✕

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

🗂 **shackemn** / **dsc-phase-1-projectA**

⚖ View license

☆ **0** stars   ⑂ **0** forks

☆ Star	⊙ Unwatch ▾

‹› **Code** ⊙ Issues ⑂ Pull requests ▷ Actions ▥ Projects 📖 Wiki 🛡 Security

⑂ master ▾   ⋯

shackemn Add files via upload …	2 minutes ago 🕐 **10**

View code

README.md   ✏

Phase 1 Project: Microsoft Movie Industry Analysis

Overview

This project follows Microsoft and their entry into the movie industry. Since this is not an area of business they have ever been in, we have been tasked with analyzing the industry and assisting with key decision-making processes.

Data

I used the following data tables in my analysis. The data comes from Box Office Mojo, IMDB, and The Movie Database

- imdb.title.basics
- imdb.title.ratings
- bom.movie_gross
- imdb.name.basics
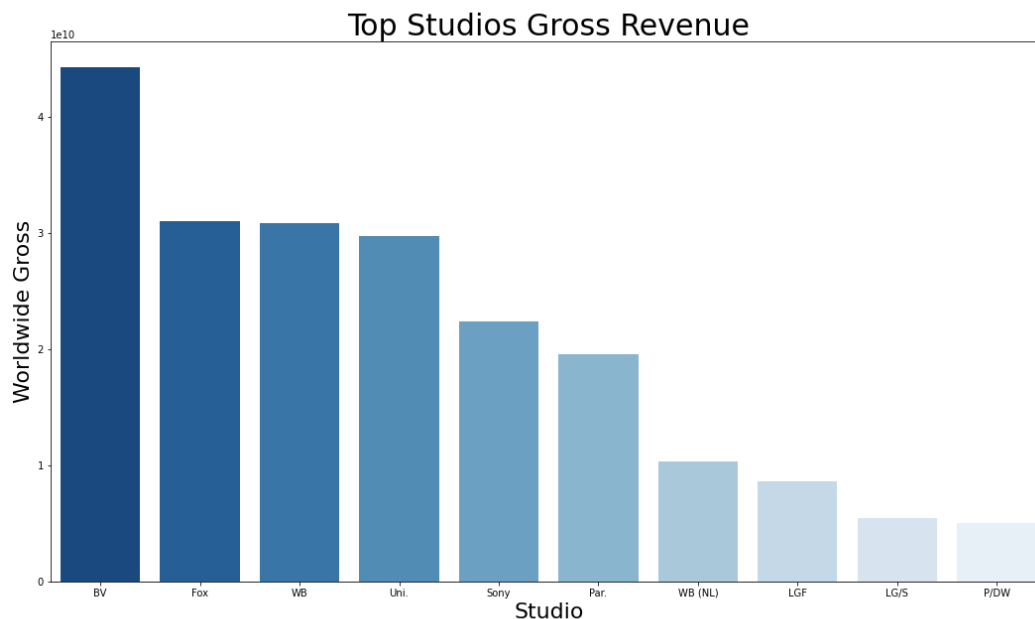- imdb.title.principals
- tn.movie_budgets

It is up to you to decide what data from this to use and how to use it. If you want to make this more challenging, you can scrape websites or make API calls to get additional data. If you are feeling overwhelmed or behind (e.g. struggled with the Phase 1 Code Challenge), we recommend you use only the following data files:

- imdb.title.basics
- imdb.title.ratings
- bom.movie_gross

Question 1: Which movie studios are the highest grossing?

In order for Microsoft to make the right decisions, I decided to look at current movie studios. I looked for the studios with the highest grossing revenue across the board. With this information Microsoft can choose to either emulate these studios or become partners with them.
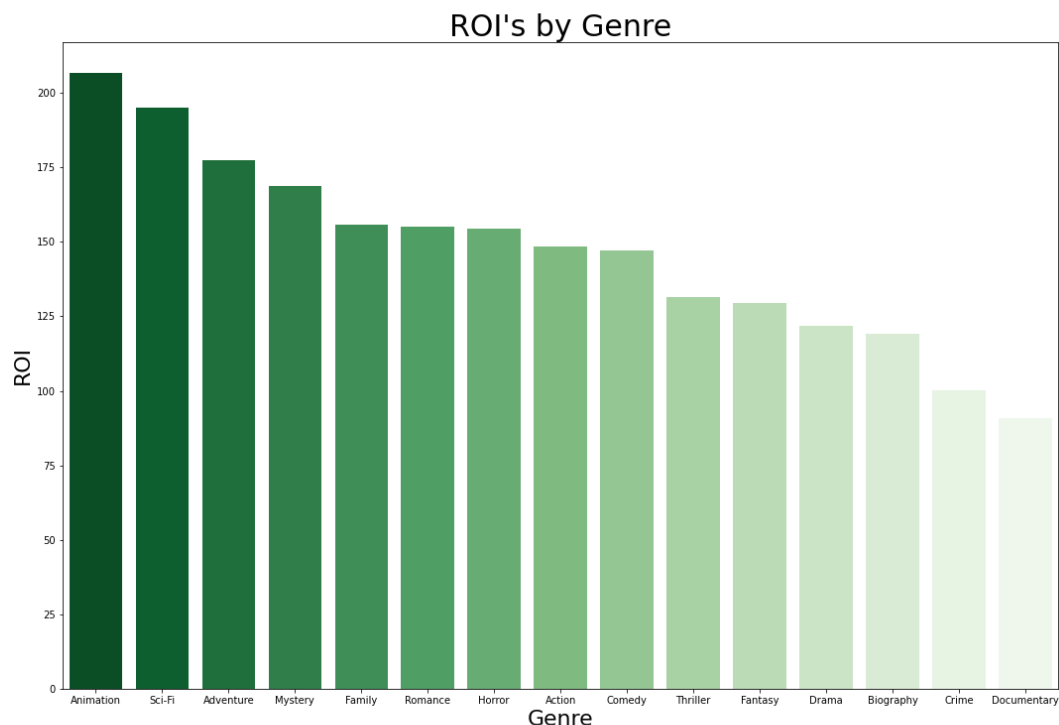
After cleaning the data from the bom.movie_gross file, I added a new column summing up the domestic and foreign gross. I then used pandas to group by studio to show the sum of total gross revenue. Then I visualized the data.

The Gross Revenue is measured in the billions, so the majority of the top 10 studios had over a billion dollar revenue from their movies. This shows that this venture is a worthwhile project and which movie studios they should emulate.
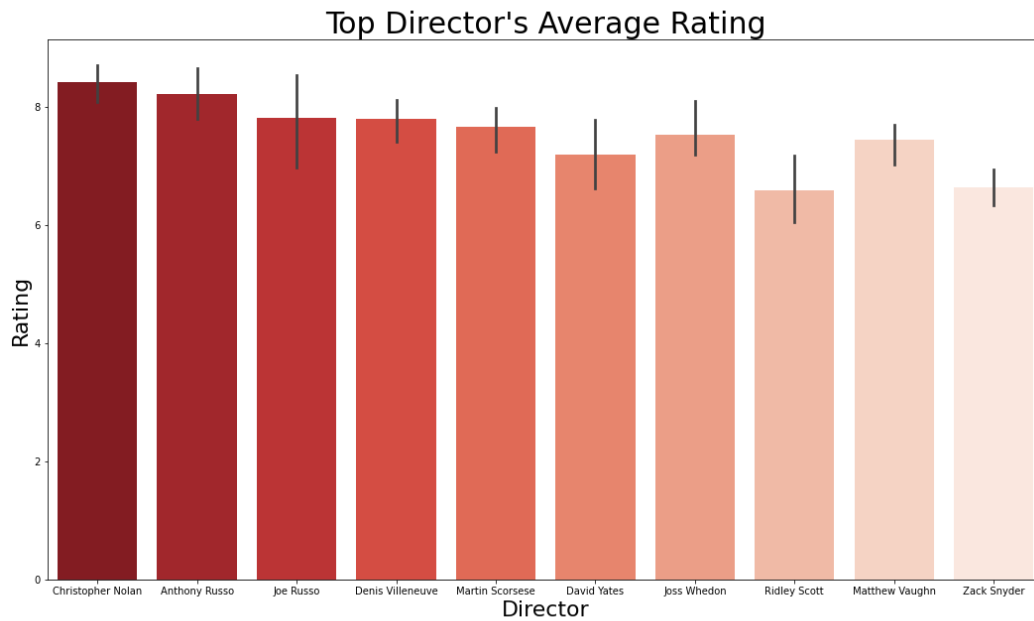
Question 2: Which movie genres had the highest ROI

In order to answer this question, I cleaned and joined the imdb.title.basics and tn.movie_budgets tables. The genres in the "basics" table needed to be split into a list in order to work with the data. The "budgets" table had all of their monetary values entered as strings including dollar signs and commas. I had to remove those and change the type into floats before I could work with the data. After doing these things I calculated the ROI for each entry using the worldwide_gross and production_budget. I then removed some outliers using the IQR method. Then I grouped the data by genre and calculated the mean of those genres.

This visualization shows that the highest return on investments come from the Animation and Sci-fi genres though there were several that were relavtively high. This data will help Microsoft choose what types of movies they should focus on first for their first venture into the movie industry. Note that almost all movies have more than one genre, and a combination between some of the highest ranking genres should probably be chosen.

Question 3: Which directors are the most popular?

To answer this question I had to join and clean several different tables including: imdb.title.basics, imdb.title.ratings, imdb.name.basics, and imdb.title.principals. I dropped many columns and rows, so I could focus on the directors and ratings themselves. After this was done I found the highest rated directors using their averaged ratings across their movies. I first had to filter out some of the data. For instace, some of the directors had a perfect 10/10 rating due to the fact that they had a total of 5 votes. I chose only to take the 10 directos who had the highest number of votes(the higest had 499,479 for reference). I then grouped the data my director and calculated the mean rating for each of them.

This visualization shows that the top directors are very highly rated from people. Microsoft should consider hiring one of these directors in order to not only interest people to come see the movies, but also increase the chance of the movies' success.

Recommendations

- Microsoft should try to partner or at least to emulate the business decisions of the top movie studios, especially Disney whose gross revenue was much higer than the others.

- Deciding which genre for the first movie is highly important. I recommend choosing some combination of the animation, sci-fi, mystery, and adventure genres.

- Think about if Steven Spielberg was coming out with a new movie, and how many people would go see it. Choosing the right director can have an extremely large effect on revenue and success of a movie. I recommend hiring one of the most popular directors to work on the first film. This will generate buzz about the new movie studio. I suggest trying to hire Christopher Nolan or one of the Russo Brothers.

Future Work

- We could further investigate directors since the most popular ones will come at a high cost. We should look at ROI's of these directors' movies to further analize the data for an informed decision.

- Check to see if there is a particular month or holiday that we should focus on to release certain genres.

- We should further look into popular studios. What genres are they producing? Do their budgets for certain genres differ from others?

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 2

DavidBraslow David Braslow

shackemn

Languages

● **Jupyter Notebook** 100.0%