johnny Depp	The Lone RangerÂ
Zack Snyder	733	143	291021565	Action|Adventure|Fantasy|Sci-Fi	Henry Cavill	Man of SteelÂ
Andrew Adamson	258	150	141614023	Action|Adventure|Family|Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince CaspianÂ
Joss Whedon	703	173	623279547	Action|Adventure|Sci-Fi	Chris Hemsworth	The AvengersÂ
Rob Marshall	448	136	241063875	Action|Adventure|Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger TidesÂ
Barry Sonnenfeld	451	106	179020854	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	Will Smith	Men in Black 3Â
Peter Jackson	422	164	255108370	Adventure|Fantasy	Aidan Turner	The Hobbit: The Battle of the Five ArmiesÂ
Marc Webb	599	153	262030663	Action|Adventure|Fantasy	Emma Stone	The Amazing Spider-ManÂ
Ridley Scott	343	156	105219735	Action|Adventure|Drama|History	Mark Addy	Robin HoodÂ
Peter Jackson	509	186	258355354	Adventure|Fantasy	Aidan Turner	The Hobbit: The Desolation of SmaugÂ
Chris Weitz	251	113	70083519	Adventure|Family|Fantasy	Christopher Lee	The Golden CompassÂ
Peter Jackson	446	201	218051260	Action|Adventure|Drama|Romance	Naomi Watts	King KongÂ
James Cameron	315	194	658672302	Drama|Romance	Leonardo DiCaprio	TitanicÂ
Anthony Russo	516	147	407197282	Action|Adventure|Sci-Fi	Robert Downey Jr.	Captain America: Civil WarÂ



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.

Answer - The Most Common Genres of Movies in the IMDb is ‘Drama’ followed by ‘Comedy’ and then ‘Thriller’.

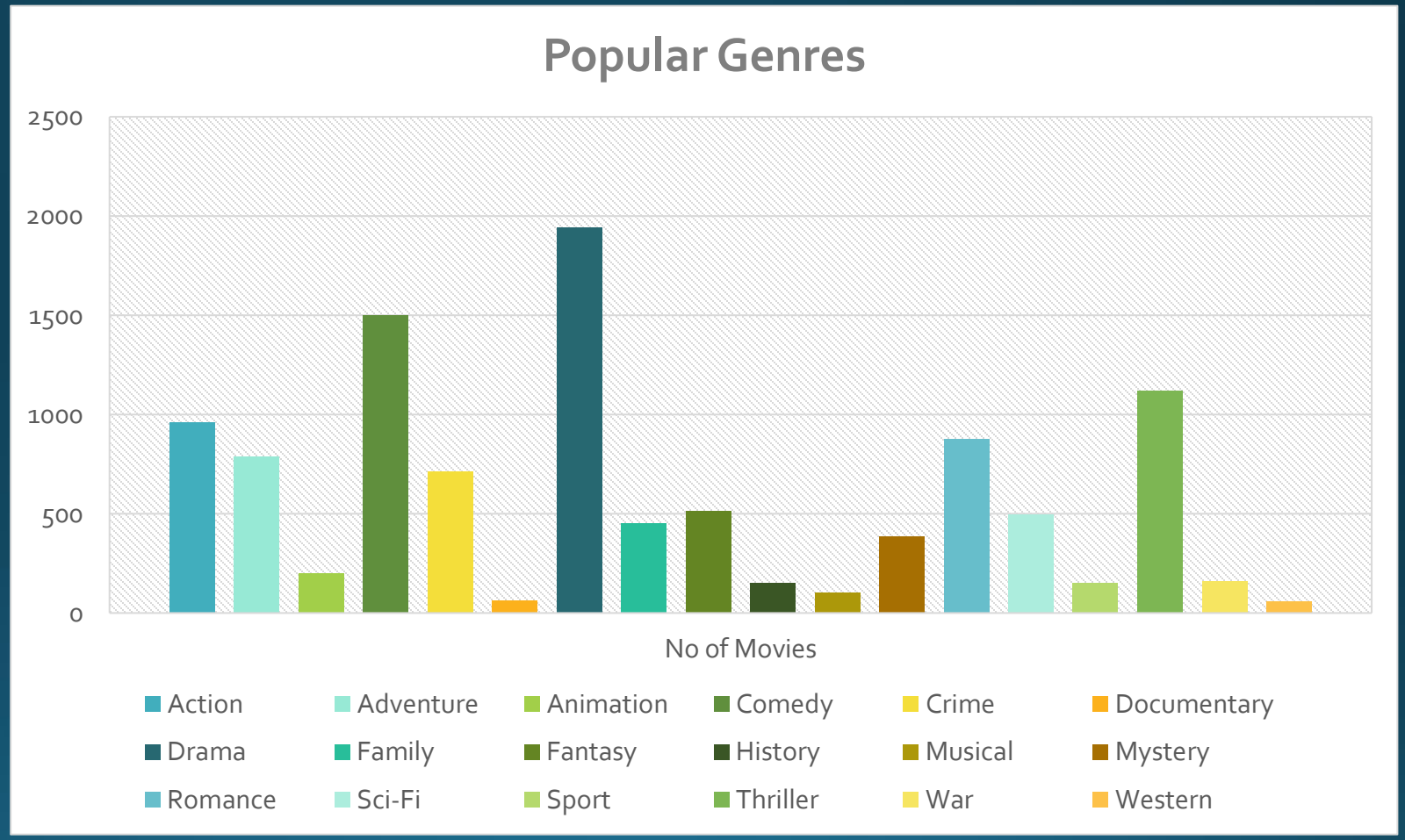
Also, T

Descriptive Analysis fo Various Genres

Mean_IMDb	Mode_IMDb	Median_IMDb	Max_IMDb	Min_IMDb	StdDev_IMDb	Var_IMDb
590.67	129.00	473.50	1941.00	58.00	531.79	282805.65

Movie Genre Analysis

Genre	No of Movies
Drama	1941
Comedy	1503
Thriller	1117
Action	962
Romance	877
Adventure	787
Crime	714
Fantasy	514
Sci-Fi	497
Family	450
Mystery	383
Animation	199
War	160
History	153
Sport	151
Musical	103
Documentary	63
Western	58

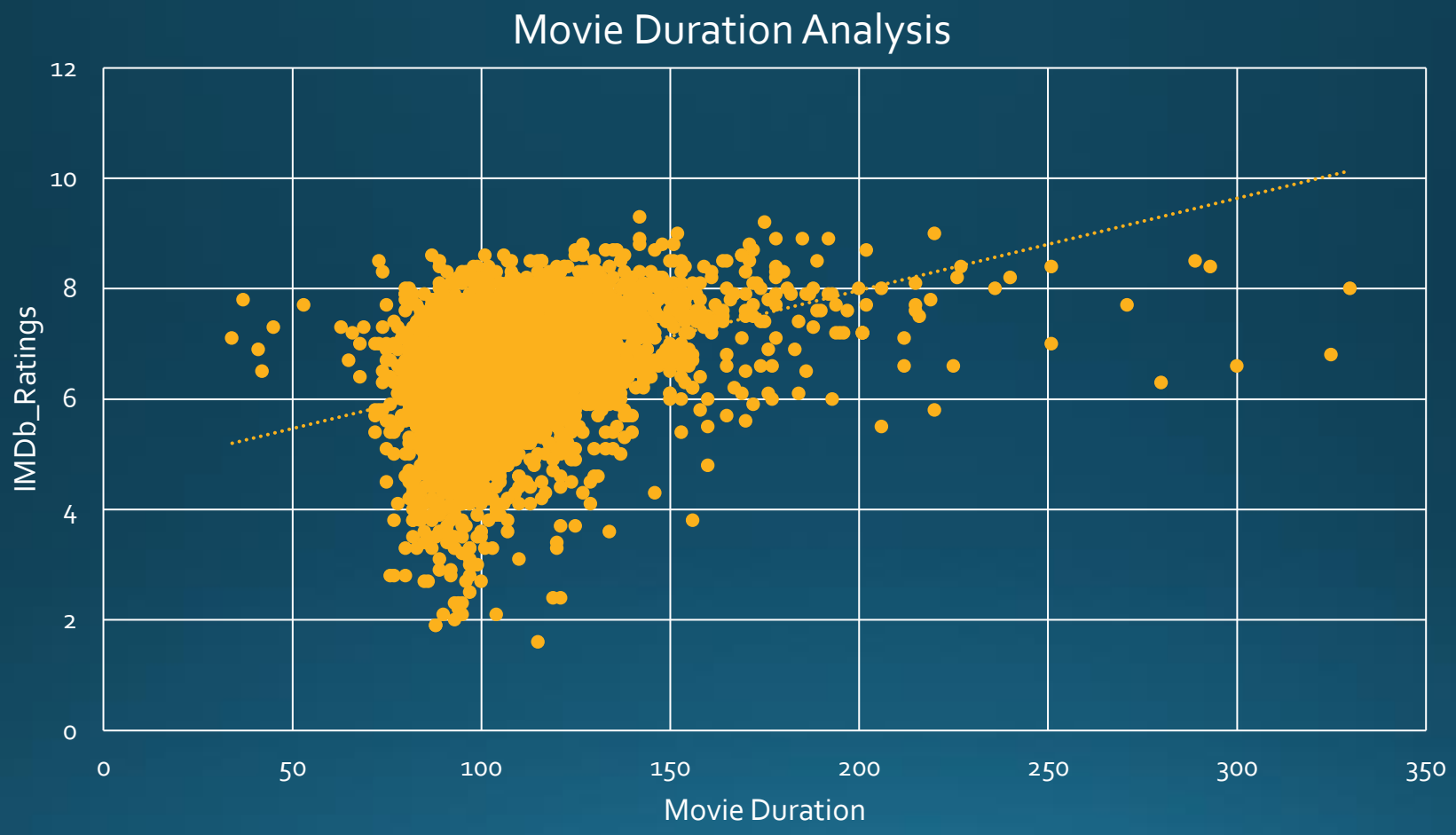


B. Movie Duration Analysis:

- Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
- Answer – The Average Movie in the Dataset is of 110 Minutes.

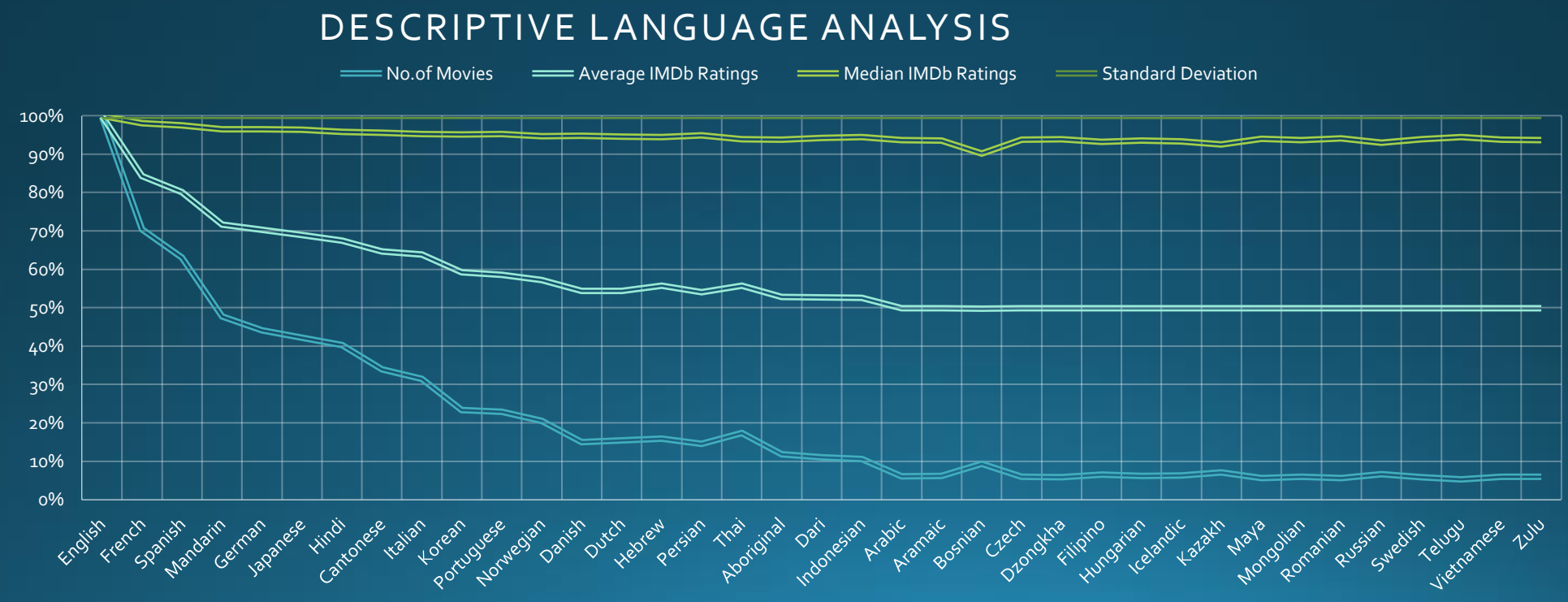
Average			Median			Standard Deviation
110			106			22.76

B. Movie Duration Analysis:



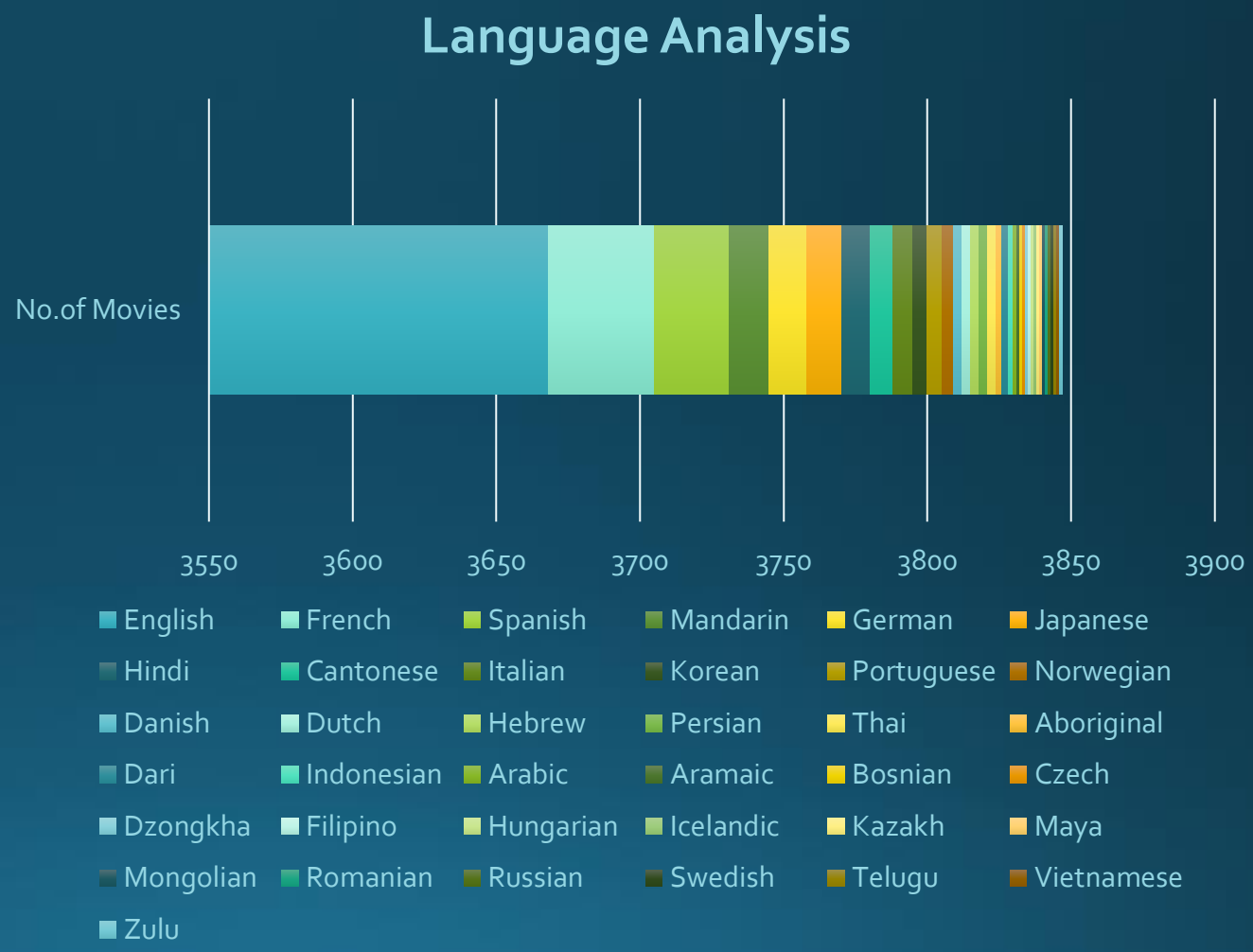
C. Language Analysis:

- Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.
- Solution – The Most Common Language used in Movies is English Language followed by French, Spanish, Mandarin



C. Language Analysis:

Language	No.of Movies	Average IMDb Ratings	Median IMDb Ratings	Standard Deviation
English	3668	6.42	6.50	1.05
French	37	7.29	7.20	1.05
Spanish	26	7.05	7.15	1.05
Mandarin	14	7.02	7.25	1.05
German	13	7.69	7.70	1.05
Japanese	12	7.63	7.80	1.05
Hindi	10	6.76	7.05	1.05
Cantonese	8	7.24	7.30	1.05
Italian	7	7.19	7.00	1.05
Korean	5	7.70	7.70	1.05
Portuguese	5	7.76	8.00	1.05
Norwegian	4	7.15	7.30	1.05
Danish	3	7.90	8.10	1.05
Dutch	3	7.57	7.80	1.05
Hebrew	3	7.50	7.30	1.05
Persian	3	8.13	8.40	1.05
Thai	3	6.63	6.60	1.05
Aboriginal	2	6.95	6.95	1.05
Dari	2	7.50	7.50	1.05
Indonesian	2	7.90	7.90	1.05
Arabic	1	7.20	7.20	1.05
Aramaic	1	7.10	7.10	1.05
Bosnian	1	4.30	4.30	1.05
Czech	1	7.40	7.40	1.05
Dzongkha	1	7.50	7.50	1.05
Filipino	1	6.70	6.70	1.05
Hungarian	1	7.10	7.10	1.05
Icelandic	1	6.90	6.90	1.05
Kazakh	1	6.00	6.00	1.05
Maya	1	7.80	7.80	1.05
Mongolian	1	7.30	7.30	1.05
Romanian	1	7.90	7.90	1.05
Russian	1	6.50	6.50	1.05
Swedish	1	7.60	7.60	1.05
Telugu	1	8.40	8.40	1.05
Vietnamese	1	7.40	7.40	1.05
Zulu	1	7.30	7.30	1.05



D. Director Analysis:

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

•Solution – Top Directors based on IMDb Score is Tony Kayne with Mean Average Rating of 8.6 and Percentile of 8.6 amongst all Top Directors and Percentile Rank is 1.

No of Movies	Top 10 Directors Name	Average IMDb Ratings	Percentile	Percentile Rank
1	Tony Kaye	8.6	8.6	1
1	Charles Chaplin	8.6	8.51	
1	Alfred Hitchcock	8.5	8.5	
1	Damien Chazelle	8.5	8.5	
1	Majid Majidi	8.5	8.44	
8	Christopher Nolan	8.4	8.4	
1	S.S. Rajamouli	8.4	8.4	
1	Richard Marquand	8.4	8.4	
3	Sergio Leone	8.4	8.4	
1	Asghar Farhadi	8.4	8.4	
1	Marius A. Markevicius	8.4	8.4	

D. Director Analysis:

TOP 10 BEST DIRECTORS - IMDB RATINGS

1	MARIUS A. MARKEVICIUS	8.4
1	ASGHAR FARHADI	8.4
3	SERGIO LEONE	8.4
1	RICHARD MARQUAND	8.4
1	S.S. RAJAMOULI	8.4
8	CHRISTOPHER NOLAN	8.4
1	MAJID MAJIDI	8.5
1	DAMIEN CHAZELLE	8.5
1	ALFRED HITCHCOCK	8.5
1	CHARLES CHAPLIN	8.6
1	TONY KAYE	8.6

E. Budget Analysis:

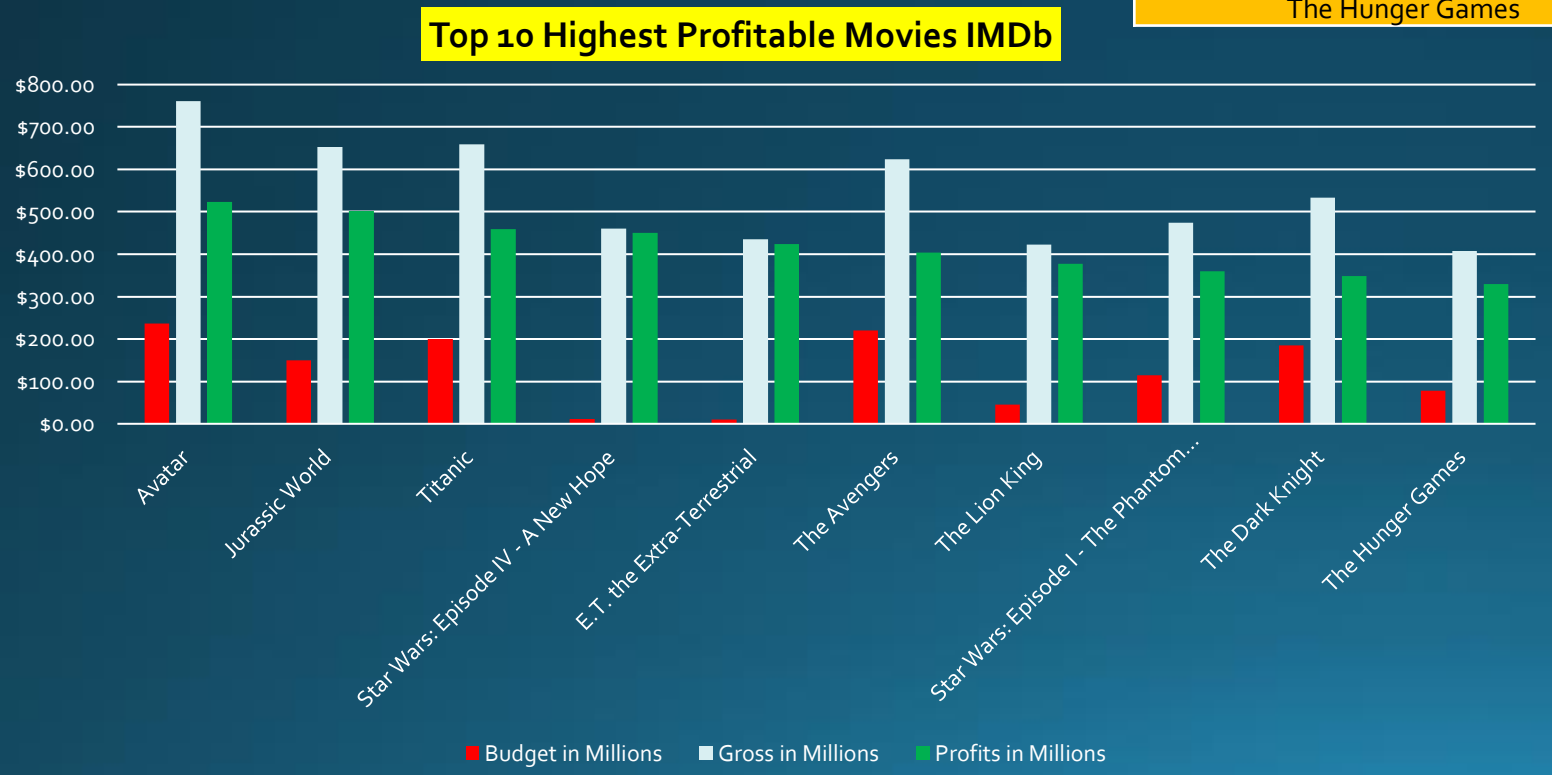
- Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

- Solution- The Correlation Coefficient is 0.100 which is Positive but If it could have near to 1,it could have been strong Possibility of High Budget Movies earns the Most.

- The Top Highest Grossing Movies of IMDb is Avatar and Jurassic World both has Earned Profit over 500Million USD+

E. Budget Analysis:

Movie Title	Budget in Millions	Gross in Millions	Profits in Millions
Avatar	$237.00	$760.51	$523.51
Jurassic World	$150.00	$652.18	$502.18
Titanic	$200.00	$658.67	$458.67
Star Wars: Episode IV - A New Hope	$11.00	$460.94	$449.94
E.T. the Extra-Terrestrial	$10.50	$434.95	$424.45
The Avengers	$220.00	$623.28	$403.28
The Lion King	$45.00	$422.78	$377.78
Star Wars: Episode I - The Phantom Menace	$115.00	$474.54	$359.54
The Dark Knight	$185.00	$533.32	$348.32
The Hunger Games	$78.00	$408.00	$330.00



Correlation Coefficient
0.100823531
Maximum Gross Profitable Movie
$523.51
Million USD

Results

- Through this project, I have gained a better understanding of data analysis and visualization techniques using Excel. It has helped me develop skills in cleaning and organizing data, identifying profitable movies, recognizing top directors, and exploring popular genres. By working on this project, I have gained practical experience in analyzing real-world data, extracting valuable insights, and presenting information effectively through charts and tables. It has improved my overall knowledge and proficiency in data analysis.

Thank You

Linkedin ID

Video Presenation

IMDb Data Analysis Raj Worksheet