Box Office Analysis: Identifying Best Performing Movie Genres for Microsoft's New Studio

A Data-Driven Approach to Informing Content Creation Decisions

Overview

This project analyzes the datasets from 3 movie websites namely; **Box Office Mojo, TheMovieDB** and **The Numbers**. The 3 datasets are merged to enable a more in-depth analysis and better findings with regards to the top/best performing movie genres in the Box Office. Microsoft can definitely use the findings from this analysis to help decide what type of films to create so as to stay at par in the movie industry.

Business Understanding

Based on the business problem, which is, Microsoft wants to get in on the fun of creating movies/original video content but they have no knowledge of creating movies, I have formulated 4 business questions whereby I will use my dataset to extract meaningful findings which can be translated into actionable insights for the Head of Microsoft's new movie studio to help him/her decide on what types of films to create. These business questions are:

1. What are the top 3 best performing genres of movies at the box office?
2. What is the relationship between production budget and the success of a movie both domestically and worldwide?
3. What is the competitive landscape of the movie industry in terms of market share?
4. How does the release time of a movie contribute to its success?

Data Understanding

I therefore have 3 separate data files:

- ➔ bom.movie_gross.csv.gz: each record represents a movie title, with attributes of that movie (eg. domestic gross).
- ➔ tmdb.movies.csv.gz: each record represents a movie title as well, with attributes such as release date.
- ➔ tn.movie_budgets.csv.gz: each record represents a movie title as well, with attributes such as production budget.

Also, note that the data may not reflect the most-up-to-date trends and performances in the movie industry since its scope is upto 2020.

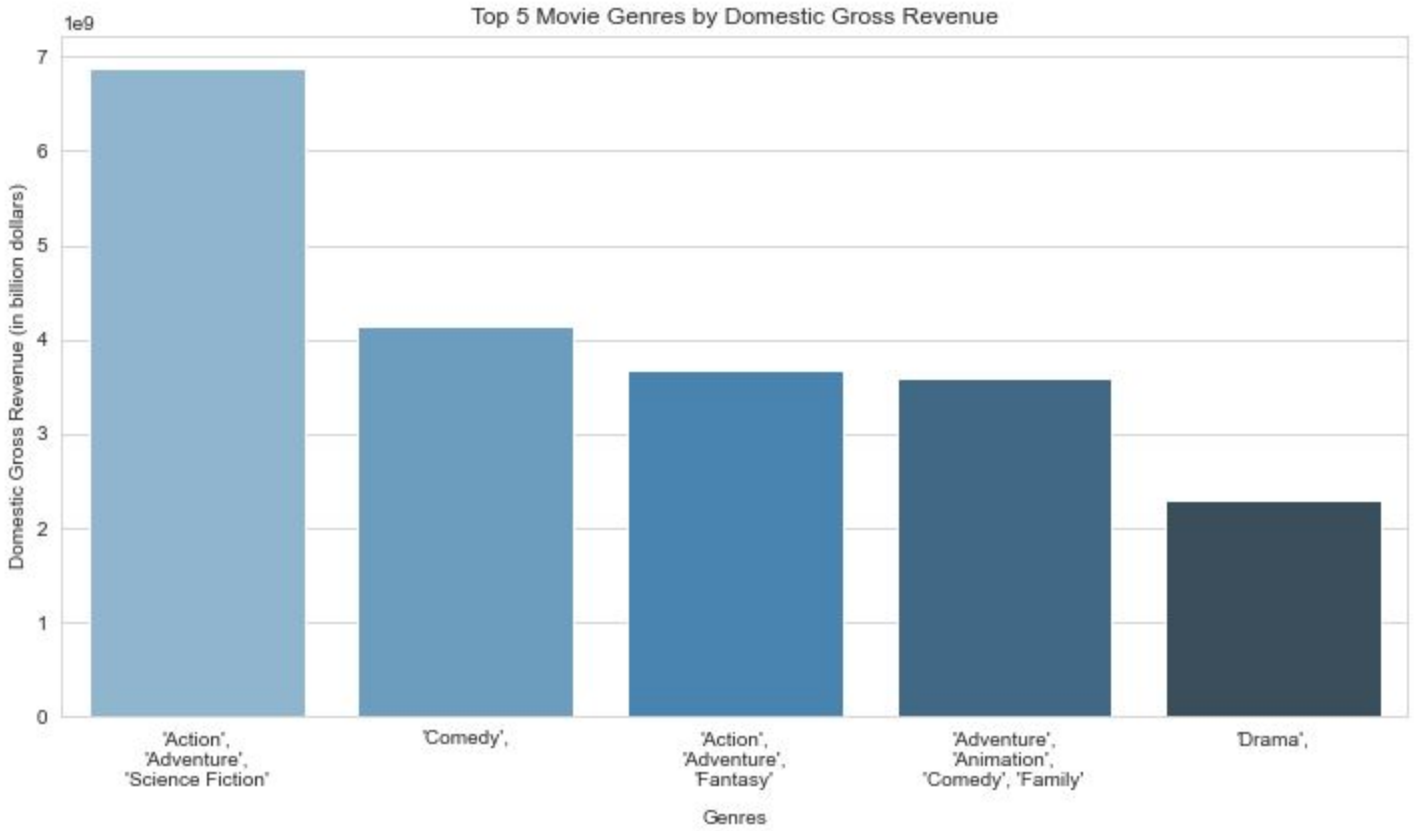
Therefore, to carry out my data analysis and manipulation, the tools I used included Python programming language in conjunction with Python libraries such as Pandas and Seaborn in a Jupyter Notebook.

Below, I highlight the steps and processes I undertook for the analysis.

Performing Aggregations to answer Business Question 1

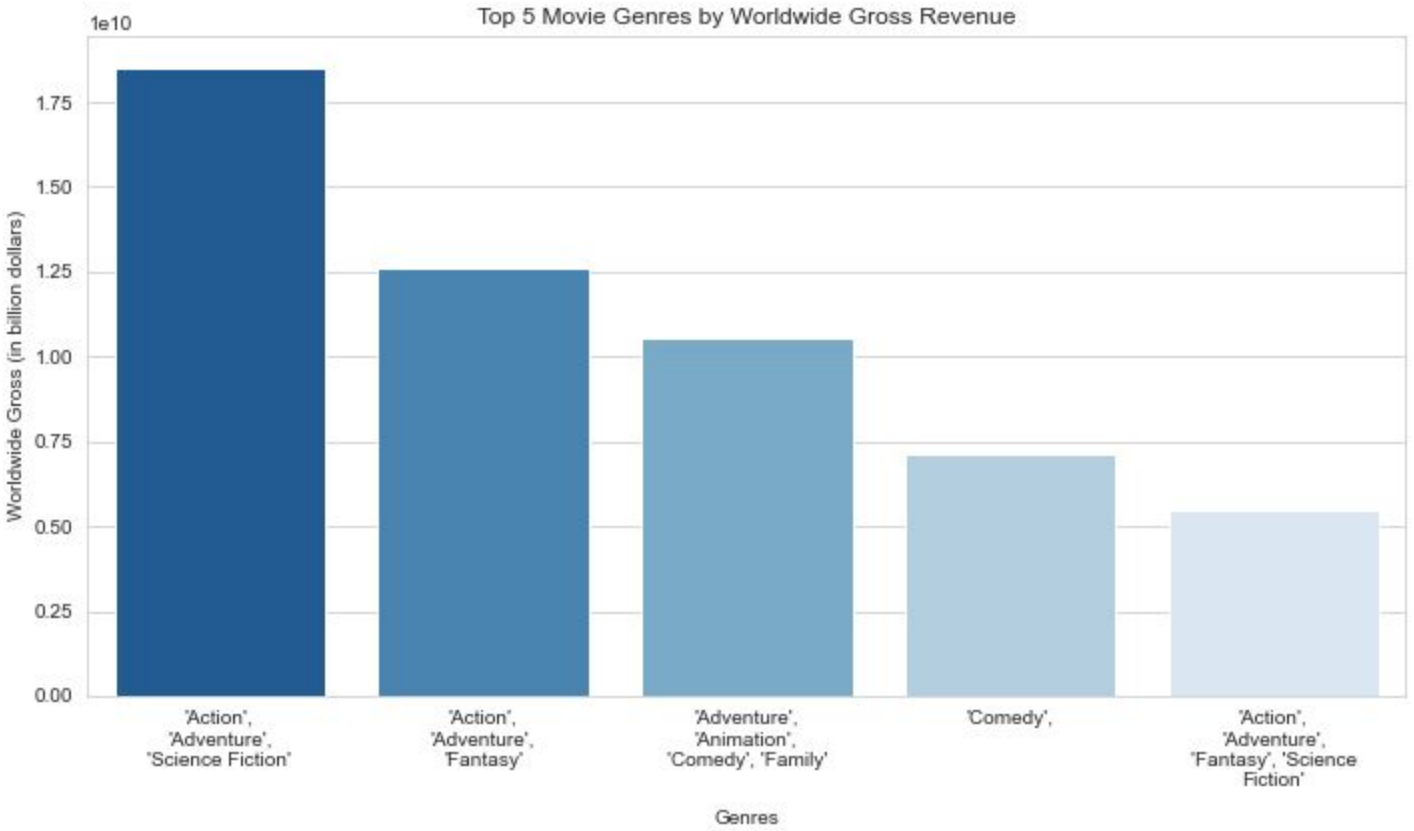
To come up with a final top 3, I considered the top genres across multiple metrics and determined which genres are consistently ranking high across the board.

I created plots that show the rankings for each genre by domestic gross, worldwide gross, popularity, vote count, and vote average, and then compared the results. My approach was mainly using descriptive statistics such as mean and sum of the variables.


Top 5 Movie Genres by Domestic Gross Revenue
1e9
Domestic Gross Revenue (in billion dollars)
0
1
2
3
4
5
6
7
'Action', 'Adventure', 'Science Fiction'
'Comedy',
'Action', 'Adventure', 'Fantasy'
'Adventure', 'Animation', 'Comedy', 'Family'
'Drama',
Genres

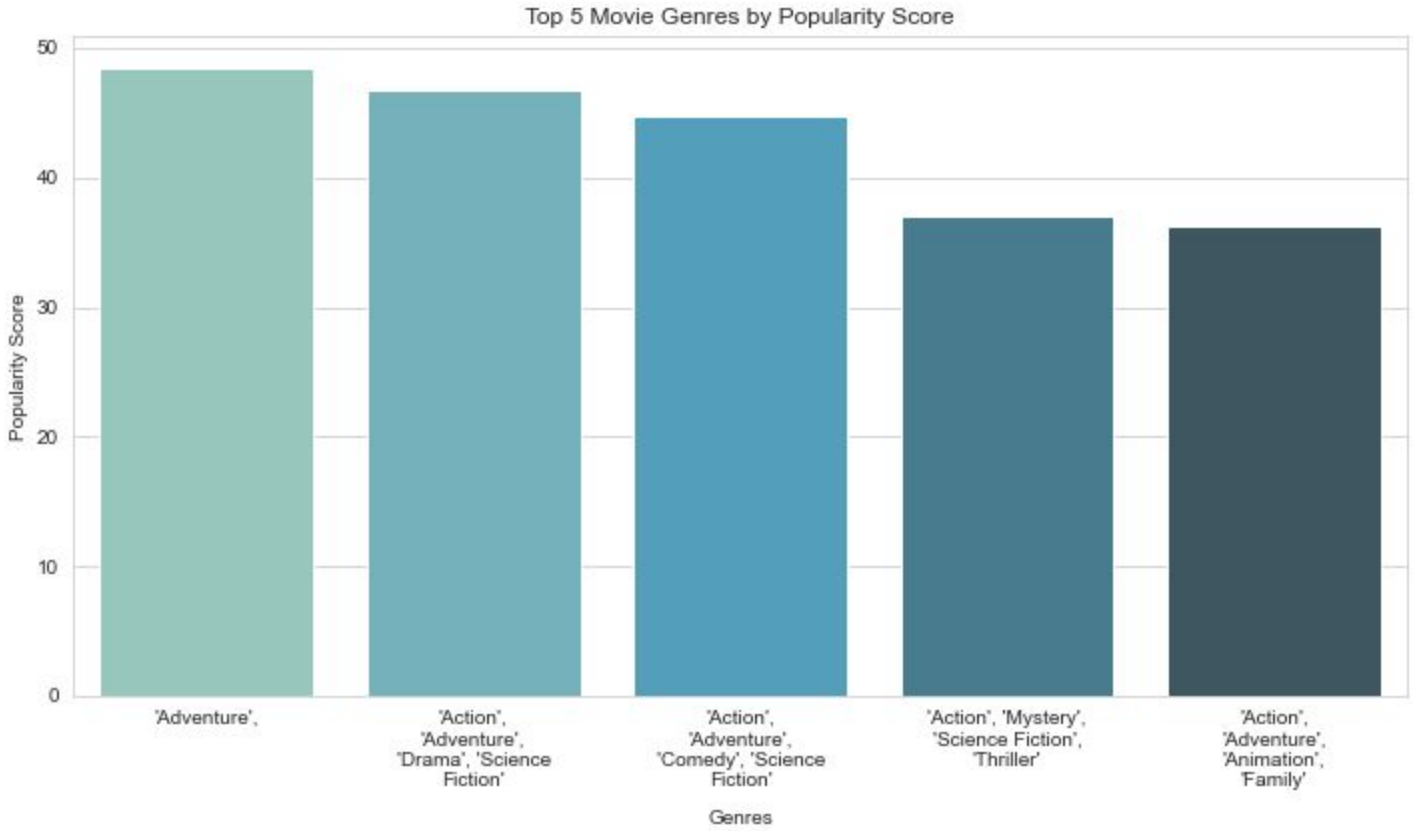
From the above results, the top 5 genres sorted by total domestic gross are:

1. Action, Adventure, Science Fiction
2. Comedy
3. Action, Adventure, Fantasy
4. Adventure, Animation, Comedy, Family
5. Drama


Top 5 Movie Genres by Worldwide Gross Revenue
1e10
Worldwide Gross (in billion dollars)
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
'Action', 'Adventure', 'Science Fiction'
'Action', 'Adventure', 'Fantasy'
'Adventure', 'Animation', 'Comedy', 'Family'
'Comedy',
'Action', 'Adventure', 'Fantasy', 'Science Fiction'
Genres

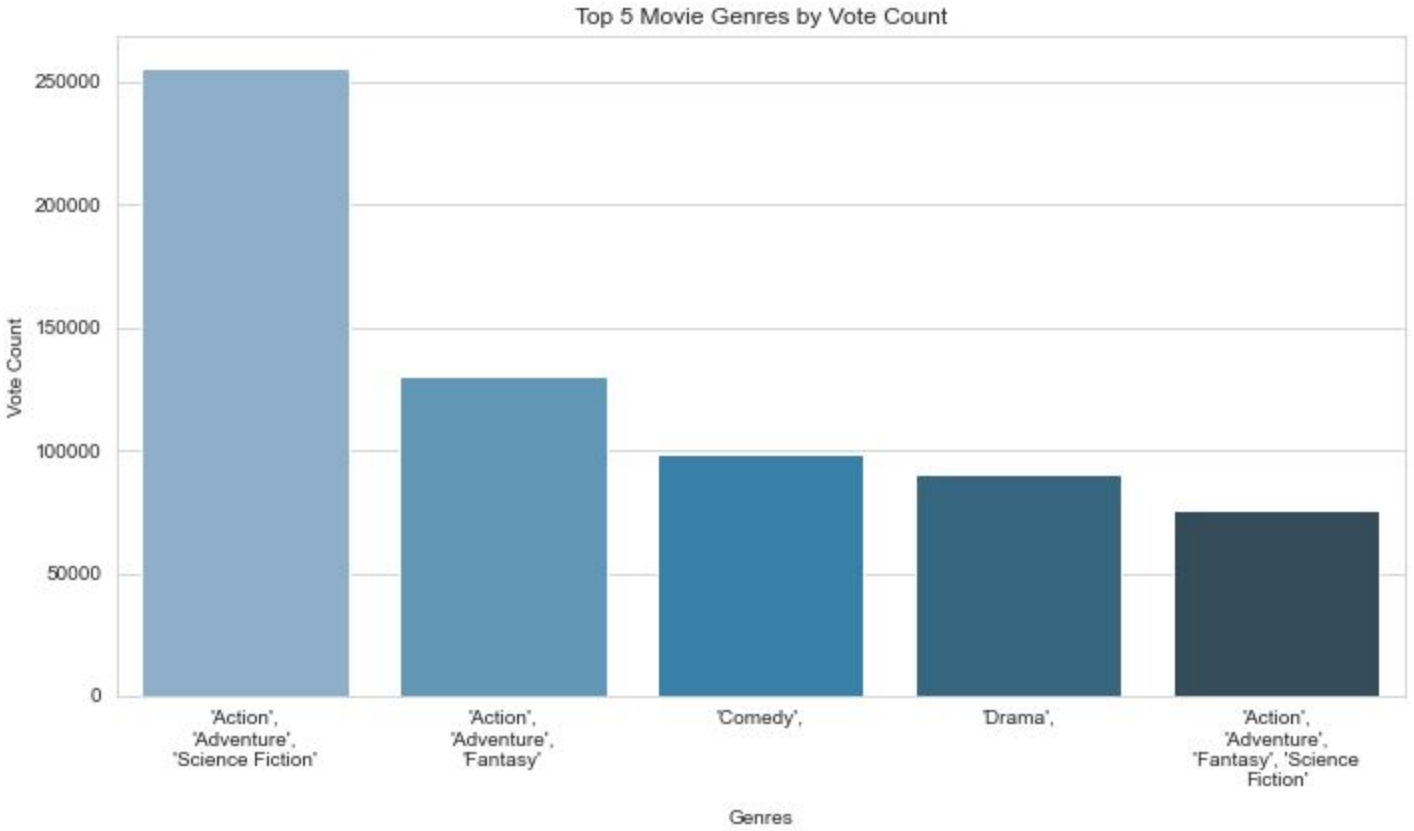
From the above results, the top 5 genres sorted by worldwide gross are:

1. Action, Adventure, Science Fiction
2. Action, Adventure, Fantasy
3. Adventure, Animation, Comedy, Family
4. Comedy
5. Action, Adventure, Fantasy, Science Fiction


Top 5 Movie Genres by Popularity Score
Popularity Score
50
40
30
20
10
0
'Adventure',
'Action', 'Adventure', 'Drama', 'Science Fiction'
'Action', 'Adventure', 'Comedy', 'Science Fiction'
'Action', 'Mystery', 'Science Fiction', 'Thriller'
'Action', 'Adventure', 'Animation', 'Family'
Genres

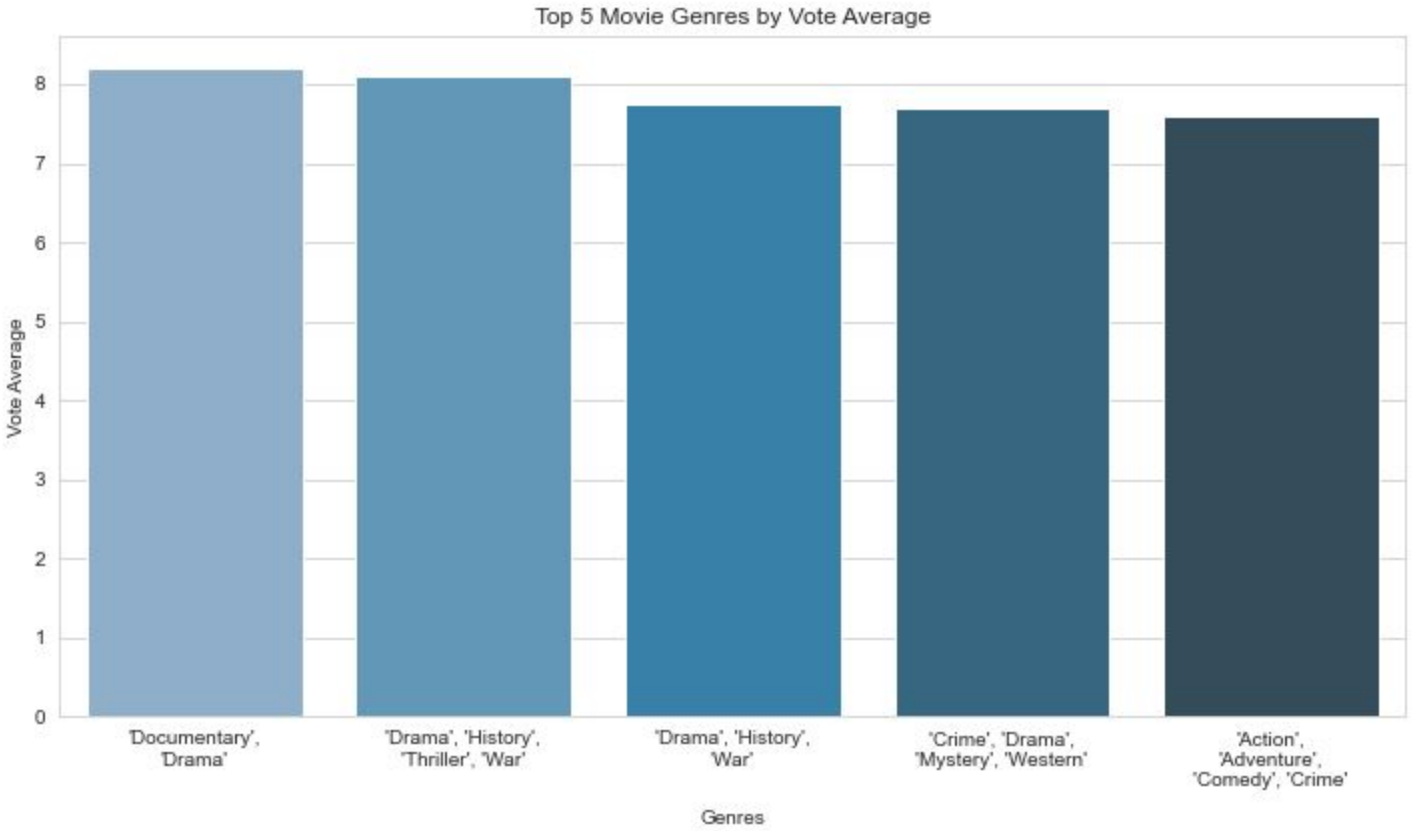
From the above results, the top 5 genres sorted by popularity are:

1. Adventure
2. Action, Adventure, Drama, Science Fiction
3. Action, Adventure, Comedy, Science Fiction
4. Action, Mystery, Science Fiction, Thriller
5. Action, Adventure, Animation, Family


Top 5 Movie Genres by Vote Count
Vote Count
250000
200000
150000
100000
50000
0
'Action', 'Adventure', 'Science Fiction'
'Action', 'Adventure', 'Fantasy'
'Comedy',
'Drama',
'Action', 'Adventure', 'Fantasy', 'Science Fiction'
Genres

From the above results, the top 5 genres sorted by vote count are:

1. Action, Adventure, Science Fiction
2. Action, Adventure, Fantasy
3. Comedy
4. Drama
5. Action, Adventure, Fantasy, Science Fiction


Top 5 Movie Genres by Vote Average
Vote Average
0
1
2
3
4
5
6
7
8
'Documentary', 'Drama'
'Drama', 'History', 'Thriller', 'War'
'Drama', 'History', 'War'
'Crime', 'Drama', 'Mystery', 'Western'
'Action', 'Adventure', 'Comedy', 'Crime'
Genres

From the above results, the top 5 genres sorted by vote average are:

1. Documentary, Drama
2. Drama, History, Thriller, War
3. Drama, History, War
4. Crime, Drama, Mystery, Western
5. Action, Adventure, Comedy, Crime

Therefore, from these 5 sets of analyses I concluded that the top 3 best performing types of movies are an assortment of:

1. **Action, Adventure, Science Fiction**
2. **Action, Adventure, Fantasy**
3. **Comedy**

Creating a correlation matrix to answer Business Question 2

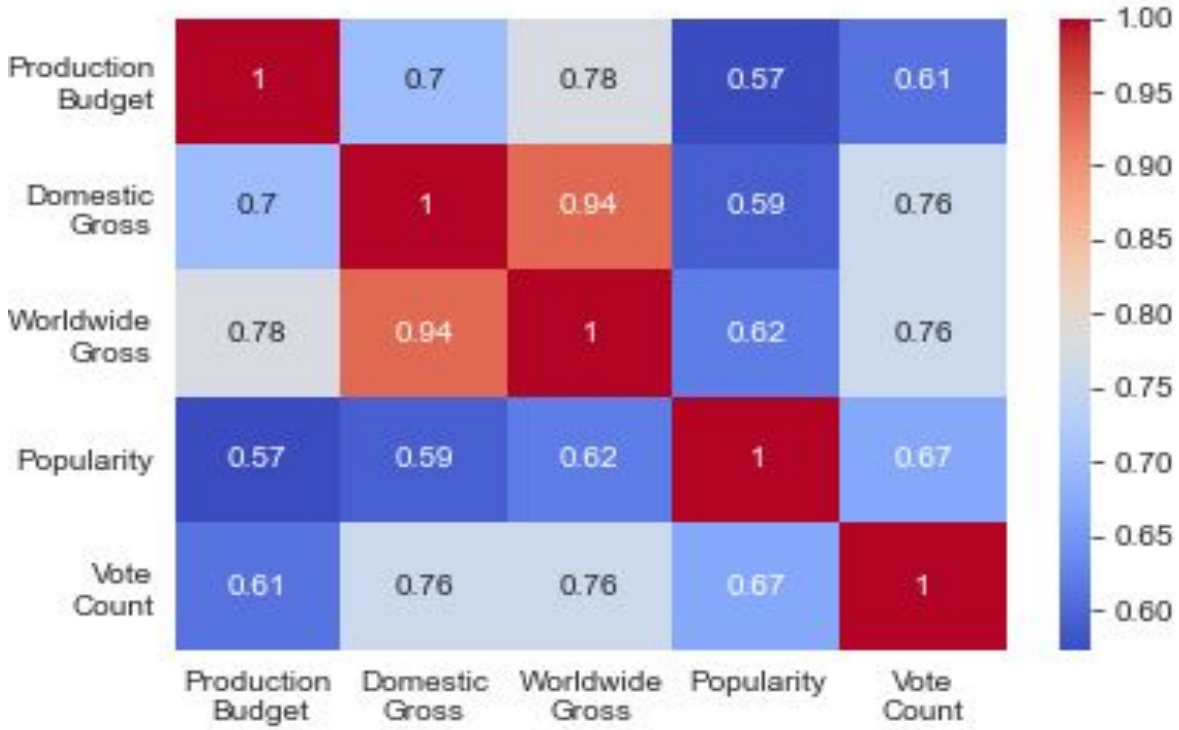
A correlation matrix is a table that shows how strongly different variables are related to each other. It is a way to measure the degree to which two or more variables move together or apart.

It contains a set of numbers that represent the strength of the relationship between each pair of variables. These numbers can range from -1 to 1, where -1 indicates a strong negative relationship (when one variable goes up, the other goes down) and 1 indicates a strong positive relationship (when one variable goes up, the other goes up as well). A value of 0 indicates no relationship between the variables.

In this section I created a correlation matrix for the features which are indicators of a successful movie in the Box Office.

These features included:

- Production budget
- Domestic gross
- Worldwide gross
- Popularity
- Vote count

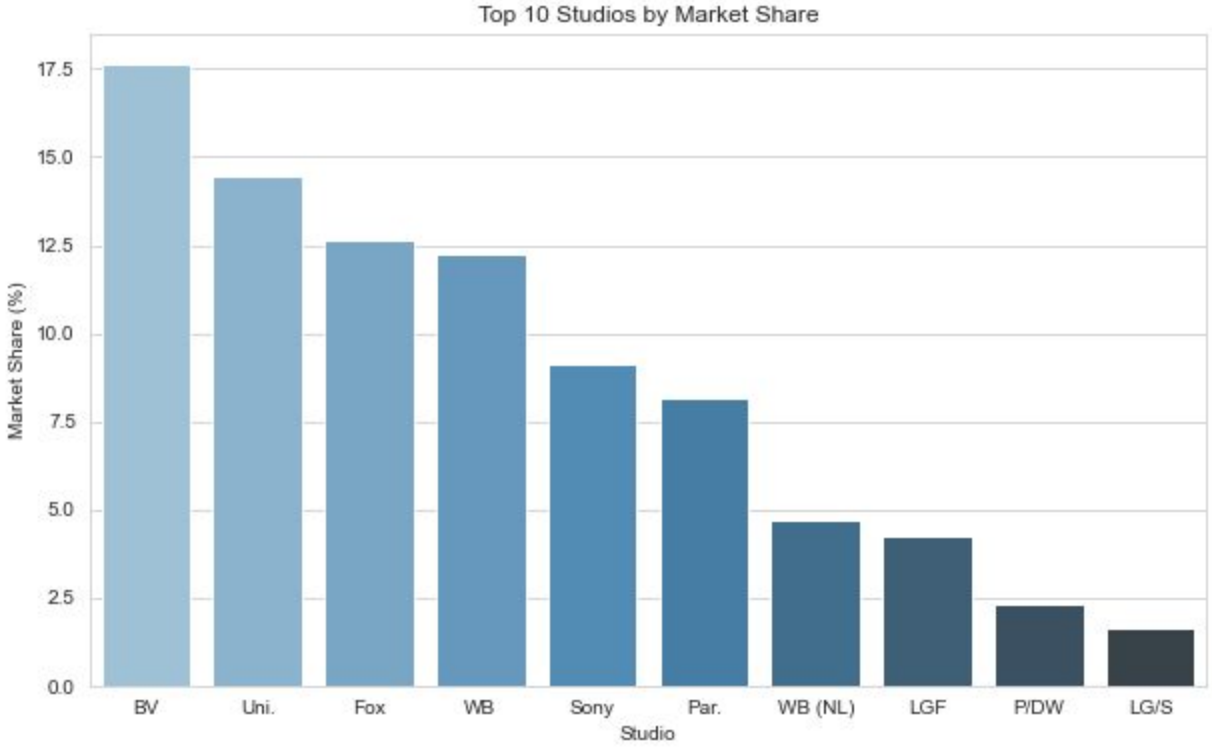

Production Budget
Domestic Gross
Worldwide Gross
Popularity
Vote Count
1
0.7
0.78
0.57
0.61
0.7
1
0.94
0.59
0.76
0.78
0.94
1
0.62
0.76
0.57
0.59
0.62
1
0.67
0.61
0.76
0.76
0.67
1
Production Budget
Domestic Gross
Worldwide Gross
Popularity
Vote Count
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60

The correlation coefficient between production budget and worldwide gross revenue is **0.78**. This figure indicates a strong positive correlation between the two features. This means that as production budget increases, worldwide gross revenue tends to increase as well.

The correlation coefficient between production budget and domestic gross is **0.7**. This figure also indicates a strong positive correlation between the two features. This means that as production budget increases, domestic gross revenue tends to increase as well.

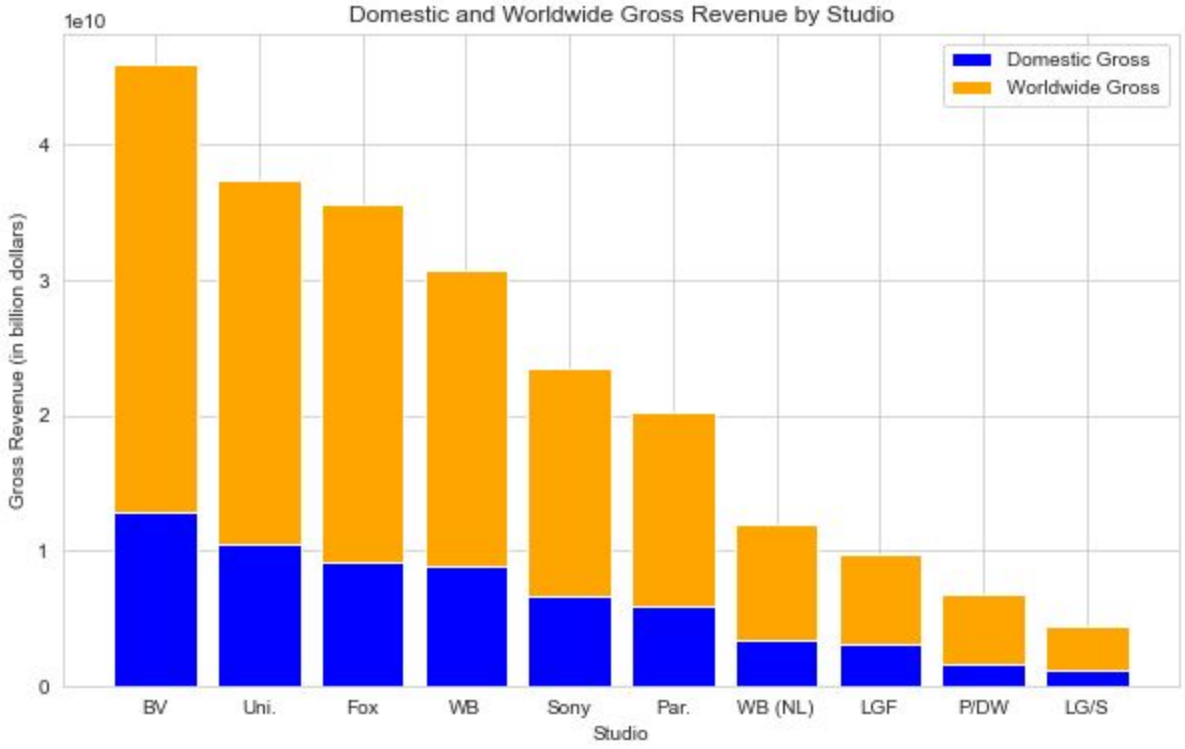
Performing Aggregations and Engineering a new feature `market share` to answer Business Question 3

Market share is basically what proportion of the market each production studio holds in the movie industry. I calculated the market share by getting the sum of a studio's domestic revenue divided by the total domestic revenue of all the studios multiplied by 100%.


Top 10 Studios by Market Share
Market Share (%)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
BV
Uni.
Fox
WB
Sony
Par.
WB (NL)
LGF
P/DW
LG/S
Studio

Based on the above analysis, BV Studio has the highest domestic and worldwide gross revenue, the highest total vote count, and the highest mean popularity. BV Studio has a market share of 17.6%, followed by Uni. with 14.4% and Fox with 12.7%. This suggests that BV Studio is the clear leader in many metrics. This provides insight into the competitive landscape of the movie industry based on the chosen metrics.

I find this to be an important aspect because as Microsoft dives into the movie industry, it's imperative they understand who they are going to be competing against in terms of capturing the audience's attention.

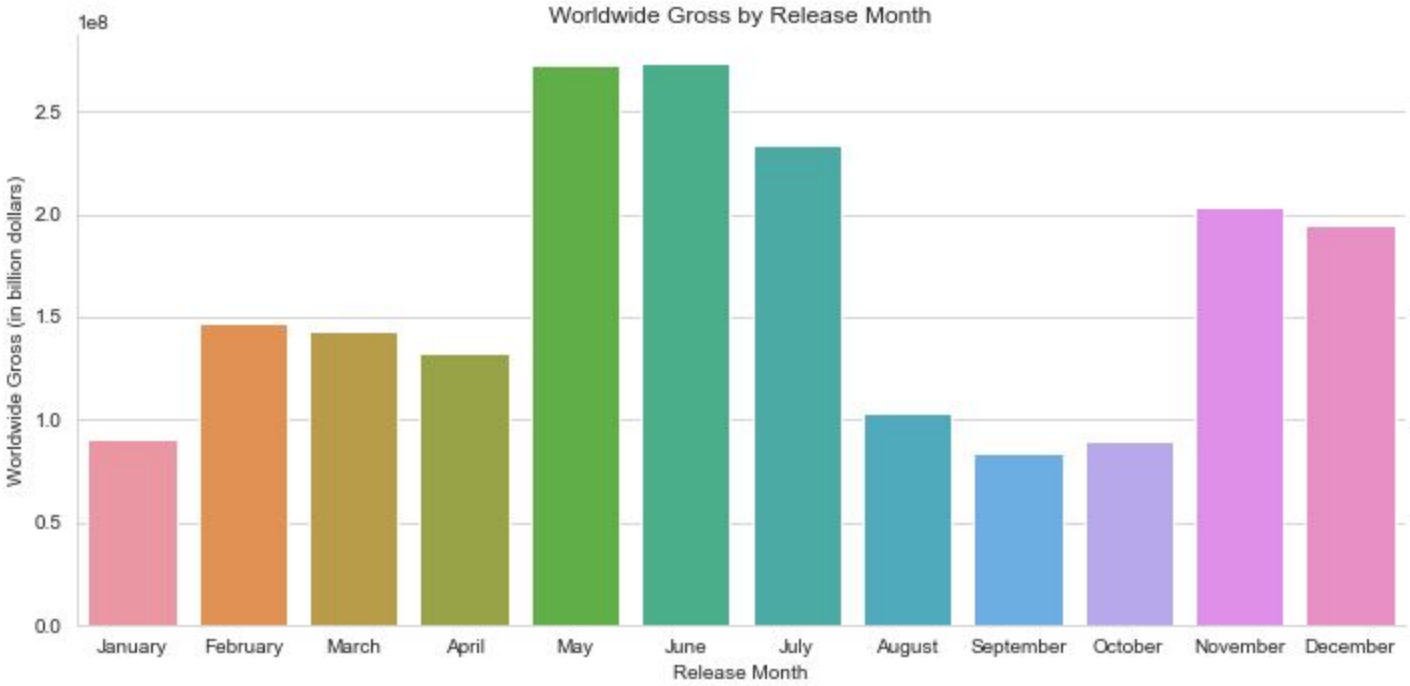

Domestic and Worldwide Gross Revenue by Studio
1e10
Domestic Gross
Worldwide Gross
Gross Revenue (in billion dollars)
0
1
2
3
4
BV
Uni.
Fox
WB
Sony
Par.
WB (NL)
LGF
P/DW
LG/S
Studio

The stacked bar plot shows the domestic and worldwide gross revenue of the top 10 movie studios, broken down by region, that is, Domestic revenue and Worldwide revenue.

Overall, this plot helps to visualize the revenue breakdown of the top movie studios, and can provide insights into the relative importance of different regions for these studios. This is also an indicator to Microsoft's new movie studio, in that they can anticipate more of their revenue generated from the international market, as compared to the domestic market.

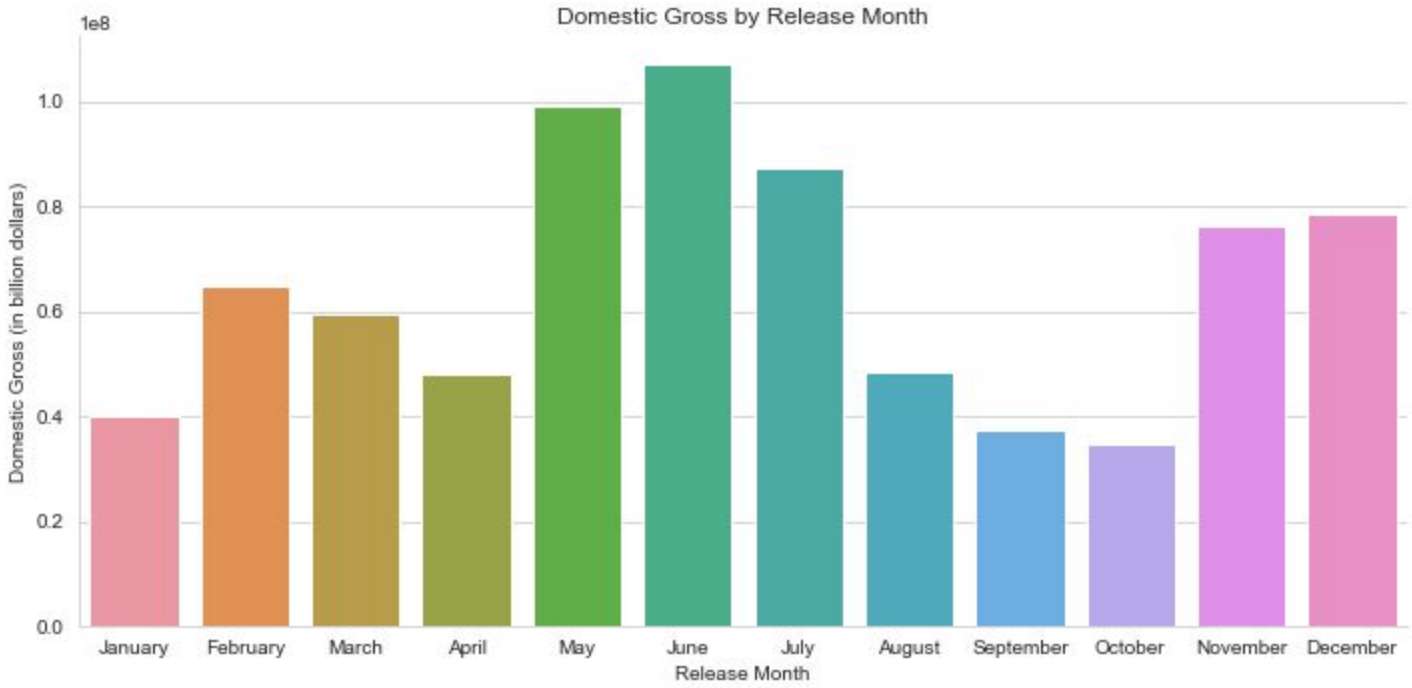
Performing Aggregations to answer Business Question 4

Here I used the worldwide gross and domestic gross variables as my units of measurement to depict the revenues generated in the different months and hence determine when it is suitable to release a movie, thereby increasing its chances of success.


Worldwide Gross by Release Month
1e8
Worldwide Gross (in billion dollars)
2.5
2.0
1.5
1.0
0.5
0.0
January
February
March
April
May
June
July
August
September
October
November
December
Release Month

The x-axis shows the months of the year, and the y-axis shows the average gross in billion dollars. Each bar represents the average gross for a particular month.

The plot shows that the months of May, June and July have the highest average worldwide gross, while the months of September and October have the lowest.


Domestic Gross by Release Month
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Domestic Gross (in billion dollars)
January
February
March
April
May
June
July
August
September
October
November
December
Release Month

The x-axis shows the months of the year, and the y-axis shows the average gross in billion dollars. Each bar represents the average gross for a particular month.

The plot shows a similar trend, with May, June and July having the highest average domestic gross, and September and October having the lowest.

Overall, these plots suggest that releasing a movie in May, June or July may lead to higher gross revenue, both domestically and worldwide, while releasing a movie in September or October may result in lower gross revenue.

Conclusion

This analysis leads to four **recommendations** that will enable Microsoft get into the movie industry with a resounding success for the movies that will be produced/created.

1.

Based on the findings of the top 3 best performing types of movies in the Box Office, Microsoft should consider producing movies around the genre combinations of:

- **Action, Adventure, Science Fiction**
- **Action, Adventure, Fantasy**
- **Comedy**

Also, they can play around the genres creatively and come up with something a bit unique, for example, a combination of **Action, Adventure, Comedy** or even **Action, Science Fiction, Fantasy** to see the response and reaction from the movie lovers.

2.

Based on the findings of strong positive correlation between production budget and domestic gross, and between production budget and worldwide gross, the Head of Microsoft's new movie studio should liaise with the finance department and ensure that sufficient budgetary allocation is made to film production.

This is because based on the correlation coefficients, I concluded that the production budget allocated to a movie is a good indicator of its success or lack thereof, both domestically and internationally.

This would enable the several aspects involved in film production to be taken care of sufficiently, for example:

- Production equipment
- Marketing and Distribution
- Labour costs

3.

Based on the findings of how competitive the movie industry is in terms of market share, Microsoft will need to differentiate itself in order to stand out, for example:

- Microsoft's new movie studio could focus on producing high-quality movies with unique and compelling storylines, or even with diverse and inclusive casts. For example, they could explore new combinations of the top performing genres as stated in the first recommendation.

- Leveraging innovative marketing and distribution strategies to reach wider audiences.

- Partnering with well-known and respected directors and actors.

4.

Based on the findings of the best months to release a movie being May, June and July, the Microsoft new movie studio should consider releasing movies around this time.

There could be various factors that contribute to the high revenues in May, June and July:

- It’s summer blockbuster season, which typically runs from May to August
- Audience availability

Therefore, if Microsoft takes advantage of this period, the movies released are likely to yield higher gross revenues.

Next Steps

Further analyses could yield additional insights to further improve decision-making for the movie genres to produce in the new Microsoft studio:

- A further analysis into the directors/actors in the film industry. This analysis could provide insights on whom to hire during production of movies. Involving highly rated/successful directors & actors in production increases the probability of success for a movie.

- A further analysis into the reviews made my the public by conducting sentiment analysis on social media and other online platforms to gauge the public's reaction to movie trailers, posters, and other promotional materials.

Thank you!

If you have any questions or would like to discuss this topic further, please don't hesitate to ask me. You can contact me via my LinkedIn profile, which is linked below.

Bryson Shitsukane.

linkedin.com/in/bryson-shitsukane-584bb7161