View license

☆ 0 stars   476 forks   ⊙ 0 watching   Activity

Public repository · Forked from learn-co-curriculum/dsc-phase-1-project-v2-4

master

Branches   Tags

This branch is 22 commits ahead of learn-co-curriculum:master.   Contribute   Sync fork

bennettandrewm Delete movie_data_erd.jpeg   ...   10 minutes ago   40

View code

≡ README.md

BREAKING DOWN THE BLOCK BUSTERS

Analysis of Top Grossing Movies in Theaters of Last 20 years

Author: Andrew Bennett

Project Overview

To aid Microsoft on its voyage into movie making movies, an analysis of approximately 2,000 movies of the last 20 years was performed to begin to understand what movies are doing well at the box office. This repository contains that analysis including all computer coding, databases, csv files, and other supporting documents. The purpose was to provide concrete recommendations that could be used to launch a movie studio. Based on two reliable industry databases (IMDB and *The Numbers*), the analysis looked at genre, production budgets, and directors and recommends the following:

- focus on movies in the Adventure Genre
- invest heavily in big budget movies ($70M)
- hire directors with a track record of Profitability doing Big Budget Adventure Movies

"The business of film is to make dreams into reality so audiences can escape for a few hours."

Business Problem

Microsoft sees all the big companies creating original video content and wants to launch a movie studio. The problem? They've never done this before, and don't know the first thing. But that's never stopped anyone before.

So how do we learn what movies to make today, that will perform in the future? Let's study the past. Specifically, the box office success of all the movies of the last 20 years. If we can glean insight into what goes into making successful movies, perhaps Microsoft can reproduce blockbuster magic.

So, we're going to access some publically available datasets and focus our analysis on three major variables in moving making.

- genre
- budget
- directors

These factors represent some of the simplest, most foundational decisions. Genre determines which scripts and creative teams you'll need. The budget is the money you'll need upfront to greenlight a project. And the director is the creative lead who will shepard the day-to-day movie-making and ultimately deliver the project to you and your audience.

Lucky for us, the information is readily available. But what specifically are we going to measure? To gauge financial success, we'll look at Return on Investment (ROI) and the profit margin on the Worldwide Gross Revenue, which we'll call Gross_Margin. We calculate ROI as the Worldwide Gross Revenue divided by the production budget. So, an ROI > 1 means a movie made a profit. ROI < 1 and the movie lost money. Gross_Margin is Worldwide Gross Revenue minus production budget. This measures how much profit, in dollars, was made compared to a movie's cost to produce.

These will be the two financial performance metrics we'll use to compare against the genre, budget, and directors. We should take a second to note that we have not factored in marketing or distribution costs into either ROI or Gross Margin. This data is not as readily accessible on this scale. These costs are also not known BEFORE a movie is greenlit, so it's hard to consider these costs as inputs, however they do affect the REAL profits of a movie studio when they're all accounted for.

So let's begin...

This Repository

This Repository Structure is straightfoward to use

Folder

- Data Folder - data (zipped and unzipped) used for this work
- Images Folder - image files you see here in this Readme, in the presentation, and in the
- PDFs Folder - Jupter Notebook and Presentation as PDFs

Files

- Microsoft_Analysis - the Jupyter Notebook with the active code
- Readme - is... what you're reading now
- License - is for Flatiron
- .Gitignore - is the ignore file
- .canvas - is related to canvas software used as part of the Flatiron curriculum

Data Understanding and Analysis

The Data Source

To solve this problem we used well-known industry resources where are included in this repository in the folder `zippedData` . These files contain movie datasets from the following publicly available websites:

- Box Office Mojo
- IMDB
- Rotten Tomatoes
- TheMovieDB

- The Numbers

As we dug into the files, we eventually used two sources for our analysis.

- `im.db.zip`
  - Zipped SQLite database (this was unzipped in the BASH window using the unzip command)
- `zippedData/tn.movie_budgets.csv.gz`
  - Compressed CSV file (opened using `pd.read_csv` in the Jupyter Notebook)

The IMDB movie database (https://www.imdb.com/) is one of the most complete, publicly box office available databases. It's a well known industry staple and incredibly robust. But... it doesn't have budget numbers. We'll get from The Numbers (https://www.the-numbers.com/) which reports daily and weekly box office sales for movies currently in release. From there, we examine the movie titles, genres, production budgets, directors and their respective gross box office receipts.

The following diagram (courtesy of Flatiron) was useful for the analysis with the IMDB database. It shows the heirarchy of the database we used.
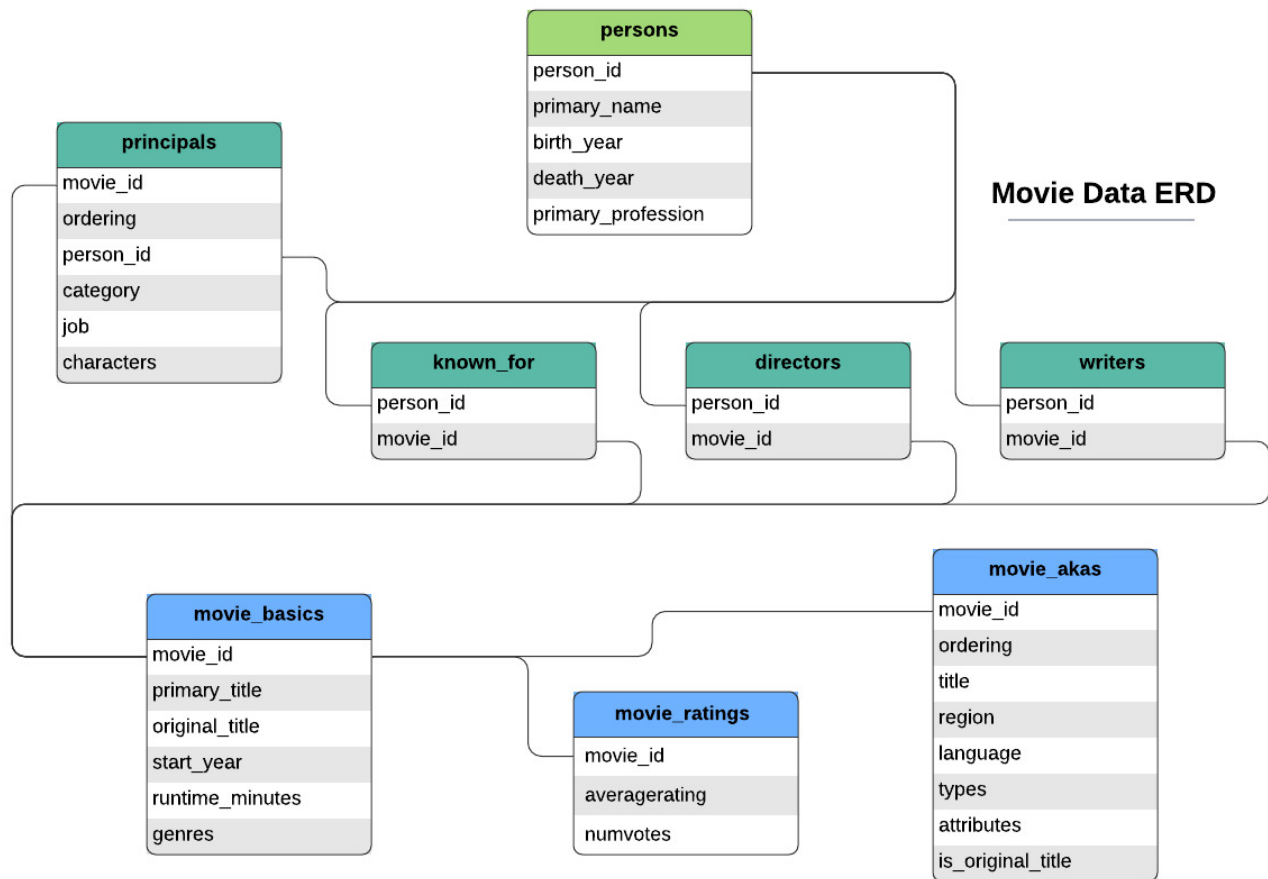
RESULTS

GENRE

To examine genre, the goal was to find a correlation between the roughly 20 or so categories and the financial metrics, Gross_Margin in particular. When we correlated, we found that the Adventure Genre correlates most strongly with Gross_Margin.
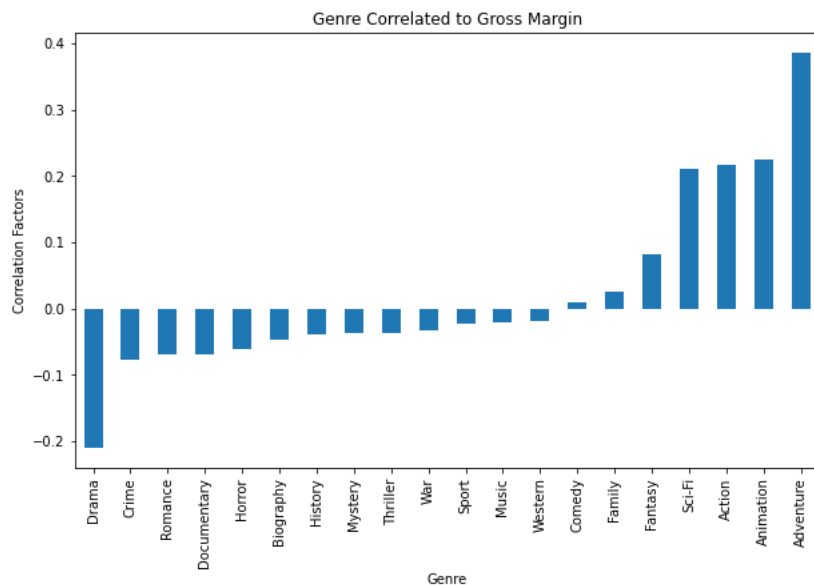
GRAPH 1 - Correlation Factors of Genre on Gross_Margin.

We can see that it's nearly double the correlation factor of the next genres which are Sci-Fi, Action, and Animation. These genres also positiviely correlate, pairing well with Adventure. When we think about these categories. it's important to note that rarely is a movie just one genre. More common to see a Sci-Fi, Adventure movie or Animated, Adventure movie. And the data bares this out, with one common theme. ADVENTURE!

BUDGET

To examine budget, we wanted to see if there was any range of budget that seemed to perform better than othes. To do this, we created budget categories and measured the distribution of these categories against the ROIs. The goal was to see if there was a trend between HOW MANY movies success or fail (based on an ROI = 1) based on their budget. What we found was...
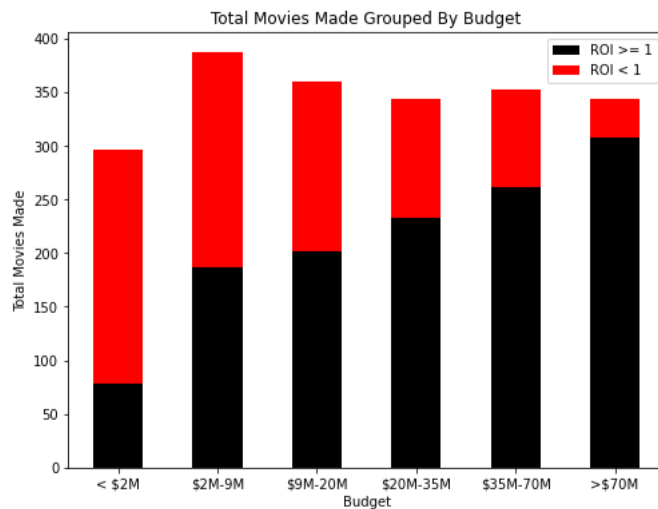

GRAPH 2 - Distribution of Movies by budget with ROI greater than 1

Many more movies with the largest budgets (>$70M) have an ROI >= 1. In fact, there appears to be a trend where as the budgets get larger, the more movies make money. It's important to know that these numbers do not factor in cost of distribution and marketing.

DIRECTORS

To determine the best directors, we wanted to see who had the best track record in the budget range. We found all of the directors who made AT LEAST 2 movies in this range, and looked at their average profit. In other words, we ranked these directors in descending order based on mean Gross Margin and Voila!

Average ROI by Director for >$70M Budget Adventure Films
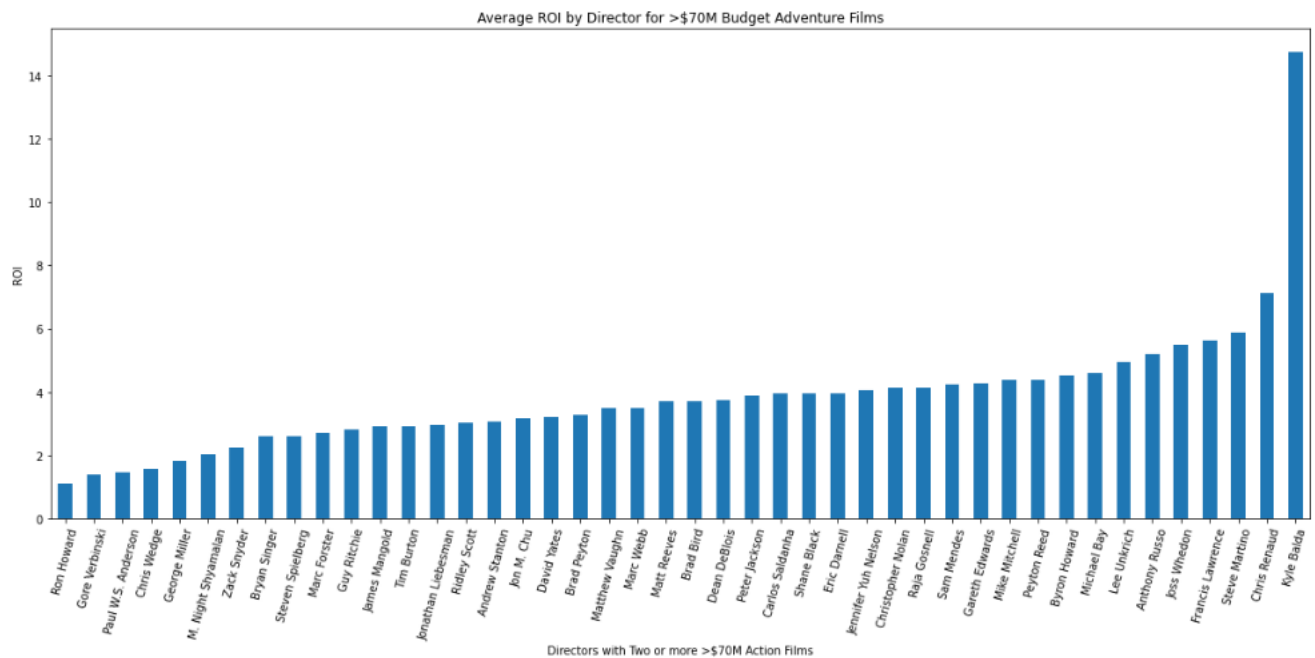
GRAPH 3 - Directors with Best Average ROI in the $>70M range

There are 44 directors who have made at least 2 successful Adventure. Kyle Banda, Chris Renaud, Steve Martino, Francis Lawrence, Joss Whedon, and Anthony Russo all have ROI > 5. Directors matter.

Conclusion - Directors with Best Average Margins in the $>70M range

While there's much more work to done to launch the movie studio, the data from the past 20 years shows, Adventure movies in (Sci-Fi, Animation, and Action) are most strongly correlated with Gross Profit Margin. More big budget movies break even than any other budget. And finding a director matters. Microsoft should target these directors (provided in the visual) on this list that have a track record of making profitable, Big Budget Adventure movies. So... let's greenlight the studio!

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%