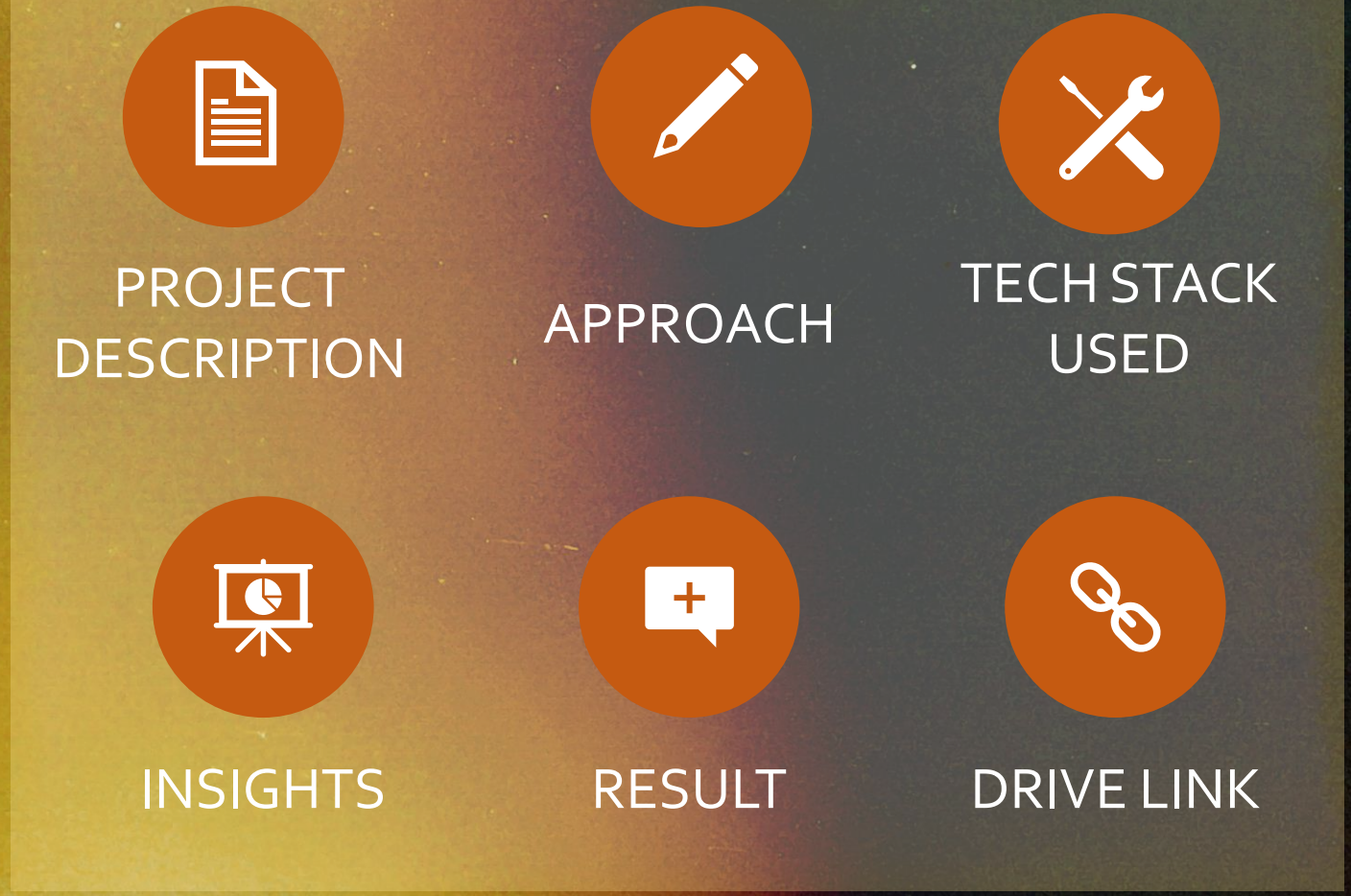
IMDB Movie Analysis

-By Aayushi

- Project Description
- Approach
- Tech Stack Used
- Insights
- Result
- Drive Link



PROJECT DESCRIPTION

The project aims to explore the influential factors determining a movie's success on IMDB, defined by high ratings. Leveraging a dataset on IMDB movies, the investigation will identify correlations between various movie attributes and viewer ratings.

Factors such as genre, budget, gross income, director, and language will be analyzed to discern their impact on movie success. Understanding these dynamics is crucial for stakeholders in the film industry, including producers, directors, and investors.

By gaining insights into the determinants of a movie's success, stakeholders can make informed decisions regarding project investments, casting choices, and strategic planning. Ultimately, this project endeavors to provide actionable insights that empower industry professionals to enhance the likelihood of producing successful movies in the future.

APPROACH

The project commenced by downloading the provided IMDB Movies dataset, followed by an in-depth exploratory data analysis (EDA) using Microsoft Excel. During the EDA phase, each column was scrutinized to identify data types, validate data accuracy, and address missing values across different attributes. This rigorous examination was fundamental for understanding the dataset's structure and ensuring data integrity.

Subsequently, Excel's functionality was leveraged to conduct a comprehensive analysis. Formulae, pivot tables, and various chart types were employed to delve into the dataset and address specific inquiries. Pivot tables facilitated dynamic summarization and aggregation of data, enabling a thorough exploration of patterns in the dataset. Additionally, the use of visualizations such as bar graphs and scatter plots provided intuitive representations of key insights derived from the data.

This systematic approach ensured an insightful analysis, offering actionable recommendations for stakeholders in the film industry to make informed decisions.

TECH STACK

In conducting data analysis, **Microsoft Excel 365** served as the central tool due to its versatility and effectiveness. This version was instrumental in managing datasets, allowing for seamless data cleaning, manipulation, and visualization. Features such as pivot tables, functions, and charts were extensively employed to extract insights, facilitating informed decision-making and streamlining analytical processes.

For the presentation of analytical findings, **Microsoft PowerPoint 365** was utilized. Its user-friendly interface and robust features were leveraged to craft an engaging analysis report. PowerPoint's capabilities in designing impactful visualizations and presenting data-driven insights proved invaluable in effectively communicating the results of the analysis to stakeholders.

EXPLORATORY DATA ANALYSIS (EDA)

Dimensions	Values
Original Column Count	28
Original Row Count	5043
Missing data represented as	blank
Column Count After EDA	10
Row Count After EDA	3786
Deleted Column Count	18
Duplicate Rows Removed	101

color
director_name
num_critic_for_reviews
duration
director_facebook_likes
actor_3_facebook_likes
actor_2_name
actor_1_facebook_likes
gross
genres
actor_1_name
movie_title
num_voted_users
cast_total_facebook_likes
actor_3_name
facenumber_in_poster
plot_keywords
movie_imdb_link
num_user_for_reviews
language
country
content_rating
budget
title_year
actor_2_facebook_likes
imdb_score
aspect_ratio
movie_facebook_likes

Initial Dataset Analysis

Firstly, we analyze the provided dataset. The dimensions of the dataset provided are listed in the screenshot alongside.

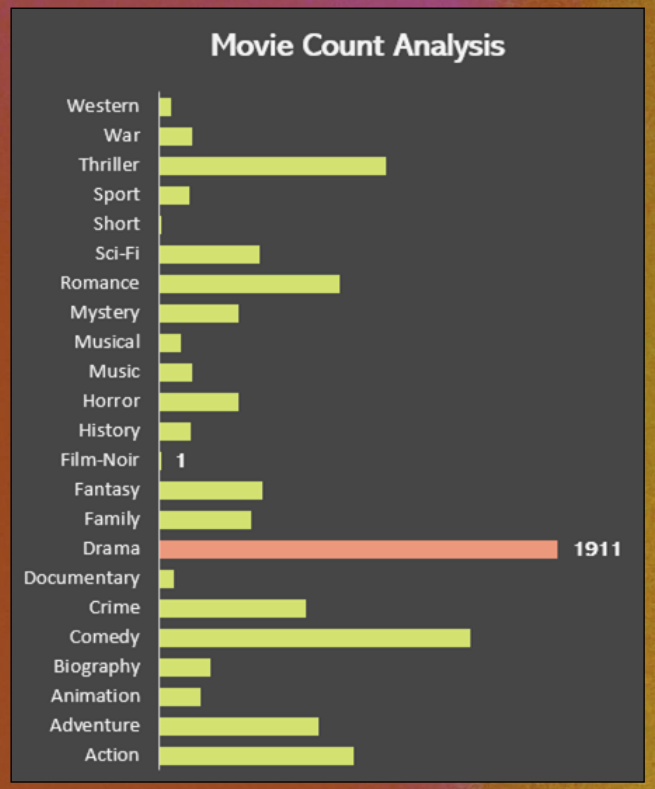
We are provided with 28 columns as listed:

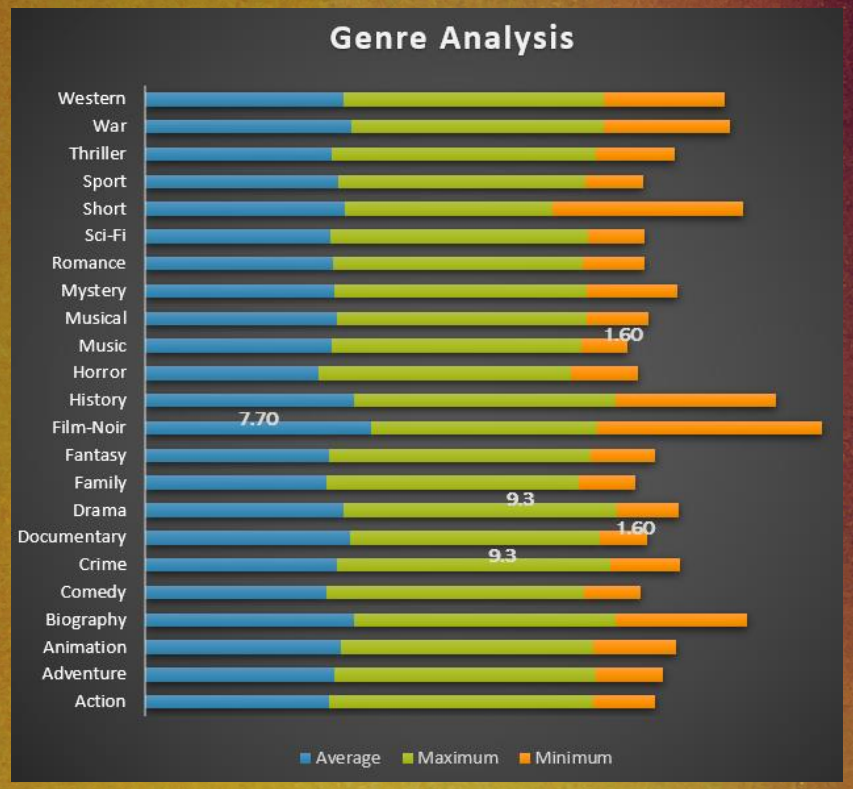
Column Value Analysis

COLUMN VALUE ANALYSIS				
Original Column	**Renamed Column**	**Data Type**	**No (Missing Values)**	**Range**
director_name	Director	Text	102	-
movie_title	Movie Title	Text	0	-
imdb_score	IMDB Score	Integer	0	3.9 - 7.8
title_year	Release Year	Integer	0	1953 - 2014
duration	Duration	Integer	15	65 - 111
genres	Genres	Text	0	-
language	Language	Text	12	-
country	Country	Text	5	-
budget	Budget	Integer	492	1100 - 1000000
gross	Gross Income	Integer	871	4584 - 11529368

- After analyzing the given tasks, we remove the unnecessary columns from the dataset.
- From the given **28 columns**, we narrow down our dataset to **10 columns** that we will work upon.
- The dimensional analysis of the selected columns is as shown alongside.
- After removing the unnecessary columns, we still have **missing values** which cannot be replaced with any descriptive measure. Hence, we **deleted** those rows.
- Subsequently, we **reordered and renamed** the selected columns as per need.
- Lastly, we scanned for **duplicate rows** and **deleted** those as well.

INSIGHTS:
Movie Genre Analysis





Movie Genre Analysis

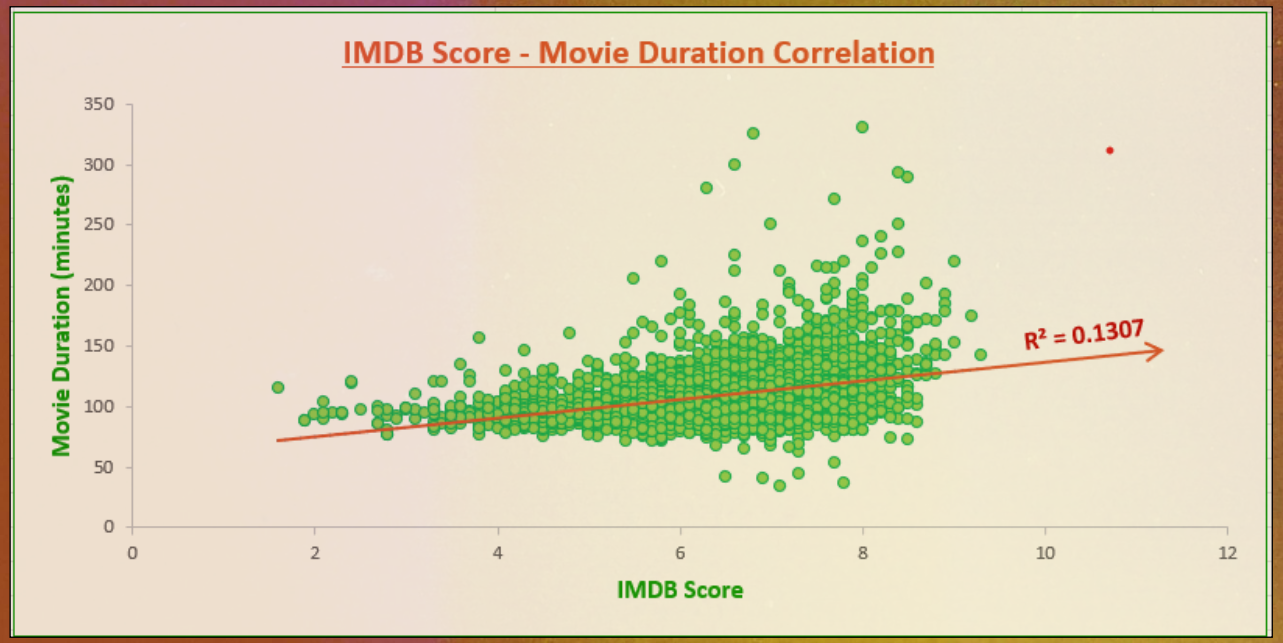
The analysis of movie genres and their impact on IMDB scores reveals several insightful findings.

- Firstly, **Drama** emerges as the **most common genre** in the dataset, with a count of **1911** movies.
- Also, **Drama** & **Crime** movies stand out as the **highest-rated genre**, with an impressive IMDB score of **9.3**.
- Conversely, **Documentary** & **Music** films exhibit the **lowest average rating**, with an IMDB score of **1.60**.
- Notably, **Film-Noir** emerges as the genre with the **highest average rating**, with an IMDB score of **7.70**, indicating a consistent level of audience appreciation for this category.

These insights underscore the diversity in audience preferences and highlight the significance of genre selection in influencing the perceived quality and success of movies.

INSIGHTS:
Movie Duration Analysis

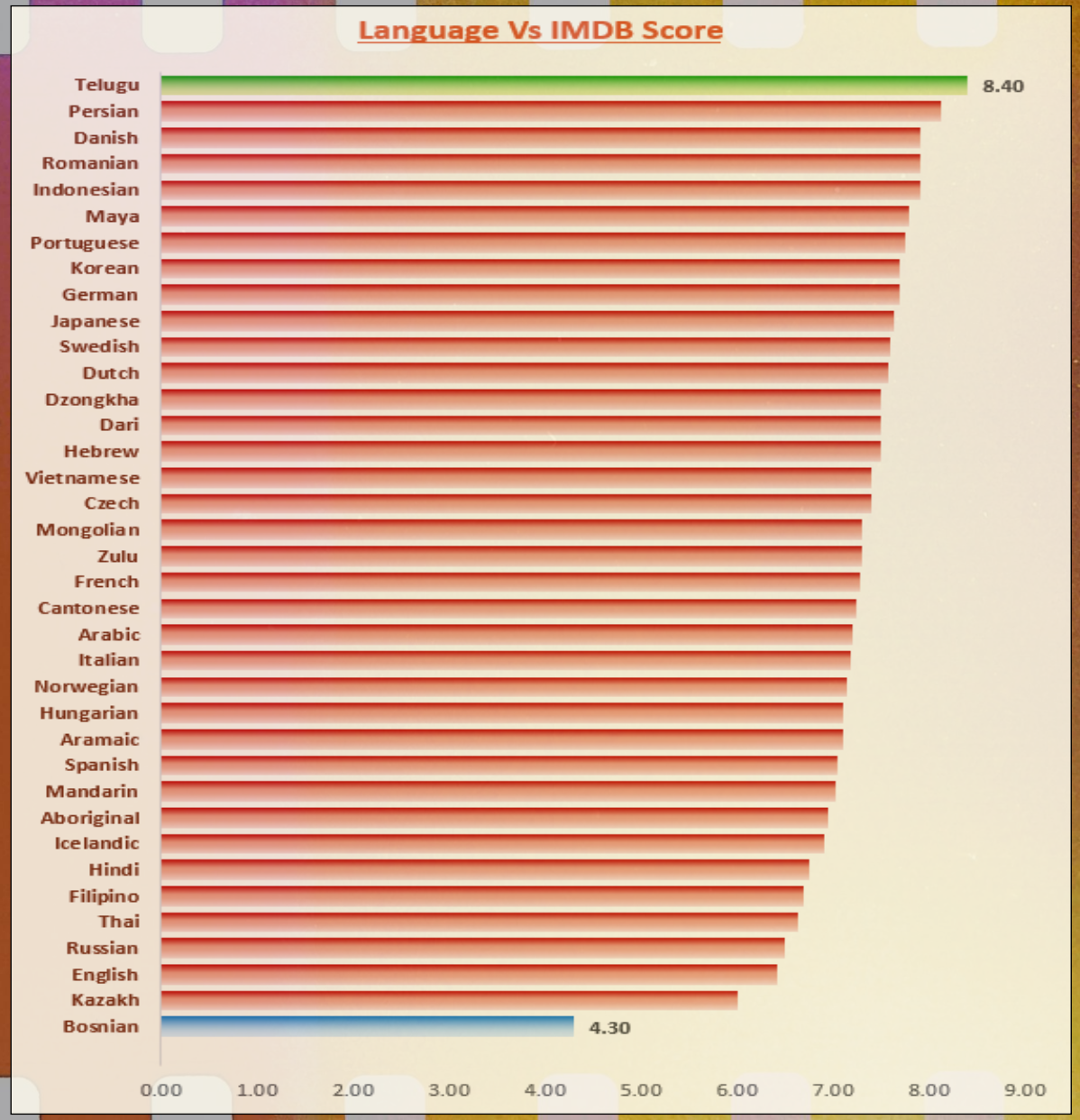
Movie Duration Analysis



The analysis of movie durations unveils several key insights regarding their distribution and impact on IMDB scores.

- On **average**, movies in the dataset have a duration of approximately **109.81 minutes**, with significant variations observed.
- The **longest movie, "Blood In, Blood Out**," spans a remarkable **330 minutes.**
- The **shortest movie**, "**Marilyn Hotchkiss' Ballroom Dancing and Charm School**," clocks in at just **34 minutes**.
- The **median duration** stands at **105 minutes**, indicating a **balanced distribution**.
- However, the **standard deviation** of **22.76 minutes** suggests notable variability across movie lengths.
- The **correlation** between IMDB score and movie duration is moderate, with a coefficient of **0.3615**, suggesting a **modest positive relationship**.
- Nonetheless, the R-squared value of **0.1307** indicates that movie duration explains only a limited proportion of the variability in IMDB scores, emphasizing the multifaceted nature of factors influencing audience ratings.

INSIGHTS:
Language Analysis



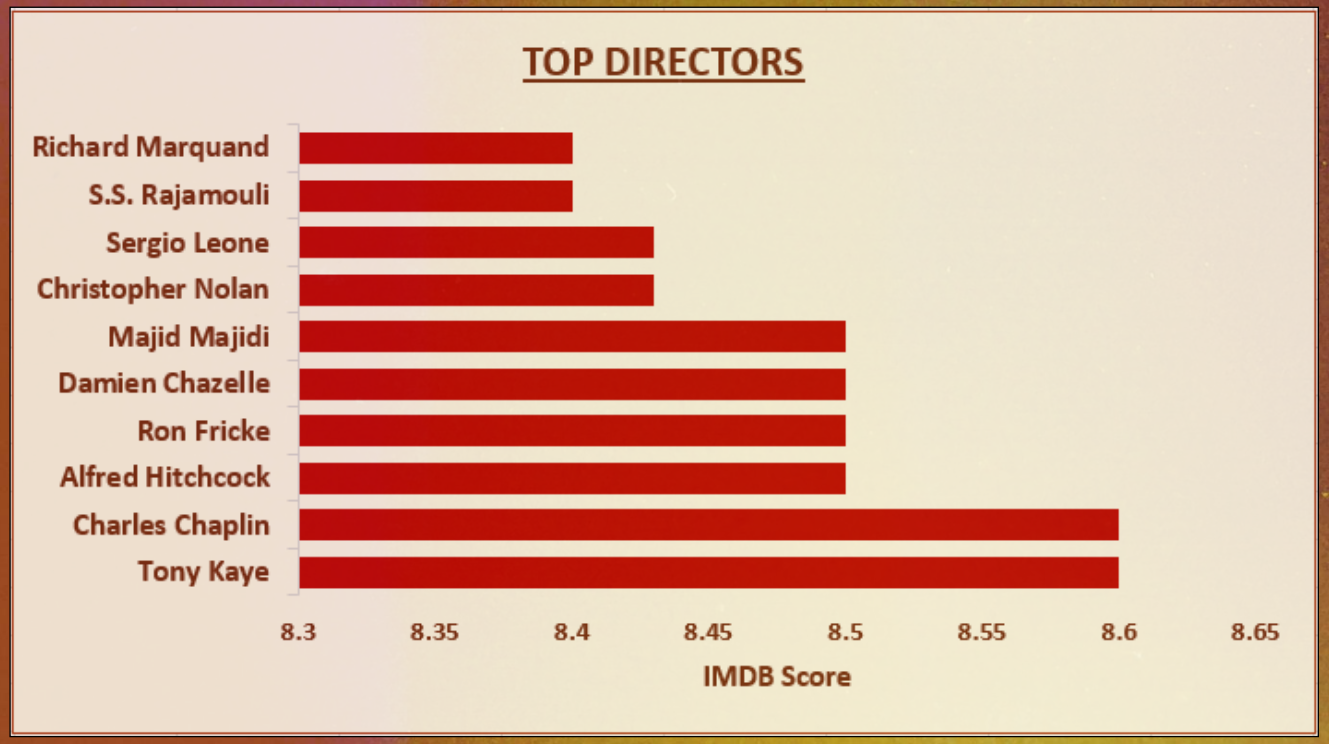
Language Analysis

The analysis of movie languages reveals intriguing patterns in distribution and their impact on IMDB scores.

- **English** emerges as the most common language, followed by **French, Spanish, Mandarin, and German**.
- Despite its prevalence, English doesn't always guarantee the **highest ratings. Telugu** movies, surprisingly, hold the top spot with an impressive IMDB score of **8.40**.
- Conversely, **Bosnian** language movies exhibit **the lowest ratings** with an IMDB score of **4.30**.
- These findings highlight the **diversity of languages** in cinema and their varying degrees of audience reception.
- The disparity in ratings underscores the influence of cultural context and regional preferences on viewer perception, shedding light on the nuanced dynamics between language and film appreciation.

INSIGHTS:
Director Analysis

Director Analysis

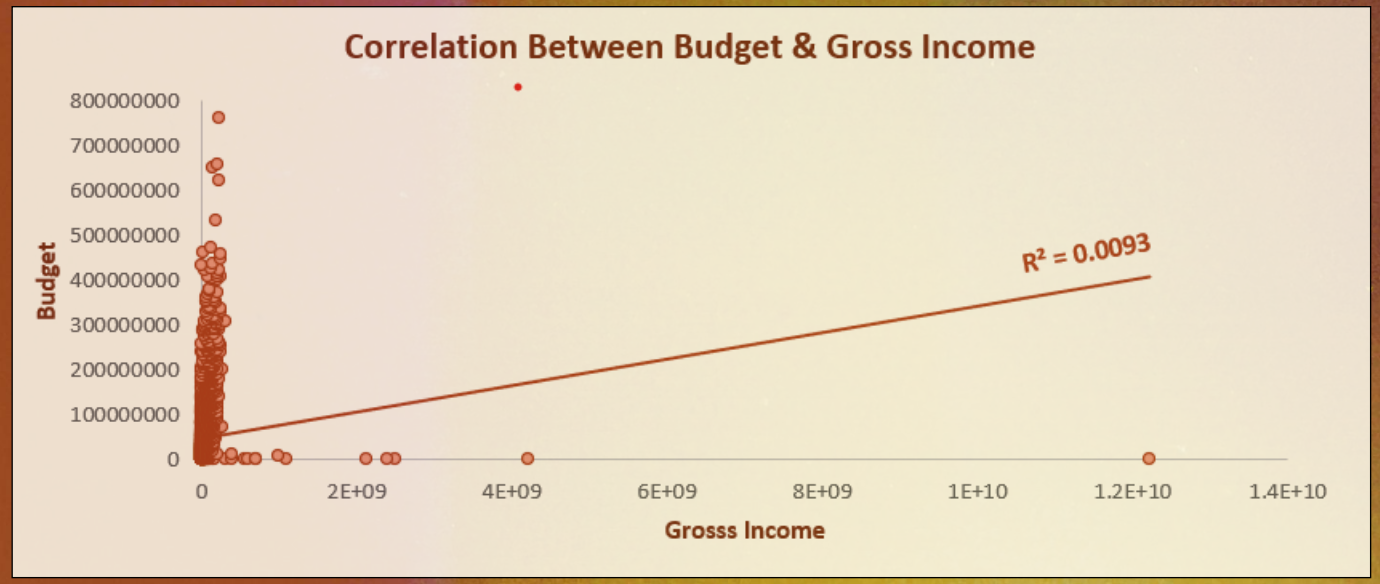


The analysis of directors' impact on movie ratings unveils significant insights into their influence on film success.

- Directors such as **Tony Kaye, Charles Chaplin, and Alfred Hitchcock** stand out with consistently **high average IMDB scores**, each achieving an impressive rating **above 8.5**.
- By employing percentile calculations, their contributions to movie success are underscored, revealing their ability to consistently deliver captivating storytelling and cinematic excellence
- The inclusion of directors like **Christopher Nolan** and **Sergio Leone** further emphasizes the enduring impact of visionary filmmaking on audience engagement and acclaim.
- These findings highlight the pivotal role of directors in shaping audience perceptions and driving positive ratings, offering valuable insights into the **correlation between directorial prowess and movie success**.

INSIGHTS:
Budget Analysis

Correlation between Budget & Gross Income	0.0966
R-Squared value	0.0093
Movie with Highest Profit Margin	Avatar
Highest Profit Margin	523505847
Lowest Gross Income Movie	Skin Trade
Highest Gross Income Movie	Avatar
Lowest Budget Movie	Tarnation
Highest Budget Movie	The Host



Budget Analysis

The exploration of movie budgets and their financial outcomes reveals intriguing insights into the relationship between investment and success.

- Despite a **low correlation coefficient** of **0.0966** and a **modest R-squared value** of **0.0093**, indicating a **weak relationship between budget and gross earnings**, certain movies have achieved remarkable profit margins.
- "**Avatar**" emerges as the movie with the **highest profit margin**, demonstrating exceptional financial success with a **profit** of **$523,505,847**.
- Conversely, "**Skin Trade**" records **the lowest gross income**, while "**Tarnation**" holds the title of the **lowest budget movie**.
- "**The Host**" stands out as the movie with the **highest budget**, suggesting that financial success is not solely determined by the amount invested but also by factors such as marketing strategies and audience reception.

RESULT

The project report provides comprehensive insights into various factors influencing the success of movies, drawn from an extensive analysis of IMDB movie data.

- Firstly, the analysis of movie genres highlighted the prevalence of **Drama** while showcasing Crime and Drama as the highest-rated genres, and Documentary and Music as the lowest-rated.
- Furthermore, the examination of movie durations revealed significant variability, with **Telugu** movies leading in ratings despite Bosnian movies having the lowest ratings.
- The exploration of directorial influence identified top directors such as Tony Kaye and Charles Chaplin, emphasizing their pivotal role in achieving high IMDB scores.
- Additionally, the budget analysis unveiled intriguing findings, indicating a **weak correlation between budget and gross earnings**, yet showcasing exceptional profit margins for movies like "Avatar."

These insights collectively underscore the multifaceted nature of movie success, influenced by factors ranging from genre selection and directorial prowess to budget allocation and audience reception.

DRIVE LINK

Please click below to find the link to the excel sheet:
IMDB Movies