**The Rise of Indian Cinema in the New Millennium (2000-2021)**



***Submitted by: -***

|  |  |
| --- | --- |
| **Name of Students** | **Examination Roll No.** |
| **Arin Ganguly** | **11600222029** |
| **Rishabh Yadav** | **11600222081** |
| **Subhadeep Pan** | **11600222121** |
| **Sudip Kundu** | **11600222124** |
| **Souvik Ghosh** | **11600222117** |
| **Akash Polley** | **11600222014** |
| **Jai Prakash Rai** | **11600222055** |
| **Soham Hazra** | **11600222107** |
| **MD Kaif** | **11600222060** |
| **Dattatreya Saha** | **11600222046** |
| **Abhijit Dey** | **11600222010** |
| **Antara Pal** | **11600222021** |
| **Bikram Majumdar** | **11600222044** |
| **Ankan Das** | **11600222017** |
| **Ayush Saha** | **11600222041** |

**Under the supervision of**

**Mr. Sourav Ghorai**

Assistant Professor, Dept. of Basic Science

**Department of Computer Science & Engineering,**

**MCKV Institute of Engineering**

**243, G.T. Road(N)**

**Liluah, Howrah – 711204**

**May 2024**

**CONTENTS**

* Introduction
* Literature review
* Abstract
* Methodology
* Solution
* Discussion
* Reference

**INTRODUCTION**

Cinema of India has influenced world cinema since second half of 20th century The Cinema of India, also known by its sobriquet Bollywood, and known for winning national and international awards, comprises the films produced across India. Indian cinema is a part of Indian art and culture, and encompasses films produced across India, Hindi cinema, born in Mumbai is one of the biggest film industries in the World, as well as multiple regional film industries.

Indian Cinema has ensured that diversified genres of storytelling were given due weight in its share dictionaries. This is especially underlined in the way in which mediums span a range of genres. A genre in film is a category or classification of a narrative film, based on its narrative elements. This, in turn, shapes the assumptions and consumption habits of the audience. Similar narratives, aesthetic style, and thematic concerns that genres have in common. Each genre carries its own set of expectations.

The kind of horror movie a viewer is about to watch will be filled with suspense and fear while a Romantic Comedy will be full of laughter and a happy ending. The predictability of these conventions puts the audience at ease, providing them with a level of comfortably which better allows them to curate content for whether or not they feel like laughing, crying, or feeling engaged.

Be it the silent era, through mythological to historical subjects, most of the films focused on such genres. But sound opened up new worlds in narrative, which gradually incorporated elements of romance, drama, and social commentary. This time also saw the rise of the 'Masala' genre of Bollywood film with as a special Indian blend of action, romance, comedy, and melodrama skillfully sewn into one film.

That he relates to — distributors and sales agents, these are things that they know. They also frequently brand by genre so there is a high degree of audience recognition from them. Understanding genre trends is important for filmmakers, producers and marketers. The findings can help us make informed decisions about future projects and marketing strategies. Furthermore, as audience preferences continue to evolve, genres that focus on niche areas and specific subcultures will gain popularity. These ‘micro-genres’ will cater to specific audience interests, providing more personalized and relatable content.

Indian cinema’s themes and stories are rooted in Indian culture and society, making it an authentic representation of Indian life and values. The future of genre in Indian films looks promising and expensive. Since our entertainment industry is evolving dynamically, formula content will go for a toss. This will push the boundaries of storytelling, as creators experiment with various combinations of genres to cater to a diverse audience. Genre trends in films will continue to evolve, offering a more diverse and vibrant landscape of storytelling.

**LITERATURE REVIEW**

Our topic is analysis of the trend of last 20 years of Bollywood movie genres preferred by people. And to predict the most and least preferable genres in the next 4 years.

Movie reviews do have an effect on the consumer thought process, which consequently leads to the decision on whether or not to purchase a ticket to see a movie. This effect is also greatly varied by the sources as well as the type of information contained by the review.

Movie preferences can vary widely among different groups of people, so it's challenging to definitively say which genre of movies is the most liked overall. However, some genres tend to be more popular than others based on box office performance, average rating, and audience reception.

Conducting a detailed year wise literature review of Bollywood movies from 2000 to 2026 is a comprehensive task. Below is a high-level overview highlighting key trends, notable films, and significant changes in the Bollywood film industry during this period.

**2000-2009: (Rise of Multiplex Culture and Globalization)**

Key Trends:

1. The rise of multiplexes transformed the business model of Bollywood.
2. Increased collaboration with international talent and settings, reflecting globalization.
3. A shift towards more urban-centric stories and complex narratives.

Notable Films:

* 2000: Kaho Naa... Pyaar Hai - Debut of Hrithik Roshan, Romance , action movie.
* 2001: Lagaan , Dil Chahta Hai - Modern storytelling and new wave of Bollywood.
* 2002: Devdas - Grand production values.
* 2003: Kal Ho Naa Ho, Munna Bhai M.B.B.S- Blend of comedy and drama.
* 2004: Swades , Dhoom - Beginning of franchise culture.
* 2005: Black, Bunty Aur Babli - Diverse genres.
* 2006: Rang De Basanti, Lage Raho Munna Bhai - Youth-centric and socially relevant films.
* 2007: Chak De! India, Taare Zameen Par - Inspirational and educational narratives.
* 2008: Ghajini, Jodhaa Akbar - Action-packed and historical epics.
* 2009: 3 Idiots - Critique of the education system.

On basis of Box office collection:-

1. It was found that the genre preference prediction on the basis of Box office collection Most favourable was Adventure movies.
2. We found that the genre preference prediction on basis of Box office collection least favourable was war movies.

On basis of Average rating :-

1. Genre preference on the basis of Box office collection, were History and war movies.
2. The genre preference prediction on basis of Box office collection, the least favourable was Thriller movies.

**2010-2019: Experimentation and Content-Driven Cinema:-**

Key Trends:

1. Increasing emphasis on content-driven cinema.
2. Experimentation with genres and non-traditional storytelling.
3. Growth of independent films and rise of digital streaming platforms.

**Notable Films:**

* 2010: Dabangg, My Name is Khan (Karan Johar) - Strong character-driven narratives.
* 2011: Zindagi Na Milegi Dobara, The Dirty Picture - Modern and biographical themes.
* 2012: Barfi! , Kahaani - Innovative storytelling.
* 2013: Bhaag Milkha Bhaag, Lunchbox - Biopics and heartfelt dramas.
* 2014: PK, Queen - Satirical and feminist narratives.
* 2015: Bajrangi Bhaijaan ,Piku - Cross-border themes and family-centric stories.
* 2016: Dangal , Neerja - Sports biopics and heroic tales.
* 2017: Baahubali 2: The Conclusion, Hindi Medium - Epic sagas and social commentaries.
* 2018: Padmaavat , Stree - Historical epics and horror-comedy.
* 2019: Gully Boy, Article 15 - Stories from the streets and socially charged narratives.

On basis of Box office collection:-

1. Found that the genre preference prediction on basis of Box office collection Most favourable was Epic movies.
2. Found that the genre preference prediction on basis of Box office collection least favourable was Mystery movies.

On basis of Average rating :-

1. We found that the most preferred genre on the basis of Box office collection was Epic movies.
2. We found that the most preferred genre on the basis of Box office collection was Dance movies.

**FROM(2000-2021):-**

On basis of Box office collection:-

1. We can observe that the genre preference prediction on basis of Box office collection , Most favourable is Epic movies.
2. The genre preference prediction on basis of Box office collection for next 4 years , least favourable genres are Mystery, war, Family &Fantasy movies.

On basis of Average rating :-

1. The analysis reveals that the genre preference prediction on the basis of Box office collection for next 4 years, is Epic movies genre.
2. Also, the on the basis of Box office collection for next 4 years , least favourable genre is dance movies.

2020-2026: Digital Transformation and Pandemic Impact

Key Trends:-

1. Rapid adoption of digital platforms for film releases due to the COVID-19 pandemic. Increased focus on web series and OTT content.
2. Continued experimentation with genres and formats.

**Notable Films:**

* 2020: Dil Bechara, Gunjan Saxena: The Kargil Girl - Direct-to-digital releases.
* 2021: Shershaah, Mimi - Biopics and strong female-led narratives.
* 2022: Gangubai Kathiawadi, K.G.F: Chapter 2 - Biographical drama and action blockbusters.
* 2023: Pathaan, Rocky Aur Rani Kii Prem Kahani - High-profile star vehicles.

**Prediction for 2024-2026:**

Predicted trends include further diversification of content, increased international collaborations, and more films tackling social issues.

The Bollywood film industry has seen significant evolution from 2000 to 2026, marked by globalization, the rise of multiplexes, the digital revolution, and the impact of the COVID-19 pandemic. The shift towards content-driven cinema , experimentation with new genres, and the growing influence of digital platforms are notable trends that have shaped Bollywood during these year.

Prediction on the basis of Box office collection:-

* It was found that the genre preference prediction on the basis of Box office collection for the most preferred in the following 4 years is ***Action movies***.
* It is clear that the least preferred genre on the basis of Box office collection for next 4 years is ***Family movies***.

Prediction on the basis of Average rating :-

* The genre preference prediction is ***Thriller movies*** on the basis of Box office collection for next 4 years
* We found that the genre preference prediction on basis of Box office collection for next 4 years , least favourable is ***Family movies***.

**ABSTRACT**

In the rapidly evolving film industry, understanding past trends and forecasting future performance is crucial for stakeholders ranging from producers to marketers. Our 2nd Year B.Tech project, focuses on analyzing movie data from 2000 to 2021 to discern patterns in box office collections and average ratings across various genres. By leveraging a robust dataset, we meticulously examine the performance metrics of different genres, identifying key factors that contribute to their success.

Our analysis employs several python packages-- pandas, Matplotlib are used to draw

* Average Box Office Collection for Each Genre from 2000 to 2021
* Average Rating for Each Genre From 2000 to 2021

Building on these insights, we developed a predictive model to forecast box office collections and average ratings for upcoming movies from 2022 to 2025. This model integrates historical data trends and external factors, providing a forward-looking perspective on the film industry's trajectory.

We have used Statsmodels.tsa.holtwinters.ExponentialSmoothing(a Python Package) for Future Forecasting Like—

* Trend Line for Box Office Collection (2000-2025)
* Trend Line for Average Rating (2000-2025)
* Genre Preference Prediction(Box Office Collection) for Next 4 years(2022-2025)
* Genre Preference Prediction (Average Rating) for the next 4 years(2022-2025)

The outcomes of this project aim to offer valuable predictions for industry professionals, enabling data-driven decision-making in movie production, distribution, and marketing strategies. By presenting a comprehensive analysis and reliable predictions, our project contributes to the academic and practical understanding of movie analytics, setting a foundation for future research and innovation in the field.

**METHODOLOGY**

The Literature Review was analyzed using trend lines based on the Box office collection and People’s review (given points).

To implement the data set to translate into trend lines, several python packages were used, here are the packages and the specific methods along with their purposes, used in our analysis of the data sets:

1. **Pandas**:

***“import pandas as pd”***

* **Purpose**: Data manipulation and analysis.
* **Usage**:
* Data-sets, including The names, genres, rating & Box office Collection were extracted from a CSV file using pandas.
* Filtering and aggregation of data was done based on genres and years.
* Calculation of average ratings and box office collections also done using pandas.

1. **matplotlib.pyplot:**

“***import matplotlib.pyplot as plt”***

* **Purpose:** Data visualization.
* **Usage:**
* Bar charts for average ratings and box office collections were plotted using matplotlib.
* Forecast and historical trend lines for ratings and box office collections.

1. **Statsmodels.tsa.holtwinters.ExponentialSmoothing:**

***“from statsmodels.tsa.holtwinters import ExponentialSmoothing”***

* **Purpose:** Time series forecasting.
* **Usage:**
* The Holt-Winters method was used for forecasting average ratings and box office collections for each genre.

**Explanation of the code with the help of packages used:**

1. **Reading the Data**:

* **pandas** is used to read the CSV file containing movie data.

***data = pd.read\_csv('movie\_data.csv')***

1. **Calculating Average Ratings and Box Office Collections**:

* **pandas** is used to filter the data based on genres and calculate the mean ratings and box office collections.

***genre\_data = data[data['Genre'].str.contains(genre, case=False)]***

***genre\_ratings[genre] = genre\_data['Rating'].mean()***

***genre\_boxoffice[genre] = genre\_data['BoxOfficeCollection'].mean()***

1. **Plotting Average Ratings and Box Office Collections**:

* **matplotlib.pyplot** is used to create bar charts for average ratings and box office collections for each genre.

***plt.bar(genre\_ratings.keys(), genre\_ratings.values(), color='skyblue')***

***plt.bar(genre\_boxoffice.keys(), genre\_boxoffice.values(), color='lightgreen')***

1. **Filtering Data by Year**:

* **pandas** is used to convert the 'Year' column to datetime and filter the data within a specific date range.

***data['Year'] = pd.to\_datetime(data['Year'], format='%Y')***

***data = data[(data['Year'].dt.year >= 2000) & (data['Year'].dt.year <= 2021)]***

1. **Calculating Yearly Averages**:

* **pandas** is used to group the data by year and calculate the average ratings and box office collections per year for each genre.

***yearly\_ratings = genre\_data.groupby(genre\_data['Year'].dt.year)['Rating'].mean()***

***yearly\_boxoffice =***

***genre\_data.groupby(genre\_data['Year'].dt.year)['BoxOfficeCollection'].mean()***

1. **Forecasting Future Data:**

* **statsmodels.tsa.holtwinters.ExponentialSmoothing** is used to apply the Holt-Winters method to forecast the average ratings and box office collections for each genre from 2022 to 2025.

***model\_ratings = ExponentialSmoothing(ratings, trend='add', seasonal=None)***

***fit\_ratings = model\_ratings.fit()***

***forecast\_ratings[genre] = fit\_ratings.forecast(steps=4)***

1. **Plotting Forecasted Data**:

* **matplotlib.pyplot** is used to plot the forecasted average ratings and box office collections, along with historical data.

***plt.plot(forecast.index, forecast, label=genre, marker='o')***

***plt.plot(yearly\_ratings.index, yearly\_ratings.values, label=genre)***

1. **Determining the Most Preferred Genre**:

* **pandas** is used to identify the genre with the highest predicted average rating and box office collection in the last forecasted year.

***most\_preferred\_genre\_rating = max(forecast\_ratings, key=lambda x: forecast\_ratings[x].iloc[-1])***

***most\_preferred\_genre\_boxoffice = max(forecast\_boxoffice, key=lambda x: forecast\_boxoffice[x].iloc[-1])***

By using these packages and methods, the code reads and processes movie data, calculates historical averages, forecasts future trends, and visualizes the results to identify genre preferences.



**SOLUTION**



**Overview of the solution:**  
The following code analyzes movie data by genre across various years. Here's a brief tour:

***Data Loading and Preprocessing:***

1. Libraries like pandas and matplotlib are used for data manipulation and plotting.
2. The code reads data from a CSV file named "movie\_data.csv".
3. It defines a list of genres to be analyzed.
4. The Year column is converted to datetime format.

***Genre-wise Analysis by Period:***

1. It calculates average rating, box office collection, and movie count for each genre in two periods (2000-2010 & 2011-2021).
2. Separate bar charts are plotted to visualize these metrics for each period.
3. Overall Genre-wise Analysis:
4. Similar calculations are done for the entire data (2000-2021).
5. Bar charts are again plotted for overall average ratings and box office collections.

***Genre-wise Prediction (2022-2025):***

1. The code filters data for movies between 2000 and 2021.
2. It creates dictionaries to store average ratings and box office collections per year for each genre.
3. A forecasting method (Holt-Winters) is used to predict these values for the next 4 years (2022-2025).
4. Line charts are plotted to show the predicted average ratings and box office collections for each genre.
5. Trend Line Analysis (2000-2025):
6. Line graphs are created to visualize the trend of average rating and box office collection across all genres from 2000 to 2025.
7. Forecasted values for the next 4 years are included with a dashed line.
8. Identifying Most Preferred Genres:
9. The code finds the genre with the highest predicted average rating and box office collection in the year 2025 (last year of forecast).
10. It then prints these genres as the most preferred ones based on the respective metrics.

***Data set:***

The data set used in this analysis is illustrated below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Movie Name | Genre | BoxOfficeCollection | Rating |
| 2000 | Kaho Naa... Pyaar Hai | Romance, Action | 62 | 6.9 |
| 2000 | Mohabbatein | Drama, Romance | 43.5 | 7.1 |
| 2000 | Mission Kashmir | Action, Drama | 34 | 6.8 |
| 2000 | Josh | Action, Drama | 30 | 6.1 |
| 2000 | Har Dil Jo Pyar Karega | Drama, Romance | 24 | 5.2 |
| 2000 | Dulhan Hum Le Jayenge | Comedy, Romance | 21 | 5.1 |
| 2000 | Fiza | Drama | 18 | 6.1 |
| 2000 | Refugee | Drama, Romance | 17 | 5.4 |
| 2000 | Raju Chacha | Drama, Family | 14.5 | 5.3 |
| 2000 | Dhadkan | Drama, Romance | 13.5 | 6.5 |
| 2001 | Gadar: Ek Prem Katha | Action, Drama | 76.88 | 7.2 |
| 2001 | Kabhi Khushi Kabhie Gham... | Drama, Romance | 55.65 | 7.4 |
| 2001 | Lagaan: Once Upon a Time... | Drama, Sport | 36 | 8.1 |
| 2001 | Dil Chahta Hai | Comedy, Drama | 28 | 8.1 |
| 2001 | Indian | Drama, Thriller | 22 | 5.8 |
| 2001 | Mujhe Kucch Kehna Hai | Romance, Drama | 20.25 | 5.3 |
| 2001 | Yaadein | Drama, Musical | 17 | 4.3 |
| 2001 | Ajnabee | Mystery, Thriller | 16.5 | 6.4 |
| 2001 | Chori Chori Chupke Chupke | Drama, Romance | 14 | 5.1 |
| 2001 | Zubeidaa | Biography, Drama | 13 | 6.5 |
| 2002 | Devdas | Drama, Romance | 41 | 7.6 |
| 2002 | Raaz | Horror, Thriller | 23 | 6.5 |
| 2002 | Saathiya | Drama, Romance | 15 | 6.8 |
| 2002 | Kaante | Action, Crime | 14 | 6.5 |
| 2002 | Humraaz | Musical, Thriller | 13.5 | 6.4 |
| 2002 | Mujhse Dosti Karoge! | Drama, Romance | 12.5 | 5.1 |
| 2002 | Aankhen | Drama, Thriller | 12.25 | 7.4 |
| 2002 | Mere Yaar Ki Shaadi Hai | Comedy, Drama | 12 | 5.4 |
| 2002 | Awara Paagal Deewana | Action, Comedy | 11.25 | 6.2 |
| 2002 | Company | Action, Crime | 10.5 | 8 |
| 2003 | Koi... Mil Gaya | Drama, Sci-Fi | 49.5 | 7.1 |
| 2003 | Kal Ho Naa Ho | Comedy, Drama | 38 | 8 |
| 2003 | Baghban | Drama | 22.5 | 7.4 |
| 2003 | Tere Naam | Drama, Romance | 14.5 | 7.1 |
| 2003 | Andaaz | Drama, Romance | 13.5 | 5.3 |
| 2003 | Chalte Chalte | Drama, Romance | 13 | 6.6 |
| 2003 | Main Prem Ki Diwani Hoon | Comedy, Romance | 12.5 | 4.1 |
| 2003 | Hungama | Comedy | 12 | 7.5 |
| 2003 | LOC Kargil | Action, Drama | 11.5 | 5.1 |
| 2003 | Qayamat | Action, Thriller | 10.5 | 4.4 |
| 2004 | Veer-Zaara | Drama, Romance | 41 | 7.8 |
| 2004 | Main Hoon Na | Action, Comedy | 33 | 7 |
| 2004 | Mujhse Shaadi Karogi | Comedy, Drama | 29 | 6.7 |
| 2004 | Dhoom | Action, Crime | 29 | 6.6 |
| 2004 | Hum Tum | Comedy, Drama | 21 | 7 |
| 2004 | Fida | Action, Romance | 14.5 | 5.5 |
| 2004 | Murder | Drama, Thriller | 14.5 | 5.5 |
| 2004 | Masti | Comedy, Romance | 13.5 | 6.1 |
| 2004 | Lakshya | Drama, War | 13 | 7.8 |
| 2004 | Aitraaz | Drama, Romance | 13 | 6.6 |
| 2005 | No Entry | Comedy, Drama | 43 | 6.5 |
| 2005 | Bunty Aur Babli | Comedy, Drama | 36 | 6.3 |
| 2005 | Garam Masala | Comedy, Drama | 33 | 6.6 |
| 2005 | Kyaa Kool Hai Hum | Comedy | 29 | 6.3 |
| 2005 | Black | Drama | 28 | 8.1 |
| 2005 | Salaam Namaste | Comedy, Drama | 27 | 6.3 |
| 2005 | Paheli | Drama, Fantasy | 27 | 6.5 |
| 2005 | Parineeta | Drama, Musical | 26 | 7.1 |
| 2005 | Waqt: The Race Against Time | Drama, Family | 23 | 7 |
| 2005 | Maine Pyaar Kyun Kiya | Comedy, Drama | 22.5 | 5.5 |
| 2006 | Dhoom 2 | Action, Crime | 82.3 | 6.3 |
| 2006 | Lage Raho Munna Bhai | Comedy, Drama | 75 | 8.1 |
| 2006 | Krrish | Action, Adventure | 69 | 6.4 |
| 2006 | Kabhi Alvida Naa Kehna | Drama, Romance | 46 | 6.1 |
| 2006 | Don | Action, Crime | 51 | 7.1 |
| 2006 | Phir Hera Pheri | Comedy, Crime | 41 | 6.9 |
| 2006 | Fanaa | Drama, Romance | 52 | 7.1 |
| 2006 | Vivah | Drama, Romance | 31 | 6.6 |
| 2006 | Omkara | Action, Crime | 23.5 | 8.1 |
| 2006 | Baabul | Drama, Family | 22 | 5.3 |
| 2007 | Om Shanti Om | Comedy, Drama | 79.42 | 6.7 |
| 2007 | Welcome | Comedy, Romance | 72.21 | 6.9 |
| 2007 | Taare Zameen Par | Drama, Family | 61.83 | 8.4 |
| 2007 | Chak De! India | Drama, Sport | 61.59 | 8.2 |
| 2007 | Partner | Comedy, Romance | 60.45 | 5.7 |
| 2007 | Bhool Bhulaiyaa | Comedy, Horror | 49.64 | 7.4 |
| 2007 | Heyy Babyy | Comedy, Drama | 47.8 | 6 |
| 2007 | Jab We Met | Comedy, Romance | 31.2 | 7.9 |
| 2007 | Guru | Drama, Biography | 27.93 | 7.7 |
| 2007 | Namastey London | Drama, Romance | 27.63 | 7.1 |
| 2008 | Ghajini | Action, Drama | 114 | 7.3 |
| 2008 | Rab Ne Bana Di Jodi | Comedy, Drama | 84.68 | 7.2 |
| 2008 | Singh Is Kinng | Action, Comedy | 69.5 | 5.7 |
| 2008 | Dostana | Comedy, Drama | 44 | 6.5 |
| 2008 | Race | Action, Drama | 60 | 6.6 |
| 2008 | Jodhaa Akbar | Drama, History | 56 | 7.6 |
| 2008 | Golmaal Returns | Comedy | 51 | 5.1 |
| 2008 | Fashion | Drama | 27 | 7 |
| 2008 | Rock On!! | Drama, Musical | 27 | 7.7 |
| 2008 | Jaane Tu... Ya Jaane Na | Comedy, Romance | 24 | 7.5 |
| 2009 | 3 Idiots | Comedy, Drama | 202 | 8.4 |
| 2009 | Love Aaj Kal | Comedy, Drama | 66.56 | 6.8 |
| 2009 | Wanted | Action, Crime | 61 | 6.6 |
| 2009 | Ajab Prem Ki Ghazab Kahani | Comedy, Drama | 63.5 | 6.3 |
| 2009 | De Dana Dan | Comedy, Drama | 48 | 5.2 |
| 2009 | Kaminey | Action, Crime | 41 | 7.4 |
| 2009 | All the Best: Fun Begins | Comedy | 42 | 6.1 |
| 2009 | Paa | Drama | 30 | 7.1 |
| 2009 | Dev.D | Drama, Romance | 24.5 | 8 |
| 2009 | New York | Drama, Thriller | 41 | 6.9 |
| 2010 | Dabangg | Action, Comedy | 141.24 | 6.2 |
| 2010 | Golmaal 3 | Comedy, Drama | 108 | 5.3 |
| 2010 | Raajneeti | Drama, Thriller | 93 | 7.1 |
| 2010 | My Name Is Khan | Drama, Romance | 82 | 8 |
| 2010 | Housefull | Comedy, Romance | 72 | 5.4 |
| 2010 | Tees Maar Khan | Comedy, Crime | 60 | 2.6 |
| 2010 | Once Upon a Time in Mumbaai | Action, Crime | 58 | 7.4 |
| 2010 | I Hate Luv Storys | Comedy, Romance | 42 | 5.6 |
| 2010 | Band Baaja Baaraat | Comedy, Drama | 31 | 7.2 |
| 2010 | Peepli Live | Comedy, Drama | 30 | 7.4 |
| 2011 | Bodyguard | Action, Romance | 148.86 | 4.6 |
| 2011 | Ready | Comedy, Drama | 119.78 | 4.8 |
| 2011 | Ra.One | Action, Sci-Fi | 113.94 | 4.7 |
| 2011 | Don 2 | Action, Crime | 112 | 7 |
| 2011 | Singham | Action, Crime | 100.3 | 6.8 |
| 2011 | Zindagi Na Milegi Dobara | Adventure, Drama | 90 | 8.2 |
| 2011 | The Dirty Picture | Biography, Drama | 80 | 6.6 |
| 2011 | Rockstar | Drama, Musical | 70 | 7.6 |
| 2011 | Yamla Pagla Deewana | Comedy, Drama | 55 | 5.7 |
| 2011 | Murder 2 | Drama, Thriller | 50 | 6.2 |
| 2012 | Ek Tha Tiger | Action, Adventure | 198.78 | 5.5 |
| 2012 | Dabangg 2 | Action, Comedy | 155 | 4.9 |
| 2012 | Rowdy Rathore | Action, Comedy | 131 | 5.7 |
| 2012 | Jab Tak Hai Jaan | Drama, Romance | 120.85 | 6.7 |
| 2012 | Agneepath | Action, Drama | 119.98 | 6.9 |
| 2012 | Barfi! | Comedy, Drama | 112 | 8.1 |
| 2012 | Housefull 2 | Comedy, Romance | 114 | 5.4 |
| 2012 | Talaash | Drama, Thriller | 90 | 7.2 |
| 2012 | Son of Sardaar | Action, Comedy | 88 | 4 |
| 2012 | Bol Bachchan | Comedy, Drama | 84 | 5.5 |
| 2013 | Dhoom 3 | Action, Crime | 282.8 | 5.4 |
| 2013 | Chennai Express | Action, Comedy | 227.13 | 6 |
| 2013 | Krrish 3 | Action, Sci-Fi | 240.5 | 5.2 |
| 2013 | Yeh Jawaani Hai Deewani | Comedy, Drama | 177.5 | 7.1 |
| 2013 | Goliyon Ki Raasleela Ram-Leela | Drama, Romance | 112.5 | 6.4 |
| 2013 | Bhaag Milkha Bhaag | Biography, Drama | 108.93 | 8.2 |
| 2013 | Grand Masti | Comedy, Drama | 102 | 4.4 |
| 2013 | Aashiqui 2 | Drama, Musical | 78.42 | 7 |
| 2013 | Special 26 | Crime, Thriller | 66 | 8 |
| 2013 | Raanjhanaa | Drama, Romance | 62 | 7.6 |
| 2014 | PK | Comedy/Drama | 854 | 8.2 |
| 2014 | Kick | Action/Comedy | 402 | 5.5 |
| 2014 | Happy New Year | Action/Comedy | 398 | 5.1 |
| 2014 | Bang Bang! | Action/Romance | 340 | 6.5 |
| 2014 | Singham Returns | Action | 216 | 5.6 |
| 2014 | Holiday: A Soldier... | Action/Thriller | 178 | 7.4 |
| 2014 | Jai Ho | Action/Drama | 183 | 5.3 |
| 2014 | Ek Villain | Thriller | 175 | 6.4 |
| 2014 | 2 States | Romance/Comedy | 175 | 6.9 |
| 2014 | Queen | Comedy/Drama | 108 | 8.2 |
| 2015 | Bajrangi Bhaijaan | Drama/Action | 969 | 8 |
| 2015 | Prem Ratan Dhan Payo | Romance/Drama | 432 | 4.8 |
| 2015 | Bajirao Mastani | Period/Drama | 356 | 7.2 |
| 2015 | Dilwale | Romance/Action | 394 | 5.3 |
| 2015 | Singh Is Bliing | Action/Comedy | 135 | 5.4 |
| 2015 | Tamasha | Romance/Drama | 136 | 7.2 |
| 2015 | Piku | Drama/Comedy | 141 | 7.6 |
| 2015 | Gabbar is Back | Action/Drama | 162 | 7.1 |
| 2015 | Baby | Action/Thriller | 125 | 7.4 |
| 2015 | Badlapur | Thriller/Drama | 78 | 7.5 |
| 2016 | Dangal | Biographical/Drama | 2024 | 8.4 |
| 2016 | Sultan | Sports/Drama | 623 | 7 |
| 2016 | Ae Dil Hai Mushkil | Romance/Drama | 237 | 5.8 |
| 2016 | MS Dhoni: The Untold Story | Biographical/Sports | 216 | 7.8 |
| 2016 | Airlift | Thriller/Drama | 231 | 8 |
| 2016 | Rustom | Crime/Thriller | 218 | 7.1 |
| 2016 | Dear Zindagi | Romance/Drama | 139 | 7.6 |
| 2016 | Neerja | Biographical/Drama | 135 | 7.8 |
| 2016 | Kapoor & Sons | Drama/Comedy | 152 | 7.7 |
| 2016 | Pink | Drama/Thriller | 107 | 8.1 |
| 2017 | Baahubali 2: The Conclusion | Epic/Action | 1810 | 8.2 |
| 2017 | Golmaal Again | Comedy | 311 | 5 |
| 2017 | Tiger Zinda Hai | Action/Drama | 570 | 6 |
| 2017 | Judwaa 2 | Comedy/Romance | 227 | 3.9 |
| 2017 | Toilet: Ek Prem Katha | Comedy/Drama | 311 | 7.2 |
| 2017 | Raees | Crime/Drama | 308 | 6.7 |
| 2017 | Fukrey Returns | Comedy | 110 | 4.8 |
| 2017 | Badrinath Ki Dulhania | Romance/Drama | 214 | 6.2 |
| 2017 | Hindi Medium | Comedy/Drama | 335 | 7.8 |
| 2017 | Secret Superstar | Musical/Drama | 965 | 7.8 |
| 2018 | Sanju | Biographical/Drama | 586 | 7.7 |
| 2018 | Padmaavat | Period/Drama | 585 | 7.1 |
| 2018 | Simmba | Action/Drama | 474 | 5.9 |
| 2018 | Thugs of Hindostan | Action/Adventure | 335 | 4.9 |
| 2018 | Stree | Horror/Comedy | 180 | 7.6 |
| 2018 | Raazi | Spy/Thriller | 194 | 7.8 |
| 2018 | Badhaai Ho | Comedy/Drama | 221 | 8 |
| 2018 | Andhadhun | Thriller/Comedy | 456 | 8.3 |
| 2018 | Hichki | Drama/Comedy | 239 | 7.5 |
| 2018 | Zero | Romance/Drama | 181 | 5.5 |
| 2019 | War | Action/Thriller | 474 | 6.5 |
| 2019 | Kabir Singh | Romance/Drama | 379 | 7.1 |
| 2019 | Uri: The Surgical Strike | Action/Drama | 342 | 8.2 |
| 2019 | Bharat | Drama/Action | 326 | 5.3 |
| 2019 | Gully Boy | Musical/Drama | 238 | 8.2 |
| 2019 | Housefull 4 | Comedy | 280 | 3.3 |
| 2019 | Mission Mangal | Drama/Sci-Fi | 290 | 6.5 |
| 2019 | Kesari | Action/Drama | 207 | 7.4 |
| 2019 | Total Dhamaal | Comedy/Adventure | 228 | 4.3 |
| 2019 | Chhichhore | Comedy/Drama | 215 | 8.2 |
| 2020 | Tanhaji: The Unsung Warrior | Action/Drama | 367 | 7.6 |
| 2020 | Baaghi 3 | Action | 137 | 2.1 |
| 2020 | Angrezi Medium | Comedy/Drama | 101 | 7.3 |
| 2020 | Street Dancer 3D | Dance/Drama | 97 | 3.6 |
| 2020 | Malang | Romance/Thriller | 84 | 6.7 |
| 2020 | Shubh Mangal Zyada Saavdhan | Romance/Comedy | 86 | 6.9 |
| 2020 | Love Aaj Kal | Romance/Drama | 52 | 2.6 |
| 2020 | Chhapaak | Drama | 55 | 5 |
| 2020 | Thappad | Drama | 44 | 6.9 |
| 2020 | Gulabo Sitabo | Comedy/Drama | 65 | 6.3 |
| 2021 | Radhe | Action | 245 | 1.9 |
| 2021 | Sooryavanshi | Action/Drama | 231 | 4.3 |
| 2021 | Bell Bottom | Spy/Thriller | 60 | 7 |
| 2021 | Shershaah | Biographical/Drama | 184 | 8.1 |
| 2021 | Chehre | Mystery/Thriller | 28 | 5.7 |
| 2021 | Chandigarh Kare Aashiqui | Romance/Drama | 38 | 7.4 |
| 2021 | Bhuj: The Pride of India | War/Drama | 32 | 4.2 |
| 2021 | Maidaan | Sports/Drama | 75 | 8 |
| 2021 | Jersey | Sports/Drama | 94 | 8.4 |
| 2021 | Satyameva Jayate 2 | Action | 86 | 1.9 |

***Implementation Code:***

import pandas as pd

import matplotlib.pyplot as plt

from statsmodels.tsa.holtwinters import ExponentialSmoothing

# Read the data from CSV file

data = pd.read\_csv('movie\_data.csv')

# Define the specified genres

specified\_genres = ['Romance', 'Action', 'Drama', 'Comedy', 'Family', 'Thriller',

'Musical', 'Mystery', 'Biography', 'Horror', 'Crime', 'Sci-Fi', 'War',

'Fantasy', 'History', 'Adventure', 'Epic', 'Period', 'Biographical',

'Sports', 'Spy', 'Dance']

# Function to calculate average rating, box office collection, and count of movies for each genre in a given period

def calculate\_averages\_and\_counts(data, start\_year, end\_year):

period\_data = data[(data['Year'].dt.year >= start\_year) & (data['Year'].dt.year <= end\_year)]

genre\_stats = {genre: period\_data[period\_data['Genre'].str.contains(genre, case=False)] for genre in specified\_genres}

genre\_ratings = {genre: stats['Rating'].mean() for genre, stats in genre\_stats.items()}

genre\_boxoffice = {genre: stats['BoxOfficeCollection'].mean() for genre, stats in genre\_stats.items()}

genre\_counts = {genre: len(stats) for genre, stats in genre\_stats.items()}

return dict(sorted(genre\_ratings.items())), dict(sorted(genre\_boxoffice.items())), dict(sorted(genre\_counts.items()))

# Function to plot bar charts with counts

def plot\_bar\_chart\_with\_counts(data, counts, title, ylabel, color):

labels = [f"{genre} ({count} movies)" for genre, count in counts.items()]

plt.figure(figsize=(12, 6))

plt.bar(labels, data.values(), color=color)

plt.title(title)

plt.xlabel('Genre')

plt.ylabel(ylabel)

plt.xticks(rotation=45, ha='right')

plt.tight\_layout()

plt.show()

# Function to prepare and forecast data using Holt-Winters method

def forecast\_data(genre\_yearly\_data, steps=4):

forecasted\_data = {}

for genre, data in genre\_yearly\_data.items():

if len(data) > 2: # Holt-Winters needs at least 3 points to work

data = data.reindex(range(2000, 2022), fill\_value=0)

model = ExponentialSmoothing(data, trend='add', seasonal=None)

fit = model.fit()

forecasted\_data[genre] = fit.forecast(steps=steps)

return forecasted\_data

# Function to plot forecasted data

def plot\_forecasted\_data(forecasted\_data, title, ylabel):

plt.figure(figsize=(14, 7))

for genre, forecast in forecasted\_data.items():

plt.plot(forecast.index, forecast, label=genre, marker='o')

plt.title(title)

plt.xlabel('Year')

plt.ylabel(ylabel)

plt.xticks(range(2022, 2026))

plt.legend(loc='upper left', bbox\_to\_anchor=(1, 1), ncol=2, fontsize='small')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Convert 'Year' column to datetime

data['Year'] = pd.to\_datetime(data['Year'], format='%Y')

# Calculate averages and counts for the periods 2000-2010 and 2011-2021

ratings\_2000\_2010, boxoffice\_2000\_2010, counts\_2000\_2010 = calculate\_averages\_and\_counts(data, 2000, 2010)

ratings\_2011\_2021, boxoffice\_2011\_2021, counts\_2011\_2021 = calculate\_averages\_and\_counts(data, 2011, 2021)

# Plotting bar charts for each period

plot\_bar\_chart\_with\_counts(ratings\_2000\_2010, counts\_2000\_2010, 'Average Rating for Each Genre from 2000 to 2010', 'Average Rating', 'skyblue')

plot\_bar\_chart\_with\_counts(boxoffice\_2000\_2010, counts\_2000\_2010, 'Average Box Office Collection for Each Genre from 2000 to 2010', 'Average Box Office Collection (INR in Crores)', 'lightgreen')

plot\_bar\_chart\_with\_counts(ratings\_2011\_2021, counts\_2011\_2021, 'Average Rating for Each Genre from 2011 to 2021', 'Average Rating', 'skyblue')

plot\_bar\_chart\_with\_counts(boxoffice\_2011\_2021, counts\_2011\_2021, 'Average Box Office Collection for Each Genre from 2011 to 2021', 'Average Box Office Collection (INR in Crores)', 'lightgreen')

# Calculate overall averages and counts for 2000-2021

ratings\_2000\_2021, boxoffice\_2000\_2021, counts\_2000\_2021 = calculate\_averages\_and\_counts(data, 2000, 2021)

# Plotting overall average charts for 2000-2021

plot\_bar\_chart\_with\_counts(ratings\_2000\_2021, counts\_2000\_2021, 'Average Rating for Each Genre from 2000 to 2021', 'Average Rating', 'skyblue')

plot\_bar\_chart\_with\_counts(boxoffice\_2000\_2021, counts\_2000\_2021, 'Average Box Office Collection for Each Genre from 2000 to 2021', 'Average Box Office Collection (INR in Crores)', 'lightgreen')

# Filter data from 2000 to 2021

data = data[(data['Year'].dt.year >= 2000) & (data['Year'].dt.year <= 2021)]

# Create dictionaries to store the average ratings and box office collections per year for each genre

genre\_yearly\_ratings = {genre: data[data['Genre'].str.contains(genre, case=False)].groupby(data['Year'].dt.year)['Rating'].mean() for genre in specified\_genres}

genre\_yearly\_boxoffice = {genre: data[data['Genre'].str.contains(genre, case=False)].groupby(data['Year'].dt.year)['BoxOfficeCollection'].mean() for genre in specified\_genres}

# Forecast the average ratings and box office collections for each genre from 2022 to 2025

forecast\_ratings = forecast\_data(genre\_yearly\_ratings)

forecast\_boxoffice = forecast\_data(genre\_yearly\_boxoffice)

# Plotting the forecasted average ratings

plot\_forecasted\_data(forecast\_ratings, 'Genre Preference Prediction (Average Rating) for the Next 4 Years (2022-2025)', 'Predicted Average Rating')

# Plotting the forecasted average box office collections

plot\_forecasted\_data(forecast\_boxoffice, 'Genre Preference Prediction (Box Office Collections) for the Next 4 Years (2022-2025)', 'Predicted Box Office Collection (INR in Crores)')

# Plotting trend line graph for average rating over years (2000 to 2025)

plt.figure(figsize=(14, 7))

for genre, yearly\_ratings in genre\_yearly\_ratings.items():

plt.plot(yearly\_ratings.index, yearly\_ratings.values, label=genre)

for genre, forecast in forecast\_ratings.items():

plt.plot(range(2022, 2026), forecast, label=f"{genre} (forecast)", linestyle='--')

plt.title('Trend Line for Average Rating (2000-2025)')

plt.xlabel('Year')

plt.ylabel('Average Rating')

plt.xticks(range(2000, 2026))

plt.legend(loc='center left', bbox\_to\_anchor=(1, 0.5), ncol=2, fontsize='small')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Plotting trend line graph for box office collection over years (2000 to 2025)

plt.figure(figsize=(14, 7))

for genre, yearly\_boxoffice in genre\_yearly\_boxoffice.items():

plt.plot(yearly\_boxoffice.index, yearly\_boxoffice.values, label=genre)

for genre, forecast in forecast\_boxoffice.items():

plt.plot(range(2022, 2026), forecast, label=f"{genre} (forecast)", linestyle='--')

plt.title('Trend Line for Box Office Collection (2000-2025)')

plt.xlabel('Year')

plt.ylabel('Box Office Collection (INR in Crores)')

plt.xticks(range(2000, 2026))

plt.legend(loc='center left', bbox\_to\_anchor=(1, 0.5), ncol=2, fontsize='small')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Determine the genre with the highest predicted average rating and box office collection in the last forecasted year

most\_preferred\_genre\_rating = max(forecast\_ratings, key=lambda x: forecast\_ratings[x].iloc[-1])

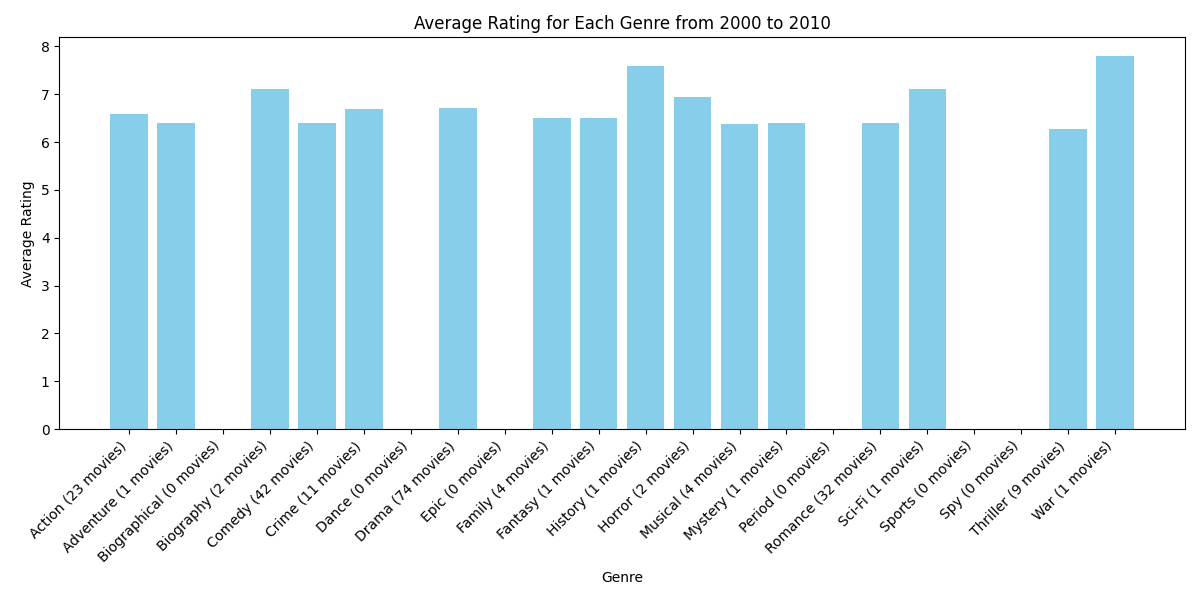
most\_preferred\_genre\_boxoffice = max(forecast\_boxoffice, key=lambda x: forecast\_boxoffice[x].iloc[-1])

print(f"The most preferred genre based on average rating in the next 4 years is predicted to be: {most\_preferred\_genre\_rating}")

print(f"The most preferred genre based on box office collections in the next 4 years is predicted to be: {most\_preferred\_genre\_boxoffice}")

***Output:***

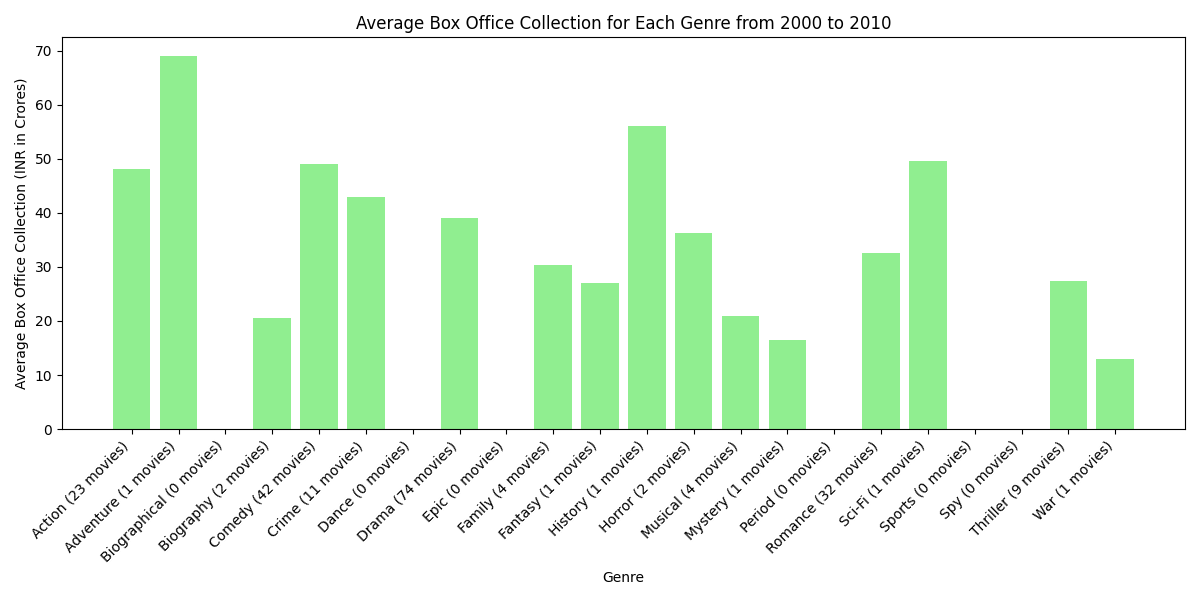
The following graphs were obtained after executing the code:

**Year 2000-2010**  


Key insights from the Bar chart:

Top Rated Genres: The genres "Musical" and "War" have the highest average ratings, each with an average rating of 8. However, it's important to note that these genres have a very low number of movies (1-2 movies), which could skew the average.

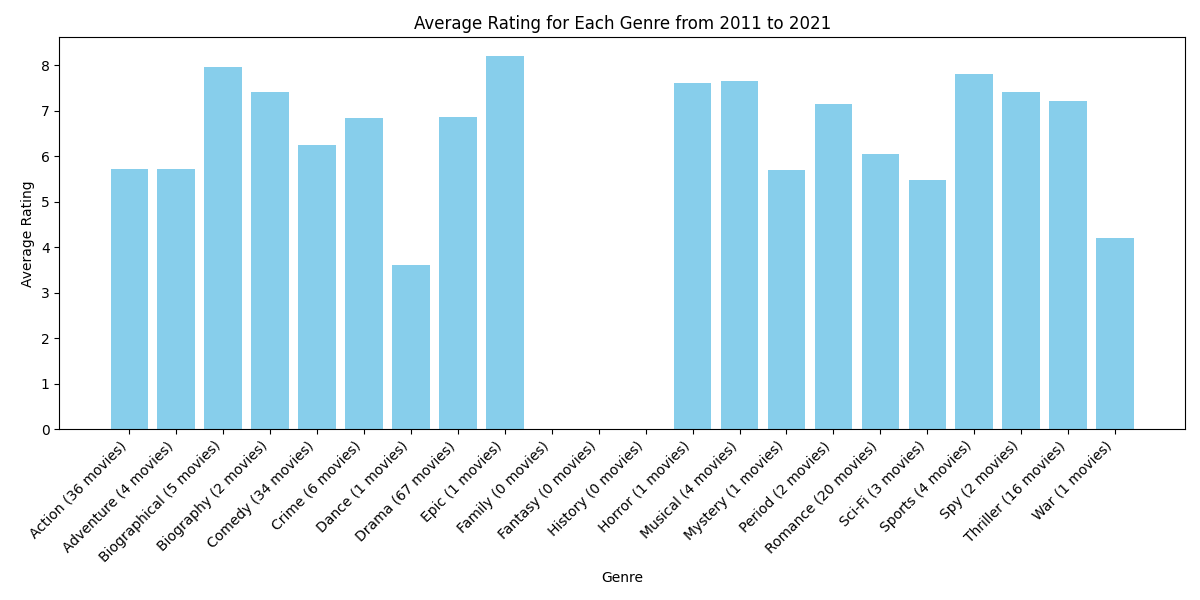
Genres with Moderate Ratings and More Movies: Genres like "Comedy" and "Drama" have lower average ratings compared to "Musical" and "War" but have a significantly higher number of movies. "Comedy" has an average rating of about 7 with 42 movies, and "Drama" has an average rating of about 7 with 74 movies, indicating a larger sample size and a consistent audience preference.



The bar chart provides the following key insights:

Top Box Office Collections: The genres "Biographical" and "Musical" have the highest average box office collections, with "Biographical" reaching around 70 crores and "Musical" also having a high collection. However, the number of movies in these genres is very low (0-1 movies), which might not provide a comprehensive view.

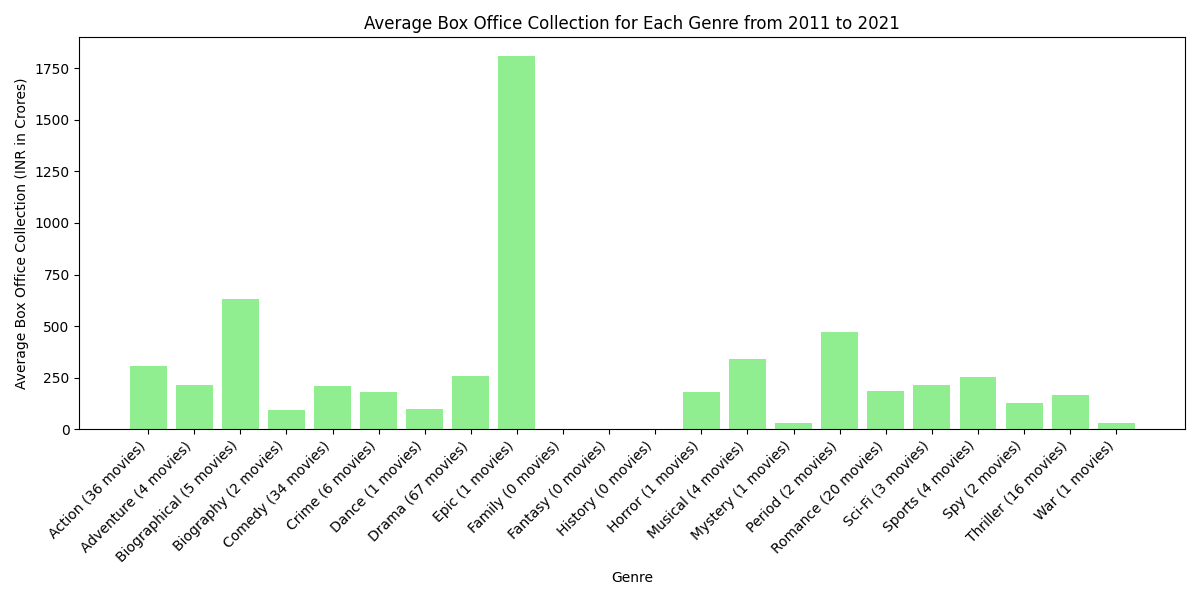
Genres with Moderate Box Office Collections and More Movies: "Action" and "Drama" have significant box office collections along with a higher number of movies. "Action" has an average box office collection of around 50 crores with 23 movies, and "Drama" has around 40 crores with 74 movies, indicating these genres are consistently popular and attract a larger audience over time.

**Year 2011-2021**

The bar chart provides the following key insights:

Top Rated Genres : The genres "Family" and "Musical" have the highest average ratings, each scoring around 8. However, these genres have very few movies (1-4 movies), which could affect the reliability of the average rating.

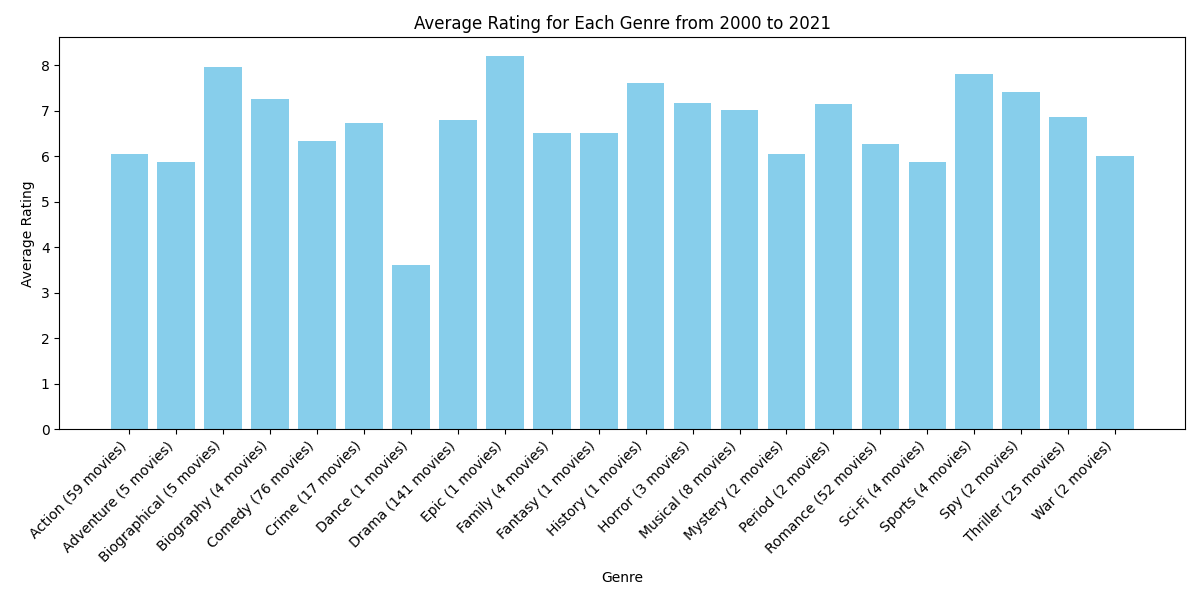
Genres with Moderate Ratings and More Movies : Genres such as "Comedy" and "Drama" have significant numbers of movies with consistent ratings. "Comedy" has an average rating of about 7 with 34 movies, and "Drama" has a similar rating with 67 movies. These larger sample sizes indicate stable audience preferences over the period from 2011 to 2021.



The bar chart shows the average box office collection for various movie genres in India from 2011 to 2021. Here are the two key takeaways:

Drama is the genre with the most movies (67) included in the dataset.

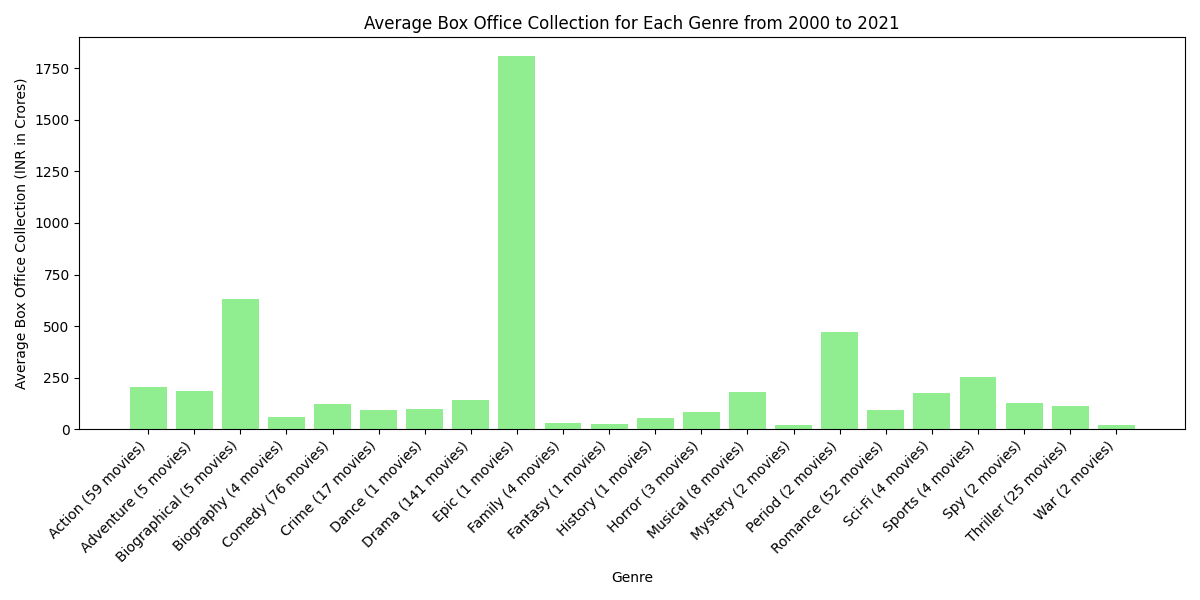
Action is the genre with the highest average box office collection (around 1500 Crore INR).

**AVERAGE FOR YEAR 2000-2021**

The most important takeaway from this graph is that the average rating for each genre has increased from 2000 to 2021. However, it is important to note that the y-axis (average rating) is not labeled, so we cannot determine the exact value of the increase.

Here are some additional observations:

1. Drama is the genre with the most movies (141) included in this dataset.
2. There are some genres with very few movies included in the dataset (e.g., Dance, Epic, Fantasy).

Here are the two key takeaways:

1. Drama is the genre with the most movies (141) included in the dataset.
2. Action is the genre with the highest average box office collection (around 1500 Crore INR).

**DISCUSSION:**

**Introduction:**

The audience interest in popularity of movie genres has always been a very interesting topic in the wide world of film. To be successful, filmmakers, film studios and investors have to understand what movies people like most. This paper examines the trends of movie genres between 2000 and 2021 in order; to find out which ones have done well and then make predictions for four years from now.

**Analyzing Past Trends:**

It is interesting to note that one can learn a lot about the genre’s performance over the past twenty years. Theatrical blockbusters in superheroic, action or fantasy style dominated cinemas paid by such franchises as Marvel Cinematic Universe or Star Wars. These genres are suited for broad audiences since they appeal both casual filmgoers and die-hard fans.

Additionally, streaming services played a significant role in making particular genres popular among viewers with diverse preferences who want different content – comedy, drama and thriller as examples. In recent years rom-coms have lost their former glory but there are still people who adore them.

**Factors Influencing Genre Popularity:**

A number of factors explain why some genres are more successful than others. Cultural movements, changing society norms, technological improvements or financial conditions all contribute towards shaping audience preferences. For instance, growing escapism demands.

Furthermore, the evolution of storytelling techniques and special effects has expanded the possibilities within genres, attracting audiences with immersive experiences. Additionally, the accessibility of streaming services has enabled viewers to explore niche genres and international cinema, contributing to the diversification of viewing habits.

Predicting the future of genre performance is a deeply complex process. Although past performances give us insightful foresight, one must combine old trends with the new styles that are coming up, and also the old audience gradually becomes the new one. It is very probable that in the next four year period we will observe both a development of certain current trends and an emergence of different dynamics.

The genres that have proven to be strong and can change themselves, like superhero films and action-adventures, will remain the most popular and expectedly so with the arrival of a number of highly awaited sequels and spin-offs. On the other hand, storytelling may also be reinvented by audiences who are looking for more interesting ways to express current issues.

Representation as well as diversity is becoming an appealing aspect in media. This under independent dramas and documentaries; hence those which have underrepresented voices or perspectives might end up being focused on. As a result of this the combination genres which merge more than one category could provide better options for individuals who are in need of coming up with fresh experiences.

**Conclusion:**

In the end of this analysis on the movie genre trends from 2000 to 2021 one can get really valuable information about the audience preferences and market dynamics. Although such genres as superhero films and action-adventures are prevalent, one must not overlook that there is still space for innovation and diversification. Through combining past information with tomorrow’s leading practices, stakeholders will gain deep insight for decision-making process and futuristic cinema scenario development. The coming years will be more about recognizing changing audience tastes as well as producing engaging content that will connect with viewers.

**REFERENCE**

1. <https://www.bollywoodhungama.com/movies/top-100-movies/>

2. <https://en.wikipedia.org/wiki/Lists_of_Hindi_films>

3. <https://www.imdb.com/list/ls056464416/>

4.<https://www.filmibeat.com/bollywood/movies/january-2000.html>

5. [www.morismedia.in](http://www.morismedia.in)

6. <https://www.theguerrillarep.com/>

7. [www.researchgate.net](http://www.researchgate.net)

8. medium.com