

PROJECT 5

IMDB Movie Analysis

FINAL PROJECT-1

BY : - PRADEEP SHET

TECH STACK USED:-

1.MS EXCEL

2.MS POWERPOINT

HYPERLINK OF EXCEL FILE:-

<https://docs.google.com/spreadsheets/d/1brd4hPq4mXhZzQo3xv-xdUSd0jf3U8vd/edit?usp=sharing&ouid=112153428545971944615&rtpof=true&sd=true>

PROJECT DESCRIPTION :

The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

DATA CLEANING

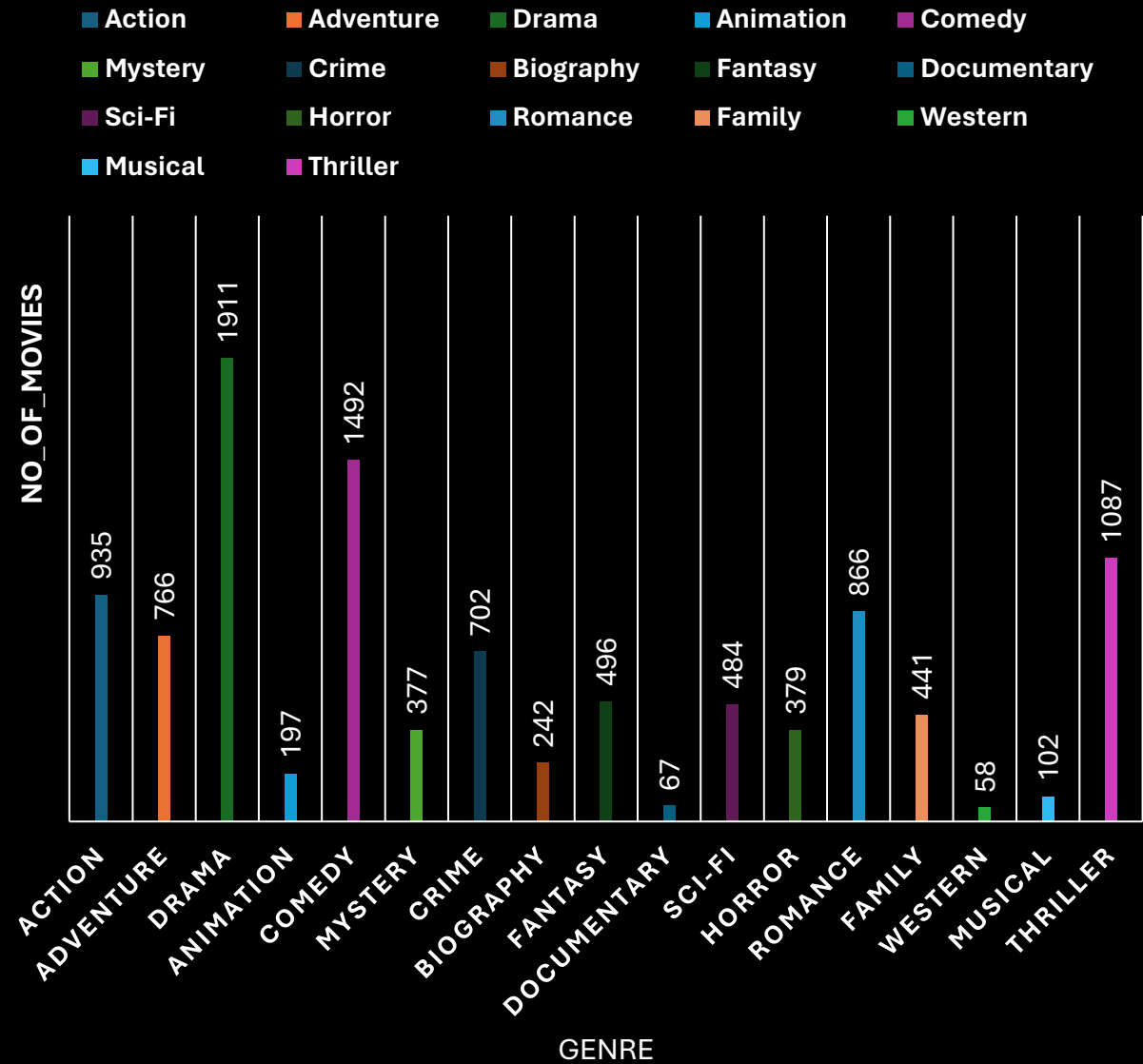
- REMOVED SOME OF THE COLUMNS FROM THE DATA SET WHICH HAD NO OR LESS USE IN DETERMINING THE TASKS.
- (Color, director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, cast_total_facebook_likes, actor_3_name, facenumber_in_posts, plot_keywords, movie_imdb_link, content_rating, actor_2_facebook_likes, aspect_ratio, movie_facebook_likes).
- REMOVED BLANK SPACES FROM THE DATASET TO LOCATE AND ADDRESS THESE BLANK ROWS, I UTILIZED THE "FIND & SELECT" FUNCTION, FOLLOWED BY SELECTING "GO TO SPECIAL" AND CHOOSING THE "BLANK" OPTION. THIS ACTION EFFECTIVELY HIGHLIGHTED ALL THE BLANK ROWS. SUBSEQUENTLY, I EMPLOYED THE SHORTCUT "CTRL + -" AND SELECTED THE "ENTIRE ROWS" OPTION TO ELIMINATE THESE BLANK ROWS FROM THE DATASET.

A. MOVIE GENRE ANALYSIS

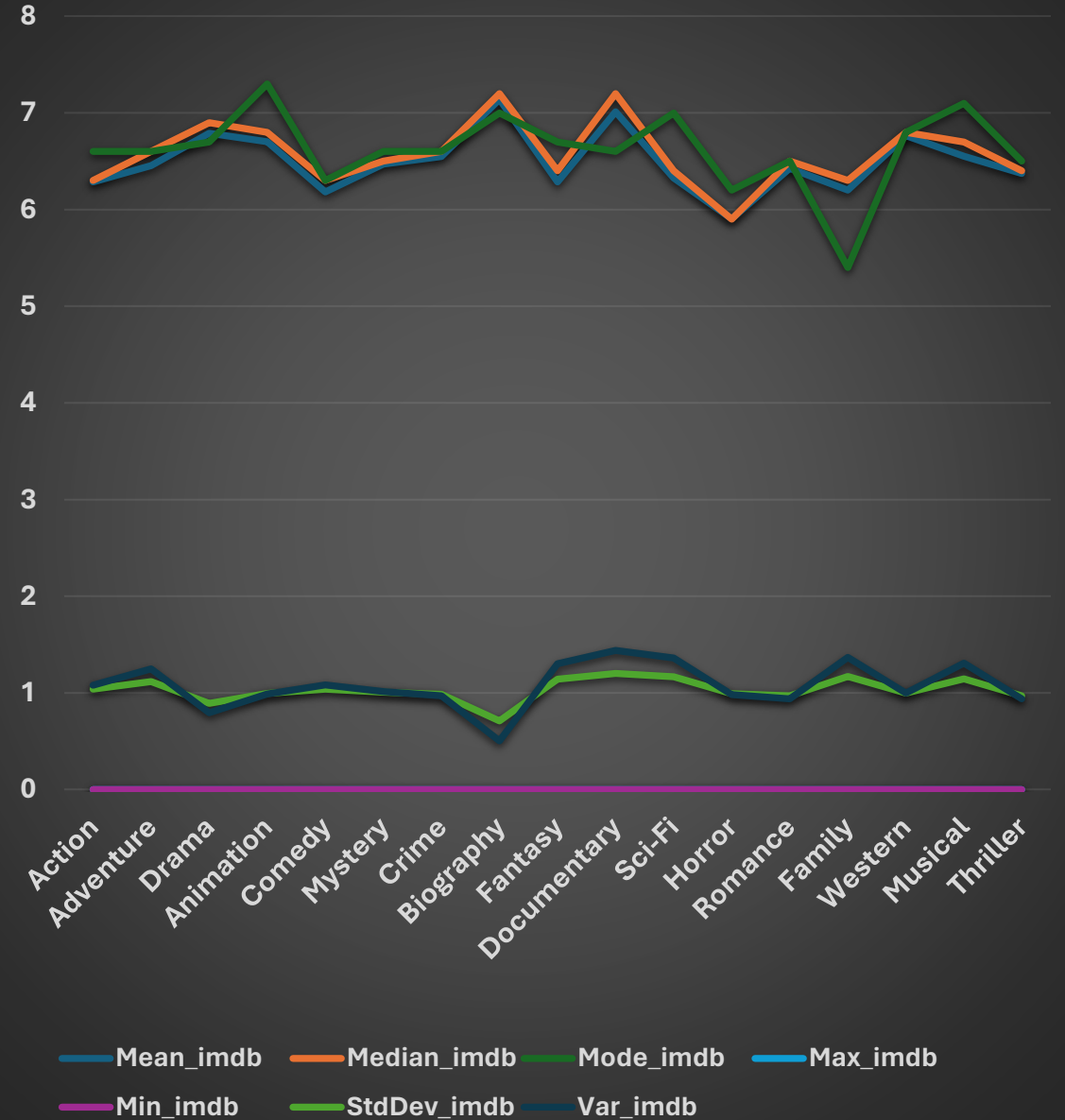
- **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.
- **Hint:** Use Excel's COUNTIF function to count the number of movies for each genre. You might need to manipulate the 'genres' column to separate multiple genres for a single movie. Use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics. Compare the statistics to understand the impact of genre on movie ratings.

GENRE	NO_OF_MOVIES	MEAN_IMDB	MEDIAN_IMDB	MODE_IMDB	MAX_IMDB	MIN_IMDB	STDDEV_IMDB	VAR_IMDB
Action	935	6.285989305	6.3	6.6	9	2.1	1.038357736	1.078186788
Adventure	766	6.454960836	6.6	6.6	8.9	2.3	1.116926308	1.247524378
Drama	1911	6.789115646	6.9	6.7	9.3	2.1	0.891064898	0.793996652
Animation	197	6.700507614	6.8	7.3	8.6	2.8	0.993627525	0.987295659
Comedy	1492	6.183310992	6.3	6.3	8.8	1.9	1.039919012	1.081431552
Mystery	377	6.469496021	6.5	6.6	8.6	3.1	1.007391835	1.014838309
Crime	702	6.548148148	6.6	6.6	9.3	2.4	0.984105199	0.968463042
Biography	242	7.140082645	7.2	7	8.9	4.5	0.71009671	0.504237338
Fantasy	496	6.285080645	6.4	6.7	8.9	2.2	1.140414241	1.30054464
Documentary	67	7.011940299	7.2	6.6	8.5	1.6	1.199939694	1.439855269
Sci-Fi	484	6.327272727	6.4	7	8.8	1.9	1.16718415	1.362318841
Horror	379	5.903957784	5.9	6.2	8.6	2.3	0.991023285	0.982127152
Romance	866	6.426212471	6.5	6.5	8.5	2.1	0.968996249	0.938953731
Family	441	6.2	6.3	5.4	8.6	1.9	1.169576458	1.367909091
Western	58	6.765517241	6.8	6.8	8.9	4.1	0.998516746	0.997035693
Musical	102	6.550980392	6.7	7.1	8.5	2.1	1.143535	1.307672297
Thriller	1087	6.372309108	6.4	6.5	9	2.7	0.969078327	0.939112803

NO_OF_MOVIES VS. GENRE



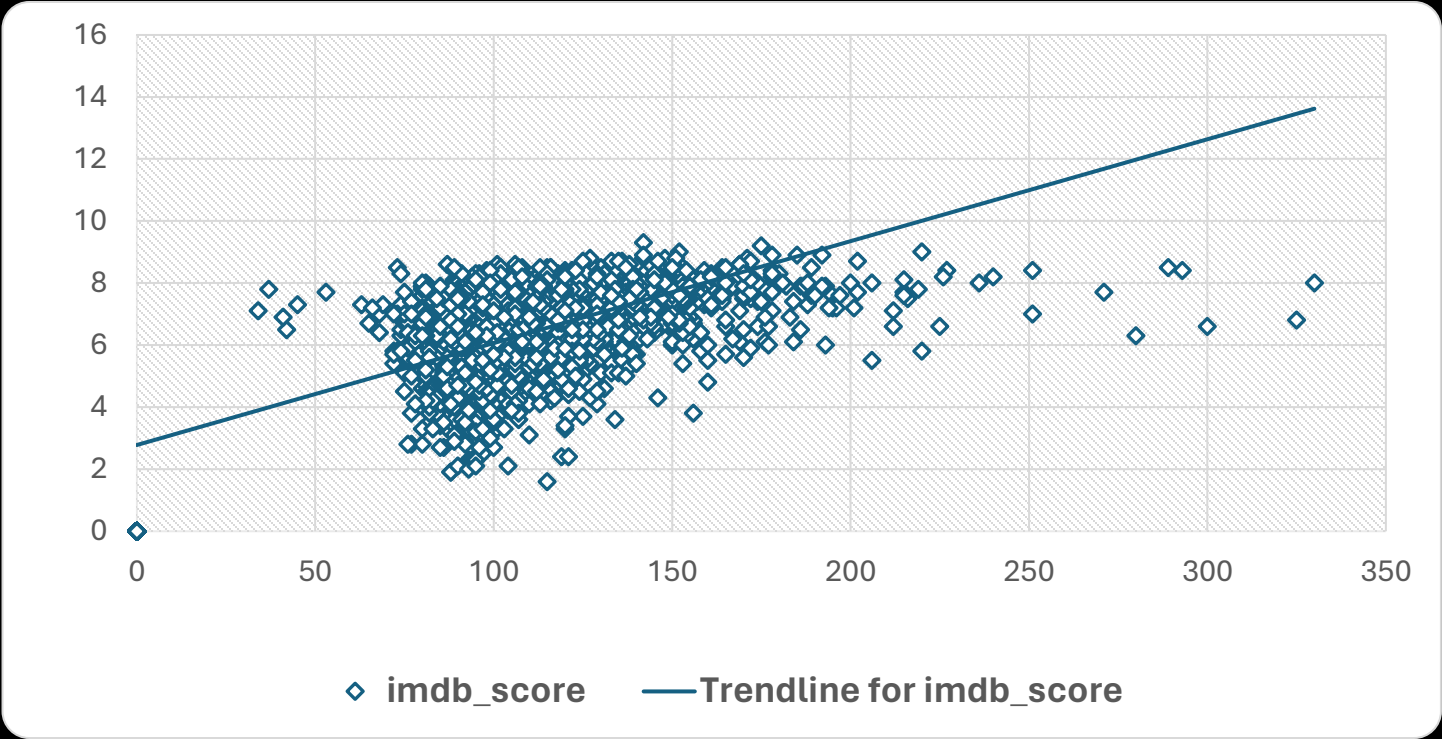
MOVIE GENRE ANALYSIS



B. Movie Duration Analysis

- **Task:** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
- **Hint:** Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV. Create a scatter plot to visualize the relationship between movie duration and IMDB score. Add a trendline to assess the direction and strength of the relationship.

OPERATIONS	VALUES
Mean	109.808505
Median	105
Mode	101
Standard Deviation	22.763201
Variance	518.16332



C. LANGUAGE ANALYSIS

- **Task**: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.
- **Hint**: Use Excel's COUNTIF function to count the number of movies for each language. Calculate the mean, median, and standard deviation of the IMDB scores for each language. Compare the statistics to understand the impact of language on movie ratings.

<u>Language</u>	<u>No_of_movies</u>	<u>Average_imdb</u>	<u>Median_imdb</u>	<u>Var_imdb</u>	<u>StdDev_imdb</u>
English	3606	6.421436495	6.5	1.107753941	1.052498903
Mandarin	14	7.021428571	7.25	0.586428571	0.765786244
Aboriginal	2	6.95	6.95	0.605	0.777817459
Spanish	26	7.05	7.15	0.6826	0.826196103
French	37	7.286486486	7.2	0.31509009	0.561328861
Cantonese	8	7.2375	7.3	0.194107143	0.440575922
Japanese	12	7.625	7.8	0.809318182	0.899621132
Italian	7	7.185714286	7	1.334761905	1.155318962
Dutch	3	7.566666667	7.8	0.163333333	0.404145188
Dari	2	7.5	7.4	0.536291667	0.732319375
German	13	7.692307692	7.7	0.410769231	0.640912811
Thai	3	6.633333333	6.6	0.203333333	0.450924975
Korean	5	7.7	7.7	0.325	0.570087713
Hindi	10	6.76	7.05	1.236	1.111755369
Danish	3	7.9	8.1	0.28	0.529150262
Portuguese	5	7.76	8	0.958	0.978774744
Norwegian	4	7.15	7.3	0.33	0.574456265
Hebrew	3	7.5	7.3	0.19	0.435889894
Indonesian	2	7.9	7.9	0.18	0.424264069
Persian	3	8.133333333	8.4	0.303333333	0.550757055

D. DIRECTOR ANALYSIS

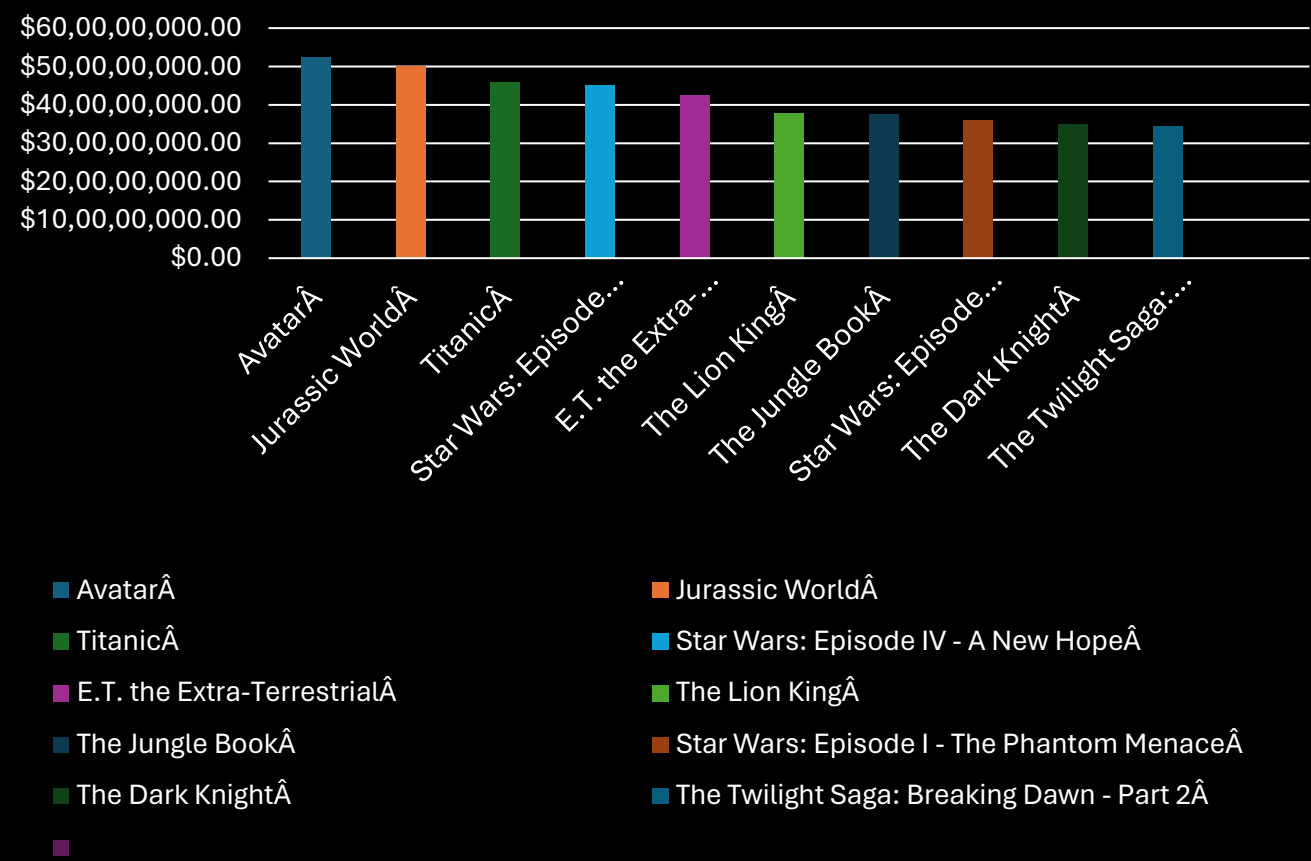
- **Task:** Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
- **Hint:** Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores

<u>DIRECTOR</u>	<u>AVERAGE_IMDB</u>	<u>PERCENTILE</u>	<u>COUNT_MOVIES</u>
James Cameron	7.914285714	0.976	7
Gore Verbinski	6.985714286	0.722	7
Sam Mendes	7.457142857	0.891	7
Christopher Nolan	8.425	0.995	8
Andrew Stanton	7.733333333	0.953	3
Sam Raimi	6.96	0.718	10
Nathan Greno	7.8	0.958	1
Joss Whedon	7.866666667	0.969	3
David Yates	7.2	0.812	3
Zack Snyder	7.142857143	0.804	7
Bryan Singer	7.2875	0.849	8
Marc Forster	7.228571429	0.841	7
Andrew Adamson	7.15	0.805	4
Rob Marshall	6.6	0.553	5
Barry Sonnenfeld	6.457142857	0.5	7
Peter Jackson	7.888888889	0.969	9
Marc Webb	7.133333333	0.801	3
Ridley Scott	7.13125	0.8	16
Chris Weitz	6.08	0.348	5
Anthony Russo	7	0.723	4
Peter Berg	6.666666667	0.592	6
Colin Trevorrow	7	0.723	2
Shane Black	7.4	0.875	2
Tim Burton	7.05	0.765	14
Brett Ratner	6.455555556	0.499	9
Dan Scanlon	7.3	0.849	1

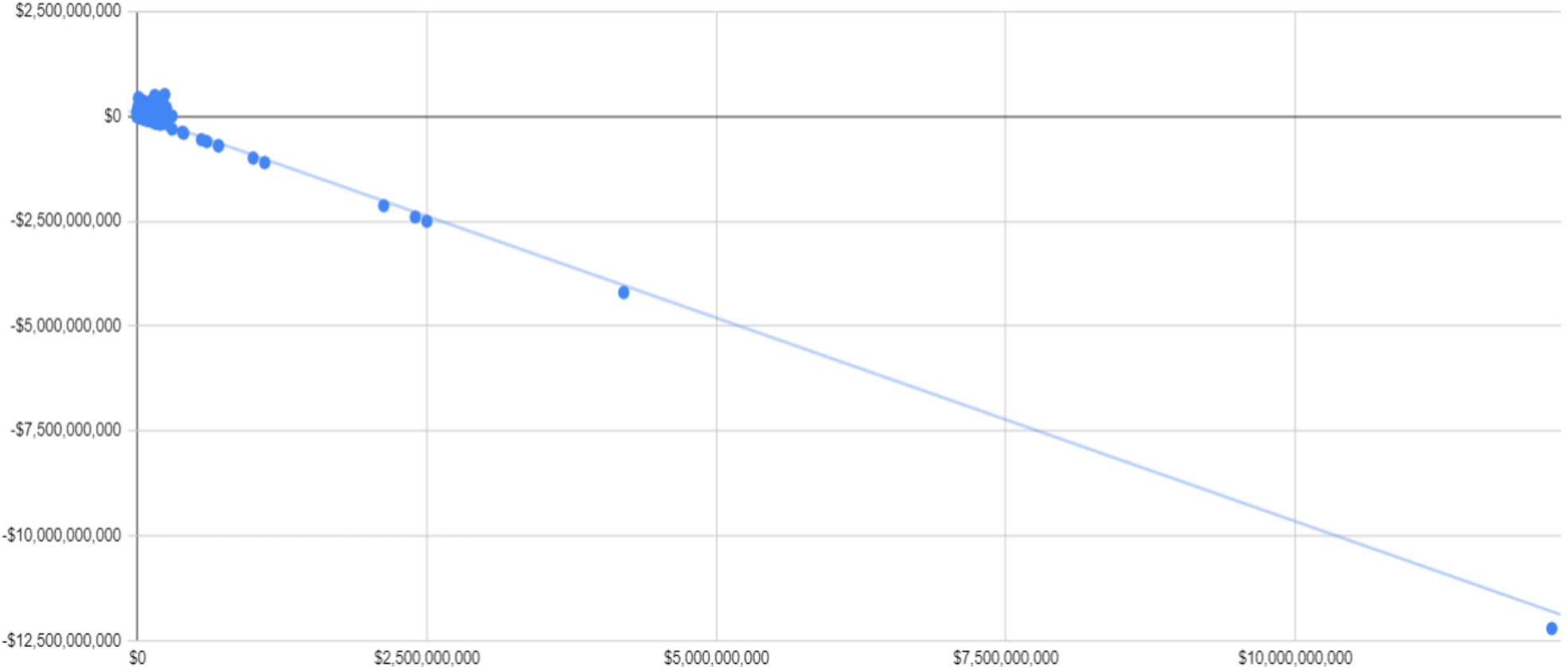
E. BUDGET ANALYSIS

- **Task**: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.
- **Hint**: Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function. Calculate the profit margin (gross earnings - budget) for each movie and identify the movies with the highest profit margin using Excel's MAX function.

MOVIES	PROFITS IN MILLIONS
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459
The Lion King	377783777
The Jungle Book	375290282
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Twilight Saga: Breaking Dawn - Part 2	344597846



Budget Vs Profit



INSIGHTS

- THE TOP 5 MOST PROFITABLE FILMS ARE AVATAR, JURASSIC WORLD, TITANIC, STAR WARS: EPISODE IV - A NEW HOPE, AND E.T. THE EXTRA-TERRESTRIAL. NOTABLY, THERE IS A POSITIVE CORRELATION BETWEEN A MOVIE'S BUDGET AND ITS GROSS EARNINGS.
- THE TOP 10 DIRECTORS WITH AN AVERAGE IMDB SCORE OF 8.4 OR HIGHER. THEY ARE TONY KAYE, CHARLES CHAPLIN, ALFRED HITCHCOCK, RON FRICKE, DAMIEN CHAZELLE, MAJID MAJIDI, SERGIO LEONE, CHRISTOPHER NOLAN, SS RAJAMOULI, AND RICHARD MARQUAND.
- THE MOST COMMONLY USED LANGUAGES IN THESE MOVIES ARE ENGLISH, FRENCH, SPANISH, MANDARIN, AND GERMAN. INTERESTINGLY, I'VE OBSERVED THAT THE LANGUAGES TELUGU AND PERSIAN HAVE THE HIGHEST AVERAGE IMDB SCORES.
- THE AVERAGE DURATION OF A FILM IS 109 MINUTES.
- THE MOST PREVALENT MOVIE GENRES FOUND IN THE DATASET INCLUDE DRAMA, COMEDY, THRILLER, AND ACTION.

RESULT OF THE PROJECT

- UNDERSTOOD EXCEL FUNCTIONS AND THEIR USAGE MUCH BETTER.
- GOT FAMILIAR WITH EXCEL.
- GOT TO KNOW ABOUT ADVANCED EXCEL FUNCTIONS.
- UNDERSTOOD HOW TO PLOT VARIOUS GRAPHS.

THANK YOU