



IMDB MOVIE RATING



DATA ANALYSIS ON IMDB MOVIE RATING



PROJECT DESCRIPTION



- ❑ Tasked to do a detailed report on IMDB Movie rating analysis on the dataset which features duplicate rows, missing values and with informative columns like Director's Name, Duration, Genre, Language, Gross, Budget, Movie Title, IMDB Rating...
- ❑ The potential aim of the project is to investigate the factors influence the movie success that can be identified by high IMDB Rating and uncover the valuable insights to provide significant solutions to all the movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.





APPROACH

1. Data Cleaning

2. Descriptive Statistical Analysis

3. Five “Why’s Approach



1

Cleaning the raw data plays a crucial role for deriving accurate valuable insights because the raw data consists of duplicate values, missing values and having more number of unwanted columns consumes the time for analysis.

2

In Descriptive Statistics Analysis the measure of central Tendency, Measure of Dispersion (Variability) helps to understand the structure of the data and describe the spread of the data to identify the patterns & trends lies within the dataset.

3

This technique will help you dig deeper into the problem. For instance, if you find that movies with higher budgets tend to have higher ratings, you can ask "Why?" repeatedly to uncover the root cause.



TECH STACK USED :



MS EXCEL



POWER POINT

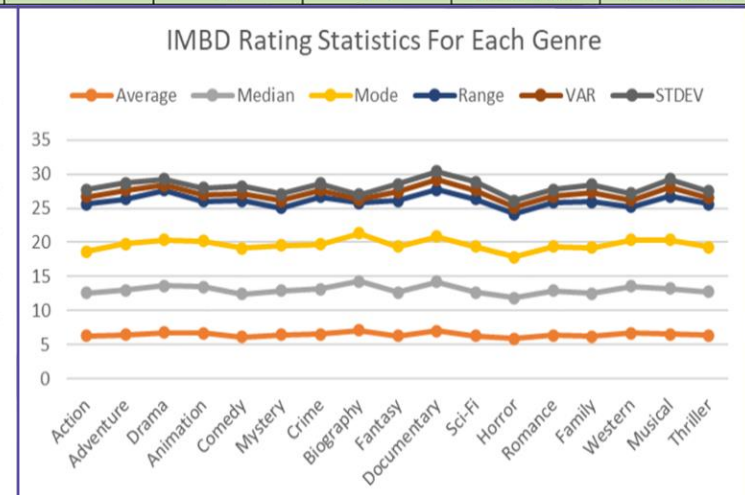
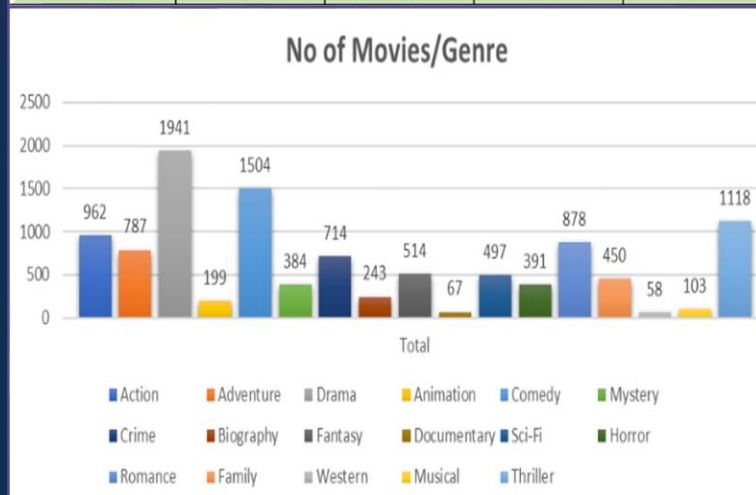


ANALYSIS AND INSIGHTS

✓ MOVIE GENRE ANALYSIS & INSIGHTS:

1. The top common genre movies are Drama, Comedy, Thriller, Action, Romance, Adventure, and crime..
2. It has been observed that both the genres of **Biography** and **Documentary** movies have a high average rating of ≥ 7 , with a decent number of movies (243 and 67, respectively). This indicates that there is currently a trend of watching movies based on real-life characters.

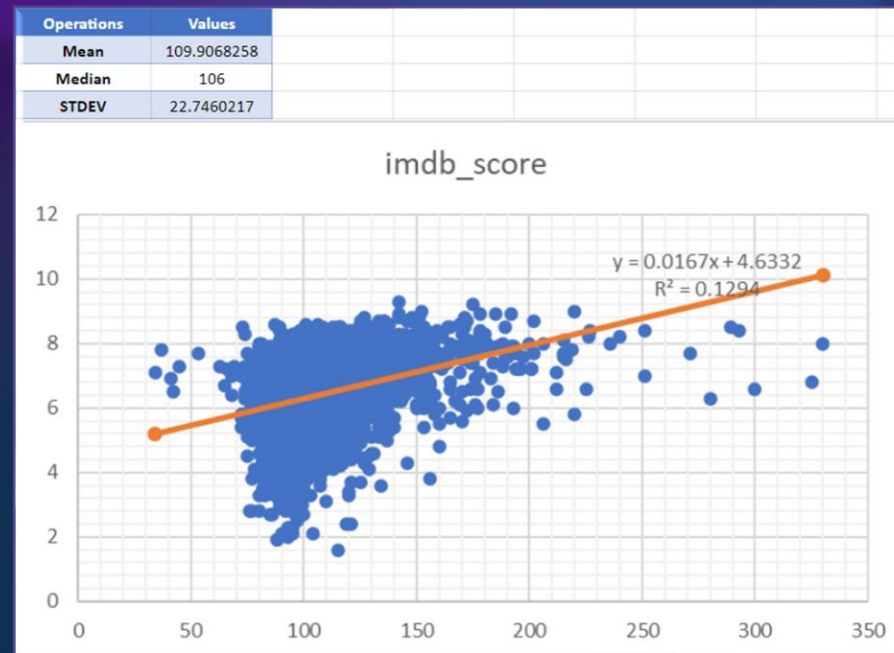
Genres	Total	Average	Median	Mode	MAX	MIN	Range	VAR	STDEV
Action	962	6.290748441	6.3	6.1	9	2.1	6.9	1.06375407	1.03138454
Adventure	787	6.458322745	6.6	6.7	8.9	2.3	6.6	1.227980898	1.108142995
Drama	1941	6.786192684	6.9	6.7	9.3	2.1	7.2	0.794056682	0.891098582
Animation	199	6.700502513	6.8	6.7	8.6	2.8	5.8	0.977322979	0.988596469
Comedy	1504	6.184441489	6.3	6.7	8.8	1.9	6.9	1.076697226	1.03764022
Mystery	384	6.472916667	6.5	6.6	8.6	3.1	5.5	1.011275022	1.005621709
Crime	714	6.544817927	6.6	6.6	9.3	2.4	6.9	0.960709434	0.980157862
Biography	243	7.141563786	7.2	7	8.9	4.5	4.4	0.502686801	0.709004091
Fantasy	514	6.290077821	6.4	6.7	8.9	2.2	6.7	1.276607011	1.129870352
Documentary	67	7.011940299	7.2	6.6	8.5	1.6	6.9	1.439855269	1.199939694
Sci-Fi	497	6.322736419	6.4	6.7	8.8	1.9	6.9	1.336558626	1.156096287
Horror	391	5.926086957	6	5.9	8.6	2.3	6.3	0.994189521	0.997090528
Romance	878	6.42881549	6.5	6.5	8.5	2.1	6.4	0.93466473	0.966780601
Family	450	6.210444444	6.3	6.7	8.6	1.9	6.7	1.351583321	1.162576157
Western	58	6.765517241	6.8	6.8	8.9	4.1	4.8	0.997035693	0.998516746
Musical	103	6.559223301	6.7	7.1	8.5	2.1	6.4	1.301850371	1.140986578
Thriller	1118	6.377370304	6.4	6.5	9	2.7	6.3	0.93317589	0.966010295



ANALYSIS AND INSIGHTS

✓ MOVIE DURATION ANALYSIS & INSIGHTS:

1. The distribution of movie duration shows clearly that it follows a normal distribution, and also the scatter plot represents a positive relation between duration and IMDB Rating
2. If we notice the movies with an IMDb rating of ≥ 7 , their duration typically lies between **130 and 200 minutes**, which indicates that viewers prefer watching movies with an average length of around 170 minutes.



UNIQUE Durations	Average	Total Movies
178	8.071428571	7
169	7.6	4
148	7.555555556	9
164	7.75	4
132	6.977419355	31
156	6.3	6
100	6.130275229	109
141	7.18	15
153	7.2	11
183	6.9	1
106	6.555319149	94
151	8.175	4
150	6.96875	16
143	7.266666667	15
173	8.1	1
136	6.88	20
186	7.2	2
113	6.531147541	61
201	7.2	3
194	7.45	2
147	7.65	8
131	6.825	28
124	6.66122449	49
135	6.948387097	31
195	7.2	1
108	6.579220779	77
104	6.245977011	87
165	7.12	5
130	6.880555556	36
142	7.546666667	15
125	6.760416667	48
123	6.860377358	53
103	6.351807229	83
118	6.625806452	62
140	7.094117647	17
149	7.6	4
114	6.494444444	54
116	6.522807018	57
154	7.333333333	9
122	6.931914894	47
93	5.806976744	86
98	6.3	105
91	5.956097561	82
158	7.14	5
96	6.143181818	88



ANALYSIS AND INSIGHTS

✓ MOVIE LANGUAGE ANALYSIS & INSIGHTS:

1. The table represents that most of the films are in English language as it is one of the most spoken language in many of the countries, so language is not one of the major influencing factor in IMDB Rating.

Languages	Total Movies	Mean IMDB	Median IMDB	STDEV IMDB
English	3673	6.424394228	6.5	1.048519845
Mandarin	14	7.021428571	7.25	0.765786244
Aboriginal	2	6.95	6.95	0.777817459
Spanish	26	7.05	7.15	0.826196103
French	37	7.286486486	7.2	0.561328861
Filipino	1	6.7	6.7	#DIV/0!
Maya	1	7.8	7.8	#DIV/0!
Kazakh	1	6	6	#DIV/0!
Telugu	1	8.4	8.4	#DIV/0!
Cantonese	8	7.2375	7.3	0.440575922
Japanese	12	7.625	7.8	0.899621132
Aramaic	1	7.1	7.1	#DIV/0!
Italian	7	7.185714286	7	1.155318962
Dutch	3	7.566666667	7.8	0.404145188
Dari	2	7.5	7.4	0.732319375
German	13	7.692307692	7.7	0.640912811
Mongolian	1	7.3	7.3	#DIV/0!
Thai	3	6.633333333	6.6	0.450924975
Bosnian	1	4.3	4.3	#DIV/0!
Korean	5	7.7	7.7	0.570087713
Hungarian	1	7.1	7.1	#DIV/0!
Hindi	10	6.76	7.05	1.111755369
Icelandic	1	6.9	6.9	#DIV/0!
Danish	3	7.9	8.1	0.529150262
Portuguese	5	7.76	8	0.978774744
Norwegian	4	7.15	7.3	0.574456265
Czech	1	7.4	7.4	#DIV/0!
Russian	1	6.5	6.5	#DIV/0!
None	1	8.5	8.5	#DIV/0!
Zulu	1	7.3	7.3	#DIV/0!
Hebrew	3	7.5	7.3	0.435889894
Dzongkha	1	7.5	7.5	#DIV/0!
Arabic	1	7.2	7.2	#DIV/0!
Vietnamese	1	7.4	7.4	#DIV/0!
Indonesian	2	7.9	7.9	0.424264069
Romanian	1	7.9	7.9	#DIV/0!
Persian	3	8.133333333	8.4	0.550757055
Swedish	1	7.6	7.6	#DIV/0!



ANALYSIS AND INSIGHTS

✓ MOVIE DIRECTOR ANALYSIS & INSIGHTS:

1. The top directors with more than 5 movies, each having an average IMDb rating of ≥ 7 , have been successfully delivering great movies.

Director Names	No_of_movies	Mean IMD	Percentile IMD
Steven Spielberg	25	7.544	0.909
Woody Allen	19	7	0.724
Clint Eastwood	19	7.205263158	0.841
Ridley Scott	17	7.070588235	0.769
Martin Scorsese	16	7.675	0.938
Richard Linklater	11	7.327272727	0.87
Rob Reiner	11	7.018181818	0.763
Francis Ford Coppola	9	7.655555556	0.937
Peter Jackson	9	7.675	0.939
Stephen Frears	8	7.0875	0.772
Danny Boyle	8	7.4375	0.888
Ang Lee	8	7.25	0.845
James Mangold	8	7.075	0.77
Quentin Tarantino	8	8.2	0.989
Edward Zwick	8	7.325	0.868
Gus Van Sant	7	7.128571429	0.801
Terry Gilliam	7	7.328571429	0.871
Guy Ritchie	7	7.114285714	0.799
Ethan Coen	7	7.114285714	0.799
David O. Russell	7	7.171428571	0.811
Wes Anderson	7	7.628571429	0.931
Kenneth Branagh	7	7.157142857	0.808
Jason Reitman	6	7.2	0.812
James Wan	6	7.2	0.812
Darren Aronofsky	6	7.483333333	0.892
Paul Thomas Anderson	6	7.516666667	0.908
Sam Mendes	6	7.5	0.892
Doug Liman	6	7.128571429	0.8

Director Names	No_of_movies	Mean IMDB	Percentile IMDB
Steven Spielberg	25	7.544	0.909
Woody Allen	19	7	0.724
Clint Eastwood	19	7.205263158	0.841
Steven Soderbergh	16	6.70625	0.621
Tim Burton	16	6.93125	0.714
Ridley Scott	17	7.070588235	0.769
Renny Harlin	15	5.746666667	0.247
Spike Lee	15	6.733333333	0.623
Martin Scorsese	16	7.675	0.938
Robert Rodriguez	13	5.692307692	0.228
Barry Levinson	13	6.576923077	0.552
Oliver Stone	13	6.907692308	0.709
John Carpenter	12	6.833333333	0.664
Joel Schumacher	12	6.341666667	0.454
Tony Scott	12	6.791666667	0.638
Wes Craven	11	6.109090909	0.379
Richard Linklater	11	7.327272727	0.87
Sam Raimi	11	6.85	0.677
Rob Reiner	11	7.018181818	0.763
Shawn Levy	11	6.090909091	0.349
Chris Columbus	11	6.654545455	0.591
Kevin Smith	10	6.78	0.636
Rob Cohen	10	5.69	0.227
Paul W.S. Anderson	10	5.99	0.31
Brian De Palma	10	6.6	0.554
John McTiernan	10	6.63	0.585
Ron Howard	13	6.930769231	0.714
Dennis Dugan	9	5.733333333	0.246
Bobby Farrelly	9	6.244444444	0.417
Richard Donner	9	6.633333333	0.587
Phillip Noyce	9	6.766666667	0.634
Francis Ford Coppola	9	7.655555556	0.937
Donald Petrie	8	5.6625	0.225
Garry Marshall	8	5.95	0.301
Lasse Hallström	8	6.9875	0.724
Ivan Reitman	8	6.075	0.345
Neil Jordan	8	6.7625	0.631
Stephen Frears	8	7.0875	0.772
Adam Shankman	8	5.9625	0.308
Harold Ramis	8	6.55	0.543
Danny Boyle	8	7.4375	0.888
Ang Lee	8	7.25	0.845
F. Gary Gray	8	6.9375	0.717
James Mangold	8	7.075	0.77
Peter Jackson	9	7.675	0.939



ANALYSIS AND INSIGHTS

✓ MOVIE BUDGET ANALYSIS & INSIGHTS:

1. The table represents that the correlation between gross and budget is positive, that indicates if the budget increases, there is a high probability that the gross collection will Increase which will turnout into a profit margin
2. When the budget is high, the quality, size and casting of the movie make it stand out among the competition and the viewers preferred to see their stars in theatres only.

Top 10 Movies	Profit	Correl Between Budget and Gross of the movies
Avatar	523505847	CORREL is designed to work with ranges, not individual cells or relative ranges.
Jurassic World	502177271	
Titanic	458672302	0.100982454
Kinsey	449935665	
Friday After Next	424449459	
The Avengers	403279547	
The Lion King	377783777	
Star Wars: Episode I - The Phantom Menace	359544677	
The Dark Knight	348316061	
The Hunger Games	329999255	



Movie Title	Profit Margins	Movie Ranking
Avatar	523505847	1
Pirates of the Caribbean: At World's End	9404152	1464
Spectre	-44925825	3594
The Dark Knight Rises	198130642	48
John Carter	-190641321	3771
Spider-Man 3	78530303	314
Tangled	-59192738	3665
Avengers: Age of Ultron	208991599	44
Harry Potter and the Half-Blood Prince	51956980	499
Batman v Superman: Dawn of Justice	80249062	300
Superman Returns	-8930592	2730
Quantum of Solace	-31631573	3461
Pirates of the Caribbean: Dead Man's Chest	198032628	49
The Lone Ranger	-125710090	3760
Man of Steel	66021565	383
The Chronicles of Narnia: Prince Caspian	-83385977	3720
The Avengers	403279547	6
Pirates of the Caribbean: On Stranger Tides	-8936125	2732
Men in Black 3	-45979146	3603
The Hobbit: The Battle of the Five Armies	5108370	1649
The Amazing Spider-Man	32030663	804
Robin Hood	-94780265	3741
The Hobbit: The Desolation of Smaug	33355354	771
The Golden Compass	-109916481	3754
King Kong	11051260	1402



RESULTS

❖ During the course of the assignment, I gained valuable insights into descriptive statistical analysis and learned how to effectively use Excel's pivot tables and functions to examine large datasets. I honed my data analysis skills and discovered how to draw meaningful conclusions from complex data. Additionally, I enhanced my ability to create data visualizations and reports in Excel, which improved the way I communicated my findings. By utilizing statistical analysis tools, I was able to identify patterns and trends within the datasets, allowing me to provide data-driven recommendations for process improvement. Overall, this project helped me build a strong foundation in data analysis and equipped me with practical skills that I can apply to future tasks and responsibilities.

- Data Cleaning
- Data Analysis
- Data Visualization
- Handling Outliners
- Descriptive Statistics
- Excel Functions



DRIVE LINKS

- HYPERLINK:
- SPREAD SHEET LINK: https://docs.google.com/spreadsheets/d/1pyTbMkSGPML8aB-aX_8ni148a6KRIU4fSmxky93qQjI/edit?usp=sharing
- GOOGLE DRIVE ASSIGNMENT LINK:
https://drive.google.com/drive/u/o/folders/1tmporV5N7_LymJ9b3-0aTIF-zyAjawm1