Capstone Project#1 Proposal

1. Problem

Today's movie watcher is spoilt for choice with increase in movie streaming platforms, DVD rental services and easy access to movie theatres. Knowing a movie's rating before watching the movie helps reduce the decision fatigue arising out of increased options. IMDb (Internet Movie database) is often a go-to source for knowing a movie's ratings, user reviews, plot, casting and other details before watching that movie. The ratings in IMDb are a function of the ratings given by the users.

But unfortunately, IMDb allows users to provide ratings and reviews only after movie release. So, if you want to select a movie for a first day show, you don't have the ratings to rely on.

2. Objective

- Develop a model to predict rating of an upcoming movie based on its cast, director, genre and release date
- Predict the revenue that an upcoming movie would earn based on the above factors

3. Client

- Movie watchers
    - Will have a rating even before a movie's release to decide whether to watch the movie or not
- Movie Theatres
    - The buyers representing the theatres can use the model to decide which movie to lease
    - They can further use the model to decide for how many weeks they should show the movie to have an optimum box-office collection
- Movie makers
    - The model will be using factors like cast, director, genre to predict revenue earned. So the movie makers can use this model to find a perfect recipe for a movie

4. Dataset

The dataset to be used will be obtained from IMDb website:
https://www.imdb.com/interfaces/

5. Approach to the Solution

- Filter and clean the data
- Find correlation between
- Identify key parameters that affect movie ratings
- Construct a model for prediction

6. Deliverables

- Power-point presentation
- Code repository on Github