PREDICTING SUCCESS OF BOLLYWOOD MOVIES

Capstone Project

For: Big Data and Analytics Program, Part time batch (PT 04/05)

By: Nikhil Gore, Alok Yadav, Ashish Patwardhan & A.J.R.Vasu

Document Contents Page No

SUMMARY OF THE PROJECT

BUSINESS CONTEXT

Can you predict if my movie will make money? Will it enter the 200 Cr club?

Every director, producer, production house executive, actor and many others concerned would go through sleepless nights before the movie launch; right from pre-production to trailer release to sales numbers coming in. The first weekend of launch is perhaps the most gut crunching time.

Imagine if we could put them at ease by giving them an early indicator of box office success?

The objective of this project is to develop an analytical framework and the associated algorithm to predict box office success.

INTENDED BUSINESS QUESTIONS TO BE ANSWERED

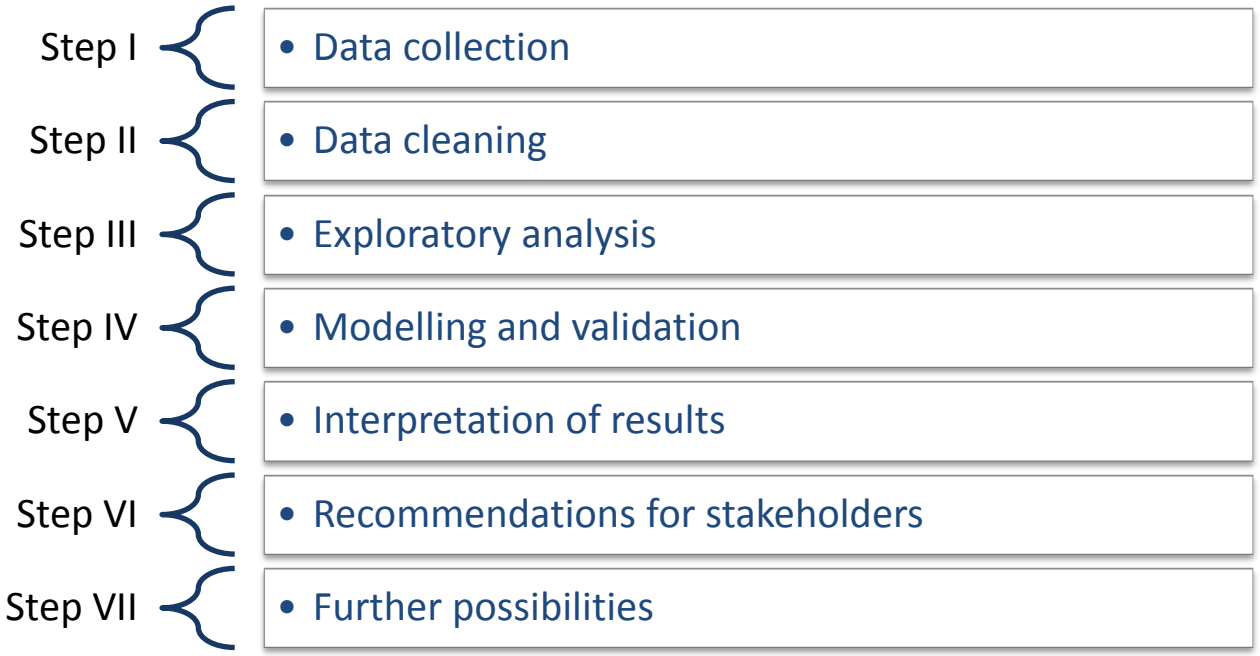
The intended key outcomes are as follows:

1. Likelihood of Box office success
    I. Analytical algorithm to predict success

2. Provide drivers of box office success. Specifically,
    I. What is the lead indicator of success? For example – is it music likeability, director ratings?
    II. What role do super stars play?

3. Insights to help stakeholders create a more effective GTM (go to market) and Intervention plan to drive $ales and success further

PROJECT LIFE CYCLE



Step I

Data collection

Our starting point was to create a data set of Bollywood movies since 2000. The key variables to be collated were movie name, launch date/ year, actor names, director name, music director name, budget for the movie, 1st day, 1st weekend and overall collection in India, if possible global collections. Additionally, we were looking to also collect qualitative inputs like viewer ratings for the movie, popularity of the actor/ director/ music director and many other such variables.

The data sources we have relied on are:

1. BOTY (Best of the year in Bollywood) for most of the information
2. IMDB for Music Director names
3. Wikipedia to verify the movie list

Method used to extract data

As expected, we used web scrapping to extract the primary set of variables. The source website was BOTY (https://bestoftheyear.in/).

We worked with Beautiful Soup package for web scrapping and parsing HTML pages (you can read the documentation here - https://www.crummy.com/software/BeautifulSoup/bs4/doc/) and LXML package (you can read the documentation here - http://lxml.de/)

While Beautiful Soup is very popular, we opted for LXML for the following reasons:

- LXML imports the whole html page as an object and allows use of CSSSELECTORS which makes the process of scraping more structured.
- It forms a tree of all the html tags allowing indexing of multiple tags within a parent tag that makes the process of scraping more target-oriented, efficient and faster.

Sample code

(Refer to Annexure 1 for the full code)

```
# import lxml html parser
import lxml.html

# import css selector
from lxml.cssselect import CSSSelector

# for sending http requests
import requests

# create a list of periods beginning 2000

period =
['2000','2001','2002','2003','2004','2005','2006','2007','2008','2009','2010','2011','2012','2013','2
014','2015','2016','2017']
#period = ['2016']


# declare empty list to which movie name and information web link will be appended. The list of
weblinks will be used
# to collect data for each movie
```

The final output was a list of **943** movies starting from **2000 till 2017**. Please **NOTE** that BOTY does not compile all the movies. However this is still one of the best sources to get a wide assortment of movies.

Step II

Data cleaning (and creation of additional variables)

The table below lists all the variables which have been collected (used as is) and additional variables which have been added, derived and for what purpose.

S. No	Label	Origin	Description	Derivation of additional variables (- / indicates variable value used as is)
1	S. No	-	Serial numbers	-
2	IMDB Id	IMDB	Movie ID from IMDB	-
3	Movie Name	BOTY	Movie name	-
4	Release Date	BOTY	Release date of the movie	-
5	Release Year	BOTY	Year of release	-
6	Actors	BOTY	Movie Star cast	-
7	Super Star	Derived	Presence of Superstar/s in a movie	1= Present, 0=Not Present
8	Multiple Super Stars	Derived	Presence of more than one Superstar in the movie	1= Multiple Superstars, 0= Single Superstar or No Superstar
9	Gender of Super Stars	Derived	Gender of the Superstar/s	1= Male, 2= Female, 3= Both (only when multiple Super stars are present)
10	Super Star /s' Names	Derived	Name of the Superstar/s	Name
11	Super Star Rating	Derived	Rating of the Superstar	Derived based on Delphi method (Scale 1-5; 5 being the highest/ for Superstar rating)
12	Character Star	Derived	Presence of Character star/s in the movie	1= Present, 0=Not Present
13	Multiple Character Stars	Derived	Presence of more than one Character star in the movie	1= Multi Character-stars ,0= Single Character star or No Character star
14	Character Star/s' Names	Derived	Name of the Character star/s	Name
15	Character Star Rating	Derived	rating of the Character-star/s	Based on Delphi method (Scale 1-5; 5 being the highest/ for Superstar, 4/ for Character star)
16	Super Star + Character Star	Derived	Presence of Superstar/s and Character star/s together in the movie	1= Superstar/s + Character star/s , 0= No Superstar/s or Character star/s
17	Overall Star Cast Score	Derived	Star cast rating	• Sum of individual actors' score divided by number of actors. • Actor score – Avg. value of business done by his/her movies
18	Overall Star Cast Score-Delphi Method	Derived	Star cast rating	• Sum of Individual actors' score divided by number of actors. • Actor score – based on Delphi method (5- Super star, 4, 3, 2, 1 – Almost unknown)
19	Director Name	BOTY	Name of the Director/s	-
20	Director Rating	Derived	Director rating	Avg. value of business done by his/her movies
21	Music Director	IMDB	Name of the Music director/s	-

S. No	Label	Origin	Description	Derivation of additional variables (- / indicates variable value used as is)
22	Movie Music Rating	Derived	Music Rating	Based on Delphi method and popularity gleaned from reviews (5 – Almost all songs were a hit, 3 – 1/2 songs were a hit, 1- Not popular at all)
23	BOTY Rating	BOTY	BOTY Rating	-
24	BOTY User rating	BOTY	BOTY User rating	-
25	Genre	BOTY	Genre of the movie	6 prominent Genres were identified (Drama, Action, Comedy, Romance, Thriller and Crime)
26	Budget Unadjusted	BOTY	Movie budget	Amount In INR Cr
27	Budget Adjusted	Derived	Movie budget	• Original budget adjusted for inflation to bring it to year 2013 equivalent (for last 4 years, no adjustment has been done) • 2013 onwards, no inflation factor has been applied with the assumption that the revenue is comparable.
28	Budget Label	Derived	Classification of Movie budget	1 = in the bottom 33% percentile, 2 = Middle 33%, 3 = in the Top 33% of Budget range
29	India BO Un-adjusted Revenue	BOTY	Box office Revenue in India	Amount In INR Cr
30	World BO Un-adjusted Revenue	BOTY	Box office Revenue World wide	Amount In INR Cr
31	Adjusted India BO	BOTY	Inflation adjusted Indian Box office Revenue	Amount In INR Cr
32	Adjusted World BO	BOTY	Inflation adjusted Worldwide Box office Revenue	Amount In INR Cr
33	Inflation Adjustment Factor	Derived	Inflation adjustment factor	Adjusted revenue divided by unadjusted revenue (was derived to update Budgets spent on a movie)
34	ROI	Derived	Return on Investment	Inflation adjusted Worldwide revenue divided by Inflation adjusted budget for the movie
35	BO Performance	Derived	Classification of Box office success	1= Flop, 2= Below Average, 3=Hit, 4=Super-hit, 5= Blockbuster (Based on ROI) , if ROI < 1 =Flop, if 1 <= ROI <= 1.99 = Below Average, if 2<= ROI <= 2.99 = Hit, if 3<= ROI <= 4.99 = Super-hit, and if ROI >= 5 then the movie = Blockbuster (Overall Global average of ROI is around 2)
36	BO Hit_Flop	Derived	Classification of Box office success	Only two groups were derived. Super Hit/ Box office = Hit and Flop/Below Average = Flop

NOTE

Additional variables such as **screenplay quality, uniqueness of the storyline** which can resonate with consumers, past success of the **story writer, marketing budget, number of screens** launched in would have been helpful. However, this is not available in the public domain in an organized manner.

- Based on various relationships to be tested, additional variables need to be created.
- For example, will presence of Multiple Super stars drive higher sales - needs to be understood, then, a new variable which captures this needs to be created.

Step III

Exploratory analysis

This section is divided into 4 parts and the following aspects are covered in each section along with the analysis tools used for each section,

1. Basic Information on Number of movies, actors et al.
   > Data summary and visualization through Tableau

2. Statistical distribution of key variables like Box office revenues and budgets and some initial observations
   > Data summary and visualizations through Seaborn package in Python and Pivot in Microsoft excel

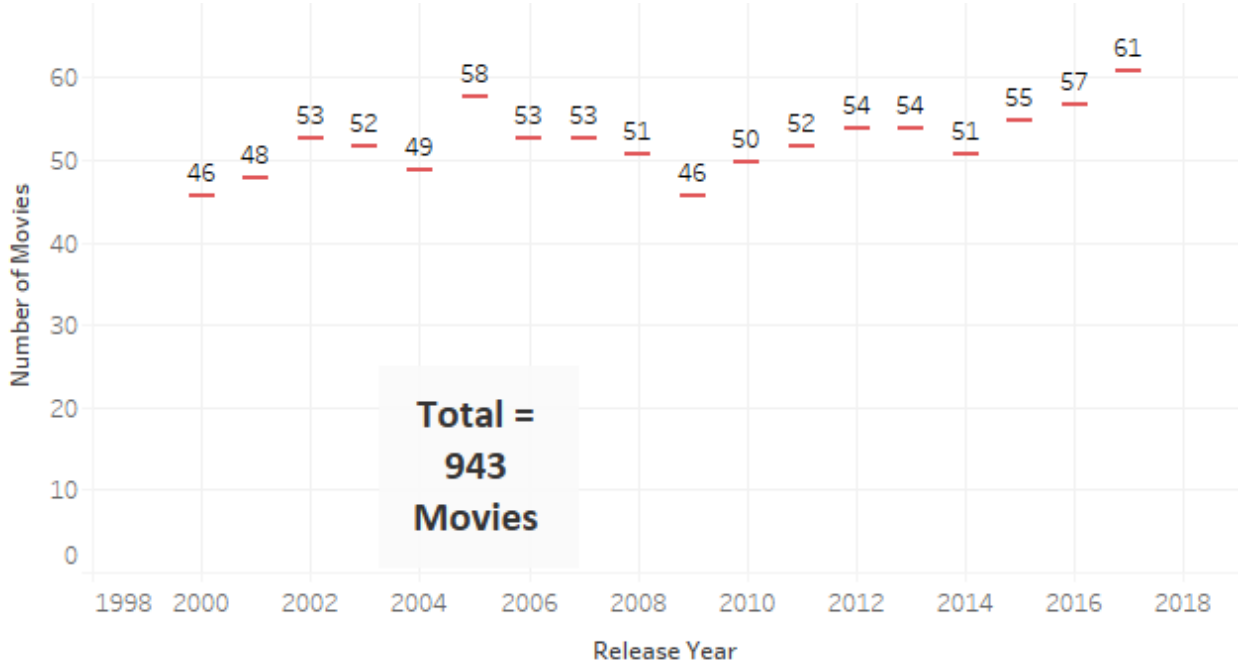
3. Exploring relationships between various variables and Box office success/ revenues and building basic hypothesis
   > Correlations and visualizations using Scipy and Matplotlib packages respectively in Python

4. Understanding what makes big movies flop and small budget movies become successes.
   > Analysing the outliers

Section 1

Basic Information

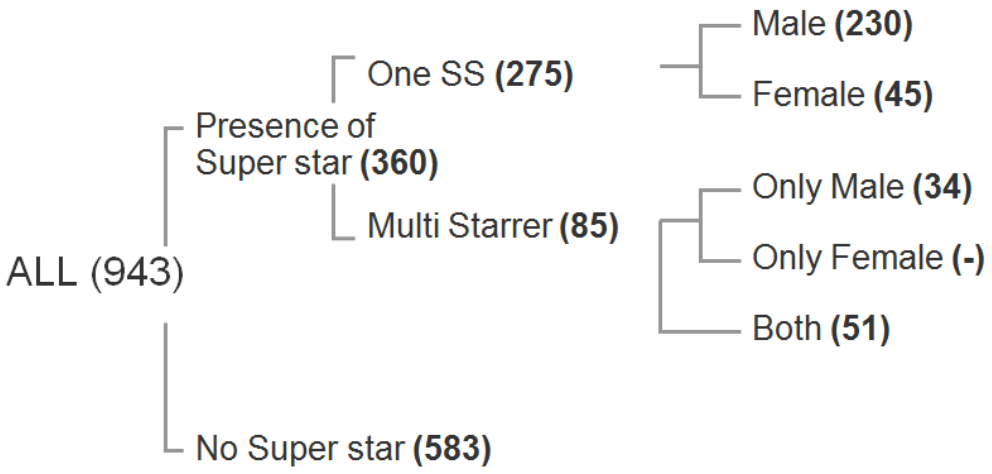
1. **Total Number of movies by year: typically we have about 50-60 movies a year**



2. Total Number of Actors, Directors and Music Directors

Description	Number
All Actors	725
- Super Stars	14
- Popular Character Actors	16
Directors	448
Music Directors	382 (including combinations of music directors working together)

3. Movies with Super stars and some details



Super Stars list

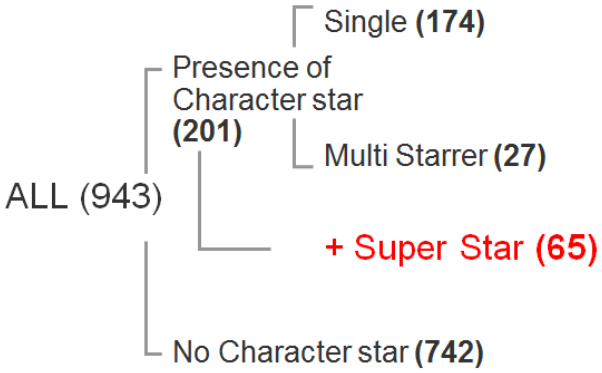
Male Super stars	Female Super stars
Aamir Khan	Aishwarya Rai Bachchan
Ajay Devgn	Deepika Padukone
Akshay Kumar	Katrina Kaif
Amitabh Bachchan	Priyanka Chopra
Hrithik Roshan	
Ranbir Kapoor	
Ranveer Singh	
Salman Khan	
Sanjay Dutt	
Shah Rukh Khan	

- About a third of the movies have Super stars.
- Among these, about 80% have one super star and that is usually a male star. Very few movies are driven by women stars alone, and the leading lady to do this is Priyanka Chopra.
- In Multi starrers – a combination of male and female super stars is quite popular.

4. **Movies with Character super stars**

Character Stars List

Anupam Kher	**Irrfan Khan**	**Nana Patekar**	**Nawazuddin Siddiqui**	**Paresh Rawal**
Arshad Warsi	Kay Kay Menon	Nandita Das	Om Puri	Rajkummar Rao
Boman Irani	Manoj Bajpayee	Naseeruddin Shah	Pankaj Kapur	Tabu
Vidya Balan				

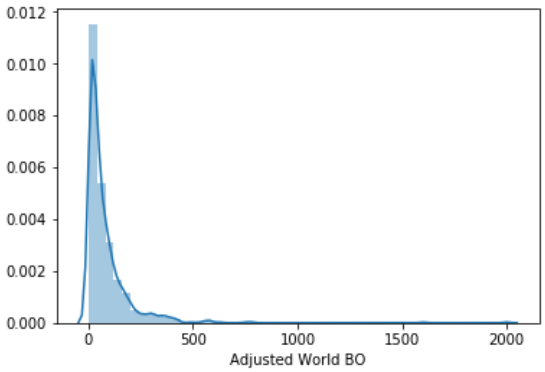


- About 20% are movies where there is presence of strong character actors and most of them engage just one male actor.
- Interestingly, less than 10% of all movies have a Matinee idol and a strong character actor.

Section 2

Some exploratory analysis of Box office performance, Top performers and Budgets

1. **Box Office Revenue Spread**



Statistical Distribution Using Seaborn in Python (Code)

```
import seaborn as sns
sns.distplot(df_all["Adjusted World BO"])
```

World Box Office (INR)

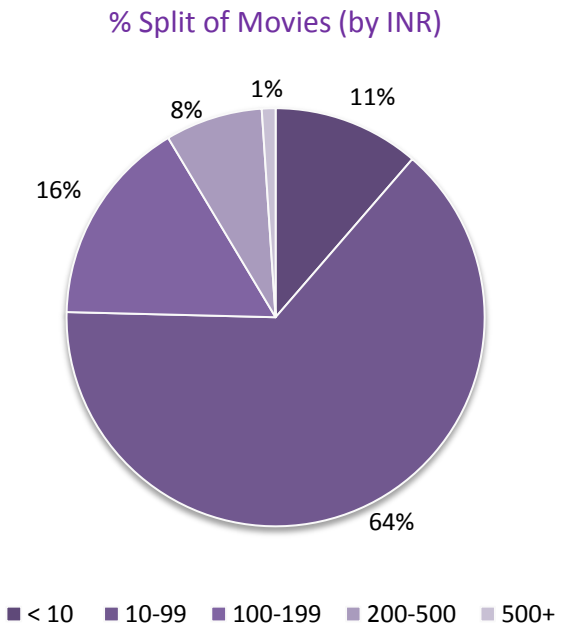
- Min : 1 Cr
- Max : 2001 Cr
- **Movie : Dangal**

India Box Office (INR)

- Min : 1 Cr
- Max : 501 Cr
- **Movie: BahuBali 2**

About half the movies do business below 30 Cr in India and their worldwide collections would be less than 45 Cr.

2. Top performing actors?



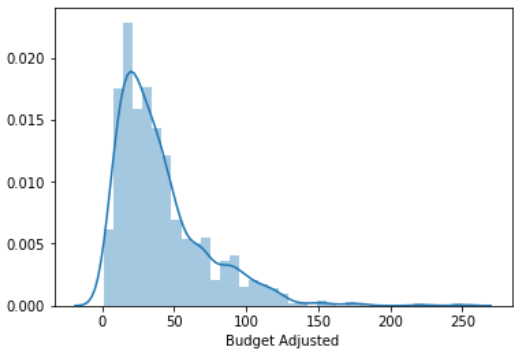
Super Star	Number of Movies (Base: 100 Cr +)	Avg. WW Revenue in INR Cr
Akshay Kumar	41	164
Shah Rukh Khan	30	273
Salman Khan	27	269
Hrithik Roshan	18	262
Amitabh Bachchan	17	193
Ajay Devgn	17	175
Priyanka Chopra	16	226
Sanjay Dutt	15	151
Aamir Khan	14	467
Katrina Kaif	14	288
Deepika Padukone	14	231

3. Most frequent and Successful Movie and Music Directors

S. No	Top 5 Movie Directors (made at least 5 movies)	Number	Avg. WW Revenue in INR Cr	Top 5 Music Directors	Number	Avg. Music Rating (Scale of 5)
1	Kabir Khan	6	270	Pritam C	89	1.8
2	Rohit Shetty	12	223	Anu Malik	76	1.7
3	Sanjay L Bhansali	6	212	Vishal Shekar	45	1.9
4	Anees Bazmee	8	193	Sajid-Wajid	39	1.7
5	Ashutosh G	6	131	A. R. Rahman	35	2.9

NOTE: Other notable mentions in the 100 Cr Avg movie revenue club are Imtiaz Ali, Abbas-Mastan, David Dhawan and Rajkumar Santoshi. Sanjay Gupta and Milan Luthria are close behind.

4. How much is being invested typically?



Budget (INR)

- Min : 1 Cr
- Max : 250 Cr
- **Movie : Babubali 2**

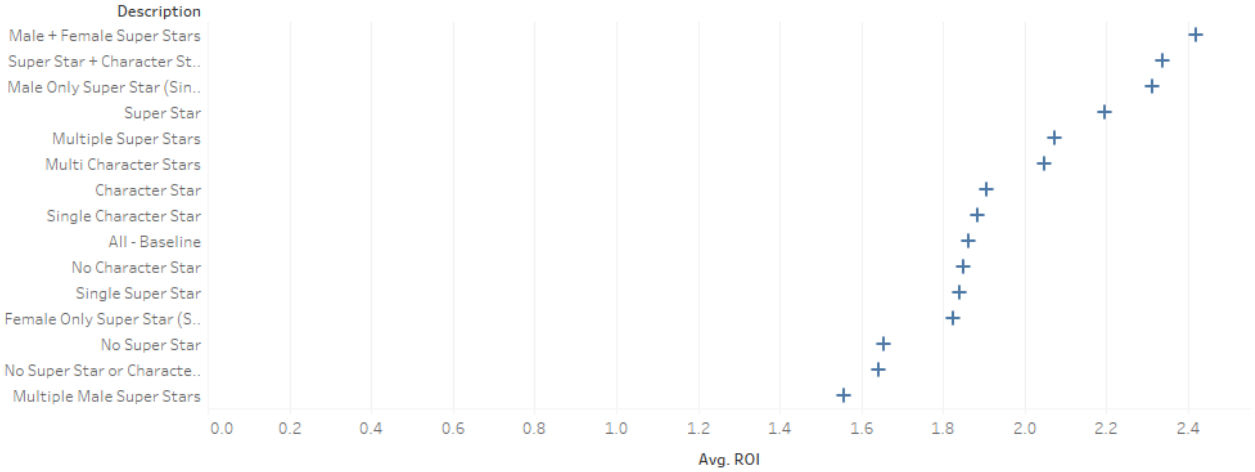
Some of the other expensive movies are : Ra One, Kites, Dhoom 3, Bang Bang, Devdas

About half the movies spend less than 30 Cr for making movies.

Section 3

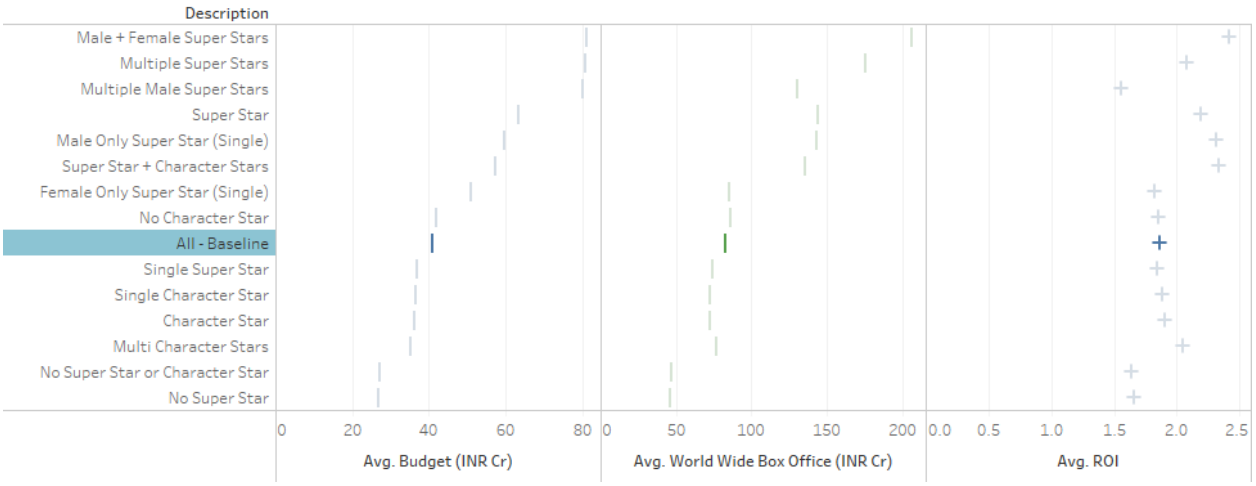
Exploring relationships between various variables and Box office success/ revenues and building basic hypothesis

1. An overview of Investment, Box office sales and ROI by presence of Movie stars



Salient points:

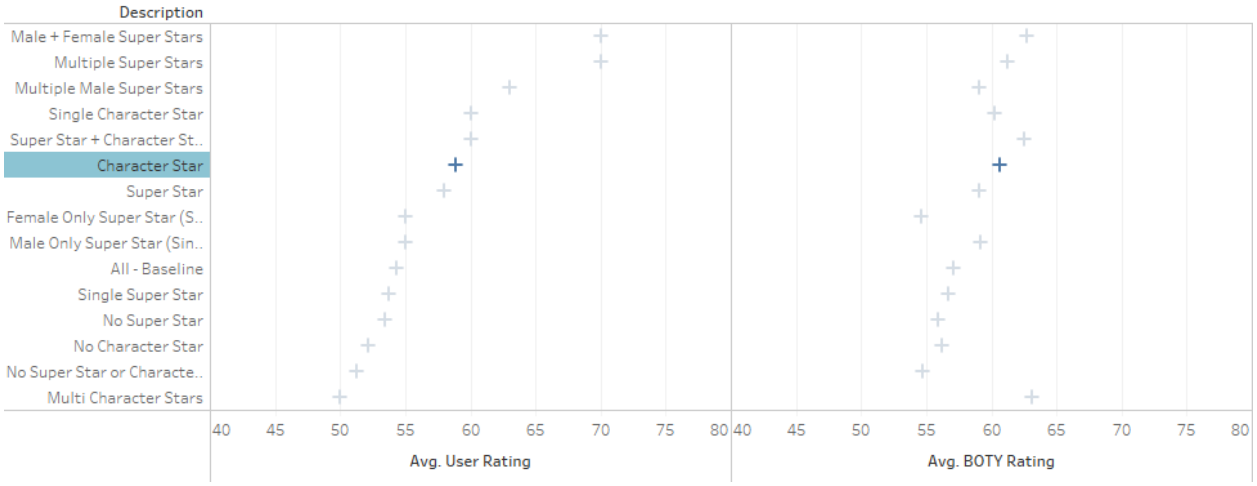
> Male and female super star pairing gives the best bang for the buck (higher ROI).
> While movies with male super stars are likely to keep the Box office ringing, the same cannot be said about movies with female only super stars.
> Movies with neither a super stars nor strong acting talent will fade away.



> Clearly movies with stars perform better than average in terms of Box office sales and the converse is true as well.
> However, when it comes to Character stars in terms of ROI, having strong acting talent can result in better returns, even though the increment vs. baseline is small.
> Also budget and sales seem to go hand in hand.

> The one unexpected finding is the low ROI of multi starrer with male super stars. There are some big flops like Ram Gopal Varma ki Aag, Department, Saawariya, Ekalavya to name a few. It is not that their budgets were high; they just didn't get in enough sales.

User Ratings



> While presence of strong acting talent doesn't yet give commensurate levels of Box office sales/ ROI, it does yield positive appreciation (significantly above average).

> One way to think of this is, over time as good and varied acting and scripts leads to higher appreciation, this could in turn lead to better returns for those who invest in it. One has to wait and watch. But for now, Stars rule the roost.

Section 4

Outliers tell a whole different story

The next logical question would be; which are those movies that have super star yet flopped and alternatively what are those low budget movies which have surpassed box office sales expectations.

What do these outliers have to offer as insight?

Big Budget Super star Movies that turned out be duds	Small budget movies without Super stars that have turned out as super hits
Khelein Hum Jee Jaan Sey	A Wednesday!
Aladin	Aashiqui 2
Knock Out	Bheja Fry
Bombay Velvet	Fukrey Returns
Zanjeer	Kahaani
Main Aur Mrs. Khanna	Kya Kehna
Drona	Malamaal Weekly
Love Story 2050	Mom
Tezz	Murder 2
Aakrosh	Neerja

Big Budget Super star Movies that turned out be duds	Small budget movies without Super stars that have turned out as super hits
What's Your Raashee	Queen
Fitoor	Raaz
Department	Vicky Donor
Family – Ties Of Blood	Vivah
The Legend of Bhagat Singh	
London Dreams	
Yuvvraaj	
Joker	
8 x 10 Tasveer	
Halla Bol	
Raju Chacha	
Saawariya	
Umrao Jaan	
Action Replayy	

> There is a clear pattern to small budget movies which have emerged as commercial hits (basis ROI). Some of them have been sleeper hits. The identifiable themes are:
  + Strong acting talent using a cluster of character actors and interesting roles
  + Unique stories that are not radically challenging the audience thinking but touch themes that they are familiar with. For example – infertility in Vicky Donor, a small town girl deciding to embrace experiences and exercise her freedom in Queen. These themes challenge the status quo but are not too radical. In some sense they play it safe.
> Big budget movies which turned out of to be box office failures are another story. No single theme emerges clearly. But here are some hypotheses:
  + Poorly defined roles and a weak storyline. For example, Main aur Mrs Khanna.
  + Riding on the bandwagon of past movies that were successes. For example if Shootout at Lokhandwala was a hit, a bunch of movies launched on similar lines (special task force to root our criminals from the underworld) were made and flopped. Case in point – Department.
  + Movies that take up historical topics and are rooted in real world incidents where the common man is a hero. The Legend of Bhagat Singh perhaps is overdone and Halla Bol had merged too many themes around common man as an activist to have a singular focus. It lacks a sense of uniqueness and sharpness in the storytelling.

Step IV

Modelling and Validation

Some background on data considered for Modelling,

To make the analysis sharper, we categorised the movie success into two buckets.

1. Flop– ROI < Less than 2
2. Hit – ROI > 3
   - > The overall average ROI for all movies is a little over 2
   - > The source variable was BO performance which was divided into 5 levels
   - > So Flop maps to Level 1,2 of BO performance and Hit maps to Level 4,5
   - > So level 3 (Average hit) was excluded to create sharper discrimination between a week performer and a strong performer
3. This brought down the list of movies to form 943 to 755

Analytical Models used,

3 algorithms/ statistical models were utilized,

1. Logistic Regression
2. RandomForest
3. Naives Bayes

Implementation of Algorithms

All 3 algorithms were implemented in R (*please refer to Annexure 2 for the full code)*

Note: The data was not split into training and test sets, as we used K-fold validation for determining model accuracy.

Quick summary of each Algorithm

S. No	Algorithm/ Package in R	Key points about the algorithm
Logistic Regression	**glm (in-built), both logit and probit link functions were used**	**Highly popular, works well when independent variables are categorical**
Random Forest	**randomForest**	**Ensemble algorithm - runs multiple Decision Trees, again works well with categorical data**
Naïve Bayes	**naiveBayes(), e1071**	**Probabilistic model, simple and very effective**

Snapshot Summary of the Code (please refer to Annexure 2 for the detailed code)

```
### Predicting likelihood of Bollywood Movies' Box office Success

## Logistic Regression for classification and Prediction

model <- glm (BO_Hit_Flop ~ SuperStar_CharacterStar + Multiple_SuperStars + SuperStar + Multiple_CharacterStars + CharacterStar + Overall_Starcast_Score +
Director_Rating + Movie_Music_Rating, data = movies, family = binomial(link ="logit"))

coef(summary(model))


## Random Forest for Classification and Prediction

fit <- randomForest( BO_Hit_Flop ~ SuperStar_CharacterStar + Multiple_SuperStars + SuperStar + Multiple_CharacterStars + CharacterStar + Overall_Starcast_Score
+ Director_Rating + Movie_Music_Rating,data=movies, importance=TRUE, ntree=2000)

plot(fit)
varImpPlot(fit)


## Naive Bayes for classification and Prediction

nb <- naiveBayes( BO_Hit_Flop ~ SuperStar_CharacterStar + Multiple_SuperStars + SuperStar + Multiple_CharacterStars + CharacterStar + Overall_Starcast_Score +
Director_Rating + Movie_Music_Rating,data=movies)

summary(nb)
```

Summary of Prediction Accuracy

The below table gives the comparison of each algorithms's effectiveness

Classifier	LR	RF	NB
Total Accuracy	***85%***	***90%***	***84%***
True Positive Rate	45%	51%	43%
False Negative Rate	**55%**	**49%**	**57%**
True Negative Rate	95%	99%	94%
False Positive Rate	5%	1%	6%

At a glance all 3 classifiers seem to be doing an equally good job. However the key problem is with the False Negative Rate.

False Negative: a Hit movie is predicted to be a Flop movie. 50% false positive is not a good result, although the overall accuracy is good.

How to fix it?

In logistic regression, the final class is determined basis the probability and the default cut off to classify it as a hit is 0.5. This is arbitrary and can be changed to balance false positive and true negative rates. With probability cut off as 0.3 a better result was achieved.

Logistic Regression	0.5 Probability cut off	0.3 Probability cut off
Total Accuracy	***85%***	85%
True Positive Rate	45%	60%
False Negative Rate	**55%**	**40%**
True Negative Rate	95%	91%
False Positive Rate	5%	9%

Final Model

LR Model:

Due to the flexibility offered by LR to adjust the false negative rate, this algorithm was zeroed-in on. Moreover interpreting the impact of independent variables on the response variable (Hit vs. Flop) is more intuitive.

Independent variables chosen:

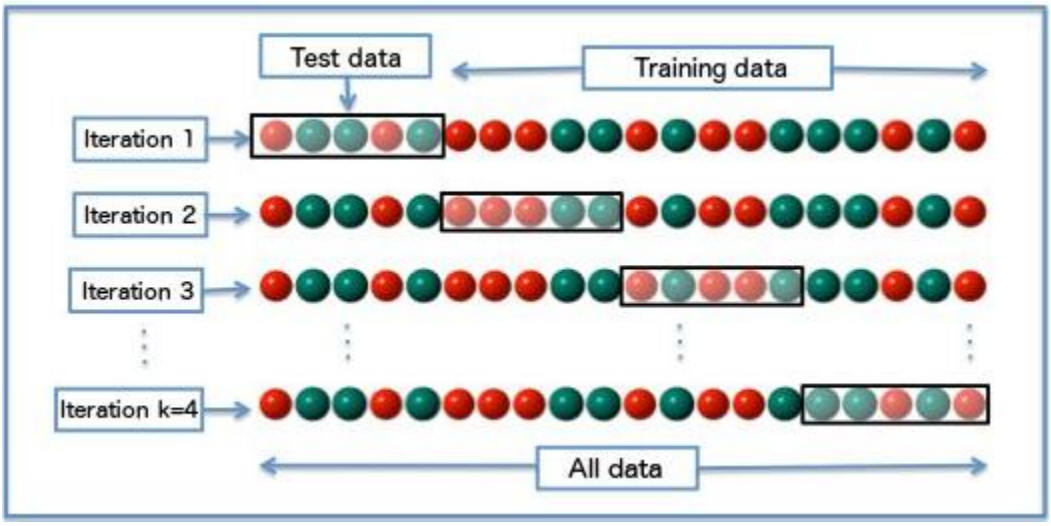
8 variables were shortlisted based on the p-value of each of the variable ($p < 0.05$). This indicates that they have a role to play in influencing the response variable (Refer to the results table for the variables)

K Fold Validation

Further, K fold validation (K = 10) was run to see the consistency of model accuracy. And the overall accuracy on an average was 83%.

K fold validation automatically breaks up the data into training and test modules repeatedly and run the models K times (in this case 10 times) and then gives the overall average accuracy.

The image gives a short overview of the method.



Step V

Interpretation of the results

The table below shows the odds-ratio of each independent variables on driving the success of a movie.

S. No	Independent Variable/ Feature	Impact (Odds- Ratio)
1	Presence of a Super Star	0.25
2	Presence of Multiple Super Stars	0.66
3	Presence of a Character Star	0.82
4	Presence of Multiple Character Stars	**1.30**
5	Presence of a Super Star and a Character Star	**7.00**
6	Overall Star cast Score	1.01
7	Director Rating	1.01
8	Music Rating	**1.72**

Interpreting the Result

Taking presence of Multiple Character stars as an example: **odds ratio of 1.3** indicates that not having multiple character stars vs. having them in the movie; increase the chances of success by 1.3 times.

Clearly the most important drivers of success are:

1. Presence of a Super star complemented by a Character star
2. Presence of multiple Character stars (even without a super star)
3. Likeability of Movie's Music

Does the data support the model results?

The below table shows the probability of success by various independent variables. Very clearly these results are aligned with the model results.

For example – Super star plus character star increases the probability of success to 43% from the base scenario of 20%. So the odds go up by 2.14. Here the odds will look higher as simple cross tabs will not take multi-collinearity into account, while the model does.

S. No		Total	Hit	Flop	Probability of Success	Index to Average
	Overall	***755***	***150***	***605***	***20%***	
1	Super Star	277	80	197	29%	**145**
2	Male only Super Star	181	56	125	31%	**156**
3	Female only Super Star	36	6	30	17%	84
4	Multiple Super Stars	60	18	42	30%	**151**
5	Male only Multiple Super Stars	27	3	24	11%	56
6	Male plus Female Super Stars	33	15	18	45%	**229**
7	Character Star	159	36	123	23%	114
8	Multiple Character Stars	24	7	17	29%	**147**
9	Super Star plus Character Star	47	20	27	43%	**214**
1	Star Cast Rating (Avg Rating)	83	139	70		
2	Director Rating (Avg Rating)	78	159	58		
3	Movie Music Rating (Avg Rating)	1.6	2.5	1.4		

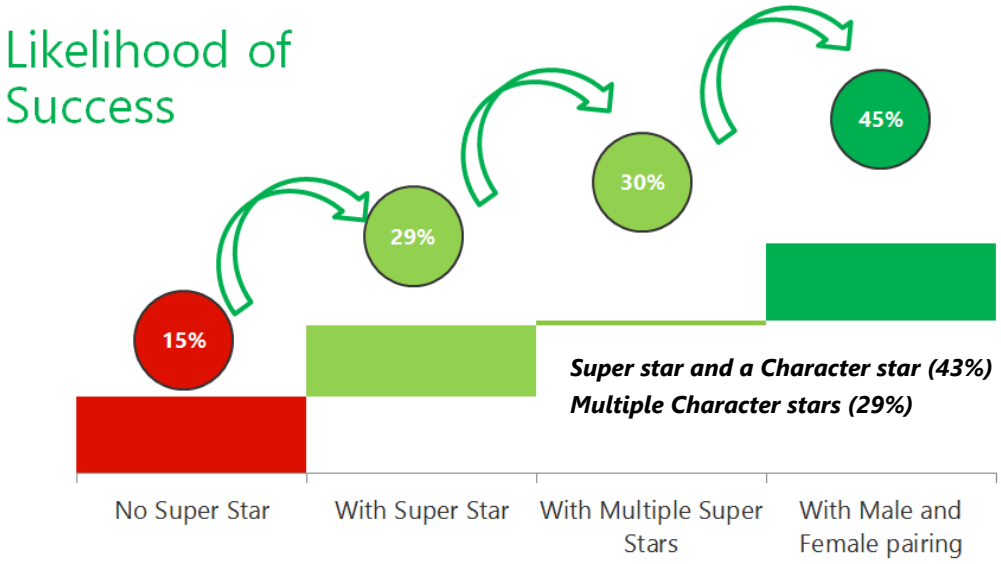
Further,

1. Super star supported by a character star stands out the most important driver of success. Presence of a super star alone (and a male super star) will also ensure success.
2. Equally important is to have a male-female super star pairing instead of male only super stars for ensuring success.
   > Movies with Multiple Male only Super stars tend not to do well. This perhaps has to do with clear definition and clarify of roles in the movie
3. Interestingly, movies with multiple character stars do significantly better than movies with just one Character star. This could be due to the overall quality of acting, that a bunch of talented actors bring together and perhaps complement each other well to ensure success.
4. Finally how well the Music engages viewers also plays a significant role.

Step VI

Recommendations for stakeholders

The below chart summarizes the high level direction one can draw from the analysis.



Key Recommendations

S for Success, S for Super stars

> If you can pay, consider a male super star.
    + And a male super star paired with a female co super star maximizes the chances of success. Weaving in a strong role for a character star can further bolster the chances of success.

Story is KING

> If one doesn't intend to work with super stars, then a stellar story is your best friend.
> Unique themes that are not radically challenging the audience's thinking but touch their hearts are likely to do well, even without the support of super stars. Pick up any sleeper hit and this trend will be visible.
> Strong story complemented with strong character roles and character stars will negate the need for big budget investments and matinee idols.
> **Note:**
    + While such movies will turn out to give good returns (ROI), they will not make the 100 Cr club, though. This is only possible with big stars.

MORE THAN ONE MASTER CHEF CAN SPOIL THE DISH

Pairing of multiple male super stars, unless the roles in the story are very clearly defined, is likely to not do well. Some of the most unsuccessful movies share this feature.

Step VII

Future Possibilities

A range of possibilities exist. A few are listed below,

> Adding more relevant data points to enhance the analytical framework/model. Data points like marketing budgets of the movies, distribution/ no of screens secured, quality of story and many other such inputs. It would only make the analytical framework richer.

> Move to the next level of analytical modeling – predicting the actual Box office sales. This is possible with the right inputs, like those mentioned above.

> Extending this framework and predictive approach to various other formats. For example - TV serials.

> A deeper evaluation and understanding of qualitative themes around story genre, cultural themes which make movies/ content successful and many of these go beyond hard numbers and models.

> Currently, we are beginning explore the possibilities of working with production houses and broadcasting companies to take this further.

ABOUT THE AUTHORS/ PROJECT CONTRIBUTORS

A.J.R.Vasu
ajrvasu@gmail.com
+91.77381.58553

- Marketing and Analytics consultant (www.linkedin.com/in/ajrvasu) with 15 years of experience spanning India, global markets and various global-local CPG companies
- Currently lead the Retail Analytics practice at The Nielsen Company (www.nielsen.com) for South Asia geography
- Experienced at P&L management, business development, analytical and statistical solution building, client consulting and talent management
- Experienced at key application areas like segmentation, forecasting, new launch success prediction, marketing mix allocation and drivers of performance
- Technical know-how spans statistical modeling, machine learning techniques, marketing research methods; beginner level hands-on experience with programming languages (Python, R, and SPSS) and familiarity with the Big data ecosystem (Hadoop, Hive, Pig, Spark and Flume)
- Bachelor's degree in Mathematics (PCM), followed by a Management degree followed by a Professional Degree in Big Data Analytics and Machine Learning

Ashish Patwardhan
ashish.patwardhan@rediffmail.com
9819008025

- Marketing and Business Development professional with 10 years of experience spanning west and north of India with various media Moguls. Interested in Machine learning and Big data analytics profile.
- Last stint was with Times of India as Senior Manager in Brand Solution (Medianet) for retail vertical for West region.
- Experienced in Business development, Brand Solution, Solution selling, MIS and Team Management.
- Technical know how spans statistical modelling, machine learning algorithms, big data technologies, moderate level programming.
- Bachelor's degree in Engineering, followed by a Management degree followed by Micro degree in Big data analytics and Machine learning.

Nikhil Gore
nikhilgore14@gmail.com
+91 9619 72 73 16

- Engineering and a Management graduate with ten years of experience in business, financial and data analysis, client and stakeholder management .
- Currently head of Analytics, MIS and Tax Reporting for all airline clients at Sutherland Global (Approximate Revenue – $1.5MM).
- Prior to this role, I have worked as a research analyst for eight years that exposed me to various sectors like retail, packaging, industrials, chemicals and cable broadcasting entailing projects such as opportunity assessment, market size modelling, business analysis and primary research surveys.
- Know-how of research techniques, machine learning algorithms and basic programming.
- Completed BE in Civil Engineering from VJTI, Mumbai and MBA in Finance from Institute of Management, Nirma University (IMNU), Ahmedabad followed by a certificate program in Big Data Analytics from SP Jain School of Global Management, Mumbai.

Alok Yadav
alok.proc@gmail.com
+91 8097 72 01 87

- Biotechnology and Management graduate with six years of experience in Clinical Data Management and Teaching.
- Currently working as Clinical Data Manager at Tata Consultancy Services, Mumbai. The role involves Data management and analytics and data interpretation activities.
- Prior to this role, I worked as a visiting faculty in various institutes affiliated to Mumbai University.
- Know-how of machine learning algorithms, basic programming and SAS.
- Completed MSc. in Biotechnology from Mumbai University, Professional Program in Management from SPJIMR, Post Graduate Diploma in Pharmaceutical Management from Garware Institute, Post Graduate Diploma in Applied Statistics from Mumbai University and a certificate program in Big Data Analytics from SP Jain School of Global Management, Mumbai.