Database Management

Class C, Team 5

Carl Hümbs   500927977
Bart Remmers   500777595
Dide Uzun   500923796
Pip Velthuijs   500831930

Table of Contents

1. Introduction

The movie industry is a significant player in the worldwide economy, and certainly one of the most entertaining. Globally contributing a little over 75 billion U.S. dollars, employing almost 390 thousand people in more than 55 thousand businesses (Statista Research Department, 2023). To compete in this business, it requires great investments and great understanding of trends and influencing factors. It was one of the most steadily growing businesses each year, until the COVID-19 outbreak in 2020, as depicted in figure 1 (Li, 2023) . However, in 2021 the global box office revenue grew again by over 80%, making it also quite the resilient industry (Statista Research Department, 2023).
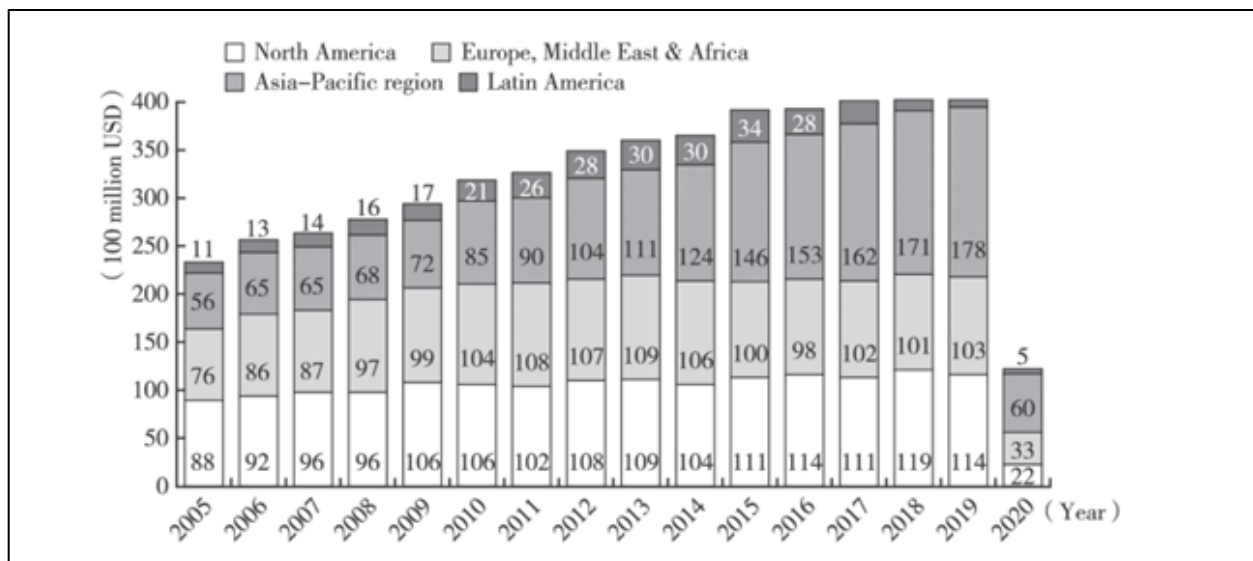

Figure 1: Growth until 2020 (Li, 2023)

As filmmakers recently have taken a big hit from the pandemic, strategic decision making has become more important than ever. Hollywood, often characterized as "The land of hunch and the wild guess," presents an intriguing landscape for researchers to find factors that lead to success, particularly those which one can control (Sharda, 2009).

This research aims to uncover the determinants of successful box office revenue, offering a comparative analysis of various factors, including cast, rating, reviews, and genre to seek the most influential factor contributing to a successful box office.

1.1 Research Question

The main research question this report will answer is:

'What are the key determinants of successful worldwide box office sales for movies?'

The answer to this research question provides producers and internal stakeholders with valuable insights into the (controllable) determinants of box office success and hopefully provides a significant strategic advantage over competitors who overlook its importance. The insights can be incorporated into the company's sales strategy or used to benchmark the competitive position.

1.2 Sub-Questions

To answer the research question, four sub-questions are formulated:

1. To what extent does graphic content (violence, explicit language, and explicit sex) influence a movie's performance in terms of box office sales?
2. To what extent do reviews contribute to the worldwide box office sales according to the LIWC parameters?
3. How does the genre of a movie influence its worldwide box office sales?
4. In which way does the budget influence movies worldwide box office sales?

2. Literature Review

2.1 Graphic content and movie ratings

"The question is, do you go for the PG-13, or do you go for what it should be, which is R?

Financially it makes quite a difference…"

- Ridley Scott

Movie ratings and their definition

Movies get rated based on the level op graphic content they include, this rating system was first introduced in 1968 by the Motion Picture Association. The rating system is created to help parents make choices of what they're children should and shouldn't watch. When it comes to the box office (which is based on ticket sales for movies), there are 5 distinct types of

ratings. The ratings are last revised in 2013, this is the exact definition the Motion Picture Association gave for each rating:

- G – General Audiences: All ages admitted. Nothing that would offend parents for viewing by children. Introduced in 1968.

- PG – Parental Guidance Suggested: Some material may not be suitable for children. Parents are urged to give "parental guidance". May contain some material parents might not like for their young children. Introduced in 1968.

- PG-13 – Parents Strongly Cautioned: Some material may be inappropriate for children under the age of 13. Parents are urged to be cautious. Some material may be inappropriate for pre-teenagers. Introduced in 1984.

- R – Restricted: Under 17 requires accompanying parent or adult guardian. Parents are urged to learn more about the film before taking their young children with them to see it. Introduced in 1968.

- NC-17 – Adults Only: No one 17 and under is admitted. Clearly adult content. Children are not admitted. Introduced in 1990.

The ratings 'G' and 'PG' are considered family friendly. The ratings PG-13', 'R', and 'NC-17' are considered graphic. Movies without a rating receive the label 'Not Rated'. (Film Ratings, Informing Parents Since 1968, Copyright 2023)

The effect of graphic content and corresponding ratings on a movie's success according to researchers
Multiple researchers like Daniel Ellis, Brooke Conaway, Pedro Garcia-del-Barrio, and Hugo Zarco have looked into the effect of explicit content in movies and the MPAA ratings that come with it, in relation to box office revenues.

Ellis & Conaway rated R movies have a significantly lower box office (between 15 and 34 million dollars) in comparison to the less restricted PG-13 rating. An interesting fact is that gun violence in PG-13 movies has tripled since 1985. Showing a trend in graphic content. It's added that, as previous literature showed, Oscar nominations, star power (except one-hit wonders), and average professional critical reviews have a very positive effect on box office revenues. (Daniel Ellis, 2015)

Garcia-del-Barrio & Zarco focus specifically on the content, working from the usual claim that certain content like violence and/or sex is essential to meet the consumers' taste. They analyzed the link between R-ratings and box office revenues and explored if offensive content is an effective ingredient for a successful box office revenue. The paper concludes that there is a statistically significant positive relationship between violence and a successful box office but notes that successful box office revenue is statistically associated with a lower concentration of sex and profanity. Most likely in order to prevent any more hesitancy about going to the theatre. (Zarco, 2017)

Looking a little more into Ellis & Conaway's finding of more violence in PG-13 movies (in 2015 thrice as much since 1985), it seems yesterday's rated-R is today's PG-13, making those MPAA ratings rather incoherent. The Federal Trade Commission's "Marketing Violent Entertainment to Children" shows that the MPAA has adopted stricter rules for ad placement, type of violence, and other content shown. But no new rules for targeting teens and younger children for R-rated and PG-13-rated films. There seems to be a conscious decision to make graphic content quite accessible to a big audience. (Afra, 2016)

2.2 Contribution of reviews on box office sales

Internet reviews created by users play a significant role as an information resource for consumers. They serve a crucial function by reducing uncertainty regarding various products and aiding consumers in evaluating the quality of those products. (Koh, N. S., Hu, N., & Clemons, E. K., 2010) It is widely recognized that consumers often turn to ratings and reviews as a primary decision-making factor when selecting their next movie to watch. (K Topal, G Ozsoyoglu, 2016) Understanding the characteristics that contribute to the efficacy of product evaluations is important from both a management and theoretical standpoint, because these reviews have a significant impact on potential consumers and, as a result, sales. (Ahmad, S. N., & Laroche, M., 2015)

Consumers who were exposed to high-quality reviews, as opposed to low-quality ones, tended to assess products more favorably consequently, this positive evaluation resulted in a stronger intention to make a purchase. (Lee, E. J., & Shin, S. Y., 2014) Review quality is frequently defined in terms of the persuasive power of the message and is commonly evaluated based on factors such as its relevance, timeliness, accuracy, and comprehensiveness. (Cheung, C. M., & Thadani, D. R., 2012)Review quality is frequently defined in terms of the persuasive power of the message and is commonly evaluated based on factors

such as its relevance, timeliness, accuracy, and comprehensiveness. (Cheung, C. M., & Thadani, D. R., 2012)

A more recent study revealed that review attributes, including the number of reviews, the rating of reviews, and the length of reviews, interact with the concept of review helpfulness and impact box office performance. The study's findings show that in circumstances where a review is seen as helpful, the volume and length of reviews have a greater impact on a film's box office performance. (Lee, S., & Choeh, J. Y., 2018)

However, the linear link between word count and review helpfulness may only be accurate up to a point before turning into a non-linear relationship or showing no relationship. When the word count exceeds a particular threshold, the impact on review usefulness may begin to diminish. (Lee, S., & Choeh, J. Y., 2018)

Review ratings exhibit a positive correlation with both the frequency and consistency of reviews. Several variables can be related to this phenomenon. Firstly, reviewers are frequently motivated to offer their thoughts when they are particularly pleased with products, driven by a desire to aid fellow consumers or support businesses. Secondly, the motivation to express constructive criticism positively impacts the volume of reviews. In simple terms, satisfied reviewers tend to contribute evaluations to maintain their positive emotional state. (Chen, C. C., Chen, X. P., & Huang, S., 2013)

2.3 Influence of Movie Genres on Box Office Performance

The term 'genre' is borrowed from the French language and meaning 'type' or 'kind' (Neale, 2005), represents a common cultural consensus: "genre is what we collectively believe it to be" (Tudor, 1976). Genres serve as a cognitive tool that simplifies perception and evaluation by grouping similar works together, aiding audiences in forming expectations about artistic products (Cerulo, 2010). Over the past three decades, genres have held a significant position in the realm of the cinema industry, playing an essential role in the description, navigation, and discovery of movies (Neale, 2005). For film producers, genres provide a template for decision-making, and their relative stability enables studios to standardize production practices (Perretti, 2007). Meanwhile, for viewers, the consistency within a genre provides familiarity with established rules for identifying that genre. This familiarity allows them to infer recurring themes and narrative structures without having to watch the movie in advance. Given that viewers employ genre as a piece of information to make choices, it is highly likely that a movie will meet their expectations and provide satisfaction (Hung & Guan, 2020).

In the context of movie genres and their influence on box office performance, research presents a mixed picture. Some studies suggest that genre serves as a driver for initial movie viewership (Chang, Nam, Chan-Olmsted, & Kim, 2016, p. 179). Conversely, other research, conducted by Collins et al. (2002) indicates that the influence of genre on box office success remains uncertain. Their findings suggest that genre should be considered more as a controlling factor than a predictive variable (Collins, Hand, & Snell, 2002). Furthermore, Ainslie et al. (2005) propose that producers should avoid releasing a movie concurrently with others of the same genre, as this can have adverse effects on box office performance, both in the short and long term (Ainslie , Drèze, & Zufryden, 2005).

When examining the type of genres, several key findings have emerged from different studies. Two studies analyzed the frequency of various genres among the top 20 grossing films at the US box office between 1967 and 2008. These studies identified five genres – action, adventure, comedy, thriller, and drama – as the most frequently occurring. Interestingly, all but the drama genre exhibited an increase in frequency within the highest box office rankings over this period (Lu, Waterman, & Zhaoxu Yan, 2005; Ji & Waterman, 2010).

A little more recent analysis, covering the period from 1991 to 2010, highlights changing genre trends. The study revealed a narrowing range of genres at the highest rankings, specifically focusing on fantasy/science fiction, crime/thriller, and drama films. Additionally, there were marginal increases in the presence of action/adventure and family films.

Surprisingly, another study by Garcia-del-Barrio and Zarco (2017) suggests that adventure films, when comparing to dramas, yield significant additional revenues from theatrical audiences. These findings align with Statista's research on the most popular movie genres in the United States and Canada. Adventure films recorded a box office revenue of more than 66 billion U.S. dollars. Action movies followed with more than 55 billion dollars in revenue (Statista Research Department, 2023). In contrast, western films generated comparatively lower revenues (Garcia-del-Barrio & Zarco, 2017).

Given the mixed findings regarding the influence of movie genres on box office performance and the dominance of genre-related research, the following hypothesis is proposed: The adventure genre has a positive influence on box office worldwide revenue.

7

2.4 Budget as influence or predictor for success

A movie budget is the outline of the costs for all the steps of making and distributing a movie. It covers filmmaking, pre-production and post-production. Including: salaries, theatrical properties (props), wardrobe, food, location costs, and permits. Generally speaking, an average movie costs 50-100 million dollars, up to 30 million dollars is considered low budget. (Lloyd Kaufman, 2012)

Budget is arguably one of the most important factors of making a movie, as it determines how much real-world items you can use in your movie. But it is also said that creativity hides within constraints. So, does investing more money equate to more revenue? Fortunately, many researchers have addressed this cliché since the early days of the movie business. About twenty years ago, John, Ravid and Sunder concluded that big budgets really have no impact on revenues at all, but nevertheless found that the average budget in Hollywood increases every year. (Sunder, 2000) (S. Abraham Ravid, 1999) A few years later though, some of the same researchers found out that the budget does influence the critical reviews, which themselves influence the box office revenues. (Suman Basuroy, 2003)

Yet the trend of growing budgets they pointed out all these years ago still applied in the years after, at that time (2000) the average budget was around 38 million dollars, ten years later it was just over 60 million dollars and by 2021 the average movie budget was 87 million dollars. (Average Movie Budget, Copyright   2023)

These numbers seem conflicting, why do all the budgets grow if they do not do much for revenue? Schuyler Moore comes up with a very simple explanation: when the exact same movie is made 2x, once for 10 million and once for 50 million, the latter will be sold for 5x the price as the first. As the budget increases, so does the price a distributor is willing to pay for it. (Moore, 2019) This shallow approach from the distributors makes it, weirdly enough, attractive to spend more money on a movie.

2.5 Hypotheses

For each of the four sub-questions, hypotheses are developed based on theories from the literature review:

1. **To what extent does graphic content (violence, explicit language, and explicit sex) influence a movie's performance in terms of box office sales?**

SQ1 Null Hypothesis 1:

Fail to reject the null, there is no evidence to support a statistically significant difference in the box office sales between movies with graphic content and movies with family-friendly content.

SQ1 Alternative Hypothesis 1: Reject the null,

there is a statistically significant difference in the box office sales between movies with graphic content and movies with family-friendly content.

SQ1 Null Hypothesis 2:

Fail to reject the null, there is a statistically significant difference between the performance in box office sales for PG-13 as opposed to other graphic ratings.

SQ1 Alternative Hypothesis 2:

Reject the null, there is a statistically significant difference between the performance in box office sales for PG-13 as opposed to other graphic ratings.

2. **To what extent do reviews contribute to the worldwide box office sales according to the LIWC parameters?**

   SQ2 Null Hypothesis 1:

   There is no significant relationship between the average word count of reviews and worldwide box office sales for movies.

   SQ2 Alternative Hypothesis 1:

   There is a strong positive linear relationship between the average word count of reviews and worldwide box office sales for movies.

   SQ2 Null Hypothesis 2:

   Review sentiment, as measured by LIWC parameters such as "idvscore", "affect", "posemo", and "negemo" have no significant relationship with worldwide box office sales.

   SQ2 Alternative Hypothesis 2:

   Review sentiment, as measured by LIWC parameters such as "idvscore", "affect", "posemo", and "negemo" significantly correlates with worldwide box office sales.

3. **How does the genre of a movie influence its worldwide box office sales?**

   SQ3 Null Hypothesis 1:

The adventure genre has no significant positive influence on the worldwide box office sales.

SQ3 Alternative Hypothesis 1:

There adventure genre has a significant positive influence on the worldwide box office sales.

SQ3 Null Hypothesis 2:

There is no significant difference in the mean of worldwide box office sales between movies of the "adult" genre and movies of all other genres.

SQ3 Alternative Hypothesis 2:

There is a significant difference in the mean worldwide box office sales between movies of the "adult" genre and movies of all other genres.

4. **In which way does the budget influence movies worldwide box office sales?**

SQ4 Null Hypothesis:

The production budget of movie productions does not have a significant positive impact on box office revenue.

SQ4 Alternative Hypothesis: The production budget of movie productions contributes significantly to higher box office results.

3. Methodology

A combination of primary and secondary research methods was used to determine the determinants of box office movie sales. In order to acquire the necessary information, the research initiated a literature review (secondary research) based on the sub questions. This aimed to establish theoretical depth and gather scientific background information about the movie industry.

3.1 Hypotheses Testing

To test the hypotheses, the research employed SQL and Python programs. SQL, which stands for Structured Query Language, is a standardized programming language utilized for managing relational databases and executing various operations on the data within them (Loshin & Sirkin, 2022). Specifically, SQL was used for creating and manipulating necessary database tables.

Python, renowned for its high-level, object-oriented programming capabilities, is recognized for its interpreted nature and dynamic semantics (Python Software Foundation, sd). In this research, Python was utilized to encapsulate each database.

For each sub question, the following methodology was followed, in order to test our hypotheses:

Sub question	Type of research	Research method
SQ1: How important is the inclusion of (graphic) content for a high box office revenue?	Primary and secondary	Sub question 1 underwent investigation through a literature review, which included academic studies, theories, reports, and books retrieved from the HvA database and Google Scholar. Additionally, primary research was conducted, involving data analysis using SQL and Python.
SQ2: To what extent do reviews contribute to the worldwide box office sales according to the LIWC parameters?	Primary and secondary	Sub question 2: A correlation analysis was conducted between movie reviews and worldwide box office sales using Python and postgreSQL, with pandas library for data manipulation and psycopg2 for database connectivity. Data was sourced from expert and user reviews, and descriptive statistics were used to find average movie sentiments. The significance of relationships between review

|		attributes (e.g., sentiment scores, word count) and box office performance was tested by using correlation matrix.
SQ3: How does the genre of a movie influence its worldwide box office revenue?	Primary and secondary	Sub question 3 utilized both a literature review and data analysis through SQL and Python.
SQ4: To what extend does the budget influence a movies worldwide box office revenue?	Primary and secondary	To address the sub question 4, a linear regression analysis has been conducted, using python and SQL to investigate whether there is a statistically significant positive relationship between production budget and worldwide box office revenue. In addition, a literature review was done.

The hypotheses formulated during the literature review were subsequently answered using four Excel data files: Sales, MetaClean, User reviews, Expert reviews. These data files were constructed based on information extracted from numbers.com and metacritic.com. The MetaClean file contained information about the movies themselves, while the Sales data file included details such as movie runtime, production budget, release year, and box office worldwide revenue. The user reviews and expert reviews files included data points such as review scores, thumbs-up ratings, word count, and emotional analysis.

Before commencing data analysis, the Excel data files underwent diligent preparation and cleaning. This process involved rigorous checks for missing values, duplicate entries, and outliers. Missing values were addressed by deleting rows based on the extent of missing data and its impact on the subsequent analysis. For instance, rows with missing values in columns such as movie titles and worldwide box office were removed. Subsequently, the column names underwent standardization to ensure uniformity in data representation. Next in the process was column selection, where only relevant columns believed to influence box office sales were chosen. The criteria for variable selection were based on their potential as determinants of box office revenues. To connect the Datafile Expert Reviews and User Reviews, a Movie Name attribute was derived from the URL. In the Datafile Sales, a Movie ID was generated by collecting and indexing all movie names to ensure unique identifiers. Similarly, the Genre table was created by gathering and indexing all available genre names while removing duplicates.

After all the cleaning was done, a new database was created in PostgreSQL and all the files were imported into this database. By using the psycopg2 library, a Python connection to this database has been made and further analysis for the research questions is conducted in Python.

3.2 Entity Relationship Diagram (ERD)

An Entity Relationship Diagram (ERD) was developed to visualize the relationships between the various entities involved in the analysis. The ERD helps in understanding the interaction and dependencies between these entities:
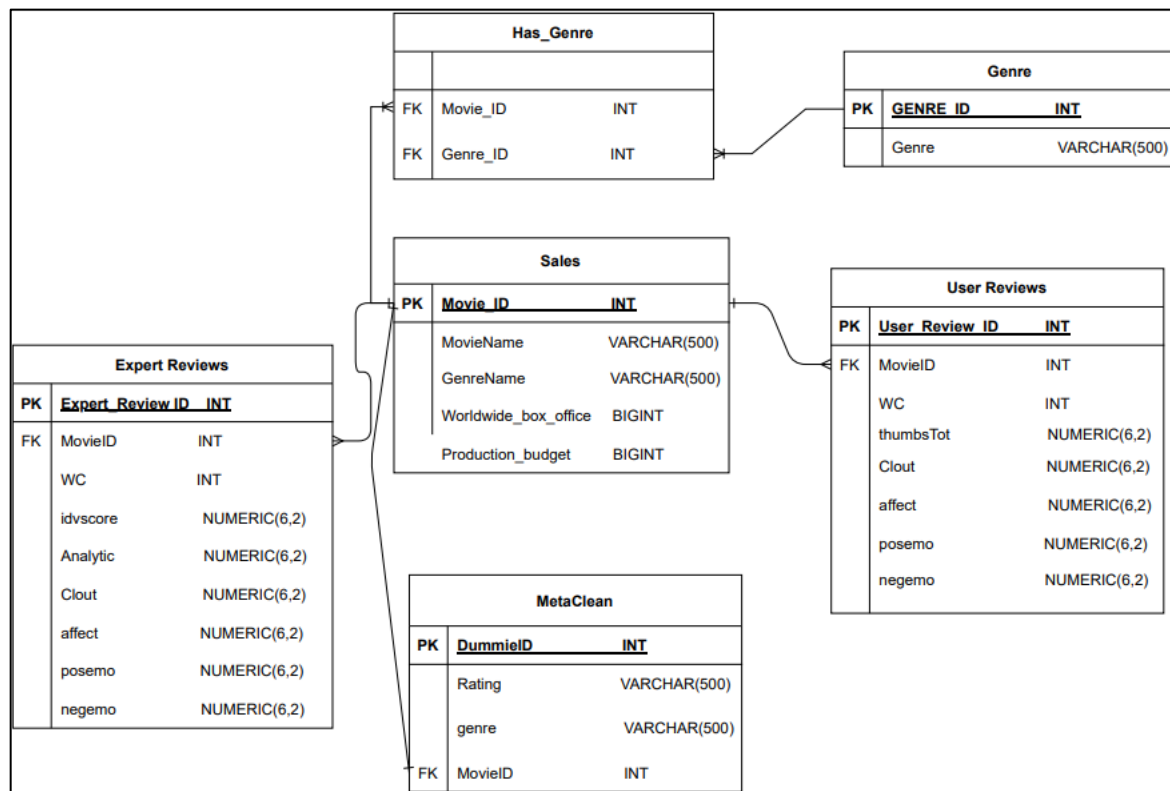
Entity relationship diagram (ERD), representing the structure of the database used in this project.

3.2.1 Cardinality ratios

The next cardinality ratios are used in the ERD. First, to connect the Sales entity with both the Expert Review and User Review entities, a one to many (1:n) cardinality ratio is used. This is because one movie (from the Sales entity) can have multiple reviews from experts or users. This also occurs in the Genre table, since one movie (from the sales entity) can have multiple genres. So this is also a one to many relationship (1:n).

There is also a one to one (1:1) cardinality ratio used when linking the Sales entity to the MetaClean entity. Since a movie (the one from the sales entity) is only able to receive one rating.

3.2.2 Entities

All the entities used in this ERD are classified as weak. ExpertReviews, UserReviews, MetaScore, and Genre do not have a significant effect when the movie is not present in the sales entity. Due to this fact, the relationship is weak because it is not assumed that every movie occurs in the ExpertReviews, UserReviews, MetaScore, and Genre.

This ERD consists of solely partial participation(s), as occurrence in the sales entity does not give any guarantee of presence in any of the old entities.

Bibliography

Afra, K. (2016). PG-13, Ratings Creep, and the Legacy of Screen Violence: The MPAA Responds to the FTC's "Marketing Violent Entertainment to Children" (2000–2009). *Cinema Journal Vol. 55, Number 3*, 40-64.

Ahmad, S. N., & Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Information Management*, 76-111.

Ainslie , A., Drèze, X., & Zufryden, F. (2005). Modeling Movie Life Cycles and Market Share. *Marketing Science*, pp. 508-517. Retrieved September 5, 2023, from https://www.jstor.org/stable/40056978

Average Movie Budget. (Copyright 2023). Retrieved from Filmlifestyle: https://filmlifestyle.com/average-movie-budget/

Cerulo, K. A. (2010, April). Mining the intersections of cognitive sociology and neuroscience. *Poetics*. Retrieved September 5, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0304422X0900062X

Chang, B.-H., Nam, S.-H., Chan-Olmsted, S. M., & Kim, H. (2016, October 1). Repeat Consumption of Media Goods: Examining Factors Affecting. *JOURNAL OF MEDIA ECONOMICS*, pp. 167–180. Retrieved September 5, 2023, from http://dx.doi.org/10.1080/08997764.2016.1239622

Chen, C. C., Chen, X. P., & Huang, S. (2013). Chinese guanxi: An integrative review and new directions for future research. *Management and Organization Review*, 167-207.

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems*, 461-470.

Collins, A., Hand, C., & Snell, M. (2002). *What Makes a Blockbuster? Economic Analysis of Film.* Managerial and Decision Economics. doi:10.1002/mde.1069

Daniel Ellis, B. C. (2015). Do MPAA Ratings Affect Box Office Revenues?

Film Ratings, Informing Parents Since 1968. (Copyright 2023). Retrieved from MOTION PICTURE ASSOCIATION, INC.: https://www.motionpictures.org/film-ratings/

Garcia-del-Barrio, P., & Zarco, H. (2017). Domoviecontentsinfluencebox-officerevenues? *Applied Economics*, pp. 1679–1688. Retrieved September 5, 2023, from http://dx.doi.org/10.1080/00036846.2016.1223828

Hung, Y.-C., & Guan, C. (2020). *Winning box office with the right movie synopsis.* Singapore: European Journal of Marketing. doi:10.1108/EJM-01-2019-0096

Ji, S. W., & Waterman, D. (2010, December). Production Technology and Trends in Movie Content: An Empirical Study. p. 25. Retrieved September 5, 2023, from https://www.researchgate.net/publication/228448250_Production_Technology_and_Trends_in__Movie_Content_An_Empirical_Study

K Topal, G Ozsoyoglu. (2016). Movie review analysis: Emotion analysis of IMDb movie reviews. *Social Networks Analysis and Mining*, 18.

Karniouchina. (2011). Impact of star movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, pp. 62-74. Retrieved September 2023, from https://doi.org/10.1016/j.ijresmar.2010.10.001

Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic commerce research and applications*, 374-385.

Lee, E. J., & Shin, S. Y. (2014). When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. *Computers in human behavior*, 356-366.

Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue. *Management Decision*, 849-866.

Li, Q. W. (2023). *The Global Film Market Transformation in the Post-Pandemic Era.* Oxfordshire: Taylor & Francis.

Lloyd Kaufman, A. W. (2012). *Produce Your Own Damn Movie!* New York City: St. Martin's Press.

Loshin, P., & Sirkin, J. (2022, February). What is Structured Query Language (SQL)? *Data Management*. Retrieved September 7, 2023, from https://www.techtarget.com/searchdatamanagement/definition/SQL#:~:text=Structured%20Query%20Language%20(SQL)%20is,on%20the%20data%20in%20them.

Lu, W., Waterman, D., & Zhaoxu Yan, M. (2005, September 8). Changing Markets, New Technology, and Violent Content: AnEconomic Study of Motion Picture Genre Trends. p. 32. Retrieved September 5, 2023, from https://www.researchgate.net/publication/228675899_Changing_markets_new_technology_and_violent_content_An_economic_study_of_motion_picture_genre_trends

Moore, S. (2019, April 13). *The importance of film budgets*. Retrieved from Forbes: https://www.forbes.com/sites/schuylermoore/2019/04/13/the-importance-of-film-budgets/

Neale, S. (2005). Part I Genre: Definitions and concepts. In *Genre and Hollywood* (Vol. 1, p. 344). Routledge. doi:10.4324/9780203980781-2

Nelson, R. A., & Glotfelty, R. (2012). Movie stars and box office revenues: An empirical analysis. *Journal of Cultural Economics*, 141-166.

Perretti, F. (2007). Mixing genres and matching people: a study in innovation and team composition in Hollywood. *Journal of Organizational Behavior*, p. 24. doi:10.1002/job.464

Python Software Foundation. (n.d.). *What is Python? Executive Summary*. Retrieved September 7, 2023, from Python.org: https://www.python.org/doc/essays/blurb/

Ravid. (1999). Information, Blockbusters, and Stars: A Study of the Film Industry. *The Journal of Business* , pp. 463-492. Retrieved September 2023, from https://doi.org/10.1086/209624

S. Abraham Ravid. (1999). Information, Blockbusters, and Stars: A study of the Film Industry. *Journal of Business, 72*, 463–492.

Statista Research Department. (2023). *Most popular movie genres in the United States and Canada between 1995 to 2023, by total box office revenue.* Statista. Retrieved September 5, 2023, from https://www.statista.com/statistics/188658/movie-genres-in-north-america-by-box-office-revenue-since-1995/

Suman Basuroy, S. C. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, 103–117.

Sunder, K. J. (2000). The Role of Termination in Employment Contracts: Theory and Evidence from Film Directors' Careers. *SSRN Electronic Journal*.

Tudor, A. (1976). Genre and critical methodology. In *Movies and Methods.* University of California Press. Retrieved September 5, 2023

Zarco, P. G.-d.-B. (2017). Do movie contents influence box-office revenues? *APPLIED ECONOMICS VOL. 49*, 1679–1688.

Appendix

This is the GitHub repository of Team 5, which includes:

- The Python code that has been used to clean the original data, in order to use it in the database.
- The Python Encapsulation, used to access the created database, as well as all the code used to analyze the data, and also the results of these analyses and hypothesis testing.
- This introduction document
- ERD diagram
- SQL queries used to create the database

Here you can find the Trello Board which has been used by the Team to organize and manage all tasks and steps of the project.

ERD