IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023

Internet Movie Database Data Analysis Using Python

Siddhesh More, Atharva Jadhav, Vedashree Bhat
Department of Computer Engineering
All India Shri Shivaji Memorial Society's College of Engineering, Pune, India

Abstract: *The film industry is highly competitive and dynamic, with thousands of movies released every year. To succeed in this industry, it's crucial for filmmakers, studios, and investors to make informed decisions. IMDB, a popular source for movie-related data, provides a wealth of information on movies, including details about genres, budgets, revenues, critical reception, and more. The problem statement aims to leverage this data to gain valuable insights and inform key decisions within the film industry.*

Keywords: Data Analysis

I. INTRODUCTION

Understanding and predicting the aspects that lead to a film's success is critical in the ever-changing world of the film industry. The Internet Movie Database (IMDb) has developed as a goldmine of information for both moviegoers and industry experts, providing a massive reservoir of movie data such as user ratings, reviews, cast and crew information, box office statistics, and more. The IMDB Movies Data Science Project aims to uncover the secrets behind the films that fascinate fans and do extremely well on this significant platform by utilizing this plethora of data.

IMDb, which is owned by Amazon, has become the go-to site for movie- related information, allowing users to explore and assess films in a variety of ways. IMDb contains everything from classic masterpieces to modern blockbusters.

IMDb has played an important part in determining how we perceive and engage with the world of movies, from classic masterpieces to contemporary blockbusters. Understanding the complicated interplay of variables that contribute to a film's success and critical acclaim on this platform is a task that data science can meet front on.The IMDB Movies Data Science Project has a number of goals. It seeks to deconstruct the factors that distinguish a film, whether through high IMDb user ratings, box office success, or both. We hope to identify hidden patterns and insights that drive a movie's IMDb rating and financial success by evaluating a wide range of movie attributes such as genre, director, actors, budget, and release date.

One of the main goals of this research is to create predictive models that can accurately anticipate a movie's IMDb rating and box office performance. These models will be useful tools for filmmakers, studios, and producers, allowing them to make educated decisions and investments in their projects.

The IMDB Movies Data Science Project will also investigate the temporal component of movie data, digging into long-term trends and patterns. We hope to uncover the temporal dynamics that influence movie popularity on IMDb by researching the impact of seasons, holidays, and other cyclical influences.

Furthermore, the research will investigate the film industry's significant figures - directors, actors, and genres - and examine their constant contributions to high-rated and genre films.

It will also look into any geographical or cultural differences in movie choices and IMDb ratings, giving a worldwide view on the art of filmmaking.

Finally, the findings and recommendations of the IMDB Movies Data Science Project will be extremely valuable to a wide range of audiences, including industry professionals looking to optimize their decision-making processes, filmmakers looking to create content that resonates with their target audience, and movie enthusiasts curious about the underlying dynamics of the film industry. This project is a journey into the heart of cinematic data, providing insights that can affect the future of filmmaking and our enjoyment of the art form itself.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023

II. OBJECTIVE

The IMDB Movies Data Science project has many goals, all of which aim to acquire insight into the aspects that determine a movie's success and popularity on the Internet Movie Database (IMDb). Among these goals are:

1. Understanding the Factors That Influence Film Success: Analyze and determine the important elements and characteristics that lead to the success of a film, such as high IMDb user ratings, box office performance, and critical praise.
2. Investigate how different movie variables, such as genre, director, actors, budget, and release date, affect IMDb ratings and financial performance.
3. Predictive Modeling: Create accurate predictive models based on movie attributes that can forecast IMDb ratings and box office results, allowing filmmakers and studios to make informed decisions

III. USED TECHNOLOGIES

Programming Languages:

- Python: Widely used for machine learning and computer vision tasks due to its extensive libraries like OpenCV and TensorFlow.
- Libraries and Frameworks:
- The provided code appears to be a Python script for data analysis and visualization. Let's break down the technologies and libraries used in this code:
- numpy (np): NumPy is a fundamental package for scientific computing in Python. It provides support for working with large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
- pandas (pd): Pandas is a popular data manipulation and analysis library in Python. It provides data structures like DataFrames for handling and analyzing structured data.
- seaborn (sns): Seaborn is a data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

IV. RESULT AND DISCUSSIONS

```
In [17]:
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")

In [20]:
data = pd.read_csv('final.csv')

In [21]:
pd.set_option('display.max_columns',None)

data.head()
```

Out[21]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	7
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	3
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	2
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	4
4	NaN	Doug Walker	NaN	NaN	131.0	NaN	Rob Walker	131.0	

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023

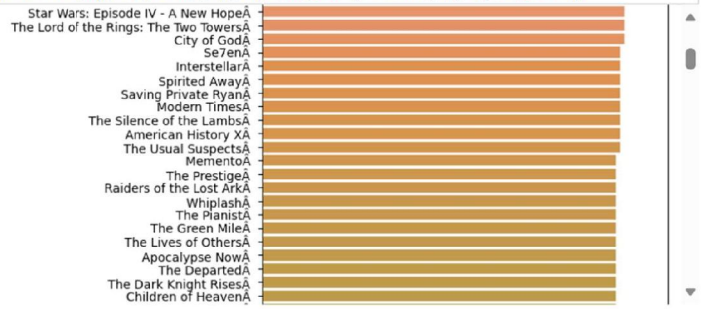
```
In [58]:
plt.figure(figsize = (6,50))

sns.barplot(data = non_english, y = non_english['movie_title'], x = non_english['imdb_score'])

Out[58]: <AxesSubplot:xlabel='imdb_score', ylabel='movie_title'>
```

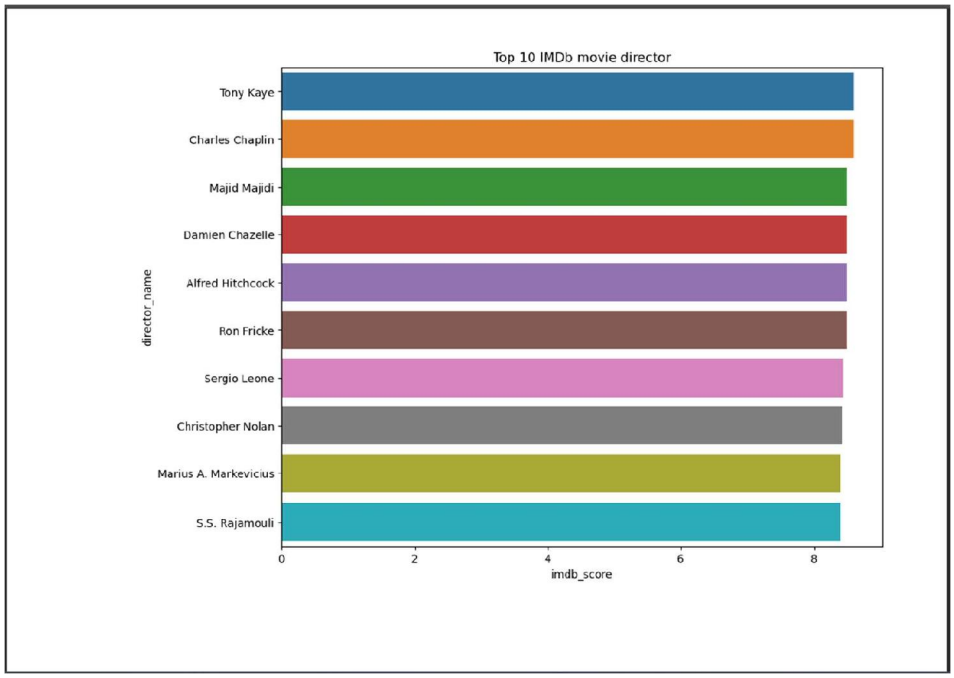

```
### Now for the best directors, we will br grouping some coloums
```

```
In [55]:
plt.figure(figsize = (6,50))
sns.barplot(data = IMDb_Top_250, y = IMDb_Top_250['movie_title'], x = IMDb_Top_250['imdb_score'])
```



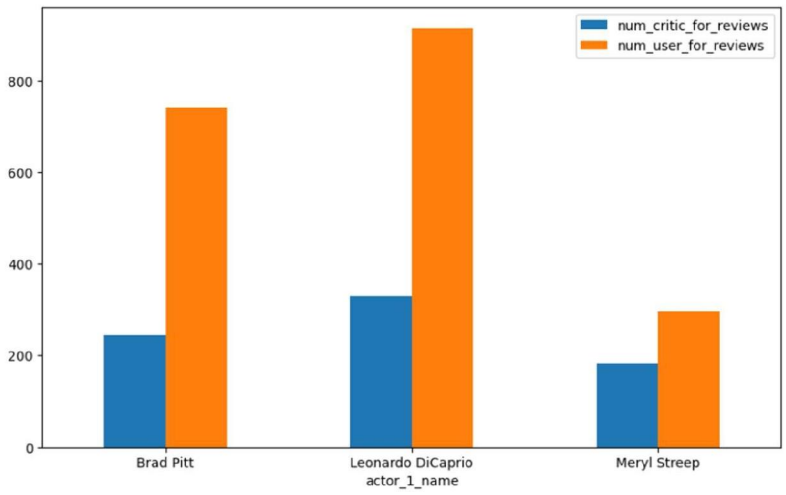

```
## Listing with language which is not english
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023



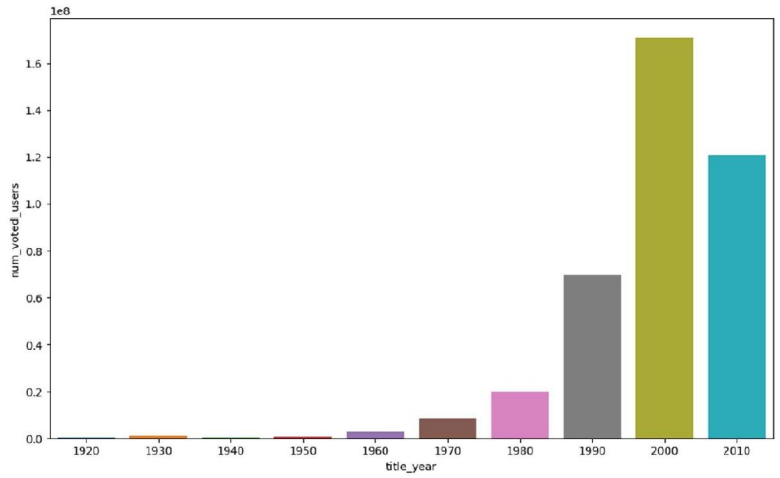

```
In [78]: 1 combined.groupby(['actor_1_name'])['num_critic_for_reviews','num_user_for_reviews'].mean().plot(kind = 'bar',
         2                                                                                              figsize = (10,6))
         3 plt.xticks(rotation = 360)

Out[78]: (array([0, 1, 2]),
          [Text(0, 0, 'Brad Pitt'),
           Text(1, 0, 'Leonardo DiCaprio'),
           Text(2, 0, 'Meryl Streep')])
```



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023

```
In [82]:  1 plt.figure(figsize = (12,7))
          2 sns.barplot(data = df_num_voted, x = df_num_voted['title_year'], y = df_num_voted['num_voted_users'])
          3
Out[82]: <AxesSubplot:xlabel='title_year', ylabel='num_voted_users'>
```



V. APPLICATIONS

Analyzing IMDB movie data has various practical applications in the film industry and beyond. Here are some examples of how the insights derived from IMDB movie data analysis can be applied:

1. Film Production and Investment:

- Genre Selection: Movie studios can use genre popularity trends to make informed decisions about which types of movies to produce.
- Budget Allocation: Analysis of budget vs. revenue can guide studios in allocating resources more effectively, optimizing production costs, and maximizing profits.
- Casting Decision: Knowledge of the impact of directors and actors on a movie's performance can help in casting decisions, potentially attracting bigger audiences and higher revenue.

2. Marketing and Release Strategies

- Release Timing: Information about the best time to release a movie can help studios strategically plan their marketing and distribution efforts to maximize box office success.
- Competition Analysis: Understanding the competitive landscape and market conditions can aid in positioning a movie effectively against other releases.

3. Content Quality Improvement:

- Critical Reception: Data on movie ratings can be used to assess the quality of movies, guiding filmmakers and studios to focus on producing content that resonates with audiences and critics.

4. Viewer Insight:

- Audience Preferences: Data analysis can reveal audience preferences for different types of movies, aiding streaming platforms and theaters in curating content to attract viewers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, December 2023

5. Academic Research

- IMDB data analysis can be valuable for academic research in fields like sociology, cultural studies, and media studies to understand societal and cultural trends through film.

6. Recommendation Systems:

- Streaming platforms can use movie data analysis to improve their recommendation systems, suggesting movies to viewers based on their preferences and past viewing habits.

7. Investment Decision

- Investors in the film industry can use data analysis to make more informed decisions about where to allocate their funds, potentially leading to higher returns on their investments.

8. Market Insights

- Film industry reports and market research agencies can use this data to provide insights to stakeholders, helping them understand the current state of the industry and make data-driven decisions.

9. Content Licensing

- Content acquisition teams can use data analysis to identify movies that may be valuable for licensing or distribution in different markets.

10. Predictive Analytic

- Predictive models can be built based on historical movie data to forecast box office performance, revenue, and other critical metrics for upcoming films.

In summary, IMDB movie data analysis can have a broad range of applications, from improving decision-making in film production and distribution to enhancing the viewing experience for audiences and contributing to academic research and market insights. It plays a significant role in the film industry's efforts to produce successful and profitable movies.

VI. CONCLUSION

The IMDB Movies dataset data science analysis has produced useful insights into the world of movies, their qualities, and their impact on the business. Several major insights have emerged from this analysis:

1. Genre Trends: According to the statistics, certain genres are more popular than others, with drama, comedy, and action being among the most often produced genres. This information can help filmmakers and studios decide on future projects.
2. Budget vs. Revenue: The analysis revealed a favorable relationship between a film's budget and its revenue. While this may come as no surprise, the data can assist studios in making educated judgments about their film production investments.
3. Critical Reception: The statistics also emphasized the importance of critical reception, since films with higher ratings do better at the box office. This emphasizes the importance of creating high-quality material.
4. Director and Actor Impact: According to the dataset, some filmmakers and performers consistently add to a film's success. This insight can help to steer casting decisions and industry cooperation.
5. Release Dates: The study found that the date of a film's release can have a substantial impact on its success. Seasonal trends, competition, and market conditions all influence how well a film performs.

Finally, the data science research of the IMDB Movies dataset gave useful insights for the film industry. This information can be used by filmmakers, studios, and investors.