ⵔ **casanave** / **Film_Analysis_Project**   (Public)

forked from **learn-co-curriculum/dsc-project-template**

<> **Code**   ⵗⵔ Pull requests   ⊙ Actions   ▥ Projects   ▢ Wiki   ⊘ Security   ⯞ Insights

ⵔ master ▾   ⋯

Film_Analysis_Project / README.md

👤 **casanave** Update README.md   ⟳ **History**

👥 **3 contributors**   👤👤👤

≔   65 lines (51 sloc)   │   4.87 KB   ⋯

Movie-Recommendation-Project

Flatiron Phase One Project: Analysis Project to Make Recommendations to Hypothetical Microsoft Movie Studios

Overview:

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.

Citation:

Much of this project and its code come directly from https://github.com/ElyLin/MovieAnalysisProject where I was a contributor. After being a contributor to the group project, I wanted to go back and re-explore my process and analysis on my own.

In my presentation, I recommend making films based on DOOM, Wolfenstein, The Elder Scrolls: Skyrim, Fallout and Dishonored, all Bethesda games that Microsoft bought the rights to in 2020. https://news.microsoft.com/2020/09/21/microsoft-to-acquire-zenimax-media-and-its-game-publisher-bethesda-softworks/

Business Understanding:

In this project, I hope to make a recommendation to Microsoft based on the Net Profit of the films. I'm using Net Profit as my main financial data metric since Microsoft is already so large and successful that they can afford to put some investment into their films. They don't need to focus on getting a high profit from a low-budget film, they don't need to be so conservative with their investments. They can instead safely invest by following my recommendations.

I will be making recommendations based on: Genre, Release Date of the Film, and the Runtime of the Film.

Data Used:

movie_budgets_df = pd.read_csv('zippedData/tn.movie_budgets.csv.gz')
title_basics = pd.read_csv('zippedData/imdb.title.basics.csv.gz')
title_ratings = pd.read_csv('zippedData/imdb.title.ratings.csv.gz')
tmdb = pd.read_csv('zippedData/tmdb.movies.csv.gz')

Methods for Data Cleaning:

1. Use a dictionary to decode the encoded Genre Ids from TMBD
2. Merged both IMDB datasets together using an inner merge
3. Combined the Title of the film with its IMDB title ID to make a unique identifier
4. Merged IMDB with TMDB datasets together using an inner merge
5. Incorporated movie budget data using an inner merge
6. Removed outliers by getting rid of films less than 45 min and more than 180 min
7. Dropped unnecessary columns
8. Reformatted financial data into integer format for manipulation
9. Checked for null values and removed small subsection of rows with null values
10. Checked for duplicates

Size of Data:

After cleaning and removing null values and duplicates, this analysis will use a dataset the size of 2,864 films.

Meathods of Analysis:

Genre:

1. Made a row for each film for each genre it fell into
2. Aggregated data based on genre and mean net profit
3. Visualized aggregated data to learn that the genres with the most net profit are: Animation, Adventure, Fantasy, and Family

Release Date:

1. Manipulated release date information into a date-time value
2. Made a new column for the month in numeric and Name values
3. Made a pivot table to aggregate data based on month with average net profit for: Animation, Adventure, Fantasy, and Family films
4. Visualized Data using a stacked bar chart where each bar is a month, broken into genres to show overall net profit
5. Visualization showed the most profitable months to release films in these genres are: June, May, March of any genre, or July with an emphasis on non-Fantasy films (which, interestingly do much better in November.)

Runtime:

1. Made initial scatterplot to show net profit based on runtime
2. When initial scatterplot was inconclusive, made runtime data into four bins: Sixty, One hundred twenty, One hundred fifty, and One hundred eighty
3. Aggregated runtime data based on four bins by net profit
4. Visualized in a bar chart four bins by net profit and learned that films between 120-150 minutes in those four recommended genres did best consistently

Conclusion

My three business recommendations to Microsoft Studios would be:

1. Focus on making films that are a combination of animation, fantasy, adventure, and family. There are many examples of the combination of these genres already, (Pixar

Studios being the best example of having luck here.)

2. Release this film in June, May, March, July or November

3. Make the movie somewhere within 120-150 minutes long since movies in this range tend to make the most net profit on average.

Future Work

I'd love to, if the project parameters were a bit larger, take a look at what studios tend to do best within the three recommendations I've already given. I'd also love an opportunity to work with review data and see how films that fit these three categories are normally rated and reviewed.