



ShamlaA /
DSF-PT07



<> Code

Issues

Pull requests

Actions

Projects

Security

Insights



DSF-PT07 / Project_P1 /

Add file



Shamla Araya created a readme file

3a9d317 · 29 minutes ago



| Name | Name | Last commit date |
|---------------------------|-----------------------------------|------------------|
| .. | | |
| Images | Created a new .ipynb file for ... | 4 hours ago |
| data | created a readme file | 29 minutes ago |
| .DS_Store | created a readme file | 29 minutes ago |
| .gitignore | created a readme file | 29 minutes ago |
| DS Project Presentatio... | created a readme file | 29 minutes ago |
| README.md | created a readme file | 29 minutes ago |

README.md



EDA FOR MICROSOFT MOVIE STUDIO

Data Analysis by: Shamla Araya

DSF-PT07 @ MORINGA

Overview

This project performs an exploratory data analysis to learn the current pattern in the movie industry to come up with recommendations that would help Microsoft's new movie studio to get into this competitive market the right way. Movie data collected from different globally renowned sources will be used for this analysis and it will cover the time period from 2013-2019. Microsoft Movie Studio will find this analysis very useful in guiding their decision-making process regarding what movie genres to focus their investment in.

Business Problem

Microsoft has decided to get into the movie business after observing that the movie business is becoming an attractive business. However they are facing a challenge on how to get into it as they don't know which movie genres are highly popular among audiences, which genres are most profitable and what are the production costs for these genres.

The purpose of this analysis is, therefore, to provide real life data based recommendations to tackle the above stated challenge. For this reason I will answer the following key points in my analysis:

- How many movies are made of each genre during 2013-2019?
- What genres are the most popular among the audience?
- Which ones are the top 5 genres that earned high profit in 2013-2019?
- What is the relationship between the production cost and the profits?

The answers to these questions will enable Microsoft enter the business with sharp competitive edge and actually produce movies that will return good profits.

Data Understanding

The source of the data for this analysis are:

1. IMDB ['movie_ratings', 'movie_basics'](movie_id, primary_title, start_year, genres, averagerating)
2. The MovieDB (title, popularity, vote_average)
3. The Numbers (movie, domestic_gross, worldwide_gros, production_budget)

My goal for this analysis is requires that I use the variables from IMDB, TMDb and TN databases. You will notice that, after carefully reviewing the data, I have changed some of the column (variable) names to make it possible for me to join and marge the different datasets into one complete data frame for my analysis. I have also filtered the data to avoid any missing values and specify the period of my analysis which is 2013-2019, during the cleaning phase.

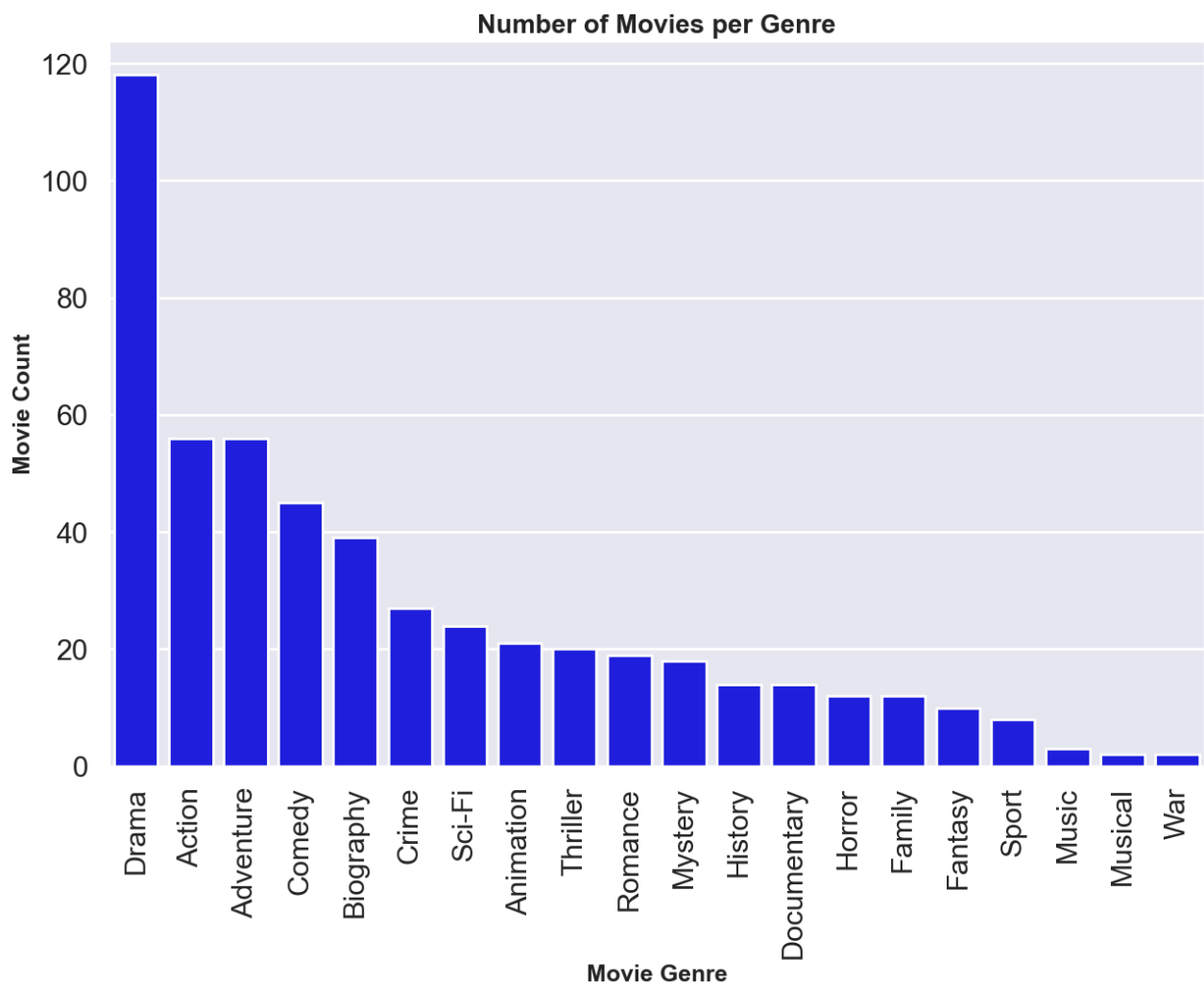
VISUALIZING THE FINDINGS

In the cells that follow are the visual representations of the analysis I carried out namely:

1. Number of movies per genre.
2. Genre popularity.
3. Total net profit per genre.
4. Relationship between production cost and profits.

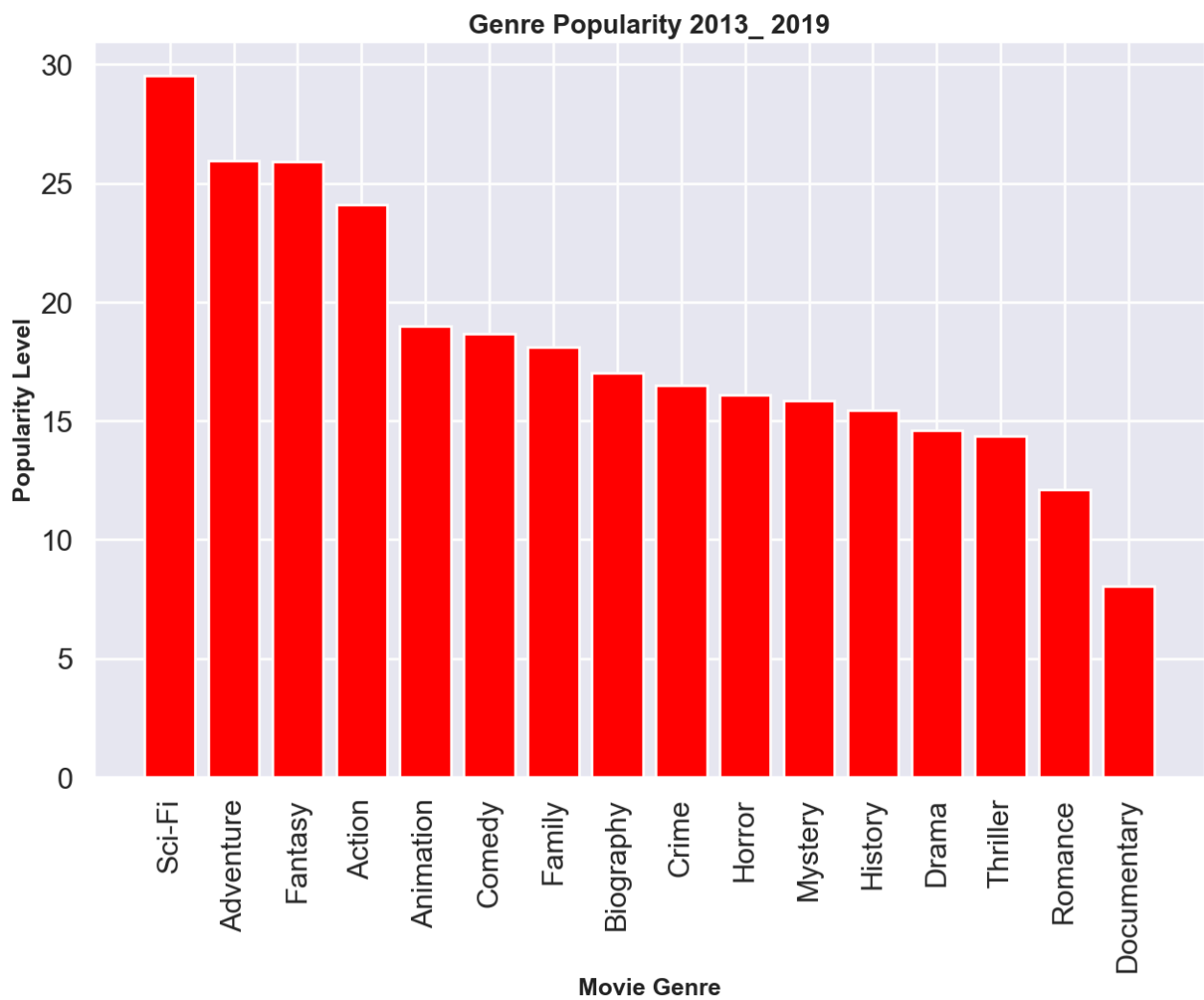
Getting Started

1. Number of movies per genre.



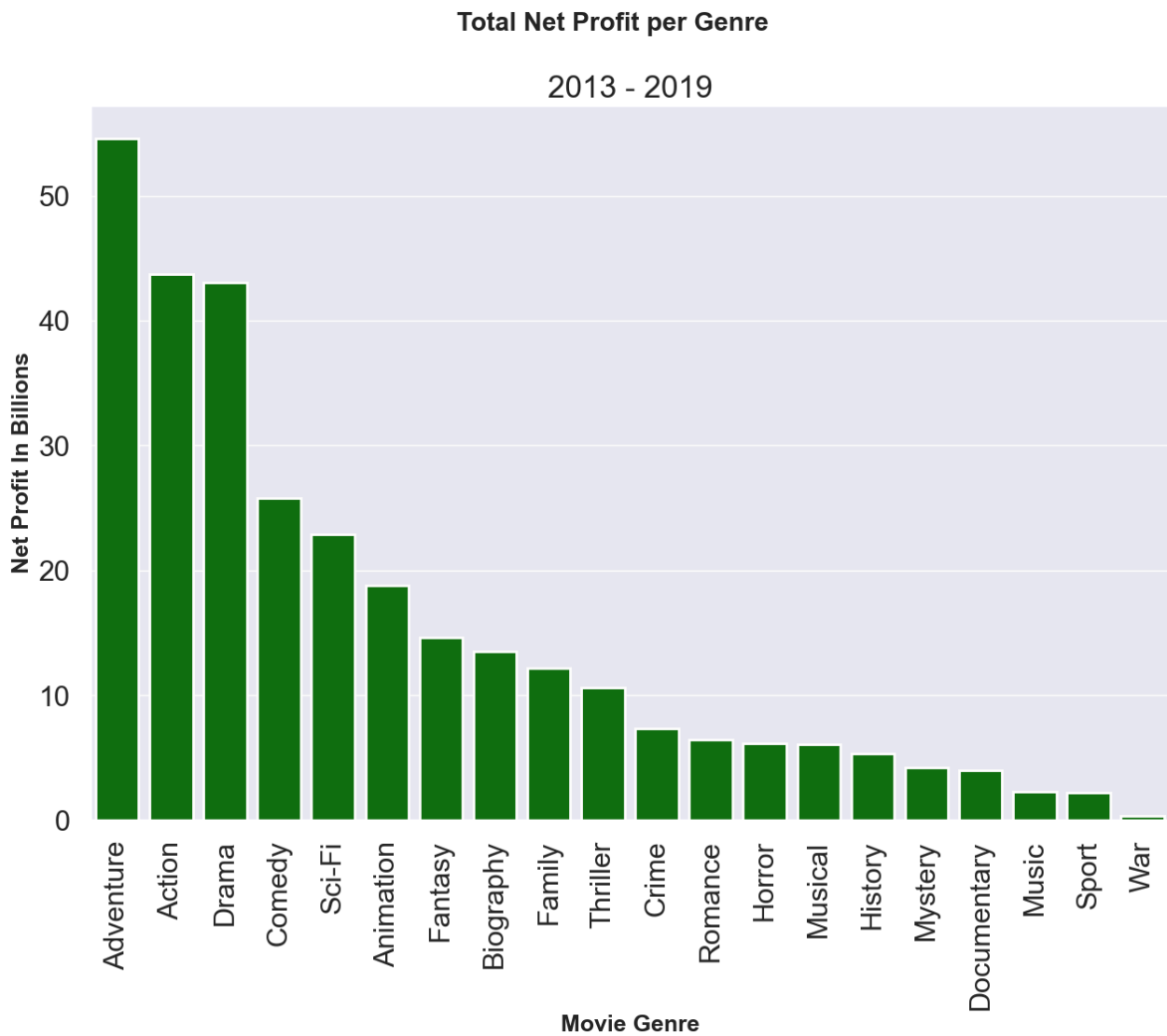
According to the graph above, majority of the movies produced between 2013-2019 are classified as Drama. The figure exceeds all other genres by at least double.

2. Genre popularity.



This plot shows that SciFi, Adventure, Fantasy, Action and Animation are the top 5 most popular genres.

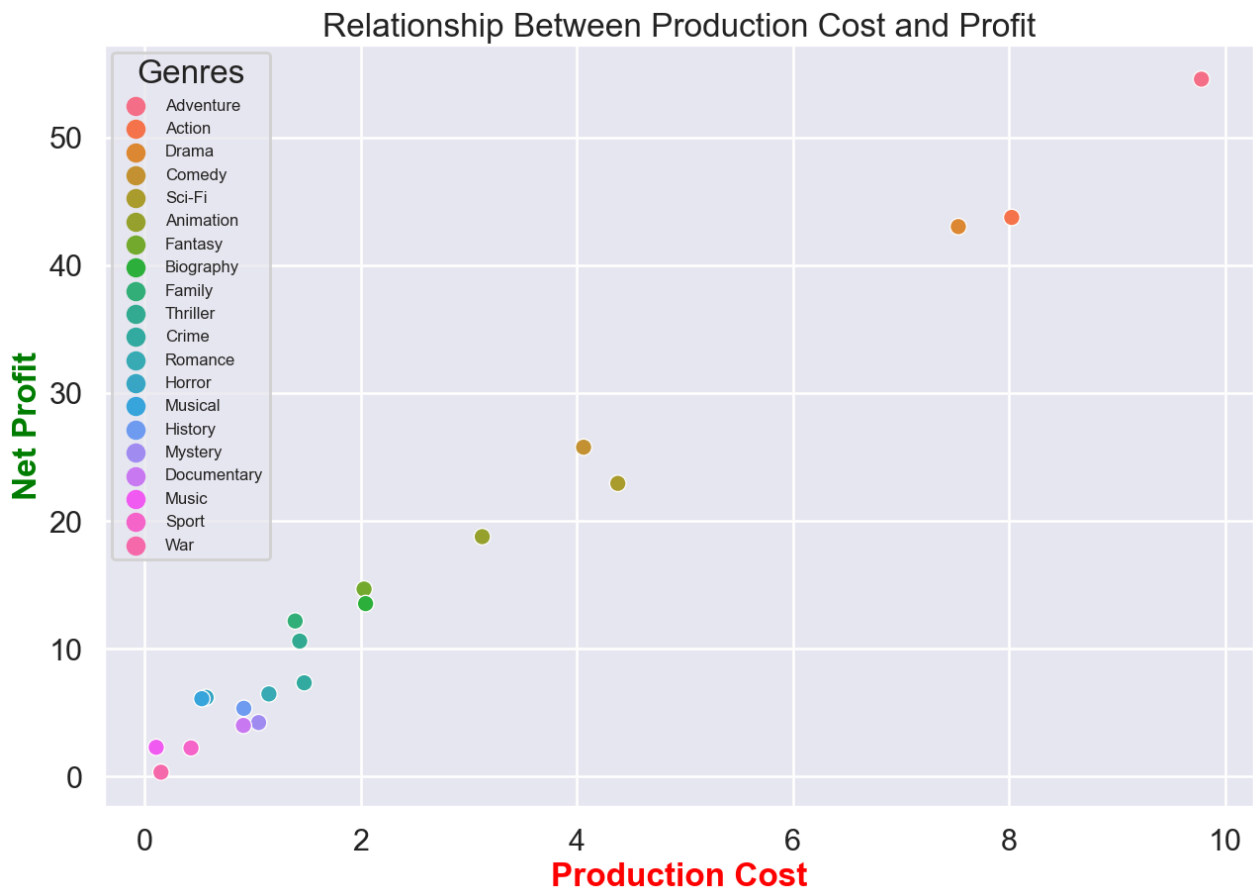
3. Total net profit per genre.



As we can see it from this analysis, the Adventure and Action genres are the top earning ones with 54.540b USD and 43.727b USD respectively. Drama on the other hand inspite of the number of movies, is only third when it comes to profitability with 43.010b USD.

Finally the relationship between the production budget and the profitability of the genres by running a scatter plot. This is the main plot of the series that will be shown in the chart below and will demonstrate how these variables are correlated to each other.

4. Relationship between production cost and profits.



As we see it in the scatter plot above, the production cost and profit are positively correlated. Which means that the most profitable genres require high production budget. On the other hand low budget movies aren't very profitable.

CONCLUSIONS

To tackle the problem presented by the Microsoft Movie Studio, I have used data collected from three of the global movies datasets namely; the IMDB, TMDB and TNDB and performed exploratory analysis. In order to come up with effective and helpful recommendations, it is crucial that the data collected be as latest as possible. Based on this understanding, this analysis covers movies produced between 2013 - 2019 and shows that:

- During this period 118 Drama movies were made.
- Sci-Fi, Adventure, Fantasy, Action and Animation were the top 5 most popular genres.
- Adventure, Action, Drama, Comedy, and SciFi were the top five most profitable genres
- The production cost is positively correlated to profit.

RECOMMENDATIONS

Based on the result of the analysis, I recommend the following:

- The Microsoft Movie Studio should focus on producing Adventure, Action, Drama, Comedy, and SciFi movies for high profit earning.

- The studio should also understand that the most profitable genres cost more to produce. With this in mind Microsoft should focus on allocating enough budget and producing the most popular movies.
- The studio should also consider producing high budget Animation movies as animation is one of the top 5 most popular genres.