

Project Luther Proposal

Predicting a Movie Sequel's Box Office Return

Phoi-Tack (Charlie) Lew

Domain

The reported revenue from Hollywood's movie industry in 2018 was close to USD 41 Billion. It may seem like a lucrative business but it is mired in risk and uncertainty. Not all movies produced by Tinseltown are sure successes and even if they are, will a movie sequel guarantee financial success? This project aims to explore and gain better insights into this question through the use of web scrapping and regression tools from at two movie database sources. Those sources are from the Internet Movie Database (www.imdb.com) and Box Office Mojo (www.boxofficemojo.com).

MVP

A Minimum Viable Product will be to at least predict the world-wide box office return as a function of release time between a sequel and its original. Other correlations to explore are movie budget and audience score for example.

DATA

As mentioned, two movie databases will be used to determine the prediction. Below are the proposed features that will be extracted from the databases.

Number	Feature	Type	Dimension
1	Time between Release	Numeric (integer)	Months
2	World-Wide Box Office Return	Numeric (float)	USD
3	Domestic Box Office Return	Numeric (float)	USD
4	Production Budget	Numeric (float)	USD
5	IMDB Rating	Numeric (float)	-
6	Metascore	Numeric (float)	-
7	Runtime	Numeric (float)	Minutes
8	Release Year	Numeric (integer)	Year
9	Release Month	Numeric (integer)	Month (1-12)
10	Weeks in Theaters	Numeric (integer)	Weeks

Table 1: Features for predicting a sequel's box office return