

IMDB MOVIE ANALYSIS

Presented by: Meghna Halder



PROJECT OVERVIEW

Exploring Factors Influencing Movie Success

This project aims to investigate the factors that influence the success of movies on IMDB, defined by high IMDB ratings.

We will analyze various aspects such as movie genres, durations, languages, directors, and budgets to uncover meaningful insights.

Tech-Stack Used

Excel

DATA CLEANING

TO CLEAN THE DATA

The data cleaning process involved several steps:

- Handling missing values by filling or removing incomplete data.
- Removing duplicate entries to ensure data accuracy.
- Separating genres into individual columns for better analysis.

DRIVE LINK(CLEANED DATASET):

https://docs.google.com/spreadsheets/d/15W1GJZM1RF2OEGJW6SXVRP_ZTI0KMPYR/edit?usp=sharing&ouid=103018090401541052831&rtpof=true&sd=true

MOVIE GENRE ANALYSIS

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

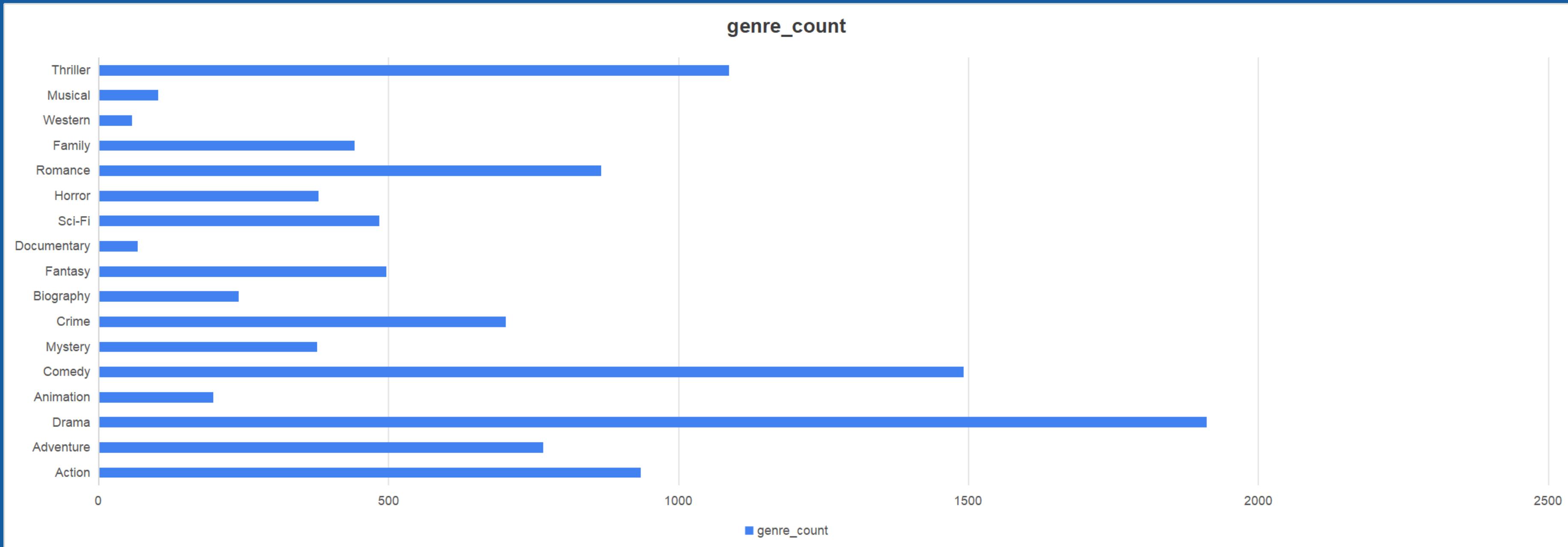
- We need to determine the most common genres using Excel's COUNTIF function.
- We will calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) for each genre using Excel functions.
- We will compare the statistics to understand the impact of each genre on movie ratings.

MOVIE GENRE ANALYSIS

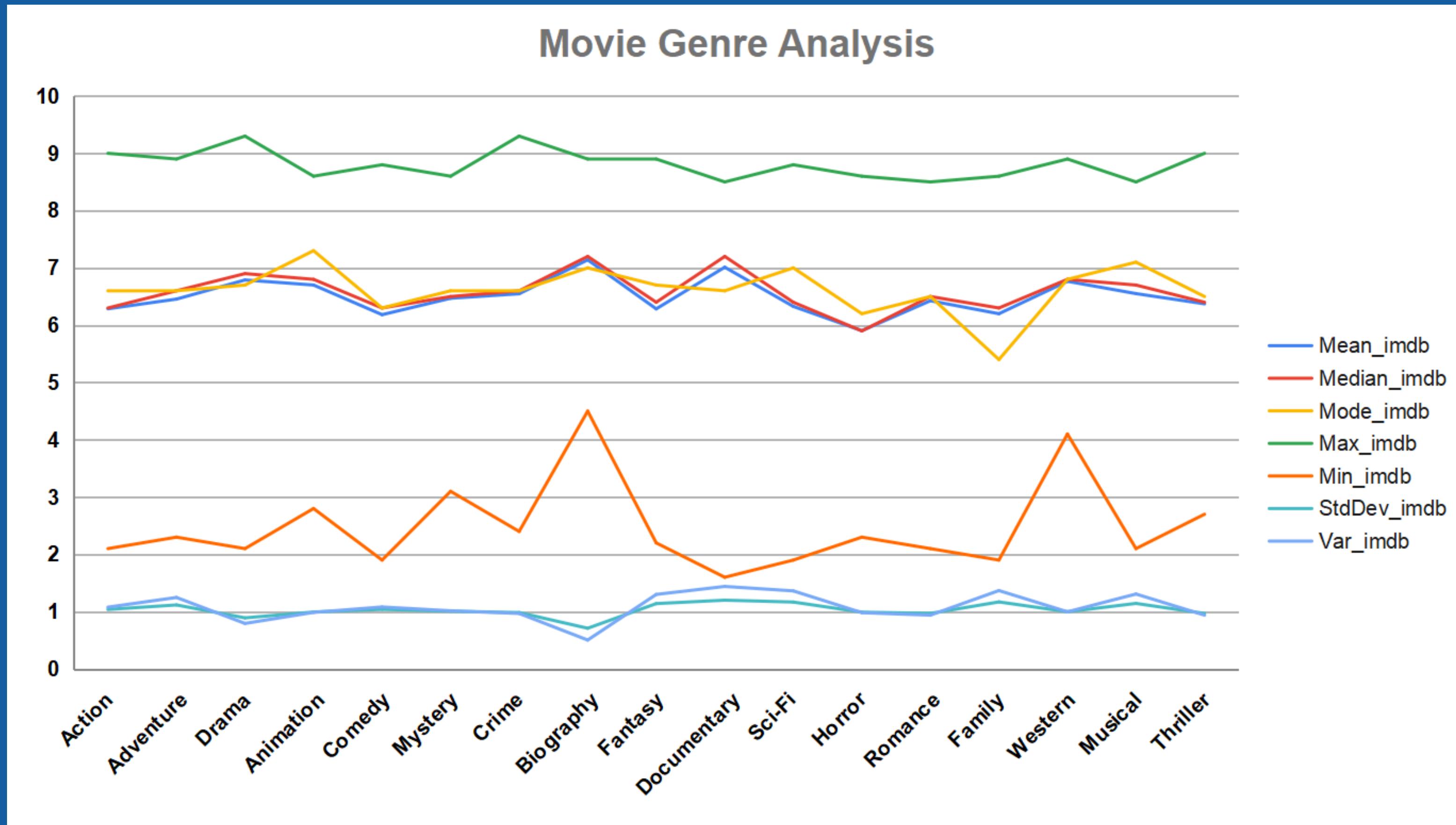
genres_unique	genre_count	mean	median	mode	max_imdb	min_imdb	stdDev_imdb	Var_imdb
Action	935	6.285989305		6.3	6.6	9	2.1	1.038357736
Adventure	766	6.561307902		6.6	6.6	8.9	2.3	1.116926308
Drama	1911	6.821745562		6.7	6.7	9.3	2.1	0.891064898
Animation	197	6.763043478		6.6	7.1	8.6	2.8	0.993627526
Comedy	1492	6.164424951		6.4	6.4	8.8	1.9	1.039919012
Mystery	377	64.01712454		6.5	6.5	8.6	3.1	1.007391835
Crime	702	6.945238095		6.6	6.7	9.3	2.4	0.984105199
Biography	242	7.151941748		7.1	7	8.9	4.5	0.71009671
Fantasy	496	75.9029694		6.5	6.1	8.9	2.2	1.140414241
Documentary	67	6.951162791		7.3	7.2	8.5	1.6	1.199939694
Sci-Fi	484	#DIV/0!		6.6	6.7	8.8	1.9	1.16718415
Horror	379	5.813461538		6.2	5.9	8.6	2.3	0.991023285
Romance	866	#DIV/0!		6.6	6.2	8.5	2.1	0.968996249
Family	441	6.5		6.5	5.9	8.6	1.9	1.169576458
Western	58	5.575		6.6	6.7	8.9	4.1	0.998516746
Musical	102	187.9699527		6.6	6.3	8.5	2.1	1.143535
Thriller	1087	462.7664575		6.5	6.6	9	2.7	0.969078327

THE RELATIONSHIP BETWEEN MOVIE DURATION AND IMDB SCORES INDICATED THAT MOVIES WITHIN A MODERATE DURATION RANGE TEND TO HAVE BETTER RATINGS, AS VISUALIZED THROUGH SCATTER PLOTS AND TRENDLINES.

MOVIE GENRE ANALYSIS



MOVIE GENRE ANALYSIS



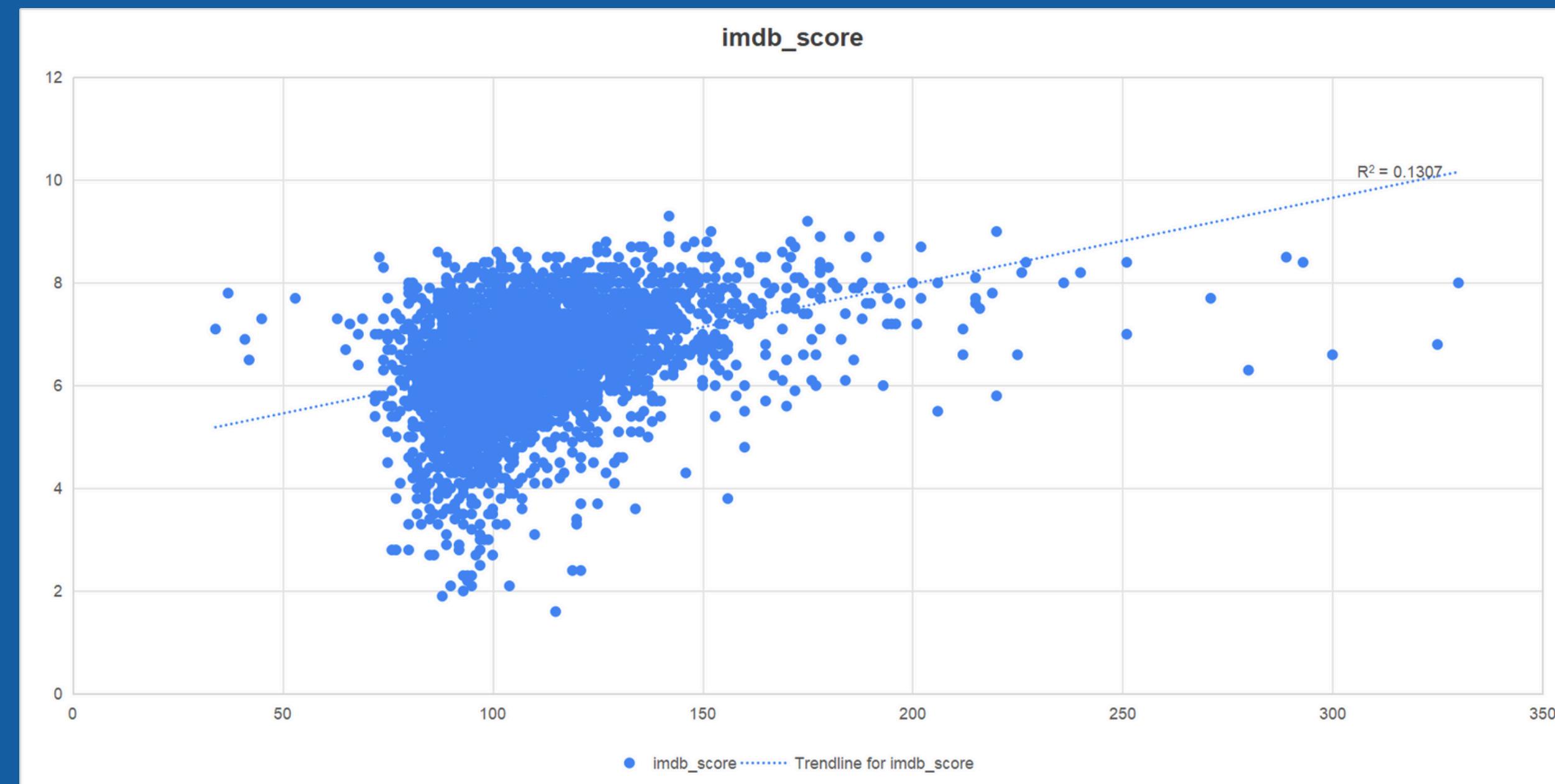
MOVIE DURATION ANALYSIS

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

- We need to determine the distribution of movie durations using Excel's functions.
- We will calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) for movie durations using Excel functions.
- We will create a scatter plot to visualize the relationship between movie duration and IMDB score.
- We will add a trendline to the scatter plot to assess the direction and strength of the relationship.
- We will interpret the trendline and descriptive statistics to understand the impact of movie duration on IMDB scores.

MOVIE DURATION ANALYSIS

Operation	Value
Mean	109.7151178
Median	105
Mode	101
Variance	514.5936088
Standard Deviation	22.6846558



LANGUAGE ANALYSIS

Examine the distribution of movies based on their language.

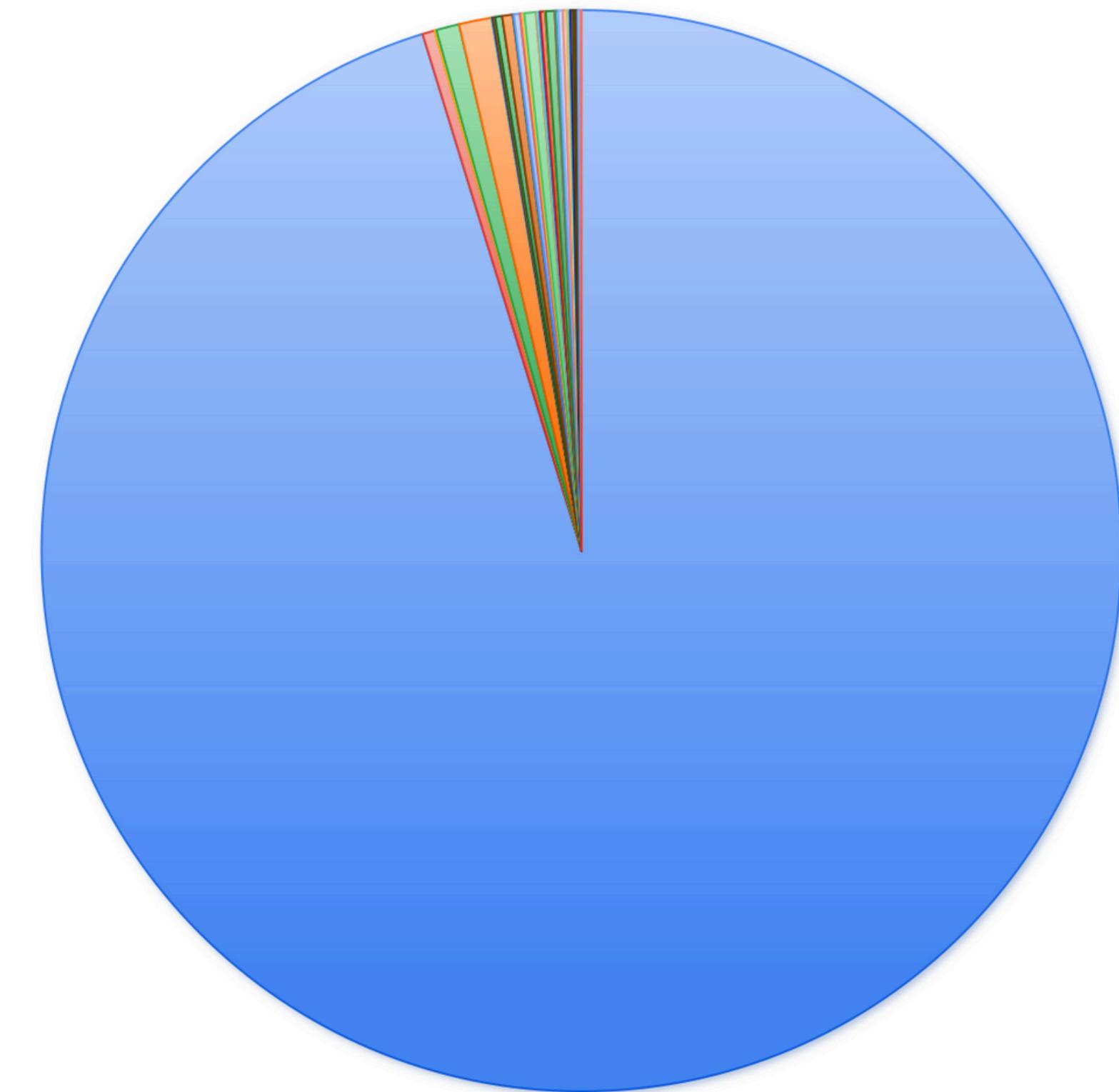
- We need to use Excel's COUNTIF function to count the number of movies for each language.
- For each language, we will calculate the mean of the IMDB scores using the AVERAGEIF function.
- For each language, we will calculate the median of the IMDB scores using the MEDIAN function combined with IF.
- For each language, we will calculate the standard deviation of the IMDB scores using the STDEVIF function.
- We will compare the mean, median, and standard deviation of the IMDB scores for each language.
- We will analyze these statistics to understand the impact of language on movie ratings.

LANGUAGE ANALYSIS

Language	movie count	avg_imdb	median_imdb	var_imdb	StdDev_imdb
English	3608	6.421436495		6.5	1.107753941
Mandarin	14	7.021428571		7.25	0.588428571
Aboriginal	2	6.95		6.95	0.605
Spanish	28	7.05		7.15	0.6828
French	37	7.288488488		7.2	0.31509009
Filipino	1	6.7		6.7	#DIV/0!
Maya	1	7.8		7.8	#DIV/0!
Kazakh	1	6		6	#DIV/0!
Telugu	1	6.4		6.4	#DIV/0!
Cantonese	8	7.2375		7.3	0.194107143
Japanese	12	7.825		7.8	0.809318182
Aramaic	1	7.1		7.1	#DIV/0!
Italian	7	7.185714286		7	1.334761905
Dutch	3	7.588888887		7.8	0.163333333
Dari	2	7.5		7.4	0.536291667
German	13	7.692307692		7.7	0.410769231
Mongolian	1	7.3		7.3	#DIV/0!
Thai	3	6.633333333		6.6	0.203333333
Bosnian	1	4.3		4.3	#DIV/0!
Korean	5	7.7		7.7	0.325
Hungarian	1	7.1		7.1	#DIV/0!
Hindi	10	6.78		7.05	1.236
Icelandic	1	6.9		6.9	#DIV/0!
Danish	3	7.9		8.1	0.28
Portuguese	5	7.78		8	0.958
Norwegian	4	7.15		7.3	0.33
Czech	1	7.4		7.4	#DIV/0!
Russian	1	6.5		6.5	#DIV/0!
None	1	8.5		8.5	#DIV/0!
Zulu	1	7.3		7.3	#DIV/0!
Hebrew	3	7.5		7.3	0.19
Dzongkha	1	7.5		7.5	#DIV/0!
Arabic	1	7.2		7.2	#DIV/0!
Vietnamese	1	7.4		7.4	#DIV/0!
Indonesian	2	7.9		7.9	0.18
Romanian	1	7.9		7.9	#DIV/0!
Persian	3	6.133333333		6.4	0.303333333
Swedish	1	7.8		7.8	#DIV/0!

LANGUAGE ANALYSIS

English Mandarin Aboriginal Spanish French Filipino Maya Kazakh Telugu Cantonese Japanese Aramaic Italian
Dutch Dari German Mongolian Thai Bosnian Korean Hungarian Hindi Icelandic Danish Portuguese Norwegian
Czech Russian None Zulu Hebrew Dzongkha Arabic Vietnamese Indonesian Romanian Persian Swedish



DIRECTOR ANALYSIS

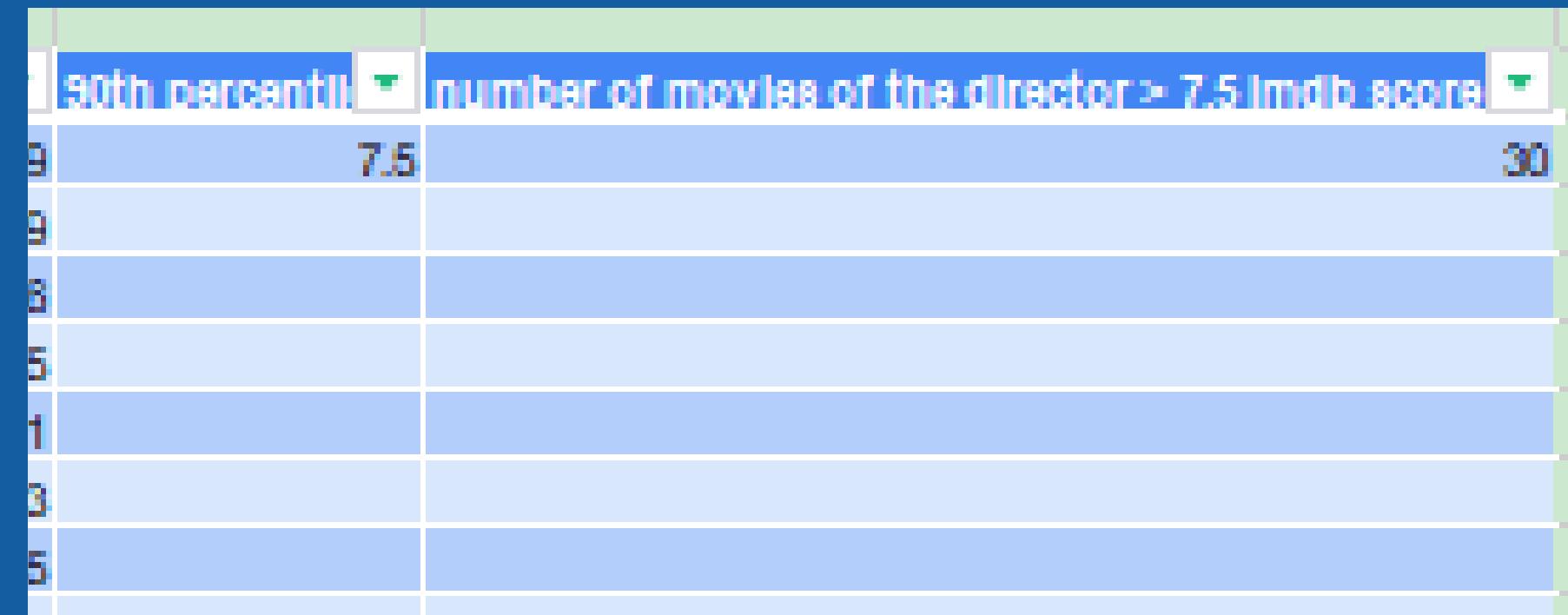
Influence of directors on movie ratings.

- We need to use Excel's AVERAGEIF function to calculate the average IMDB score for each director.
- We will use Excel's PERCENTILE function to identify the threshold for top directors.
- We will create a list of directors whose average IMDB scores are above the 90th percentile.
- We will compare the average scores of the top directors to the overall distribution of average IMDB scores.
- We will create visualizations, such as box plots or bar charts, to illustrate the comparison.

DIRECTOR ANALYSIS

director_name	avg_imdb	T	S0
Doug Lefler	8.6		
Stevan E. de Souza	8.5		
Damien Chazelle	9.3		
Lorene Scafaria	8.6		
David Gelb	8.5		
Agnieszka Wojtowicz	8.7		
Dave McKean	8.5		
Clark Gregg	8.5		
Damien Chazelle	8.7		
Jeff Garlin	8.6		
Molly Bernstein	8.7		
Ricki Stern	8.6		

TOP 10 DIRECTORS



PERCENTILE CALCULATIONS HELPED IN DISTINGUISHING DIRECTORS WITH CONSISTENTLY HIGH RATINGS, SUGGESTING THEIR SIGNIFICANT INFLUENCE ON MOVIE SUCCESS.

BUDGET ANALYSIS

Explore the relationship between movie budgets and their financial success.

- We need to use Excel's CORREL function to calculate the correlation coefficient
- We will create a new column for profit margin and calculate it as the difference between gross earnings and budget for each movie.
- We will use Excel's MAX function to find the highest profit margin
- We will identify the movie(s) corresponding to the highest profit margin.
- We will use the profit margin data to highlight the movies with the highest profit margins in the scatter plot.

BUDGET ANALYSIS

Movie Title	Profit Margin (%)	Total Profit (\$)
Avatar	76.05	56669
The Dark Knight Rises	44.81	30478
Avengers: Age of Ultron	45.89	91458
The Avengers	62.32	79374
Titanic	65.86	72108
Jurassic World	65.21	77147
The Dark Knight	53.33	15909
Star Wars: Episode I - The Phantom Menace	47.45	44541
Shrek 2	43.64	70943
Dirty Grandpa	46.09	35540

TOP 10

90th percentile	number of movies of the director > 7.5 Imdb score
7.5	30

DRIVE LINK(FINAL SHEET):

HTTPS://DOCS.GOOGLE.COM/SPREADSHEETS/D/15W1GJZM1RF2OEGJW6SXVRP_ZTIOKMPYR/EDIT?USP=SHARING&OUID=103018090401541052831&RTP_OF=TRUE&SD=TRUE

RESULTS

- Enhanced Understanding of Movie Data: The systematic approach to analyzing genres, durations, languages, directors, and budgets provided a deep understanding of the factors influencing movie ratings and financial success.
- Proficiency in Excel for Data Analysis: The project showcased advanced Excel techniques, including the use of COUNTIF, AVERAGEIF, MEDIAN, STDEV, CORREL, and MAX functions, as well as the creation of scatter plots and trendlines for data visualization.
- Identification of Key Trends: The analysis highlighted significant trends, such as the impact of genre, language, and director on movie ratings, and the relationship between budgets and earnings.
- Practical Insights for Movie Production: The findings can inform movie producers and marketers about audience preferences, effective budget allocation, and the potential influence of directors and languages on movie success.

Overall, the project contributed to a thorough understanding of the IMDB movie dataset, demonstrating the ability to derive meaningful insights through data analysis and statistical techniques using Excel.