✕

Learn Git and GitHub without any code!

Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

Read the guide

ᛦ **tiaplagata** / **dsc-phase-1-project-online**

forked from learn-co-curriculum/dsc-phase-1-project-online

DS Phase 1 Final Project - Online

⚖ View license

☆ **0** stars   ᛦ **22** forks

☆ Star	⊙ Watch ▾

Code	Pull requests	Actions	Projects	Wiki	Security	Insights	Settings

ᛦ master ▾   • • •

This branch is 8 commits ahead of learn-co-curriculum:master.

ᛦ Pull request   ± Compare

👤 **tiaplagata** ⋯	23 minutes ago ↻

View code

README.md   ✎

Welcome to this film industry exploration!

Navigation

- Project Overview
- Question 1
- Question 2
- Question 3
- Question 4
- Conclusion
- Future Work

Important Links

- Slideshow Presentation
- Non-Technical Video Walkthrough
- Jupyter Notebook with Cleaning & Questions

Project Overview

Microsoft is creating a new movie studio and has hired some new data scientists to explore the film industry. Our goal is to perform some preliminary analysis so Microsoft can take away a few actionable insights for their new movie studios in order to create the best movies possible.

In the process, I explored the following questions:

Question 1

What are the qualities of the highest-grossing films?

Scope & Data Used

To find this conclusion, we can look at the bom_movie_gross dataset from Box Office Mojo. The top-grossing films in this case are those with the highest domestic gross. We are looking at domestic gross over foreign gross because as a new movie studio, we should conquer our home market before trying to overstretch to foreign and domestic markets all at once.

Findings

The 4 studios with the top-grossing films are:

- Buena Vista (BV) | total domestic gross: $18.4 billion

- Universal Studios (Uni.) | total domestic gross: $12.9 billion

- Warner Bros. (WB) | total domestic gross: $12.1 billion

- 20th Century Fox (Fox) | total domestic gross: $1.1 billion

We can further investigate these top-grossing films by looking at the top 10 films for each of the top studios.

We can easily see a pattern in the qualities and genres of these films. Most of them are action films. They are superhero films (e.g. Avengers, Deadpool, The Dark Knight, etc), sci-fi/fantasy franchises (e.g. Star Wars, Jurassic World, Harry Potter, Dawn of the Planet of the Apes, etc), and animated films for kids & families (e.g. Incredibles, Despicable Me, etc).

Recommendations

Based on these findings, I would recommend making films that reflect these qualities (superhero films, animated films for families, sci-fi/fantasy films, etc.

Investing in a sci-fi/fantasy or superhero franchise seems to be a great investment. We can see a trend in these films over the past 10 years.

We can also conclude that buying the licensing rights to an popular sci-fi saga/series to adapt into a movie would be a good investment as well.

Question 2

Which films made the most money?

Scope & Data Used

To answer this question, we can look at the tn.movie_budgets database, and narrow down our data set to find all films between 2010 and 2020.

Findings

The first thing we should ask ourselves is: is there a relationship between the movie's production budget and the movie's gross earnings? When mapped in Seaborn's relplot, we can easily see this relationship.

In "Production Budget and Domestic Gross", we can see that as the production budget increases, so does the domestic gross. We can see that most of the movies with budgets less than 100 million dollars, do not make any more than 500 million dollars in domestic gross earnings. Movies that made over 600 million dollars all had production budgets of 200 million dollars or more. Moreover, if we look at the correlation coefficient between production budget and domestic gross, we can see that there is somewhat strong correlation between the 2 variables (0.73).

In "Production Budget and Worldwide Gross", we can see a similar trend. Our correlation coefficient between these 2 variables was 0.80, which indicates a stronger correlation than with domestic gross. It is interesting to note, that this correlation is not without exception. There were still various films with production budgets between 250 million and 350 million dollars with worldwide gross earnings less than 1 billion dollars.

We can therefore conclude that there is a strong relationship between production budget and gross earnings, but what about your return on investment? We can also investigate the type of movies that would produce the highest returns.

Calculating the ROI using the worldwide gross, we can see that many movies lost money (indicated by a negative ROI), but there were also many movies that made 50x or even 400x on their production budget. When we look at the top 20 movies with the highest ROIs, we can see right away that most of them are horror films.

Recommendations

Based on these findings, I would recommend that we should ask ourselves a follow-up question:

- *How big is our production budget?*

If we have a lot to invest, we should take into consideration our findings from question 1, and conclude that we can invest in a sci-fi/fantasy/superhero franchise or a film remake, use a large production budget, and make more gross earnings.

However, if we do not want to invest so much, we should make a horror film, in order to receive a large ROI.

Question 3

Which film genres have the highest ratings?

Scope & Data Used

To discover which types of films receive the highest ratings, we can look at the average ratings per genre. This question was answered using the imdb.title.basics and imdb.title.ratings databases.

Findings

Upon joining these two tables, we can see a large variety of films ranging from 2010 to 2019, which is the same range we have looked at for the last 2 questions. I decided to limit the data set to include only the ratings from movies with 50,000 votes or more, since the average rating is highly skewed for movies with less votes (i.e. a movie with 5 votes of all 9.0 ratings and up, will give a very high average rating based off little data).

Once we separate the movie genres into single-genre categories, we can look at the medians and distributions of average ratings for each movie genre. This information tells us which movie genres generally receive higher ratings, and when looking at the violin plot, "Average Ratings Per Genre", we can also see the distribution of each genre's average rating.

The movie genre with the highest median for "average rating" (the column name is average rating) is "Documentary", with a median rating of 8.10. Second highest is "Biography", with a median "average rating" of 7.45. And tied for second highest is "War", with a median "average rating" of 7.45. However, there is not much variation between each median "average rating", as all of the medians fall between 6.30 and 8.10. Therefore, even though "Documentary", "Biography" and "War" movies are the highest rated movies, their ratings are not much higher than "Animation", for example, with a median "average rating" of 7.20.

What the violin plot, "Average Ratings Per Genre", does tell us, is the following:

- the long tails on violins indicate outliers in their ratings (e.g. Documentary, Biography, Music)
- the taller violins indicate genres with more varied ratings (e.g. War, Western, Sci-Fi)
- the wider violins indicate a higher probability that the rating will be in that range (e.g. History, Animation)

Recommendations

Based on these findings, I would recommend making films with less varied ratings if we are looking to make crowd-pleasing films.

Documentaries seem to be extremely varied, so we should avoid making those films, whereas Animated films seem to be well-liked with less variation in ratings.

Question 4

Scope & Data Used

To answer this question, the tn.movie_budgets database was helpful in providing movie release dates, as well as gross domestic earnings. We used domestic gross over worldwide gross since holidays vary from country to country. With this database, I included movies from 2010-2020, which gives us data from 10 occurrences of each month.

Findings

When we look at the domestic gross by month over the past 10 years, we can find which months have the highest domestic gross. We can use the median domestic gross per month as a good indicator to rule out outliers. Thus, the highest grossing months are:

- July - $31,206,263

- November - $30,659,817

- August - $21,295,021

- April - $20,316,694

- February - $19,452,138

Whether or not you release the movie directly on a holiday (i.e. Thanksgiving Day) does not really affect it's gross. In the plot, "Movie Releases by Month", we can see lots of green dots in December, indicating that the movie was released on a holiday, yet December had the lowest median domestic gross earnings.

Recommendations

Based on these findings, I would recommend releasing our films over the summer, or November. February also seems to be a great time for releases.

I would also recommend NOT releasing a film in December, because it had the lowest domestic gross by month. This could be due to the fact that many other studios release films during this time, or because people are generally busy with family, travel, etc over this time, but it would take more research and data to understand why this happens.

Conclusion

In conclusion, I discovered that the best films to create should reflect the following recommendations:

- If we want to start with a small budget, we should focus on horror films, which produce the highest returns.

- If we want to jump into the industry with a large budget, we should either:

  - Invest in licensing rights to a sci-fi/fantasy/superhero franchise and create films based on that
  - Invest in an animation studio and create animated family films, which will also likely garner positive reviews.

- We shouldn't worry too much about the ratings because each genre seems to receive a similar range of ratings. However, if we are aiming to please, it might be a good idea to avoid documentaries, on which people have more extreme opinions.

- We should plan on releasing our films in the summer, specifically July or August, or in November, in order capture the highest gross earnings.

Following these recommendations should lead us on a path for a high probability of success.

Future Work

If I had time to explore further, I would investigate the following:

- Build a predictive model for movie trends

- Calculate a more accurate ROI model based on all costs (i.e. marketing budget, distribution costs, etc)

- Dig deeper into film ratings, pulling the Rotten Tomatoes data as well

- Figure out which genres to release during each month -- Is there a seasonality to certain genres' release dates?

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%