IMDB MOVIE ANALYSIS

A) PROJECT DESCRIPTION:

- Dataset having 28 columns,5044 rows of different IMDB Movies.
- Defined a problem, clean the data as necessary remove null and duplicate values.
- Explore the dataset and derive insight.
- Use root cause analysis five why's approach.
- Create chart pivot table and graphs to answer given questions.
- Analyze answer to given question and make a report for data driven decision.

B) PROJECT APPROACH:

- ❖ Download all Data provided
- ❖ Understanding the data
- ❖ Find duplicate and null values
- ❖ Data processing and solving the asked problems
- ❖ Create charts for easy representation

C) TECH STACK USED:

The tech stack used included:

- ❖ Microsoft excel

D) PROJECT INSIGHT:

I. Movie Genre Analysis: Determine the most common genres of movies in the dataset.
SOLUTION: PIVOT TABLE > GENRES, COUNT OF GENRES>

AFTER THAT CALCULATED:

AVERAGE: =AVERAGE(E4:E748)

MEDIAN =MEDIAN(E4:E748)

MODE =MODE(E4:E748)

MAXIMUM=MAX(E4:E748)

MINIMUM =MIN(E4:E748)

VARIANCE=VAR(E4:E748)

AVERAGE	4.997315
MEDIAN	1
MODE	1
MAX	147
MIN	1
VAR	188.0269
SD	13.71229

STANDARD DEVIATION=STDEV.S(E4:E748)

GENRES	Count of genres
Comedy|Drama|Romance	147
Drama	140
Comedy|Drama	137
Comedy	137
Comedy|Romance	131
Drama|Romance	114
Crime|Drama|Thriller	81
Action|Crime|Thriller	55
Action|Crime|Drama|Thriller	50
Action|Adventure|Sci-Fi	46
Comedy|Crime	45
Action|Adventure|Thriller	45
Horror	43
Drama|Thriller	42

II. Movie Duration Analysis: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

SOLUTION:

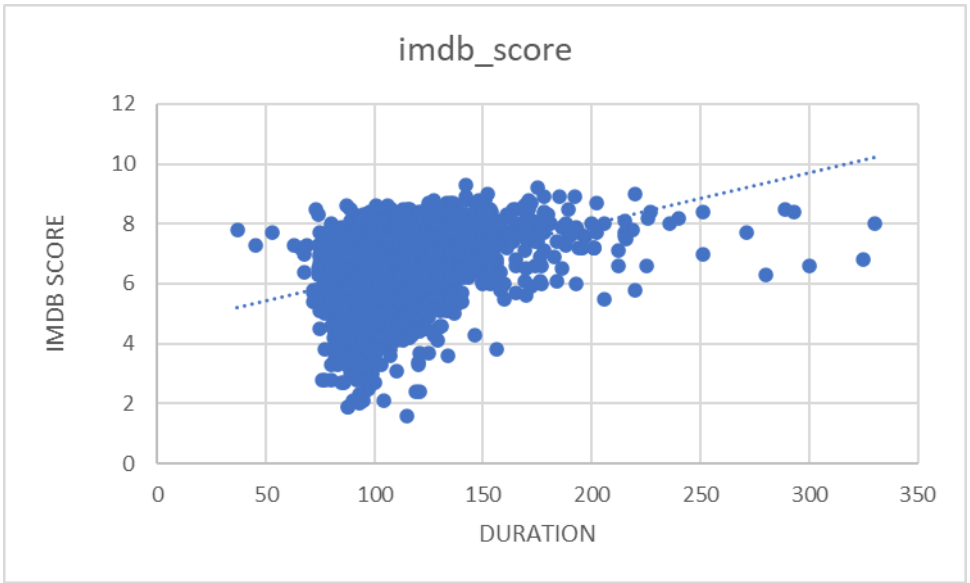
AFTER THAT CALCULATED:

AVERAGE =AVERAGE(A:A)

MEDIAN=MEDIAN(A:A)

STANDARD DEVIATION=STDEV.S(A:A)

AVG	MEDIAN	SD
110.2635	106	22.67832



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

SOLUTION: PIVOT TABLE > LANGUAGE ,COUNT OF LANGUAGE>

The most common language used in the movies is English.

AFTER THAT CALCULATE

Language	Count of language
Aboriginal	2
Arabic	1
Aramaic	1
Bosnian	1
Cantonese	7
Czech	1
Danish	3
Dari	2
Dutch	3
English	3566
Filipino	1
French	34
German	10
Hebrew	1
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
Grand Total	**3723**

CALCULATIOOF MEAN, MEADIAN AND SD

LANGUAGE	MEAN	MEDIAN	S.D
Aboriginal	6.95	6.6	1.053644
Arabic	7.2	6.6	1.053644
Aramaic	7.1	6.6	1.053644
Bosnian	4.3	6.6	1.053644
Cantonese	7.342857143	6.6	1.053644
Czech	7.4	6.6	1.053644
Danish	7.9	6.6	1.053644
Dari	7.5	6.6	1.053644
Dutch	7.566666667	6.6	1.053644
English	6.95	6.6	1.053644
Filipino	6.7	6.6	1.053644
French	7.355882353	6.6	1.053644
German	7.77	6.6	1.053644
Hebrew	8	6.6	1.053644
Hindi	7.22	6.6	1.053644
Hungarian	7.1	6.6	1.053644
Indonesian	7.9	6.6	1.053644
Italian	7.185714286	6.6	1.053644
Japanese	7.66	6.6	1.053644
Kazakh	6	6.6	1.053644
Korean	7.7	6.6	1.053644
Mandarin	7.021428571	6.6	1.053644
Maya	7.8	6.6	1.053644
Mongolian	7.3	6.6	1.053644
None	8.5	6.6	1.053644
Norwegian	7.15	6.6	1.053644
Persian	8.133333333	6.6	1.053644
Portuguese	7.76	6.6	1.053644
Romanian	7.9	6.6	1.053644
Russian	6.5	6.6	1.053644
Spanish	7.082608696	6.6	1.053644
Thai	6.633333333	6.6	1.053644
Vietnamese	7.4	6.6	1.053644
Zulu	7.3	6.6	1.053644

IV. Director Analysis: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

SOLUTIONS: PIVOT TABLE> DIRECTOR NAME > AVG OF IMBD SCORE>

AFTER THAT CALCULATE:

LARGE =LARGE(K4:K13,1)

PERCENT=PERCENTRANK(K4:K13,8.7)

PERCENTILE=PERCENTILE(K4:K13,K18)

LARGE	8.7
PERCENT	1
PERCENTILE	8.7

top 10 Director name	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4

V. Budget Analysis: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

SOLUTION: CALCULATED

PROFIT=(GROSS – BUDGET)

AFTER THAT CALCULATED

CORRELATION=CORREL(A:A,B:B)

HIGH PROFIT MARGIN=MAX(C:C)

CORRELATION	0.098318102
HIGH PROFIT MARGIN	523505847

Vi. Result: Through this project, I have achieved a better understanding of data analysis and visualization techniques using Excel. It has helped me develop skills in cleaning and organizing data, identifying profitable movies, recognizing top directors, and exploring popular genres. By working on this project, I have gained practical experience in analyzing real-world data, extracting valuable insights, and presenting information effectively through charts and tables. It has improved my overall knowledge and proficiency in data analysis.

HYPERLINK: excel sheet solution link