

CS685A Data Mining project : Movies Analysis and Recommendation System

M.tech

Submitted by :-

Abhinav Dudeja (21111001) (abhinavd21@iitk.ac.in)
Malkeet Singh Dhalla (21111035) (malkeetsd21@iitk.ac.in)
Manthan Kojage (21111039) (manthank21@iitk.ac.in)
Sandeep Saket (21111055) (sandeeps21@iitk.ac.in)
Shivam Kharwar (21111058) (skharwar21@iitk.ac.in)

Under the guidance of: Dr. Arnab Bhattacharya sir



Department of Computer Science & Engineering
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
Kalyanpur, Kanpur, Uttar Pradesh 208016, INDIA.2021-2022

Acknowledgements

We are profoundly grateful to **Dr. Arnab Bhattacharya** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

We would like to express deepest appreciation towards **Dr. Arnab Bhattacharya**, **Mr. Mainak Chaudhari** HOD Computer Science & Engineering Department and **All TAs** whose invaluable guidance supported us in completing this project.

At last we must express our sincere heartfelt gratitude to all the members of Computer Science & Engineering Department who helped us directly or indirectly during this course of work.

Abhinav Dudeja (21111001)

Malkeet Singh Dhalla (21111035)

Manthan Kojage (21111039)

Sandeep Saket (21111055)

Shivam Kharwar (21111058)

Contents

1	Intention of the Project	4
2	Introduction	5
3	Accounted Data-Sets and Data-Frames	6
4	Methodology and Analysis	11
4.1	Analysis System	11
4.2	Recommendation System	13
5	Visual Aided Results	14
5.1	Movies	14
5.2	Genre And Time	16
5.3	Crew	18
6	Conclusion	22

Abstract

The Bollywood industry plays a vital role in India's economy. Every year hundreds of Bollywood movies are released in India with the hope of becoming the next blockbuster, but only one-third of them gets success. The producers, actors, investors, and sponsors are all interested in predicting the movie's box office success.

Movies from several genres like comedy drama, thriller, documentary, biography, crime, family, fantasy, mystery, musical, horror, animation, action, romance, adventure and so on has their own rise and falls and success rate. There is a lot of data associated with the movies and history of Bollywood.

Analysing this data properly can give us all some vital information regarding the trends in Bollywood like a patriotic movies is less probable to go flop or an action movies has highest chances of success. But to as the data storing and data handling techniques have evolved from the time Bollywood has started hence the collection of data haven't been the same and consistent.

For developing an efficient and precise recommendation system that can suggest good movies complying with the users taste we need to integrate the data into a consistent dataframe that can be loaded, accessed, evaluated easily and efficiently.

Our project is focused on that analyzing effect of the genres, the release date around holidays, the release month, actors, directors, ratings on the box office success of a movie and several other attributes on that integration of several datasets.

1 Intention of the Project

Recommendation System is a computer program that is associated with two intervening fields in computer science and they are Data Science and Machine Learning.

Data Science and Analysis helps in determining the useful datasets and helps in predicting the rise and falls in the trends for the area in which system is working. Before creating a Machine Learning model, the model needs to be provided with the accurate datasets with precised and non-outlying values. Our main aim was to focus on this area where we improve the information extraction and modelling method.

Machine Learning helps to develop the model that takes this background prepared data and use this data in its algorithm to predict or recommend the services or products that has a probable chance of meeting users taste.

We proposed a Movies Analysis and Recommendation System from the available and web scrapped datasets and study the trends in bollywood movies era. We depicted many presentable results through graphs and other visual aids wherever it was possible.

We also developed a demo Recommendation System for sake of completion of project and showing how a naive recommendation system uses datasets to predict movies based on given query.

2 Introduction

Recommender Systems are nowadays an indispensable component in the e-commerce applications and environment. They have the capability to predict or suggest a variety of products and services based on user or group of users past history. There are various types of Recommender Systems and we come across them in our daily lives without noticing them.

For instance the autoplay button in the YouTube tries to pick the next song based on users taste and in most case it picks the best song that even user cannot think of, another scenario is where Netflix movie recommendation system chooses the thumbnails and movies in an order best for the user.

These recommender system requires two complex components : a Machine Learning model and DataSets and DataFrames on which the model works.

Developing an efficient machine learning model requires quite a study and analysis and is a part of Machine Learning field but extracting and modelling data into a data structure that can be easily accessible to the model is the part that comes under Data Mining and Data handling.

Our project, "Movie Analysis and Recommendation System" is a demonstration of such a recommender system focusing on data handling and mining aspect of it.

3 Accounted Data-Sets and Data-Frames

As our project focuses on the analysis and solving the queries in regard with the bollywood movies. So the availability of the relevant data from which we can extract important information is necessary. After some research we have found those datasets which were available online so that we could extract information from them and analyze it in order to show some useful findings which will be helpful to show some recent and old trends in the world of cinema. Some of the datasets were available online and some are obtained from web scrapping. Followings are the Datasets used for the project:

1.bollywood_full_1950-2019.csv : This dataset is found online which contains attributes related to bollywood movies such as :
title_x, imdb_id, poster_path, wiki_link, title_y, original_title, is_adult, year_of_release, runtime, genres, imdb_rating, imdb_votes, story, summary, tagline, actors, wins_nominations, release_date
For our analysis we have extracted important attributes from this set of attributes which are :
title_x : The title of the movie
imdb_id : IMDB ID of the movie
year_of_release : Year in which the movie was released
genres : Genres of the movie
imdb_rating : IMDB Rating of the movie
actors : Lead actors from the movie

The dataset can be found here :
<https://www.kaggle.com/pncnmnp/the-indian-movie-database?select=1950-2019>

2. bollywood_crew_data_1950-2019.csv : This dataset is found online which contains attributes related to bollywood movies such as:
crew_id, name, born_year, death_year, profession, known_for
For our analysis we have extracted important attributes from this set of attributes which are :
crew_id : A unique id for a crew member(including actors,directors and producers)
name : Name of crew member
profession : His/Her Profession(i.e actor,director etc.)

The dataset can be found here:
<https://www.kaggle.com/pncnmnp/the-indian-movie-database?select=1950-2019>

3. bollywood_crew_1950-2019.csv : This dataset is found online which contains attributes related to bollywood movies such as:

imdb_id, directors, writers

For our analysis we have extracted important attributes from this set of attributes which are :

imdb_id: IMDB ID of the movie

directors: A unique id for the director(crew member)

The dataset can be found here:

<https://www.kaggle.com/pncnmp/the-indian-movie-database?select=1950-2019>

4. indian movies.csv : This dataset is found online which contains attributes related to bollywood movies such as:

ID, Movie Name, Year, Timing(min), Rating(10), Votes, Genre, Language

For our analysis we have extracted important attributes from this set of attributes which are:

ID: IMDB ID of the movie

Movie Name: The title of the movie

Year: Year in which the movie was released

Genre: Genre of the movie

Language: Language in which the movie is made

The dataset can be found here:

<https://www.kaggle.com/nareshbhat/indian-moviesimdb>

5. BollywoodMovieDetail.csv : This dataset is found online which contains attributes related to bollywood movies such as:

imdbId, title, releaseYear, releaseDate, genre, writers, actors, directors, sequel, hitFlop

For our analysis we have extracted important attributes from this set of attributes which are:

imdbId: IMDB ID of the movie

title: The title of the movie

releaseYear: Year in which the movie was released

genre: Genre of the movie

actors: Lead Actors of the movie

directors: Director of the movie

The dataset can be found here:

<https://www.kaggle.com/mitesh58/bollywood-movie-dataset?select=Bollywood\MovieDetail.csv>

6. BollywoodActorRanking.csv : This dataset is found online which contains attributes related to bollywood movies such as:

actorId, actorName, movieCount, ratingSum, normalizedMovieRank, googleHits, normalizedGoogleRank, normalizedRating

For our analysis we have extracted important attributes from this set of attributes which are:

actorId: A Unique Id of an Actor

actorName: Name of the Actor

movieCount: Total Movies done by the Actor

googleHits: The Number of Google Hits for that Actor

normalizedRating: Rating of an Actor(Normalised 0-10)

The dataset can be found here:

<https://www.kaggle.com/mitesh58/bollywood-movie-dataset?select=Bollywood\ActorRanking.csv>

7. BollywoodDirectorRanking.csv : This dataset is found online which contains attributes related to bollywood movies such as:

directorId, directorName, movieCount, ratingSum, normalizedMovieRank, googleHits, normalizedGoogleRank, normalizedRating

For our analysis we have extracted important attributes from this set of attributes which are:

directorId: A Unique Id of a Director

directorName: Name of the Director

movieCount: Total Movies done by the Director

googleHits: The Number of Google Hits for that Director

normalizedRating: Rating of a Director(Normalised 0-10)

The dataset can be found here:

<https://www.kaggle.com/mitesh58/bollywood-movie-dataset?select=Bollywood\DirectorRanking.csv>

8. boxoffice.csv : This dataset contains attributes related to bollywood movies such as:

movie_id, name, release_date, total_nett_gross, first_week, budget, india_gross, overseas_gross, worldwide_gross, all_time_rank, footfalls, adjusted_nett_gross, imdb_id

For our analysis we have extracted important attributes from this set of attributes which are:

movie_id: A unique Id for a movie

name: Name of the Movie

release_date: The Release date of the Movie

first_week: The first week collections of the Movie

total_nett_gross: The total nett collection of the movie in domestic markets(India)

budget: The budget of the movie

overseas_gross: The overseas nett collection of the movie
worldwide_gross: The worldwide total gross figure for the movie
footfalls: The tickets sold for the Movie
adjusted_net_gross: The adjusted worldwide total gross figure for the movie
imdb_id: The IMDB ID of the movie found out from bollywood_full_1950-2019.csv
through longest common subsequence and fuzzy match logic

This dataset is extracted from:

<https://www.boxofficeindia.com/> using python library BeautifulSoup (WEB
SCRAPPED)

9. boxoffice_actors.csv : This dataset contains attributes related to bollywood movies such as:

actor_id, actor_name, role, movie_id, imdb_id, movie_name, nett_gross, verdict
For our analysis we have extracted important attributes from this set of attributes
which are:

movie_id: A unique Id for a movie

movie_name: Name of the Movie

actor_id: A unique Id for a Crew Member(Lead Actor,Side Actors,Directors,Producers
etc)

actor_name: Name of a Crew Member(Lead Actor,Side Actors,Directors,Producers
etc)

role: Role that the Crew Member plays in making of the movie

nett_gross: The total nett collection of the movie in domestic markets(India)

verdict: Verdict of the Movie (Hit,Flop etc.)

imdb_id: The IMDB ID of the movie found out from bollywood_full_1950-2019.csv
through longest common subsequence and fuzzy match logic.

This dataset is extracted from:

<https://www.boxofficeindia.com/> using python library BeautifulSoup (WEB
SCRAPPED)

10. Data for repository.csv : This dataset is found online which contains
attributes related to bollywood movies such as:

Movie Name, Release Period, Whether Remake, Whether Franchise, Genre, New
Actor, New Director, New Music Director, Lead Star,Director, Music Director,
Number of Screens, Revenue(INR), Budget(INR)

For our analysis we have extracted important attributes from this set of attributes
which are:

Movie Name: Name of the Movie

Release Period: Whether movie released on holiday or non holiday

Genre: Genre of the movie

Lead Star: Lead Actor of the Movie
Director: Director of the Movie
Number of Screens: No of Screens on which the movie has been released on
Revenue(INR): The total nett collection of the movie in domestic markets(India)
Budget(INR): Budget in which the movie is made

The dataset can be found here:

<https://data.mendeley.com/datasets/3c57btcxy9/1>

11. bollywood_box_clean.csv : This dataset is found online which contains attributes related to bollywood movies such as:

movie_name,
movie_opening, movie_weekend, movie_firstweek, movie_total, movie_total_worldwide,
movie_genre, movie_director, release_date, release_day, release_month, release_year,
runtime, producer, banner, actors, movie_details

For our analysis we have extracted important attributes from this set of attributes which are:

movie_name: Name of the Movie
movie_genre: Genre of the movie
actors: Lead Actors of the Movie
movie_director: Director of the Movie
banner: Banner under which the movie is released
producer: Producer of the Movie
movie_total: The total nett collection of the movie in domestic markets(India)
movie_total_worldwide: The total nett collection of the movie worldwide
movie_opening: The opening day nett collection of the movie in india
movie_weekend: The opening weekend nett collection of the movie in india
movie_firstweek: The first week nett collection of the movie in india

The dataset can be found here:

https://www.kaggle.com/sugandhkhobragade/bollywood-box-office-20172020?select=bollywood_box_clean.csv

4 Methodology and Analysis

4.1 Analysis System

For data analysis, We worked on the following problem statements:

1. Rank the movies (from 1994) based on:
 - (a) Gross collection
 - (b) Profit made
 - (c) Profit percentage
2. Make a dataset with the following columns: *Movie Name, Genre, Collection, Ratings*. Compare the *Collection* and *Ratings* and compare the *Genre* and *Collection*. Only take the common movies (2017-2019).
3.
 - (a) Make a dataset with the following columns: *Year of release, number of movies, Genre most used*.
 - (b) Show the top 5 genres for each year.
 - (c) Make a dataset with the following columns: *Year of release, number of movies, collection*. Show the growth/decline of number of movies and collection with each year.
4. Make a dataset with the following columns: *Language, Number of Movies, Highest Rating, Avg Rating, Lowest Rating*.
5. Compare number of movies, google hits and normalized rating of the actors and directors.
 - (a) Find the 10 most popular actors and directors.
 - (b) Find the 10 actors and 10 directors who have worked in maximum number of movies.
6. For the past 13 years
 - (a) Rank the Genre of The Movies based on:
 - Popularity
 - Cumulative profit made
 - Average profit made
 - Cumulative gross collection
 - Average gross collection
 - (b) Show the significance of the Screen Count of the movie on which it was released to the gross collection made by the movie (past 13 years).

7. Find the number of movies by each actor, also find their highest rated movie and average rating of all the movies done by the actor. Make a dataset with columns: *Actor name, no. of movies, highest rating, highest rated movie, avg rating*. Repeat for the directors.
8. Find the genre in which an actor works the most. Make a dataset with columns: *Actor name, Genre Most Performed*. Repeat for the directors.
9. Compare the collection of movies within India and worldwide. Find the movies with maximum collection in India and worldwide. Find the percentage of revenue collected on opening day and opening weekend to the total_collection (recent 3 years).
10. (a) Find the banner, producers, actors, directors which has earned the most money at the domestic box office collection in recent 3 years.
(b) Find all of it based on total worldwide box collections also.
11. Rank the movies based on both IMDb Ratings and Box office Collections from the past 13 years i.e. Find the Rankings of the most successful movies(both box office collection and Critical acclamation wise). Give equal weightage to both in the process of evaluation. Normalize the ratings and collection on a scale of 0-1 and then find the rankings of the most successful movies.
12. Rank the most successful actors (here also the most successful is evaluated based on box office collection and imdb ratings as above) and similarly do this for directors as well.
13. Compare the ratio of revenue to the budget for the movies released on a normal day to the movies released on a holiday. Also do this comparison for movies with a new lead actor to movies with a veteran actor(for the last 13 years).
14. Find the month from each calendar year(from the year 1994) in which the cumulative earnings of all the movies released is the most from that particular calendar year. Output columns as *year, bestmonth, totcollection(total_net_gross)*. Do this for footfalls as well(i.e. most cumulative footfalls of the movies released in the month). Similarly output as *year, month, footfalls*.
15. For each year from 1994, find the movies and actors with highest nett India collection (total_net_gross), highest overseas collection and footfalls. Output columns as *year, movie with highest total_net_gross, movie with highest overseas collection ,highest footfall, actor with highest nett India collection, highest overseas collection, highest footfall*.

16. Find the top 20 actors with highest cumulative earnings of their movies (from 1994). Output columns as *rank*, *actor*, *totalearning*. Do this for directors and producers as well.
17. Find the debut movie for each actor with their debut year. Output columns as *actor*, *debutmovie*, *debutyear*. Do this for producers and directors as well.
18. Find the top 5 actors who have given the most all time blockbusters, blockbusters, superhits, hits, average, flops and disasters. Output columns as *most All Time Block Busters*, *most BlockBusters*, and so on. (The rows should contains the names of the actors underscore the movies for ex. the most BB movies given). Do this for producers and directors as well.
19. Find top 5 actor-director combination that has done most business at the box office. Output as *rank*, *actor*, *director*, *total collections*. Find the top 5 director-producer combination and actor-producer combination as well.
20. (a) Find the total movies in 100cr club, 200cr club and 300cr club. Find actors, directors, producers which got the most movies in the above clubs.
 (b) Find the top 20 actors which are most watched i.e cumulative footfalls. Similarly, find directors and producers as well. Also mention the cumulative footfalls of each evaluation.

4.2 Recommendation System

Our recommendation system is just a demonstration that performs scrapping on the IMDb MovieLens dataset and arranges data into usable structures. It does not implement any ML algorithm or its libraries. It just recommends the top rated movies of the actor and entered genres. Our focus is on Information extraction from the datasets rather than ML model.

5 Visual Aided Results

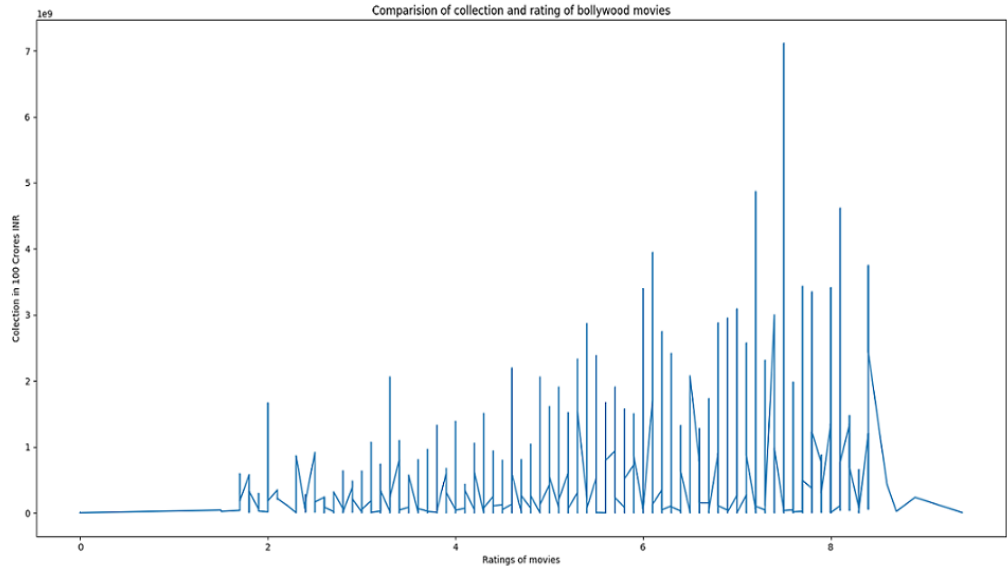
Many results and insights were gathered, after doing a thorough analysis of Bollywood movies and crew (actors, directors and producers) based on the ratings, popularity and business data. It was discovered that 'Akshay Kumar' is the most successful actor for the past 3 three years based on boxoffice collection within India whereas, 'Salman Khan' takes the lead when worldwide collection is considered. Fox Star Studios is the best production house and 'Ali Abbas Zafar' is the most successful director. Other findings were also derived. These findings are divided into three categories:

1. Movies
2. Genre And Time
3. Crew

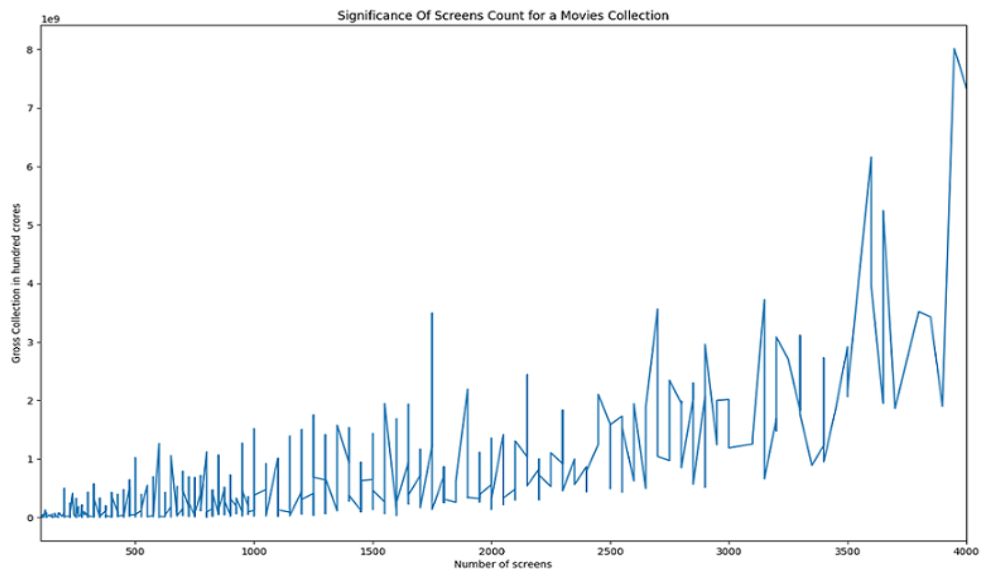
5.1 Movies

Analysis of movies was done on ratings, budget, revenue, footfalls, domestic and worldwide success and following results can be derived:

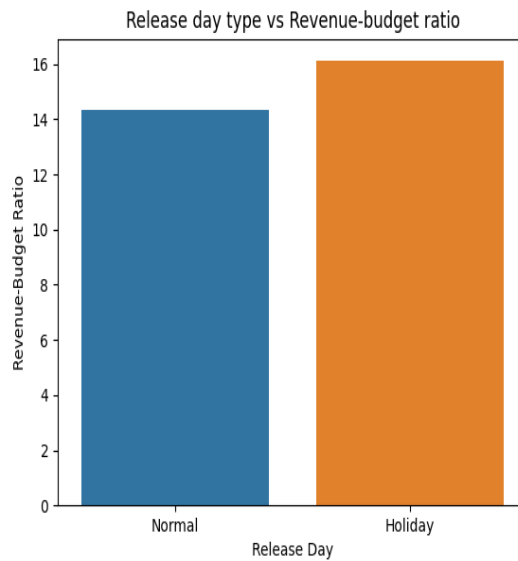
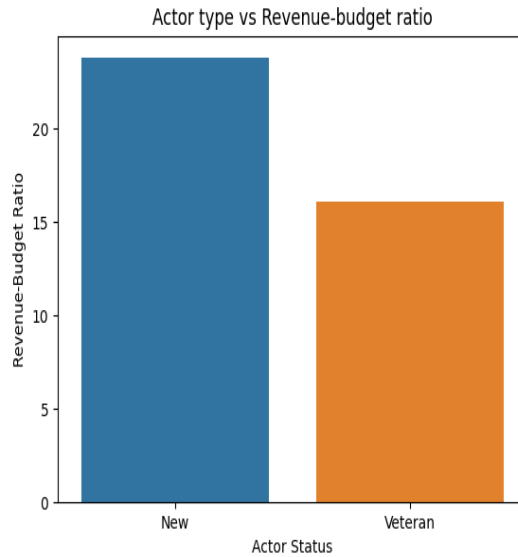
1. 'Hum Aapke Hain Koun..!' was discovered to be the best movie overall when the imdb ratings and revenue was given equal weightage. It was way ahead of the runner-up 'Dilwale Dulhania Le Jaayenge' with a score of 10, whereas DDLJ could only get a score of 7.66. Complete list can be found in the file movie-score.csv.
2. 'Tanhaji' was the most successful movie on the boxoffice last year.
3. After hindi the languages in which most movies are released are malayalam, tamil and telugu. Other language movies like bengali, urdu, kannada are also release but are lesser in number.
4. 'Bahubali 2' has the highest boxoffice collection of all time with a collection of over Rs.800 crores. 'Dangal' and 'PK' are close runner ups with revenue over Rs.700 crores and Rs. 600 crores. 'Bheja Fry' has best return on investment with highest profit percentage.
5. A total of 7 movies are in 300 crore club, 12 are in 200 crore club and 63 are in 100 crore club.
6. It was found with this analysis that rating of a movie has low impact on the revenue made by the movie on the boxoffice. Though few movies with high imdb rating make more revenue but, generally there is no strong evidence that a better rating means a better boxoffice performance. Following graph confirms this.



7. Screen count has a significant affect on the amount of money a movie makes as we can see in the figure.



8. Movies with new actors have had a better revenue to budget ratio for the past 13 years. Also movies released on a holiday have a slightly better ratio than movies released on normal days. Following graphs shows the comparison:

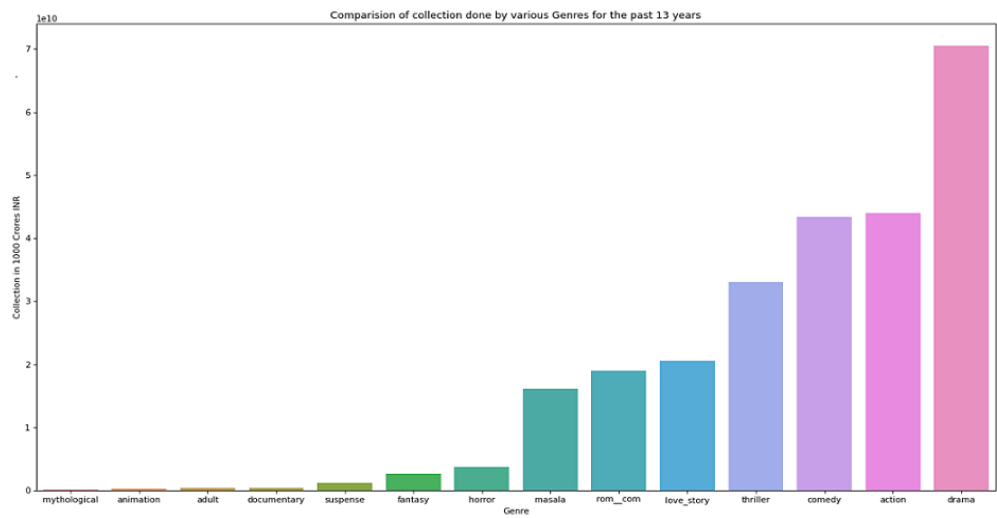


5.2 Genre And Time

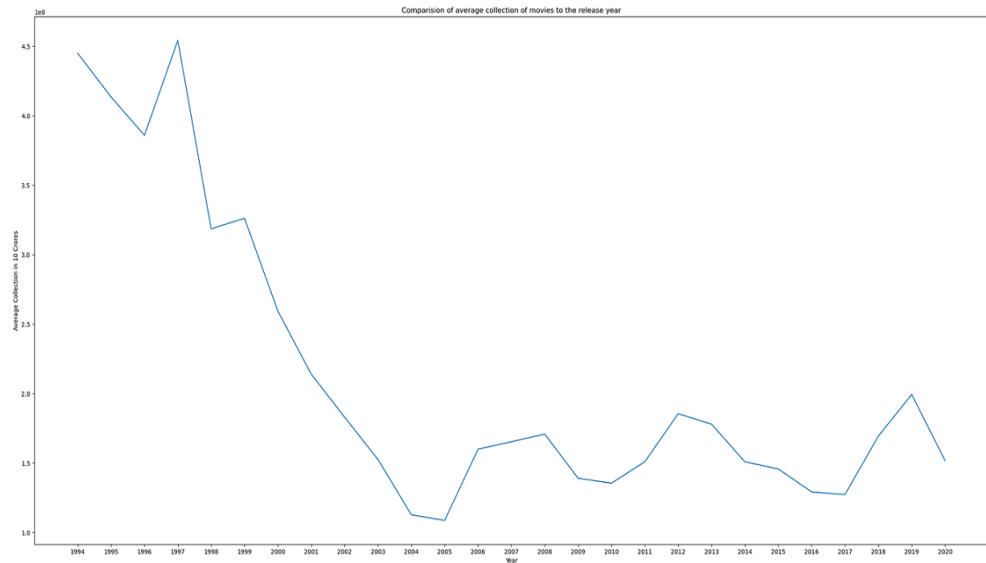
1. An interesting result was found when comparing the footfalls monthly for each year. It was found that for most of the years from 1993 to 2020 the footfalls count and total collection was highest in the month in the second half of the year. Exact list of months for each year can be found in the file month-footfalls.csv and month-total-collection.csv.
2. Drama is the most successful genre of Indian cinema based on total collection on the boxoffice by a huge margin. Action and comedy are 2nd and 3rd respectively. Whereas masala is the genre with best average collection per

movie. This explains the lack of sci-fi, animated and mythological movies in India.

3. For the last 13 year a total of 639 drama, 284 comedy, 212 thriller and 127 action movies have been made. On the other hand only 13 fantasy, 3 animation and 1 documentary have been made. The following figure shows the comparison of revenue made by different genres.

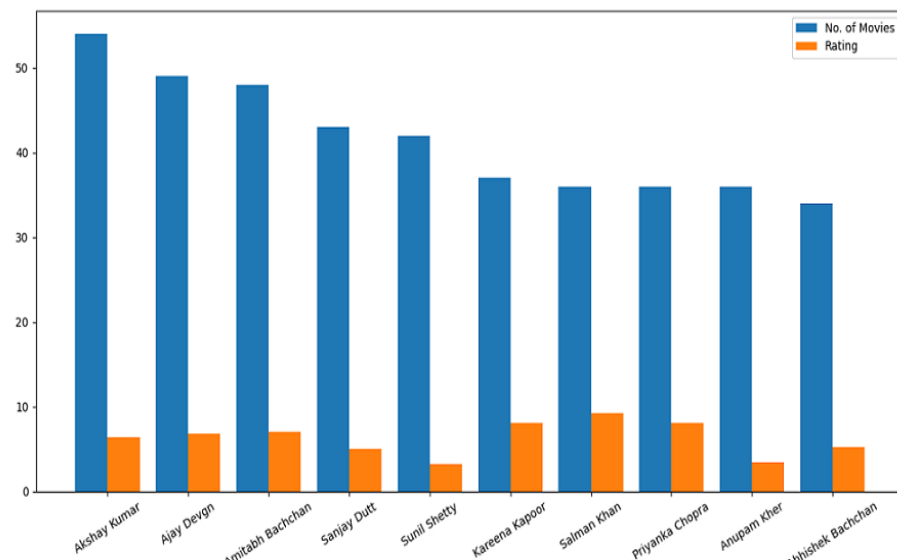


4. Surprisingly average collection by movies for past 27 years was best in 1997. Which shows that though a single movie might be making more money at the boxoffice but, many movies released nowadays are not doing well. As can be seen in the figure.

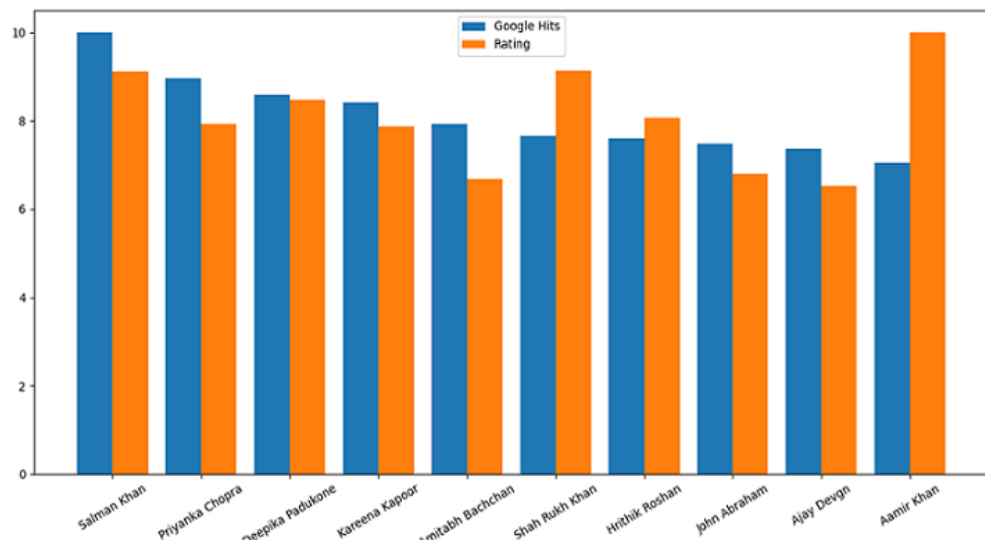


5.3 Crew

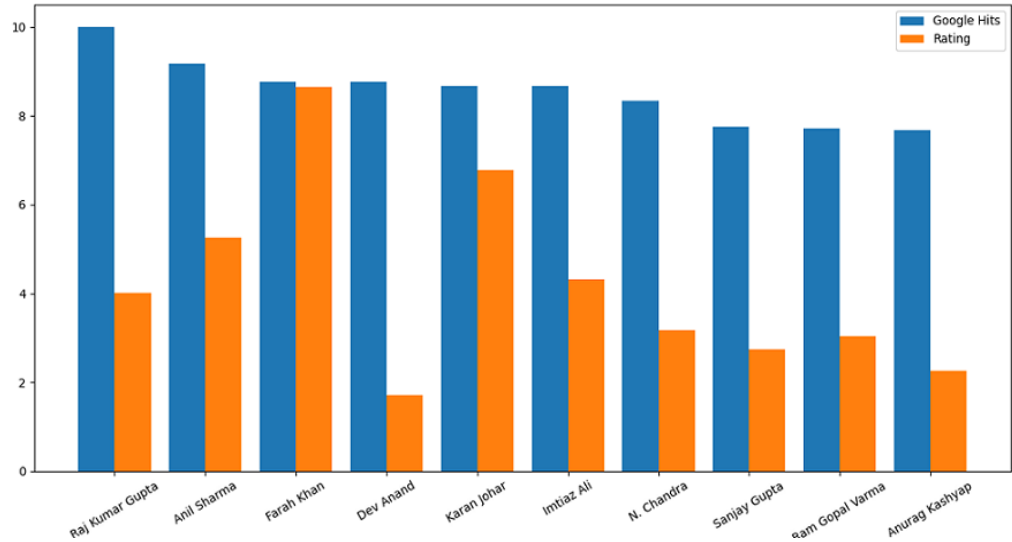
1. Based on the most revenue made from all movies done together 'Salman Khan' and 'Ali Abbas Zaffar' is best Actor-Director team, they have made a total of more than Rs.837 crores. 'Rajkumar Hirani' and 'Vidhu Vinod Chopra' is the best Director-Producer team and 'Shahrukh Khan' and his wife 'Gauri Khan' is the best Actor-Producer team.
2. 'Akshay Kumar' has worked in the most number of movies as Lead Actor with 54 and has 13 movies in the 100 crore club. 'Kareena Kapoor' has worked in 37 movies as the Lead Actress.



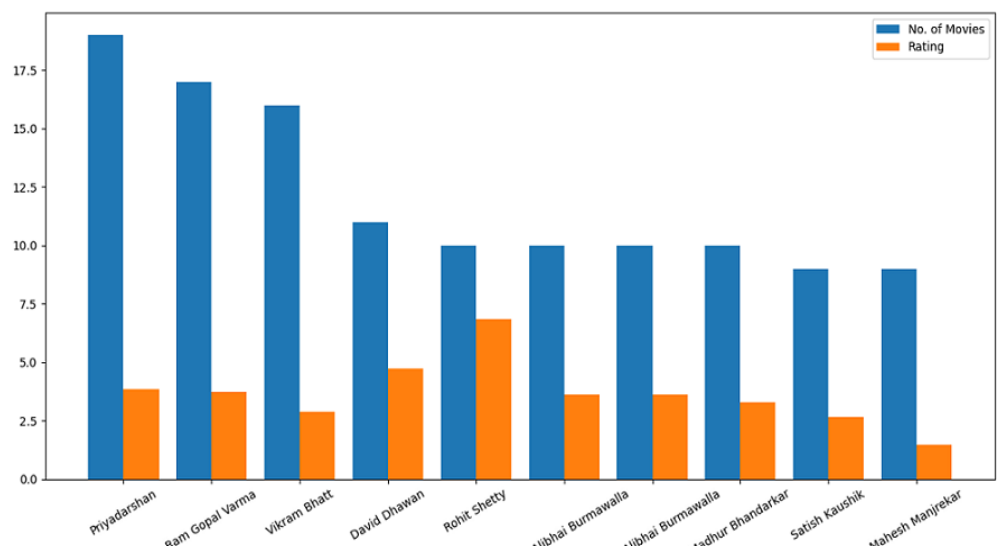
3. 'Shakti Kapoor' has worked in 281 movies which is the most by any actor. Other actor like 'Aruna Irani', 'Anupam Kher', 'Kader Khan' have also done 200+ movies. This is a milestone reached by only a few actor in the industry. actor-movierating.csv also gives the highest rating ever achieved by these actor's movies.
4. When compared on the basis of imdb rating and boxoffice collection 'Akshay Kumar' turns out to be the most successful actor, closely followed by 'Aamir Khan', 'Salman Khan' and 'Shahrukh Khan'.
5. Most number of all time blockbusters have been delivered by 'Aamir Khan' with a total of 5. Most number of flop and disaster movies have been delivered by 'Mithun Chakraborty' and 'Shakti Kapoor' respectively. actors-movies-verdict.csv gives more insights on this.
6. 'Salman Khan' is the most popular actor based on google hits. He also has the most number of tickets sold for his movies. He is followed by 'Priyanka Chopra', 'Deepika Padukone' and 'Kareena Kapoor' on popularity scale. 'Aamir Khan', who has the best rating is number 10th in case of popularity.



7. 'Saif Ali Khan' was the most successful actor for the year 2020.
8. 'Raj Kumar Gupta' is the most popular director based on google hits. 'Farah Khan' is the 3rd most popular director and has a rating of 8.65 which is the best among top 10.



9. 'Rajkumar Hirani' is the most successful director based on imdb ratings and revenue made. He has 2 all time blockbusters which is the most by any director in that category and also has the most number of movies in 300 crore club. 'Rohit Shetty' is the second most successful director with most number of blockbusters, 4 and has most number of movies in 200 and 100 crore club. Priyadarshan is the most featured director with 19 movies.
10. 'Hrishikesh Mukherjee' has worked in 41 movies with a average rating of 7.3. 'David Dhawan' has worked in 39 movies and has the highest footfalls of all time combined. He has almost twice the number of tickets sold for his movies than any other director.



11. 'Vidhu Vinod Chopra' has produced 2 all time blockbusters which is the same number of all time blockbusters directed by 'Rajkumar Hirani'. They also make the most successful director-producer team.
12. 'Aditya Chopra' is the most successful producer on boxoffice and has 2 movies in the 300 crore club

Other results like crews' debut year, debut movie, genre most worked in, can also be found in the output files.

6 Conclusion

Analyzing bollywood movies and stars using various grading points, this project provided insights into Indian cinema. Many interesting facts were discovered through this analysis. Comparing movies, genres, cast and crew helps in better judgement of which type of movies are better suited for Indian audience. Drama turned out to be the most earning genre for the past 13 years, this shows why there is a lack of movies in other genre. Other facts which are probably known, like Akshay Kumar is the highest grossing actor, Aamir Khan is the highest rated actor and so on were also found.

But what makes this analysis interesting is that some very surprising facts were also discovered. 'Hum Aapke Hain Koun..!' a 1990s movie is the best movie taking into consideration imdb rating and revenue made by the movie. One very interesting fact, that generally, most revenue is generated during the second half of the year. This shows that more people are watching movies in the theaters during the second half of the year and also that most popular actors'/actresses' movies are released during this time period. This can be due to the holiday season in India that start during the months of September-October due to Diwali. It was also found that movies released during the holidays make more revenue as compared to movies released on normal days which also strengthen the fact stated above.

This research of bollywood data can be really helpful in recommending better options to the right audience. Analyzing movies and generating better recommendations in this world with abundance of data can really help save time in finding the right movie. We also made this demo recommending system without any ML algorithm just to give view how recommendation system uses these dataset to predict movies.

The present scenario has established that films are no longer being produced in the conventional manner they used to be. A great degree of rationale and scrutiny is invested behind determining the type of movie to be created, who should create it as well as how long it is supposed to be. This analysis can also help movie makers better understand the audience and produce better movies with the right cast and story. This analytics can aid producers, production firms, and executives in determining accurate decision-making, predicting trends, and in comprehending the preferences of the viewers in an advanced manner.

Analyzing data to extract out important information, especially in the field of movies and in this modern era helps people save time and money. Time and money are two of the most scarce resource in a common person's life which one

wants to invest in a way one enjoys the most. Thus, by providing them with recommendations better suited for them, our project helps use these two resources in that way.