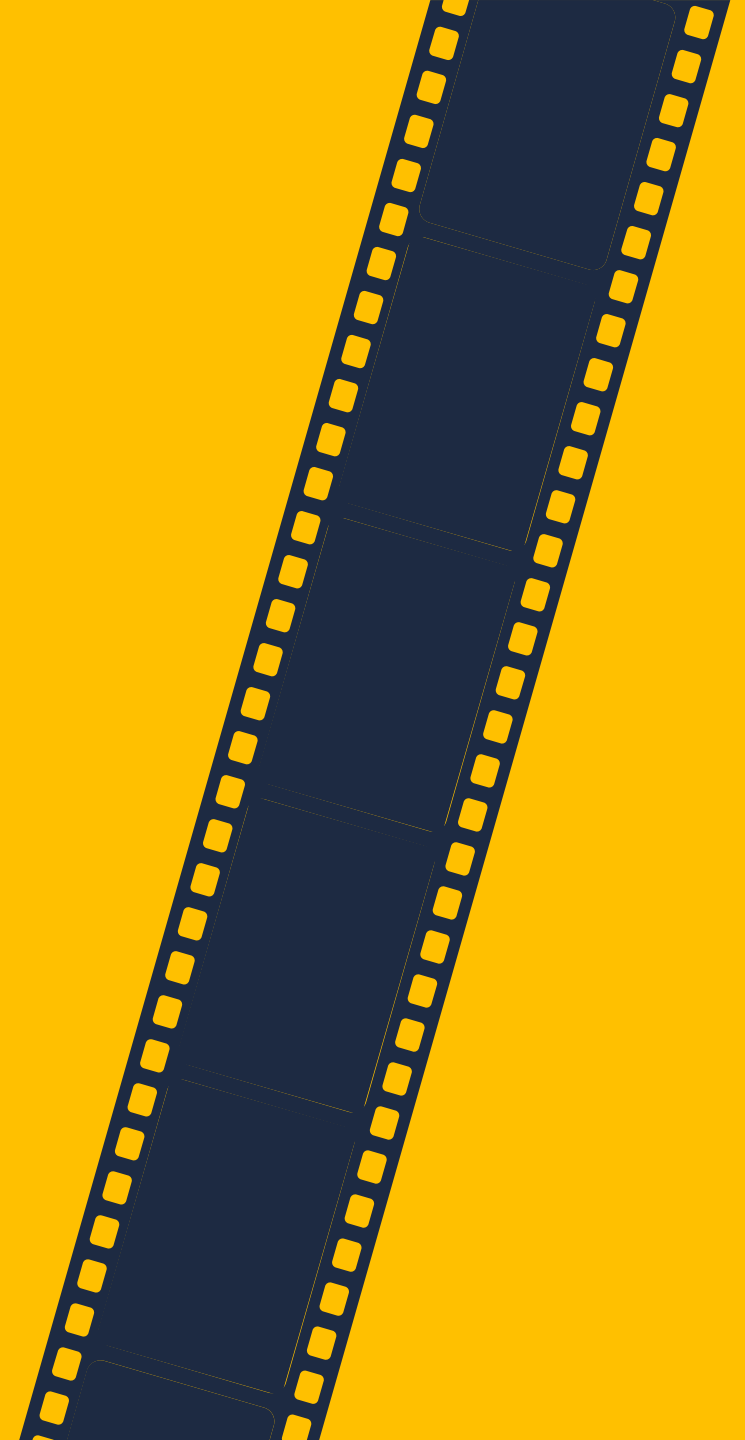
IMDB Movie Analysis

Trainity Project

Shahnawaz Akhtar
shahnawazakhtar2@gmail.com

Contents:

Excel Tasks:



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 2019**

My Approach:

I thoroughly reviewed the dataset, which consists of 28 columns and 5043 rows. I noticed the presence of unwanted columns, null values, and blank rows. Therefore, I decided to clean the dataset comprehensively.

1.Column Selection: I removed columns that were irrelevant to our project and did not provide any valuable insights. After this step, I retained only 9 columns: director’s name, duration, movie title, genre, budget, gross, IMDb rating, language, and country.

2.Blank Row Removal: I identified many blank rows in the dataset. To locate them, I used the "Find & Select" feature, clicked on "Go to Special," and selected the "Blanks" option. This highlighted all the blank rows. I then used the shortcut "CTRL + -" and chose the "Entire row" option to delete all the blank rows from the dataset.

3.Duplicate Row Removal: Finally, I removed any duplicate rows present in the dataset. After completing these steps, I was left with a cleaned dataset consisting of 9 columns and 3786 rows. The cleaned dataset is provided below.

DATA ANALYSIS

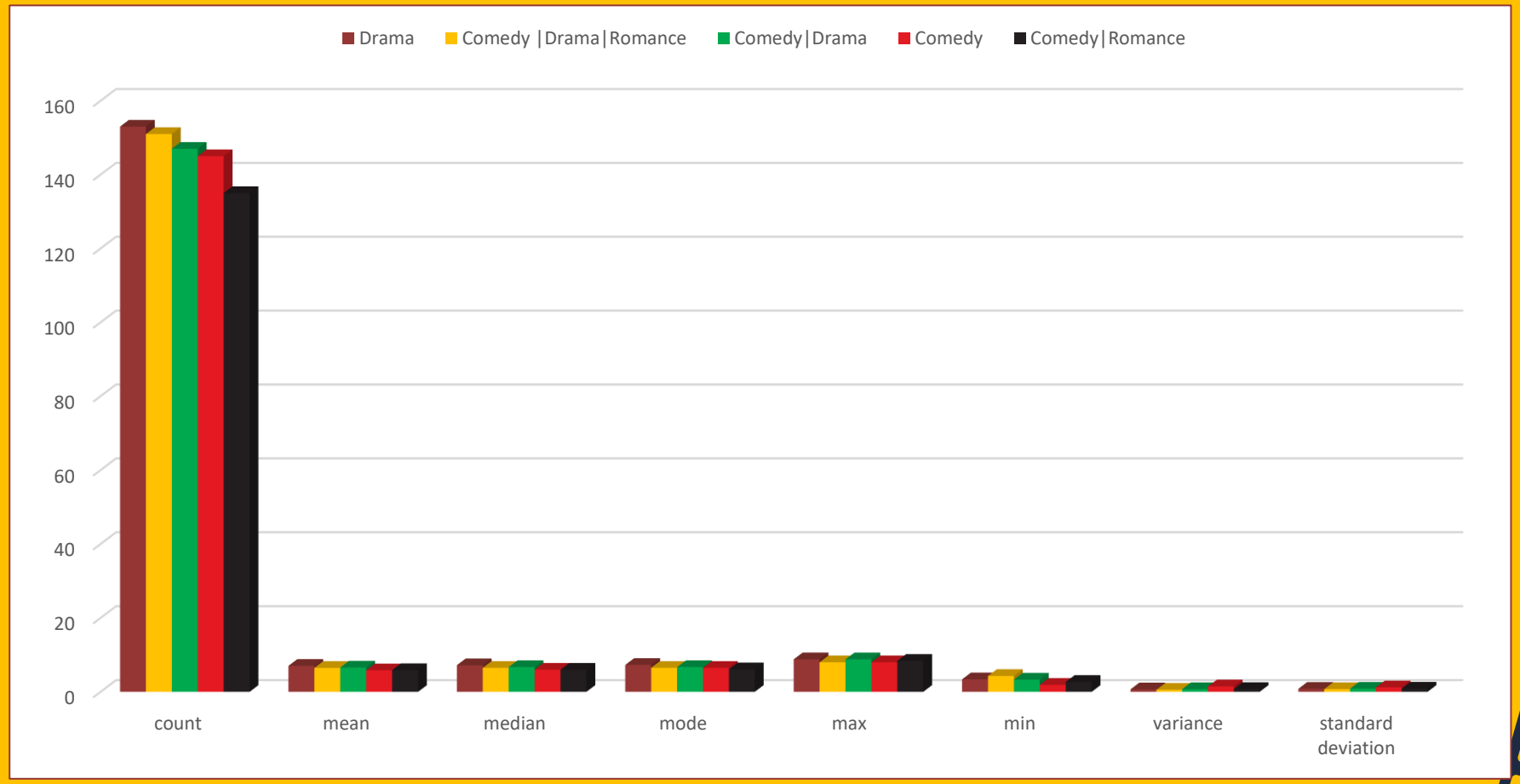
Task-A) Movie Genre Analysis:

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:

genres	count	mean	median	mode	max	min	variance	standard deviation
Drama	153	7.04183	7.2	7.3	8.8	3.4	0.687055	0.828887522
Comedy|Drama|Romance	151	6.494702	6.5	6.5	8	4.3	0.562772	0.750181141
Comedy|Drama	147	6.583673	6.7	6.7	8.8	3.3	0.7348	0.857204825
Comedy	145	5.84069	6	6.5	8	1.9	1.481875	1.217322686
Comedy|Romance	135	5.896296	6	6.1	8.4	2.7	0.76827	0.87650999

Graph representation:



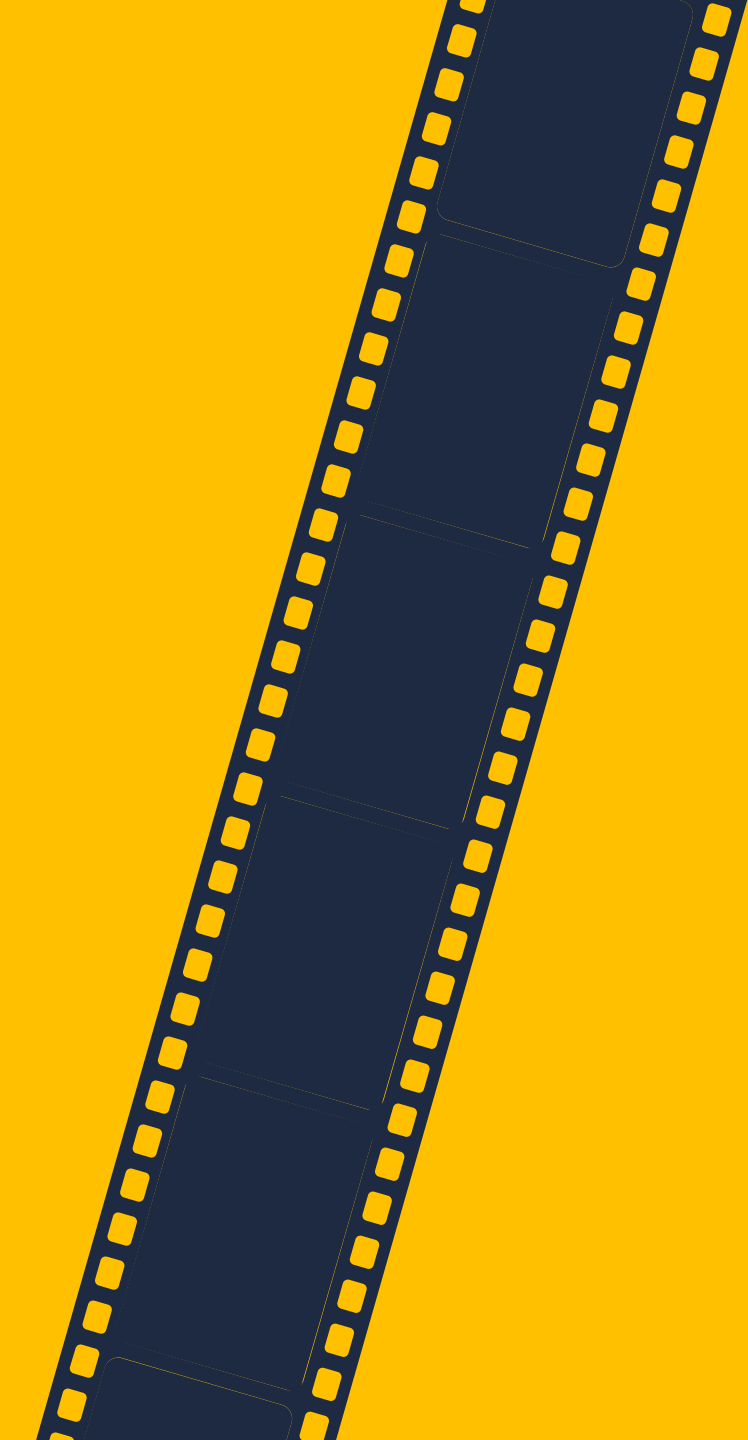
DATA ANALYSIS

Task-B) Movie Duration Analysis:

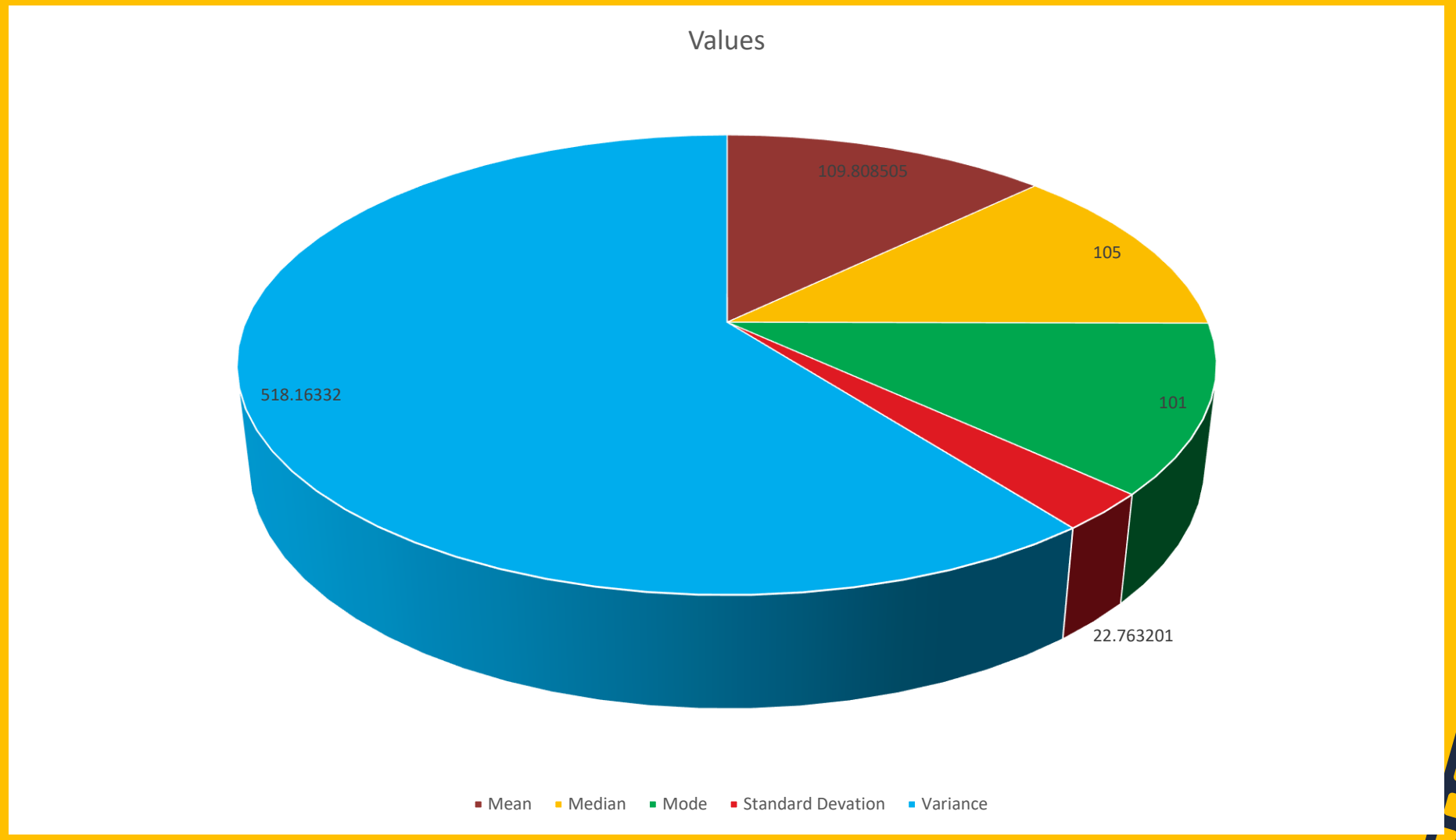
Analyze the distribution of movie durations and its impact on the IMDB score.
Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Results:

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



Graph representation:



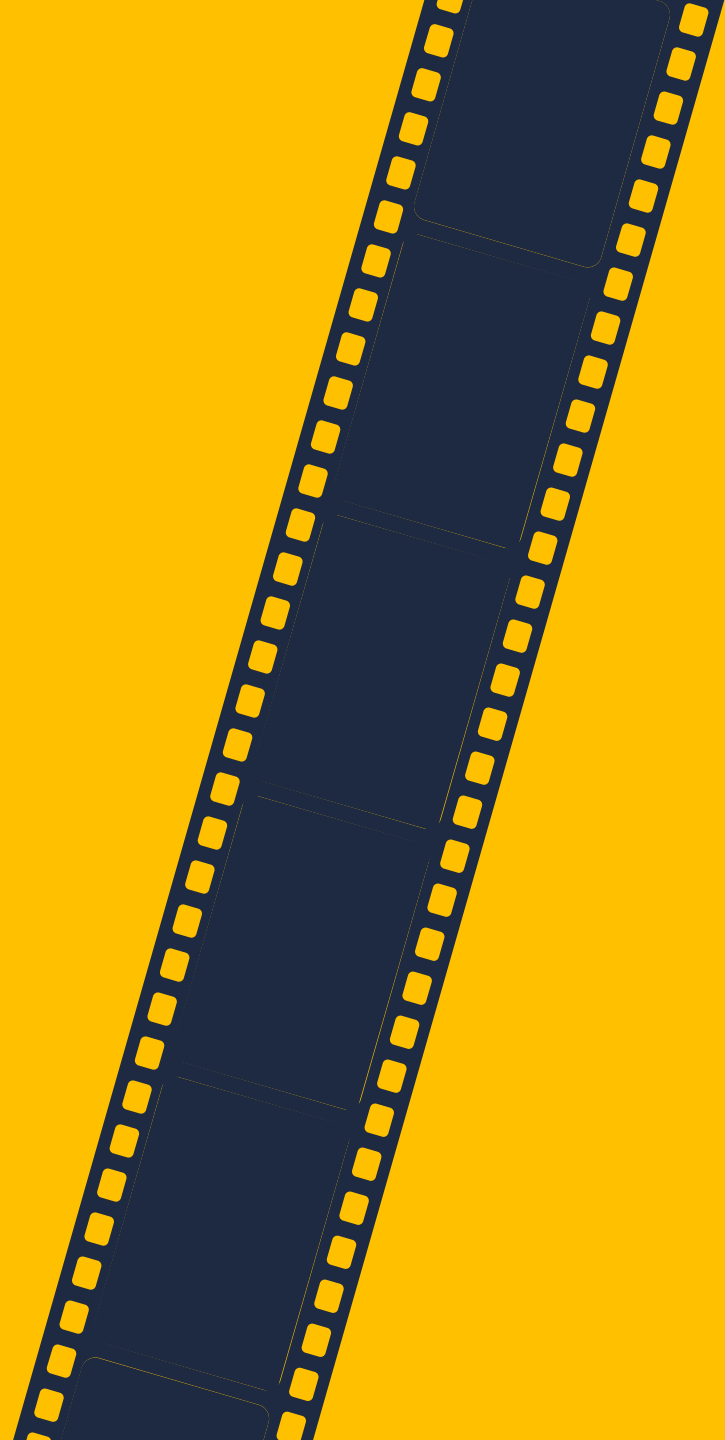
DATA ANALYSIS

Task-C) Movie Language Analysis:

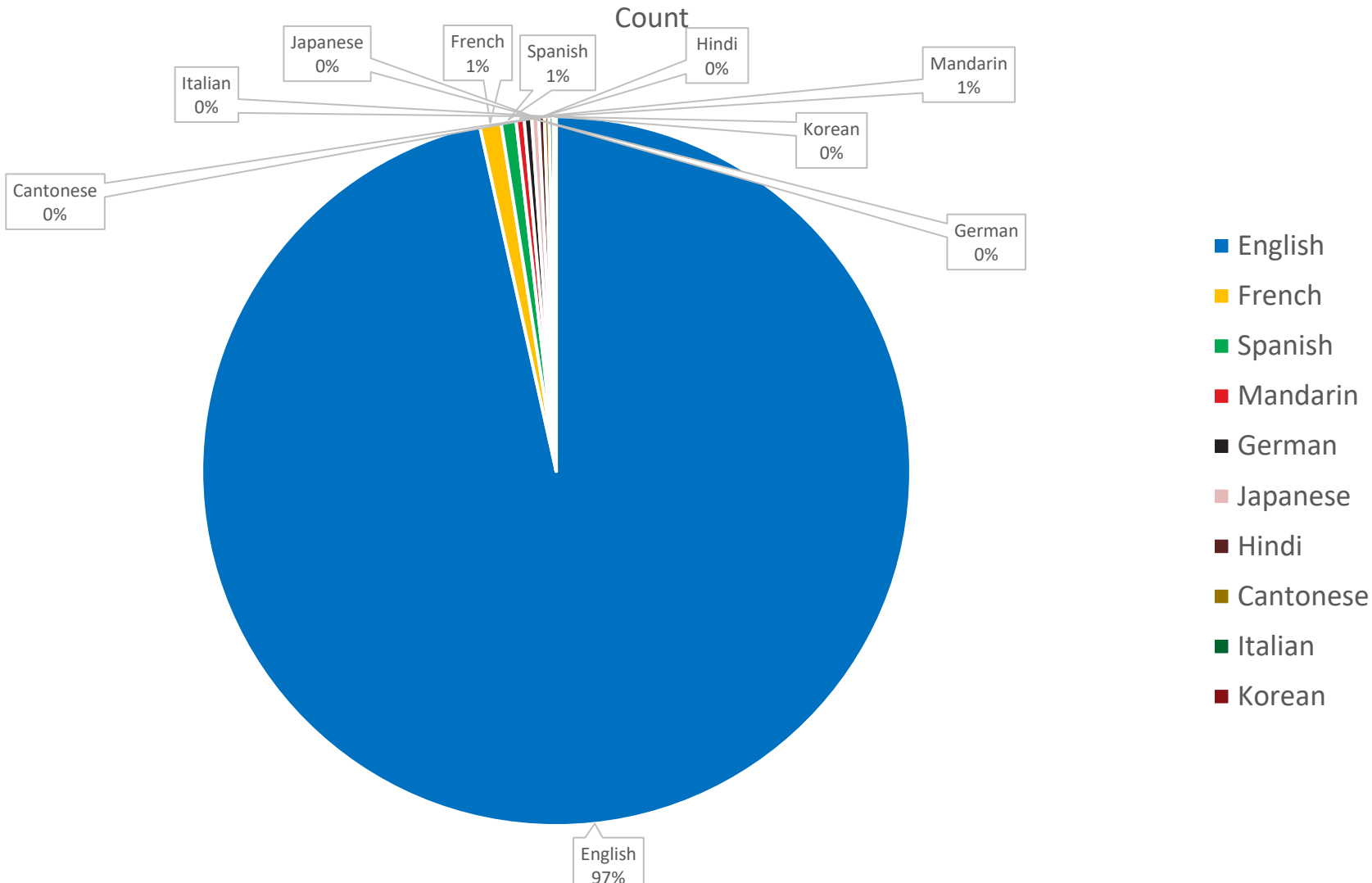
Situation: Examine the distribution of movies based on their language.
Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Results:

Language	Count	Mean	Median	Standard Deviation
English	3668	6.423909	6.5	1.048750752
French	37	7.286486	7.2	0.561328861
Spanish	26	7.05	7.15	0.826196103
Mandarin	14	7.021429	7.25	0.767586244
German	13	7.692308	7.7	0.640912811
Japanese	12	7.625	7.8	0.899621132
Hindi	10	6.76	7.05	1.111755369
Cantonese	8	7.2375	7.3	0.44057592
Italian	7	7.185714	7	1.155318962
Korean	5	7.7	7.7	0.570087713



Graph representation:



DATA ANALYSIS

Task-D) Movie Director Analysis:

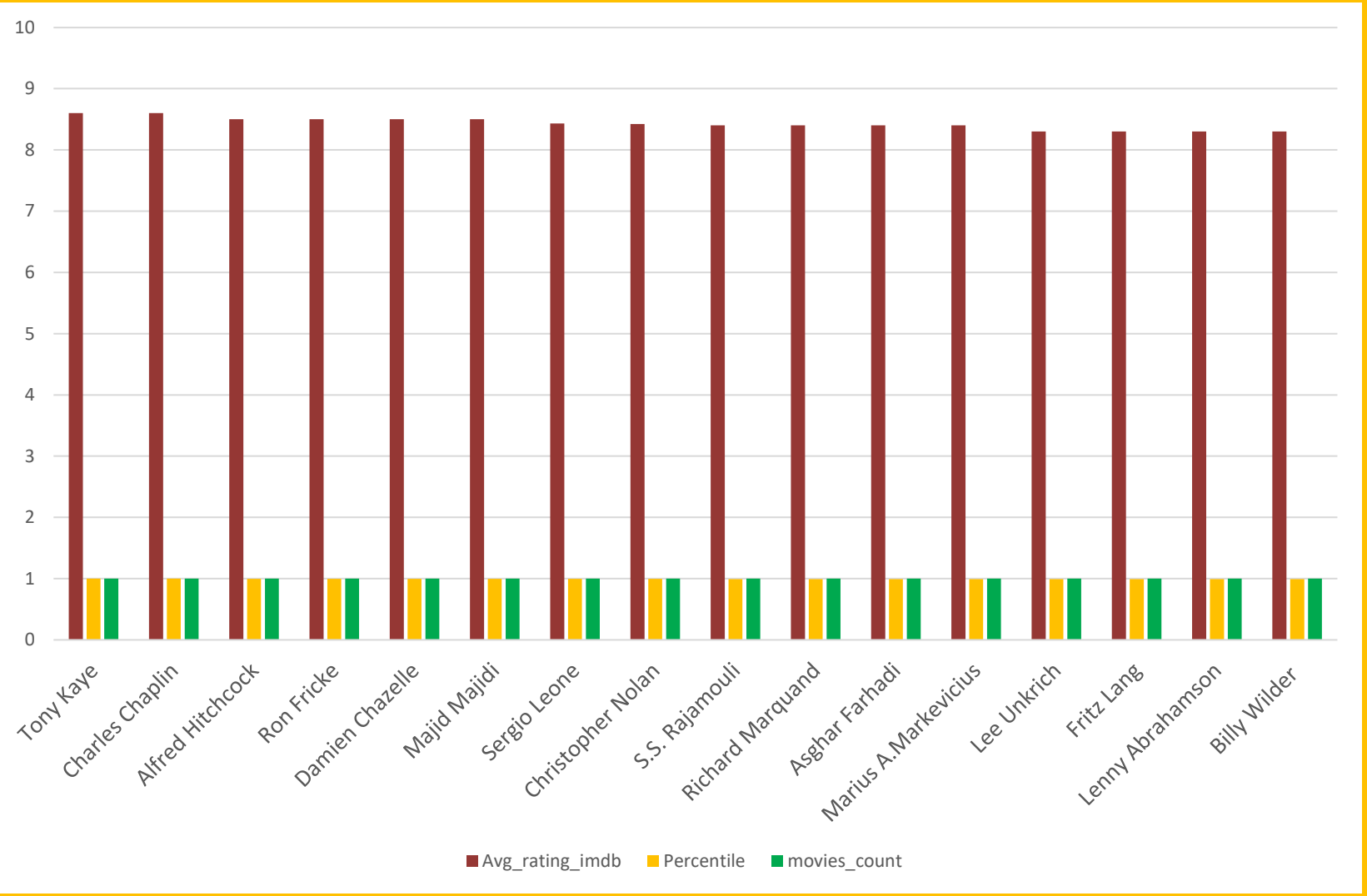
Influence of directors on movie ratings.

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

Results:

Director	Avg_rating_imdb	Percentile	movies_count
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	1
Christopher Nolan	8.425	0.995	1
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

Graph representation:



DATA ANALYSIS

Task-E) Movie Budget Analysis:

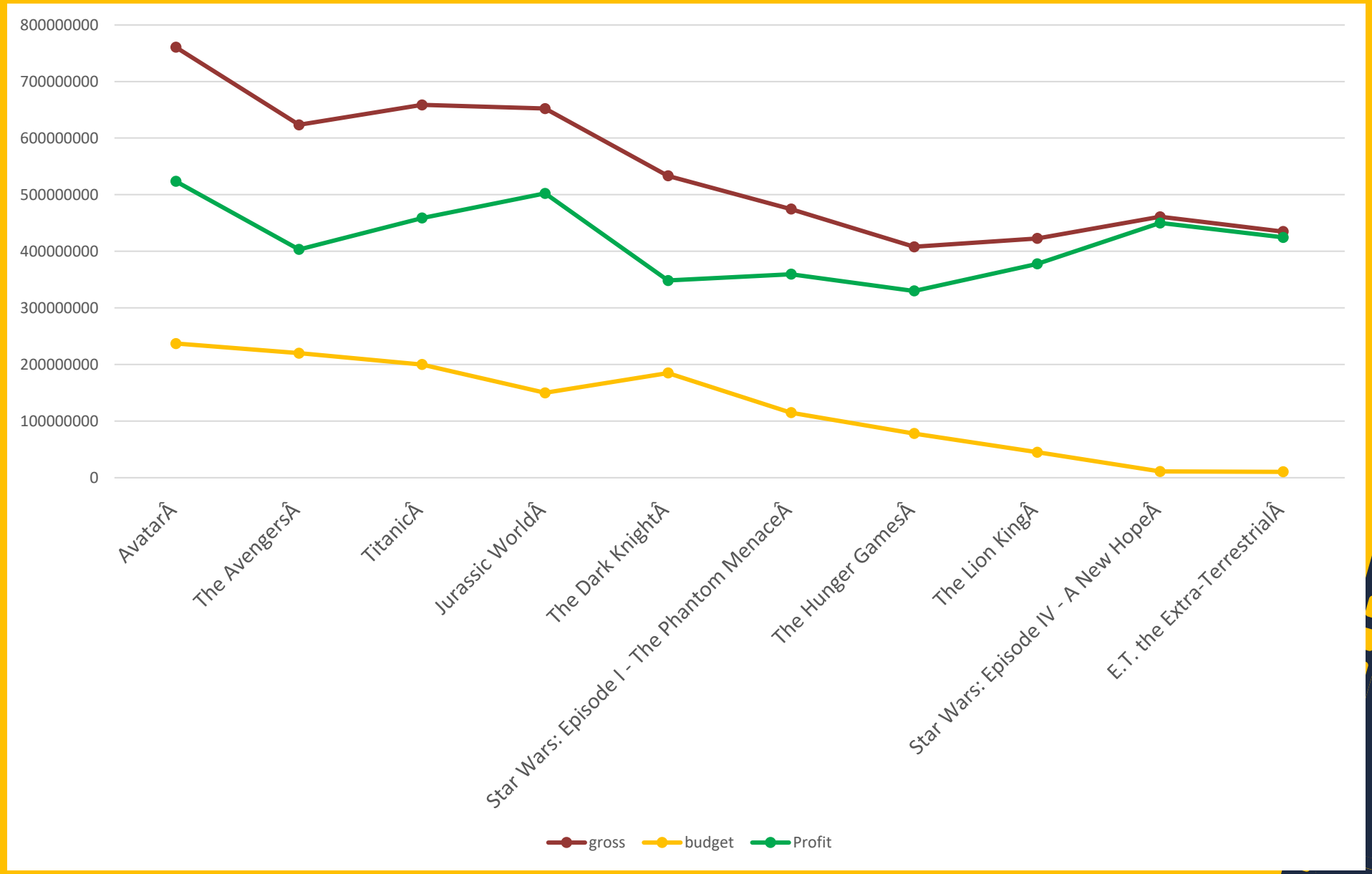
Explore the relationship between movie budgets and their financial success.

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

Results:

movie_title	gross	budget	Profit
AvatarÂ	760505847	237000000	523505847
The AvengersÂ	623279547	220000000	403279547
TitanicÂ	658672302	200000000	458672302
Jurassic WorldÂ	652177271	150000000	502177271
The Dark KnightÂ	533316061	185000000	348316061
Star Wars: Episode I - The Phantom MenaceÂ	474544677	115000000	359544677
The Hunger GamesÂ	407999255	78000000	329999255
The Lion KingÂ	422783777	45000000	377783777
Star Wars: Episode IV - A New HopeÂ	460935665	11000000	449935665
E.T. the Extra-TerrestrialÂ	434949459	10500000	424449459

Graph representation:



Insights

- **Drama Dominates: The drama genre is the most prevalent, with 153 movies. It also maintains a solid average IMDb rating of 7.04, indicating a generally favorable reception.**
- **Comedy Genre Variations: Various combinations of comedy (Comedy|Drama, Comedy|Romance) show mixed performance, with average IMDb ratings ranging from 5.84 to 6.58. The standard deviations are higher in pure comedy, suggesting varying audience reactions.**
- **Movie Duration Impact: The average movie duration is approximately 110 minutes, with a median of 105 minutes. Movies around this duration tend to perform well on IMDb, showing a correlation between standard feature-length films and audience satisfaction.**
- **Language Influences Ratings: English-language films dominate the dataset, but movies in German and Japanese have notably higher average IMDb ratings (7.69 and 7.62, respectively), suggesting strong audience approval despite fewer entries.**
- **Top Directors: Directors like Tony Kaye and Charles Chaplin have an average IMDb rating of 8.6, placing them in the 99th percentile, indicating that their films are highly regarded by audiences.**
- **Budget vs. Profitability: Blockbuster movies like *Avatar* and *Titanic* show that higher budgets correlate with massive gross earnings, but smaller budget films like *E.T. the Extra-Terrestrial* can achieve impressive profit margins due to lower initial investments.**
- **Romantic Comedies Struggle: Despite being common, genres like Comedy|Romance have lower average IMDb ratings (around 5.89), suggesting that these films may not resonate as well with IMDb users.**
- **Cultural Appeal of Non-English Films: Non-English movies, particularly those in French and German, tend to have higher average ratings (7.28 and 7.69, respectively), indicating their strong appeal to niche audiences.**
- **Success of Classic Directors: Renowned directors like Alfred Hitchcock and Sergio Leone maintain very high average IMDb ratings (8.5+), showing the enduring legacy and influence of classic cinema.**
- **Blockbuster Formula: The analysis shows a clear pattern where high-budget, franchise-driven films (e.g., *The Avengers*, *Jurassic World*) consistently generate high profits, reinforcing the blockbuster formula's success.**

Thank You !