UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

Group Project

Data Science Programming II (BSD2223)
Semester II, 2023/2024

Movie Marvericks

Group 7

STUDENT ID	NAME	SECTION
SD22002	ALMIRA DAMIA BINTI SYAHNIZAM (Leader)	
SD22005	NURKHAIRUL IZZATI BINTI MOHD SALLEHAN	
SD22016	SRI SHAMNEE SAI A/P RAJAH	02G
SD22051	MOHAMMAD BASHARIL AIMAN BIN MOHAMMAD BAKHTIAR	

TABLE OF CONTENTS

1.0   Introduction

1.1 Background of Study

In the movie industry, entrepreneurship is the application of creative entrepreneurial ideas to different aspects of distribution, promotion, and production. This gives industry professionals and filmmakers the ability to manage their careers and adjust to the ever-changing market landscape. The importance of this approach is growing as a result of advancements in technology and shifts in consumer behavior. Thanks to digital channels, filmmakers now enjoy unmatched access to global audiences and creative flexibility. As a result, entrepreneurship in the film business provides financial security and access to several revenue streams. Comprehending this notion is crucial for prospective filmmakers and experts in the field, as it offers valuable perspectives on the dynamic nature of the business and avenues for achievement. Moreover, entrepreneurship propels innovation in the film industry and sets its trajectory for the twenty-first century.

The project's dataset consists of a thorough compilation of 6,820 films that were released within a 35-year period, from 1986 to 2020. The financial, production, and reception details for every movie are abundant in this dataset, making it valuable for predicting trends and patterns in the cinema business. Through the use of this data, a comprehensive knowledge of the movie industry is provided by illuminating the various facets that influence the growth and commercial success of films over a 35-year span.

1.2 Objectives

1. Evaluate the elements that lead to a movie's box office or critical success

2. Analyze genre trends over time to identify emerging genres and opportunities for innovation in storytelling.

3. Explore the relationship between movie budgets and box office revenues to identify patterns of financial success.

2.0   Project Description

2.1 Why this project?

This project is chosen for several compelling reasons. Firstly, it aligns with Sustainable Development Goals (SDGs), offering an exploration of how the influential platform of movies can either support or obstruct global sustainability efforts. Understanding how movies relate to the Sustainable Development Goals (SDGs) is crucial for promoting beneficial social, economic, and environmental effects since movies have a significant influence on cultural norms and behaviors. Additionally, the project offers a chance for in-depth study, drawing on a strong dataset covering thirty years to explore several aspects of the film business, such as audience demographics and production spending.This volume of information makes it easier to get detailed insights into the ways that various aspects of the sector interact with sustainability issues, which promotes a comprehensive comprehension of its impact.

Moreover, the project is both timely and relevant, occurring amidst a period when social and environmental concerns are receiving more attention, technology is developing quickly, and consumer preferences are changing due to the rise of streaming services like Netflix. This context underscores the importance of scrutinizing the evolving dynamics of the movie industry, providing an opportunity to examine these shifts in depth and evaluate their implications for sustainability. Finally, the project holds potential for informing industry practices, offering insights, patterns, and best practices to stakeholders ranging from filmmakers and production companies to policymakers and audiences. It aims to add to the existing conversations on how the film industry can better align with the SDGs and act as a catalyst for good change by offering data-driven recommendations.

2.2 Importance of project and innovative values

In the movie industry, the importance of project management and innovative values cannot be overstated. These elements play crucial roles in ensuring the successful production, distribution, and reception of films. In the film industry, production value refers to the overall quality of the movie as determined by its technical merits, such as set design, special effects, and costumes. A larger budget also usually corresponds with a better production value, allowing filmmakers to produce a more polished and professional result.

Besides, innovative values are also important because they refer to the creative and original ideas that set a film apart from others. These values are essential in the movie industry, as they allow filmmakers to tell unique and engaging stories that captivate audiences. For example, "Parasite" innovative storytelling approach, which combines social commentary and dark comedy, makes it a standout in the thriller genre. Its exploration of themes like class and inequality also adds a layer of relevance and complexity.

Having high production value and innovative values in a movie project is crucial for several reasons. Firstly, it enhances the overall viewing experience for the audience, making the film more engaging and immersive. Secondly, it increases the film's chances of commercial success, as high-quality productions are more likely to attract a larger audience and generate more revenue. Moreover, innovative values in filmmaking encourage creativity and risk-taking, which can lead to the development of new and exciting storytelling techniques, genres, and styles. This, in turn, can help to push the boundaries of the film industry and drive its evolution forward.

3.0   Data Description

This dataset provides information about the movie industry, encompassing 6,820 movies released between 1986 and 2016, with an average of 220 movies per year. It includes 15 columns and 5,422 rows.

| No.	Title	Explanation	Type
1	Name	Name of the movie	Qualitative
2	Rating	Rating of the movie (R, PG-13, etc.)	Qualitative
3	Genre	Main genre of the film	Qualitative
4	Year	Release Year	Qualitative
5	Release	Release Date	Qualitative
6	Score	IMDb score	Qualitative
7	Votes	Count of IMDb user votes	Quantitative
8	Director	The director of the film	Qualitative
9	Writer	The writer of the film	Qualitative
10	Star	The main actor/actress of the film	Qualitative
11	country	Country of origin	Qualitative
12	budget	the budget of a movie. Some movies don't have this, so it appears as 0	Quantitative
13	gross	Revenue of the movie	Quantitative
14	company	The production company	Qualitative
15	runtime	Duration of movie	Quantitative

4.0   Data Preparation

The process of data preparation includes data cleaning, as it is a crucial step in data analysis to ensure the quality and consistency of the dataset. In this process , we first load the necessary libraries, dplyr and tidyr, which provide functions for data manipulation and tidying. We then import the dataset using read.csv, specifying the file path to the CSV file. To understand the dataset, we employ head, str, and summary functions. These commands allow us to preview the first few rows, check the structure, and obtain summary statistics, respectively. This initial exploration helps us identify any immediate issues or inconsistencies within the dataset.

Next, we check for missing values and duplicates. Using the colSums function, we count the number of missing values in each column, which informs us about the extent of missing data that needs to be addressed. To identify duplicate records, we use the duplicated function and filter out these rows. Addressing duplicates is essential to avoid skewed results in our analysis. After identifying the duplicates, we can decide whether to remove them or handle them based on the context and significance of the data, as we did not have any duplicate values we did not do anything about it.

Finally, we clean and transform the dataset by converting columns to appropriate data types using the mutate function from the dplyr package. For instance, we convert the year column to an integer, genre and rating to factors, and score, votes, budget, and gross to numeric types. We also rename the runtime column to duration for better clarity. This transformation ensures that each column is in a suitable format for analysis. After making these adjustments, we save the cleaned dataset to a new CSV file using write.csv, specifying the desired file path. This clean dataset is now ready for further analysis, providing a solid foundation for accurate and reliable insights.

5.0   Data Analysis, Results and Discussion

5.1 Visualization 1: Number of Movies by Year
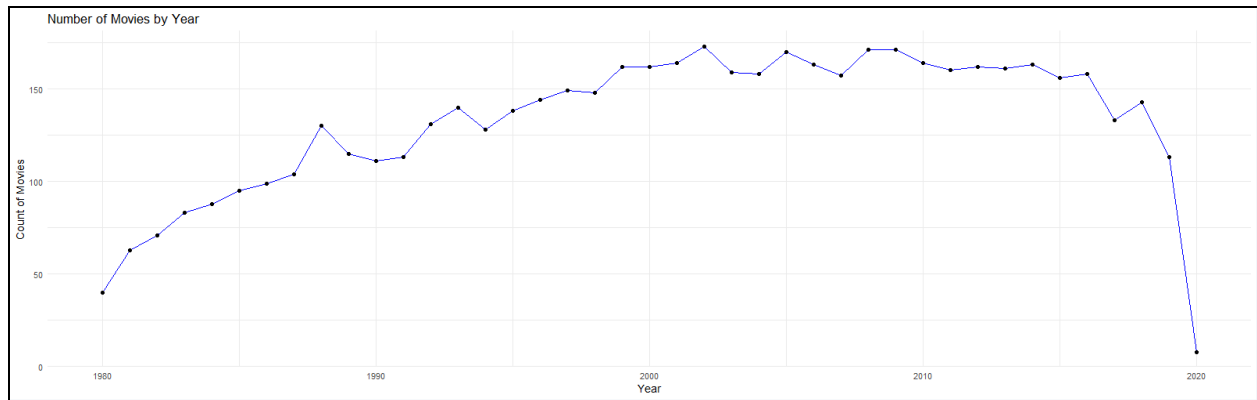
Result:

The number of movies produced showed an increasing trend from the 1980s, peaking in the late 1990s and early 2000s. This peak was followed by a period of relative stability until around 2010. After 2010, there was a noticeable decline in movie production, with the count reaching its lowest level in 2020 during the entire period analyzed (1980-2020).

Discussion:

From the graph described above, it's also possible to outline some significant patterns and probable effects for the motion picture business in the last few decades. The increase in the production of the movies from the 1980s to the early 2000s can be attributed many reasons for instance due to increased demand for the movies by the audience or because the new technologies came in which helped in the production of the new movies or because of the favorable economic conditions that prevailed during that period that encouraged people to invest. This points out that there was a period of growth and development of the film industry in the late 1997/ early 2000 that might have been informed by factors such as; increase in the number of blockbuster films, advancement in technologies through digital cinemas or the globalization in movie markets.

However, the latter period from around 2010 and seeing a reduced number of movies made per year do make one ask what the problems have been with the industry in the recent past. Such a decline, which is most pronounced in 2020 and is shown below, can be attributed to a host of factors that range from shifting consumer preferences to the growth of streaming services and other forms of entertainment, economic conditions affecting the feasibility of theatrical releases or production budgets of films, or changes in the structure of the film industry financing.

As seen from the above graphical representation it is also worth to understand that this graph on its own does not enable the understanding of the key reasons that led to the observed trends. Population of variables such as industry trends, consumers' preferences and behavior, technology and economic condition would require an additional examination for a more refined understanding of the causal factors of the observed oscillations in the frequency of movies production.Also, further study must be done on the quality and acceptance of the movies created throughout the periods, or the financial achievement of the industry in the correlation of the production rates and success or profitability.

Lastly, the drastic drop in the number of produced movies in the recent past as highlighted in the above graph may imply that the industry is facing some difficulties or is in the process of evolving in a way that require the sectors to formulate plans on how to respond to the challenges requiring transformation to maneuver through the environment properly.

5.2 Visualization 2: Number of Movies by Genre
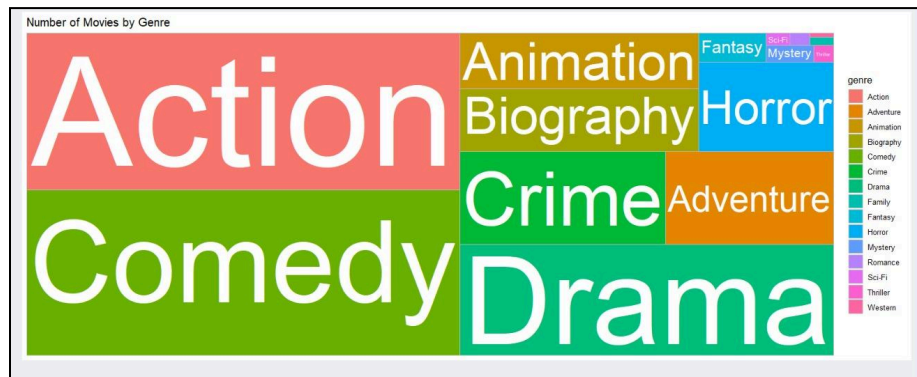
Results :

The treemap visualization illustrates the number of movies categorized by genre. The most well-known genres, indicated by the largest squares and font size, are Action, Comedy and Drama. These three genres dominate the dataset, suggesting that they are the most prevalent types of movies produced. Other notable genres, though represented in smaller fonts, include Crime, Adventure, Horror, and Biography. Smaller genres such as Animation, Fantasy, Sci-Fi, Mystery, Thriller, and Western are also present but appear less frequently in the dataset.

Discussion:

The visualization highlights the dominance of Action, Comedy, and Drama genres in the film industry. This prevalence could be attributed to the broad appeal these genres have among audiences. Action films are known for their thrilling sequences and high energy, often drawing large crowds seeking entertainment and excitement. Similarly, Comedy movies attract viewers looking for humor and light-hearted enjoyment, while Drama films appeal to those interested in storytelling and character development. The large representation of these genres indicates a significant demand and a robust market for them.

In contrast, genres like Animation, Fantasy, Sci-Fi, Mystery, and Thriller are less prominent. This could suggest that while these genres have dedicated fan bases, they may not appeal as widely to the general audience as Action, Comedy, and Drama. Animation, often targeted towards younger audiences or families, has a distinct but narrower market. Sci-Fi and Fantasy, though popular, might require larger budgets and thus, fewer films are produced compared to the more universally appealing genres. The smaller representation of Mystery and Thriller could indicate a niche audience that prefers these genres.

The diversity of genres displayed in the visualization reflects the varied tastes of moviegoers and the film industry's attempt to cater to these different preferences. However, the predominance of certain genres over others highlights market trends and audience preferences that significantly influence film production. Understanding these trends can help industry stakeholders, such as producers and marketers, to make informed decisions about future projects and investments, ensuring they align with audience demands and maximize potential returns.

5.3 Visualization 3 : Top 5 Comedy Movies by Gross
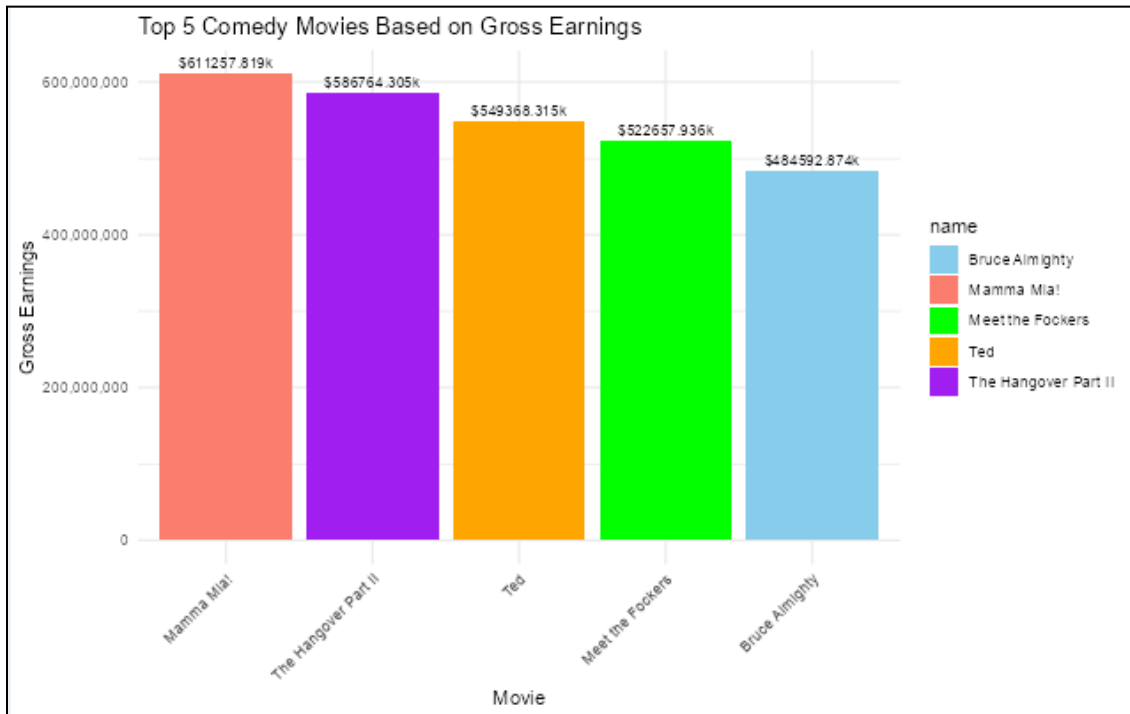
Results:

The bar chart provided illustrates the top 5 comedy movies based on their gross earnings.
Here is the detail and analysis that we get from the visualization:

- **"Mamma Mia!":** $611,257,319
- **"The Hangover Part II":** $586,764,305.
- **"Ted":** $543,368,315.
- **"Meet The Fockers":** $522,657,356.
- **"Bruce Almighty":** $484,592,874.

Discussion:

Based on our analysis, the musical comedy "Mamma Mia!" achieved blockbuster status due to several factors. Effective marketing strategies capitalized on the existing popularity of the ABBA-themed stage musical, whose timeless songs resonate across generations. The film's casting of renowned actress Meryl Streep added significant appeal, attracting a broad audience through her star power and reputation for strong performances. Additionally, the film's feel-good storyline, having a good backdrop, and provided joyful escapism. These elements that include strategic marketing, beloved musical content, and outstanding casting are combined to drive "Mamma Mia!" to impressive box office success.

As for "Bruce Almighty," the movie's lower gross earnings compared to the other top comedies might be attributed to several factors. One significant challenge could have been the competition from other major releases at the time, which might have overshadowed its box office performance. The fantasy comedy genre, while appealing, may not have resonated as broadly with audiences as the vibrant and cheerful musical comedy of "Mamma Mia!" Additionally, viewers may have had a preference for more lighthearted and uplifting content, which "Bruce Almighty" did not fully provide despite its humorous premise. Moreover, the marketing strategies for "Bruce Almighty" might not have been as effective or far-reaching as those employed for "Mamma Mia!," resulting in less audience engagement and lower overall earnings. These combined factors likely contributed to "Bruce Almighty" having the lowest gross among the top five comedy movies.

These insights underscore the film industry's need for a balanced approach that combines creative innovation with strategic marketing and timely releases to maximize audience appeal and box office returns. The information gained from these findings can be valuable to audiences who want to search and identify the best movie to watch with their preferences. This analysis can be used further by directors, website creators and movie streaming platforms to make research and marketing strategies for their future movies.

5.4 Visualization 4 : Top 5 Action Movies by Gross
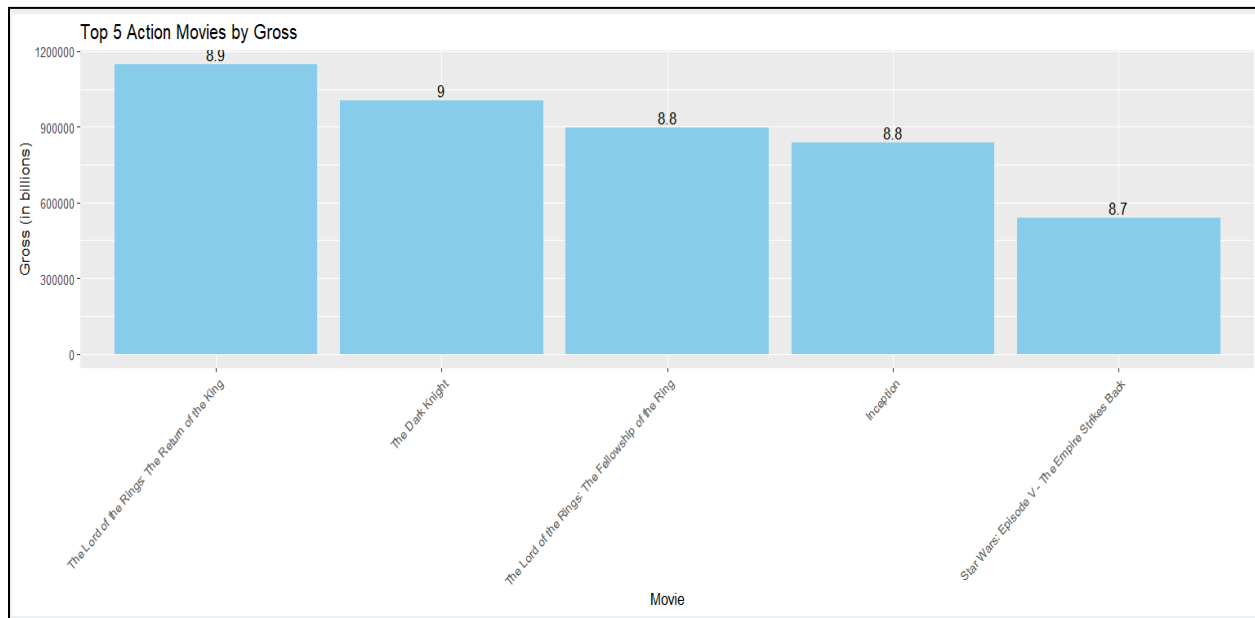
Results :

This visualization shows a bar graph for the top 5 action movies by gross and score.

- The Lord of the Rings : The Return of the King - 1100000 billion and the score is 8.9.

- The Dark Knight -1050000 billion and the score is 9

- The Lord of the Rings : The Followship of the King - 900000 billion and the score is 8.8

- Inception- 825000 billion and the score is 8.8

- Star Wars Episode V: The Empire Strikes Back- 525000 billion and the score is 8.7

Discussion :

The gross is the revenue of the movie and the score is given by the audience after watching the movie. Here we can see that all the top 5 action movies scored more than 8.5. This score shows that the movie listed here is highly liked by people. That is the reason why the movie got high gross revenue. The highest revenue for the action movie is for the film called The Lord of the Rings : The Return of the King. Even though the highest score is for the movie The Dark Knight, it did not have the highest revenue. The revenue shows how many people have paid to watch the movie. The difference between the revenue and the score is not very serious to consider because the score is still 8.5 where it shows that the movie is highly liked by people.

5.5 Visualization 5 : Movie Budget vs Gross
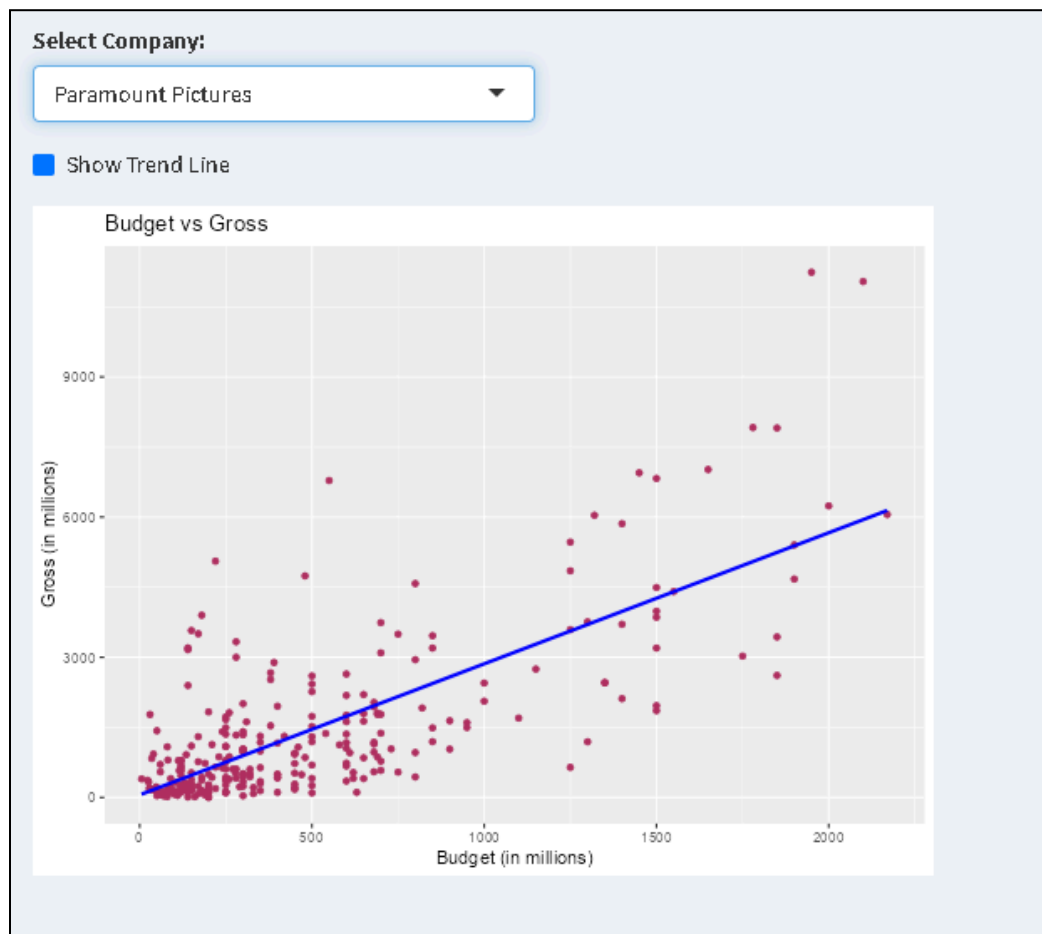
Results:

Above visualization shows a scatter plot to display the correlation between movie budget and gross by company. Here, we use a dropdown filter button to make our visualization more appealing and interactive for viewers to choose which company they want to see or analyze. The red dot represents movie releases by the company and the blue line indicates the overall trend between gross and budget. We can see that as the gross earning's increase, the budget of the movie also increases.

Discussion:

As an example, we choose Paramount Pictures to analyze their correlation. Here we can analyze it as having a positive slope and increasing trend line which indicate a positive relationship between gross earnings and budget. There are some factors that can lead to this positive relationship. One of them is production quality. Paramount Picture provides higher budgets for some movies allowing for better production quality, advanced special effects or computer graphics. This can help in attracting a larger amount of audience to watch the movie produced by them.

Next, well-known stars they hired also can be a contributing factor. With the larger budgets, the company can afford to hire a top-tier star or director who has a good background and skills to get more audiences watching the movie. Star power can significantly boost the film's box office performance as we know some viewers will watch the movie based on a famous star or their favorite star.

Lastly, based on their marketing and promotion perspectives. Large amounts of budgets can help to allocate for good marketing and promotions. From doing advertising, campaign, trailer, poster and social media promotion can create higher ticket sales. With this initiative, higher gross earnings will be achieved by the company.

By understanding the correlation between gross and budget by company it can provide a better insight for other companies, movie producers, stars and viewers. As for companies, they can make a good company with positive relationships as their references to produce a good blockbuster in a legal way. Movie producers get to choose which company with a better budget so that they can give more contribution to their idea in making a high quality movie. From the star's perspective, they get to see which company can provide a good platform for their talents to shine and negotiate better deals. And for viewers, understanding the correlation between gross and budget by company can help them anticipate the quality of the movie they're about to watch and make more informed choices about where to invest their time and money. This insight can also aid in predicting box office success and potentially uncovering hidden gems that might otherwise go unnoticed. Overall, a deeper understanding of this correlation can benefit all stakeholders in the film industry, leading to better collaborations, more successful productions, and ultimately, more satisfying cinematic experiences for audiences worldwide.

5.6 Visualization 6 : Top 5 Stars by Score

Results:

The bar chart showcases the distribution of stars across the top 5 stars based on score. The data spans from 1980 to 2020 and shows the stars based on the selected year. The total score is in sum.

Discussion:

A bar chart displaying the Top 5 Stars by Total Score in the year 2012. The data is presented with the star's name on the left and their corresponding Total Score value represented by the length of the horizontal blue bars. For example, the chart displays the year 2012. Hugh Jackman had the highest Total Score of 14.9. Bradley Cooper and Denzel Washington had the second and third highest scores of 14.7 and 14.0, respectively. Joseph Gordon-Levitt and Colin Farrell rounded out the top 5 with scores of 13.9 and 13.4, respectively. The chart provides a visual representation of the top-performing stars based on a total score. Since the year can be selected from 1980 to 2020, this chart likely represents just one snapshot in time, and the rankings and scores may vary when viewing data from different years within that range.

5.7 Visualization 7 : Movies Count by Rating and Year

Select Year:
2017

Dot Plot of Movie Counts by Rating for 2017
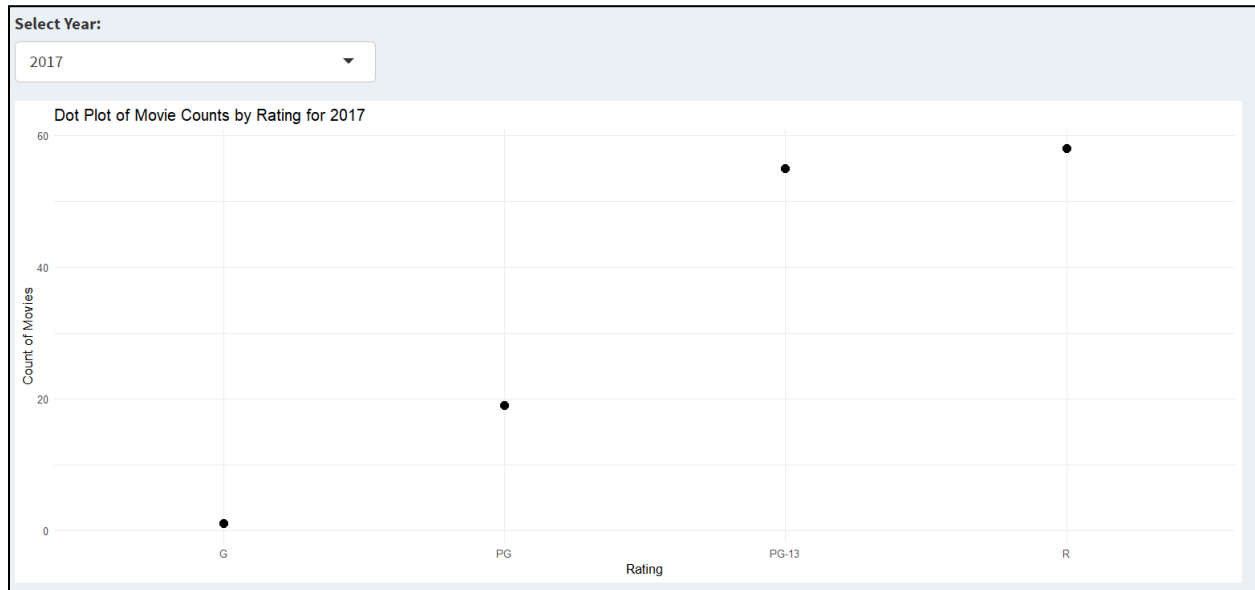
Results :

- G- 2 movies
- PG- 20 movies
- PG-13 - 55 movies
- R- 59 movies

Discussion:

The visualization here shows the count for movies by rating and year. For this visualization we have a drop down button where we can select the year we want to analyze. For example, the year 2017 is selected. There are 4 types of movie rating shown in the output. G refers to General Audience, PG refers to Parental Guidance, PG-13 is where the kids under 13 years old cannot watch the movie and lastly the R refers to Restricted Movies.On the year 2017, restricted rating movie is the highest in movie count. Restricted here shows that the movies only adults can watch. This type of movie usually has creative freedom. This type of movie will

usually target the adult audience to match their maturity level. For this type of movies, the filmmaker can maintain the artistic elements to achieve their goal in making the movies.Moreover, in restricted type movies many genres fall under this such as horror, action, thriller and so on. Mostly restricted type movies are preferred by many people because they can relate to the movie content.

The least count is for the general audience movie due to the limited audience. This is because most kids and families will prefer this type of movie. Since the population that prefers general audience rating movies is less, so the count for movie releases is also less. Other than that , the commercial success for this type of movie will not match the expectation of the producers since it is not mainly preferable . General audience type movies also have some restrictions in terms of content because the kids should be comfortable while watching it. All these reasons make the filmmakers avoid doing general audience movies.

5.8 Visualization 8 : Top 5 Movies based on Votes and Genre
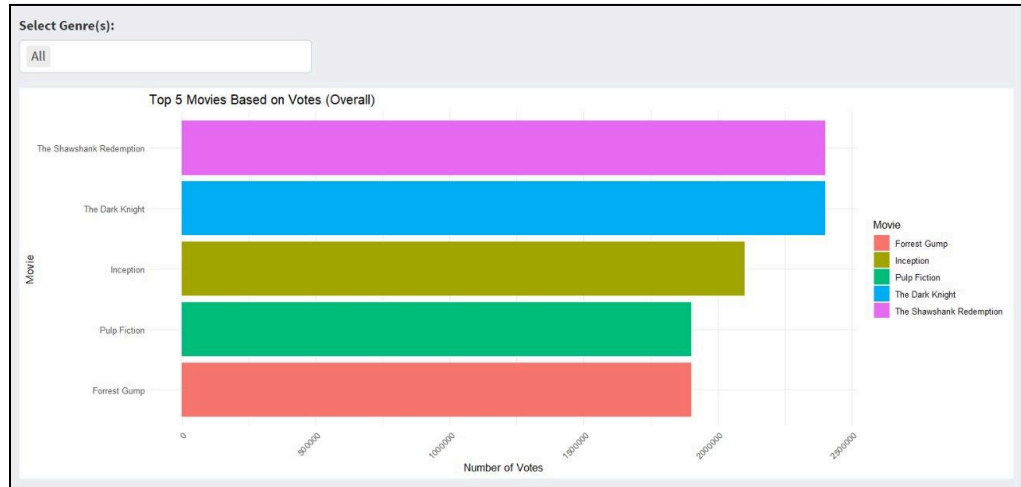
Results:

Based on the data provided, the visualization shows the top 5 movies by genre and votes, where the genre can be filtered by selecting different genres in the selective input box. The top 5 movies for overall genre based on votes give the following results:

- "The Shawshank Redemption" and "The Dark Knight" - 2400000 votes
- "Inception" - 2100000 votes
- "Forrest Gump" and " Pulp Fiction"- 1900000 votes

Discussion:

The high number of votes for "The Shawshank Redemption" reflects its enduring popularity and critical acclaim. Often considered one of the greatest films ever made, it has a broad appeal due to its powerful storytelling, memorable performances, and profound themes of hope and redemption. Its top position in this list underscores its lasting impact on audiences and its status as a cinematic masterpiece.

"The Dark Knight" and "Inception," both directed by Christopher Nolan, highlight his influence in modern cinema. "The Dark Knight," with its groundbreaking portrayal of the superhero genre and a legendary performance by Heath Ledger as the Joker, continues to captivate audiences. "Inception," known for its innovative narrative structure and visual effects, also garners significant votes, showcasing the audience's appreciation for thought-provoking and visually stunning films. Nolan's works are celebrated for their complexity and ambition, drawing a dedicated fan base.

"Pulp Fiction" and "Forrest Gump" also demonstrate the diversity of popular films. "Pulp Fiction," directed by Quentin Tarantino, revolutionized filmmaking with its non-linear narrative and eclectic dialogue, earning it a cult status. "Forrest Gump," with its heartwarming story and iconic performances, resonates with a wide audience due to its emotional depth and cultural references. The presence of these films in the top five indicates a preference for both innovative storytelling and emotionally engaging narratives, highlighting the varied tastes of moviegoers.

These visualizations provide valuable insights into audience preferences, showing that while certain genres and directors consistently attract high engagement, there is also a strong appreciation for diverse storytelling styles and themes. This information can guide filmmakers and producers in understanding what resonates most with audiences, helping to shape future projects that align with these interests.

6.0   Conclusion

In summary, the thorough statistical analysis conducted has shed light on various aspects of the movie industry, offering valuable insights for both movie enthusiasts and aspiring entrepreneurs. By understanding audience preferences and market trends, this data can guide movie fan's in making informed decisions about their entertainment choices. Likewise, entrepreneurs can utilize these findings to identify potential opportunities and tailor their strategies accordingly. By doing this research also helps to indicate what the factor contributing to high and low data in our analysis. Overall, this research contributes to the ongoing development and success of the movie industry by providing valuable information for stakeholders to navigate its complexities and capitalize on emerging trends.

Based on the objective to evaluate the elements contributing to a movie's box office or critical success, we can conclude that multiple factors play critical roles in determining a film's performance. These include the type of rating, score based on the viewer, effective marketing strategies, timing of release, genre of the movie, budget used and gross they earn . Additionally, factors such as publicity, critical reviews, and cultural relevance also influence a movie's success. By comprehensively analyzing these elements, filmmakers and industry professionals can gain insights into the preferences and expectations of their target audiences, thus increasing the possibility of producing commercially successful and critically successful films.

As for objective analyzing genre trends over time to identify emerging genres and opportunities for innovation in storytelling. We have observed changes in what kinds of movies people prefer to watch, with some genres becoming more popular while others lose their popularity. By spotting emerging genres and trends, filmmakers can grab new opportunities for creative storytelling. This understanding helps them produce content that is synonymous with modern audiences, keeping the film industry vibrant and innovative. Recognizing these emerging genres also gives filmmakers a chance to explore new ideas and push the limit in storytelling, enriching the movie-watching experience for audiences everywhere.

Lastly, for objective exploration the relationship between movie budgets and box office revenues to identify patterns of financial success. We find that spending more money on a film doesn't always mean it'll make more money back. While big budgets can help make movies look impressive and attract big stars, other things like how the movie is marketed, when it's released, and what type of movie it is also play a big role in how much money it makes. So, filmmakers need to be smart about how they spend their money and make sure they're making movies that people actually want to see in order to make a profit.

As for the proposed solution, the movie recommendation system holds significant promise for entrepreneurship within the film industry by offering personalized recommendations tailored to individual preferences and market trends. By leveraging data analytics and machine learning algorithms, entrepreneurs can enhance customer engagement, increase user satisfaction, and drive revenue growth through targeted marketing and content delivery. This solution not only streamlines the consumer experience but also provides valuable insights for entrepreneurs to optimize their product offerings, identify niche markets, and capitalize on emerging trends, ultimately fostering innovation and competitiveness within the industry. It also can give valuable insight for the company, producer, director and star to see the trend in the movie based on many aspects.

7.0   Limitation of the Study

Firstly,time constraints are the limitation of this study because the dataset spans 35 years only, from 1986 to 2020. Although this duration provides ample data for analysis, it might not encompass the latest trends and advancements in the film industry. The movie industry is constantly changing, with ongoing developments in technology, shifting audience preferences, and evolving market dynamics. Consequently, the insights derived from this dataset may not completely represent the present condition of the industry.Moreover, although the dataset offers valuable insights into different facets of the film industry, it may be lacking in details for a lot of countries. The dataset only focuses on United State movies.

Additionally, the research concentrates mostly on the quantitative evaluation of data, including box office earnings, ratings, and genre patterns. While quantitative analysis provides important information, it may not completely encompass the qualitative elements of movie production, such as artistic originality, cultural importance, and audience response. Including qualitative research techniques, like interviews with professionals in the industry or surveys of the audience, could offer a more thorough insight into the film industry.
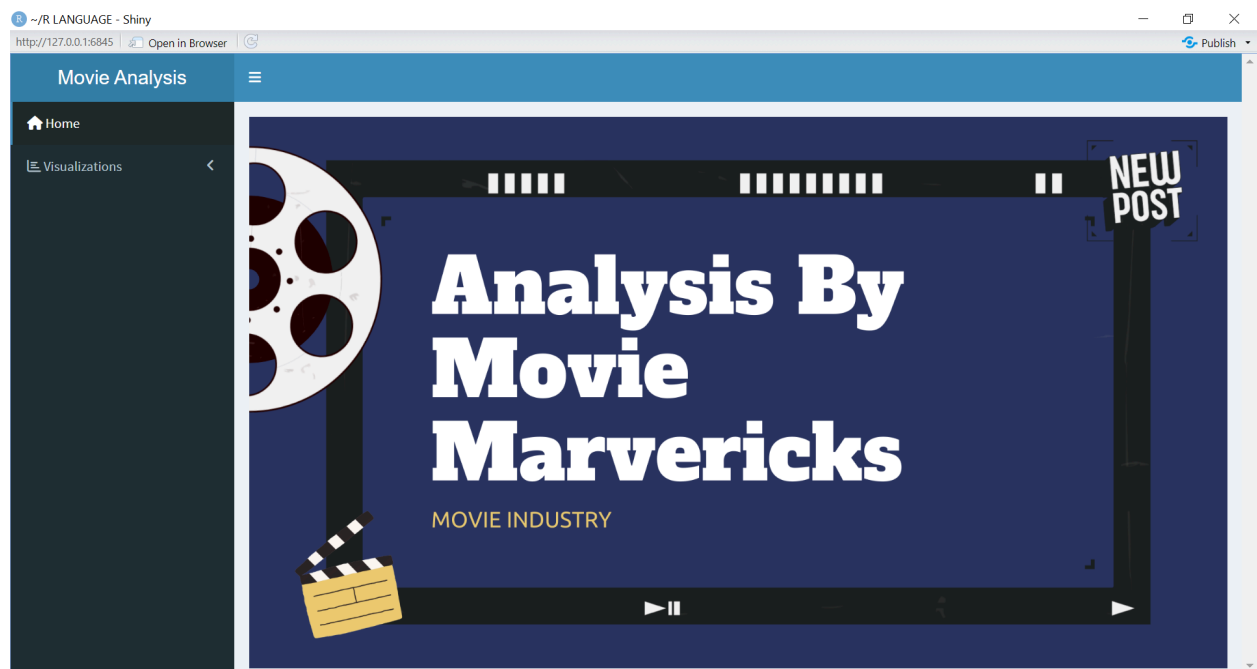
In general, although this research provides important understanding of different facets of the film industry, it's important to acknowledge its constraints and analyze the results carefully. Subsequent studies could overcome these limitations by utilizing more up-to-date and thorough datasets, considering regional differences, integrating qualitative research techniques, and investigating further factors impacting the triumph of movies.
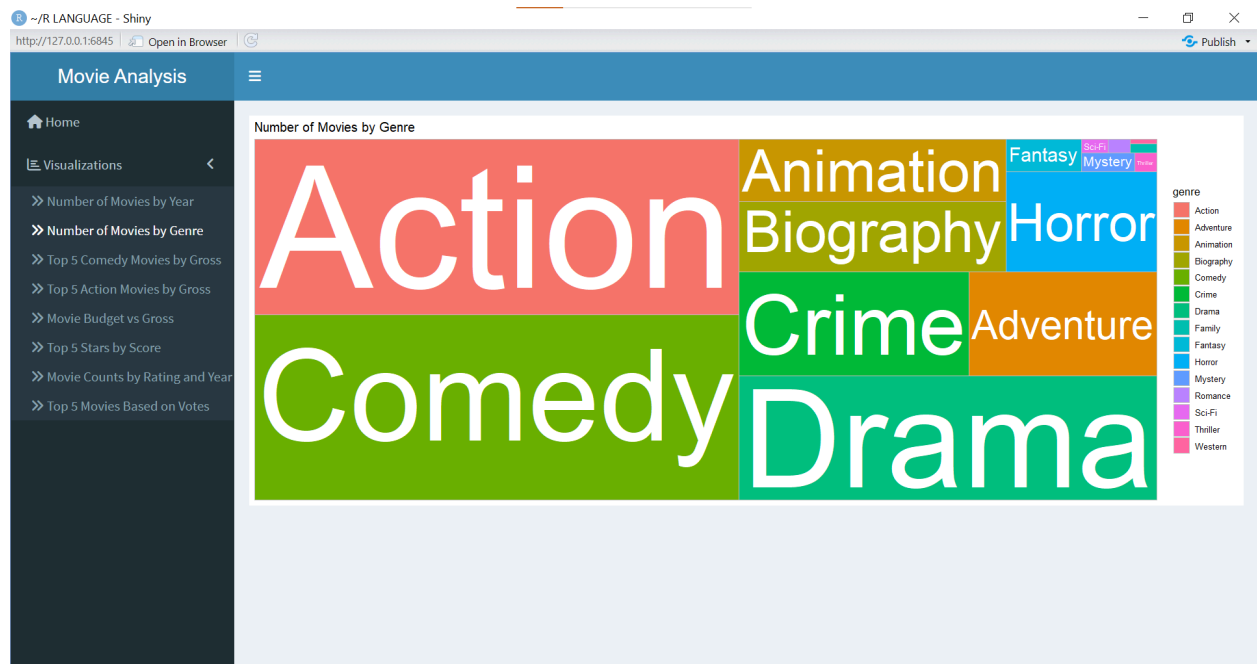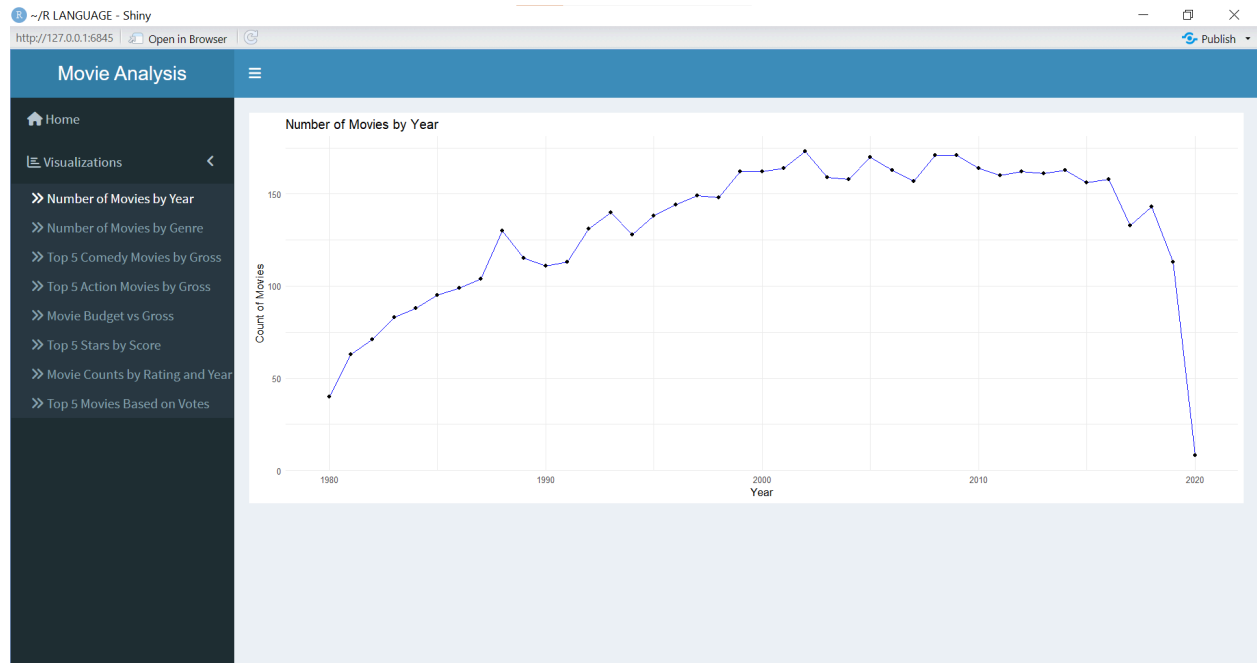
8.0   Appendix

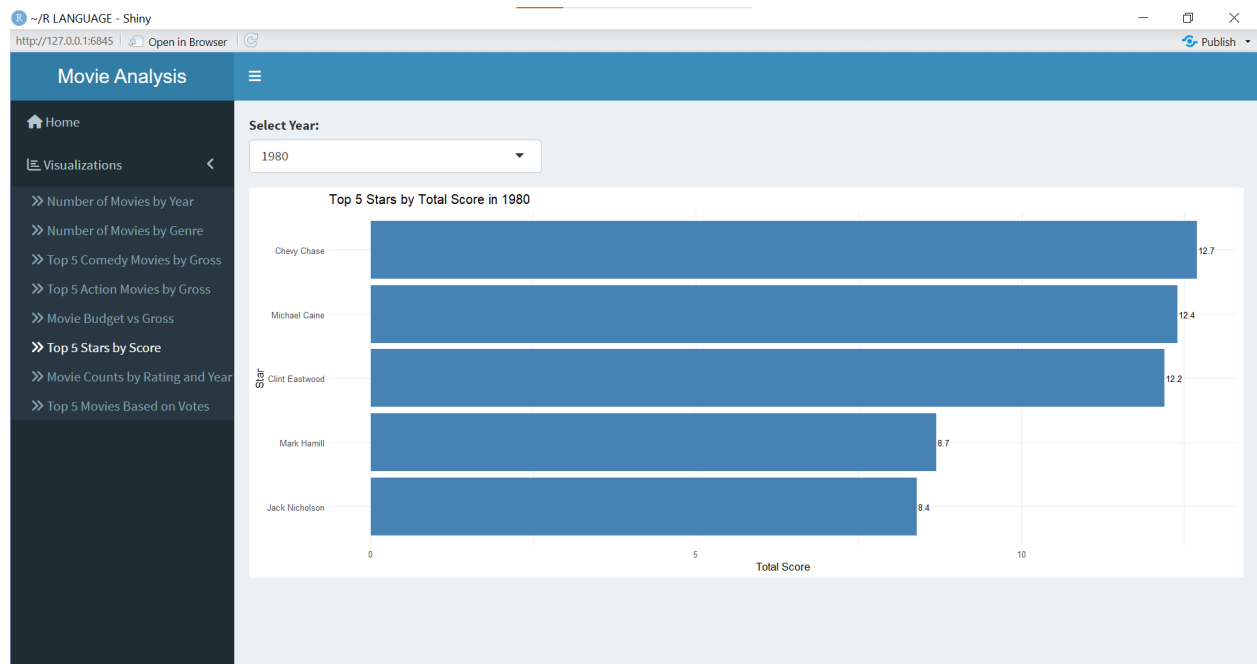Link for Dataset, Presentation Slide, Presentation Video and R Coding:

https://drive.google.com/drive/folders/1t8qhB1wyUYnRCycUyHtgnzb2_9HQxeQs?usp=sharing

GUI screenshot:
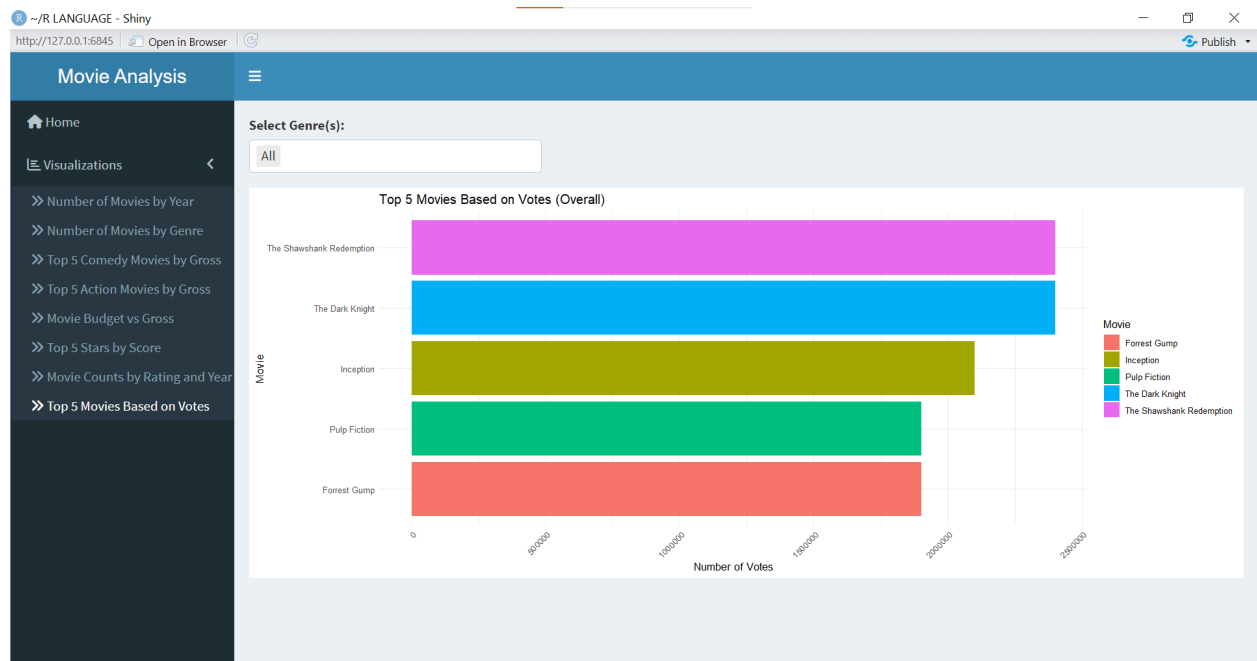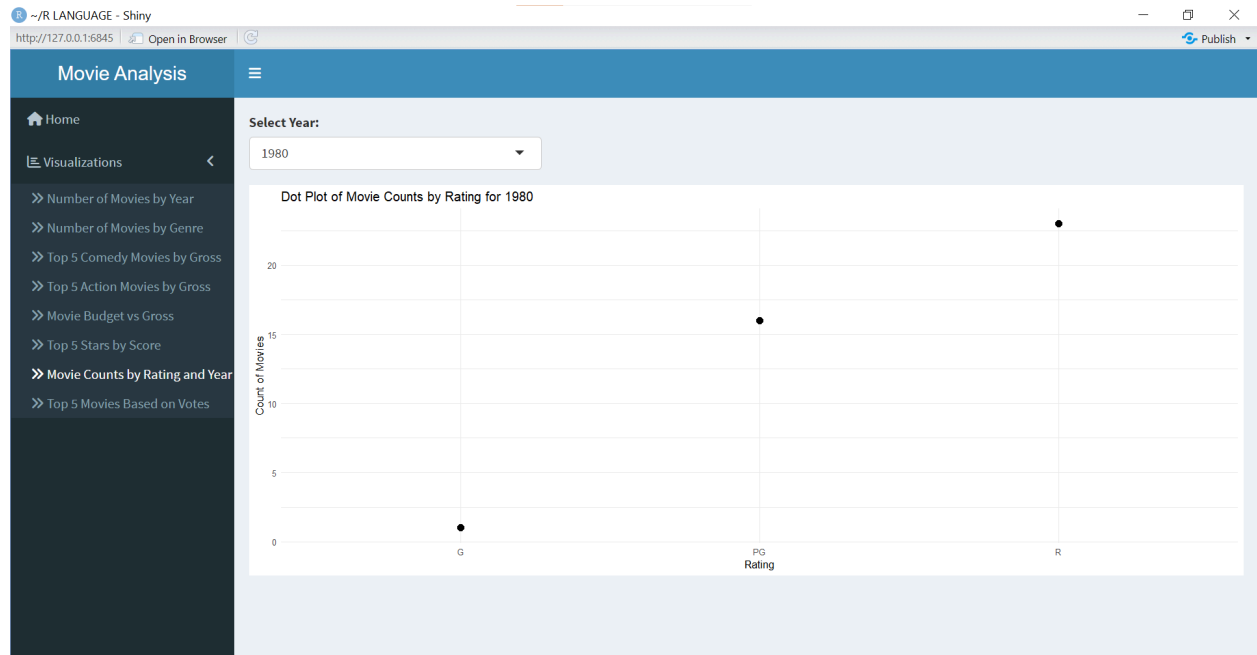
GROUP PROJECT PLAN & APPROVAL

SECTION NO.	02G	
GROUP NAME	Movie Marvericks	
GROUP MEMBERS **(Leader's name is the first in the list)**	**ID NO.**	**NAME**
	SD22002	ALMIRA DAMIA BINTI SYAHNIZAM
	SD22016	SRI SHAMNEE SAI A/P RAJAH
	SD22051	MOHAMMAD BASHARIL AIMAN BIN MOHAMMAD BAKHTIAR
	SD22005	NURKHAIRUL IZZATI MOHD SALLEHAN BINTI
PROJECT TITLE	MOVIE INDUSTRY	
PROJECT DESCRIPTION	As movies are a significant cultural and entertainment medium, that is why we are Interested in analyzing the movie industry. This project can uncover hidden patterns and correlations that traditional methods may overlook. It can potentially inform decision-making processes and drive innovation within the movie industry. Focused on movie revenue and analyze over the lost decades, market trends and competition, production budgets and costs.	
DATA DESCRIPTION	Quantitative data - votes, budget, gross, runtime Qualitative data - name, rating-genre, year, released, score, director, writer, star, country, company	
APPROVED BY	Associate Professor Dr. Roslinazairimah Zakaria	
DATE	3/4/2024	

Marking Sheet (For Lecturer's use only)

	DATA SCIENCE PROGRAMMING II (BSD2223)		MARKS: 120 (30%)
UNIVERSITI MALAYSIA PAHANG PUSAT SAINS MATEMATIK	GROUP LEADER: ALMIRA DAMIA	ID NO : SD22002	
	GROUP PROJECT	SECTION NO: 01G/02G/03G/04G DUE DATE: 19/6/2023	/120

RUBRICS FOR CLO2/PLO2

CLO2: Analyse and summarise data using appropriate programming tools	PLO2: Cognitive Skills and Functional work skills with focus on Numeracy skills. **C4: Analysis**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **obtain appropriate data** for the project	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **summarise data numerically and graphically using R**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **analyse data using R and obtain data insights**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **provide conclusion** and recommendation from the project.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10

RUBRICS FOR CLO3/PLO3

CLO3: Develop programming codes to solve problems	PLO3: Functional work skills with focus on Practical, and Digital skills. **P4: Mechanism**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **construct R codes to summarise data numerically and graphically**.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **construct R codes for data analysis**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **develop a dashboard (GUI) to present the data using Rshiny.**	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	4	20

RUBRICS FOR CLO4/PLO5

CLO4: Demonstrate verbal and written communication skills	PLO5:Functional work skills with focus on communication skills. **A3: Valuing**	/20	/5

	Criteria	Achievement Level						Weightage	Score
		0	1	2	3	4	5		
		Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Written	Ability to write the report findings coherently.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to present the project report in the given format which include data description, data analysis, results and discussion.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to present the project proficiently by organizing and communicating the results in a clear, logical, and easy-to-follow manner.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to deliver the dashboard to summarise the project findings.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5

RUBRICS FOR CLO5/PLO8

CLO5: Relate entrepreneur skills in assigned task	PLO8:Entrepreneural skills **A4: Organising values**	/20	/5

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to articulate the given/ chosen case study related to entrepreneurship.	Unable to articulate the given/ chosen case study.	Able to articulate the given/ chosen case study fairly weak.	Able to articulate the given/ chosen case study fairly well.	Able to articulate the given/ chosen case study well.	Able to articulate the given/ chosen case study reasonably well.	Able to articulate the given/ chosen case study excellently.	2	10
Ability to deliver entrepreneur ideas.	Unable to deliver any entrepreneur idea.	Delivery of idea is unclear, vague and not systematic.	Delivery of idea is less clear, vague and systematic.	Delivery of idea is moderately clear, vague and systematic.	Delivery of idea is clear and systematic.	Delivery of idea is very clear and systematic.	2	10