Challenge Proposal

Predicting Movie Earnings

Thomas van der Molen

S4-AI41

Make this into a notebook instead

|  |  |
| --- | --- |
| **Project Information** | |
| Project members | Thomas van der Molen (4168003) |
| Project name | Challenge Proposal |
| Version | 1.2 |

Table of Contents

[Version History 2](#_Toc97547208)

[Domain Understanding (Chapter 1) 3](#_Toc97547209)

[Introduction 3](#_Toc97547210)

[Problem Statement 3](#_Toc97547211)

[Target Audience 4](#_Toc97547212)

[Expert Domain Clarification 4](#_Toc97547213)

[Data Sourcing (Chapter 2) 4](#_Toc97547214)

[Exploratory Data Analysis 4](#_Toc97547215)

[Analytic Approach (Chapter 3) 5](#_Toc97547216)

[Evaluation (Chapter 4) 6](#_Toc97547217)

# Version History

|  |  |  |
| --- | --- | --- |
| **Version** | **Date** | **Change** |
| 1.0 | 17-02-2022 | Created document |
| 1.1 | 18-02-2022 | Started work on the introduction |
| 1.2 | 25-02-2022 | Added more structure based on SI lecture |

# Domain Understanding (Chapter 1)

## Introduction

As someone who enjoys watching movies, there often seems to be a polarizing difference between what movies perform well and which ones do not. With movies being [a billion-dollar industry](https://www.boxofficepro.com/mpa-2019-global-box-office-and-home-entertainment-surpasses-100-billion/), there seems to be a lot of risk and reward involved when investing and making movies.

I am interested in finding possible patterns to predict how much a movie will earn in the box office compared to the amount of money spent on budget. This could give movie companies a clearer vision of how much they should spend for the best result, reducing the amount of risk taken when investing in movies with budgets of [100 million+ dollars](https://www.the-numbers.com/movie/budgets/all).

## Problem Statement

For this project I Will be trying to predict:

*How much will a movie earn compared to its budget based on publicly available data.*

## Target Audience

There are a lot of different parties involved during the process of making a movie, for this reason there are many different groups that could benefit from the given predictions.

The most obvious ones are the production and distribution companies such as: [Paramount Pictures Studios](http://www.paramountstudios.com/), [Warner Bros. Pictures](https://www.warnerbros.com/company/divisions/motion-pictures), [Sony Pictures](https://www.sonypictures.com/), [Lionsgate](https://www.lionsgate.com/), etc.

Further from the actual production and distribution process are companies such as [Peacock Film Finance](https://peacockfilmfinance.com/), who as the name suggests help finance the production of movies.

Besides these you could also include Bond companies, these companies are a form of insurance agency such as [Surety Bonds Direct](https://www.suretybondsdirect.com/).

I am going to focus my domain on the production companies, because these are the companies that will originally set the budget and could use this prediction to get investments and get more people interested in their movie.

## Expert Domain Clarification

I have tried reaching out to several (smaller) movie production companies to try and get an interview with them such as [24fps productions](https://www.24fpsproductions.com/). Sadly, there has been no responses to my requests.

# Data Sourcing (Chapter 2)

One of the other reasons I chose to try and work with movies, is because movies seem to have a lot of public data attached to them for example, most movies share their box office revenue ([or are tracked by third parties)](https://movies.stackexchange.com/questions/27777/why-are-box-office-numbers-made-public) and in most cases movie budgets will also be provided. Other than this information such as the cast, genre, runtime, release date and critic scores are also widely available.

For sourcing data there are a few well known sources to get large amounts of data from movies, such as [IMDb](https://www.imdb.com/), [TMDB](https://www.themoviedb.org/) and [Rotten Tomatoes](https://www.rottentomatoes.com/).

## Exploratory Data Analysis

# Analytic Approach (Chapter 3)

What models can I use

# Evaluation (Chapter 4)

Delivering / deploying the product