PROJECT 5

IMDB MOVIE ANALYSIS

TECH-STACK USED

Microsoft Excel

A MADHAVA VARMA

ANALYSIS ON

A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

C. Language Analysis: Examine the distribution of movies based on their language.

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

E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

A. MOVIE GENRE ANALYSIS:

Objective: Analyze the distribution of movie genres and their impact on the IMDB score.
Your 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.

Formulae:-

To count : =COUNTIF('cleaned data'!E$2:$E$3849, K2)
Mean : =AVERAGE(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849))
Median: =MEDIAN(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849))
Mode: =MODE(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849)) **Max:** =MAX(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849))
Min: =MIN(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849)) **Variance:** =VAR(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849)) **Standard Deviation:** =STDEV.S(IF('cleaned data'!$E$2:$E$3849=A2, 'cleaned data'!$N$2:$N$3849))

A. MOVIE GENRE ANALYSIS:

RESULT

Most common genres are:-								
genres	Count	Mean	Median	Mode	Max	Min	Variance	Standard Deviation
Drama	153	7.041830065	7.2	7.3	8.8	3.4	0.687054524	0.828887522
Comedy|Drama|Romance	151	6.494701987	6.5	6.5	8	4.3	0.562771744	0.750181141
Comedy|Drama	147	6.583673469	6.7	6.7	8.8	3.3	0.734800112	0.857204825
Comedy	145	5.840689655	6	6.5	8	1.9	1.481874521	1.217322686
Comedy|Romance	135	5.896296296	6	6.1	8.4	2.7	0.768269762	0.87650999

B. MOVIE DURATION ANALYSIS:

Objective: Analyze the distribution of movie durations and its impact on the IMDB score.

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

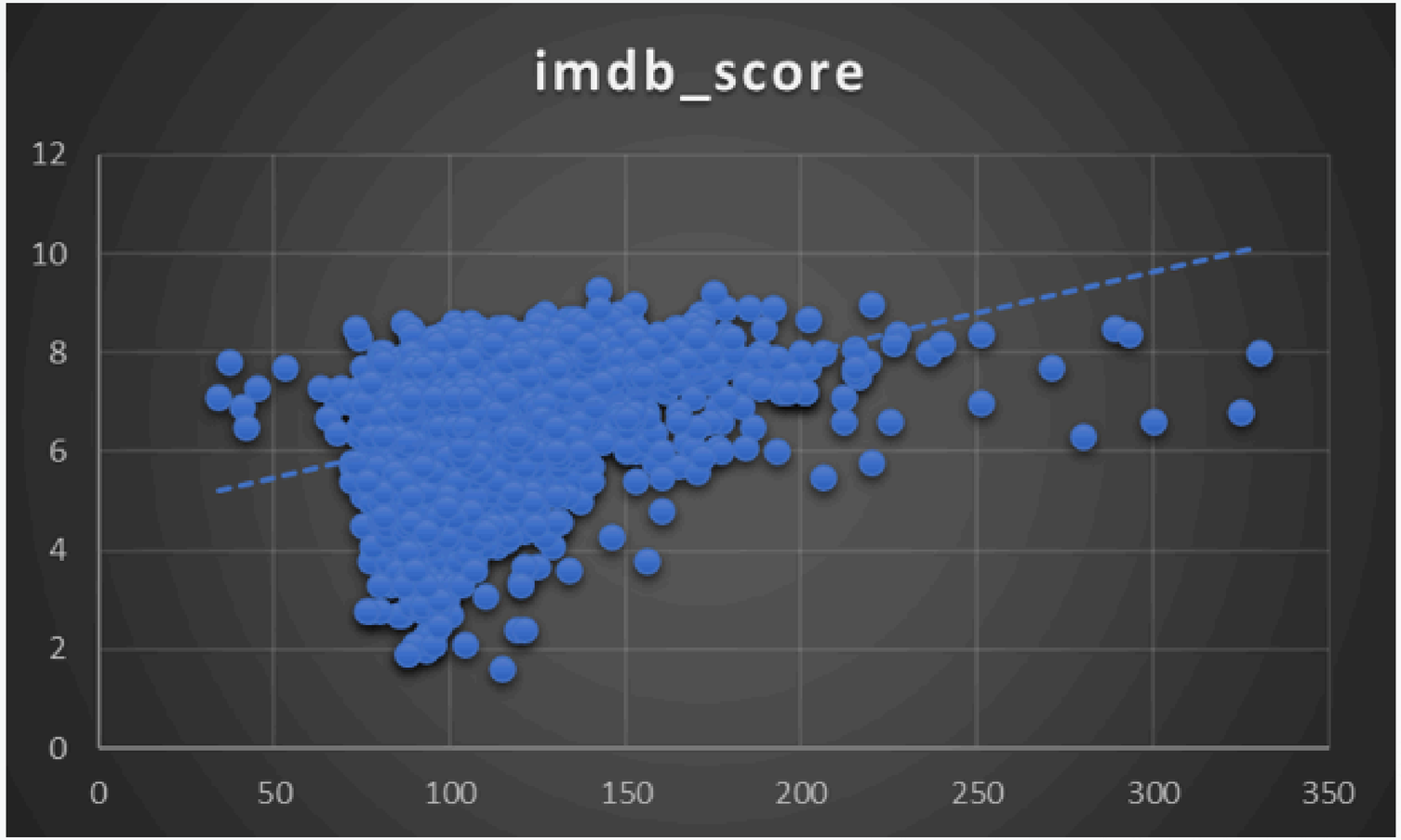
Formula:

Mean: =AVERAGE(A:A)
Median: =MEDIAN(A:A)
Standard deviation: =STDEV.S(A:A)

B. MOVIE DURATION ANALYSIS:

RESULT:

Average	109.9241164
Median	106
Standard Deviation	22.75364979



C. LANGUAGE ANALYSIS:

Objective: Examine the distribution of movies based on their language.
Your Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Formulae:

Count: =COUNTIFS('cleaned data'!$J$2:$J$3849, J2)

Mean: =AVERAGE(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

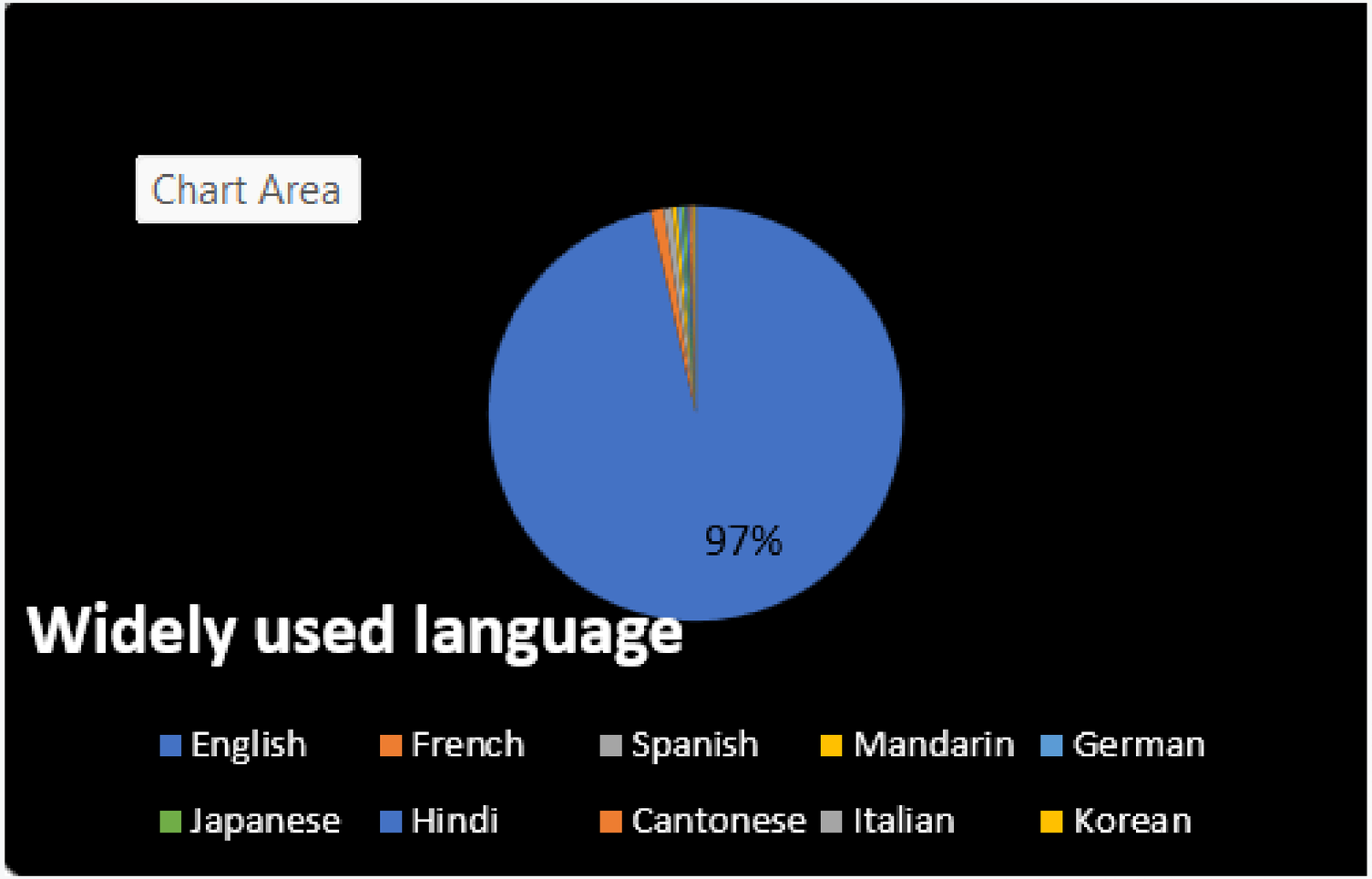
Median: =MEDIAN(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

Standard Deviation: =STDEV.S(IF('cleaned data'!$J$2:$J$3849=J2, 'cleaned data'!$N$2:$N$3849))

C. LANGUAGE ANALYSIS:

RESULT:

Most common Languages are:-				
Language	**Count**	**Mean**	**Median**	**Standard Deviation**
English	3668	6.42391	6.5	1.048750752
French	37	7.28649	7.2	0.561328861
Spanish	26	7.05	7.15	0.826196103
Mandarin	14	7.02143	7.25	0.765786244
German	13	7.69231	7.7	0.640912811
Japanese	12	7.625	7.8	0.899621132
Hindi	10	6.76	7.05	1.111755369
Cantonese	8	7.2375	7.3	0.440575922
Italian	7	7.18571	7	1.155318962
Korean	5	7.7	7.7	0.570087713



D. DIRECTOR ANALYSIS:

Objective: Influence of directors on movie ratings.
Your Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Formulae:

Average: =AVERAGE(IF('cleaned data'!$A$2:$A$3849=A2, 'cleaned data'!$N$2:$N$3849))
Percentile: =PERCENTILE(H2:H11, H15)

D. DIRECTOR ANALYSIS:

RESULT:

director_name	Average
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

E. BUDGET ANALYSIS:

Objective: Explore the relationship between movie budgets and their financial success.
Your Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Procedure:

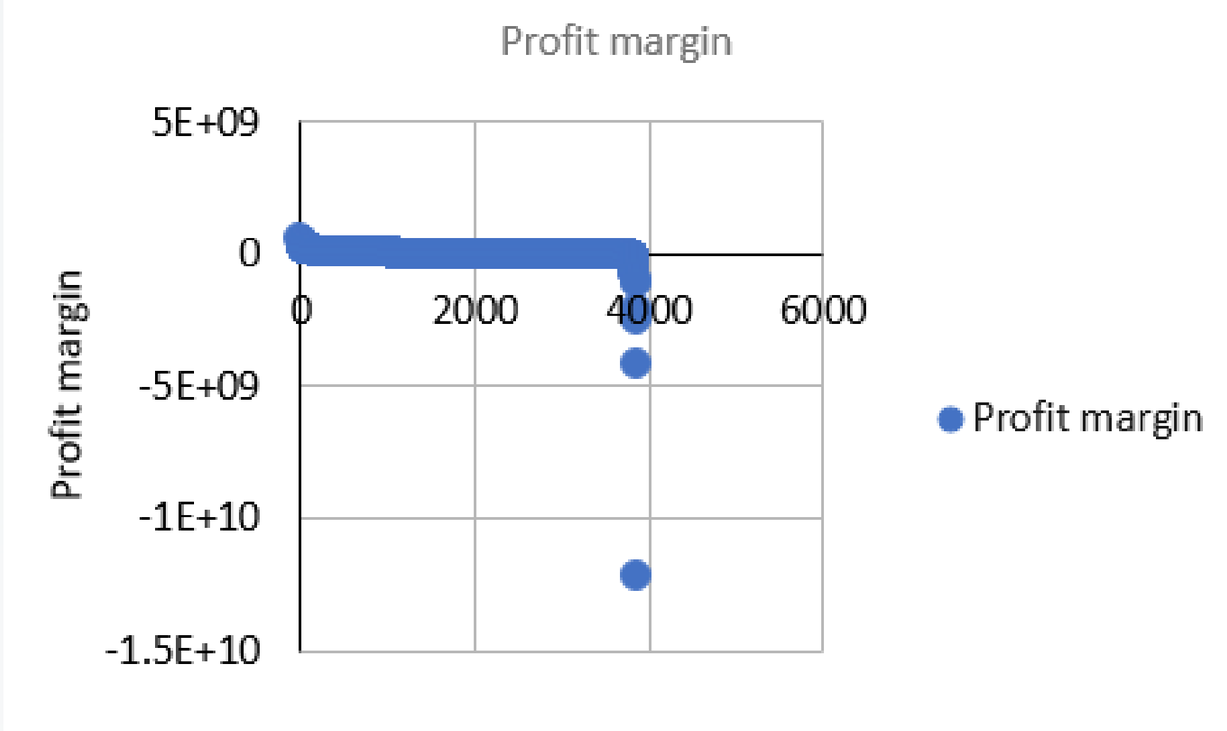
1. First calculate the profit margin for each movie by subtracting budget value from gross value.
2. Then use **CORREL** function to calculate correlation coefficients between movie budgets and gross earnings.
3. Use **MAX** function to get highest profit margin.
4: Use:
=INDEX(B2:B3849, MATCH(1,IF(D2:D3849=G11, 1),0)) to get the title of the movie.

E. BUDGET ANALYSIS:

RESULT:

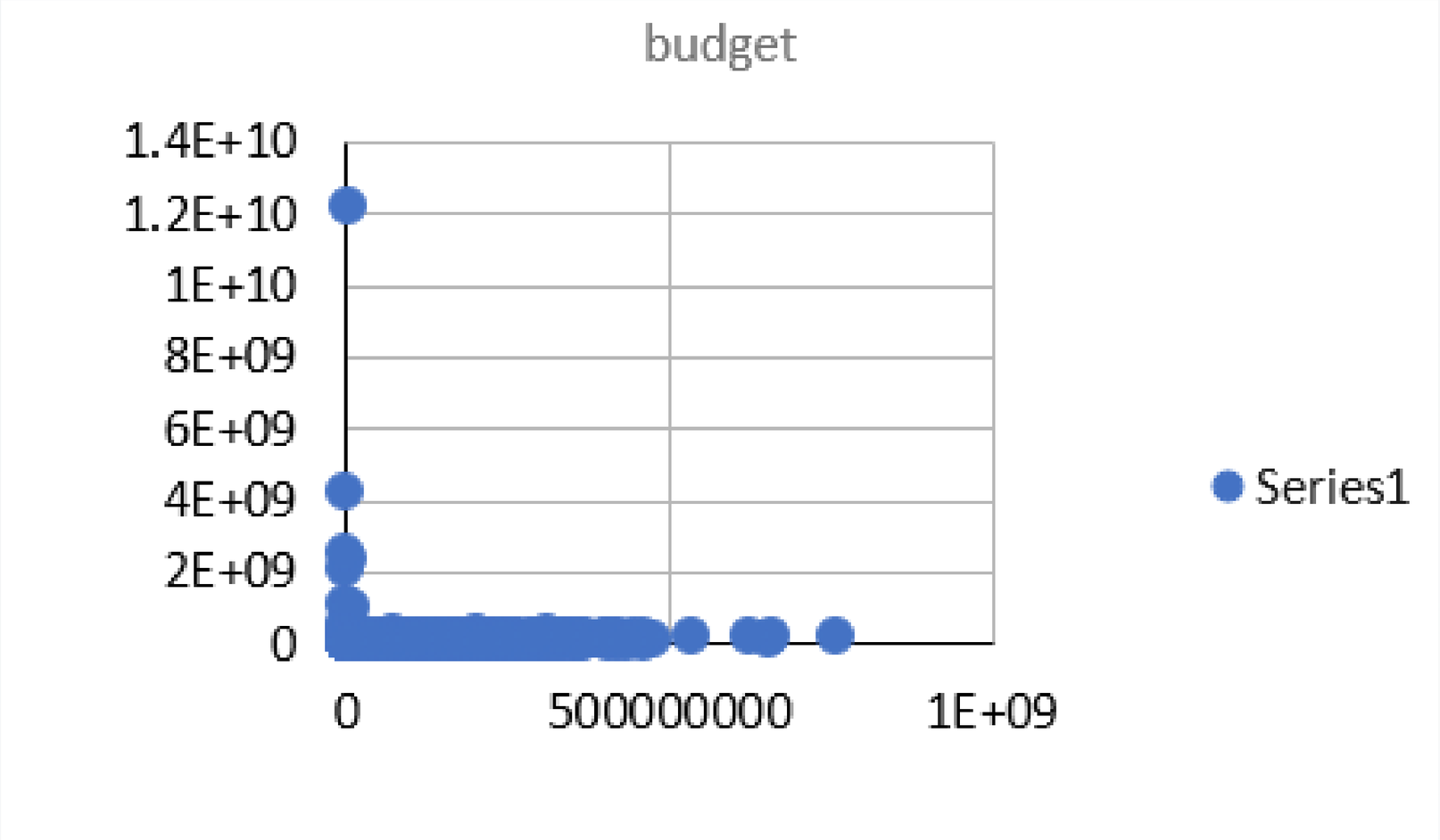
CORRELATION
0.100850218

MAX PROFIT	MOVIE TITLE
523505847	AvatarÂ



E. BUDGET ANALYSIS:

RESULT:



EXCEL SHEET LINK

The drive link for the excel sheet is:

https://docs.google.com/spreadsheets/d/1yePbc0Q-fDm96HRqq39FV3sKnuz0_BhI/edit?usp=drive_link&ouid=110721591021029757514&rtpof=true&sd=true

Working on this project helped me understand and use Microsoft Excel more better. It helped me gain experience on handling with graphs, charts, how to use statistics effectively and how Data Analytics is implemented using it in the real world and obtain insights with the data provided as a Data Analyst. It also helped me gain experience in handling large sets of data.

THANK YOU.