IMDb MOVIE ANALYSIS

PROJECT DESCRIPTION

IMDb stands for the **Internet Movie Database**. It is an online database that was founded in 1990 and is now one of the most comprehensive sources of information related to movies, TV shows, actors, directors, and other entertainment content.

I was given a dataset with information related to movies from 1916 to 2016. My task was to clean the data and analyze the data related to **genre**, **budget**, **movie duration**, **director**, and **language**.

APPROACH

STEP 1: I cleaned the data that was provided to me.

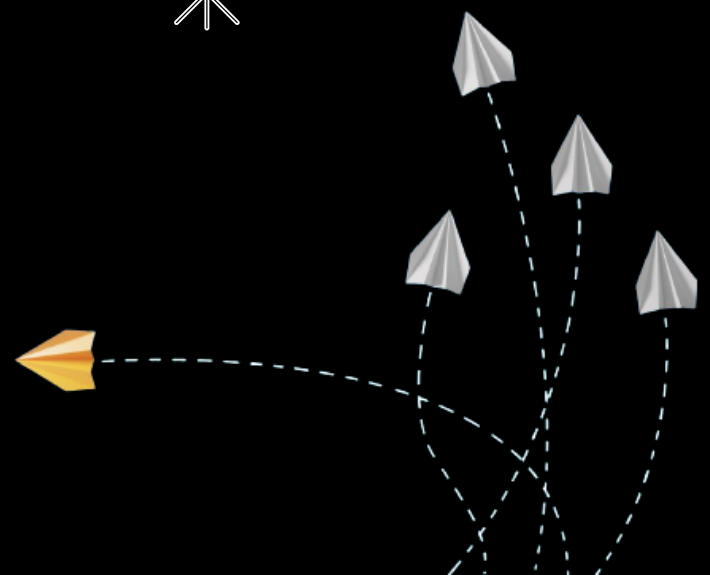
STEP 2: Analyzed the cleaned data.

STEP 3: Gained insights from the analysis.

STEP 4: Created charts and graphs for those insights.

STEP 5: Created a presentation to outsource my information.

STEP 6: Presented my presentation.



TECH-STACK USED

Excel(Microsoft 365 version)

Used to analyze the dataset provided to me.

PowerPoint(Microsoft 365 version)

Used to create and present the presentation.

INSIGHTS

A. Movie genre analysis:

Determine the most common genres of movies in the dataset.

B. Movie duration analysis:

Analyze the distribution of movie durations and its impact on the IMDb scores.

C. Language analysis:

Determine the most common languages used in movies and analyze the impact on IMDb scores.

D. Director analysis:

Identify the top directors based on their average IMDb scores.

E. Budget analysis:

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

A. MOVIE GENRE ANALYSIS

INSIGHTS: Movie genre analysis

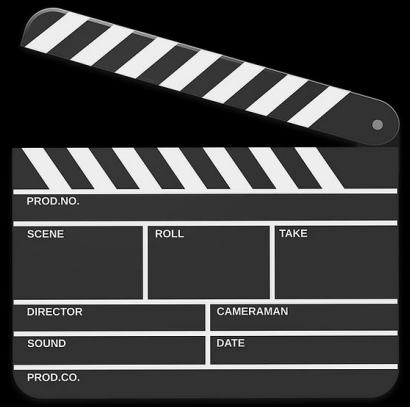
STEPS PERFORMED

- Firstly, I extracted movie_title, imdb_scores, and genres columns from the cleaned data.
- Then I separated genres into different columns using text to columns command and identified all the unique genres. There were a total of 24 genres.
- Then I created a table with having movie name and all 24 genres as my columns. For each movie, if the genre is present among 24 then it will be 1 else it will be blank.
- After that using the COUNTIFS function I created a table having information about how many movies were created for a genre and IMDb scores.
- This table helped me to gain the descriptive statistics i.e., mean, median, mode, max, min, variance, and standard deviation of the IMDb scores.

INSIGHTS: Movie genre analysis

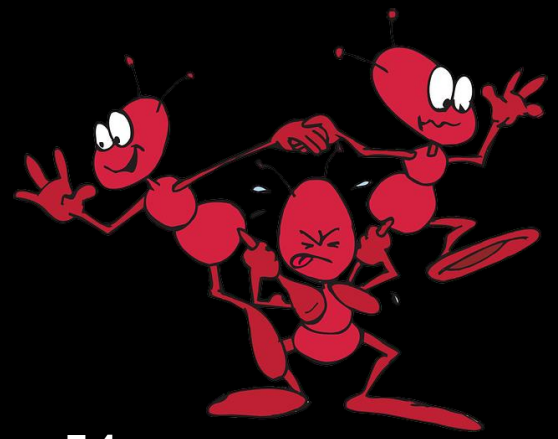
INSIGHTS

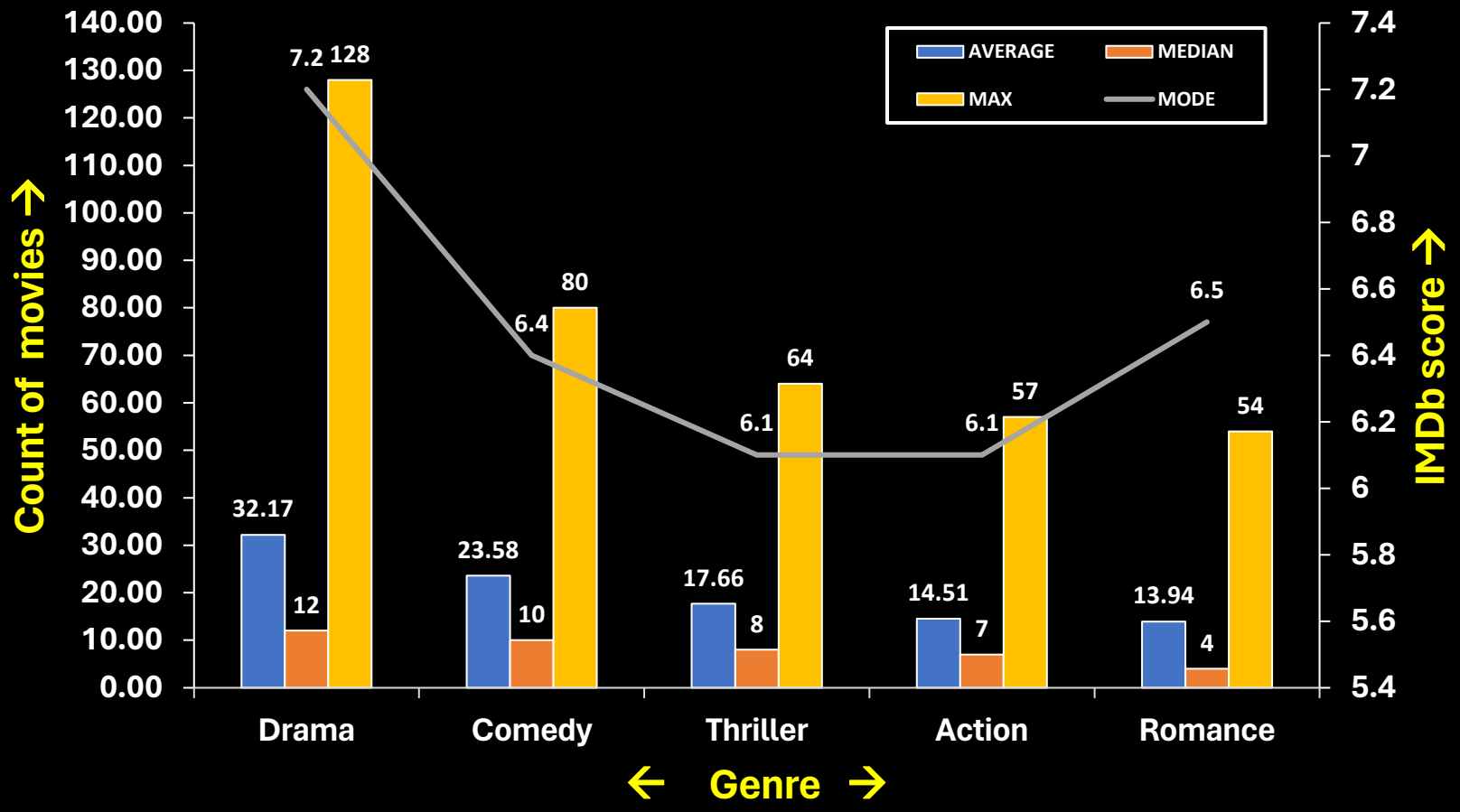
- There are a total of **24 genres**.
- **Drama**, **Thriller**, **action**, **comedy**, and **romance** are the **top 5 genres** that are performing well.
- Genre **drama has the maximum count of movies**(128 movies) for a 7.2 IMDb score.
- **Drama**, **Comedy**, and **Thriller** have an **average count of 17-33 movies** based on their IMDb scores.
- **Drama has comparatively greater variance and standard deviation** compared to any other genre.



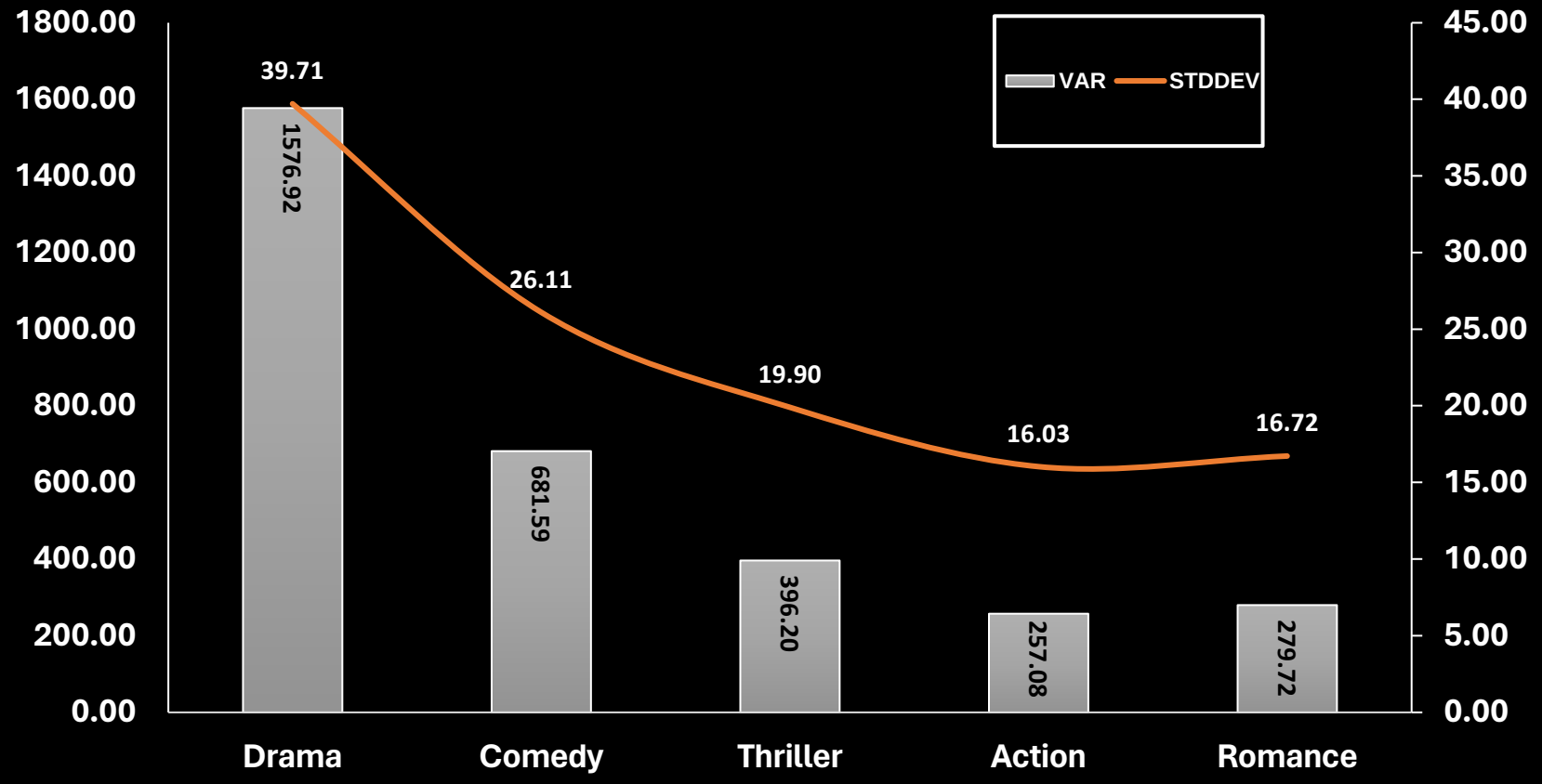
Unique genres
Drama
Comedy
Thriller
Action
Romance
Adventure
Crime
Sci-Fi
Fantasy
Horror
Family
Mystery
Biography
Animation
Music
War
History
Sport
Musical
Documentary
Western
Film-Noir
Short
News

INSIGHTS: Movie genre analysis





INSIGHTS: Movie genre analysis



B. MOVIE DURATION ANALYSIS

INSIGHTS: Movie duration analysis

STEPS PERFORMED

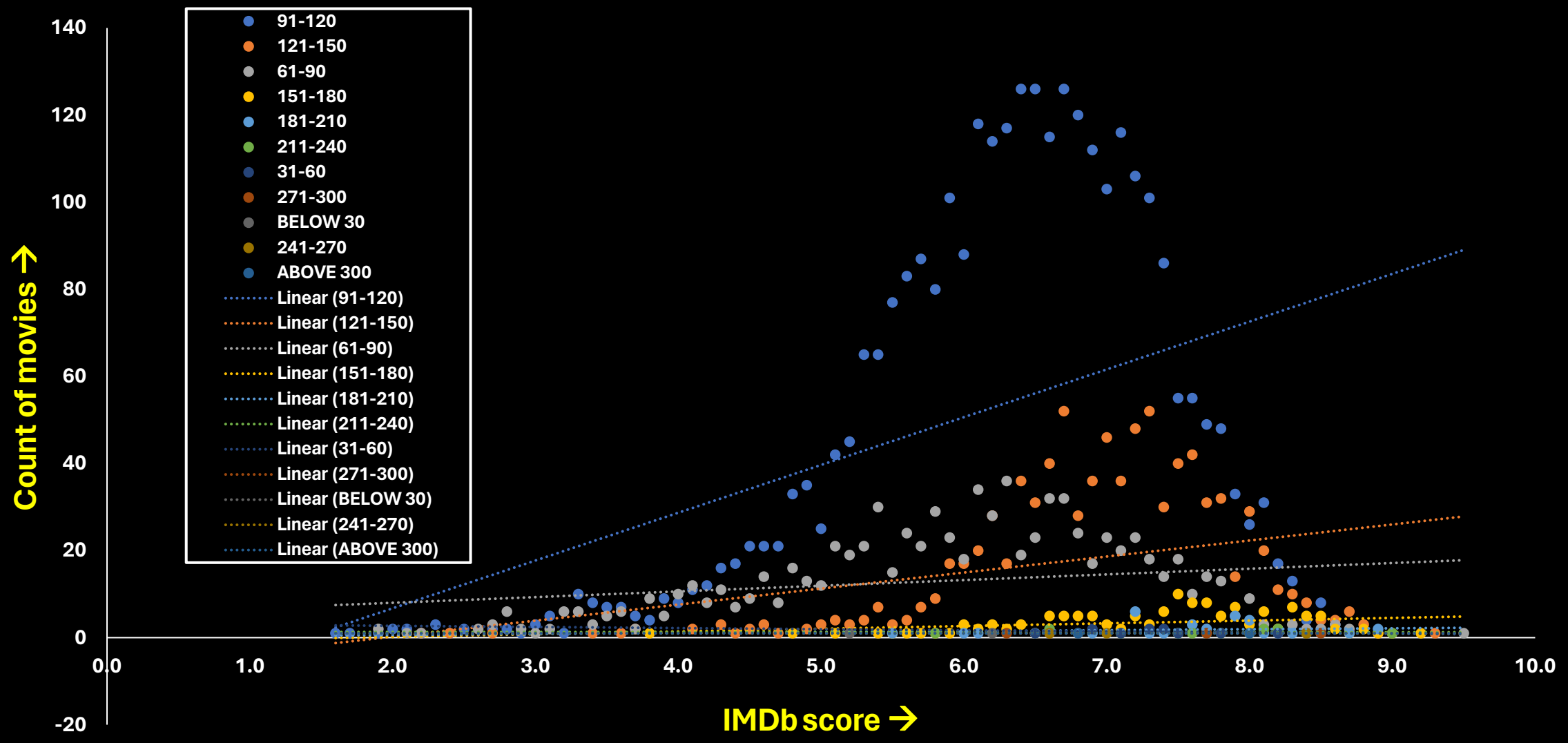
- At first, I extracted movie_title, IMDb_score and duration from the cleaned data.
- Then I created class intervals for the duration column which I used to show the relationship with IMDb scores.
- I created descriptive statistics for the duration column using the Data analysis command in the Analysis group which is in the data tab.
- Then I created a separate table using class interval and IMDb scores which denotes how many movies were created within that duration and what is their IMDb score.

INSIGHTS: Movie duration analysis

INSIGHTS

- Movies that run for 91-120 minutes have the highest number of films and tend to receive an IMDb score between 5.5 and 7.5.
- Most of the movies are created with a duration of 61-120 minutes.
- Shortest duration of a movie is 7 mins, and the longest duration is 330 mins.

INSIGHTS: Movie duration analysis



C. LANGUAGE ANALYSIS

INSIGHTS: Language analysis

STEPS PERFORMED

- At first, I extracted movie_title, Language, and IMDb score columns from the cleaned data.
- Using the data from the table, I created a pivot table to analyze the number of movies produced in each language and their corresponding IMDb scores.
- I created one more table that denotes the number of movies produced in each language.
- Then using the pivot table, I created a table that gives the mean, mode, median, max, min, variance, and standard deviation for each language.

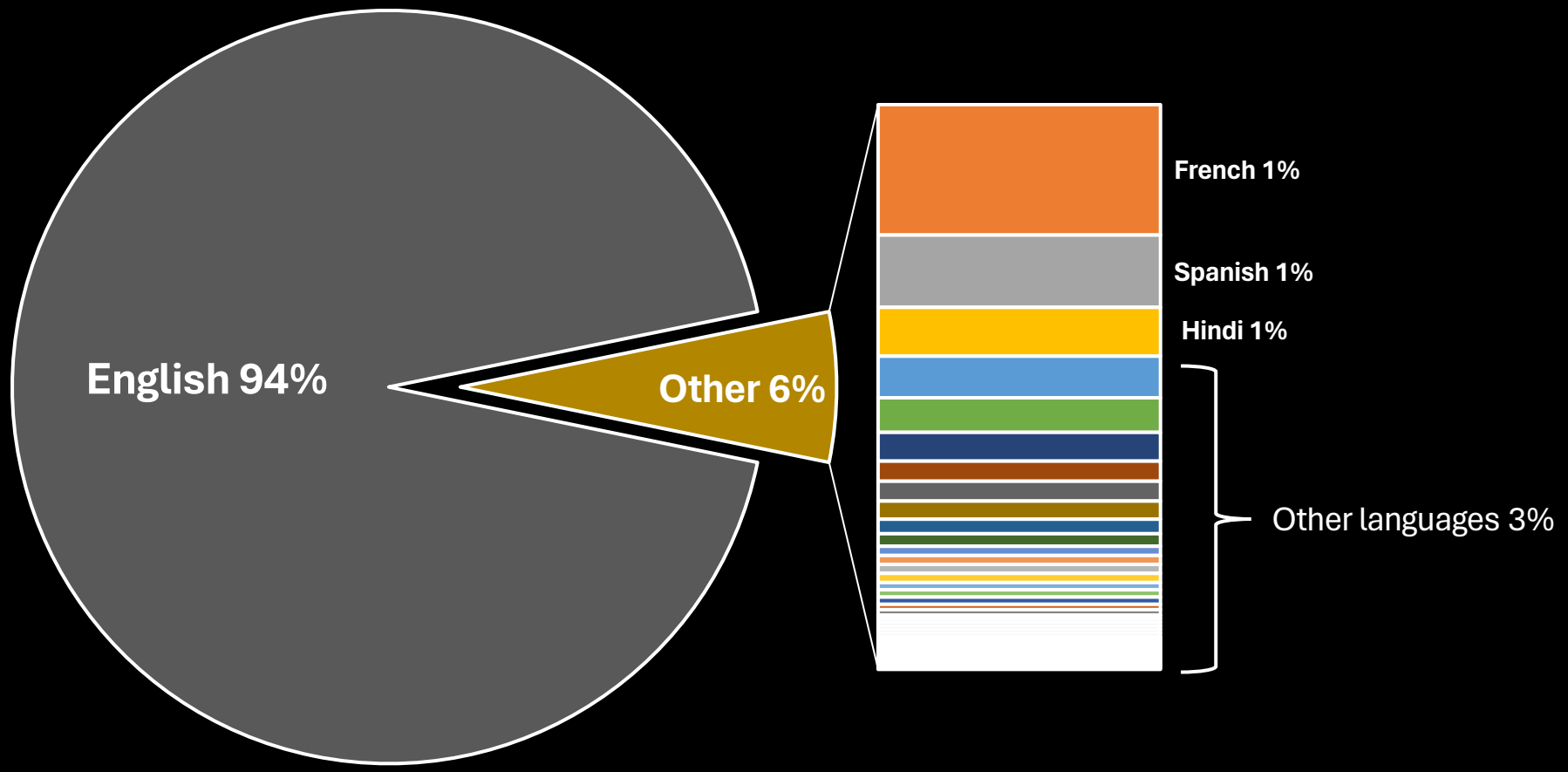
INSIGHTS: Language analysis

INSIGHTS

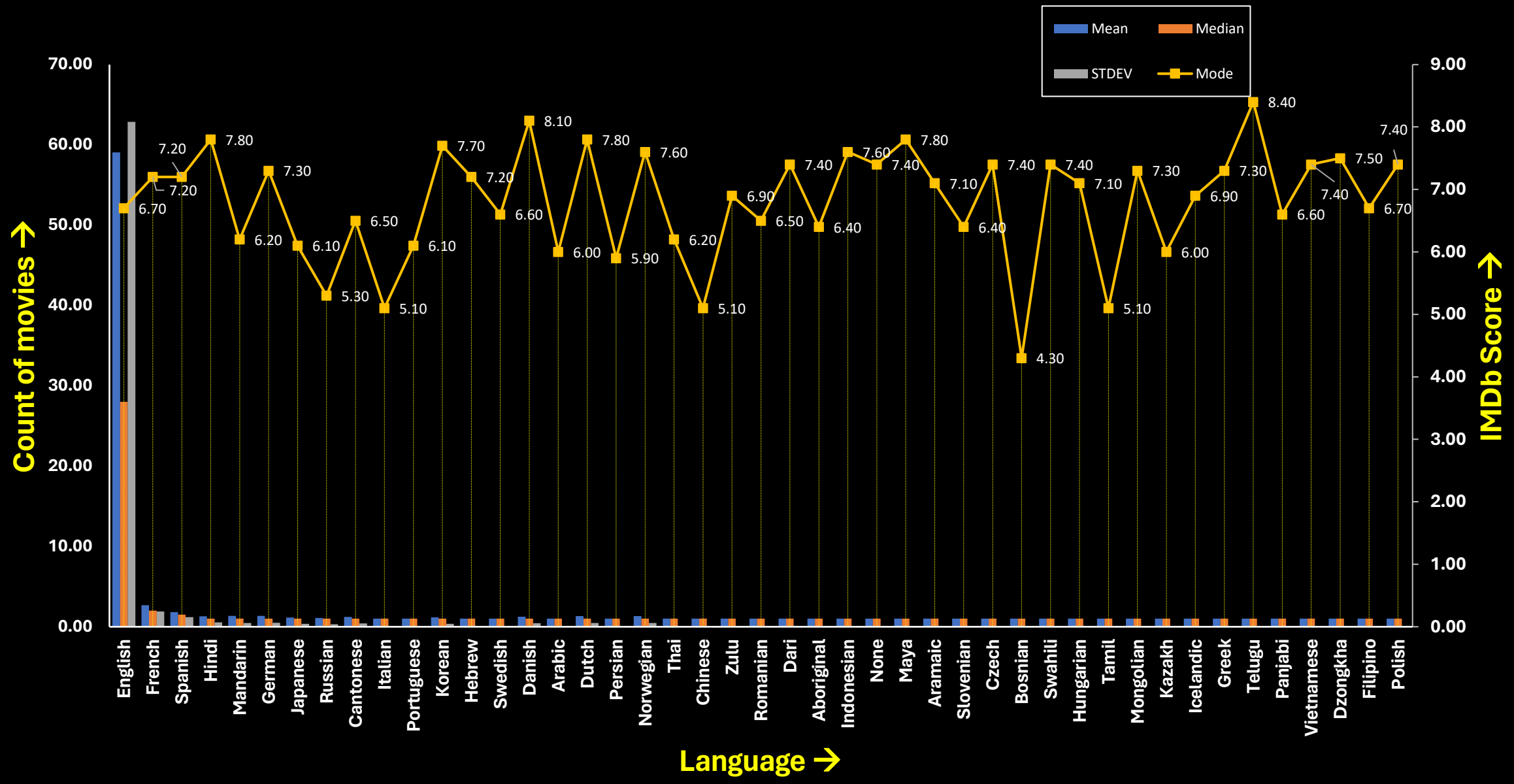
- 94% are English movies.
- Mean, median, and standard deviation of English movies are much greater than other movies.
- 207 English movies have an average IMDb score of 6.7.

INSIGHTS: Language analysis

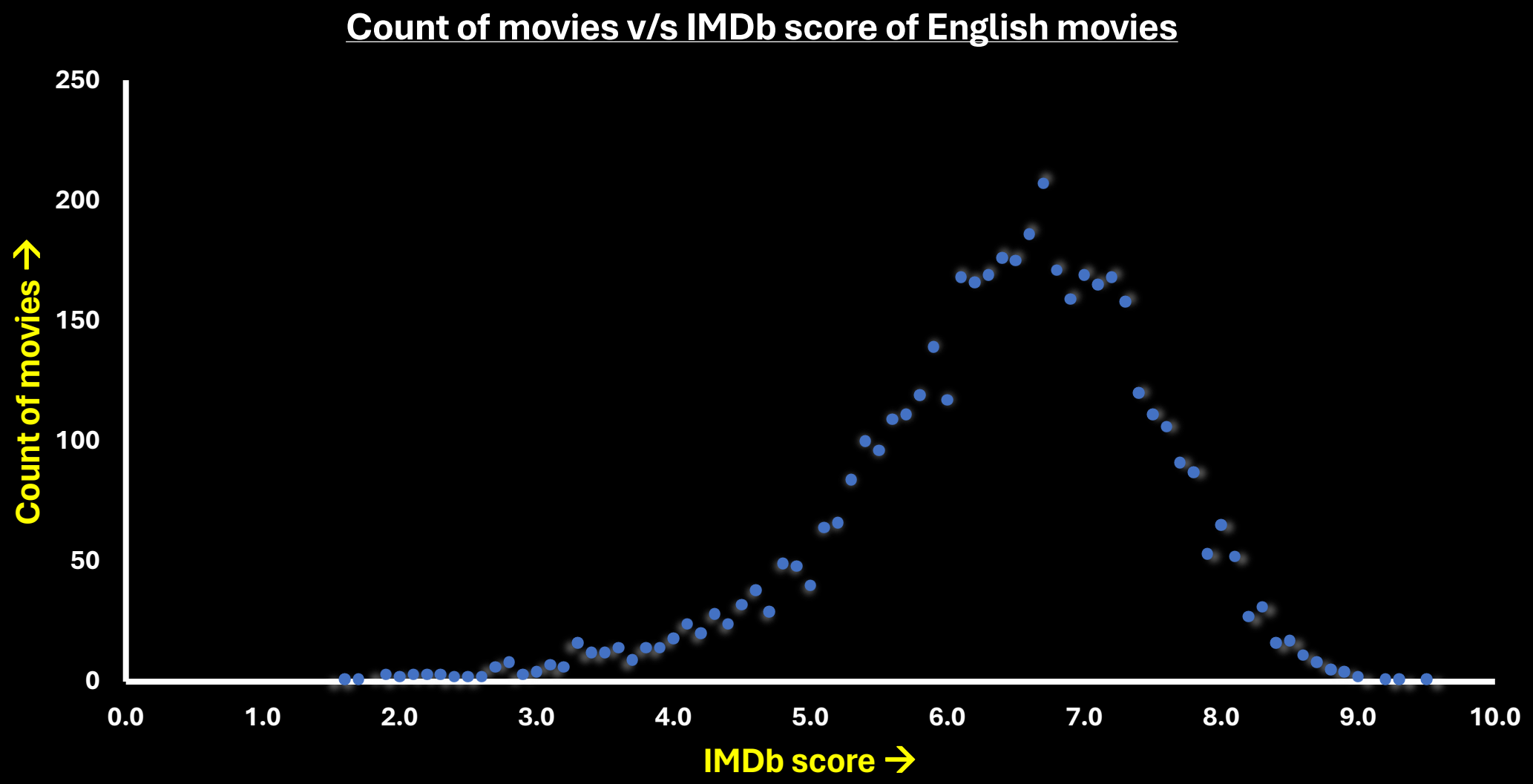
Distribution of movies based on language



INSIGHTS: Language analysis



INSIGHTS: Language analysis



D. DIRECTOR ANALYSIS

INSIGHTS: Director analysis

STEPS PERFORMED

- At first, I extracted movie_title, director_name, and imdb_score columns from the cleaned data.
- Using the data from the table, I created a pivot table that tells us how many movies were directed by each director and what is their average IMDb score.
- From that table I was able to know the percentile of each director based on their IMDb score.
- Using the data, I extracted the directors whose percentile is equal to or greater than 99%.

INSIGHTS: Director analysis

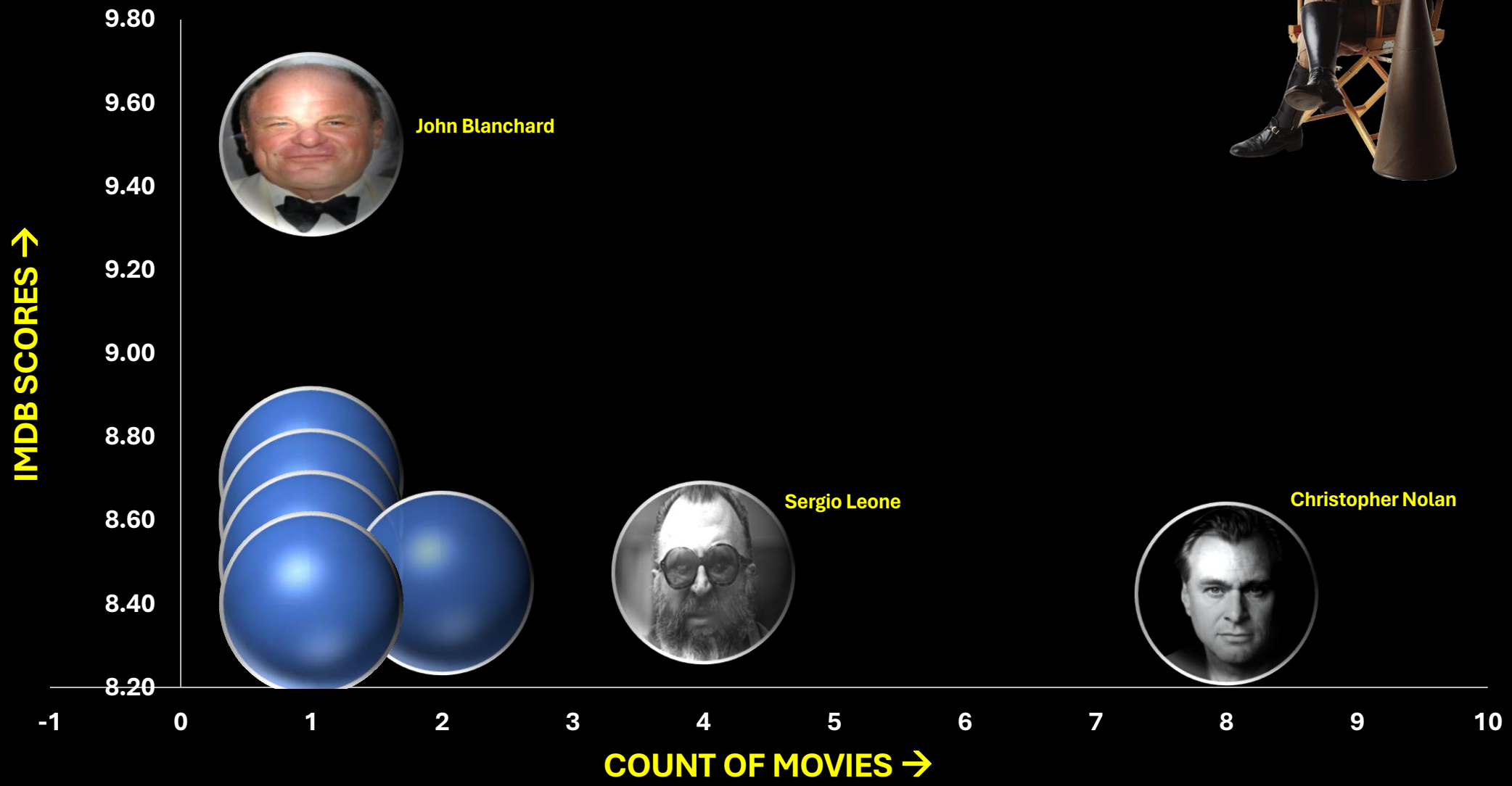
INSIGHTS

- John Blanchard who has directed Towering Inferno which has an IMDb rating of 9.5 has a percentile of 100%.
- Christoper Nolan's has directed a total of 8 movies and its average IMDb score is 8.43 and has a percentile of 99.40%.
- There are a total of 23 directors whose percentile is equal to or greater than the 99% percentile.

INSIGHTS: Director analysis

Row Labels	Count of movie_title	Average of imdb_score	Percentile
John Blanchard	1	9.50	100.00%
Sadyk Sher-Niyaz	1	8.70	99.80%
Mitchell Altieri	1	8.70	99.80%
Cary Bell	1	8.70	99.80%
Mike Mayhall	1	8.60	99.70%
Charles Chaplin	1	8.60	99.70%
Ron Fricke	1	8.50	99.60%
Raja Menon	1	8.50	99.60%
Majid Majidi	1	8.50	99.60%
Damien Chazelle	1	8.50	99.60%
Sergio Leone	4	8.48	99.50%
Tony Kaye	2	8.45	99.50%
Christopher Nolan	8	8.43	99.40%
S.S. Rajamouli	1	8.40	99.00%
Rakeysh Omprakash Mehra	1	8.40	99.00%
Richard Marquand	1	8.40	99.00%
Robert Mulligan	1	8.40	99.00%
Moustapha Akkad	1	8.40	99.00%
Marius A. Markevicius	1	8.40	99.00%
Jay Oliva	1	8.40	99.00%
Catherine Owens	1	8.40	99.00%
Asghar Farhadi	1	8.40	99.00%
Bill Melendez	1	8.40	99.00%

INSIGHTS: Director analysis



E. BUDGET ANALYSIS

INSIGHTS: Budget analysis

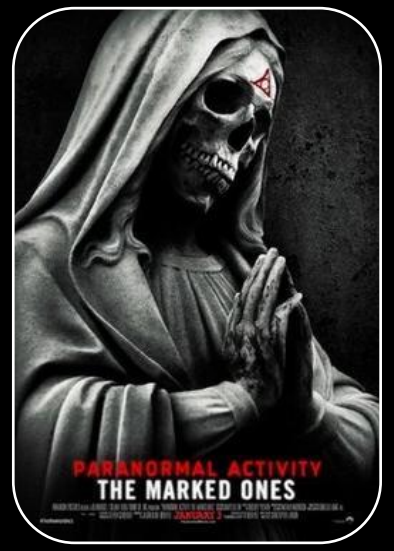
STEPS PERFORMED

- At first, I extracted movie_title, gross, and budget columns from the cleaned data.
- Then I created a column for the profit margin on each movie.
- I also created a profit margin % column which tells the percentage increase of gross with respect to its budget.
- Then by using the CORREL function I found the correlation coefficient between gross and budget.

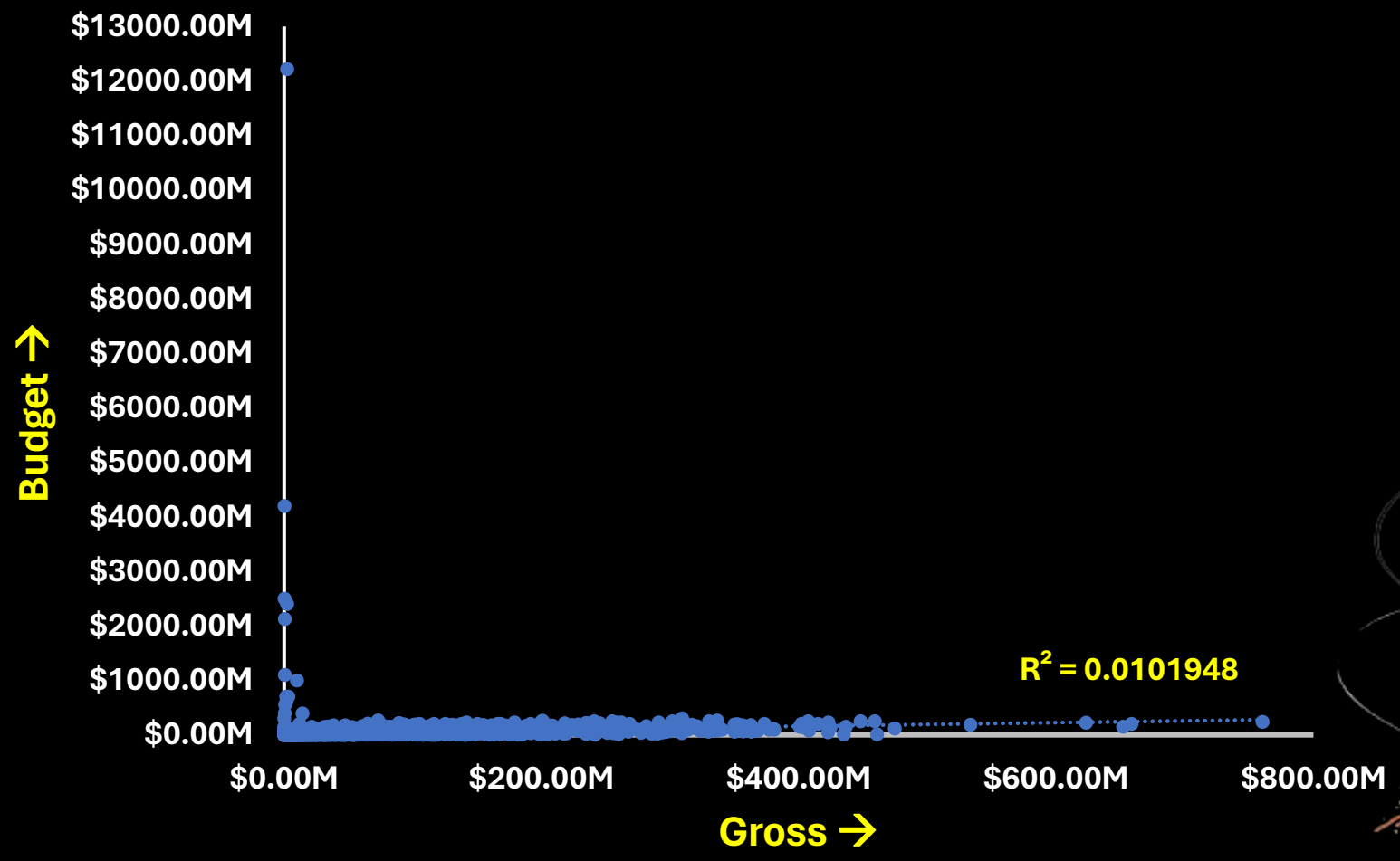
INSIGHTS: Budget analysis

INSIGHTS

- The correlation coefficient of gross and budget is 0.100969202. It means that they have a weak positive correlation which implies that if the budget increases it does not mean that the gross will also increase.
- Avatar movie has the highest profit margin of $523 Million(approx.).
- Paranormal Activity movie has the highest profit margin percentage of 719349% because it has a budget of $15000 but the gross reached $107 Million(approx.).



INSIGHTS: Budget analysis



Excel Link: click here