MOVIE DATA ANALYSIS

Cindy Tumaini

24 July 2024

Outline

- Overview
- Business Problem
- Data
- Methods
- Results
- Conclusions

1. Overview

- This project analyzes data from multiple sources from the film industry in order to **determine which films yield the highest ROI (Return on Investment/ profitability)** based on genres and the release month in order to advise a company planning on starting a film studio. Descriptive analysis enables us to dtermine which films perform the best based on ROI and average rating of the films to enable the company to decide what films to create.

2. Business Problem

- Your company now sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with **exploring what types of films are currently doing the best at the box office.** You must then *translate those findings into actionable insights* that the head of your company's new movie studio can use to **help decide what type of films to create.**

3.Data

- The data sources were from the zipped file that contained the 'im.db' file that contained the **movie information such as title, genres, directors, rating** etc and the 'tn.movie_budgets.csv.gz' that **contained the numbers data such as production budget and worldwide gross**.

- From the 'im.db' an sql query was used to select relevant data from the tables within the database.

- After cleaning and mergng the data, a new feature was generated called **ROI** (*return on investment*) that calculated the net profit of the movies. This feature was then used to analyse the data by aggregating values based on genre, directors, release months in order to determine profitability.

4. Methods

4.1.Data Cleaning

- The data was cleaned by renaming thecolumns to enable readability and to make it easy to merge the dataframes.
- Columns were converted into their required data types e.g **worldwide_gross** column from text to numeric data.
- The records with null values in the required columns were dropped.

4.2.Analysis

- Descriptive analysis was performed where aggregate values of mean, count and median of the ROI column was conducted.
- Exploratory analysis wascarried out to take a deep dive into the data to find out inconsistencies as well as connections
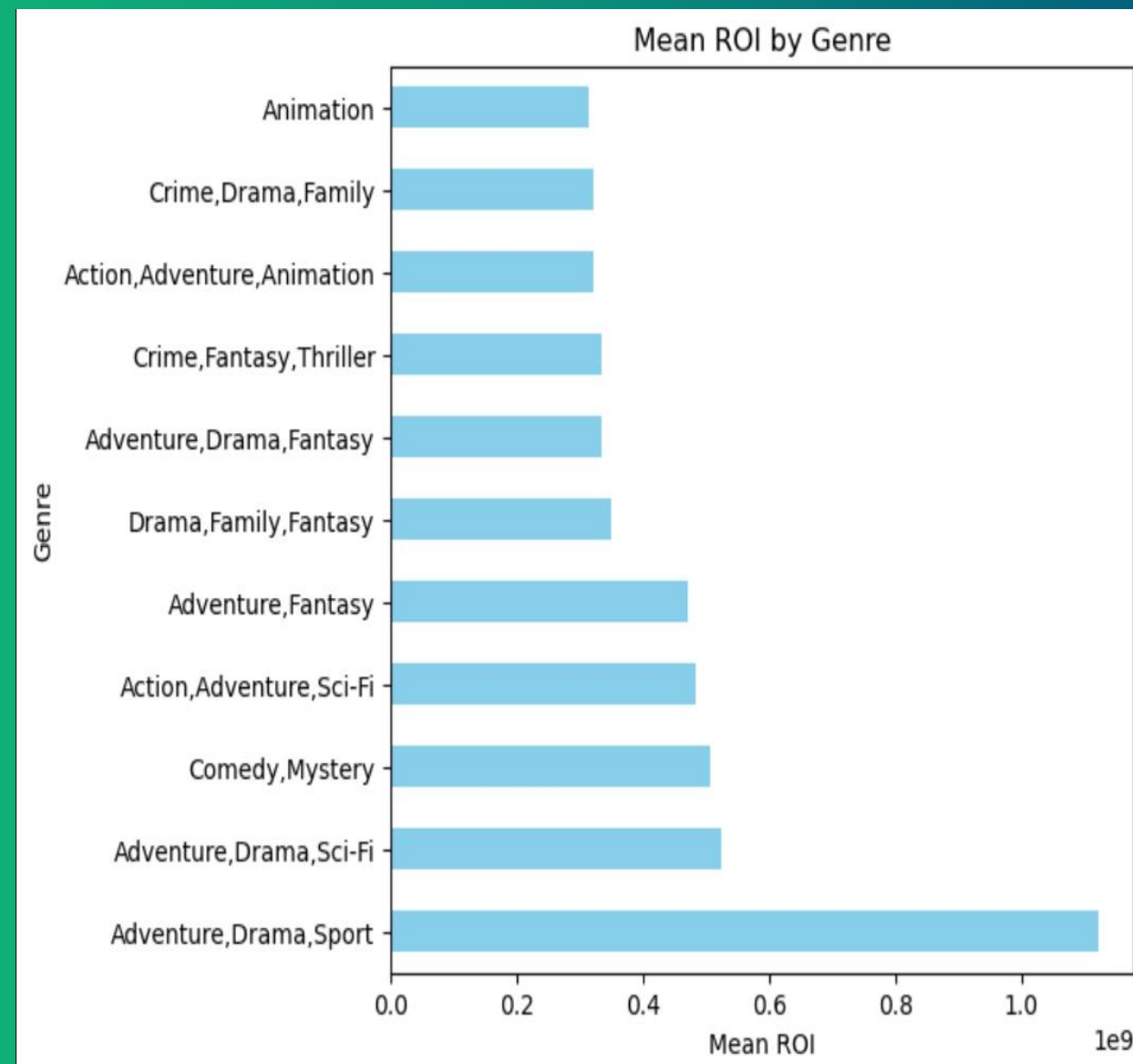
5. Results

Box Office Perfomance

- From the data, the highest profiting genres seem to be **Sci-Fi** and **Adventure** as the most profitable genres appearing in different combinations.

- The most profitable genre combinations in order are:

1. **Action, Adventure, Sci-Fi**

2. **Adventure,Drama,Sci-Fi**

3. **Comedy,Mystery**

I ignored the first value suggested by the data in the means beacuse I consider it an outlier and could skew the expected results and I chose

- **Drama**, **Comedy** and **Action** also seem to be highly profitable in different combinations. The **Documentary** genre was ignored because it only appeared in freqency rather than mean and median ROI.


Mean ROI by Genre

5. Results

5.1 **Release month against ROI**

- In this analysis I invetigate which would be the most profitable months to release a movie.

- Most movies are released within the lasy three months of the year i.e **October, November** and **December**. However the months with the **highest average ROI** seem to be in the **middle of the year** between **May, June** and **July** with **November** following closely behind.

- The median average ROI of films relased by months emphasizes that although there are many films that do not provide a high ROI (some under $10 million) films typically released in **July** or **November** are bound to do well.

Mean and Median ROI for each month
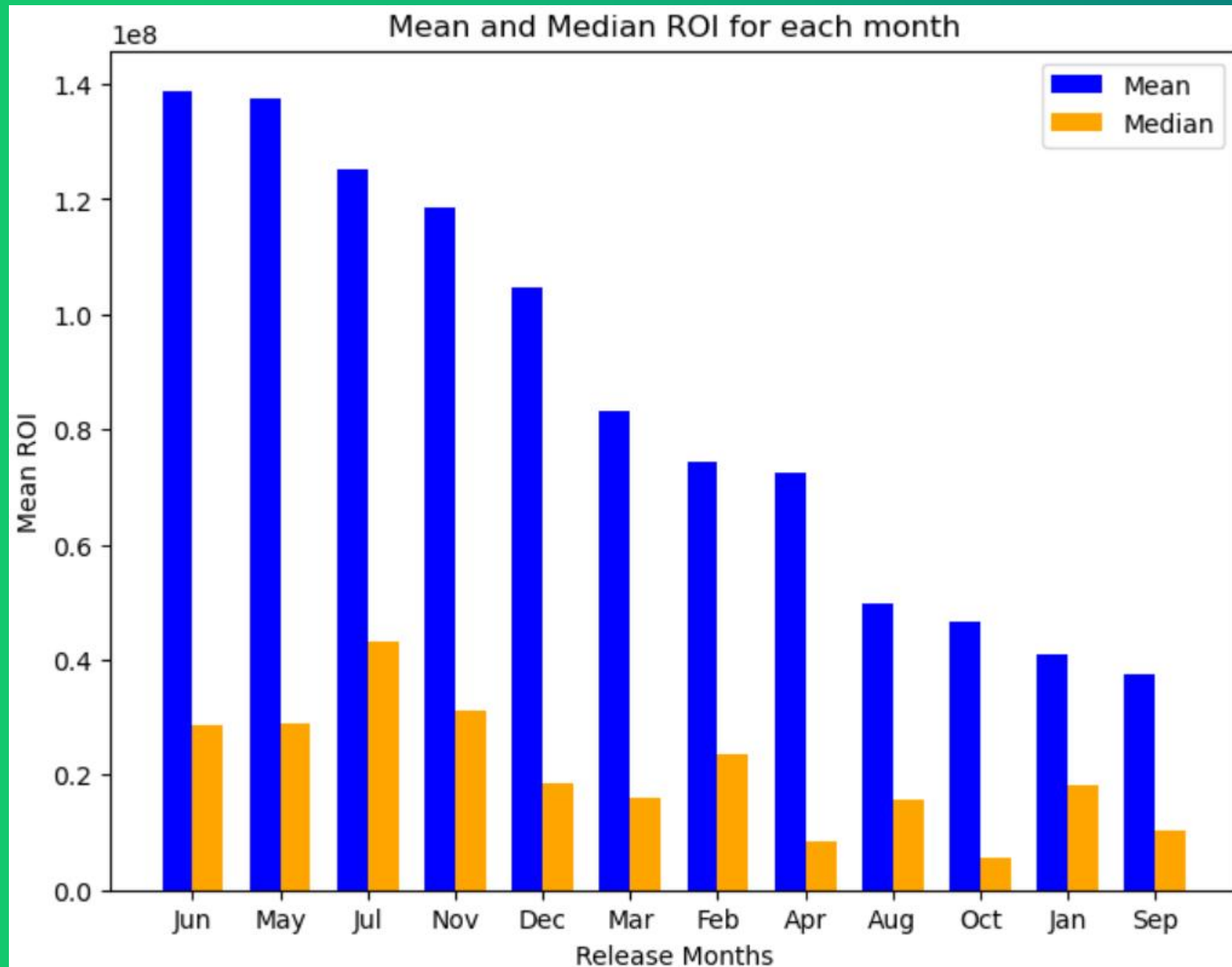
Directors Who are Most Likely to create a film with a High ROI and Positive Rating

- Right of the bat we can seen that the top 5 most common director names in the film industry are: _David Gordon Green, Steven Soderbergh, Steven Spielberg, Ridley Scott and Jon M. Chu_. What's that thing they always say? Practice makes perfect right?

- These money makers in the industry are likely to get the audience exited as they know what direction the film will take and thay have directed films that are above the average rating of 6.27, well except _Jon M. Chu_. He is not to be discared so lightly based of public opinion because he happens to direct films **that generate an impresive High ROI**.

- The directors I would recommend to the company moving forward would be as follows:

1. **Steven Spielberg**
2. **Ridley Scott**
3. **Jon M. Chu**

- These directors are most likely to direct a film that would genreate double the ROI that other directors in 75% quantile could not generate. However, if the matter of rating is to be considered, _Jon M. Chu_ is one to eliminate from the list.

Comparison of Mean and Median Earnings of Movie Directors
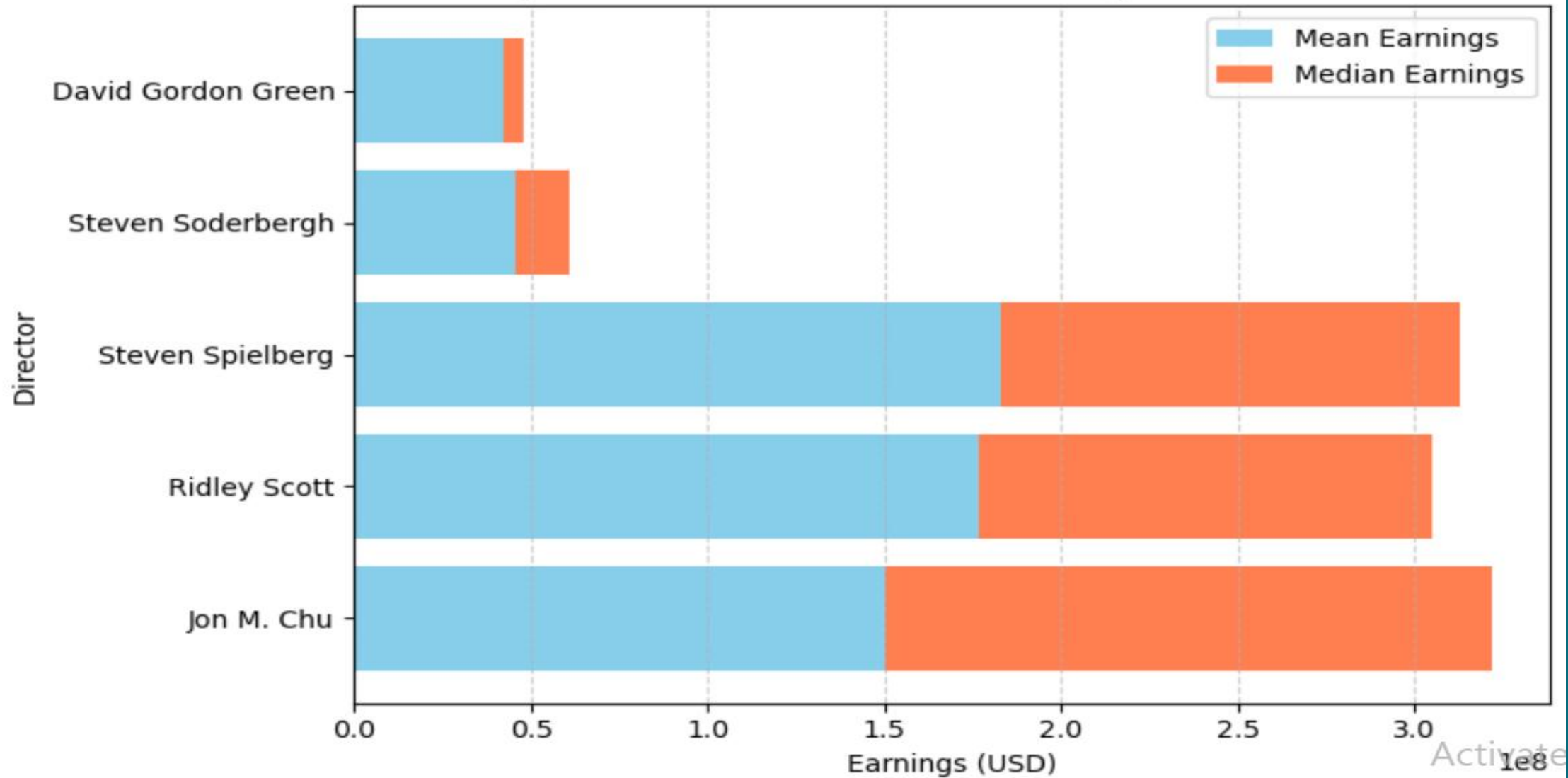
6. Conclusions

This analysis has yielded the following conclusions in order to generate a high ROI:

- **A film of the genre combination 'Action,Adventure,Sci-Fi' should be among the first to be created.** This genre combinationhas proved to frequently yeild a high ROI along with the genre combinations of '**Adventure,Drama,Sci-Fi'** and '**Comedy,Mystery'** as they also have a high profitability in the box office.

- **The films should be released in the months of May, June or July.** It should also be noted that films typically do well in *November* as well. The top three months have shown the highest profitability in this regard.

- **The directors to be contacted to direct the new films should be *Steven Spielberg, Ridley Scott* and *Jon M. Chu.*** Their names are well known in the industry and they audience loves them as their films have a general positive rating. Their skill in scrpting and working the cast is seen in the return they hae generated through a high ROI.

Next Steps

- **Conducting further analysis on specific genres.** The data shows results based on a combination of genres and a further analysis to determine the profitability of these genres and hence influence the film making process. Some things to look out for would be the target audience for these genres and probably their purchasing power. I would also consider what specific elements of this genres enable them to perform better in terms of ROI.

- **Investigating the periods in which movies were released.** The data suggests that most movies are most profitable between the middle of the year but other factors such as relatablilty of the film, current events and trends and audience type could be factors to consider that would influence the ROI.

- **Checking how movie ratings(G to R-rated films) affect the ROI.** This could possibly influence the decision on the type of films to be created based on the target audience.

THANK YOU

Email: mtumainicindy@gmail.com

Github: @TumainiC

LinkedIn: https://www.linkedin.com/in/cindy-tumaini/