Task-5: IMDB Movie Analysis

Contents:

Excel Tasks:

1. Movie Genre Analysis
2. Duration Analysis
3. Language Analysis
4. Director Analysis
5. Budget Analysis

Project Description:

- The IMDb Movie Analysis project aims to explore and analyze a comprehensive dataset of movies available on the IMDb platform.
- This dataset contains essential information about movies, including director names, movie titles, duration, genre, budget, gross earnings, IMDb ratings, and more.
- Through in-depth data analysis using Excel, Data Visualization and Statistics techniques this project seeks to extract valuable insights and trends that contribute to a movie's success.

- **Software Used: Microsoft Excel 365**

- **NOTE: ALL THE LINKS FOR CLEANED DATASET AND SOLUTIONS DATASET ARE PROVIDED BELOW !!!**

DATA HANDLING

My Approach:

- I have gone through the dataset and understood all the given columns. Then I have observed that there are a total of 28 Columns and 5043 Rows. This dataset consists of unwanted columns, Null values and Blank rows. So, I have decided to Clean this dataset thoroughly.

1. First, I have deleted the columns which have no relation to our project and don't provide any valuable insights. In the end, I only left with 9 Columns which are director's name, duration, movie title, genre, budget, gross, IMDB rating, language and country.
2. Then, I noticed that there were many blank rows. To find them I first clicked on "Find & Select" then clicked on "go to special" and selected the "blank" option. It highlighted all the blank rows. Then I clicked the shortcut "CTRL + - " and selected the "Entire rows" option. This process deleted the entire blank rows in the dataset.
3. Finally, I also deleted the duplicate rows present in the dataset. Now, I left with a total of 9 Columns and 3786 Rows.

DATA ANALYSIS

1) Movie Genre Analysis:

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

Results:

GENRE	No_of_movies	Mean_imdb	Median_imdb	Mode_imdb
Action	935	6.285989305	6.3	6.6
Adventure	766	6.454960836	6.6	6.6
Drama	1911	6.789115646	6.9	6.7
Animation	197	6.700507614	6.8	7.3
Comedy	1492	6.183310992	6.3	6.3
Mystery	377	6.469496021	6.5	6.6
Crime	702	6.548148148	6.6	6.6
Biography	242	7.140082645	7.2	7
Fantasy	496	6.285080645	6.4	6.7
Documentary	67	7.011940299	7.2	6.6
Sci-Fi	484	6.327272727	6.4	7
Horror	379	5.903957784	5.9	6.2
Romance	866	6.426212471	6.5	6.5
Family	441	6.2	6.3	5.4
Western	58	6.765517241	6.8	6.8
Musical	102	6.550980392	6.7	7.1
Thriller	1087	6.372309108	6.4	6.5

DATA ANALYSIS

1) Movie Genre Analysis:

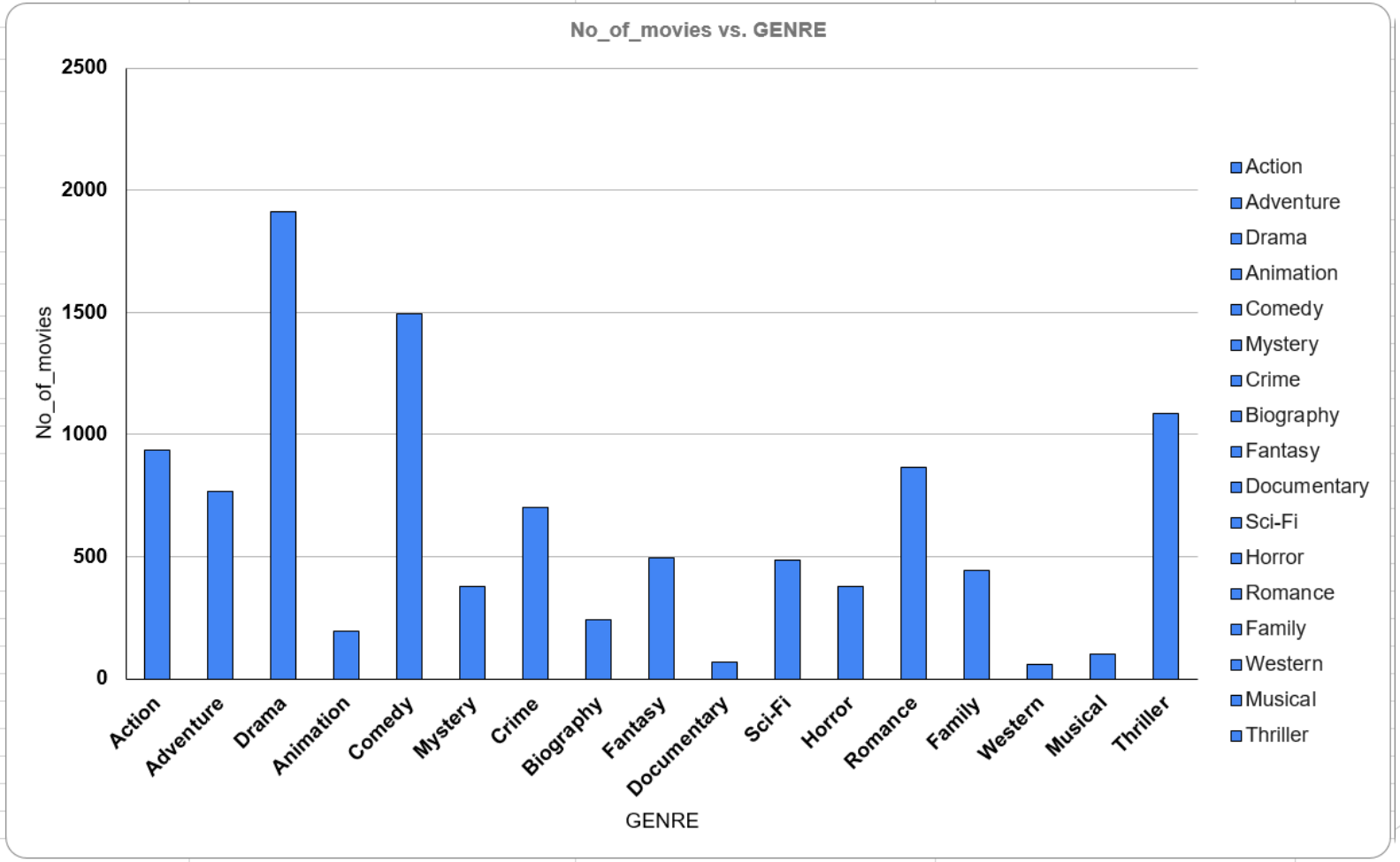
Result:

Max_imdb	Min_imdb	StdDev_imdb	Var_imdb
9	2.1	1.038357736	1.078186788
8.9	2.3	1.116926308	1.247524378
9.3	2.1	0.891064898	0.793996652
8.6	2.8	0.993627526	0.987295659
8.8	1.9	1.039919012	1.081431552
8.6	3.1	1.007391835	1.014838309
9.3	2.4	0.984105199	0.968463042
8.9	4.5	0.71009671	0.504237338
8.9	2.2	1.140414241	1.30054464
8.5	1.6	1.199939694	1.439855269
8.8	1.9	1.16718415	1.362318841
8.6	2.3	0.991023285	0.982127152
8.5	2.1	0.968996249	0.938953731
8.6	1.9	1.169576458	1.367909091
8.9	4.1	0.998516746	0.997035693
8.5	2.1	1.143535	1.307672297
9	2.7	0.969078327	0.939112803

DATA ANALYSIS

1) Movie Genre Analysis:

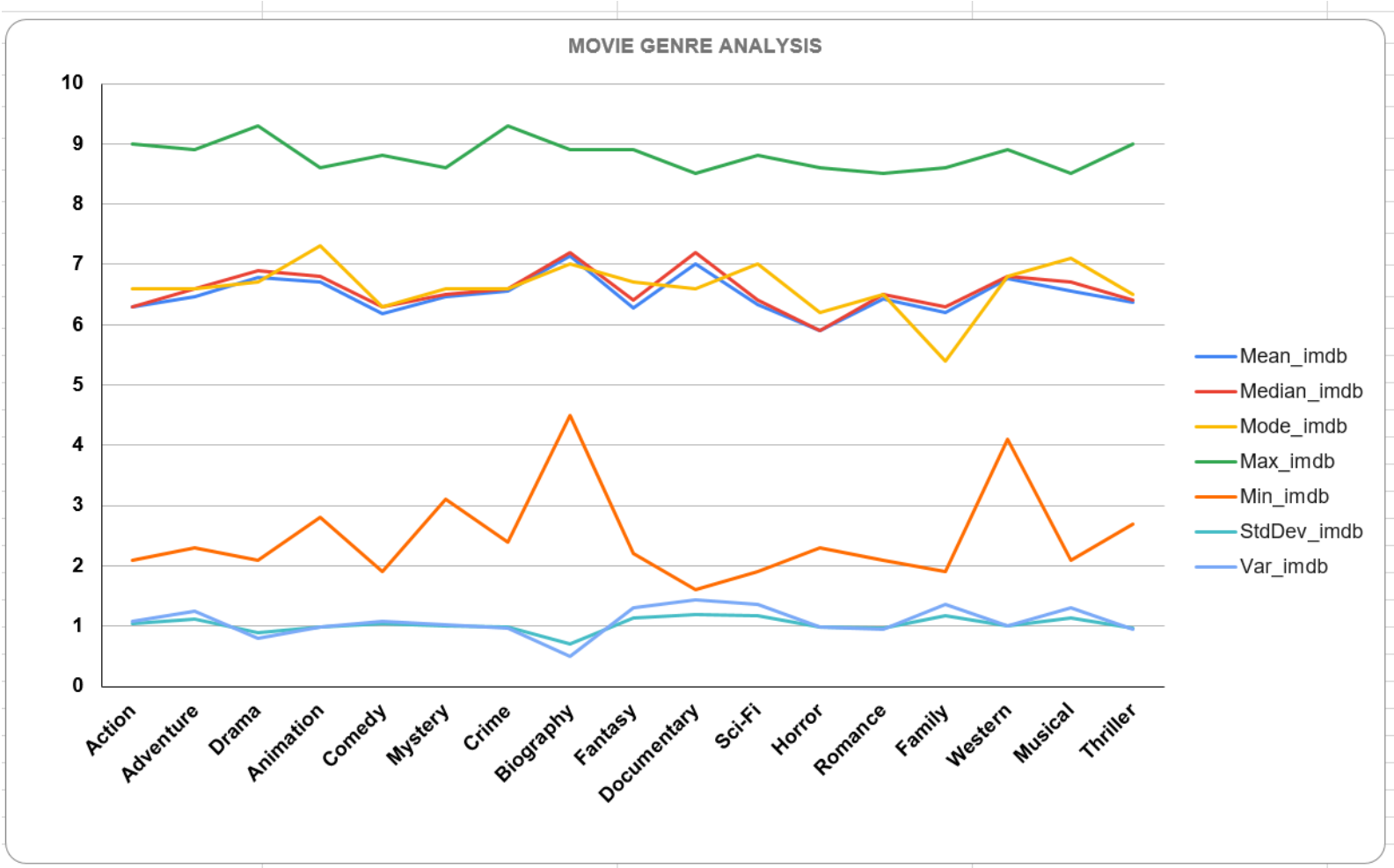
Result:



DATA ANALYSIS

1) Movie Genre Analysis:

Result:



DATA ANALYSIS

2) Movie Duration Analysis:

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

Results:

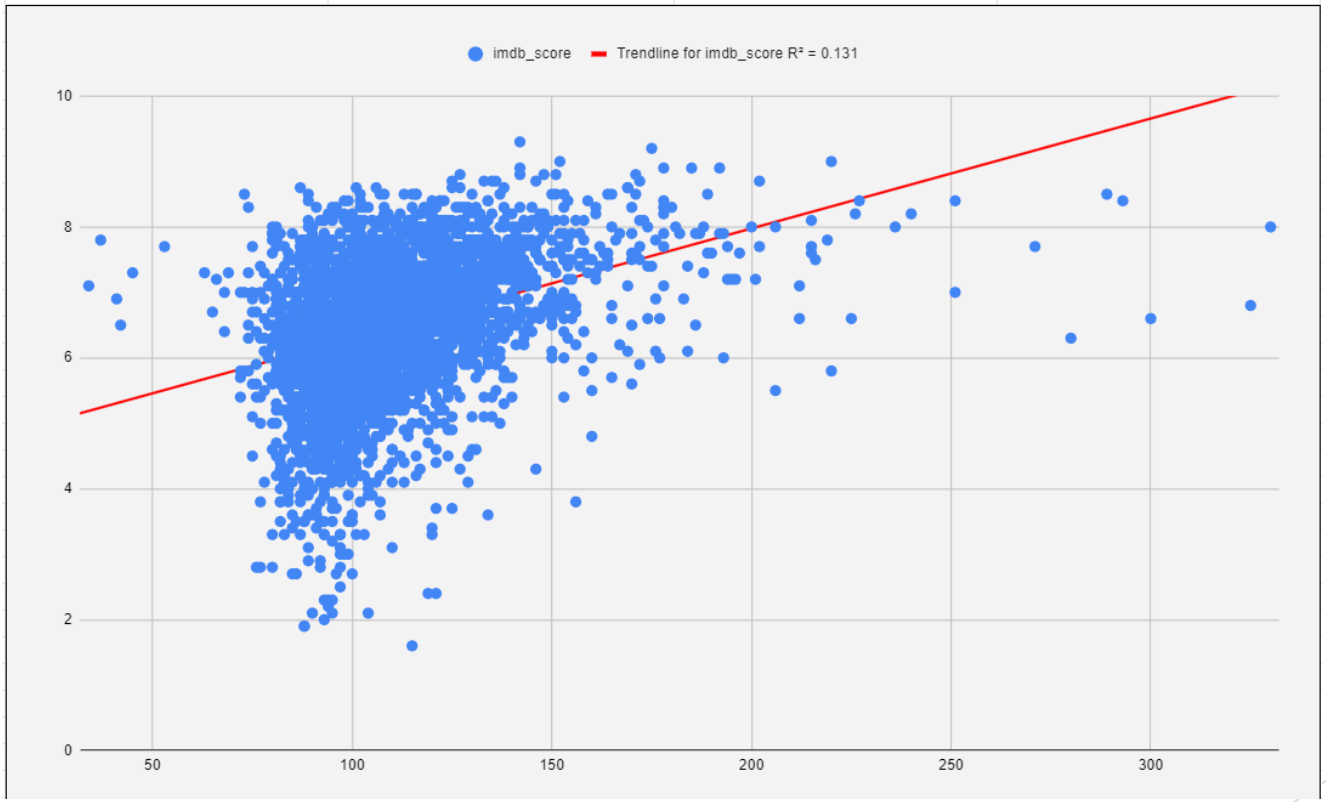
Operations	Values
Mean	109.808505
Median	105
Mode	101
Standard Devation	22.763201
Variance	518.16332

DATA ANALYSIS

2) Movie Duration Analysis:

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

Results:



DATA ANALYSIS

3) Movie Language Analysis:

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

Results:

Language	No_of_movies	Average_imdb	Median_imdb	Var_imdb	StdDev_imdb
English	3606	6.421436495	6.5	1.107753941	1.052498903
French	37	7.286486486	7.2	0.3150900901	0.5613288609
Spanish	26	7.05	7.15	0.6826	0.8261961026
Mandarin	14	7.021428571	7.25	0.5864285714	0.765786244
German	13	7.692307692	7.7	0.4107692308	0.6409128106
Japanese	12	7.625	7.8	0.8093181818	0.8996211324
Hindi	10	6.76	7.05	1.236	1.111755369
Cantonese	8	7.2375	7.3	0.1941071429	0.4405759218
Italian	7	7.185714286	7	1.334761905	1.155318962
Korean	5	7.7	7.7	0.325	0.5700877125
Portuguese	5	7.76	8	0.958	0.9787747443
Norwegian	4	7.15	7.3	0.33	0.5744562647
Dutch	3	7.566666667	7.8	0.1633333333	0.4041451884
Thai	3	6.633333333	6.6	0.2033333333	0.4509249753
Danish	3	7.9	8.1	0.28	0.5291502622
Hebrew	3	7.5	7.3	0.19	0.4358898944

DATA ANALYSIS

4) Movie Director Analysis:

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

Result:

Director	Average_imdb	percentile	Count_movies
Tony Kaye	8.6	0.999	1
Charles Chaplin	8.6	0.999	1
Alfred Hitchcock	8.5	0.997	1
Ron Fricke	8.5	0.997	1
Damien Chazelle	8.5	0.997	1
Majid Majidi	8.5	0.997	1
Sergio Leone	8.433333333	0.996	3
Christopher Nolan	8.425	0.995	8
S.S. Rajamouli	8.4	0.993	1
Richard Marquand	8.4	0.993	1
Asghar Farhadi	8.4	0.993	1
Marius A. Markevicius	8.4	0.993	1
Lee Unkrich	8.3	0.991	1
Fritz Lang	8.3	0.991	1
Lenny Abrahamson	8.3	0.991	1
Billy Wilder	8.3	0.991	1

DATA ANALYSIS

5) Movie Budget Analysis:

Explore the relationship between movie budgets and their financial success.

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Results:

Movies	Profits in Millions
AvatarÂ	523505847
Jurassic WorldÂ	502177271
TitanicÂ	458672302
Star Wars: Episode IV - A New HopeÂ	449935665
E.T. the Extra-TerrestrialÂ	424449459
The Lion KingÂ	377783777
The Jungle BookÂ	375290282
Star Wars: Episode I - The Phantom Mena	359544677
The Dark KnightÂ	348316061
The Twilight Saga: Breaking Dawn - Part 2	344597846

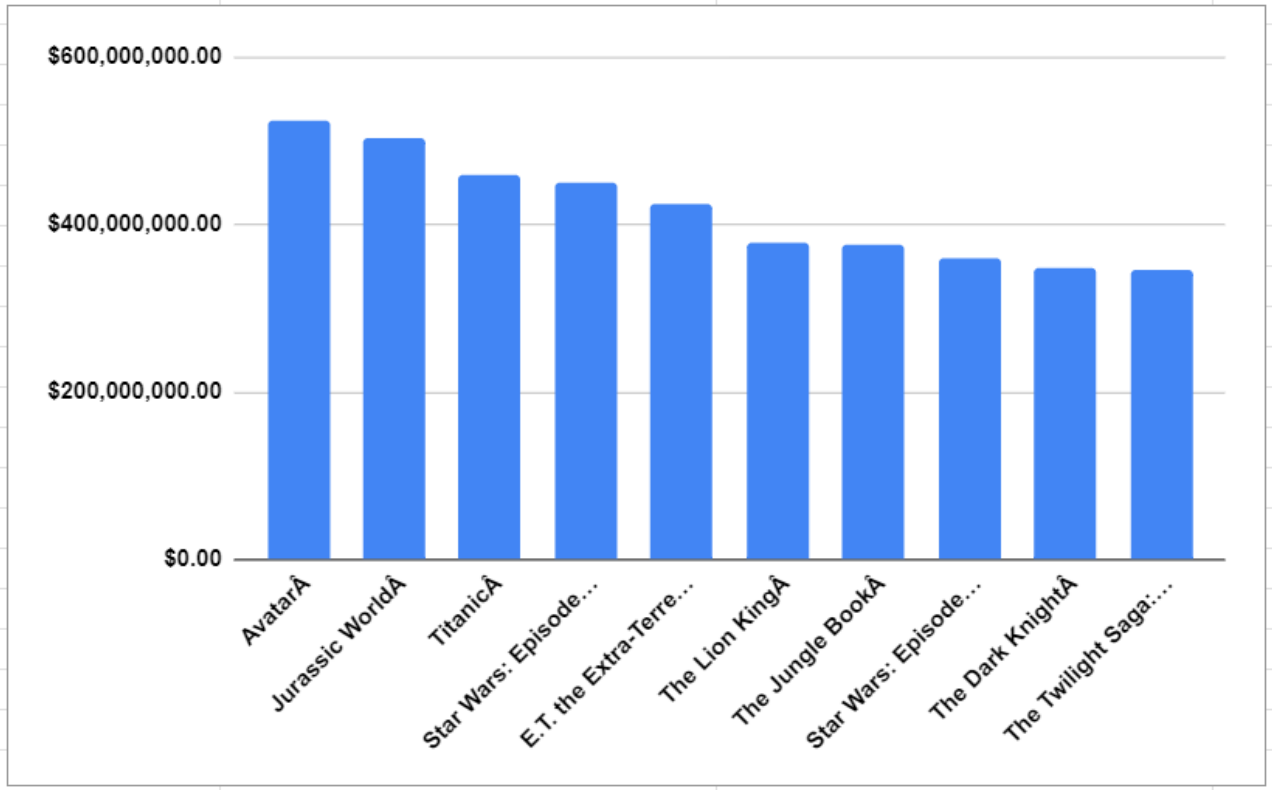
DATA ANALYSIS

5) Movie Budget Analysis:

Explore the relationship between movie budgets and their financial success.

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Results:



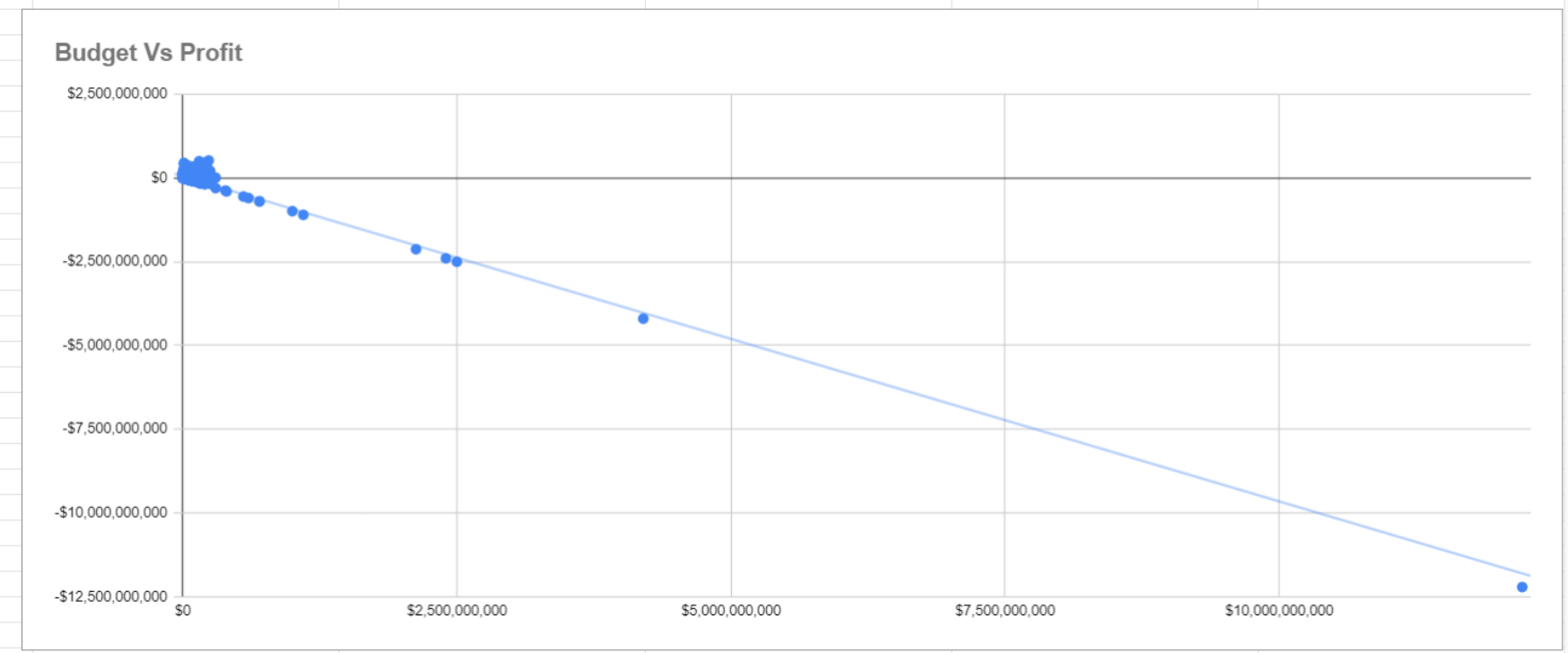
DATA ANALYSIS

5) Movie Budget Analysis:

Explore the relationship between movie budgets and their financial success.

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Correlation Graph:



Insights

- The Most common movie genres from the dataset are Drama, Comedy, Thriller and Action.
- The Average duration of a Movie is 109 minutes. The trendline between the duration vs imdb score is elevated upward with R^2 = 0.131
- The Most common languages used in the movies are English, French, Spanish, Mandarin and German. I have also Observed that the languages Telugu and Persian have the highest average imdb score.
- I have identified that Tony Kaye, Charles Chaplin, Alfred Hitchcock, Ron Fricke, Damien Chazelle, Majid Majidi, Sergio Leone, Christopher Nolan, SS Rajamouli and Richard Marquand are the top 10 directors with average imdb score >=8.4
- The Top-5 with highest profits are Avatar, Jurassic World, Titanic, Star Wars: Episode IV - A New Hope and E.T. The Extra-Terrestrial. The Correlation between budget and gross is positive.

Conclusion

- The Cleaned Dataset:

https://docs.google.com/spreadsheets/d/1QZcrT5BZhKOTA9_pnpaorlPPRI7wW4BCzT_FyVd0YQY/edit?usp=sharing

- The Results Dataset:

https://docs.google.com/spreadsheets/d/1X-ak_kajhbePb1_NzvtnA8kmErCSwUN9yEJqvIdsac0/edit?usp=sharing

Thank You