

Microsoft Movies Project

Path to becoming a leader in original video content

Summary

Exploring the IMDB database and Box Office Mojo CSV file, we have articulated our research around three axes, Revenues, Rating and Genres, in order to establish our investment strategy into movies production.

In a crowded, declining market, we have identified Animation, Adventure, Mystery and Action as the Genres most likely to harbor high Revenues projects.

We will stay away from crowded Genres and low to middle quality productions, focusing instead on high quality projects in the selected Genres.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- **An opportunity too good to miss**

Size of the Movie & Video Production market in the US is in excess of \$18 billion per year, of which Microsoft is in a position to command a significant market share.

- **Requiring, sound, discriminate choices**

While the market grew in 2022 by 1.9%, for the last 5 years it has contracted by 8%. Since we cannot count on industry growth alone, we need to make sound, data centric, investment choices.

- **Supported by in depth data analysis**

Our data analysis helps identify and address market trends and opportunities, leading to optimal investment decisions.

Data

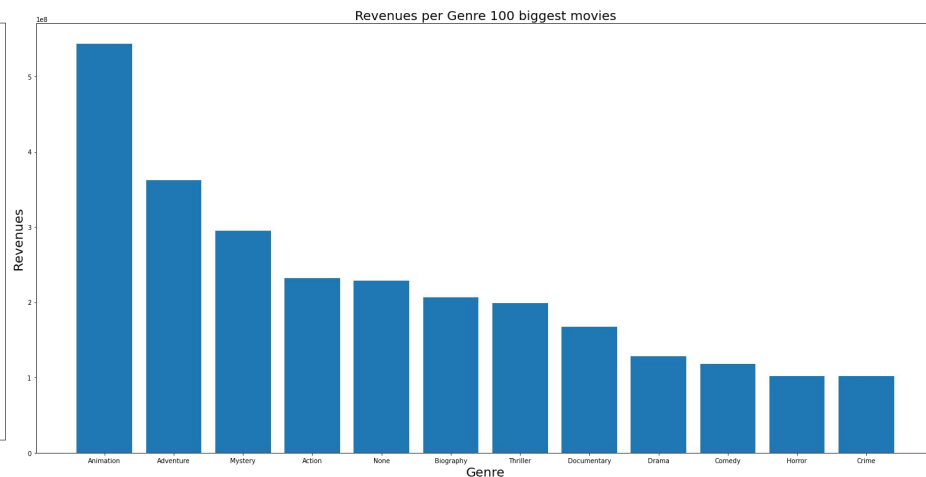
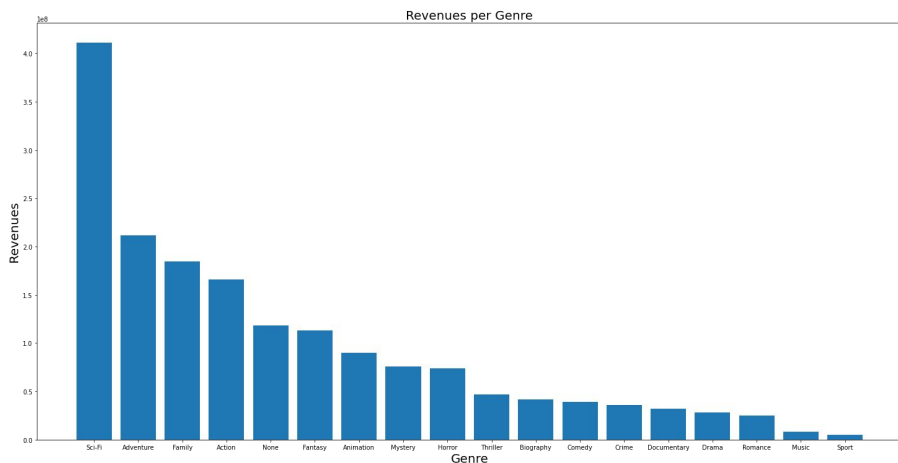
- We have focused on two primary sources:
 1. The IMDB database, covering more than 146,000 movie titles, allowing an in depth analysis of their Genres and Ratings.
 2. The Box Office Mojo CSV file, providing Revenues numbers, both foreign and domestic, for 3356 major movie titles.

Methods

- The Box Office Mojo CSV file was transformed into a Pandas DataFrame, cleaned (movies without domestic revenues have been discarded and all revenues transformed into numerical values), while a new column has been created to reflect each movie's Global revenues. This provides an ordered list of movies by Global Revenues.
- From the IMDB database, we have LEFT JOINed the movie_ratings and movie_basics tables in order to obtain a table combining movie ratings and genres. After preparing the data (dealing with Null values and selecting the first stated Genre as representative for each movie), we have produced a Pandas DataFrame, articulating the viewers' rating to each movie Genre.
- Finally, we have combined the 2 DataFrames in order to analyse a subset of data articulating Genres, Rating and Revenues. We have analysed the full subset, as well as the top 100 and 500 movies by Revenues.

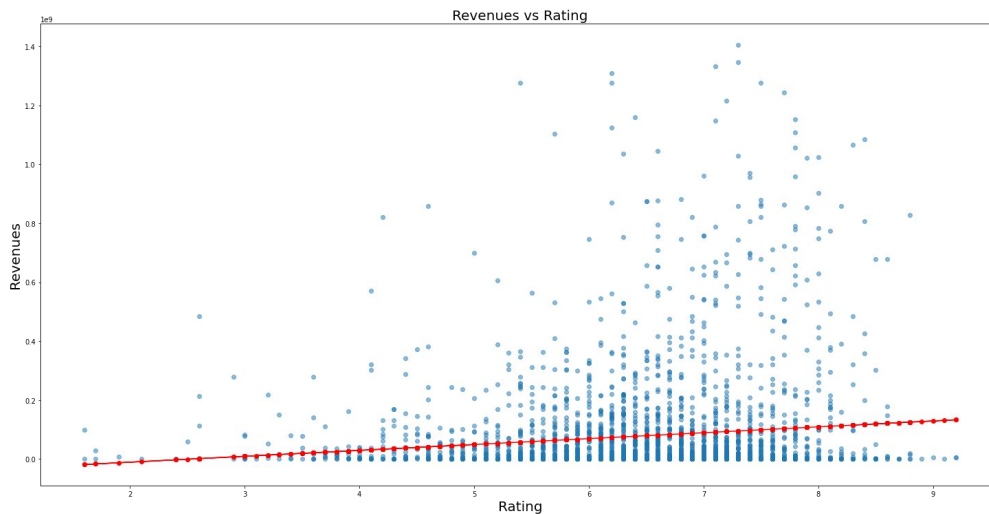
Results (1)

- The Sci-Fi Genre is by far the one generating the highest Revenues, followed by Adventure, Family and Action. Yet, while looking at the top 100 movies in term of Revenues, the Animation Genre comes on top, again followed by Adventure.



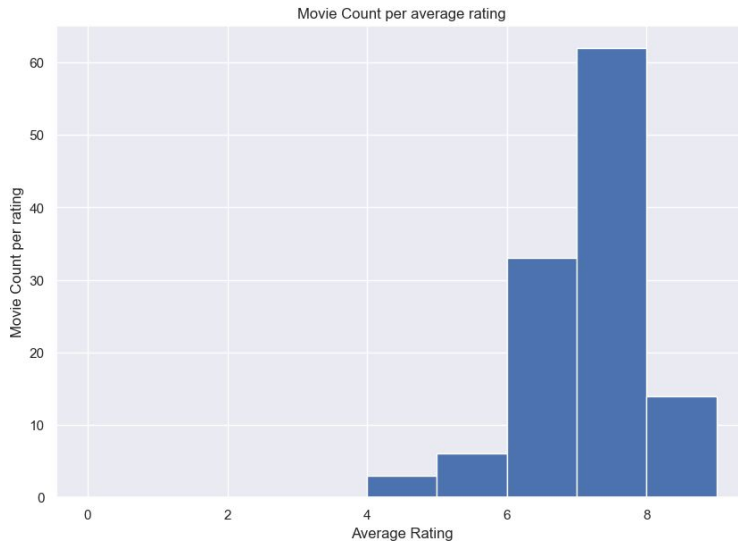
Results (2)

- Viewers' rating is a poor predictor of movie Revenues. None of the top Genres in Rating terms are in the top 5 Genres by Revenues. More importantly, the correlation between Ratings and Revenues is low (though positive).



Results (3)

- Nonetheless, while Rating overall doesn't predict Revenues, the immense majority of movies with Revenues above \$500 million has ratings above 6 out of 10.



Conclusions

- Our investments should focus on Genres with the highest propensity to generate high Revenues per project: Animation, Adventure, Mystery and Action.
- We should avoid investing in crowded genres (Sci-Fi) and average ratings productions, low predictors of both Revenues overall and per project.
- We should position ourselves as producers of quality entertainment, aiming for viewers' rating above average (at least 6 out of 10)

Thank You!

Email: paul.poziumschi@gmail.com

GitHub: @Paul11one

LinkedIn: [linkedin.com/in/paul-p-754b4954/](https://www.linkedin.com/in/paul-p-754b4954/)