



Phase one project

By John Mark Ang'awa



Introduction

This project is conducted for Microsoft company to understand the movie industry that they want to venture into. The project involves collecting of data from various sources and analyzing them to come up with a solution to make them top in this market.

Overview

- Business understanding
- Data understanding
- Data preparation
- Data analysis
- Data visualization
- Recommendation
- Evaluation and future improvement ideas
- Conclusion

Business Understanding

Microsoft, a technology giant, has recognized the success of original video content produced by big companies and wants to venture into the movie industry by creating their own movie studio. However, they lack knowledge and insights about creating successful movies in this competitive environment.



Data Understanding

In analyzing the performance of box office movies, we will utilize multiple datasets from various sources including Box Office Mojo, IMDb, Rotten Tomatoes, TheMovieDB, and The Numbers. These datasets contain comprehensive information about movies, including box office revenue, ratings, release dates, production budgets and genres. By combining and analyzing these datasets, we can gain insights into the factors that contribute to the success of movies.

Goal of the project

To analyze the current box office trends and identify the types of films that are best performing. By doing so, we aim to provide actionable insights to the head of Microsoft's new movie studio, helping them make informed decisions on the types of films to create and maximize their chances of success.



Methods

The size of the datasets are wide allowing for a robust analysis and provides a comprehensive view of the movie landscape. We will perform a statistical data analysis using various data science analytical tools and come up with visualizations that can describe out data better to the stakeholders. This will help identify key trends and characteristics of movies successes.

Data preparation

For the data preparation, we collected data from vast movie sites including,

- Box Office Mojo
- IMDb, Rotten Tomatoes
- TheMovieDB
- The Numbers

This data hugesome amount of data was then converted to csv files and loaded various datasets for analysis purposes.

Data analysis process

Here we used various libraries for analysis purposes including pandas and matplotlib libraries which are key for a good analysis.

The analysis of this data was a quantitative analysis which explains data using figures and plots.

Then we visualized the data and came up with concrete conclusions.

The data analysis process had various steps;



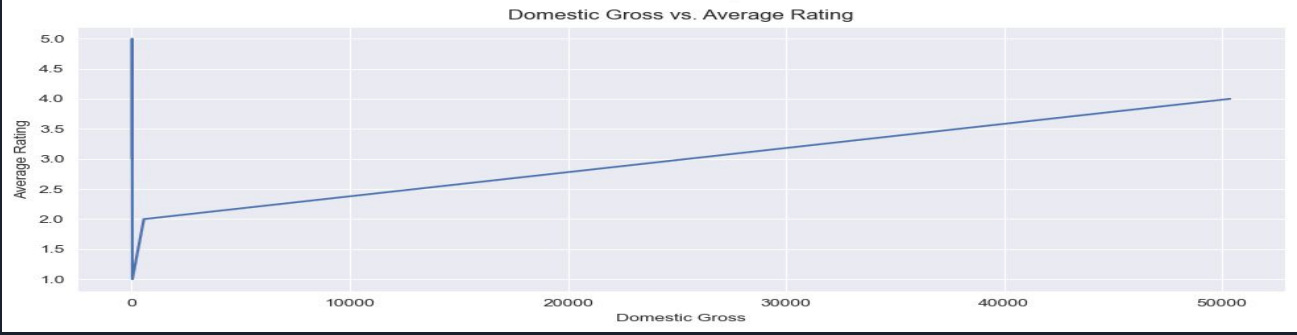
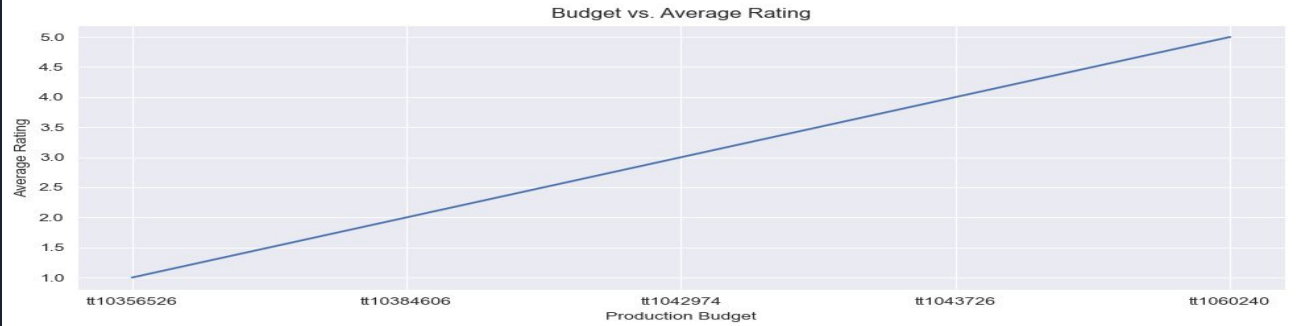
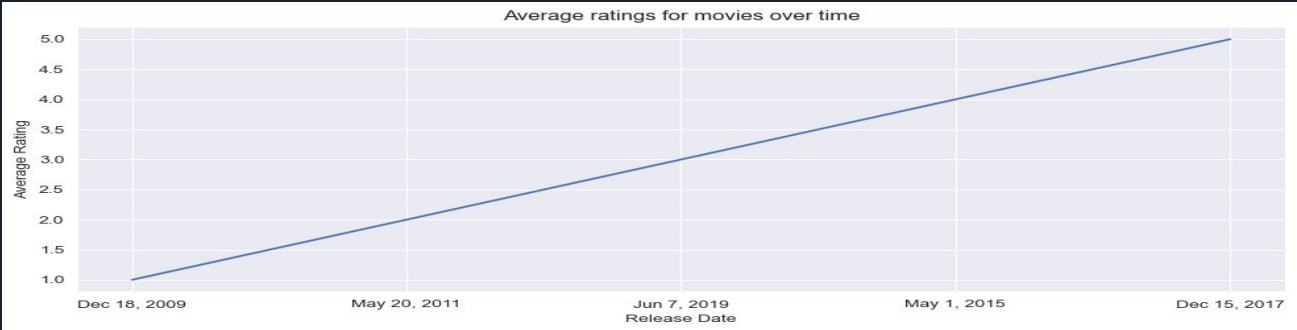
Steps in data analysis;

- ❖ Importing various libraries to use
- ❖ Loading of the datasets to be used
- ❖ Exploring the datasets by viewing them
- ❖ Data cleaning process which involves checking for duplicates, missing data and dropping them.

Data visualization

After analysing the data and ascertaining that we have the data that we have, we came up with three line graph visualizations that were able to predict movie trends through;

- 1) their average ratings overtime
- 2) budget averages against ratings
- 3) domestic gross against their ratings





Recommendations

1. From this analysis it is clear that the movie industry is very competitive.
2. This is because quality of movies go up by rating based on the the budget allocated for them.
3. The more the funds spent on creating the movies which increase the quality of the movies, the more their ratings. It is therefore prudent that the head of Microsoft's office should focus more budget allocation on production for best movie production to up with this market.



Evaluation

At the beginning of the project Microsoft did not have any insights or data for how to venture into the movie industry. I was able to explore, generate and analyse data that resulted in a way forward for Microsoft company.

It is important to note that the movie industry is very competitive and Microsoft should consider getting into it. This is because there is high profit generation for high quality movies produced.



Conclusion

The aim of this project is to provide valuable insights for Microsoft's new movie studio through exploring the box office performance of films. Through these comprehensive movie datasets analysis, we can identify the types of films that are currently doing well at the box office. The findings of this analysis will enable stakeholders to make informed decisions concerning the types of movies to create, the genres to focus on, and the investment strategies to take on. Through these insights, Microsoft can become successful in the movie industry and establish a strong presence as a creator of original video content.