

Capstone 1 Project Proposal

This capstone project aims to build a predictive model that can accurately the final North American box office performance of a film. Factors the model will be taking into account include budget, genre, opening weekend gross, aggregated critics as well as audience ratings. The model will only take into account North American box office numbers, due to many films, not releasing simultaneously or at all in certain regions. This results in the inability to fairly compare a film's worldwide box office performance to another. There is a larger pool of data for the North American box office which remains consistent, therefore this capstone will favour the analysis of North American box office gross instead of worldwide numbers. While the opening weekend will be used to analyze the performance of the film, it will not be predicted by this model. A film's opening weekend gross is greatly dependent on various factors such as the cost of marketing, which most studios do not report to the public.

This project will gather and combine data from multiple sources providing different factors to make accurate and precise predictions. The box office numbers for final and opening weekend domestic gross will be taken from Box Office Mojo, the industry leader for reported box office figures. The MoviesLens dataset containing user reviews from over a quarter of a million different users will be combined with the Open Movie Database (OMDB) API, which provides the critics' ratings of films from websites such as Rotten Tomatoes and Metacritic. The OMDB data also includes IMDB ratings from its users which will be used alongside the user reviews from the earlier dataset. The number of user reviews for each film will prove useful in avoiding overreliance on user ratings if there is not a wide range of reviews available.

This model alongside predicting final gross numbers will compare and contrasts the importance of each factor provided. Do critics rating on big-budget blockbusters have the same impact as the audience ratings? An example of the divisive nature of critics and audience ratings is *Star Wars: The Last Jedi* which garnered heaps of critical praise but left many in the Star Wars fan community visibly divided, yielding a significantly lower audience rating. The opposite is true for the recent release of *Joker*, despite mixed reviews from critics the film has scored high ratings from the audience and has made a big splash at the box office. How do films of different genres but similar budgets perform in contrast to the same genres and different budgets? This model aims to provide useful insights to both trackers of box office numbers as well as film studios. Film studios often test screen with audiences as well as screen and permit critics to write about their films often weeks in advance to help build anticipation for their films. This model can help studios better understand how impactful audience and critics ratings will be to their film's overall box office performance.

Dataset Links:

https://www.boxofficemojo.com/chart/top_lifetime_gross/?ref=bo_cso_ac

https://www.boxofficemojo.com/chart/top_opening_weekend/?ref=bo_csw_ac

https://www.kaggle.com/rounakbanik/the-movies-dataset#movies_metadata.csv

<http://www.omdbapi.com/>