

Dmvinedata / Phase1_Project Public

This is the Phase 1 project repo. This project was completed June 2022

 View license

 1 star  0 forks

 Starred

 Unwatch

 Code

 Issues

 Pull requests

 Actions

 Projects

 Wiki

 Security



 main



Dmvinedata Final1 ...

4 minutes ago

 25

[View code](#)

 README.md



Microsoft Movie Studio Analysis

Authors: Deztiny Jackson

[Presentation Link](#)

[Jupyter Notebook Link](#)

Overview

This analysis directs Microsoft's potential studio head with actionable insights for their movie studio development. These recommendations are determined from insights from box office movie "Ratings" and depends on other attributes such as "Genre", "Directors" and "Movie Budget" data. The datasets given were filtered to analyze movies and their associated attributes that have a minimum "8/10" ("B") rating. The analysis observed and extracted patterns from data of some of the "best" movies.

Business Problem

"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."

Ref: [Phase 1 Project Description](#), 2022

Microsoft is in competition with current major studios who have been in business for many years and already have a wealth of experience. The questions driving the analysis, support Microsoft making choices based on direct data output from their competition as well as potential growth areas that may be overlooked. The analysis questions and main variable of success were chosen with the mindset that Microsoft is a major tech company that has influence and skillsets in other areas of tech. This analysis is for a movie studio which is a part of a greater ecosystem.

Data Understanding

Box office related datasets (SQL and CSV) with the box office movie target "Rating" variable and associated independent "Genre, Director and Budget" variable information were taken from [IMBD](#) and [the-numbers](#).

IMBD Tables Used:

- **movie_basics**
 - The movie_basics dataset describes just over 146,000 movies with their associated genres, years.
- **movie_ratings**
 - The movie_ratings dataset describes 73,856 movie's average rating (from average ratings that range from "1" to "10") and the number of votes the average is calculated (ranging from "5" to "1.84 Mil").
- **directors**

- The directors dataset describes 291,174 movies and their associated director's id.
- **persons**
 - The persons dataset describes 606,648 people and their associated movie related profession(s).

Budget CSV:

movie_budgets

- This dataset describes the movie budget and movie gross from 5782 movies. The movie name values are the only link to joining with other movie datasets. This can potentially be a single point of failure or limitation in the data analysis.

Data Analysis

The data is modeled using scatter plots and colormaps. This modeling method allowed the use of 3 main variables depending. I could show the correlation between two independent variables as well as the a Rating (using color) on each visual.

The initial plan was to use a different type of graph for the different independent variables. Because of complexity and the ability to show the correlation between more than two variables a scatter plot with a colormap was used each time.

The three main independent variables chosen where "Genre, Director & Budget". When looking through the data, other variables impacted the interpretation of this data in relation to the Rating (e.g. number of votes).

Ref:[Austin Animal Center Needs Project Example](#)

Genre Questions:

- What are the top rated genres?
Consider the number of votes to determine he average.
- What is the number of movies each genre contain.

Director Questions:

- How many different type of genres does each director have?
- How many movies does each director have?

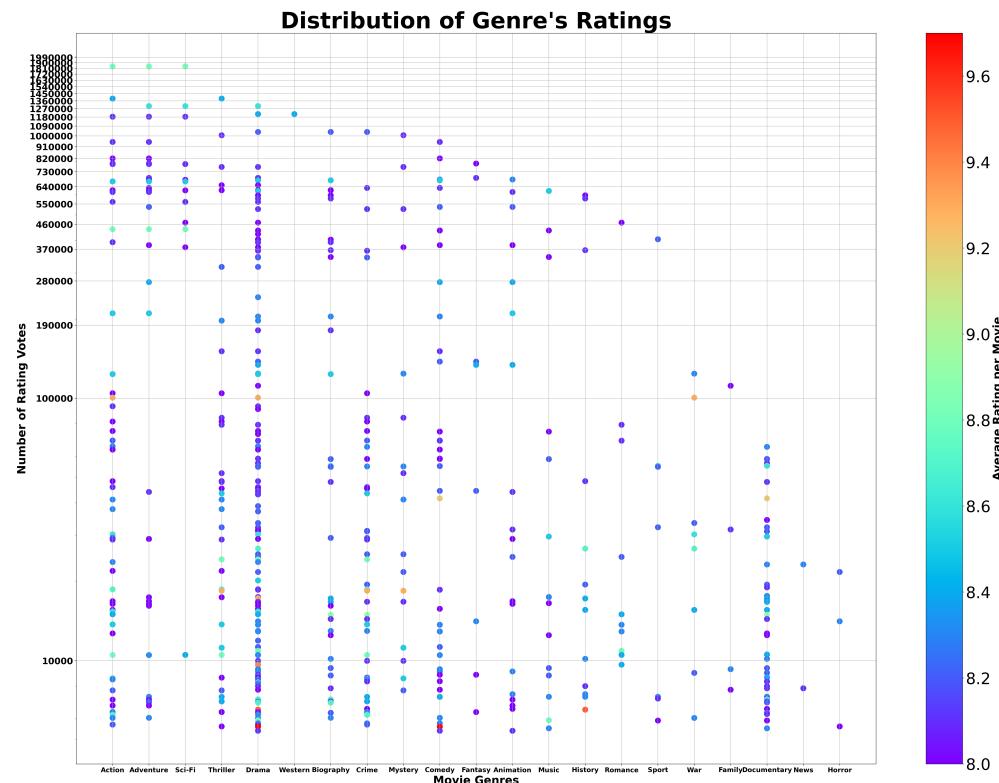
- Which directors does each genre have movies in?
- Who are the top 20 directors that have more than 2 movies with top ratings?

Budget Questions:

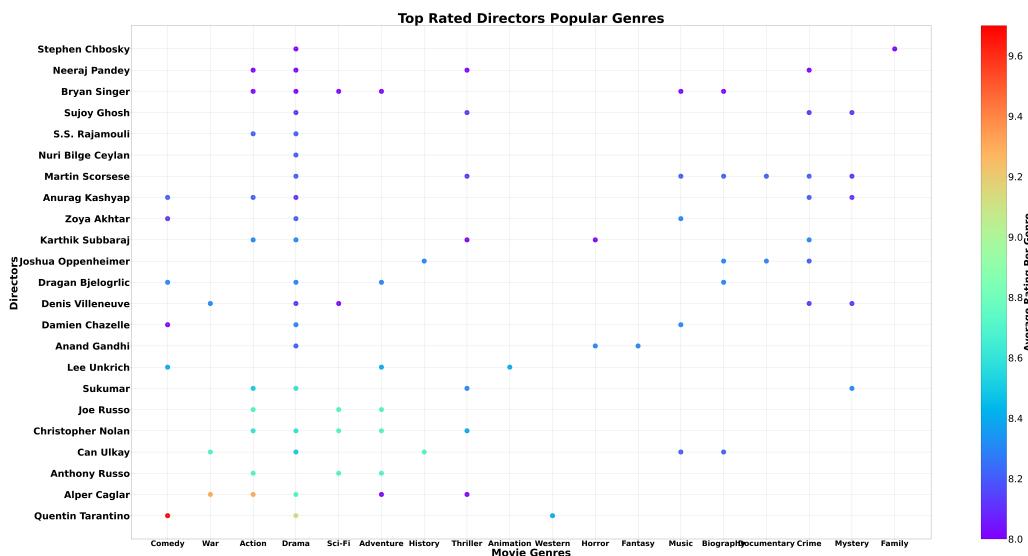
- How does the different dataset impact the findings?
- What types of movies have the highest, medium and lowest budgets?
- Which directors and genres are associated with the highest budgets? Patterns?

Visualizations

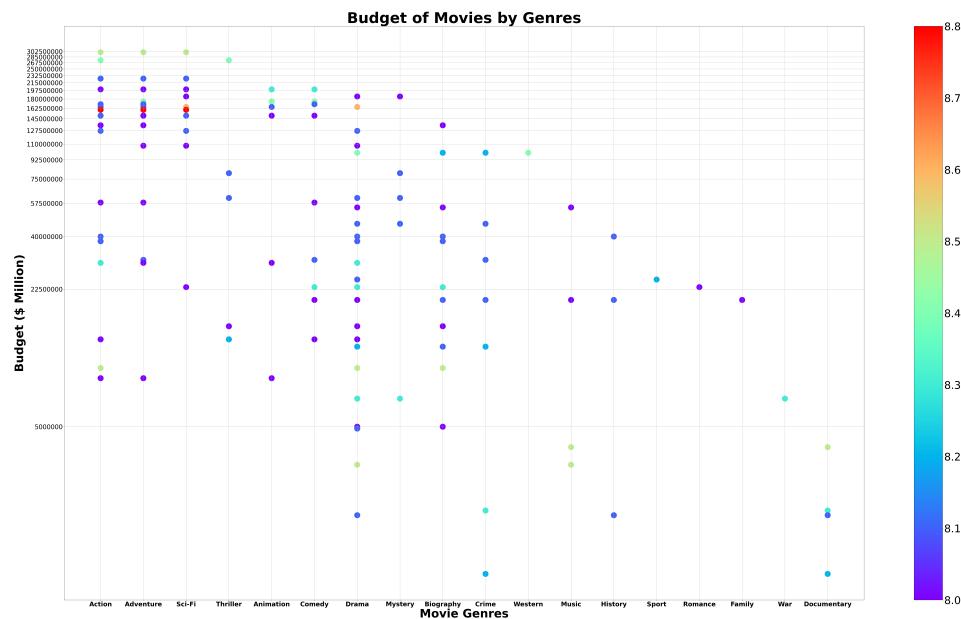
Genre Analysis



Director Analysis



Budget Analysis



Conclusions

As a for profit company, high revenue is always a goal. The "rating" was chosen as a target variable because it can help support the goal of revenue, but also provide other insights that can be further explored and explained. The following are recommendations for Microsoft from the analysis:

1. Invest in at least one of high budget films in the category of "Action, Adventure or Sci-Fi".

- If the action is taken, I highly recommend the one of the Russo brothers or Christopher Nolan directs this movie. They have some of the highest ratings in these categories multiple times.
- Movie studios may have major movies that they can be associated with. If Microsoft wants to establish themselves in the industry they need one (or more) as well.

2. Invest in multiple movies in "Drama, Documentary and/or Crime". If one genre, choose "Drama". These types of movies are popular and lower cost risk.

- From the movies that had at least a 8/10 average rating, dramas ranked first(124/173 movies), 3rd was documentaries (53/173 movies) and 4th crime (50/173 movies).
- All three of these genres were across the spectrum in terms of budget and the number of votes for the rating.
- Most of the directors who had multiple movies credits had a movie categorized as "Drama."

3. For the *initial* start of the movie studio, *DO NOT* invest resources in genres focused on "Family, Sports or News". These are higher risk due to lack of highly rated rates compared to the other genres.

- There is a substantially lower level of movies in these genres "Family (4) and Sports (9) and News (4)"
- Their budgets are a lot lower overall and pose a risk in the number of viewers leading to engagement and profit.
- A lot of multi movie directors do not work in this genre.
- The low amount of budgets, ratings and directors, increase the risk of their reception compared to other genres.

Limitations and Improvements

The following are analysis limitations and analysis future analysis ideas for Microsoft:

1. Microsoft needs to understand and communicate their constraints and vision for their movie experience. The results given were focused only a few parameters. For more fidelity, the recommendation is to add parameters and/or dive deeper into some of the recommendation's patterns. I would add parameters such as regions where the movie plays as well as languages. These can give more information and nuances as to where to place theatres and certain movies

2. The data analysis was limited in term of the methods, scope of the analysis and data parameter tuning. The problem given was a one paragraph statement. In a usual business analysis case, there would be more iterative communication with the stakeholder to clarify assumptions, considerations and the specific problem need. Even if Microsoft doesn't know the exact problem, continuous communication would refine the vision, values, and initial needs. This analysis was an initial low fidelity partial solution to the problem.

3. Future improvements:

- Add more parameters to analysis (e.g. region, language). These can give more information and nuances such as where to place theatres and certain movies from the genre recommendations.
- Use the recommendations and do more detailed work into the genres (causality and correlation).
- Analyze the combinations of genres with respect to the target variable and independent variables. For example, movies that were categorized as "Comedy and Animation or Sci-Fi" had higher budgets than "Comedies" that were not.
- Research more on Microsoft's current technical ecosystem and their holistic vision as a company. The potential movie studio is a part of a great system and it is imperative this node harmonizes with the rest of the ecosystem. For example, Microsoft recently invested millions of dollars into the metaverse. The movie studio and the metaverse have many connection points (e.g. brands, storylines, characters, stakeholders). Understanding these points, can help scope and clarify the analysis entities area for a better outcome.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

https://github.com/Dmvinedata/Phase1_Project

[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%