

Does Sentiment effect the Gross income of a movie?

Krishma

Dept. Computer Science & Technology (M.Tech)
Guru Nanak Dev University
Amritsar

Amit Chhabra

Dept. Computer Science & Technology
Guru Nanak Dev University
Amritsar

Abstract: - Now days millions of people are using social network sites and through these social sites they can express their opinion, emotions and other advices. Now days growing of social media most beneficial for many business, educational, box-office movie and so on. In this paper we discuss about Hollywood movies data of previous two years 2014 and 2015. This dataset of movie divides into two types of features: first is the conventional feature and second one is social media features. In this number of attributes such as budget, genre, gross, views and comments, sentiments affect the popularity of a movie. Some social media engine such as Twitter, YouTube etc. are main platforms where public can share their opinions about the movies.

Keywords: Social Media, Sentiments, Gross Income, Hollywood Movie's.

I. INTRODUCTION

In success of movie box office, social media plays a very most important role. Because people too much involved in social network such as Twitter, Facebook, YouTube etc. [12] Prediction of achievement in business has been of great interest to the economists and financial professionals.[9] Today social media has shown its analytical control in several domains, which inspires us to exploit social media content to predict box-office revenues. With initiation of data analytics, the prediction method has been made intelligent by considering the historical data and employing various data analytical techniques to infer the future events. The movie commerce international produces a huge widespread variation of movies per year. However, very few movies are a success and are ranked high. Given the low success rate, modes to predict reliably the box office collections of a movie can contribution by way of refining the occupational significantly and growth usual returns. The common variation of cinemas produced per 12 months is increased than a thousand So, to make the film profitable, it goes into a material of condition that the movie succeeds. [4] In the skill of Web 2.0, clients share their rankings or response simply with additional people after spotting a movie. User grade simplified the structure which customers prompt their opinions about a product and is a attractive indicator to predict the container office. More than 4000 movies produced exclusive one year in the whole world. [14] The task of predicting Box-Office revenues for movies the usage of the chatter from Twitter, one of the fastest emerging social networks in the Internet. Twitter, a micro blogging website, now plays an important role in the research of social network [14]. People share their preferences on Twitter using free-format, limited-length texts, and these texts (often called "tweets") provide rich information for companies/institutes who want to know about whether people like a certain product, movie, or service. We have intensive on movies in this learning for two key reasons.

The issue of movies is of significant interest amongst the social media being community, categorized individually via massive variation of clients discussing movies.

- i. The real-world moments can be simply placed from box-office income for movies.
- ii. Movie producers spend a lot of effort and currency in publicizing their movies and have furthermore embraced the Twitter medium for this purpose.

Here, a we said about how sentiments are created, how great and dreadful opinions propagate and how they influence people. For a bad movie, the opening reviews possibly will be sufficient to discourage others from perceiving it, although on the additional, it is reasonable for concentration to be generated via earnings fantastic calculations and opinions over time. We got our dataset from the Internet Movie Database (IMDb). Hollywood is a multi-billion-dollar industry which releases more than a hundred films a year, with large variations in the budgets and box office grosses of the movies. Identifying which factors are important to a movie's profitability and subsequently predicting the success of a movie given its relevant parameters could save movie studios hundreds of millions of dollars a year. Some social media engines such as Twitter, YouTube etc. are important structures the place people can share their opinions about the movies [13]. Most of the studies done for prediction of cinemas success use predictable attributes, grown from on line movies databases. However, with overview of social media, public opinion has been harnessed about a series of dual events/entities from panels such as YouTube and Twitter [12]. Similarly, for movies, social media web sites have contributed a remarkable quantity to the credit of movies. Now absolutely everybody can review, rate, comment or share their opinions about a movie online. Thus, social media plays an important role in predicting the success of a movie.[10] With the development of various social media stations, electronic word-of-mouth (eWOM) serves as a reliable foundation of data for movie customers. Several revisions on the effect of eWOM on sales have been conducted due to the production of eWOM on social media. eWOM Traditional word-of-mouth (WOM) is defined as informal, Person to person communication between a supposed non-commercial communicator and a receiver about a brand, a product, an organization, or a service. This has developed into electronic Word-Of-Mouth (eWOM) through the growing occurrence of the Internet and may be defined as any +ve/-ve statement made by likely, real, and former customs about a product or a company via the Internet. The impacts of blogs and YouTube on box office revenues were

consistent through the initial stage until the late stage of opening. So, they represent mass media and interpersonal communication. Blogs have characteristics of interpersonal communication with strong persuasive effects; users can frequently visit blogs by searching for the desired information. [11] In recent years, social media has played a huge role in how we share and communicate our thoughts and opinions. This information can be very valuable for companies and governments as it can be used to analyse public mood and opinion which is a very powerful tool. Specifically, it uses chatter from Twitter to predict box office revenue of movies by extracting features such as tweets and their sentiments.

II. RELATED WORK

[1] Apala, Krushikanth R. et al. in 2013 "Prediction of movies box office performance using social media." In this paper the box office movie presentation of the cinemas was characterized by issues resultant from social media and IMDb. In this sentiment analysis, twitter, data mining, you tube and IMDb keywords are used. In this work of research create a predictive model involving of genre of successful films classified by user scores of the IMDb movie database, via Twitter's followers total, popularity of box office denoted by the number of comments and views of official movie trailers accessible by YouTube, and sentiment to a movie derived from YouTube viewers' comments.

[2] Chintagunta, Pradeep K. et al. in 2010 "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets." In this work online word of mouth, new product release, motion pictures etc keywords are used. The general technique of moments-based approach to computing the impression of the valence, volume, and variance of online user ratings from Yahoo! Movies on box office presentation while accounting for together these forms of biases.

[3] Dooms, Simon et al. in 2013 "Movietweetings: a movie rating dataset collected from twitter." In this work dataset used by authors the MovieTweatings dataset in which collect automatically from structured social media posts (i.e., Twitter). In this paper the number of ratings per user and the cumulative percentage of the total amount of ratings. In this paper, tap into the huge accessibility of social media and paradigm a new movie rating dataset 'MovieTweatings' based on public and well-structured tweets.

[4] Ping-Yu Hsu, Yuan-Hong Shen et al. 2014 "Predicting Movies User Ratings with IMDb Attributes." The idea to integrate classical and social media factors to improve the prediction accuracy of the movie success. They collected classical attributes (genre, budget etc.) from IMDb and social attributes (Tweets, views) from social websites like YouTube, Twitter. IMDb is the largest movie database in the world. The website had 2.8 million titles (includes episodes) and 5.9 million personalities in its database on May 2014. In this dataset of movie different types of attributes like movie name, gross, sentiments, likes and dislikes etc. User can rate one movie as many times as they want but each rating overwrites the previous rating for the same movie. The rating shown in IMDb is not an average rating of the original data by every voting user but a kind of weighted average of an undisclosed calculation method.

[5] Lu Yafeng et al. 2014 "Business intelligence from social media: A study from the vast box office challenge." In this research the application of visual analytics for social media analysis has verified comparatively effective. There are still various tasks in applying this to all domains of business intelligence. Box office movie estimates effort fine as one can track the usefulness of Ad campaigns. As the analysis from Twitter it becomes difficult to choose effective keywords. Due to the ever-changing stream of social media sources and users, it is likely that any automated system for data collection and prediction will ultimately be directed off course. Here social media, box office, visualization and prediction keywords are used.

[7] Karniouchina et al. 2011 "Impact of star and movie buzz on motion picture distribution and box office revenue." In this index terms star buzz, movie buzz, cyber buzz, motion pictures etc are used. This study donates to research on the impression that consumes buzz on motion picture delivery & box office achievement by examining the effect of buzz generated about the individual stars & about the movie itself.

[9] Liu, Ting, Xiao Ding et al. 2016 "Predicting movie Box-office revenues by exploiting large-scale social media content." In this study both linear and non-linear regression models, which are based on the crowd wisdom of social media, especially the posts of users, to predict movie box-office revenues. More specifically, the attention and popularity of the movie, purchase intention of users, and comments of users are automatically mined from social media data. In our model, the use of Linear Regression and Support Vector Regression in predicting the box-office revenue of a movie before its theatrical release is explored. To evaluate the effectiveness of the proposed approach, a cross-validation experiment is conducted. The experimental results show that large-scale social media content is correlated with movie box-office revenues and that the purchase intention of users can lead to more accurate movie box office revenue predictions. Both the linear and non-linear prediction models have the advantage of predicting movie grosses in our experiments. Keywords are Movie box-office revenue Social media Prediction Purchase intention mining

[12] Mehreen Ahmed et al. in 2015 "Using Crowd-source based features from social media and Conventional features to predict the movies popularity". This paper presents the contrast of Conventional Features with Social Media features in defining the popularity of movies. In this paper researches presented that social media features such as Sentiment Score of tweets related to movies, Number of Opinions and Comments of movies' promos on YouTube and supporter following on twitter can helpfully be utilized to predict the popularity of movie. In this paper presents the contrast of Conventional Features with Social Media features in defining the status of movies.

[17] Subramaniaswamy et al. in 2017 "Predicting movie box office success using multiple regression and SVM." This paper studies the effectiveness of using multiple linear regression and Support Vector Machine Classification to predict the box-office success of movies, although analysing the impact of variables like trailer views, search engine views, rankings and time of release.

In this paper Data Mining, Box-Office Gross, Regression; SVM Classification; Machine Learning and Data Analytics key terms used.

[18] Huiyu Bi et. Al in 2016 "Classify user's rating based on IMDB data." In this research of paper authors used machine learning methods to classify a user's rating for a movie into -ve and +ve sentiment based on comments for movie. In this paper Machine learning techniques Logistic Regression, Random Forest, XGBoost used. SVM best classification model. In this more features by discovering other structures i.e. movie's box office & the production country which related to the movie rating. For example: some directors are good at certain movie genre, so the combination of director and genre could influence the movie quality and movie rating. In machine learning problem, features are important for the presentation of the model. They are not only representative the points but also helping us understanding the context of problem. In this paper key-words Sentiment Analysis, Logistic Regression, Random Forest, XGBoost, RBF SVM, IMDb are used.

[19] Ritu Khandelwal et Al "Comparative Analysis for Prediction of Success of Bollywood Movie" in 2019. In this paper different types are used for the prediction such as AdaBoost, Naïve Bayes, decision tree, KNN, and Logistic Regression to get the greatest correct predicted results. The results are likened on the basis of several methods of sampling from test and score. Based on Cross-Validation with 2 folds, Decision tree gives the most accurate results. Because of Random sampling, KNN gives the most accurate results. Based on Leave one out, the most accurate results are given by Decision tree algorithm. In this paper keywords decision tree, data mining, Prediction and Orange are used.

[21] Sarlan et Al "Twitter sentiment analysis" in 2014. In this research paper sentiment analysis report design in which huge Amount of tweets. In this paper user's perception through tweets into +ve tweets and -ve tweets i.e. represents in the form of pie chart. In this paper keywords Twitter, Sentiments, opinion mining, social media networks, NLP.

[25] Krauss et. Al. "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis" In 2008. In this paper familiarizes a new Web mining method that associations social network analysis and automatic sentiment analysis. This paper experiments analysing online forum thoughts on the Internet movie database (IMDb) by exploratory the correlation of the social network structure. In this paper keyword used Trend Prediction, Social Network Analysis, Online Forum, Internet Movie Database, Oscar Awards etc. This paper experiment shows that a high strength of argument about a certain movie at IMDb is a strong pointer of success of that movie at the box office.

III. PROPOSED METHODOLOGY

1. Data collection: The dataset must include parameters such as Actor, etc. of past Bollywood movies.

The first step is Data Collection. we collect the data from UCI dataset. The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. The IMDb (Internet Movie Database) is an online database of information related to films, video games, television programs, fictional characters, biographies, plot summaries and reviews. IMDb is the world's most popular and authoritative source for box office movie content. Find ratings and reviews for the movie and TV shows. Data Collector is the major module as it retrieves information about movies from diverse sources including movies web sites i.e. IMDB and social media including YouTube and Twitter.

1.1. Feature extraction: - In our dataset collected data of 15 features for each movie. The feature can further divide into conventional features and social media features. Here conventional features are those feature that are available in movie resources like IMDb.

1.1.1 Conventional features are following discussed: -

- Budget
- Gross
- Genre
- Sequel
- Rating
- Screens

Budget defines as total production cost of movie.

Gross: Gross like Earning of a film from all revenue sources.

Genre: genre defines as Adventure, Action, Drama, sequel as Film, theatre that continues the story or expands upon, some earlier work. In dataset 15 values available those are mapped into numeric values like 1-15.

Screens: total number of screens in cinemas etc.

1.1.2 Social media features as follows: -

- Total Number of views
- Total Number of likes
- Total Number of dislikes
- Total Number of comments
- Total Number of Aggregate followers
- Total sentiment score

Number of Views and Comments: The number of views and comments of promo of movies on YouTube etc social media sites calculated.

Number of likes and dislikes: Number of Likes and Dislikes of trailers on YouTube are considered. Its like as same as number of views and comments of social media.

Total Number of Aggregate followers: Here number of followers are used social media.

Sentiment score: sentiment score defined the opinion of users about a movie. These sentiments are represented by positive (+Ve) and negative (-Ve). It is calculated as below:

$$\text{Sentiment (M}_i\text{)} = S_1, S_2, S_3, \dots, S_n$$

Here, $S_1, S_2, S_3, \dots, S_n$ are total number of tweets of sentiment score. Sentiment (M_i) represents sentiments score of a movie.

2. Pre-processing:

After data transformation, we found there are some NA's in the numeric variable's "budget", "screens" and "aggregate followers".

2.1 "Budget" variable: Since there are only one NA's in it and it is not reasonable to replace it with the mean or the median, we delete the row. Because we can't analyse the budget of this movie of year 2014 according to another movies budget. So, it removes from dataset.

2.2 "Screens" variable: In our dataset screens variable have 10 missing values. These missing values are not affected to gross and budget of movies in year 2014. So, it's removed from dataset.

2.3 "Aggregate Followers" variable: we remove these values because aggregate followers are valued in only social media networks. In box-office movie have doesn't matter about followers because in this movie box office. Because in this paper we discussed more than social media networks like YouTube, Twitter etc.

Table 1. show all about of data in which different attributes shows their different factors. In above table show's that their attributes minimum values, maximum values, average values and median. This table used for simply understand the dataset of movies. In our dataset have 15 different attributes in which some attributes are conventional features and other's attributes are social media features.

Table 1: Attributes/Parameter under study

S. No.	Keyword	Description	Min	Max	Average	Median
1	Gross	Earning of a film from all revenue sources	2470	643000000	77234127.13	46650000
2	Budget	Total production cost	7000	250000000	53584562.66	30000000
3	Genre	Adventure, Action, Drama	1	15	5.255319	3
4	Rating	IMDb rating	3.1	8.7	6.431915	6.5
5	Comments	Something said about a topic	0	38363	1821.495	883.5
6	Likes	liking something	1	370552	12941.98	6734
7	Dislikes	something is unpleasant and not like	0	13960	714.8989	373.5
8	Screens	total number of screens in cinemas	2	4324	2341.33	2875
9	Sentiment	View or opinion. It can also mean an emotion	-38	29	2.829787	0
10	Sequel	Film, theatre that continues the story or expands upon, some earlier work	1	7	1.43617	1

IV. RESULT AND DISCUSSION

In fig.1 consists of highest ten entries of budget of both 2014 and 2015 year. In both year's in both year's highest budget was 250000000 and lowest budget of 2014 was

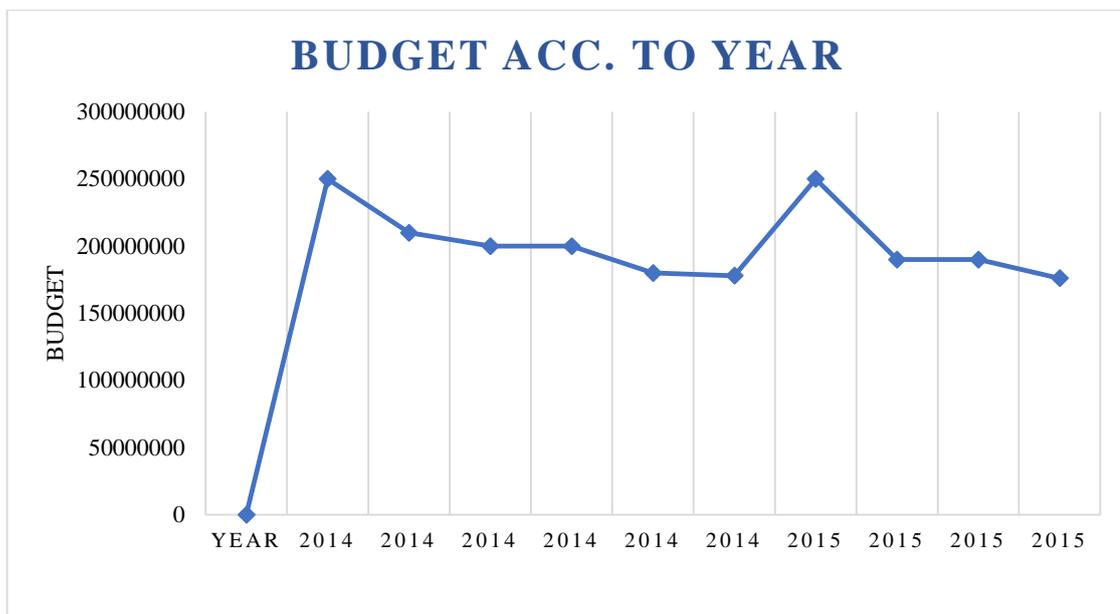


Fig 1. Highest budget of year 2014 & 2015

70000 and other side 2015 lowest budget was 700000.



Fig 2. Top Ten Rating according to Sequel of movie's

In fig.2 discuss that sequel doesn't affect to rating of a movie because here highest rating 8.7 of 2014 and 8.6 of 2015 and lowest rating of both years like 3.1 and 4. Another side of sequel highest sequel 7 and lowest sequel 1. But in this fig show that highest sequel 7 have rating like 7.4, 8.1. this means sequel doesn't matter in rating system of movie.

In fig.3 The aim of movie makers is to provide audience two to three hours of great entertainment. In addition to this they also aim to make good profit. This paper examines that does positive or negative sentiment of people towards a movie shared on social media websites affects its gross income? Well, the results depicted shows that it is not so significant factor. Instead budget is the most significant factor affecting a movie's gross income, followed by screens on which movie was screened and if the movie is a sequel or a prequel of an earlier.

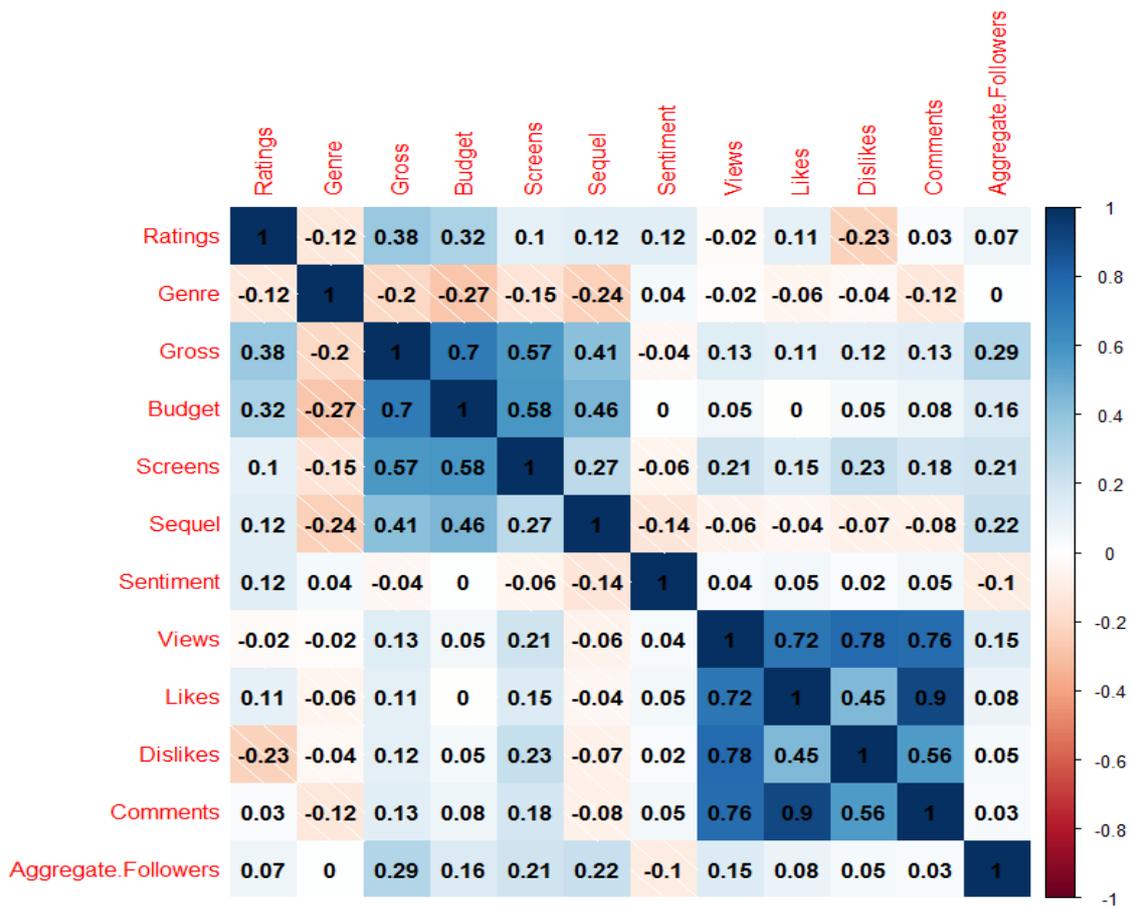


Fig. 3. Co-relation on different factors

movie. Further factors like expert ratings and aggregate online social media followers also affect the gross income of a movie.

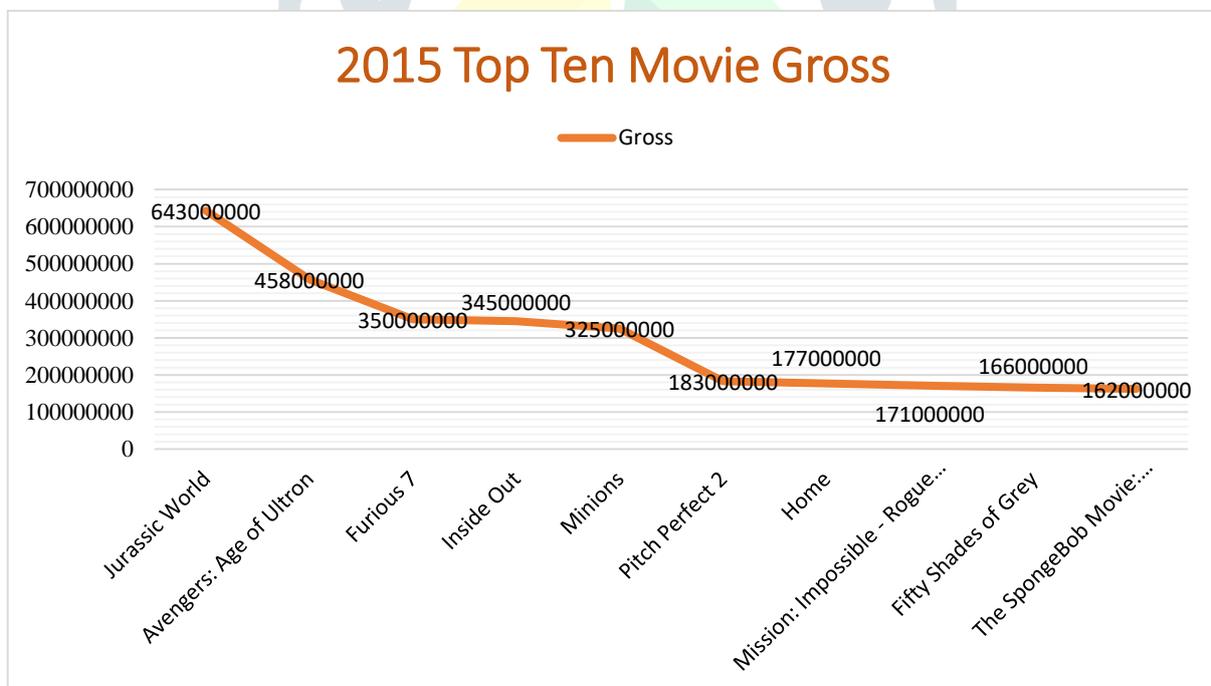


Fig. 4. Top Ten gross list of 2015 movie's

In fig 4. Shows the gross income of top ten movies of 2015. In which highest gross of movie Jurassic World with 643000000 Gross and this movie budget is 150000000. Like this others movie have their own budget and gross.

Movie	Year	Ratings	Genre	Gross	Budget	Screens	Sequel	Sentiment	Views	Likes	Dislikes	Comments	Aggregate Followers
1 Jurassic World	2015	7.3	1	643000000	150000000	4274	4	1	9143740	34746	1074	5107	6180000
2 Avengers: Age of Ultron	2015	7.8	1	458000000	250000000	4276	2	21	10366624	31552	989	3843	10070000
3 Furious 7	2015	7.4	1	350000000	190000000	4004	7	-17	59056	330	8	39	11890000
4 Inside Out	2015	8.6	12	345000000	175000000	3946	1	2	1438926	4632	262	496	232000
5 Minions	2015	6.6	1	325000000	74000000	4301	2	-2	1341909	1607	764	48	250000
6 Pitch Perfect 2	2015	6.7	8	183000000	29000000	3473	2	1	9214467	39824	998	1987	7336000
7 Home	2015	6.7	8	177000000	135000000	3708	1	-2	10341783	24413	1675	3426	5070000
8 Mission: Impossible - Rogue Nation	2015	7.8	3	171000000	150000000	3956	5	3	8748596	20352	649	1842	6605000
9 Fifty Shades of Grey	2015	4.2	3	166000000	40000000	3646	1	9	31859569	49900	13960	9119	946000
10 The SpongeBob Movie: Sponge Out of Water	2015	6.1	8	162000000	74000000	3641	2	0	5536822	29411	1840	1281	184000
11 San Andreas	2015	6.4	1	154000000	110000000	3777	1	3	12632836	36508	2210	7559	9737600
12 Mad Max: Fury Road	2015	8.3	1	153000000	150000000	3702	4	-4	2732371	13030	497	1774	768700
13 Straight Outta Compton	2015	8.3	9	135000000	28000000	2757	1	5	848970	12607	237	1560	55618
14 Insurgent	2015	6.4	2	130000000	110000000	3875	2	-6	2834800	5664	5746	66	4240000
15 Spy	2015	7.3	1	110000000	65000000	3711	1	0	6649290	20750	750	1666	265000
16 Trainwreck	2015	6.7	12	105000000	35000000	3158	1	1	9511288	19903	2581	2955	2014000
17 Tomorrowland	2015	6.6	12	93200000	190000000	3972	1	1	999867	4212	66	250	1198000
18 Get Hard	2015	6.1	8	90400000	40000000	3175	1	6	10078326	26565	1418	2395	2939000
19 Terminator Genisys	2015	6.8	1	89400000	155000000	3758	5	1	84870	265	13	63	3877901
20 Ted 2	2015	6.6	8	81300000	68000000	3442	2	5	6711914	29903	984	1767	10988000
21 Paul Blart: Mall Cop 2	2015	4.3	1	71000000	30000000	90	2	15	2389347	8809	935	892	1618000
22 Pixels	2015	5.6	1	71000000	88000000	3723	1	10	5340100	26134	2007	3717	2466000
23 Magic Mike XXL	2015	6.2	8	65900000	14800000	3355	2	7	5128288	18475	858	1579	8392000
24 The Wedding Ringer	2015	6.7	8	64500000	23000000	3003	1	-8	4826940	10521	478	755	2284000
25 Vacation	2015	6.3	2	54900000	31000000	3411	5	13	2554327	10062	464	871	2347000
26 Fantastic Four	2015	4	1	52700000	12000000	3995	1	20	7560211	24168	3524	7139	881000
27 Poltergeist	2015	5	15	47400000	35000000	3240	1	-5	3651828	13998	969	2205	1066

Fig 5. Top ten 2015 movie’s gross v/s sentiment

Fig 5. shows that movie gross not fully-dependent on sentiment’s because in this fig shows highest gross of movie’s have positive sentiment’s as well as negative sentiments. So, we can say that sentiment not affected the gross of movie. If we refers the sentiments for estimates the grossing of movie then not exact values are there because above example we see that if movie no. 4 furious 7 top 3 grossing movie of year 2015 but another side if we see this movie sentiments then its have negative sentiments. So we finally say that not only sentiments those effected to the gross of any movie in movie industry.

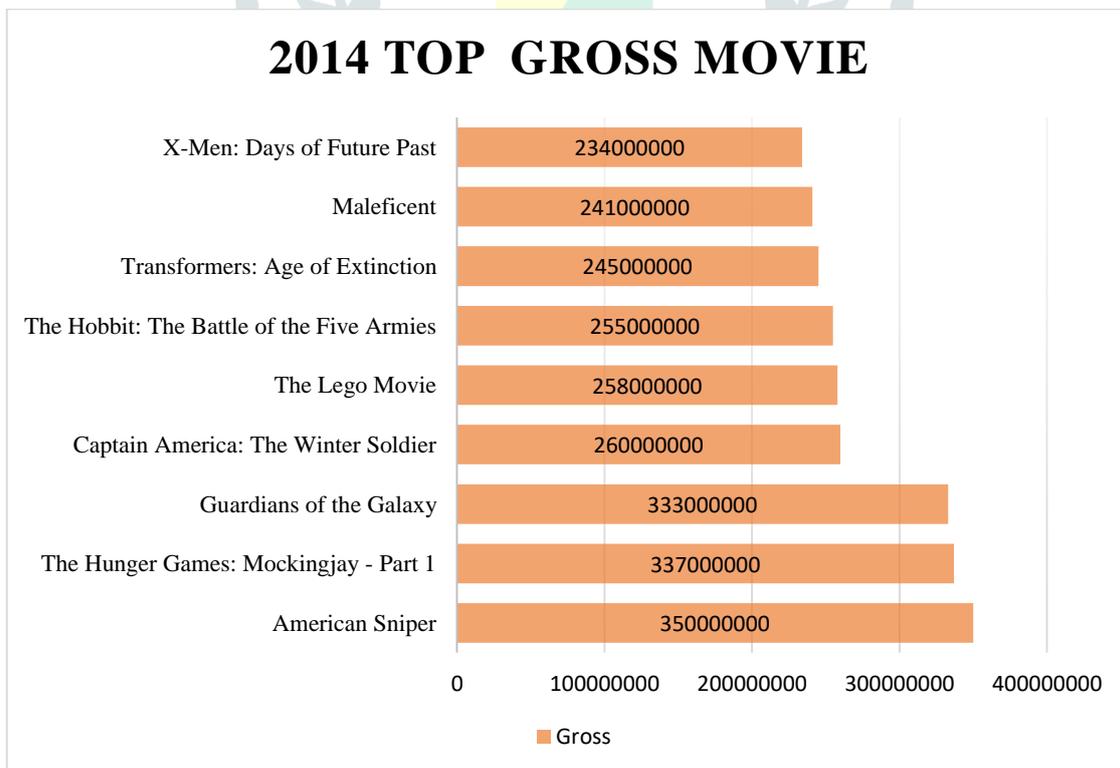


Fig. 6. Top gross list of 2014 movie’s

Fig 6. Show same as fig 4. Because in fig no. 4 shows the top ten movie's gross of year 2015 and another side fig 6. Shows the top ten grossing movies according to year 2014.

V. CONCLUSION

This paper presents the comparison of social media features and conventional features. In this paper shows that in previous year of work only views effected on movie gross income but in this era of social media sentiments also effected to the gross of movie. Other side we can say that not only single attribute like sentiments effected the grossing of movie. For success of movie box office sequel, views, likes etc played main role in success of any movie. In our work discussion about co-relation on different factors.

REFERENCES

- [1] Apala, Krushikanth R., Merin Jose, Supreme Motnam, C-C. Chan, Kathy J. Liszka, and Federico de Gregorio. "Prediction of movies box office performance using social media." In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1209-1214. ACM, 2013.
- [2] Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman. "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets." *Marketing Science* 29, no. 5 (2010): 944-957.
- [3] Dooms, Simon, Toon De Pessemier, and Luc Martens. "Movietweetings: a movie rating dataset collected from twitter." In *Workshop on Crowdsourcing and human computation for recommender systems, CrowdRec at RecSys*, vol. 2013, p. 43. 2013.
- [4]] Oh, Chong, Yaman Roumani, Joseph K. Nwankpa, and Han-Fen Hu. "Beyond likes and tweets: Consumer engagement behavior and movie box office in social media." *Information & Management* 54, no. 1 (2017): 25-37.
- [5] Lu, Yafeng, Feng Wang, and Ross Maciejewski. "Business intelligence from social media: A study from the vast box office challenge." *IEEE computer graphics and applications* 34, no. 5 (2014): 58-69.
- [6] Sharda, Ramesh, and Dursun Delen. "Predicting box-office success of motion pictures with neural networks." *Expert Systems with Applications* 30, no. 2 (2006): 243-254.
- [7] Karniouchina, Ekaterina V. "Impact of star and movie buzz on motion picture distribution and box office revenue." *International Journal of Research in Marketing* 28, no. 1 (2011): 62-74.
- [8] Gopinath, Shyam, Pradeep K. Chintagunta, and Sriram Venkataraman. "Blogs, advertising, and local-market movie box office performance." *Management Science* 59, no. 12 (2013): 2635-2654.
- [9] Liu, Ting, Xiao Ding, Yiheng Chen, Haochen Chen, and Maosheng Guo. "Predicting movie Box-office revenues by exploiting large-scale social media content." *Multimedia Tools and Applications* 75, no. 3 (2016): 1509-1528.
- [10] Baek, Hyunmi, Sehwan Oh, Hee-Dong Yang, and JoongHo Ahn. "Electronic word-of-mouth, box office revenue and social media." *Electronic Commerce Research and Applications* 22 (2017): 13-23.
- [11] Choudhery, Deepankar, and Carson K. Leung. "Social media mining: prediction of box office revenue." In *Proceedings of the 21st International Database Engineering & Applications Symposium*, pp. 20-29. ACM, 2017.
- [12] Ahmed, Mehreen, Maham Jahangir, Hammad Afzal, Awais Majeed, and Imran Siddiqi. "Using Crowd-source based features from social media and Conventional features to predict the movies popularity." In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (Smart City)*, pp. 273-278. IEEE, 2015.
- [13] Predicting Movie Revenue Nikhil Apte, Mats Forssell, Anahita Sidhwa December 16, 2011
- [14] "Predicting the Future with Social Media" Sitaram Asur. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology
- [15] Dhar, Tirtha, Guanghui Sun, and Charles B. Weinberg. "The long-term box office performance of sequel movies." *Marketing Letters* 23, no. 1 (2012): 13-29.
- [16] Kaur Arundeeep, and AP Gurpinder Kaur. "Predicting Movie Success: Review of Existing Literature." *International Journal of Advanced Research in Computer Science and Software Engineering* 3, no. 6 (2013).
- [17] Subramaniaswamy, V., M. Vignesh Vaibhav, R. Vishnu Prasad, and R. Logesh. "Predicting movie box office success using multiple regression and SVM." In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 182-186. IEEE, 2017.
- [18] Imdb report ref Bi, Huiyu, Miao Wang, and Yuan Tian. "Classify user's rating based on IMDB data." (2016).
- [19] Khandelwal, Ritu, and Harshita Virwani. "Comparative Analysis for Prediction of Success of Bollywood Movie." *Available at SSRN 3350907* (2019).

- [20] Nithin, V. R., M. Pranav, and P. B. Sarath Babu. "Lijiya: A predicting movie success based on IMDB data." *Int. J. Data Min. Tech* 3 (2014): 365-368
- [21] Sarlan, Aliza, Chayanit Nadam, and Shuib Basri. "Twitter sentiment analysis." In *Proceedings of the 6th International Conference on Information Technology and Multimedia*, pp. 212-216. IEEE, 2014.
- [22] Kaur, Amandeep, Deepesh Khaneja, Khushboo Vyas, and Ranjit Singh Saini. "SENTIMENT ANALYSIS ON TWITTER USING APACHE SPARK." (2016).
- [23] Pawar, Kishori K., Pukhraj P. Shrishrimal, and R. R. Deshmukh. "Twitter sentiment analysis: A review." *International Journal of Scientific & Engineering Research* 6, no. 4 (2015): 957-964.
- [24] Koschat, Martin A. "The impact of movie reviews on box office: media portfolios and the intermediation of genre." *Journal of Media Economics* 25, no. 1 (2012): 35-53.
- [25] Krauss, Jonas, Stefan Nann, Daniel Simon, Peter A. Gloor, and Kai Fischbach. "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis." In *ECIS*, pp. 2026-2037. 2008.
- [26] Rui, Huaxia, Yizao Liu, and Andrew Whinston. "Whose and what chatter matters? The effect of tweets on movie sales." *Decision Support Systems* 55, no. 4 (2013): 863-870.

