Strategic Movie Production Analysis

Goals

1. Identify top-movies genres in the box office

This can be done by analyzing which movie genres are the most successful in the box office grossing a lot of money and its popularity among the audience.

2. Gather factors that make a movie be a success or popular

This can be done by investigating characteristics that contribute to top performing films.

Examining factors such as the movie budget, special effects used, storyline of the movie that engages the audience and the cast of the movie as some audience prefer watching movies with a popular cast and director.

3. Understand the market insights

Understand which movie genres are popular with which type of age group and gender as different audiences have different preferences. Example is child under the age 10 primarily girls would prefer to watch animated barbie movies.

Recommendations

1. Budget vs Revenue - Understand the relationship between the budget used to make the movies and the revenue received by the movie studio after the movie sales. We want to understand which aspect influences the other and how can it used by the stakeholders to make the movie studio a success.

2. Popularity vs Audience - Understand the relationship between popularity and the given audience. In the case popularity refers to how popular a certain film is and audience refers to the number of people who watch the certain films based on their popularity. We want to understand how the popularity of a film affects the number of people who go watch the certain films, and how it can be used by the movie studio stakeholders to make the movie studio succeed.

3. Rating - Understand of a rating of a movie influences the success of a movie. Rating of the movies are made by the audience and influence the number of people who watch the movies. We want to understand how this aspect can be used the stakeholders to make the movie studio a success.

4. Genres - This is the categories of movies which can either be romance action etc. Every audience have different preferences in terms of genres. As a movie we want to understand how different genres influence the success of a movie and with other aspects such as rating and popularity.

Business Understanding

Introduction

With the competitive edge in the entertainment industry, many companies are seeking in creating original video content as popular way to secure audience attention and use it as a way of earning more revenue.

Our company is interested into venturing to this dynamic market as it been seen and proven as a great source of income.

The real-life problem is that our organization is opening a new movie studio as a step into the film production industry. But we don't have the knowledge and understanding needed to make movies that will do well at the box office. This research attempts to study current box office statistics in order to determine the kinds of films that are doing the best in order to address this difficulty. Through the conversion of these insights into practical plans, we can direct the production of films that both connect with viewers and bring in a significant amount of money for our company's new film studio.

Stakeholders

1. Company executives

Company executives provide insights and input on how the movie studio is to be managed.

They provide strategic moves such as hiring of the intended staff such as directors and cast members of a certain movie.

They oversee the production of the movie in terms of budget required and as well as producers to ensure the movie is a success.

Another role is they approve which type of films to be produced and allocate resources to them.

Problem faced by company executives lack of market insight, company executives cannot approve production of a movie without knowing the intended market and what their needs and wants includes.

Stakeholders

2. Marketing and sales teams

A problem faced by marketing and sales teams is the ineffective marketing and sales strategies.

Market and sales teams require a deep understanding of the audience preferences and market trends to develop a compelling campaign to promote the movies and increase revenue.

Without the knowledge, they incur a loss and money wastage in promoting the movie and lower return in revenue in the box-office.

By identifying which genre of movies gross a lot of money in the box office, marketing and sales teams identify the target market which is the audience depending on their preferences and promote the movies to them increasing the chances of a successful film release.

From this, the movies end up grossing a lot of money in the box office.

Stakeholders

3. Film and production team

The film production team are entrusted from the analysis of movies that generate the highest amount of money in the box office, to identify the key elements that make the successful.

The key elements could include a certain movie genre with its preferred audience, the movie rating, the cast of the movie, the storyline of the movie and its popularity with the audience and the special effects used in the movie.

The film production could then adopt such ideas and create original video content from those elements.

The problem faced by the film production team is lack of knowledge about the key elements, without this knowledge they are at a higher risk to

making a bad movie that does not fit the audience preferences leading to poor performance at the box office incurring the movie studio a huge loss.

Stakeholders

4. Investors

Investors are a very important part in the movie making process as they provide financial aid that is essential with potential return on their investment. This encourages them to invest in more movies if the movies they invest in become huge successes.

Without the proper market analysis and reports about what the target audience may want, many investors may be hesitant in getting involved with the movie making limiting the movie studio financial resources.

Conclusion

The implications set by the project to the stakeholders include:

1. Through analysis and identifying the genres and characteristics of higher grossing films in the box office, the data derived from that provides easy decision making to the film production team who are able to meet the elements and provide film productions that satisfy consumer wants and generate revenue.

2. Company executives will be able to approve movies that meet the consumer wants and needs by allowing film production that meets the elements of grossing a high rating movie.

3. Market and sales teams will be able to design campaigns that are attracting that draw in the audience as a way of promoting the movie.

4. Investors confidence will due to the profit the studio is gaining and invest in more movie productions.

5. The audience will be able to watch and enjoy movies tailored to their needs and promote it by giving it a high rating.

Data Understanding

Data Sources

1. Box-office Mojo

This is a detailed database that tracks box-office revenues of different types of movies world-wide and their overall financial performance of the movies. This data source is relevant to our project as it will enable us to identify which movies perform well in certain countries and how much revenue it attracts in a certain period of time and their overall performance after.

2. IMDb

This website provides information on film genres, cast and crew, storyline context of the film and the rating of the film which enables us to understand the context of successful films.This data is relevant to our project as studying and analyzing different movie genres provides us with certain audience preferences and other information that contribute to a movie success such as captivating storyline context that draws in the audience.

Data Sources

3.Rotten Tomatoes

This is a website that provides the audience with a space to be able to critic and post reviews about certain movies. This data is relevant as audience opinion is critical in the movie industry such that a bad rated movie with negative reviews leads to poor performance in the box office and vice versa. Through rotten tomatoes some audience may be compelled to watch the movies based on the rating and reviews of the people who have already watched the film.

4. TheMovieDB

This is a website that engages the audience to know which films are trending on a certain day of the week or on a certain week time period by showcasing in terms of percentage to show what percentage of the audience is watching these films. Another feature of this website is that, it shows which popular films in categories which include those that can be streamed, are on tv, can be rented and in theaters in terms of percentage to show what percentage of the audience is watching these films in the mentioned categories. This data is relevant as it shows which films are the audience attracted to the most in terms of the percentage such that a movie with a high percentage of 89% is popular among the audience and that many of them watch it. Through this we can be able to extract key characteristics from the films to know why the high percentage of the audience watch such films.

Data Sources

5. The numbers

Includes data on film budgets, revenues, and financial performance, which is crucial for understanding profitability. Its relevance is that it gives insights into budget and profitability, which are important for evaluating financial success.

Data understanding entails:

1. Data descriptive statistics

2. Data information and size

Justification for feature inclusion

From the analysis of the 6 datasets we decided to use 3 out of the 6 datasets.

These datasets include:

- Dataset tmdb.movies.csv.gz

- Dataset tn.movie_budgets.csv.gz

- Dataset im.db.zip

These datasets contained the features needed for the analysis and understanding of our project.

These feature include:

- vote count, rating and popularity - this feature is essential for this project as through vote count and rating we can be able to identify which films are popular among the audience this can be seen through database tmbd.movies.csv.gz

- Budget - This is essential for the project as it allows us to understand the relationship between production budget and revenue to know how much profit and loss the movie studio has incurred. This feature is found in the dataset tn.movie_budgets.csv.gz

- Revenue - It is the main indicator that shows us the financial success the movie has made or the financial loss the movie has made when it is compared to the budget. Revenue can be domestic gross which is revenue income from a certain region or country or world-wide gross which is revenue generated across the globe. This feature is found in the dataset tn.movie_budgets.csv.gz

- Genres - This feature enables us to identify which movie genre is popular and successful in the box office based on its average rating and number of votes made by the audience. This feature is found in the dataset im.db.zip.

Limitations

- Missing values - some features may contain missing values that may affect the analysis of the data.

- NAN values - Some key features in the dataset may contain NAN values that might reduce the accuracy of the analysis and may cause errors during analysis if not acted upon.

- NULL values

- Duplicated values

Data Preparation

Under this section, we will go each chosen dataset and clean in preparation for analysis.

Cleaning of a dataset is very crucial to maintain accuracy as it removes errors and inconsistencies that may cause issues during analysis.

Another reason is to maintain consistency and efficiency of the data.

We would also want to maintain the integrity of our data.

Doing so, we will be looking mainly for missing values, NAN values, duplicate values and dropping unnecessary columns that could affect our accuracy during analysis.

Pandas library will be used as tool to clean the selected datasets.

1. Handling missing values

In data preparation we handle missing values to maintain the integrity of our data, accuracy to ensure our data is not biased and consistency to make it easier for analysis.

In this subsection, missing values will be handled by:

- Identifying the number of missing values in each column of each dataset.

2.Handling NAN Values

In this section we will be handling NAN values.

It is important to handle NAN values they reduce data quality indicating missing values, affect statistical analysis such as mean and standard deviation which is important during the analysis stage of this project.

It also affects visualization which is an important aspect of this project as they lead to misleading and incomplete visualizations.

The NAN values will be handled by:

- Calculating the percentage of NAN values for each column in each selected dataset.

3. Handling NULL values

Since we have database, it is important to check if the database ha NULL values.

We check for NULL values to maintain data integrity as NULL values indicate incomplete or missing values.

Maintain accuracy during analysis as NULL values affect accuracy of a model.

4. Handling Duplicated values

We handle duplicated values to maintain accuracy during analysis especially during statistical analysis of means, variances and standard deviation.

We also handle duplicates to maintain data consistency and integrity.

Data Analysis

In this section we will be applying statistical measures and logical techniques to evaluate and analyze data.

The chosen datasets specific for this project include tn.movie_budgets.csv.gz, tmdb.movies.csv.gz and database im.db.zip.

Each of these datasets will be analyzed through various methods such as hypothesis testing and linear regression.
Our recommendations in this case is what we stated under our business goals, to help understand and implement ways in which each stakeholder can take the movie studio a success.

Our findings will be the results achieved after analysis.

These methods will help us create findings to our business problem and use our analysis and findings as recommendations to our business problem.

We will also be able to explain why our findings support our recommendations and how the recommendations will help the new movie studio succeed.

Methods

1. Hypothesis testing

This is one of the methods that will be used for analysis. Under hypothesis testing, we will be able to analyze each of the datasets to achieve our findings. The findings will be used to explain why it supports our recommendations.

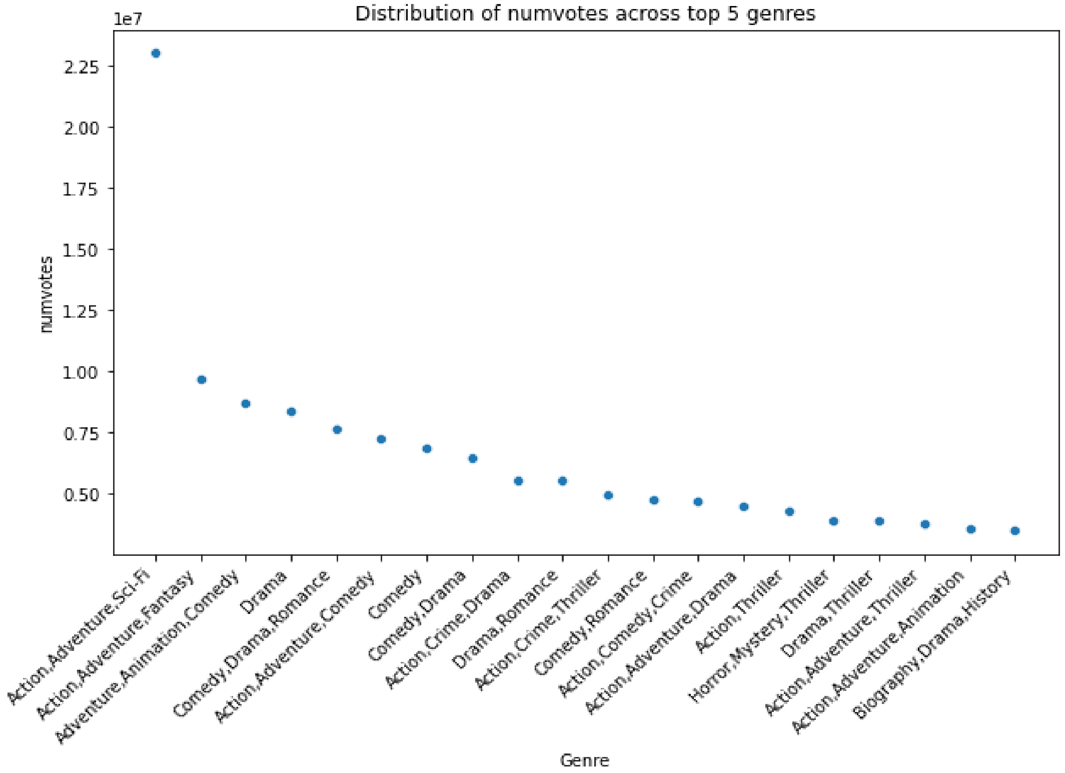
In hypothesis testing, we will be defining the null and alternative hypothesis of each of the datasets and concluding if we are able to reject the null hypothesis or if we fail to reject the null hypothesis.

Statistical tests such as t_tests, correlation will be used with, p-values and alpha values will be used to help us either reject or fail to reject the null hypothesis.

The analysis will be either one-tailed or two-tailed depending on the specific finding.

We will be using an alpha value of 0.05.

Example of hypothesis testing



The visualization below is an example of a hypothesis testing that was done for database im.db.zip.

- Null hypothesis

The null hypothesis states that there is no significant difference between the number of vote across different genres.

- Alternative hypothesis

The alternative hypothesis states there is a significant difference between the number of votes across different genres. Meaning the number of votes are different across different genres.

For the relationship between genres and number of votes, we conclude our findings to reject the null hypothesis starting there is a significance number of votes across different genres.

From the visualization below we can see that the genre with the highest number of votes is action, adventure and Sci-fi as per the analysis showing that this genre is the most watched.

Methods

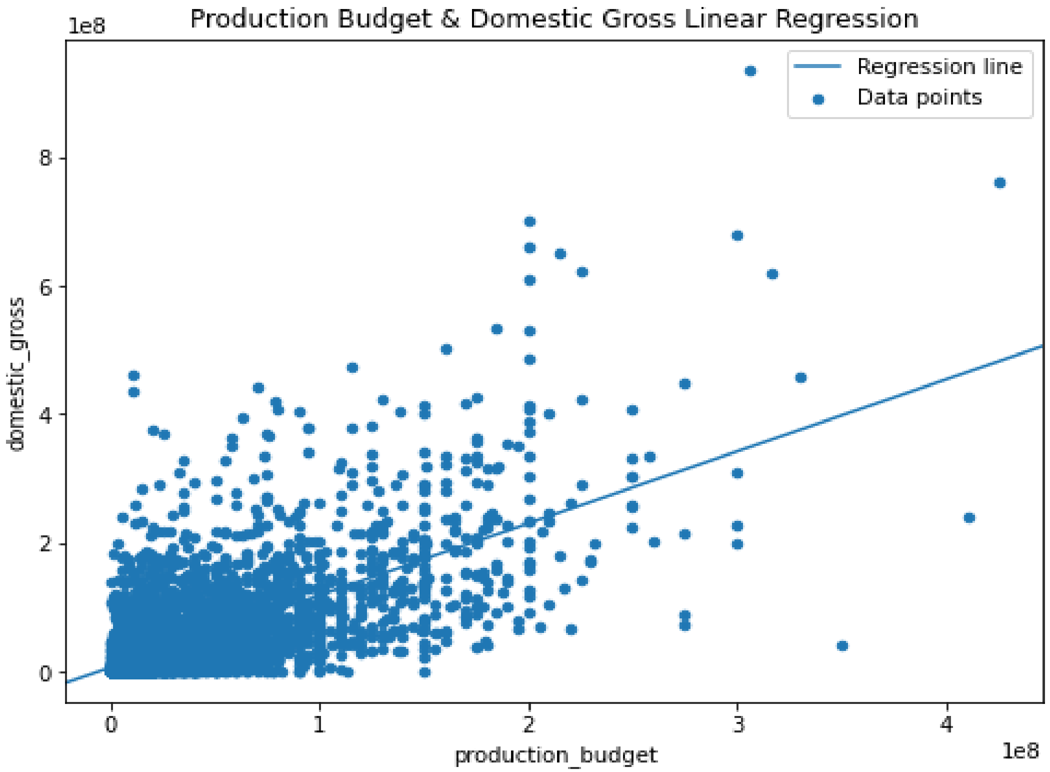
2. Linear Regression

This is another method that will be used for analysis. In linear regression, we will be able to identify independent and dependent variables and identify if there is a linear relationship between them.

From this we want to know if we will be able to predict the values of the dependent variable from the independent variable and find out if there is a linear relationship between the two variables.

Independent and dependent variables will be identified from the various datasets. Methods such as r-squared will be used to interpret what percentage of the dependent variable is explained from the prediction, f-statistic and p-value.

Examples of linear regression



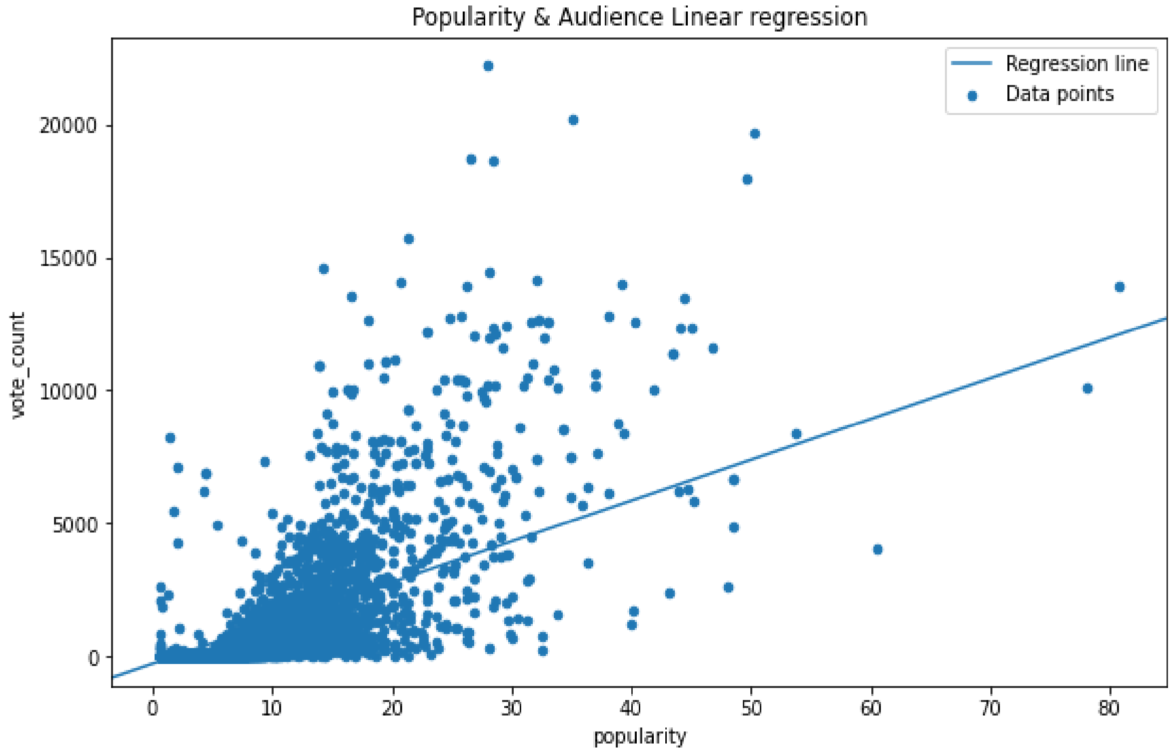
This visualization represents a linear relationship between production budget and domestic gross,

From the relationship between production budget and domestic gross, the rsquared results show that our model explains 47% of the variance in domestic gross the dependent variable.

The visualization of the relationship between production budget and domestic gross, it shows that the linear relationship between the two is quite minimal. This means that a higher production budget can be able to gross a high domestic gross but that is not a guarantee as not every time a film with a high production budget generates a high domestic gross.

This is seen as some movies with low production budget are able to generate a high domestic gross. This means it is not a guarantee that a high production budget predicts a high domestic gross.

Examples of linear regression



This visualization represents the relationship between popularity and vote count.

From the relationship between popularity and vote count, the rsquared results show that our model explains 48% of the variance vote count the dependent variable.

The visualization shows a linear regression relationship between popularity and vote count. The trend shows that higher popularity sometimes lead to higher audience watch time in watching the films.

However the relationship is not very strong as some movies which are quite popular do not have many audience watch time and some movies that are not that popular receive many audience watch time.

Our linear regression between the two variables can be to predict that some popular films are able to predict a high number of watch time from the audience.

Business Recommendations

Recommendation 1

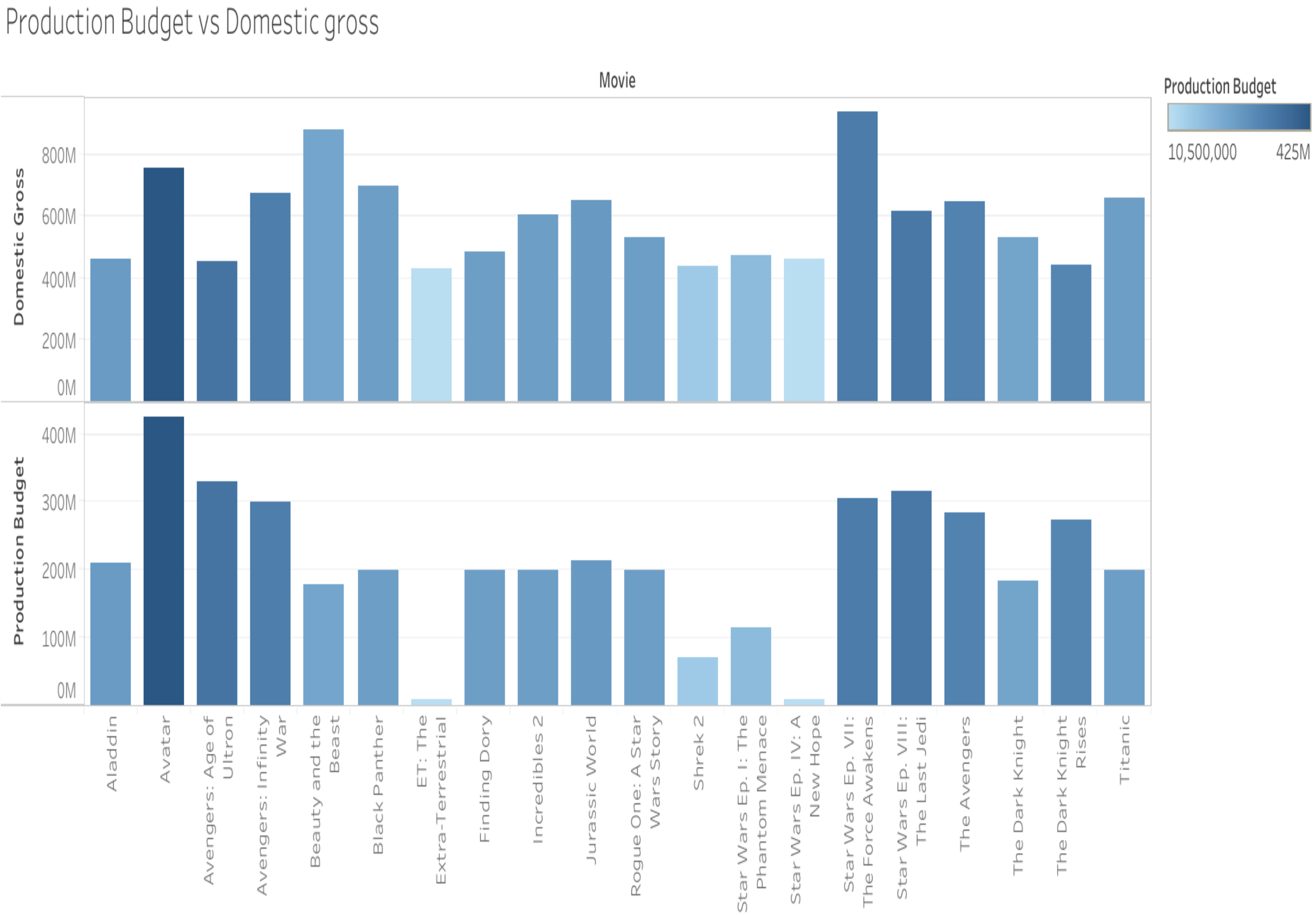
My first business recommendation for the new movie studio, is to emphasize more on what the audience like to see.

We all know that every audiences have different preferences according to age, hobbies and characteristics as they watch films that relate to them.

As a new movie studio, I would engage the movie studio film and production who are one of the stakeholders to prioritize genres as a way of attracting different audiences to watch the films. As different audiences have different preferences when it comes to genres.

Prioritizing genres can lead to popularity of films making the popular at the box office.

Example Production Budget vs Domestic Gross



This visualization shows the relationship between production budget and domestic gross.

It shows that some movies with high production budget tend to gross a lot of money in terms of domestic gross but not all the time as some movies with low production tend to gross more money than movies with higher production budget.

Recommendation 2

Another business recommendation, I would recommend to the new movie studio is a way of promoting the films. This can be done through the marketing and sales team.

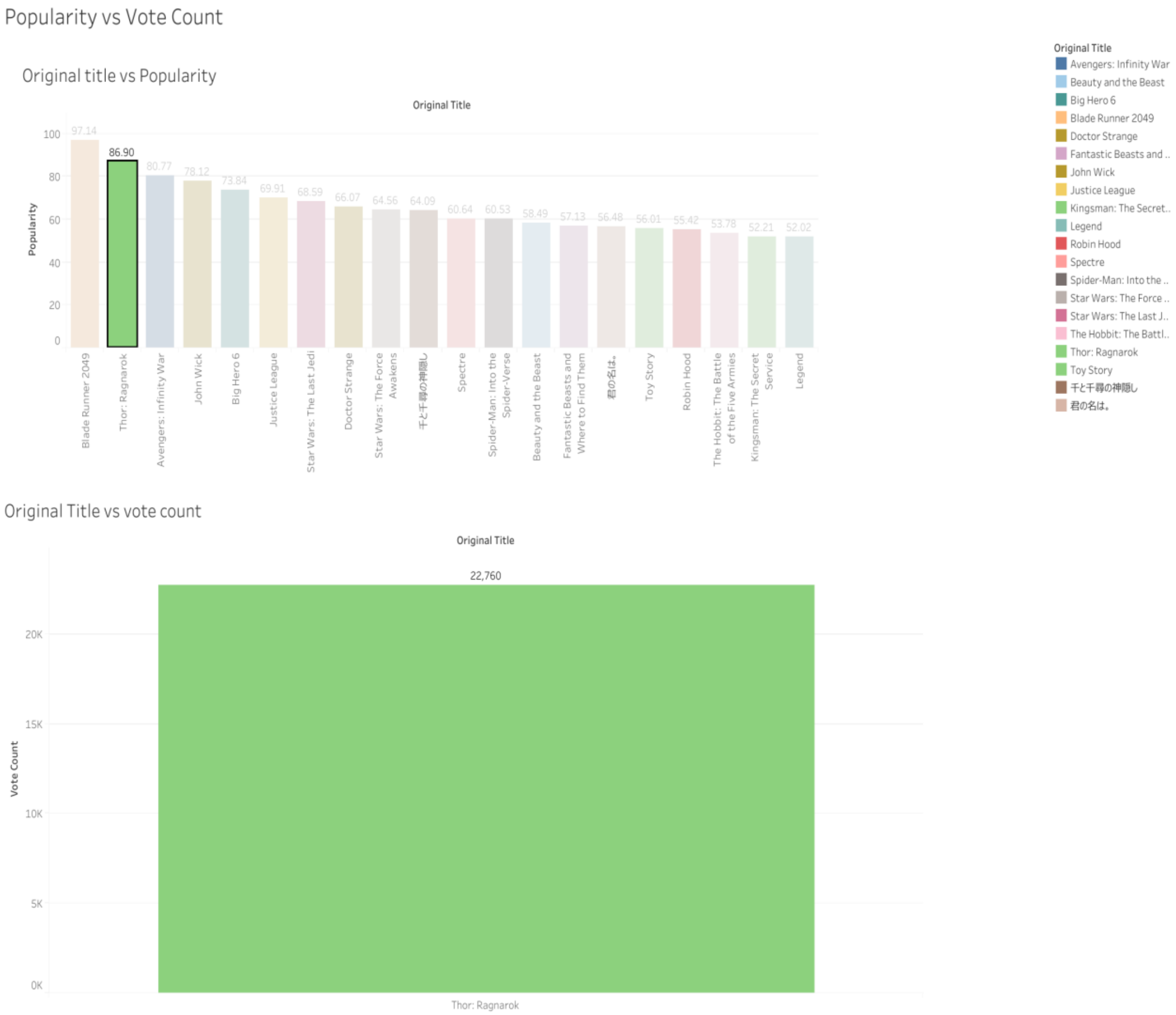
By doing this, this makes the films become popular among the audience. As seen through the linear regression between popularity and vote count, popularity of a film influences the number of people who end up watching the film based on the popularity.

This leads to the movie generating money in the box office. The number of people who end up watching the movie also influences the rating of the movies.

Movies that have a high number of audiences receive a high rating from the audience influencing more people to watch the movies.

Increasing the gross income of the movies in the box office.

Example Popularity vs Vote Count



The visualization shows the relationship between popularity and vote count.

As seen thor rangnorok is the second popular film with a popularity vote of 86.90 and has a total vote count of 22,760 making the 4th most watched film, showing that the popularity of a film influences the number of people who go to watch it.

Recommendation 3

The last recommendation I would the movie studio to prioritize production budget which is used by the film production team. Production budget is essential as it helps the production team to create a team that meets the audience requirements.

The production budget is approved by the company executives and provided by the investors who are stakeholders of the movie studio.

A high production sometimes guarantees that a movies grosses a lot of income in the box office in the terms of domestic and world wide gross, but sometimes not all the time as some movies with a low production budget grosses more than movies with a high production budget as per the analysis.