MOVIE INDUSTRY DATA REPORT

TEAM: Jupiter III

1. <u>Business Understanding</u>

Business Overview

As of 2018, 54 percent of U.S.-based respondents stated that their favorite movie watching location was at home. Only around 13 percent of respondents stated that they preferred to watch movies in a theater, while 22 percent stated that they like the theater and home options equally.

Movies have long been a staple of the entertainment industry in both the United States and across the globe both for streaming services and movie theaters. According to statista, Netflix generated total revenue of nearly 7.7 billion U.S. dollars, up from about 6.64 billion in the corresponding quarter of 2020. Meaning that more and more people are turning to streaming services.

Business Objective

In this data analysis project, we aim to discover if the movie industry is in decline because of the emergence of netflix and other online streaming services.

We are trying to find out with the addition and improvement of streaming services, is the movie industry on the decline or is it stronger than ever.

Business Success Criteria

- Identify the genre with the highest income to budget ratio
- Identify the country with highest movie production

Assessing the situation

1. **Data Inventory**

For this analysis and research we are going to use the movies.csv

2. **Assumptions**

The data provided is correct and up to date

3. Constraints

The dataset only provides an accumulation of data from the year 1980-2020

Data Mining Goals

1. To analyze genre makes more money and is highly watched
2. To determine which genre has the highest income to budget ratio
3. To determine which company has the best movie investments
4. To determine which star's and director's movies have the highest possibility of income success based on previous successes
5. To determine which year was a good movie year and why
6. To determine highest grossing film and get its attributes
7. To determine the country of origin that produces the movies with the highest gross
8. Relationship between the user rating and the income of a movie.

2. Data Understanding

Data Understanding overview

For this data analysis project, we are using data from kaggle and can be acquired here [here]. This datasets is:

★ **Movie Industry -** This dataset provides information on movies found from imdb together with their attributes such as year released.

Data Description

This data set is named Movie Industry and it has 10 columns and 7668 rows in the dataset. Column names and definition:

- **budget**: the budget of a movie. Some movies don't have this, so it appears as 0
- **company**: the production company
- **country**: country of origin
- **director**: the director
- **genre**: main genre of the movie.
- **gross**: revenue of the movie
- **name**: name of the movie
- **rating**: rating of the movie (R, PG, etc.)
- **released**: release date (YYYY-MM-DD)

- **runtime**: duration of the movie
- **score**: IMDb user rating
- **votes**: number of user votes
- **star**: main actor/actress
- **writer**: writer of the movie
- **year**: year of release

Verifying Data Quality

For verifying data , we checked for null/missing values in the dataset and this was the outcome:

>**rating:** 1.004173,
>**released:** 0.026082
>**score:** 0.039124
>**votes:** 0.039124
>**writer:** 0.039124
>**star:** 0.013041
>**country:** 0.039124
>**budget:** 28.312467
>**gross:** 2.464789
>**company:** 0.22170
>**Runtime:** 0.052165

3. Data Preparation

Loading Data

1. Importing the necessary libraries for this analysis. (Pandas, Numpy, Matptolib, seaborn)
2. Computing the data description in rows and columns dimensions
3. Inspecting the data types in the various columns.

>**name**   object
>**rating**   object
>**genre**   object
>**year**   int64
>**released**   object
>**score**   float64

votes   float64

director   object

writer   object

star   object

country   object

budget   float64

gross   float64

company   object

runtime   float64

Cleaning data

1. Filtering out the null values in the budget column
2. Filling in the missing values in the rating , country and company columns respectively- the columns key to our analysis
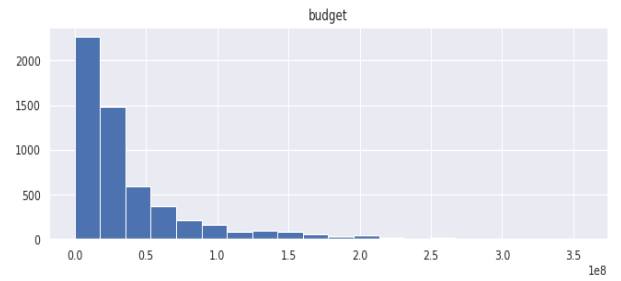3. Checking for duplicates in the data
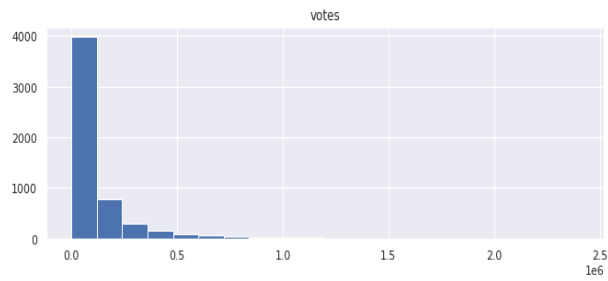4. Inserting space in between the columns

4. Analysis

1. Computing the gross revenue range of the movies

Gross Numeric Attributes
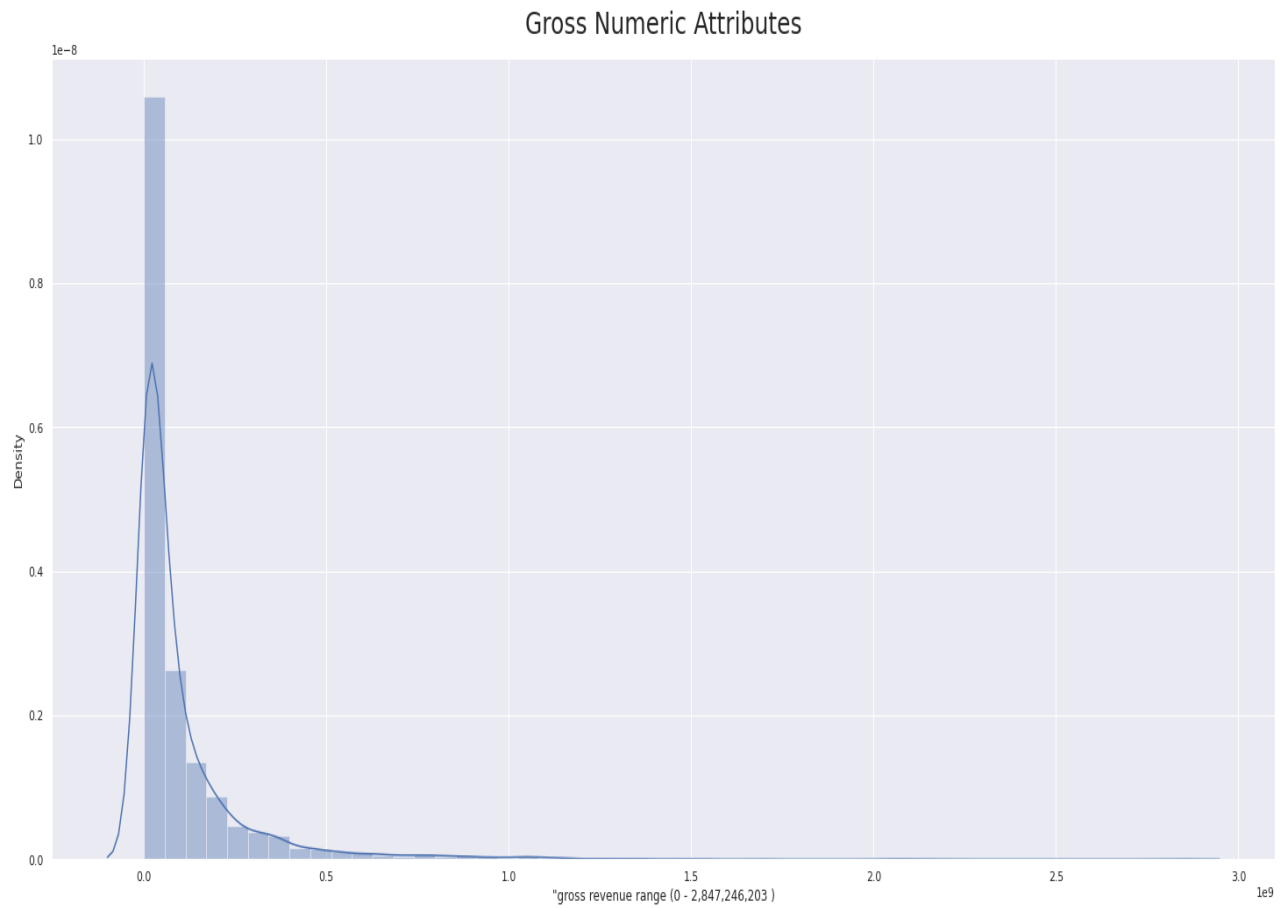
- Revenue distribution across our dataset. We can observe that it's skewed to the left.

2. Distribution across all numerical attributes

Distribution Across all Numerical Attributes
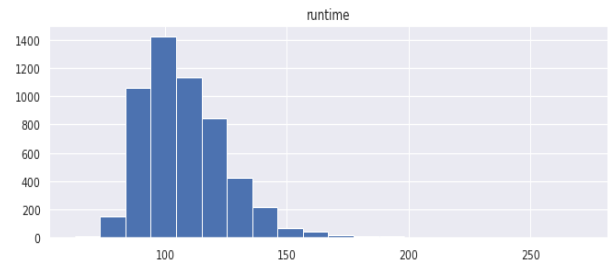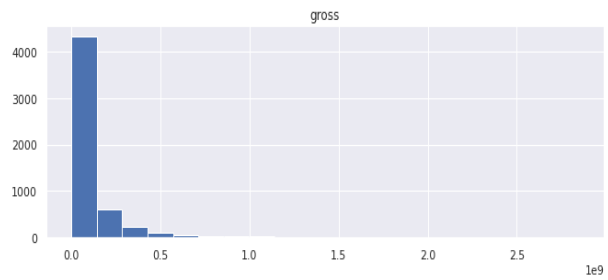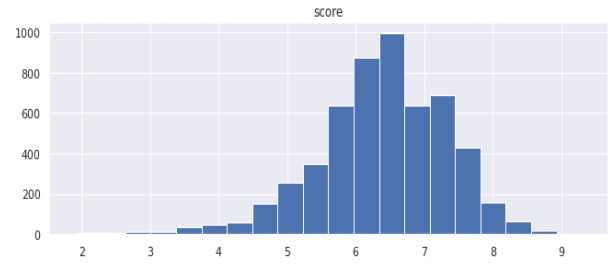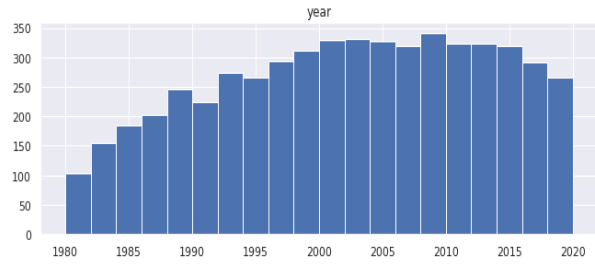
3. Checking for outliers in the dataset

Outliers in Genres
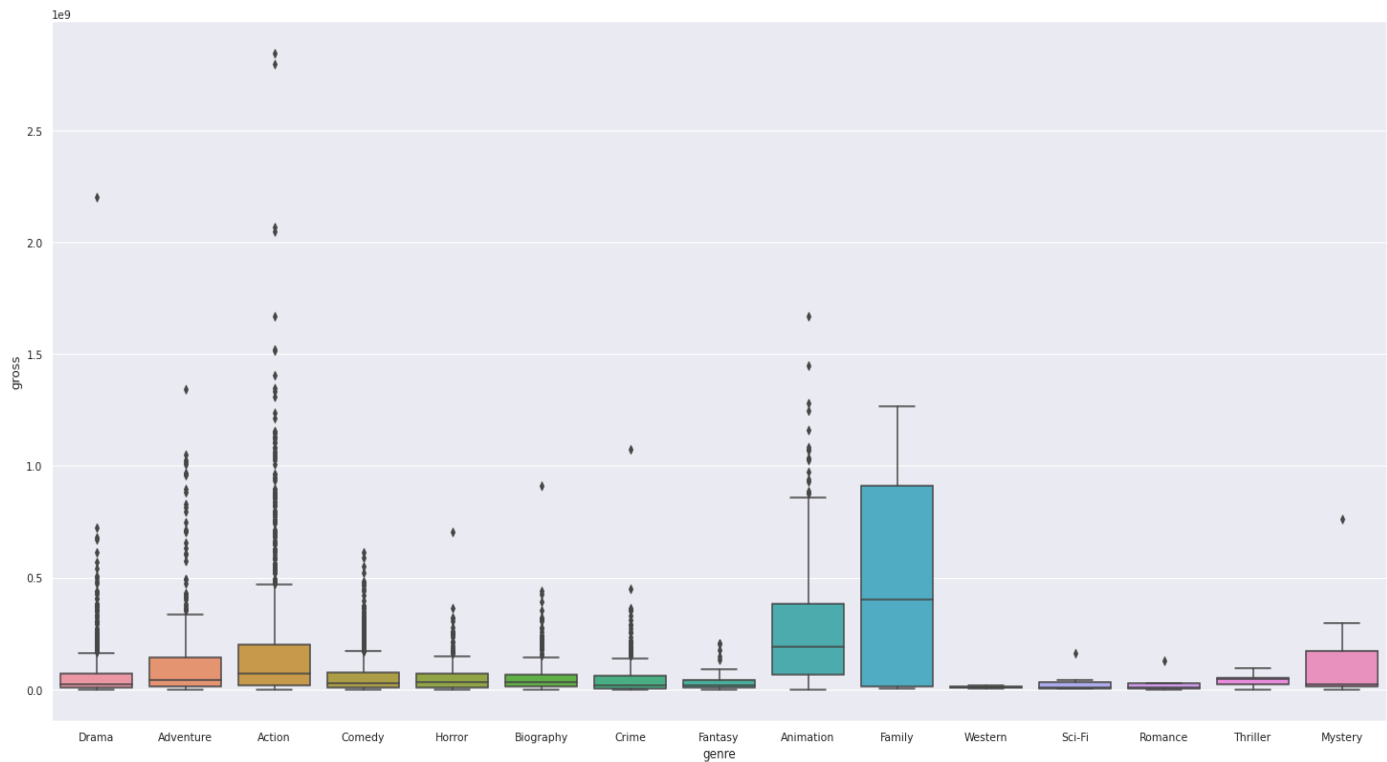
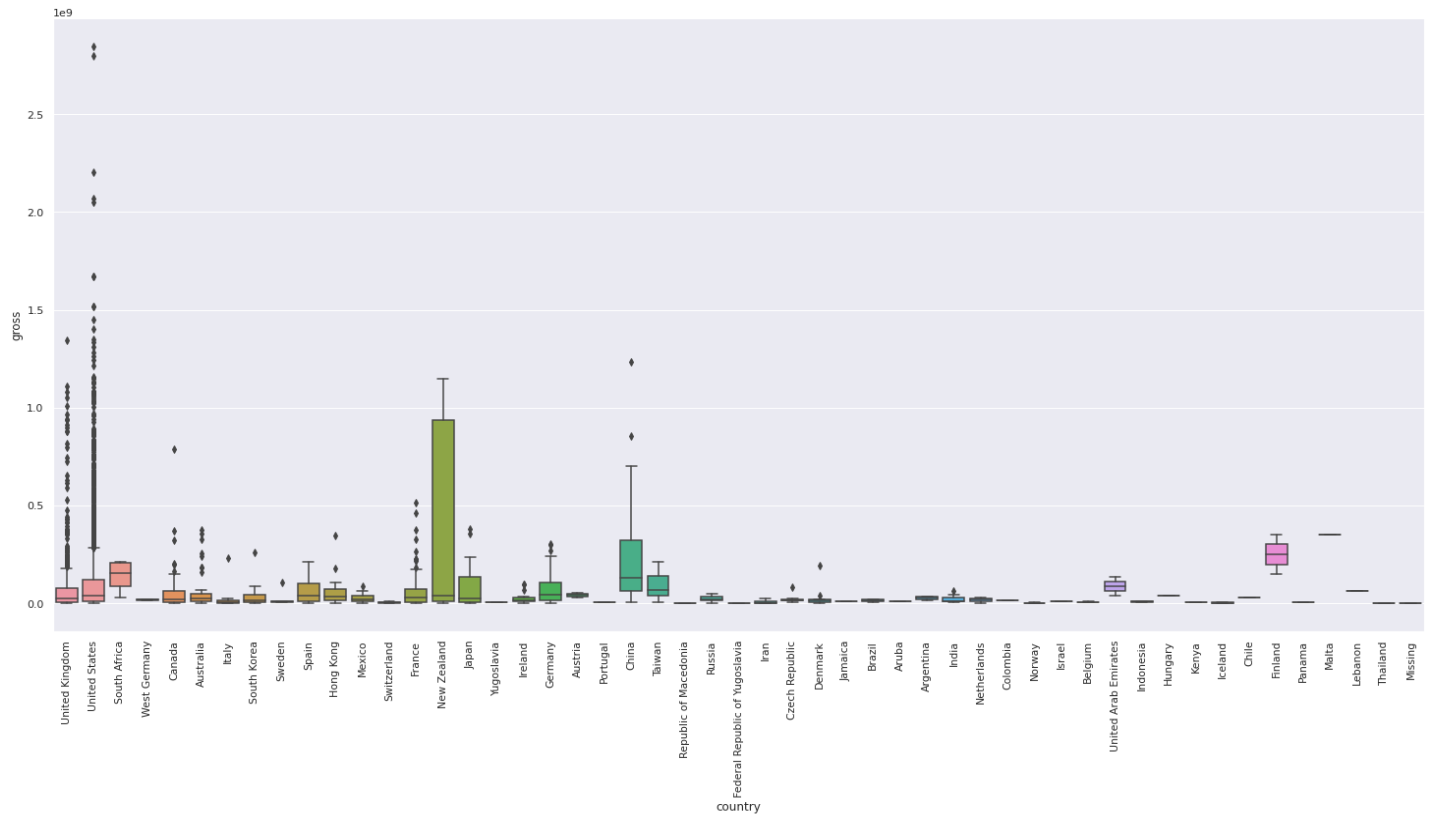- Action genre has the most outliers
- Western Genre has the least outliers
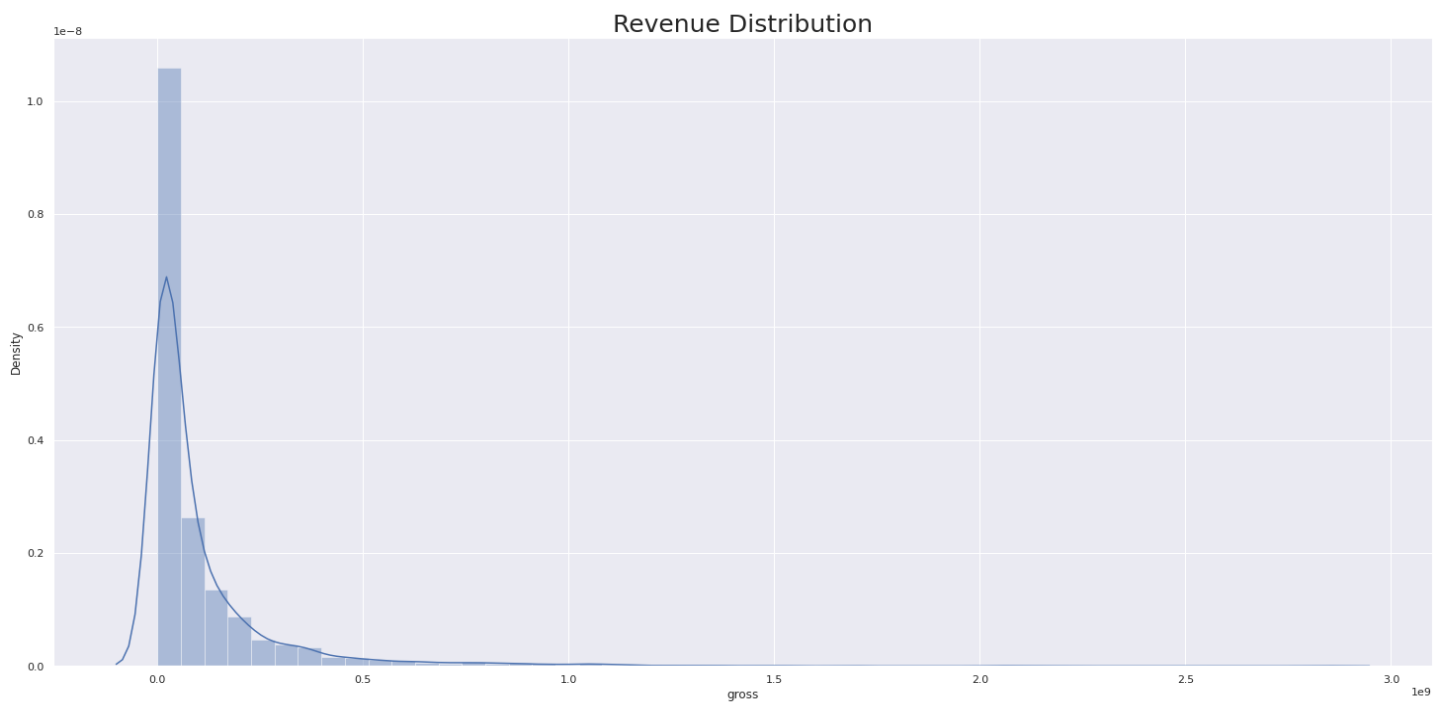
5. Outliers in rating

Outliers in Rating
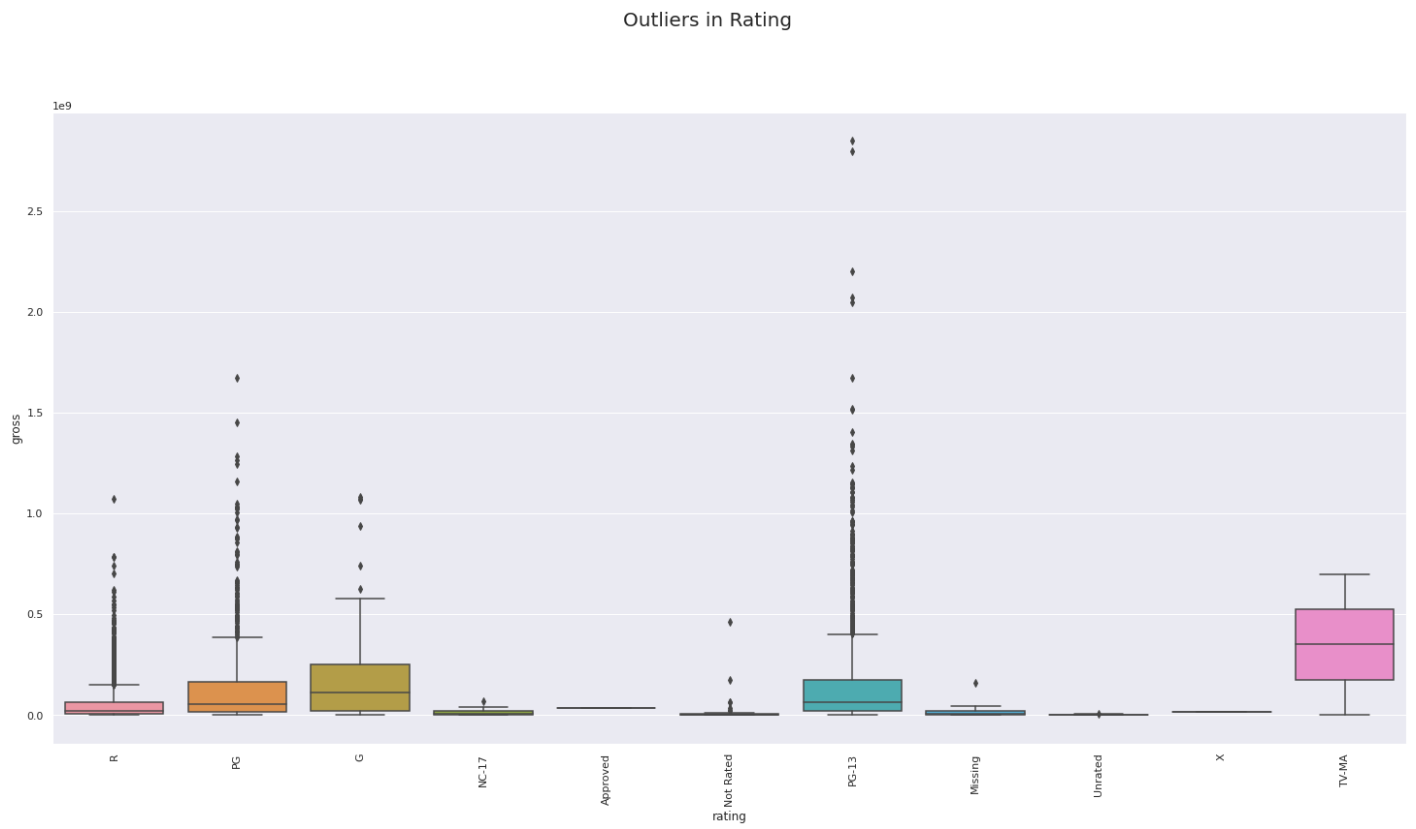
- PG-13 has most outliers

6. Outliers in country

Outliers in Country

● United states has the most outliers followed by the UK

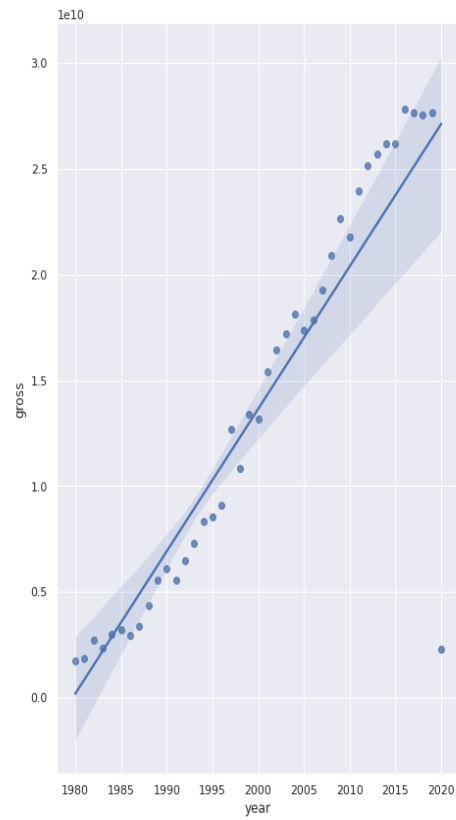## 7. Computation of Univariate analysis

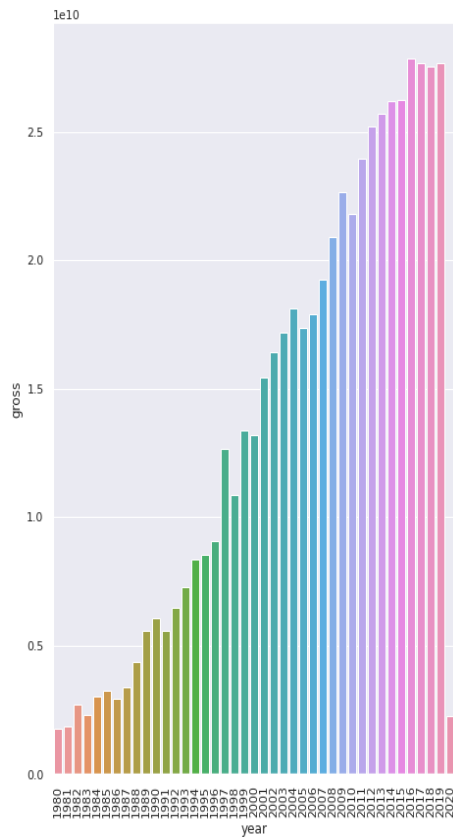### I. Revenue distribution

Revenue Distribution

II. Kernel Distribution for genres

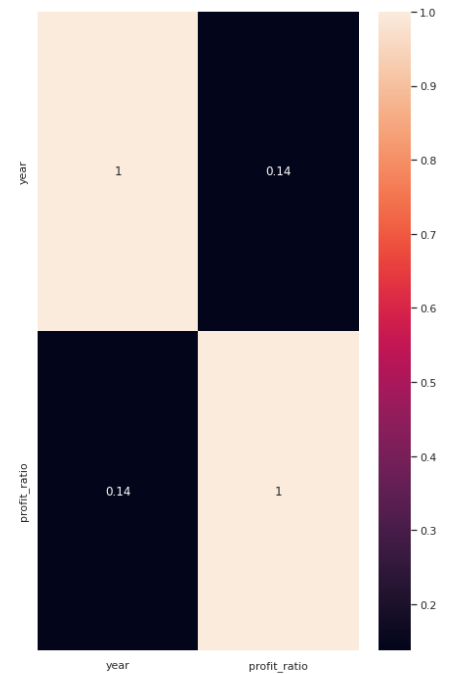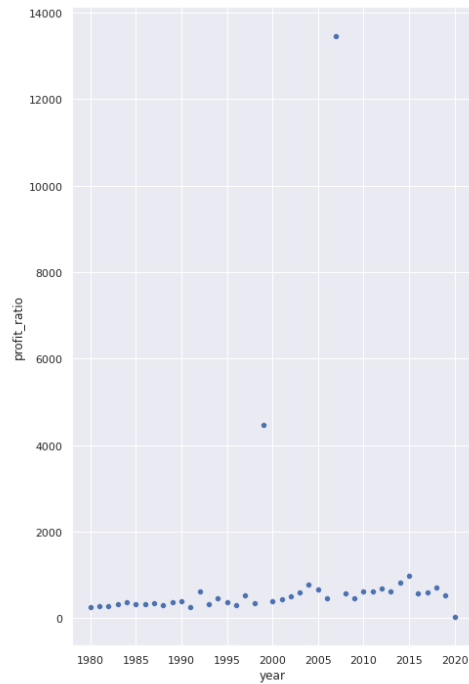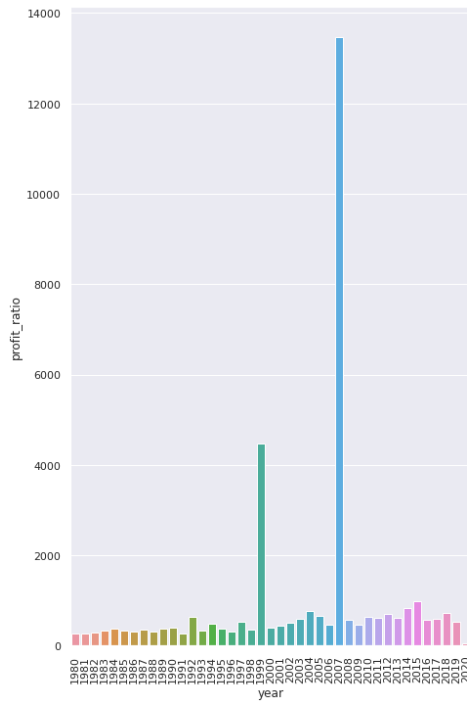Kernel Distribution curve for Genres
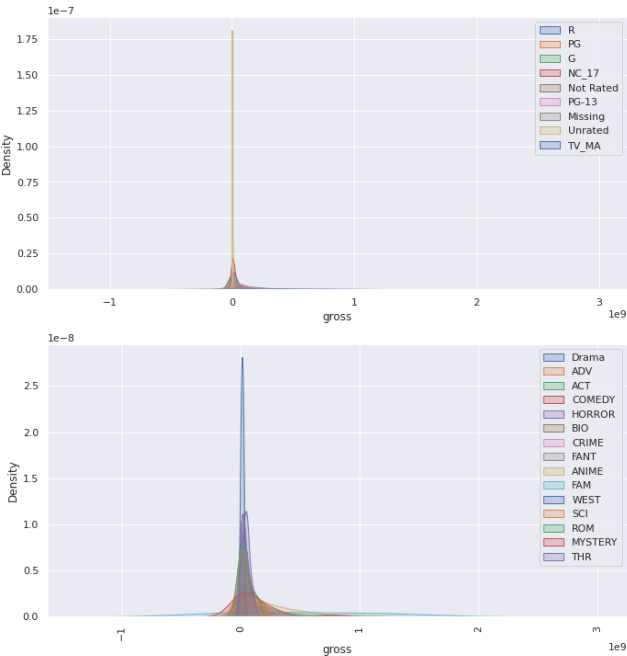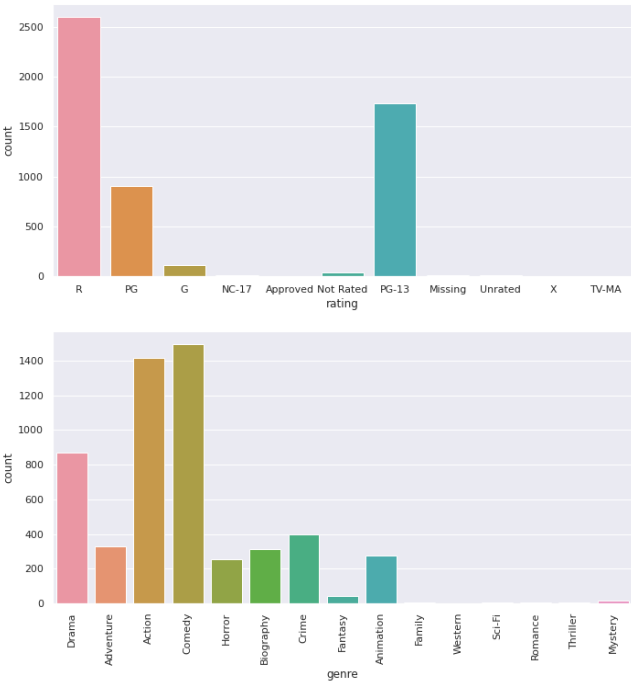
8. Bivariate Analysis
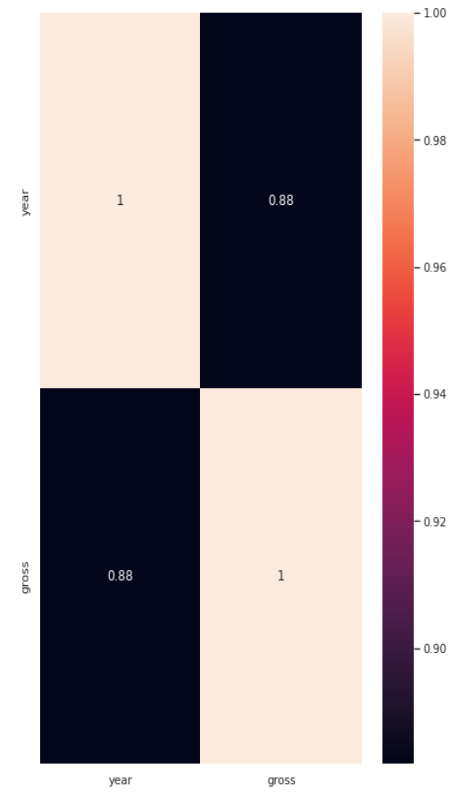
- The amount of gross revenue increases over time, strong positive correlation between gross revenue and years

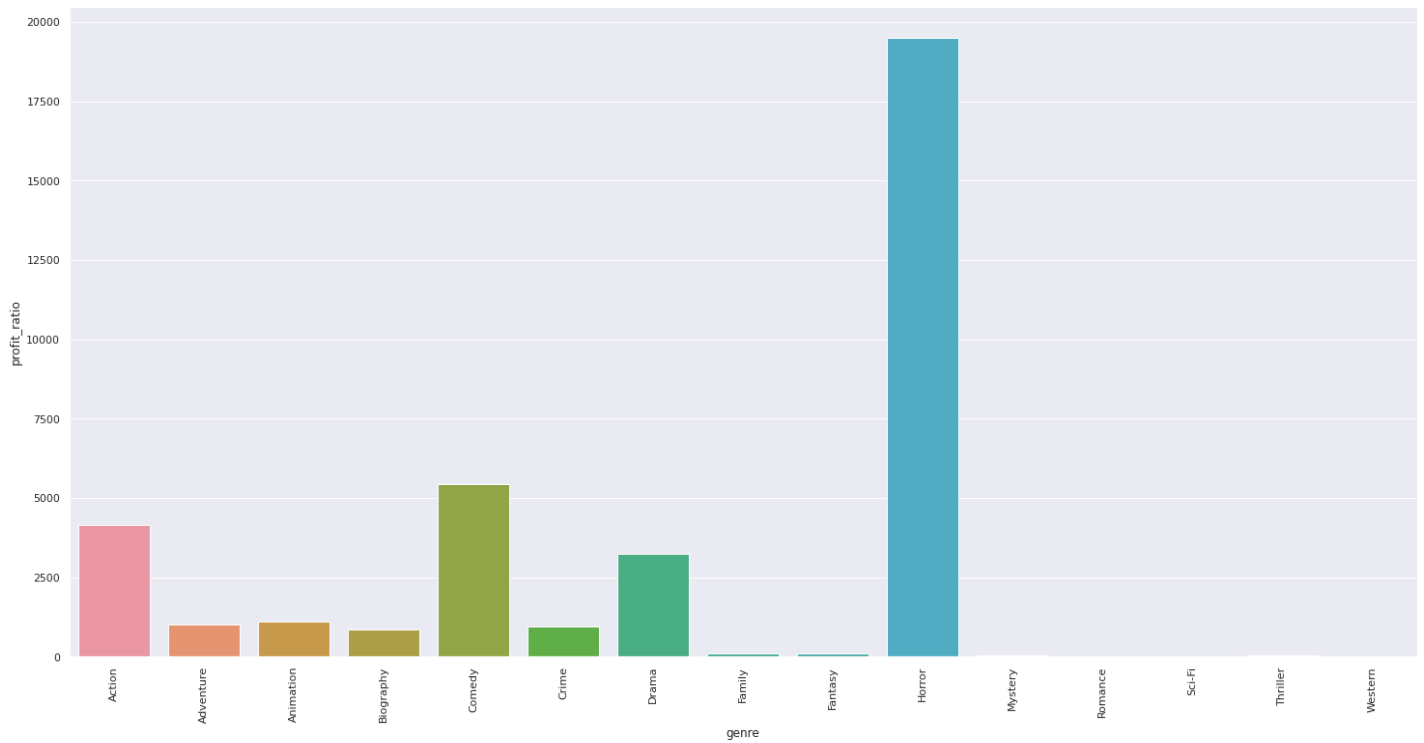## 9. Analysis of the revenue of the movie industry over the years

Revenue vs Year

- The years 2007 and 1997 were the most profitable
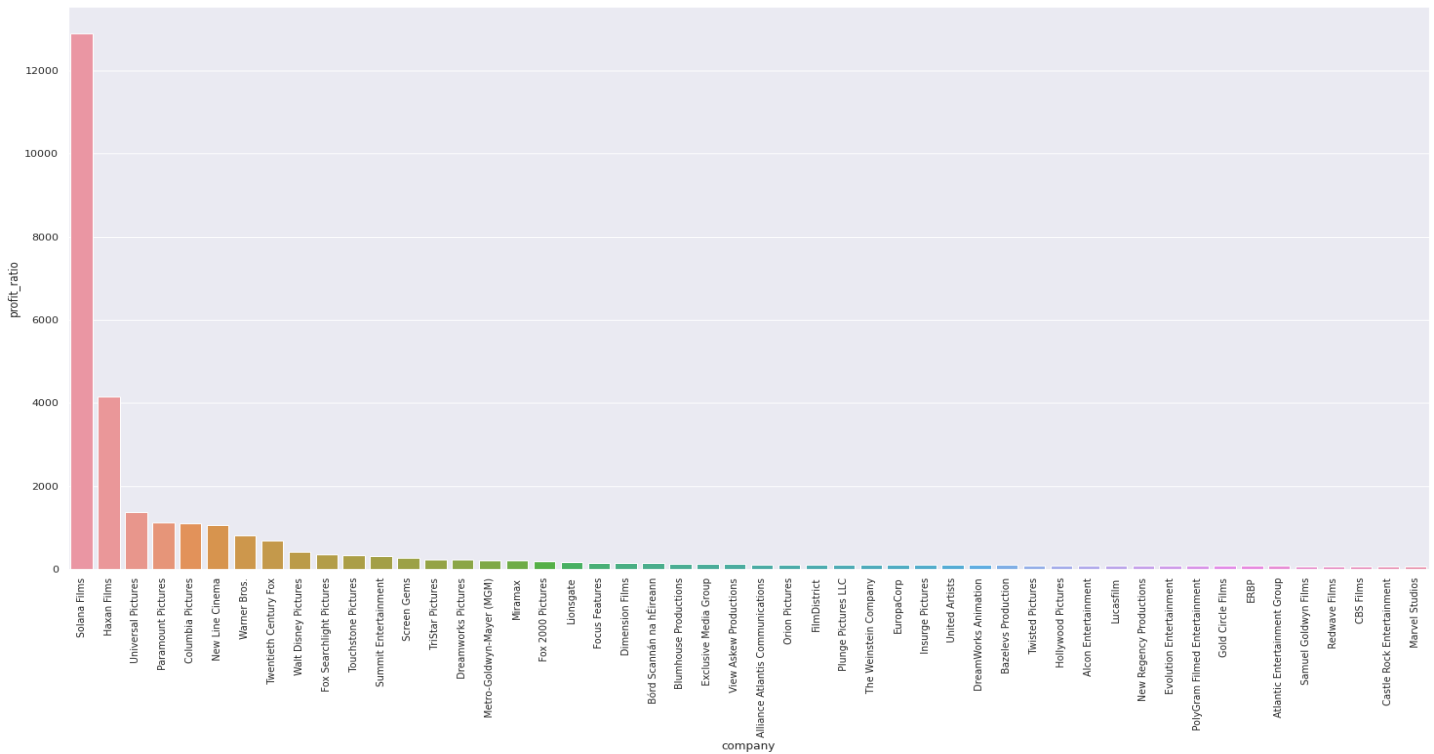
10. Analyzing the genre that makes the most money

Genres vs Revenue

- The genre most profitable according to our analysis is Horror

## 11. Determining which company has best movie investments

Revenue vs Production Company

- The company with the highest return on investment on our data set is Solana films. Solana films has the best movie investment being the company producing Horror movies which as seen above had a great return on investment
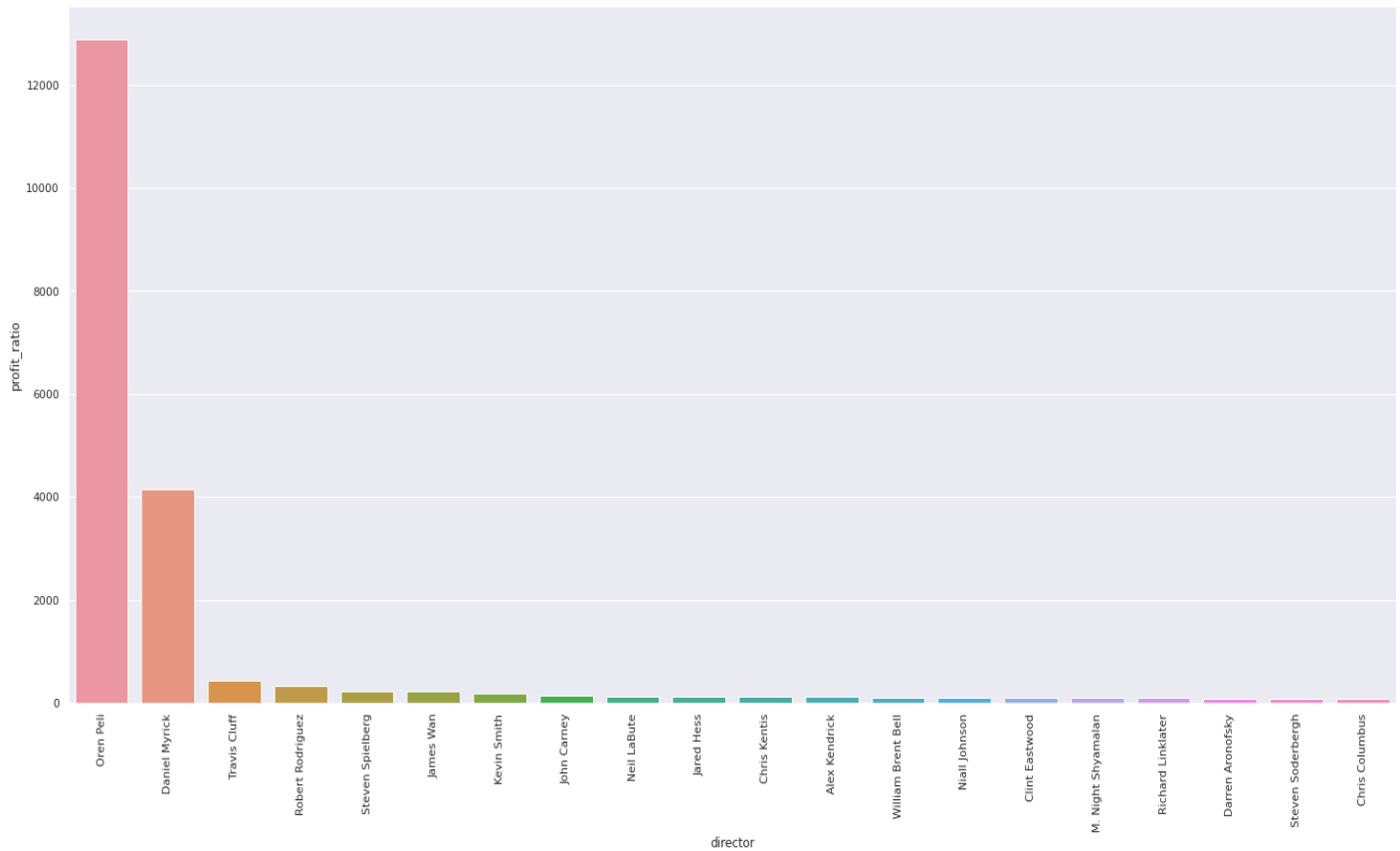
12. Company with the highest movie investment is Warner Bros.

Revenue vs Production Company
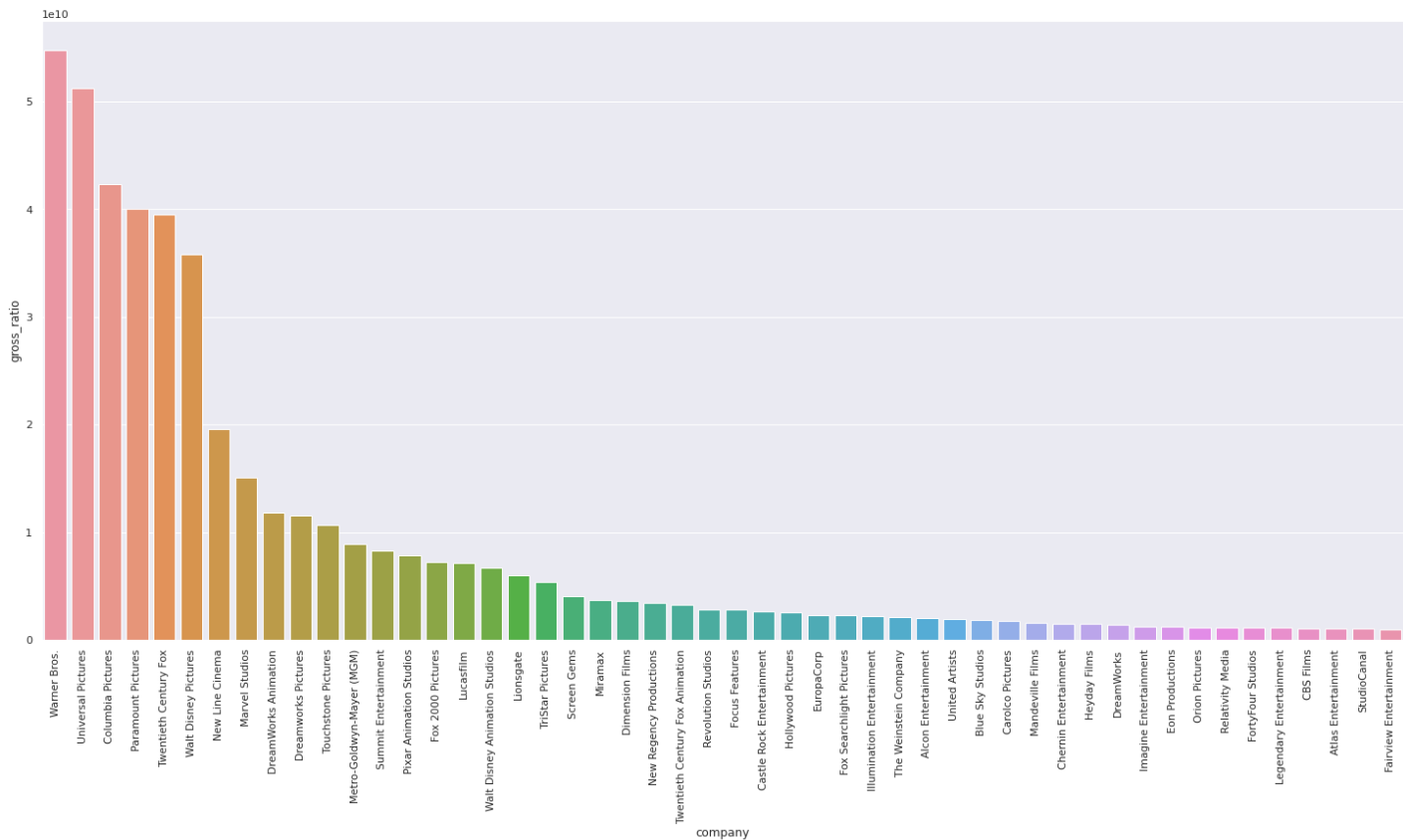
12. Determining which star's and director's movies have the highest possibility of income success based on previous successes
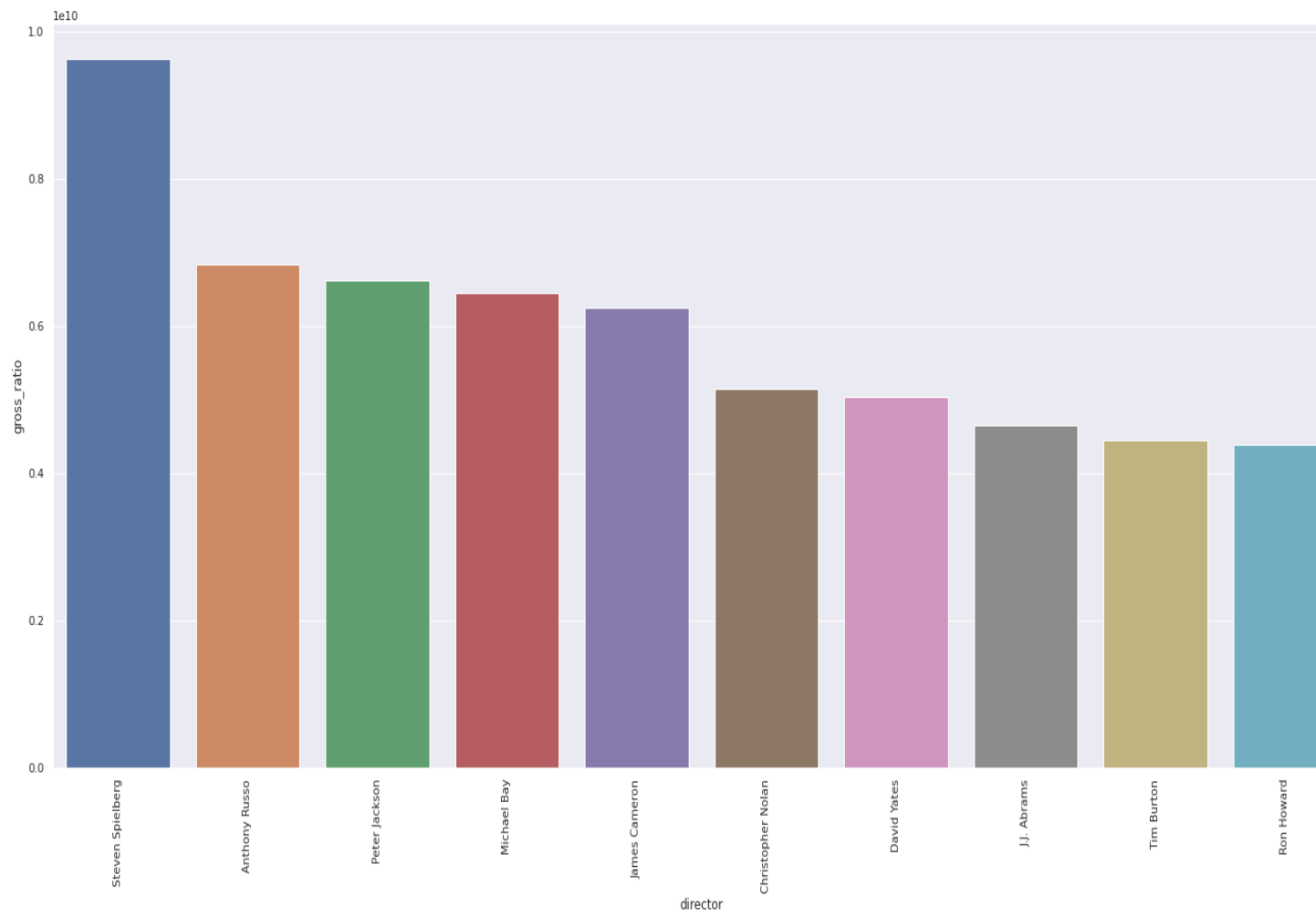
Directors possibility of income success based on previous successes

- The director with the highest possibility of income success based on previous successes is Oren Peli in terms of profit ratio.

Director vs Gross revenue

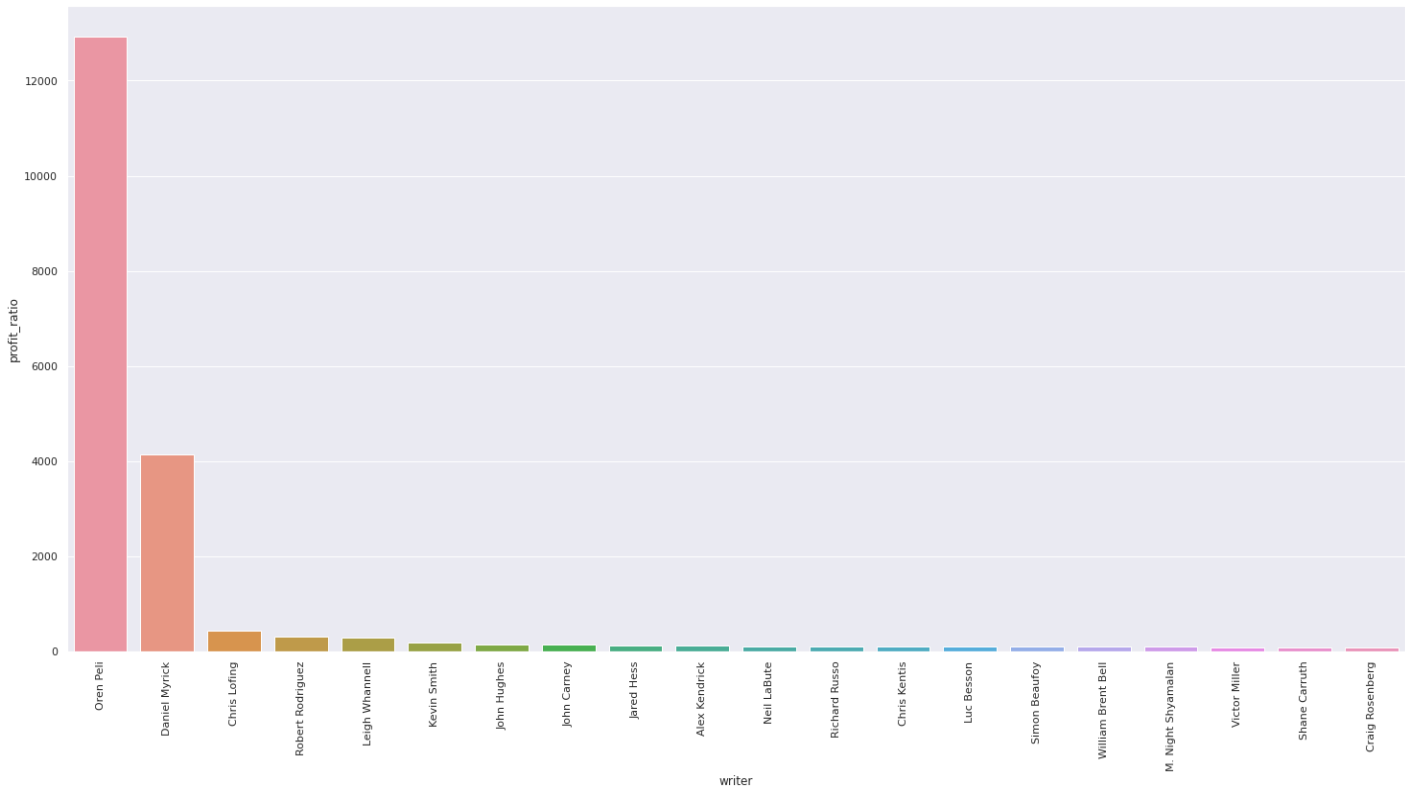Directors possibility of income success based on previous successes

Steven Speilsberg had the highest gross revenue overall

13. Writers possibility of income success based on previous successes
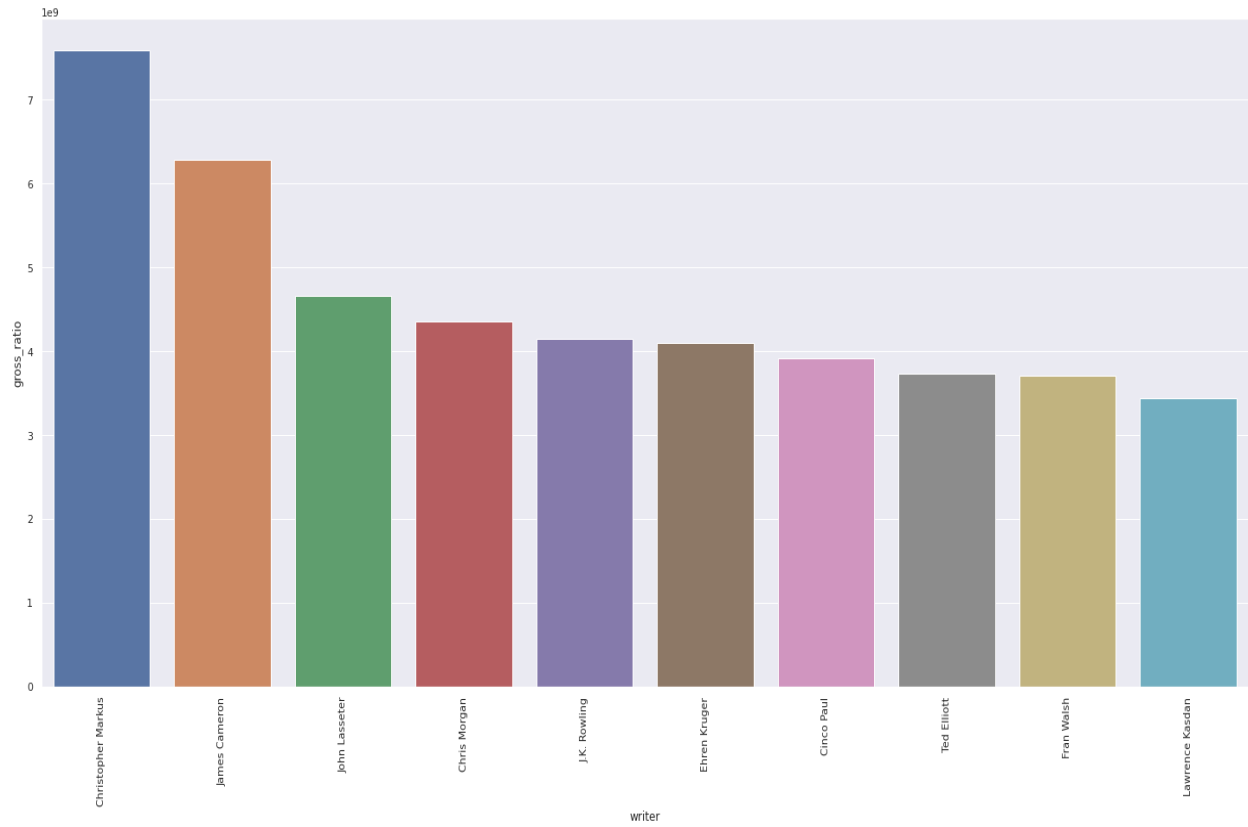
Writer vs Profit Ratio

Writers possibility of income success based on previous successes

- The writer with highest possibility of income success based on previous successes is Oren Peli

Writer vs Gross revenue

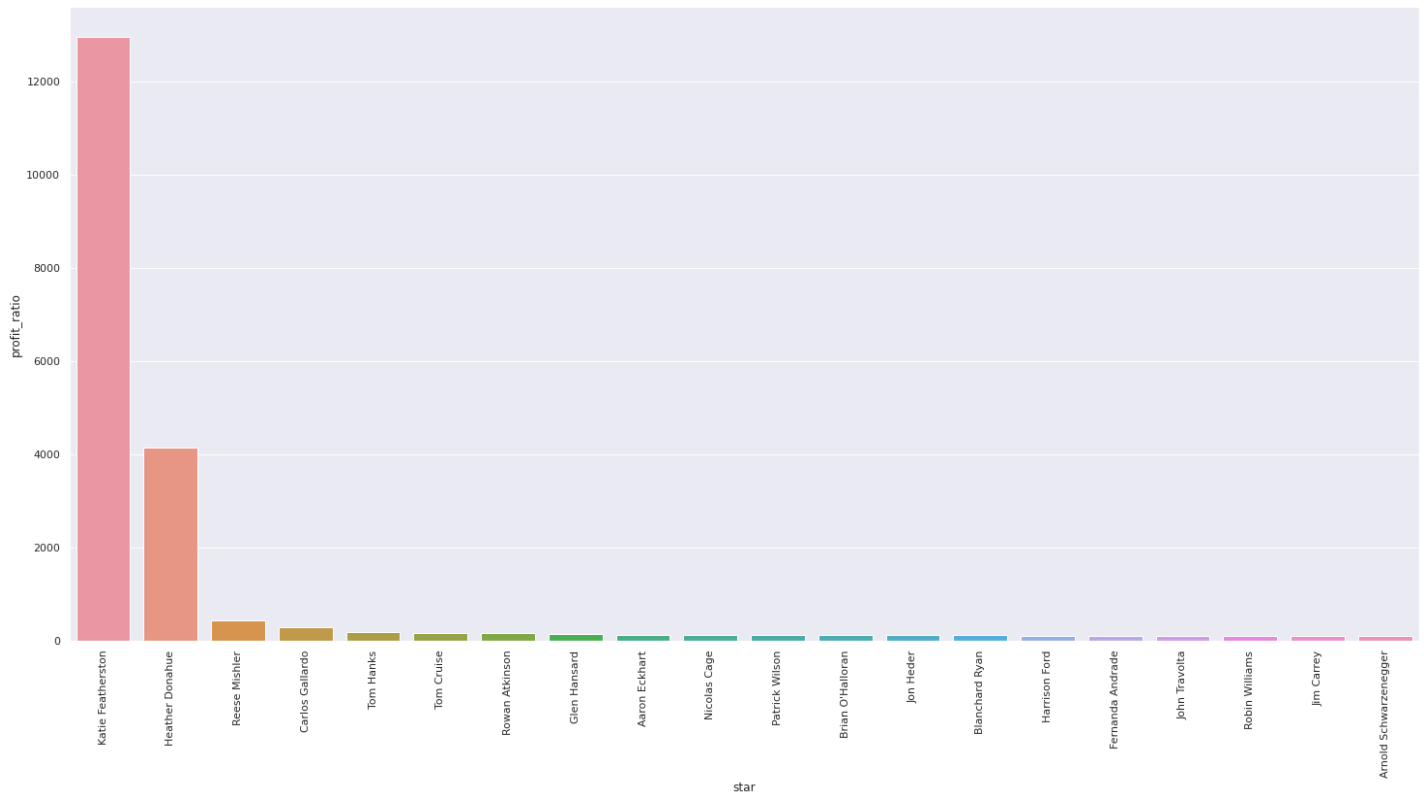Writer possibility of income success based on previous successes

Chrisptopher markus overall had the highest gross revenue

14. Actor's possibility of income success based on previous success

Actors vs Profit Ratio

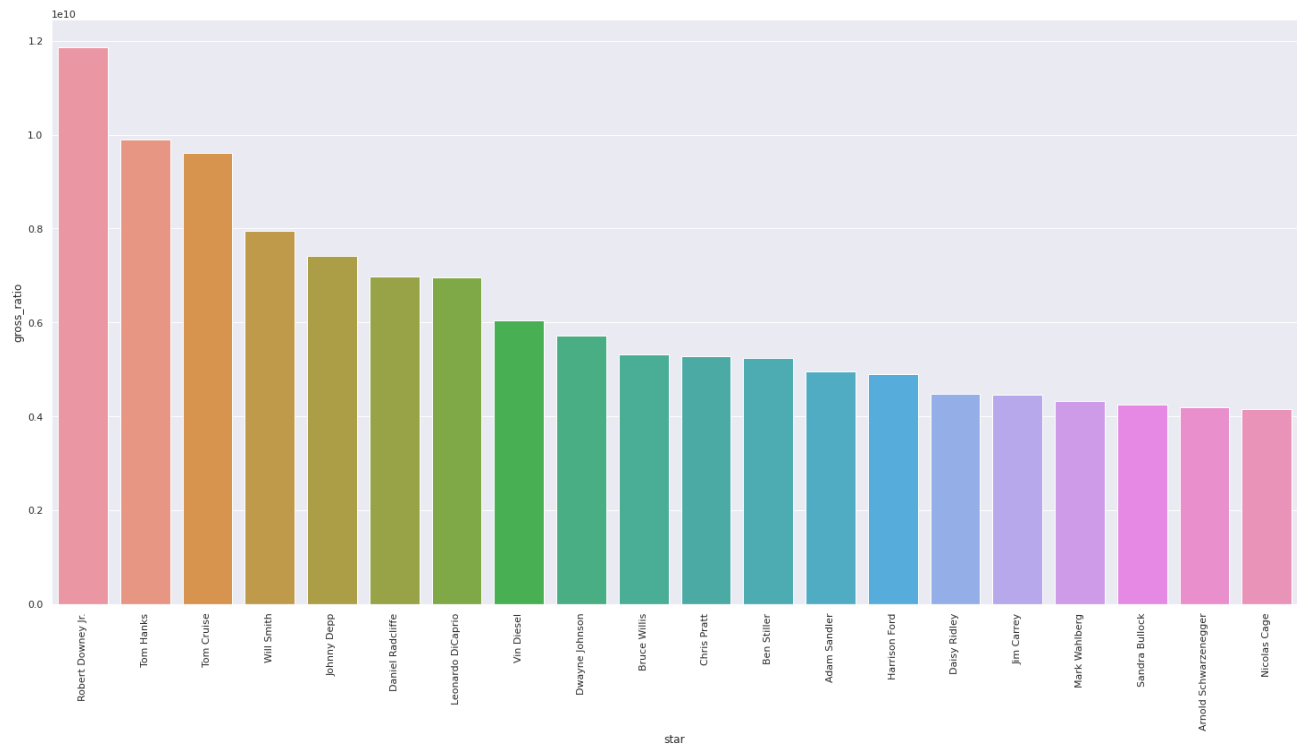Actors possibility of income success based on previous successes

- The actor with the highest possibility of income success based on previous successes is Katie Heatherstone.

14. Actors possibility of income success based on previous successes
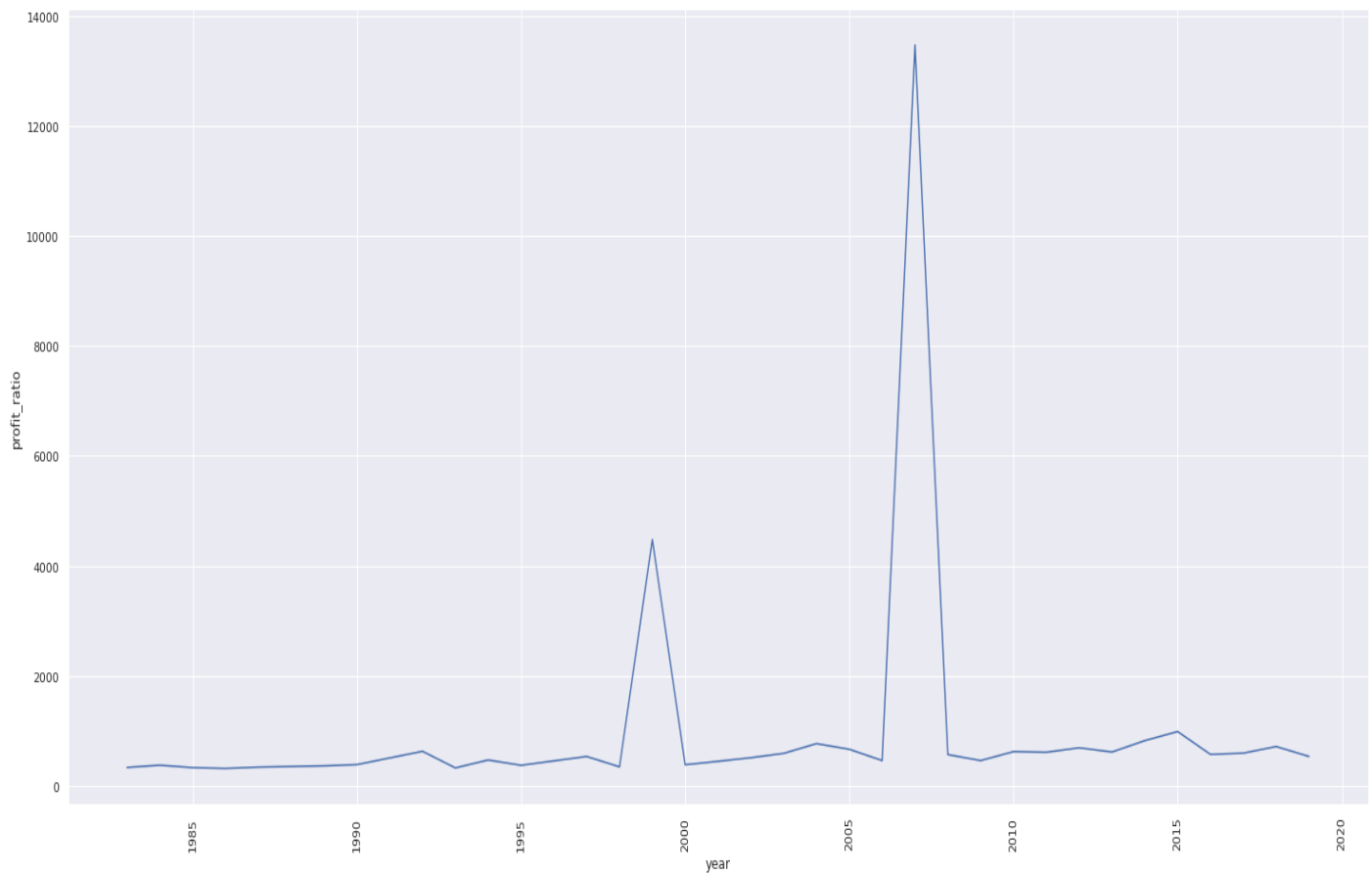
Actor vs Gross revenue

Actors possibility of income success based on previous successes

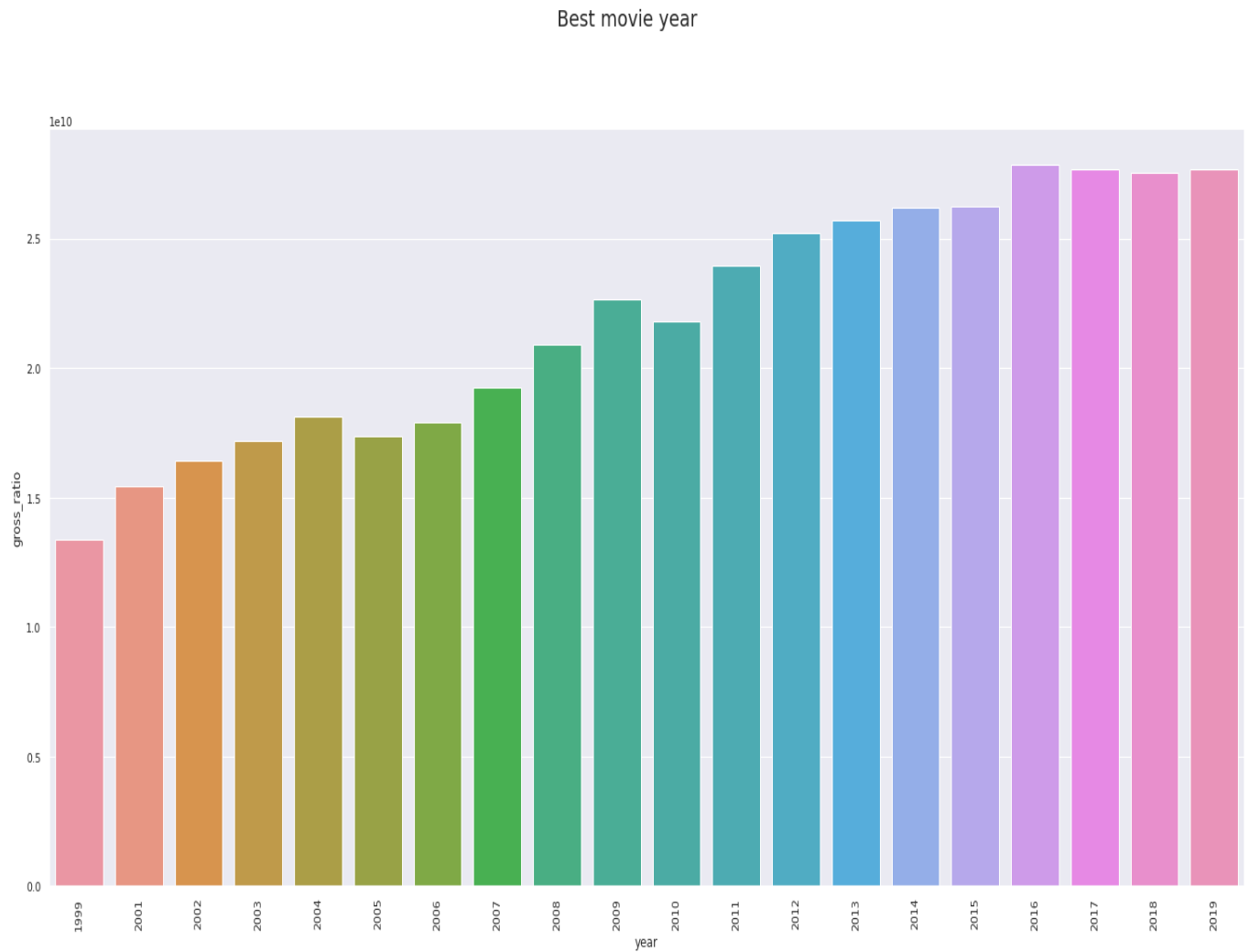- Robert Downer Jr was the actor with the highest gross revenue overall

15. Determining which year was a good movie year and why in terms of Profit ratio

Year revenue across the board

The best year is with the highest profit ratio was 2007 reason it was the premier year for the Paranormal activity movie
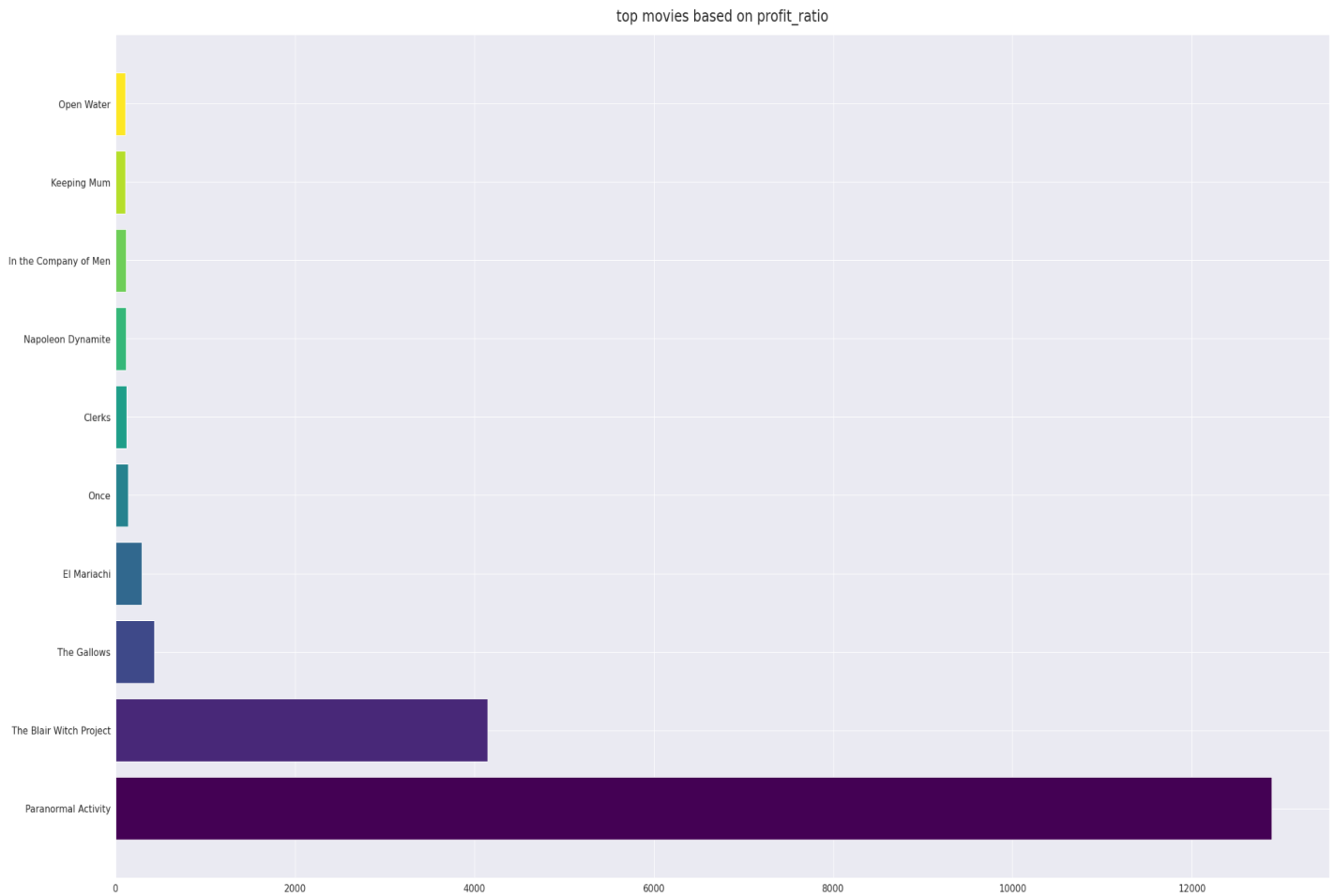
16. Best movie year in terms of Gross revenue

Best movie year

- 2016 has the highest gross revenue. It had the highest release of action movie

17. Determining highest grossing film by different attributes

I. Top movies based on profit ratio

top movies based on profit_ratio

- The highest movie by profit ratio was Paranormal activity. (had a $15000 investment which resulted to $1.3M profits)
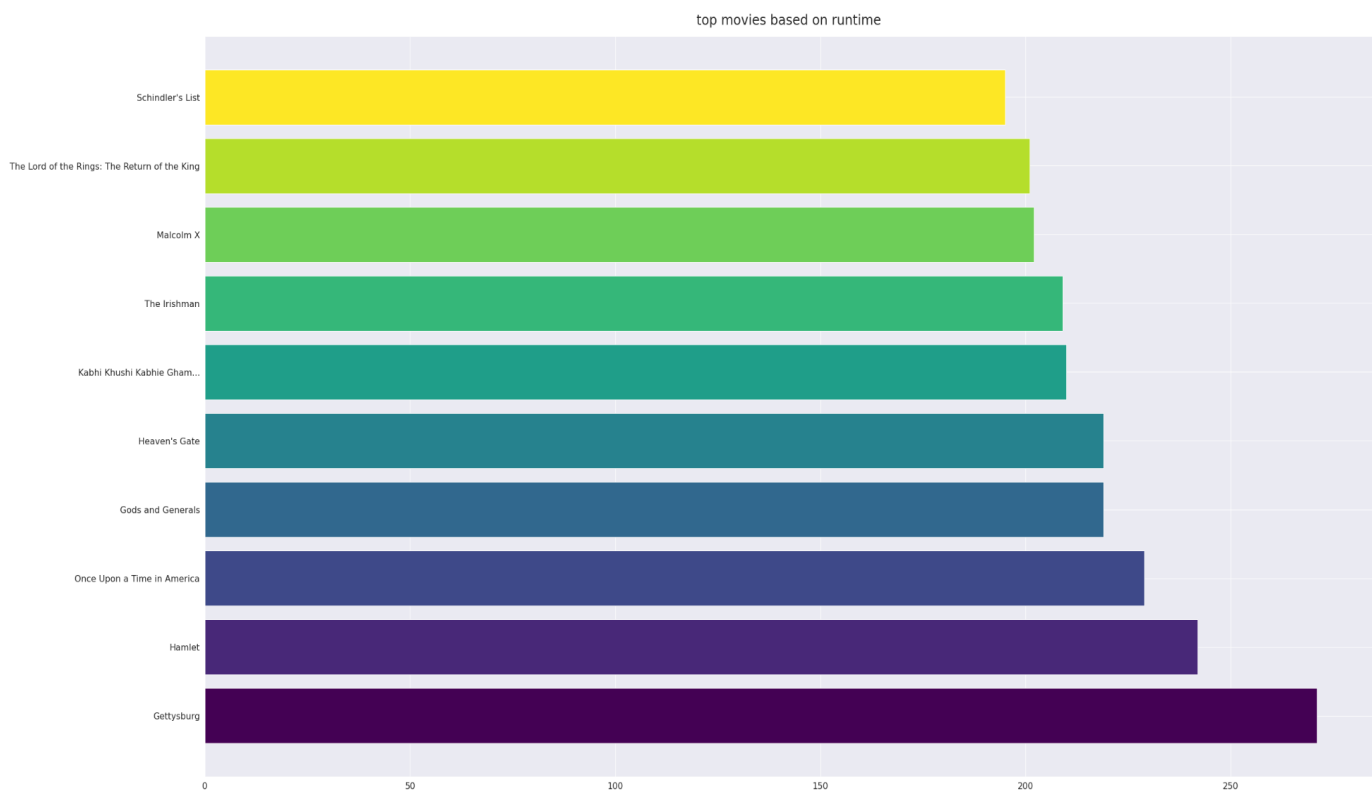
II. Based on votes

top movies based on votes
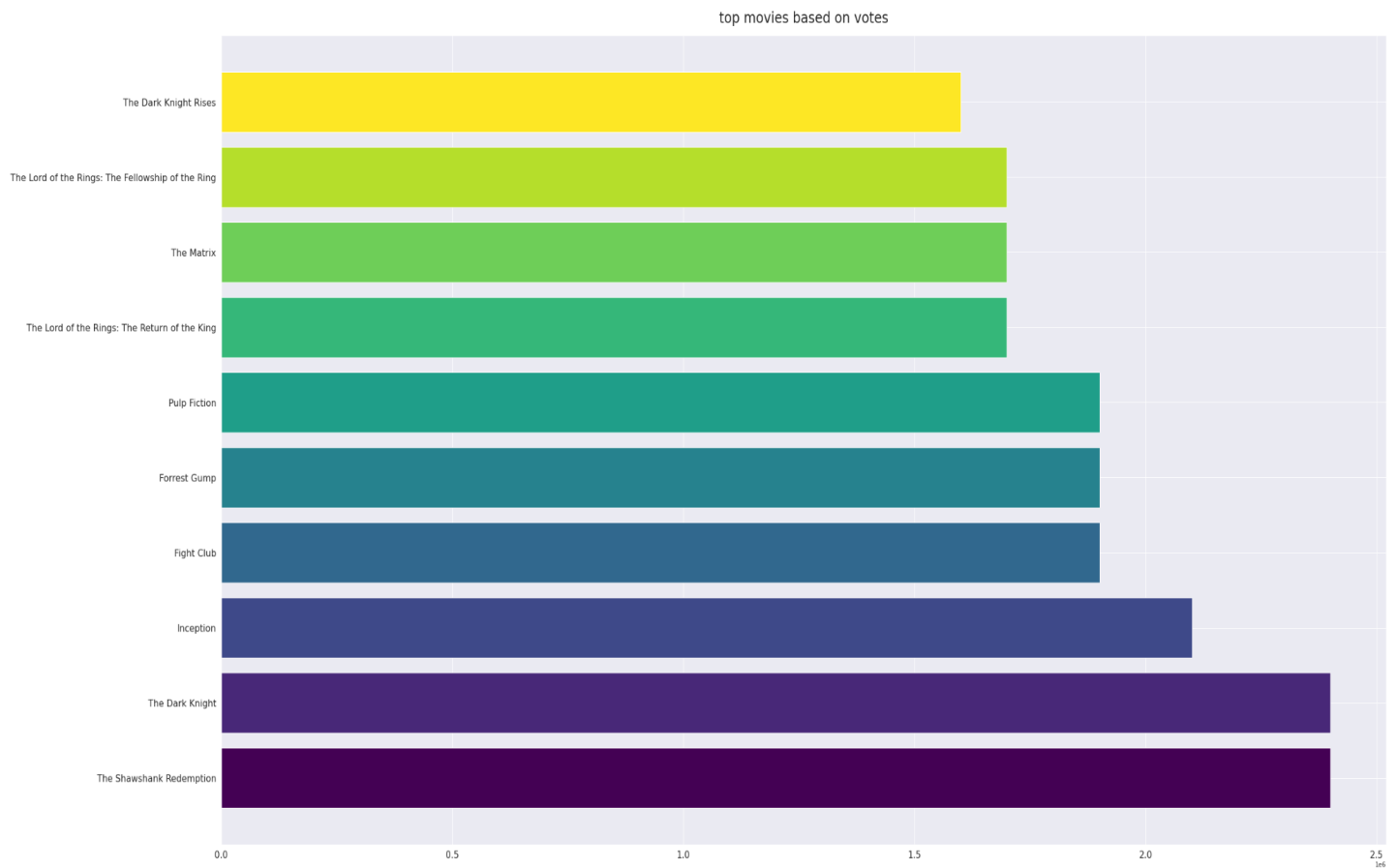
- The highest movie by votes was The Shawshank redemption. (It's the best or highest vote movie in IMDB)

III. Top movies based on rating

top movies based on rating

- Top movies had the same rating which is R rating.
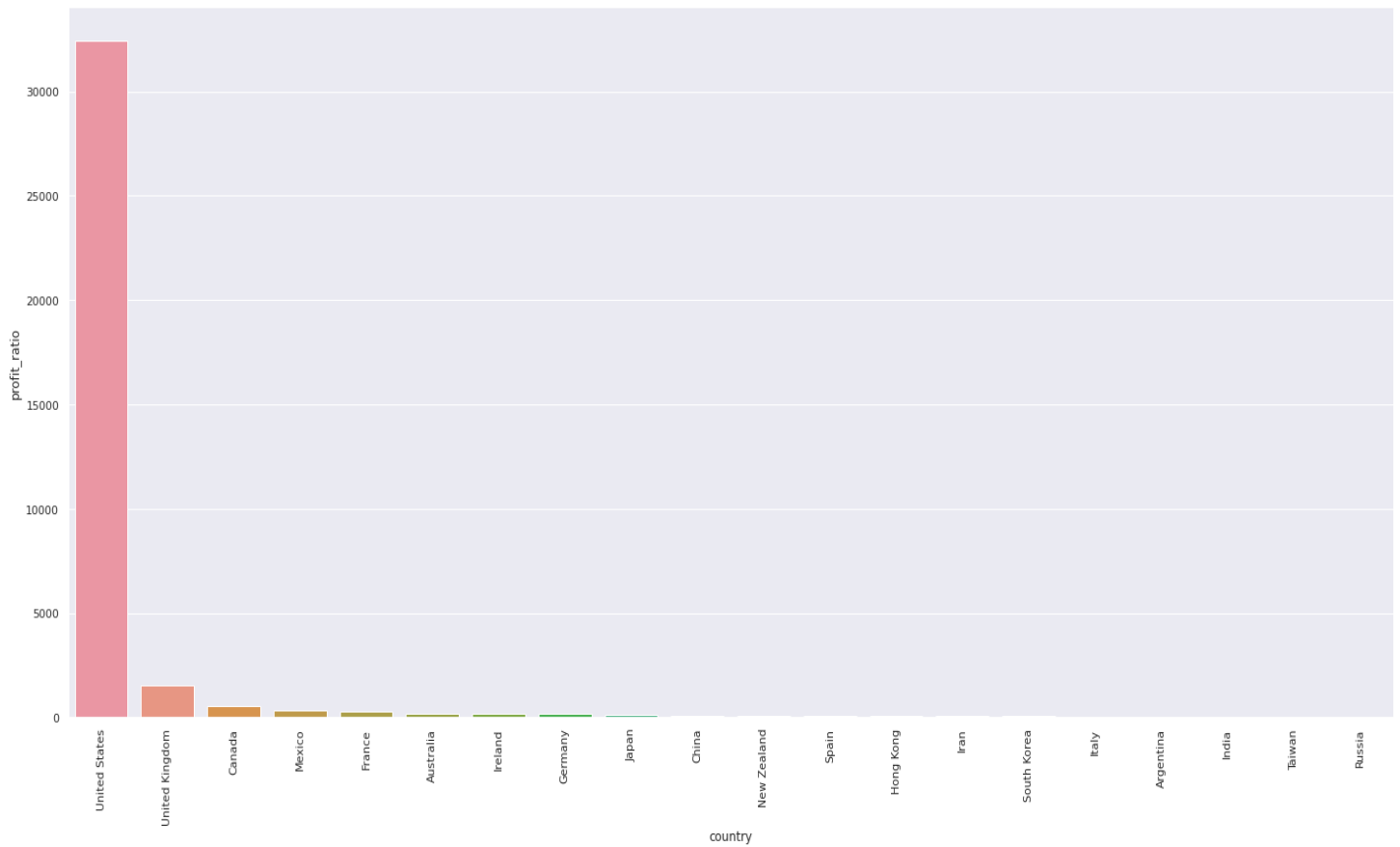
IV. Top movies based on runtime

top movies based on runtime

- The highest movie by runtime was Gettysburg. (It was a 4hour runtime Film)

18. Determining the country of origin that produces the movies with the highest film
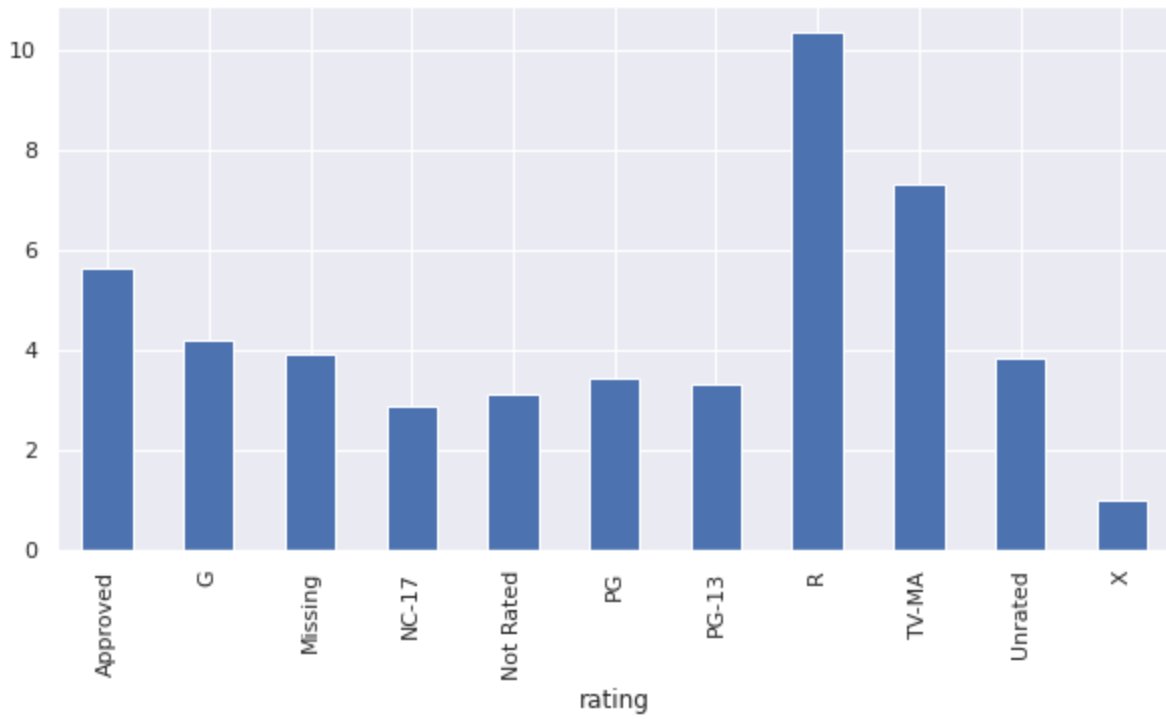
country of origin that produces the movies with the highest film

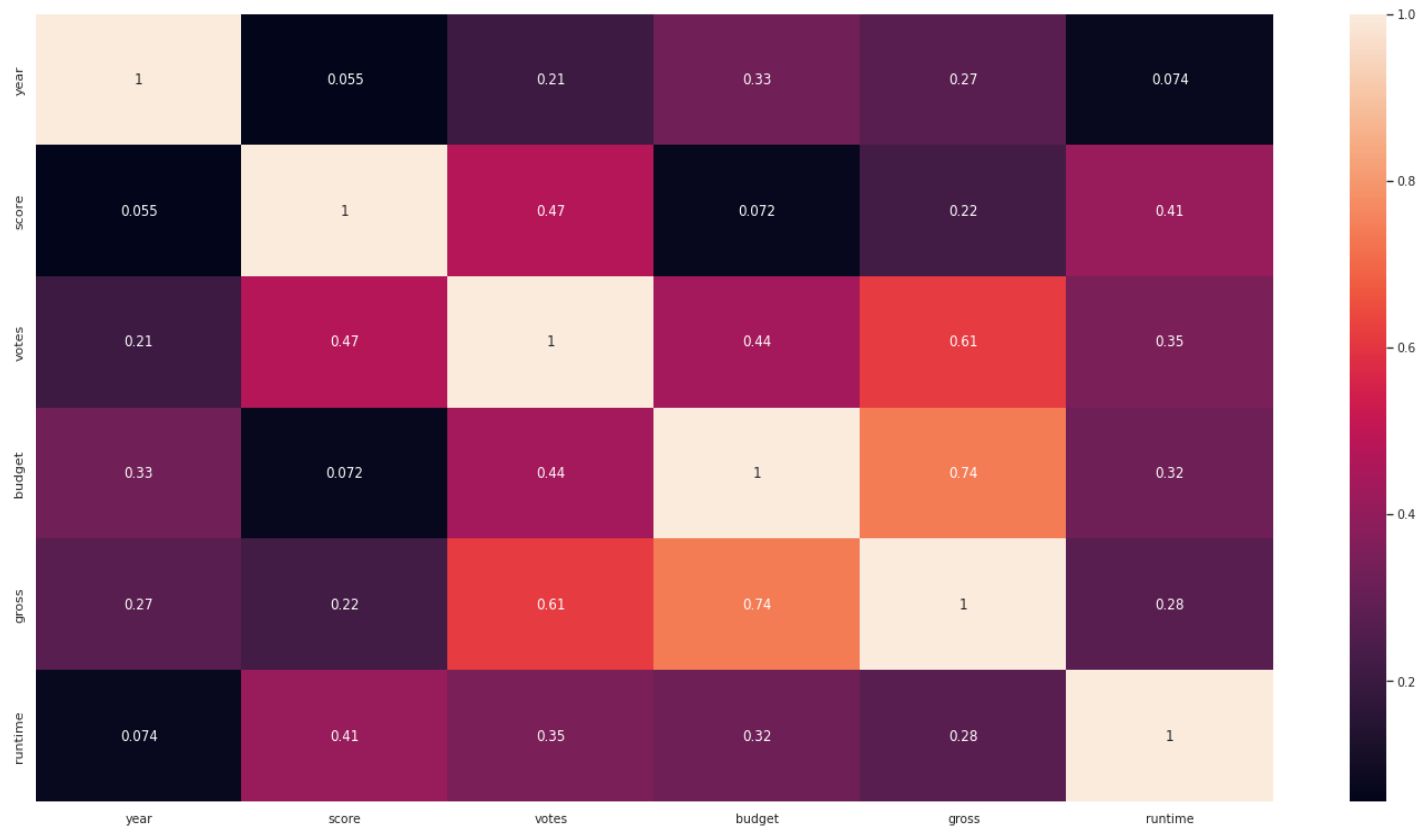- United states is the country of origin that produces the movies with the highest film

19. Relationship between the user rating and the income of a movie.

Rating vs revenue

- R rating generates more revenue

20. Multivariate Analysis

- The gross revenue and the budget are positively correlated
- The votes and gross revenue are positively correlated
- The gross and year are positively correlate

5. Conclusion

During our analysis we determined that the movie industry was stronger than ever with a rising trajectory since 1980, with the peak year as 2016 which had the highest revenue ever. Although with the addition of streaming services it would seem that there would be decline in revenue but with this analysis we can conclude that the movie industry was not affected based on our dataset.

The Actors that had the highest number of eyeballs looking ahead for their films are Robert Downey Junior, Tom Hanks and Tom Cruise since they grossed the highest in the Box Office.

6. Recommendations

- We recommend streaming services invest heavily on Action and Comedy Films since they have a high gross revenue to profit ratio.
- Streaming services should partner with these production companies i.e. Warner Bros, Universal Studios, Columbia Pictures and 20th Century Fox
- We recommend working with Steven Spielberg and Aaron Sorkin since he grossed a high revenue to profit ratio through their Action and Comedy films respectively
- The highest performing writers were Christopher Markus James Cameron.