PROJECT -5

IMDB Movies Analysis

[IMDB DATASET LINK](#)

Project Description: The aim of this project was to analyze a dataset of IMDB movies to uncover factors that influence the success of a movie on IMDB, as measured by high IMDB ratings. This analysis is significant for movie producers, directors, and investors seeking to understand the elements that contribute to a movie's success, enabling them to make more informed decisions for future projects.

Approach

Data Cleaning and Preprocessing:

- Handling Missing Values： Identified and imputed missing values using the median for numeric columns and the mode for categorical columns.
- Converting Data Types**:** Ensured each column had the appropriate data type for analysis.
- Removing Duplicates**:** Checked and removed any duplicate records to ensure data integrity.

Data Analysis:

- Movie Genre Analysis: Determined the most common genres and calculated descriptive statistics for IMDB scores for each genre.
- Movie Duration Analysis: Calculated descriptive statistics for movie durations and visualized the relationship between movie duration and IMDB scores using scatter plots and trendlines.
- Language Analysis: Identified the most common languages and calculated descriptive statistics for IMDB scores for each language.
- Director Analysis: Calculated average IMDB scores for directors and identified top directors using percentile calculations.
- Budget Analysis: Analyzed the correlation between movie budgets and gross earnings, and identified movies with the highest profit margins.

Tech-Stack Used

- **Microsoft Excel 2019**

Insights

- **Genre Analysis**: Certain genres, such as History, Biography, War and Drama, tend to have higher IMDB ratings compared to others. Descriptive statistics revealed that these genres often have higher mean and median ratings.
- **Duration Analysis:** There is a positive correlation between movie duration and IMDB score, indicated by the upward-sloping trendline. This suggests that, generally, as the duration of a movie increases, its IMDB score tends to be higher. Movies with durations between 90 to 150 minutes generally received higher IMDB ratings. A scatter plot showed a positive correlation between movie duration and IMDB scores.
- **Language Analysis:** English-language movies dominate the dataset and tend to have higher IMDB ratings on average. Other languages also showed varying impacts on ratings.
- **Director Analysis:** Directors like Christopher Nolan and Steven Spielberg consistently produced high-rated movies, as revealed by their high average IMDB scores.
- **Budget Analysis:** There is a moderate positive correlation between movie budgets and gross earnings. Movies with higher budgets generally achieved higher gross earnings, and the highest profit margins were observed in movies with balanced budget-to-earnings ratios with movie titled "Paranormal Activity".

Results

The project successfully identified several factors influencing the success of movies on IMDB. Key achievements include:

- Understanding the impact of movie genres on IMDB ratings.
- Identifying the optimal range for movie durations to achieve higher ratings.
- Highlighting the significance of language in movie ratings.
- Recognizing top-performing directors based on average IMDB scores.
- Establishing a correlation between movie budgets and financial success.

Overall, this analysis provided valuable insights into the factors that contribute to a movie's success on IMDB, offering actionable knowledge for movie industry stakeholders to make informed decisions in their future projects.