

Intelligence Analytics Challenge – 4.0

“IN GOD WE TRUST. ALL OTHERS MUST BRING DATA.”

– [W. Edwards Deming](#)

Cleaning Steps

- 1) The entries where budget or box-office gross were not available were removed.
- 2) The multivalued fields were split into multiple fields.
 - a. Actor
 - b. Director
 - c. Production
 - d. Genre
 - e. Keywords
- 3) The budget and box-office currency was in EUR, GBP at places. These were converted to USD values.
- 4) The fields where entries were N/A, the rows are either removed, old replaced with -1 as suited.
- 5) Two calculated fields were created.
 - a. Profit = BoxOffice_Gross – Budget
 - b. Profit % = (Profit/Budget) * 100
- 6) The fields which are irrelevant to this analysis were removed:
 - a. Poster
 - b. DVD release date
 - c. Movie votes
 - d. Run time

Approach

The dataset was initially cleaned and then was modelled in [SAP Analytics Cloud](#). The modelled dataset was then supplied to [Smart Discovery \(Machine Learning + Predictive Analytics\) tools](#) that comes with SAP Analytics Cloud. Using the insights, multiple dashboards were built to answer each question.

The visualization files that comes with this submission is a [merger of multiple stories and dashboards](#) that addresses each specific question.

Tools

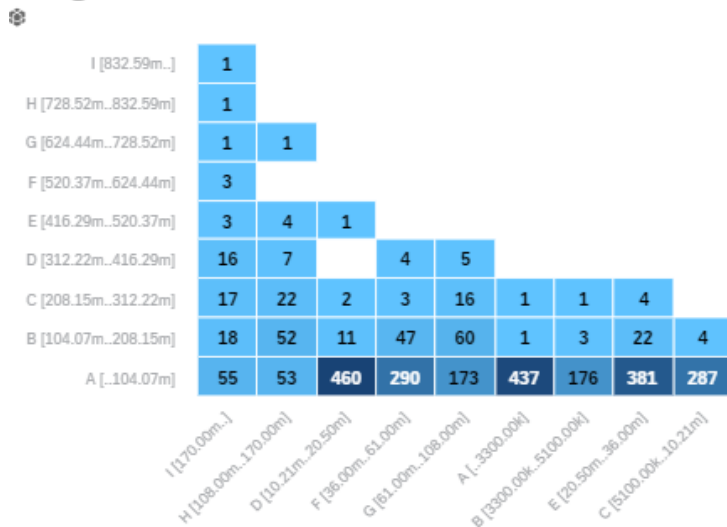
- Microsoft Excel
- SAP Analytics Cloud
- Tableau

Question #1

Why do some small budget films become blockbuster hits? Alternatively, why do some large budget films fail?

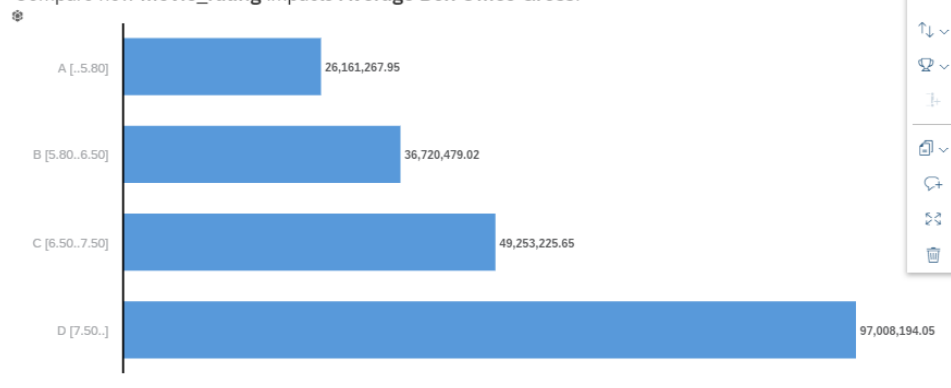
- The key influencers that affect the Box office Gross are as follows with the decreasing order of their importance:
 - Budget, Movie_rating [voters score], Genre, production, MPAA rating, actors.
- Among these Budget has the highest influence followed by movie_rating.
- From the below screenshot, it is clear that movies with low budget become blockbuster hits, while the movies with larger budget fail. There is only 1 film that is a hit with the highest budget and around 55 films even though having the highest budget have failed.

Compare the records per Box Office Gross and Budget.



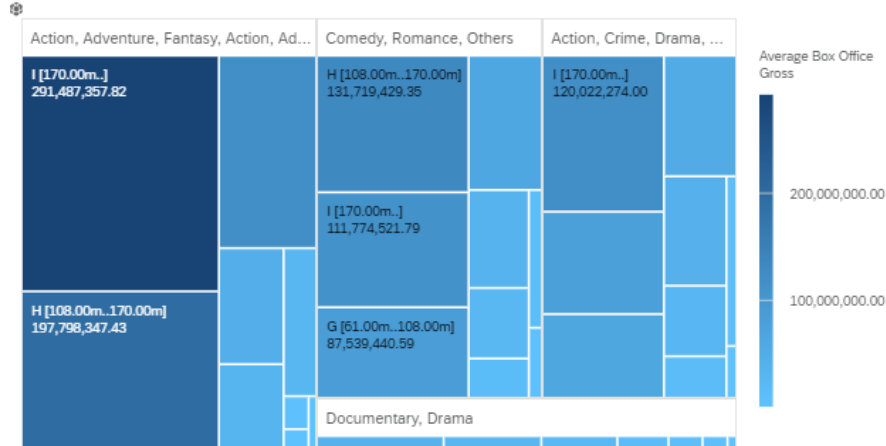
- By Comparing movie rating with gross budget, we can get insights that gross is greater for movies with higher ratings

Compare how movie_rating impacts Average Box Office Gross.



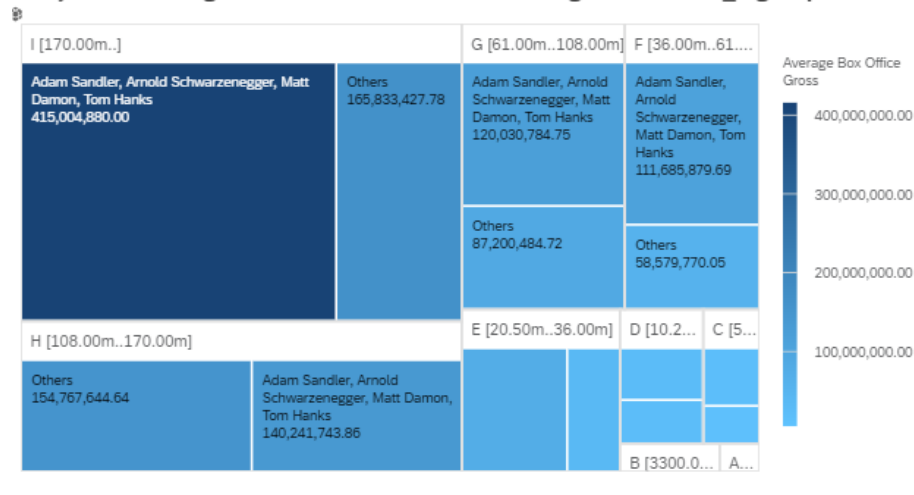
- **By Analyzing genre, budget and gross:** There are quality movies with genre Comedy, Romance and others, having more gross even if their budget is less. Similar is the case with the genres, Action, Crime and Drama.

Analyze the **Average Box Office Gross** between **genre group** and **Budget**.



- Actors also have a tremendous influence on a movie success, even if the budget is not so high. From the below graph we can get an idea that even if the budget is not high for films with actors like Adam Sandler, Arnold Schwarzenegger, Matt Damon, Tom Hanks, movies have gained tremendous profits.

Analyze the **Average Box Office Gross** between **Budget** and **actors_1 group**.



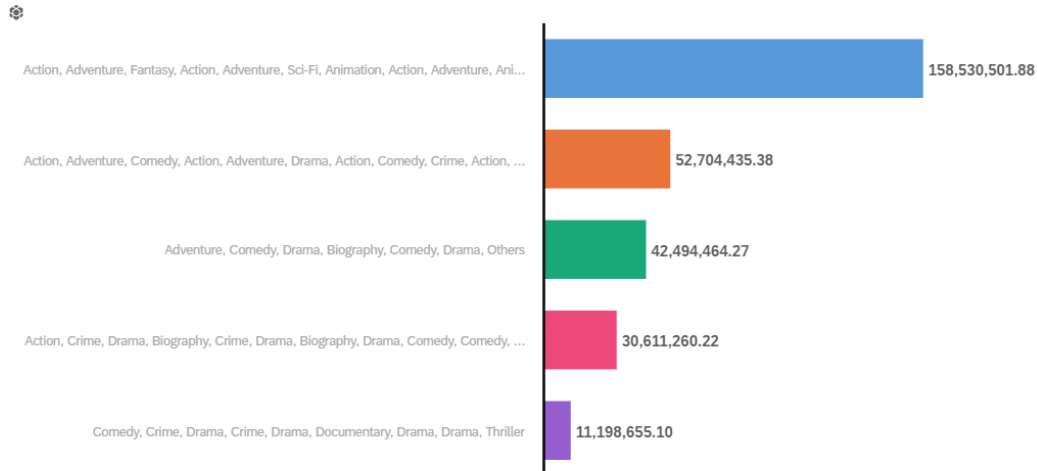
Even though, budget is a criterion, there are other factors that influence the success of the film immensely.

Question #2

How do movies fare in terms of genre? Comedy, science fiction?

- Comedy is a critical genre in the movies. It is a major factor affecting almost all the categories except for the Documentary and Drama Genre, rest it is present in every bin of categorization.
- The first bin is the highest genres combinations for gross income, it combines with sci-fi, Action, Adventure, fantasy, Crime, thriller and animation.

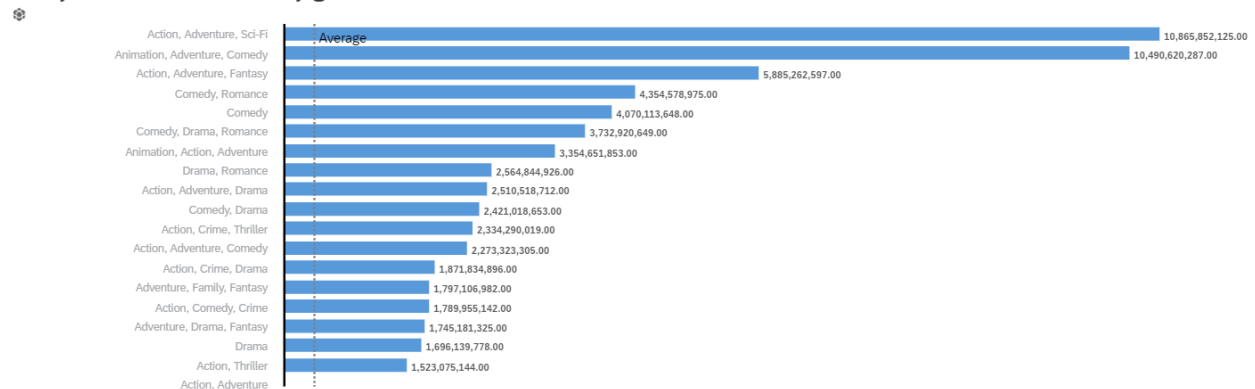
Compare the **Average Box Office Gross** of each **genre group**.



Note: The combination of multiple genre has a different influence as compared to a type of genre. So, the genre field is kept as it is so that the interplay and combination of genre and the corresponding factors can be analyzed.

The total gross for each genre is as follows:

Analyze **Box Office Gross** by **genre**.



The combination of Action, Adventure, Sci-Fi is the top contributor, 2802% above the average

Question #4

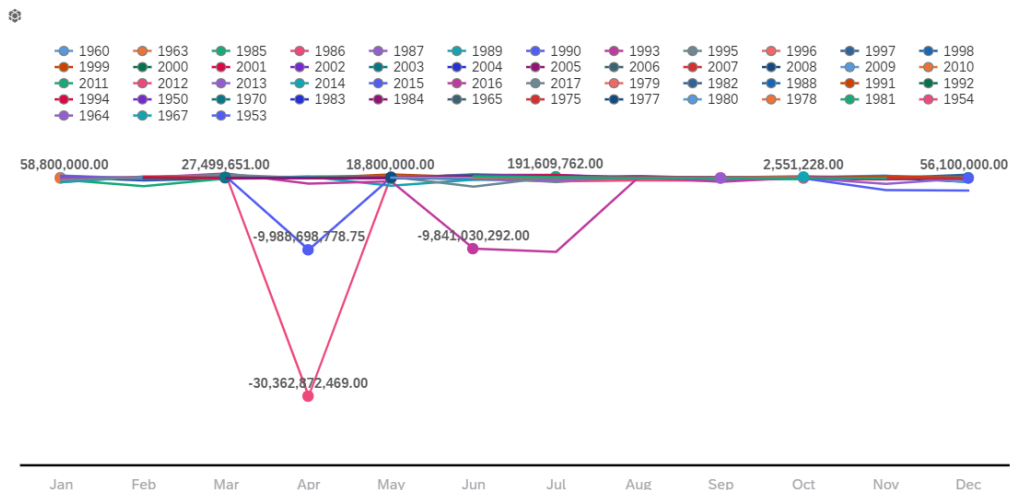
Does release date influence box office performance? Do movies released at the same time perform differently?

- The date field is split as day, month, year and aggregated to study the various trends on movie release on a particular day, month and year.



a) Trend Across days – Aggregated Profit

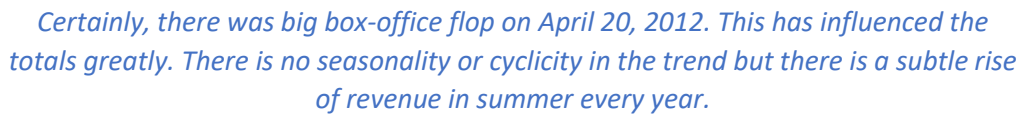
Profit per release_month, release_year for Actual



b) Trend Across Months over years – Aggregated Profit

The metric chosen to run the comparative study is profit (calculated as: Gross – Budget).

The graph shows some evident outliers. After removing them, the above graph looks so.

 | **release_year + release_month Filter**

Question #6

Does rating of a movie have any effect on movie earnings?

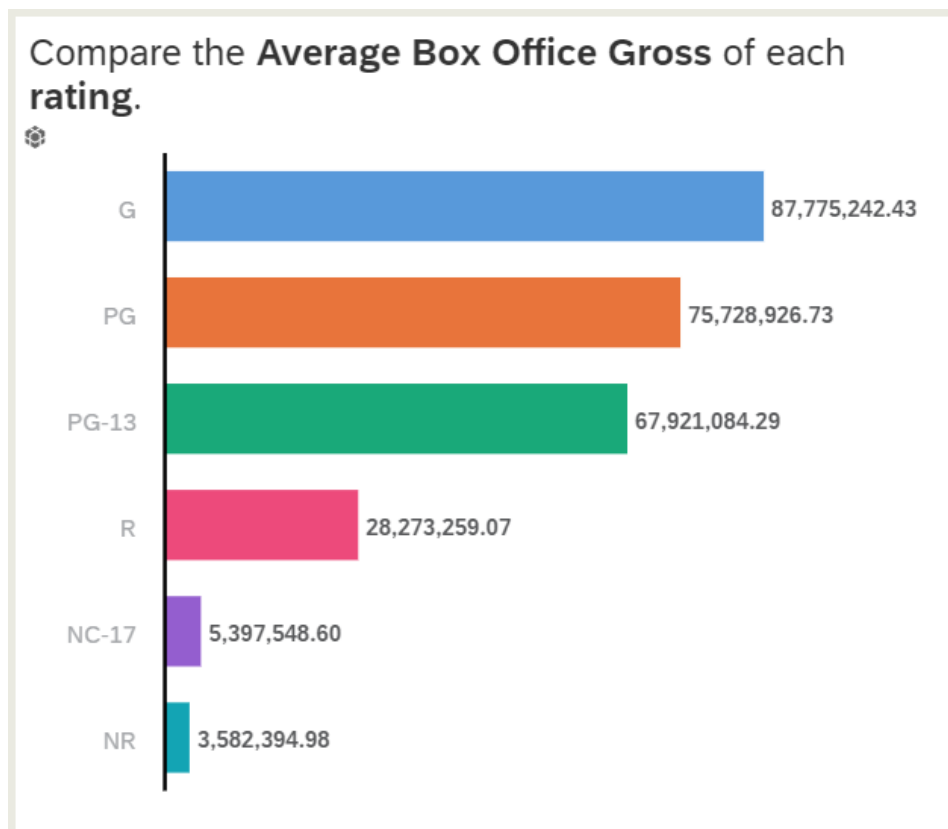
The question does not tell us if it was talking about the IMDB rating or movie rating. Hence, a check on the influence of these both on the movie earnings (Box Office Gross) was conducted.

We found 2 key influencers for Box Office Gross.

🔗 | Mar 24, 2019 13:40:49 | High Quality (4/5)

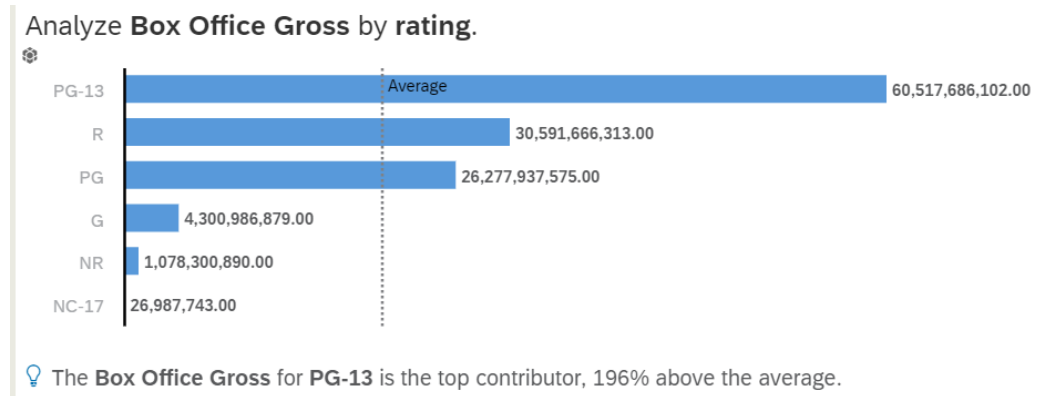


Even though, both have influence, rating stood out. Analysis was carried out for both the metrics and the results are as below.

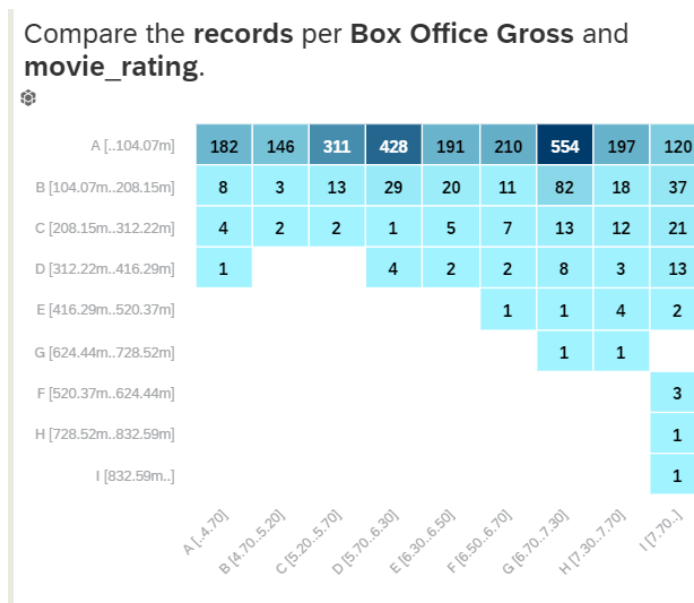


a) The average is high for G

However, the PG-13 category stands out by earning the most. The total box office gross of PG-13 category is around 200% above the average.



The effect of movie-rating is as follows:



- There are movies with high movie-rating which have ended up being flops.
- There are low-rated movies that have went on to do well.
- However, the movies with high rating mostly do well.
- The bin of **[6.7 – 7.3]** is a safe zone and most of the movies which are part of that bin seems to be doing well.

The PG-13 and G categories have been successful. It is safer to make movies in this category. A mid-range to high range in movie-rating ascertains the success of a movie with some confidence. However, movies are subject to market risks. There is no one successful formula.

Question #7

If you are running a movie production house, which Actor(s)/Director would you like to cast in your movie?

The answer can be given from two perspectives: –

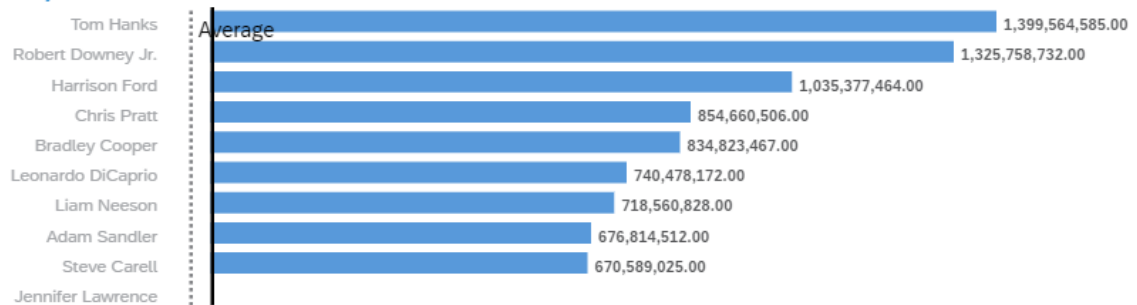
a) Profit

b) Profit %.

When we look from the profit perspective,

Analyze Profit by actors_1.

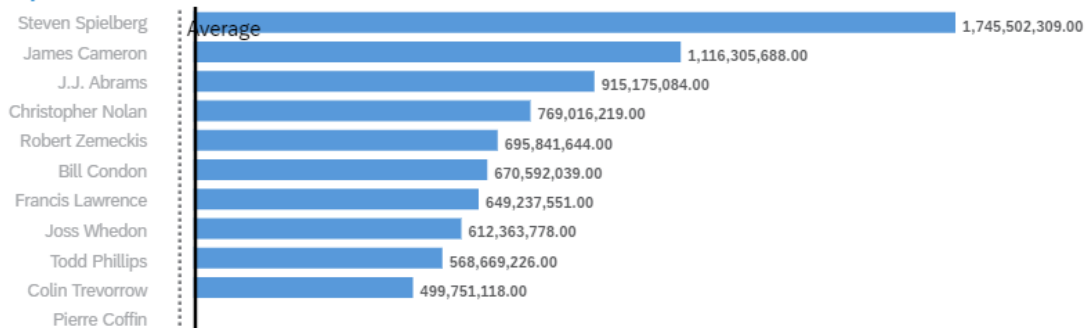
Top N (AUTO)



The Profit for Tom Hanks is the top contributor. The Profit variation was driven significantly by

Analyze Profit by director_1.

Top N (AUTO)



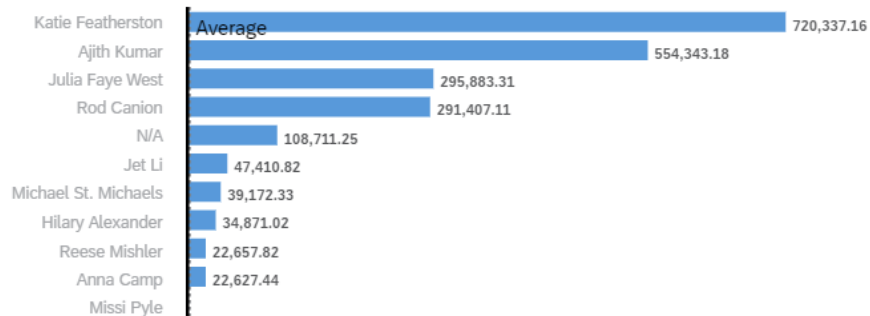
The Profit for Steven Spielberg is the top contributor.

It is highly profitable in doing movies with actors like Tom Hanks, Robert Downey Jr., Harrison Ford, Chris Pratt, Bradley Cooper, Leonardo DiCaprio and their likes. The ventures of Steven Spielberg, James Cameron, JJ Abrams and Christopher Nolan assures success to a degree.

The conundrum here is that this is all possible only if the production house is able to bear a big budget. So, an analysis was conducted to figure out if there are actors or directors who can make low budget movies huge success. Below are the results:

Analyze **Profit %** by **actors_1**.

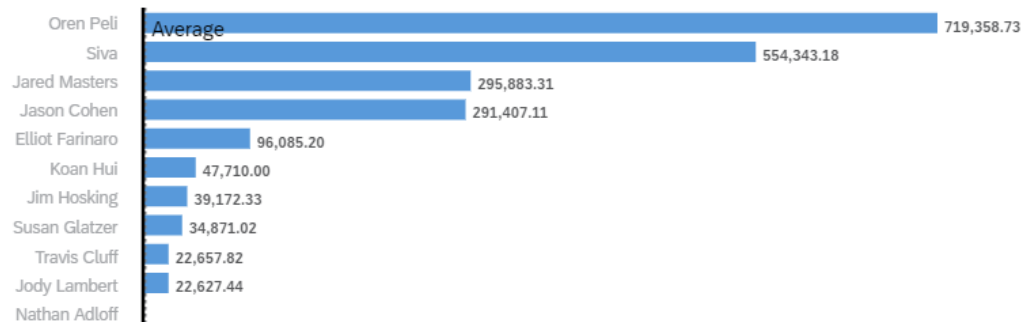
⚙️ | [Top N \(AUTO\)](#)



💡 The **Profit %** for **Katie Featherston** is the top contributor, 43641% above the average.

Analyze **Profit %** by **director_1**.

⚙️ | [Top N \(AUTO\)](#)



💡 The **Profit %** for **Oren Peli** is the top contributor, 48675% above the average.

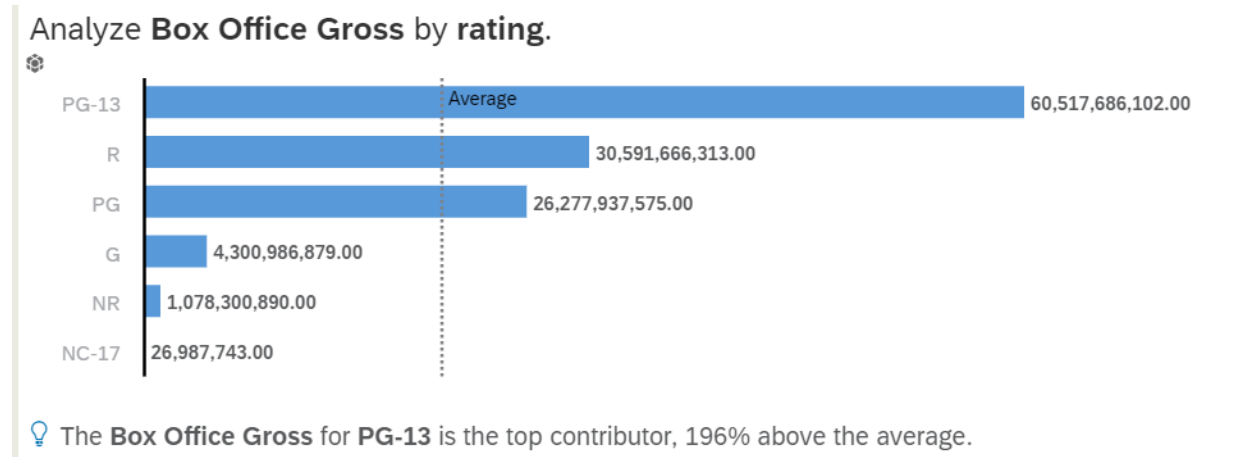
The measurement metric was profit %.

Horror movie genre starring Katie Featherston, directed by Oren Peli seems to be apt for production houses who want to take comparatively lower budget films.

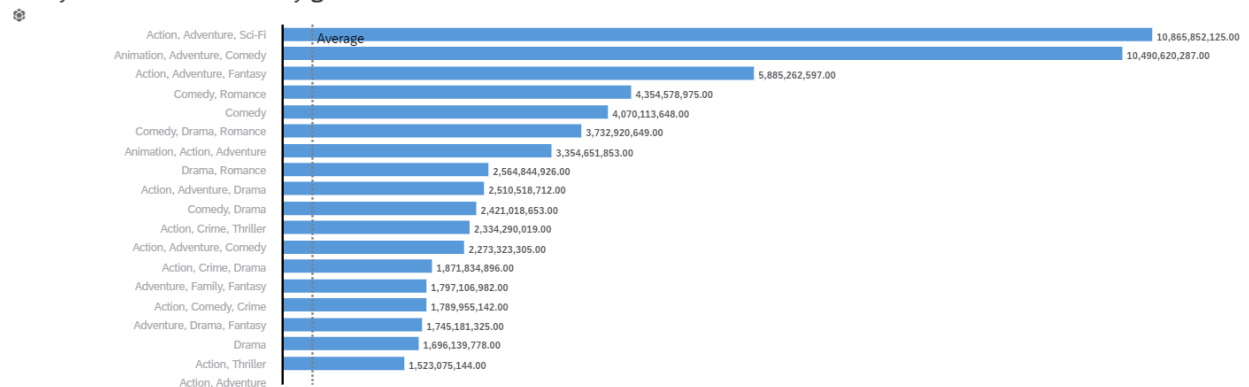
Question #8

Does a plot of a movie affect its earnings in any way?

The plot of a movie is keyword tagged and given in the dataset as keywords. However due to the complexity of the problem, we chose to go forward with fields of genre and rating which represents the plot of the movie. The below visualization throws more light into the effect.



Analyze **Box Office Gross** by genre.



The Action-Adventure-SciFi genre movie plots assures success

Question #5

If you were building a model to predict profitability of a movie at the box office. Which key variables would you consider? Bring out the relationships of these variables and their effect on profitability of a movie. Once you have studied the various relationships and have a robust model, what kind of model would you advise to a movie distributor on how he can maximize profit?

As per the analytics conducted in previous question, the key variables that should be considered are:

- Budget
- Movie_rating
- Production
- MPAA rating
- Actors
- Directors

The effect of the above variables on the movie success are:

- **Budget:** There is a chance that with high budget, the movie would either be a big flop or big success. Low budget movies and medium budget movies are strong contenders for assured movie earning.
- **Genre:** SciFi-Adventure themed movie and horror thrillers are the ones that make the most.
- **Movie_rating:** The bin [6.7 – 7.3] guarantees the commercial success of the movie with some confidence.
- **MPAA rating:** PG-13 and G are the ones which do well in box office generally.
- **Actors:** It is safer to go with actors like Tom Hanks and Robert Downey Jr. for big budget films. For low budget movies, Katie Featherston is the perfect actor.
- **Directors:** The likes of James Cameroon and Steven Spielberg are suited for high budget and Oren Peli does wonders with low budget flicks.

Two successful formulae are as below:

- Budget: High
- Genre: SciFi - Action – Adventure
- Rating: G
- Actor: Tom Hanks/Robert Downey Jr
- Director: Steven Spielberg/James Cameroon

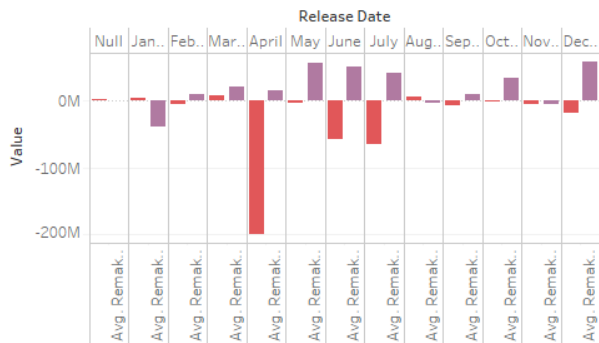
- Budget: Medium to Low
- Genre: Horror
- Rating: PG-13
- Actor: Katie Featherston
- Director: Oran Peli

Question 3

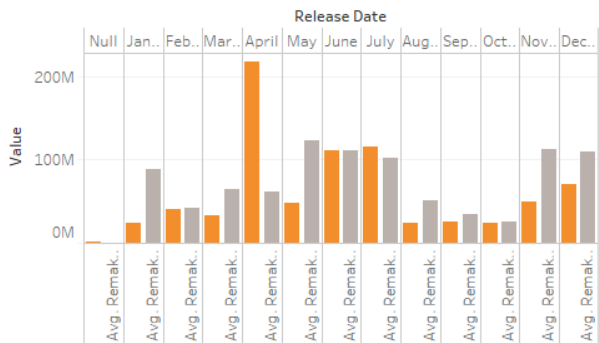
- As per the data trend the budgets of Remakes-Sequels-Tentpoles is slightly on higher side than regular movies
- Average box office gross for the remakes is low when compared with other movies whereas they fair well when we look at actual revenues earned
- In terms of overall ratings, they were at similar level to other regular flicks



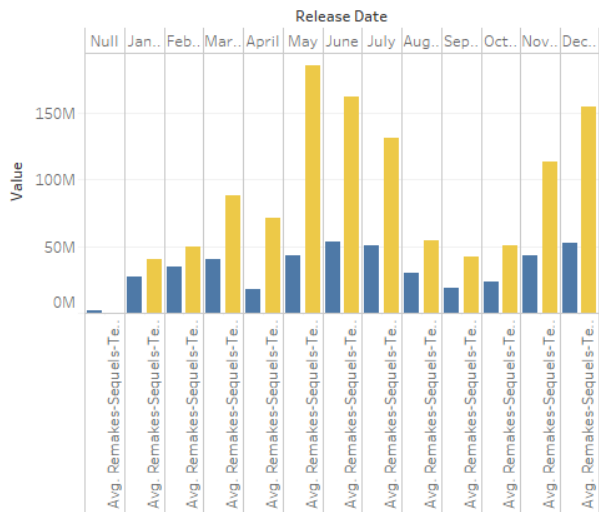
Revenue



Budget



Gross



AVERAGE RATINGS

