IMDB Movie Analysis

- **Project Description:** Investigate the factors that influence the success of a movie on IMDB, with success defined by high IMDB ratings. The aim is to provide actionable insights for movie producers, directors, and investors to help them make informed decisions in future projects.

- **Approach:**

1. **Data Cleaning:**

   - **Handling Missing Values**: Identify and address any missing data in the dataset, either by filling in values, removing rows, or using other imputation methods.
   - **Removing Duplicates**: Ensure there are no repeated entries in the dataset to maintain data integrity.
   - **Converting Data Types**: Convert columns to appropriate data types, such as date formats for release dates or numeric formats for budget and ratings.
   - **Feature Engineering**: Create new variables that may provide additional insights, such as categorizing movies into different budget ranges.

2. **Data Analysis:**

   - **Exploratory Data Analysis (EDA)**: Visualize and analyze the distribution of ratings, budgets, genres, and other key variables.
   - **Correlation Analysis**: Examine the relationships between IMDB ratings and other factors like genre, director, budget, cast, year of release, etc.
   - **Trend Analysis**: Identify patterns over time, such as changes in ratings or preferences for certain genres.

3. **Five 'Whys' Approach:**

   - Apply this technique to delve deeper into the root causes of your findings. For example, investigate why higher budgets might correlate with higher ratings and continue probing to uncover the underlying reasons.

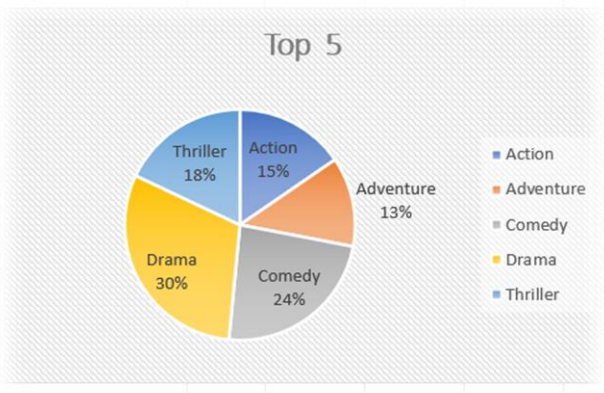
4. Report and Data Story:

- **Introduction**: Present the problem statement and objectives.
- **Methodology**: Describe the data cleaning, analysis techniques, and the rationale behind them.
- **Findings**: Summarize key insights, such as the impact of budget, genre, or director on movie success.
- **Visualizations**: Use charts and graphs to illustrate your findings clearly.
- **Conclusions and Recommendations**: Offer actionable insights that can guide future decisions for stakeholders in the movie industry.

➢ **Tech-Stack Used:** Used for initial data cleaning, exploration, and visualization. Excel's built-in functions and pivot tables help in summarizing data, performing basic statistical analysis, and creating charts to identify trends and patterns in the dataset.

➢ **Insights:**

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.
Conclusion: I identified the most common genres in the dataset and analyzed their impact on IMDB scores. Action, Biography, Crime, Comedy and Drama emerged as the most prevalent genres.



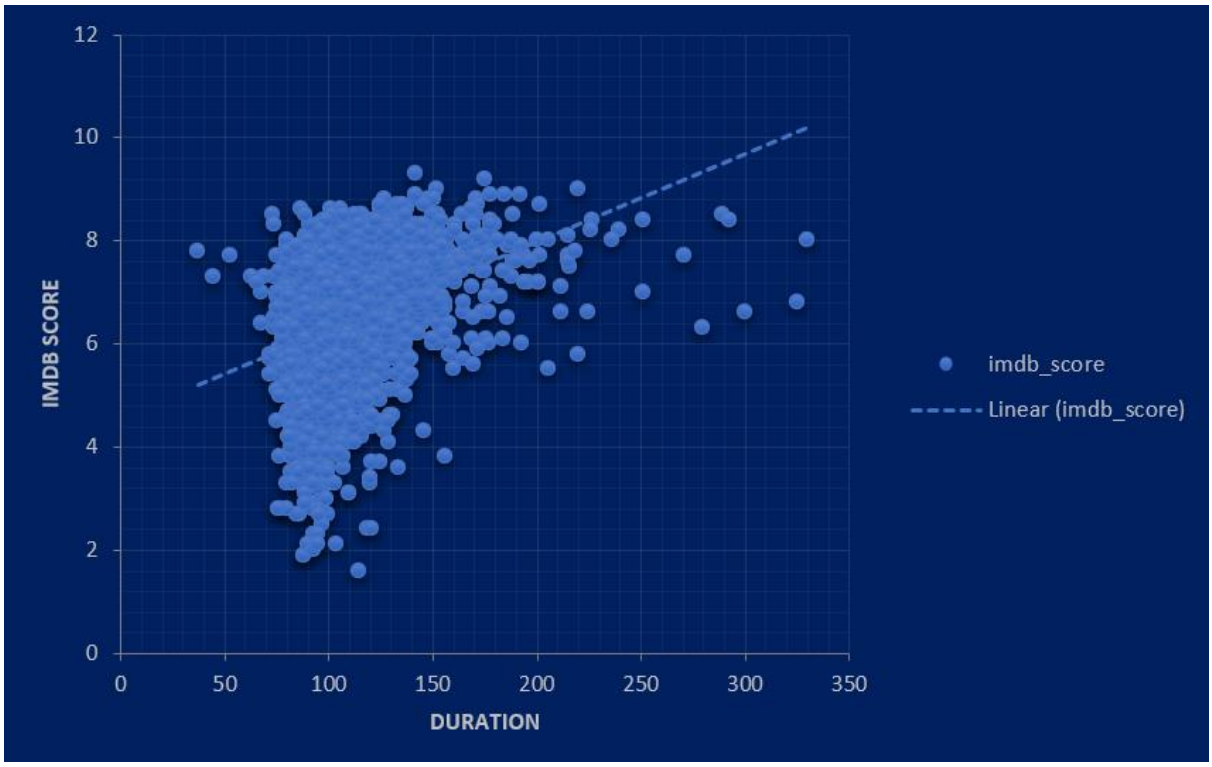
Geners Count	
Median	442
Max	1894
Min	0
Mode	15
Var	202500
Sd	450

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.

Conclusion: The distribution of movie durations varied widely, ranging from short films to lengthy epics. However, we observed a weak correlation between movie duration and IMDB score, suggesting that movie length alone does not significantly influence audience ratings.

From the below scatter plot there is a slight possibility to score 6- 8.5 IMDB score if movie duration is more than 250 minutes. With the help of trend line, we can also predict IMDB score can increase with respect to duration of the movie.

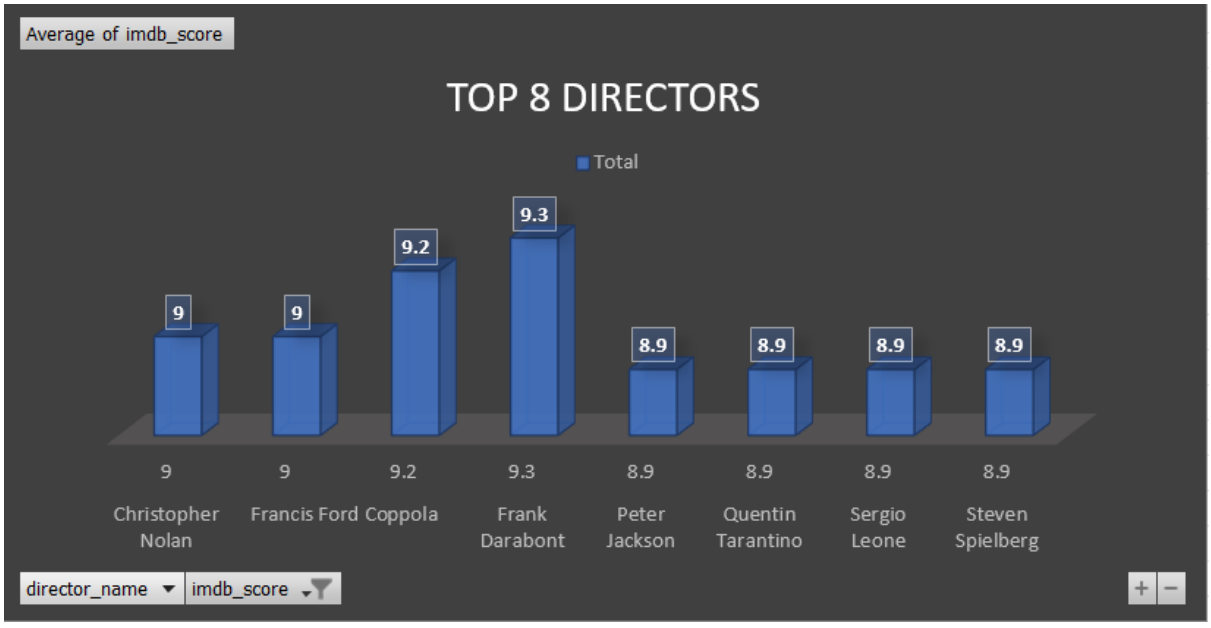


Statistics	
Average	110.26
Median	106.00
Std	22.64
Min	37
Max	330

C. **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.
Conclusion: In this IMDB data most of the movies are make in English Language and then other like

Langauge	Count of Language
English	3599
French	34
German	10
Japanese	10
Mandarin	15
Spanish	23
Grand Total	**3691**

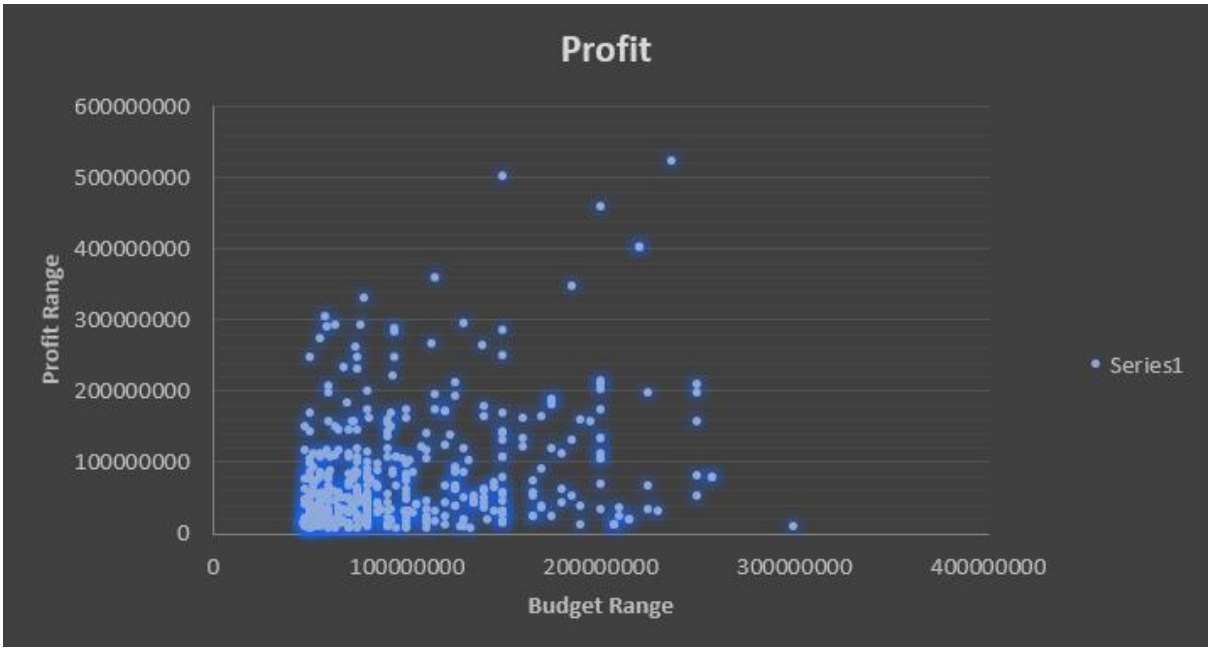
D. **Director Analysis:** Influence of directors on movie ratings.
Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
Conclusion: Top directors based on their average IMDB scores and analyzed their contribution to movie success. **Frank Darabont, Francis Ford Coppola, & Christopher Nolan consistently produced movies with high IMDB ratings**, reflecting their strong influence on audience perception.



Directors	Average of imdb_score
⊟ Christopher Nolan	**9**
9	9
⊟ Francis Ford Coppola	**9.1**
9	9
9.2	9.2
⊟ Frank Darabont	**9.3**
9.3	9.3
⊟ Peter Jackson	**8.9**
8.9	8.9
⊟ Quentin Tarantino	**8.9**
8.9	8.9
⊟ Sergio Leone	**8.9**
8.9	8.9
⊟ Steven Spielberg	**8.9**
8.9	8.9
Grand Total	**9.0125**

E. **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.
Conclusion: I analyzed the relationship between movie budgets and financial success by examining the correlation between budgets and gross earnings. The results showed a positive correlation between these two variables, but the strength of this relationship fluctuated based on factors like genre and director. Notably, films with budgets exceeding $200 million tend to have a higher likelihood of achieving significant profit margins.

Max	523505847
Min	-12213298588



Result

- ✍ Analyzed Key Factors: We looked at various aspects like genres, movie lengths, languages, directors, and budgets to understand their impact on a movie's performance.
- ✍ Identified Trends and Patterns: By examining these factors, we noticed certain trends and patterns that reveal what audiences tend to prefer and how the industry operates.
- ✍ Provided Actionable Insights: Our findings offer useful recommendations for filmmakers, producers, and studios. These insights can help them fine-tune their production strategies to create movies that are more likely to succeed and resonate with audiences.

Excel Sheet Link:
https://docs.google.com/spreadsheets/d/1gy3M8hAZixbcje25yUl7JQY0aJMKYLPb/edit?usp=sharing&ouid=116406143301160000153&rtpof=true&sd=true