IMDB MOVIE GENRE ANALYSIS

- RUKMINI POOJARI

Objective

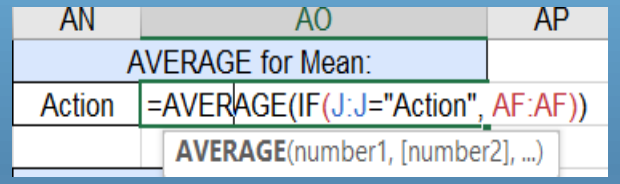
- The objective of this presentation is to analyze the key factors that influence the success of movies on IMDB, with success defined by high IMDB ratings. The analysis aims to provide insights into how various attributes—such as movie genres, duration, language, directors, and budgets—affect movie ratings and financial performance. By exploring the relationships between these factors, we hope to offer valuable insights for movie producers, directors, and investors, helping them make informed decisions to maximize both critical acclaim and financial success in future film projects.

Movie Genre Analysis: Distribution and Impact on IMDB Scores

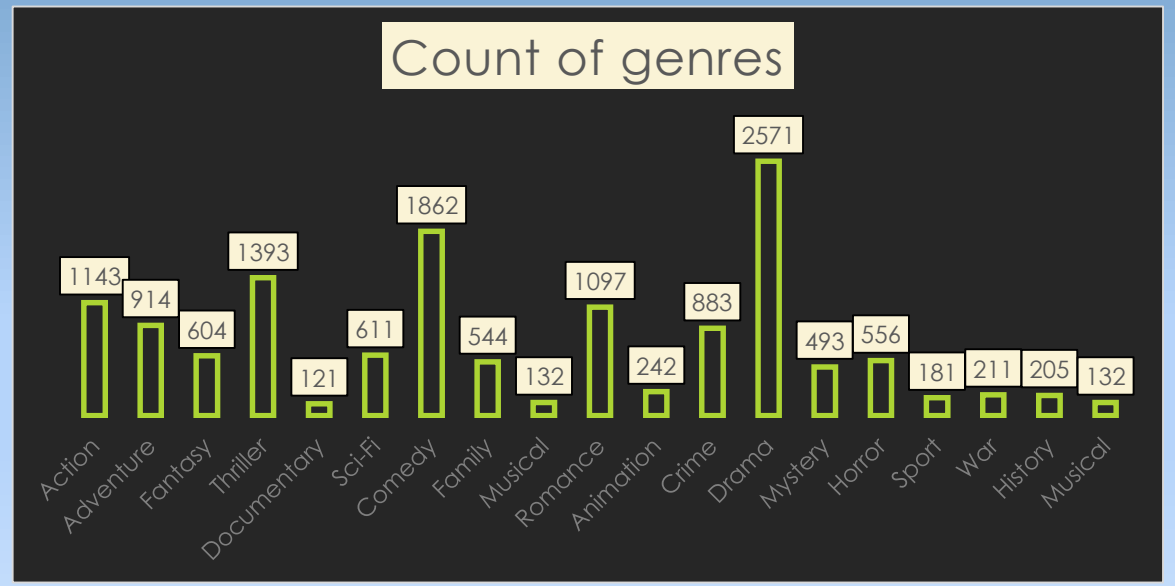
Objective : Analyze the distribution of movie genres in the dataset and their impact on IMDB scores.
•Provide insights into which genres are the most common and how they perform in terms of ratings.
•Help stakeholders make genre-based decisions for future film projects.

Steps Taken:
1. Genre Splitting: Used Excel's Text to Columns and Power Query to separate multiple genres.
2. Genre Count: Applied COUNTIF function to count occurrences of each genre.
3. Descriptive Statistics: Calculated mean, median, mode, range, variance, and standard deviation for IMDB scores using Excel functions like AVERAGE, MEDIAN, MODE.SNGL, VAR, STDEV.S.
Visual: Bar chart or showing the frequency of movies by genre.

Most Common Genres: Action and Drama are the most frequent genres, but they differ in terms of average rating.



AVERAGE for Mean:	
Action	6.441056423



Movie Duration Analysis: Distribution and Impact on IMDB Score

Objective :Analyze the distribution of movie durations in the dataset.
Explore how movie duration affects the IMDB score.
Provide insights into the ideal duration range for better ratings.

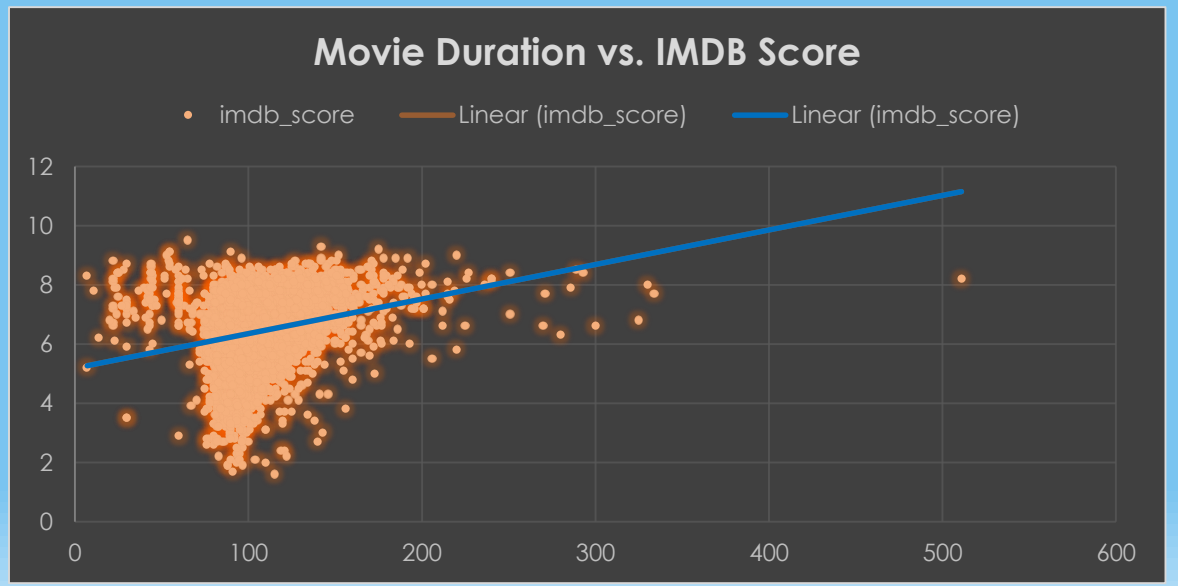
AVERAGE function to calculate the mean movie duration:			
Mean movie durations	107.201074		

MEDIAN function to calculate the median movie duration:			
Meadin movie durations	103		

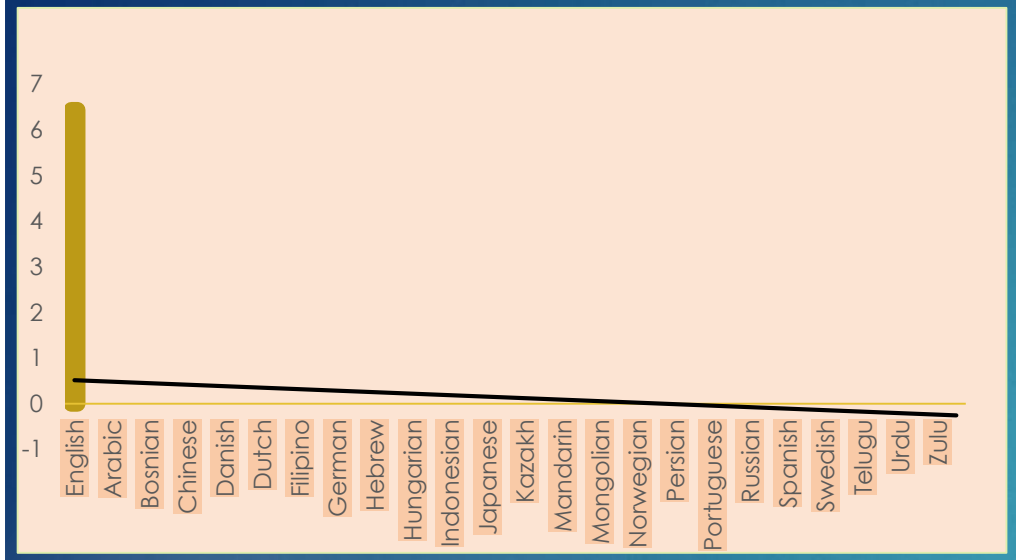
range, variance, or mode of movie duration		
Range	146	
Variance	634.7847482	

Steps Taken:
Descriptive Statistics: Used Excel's functions like AVERAGE, MEDIAN, and STDEV.S to calculate descriptive statistics for movie durations.
Relationship with IMDB Score: Created a scatter plot to visualize the relationship between movie duration and IMDB score.
Trendline :Added a trendline to assess the strength and direction of the relationship.

Key Findings
The distribution of movie durations shows that most movies fall in the 90-120 minute range.
There is a slight positive correlation between movie duration and IMDB score, but it is not strong enough to dictate that longer movies are always better.
The optimal movie length appears to be around 100-130 minutes for higher ratings.



IMDB Score for each language



=COUNTIF(T:T,AD2)

AC	AD	AE
	Count of language	
	English	4662
	Aboriginal	2
	Arabic	5
	Aramaic	1
	Bosnian	1
	Cantonese	11
	Chinese	3
	Czech	1
	Danish	5
	Dari	2
	Dutch	4
	Dzongkha	1

Objective : Analyze the distribution of movies based on language in the dataset.
•Investigate the relationship between movie language and IMDB ratings.
•Identify if certain languages lead to higher or lower movie ratings.

Steps Taken: Language Distribution: Used Excel's COUNTIF function to count the number of movies for each language.
Descriptive Statistics: Calculated mean, median, and standard deviation of IMDB scores for each language using AVERAGE, MEDIAN, and STDEV.
Comparison: Compared these statistics across different languages to assess their impact on movie ratings.

Visual: Bar chart of movie count per language.

Key Insight: English-language movies constitute the majority of the dataset

Director Analysis

Objective:

Analyze how movie directors influence IMDB ratings.
Identify the top directors based on their average IMDB score.
Use percentile calculations to rank directors and assess their contributions to movie success..

Steps Taken:

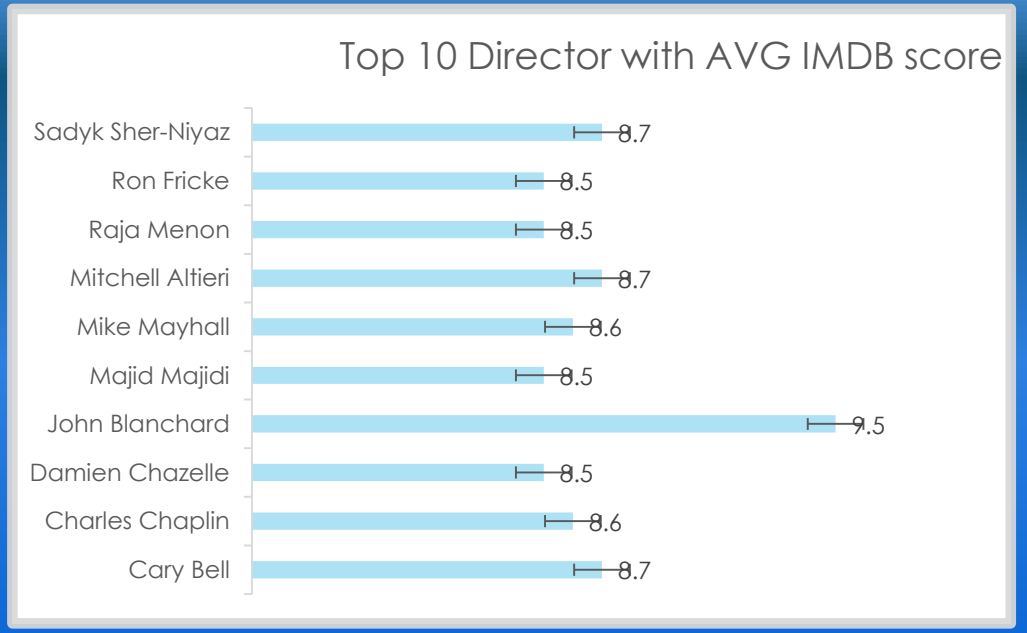
- ❖ Average IMDB Scores: Calculated the average IMDB score for each director using Excel's AVERAGE function.
- ❖ Percentile Calculation: Used Excel's PERCENTILE function to rank directors based on their average scores.
- ❖ Comparison: Compared top directors to the overall distribution of IMDB scores for insights into their impact on movie success.
- ❖ Visual: Bar chart showing the top 10 directors with the highest average IMDB scores.

Key Findings:

- Directors in the top percentiles tend to have consistent success across multiple genres.
- Directors with higher IMDB scores often work with high-quality production teams and actors.
- The top 5% of directors are more likely to produce critically acclaimed films that resonate with both audiences and critics

Large	9.5
Rank	1
percentile	9.5

Top 10 Director na	Average of imdb_score
Cary Bell	8.7
Charles Chaplin	8.6
Damien Chazelle	8.5
John Blanchard	9.5
Majid Majidi	8.5
Mike Mayhall	8.6
Mitchell Altieri	8.7
Raja Menon	8.5
Ron Fricke	8.5
Sadyk Sher-Niyaz	8.7
Grand Total	**8.68**



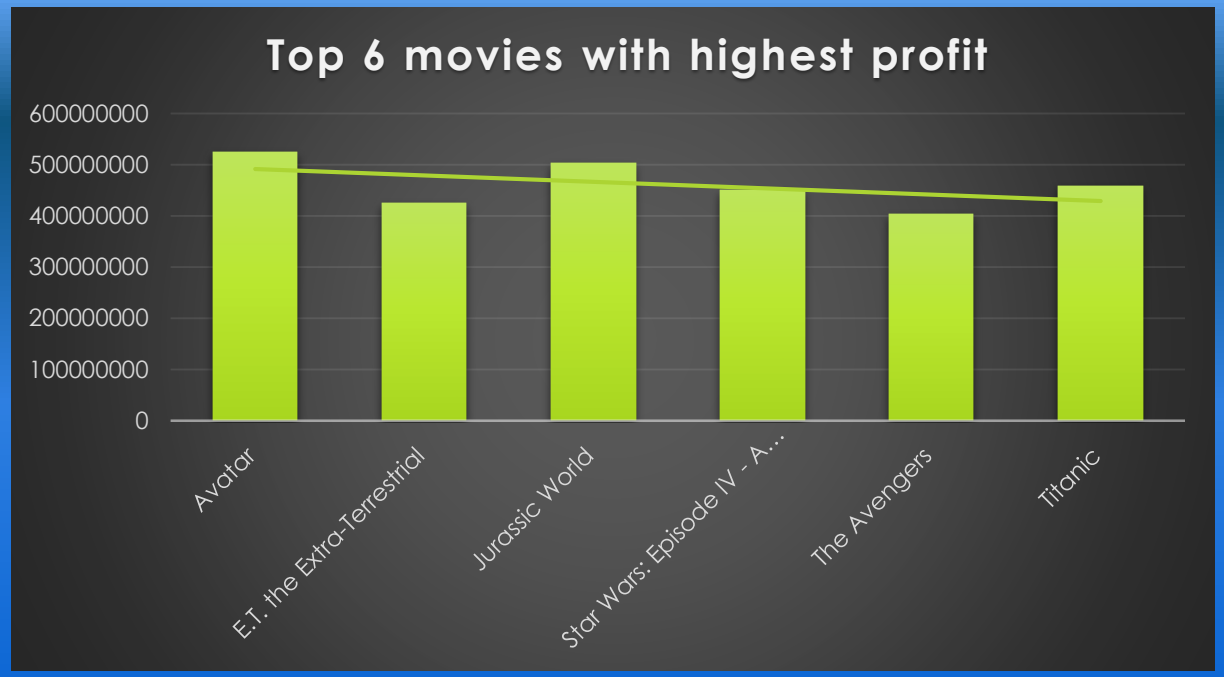
Budget Analysis

Objective: Explore the relationship between movie budgets and financial success.

Step Taken:

- Calculate the correlation coefficient to understand the relationship between movie budgets and gross earnings.
- Determine the profit margin for each movie.
- Identify the movies with the highest profit margin, helping you to analyze which films were the most financially successful relative to their budgets.
- visualize the distribution of profit margins using charts.

Movies with the Highest Profit Margin	523505847
Movies Name with the Highest Profit Margin:	Avatar
Movie Name	Average of Profit
Avatar	523505847
E.T. the Extra-Terrestrial	424449459
Jurassic World	502177271
Star Wars: Episode IV - A New Hope	449935665
The Avengers	403279547
Titanic	458672302
Grand Total	**460336681.8**



Key Insights:

- There are other factors beyond the budget that contribute to a movie's profitability.
- State that careful budget allocation and smart production strategies can help achieve higher financial success.

Project Description

This project focuses on analyzing the IMDB Movies dataset to understand the factors that contribute to a movie's success, measured by its IMDB rating. The goal is to provide actionable insights for movie producers, directors, and investors to make informed decisions in future movie projects. By investigating variables such as genre, duration, language, director, and budget, the analysis aims to uncover relationships and trends that influence movie ratings and financial success.

Approach

The approach is structured into several key tasks, each addressing different aspects of the dataset:

1.Data Cleaning: The dataset was cleaned to handle missing values, remove duplicates, and convert data types where necessary. For instance, the 'genres' column, which often contained multiple genres, was split using Excel Power Query to allow for more granular analysis.

2.Data Analysis:

Genre Analysis: Distribution of movie genres was analyzed, and descriptive statistics (mean, median, mode, range, variance, standard deviation) for IMDB scores were calculated for each genre. Excel functions like COUNTIF, AVERAGE, STDEV, etc., were used.

Duration Analysis: Descriptive statistics of movie durations were calculated, and a scatter plot with a trendline was created to observe the relationship between movie duration and IMDB scores.

Language Analysis: The distribution of movies by language was studied, and the impact of language on IMDB ratings was explored using functions like COUNTIF and AVERAGEIF.

Director Analysis: The top directors were identified based on their average IMDB scores. Percentile calculations using the PERCENTILE.INC function helped analyze which directors consistently contributed to successful movies.

Budget Analysis: The relationship between movie budgets and gross earnings was analyzed using the CORREL function to calculate the correlation coefficient. The profit margin (gross earnings - budget) was computed, and the movies with the highest profit margins were identified using the MAX function.

Tech-Stack Used

Microsoft Excel 2016:

Data Cleaning: Handled missing values, split genres using Power Query, and manipulated data types.
Data Analysis: Performed calculations for descriptive statistics (e.g., AVERAGE, MEDIAN, STDEV), used COUNTIF for counting, CORREL for correlation analysis, and PERCENTILE for ranking directors.
Visualizations: Created scatter plots, histograms, Bar and trendline charts to visualize the relationships between variables.

Insights

•**Genre Impact**: Action and drama emerged as common genres, but genres like documentary and biography showed a tendency to have higher IMDB scores. Movies with multiple genres also tended to perform well.

•**Duration vs. Rating**: Movies with moderate durations (90-120 minutes) generally received higher ratings, while excessively long or short films tended to have lower ratings. This trend was supported by a positive correlation seen in the scatter plot.

•**Language**: English was the most common language, but some less common languages like Spanish and French had slightly higher average ratings. The language of the movie can influence its appeal and rating, depending on the target audience.

•**Director Influence**: Top directors had consistently high IMDB scores, with some directors regularly scoring in the top percentiles. Directors with a strong reputation and established track record influenced the movie's overall success.

•**Budget vs. Financial Success**: A moderate positive correlation was observed between budgets and gross earnings. However, some low-budget movies had disproportionately high profits, proving that financial success doesn't always require a huge budget.

Result

Through this analysis, several meaningful insights were gained:

- **Genres** such as documentary and biography tend to have higher ratings on average.
- **Movie duration** impacts ratings, with mid-length films performing better.
- **Directors** with consistent high-performing movies can drive a film's success, as evidenced by percentile calculations.
- **Budgets** and **gross earnings** are positively correlated, but high profit margins can still come from lower-budget films.

THANK YOU !!!