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 Exel formula to separate the number of genera, categorize them into their respective sizes, and find the mean, median, and mode for descriptive statistics.
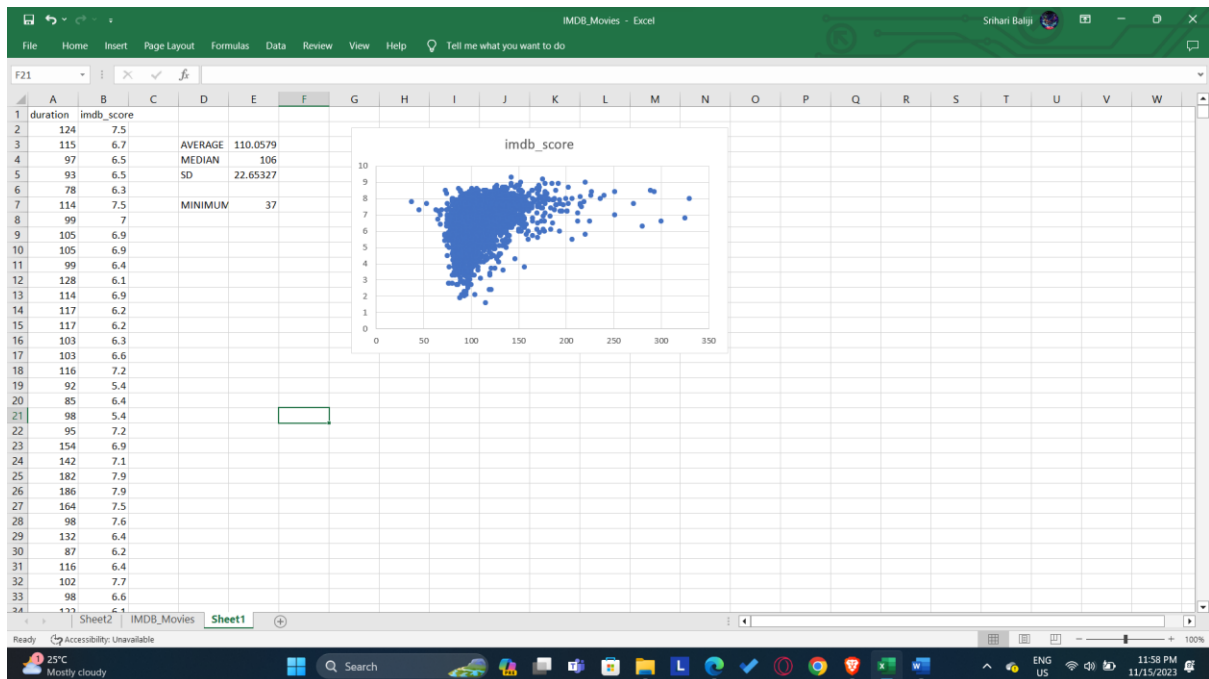
Formula used:

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

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