Electronic Integrated Computer Algorithm Journal
Vol. 2, No. 1, October 2024, pp 29~39

ISSN: 3031-0350, DOI: 10.62123/enigma.v2i1.45

Uncovering the Secrets of the 2023 Box Office: Analysis of Factors Affecting the Success of the Top 200 Films Using the Linear Regression Method

Yuyun Yusnida Lase[1*], Sindi Claudia Sianturi[1], Ichan Lingga[1], Dimas Ariyudha[1]

[1] *Politeknik Negeri Med*an, Medan, Indonesia.

DOI : 10.62123/enigma.v2i1.45

Received : September 12, 2024
Revised : September 22, 2024
Accepted : September 28, 2024

Keywords:
Box Office, Film, Linear Regression, Gross Revenue, Distributor, Number of Theatres, Data Analysis

ABSTRACT

The film industry plays a crucial role in the global economy and popular culture, realizing creative outcomes through complex processes involving production to marketing. This research analyzes the factors influencing the success of the top 200 films at the Box Office in 2023 using linear regression. Independent variables such as film ratings, number of theaters, and distributors are examined in relation to total gross revenue. The dataset processed for this research consists of 200 data points specifically for the year 2023, sourced from Kaggle and processed with Python in Google Colab. The analysis revealed that films with a rating of 7.0 and above averaged a total gross revenue of approximately $100 million, while those rated below 7.0 averaged around $50 million, indicating a negative correlation between film ratings and gross revenue. Additionally, films shown in an average of 2,000 theaters grossed approximately $150 million, demonstrating a positive correlation between the number of theaters and gross income. The analysis also indicated that films distributed by major companies tend to have higher grosses, with the top distributors achieving an average gross of $120 million compared to $70 million for smaller distributors. Nonetheless, this analysis highlights the complexity of other factors influencing a film's success. Further research is needed for a better understanding. The relevance of these findings for the film industry lies in supporting strategic decision-making and the development of more sophisticated analytical methodologies.

1. INTRODUCTION

In addition to being a vital component of our cultural life, the film business is also significant to the world economy [1]. With the rapid rise of technology and global market penetration, the industry has become one of the most dynamic and competitive areas in the world of entertainment [2]. A film's ability to succeed is now largely based on its producers' inventiveness as well as a thorough knowledge of the industry, current trends, and viewer preferences [3]. Significant failures and in certain cases bankruptcy have resulted from the difficulties in forecasting success and enormous production expenses [4]. To maximize the performance of their films, directors and other industry players are now using data analysis as a crucial tool [5]. Producers, distributors, and marketers can make better choices regarding a film's creation, distribution, and marketing by utilizing data on audience behavior, market trends, and other relevant elements [6]. The purpose of this study is to look into the variables that affect a movie's box office performance in 2023. We gather information about films from a variety of secondary data sources using a linear regression analysis technique [7], a company with headquarters in the West that manages production, content creation, and international distribution for the film and television industries The relationship between a number of variables, including distributors, theater count, and movie ratings, will be investigated in this study [8], utilizing gross total income. As a result, our research advances knowledge of the dynamics of the film industry as a whole in addition to offering insight into the elements that affect film performance. The methodical process of comprehending and interpreting information from data is known as data analysis, and it forms the foundation of this study [9]. The link between one or more independent variables and one dependent variable is understood in this context through the use of multiple linear regression as an analytical technique [10]. This research will attempt to determine the primary determinants of a film's box office success in 2023 by using this technique. The open data source portal Kaggle, which offers access to a variety of datasets for analysis, is where the data used in this study was found [11]. To determine the association between independent and dependent variables, this study will employ the linear regression approach on a dataset comprising the top 200 box office films of 2023 [12].

The findings of this study should give the film business important information for strategic choices like marketing strategy, distribution options, and film creation [13]. Furthermore, the results of this study may serve as a foundation for the creation of more complex analytical techniques for forecasting future box office success [14]. The film business may enhance their tactics, produce more and better films, and realize higher profit potential if they have a better grasp of the elements that affect a film's commercial success [15]. As a result, this research bears substantial implications for the entertainment industry's ongoing development and transformation.

*Corresponding Author Email: yuyunlase@polmed.ac.id

2. RESEARCH METHODS

2.1 Linear Regression

Finding a functional link between one or more independent variables (also referred to as predictor variables or input factors) and one dependent variable (sometimes referred to as response variables or output variables) can be done statistically using linear regression [9]. Understanding and modeling the effects of independent factors on the dependent variable is the primary objective of linear regression [16]. In linear regression, it is assumed that the relationship between the independent variable and the dependent variable can be explained through a linear function. This linear function is expressed in the form of the mathematical equation $y = \beta0 + \beta1x1 + \beta2x2 + ... + \beta nxn + \varepsilon$, [17]. where: y is the dependent variable or the predicted outcome (in this study, it represents the total gross revenue of a film). β0 is the intercept term, representing the baseline value of y when all other predictors are zero. β1,β2,…,βn are the regression coefficients or slopes, which indicate the degree of change in y for a unit change in each corresponding independent variable x1,x2,…,xn. x1,x2,…,xn are the independent variables or predictors (in this case, they represent factors such as the number of theatres, distributor presence, and film ratings). ϵ epsilonϵ represents the error term, capturing the variation in y that cannot be explained by the independent variables.

Finding the regression coefficients that most closely match the observed data is the first step in the linear regression procedure. The least squares method is typically used for this, and the model that fits the data the best is the one with the smallest mean square error between the model's projected values and the observed values [18]. There are many uses for linear regression, such as hypothesis testing, identifying correlations between variables, and making predictions. It is one of the most widely used analytical methods in statistics and has applications in many different domains, such as the natural sciences, social sciences, economics, and medicine [19].

2.2 Box Office

Even though a successful box office run is frequently regarded as an indicator of a movie's success, American distributors predominate in both domestic and international film markets [7]. The media and the film industry typically pay close attention to films that do well at the box office [20]. The number of subscribers and box office success are likewise influenced by the distribution of films [21]. Future opportunities, including sequels, follow-up projects, or more lucrative distribution arrangements, may arise as a result of the box office performance. Certain movies that do well at the box office sometimes have a lasting effect on the film business as well as trends and popular culture. Thus, the box office performance of a movie is frequently the center of interest for producers, studios, and other entertainment business participants [22].

2.3 Total Gross

A phrase that is frequently used in the film business to describe the total gross receipts or total money received from the sale of movie tickets for a picture during a given time frame. This represents the total gross revenue from all showings at different theaters and locales. TotalGross is the total amount of money made from the sale of movie tickets, including standard tickets, IMAX and 3D tickets, and any additional money made from special screenings or film festivalsThe phrase "TotalGross" is typically used in relation to a movie's box office success and financial analysis. This is a crucial metric for assessing a movie's commercial viability and is frequently used as a benchmark in the business to compare how well each movie is doing in terms of earnings [23].

2.4 Distributor

An organization or business that distributes movies to different markets or platforms so that people can see them is known as a distributor of motion pictures. A film distributor's primary responsibility is to handle every facet of distribution, from arranging the movie's marketing and promotion to negotiating contracts with theaters, streaming services, and television. Film distributors typically have strong ties to theaters, streaming services, and television networks in addition to vast networks inside the film industry [24]. They are also in charge of organizing efficient distribution plans to increase the number of viewers and money made from the released movies. In addition, by buying the distribution rights from film creators, distributors might incur financial risks or contribute to the finance of the film industry. To raise public awareness and interest in the movies they distribute, they are frequently involved in the marketing and promotion of the movies. Film distributors may occasionally operate globally, which means that they are in charge of setting up film distribution in a variety of international markets. In order to satisfy changing market demands and tastes, distribution and promotional techniques must be modified [25].

3. RESEARCH METHODS

3.1 Datasets

The dataset processed for this research consists of 200 data points and this data is only for the year 2023, taken from Kaggle. The 200 data points in the insurance.csv dataset are organized into 6 columns. The dataset consists of two numeric columns (Rank and Theaters) and four categorical columns (Distributor, TotalGross, Title, and ReleaseDate). Each of these variables was carefully selected to analyze their potential impact on the box office performance of films. Below is a detailed explanation of these variables and the rationale behind their selection:

Rank: This numeric variable represents the ranking of films based on their total gross earnings at the box office. The aim of including this variable is to analyze whether a film's position on the ranking list correlates with other factors like the number of theaters or distributor presence.

1. Theaters: This numeric variable reflects the number of theaters in which a film was screened. Theaters are a critical factor for assessing distribution reach, as wider distribution is often linked to higher gross revenue. By including this variable, the study aims to explore the relationship between theater count and box office success.
2. Distributor: This categorical variable indicates the company responsible for distributing the film. Distributors play a significant role in determining a film's marketing and reach. The goal is to investigate whether films distributed by larger, more established companies tend to perform better in terms of gross earnings.
3. TotalGross: This numeric variable represents the total revenue generated by a film during its theatrical run. It serves as the dependent variable in the regression model, with other variables being analyzed for their influence on this outcome.
4. Title: Although this categorical variable captures the name of each film, it was not used in the analysis as it does not provide numeric or categorical data useful for regression.
5. ReleaseDate: This variable records the date when the film was released, which could impact its box office performance due to seasonal factors or competition. However, it was excluded from this specific analysis as the focus was on distributor, theater count, and rank.

The purpose of including these variables in the Python data processing is to model and predict how factors like distribution strategy, theater coverage, and film ranking influence the overall box office performance. The linear regression model helps quantify the degree of influence each variable has on the total gross revenue. The distribution numbers were reorganized into a range from 1 to 50, representing 50 unique distributors within the dataset. Each distributor's name was grouped and assigned a corresponding numerical value for analysis. This step was crucial to facilitate the regression analysis, as categorical data like distributor names cannot be directly used in linear regression without first being encoded into numerical values. To determine the range and grouping of the data, the researchers did not rely on formal Distribution Tests such as Probability Plots or normality testing. Instead, a categorical grouping technique was employed, where each distributor was given a unique identifier based on the frequency of occurrence within the dataset. This approach ensures that each distributor is represented equally and allows for a clearer analysis of how distributor presence correlates with box office performance.

3.2 Data Pre-processing

Pre-processing is carried out in processing data so that the data used can be processed properly and avoid incorrect data. In the processing process carried out, the initial data used is still raw data. In the process, the required data will be formatted in a certain way and according to needs. The research stages of the system to be built can be seen in Figure 1.

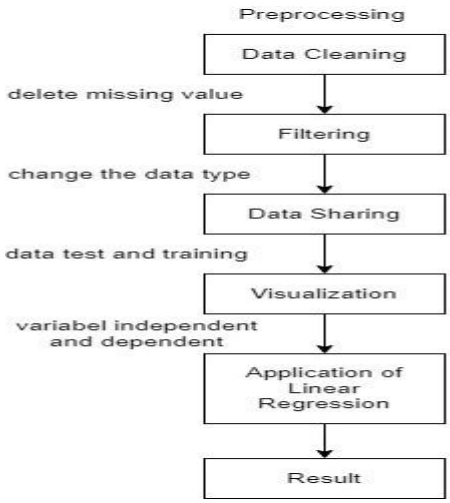

Figure 1. Research Stages

3.3 Data Analysis

This research uses multiple linear regression data analysis, which is processed using Python and implemented with Google Colab Notebook. The first job in data analysis is to remove unwanted data, such as empty data, data outside the threshold, and inappropriate data types. The results of the data cleaning process will be divided into two parts: data used for training and data used for testing. The training process involves utilizing a portion of the data to teach the algorithm how to recognize patterns and relationships within the dataset, and testing uses data to find out how the algorithm worked before. 20% of the test data and 80% of the training data consist of this split. Among the software required are datasheets used in the data simulation process and libraries or packages used in the data process. Datasheet checks data and deletes data that is not needed. An additional step is to determine percentages for training and testing data. As a result, the data accuracy testing process was carried out with data that was not included in the datasheet.

Rank	Title	Theaters	TotalGross	ReleaseDate	Distributor
1	Barbie	4337	594254460	21/07/2023	1
2	The Super Mario Bros Movie	4371	574759600	05/04/2023	2
3	Spider Man Across the Spider Verse	4332	381178195	02/06/2023	3
4	Guardians of the Galaxy Vol 3	445	358995815	05/05/2023	4
5	Oppenheimer	3761	300144670	21/07/2023	2
6	The Little Mermaid	432	297895447	26/05/2023	4
7	Avatar The Way of Water	434	684075767	16/12/2023	5
8	Ant Man and the Wasp Quantumania	4345	214504909	17/02/2023	4
9	John Wick Chapter 4	3855	187131806	24/03/2023	6
10	Sound of Freedom	3411	180587629	04/07/2023	7

Figure 2. Research Dataset

Pre-processing is carried out to reduce errors and homogenize data to facilitate the training process and testing process. The pre-processing stages that must be carried out are letter folding, filtering, word normalization, tokenization, and stopword removal. An overview of the pre-processing stages can be seen in Figure 3.

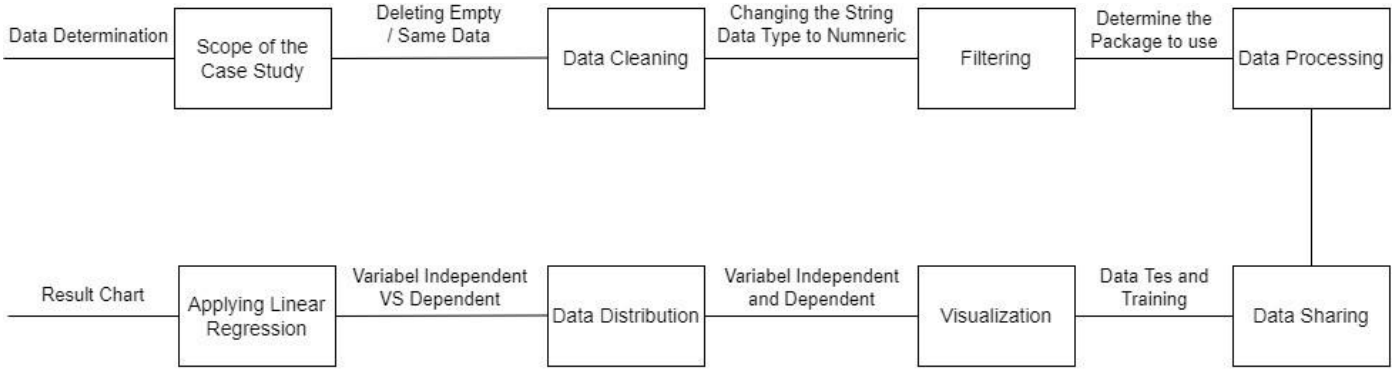

Figure 3. Preprocessing stages

The following image is the raw dataset of top-grossing films obtained from Kaggle, which will undergo a filtering process as illustrated below:

Rank	Title	Theaters	Total Gross	Release Date	Distributor
1	Barbie	4,337	$594,254,460	21/07/2023 00:00	Warner Bros.
2	The Super Mario Bros. Movie	4,371	$574,759,600	05/04/2023 00:00	Universal Pictures
3	Spider-Man: Across the Spider-Verse	4,332	$381,178,195	02/06/2023 00:00	Columbia Pictures
4	Guardians of the Galaxy Vol. 3	4,45	$358,995,815	05/05/2023 00:00	Walt Disney Studios Motion Pictures
5	Oppenheimer	3,761	$300,144,670	21/07/2023 00:00	Universal Pictures
6	The Little Mermaid	4,32	$297,895,447	26/05/2023 00:00	Walt Disney Studios Motion Pictures
7	Avatar: The Way of Water	4,34	$684,075,767	16/12/2023 00:00	20th Century Studios
8	Ant-Man and the Wasp: Quantumania	4,345	$214,504,909	17/02/2023 00:00	Walt Disney Studios Motion Pictures
9	John Wick: Chapter 4	3,855	$187,131,806	24/03/2023 00:00	Lionsgate
10	Sound of Freedom	3,411	$180,587,629	04/07/2023 00:00	Angel Studios

Figure 4. Raw Data

Figure 4 shows whether the data is still intact or not cleaned so that the data type is still the default. The following are the results of the dataset filtering that has been carried out in Figure 5.

Rank	Title	Theaters	TotalGross	ReleaseDate	Distributor
1	Barbie	4337	594254460	21/07/2023	1
2	The Super Maric	4371	574759600	05/04/2023	2
3	Spider Man Acrc	4332	381178195	02/06/2023	3
4	Guardians of the	445	358995815	05/05/2023	4
5	Oppenheimer	3761	300144670	21/07/2023	2
6	The Little Merma	432	297895447	26/05/2023	4
7	Avatar The Way	434	684075767	16/12/2023	5
8	Ant Man and the	4345	214504909	17/02/2023	4
9	John Wick Chap	3855	187131806	24/03/2023	6

Figure 5. Results of The Filtering Process

4. DISCUSSION AND RESULTS

4.1 System Implementation

The proposed method has been tested through experiments; however, the results show that classifying related software in software analysis can be achieved effectively through linear regression. In conclusion, software analysis can easily use well-designed machine learning models. Understanding of software functionality will also be enhanced by the application of machine learning in information analysis. Basic Linear Analysis A case model with one independent variable is called simple linear regression. The dependent variable is defined using simple linear regression. $Y = \beta 0 + \beta 1 + €$. The influence of independent variables is differentiated from the interaction of dependent variables by simple regression. The author of this research will provide results from the application of multiple linear regression, the Python programming language, for the development of a profit prediction system. The following description describes the steps for designing a system using the Python programming language.

4.2 Preparing Libraries and Data

The first step that must be taken to design a profit prediction system in Python is to prepare libraries and research data in CSV format. Library preparation and research data can be seen in Figure 6 below.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as plt

with open('/content/drive/MyDrive/jurnal2.csv') as file:
    data = file.read()
```

Figure 6. Library and Data Settings

4.3 Data Selection

Determine the variables to be analyzed.

```
df1=df[['Rank', 'Theaters', 'TotalGross', 'Distributor']]
df1
```

	Rank	Theaters	TotalGross	Distributor
0	1	4337	684264460	1
1	2	4371	574756600	2
2	3	4332	381176195	3
3	4	445	368986815	4
4	5	3761	300144670	2
...	...	...	...	...
174	195	7	180135	48
175	196	8	173816	49
176	197	274	168108	50
177	198	114	161290	28
178	199	90	161222	11

178 rows × 4 columns

Figure 7. Selection Data

4.4 Data Visualization

To visualize/plot data, first import matplotlib. Also import the Scatter matrix as follows:

```
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
```

Figure 8. Library visualization

After that, continue with data visualization according to the variables that have been determined first

```
viz = df1[['Rank', 'Theaters', 'TotalGross', 'Distributor']]
viz.hist()
plt.show()
```

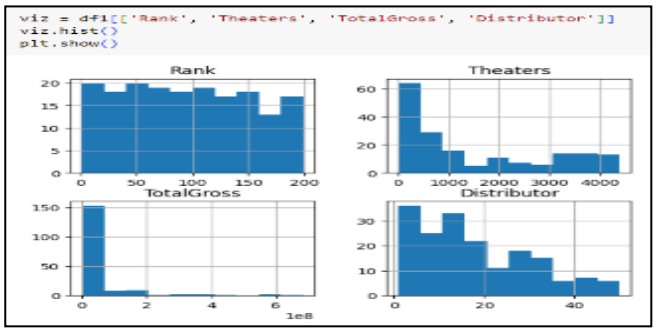

Figure 9. Visualization Diagram

From the visualization results above, it can be concluded that there is a slight upward trend, indicating that films with higher ratings tend to be shown in more cinemas. There is a slight upward trend, indicating that films with higher ratings tend to earn higher total grosses. There is a slight upward trend, indicating that films shown in more theaters tend to generate higher total grosses.

4.5 Scatter Plot

See the distribution of data between the independent and dependent variables being analyzed

1. Scatter Plot with Rank variable.

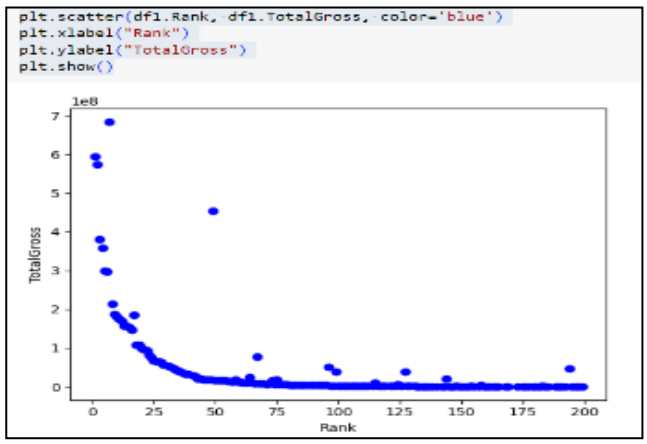

Figure 10. Scatter Plot of Rank Variables

TotalGross vs Ranking
X-axis Movie rating (Rating)
Y-axis: Film gross (TotalGross)
Each point on the graph represents one movie. The position of the point is determined by the film's rating (X-axis) and the total gross revenue it generated (Y-axis).

There appears to be a weak positive trend between a film's ratings and its grosses. Films with higher ratings generally have higher grosses. However, the data points are scattered around the trend line, indicating that ratings are not the only factor influencing a film's gross.A film's rating can affect its gross. Films with higher ratings are likely to be more popular or well received, which can result in larger audiences and higher total grosses.

2. Scatter Plot with variables Theaters

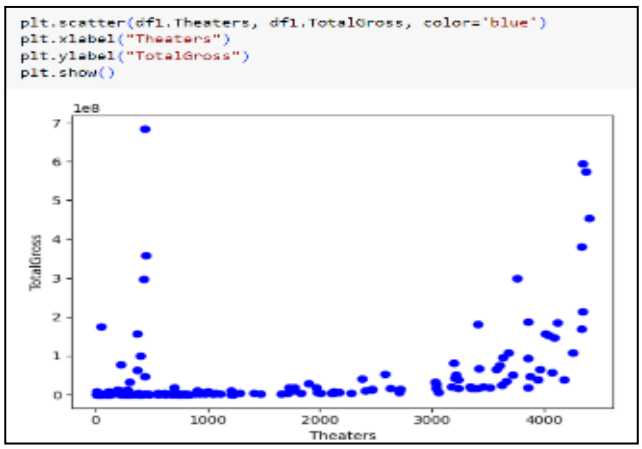

Figure 11. Scatter Plot Variable Theaters

TotalGross vs Cinema
X-axis: Number of cinemas showing films (Cinema)
Y-axis: Film gross (TotalGross)
Films that play in more theaters generally have higher grosses. Data points do not cluster tightly around the trend line, indicating that the number of theaters is not the sole determinant of a film's total gross.

3. Scatter Plot with Distributor variable.

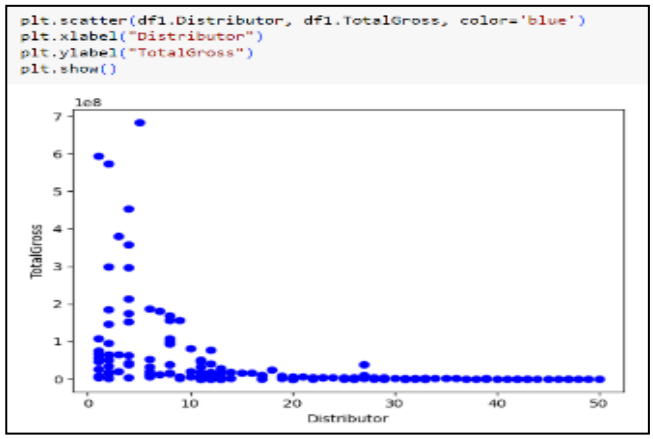

Figure 12. Scatter Plot of Distributor Variables

TotalGross vs. Distributor
X-axis: Film publisher/distributor (Distributor)
Y-axis: Gross film revenue (TotalGross)
This scatter plot illustrates the relationship between the number of theaters showing a film and its total gross revenue. There is a slight upward trend, suggesting that films screened in a larger number of theaters tend to achieve higher gross revenue.

4.6 Linear Regression

After seeing the data distribution in the previous image, linear regression modeling can be carried out, here are the steps. Make sure you have installed Scikit Learn first, and another one for testing. This research uses an 80:20 split, with 80% of the dataset allocated for training and the remaining 20% reserved for testing. A linear regression machine learning model is then created and trained using the decoupled data. To simplify the process, sci-kit-learn is used. Next, the machine learning model is applied to the test data set to generate predictions. A scatter plot is created to compare predicted values with actual values. To visually evaluate the performance of the model, the residuals are plotted. The Python code has been developed keeping these three parameters in mind.

```
!pip install scikit-learn

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.25.2)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.3.0)

from sklearn import linear_model

msk = np.random.rand(len(df1)) <0.8
train = df1[msk]
test = df1[~msk]
```

Figure 13. Coefficients and Intercept of Rank Variables

Execute according to the results of the coefficients and intercept above by adding the regression line

1. Rank variable regression

```
regr = linear_model.LinearRegression()
train_x = np.asanyarray(train[['Rank']])
train_y = np.asanyarray(train[['TotalGross']])
regr.fit (train_x, train_y)
print ('Coefficients: ',regr.coef_)
print ('Intercept: ',regr.intercept_)

Coefficients:  [[-984981.3975754]]
Intercept:  [1.3772736e+08]
```

Figure 14. Coefficients and Intercept Variables Rank

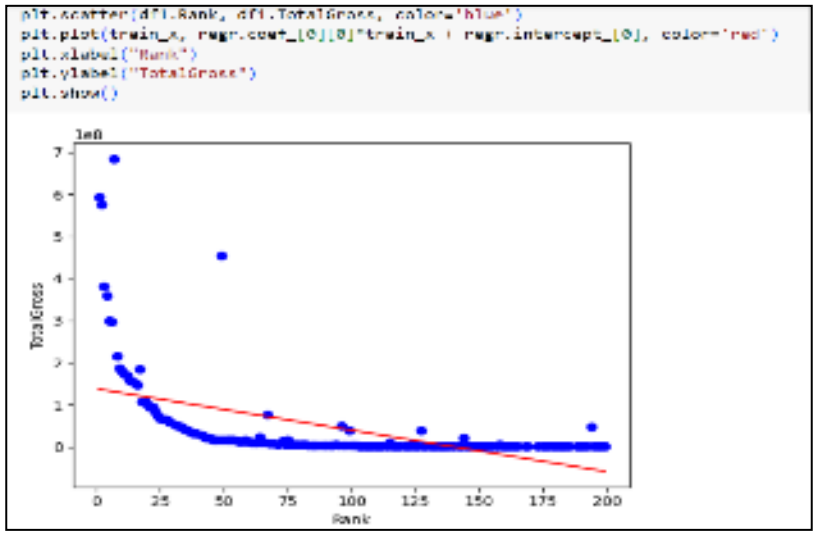

Figure 15. Linear Regression Rank Variables

TotalGross vs. Rank
Blue Dot (Data Point):
The blue dots scattered around the trend line indicate that a film's rating is not the sole determinant of its total gross. Other factors likely played a role in the film's financial success.

a. This scatter plot shows a negative correlation between a film's rating and its total gross.

Red Line (Regression Line):
The descending red line indicates a negative correlation between film ranking (rank) and total gross income (total gross).

a. This means that, on average, films with higher ratings (lower numbers) tend to have lower total grosses.
b. Conversely, films with lower ratings (higher numbers) tend to have higher total grosses.

- Theaters variable regression

```
regr = linear_model.LinearRegression()
train_x = np.asanyarray(train[['Theaters']])
train_y = np.asanyarray(train[['TotalGross']])
regr.fit (train_x, train_y)
print ('Coefficients: ',regr.coef_)
print ('Intercept: ',regr.intercept_)

Coefficients:  [[31790.5786433]]
Intercept:  [-3774781.92514459]
```

Figure 16. Coefficients and Intercept Variables Theaters

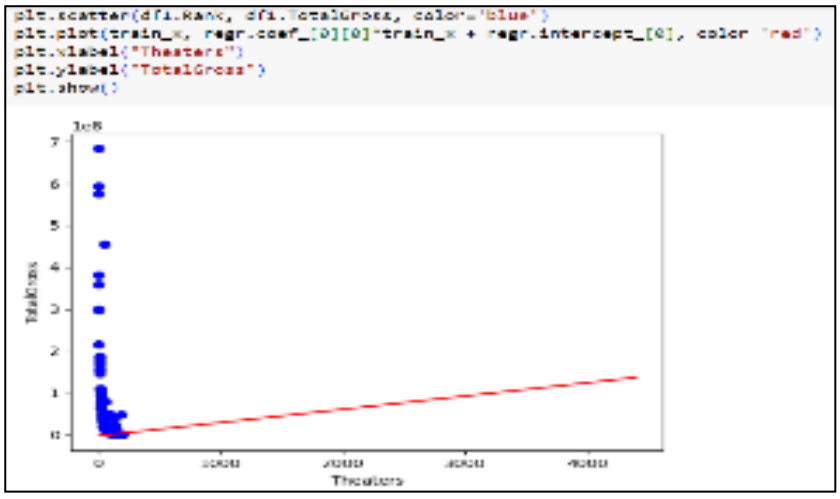

Figure 17. Linear Regression Variable Theaters

TotalGross vs. Theatres
Blue Dot (Data Point):
Each point is plotted on the graph based on total gross (Y-axis) and the number of theaters where the film was shown (X-axis).

a. The horizontal position of the blue dot indicates the number of theaters where the film is showing.
b. The further to the right the dot is, the more theaters are showing the film.
c. The vertical position of the blue dot indicates the total gross of the film.
d. The higher the dot position, the higher the film's gross income.

Red Line (Regression Line):
The slope of the red line shows the direction and strength of the relationship between the number of theaters and total gross revenue.

a. A positive slope indicates a positive correlation.
b. This means that, in general, films that play in more theaters tend to gross higher.

- Distributor variable regression

```
regr = linear_model.LinearRegression()
train_x = np.asanyarray(train[['Distributor']])
train_y = np.asanyarray(train[['TotalGross']])
regr.fit (train_x, train_y)
print ('Coefficients: ',regr.coef_)
print ('Intercept: ',regr.intercept_)

Coefficients:  [[-3221500.53527827]]
Intercept:  [99089308.95793954]
```

Figure 18. Coefficients and Intercept of Distributor Variables

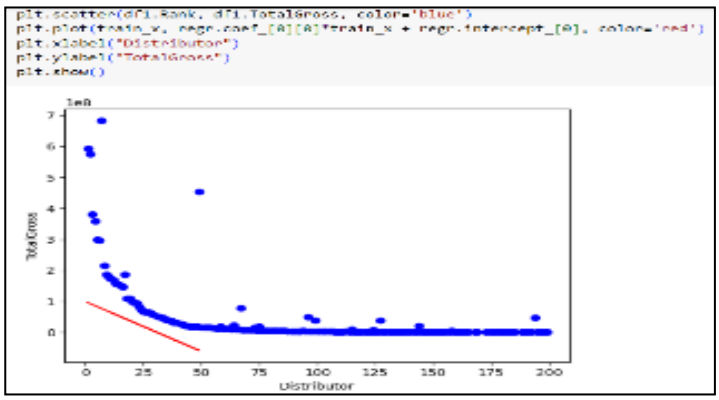

Figure 19. Linear Regression for Distributor Variables

Blue Dot (Data Point):
Each blue dot represents an individual movie in the dataset.

a. The horizontal position of the dot indicates the film's distributor, possibly represented by a code or abbreviation.
b. The vertical position of the dot indicates the total gross of the film.
c. The higher the dot position, the higher the film's gross.

Red Line (Regression Line):
The slope of the red line indicates the direction and strength of the relationship between distributors and total gross revenue.

a. This shows a positive correlation.
b. In this case, films distributed by companies positioned higher on the X-axis (likely large distributors) tend to have higher total grosses.

5. CONCLUSION

Based on the results of the analysis carried out on the top 200 films at the Box Office in 2023 using the linear regression method, several important findings were found which have become a focal point for the film industry and strategic decision making: ***RANK Effect*** **:** There is a negative correlation between a film's rating and gross earnings, indicating that films with higher ratings tend to have lower earnings, but factors such as narrative quality and creativity also influence a film's financial success. ***THEATER Impact*** There is a positive correlation between the number of theaters showing films and gross revenues. Films that play in more theaters

tend to gross higher, underscoring the importance of wide distribution in achieving commercial success. ***Role of DISTRIBUTOR*** **:** Analysis shows a positive correlation between film distributors and gross revenues. Films distributed by large companies or those with a strong market share tend to have higher gross potential.

Example Analysis: The Movie "Barbie"

```
Rank: 1
Title: Barbie
Theaters: 4337
TotalGross: 594254460
ReleaseDate: 21/07/2023
Distributor: 1
```

Figure 20. Top Movies

The film "Barbie," one of the most popular films in 2023, can be analyzed based on the above findings:

1. From the data, it can be seen that the Barbie film rank has the highest ranking in terms of popularity or ticket sales.
2. The Barbie film also played in many different theaters, supporting the finding that wide distribution correlates with higher grosses.
3. For distributors, Barbie films are distributed by large companies or have strong distributor support, which can be a factor that contributes to its success in the market.

Analysis of the film "Barbie" shows that findings about the influence of film ratings, number of theaters and the role of distributors can be applied to predicting the financial success of popular films. This strengthens the validity of the analysis results and shows the relevance of the findings in the film industry at large. While this research provides valuable insight into the factors that influence a film's success at the Box Office, its limitations lie in using data from 2023 only. Future research could expand the analysis to include data from previous years to see broader trends and patterns. Additionally, future research could explore other factors that may influence film success beyond those examined here, such as marketing strategy and film genre

REFERENCES

[1] Z. Liu, “The Influence Of Film Marketing Strategy On The Film Box Office,” 2023.

[2] S. Helma Putri, “Factors Affecting Economic Growth In West Sumatera Province Using Panel Data Regression Analysis,” 2023. [Online]. Available: Https://Mjomaf.Ppj.Unp.Ac.Id/44

[3] “Young Lc Film Box Office Charts And The Metadata Of Culture (1)”.

[4] X. Xie, “Analysis Of The Influence Factors Of Global Film Box Office Based On A Log-Linear Model,” *Bohr International Journal Of Smart Computing And Information Technology*, Vol. 3, No. 1, Pp. 17–24, 2022, Doi: 10.54646/Bijcit.2022.23.

[5] S. Yang, “Analysis Of Top Box Office Film Poster Marketing Scheme Based On Data Mining And Deep Learning In The Context Of Film Marketing,” *Plos One*, Vol. 18, No. 1 January, Jan. 2023, Doi: 10.1371/Journal.Pone.0280848.

[6] C. Yu, “5 An Analysis Of The Box Office Success Of Wolf Warrior 2 (2017).” [Online]. Available: Www.Cbooo.Cn/

[7] A. Bartosiewicz And A. Orankiewicz, “The Market Of Major Film Distributors In Poland In 2002-2018,” *Journal Of Economics And Management*, Vol. 39, Pp. 5–20, 2020, Doi: 10.22367/Jem.2020.39.01.

[8] E. Sifaki And M. Papadopoulou, “Exploring Film Marketing In The New Age Digital Era. Four Cases Of Marketing European Art House Film Productions,” 2022.

[9] L. Quaranta, F. Calefato, And F. Lanubile, “Kgtorrent: A Dataset Of Python Jupyter Notebooks From Kaggle,” In *Proceedings - 2021 Ieee/Acm 18th International Conference On Mining Software Repositories, Msr 2021*, Institute Of Electrical And Electronics Engineers Inc., May 2021, Pp. 550–554. Doi: 10.1109/Msr52588.2021.00072.

[10] A. Jaber, J. Juhász, And B. Csonka, “An Analysis Of Factors Affecting The Severity Of Cycling Crashes Using Binary Regression Model,” *Sustainability (Switzerland)*, Vol. 13, No. 12, Jun. 2021, Doi: 10.3390/Su13126945.

[11] Y. Gao, “Analysis Of The Factors Influencing The Box Office Of Domestic Animated Films,” *Advances In Economics, Management And Political Sciences*, Vol. 10, No. 1, Pp. 171–179, Sep. 2023, Doi: 10.54254/2754-1169/10/20230461.

[12] A. Collins, “Predicting Student Success At A Large State University Using Multiple Linear Regression And Hierarchical Clustering,” 2022.

[13] “Analysis Of Factors Affecting The Import Demand For Wheat And Corn In Türkiye Using The Seemingly,” 2024, Doi: 10.21203/Rs.3.Rs-4071912/V1.

[14] J. Bhatt, S. Verma, And A. Professor, “Box Office Success Prediction Through Artificial Neural Network And Machine Learning Algorithm.” [Online]. Available: Https://Ssrn.Com/Abstract=3753059

[15] “08311063 Novita Ari Prabandini”.

[16] F. Farsa, M. T. Elektro, S. Tinggi, And T. Pln, “Implementasi Metode Regresi Linier Berganda Untuk Proyeksi Kebutuhan Listrik Dalam Rangka Analisis Strategi Bisnis Di Pt Xyz,” Vol. 8, No. 2, 2019.

[17] K. Yohanes Klau, “Kondradus Yohanes Klau Penggunaan Regresi Linear Multipel Dan Metode Kuadrat Terkecil Untuk Menganalisisfaktor-Faktor Yang Mempengaruhi Hasil Produksi Jagung Di Kabupaten Belu,” *Jurnal Pendidikan Matematika*, Vol. 1, 2019.

[18] A. Mimenbayeva *Et Al.*, “Development Of A Linear Regression Model Based On Vegetation Indices Of Agricultural Crops,” *Scientific Journal Of Astana It University*, Pp. 101–110, Sep. 2023, Doi: 10.37943/15emub4283.

[19] Aziz Muslih, Sumarsono, And Andhika Mayasari, "Analisis Faktor-Faktor Yang Mempengaruhi Kepuasan Konsumen Produk Susu Murni Kpub Sapi Jaya Menggunakan Metode Regresi Linear Berganda," *Jurnal Penelitian Bidang Inovasi & Pengelolaan Industri*, Vol. 3, No. 1, Pp. 13–24, Sep. 2023, Doi: 10.33752/Invantri.V3i1.5003.

[20] B. Penelitian *Et Al.*, "Analisis Faktor-Faktor Yang Mempengaruhi Perbedaan Upah Di Kabupaten Tuban Menggunakan Pendekatan Regresi Linier Berganda," Vol. 6, No. 1, Pp. 29–36, 2024.

[21] R. Miedziak And F. Wójcik, "Film Distribution In The Time Of The Covid-19 Pandemic. Will Streaming Take Over The Entire Market? Vědecké Stati {4/8}." [Online]. Available: Https://Www.Researchgate.Net/Publication/361923761

[22] F. Mannion, "The Role Of Entrepreneurship And The Experiential In The Uk And Irish Theatrical Film Sector."

[23] D. Crane, "Cultural Globalization And The Dominance Of The American Film Industry: Cultural Policies, National Film Industries, And Transnational Film," *International Journal Of Cultural Policy*, Vol. 20, No. 4, Pp. 365–382, 2014, Doi: 10.1080/10286632.2013.832233.

[24] "Jooyoung Kwak**& Liyue Zhang 2021(Does China Love Hollywood? An Empirical Study On The Determinants Of The Box-Office Performance Of The Foreign Films In China)".

[25] T. M. Wut, "Review Of: 'Marketing Strategy Of "Tilik" Short Film Analyzed By Social Network'", Doi: 10.32388/Slqqtd.