: -IMDB MOVIE ANALYSIS: -

PROJECT DESCRIPTION: The project was related to IMDB Movies. The objective of this project is to conduct a comprehensive data analysis of the IMDB Movie datasets. The analysis mainly focuses on several key aspects such as Movie genres, Duration, Language, Directors and Budget and its impact on IMDB scores and financial success of the movies.

APPROACH: Following steps are being followed

DATA CLEANING:

	DATA CLEANING STEPS
1	FULL DATASET INTO TABLE
2	REMOVING ALL BLANK ROWS
3	SEPARATED GENRES COLUMN
4	REMOVING DUPLICATES AND MAKE UNIQUE GENRE ITEMS
5	REPLACING NON PRINTABLE CHARACTER IN MOVIE TITLE COLUMN
6	REMOVING NON-NECESSARY COLUMNS FOR PROJECT

genres						
Action	Adventure	Fantasy	Sci-Fi	Fantasy	Musical	Romance
Adventure	Thriller	Thriller	Family	Thriller	Sci-Fi	Sci-Fi
Drama	Animation	Sci-Fi	Mystery	Sport	Fantasy	War
Animation	Family	Romance	Fantasy	Family	Family	Sport
Comedy	Fantasy	Comedy	History	Romance	Romance	Fantasy
Mystery	Romance	Western	Romance	Sci-Fi	Mystery	Musical
Crime	Crime	Family	Thriller	Drama	Thriller	Mystery
Biography	Comedy	Drama	Comedy	War	War	Thriller
Fantasy	Drama	Horror	Western	History	Music	
Documentary	Sci-Fi	Crime	Horror	Western	Western	
Sci-Fi	Horror	Animation	Drama	Musical	Sport	
Horror	Mystery	History	War	Mystery		
Romance	Western	Biography	Musical	Music		
Family	History	Mystery	Crime	Crime		
Western	Musical	Sport	Music	Horror		
Musical	Music	War	Sport			
Thriller	War	Musical				
History	Biography	Music				
Music	Sport	Documentary				
War	Documentary	Film-Noir				
Sport						
Film-Noir						

DATA ANALYSIS: Using Microsoft Excel various analysis on the detests are performed to get the output of the tasks given. This includes data manipulation to extract information, calculating descriptive statistic and visualizing relationship between variables

REPORT GENERATION: Reports are generated based on the result of analysis which includes insights from each analysis tasks, supported by visualizations and statistical summaries

TECH-STACK USED: Software: Microsoft Excel 2021
Purpose: Excel was chosen for its extensive data analysis functions, including pivot tables, charts, and statistical functions and visualization of the IMDB data

INSIGHTS: various key insights were uncovered through the data analytics process. Meaningful trends and patterns were observed in the data

➢ Movie genre Analysis

➢ Movie Duration Analysis

➢ Language Analysis

➢ Director Analysis

➢ Budget Analysis

A. **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.

Genre Statistics

	Action	Adventure	Drama	Animation	Comedy	Mystery	Crime	Biography	Fantasy	Documentary	Sci-Fi	Horror	Romance	Family	Western	Musical	Thriller	History	Music	War	Sport	Film-Noir
MEAN	6.3	6.4	6.8	6.7	6.2	6.5	6.5	7.2	6.3	7	6.3	5.9	6.4	6.2	6.8	6.6	6.4	7.2	6.5	7.1	6.6	7.7
MODE	6.6	6.6	6.7	6.7	6.7	6.6	6.6	7	6.7	6.6	6.7	5.9	6.5	5.4	6	7.1	6.5	7.7	6.2	7.1	7.2	
MEDIAN	6.3	6.6	6.9	6.8	6.3	6.5	6.6	7.2	6.4	7.4	6.4	6	6.5	6.3	6.8	6.75	6.4	7.2	6.6	7.1	6.8	7.7
MAX	9	8.9	9.3	8.6	8.8	8.6	9.3	8.9	8.9	8.5	8.8	8.6	8.5	8.6	8.9	8.5	9	8.9	8.5	8.6	8.3	7.7
MIN	2.1	2.3	2.1	2.8	1.9	3.1	2.4	4.5	2.2	1.6	1.9	2.3	2.1	1.9	4.7	2.1	2.7	5.5	1.6	4.3	2	7.7
STD DEV	1.04	1.11	0.9	0.99	1.04	1.02	0.98	0.69	1.13	1.38	1.16	1	0.95	1.16	0.93	1.1	0.97	0.67	1.18	0.8	1.04	
VAR	1.08	1.24	0.8	0.98	1.07	1.03	0.97	0.48	1.29	1.92	1.35	1	0.91	1.35	0.87	1.21	0.94	0.45	1.4	0.64	1.09	

Genre Data

Action	Adventure	Drama	Animation	Comedy	Mystery	Crime	Biography	Fantasy	Document	Sci-Fi	Horror	Romance	Family	Western	Musical	Thriller	History	Music	War	Sport	Film-Noir
7.9	7.9	6.7	7.8	7.8	7.5	7.2	5.5	7.9	7.8	7.9	7	6.2	7.8	6.5	7.8	6.8	6.7	7.8	6.1	6.3	7.7
7.1	7.1	7.2	7.3	6.8	7.5	6.2	7.5	7.1	7.3	6.6	5.2	7.8	7.5	4.8	7.6	8.5	6.1	7.6	5.5	6.5	
6.8	6.8	7.7	6.3	7.3	6.5	9	6.8	7.8	7.1	7.5	7.2	7.2	6.6	8.1	5.9	5.9	5.5	5.9	7.7	6.1	
8.5	6.6	7.3	8.3	6.3	7.5	6.7	7	7.5	1.6	6.9	6.2	7.7	6.8	7.2	6.4	7	7.7	6.4	7.4	7.1	
6.6	6.2	6.8	7.2	8.3	7.6	6.1	7.8	7.3	5.9	6.1	5.8	7.3	6.1	5.4	7.1	7.8	7.4	5.4	6.6	6.8	
6.2	7.8	6.6	8.4	7.2	6.7	6.6	8.2	7.2	4.1	7.2	5.7	7.3	6.5	8.5	6.9	6.8	6.6	7.1	6	5.3	
7.5	7.5	9	6.8	6.2	6.1	6.6	7	6.6	8	8.1	6.6	6.6	7.3	6	7.5	7.2	7.1	6.9	7.1	6.8	
6.9	7.5	7.5	8.3	8.3	7.7	6.4	8	6.7	5.4	6.8	4.9	6.6	6.4	7	7	7	7.2	7.5	6.2	8	
6.1	6.9	8.3	6.5	6.5	8.1	7.3	8.2	6.8	7.5	8.2	5.7	7	6.3	5.9	4.4	8	7	5	7.2	7.3	
6.7	6.1	7.8	8.3	4.8	6.4	7.5	7.1	7.5	6.6	5.9	7	7.8	8.3	6.7	7.1	7.5	7	7	7.2	6.4	
7.3	6.7	6.1	6.4	6.9	7.4	4.8	8.4	7	2.7	7	4.9	6.1	7.2	6.7	5.8	6.2	7.6	4.4	6.9	6.3	
6.5	7.3	7.6	7.9	5.4	7.8	6.2	7	7.9	8.5	7.2	4.2	5.5	8.4	6.6	8.5	9	7.7	6.5	7.6	6.6	
7.2	6.5	6.3	7.8	8.3	6.6	6.4	7.9	6.1	7.6	6.8	5.1	6.6	6.8	7.9	6.7	5.2	6	7.1	7.7	4.5	
6.6	7.2	7.9	6.6	6.4	7.5	6.5	7.4	7.3	6.6	6	5.6	7.3	6.9	6.5	6.5	6.1	7.1	4.4	6	7.2	
8.1	6.6	8.6	8.2	7.9	7	5.8	7.2	6.5	8	5.7	6.2	6.4	8.3	6.6	5.9	5.8	6.7	7.3	7.5	3	
6.7	8.1	7.8	6.1	7.8	7	3.7	7.1	6.8	5.4	6.7	6.3	5.5	6.5	7.8	7.6	8.8	7.3	5.8	7.1	7.2	
6.8	6.7	5.6	8	6.1	5.8	6.5	6.6	7.3	7.2	6.8	5.2	5.5	7.5	4.7	5.2	6.7	7.6	8.5	6.7	7.1	
7	6.8	6.1	6.7	6.7	7.1	7	7	6.4	6.6	5.6	6.4	6.6	5.4	5.8	7	5.6	7.7	6.7	7.6	7.1	
6.7	7.5	5.5	7.6	8	7.7	7.8	6.6	6.7	5.1	6.6	5.9	5.3	8.3	6.8	6.6	6.7	7.2	6.5	7.2	3.8	
7.2	7	6.4	6.9	6.7	7.4	6.4	7.4	8.3	6.7	7	5.5	6.3	7.8	6.3	6.3	8.1	7.3	5.7	7.6	6.3	
8.2	6.7	7.2	5.1	5.9	6.1	6.7	6.7	8	7.7	8	5.9	6.5	7	7.4	7.4	6.7	5.6	4.5	6.6	5.9	
5.9	7.9	6.9	6.2	7.6	6.9	8.2	7.3	6.3	7.6	7.8	5.4	3.7	6.4	6.1	6.4	7.4	7.1	5.9	8.4	6.8	
7	6.1	5.8	7.3	6.9	5.7	3.3	7.9	6.6	7.1	7	6.8	7.1	6.5	5.2	6.4	5.8	7.6	7.6	7.1	7.4	
7.8	7.2	6.6	5.4	5.1	7.3	5.3	8	6.2	8	6.3	6.3	6.9	7.9	7.7	7.1	6.9	6.7	5.2	8.3	6.7	
7.3	8.2	6.4	6.7	6.2	6.3	7.5	6.3	7.2	8.1	7.5	5.7	4.4	7.8	6	7.6	6.4	8.4	7	7.1	6	
7.2	5.9	7.7	6.9	6.2	7.3	6.6	6.9	6.8	6.8	8.4	5.7	8.5	6.6	6.8	6.3	6.1	7.1	7.3	7.2	5.8	
6.8	7	6.1	6.9	7.3	6.6	5.5	7	6.9	7.8	5.8	6.7	7	8.2	6.4	6	6.6	7	6.7	6.1	5.8	
6	7.8	7.7	6.3	5.4	7.8	8.5	8	5.2	7.4	5.4	4.8	5.5	6.1	7.5	7.1	7.4	7.2	7.9	6.3	5.9	
5.7	7.3	7.4	7.2	6.7	7.6	7.8	7.4	5.4	7.7	7.3	6	6.9	6.4	7.5	7.4	5	7.4	5.9	7	7.3	
6.7	7.2	5.5	7.3	6.9	4.9	7.6	8.2	8.3	8.2	6.5	5.6	6	6.1	6	6.4	6.6	7.6	5.2	6.3	7.6	
6.8	6.5	8.1	7.9	6.9	7.4	6.4	6.7	7.8	7.7	6.4	3.3	6.3	7.5	7.5	5.8	6	6.6	5.4	4.7		
5.6	6.8	5.5	6.4	5.5	6.4	6.9	7.3	6.1	7.3	4.8	5.9	7.6	7.5	7.2	6.9	6	7.1	6.3	8.6	6.5	
6.6	7.3	5.8	5.9	6.1	6.9	7	7.6	7	7.6	6.9	6.4	4.3	7.6	7.8	7.2	6.4	6.1	7.4	7.4	6.5	
7.2	6	6.6	6.5	7.2	7.8	6.1	6.7	6.3	8.3	7.9	6.4	6.9	6.7	5.8	7	8.1	7.4	6.4	7.2	7.2	
7	5.7	6.4	7.2	6.1	6.8	6.7	6.4	6.4	8.1	6.1	6.8	7	8	6	6.2	8.1	6.6	6.4	6.6	6.3	

genres / imdb_score

genres	imdb_score	genres
Action	7.9	Action|Adventure|Fantasy|Sci-Fi
Adventure	7.1	Action|Adventure|Fantasy
Drama	6.8	Action|Adventure|Thriller
Animation	8.5	Action|Thriller
Comedy	6.6	Action|Adventure|Sci-Fi
Mystery	6.2	Action|Adventure|Romance
Crime	7.8	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romance
Biography	7.5	Action|Adventure|Sci-Fi
Fantasy	7.5	Adventure|Family|Fantasy|Mystery
Documentary	6.9	Action|Adventure|Sci-Fi
Sci-Fi	6.1	Action|Adventure|Sci-Fi
Horror	6.7	Action|Adventure
Romance	7.3	Action|Adventure|Fantasy
Family	6.5	Action|Adventure|Western
Western	7.2	Action|Adventure|Fantasy|Sci-Fi
Musical	6.6	Action|Adventure|Family|Fantasy
Thriller	8.1	Action|Adventure|Sci-Fi
History	6.7	Action|Adventure|Fantasy
Music	6.8	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi
War	7.5	Adventure|Fantasy
Sport	7	Action|Adventure|Fantasy
Film-Noir	6.7	Action|Adventure|Drama|History
	7.9	Adventure|Fantasy
	6.1	Adventure|Family|Fantasy

B. 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.

MOVIE DURATION	
MEAN	110.26
MEDIA	106.00
STD	22.65

Duration VS imdb_score
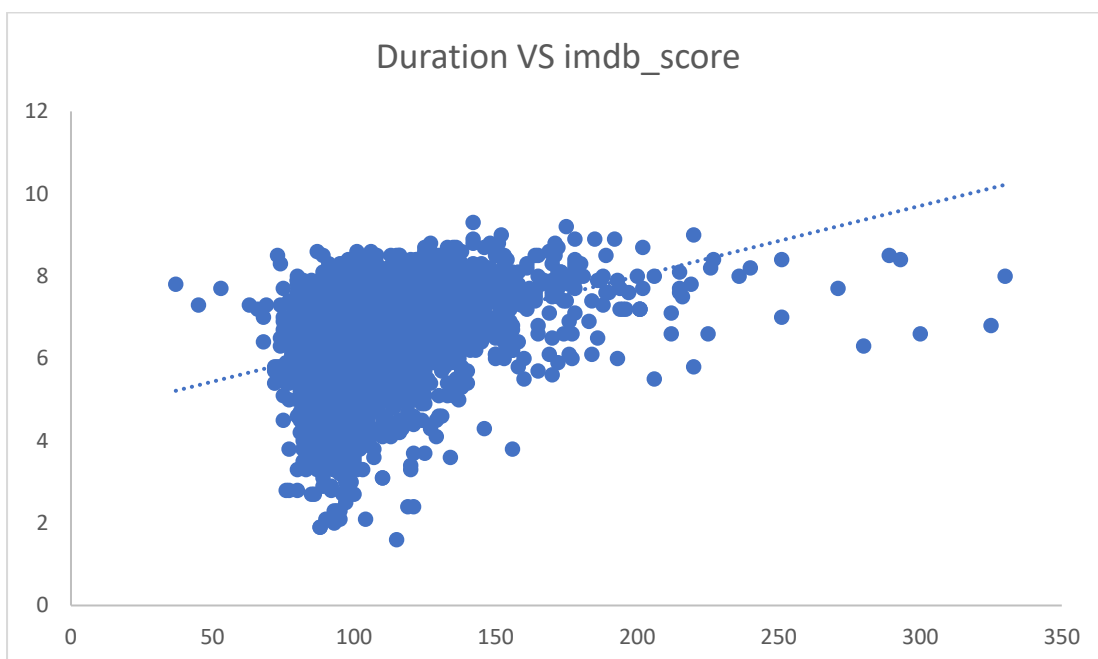
duration	imdb_score
178	7.9
169	7.1
148	6.8
164	8.5
132	6.6
156	6.2
100	7.8
141	7.5
153	7.5
183	6.9
169	6.1
106	6.7
151	7.3
150	6.5
143	7.2
150	6.6
173	8.1
136	6.7
106	6.8
164	7.5
153	7
156	6.7
186	7.9
113	6.1

C. 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.

Top 5 most common languages	
Language	**No of Movies**
English	3598
French	34
Spanish	23
Mandarin	15
Japanese	10
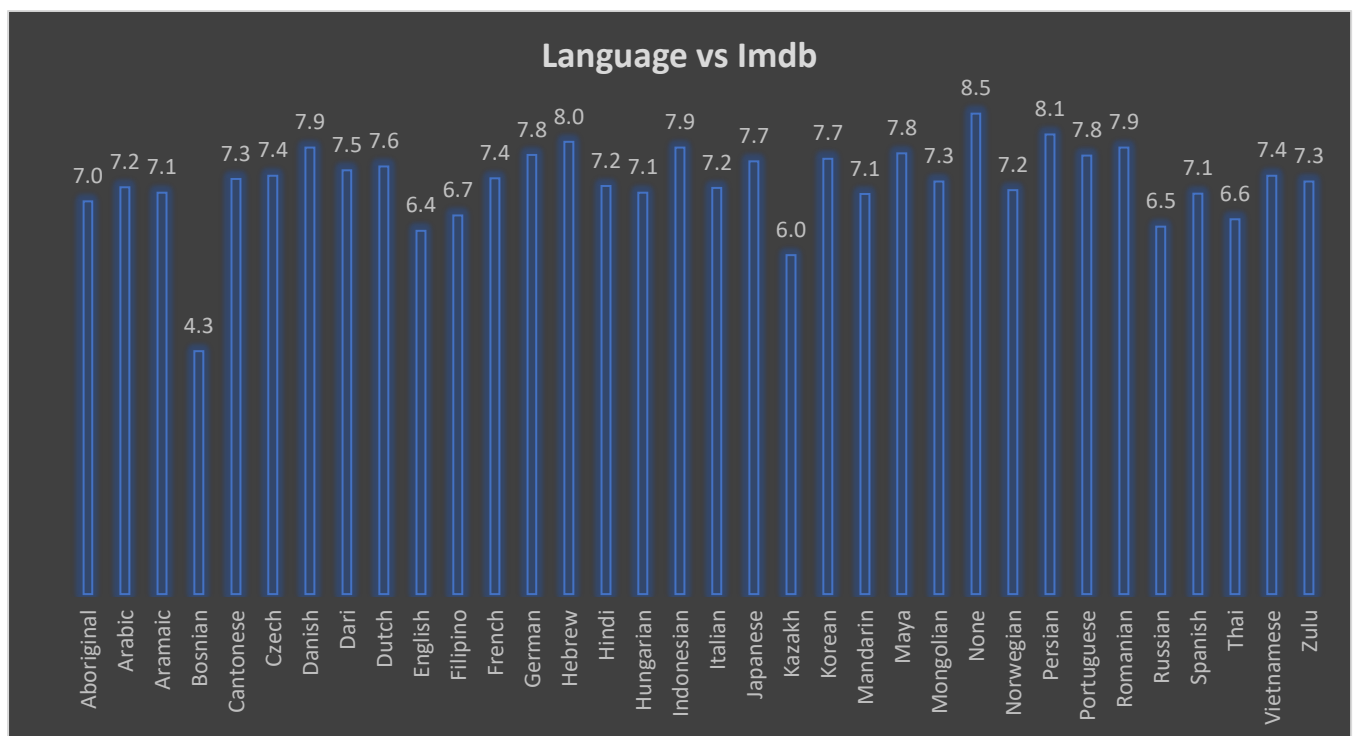
Language	Average of imdb_score
Aboriginal	7.0
Arabic	7.2
Aramaic	7.1
Bosnian	4.3
Cantonese	7.3
Czech	7.4
Danish	7.9
Dari	7.5
Dutch	7.6
English	6.4
Filipino	6.7
French	7.4
German	7.8
Hebrew	8.0
Hindi	7.2
Hungarian	7.1
Indonesian	7.9
Italian	7.2
Japanese	7.7
Kazakh	6.0
Korean	7.7
Mandarin	7.1
Maya	7.8
Mongolian	7.3
None	8.5
Norwegian	7.2
Persian	8.1
Portuguese	7.8
Romanian	7.9
Russian	6.5
Spanish	7.1
Thai	6.6
Vietnamese	7.4
Zulu	7.3

D. Director Analysis: Influence of directors on movie ratings.

Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations.
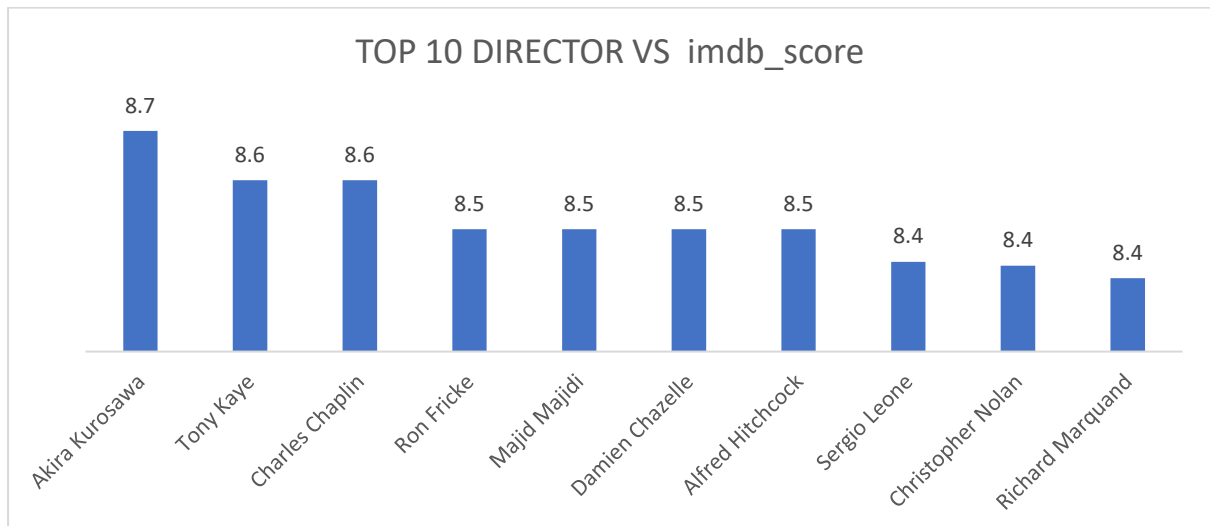
director_name	Average of imdb_score	percentaile
Akira Kurosawa	8.7	99.40%
Tony Kaye	8.6	99.20%
Charles Chaplin	8.6	99.20%
Ron Fricke	8.5	98.70%
Majid Majidi	8.5	98.70%
Damien Chazelle	8.5	98.70%
Alfred Hitchcock	8.5	98.70%
Sergio Leone	8.4	98.70%
Christopher Nolan	8.4	98.70%
Richard Marquand	8.4	98.30%

E. 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.

gross	PROFIT MARGIN	CORRELATION	MAX	MIN
760505847	523505847	0.10	523505847	-12213298588
309404152	9404152			
200074175	-44925825			
448130642	198130642			
73058679	-190641321			
336530303	78530303			
200807262	-59192738			
458991599	208991599			
301956980	51956980			
330249062	80249062			
200069408	-8930592			
168368427	-31631573			
423032628	198032628			
89289910	-125710090			
291021565	66021565			
141614023	-83385977			
623279547	403279547			
241063875	-8936125			
179020854	-45979146			
255108370	5108370			
262030663	32030663			
105219735	-94780265			
258355354	33355354			
70083519	-109916481			

E) BUDGET ANALYSIS

budget	gross	PROFIT MARGIN
237000000	760505847	523505847
220000000	623279547	403279547
200000000	658672302	458672302
150000000	652177271	502177271
185000000	533316061	348316061
115000000	474544677	359544677
78000000	407999255	329999255
45000000	422783777	377783777
220000000	623279547	403279547
11000000	460935665	449935665
10500000	434949459	424449459

MOVIES WITH HIGHEST PROFIT MARGIN

MOVIE DURATION ANALYSIS | LANGUAGE ANALYSIS | DIRECTOR ANALYSIS | BUDGET ANALYSIS

EXCEL REFER LINK: – IMDB –PROJECT 5.xlsx