

 [hixtape1212](#) / [Group6Consulting\\_CapstoneRepo](#) Public

Capstone repository for the group

☆ 0 stars    🔗 1 fork

☆ Star

👁 Unwatch ▼

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

🔗 main ▼

...



hixtape1212 Update powerpoint link pt. 2 ...

2 minutes ago

🕒 22

[View code](#)

☰ README.md



# Group 6 Consulting

## Computing Vision Movie Studio Analysis



# Computing Vision

- [Project Overview](#): The overall project goal and approach.
- [Navigation Instructions](#): Guidance for how to navigate the repo.
- [Data Science Steps](#): Summary of the steps our team took to arrive at our recommendation.

## Relevant Files

- **Jupyter Notebook:** Here's the link to our teams Jupyter Notebook. This notebook will walk through our analysis.
- **Presentation:** Here's the link to our teams presentation. It includes our recommendations and analysis findings.

## Project Overview

---

Our team was tasked with helping Computing Vision startup their own movie studio. The firm hasn't had a lot of experience in the movie industry, and wanted our firm to run the analytics, and recommend what types of films to create as well as give Computing Vision a good idea of what to expect when opening their new studio.

## Navigating the Repo

---

In this repo you will find the following:

- This ReadMe.md file
- Jupyter Notebook – Contains all relevant code used for our data analysis.
- PowerPoint.pptx – This is the presentation of our analysis and findings.

What is not contained in this repo:

- Data File – Our team used a collection of different sources, some too big to be put on this repository. The 2 files our team used were the IMDB SQL database as well as The Numbers csv. With that being said, the data files that we used were in path 'data/\*'.

## Data Science Steps

---

Below is a list of steps that our team took as well as the corresponding description. Since our task was to discover what types of films to create as well as give the firm an idea of what to expect, we based our search around a central metric (Return-on-Investment) to help the firm realize as much return as possible. This included analysis on ROI by genre, rating, and overall production budgets effects on outcome.

### Data Gathering and Cleaning

To start our team conducted EDA on all the source files to identify those which seemed to be the best for our approach to analysis. We were especially looking for files that contained movie names, genres, revenue, and costs. We settled on two of the datasets:

- IMDB SQLite database (data/im.db)
- The Numbers Movie Budgets dataset (data/tn.movie\_budgets.csv)

Combined they both contained the necessary information that we needed. The Numbers included gross revenue worldwide, domestic, and a product budget which our team considered cost. IMDB contained information on genre as well as additional attributes if we needed to get more granular. Altogether, the two datasets were combined on movie titles.

Domestic gross, worldwide gross, and production budget were cleaned to be integers, and a column for ROI calculated as  $(\text{Worldwide Gross} - \text{Production Budget}) / \text{Production Budget}$ , was added to the dataset. Duplicate movie titles were dropped, and we now had a workable data set to run analysis from.

## Genre Analysis and Hypothesis Test

We used IMDB and The Numbers data sets to identify which movie genres we can recommend based on analysis from the average worldwide gross, domestic gross, budget gross and compared this factors to a calculated ROI, we did some hypothesis testing to judge whether a genre in the population is compatible with the observed ROI of a sample of that genre, this in order to show an idea of which films generate the most money, the most expensive to make and, of course, the best ones that are worth investing in producing.

## Diminishing Returns Based off Budget

We used the IMDB database to determine the ideal budgeting strategy to optimize Computing Vision's ratings without braking the bank. We found that there was a sharp uptick in return on investment after the 10-17.5 million dollar mark for the budget in ratings. This is where the most value is and a great way for starting directors/staff/artistic directions to be tested out without committing to a massive budget. We also found there was a lul in marginal ratings inprovement and ROI based on budget until the 65 Million dollar mark. Movies produced with 50 million dollars were performing the same as ones with a 15 million dollar budget. The best budget to optimize ratings was between 95 and 105 million dollars since there was a steep increase after the 65 million dollar mark. After the 105 mark, there were deminishing returns which lead us to the conclusion that there should be two types of movies produced by Computing Vision: 10-17.5mil films and 95-105mil films.

## Ratings to Worldwide Revenue Assessment

We used the IMDB database to analyze if greater ratings truly led to higher revenues. This was completed by compiling thousands of movies and plotting their respective ratings and revenues to generate a line of best fit. There was a positive slope and correlation. As a result we were able to conclue that higher ratings in general generate greater returns in the box office. More specifically movies with a 6 or greater rating performed significantly better that the others. We also noticed a diminishing return after the rating of 8, so the investment of time and money to get above that rating are most likely not worth it.

## RunTime to Movie Rating

We used the IMDB database to identify whether there is a correlation between a movie's runtime and its rating. We used the process of first finding the mean rating of all the movies in the database, then we created a dataframe that dropped the movies that had a rating below the population's mean. From there we created a density graph and then made a recommendation of ideal runtime to optimize the movie's rating.

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Contributors 4



**hixtape1212** Sam Martin



**psolorio95** patricio solorio



**deapillai**



**mschliesman**

## Languages

● **Jupyter Notebook** 100.0%