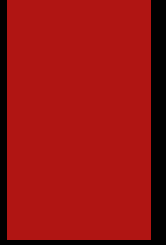# IMDB MOVIE analysis

FINAL PROJECT 1

Description: The purpose of the project is to analyze the movie related data based on genres, language, duration and so on.

Approach: The data given is not consistent and had many anomalies, in order to get the cleaned data I followed below steps

- Removed the columns that are not required. Among 28 columns 20 columns are removed.
- Removed 112 duplicate rows
- Removed 102 records where director column had blank fields
- Removed 13 records where duration column had blank fields
- Removed 10 records where language column had blank fields
- Since the records in budget and gross column had more blank fields which are greater so instead of removing the records I removed 2 outliers found in budget column and filled rest other blank fields with average values
- After data cleaning the total records : 4474 rows in Ready data sheet
- Later I calculated profit column (gross-budget) as it is required In one of the analysis

Tech Stack used: MicroSoft Excel

Insights: The insights that we can get from the project are most used genres, highest profit margin, most used language, duration analysis and so on

Result: I was able to get the data as per the project requirement. The result of each analysis is mention in further pages.

A. Movie Genre Analysis:

Task: *Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics of the IMDB scores.*
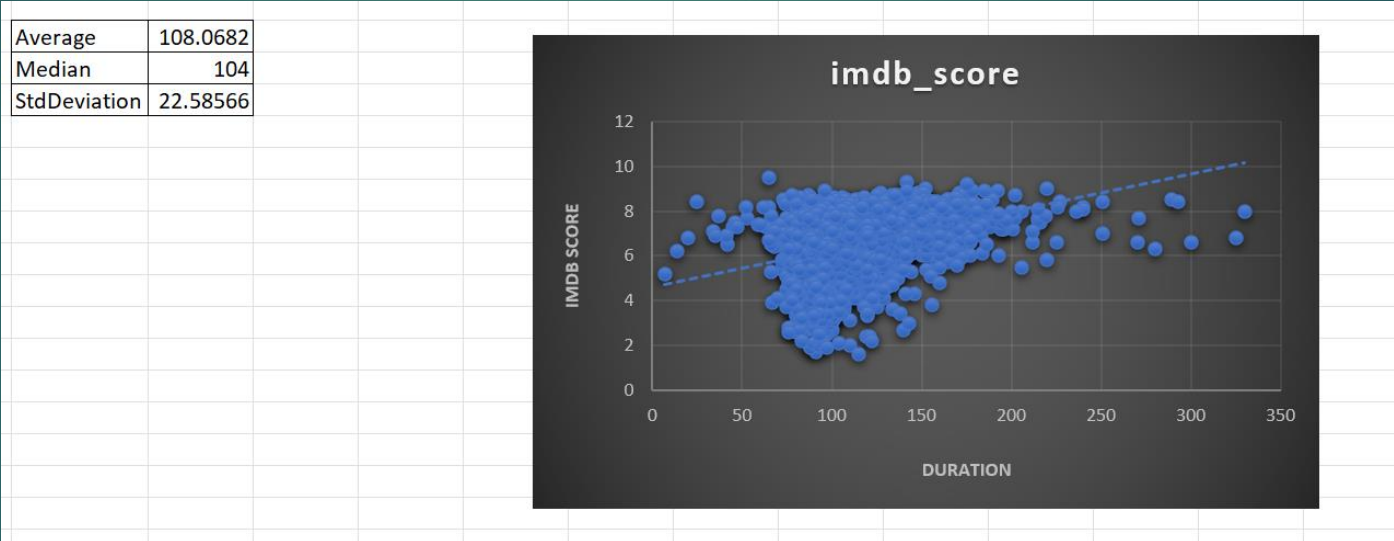
Genre	TotalCount	Median_IMDB	Mode_IMDB	Variance_IMDB	StdDeviation_IMDB	Average_IMDB	Max_IMDB	Min_IMDB
Action	1090	6.3	6.6	1.22	1.10	6.21	9	1.7
Adventure	878	6.5	6.6	1.28	1.13	6.42	8.9	1.9
Comedy	2450	6.8	6.7	0.88	0.94	6.74	9.3	2
Crime	520	6.3	6.7	1.45	1.21	6.21	8.6	1.7
Drama	232	6.7	6.7	1.30	1.14	6.54	8.6	1.7
Family	289	7.2	7	0.52	0.72	7.14	8.9	4.5
Fantasy	119	7.4	7.5	1.12	1.06	7.19	8.7	1.6
Animation	1804	6.3	6.3	1.16	1.08	6.17	9.5	1.7
Biography	834	6.6	6.6	1.01	1.00	6.52	9.3	2.4
Documentary	565	6.4	6.7	1.35	1.16	6.27	8.9	1.7
Film-noir	6	7.65	Not repeated	0.16	0.39	7.63	8.2	7.1
Horror	1060	6.5	6.5	0.96	0.98	6.44	8.6	2.1
Romance	131	6.7	7	1.50	1.22	6.50	8.5	2.1
Sci-Fi	461	6.5	6.4	1.13	1.06	6.42	8.6	2.2
Thriller	176	6.8	7.2	1.21	1.10	6.59	8.4	2
History	532	5.9	6.2	1.22	1.10	5.78	8.7	2.2
Musical	576	6.3	6.7	1.45	1.20	6.24	8.8	1.9
Mystery	1334	6.4	6.4	1.08	1.04	6.29	9	2.2
News	199	7.2	7.5	0.77	0.88	7.07	8.9	2
Short	3	7.4	Not repeated	0.18	0.42	7.53	8.1	7.1
Sport	5	6.5	Not repeated	0.45	0.67	6.38	7.1	5.2
War	204	7.1	7.1	0.76	0.87	7.04	8.6	2.7
Western	92	6.8	6.5	1.12	1.06	6.71	8.9	3.8

- Top 10 Movie genres are : Action, Adventure,Comedy,Crime,Animation,Biography,Documentry,Horror,History,Musical,Mystery
- Comedy genre I the most used among other genres and has maximum IMDB rating of 9.3
- Film-noir genre has highest IMDB rating median value : 7.65
- Fantasy and news genres have highest IMDB rating mode value :7.5
- Romance genre has highest IMDB rating Variance of value : 1.50 and Standard deviation value : 1.22
- Film-noir genre has the highest IMDB rating average value : 7.63
- Fantasy genre has lowest IMDB rating value : 1.6

B. Movie Duration Analysis:

Task: *Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score*

C. Language Analysis:

Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Language	TotalNumberOfMovies	Average_IMDBScore	Mean_IMDBScore	Median_IMDBScore	StdDeviation_IMDBScore
English	4482	6.37	6.40	6.50	1.11
Japanese	16	7.37	7.30	7.60	1.00
French	72	7.02	6.98	7.20	0.71
Mandarin	24	6.79	6.69	7.05	1.02
Aboriginal	2	6.95	6.93	6.95	0.55
Spanish	40	6.94	6.88	7.15	0.84
Filipino	1	6.70	6.70	6.70	0.00
Hindi	27	6.77	6.66	7.00	1.18
Russian	11	6.36	6.22	6.50	1.32
Maya	1	7.80	7.80	7.80	0.00
Kazakh	1	6.00	6.00	6.00	0.00
Telugu	1	8.40	8.40	8.40	0.00
Cantonese	11	6.95	6.92	7.20	0.67
German	19	7.34	7.28	7.60	0.93
Aramaic	1	7.10	7.10	7.10	0.00
Italian	10	7.08	6.98	7.15	1.14
Dutch	4	7.43	7.42	7.45	0.38
Dari	2	7.50	7.50	7.50	0.10
Hebrew	5	7.58	7.57	7.60	0.30
Chinese	3	5.67	5.65	5.70	0.45
Mongolian	1	7.30	7.30	7.30	0.00
Swedish	5	7.44	7.41	7.60	0.68
Korean	6	7.40	7.34	7.50	0.87
Thai	3	6.63	6.62	6.60	0.37
Bosnian	1	4.30	4.30	4.30	0.00
None	2	7.95	7.93	7.95	0.55
Hungarian	1	7.10	7.10	7.10	0.00
Portuguese	8	7.49	7.44	7.70	0.83
Icelandic	1	6.90	6.90	6.90	0.00
Danish	5	7.50	7.43	8.10	0.96
Arabic	5	7.38	7.34	7.40	0.79
Norwegian	4	7.15	7.13	7.30	0.50
Czech	1	7.40	7.40	7.40	0.00

- English language movies count is higher : count = 4482
- Descriptive analysis data is calculated in the excel file that is attached along with the report

D. Director Analysis:

Task: *Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations*

90th percentile	7.5	
total count >7.5 = 218	218	

- The percentile value is 7.5, there are total 218 directors with the rating greater than 7.5

E. Budget Analysis:

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Correlation	0.290646447		
Highest Profit Margin	523505847	Avatar	

- Avatar movie has the highest profit margin

Thank you