IMDB MOVIE ANALYSIS

Final project 01 by: PREETY MOHANTA

CONTENT

- Introduction
- Approach
- Tech Stack Used
- Tasks
- Insight
- Summary
- Links

INTRODUCTION…

- In this project we are working on large dataset of IMDB_Movies.
- With the help of the given data we will analyse the movies dataset and understand the factors that lead to movies success.
- With this we get to know about the impact of various factors on genre, duration, budget etc on imdb ratings
- At the end with all these analysis filmmakers and directors can get insight of the “success of the movies” and make a good call.

APPROACH…

Here are the steps we will follow for executing our project:

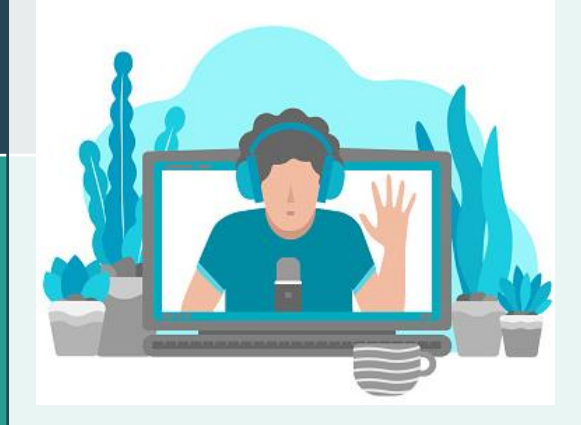
- Download the data set
- Then clean the data by removing unwanted information, duplicate, handle the missing data.
- After cleaning the data convert the data types and make it ready to use.
- Now understand the data and use the excel and perform the tasks
- With the help of excel we will perform descriptive statistics, correlation, percentile, etc.
- And at the end we get insight from it.

TECH STACK USED...

Excel

PowerPoint

Presentation



TASKS....

You are required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

1st Task	2nd Task	3rd Task	4th Task	5th Task
Movie Genre Analysis	**Movie Duration Analysis**	**Language Analysis**	**Director Analysis**	**Budget Analysis**
Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics of the IMDB scores.	Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.	Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.	Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.	Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

INSIGHT

Movie Genre Analysis

Row Labels	Count of Unique Genres	Average of Imdb_Score	Max. of Imdb_Score	Min. of Imdb_Score	StdDevp of Imdb_Score	Varp of Imdb_Score
Action	935	6.286	9	2.1	1.038	1.077
Adventure	367	6.561	8.6	2.3	1.123	1.261
Animation	46	6.763	8	4.5	0.962	0.925
Biography	206	7.152	8.9	4.5	0.698	0.487
Comedy	1026	6.164	8.8	1.9	1.036	1.074
Crime	252	6.945	9.3	3.3	0.867	0.752
Documentary	43	6.951	8.5	1.6	1.400	1.959
Drama	676	6.822	8.8	2.1	0.905	0.820
Family	3	6.500	7.9	5.7	0.993	0.987
Fantasy	35	6.234	7.9	4.3	0.881	0.777
Horror	156	5.813	8.5	2.3	1.005	1.009
Musical	2	6.750	7.2	6.3	0.450	0.203
Mystery	23	6.587	8.5	3.3	1.085	1.177
Romance	2	6.650	7.1	6.2	0.450	0.202
Sci-Fi	8	6.588	8.2	5	0.965	0.931
Thriller	3	5.300	6.3	4.8	0.707	0.500
Western	3	6.767	8.9	4.1	1.996	3.982
(blank)						
Grand Total	3786	6.462	9.3	1.6	1.057	1.118

RESULT:

The data reveals insights into movie genres based on IMDb scores:

•Prolific genres like Drama, Action, and Comedy have numerous entries, but varied average IMDb scores (6.822, 6.286, and 6.164, respectively).

•Genres like Biography and Animation have higher average IMDb scores (7.152 and 6.763) despite fewer entries.

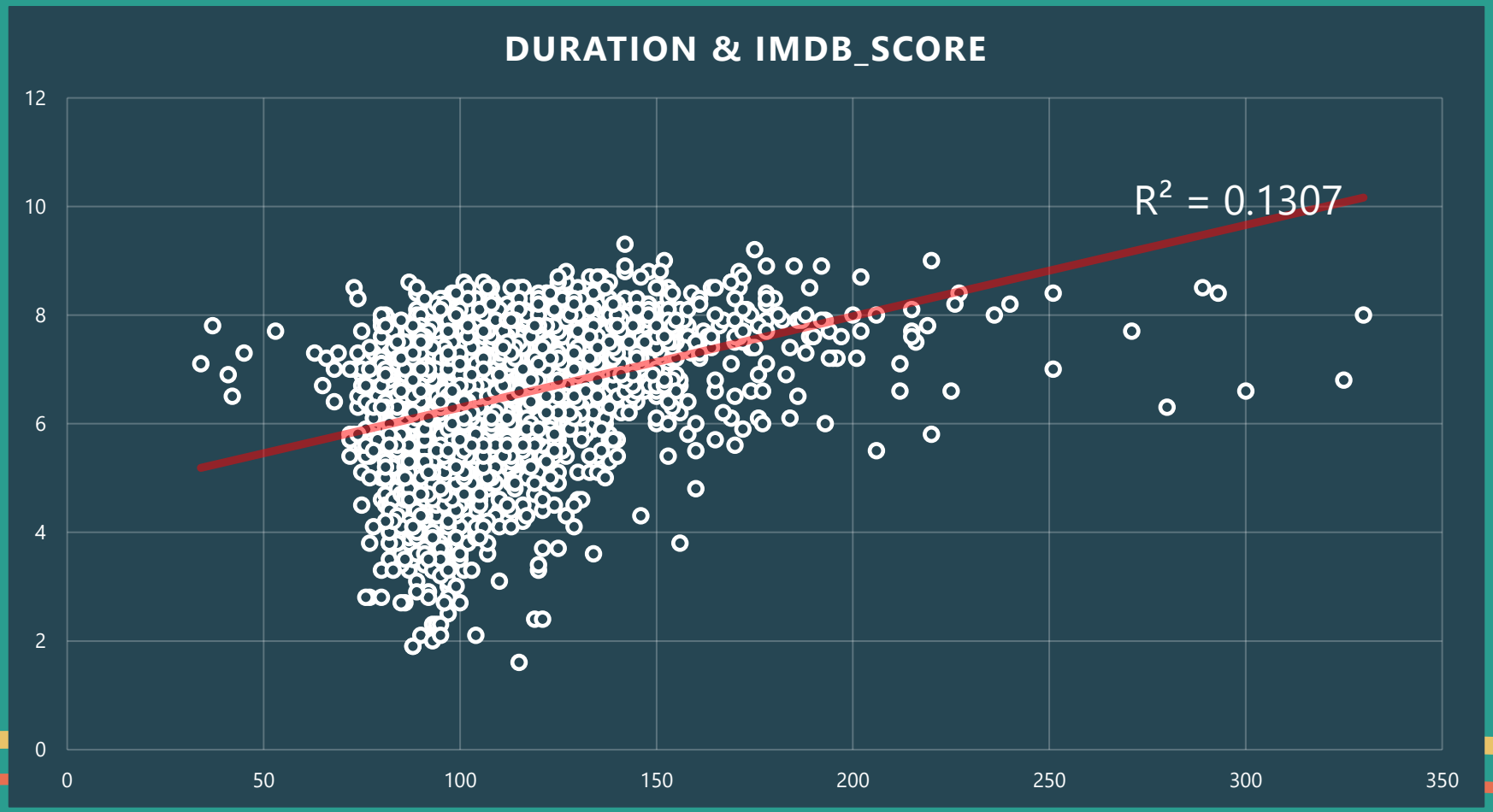
•Maximum and minimum scores shows the differences within each genre, and the standard deviation and variance of metrics showing the score dispersion.

•The insight of the dataset shows the vast world of filmmaking and audience preferences across genres.

Movie Duration Analysis

Column1	Column2
Duration	
Mean	109.808505
Standard Error	0.369949997
Median	105
Mode	101
Standard Deviation	22.763201
Sample Variance	518.16332
Kurtosis	12.40512587
Skewness	2.347508256
Range	296
Minimum	34
Maximum	330
Sum	415735
Count	3786
Largest(1)	330
Smallest(1)	34
Confidence Level(95.0%)	0.725320612

	Duration	Imdb_Score
Duration	1	
Imdb_Score	0.361506958	1



RESULT:

Insights:

- Tendency shows shorter durations.
- Considerable variability with a wide range.
- Right-skewed distribution.
- Heavy-tailedness.

These statistics shows the insight of central tendency, dispersion, and shape of the duration.

- Correlation coefficient: Approximately 0.3615.
- Indicates a moderate positive correlation.
- Longer movie durations tend to have slightly higher IMDB scores.
- Correlation doesn't imply causation.
- Other factors influence IMDB scores.
- Considerable variation exists around this trend.

Based on this, it shows that there is some relationship between movie duration and IMDB score, it is influenced by various factors, given the considerable variability and distribution shape of the duration.

Language Analysis

Row Labels	Count of Language	Average of Imdb_Score	StdDevp of Imdb_Score2	Varp of Imdb_Score
Aboriginal	2.000	6.950	0.550	0.303
Arabic	1.000	7.200	0.000	0.000
Aramaic	1.000	7.100	0.000	0.000
Bosnian	1.000	4.300	0.000	0.000
Cantonese	8.000	7.238	0.412	0.170
Czech	1.000	7.400	0.000	0.000
Danish	3.000	7.900	0.432	0.187
Dari	2.000	7.500	0.100	0.010
Dutch	3.000	7.567	0.330	0.109
Dzongkha	1.000	7.500	0.000	0.000
English	3606.000	6.421	1.052	1.107
Filipino	1.000	6.700	0.000	0.000
French	37.000	7.286	0.554	0.307
German	13.000	7.692	0.616	0.379
Hebrew	3.000	7.500	0.356	0.127
Hindi	10.000	6.760	1.055	1.112
Hungarian	1.000	7.100	0.000	0.000
Icelandic	1.000	6.900	0.000	0.000
Indonesian	2.000	7.900	0.300	0.090
Italian	7.000	7.186	1.070	1.144

Row Labels	Count of Language	Average of Imdb_Score	StdDevp of Imdb_Score2	Varp of Imdb_Score
Japanese	12.000	7.625	0.861	0.742
Kazakh	1.000	6.000	0.000	0.000
Korean	5.000	7.700	0.510	0.260
Mandarin	14.000	7.021	0.738	0.545
Maya	1.000	7.800	0.000	0.000
Mongolian	1.000	7.300	0.000	0.000
None	1.000	8.500	0.000	0.000
Norwegian	4.000	7.150	0.497	0.247
Persian	3.000	8.133	0.450	0.202
Portuguese	5.000	7.760	0.875	0.766
Romanian	1.000	7.900	0.000	0.000
Russian	1.000	6.500	0.000	0.000
Spanish	26.000	7.050	0.810	0.656
Swedish	1.000	7.600	0.000	0.000
Telugu	1.000	8.400	0.000	0.000
Thai	3.000	6.633	0.368	0.136
Vietnamese	1.000	7.400	0.000	0.000
Zulu	1.000	7.300	0.000	0.000
(blank)				
Grand Total	**3786.000**	**6.462**	**1.057**	**1.118**

RESULT:

The data shows the count of movies for each language along with the average IMDb score, standard deviation, and variance of IMDb scores.

•"English" dominates the data with 3606 movies, with an average IMDb score of 6.421, standard deviation of 1.052, and variance of 1.107.

•Languages like "Arabic," "Aramaic," and "Kazakh" have only one movie, making it difficult to understand the statistical data.

•"English" shows a high standard deviation and variance, indicating significant variability in IMDb scores.

•Languages with less counts, like "Aboriginal" and "Bosnian," have lesser standard deviations and variances due to limited data.

•Some languages, such as "Telugu" and "Swedish," have only one movie each, resulting in a standard deviation and variance of zero.

•Overall, the total includes 3786 movies, with an average IMDb score of 6.462, a standard deviation of 1.057, and a variance of 1.118.

Languages in movies affect IMDb scores based on cultural resonance, translation quality, and audience preferences. Dominant languages like English may influence scores positively, while regional languages may connect more deeply with local audiences. Genre, cultural significance, and biases also impact language's effect on scores

Director Analysis

Row Labels	Count of Director_Name	Average of Imdb_Score
Akira Kurosawa	1	8.7
Bryan Singer	1	8.6
Charles Chaplin	1	8.6
Christopher Nolan	3	8.8
David Fincher	2	8.7
Fernando Meirelles	1	8.7
Francis Ford Coppola	2	9.1
Frank Darabont	1	9.3
George Lucas	1	8.7
Hayao Miyazaki	1	8.6
Irvin Kershner	1	8.8
Jonathan Demme	1	8.6
Lana Wachowski	1	8.7
Martin Scorsese	1	8.7
Milos Forman	1	8.7
Peter Jackson	3	8.8
Quentin Tarantino	1	8.9
Robert Zemeckis	1	8.8
Sergio Leone	1	8.9
Steven Spielberg	2	8.75
Tony Kaye	1	8.6
(blank)		
Grand Total	28	8.778571429

Percentile	9.3

Point	Column1	Rank	Percent
8	9.3	1	100.00%
7	9.1	2	95.00%
17	8.9	3	85.00%
19	8.9	3	85.00%
11	8.8	5	70.00%
16	8.8	5	70.00%
18	8.8	5	70.00%
4	8.8	8	65.00%
20	8.75	9	60.00%
1	8.7	10	25.00%
5	8.7	10	25.00%
6	8.7	10	25.00%
9	8.7	10	25.00%
13	8.7	10	25.00%
14	8.7	10	25.00%
15	8.7	10	25.00%
2	8.6	17	0.00%
3	8.6	17	0.00%
10	8.6	17	0.00%
12	8.6	17	0.00%
21	8.6	17	0.00%

RESULT:

The data provides the count of movies directed by various filmmakers and their average IMDB scores:

•Variability in directors' IMDB scores ranges from 8.6 to 9.3, with the number of movies directed ranging from 1 to 3.

•Highlights include Francis Ford Coppola leading with an average score of 9.1 for his 2 movies, followed by Christopher Nolan and Peter Jackson, each directing 3 movies with an average score of 8.8.

•A trend emerges where certain directors consistently achieve high scores, such as Coppola, Nolan, and Jackson.

This summary provides insights into the correlation between director and IMDB score, emphasizing key points. In simpler terms, if a movie has an IMDB score at the 9.3 percentile, it means it outperformed around 9.3% of other movies in the dataset in terms of IMDB score.This suggests that the movie achieved a relatively high ranking compared to the majority of other movies in the dataset.

Budget Analysis

	Correlation between movie budgets and gross earnings						Gross	Budget
						Gross	1	
						Budget	0.09656892	1
	Movies with the highest profit margin				523505847			

RESULT:

- Correlation coefficient: Approximately 0.0966.
- The data shows that there is a small connection between the amount of money a movie makes and its budget.
- There's a minor trend indicating that movies with bigger budgets tend to earn more, but it's not strong enough to accurately predict financial success.
- Factors like advertising, actors, and plot are probably more influential in determining how well a movie does at the box office.
- Correlation does not imply causation; other variables may influence both gross earnings and budget simultaneously

The connection between the terms "Gross" and "Budget" is not very strong, but there is a slight trend for movies with higher budgets to make more money at the box office. It's important to note that this correlation is not a reliable indicator of a movie's financial success, as other factors such as marketing and the cast also have a significant impact.

SUMMARY

The research goes into many aspects that affect IMDb ratings (the types of movies, the duration, language, director and money issues like budget and gross earnings. These are just some links that give us a hint (for example, the link between movie length and IMDB score). Such connections may remind one of budgets and gross earnings. However, they do not provide any definite predictions because if certain variables coincide, there might still be no causation. Examples include audience preference, cultural resonance and directing abilities. The analysis demonstrates the complexity of film assessment as well as the importance of a wide range of factors beyond sheer numbers for understanding cinematic performance in its totality.

LINKS:

- Source #1 **Excel Spreadsheet**
- Source #2 **LOOM Video**

THANK YOU