A PROJECT REPORT ON

IMDB MOVIE ANALYSIS

Submitted By: -

Pritesh Kumar Bag

Trainity

C-97, C-97, Ahinsa Cir, Panch Batti, C Scheme,

Ashok Nagar, Jaipur, Rajasthan 302001

AUGUST 29,2023

PROJECT DESCRIPTION

IMDb, short for the Internet Movie Database, is a widely recognized online platform that provides a comprehensive database of information related to films, television shows, actors, directors, and other entertainment industry professionals. Users can access details such as cast and crew lists, plot summaries, user reviews, ratings, and trivia for a vast array of movies and TV shows from around the world. IMDb serves as a valuable resource for film enthusiasts, industry professionals, and casual viewers seeking information about their favourite media. It has become a go-to platform for discovering, exploring, and learning about the world of cinema and television.

APPROACH

- Downloaded the given datafile
- Understanding the datafile
- Checking the blanks and outliers
- Removing unwanted data
- Drawing summary from the data
- Used formulas, pivot table and charts

TECH STACK USED

- Microsoft Excel 2016

INSIGHTS

1. **Movie Genre Analysis: Analyse 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. Hint: Use Excel's COUNTIF function to count the number of movies for each genre. You might need to manipulate the 'genres' column to separate multiple genres for a single movie. Use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics. Compare the statistics to understand the impact of genre on movie ratings.

Generes	Count of movies	Mean OF IMDB SCORE		Generes	Count of movies	Mean OF IMDB SCORE
Action	957	6.287565308		Animation	198	6.702525253
Adventure	782	6.453964194		Family	446	6.210538117
Fantasy	510	6.289607843		War	156	7.037179487
Western	58	6.765517241		Sport	150	6.602
Romance	866	6.429099307		Mystery	382	6.472774869
Documentary	54	6.990740741		Film-Noir	1	7.7
Drama	1918	6.782221064		History	151	7.138410596
Comedy	1484	6.180188679		Crime	710	6.541408451
Sci-Fi	495	6.323434343		Musical	98	6.587755102
Thriller	1111	6.377047705				
Horror	388	5.922680412				

DESCRIPTIVE STATISTICS OF IMDB SCORE										
	Action	Adventure	Fantasy	Western	Romance	Documen	Drama	Comedy	Sci-Fi	Thriller
MEDIAN	6.3	6.6	6.6	6.8	6.5	7.35	6.9	6.3	6.4	6.4
MODE	6.1	6.7	6.7	6.8	6.5	6.6	6.7	6.7	6.7	6.5
VARIANCE	1.063977016	1.229196813	1.27276016	0.997035693	0.928632	1.731799	0.798311809	1.075197	1.341757	0.933338
STANDARD DEV	1.031492616	1.108691487	1.128166725	0.998516746	0.963656	1.315979	0.893482965	1.036917	1.158343	0.966094
MAX	9	8.9	8.9	8.9	8.5	8.5	9.3	8.8	8.8	9
MIN	2.1	2.3	2.2	4.1	2.1	1.6	2.1	1.9	1.9	2.7
RANGE	6.9	6.6	6.7	4.8	6.4	6.9	7.2	6.9	6.9	6.3

	Horror	Animation	Family	War	Sport	Mystery	Film-Noir	History	Crime	Musical
MEDIAN	6	6.8	6.3	7.1	6.8	6.5	7.7	7.2	6.6	6.7
MODE	5.9	6.7	6.7	7.1	7.2	6.6	#N/A	7.7	6.6	7.1
VARIANCE	0.98940675	0.981465672	1.11293556	0.646866832	1.095634	1.012328	#DIV/0!	0.457715	0.963022	1.199436
STANDARD DEV	0.994689273	0.990689493	1.166344978	0.804280319	1.046725	1.006145	#DIV/0!	0.676546	0.981337	1.095188
MAX	8.6	8.6	8.6	8.6	8.4	8.6	7.7	8.9	9.3	8.5
MIN	2.3	2.8	1.9	4.3	2	3.1	7.7	5.5	2.4	2.1
RANGE	6.3	5.8	6.7	4.3	6.4	5.5	0	3.4	6.9	6.4

INFERENCE-Maximum count of movies is Drama followed by Comedy genre. Lowest is film Noir Genre. War and History genre though have less count in movies but their rating s are more than 7 on average. It means people like movies based on facts and history.

2. **Movie Duration Analysis: Analyse the distribution of movie durations and its impact on the IMDB score.** Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Hint: Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV. Create a scatter plot to visualize the relationship between movie duration and IMDB score. Add a trendline to assess the direction and strength of the relationship.

DESCRIPTIVE STATISTICS	
MEAN	110.0163
MEDIAN	106
MODE	101
STDEV	22.6224
VARIANCE	511.7729
MINIMUM	37
MAXIMUM	330
RANGE	293
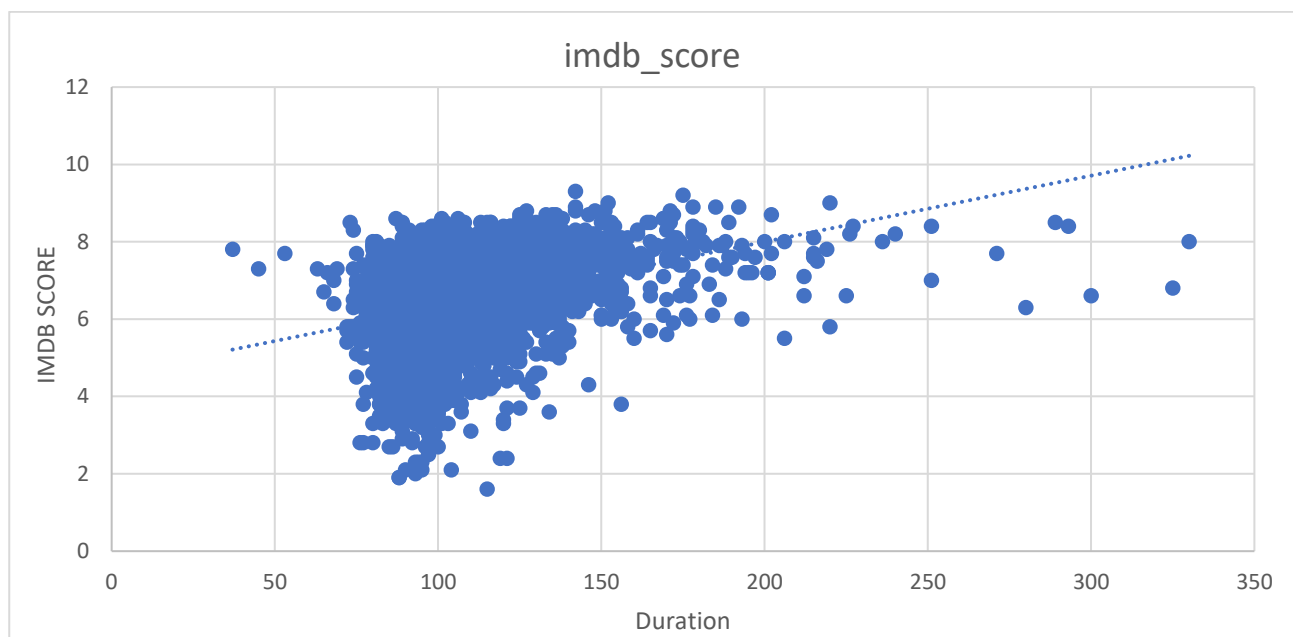
INFERENCE-On average duration of movie is 110 minutes. Median is 106 and mode 101.Maxiumum time duration of a movie is 330 minutes and the minimum duration is 37.There is little positive correlation between IMDB score and duration of the movie.

3. **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.

Hint: Use Excel's COUNTIF function to count the number of movies for each language. Calculate the mean, median, and standard deviation of the IMDB scores for each language. Compare the statistics to understand the impact of language on movie ratings.

Language	No. of Movies	IMDB Scores Mean	Median	Mode	SD	Variance
Aboriginal	2	6.95	0	#N/A	0.777817	0.605
Arabic	1	7.2	7.2	#N/A	#DIV/0!	#DIV/0!
Aramaic	1	7.1	7.1	#N/A	#DIV/0!	#DIV/0!
Bosnian	1	4.3	4.3	#N/A	#DIV/0!	#DIV/0!
Cantonese	7	7.342857	7.3	7.3	0.35051	0.122857
Czech	1	7.4	7.4	#N/A	#DIV/0!	#DIV/0!
Danish	3	7.9	8.1	#N/A	0.52915	0.28
Dari	2	7.5	7.5	#N/A	0.141421	0.02
Dutch	3	7.566667	7.8	7.8	0.404145	0.163333
English	3641	6.422247	6.5	6.7	1.049332	1.101098
Filipino	1	6.7	6.7	#N/A	#DIV/0!	#DIV/0!
French	34	7.355882	7.3	7.2	0.519435	0.269813
German	11	7.763636	7.8	7.4	0.675681	0.456545
Hebrew	2	7.65	7.65	#N/A	0.494975	0.245
Hindi	5	7.22	7.4	#N/A	0.801249	0.642
Hungarian	1	7.1	7.1	#N/A	#DIV/0!	#DIV/0!
Indonesiar	2	7.9	7.9	#N/A	0.424264	0.18
Italian	7	7.185714	7	#N/A	1.155319	1.334762
Japanese	10	7.66	8	#N/A	0.990174	0.980444
Kazakh	1	6	6	#N/A	#DIV/0!	#DIV/0!
Korean	5	7.7	7.7	#N/A	0.570088	0.325
Mandarin	14	7.021429	7.25	7.6	0.765786	0.586429
Maya	1	7.8	7.8	#N/A	#DIV/0!	#DIV/0!
Mongolian	1	7.3	7.3	#N/A	#DIV/0!	#DIV/0!
None	1	8.5	8.5	#N/A	#DIV/0!	#DIV/0!
Norwegiar	4	7.15	7.3	7.6	0.574456	0.33
Persian	3	8.133333	8.4	#N/A	0.550757	0.303333
Portugues	5	7.76	8	#N/A	0.978775	0.958
Romanian	1	7.9	7.9	#N/A	#DIV/0!	#DIV/0!
Russian	1	6.5	6.5	#N/A	#DIV/0!	#DIV/0!
Spanish	24	7.045833	7.15	5.9	0.860727	0.740851
Thai	3	6.633333	6.6	#N/A	0.450925	0.203333
Vietnames	1	7.4	7.4	#N/A	#DIV/0!	#DIV/0!
Zulu	1	7.3	7.3	#N/A	#DIV/0!	#DIV/0!

INFERENCE- More number of movies are made in English language. A non-verbal movie and some movies by Persian language and Indonesen language movies though produce less movies have IMDB score near and above 8. So, the stories of such language's night have attracted the audience and reviewers' attention. More movies and remakes can be made from them.

4. **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.

Hint: Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores.

director_name	avg imdb_score
Akira Kurosawa	8.7
Alfred Hitchcock	8.5
Asghar Farhadi	8.4
Charles Chaplin	8.6
Christopher Nolan	8.425
Damien Chazelle	8.5
Majid Majidi	8.5
Marius A. Markevicius	8.4
Richard Marquand	8.4
Ron Fricke	8.5
Sergio Leone	8.433333333
Tony Kaye	8.6
LARGE	8.7
CONTRIBUTION	0.92307
PERCENTILE	7.6

INTERPRETATION-The top ten directors with highest scores were taken. On the basis of IMDB score the success rate of a move from the director Akira Kurosawa is 92 percent. He is the director with highest average IMDB score.

5. **Budget Analysis: Explore the relationship between movie budgets and their financial success**. Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Hint: Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function. Calculate the profit margin (gross earnings - budget) for each movie and identify the movies with the highest profit margin using Excel's MAX function.

CORRELATION	
0.10050095	
HIGH PROFIT MARGIN	
523505847	

INFERENCE- there is aa relatively weak positive correlation between budget and gross income. The highest profit margin is around 52 crore.

RESULTS

- Maximum count of movies is Drama followed by Comedy genre. Lowest is film Noir Genre. War and History genre though have less count in movies but their rating s are more than 7 on average. It means people like movies based on facts and history.

- On average duration of movie is 110 minutes. Median is 106 and mode 101.Maxiumum time duration of a movie is 330 minutes and the minimum duration is 37. There is little positive correlation between IMDB score and duration of the movie.

- More number of movies are made in English language. A non-verbal movie and some movies by Persian Language and Indonesian language movies though produce less movies have IMDB score near and above 8. So, the stories of such language's night have attracted the audience and reviewers' attention. More movies and remakes can be made from them.

- The top ten directors with highest scores were taken. On the basis of IMDB score the success rate of a move from the director Akira Kurosawa is 92 percent. He is the director with highest average IMDB score

- There is aa relatively weak positive correlation between budget and gross income. The highest profit margin is around 52 crores.