MICROSOFT MOVIE STUDIO

The Future of Film Making

Created by Richard Taracha

Understanding The Context

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Specifying The Data Analysis Question

- Explore what types of films are currently doing the best at the box office.

- Translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Defining The Metrics of Success

The project will be considered a success when I'm able to provide three concrete business recommendations to the head of Microsoft's new movie studio.

While performing Exploratory Data Analysis I was required to carry out the following steps:

- **Showcase my work value with clear communication.**
- **Craft a compelling story for my movie studio project.**
- **Use visuals to explore data and present findings. Choose simple and relevant charts with good formatting.**

Recording The Experimental Design

Created by Richard Taracha

Strategic Project

METHODOLOGY

1. Load the necessary libraries and datasets for our analysis.
2. Perform data cleaning and pre-processing where necessary.
3. Carry out our analysis
4. Interpret and summarize findings.
5. Provide recommendations.

Data Relevance

These datasets provided were aggregated from a number of sources listed below:

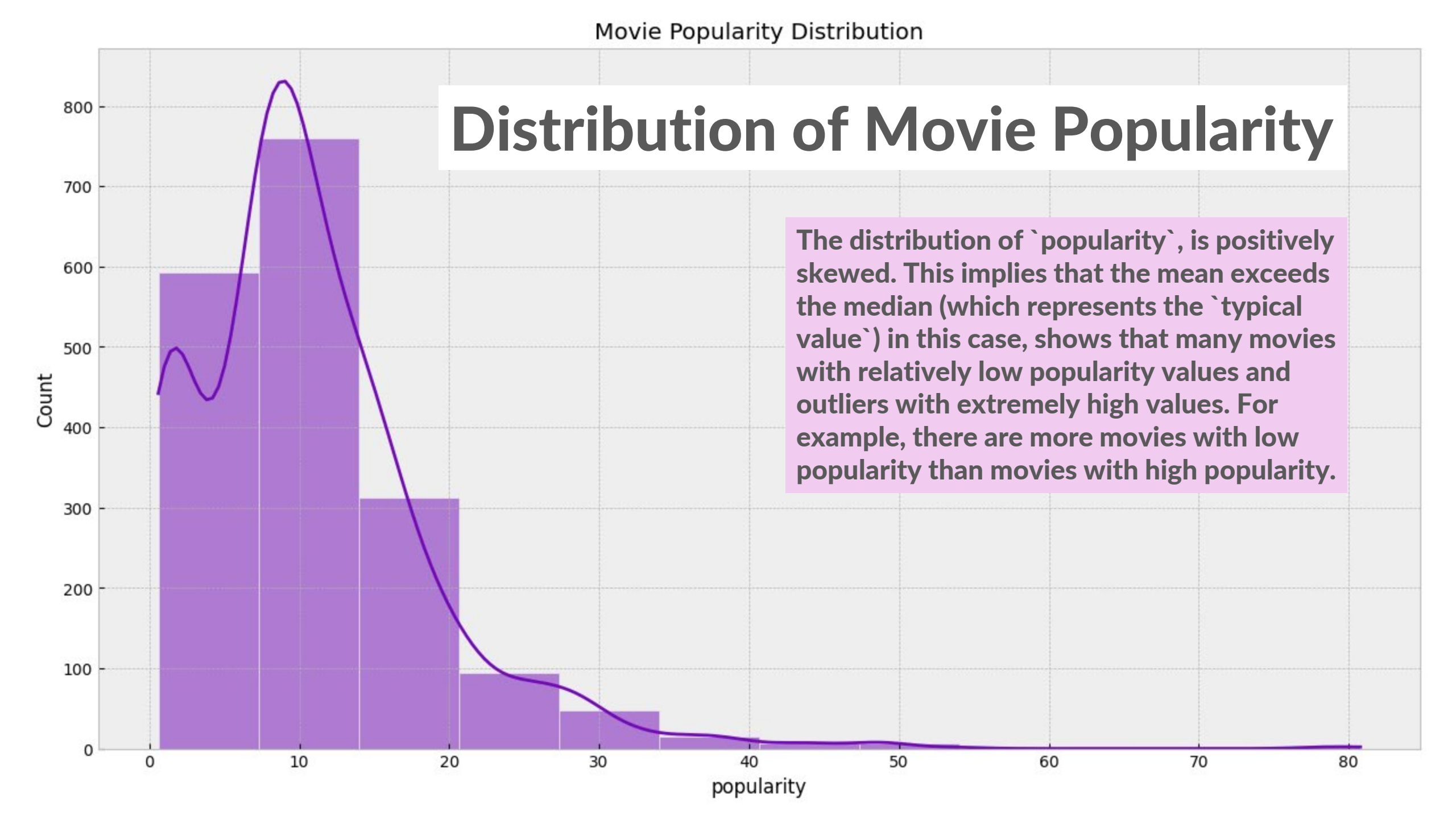
- Box Office Mojo: https://www.boxofficemojo.com/
- IMDB: https://www.imdb.com/
- Rotten Tomatoes: https://www.rottentomatoes.com/
- The MovieDB: https://www.themoviedb.org/
- The Numbers : https://www.the-numbers.com/

The data was relevant to answering our data analysis question.

Exploratory Data Analysis.

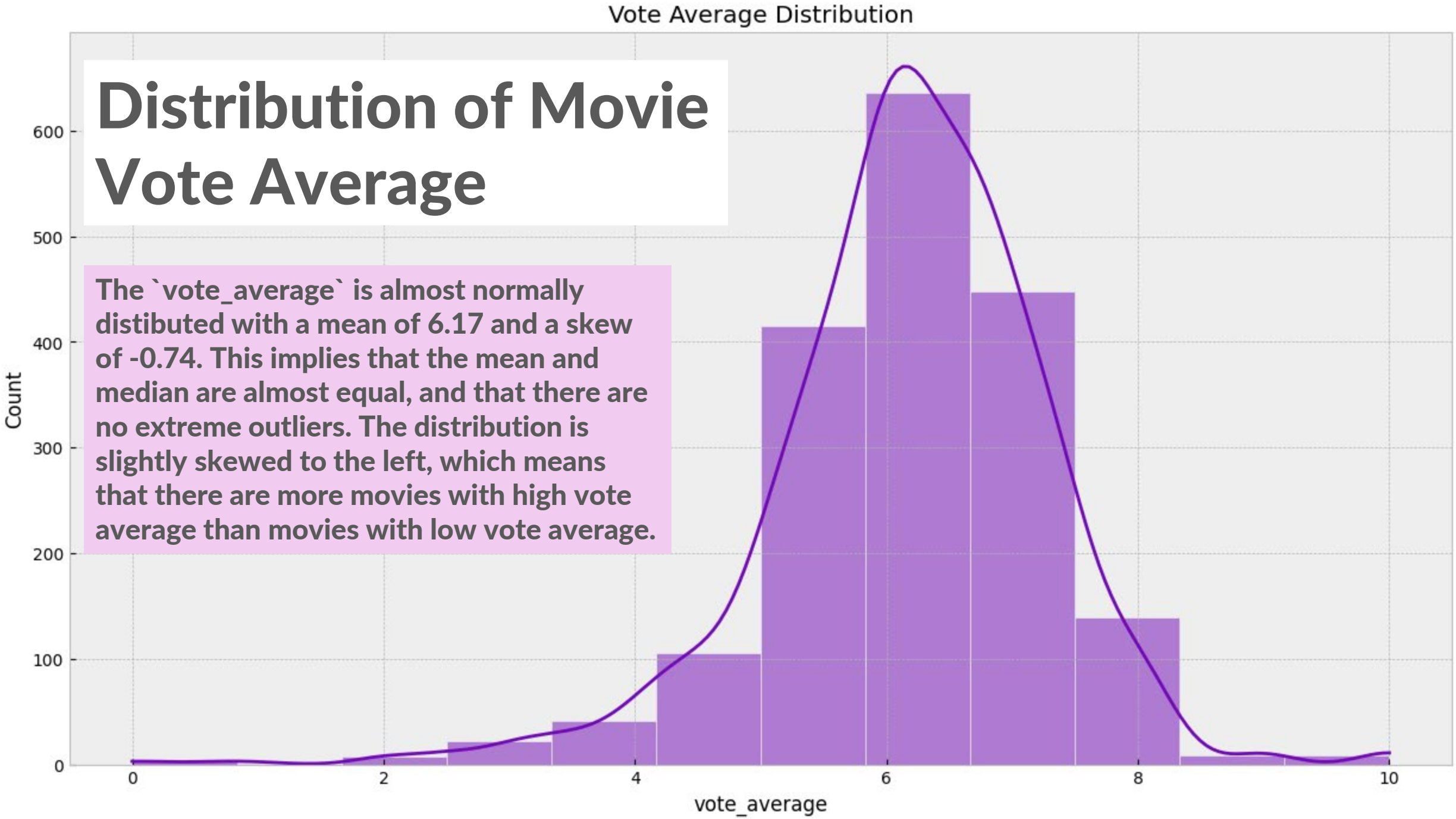
This process of examining and summarizing data sets using various techniques such as visualization and descriptive statistics to help to identify patterns and relationships in the data.

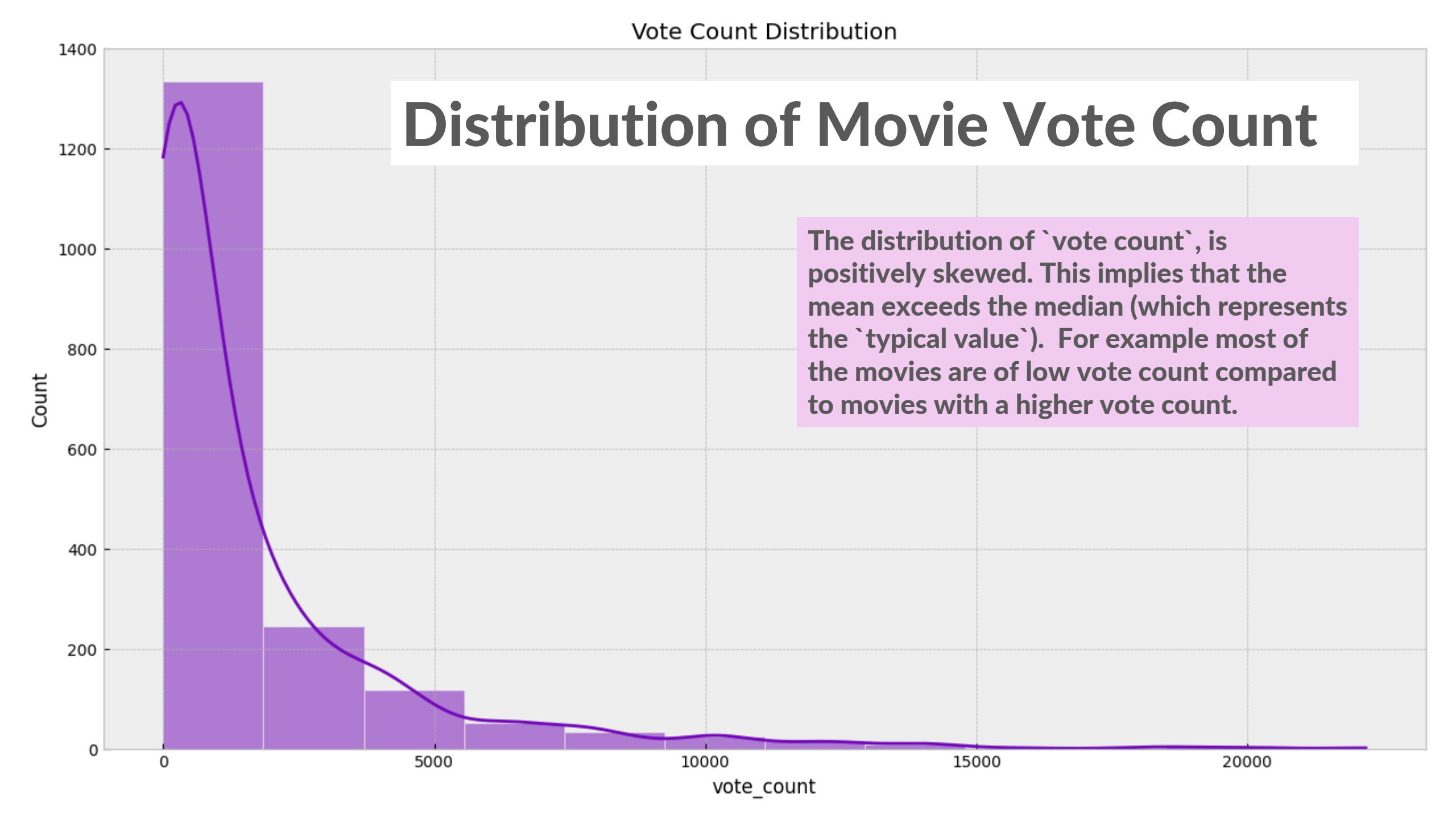
- Univariate Data Analysis.
- Bivariate Data Analysis.
- Multivariate Data Analysis.

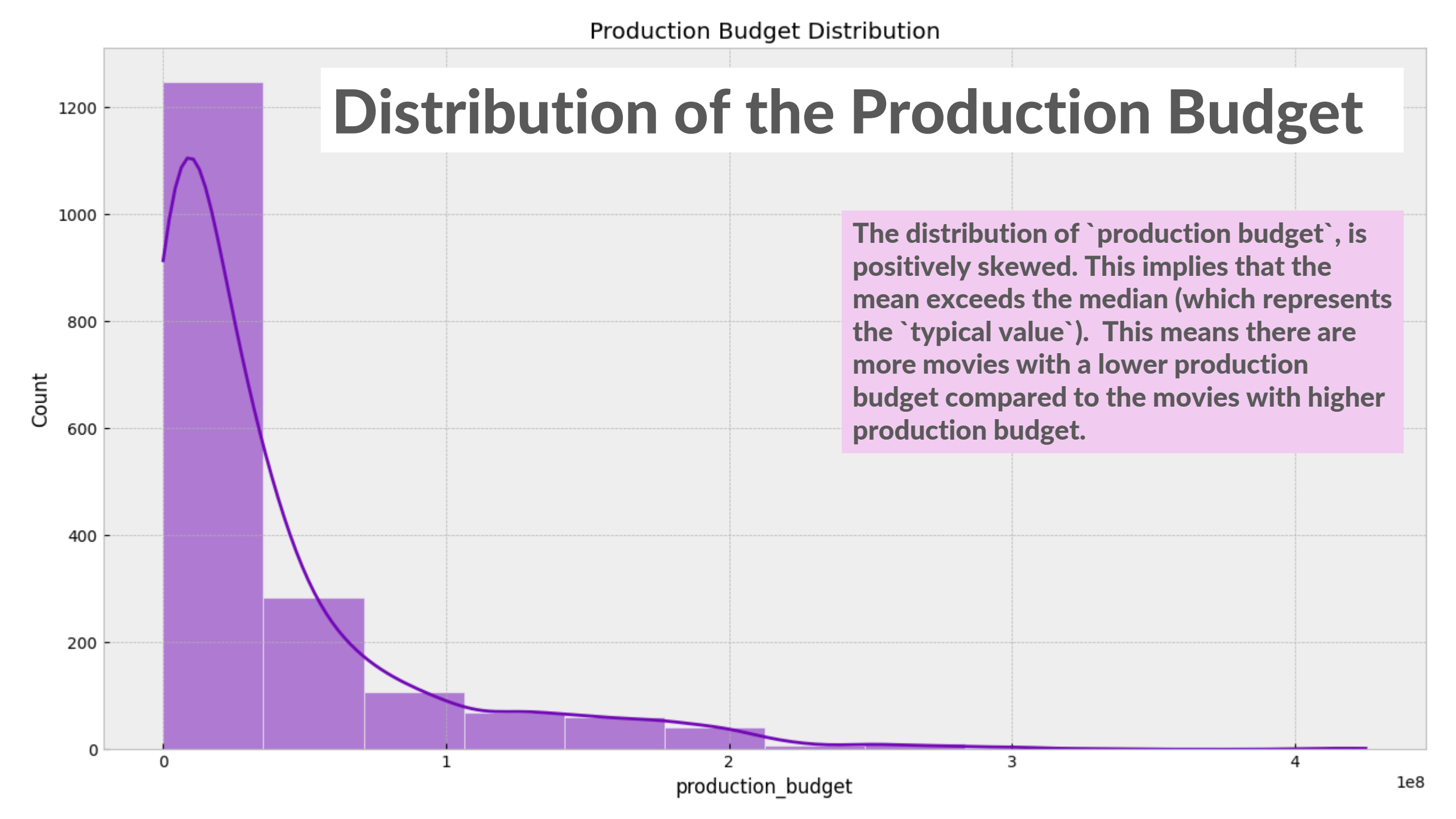

Movie Popularity Distribution
Distribution of Movie Popularity
The distribution of `popularity`, is positively skewed. This implies that the mean exceeds the median (which represents the `typical value`) in this case, shows that many movies with relatively low popularity values and outliers with extremely high values. For example, there are more movies with low popularity than movies with high popularity.
Count
popularity
0
100
200
300
400
500
600
700
800
0
10
20
30
40
50
60
70
80

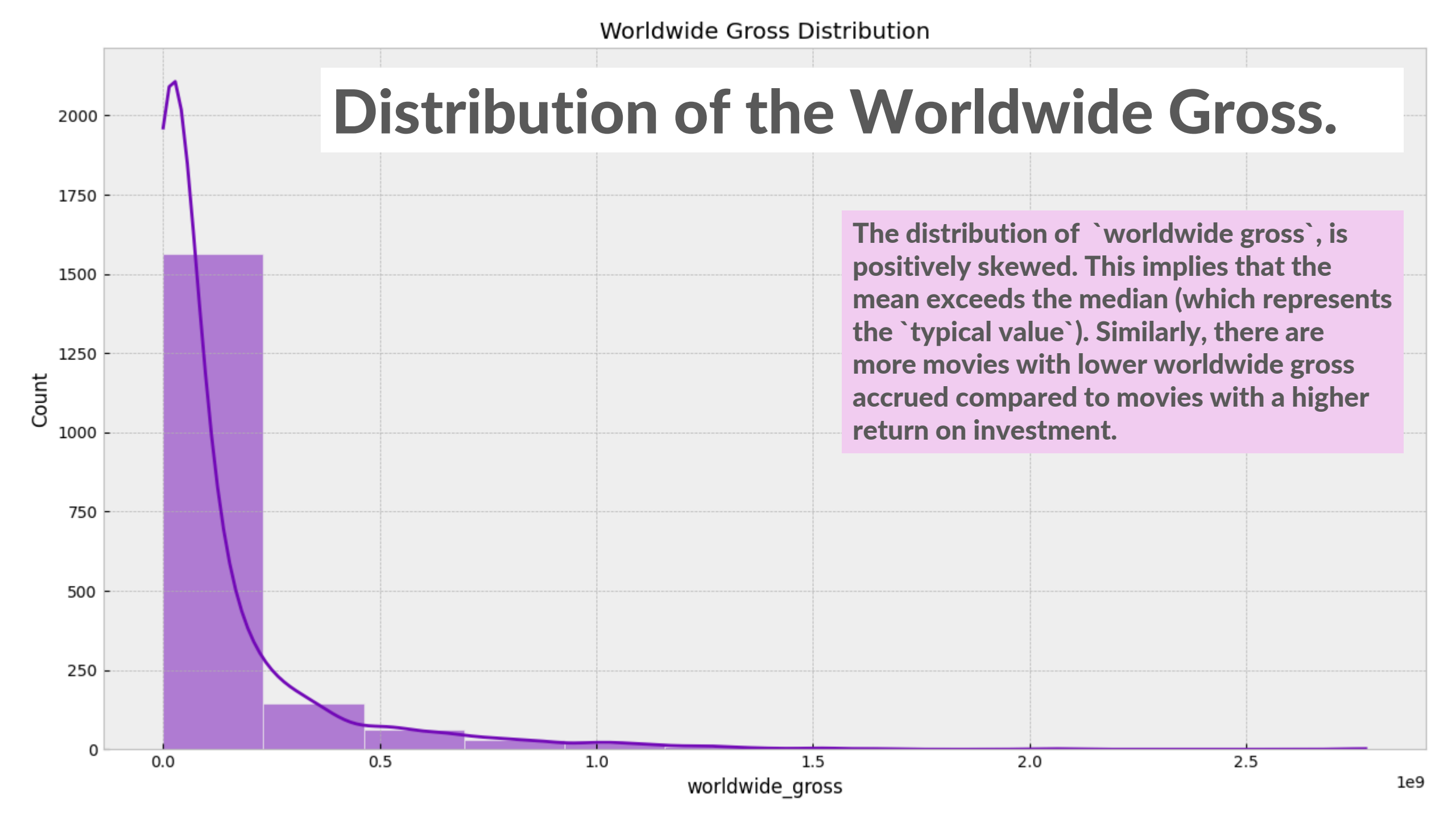
Distribution of Movie Vote Average

The `vote_average` is almost normally distibuted with a mean of 6.17 and a skew of -0.74. This implies that the mean and median are almost equal, and that there are no extreme outliers. The distribution is slightly skewed to the left, which means that there are more movies with high vote average than movies with low vote average.



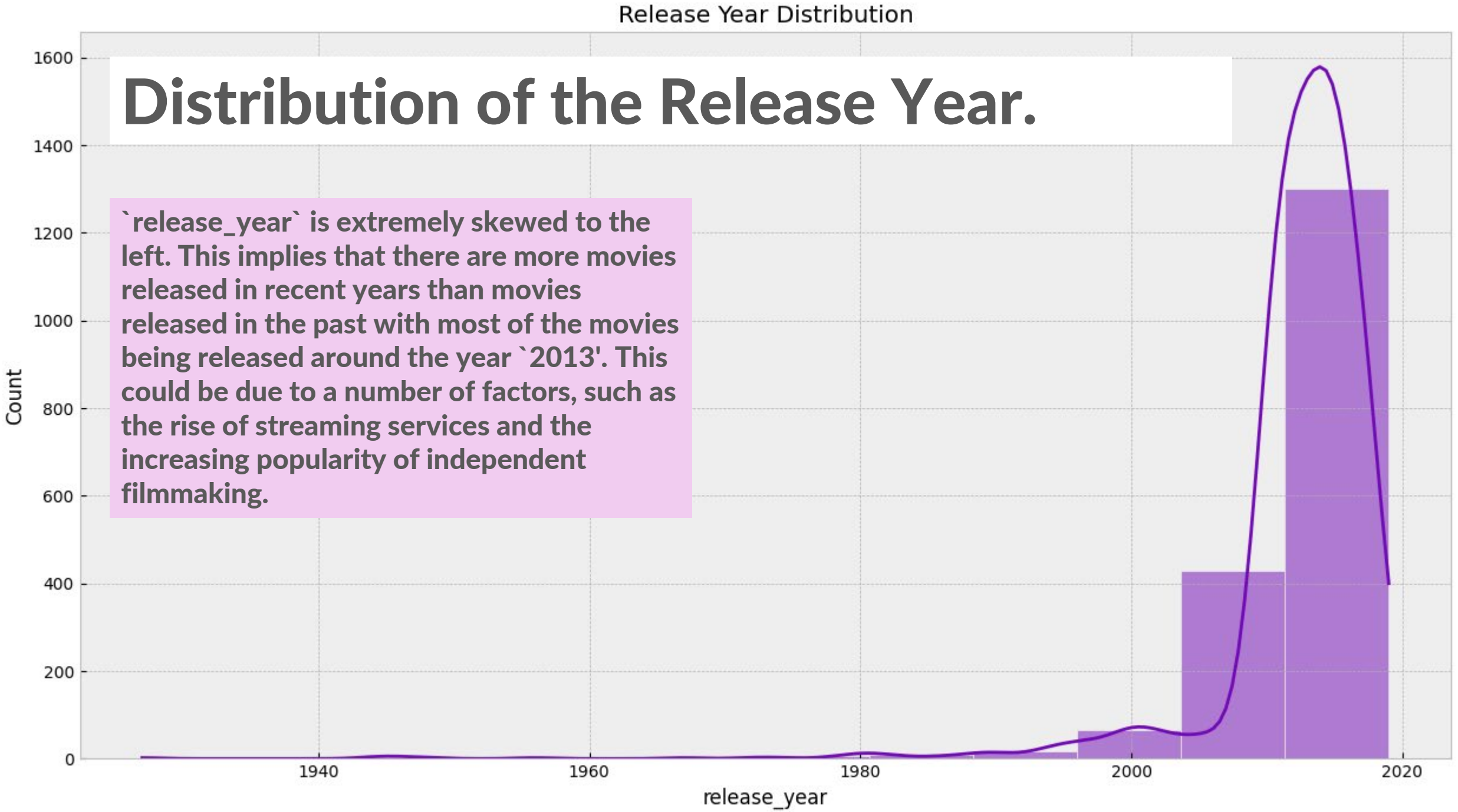

Vote Count Distribution
Distribution of Movie Vote Count
The distribution of `vote count`, is positively skewed. This implies that the mean exceeds the median (which represents the `typical value`). For example most of the movies are of low vote count compared to movies with a higher vote count.
Count
vote_count
0
200
400
600
800
1000
1200
1400
0
5000
10000
15000
20000

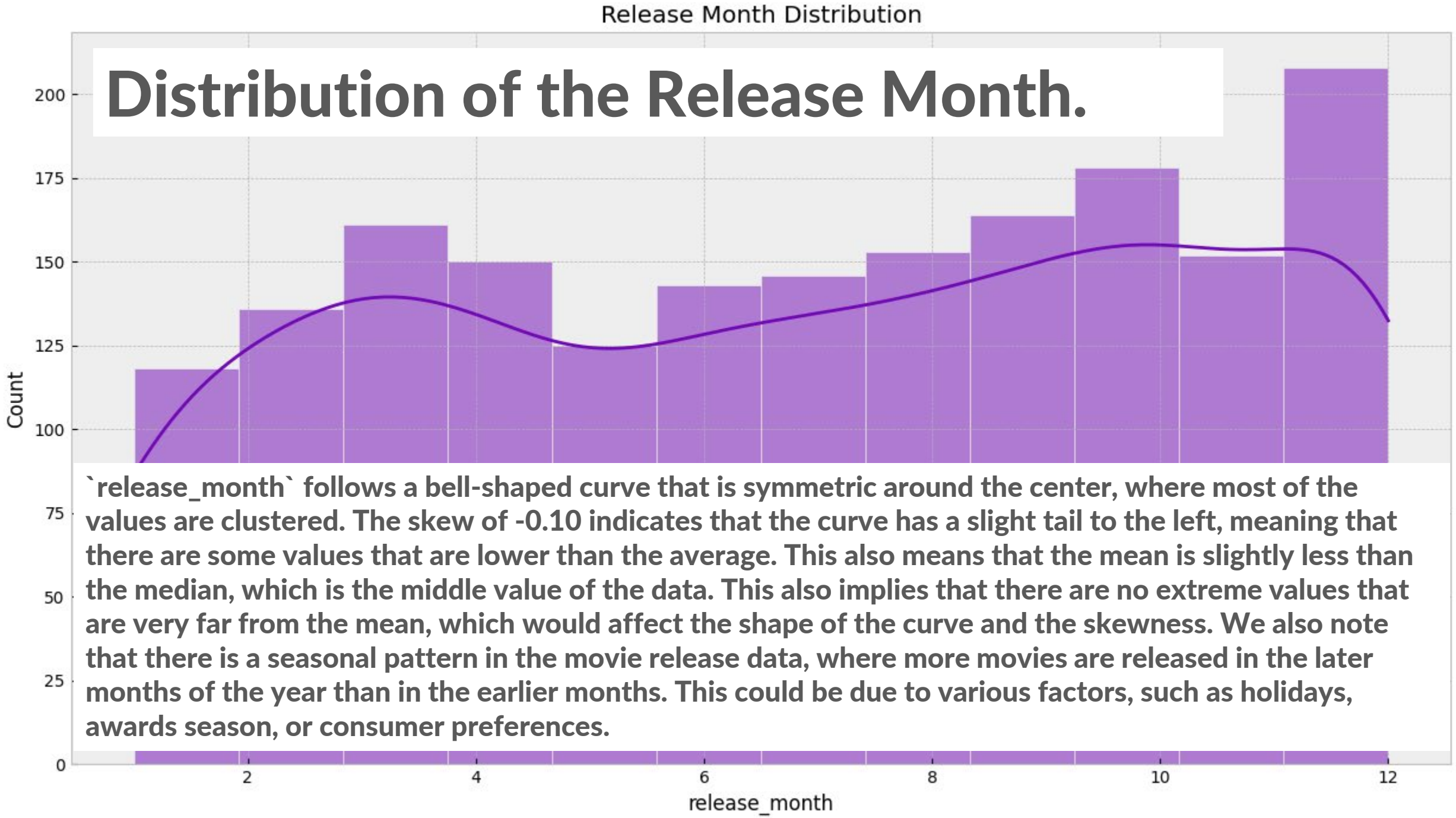

Production Budget Distribution
Distribution of the Production Budget
The distribution of `production budget`, is positively skewed. This implies that the mean exceeds the median (which represents the `typical value`). This means there are more movies with a lower production budget compared to the movies with higher production budget.
Count
1200
1000
800
600
400
200
0
0
1
2
3
4
production_budget
1e8


Worldwide Gross Distribution
Distribution of the Worldwide Gross.
The distribution of `worldwide gross`, is positively skewed. This implies that the mean exceeds the median (which represents the `typical value`). Similarly, there are more movies with lower worldwide gross accrued compared to movies with a higher return on investment.
Count
2000
1750
1500
1250
1000
750
500
250
0
0.0
0.5
1.0
1.5
2.0
2.5
worldwide_gross
1e9

Distribution of the Release Year.

`release_year` is extremely skewed to the left. This implies that there are more movies released in recent years than movies released in the past with most of the movies being released around the year `2013'. This could be due to a number of factors, such as the rise of streaming services and the increasing popularity of independent filmmaking.



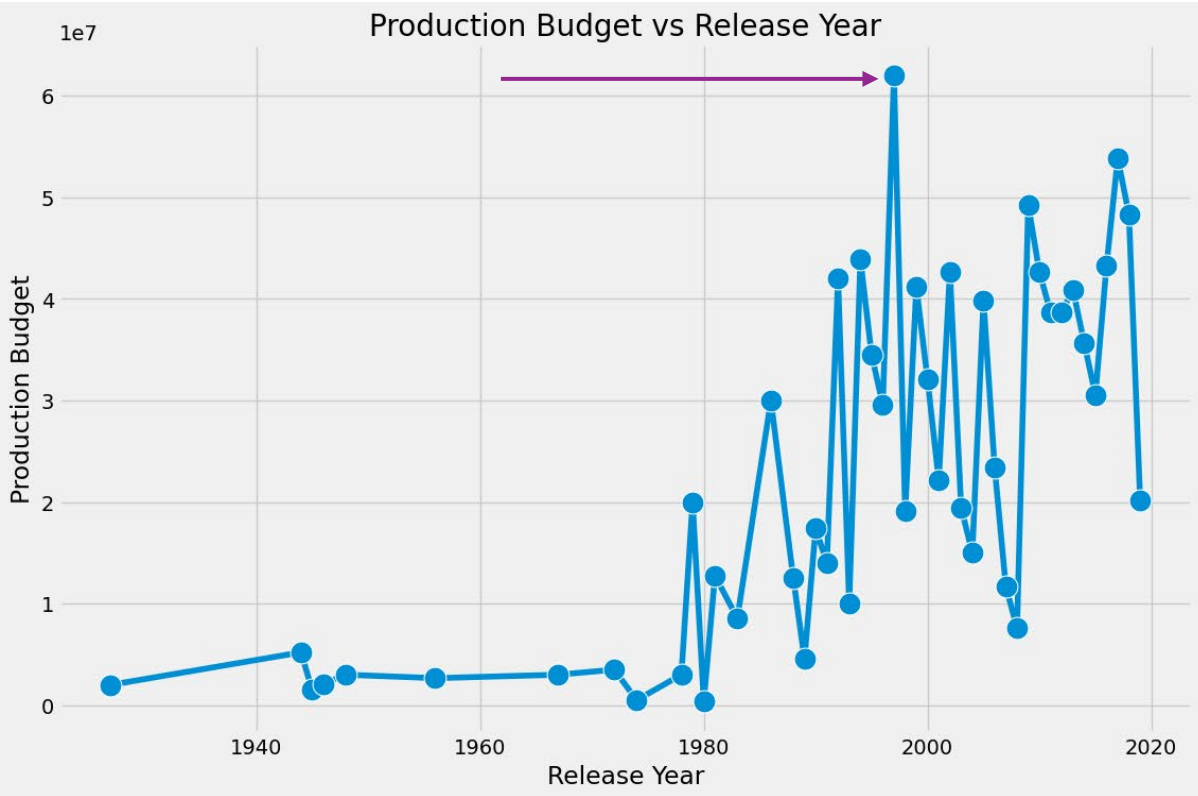

Release Month Distribution
Distribution of the Release Month.
Count
release_month
`release_month` follows a bell-shaped curve that is symmetric around the center, where most of the values are clustered. The skew of -0.10 indicates that the curve has a slight tail to the left, meaning that there are some values that are lower than the average. This also means that the mean is slightly less than the median, which is the middle value of the data. This also implies that there are no extreme values that are very far from the mean, which would affect the shape of the curve and the skewness. We also note that there is a seasonal pattern in the movie release data, where more movies are released in the later months of the year than in the earlier months. This could be due to various factors, such as holidays, awards season, or consumer preferences.

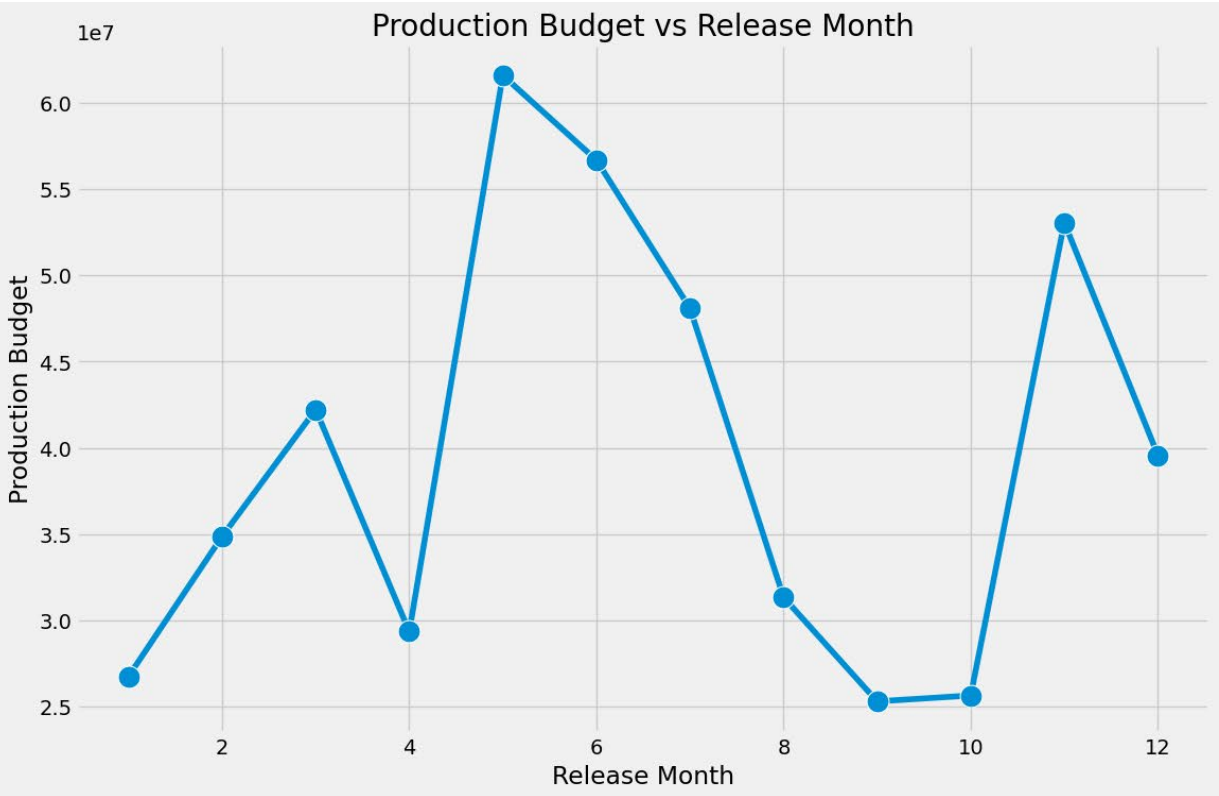
Exploratory Data Analysis.

- Univariate Data Analysis.
- Bivariate Data Analysis.
- Multivariate Data Analysis.

This process of examining and summarizing data sets using various techniques such as visualization and descriptive statistics to help to identify patterns and relationships in the data.

Average Production Budgets by the Release Month & Year.





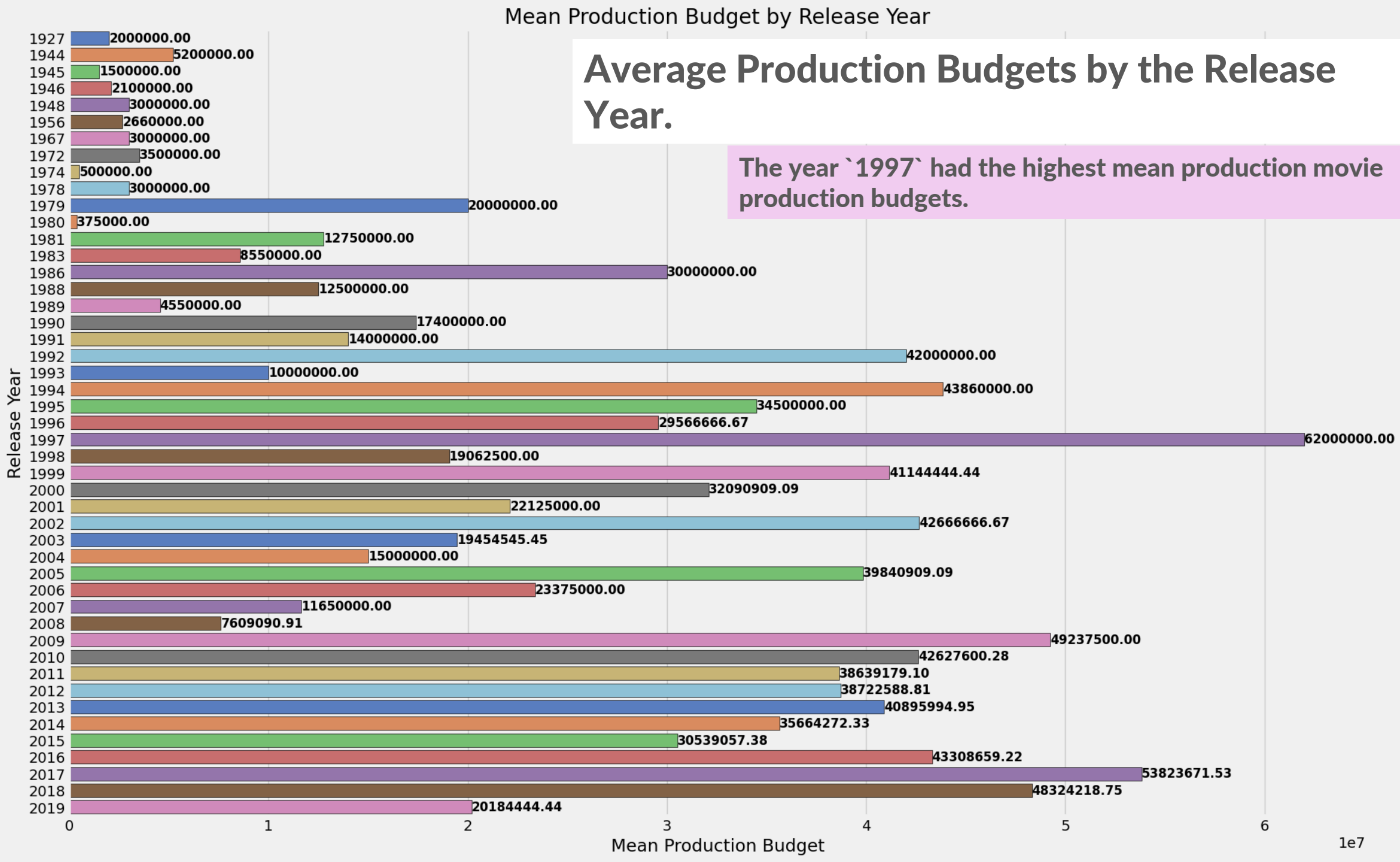
There seems to be a trend with the production budgets over the years. We can note that in the past, product budgets were relatively low but as the years went by, the budgets increased. This could be due to the fact that the movie industry has become more competitive and the need to produce high quality movies has increased. This could also be due to a number of factors, including the rising cost of talent, the increasing demand for special effects, and the decline of the market for low-budget films. However, there are still a number of low-budget films being made today, and some of these films have been successful at the box office.

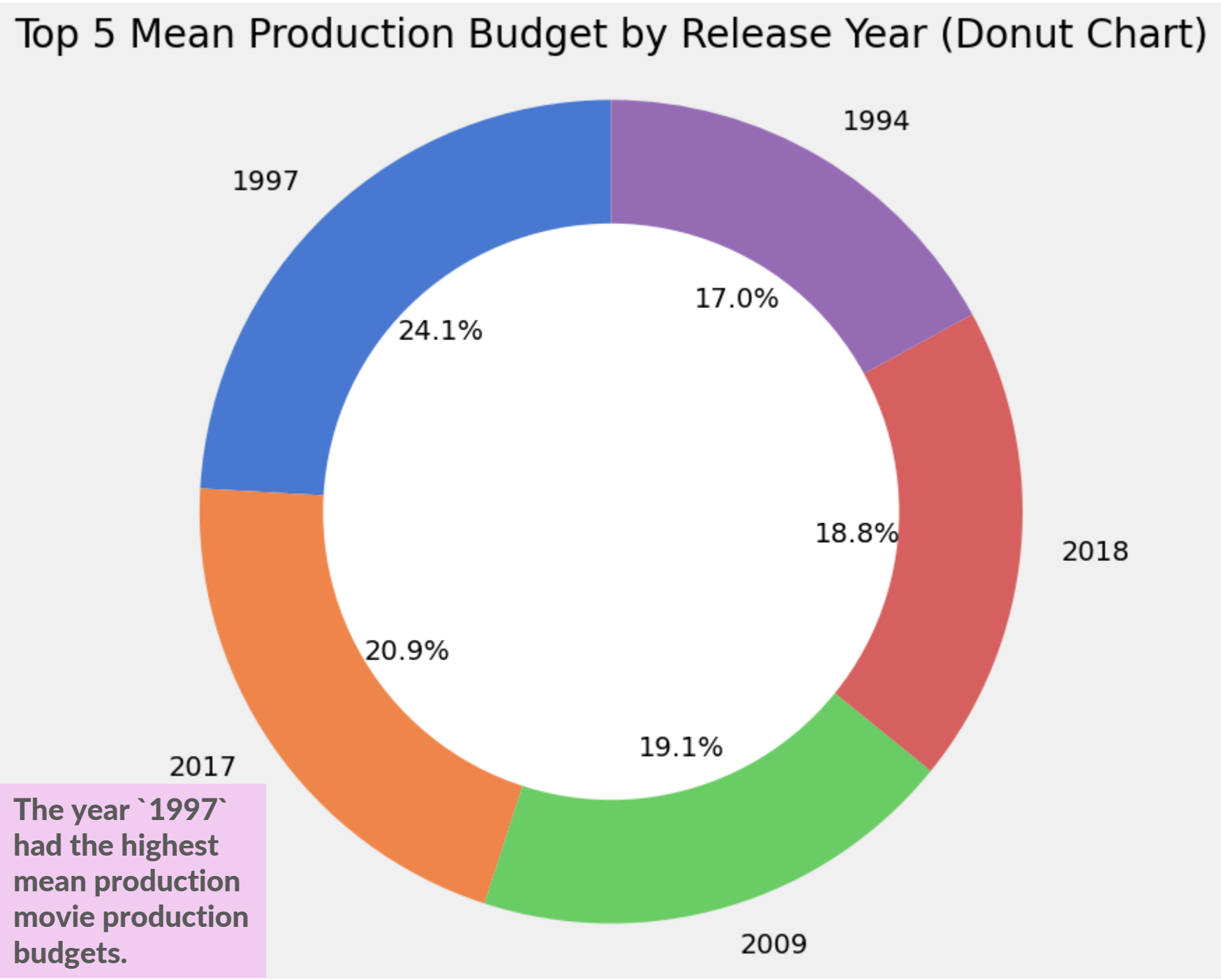
We can see just before we get into the year 2020 there was another sharp decline in the production budgets. The COVID-19 pandemic, which caused widespread economic disruption and led to a decline in consumer demand could be a reason for this decline. This could also be due to the geopolitical tensions, such as the US-China trade war, which created uncertainty and led to businesses to delay or cancel investment projects.

Production budgets for movies tend to increase between the months of may to June. This could be due to the fact that the summer season is the most popular time for movie releases. This is because people are more likely to go to the movies during the summer months, when the weather is nice and they have more free time. Another reason could be that many major film festivals take place in May and June, such as the Cannes Film Festival and the Tribeca Film Festival. These festivals can be a great way to generate buzz for a film and attract investors, which could lead to an increase in the production budget.

Average Production Budgets by the Release Year.

The year `1997` had the highest mean production movie production budgets.

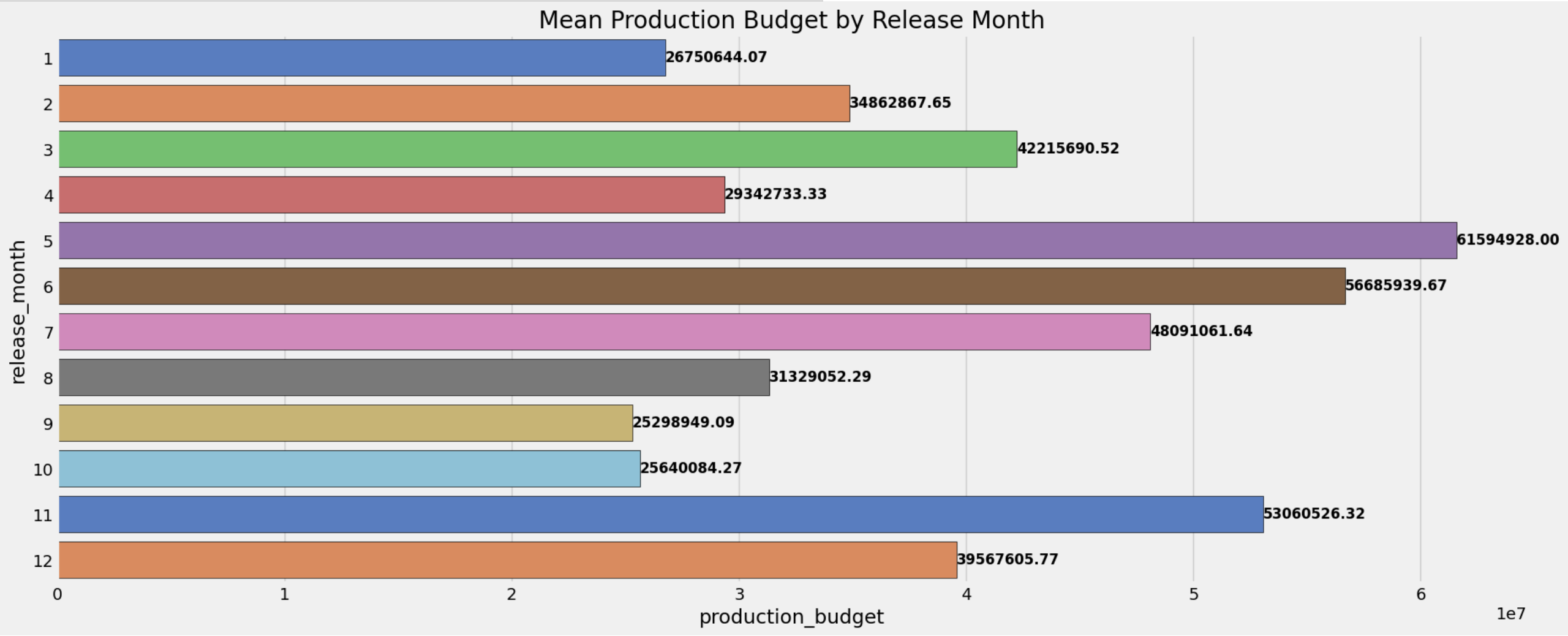




Top 5 years by average production budgets.

The year `1997` had the highest mean production movie production budgets.

Average Production Budgets by the Release Month.

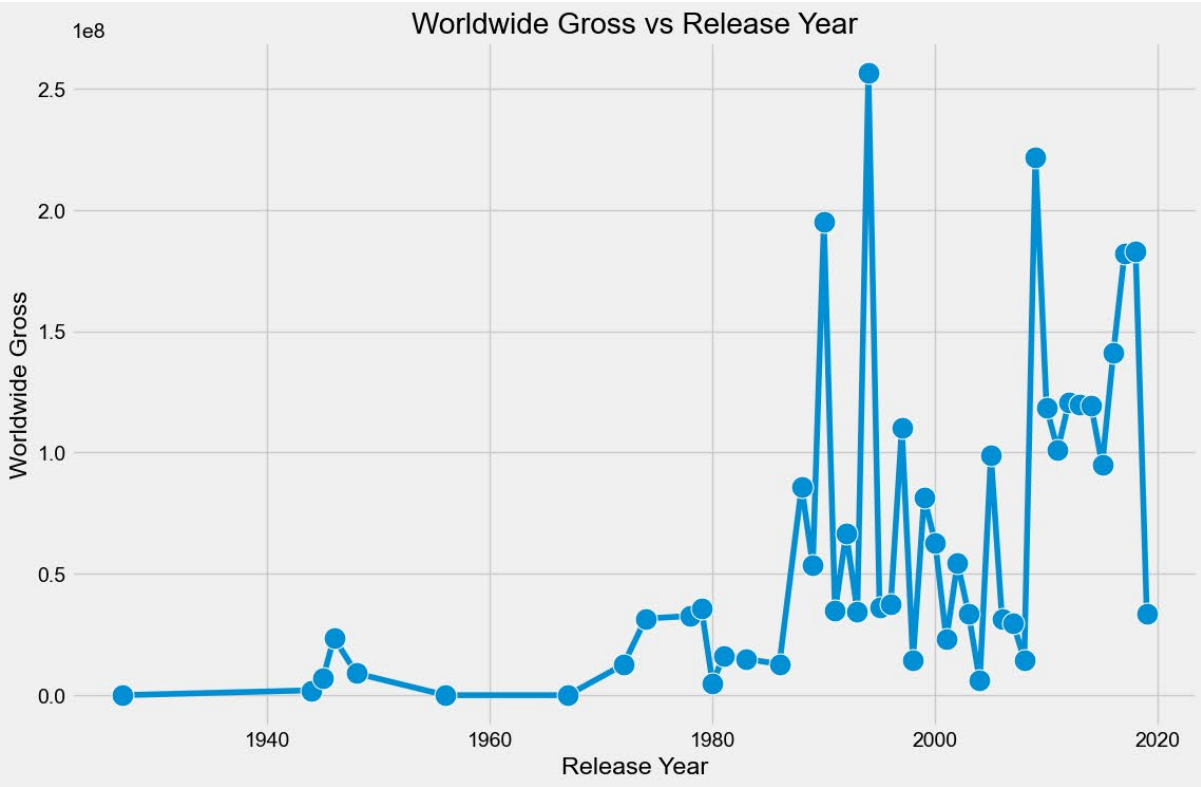


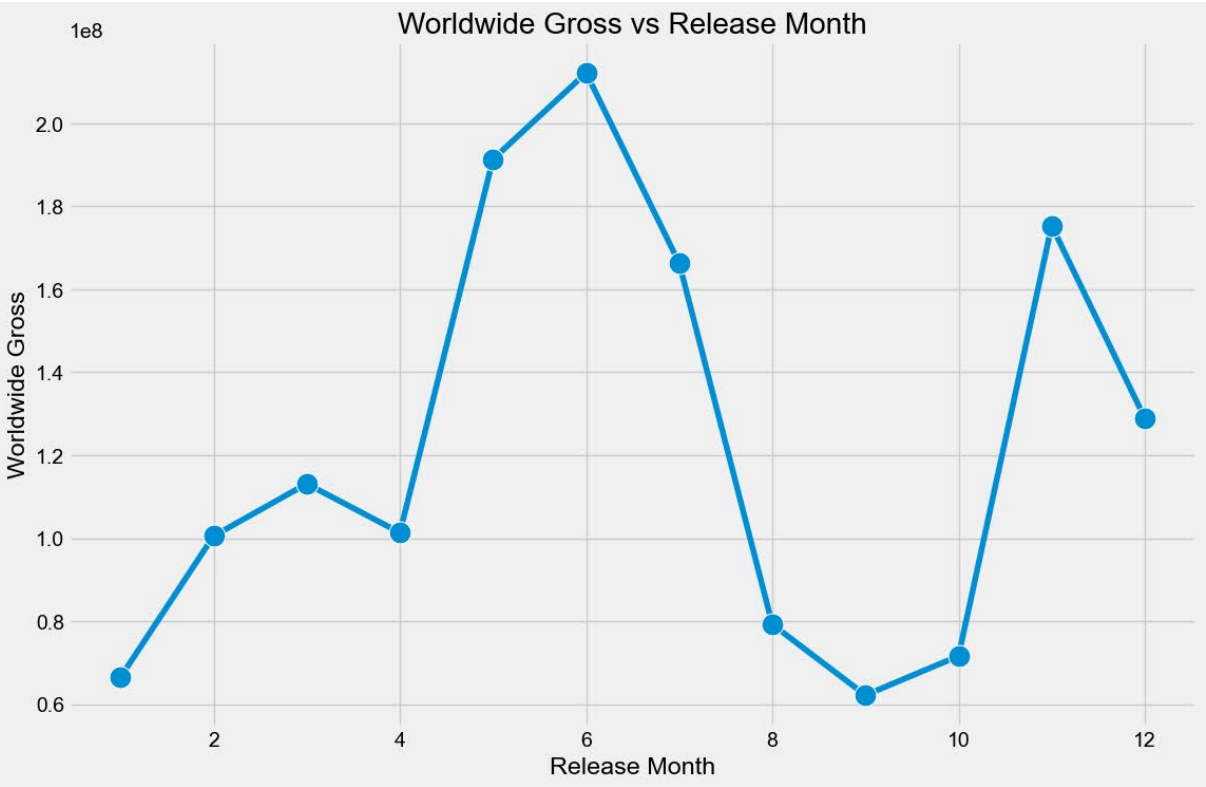
As noted earlier the second half of the year had the most months with the highest mean movie production budgets.

Question:

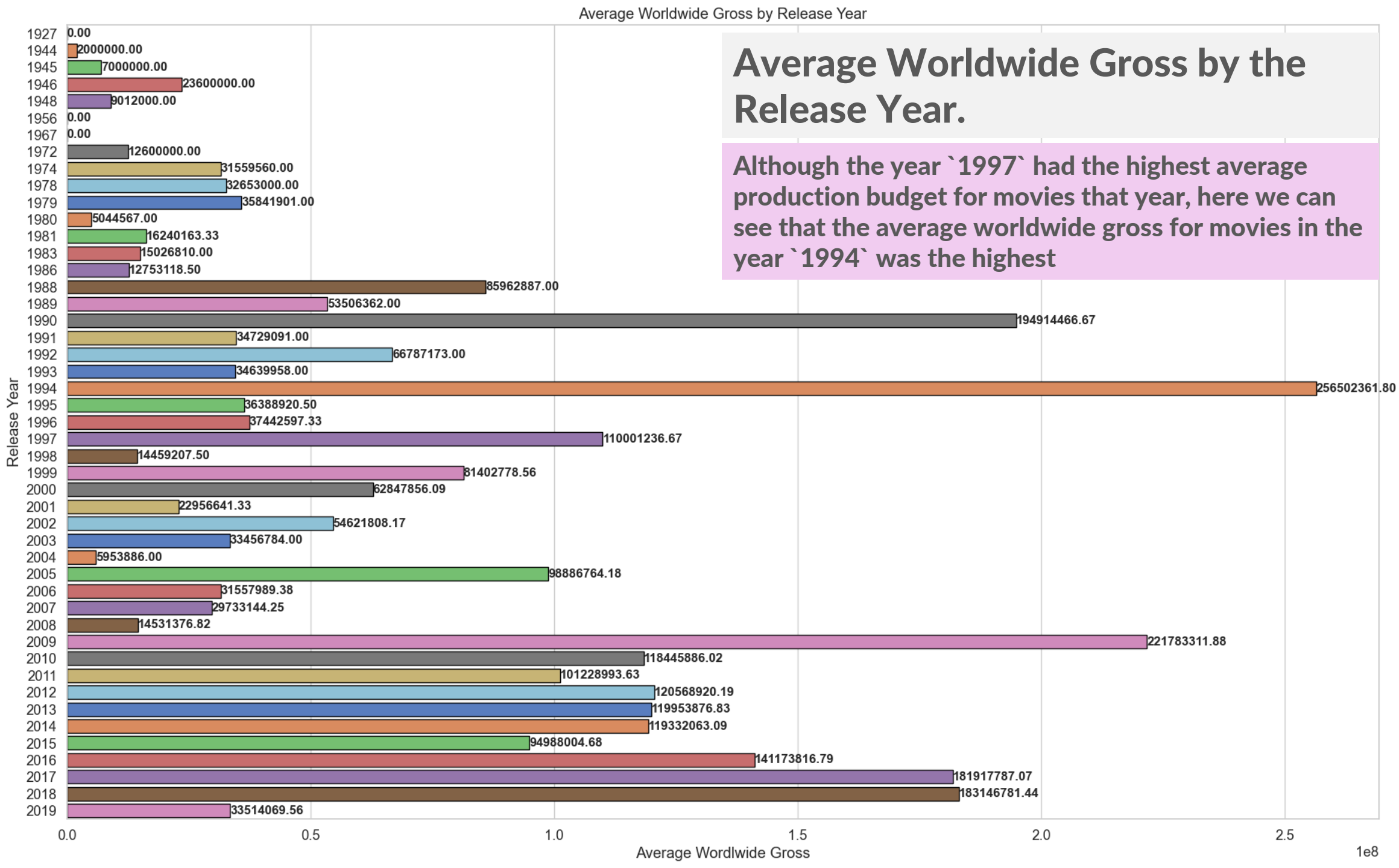
What is the average worldwide gross per month / year?

Average Worldwide Gross by the Release Month & Year.





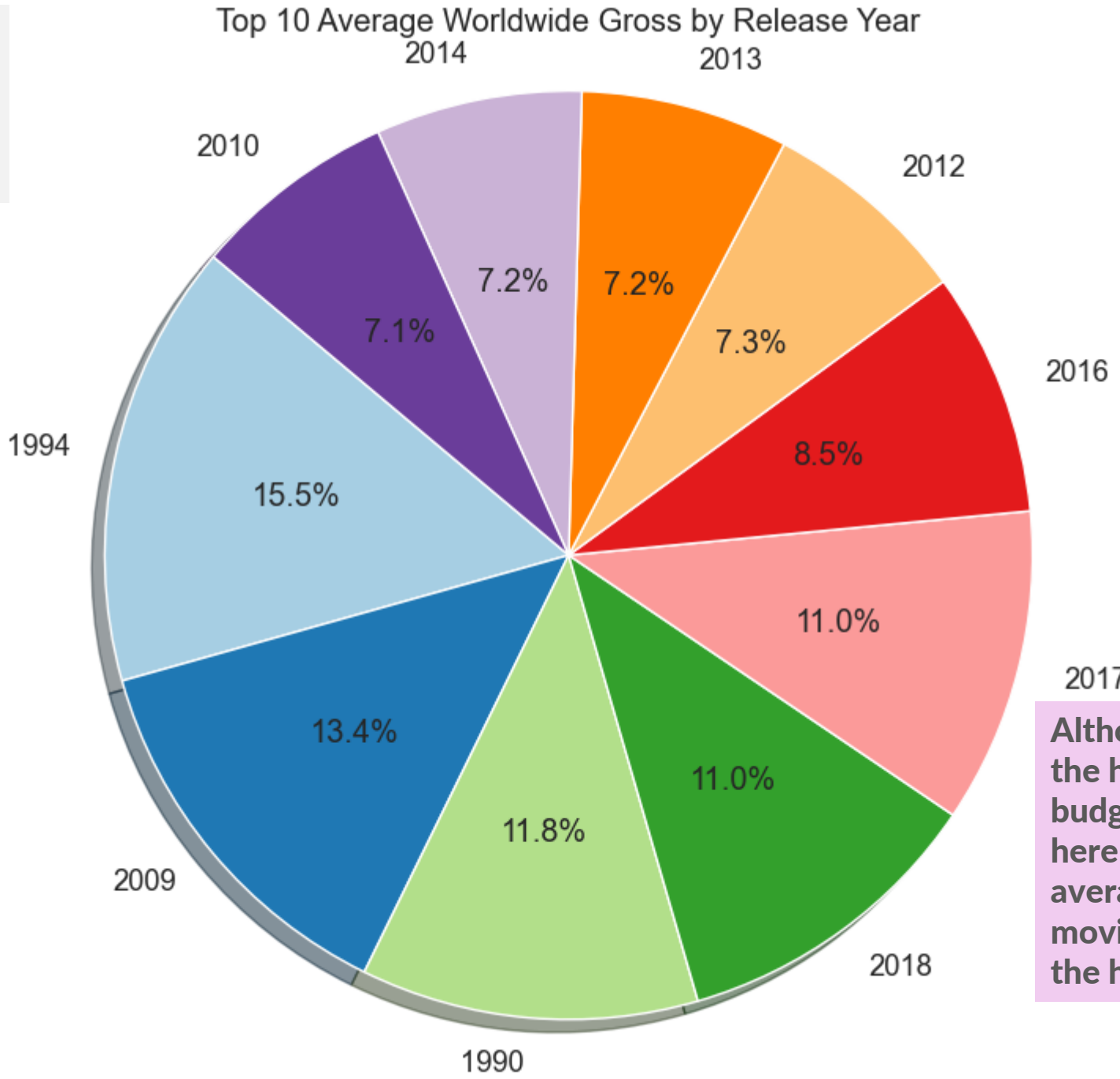
Right off the bat we can see that `production budgets` for movies as well as their `worldwide gross revenue` follow a similar trend that seem increase over the years and have a similar trend observed for both aspects over the months. There could be a correlation between the `production budgets` and the `worldwide gross` for each movie produced and released.



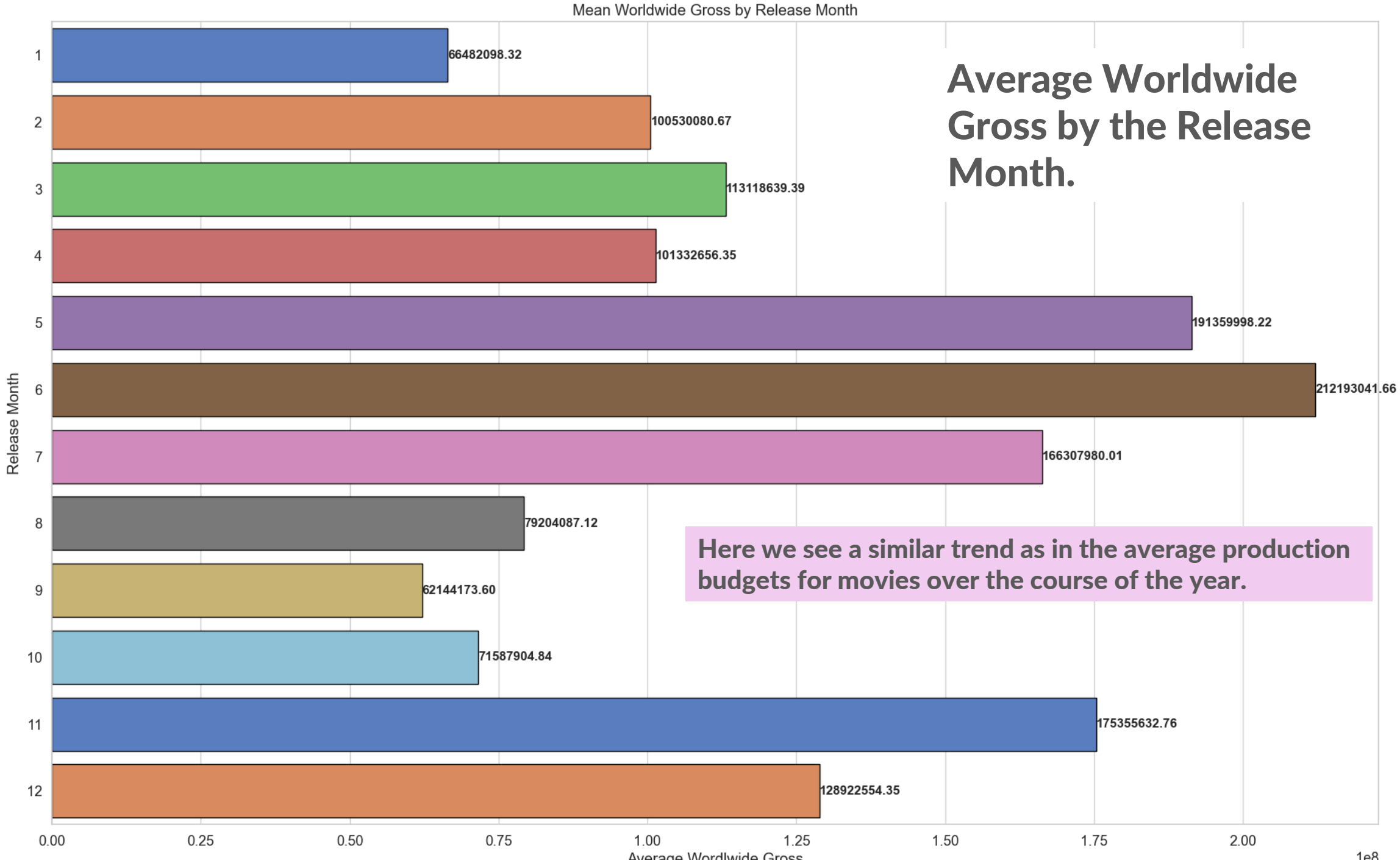
Average Worldwide Gross by the Release Year.

Although the year `1997` had the highest average production budget for movies that year, here we can see that the average worldwide gross for movies in the year `1994` was the highest

Top 5 years by average worldwide gross.



Although the year `1997` had the highest average production budget for movies that year, here we can see that the average worldwide gross for movies in the year `1994` was the highest

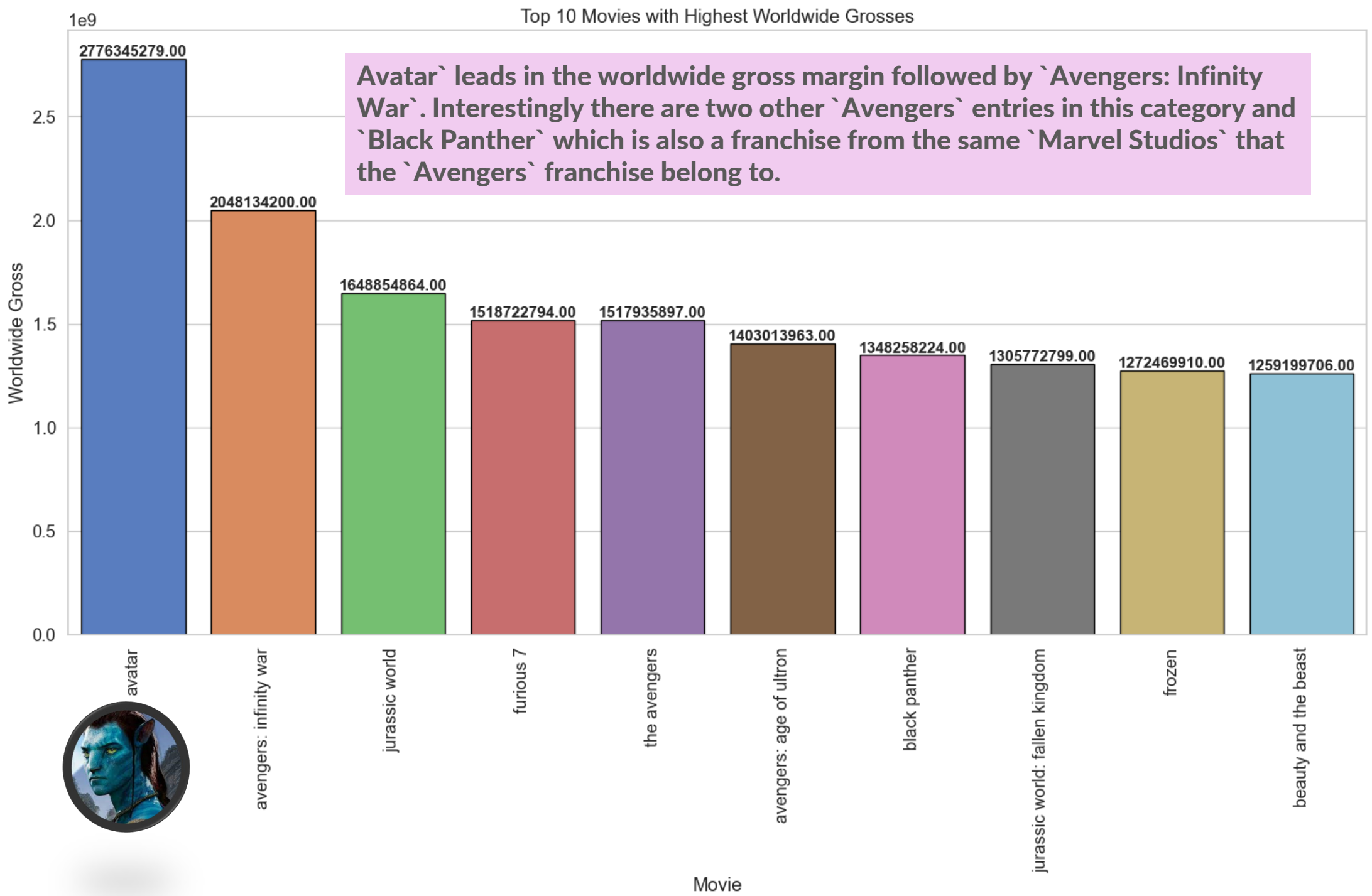


Average Worldwide Gross by the Release Month.

Here we see a similar trend as in the average production budgets for movies over the course of the year.

Question:

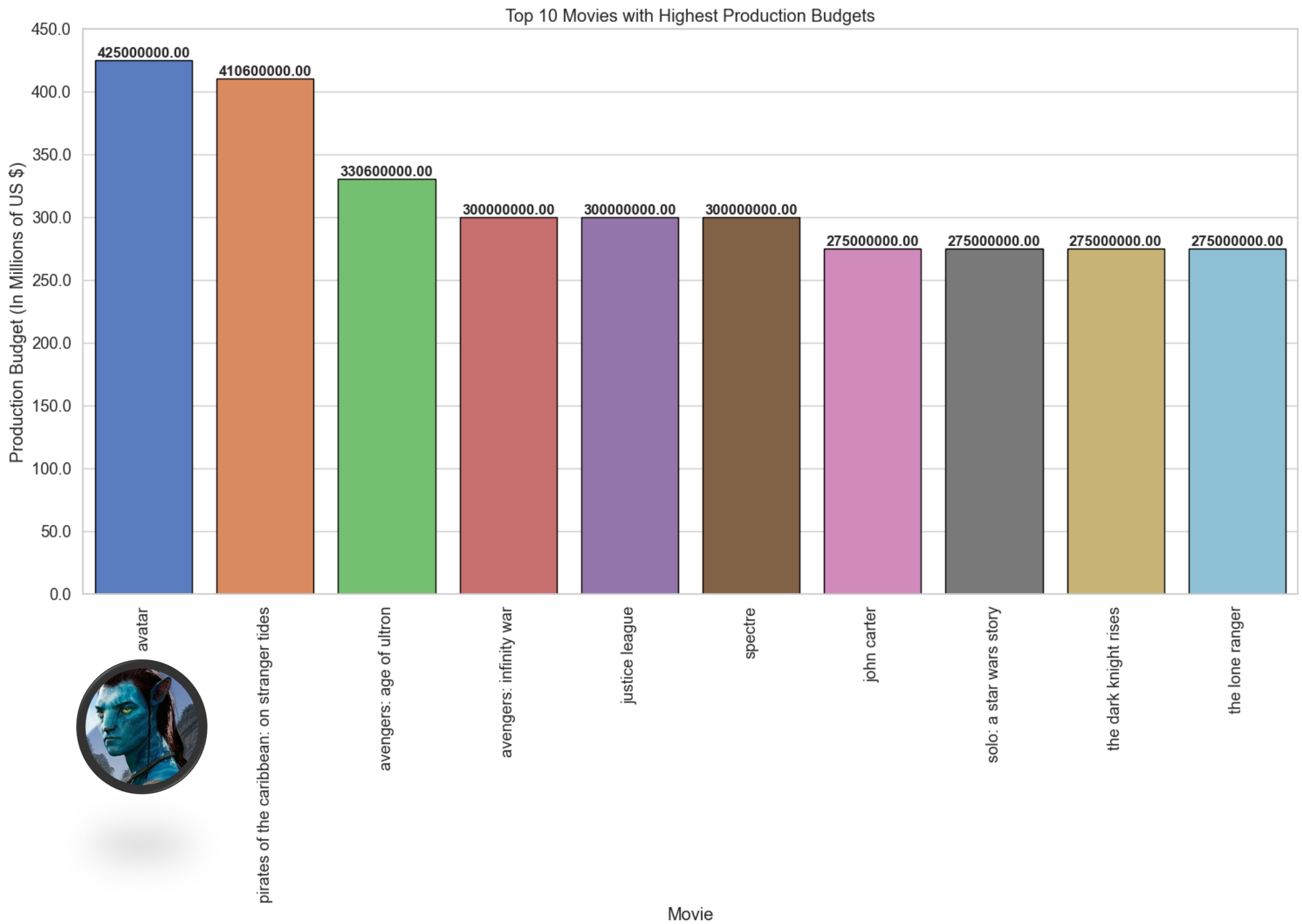
What are the top ten grossing movies worldwide (US $)?



Avatar` leads in the worldwide gross margin followed by `Avengers: Infinity War`. Interestingly there are two other `Avengers` entries in this category and `Black Panther` which is also a franchise from the same `Marvel Studios` that the `Avengers` franchise belong to.

Question:

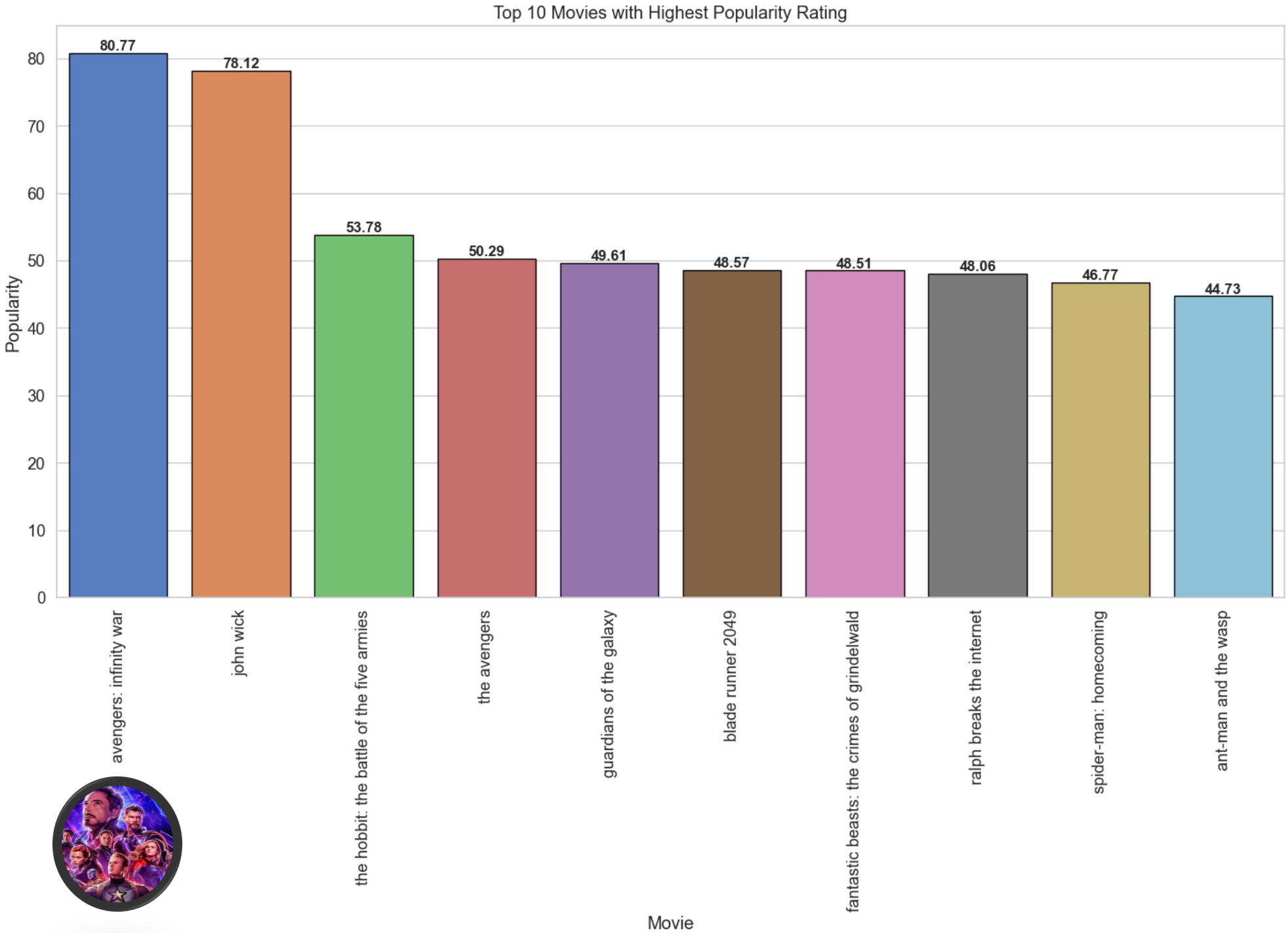
What are the top ten highest project budgets (US $)?



`Avatar` once again leads in the projects with the highest production markets. There are also two entries of the `Avengers` franchise although `Disney's Pirates of the Caribbean` takes the second place in this category.

Question:

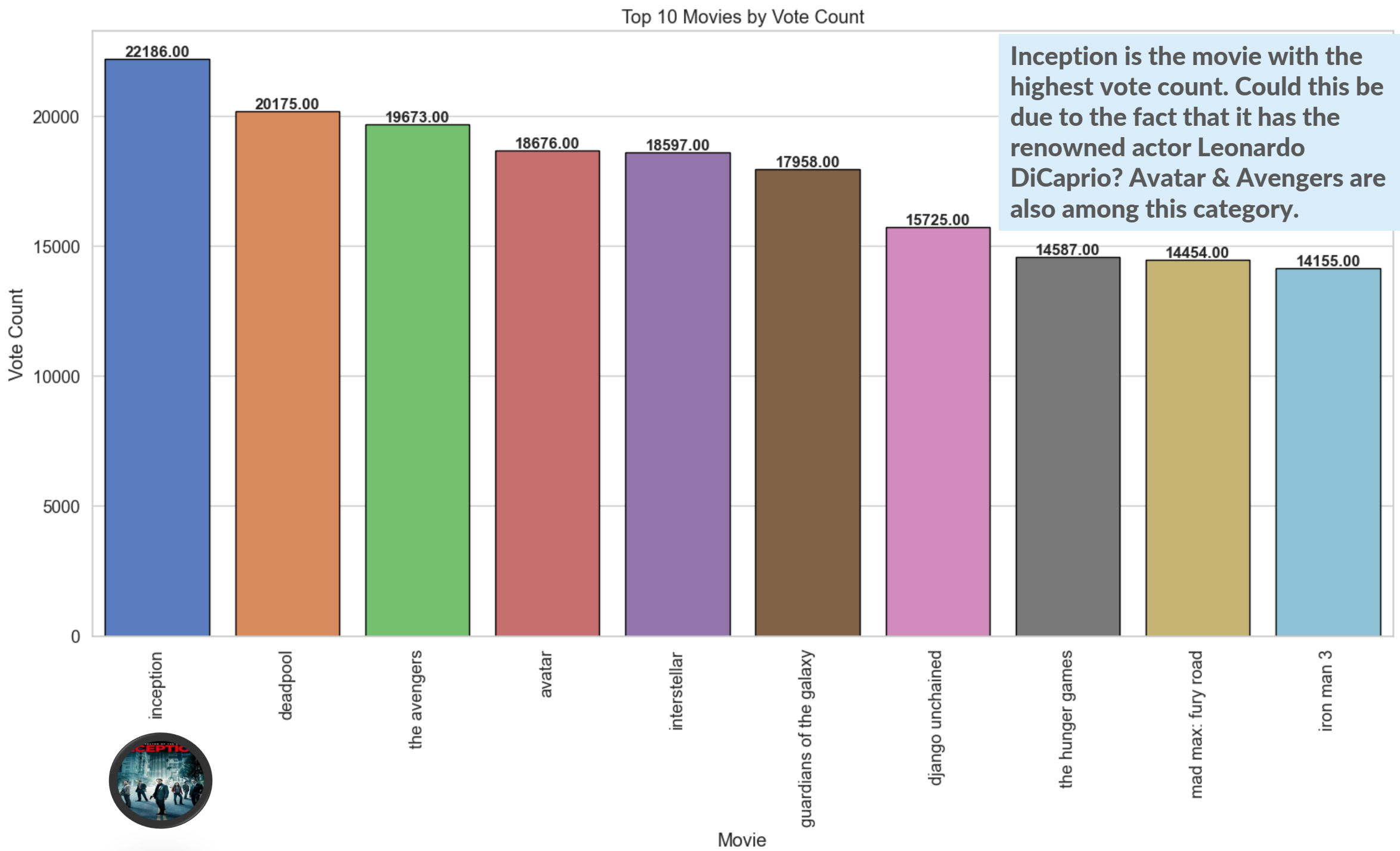
What are the top ten most popular movies?



`Avengers Infinity War` has the highest popularity score followed by `John Wick` then the `Hobbit: The Battle of The Five Armies`.
`John Wick` and `Hobbit: The Battle of The Five Armies` were not among the highest grossing films but still managed to get a very high popularity score. They also were not among the movie projects with the highest production budgets.
Also we notice that there is another `Avengers` entry in the top ten popularity score. This seems to be franchise that has done well throughout it's lifetime.

Question:

What are the top ten most popular movies by Vote Count?

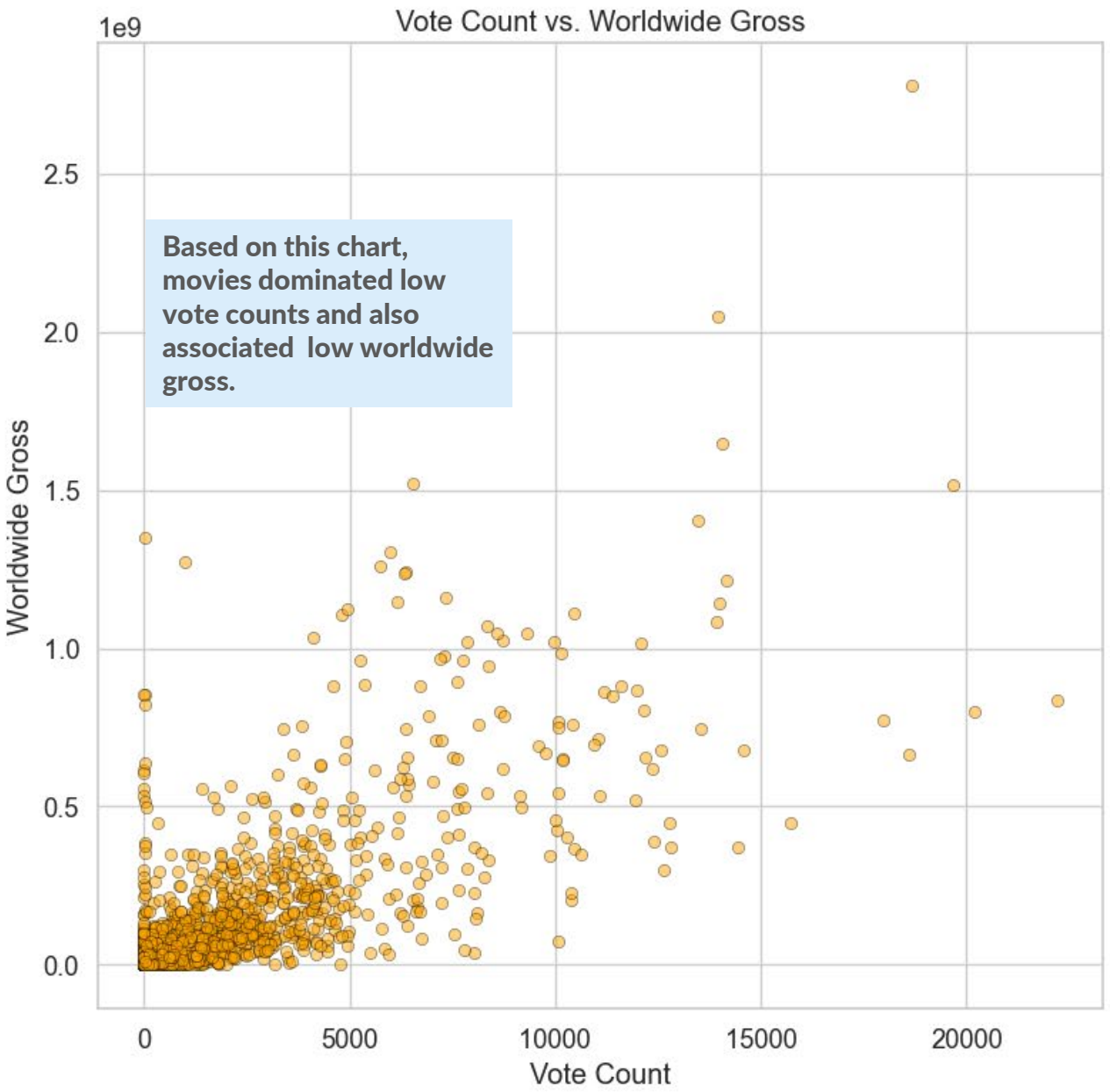


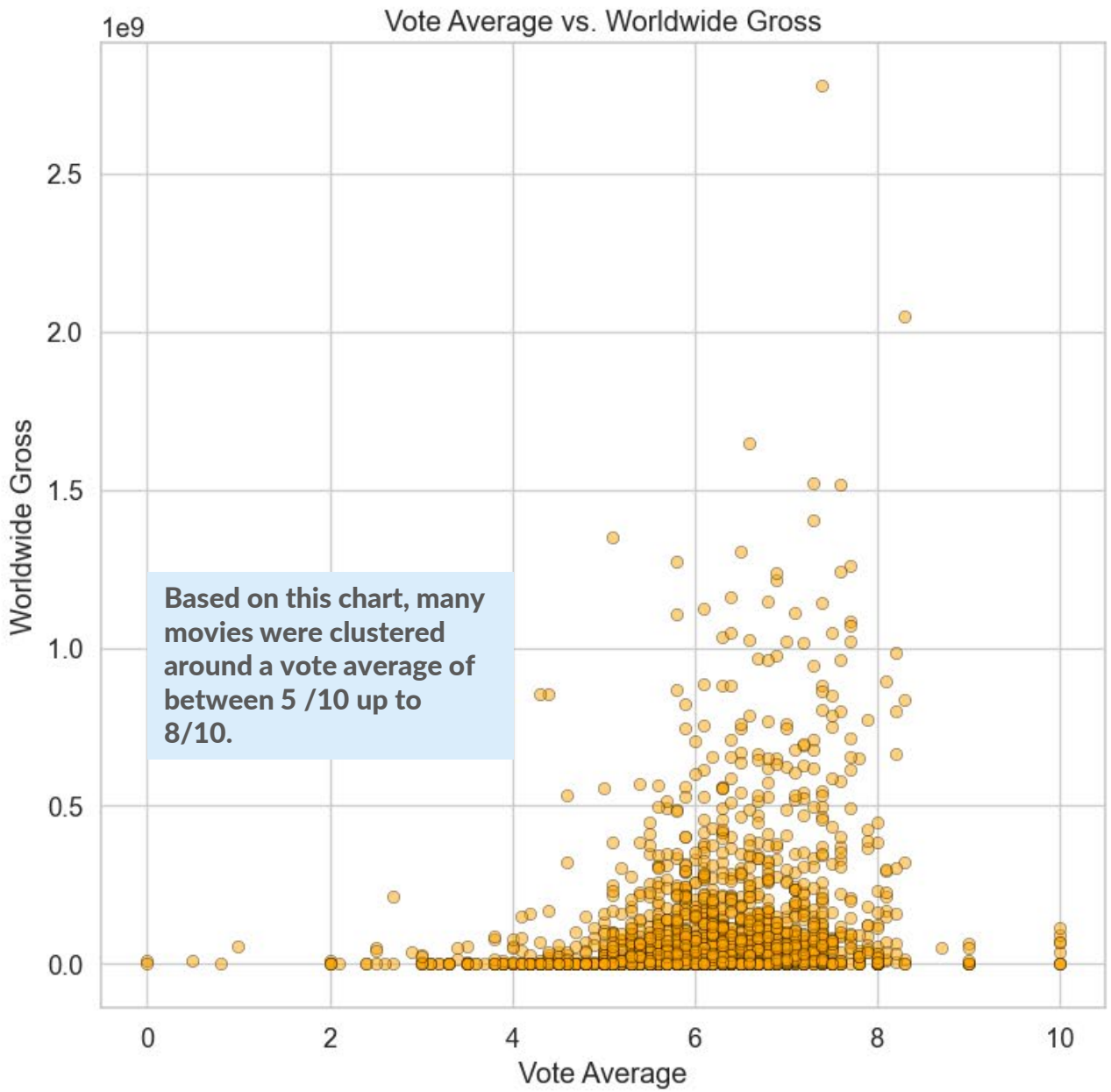
Inception is the movie with the highest vote count. Could this be due to the fact that it has the renowned actor Leonardo DiCaprio? Avatar & Avengers are also among this category.

Question:

Is there strong correlation with the Worldwide Gross?

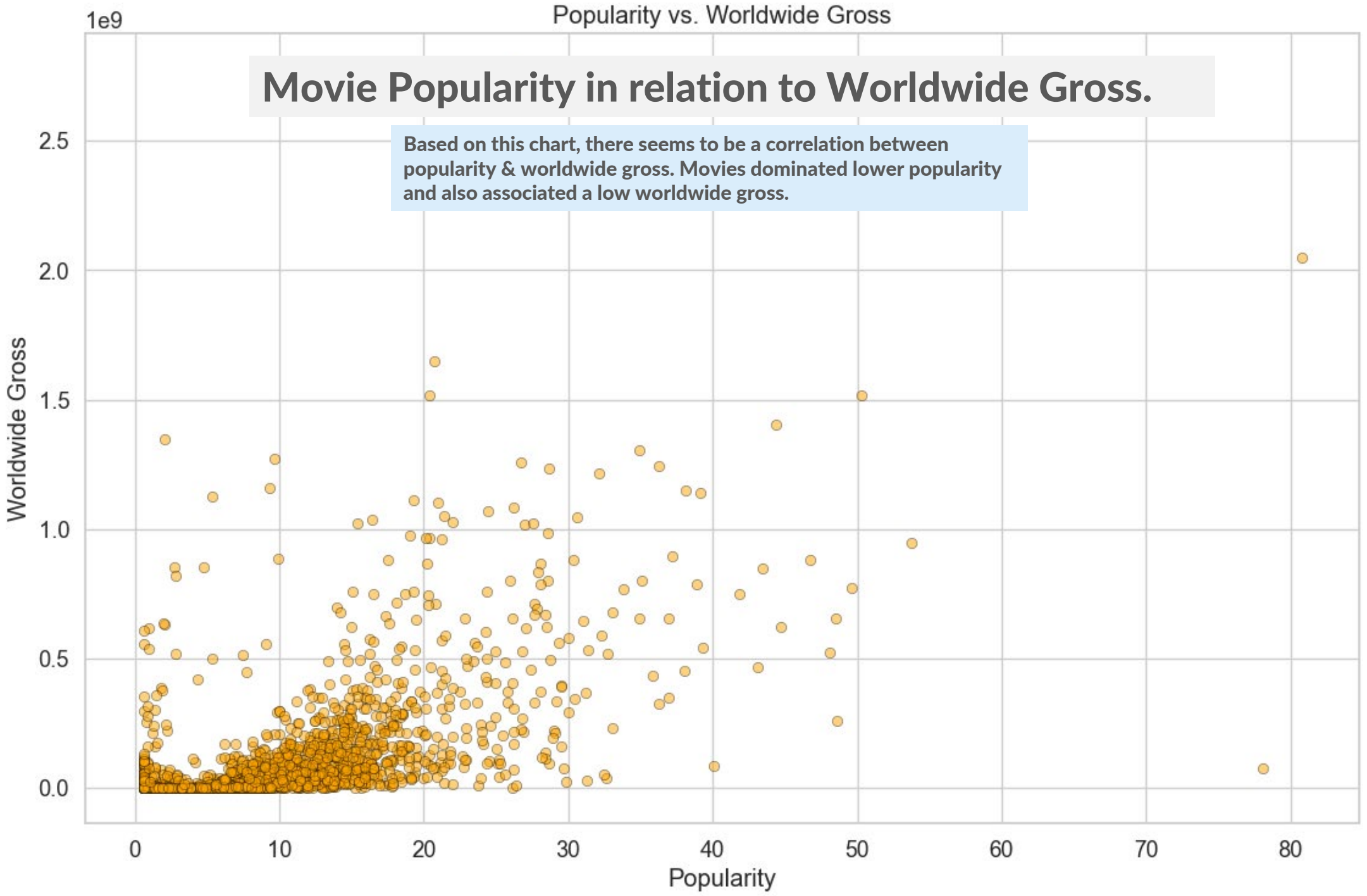
Relationship with the Worldwide gross.

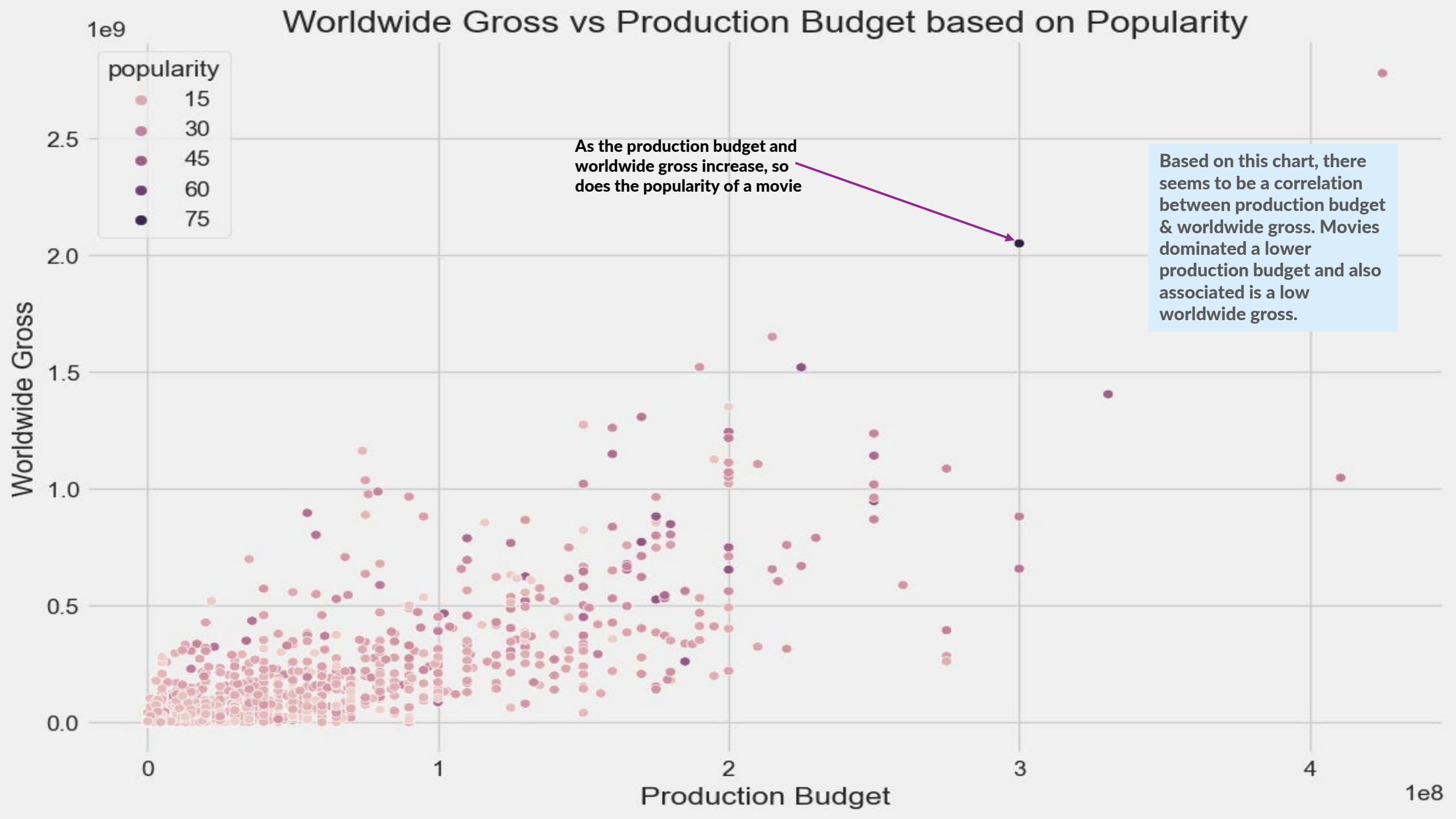




Movie Popularity in relation to Worldwide Gross.

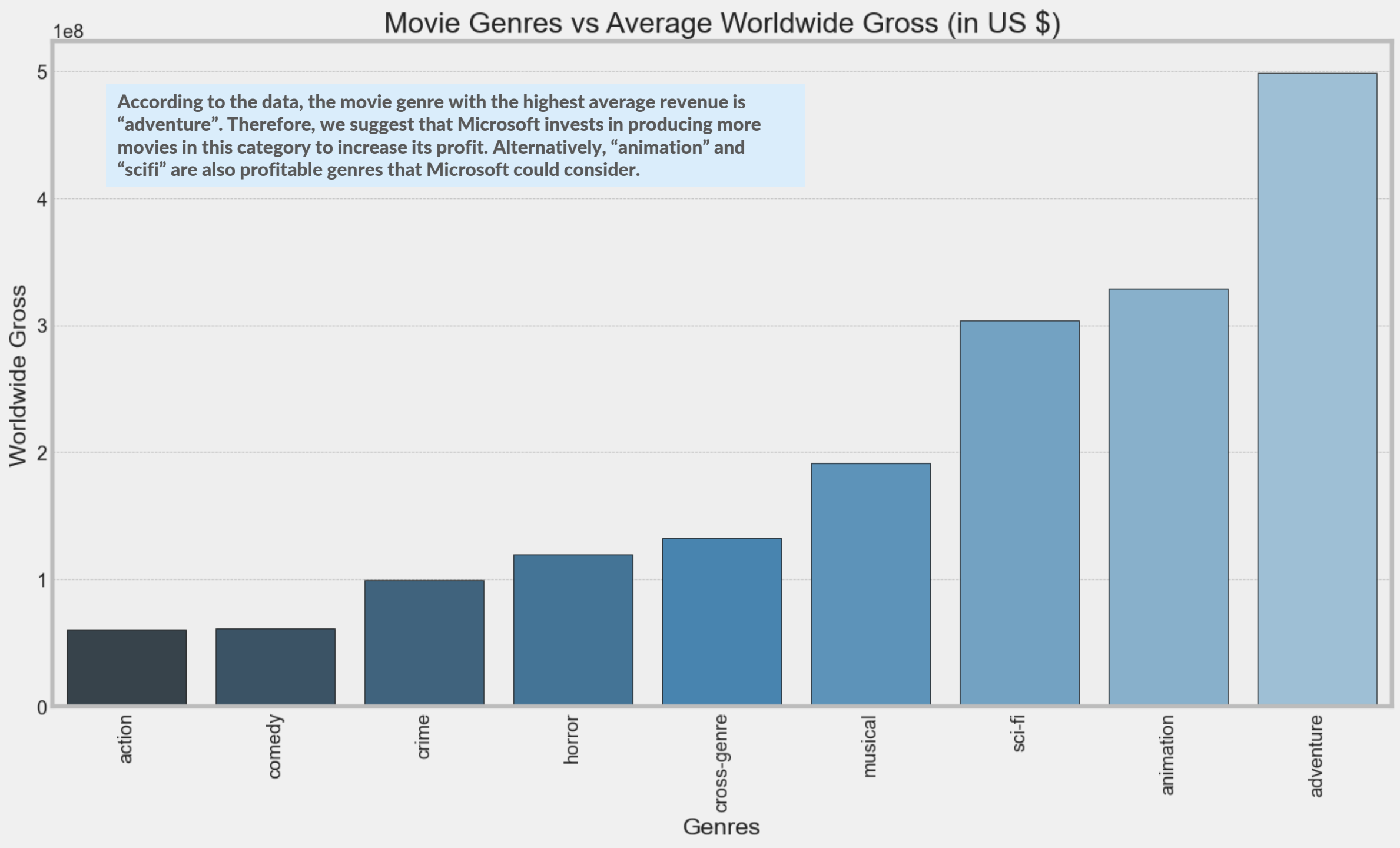
Based on this chart, there seems to be a correlation between popularity & worldwide gross. Movies dominated lower popularity and also associated a low worldwide gross.




Worldwide Gross vs Production Budget based on Popularity
1e9
popularity
15
30
45
60
75
As the production budget and worldwide gross increase, so does the popularity of a movie
Based on this chart, there seems to be a correlation between production budget & worldwide gross. Movies dominated a lower production budget and also associated is a low worldwide gross.
Worldwide Gross
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
Production Budget
1e8

Question:

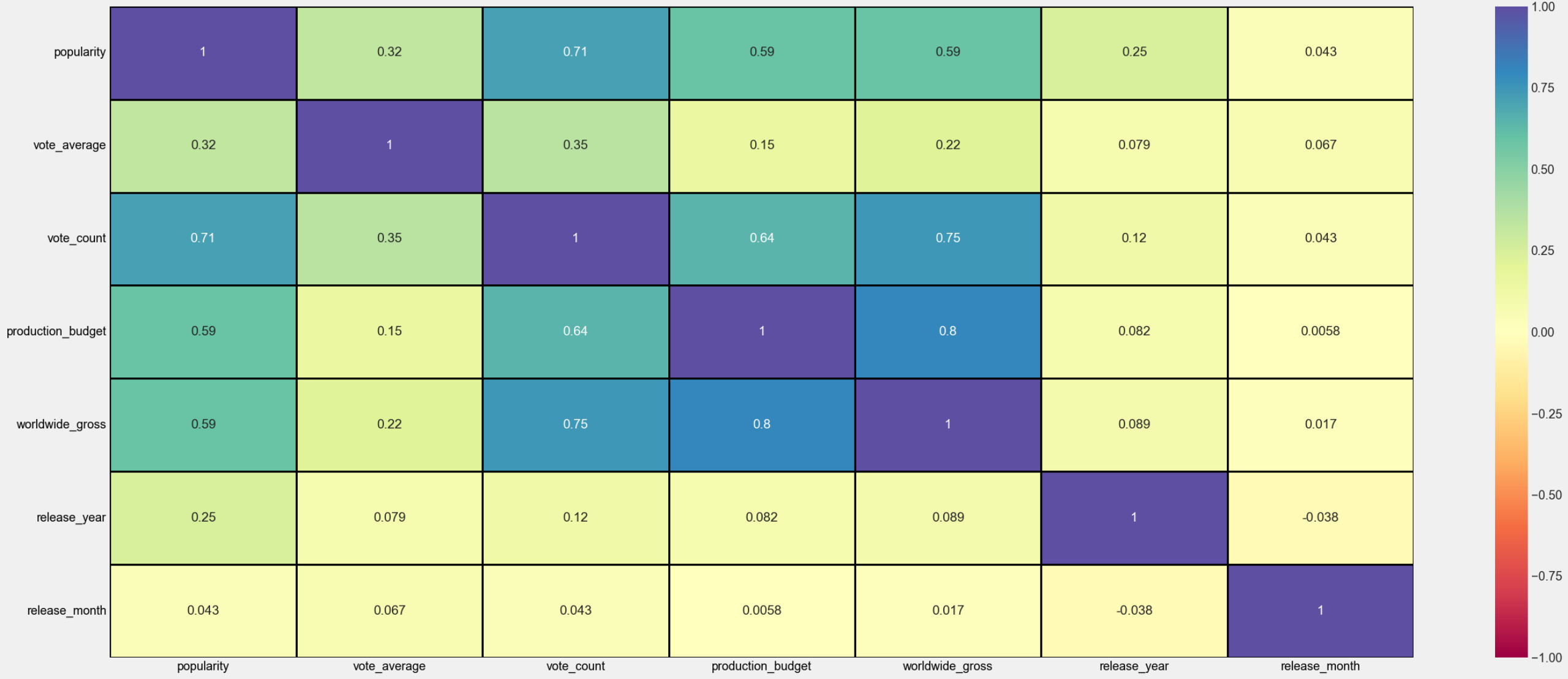
Is the relationship between the movie genre and how much it will gross worldwide that year?


Movie Genres vs Average Worldwide Gross (in US $)
1e8
5
4
3
2
1
0
Worldwide Gross
According to the data, the movie genre with the highest average revenue is "adventure". Therefore, we suggest that Microsoft invests in producing more movies in this category to increase its profit. Alternatively, "animation" and "scifi" are also profitable genres that Microsoft could consider.
action
comedy
crime
horror
cross-genre
musical
sci-fi
animation
adventure
Genres

Exploratory Data Analysis.

This process of examining and summarizing data sets using various techniques such as visualization and descriptive statistics to help to identify patterns and relationships in the data.

- Univariate Data Analysis.
- Bivariate Data Analysis.
- Multivariate Data Analysis.



We see a strong positive correlation between `vote_count` and the `popularity` of a movie.
We equally see a strong positive relationship between the `production_budget`,`wordwide gross` of a movie with the `popularity` of a movie.
A strong positive relationship is also witnessed between the `popularity` of a movie and the `vote_count` of a movie.
We equally see a strong positive relationship between the `production_budget`,`wordwide gross` of a movie with the `vote_count` of a movie.
There is also a very strong positive correlation between the `production_budget` and the `worldwide_gross` of a movie. It is actually the highest correlation in the correlation matrix with t correlation value of `0.8`

Summary of Findings

A summary of the key findings, organized by research questions or objectives, and supported by evidence from the data.

Please move to the next page

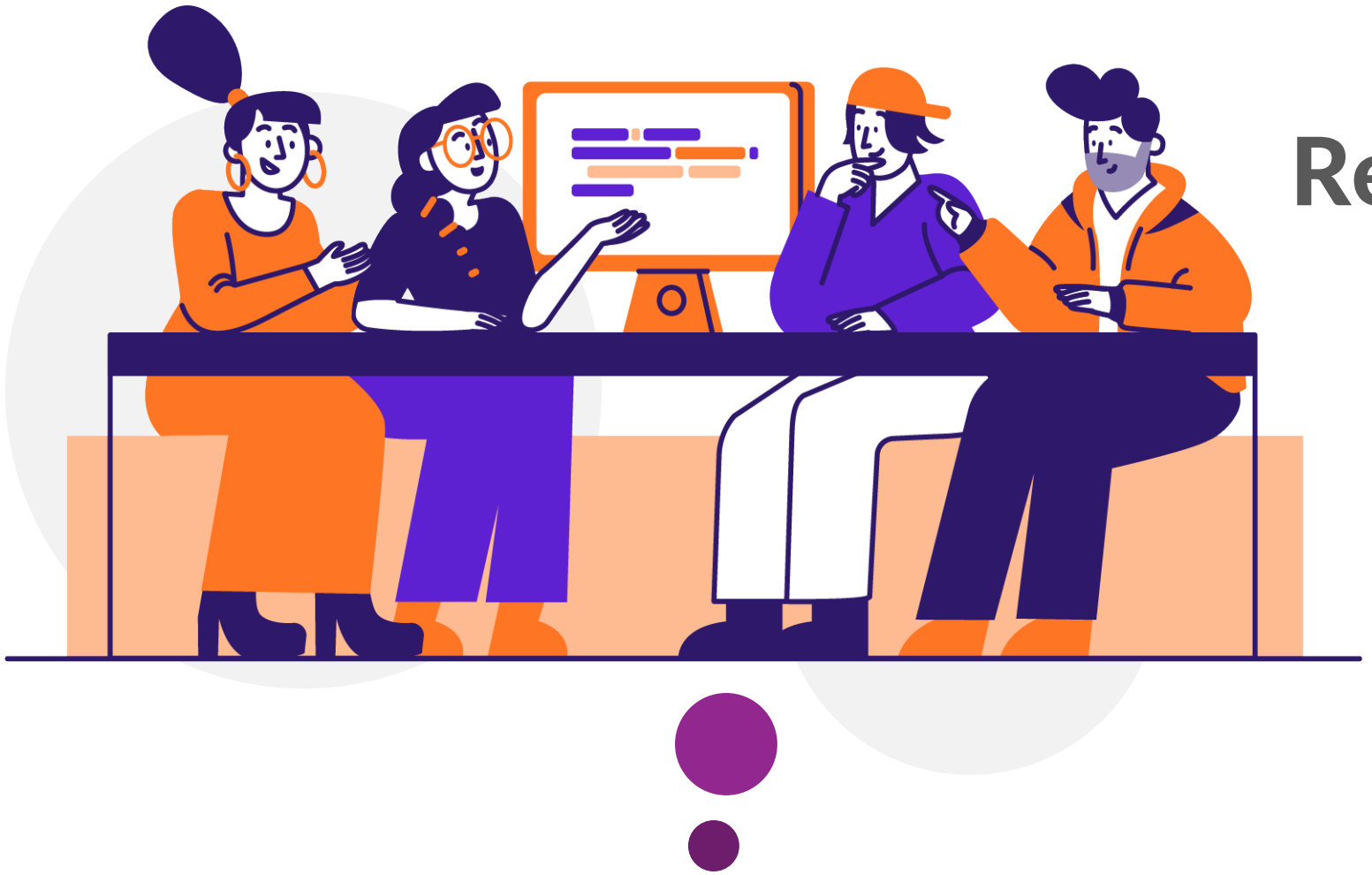
1. There seems to be a trend with the production budgets over the years. We can note that in the past, product budgets were relatively low but as the years went by, the budgets increased. This could be due to the fact that the movie industry has become more competitive and the need to produce high quality movies has increased. This could also be due to a number of factors, including the rising cost of talent, the increasing demand for special effects, and the decline of the market for low-budget films. However, there are still a number of low-budget films being made today, and some of these films have been successful at the box office.

2. We can see just before we get into the year 2020 there was another sharp decline in the production budgets. The COVID-19 pandemic, which caused widespread economic disruption and led to a decline in consumer demand could be a reason for this decline. This could also be due to the geopolitical tensions, such as the US-China trade war, which created uncertainty and led to businesses to delay or cancel investment projects

3. Production budgets for movies tend to increase between the months of may to June. This could be due to the fact that the summer season is the most popular time for movie releases. This is because people are more likely to go to the movies during the summer months, when the weather is nice and they have more free time. Another reason could be that many major film festivals take place in May and June, such as the Cannes Film Festival and the Tribeca Film Festival. These festivals can be a great way to generate buzz for a film and attract investors, which could lead to an increase in the production budget.

4. Movies with low popularity outnumber those with high popularity, and movies with high vote averages surpass those with low vote averages. Additionally, most movies have low vote counts compared to those with higher vote counts. Furthermore, there are more films with lower production budgets than higher ones, and likewise, more movies have lower worldwide gross compared to those with higher return on investment. Recent years have seen a higher number of movie releases, especially around the year 2013, possibly due to streaming service rise and growing interest in independent filmmaking. There is also a seasonal pattern, with more movies released in later months, possibly influenced by holidays, awards season, and consumer preferences.

5. Based on the data, "adventure" is the most lucrative movie genre. Microsoft should invest in producing more of these films for higher profits. "Animation" and "sci-fi" are also profitable options. In 1994, movies had the highest average worldwide gross, surpassing 1997's highest production budget. "Avatar" leads in worldwide gross margin, followed by "Avengers: Infinity War." Both "John Wick" and "The Hobbit: The Battle of The Five Armies" had high popularity scores despite not being top-grossing films or having the highest production budgets. Another "Avengers" entry ranks high in popularity, showing the franchise's ongoing success. There's a correlation between popularity and worldwide gross, as lower popularity correlates with lower gross revenue.



Recommendations

My actionable insights for the stakeholders at Microsoft Studio relevant to the business problem supported by evidence from the data and aligned with the project objectives and scope.

1. For the most favorable outcome, it is recommended to launch a movie during the month of May - June. Generally, releasing a movie during the summer months is considered advantageous. However, if circumstances dictate that the movie must be released in the fourth quarter (Q4) to capitalize on a last-minute opportunity for profit, November stands out as the most promising option for the release date.

2. Release in non-summer months - While summer sees big box office, other months have done well too like March for The Hunger Games. Spreading releases across the year is beneficial.

3. It is suggested to produce additional movies within the "adventure" genre since it proves to be the most lucrative. Alternatively, "animation" and “Sci-Fi" present themselves as the next favorable choices.

4. Draw major stars - Having big name actors like Leonardo DiCaprio or Jennifer Lawrence attached boosts marketability, especially overseas.

5. As the budget of a movie rises, there is typically a corresponding increase in profit. Nevertheless, in order to provide a more assured recommendation concerning the relationship between budget and return on investment, it is essential to gather additional data on movies with budgets exceeding $200 million.

6. Keep budgets modest - Many of the top grossing films have reasonable budgets under $100 million. Filming for less can increase profits.

7. Develop a streaming strategy - Building exclusive streaming content can help attract talent and provide an alternate distribution channel.

Questions??

Made with ♡ by Richard Taracha

taracharichard@gmail.com