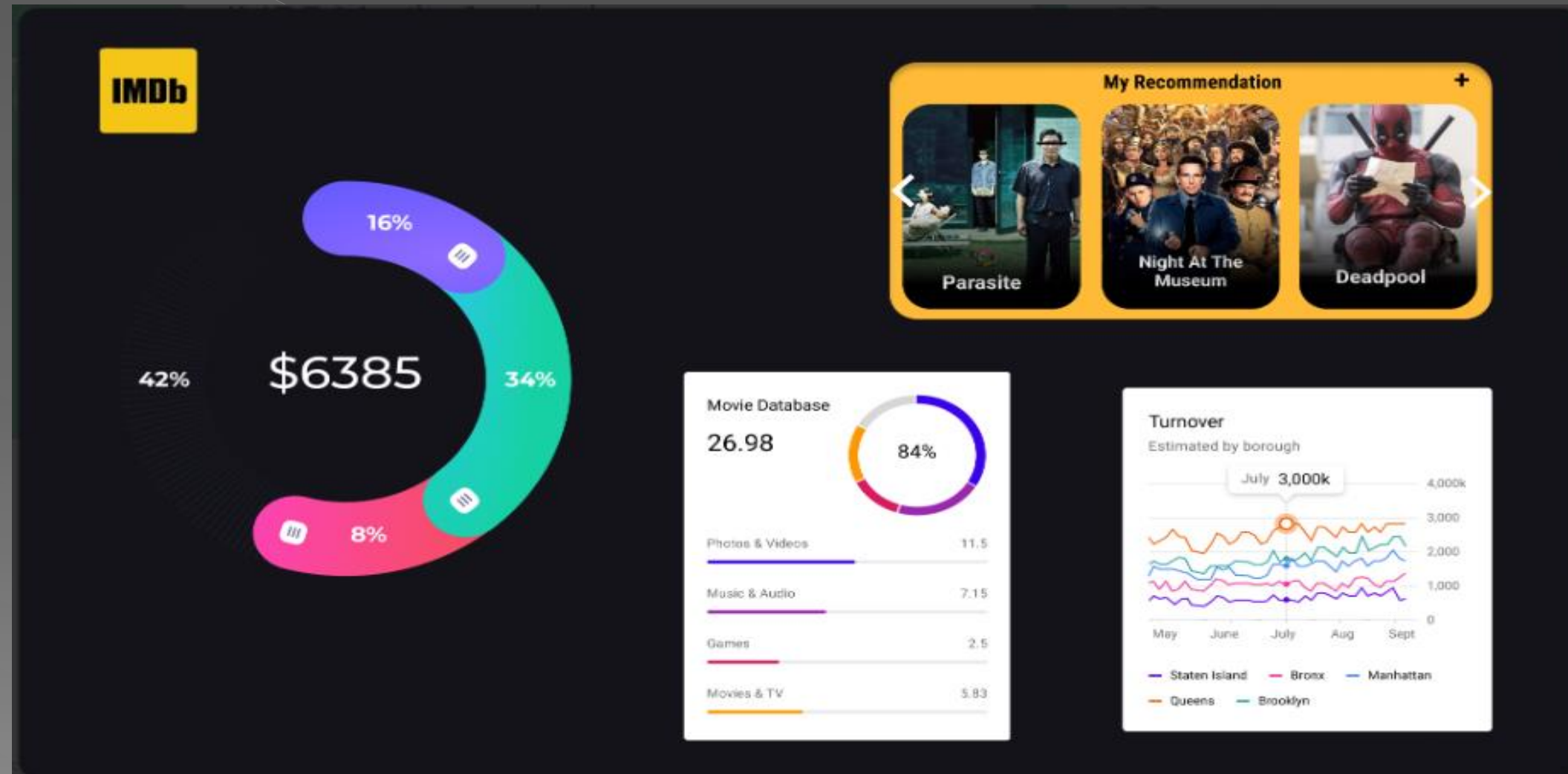# IMDB Movie Analysis

- **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.

- **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. First, I have deleted the columns which have no relation to our project and don't provide any valuable insights. Then, I have deleted all the blank and duplicate rows from the dataset.

- **Tech-Stack Used:** Used Microsoft Excel for data cleaning and visualization.

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.

| Task-1: Movie Genre Analysis									
Genre	No_of_movies	Mean_imdb	Median_imdb	Mod_imdb	Max_imdb	Min_imdb		StdDev_imdb	Var_imdb
Action	935	6.285989305	6.6	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
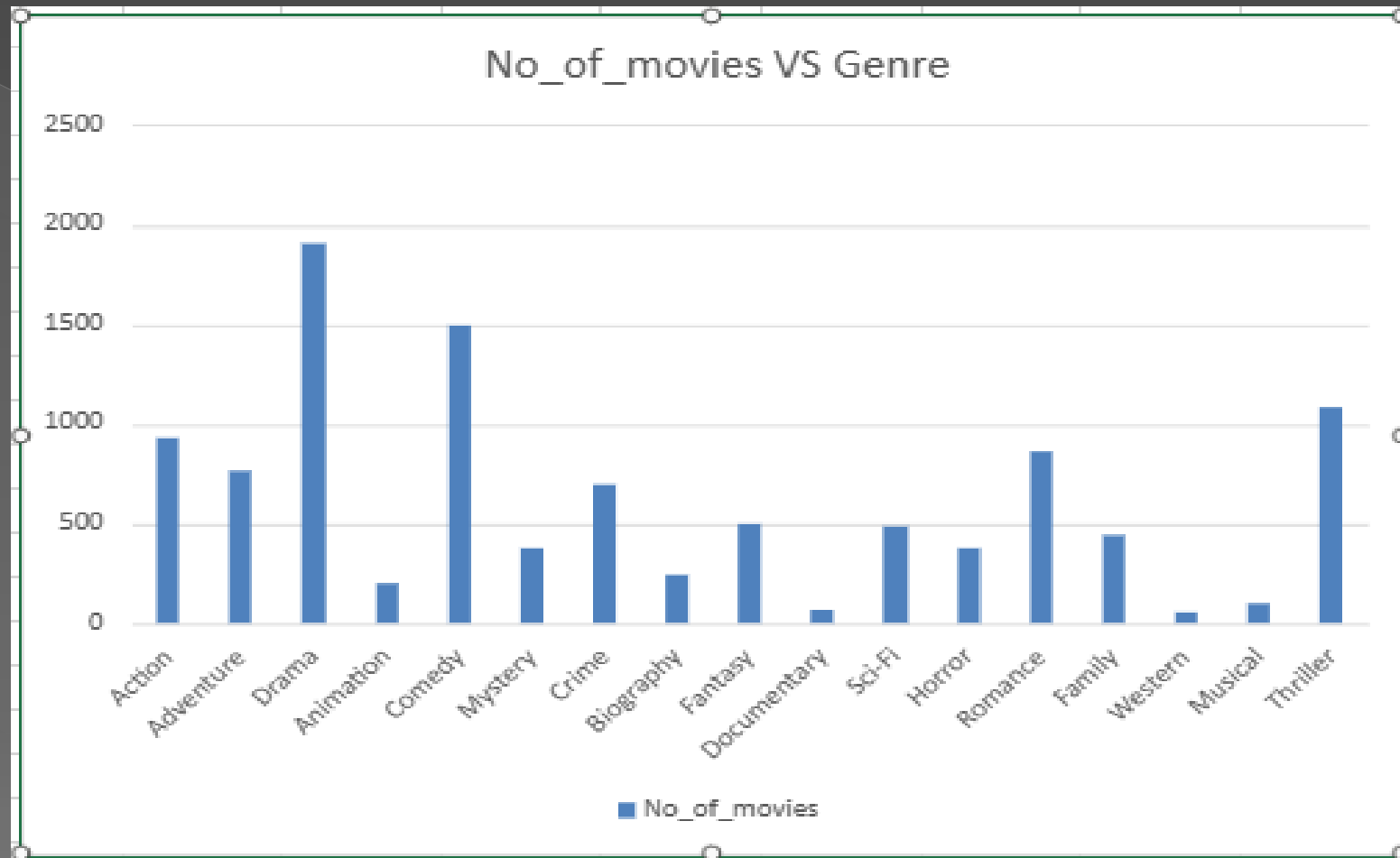
⊙ **Insights:** The Most common movie genres from the dataset are Drama, Comedy, Thriller and Action.

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.
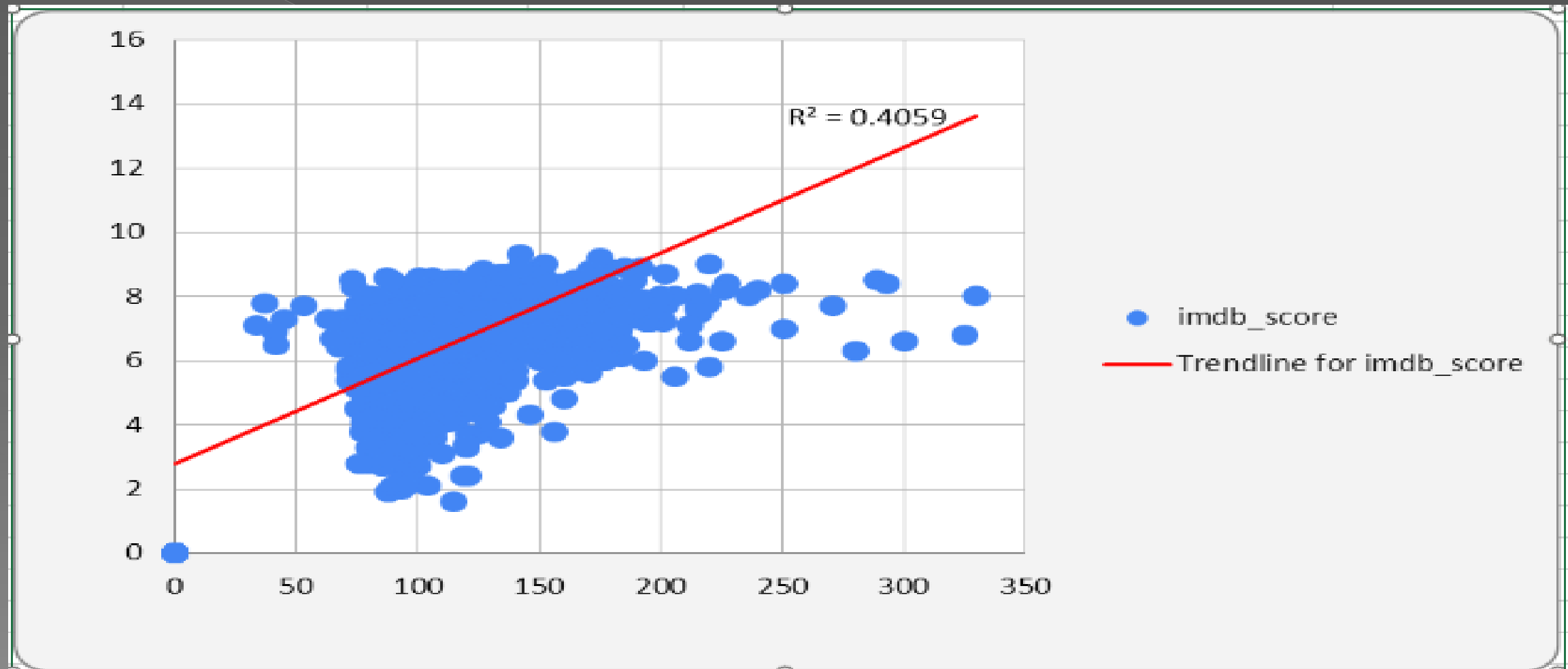
Task-2 : Movie Duration Analysis	
Operations	Values
Mean	109.808505
Median	105
Mode	101
Standard Deviation	22.763201
Variance	518.16332

◉ **Insights:** The Average duration of a Movie is 105 minutes. The trendline between the duration vs imdb score is elevated upward.

3. Language Analysis:

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.

Task-3: Language Analysis

Language	No._Of_Movies	Mean	Median	StdDev
English	3606	6.421436495	6.5	1.052498903
French	3	7.692307692	7.2	0.561328861
Spanish	14	7.021428571	7.15	0.826196103
Mandarin	37	7.286486486	7.25	0.765786244
Japanese	5	7.7	7.8	0.899621132
Hindi	13	6	7.05	1.111755369
Cantonese	12	7.76	7.3	0.440575922
Portuguese	3	7.4	8	0.978774744
Norwegian	3	4.3	7.3	0.574456265
Thai	8	6.95	6.6	0.450924975
Dutch	7	7.566666667	7.8	0.404145188
Danish	3	6.7	8.1	0.529150262
Aboriginal	26	7.05	6.95	0.777817459
Hungarian	10	7.1	7.1	0.555344332
Kazakh	5	7.2375	6	0.450924933
Icelandic	4	6.9	6.9	0.222468944
Filipino	3	7.625	6.7	0.658583373

◉ **Insights:** The Most common languages used in the movies are English, French, Spanish, Mandarin and German.

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.

Task-4: Director Analysis

Director	Average_imdb	Percentile	Count_movies
Christopher Nolan	8.425	0.995	8
S.S. Rajamouli	8.4	0.993	1
Richard Marquand	8.4	0.993	1
Sergio Leone	8.433333333	0.996	3
Alfred Hitchcock	8.5	0.996	1
Tony Kaye	8.6	0.998	1
Ron Fricke	8.5	0.996	1
Damien Chazelle	8.5	0.996	1
Charles Chaplin	8.6	0.998	1
Asghar Farhadi	8.4	0.993	1
Marius A. Markevicius	8.4	0.993	1
Majid Majidi	8.5	0.996	1

◉ **Insights:** I have identified that Christopher Nolan, S.S. Rajamauli, Richard Marquand, Sergio Leone , Alfred Hitchcock, Tony Kaye, Ron Fricke, Damien Chazelle, Charles Chaplin , Asghar Farhadi, Marius A. Markevicious and Majid Majidi are the top 10 directors with average imdb score >=8.4.

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.

Task-5: Budget Analysis

Top Movies	High Profits
AvatarÂ	264706375
Jurassic WorldÂ	262029560
TitanicÂ	253501675
Star Wars: Episode IV - A New HopeÂ	252000000
E.T. the Extra-TerrestrialÂ	250736600
The AvengersÂ	248505540
The Lion KingÂ	246623634
Star Wars: Episode I - The Phantom MenaceÂ	246478898
The Dark KnightÂ	242313371
The Hunger GamesÂ	236437427

Insights: The Top-5 with highest profits are Avatar, Jurassic World, Titanic, Star Wars: Episode IV - A New Hope and E.T. The Extra-Terrestrial.

- While working on this project, I have achieved several accomplishments. I successfully analyzed the IMDB movies data, identified the most liked genre of movies, as well as average IMDB score by each movie. Additionally, I visualized the data through charts and graphs to showcase genre vs No. Of Movies, trenline for imdb score and which movie got high profit. This project has enhanced my skills in data analysis, statistical calculations, and data visualization techniques. It has also provided me with valuable experience in deriving insights from real-world datasets, which will further strengthen my capabilities as a data analyst.

- Link for cleaned dataset:

https://docs.google.com/spreadsheets/d/1-9kyBWIbwyRGkFYTY5vBjTZrnl-4IaES/edit?usp=sharing&ouid=101431809048624548912&rtpof=true&sd=true

- **Link for resultant dataset:**

https://docs.google.com/spreadsheets/d/1wv-ozz3JOUssP6KJgV-dYtLN5OBioItV/edit?usp=sharing&ouid=101431809048624548912&rtpof=true&sd=true