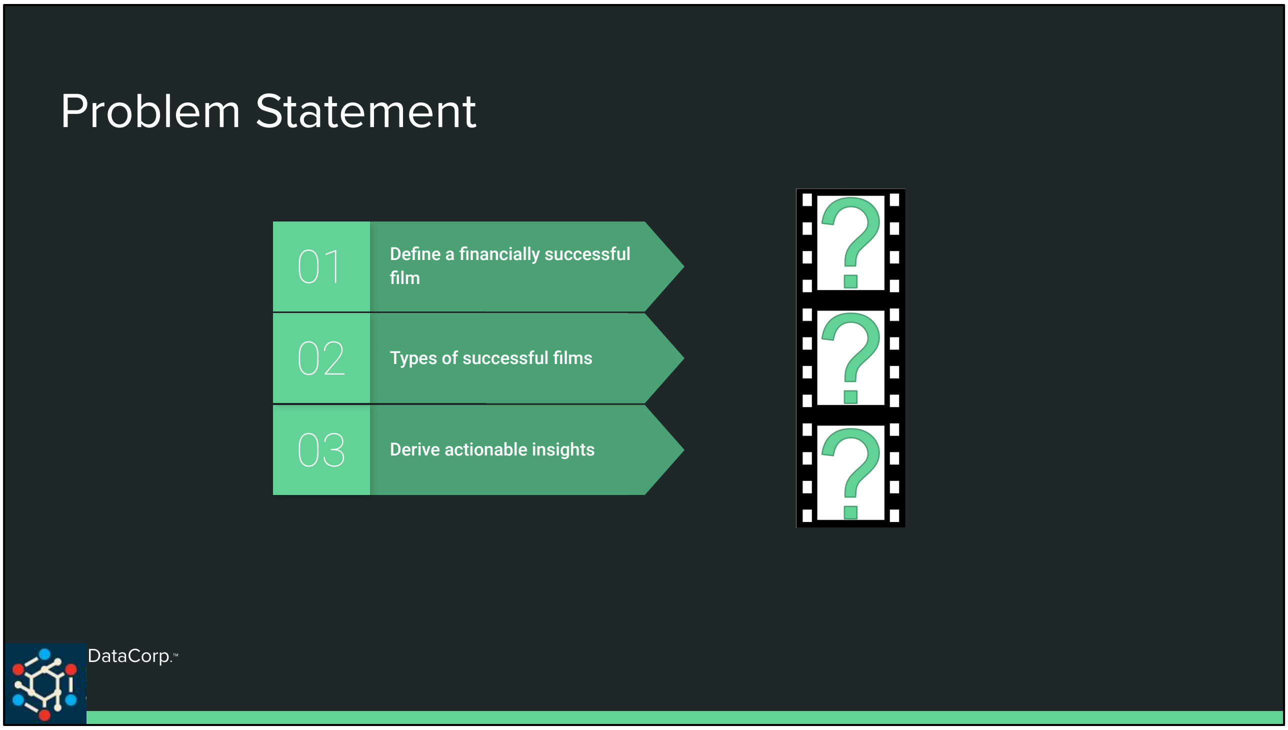
DataCorp.™

Film Industry Analysis & Insights

John Paul Hernandez Alcala

Hello, my name is John Paul, and I am a data scientist with DataCorp. Today my goal is to give you a better understanding of the movie industry. Let’s get started!



Problem Statement

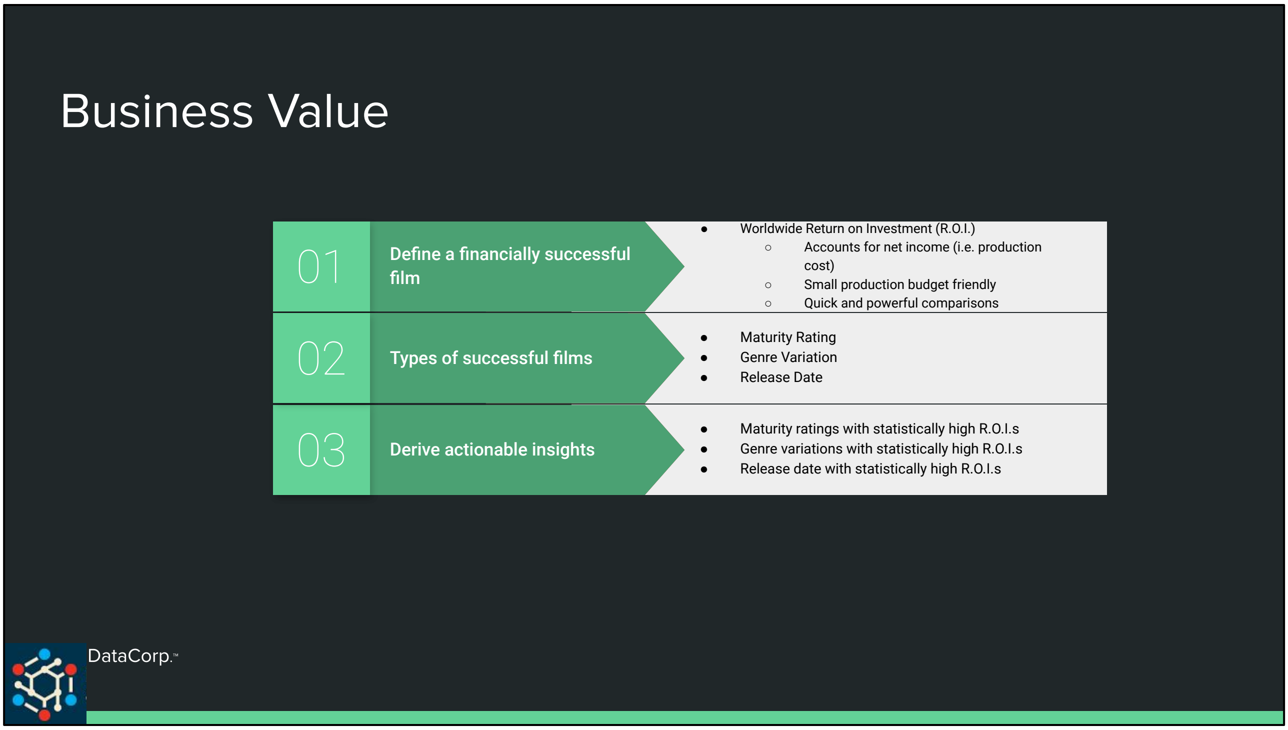
01 Define a financially successful film

02 Types of successful films

03 Derive actionable insights

DataCorp.™

Before we dive into the analysis of the film industry, we have to present ourselves with three questions that make up our problem statement: define a financially successful film, types of successful films, and derive actionable insights.



Business Value

01 Define a financially successful film

- Worldwide Return on Investment (R.O.I.)
  - Accounts for net income (i.e. production cost)
  - Small production budget friendly
  - Quick and powerful comparisons

02 Types of successful films

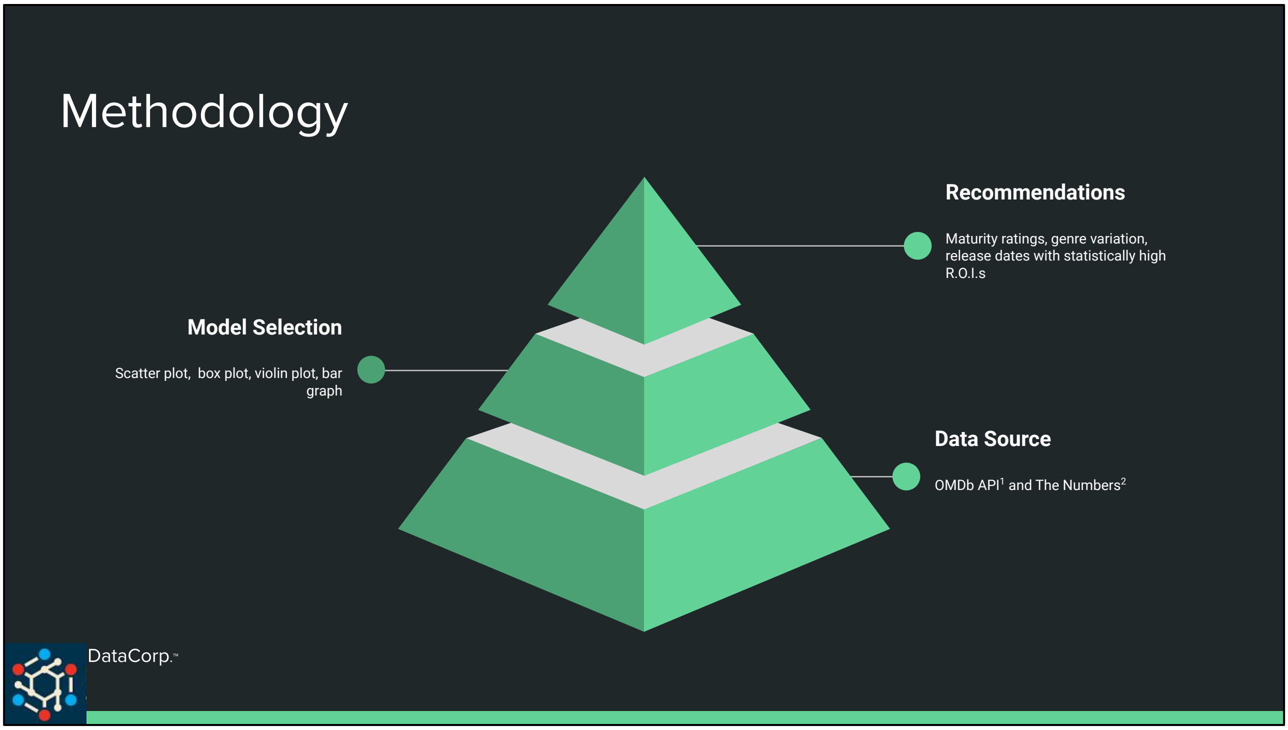
- Maturity Rating
- Genre Variation
- Release Date

03 Derive actionable insights

- Maturity ratings with statistically high R.O.I.s
- Genre variations with statistically high R.O.I.s
- Release date with statistically high R.O.I.s

DataCorp.™

To better answer these questions, we must equip ourselves with the right tools to build towards a viable and reasonable conclusion. For the 1st question, we will use the overall median value to the top 25% Worldwide Return on Investment (R.O.I.) because it takes into account net income, it is friendly to small production budgets, and it allows for quick and powerful comparisons. For the 2nd question, we will limit ourselves to the maturity rating, genre variation, and release date of successful movies. Finally for question 3, we will obtain recommendations for specific maturity ratings, genre variations, and release dates.



Methodology

Recommendations

Maturity ratings, genre variation, release dates with statistically high R.O.I.s

Model Selection

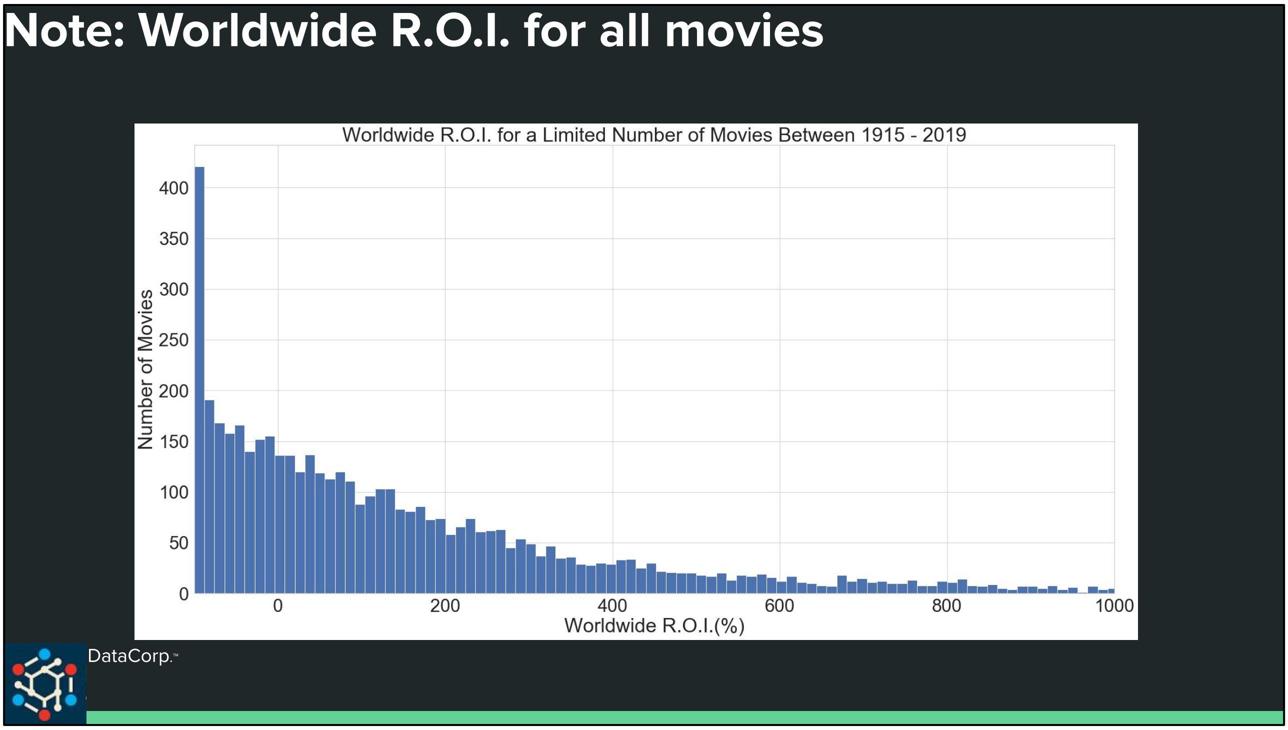
Scatter plot, box plot, violin plot, bar graph

Data Source

OMDb API[1] and The Numbers[2]

DataCorp.™

With the tools in place, we are able to build our foundation off the data we collect from OMDb and The Numbers, work our way up with statistical models such as scatter plots, box plots, violin plots, and bar graphs, and finally arrive to our recommendations.

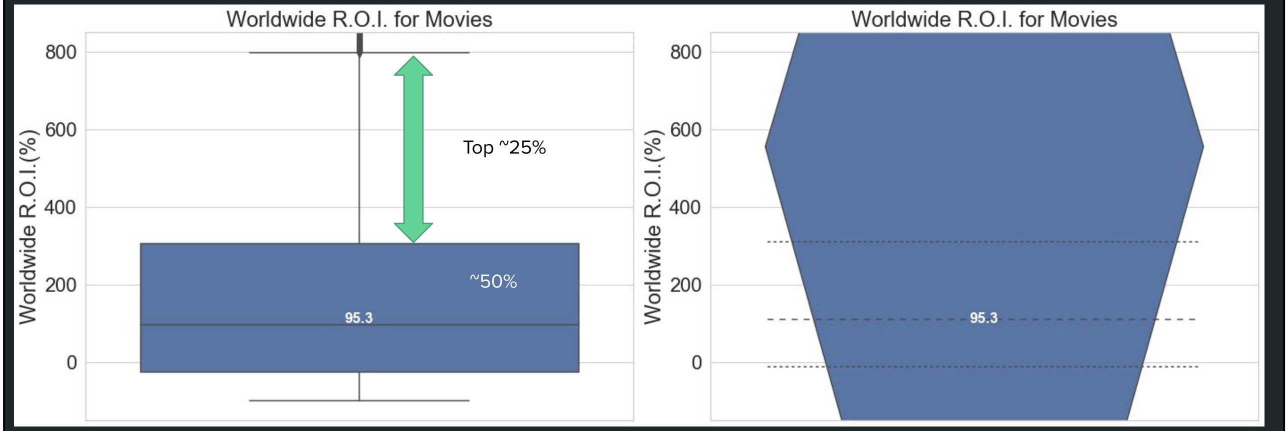


Note: Worldwide R.O.I. for all movies

DataCorp.™

To start our analysis, we need to figure out how Worldwide R.O.I. is distributed, and, from the figure here, we can tell it is not evenly distributed at all.

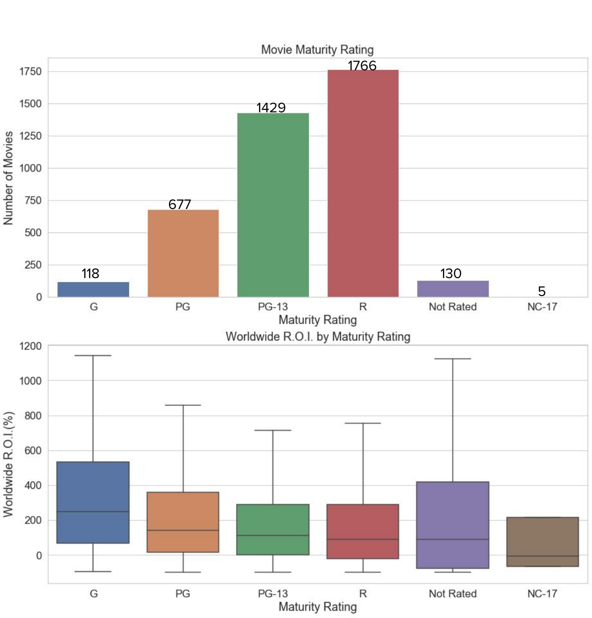
Note: Worldwide R.O.I. for all movies



DataCorp.™

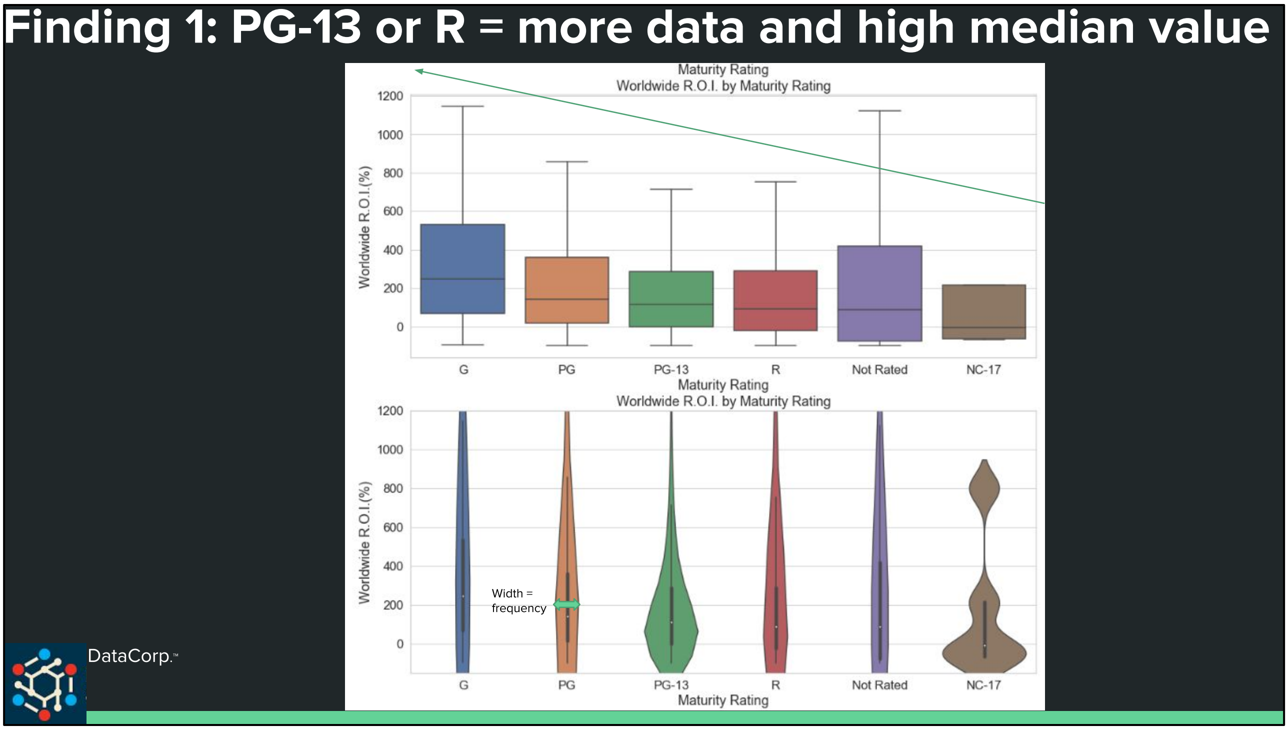
So we must look to quantifying our data with a box plot. With this plot, we can establish R.O.I. values that will be used for determining a statistically reasonable selection. The 95.3% is the median value of the data. The 50% signifies that 50% of our data resides in the box which starts slightly below zero and around 350%. Again the goal is to shoot for the median value to top 25% which is between 95.3% and 800%

Finding 1:



DataCorp.™

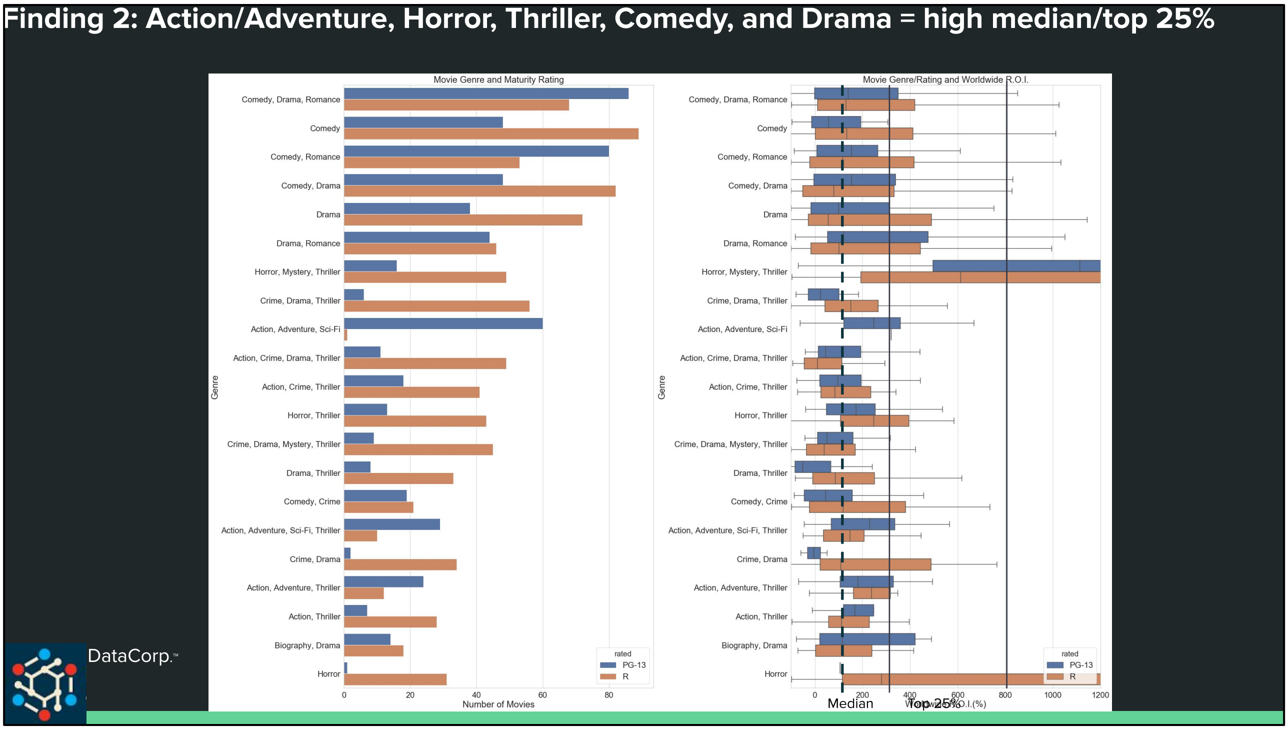
From our analysis of maturity ratings, 'PG-13' and 'R' has the most data, and 'G' rated has the highest median R.O.I. value and 75% of its movies have a R.O.I. value above ~75% followed by 'PG', 'PG-13', 'R'-tied-'Not Rated', and 'NC-17'; however, 'R' rated movies are made the most followed by 'PG-13', 'PG', 'Not Rated', 'G', and 'NC-17'



Finding 1: PG-13 or R = more data and high median value

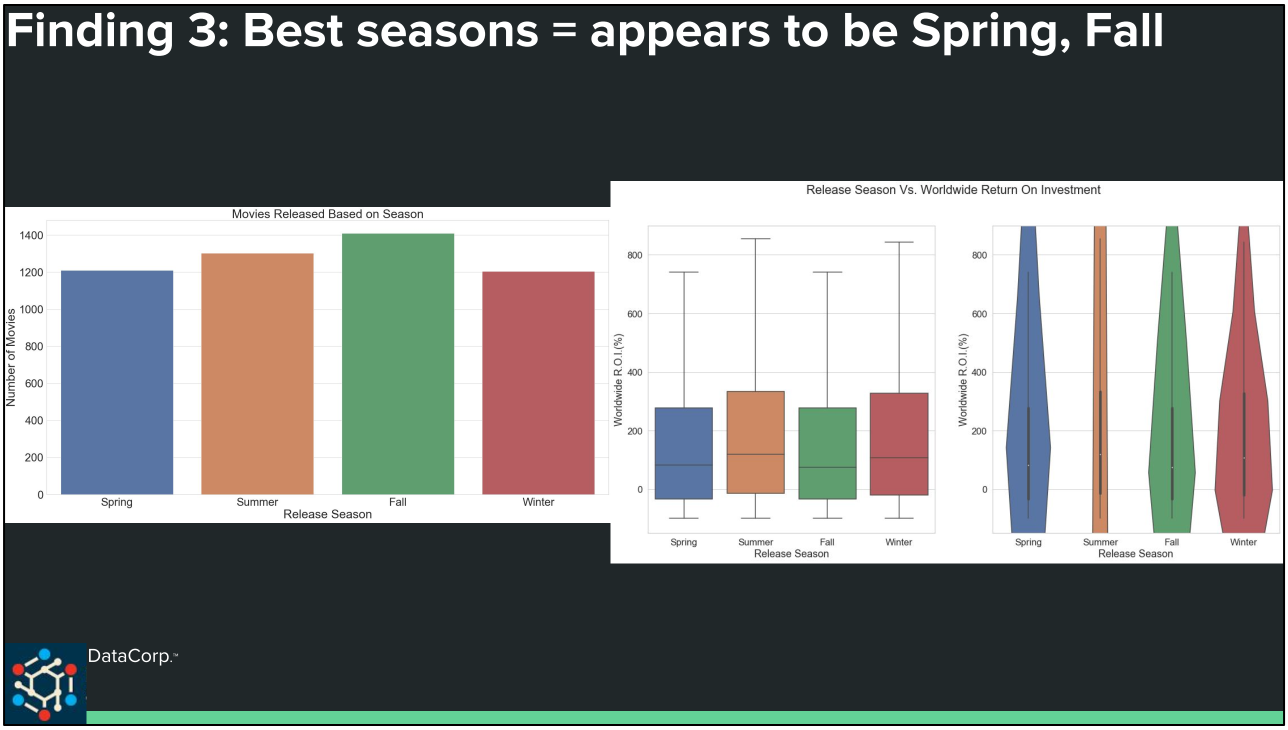
DataCorp.™

On this next set of graphs, we observe that rated 'PG-13' and 'R' movies have more R.O.I.'s values around their respective median value which suggest predictability. Although it would seem logical to select ‘G’ or ‘PG’, the amount of movies released for these two ratings only accounts for ~20% of the data; therefore, it is recommendable to design a movie with 'PG-13' or 'R' rating because of its predictability and amount of movies released with these ratings. This means Microsoft can expect an R.O.I. value around 100% - 150% for movies rated at ‘PG-13’ or ‘R’ which is within our goal.



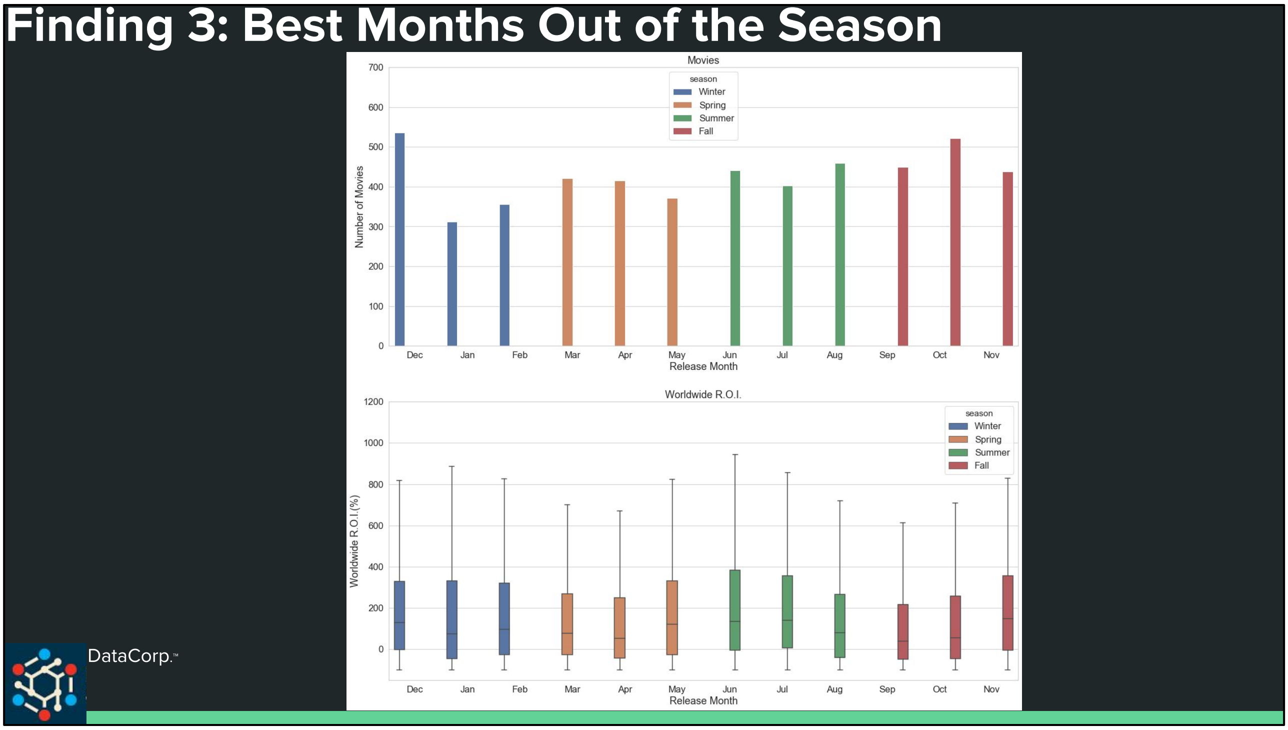
Finding 2: Action/Adventure, Horror, Thriller, Comedy, and Drama = high median/top 25%

We now take a look at genre, and we immediately notice genre for a movie is rarely limited to only one. Therefore, genre variations were analyzed instead. Of the genre variations, common genres that have more than the median R.O.I. movie value of 95.3% are Action/Adventure, Horror, Thriller, Comedy, and Drama. It is recommendable to design a movie with Action/Adventure, Horror, Thriller, Comedy, and Drama features in it because these genres have more than the median R.O.I. movie value of 95.3% or reach into the top 25%. This means Microsoft can expect an R.O.I. value around 95.3% - 350% for movies with Action/Adventure, Horror, Thriller, Comedy, and Drama features which is within our goal.



Finding 3: Best seasons = appears to be Spring, Fall

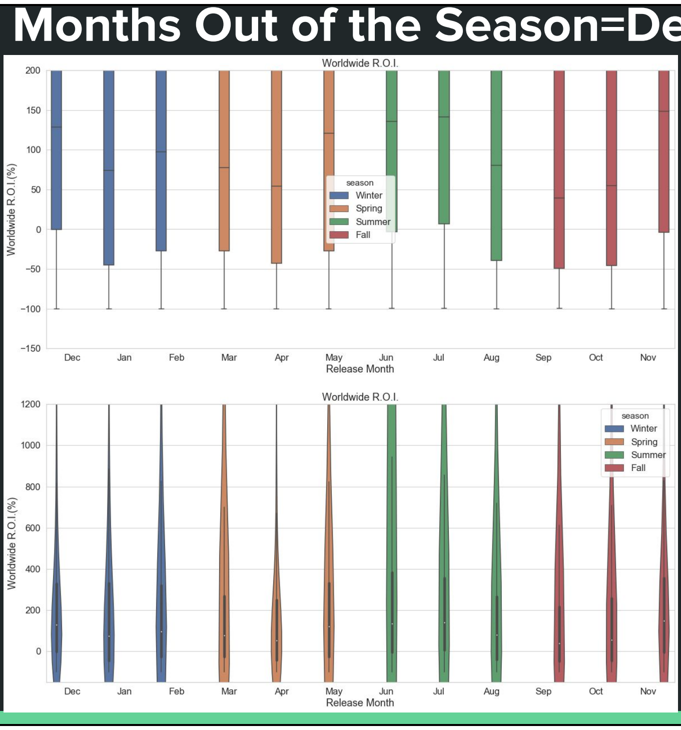
Moving next to seasons, we see that all seasons release about the same number of movies; Thus, R.O.I. distribution could be used for comparison. The season with highest median R.O.I. value is Summer followed by Winter, spring, and Fall. Although Summer and Winter have high median R.O.I. values, the frequency of movies that achieved around their respective median R.O.I. value is less certain than with the movies from Spring and Fall; in other words, movies released in Spring and Fall are more likely to have a R.O.I. that is close to the median R.O.I. value for Spring and Fall. This means Microsoft can expect an R.O.I. value ~95% - ~150% for movies released in Spring and Fall which is within our goal.



Finding 3: Best Months Out of the Season

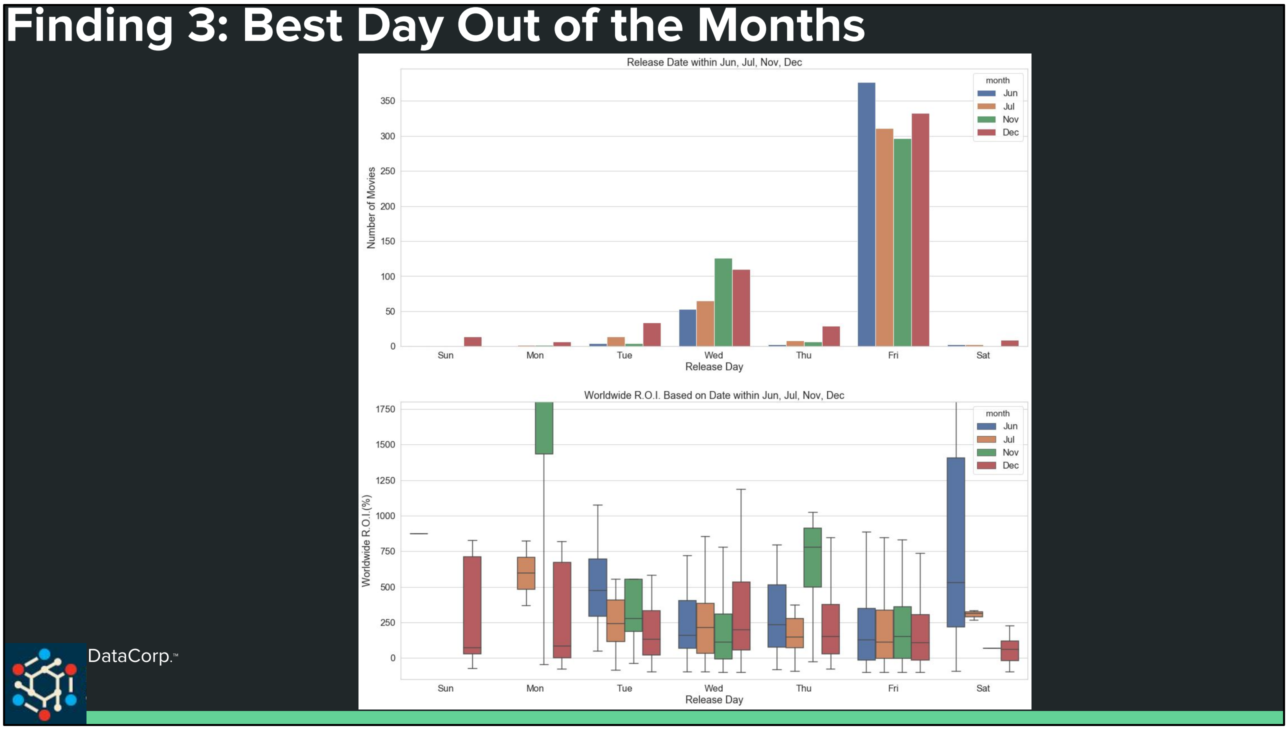
We push deeper into seasons by breaking down into months. The months release roughly about the same number of movies; therefore, just like seasons, the R.O.I. distribution could be used for comparison. The month with the highest median R.O.I. value is Nov, followed by Jul, Jun, Dec, May, Feb, Aug, Mar, Jan, Oct, April, and Sep.

Finding 3: Best Months Out of the Season=Dec, Nov



DataCorp.™

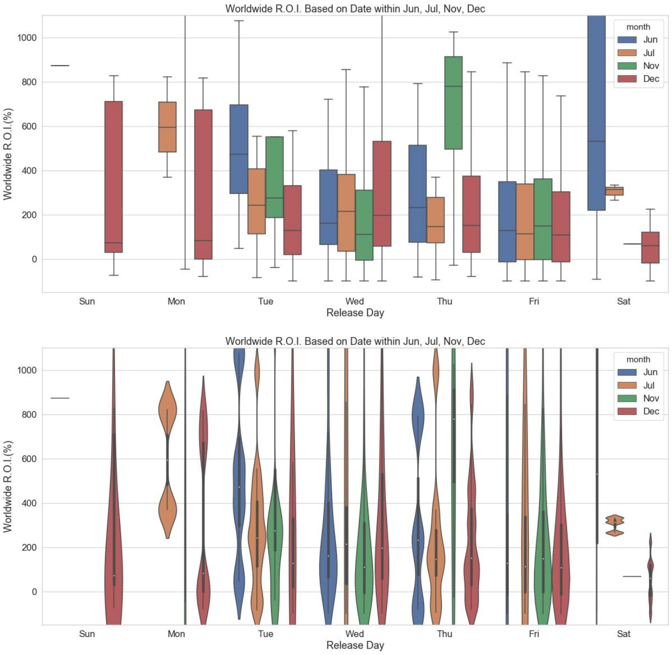
Additionally, we found that Nov and Dec not only have high median R.O.I.s, but also the most reliable R.O.I. distributions because they are close to their respective median R.O.I. values. This means Microsoft can expect an R.O.I. value ~150% - ~180% for movies released in Nov and Dec which is within our goal.



Finding 3: Best Day Out of the Months

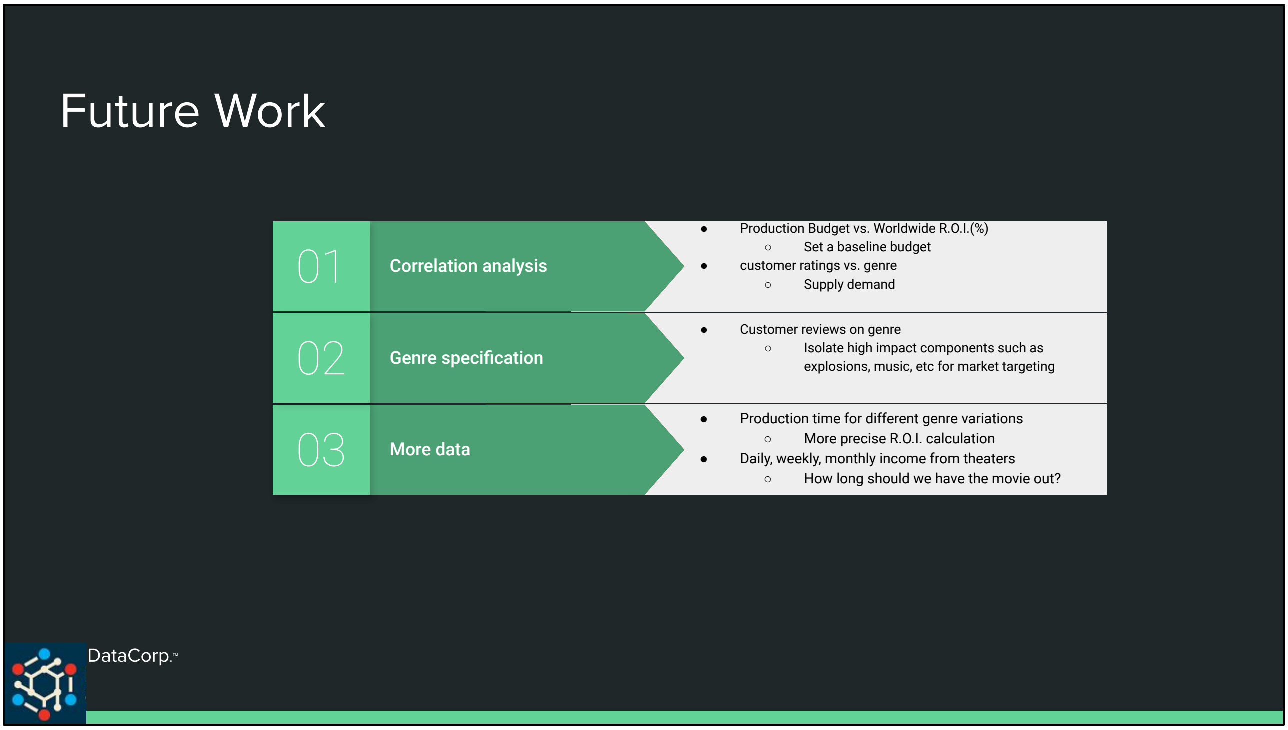
We analyze the days more closely in Jun, Jul, Nov, and Dec. We immediately notice that movies are predominantly released on Wed and Fri.

Finding 3: Best Day Out of the Months = Wed or Fri



DataCorp.™

From the previous graphs and these graphs, we recommend that the best days to release a movie are on Wednesday or Friday because of high R.O.I. median values coupled with reliable R.O.I distributions that accumulate around their respective R.O.I. median values. This means Microsoft can expect an R.O.I. value ~150% - ~180% for movies released in Wed and Fri which is within our goal.



Future Work

01 Correlation analysis

- Production Budget vs. Worldwide R.O.I.(%)
    - Set a baseline budget
- customer ratings vs. genre
    - Supply demand

02 Genre specification

- Customer reviews on genre
    - Isolate high impact components such as explosions, music, etc for market targeting

03 More data

- Production time for different genre variations
    - More precise R.O.I. calculation
- Daily, weekly, monthly income from theaters
    - How long should we have the movie out?

DataCorp.™

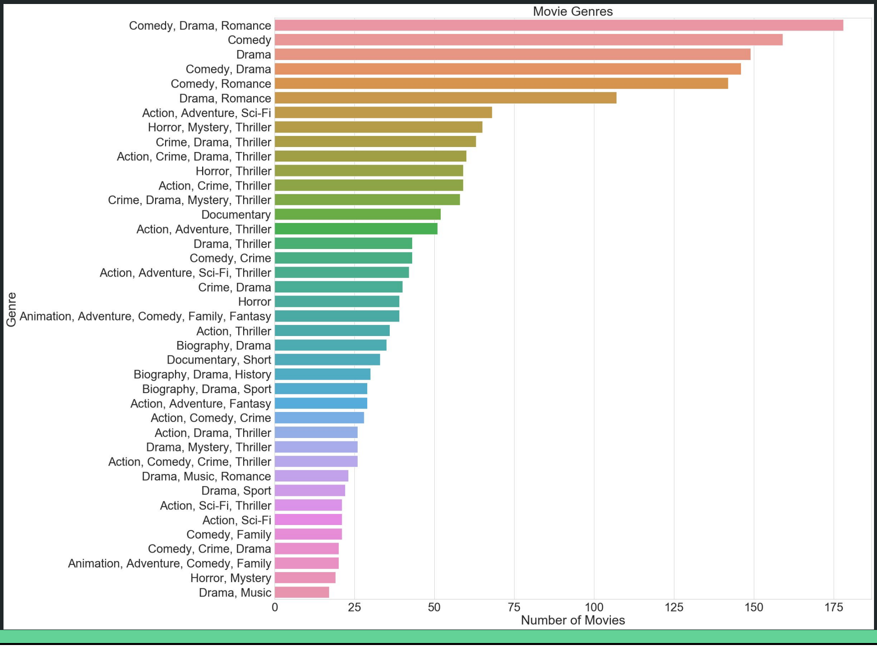
Now that we have heard all these recommendations, what now? We can do correlation analysis on production budget vs. worldwide R.O.I. (%), so we can establish a baseline budget for financial resource allocation. We can also do a correlation analysis on customer ratings vs. genre to address customer demands. Additionally, we can look at what specifics from a genre customers are looking for explosions, music, etc for market targeting. Finally, with more data, we can look at production time for different genre variations to make our R.O.I. calculation more accurate with net profit, and we can look at daily, weekly, monthly income from theaters so we can have a better idea when to pull a movie from the theater before suffering financial loss.

Thank You

DataCorp.™

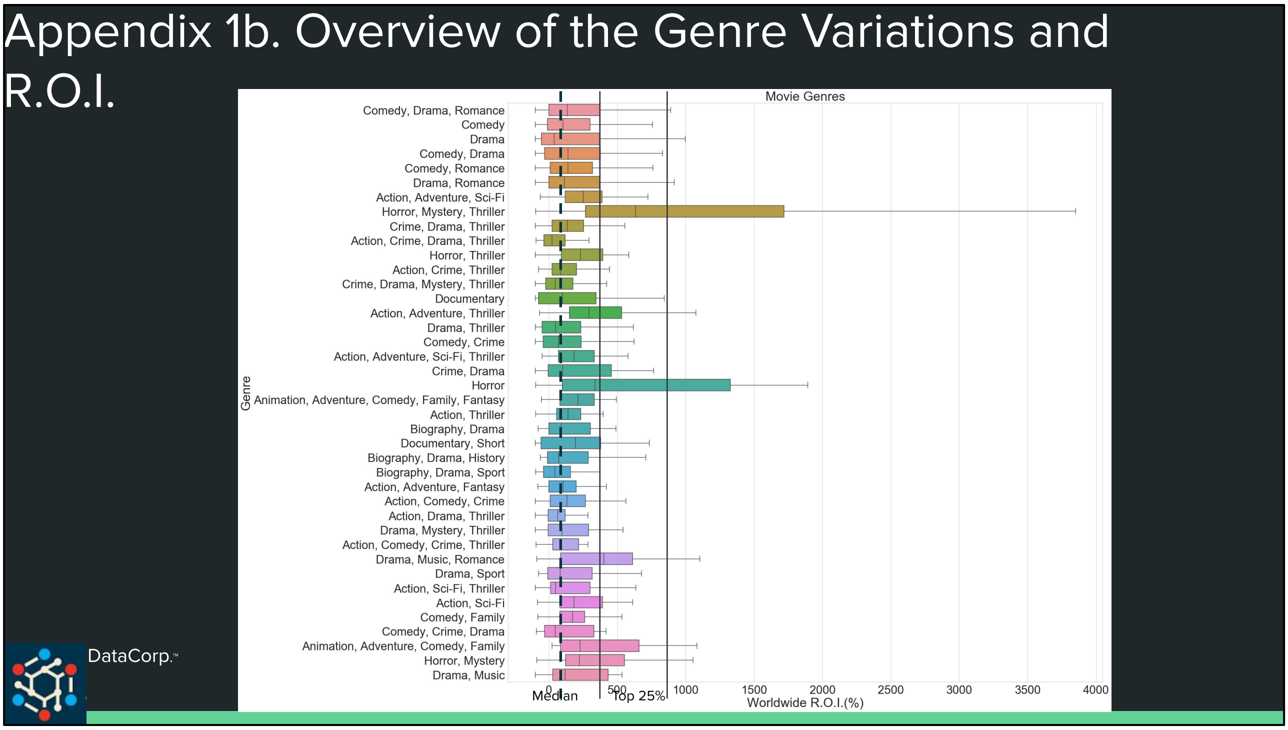
Thank you for your time! Please, feel free to ask me any questions at this time.

Appendix 1a. Overview of the Genre Variations



DataCorp.™

Here are the top 40 genres not constrained by maturity rating.



Appendix 1b. Overview of the Genre Variations and R.O.I.

Here are the top 40 genres not constrained by maturity rating. The only genre with a median value in the top 25% is horror/mystery/thriller