

EDA FOR MICROSOFT MOVIE STUDI

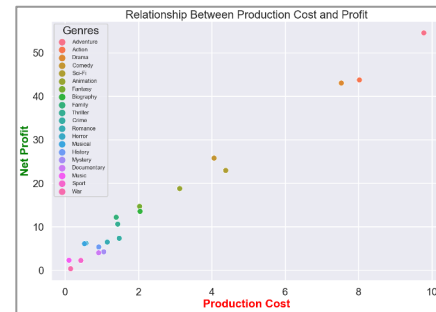
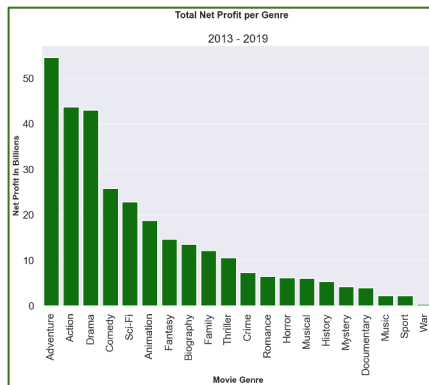
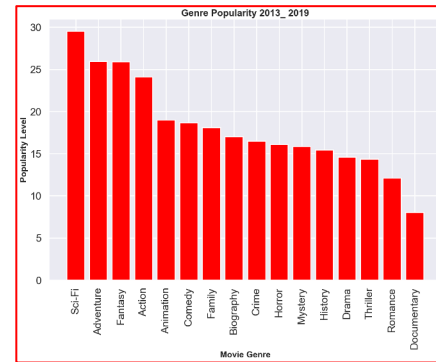
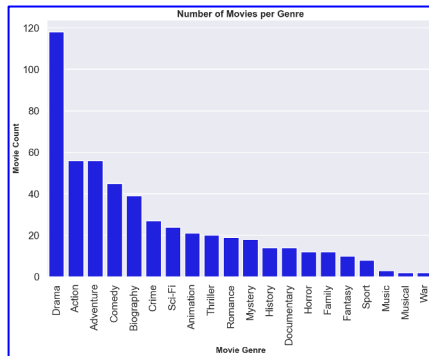
Data Analysis by: Shamla Araya

DSF-PT07 @ MORINGA

Summary

This project performed an exploratory data analysis to learn the current pattern of the movie industry aiming 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 is used for this analysis and covered movies produced from 2013-2019.

I have checked, cleaned, analyzed and visualized the data. It is my strong belief that Microsoft will find the outcome of this analysis and the recommendation very helpful in their decision making process.



Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

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?



Data

The data for this analysis is collected from different globally renowned sources. The source are:

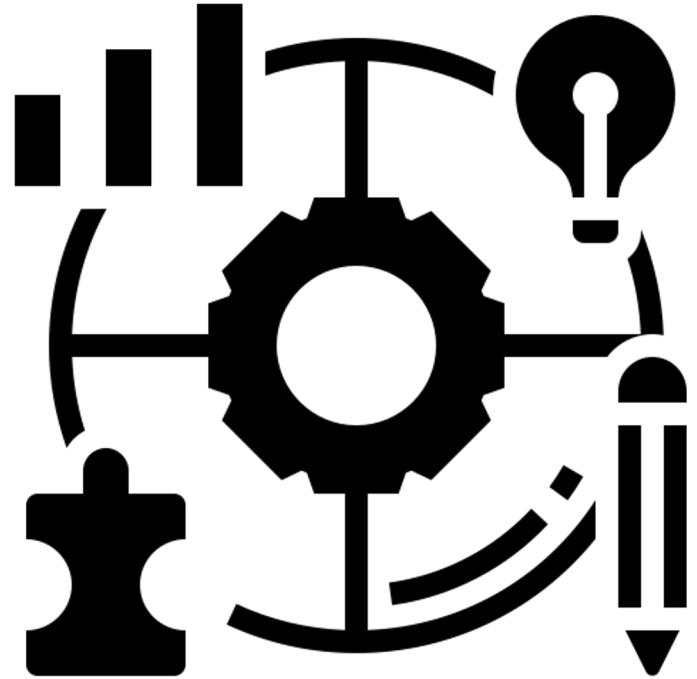
1. IMDB ['movie_ratings', 'movie_basics'](movie_id, primary_title, start_year, genres, averagerating)
2. Themoviedb (title, popularity, vote_average)
3. The Numbers (movie, domestic_gross, worldwide_gros, production_budget)



Methods

In doing this analysis I have followed a 5 step methodology.

1. I have first loaded the libraries that I used to read the datasets
2. The second method is to check and prepare the data. This involves cleaning, filtering, joining and merging different data frames.
3. The third step is analyzing the data which involves further filtering, running mathematical operations and arriving at data frames ready for visualization.
4. The fourth step is visualizing the analysis using bar plots and scatter plots.
5. The fifth step is a statement of the findings, recommendations, and final note.



Results

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 TNDDB 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.



Conclusions

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.

The movie business is a very competitive and ever evolving sector. Even though this analysis has yielded some good results, it still could be improved by using the latest datasets and also data that takes demographics of the audiences and the number of demands per movie in a given platform into the equation. I hope to be able to do such analysis for Microsoft with these data in the future.



Thank You!

Email: shamla.araya@student.moringaschoo.com

GitHub: @ShamlaA

LinkedIn: [linkedin.com/in/Shamla Tdese](https://www.linkedin.com/in/ShamlaTdese)