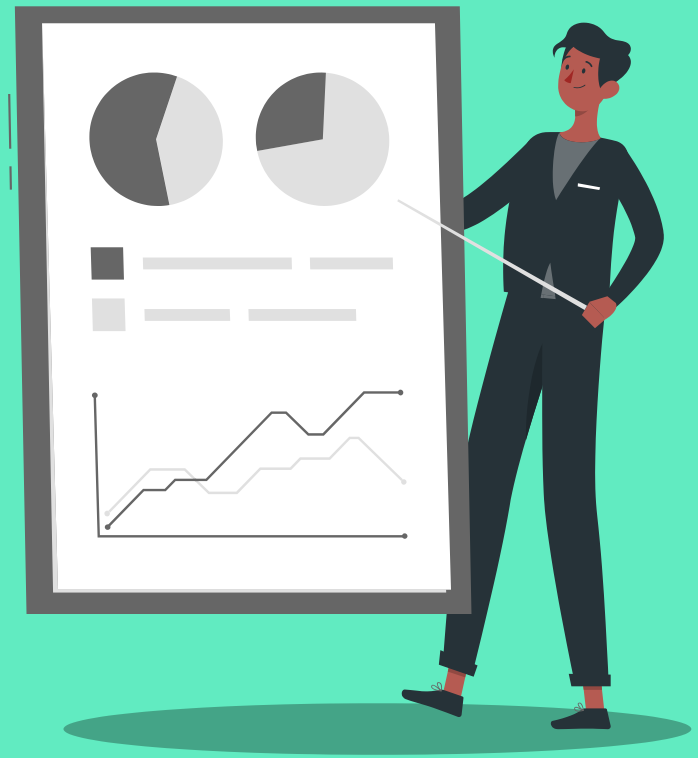


IMDB Movie Analysis Report

By Yash Shinde

Excel Sheet:- Click Here to Download

Video Presentation:- Click Here to Watch Video Presentation

Project Description!

In this project, we analyze an IMDB Movies dataset to uncover the factors that lead to high IMDB ratings, defining what makes a movie successful. Our findings will provide valuable insights for movie professionals, helping producers, directors, and investors make informed decisions in their future projects, ultimately enhancing the quality and appeal of their films.

Table of Contents

"Through a Lens of Data: Illuminating the Factors Behind Cinematic Excellence with IMDB, Empowering Industry Stakeholders to Shape Tomorrow's Blockbusters."

A Movie Genre Analysis

for each genre, calculated descriptive statistics

Movie Genre Analysis

- First I created a new sheet and copied IMDb rating and genre column.
- Then I used “Text-to-Columns” feature to make a separate column for the genre.
- Then by Using the “UNIQUE” formula selected distinct genres.
- Then in the next sheet created a table of those genres as a row & having columns as Average, Median, Mode, Max, Min, Variance and StdDev using the following formulas, respectively.
- =AVERAGEIFS('Movie Genre Analysis'!$K$2:$K$5044, 'Movie Genre Analysis'!$J$2:$J$5044, "*" & B6 & "*")
- =MEDIAN(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))
- =MODE(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))
- =MAX(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))
- =MIN(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))
- =VAR.P(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))
- =STDEV.P(IF(ISNUMBER(FIND(B6,'Movie Genre Analysis'!$J$2:$J$5044)),'Movie Genre Analysis'!$K$2:$K$5044))

Result Table

GENERE	AVERAGE	MEDIAN	MODE	MAX	MIN	VARIENCE	STANDARD DEVIATION
Action	6.23989592	6.3	6.1	9.1	1.7	1.250706667	1.118349975
Documentary	7.18016529	7.4	7.5	8.7	1.6	1.107044601	1.05216187
Adventure	6.4411701	6.6	6.7	8.9	1.9	1.278218349	1.13058319
Drama	6.76376253	6.9	7.2	9.3	2	0.916173353	0.957169448
Animation	6.57603306	6.7	6.7	8.6	1.7	1.293309883	1.137237831
Comedy	6.19524573	6.3	6.7	9.5	1.7	1.1890212	1.090422487
Mystery	6.4864	6.6	6.6	8.6	2.2	1.18737504	1.089667399
Fantasy	6.30704918	6.4	6.7	8.9	1.7	1.344983096	1.159734063
Crime	6.5647919	6.6	6.6	9.3	2.4	1.052427431	1.025878858
Biography	7.15017065	7.2	7	8.9	4.5	0.520247411	0.721281783
Sci-Fi	6.28181818	6.4	6.7	8.8	1.9	1.463695396	1.209832797
Horror	5.84353982	5.9	6.2	8.7	2.2	1.275697204	1.129467664
Romance	6.45058717	6.5	6.5	8.6	2.1	0.991189809	0.995585159
Thriller	6.31424522	6.4	6.1	9	2.2	1.110831801	1.053960057
Game-Show	2.9	2.9	#N/A	2.9	2.9	0	0
Family	6.24505495	6.4	6.7	8.7	1.7	1.441193495	1.200497187
Music	6.45797546	6.7	7.1	8.5	1.6	1.431638846	1.196511114
Western	6.68969072	6.8	6.5	8.9	3.8	1.075563822	1.037093931
Musical	6.50757576	6.7	7	8.5	2.1	1.491003214	1.221066425
Film-Noir	7.63333333	7.65	#N/A	8.2	7.1	0.155555556	0.394405319
History	7.08357488	7.2	7.5	8.9	2	0.784561133	0.885754556
Reality-TV	4.75	4.75	#N/A	6.6	2.9	3.4225	1.85
Short	6.38	6.5	#N/A	7.1	5.2	0.4456	0.667532771
Sport	6.60604396	6.8	7.2	8.7	2	1.207600833	1.098908929
War	7.07042254	7.1	7.1	8.6	2.7	0.76151954	0.872650869
News	7.53333333	7.4	#N/A	8.1	7.1	0.175555556	0.418993503

B

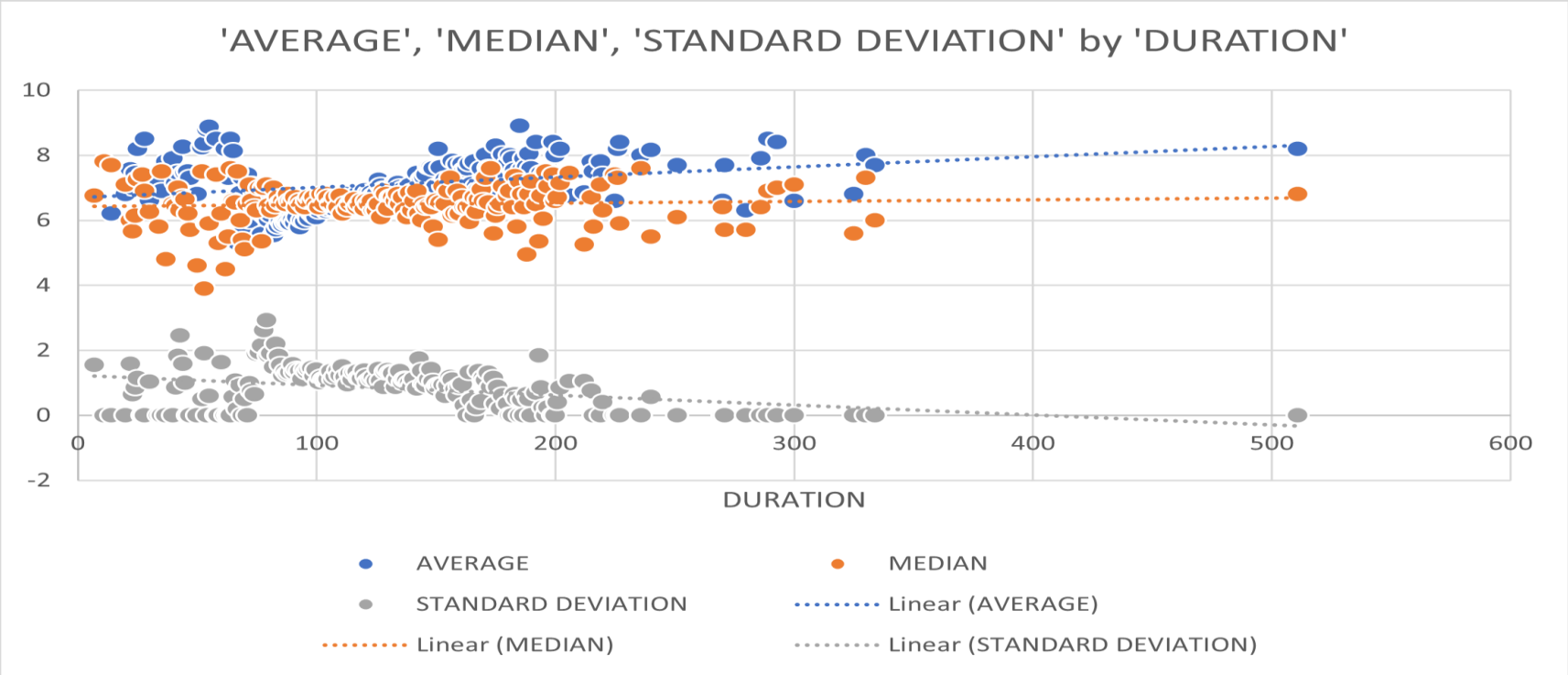
Movie Duration Analysis

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

Movie Duration Analysis

- First I created a new sheet and copied IMDb rating and duration column.
- Then by Using the “UNIQUE” formula selected distinct duration as follows.
- =SORT(UNIQUE(A2:A5029))
- Then I created a table of those durations as a row & having columns as Average, Median and Standard Deviation using the following formulas, respectively.
- =AVERAGEIFS($B$2:$B$5029,$A$2:$A$5029, E5)
- =MEDIAN(IF($A$2:$A$5029=E5,B2:B5029))
- =STDEV.P(IF($A$2:$A$5029=E5,B2:B5029))
- Then I created scatter plot named “'AVERAGE', 'MEDIAN', 'STANDARD DEVIATION' by 'DURATION'” in which I also included trend lines.
- Scatter Plot is given in the next slide
- Trend line shows that

Scatter Plot



Trendline and ScatterPlot show that movies having a duration between 25-200 have more and consistent IMDB score.

C

Language Analysis

Determined the most common languages used in movies and analyzed their impact on the IMDB score using descriptive statistics.

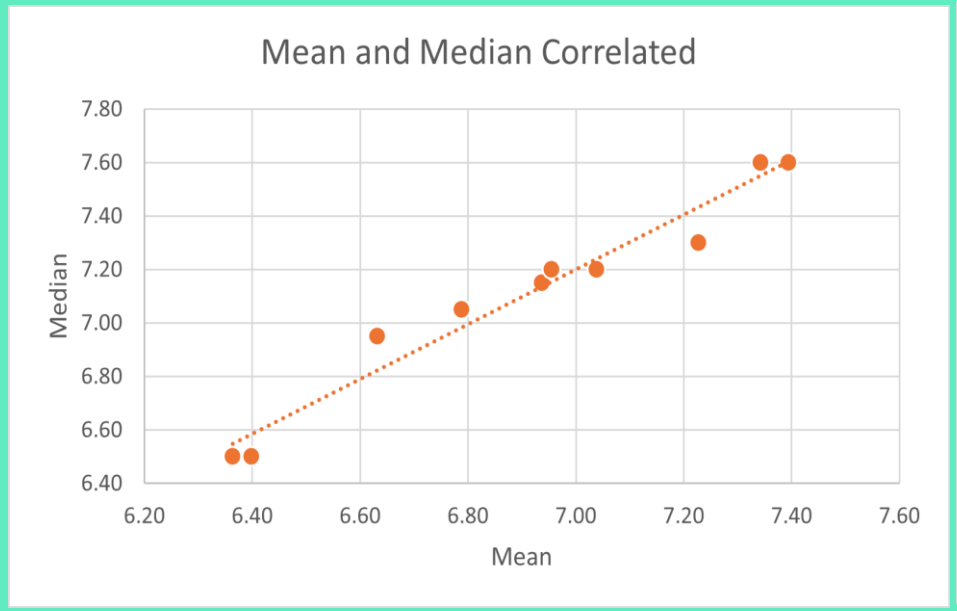
Language Analysis

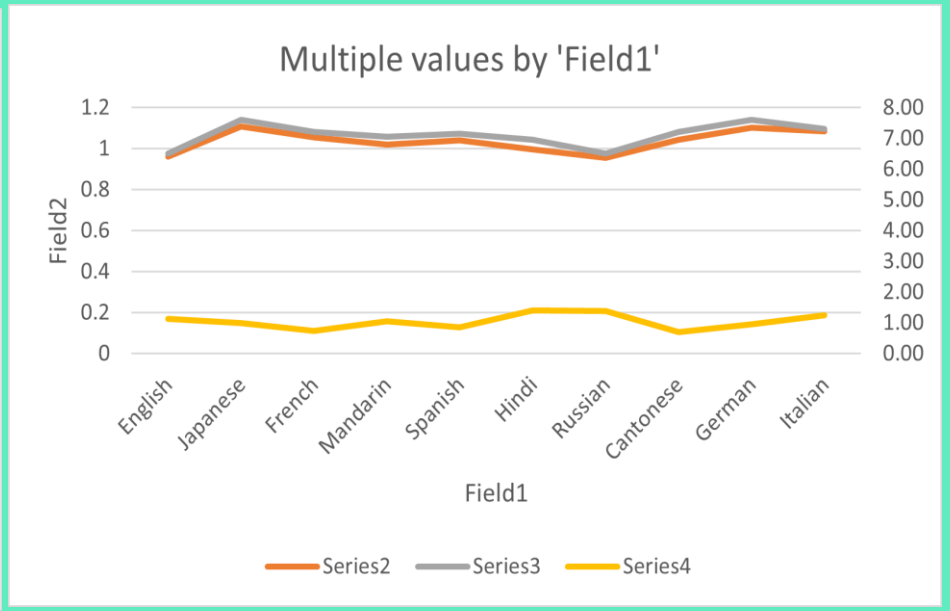
- First I created a new sheet and copied IMDb rating and Language column.
- Then by Using the “UNIQUE” formula selected distinct Languages as follows.
- =UNIQUE(A2:A5032)
- Then by using count function I have counted total count of each movie.
- =COUNT(IF(A2:A5032=D5,1))
- Then I created a table of top 10 languages as a row & having columns as Count, Mean, Median and Standard Deviation using the following formulas, respectively.
- =AVERAGE(IF($A$2:$A$5032=[@Language],$B$2:$B$5032))=MEDIAN(IF($A$2:$A$5029=E5,B2:B5029))
- =MEDIAN(IF($A$2:$A$5032=[@Language],$B$2:$B$5032))
- =STDEV(IF($A$2:$A$5032=[@Language],$B$2:$B$5032))
- Then created following graphs which shows that Mean and Median are correlated.
- Table and graph is given below.

Result Table

Language	Count	Mean	Median	Standard Deviation
English	4704	6.40	6.50	1.12
Japanese	18	7.39	7.60	0.99
French	73	7.04	7.20	0.73
Mandarin	26	6.79	7.05	1.04
Spanish	40	6.94	7.15	0.86
Hindi	28	6.63	6.95	1.40
Russian	11	6.36	6.50	1.38
Cantonese	11	6.95	7.20	0.70
German	19	7.34	7.60	0.95
Italian	11	7.23	7.30	1.24

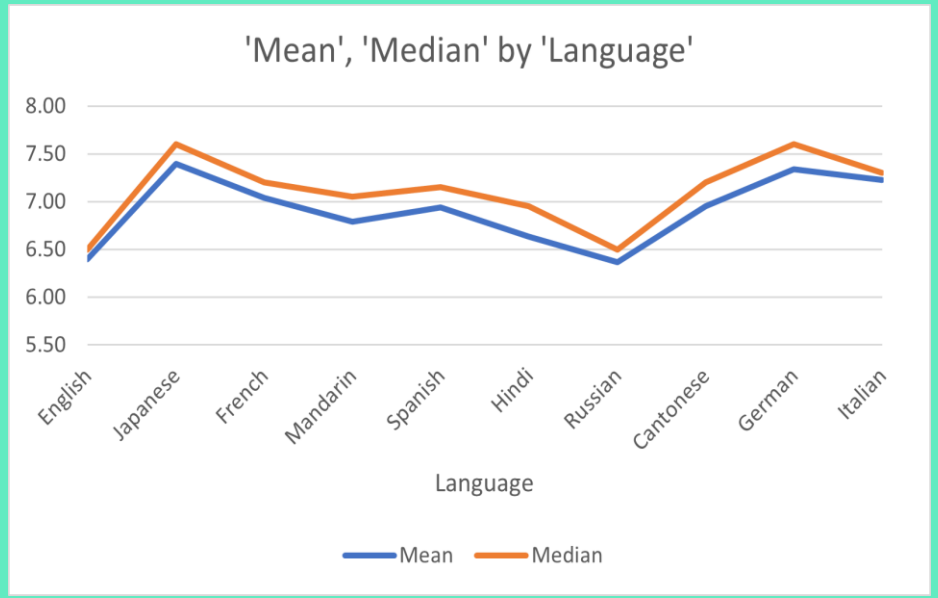
Graphs

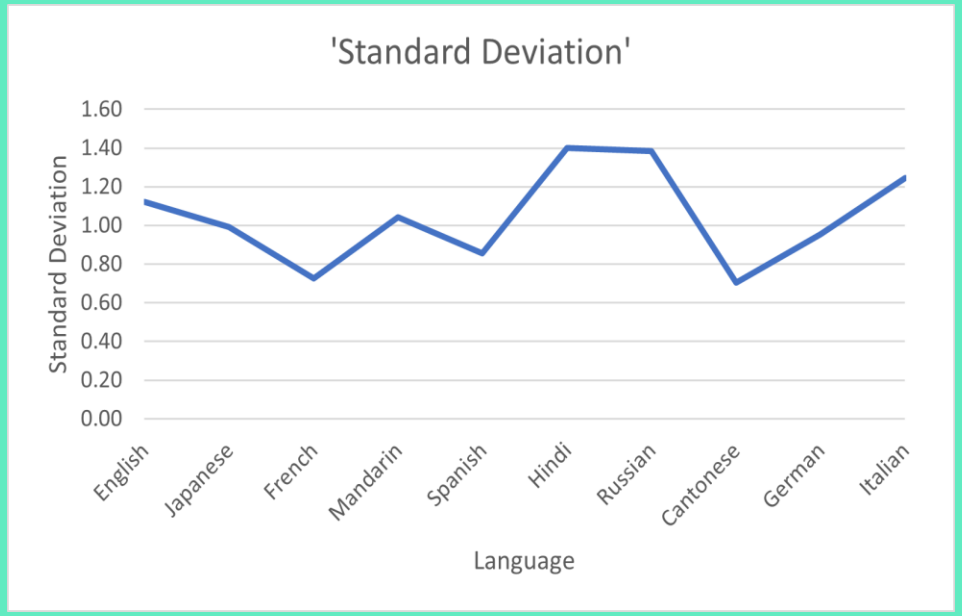




Above graphs shows that mean and median are correlated.

Graphs





Above graph shows that Hindi and Russian language movies have most standard deviation. i.e. IMDB score of Movies in this language are vary too much .

D
Director Analysis

Influence of directors on movie ratings.

Director Analysis

- First I created a new sheet and copied IMDb rating and Director name column.
- Then by Using the “UNIQUE” formula selected distinct Director Name as follows.
- =UNIQUE(A2:A5044)
- Then by using AverageIf function I calculated the average IMDB score of each director.
- =AVERAGE(IF($A$2:$A$5044=E4,$B$2:$B$5044))
- Then I created a table of directors who have at least 3 movies as a row & having columns as Count_of_directors, Rank and Percentile using the following formulas, respectively.
- =COUNT(IF($A$2:$A$5044=E4,1))
- Then by using Analysis Toolpack in the options founded rank and percentile
- Then created following graphs which shows that Mean and Median are correlated.
- Table and graph is given below.

Result Table

Director	average_imdb_rating	count_of_director	*Rank*	*Percentile*
Sergio Leone	8.475	4	1	100.00%
Christopher Nolan	8.425	8	2	99.80%
Pete Docter	8.233333333	3	3	99.60%
Hayao Miyazaki	8.225	4	4	99.40%
Quentin Tarantino	8.2	8	5	99.20%
Milos Forman	8.133333333	3	6	99.00%
Frank Capra	8.06	5	7	98.80%
Tony Kaye	8.033333333	3	8	98.60%
Stanley Kubrick	8	7	9	98.20%
David Lean	8	4	9	98.20%
Frank Darabont	7.975	4	11	97.90%
Billy Wilder	7.975	4	11	97.90%
Denis Villeneuve	7.966666667	3	13	97.70%
Joss Whedon	7.925	4	14	97.50%
James Cameron	7.914285714	7	15	97.30%
Charles Ferguson	7.866666667	3	16	96.90%
Elia Kazan	7.866666667	3	16	96.90%
Alfonso CuarÃ³n	7.8	4	18	96.50%
Richard Brooks	7.8	4	18	96.50%
Alejandro G. IÃ±Ã¡rritu	7.783333333	6	20	96.30%
Dean DeBlois	7.766666667	3	21	96.10%
Mel Gibson	7.766666667	3	22	95.80%
Victor Fleming	7.766666667	3	22	95.80%
Fred Zinnemann	7.76	5	24	95.60%
David Fincher	7.75	10	25	95.40%
Andrew Stanton	7.733333333	3	26	95.00%

Analysis | Director Analysis

Full table Drive Link :-
https://docs.google.com/spreadsheets/d/1ipGyTLqC435WaC7R35tTVDKJv3OlgmX2/edit?usp=drive_link&ouid=104957742252162470359&rtpof=true&sd=true

E

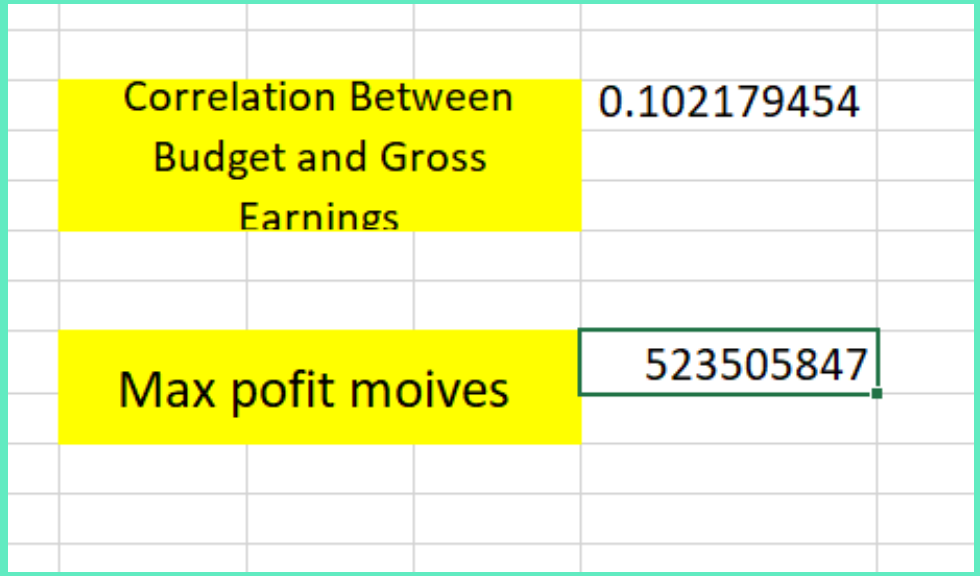
Budget Analysis

The relationship between movie budgets and their financial success

Budget Analysis

- First I created a new sheet and copied Budget, Director name and Gross column.
- Then I removed the blank entries in the table and performed the following steps.
- Then by Using the “CORREL” formula selected calculated the correlation between budget and gross.
- =CORREL(Table2[budget],Table2[gross])
- Then by using the “=[@gross]-[@budget]” formula I calculated the profit margin of each movie.
- Then by using the Max function I calculated the maximum Profit Margin.
- =MAX(Table2[profit margin])

Result



- A correlation coefficient close to 0 i.e. 0.102 indicates a weak or no linear relationship between the budget and Gross Earnings.
- Highest profit margin is 523505847 for James Cameron’s movie “AVATAR”.

Conclusion

- Thus, I have explored provided IMDB movies Dataset.
- Given key findings and all meaningful trends or patterns I have discovered.
- I have learned to use Excel to analyze the dataset.
- I have learned to gain insights by using Excel Formulas.
- All the respective Charts and their output is attached to this report.
- GitHub Repository and drive links are given as follows.

GitHub Repository:- https://github.com/ShindeYash/IMDB_Movie_Analysis

Excel Sheet:- https://docs.google.com/spreadsheets/d/1-Z8DUqb3e9oM1rhaQFGrWpMst9T6jaRQ/edit?usp=sharing&ouid=104957742521624703 59&rtpof=true&sd=true

Drive Link:- https://drive.google.com/drive/folders/1LL-qQYuRnCb_QW8dXle5u4_k-9tU_jJS?usp=drive_link

Video Presentation:- https://www.loom.com/share/b798e2adfd6b4e20a75ebe48ac3739f4?sid=0a9dd469-1fa9-481a-a394-6198a1c28c42

Thanks!

Do you have any questions?

yashpradeepshinde@gmail.com
Yash Shinde