Project Title: **Predicting Box Office Success Using Movie Data**

Proposal:

My project will attempt to create a predictive model to estimate a movie’s box office revenue based on a range of data available prior to the movies release. Using features like genre, budget, cast, director, length and rating, the project will explore the factors that are most influential towards a movies success. The data will be taken from Kaggle that originated from sources such as IMDB, The Movie Database and Box Office Mojo. I will be cleaning and merging datasets to perform data analysis and regression models to predict its revenue using Python. The challenges I might face are handling missing or inconsistent data as well as ethical considerations such as biases such as which types of stories get funded or which actors and directors get the most opportunities. I will attempt to make sure that these types of patterns are not reinforced. The goal will be to find patterns in the data to predict success rather than decide which types of movies should necessarily be made.

Data Sources:

<https://www.kaggle.com/datasets/michaelmatta0/movies-ultimate-metrics-features-and-metadata>

<https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>

<https://www.kaggle.com/datasets/asaniczka/tmdb-movies-dataset-2023-930k-movies>

(Possibly too large)

<https://www.kaggle.com/datasets/ggtejas/tmdb-imdb-merged-movies-dataset>

<https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata>