**Project 1: Proposal**

Avani Kala (NUID: 002772623), Yashvi Bhandari (NUID: 002122037)

September 25th, 2023

Prof. Satwik Kamarthi (IE 6600: Computation and Visualizations)

**Project Background**

The world of cinema is a fascinating realm with the capacity to enthrall, motivate, and evoke emotion. Recent studies show that nearly 79% of adults in the US watch movies and series every day. Our effort on IMDb movie analysis is motivated by a desire to comprehend the dynamics of the entertainment industry, and to find hidden insights inside IMDb's massive database. We believe that this project will not only enhance our own knowledge and personal development but also offer valuable insights to anyone interested in the realm of cinema.

**Project Goals**

Our motivation to undertake a project focused on IMDb movie analysis is underpinned by several compelling factors. Using IMDb's huge collection of movie ratings, reviews, and metadata, we first want to promote a greater understanding of the art of filmmaking itself by identifying the aspects of certain movies—such as direction, storyline, and performances—that stand out. Through this project, we aspire to gain a profound understanding of various aspects of movies such as industry insights, preferences of audience, critical reception, and production budgets. Along with the appreciation of cinema we aim to unearth the trends of the entertainment industry, and factors that contribute to a movie's success.

**Dataset**

**IMDB Movie Dataset** [1] is a vast collection of data related to movies and TV shows. It consists of 2.5 million entries listed on the official website of IMDB. With this information we can successfully perform data analysis and apply machine learning algorithms to explore and predict the trends of the entertainment industry. Some of the key columns of the dataset are described below:

* IMDb Rating: The rating of movies on IMDb
* Genre: The genres of the movie (e.g., action, drama, comedy, etc.).
* Revenue: The revenue or earnings generated by the movie.
* Votes: The number of votes or user ratings the movie has received on IMDb
* Cast: Information about the actors and actresses in the movie.
* Release: The year in which movies/shows were released.
* Certification: Category of certification provided by CBFC

**Sources**

[1]<https://www.kaggle.com/datasets/ashishjangra27/imdb-movies-dataset?select=Genre>

[2]<https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows>