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 goal is to provide insights into the popularity of movies, helping production companies make informed decisions on which types of movies to invest in. The data questions focus on identifying the highest-rated movies and genres, understanding the relationship between budget and revenue, identifying top directors, and tracking trends in ratings and revenue over time. Answering these questions can help production companies allocate resources effectively and make data-driven decisions for success.

Data Understanding

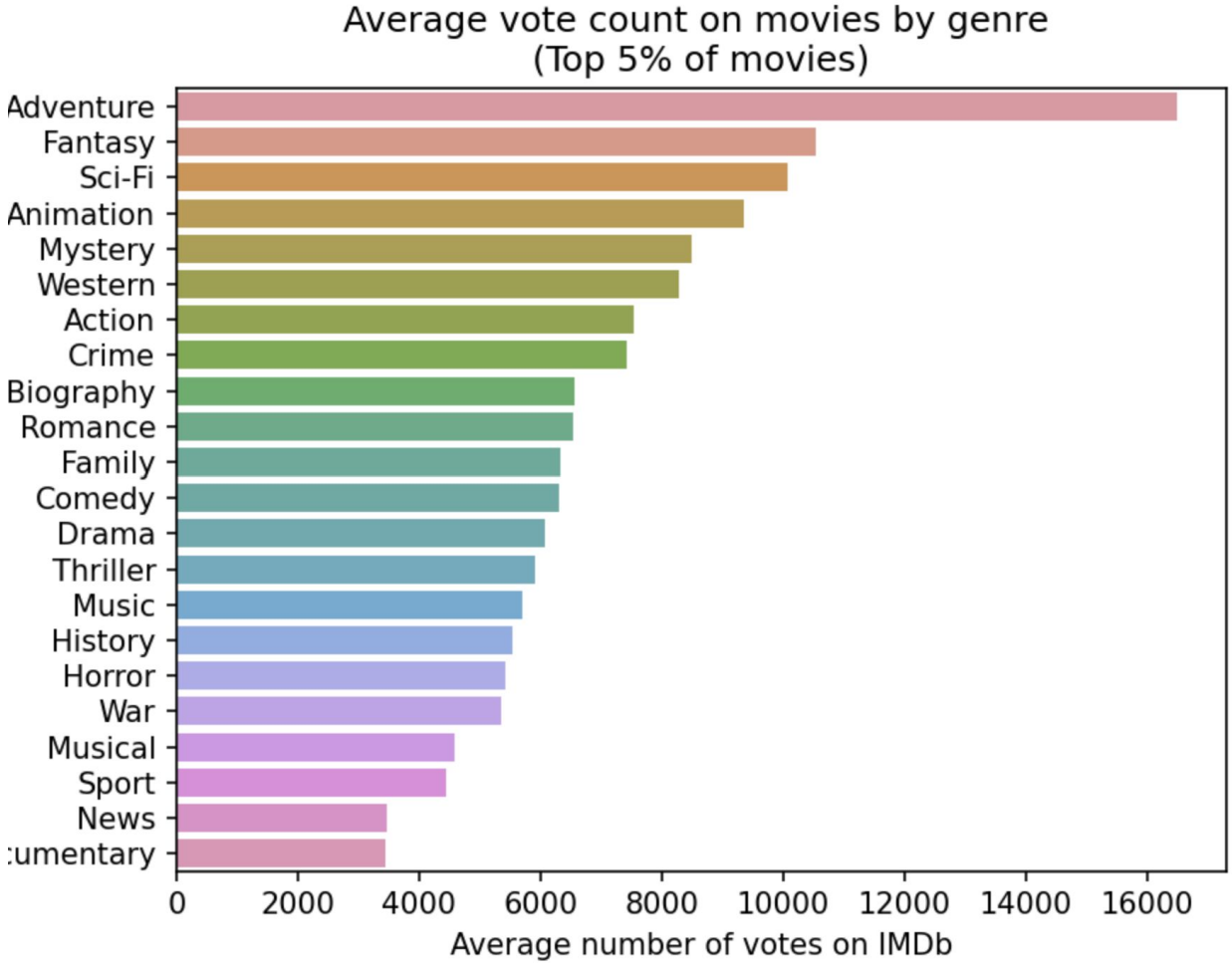
The project uses data from the zipped IM.db database to answer questions about movie ratings, genres, budget, revenue, directors, and time. The dataset includes a mix of categorical and numerical variables, representing a sample of movies from different countries and languages. The target variable depends on the specific question being asked, such as movie ratings for the highest-rated movies question and runtimes for the genre and average ratings question. The range and distribution of values for each variable will also be considered during the data analysis.

Data Preparation

Data preparation involves cleaning and transforming the data to make it suitable for analysis, integrating data from multiple sources, and handling missing values or outliers. This ensures the accuracy and quality of the analysis. By dropping irrelevant variables and creating new ones, we focus on the important variables. Handling missing values and outliers reduces bias and improves reliability. Integrating data from multiple sources creates a more comprehensive dataset for a complete view of the problem.

Measure of success.

The preferred method of measuring a movie's success is through the number of votes, which is more reliable than box office returns in the era of streaming. Research shows that there is a positive correlation between the number of votes and a movie's average rating, making it a good measure of success. The repository includes a list of contents, and instructions for using them are provided in the next section.


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.