MOVIE INDUSTRY RESEARCH

By Susan Nduta Kanyora

Overview

The project is based on the entertainment industry specifically movie production.The entertainment industry is fast evolving and it is only prudent for stakeholders such as production companies to keep up with the current trends.It has been seen in recent years that consumers are moving away from going to movie theatres to streaming movies.The creation of movies streaming sites as Netflix,Amazon Prime Video just to mention a few has made it possible to analyze the behaviour of consumers.The analysis helps production companies to make decisions based on which movie genres are popular,average ratings and how many minutes an average consumer spends on the streaming platforms.

Business Problem

The business problem being addressed is to provide insights into the success and popularity of movies, which can help production companies make informed decisions regarding the type of movies to produce and invest in.

The data questions that we plan to answer to solve this problem are:

1. What are the highest-rated movies of all time?
2. Which genres have the highest average rating?
3. What is the relationship between a movie's budget and its box office revenue?
4. Which directors have the highest average rating for their movies?
5. What is the trend in movie ratings and box office revenue over time?

Overall, answering these data questions can help production companies make data-driven decisions that lead to more successful and profitable movies.

Data Understanding

The data being used for this project comes from the zipped IMDb database.The data analysis questions are related to movie ratings, genres, budget, revenue, directors, and time, and the database contains information on all of these variables. The dataset includes a range of information related to movies, including movie title, start year, genres, average ratings and runtimes in minutes.

The data represent a sample of movies that have been released in recent years. The sample includes movies from various countries and in various languages, although there may be some bias towards English-language films. The variables included in the dataset are a mix of categorical and numerical variables, such as movie title (categorical), start year (numerical), and runtimes (numerical).

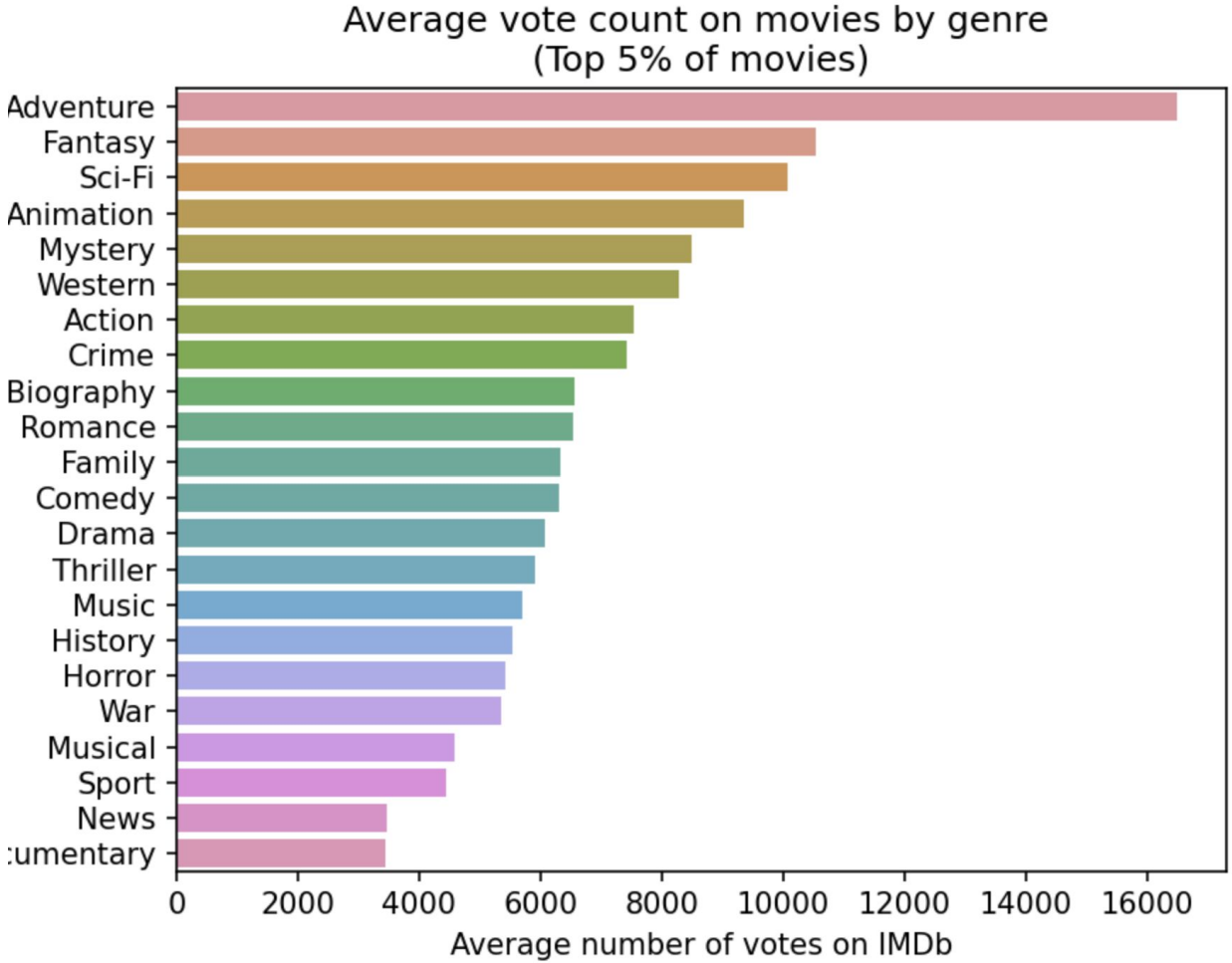
The target variable for this project will depend on the specific data analysis question being asked. For example, the highest-rated movies of all time question will use movie ratings as the target variable, while the relationship between a movie's genre and its average ratings question will use runtimes as the target variable.The range and distribution of

To measure the success of films,we look at the average ratings of specific genres and the number of votes the movies have received.

The Analysis

The graph below shows the top five movie genres in terms of average number of votes on IMDb.They include:

1. Adventure
2. Fantasy
3. Science Fiction(Sci-Fi)
4. Animation
5. Mystery


Average vote count on movies by genre
(Top 5% of movies)
Adventure
Fantasy
Sci-Fi
Animation
Mystery
Western
Action
Crime
Biography
Romance
Family
Comedy
Drama
Thriller
Music
History
Horror
War
Musical
Sport
News
:umentary
0
2000
4000
6000
8000
10000
12000
14000
16000
Average number of votes on IMDb

Recommendation

Given the output from the analysis above,the following are the proposed recommendations:

Movie production companies should concentrate on producing movies that consist of the top genres.The top genres include:Adventure,Fantasy,Sci-Fi,Animation and Mystery.

However,the production should also pay attention to other genres.This avoids customer migration to other movie streaming sites.