Business Intelligence: An EDA for the new launch of a movie studio

Robert Wong
Part-time Data Science Program, Flatiron School
Date: 23rd June 2020 Instructor: James Irving

Problem Statement

The problem statement is relevant to Microsoft's new endeavor to expand its various business segments. For instance, with years of increasing profits for its Xbox gaming console entertainment division, game development operational segment, movie streaming services, etc. Microsoft has strategized a new project to continue building relationships with its clientele with the launch of grand opening of a film studio.

Microsoft has tasked the data science department to learn about the movie industry, to see what are some of the most popular films that moviegoers would watch, and to provide movie recommendations to meet the audience's film preferences.

In this exploratory data analysis, three questions were answered in that sequential order:
What does the imdb scores and gross revenues say about certain types of films?, How does the duration and the budget play a significant role in the types of movies to produce?, and Does the directors or actors/actresses impact the return on investments (ROI) from the gross revenues, budget, and profits from the box office outcomes?

Methodology

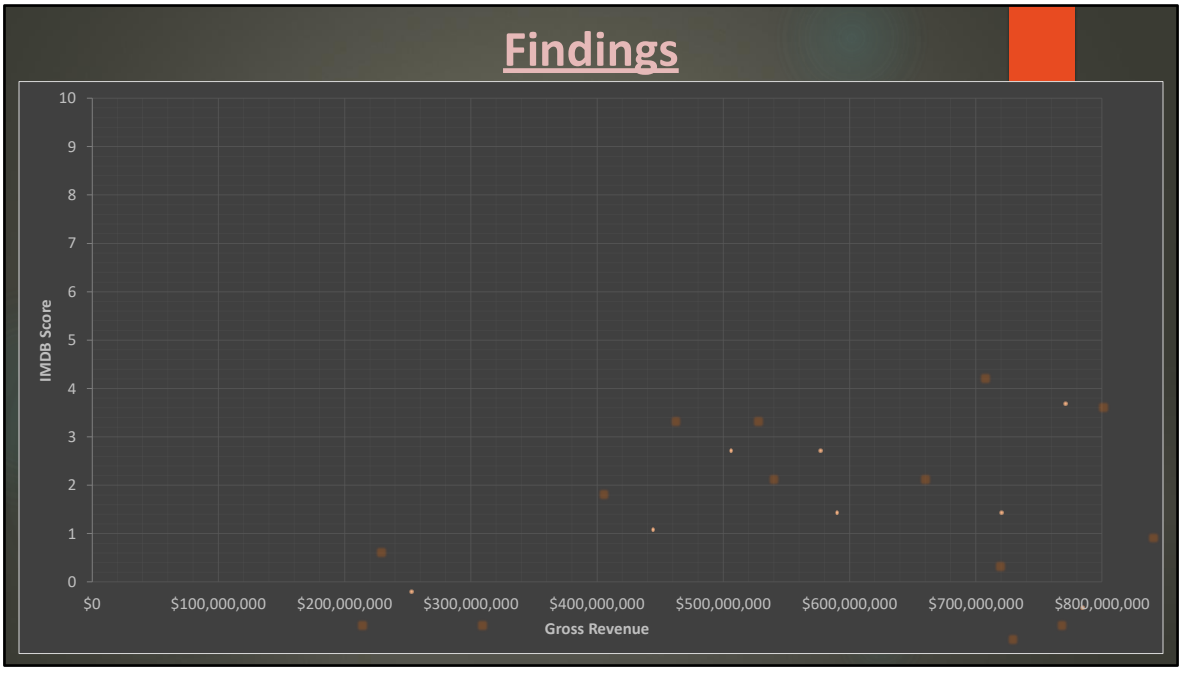
- Problem Statement
- Data Collection
- Data cleaning
- EDA Findings
- Conclusion
- Thank You

Exploratory Data Analysis on the Findings

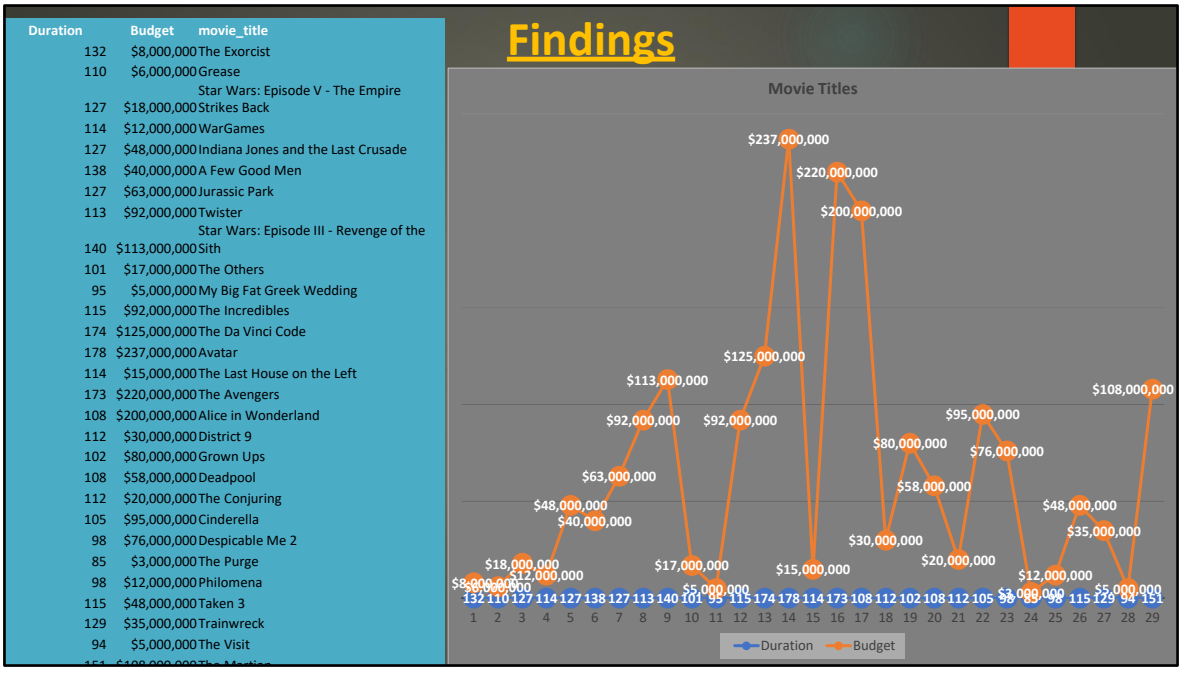
Movie Titles

Comedy	movie_title
Drama|Horror|Thriller	The Exorcist
Horror|Sci-Fi|Thriller	Grease
Horror|Mystery|Thriller	Star Wars: Episode V - The Empire Strikes Back
Mystery|Thriller	WarGames
Action|Horror|Sci-Fi|Thriller	Indiana Jones and the Last Crusade
Horror|Thriller	A Few Good Men
Action|Crime|Drama|Romance|Thriller	Jurassic Park
Action|Adventure|Sci-Fi|Thriller	Twister
Action|Adventure|Fantasy	Star Wars: Episode III - Revenge of the Sith
genres	The Others
Crime|Thriller	My Big Fat Greek Wedding
Action|Adventure|Sci-Fi	The Incredibles
Drama|Horror|Sci-Fi|Thriller	The Da Vinci Code
Action	Avatar
Comedy|History	The Last House on the Left
Action|Adventure	The Avengers
Drama|Thriller	Alice in Wonderland
Adventure|Sci-Fi|Thriller	District 9
Action|Comedy|Crime	Grown Ups
Action|Crime|Thriller	Deadpool
Fantasy|Horror|Thriller	The Conjuring
Horror	Cinderella
Action|Comedy	Despicable Me 2
Crime|Horror|Thriller	The Purge
Sci-Fi|Thriller	Philomena
Action|Horror|Thriller	Taken 3
Adventure|Comedy|Sci-Fi	Trainwreck
Drama	The Visit
Action|Adventure|Fantasy|Sci-Fi	The Martian

Gross	IMDB Score	movie_title
$137,387,272	8	The Exorcist
$241,437,427	7.2	Grease
$334,185,206	8.8	Star Wars: Episode V - The Empire Strikes Back
$37,707,719	7.1	WarGames
$197,171,806	8.3	Indiana Jones and the Last Crusade
$181,360,000	7.6	A Few Good Men
$356,784,000	8.1	Jurassic Park
$241,688,385	6.3	Twister
$380,262,555	7.6	Star Wars: Episode III - Revenge of the Sith
$141,340,178	7.6	The Others
$217,536,138	6.6	My Big Fat Greek Wedding
$261,437,578	8	The Incredibles
$64,423,650	6.6	The Da Vinci Code
$760,505,847	7.9	Avatar
$32,721,635	6.6	The Last House on the Left
$623,279,547	8.1	The Avengers
$204,565,000	6.5	Alice in Wonderland
$115,646,235	8	District 9
$162,001,186	6	Grown Ups
$363,024,263	8.1	Deadpool
$96,471,845	7.5	The Conjuring
$201,148,159	7	Cinderella
$368,049,635	7.5	Despicable Me 2
$79,568,000	5.7	The Purge
$290,158,751	7.6	Philomena
$89,253,340	6	Taken 3
$110,008,260	6.3	Trainwreck
$65,069,140	6.2	The Visit
$228,430,993	8.1	The Martian



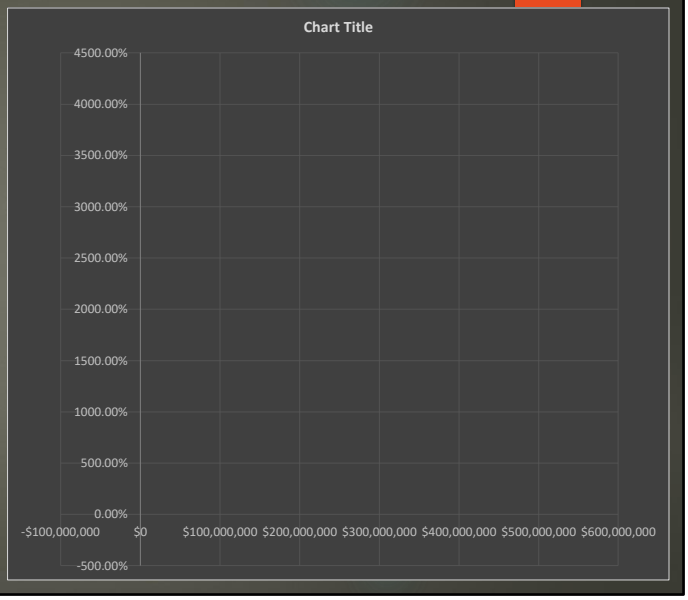
In this EDA, I analyzed the various elements concerning the relationship between the IMDB ratings and box office revenues in answering the question: What does the IMDB scores and gross revenues say about certain types of films? For instance, of the two highest grossing films, both of Avatar and Avengers from slide 6 had an IMDB rating of 7.9 and 8.1, respectively. Now as the scatterplot displays those two variables, both are outliers compared to the rest of the other distributions of the two variables which had good results in their respective genres. Also, in terms of using the ratings and gross revenues as a method to predict moviegoers film preferences, I analyzed that higher ratings does not necessarily mean a more profitable box office success. This could mean that the audiences prefer a certain genre, other user ratings from IMDB, marketing and reviews, who the director is, and who the actors/actresses playing in the films, etc. In conclusion, the scatterplot can be used as a decision-making strategy by Microsoft to accurately predict what kinds of films to produce.



For this slide, I chose a line graph to analyze the relationship among the duration and the budgeting for the top 30 grossing films. The question is, How does the duration and the budget play a significant factor in the types of movies to produce? Other than the movie, The Exorcist, I realized that the distributions for the budget costs in making the movie increases as the length of the film is longer. The line also shows that both Avatar and The Avenger films had the two highest budget spending in its production as the length of those films are the longest durations on the lists. Through this data analysis, other variable factors does play a major role in gaining additional information. For example, the venue(s), types of actors/actresses, directors, equipment, machineries, wardrobe, payroll, etc. are attributes that increases the cost of making a movie. So, with producing movies, it is significant to include these factors along with the moviegoer's preference, genre, return of investments, etc.

Findings

Gross	Budget	Profit	ROI
$137,387,272	$8,000,000	$129,387,272	1617.34%
$241,437,427	$6,000,000	$235,437,427	3923.96%
$334,185,206	$18,000,000	$316,185,206	1756.58%
$37,707,719	$12,000,000	$25,707,719	214.23%
$197,171,806	$48,000,000	$149,171,806	310.77%
$181,360,000	$40,000,000	$141,360,000	353.40%
$356,784,000	$63,000,000	$293,784,000	466.32%
$241,688,385	$92,000,000	$149,688,385	162.70%
$380,262,555	$113,000,000	$267,262,555	236.52%
$141,340,178	$17,000,000	$124,340,178	731.41%
$217,536,138	$5,000,000	$212,536,138	4250.72%
$261,437,578	$92,000,000	$169,437,578	184.17%
$64,423,650	$125,000,000	-$60,576,350	-48.46%
$760,505,847	$237,000,000	$523,505,847	220.89%
$32,721,635	$15,000,000	$17,721,635	118.14%
$623,279,547	$220,000,000	$403,279,547	183.31%
$204,565,000	$200,000,000	$4,565,000	2.28%
$115,646,235	$30,000,000	$85,646,235	285.49%
$162,001,186	$80,000,000	$82,001,186	102.50%
$363,024,263	$58,000,000	$305,024,263	525.90%
$96,471,845	$20,000,000	$76,471,845	382.36%
$201,148,159	$95,000,000	$106,148,159	111.73%
$368,049,635	$76,000,000	$292,049,635	384.28%
$79,568,000	$3,000,000	$76,568,000	2552.27%
$290,158,751	$12,000,000	$278,158,751	2317.99%
$89,253,340	$48,000,000	$41,253,340	85.94%
$110,008,260	$35,000,000	$75,008,260	214.31%
$65,069,140	$5,000,000	$60,069,140	1201.38%
$228,430,993	$108,000,000	$120,430,993	111.51%



For this slide, I used the scatterplot to investigate on the relationships amongst the gross revenue, budget, profit, and return on investments. By analyzing these items, it provided additional feedbacks on the what types of movies are feasible and profitable to produce. In addition, it helped to answer the question, Does the directors or actors/actresses impact the return on investments (ROI) from the gross revenues, budget, and profits from the box office outcomes? A closer look at the scatterplot chart shows a positive correlation among the distribution of the variables between box office revenue, budgeting, profits, and ROI. The outliers provides a scope variant regarding the genres of movies such as how long the film was in theaters. Also, this scatterplot does display that the directors and actors/actresses impact the return on investments and profits. For example, for the lists of movies in the dataset, shows that the correlation among top grossing films such as Avatar and Avengers had two out of the ten lowest ROI percentages. This is because they are the two movies with the highest budget costs, however, both are still the highest in profits generated. This, most likely, due to

the directors and cast members in those respective films. These categorical features tells us a variety of factors relating to the potential business values in making movies. This has to do with how the profits and ROI will progress the business's bottom line, balance sheet asset items such as intangible assets under long-term assets section of the balance sheet. Moreover, profits from box office revenues will allow Microsoft to reinvest into its business operations to the stakeholders, shareholders, Xbox gaming entertainment, movie streaming services, etc. Also, this scatterplot does display that the directors and actors/actresses impact the return on investments and profits. For example, for the lists of movies in the dataset, shows that the correlation among top grossing films such as Avatar and Avengers had two out of the ten lowest ROI percentages. Lastly, this chart including the ones above, helps Microsoft in decision-making strategies under this business segment as well as knowing what genres of films to produce.

Future Work

For future work, I would explore other aspects of this project if there was more time. For instance, I would research on international box office revenues. Also, that deals with the ratings, which countries where the film played, how the movie performed in terms of profits, etc.

Conclusion

This Photo by Unknown Author is licensed under CC BY-NC-ND

Throughout this new project endeavor, three recommendations has been concluded. This exploratory data analysis conclusion would assist in Microsoft's new opening of a film production studio. In addition, these business recommendations were made possible by a few statistical models and research.

The first recommendation comes from the answer to the first question both the IMDB scores and box office revenues are relevant factors in Microsoft's endeavor. For example, predicting and researching about the preferences and ratings by moviegoers will assist in making strategic decisions about which genre of films to make. Lastly, I recommend that Microsoft gather more data on its own client base from the Xbox gaming console/game development divisions and movie streaming services in order to decide what genres of films will meet its business objectives.

Secondly, in this recommendation, I analyzed the data from the second question. The factors here include budgeting costs, durations, and gross revenues play a major role in the success or failure of a movie. In the production of a film, capital must be distrusted to individuals prior to filming. For instance, salary negotiations, wardrobes, filming locations, equipment, machineries, etc. So, even prior to filming scenes and plots, the costs is already lowering the budget to make the movie. Now, in terms of the length of time the movie runs, the time it takes for a movie to finish filming is also a major impact on budgeting as well. This deals with factors such as the scope of the movie script, who the actors/actresses playing in the film, and the director. So, in this business recommendation, I insist that Microsoft should strategize on the screenplay movie script, where the film will be made, how the length of the movie should coincide with the plots in the movie script, and most of all the genre of movie to produce.

Lastly, in this final business recommendation, I utilized the answer from the third question. This entails the variable factors such as box office revenues, budget, profits, and return on investments. Those features are of great importance with regards to Microsoft's business operations and financials as well. For example, in terms of ROI, there is a wider scope regarding Microsoft's annual reports on its business segments. On the other hand, regarding the gross, budget, and profits, those are included in the income statement of Microsoft's financial statements. Whether or not this new business endeavor adds growth or loss to Microsoft's overall business objectives, it has to do with its strategies used in building what it already has established. In saying that, I recommend that Microsoft utilize what they have already establish in its business segments such entertainment, artificial intelligence, services, etc. to form this new film studio as an entry point into the movie industry. In addition, Microsoft must incorporate its financials as well other research to create a buffer from risk and potential losses as this is new space for Microsoft to enter.

Thank You

Lastly, I want to say, Thank You!!! I appreciate all of you taking the time to be here for the new launch of Microsoft's film production studio division. This new endeavor is due to our continuing success of the Xbox gaming console and game development division, movie streaming services, etc. Today's presentation consists of data that provides information on the genres of films that moviegoers are most likely to watch. In addition, the data is also used for recommending which types of movies to produce. Once again, thank you for time.