DATA602 Final Project Proposal

Key Factors Influencing Box Office and Critical Ratings of Movies

Team Members

- Fomba Kassoh
- Souleymane Doumbia

1. **Research Question**: How do factors such as box office success, genre, and critical ratings influence the likelihood of a movie winning major awards like the Oscars?

2. **Justification**: This analysis is relevant both to the film industry and marketing agencies to better understand what influences the success of movies. This can inform decisions on movie production, marketing strategies, and can provide insights into consumer preferences.

3. **Data Sources**: The dataset was obtained from IMDb (Internet Movie Database), a comprehensive source of movie data. We collected this data by web scraping IMDb's website, considering the nature of the fields present. The specific URL is provided below.

   [Top 100 Greatest Movies of All Time (The Ultimate List) - IMDb](#)

4. **Libraries Potentially Being Used**: Analysis will be conducted using Python with the following libraries:

   - Pandas: Data manipulation and analysis

   - NumPy: Numerical operations

   - Matplotlib/Seaborn: Data visualization

   - BeautifulSoup for web scraping

   - Requests and lxml

5. **EDA and Summary Statistics**: Preliminary exploration will focus on distributions of numerical features such as ratings and revenue, as well as categorical features like genres and certificates, directors, and stars. Summary statistics will provide insights into mean, median, and mode for critical scores and box office revenue. Below is the link to the EDA.

[DATA602Project/Project_Proposal_EDA.ipynb at main · hawa1983/DATA602Project (github.com)](#)