

IMDB Movie Analysis

Insights: In this data analysis project, we are using the data set provided and analyzing the movies by their categorical

Cleaning of Data:

In this process, the data is cleaned using various filtering techniques.

First of all, remove duplicate data by using the Remove Duplicate option available in Excel.

Now, by using the filter option, we will be removing blank rows from the data set.

After cleaning the data set, the number of movies dropped from 5044 to 3724.

1) Movie Genre Analysis:

Here we are using the Excel formula to separate the number of genres, categorize them into their respective sizes, and find the mean, median, and mode for descriptive statistics.

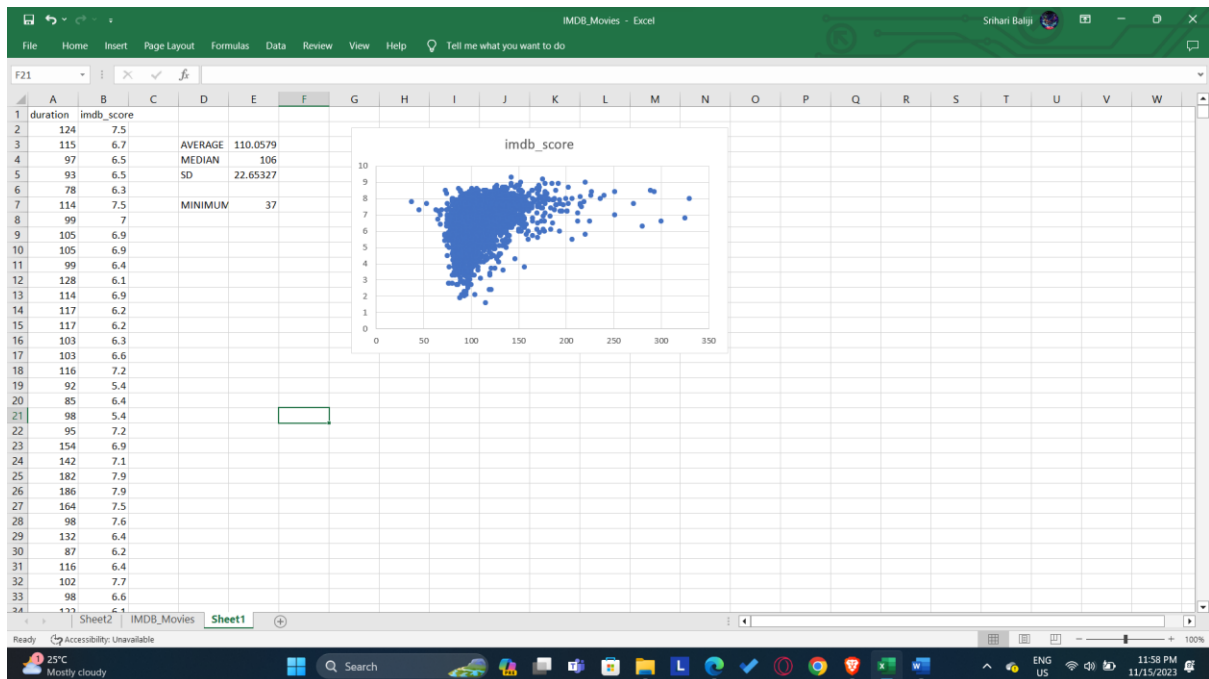
Formula used:

Used countif ,

Action	250
Adventure	254
Biography	84
Animation	94
Comedy	302
Drama	357
Document	27
Family	183
Horror	94
Mystery	109
Musical	70
Sci-Fi	149
Thriller	203
Western	40
Sport	47
Romance	192
crime	147
history	63
war	73

average	230.5
median	43
mode	3
maX	1041
MIN	2
VAR	120073.8
SD	346.5

Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score. Hear the movie duration is analysed and compared with the IMBD score and the duration of the movie



AVERAGE	110.0579
MEDIAN	106
SD	22.65327
MINIMUM	37

So the minimum is 37 minutes with a 7.8 score in IMDB
And the maximum is 330 minutes with a score of 8

Language Analysis: Examine the distribution of movies based on their language. Here by using povert table we will be separating the languages and count them

Row Labels	Count of language
Aboriginal	2
Arabic	1
Aramaic	1
Bosnian	1
Cantonese	7
Czech	1
Danish	3
Dari	2
Dutch	3
English	3621
Filipino	1
French	34
German	10
Hebrew	2
Hindi	5
Hungarian	1
Indonesian	2
Italian	7
Japanese	10
Kazakh	1
Korean	5
Mandarin	14
Maya	1
Mongolian	1
None	1
Norwegian	4
Persian	3
Portuguese	5
Romanian	1
Russian	1
Spanish	23
Thai	3
Vietnamese	1
Zulu	1
(blank)	
Grand Total	3779
AVERAGE	6.463343
MEDIAN	6.6
SD	1.053746

From the above, we can conclude that the English is the most languages used in the movies at 3621 movies out of 3779

Director Analysis: we will be taking the name of the directors and the IMDB score and doing the percentile analysis

Name of the director Average IMDB score

Name of the director	Average IMDB score
Charles Chaplin	8.60
Tony Kaye	8..60
Alfred Hitchcock	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
Asghar Farhadi	8.40
Marius A. Markevicius	8.40

hights score	8.6
rank persentail	0.89
imdb score pertantail	8.6

From the above we can tell that the movie success rate is 89 percent and the highest is 8.6 imdb score

Budget Analysis: Explore the relationship between movie budgets and their financial success. Here we will be getting the profit by subtracting gross from budget to find the profit

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	gross	budget	profit																		
2	760505847	237000000	523505847		CORRELATION	0.09999															
3	658672302	200000000	458672302																		
4	652177271	150000000	502177271		HIHG PROFIT	523505847															
5	623279547	220000000	403279547																		
6	533316061	185000000	348316061																		
7	474544677	115000000	359544677																		
8	460935665	110000000	449935665																		
9	458991599	250000000	208991599																		
10	448130642	250000000	198130642																		
11	436471036	150000000	286471036																		
12	434949459	105000000	424494959																		
13	424645577	130000000	294645577																		
14	423032628	225000000	198032628																		
15	422783777	450000000	377783777																		
16	414984497	200000000	214984497																		
17	408992272	200000000	208992272																		
18	407999255	780000000	329999255																		
19	407197282	250000000	157197282																		
20	403706375	139000000	264706375																		
21	402076689	200000000	202076689																		
22	400736600	150000000	250736600																		
23	380838870	940000000	286838870																		
24	380262555	113000000	267262555																		
25	377019252	940000000	283019252																		
26	373377893	200000000	173377893																		
27	368049635	700000000	292049635																		
28	363024263	580000000	305024263																		
29	362645141	175000000	187645141																		
30	362645141	175000000	187645141																		
31	356784000	630000000	293784000																		
32	356454367	175000000	181454367																		
33	352358779	195000000	157358779																		
34	350133553	500000000	201233553																		

The top 5 most profit movies are:

- 1)AVATAR
- 2)TITANIC
- 3)JURASSIC PARK
- 4)STAR TREK INTO THE DARKSIDE
- 5)THE DARK NIGHT

DATA SHEET LINK:

<https://docs.google.com/spreadsheets/d/1UrQ6Lv9BNqsiA-iM0RcrIQsClpDZBhzj/edit?usp=sharing&ouid=117660180828158558651&rtpof=true&sd=true>