



SVKM'S

NMIMS

Deemed to be UNIVERSITY

Project Report on
EXPLORATORY DATA ANALYSIS OF MOVIES
B.TECH 3RD YEAR
(Branch – CSE)

Submitted To:

MR. SURAJ PATIL
(Assistant Professor)

MPSTME, NMIMS
Shirpur Campus

Submitted By:

VED BHATKAR(B205)
ARIHANT JAIN(B220)
VARAD NILAKHE(B238)

Course: B TECH CS
Batch: 2019-2023

ACKNOWLEDGEMENT

This Project report was completed because of support from the groupmates, although not all of them can be mentioned.

We are greatly indebted to our good supervisor **Suraj Patil sir** for his/her useful and necessary observation, suggestions, contribution, and corrections. We would not have been able to achieve anything in this research without your supervision.

Ved Bhatkar

Arihant Jain

Varad Nilakhe

INTRODUCTION

- The film industry plays a major role in the planetary or world-wide economy. It is the symbolic contributor to the global economy.
- Every year more than hundreds to thousands of movies are released to the public audience with the hope that the movies getting released will be the next block buster.
- According to the movie industry statistics, six to seven movies out of ten movies gets unprofitable, only one third of the movie gets success.
- The producers, studios, investors, sponsors in the movie industry are alike interested in predicting the box office success of the movie.
- This paper work is on analyzing the film genre, the releasedate around holidays, the release month of movies, the languages and country with more movies from the movie review dataset.
- There are attributes (country, languages, genre, movie release date, budget and revenue) taken from the dataset and the derived attributes (release monthof the movie derived from release date of movie and profit from budget and revenue) is analysed to determine the movie performance.
- The analysed data is plotted in graphs for statistical observation of the moviesuccess.

Problem Statement:

- To automate the process of predicting whether the person has liked the movie or not. In order to do so, we will be using Naïve Bayes Classification.
- Another objective of this study is to investigate that how much a movie is liked by the people so the production should do a remake of that movie or not .
- Our goal is to compare the movies which are more liked and less liked by the people and on the basis of that what can be done to remake it or not.

TECHNIQUE USED

- It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.
- Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

ALGORITHM

- Step 1: Convert the data set into a frequency table
- Step 2: Create Likelihood table by finding the probabilities like
Overcastprobability = 0.29 and probability of playing is 0.64.
- Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction

ABOUT THE DATASET:

- The data contains the name of the movies, their release period, type of genre, music director, the movie director, the budget of the movie and the revenue of the particular movie and the number of screens its being played.

IMPLEMENTATION

Write the steps with each code and provide output snapshot.

Source Code:

```
#a Importing the dataset
library(dplyr)
movie<-read.csv("D:\\DM\\PROJECT\\movie.csv")
View(movie)
str(movie)

summary(movie)

#Data Pre Processing
#b Handling the missing data
movie$Budget.INR. =
ifelse(is.na(movie$Budget.INR.),ave(movie$Budget.INR., FUN
= function(x) mean(x, na.rm = 'TRUE')),movie$Budget.INR.)
```

```
movie$Revenue.INR. =  
ifelse(is.na(movie$Revenue.INR.),ave(movie$Revenue.INR.,  
FUN = function(x) mean(x, na.rm =  
'TRUE')),movie$Revenue.INR.)
```

```
#Designing the histogram  
hist(movie$Number.of.Screens, col = "red")
```

```
#Designing All the Plots  
library(ggplot2)  
#9  
#a Display bar chart  
ggplot(data=movie,aes(Genre))+geom_bar()
```

```
#b Design Histogram  
ggplot(data = movie, aes(Budget.INR.))+geom_histogram(bins  
= 10)
```

```
#c Design Scatter Plot  
ggplot(data =movie, aes(x = Budget.INR., y = Revenue.INR.,  
col = Genre))+geom_point()
```

```
#d Design Box Plot  
ggplot(data =movie, aes(fill = Genre, x = Budget.INR., y =  
Revenue.INR.))+geom_boxplot(notch=TRUE)
```



```
#Factoring target feature
movie$Whether.Remake = factor(movie$Whether.Remake,
                              levels = c('No','Yes'),
                              labels = c(0,1))
```

```
#Transforming data into class variable
Remake_class<-ifelse(movie$Whether.Remake=="1",
"Remaked", "Not Remaked");
movie <- data.frame(movie, Remake_class)
View(movie)
```

```
#Removing Whether.Remake from Data Frame
movie <- movie[,-3]
```

```
#Splitting the Dataset
library(caTools)
set.seed(2)
id <- sample(2, nrow(movie), prob=c(0.7, 0.3), replace=TRUE)
print(id)
```

```
#Training and Testing
movie_train <- movie[id==1,]
movie_test <- movie[id==2,]
print(movie_train)
```

```
#Building the model
library(e1071)
model <- naiveBayes(as.factor(Remake_class)~., movie_train)
print(model)
```

```
pmodel <- predict(model, movie_test)
```

```
#Building the confusion matrix
```

```
table(pmodel, movie_test$Remake_class)
```

OUTPUT SNAPSHOTS:

1) Importing dataset

Firstly, we will import the dataset.

```
> #a Importing the dataset
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

> movie<-read.csv("D:\\DM\\PROJECT\\movie.csv")
> View(movie)
> str(movie)
'data.frame': 1698 obs. of 14 variables:
 $ Movie.Name      : chr "Golden Boys" "Kaccha Limboo" "Not A Love Story" "Qaidi Band" ...
 $ Release.Period  : chr "Normal" "Holiday" "Holiday" "Holiday" ...
 $ Whether.Remake   : chr "No" "No" "No" "No" ...
 $ Whether.Franchise : chr "No" "No" "No" "No" ...
 $ Genre           : chr "suspense" "drama" "thriller" "drama" ...
 $ New.Actor       : chr "Yes" "Yes" "No" "Yes" ...
 $ New.Director    : chr "No" "No" "No" "No" ...
 $ New.Music.Director: chr "No" "Yes" "No" "No" ...
 $ Lead.Star       : chr "Jeet Goswami" "Karan Bhanushali" "Mahie Gill" "Aadar Jain" ...
 $ Director        : chr "Ravi Varma" "Sagar Ballary" "Ram Gopal Verma" "Habib Faisal" ...
 $ Music.Director   : chr "Baba Jagirdar" "Amardeep Nijjer" "Sandeep Chowta" "Amit Trivedi" ...
 $ Number.of.Screens : int 5 75 525 800 1 10 2 30 1550 500 ...
 $ Revenue.INR     : int 5000000 15000000 75000000 210000000 1000000 5000000 1500000 15000000 520000000 180000000 ...
 $ Budget.INR      : num 85000 825000 56700000 4500000 1075000 ...

>
> summary(movie)
  Movie.Name      Release.Period  Whether.Remake  Whether.Franchise
Length:1698      Length:1698      Length:1698      Length:1698
Class :character  Class :character  Class :character  Class :character
Mode :character   Mode :character   Mode :character   Mode :character

  Genre           New.Actor       New.Director    New.Music.Director
Length:1698      Length:1698      Length:1698      Length:1698
```

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.
1 Golden Boys	Normal	No	No	suspense	Yes	No	No	Jeet Goswami	Ravi Varma	Baba Jagirdar	5	5,00e+06	85000
2 Kaccha Limboo	Holiday	No	No	drama	Yes	No	Yes	Karan Bhargava	Sagar Bhatnagar	Amandeep Nijjar	75	1,50e+07	425000
3 Neil A Love Story	Holiday	No	No	thriller	No	No	No	Mahar Gil	Ram Gopal Verma	Sandeep Chawla	525	7,50e+07	5620000
4 Qaidi Band	Holiday	No	No	drama	Yes	No	No	Aadar Jan	Habib Faisal	Amir Taseef	800	2,10e+08	4500000
5 Chashmi	Holiday	No	No	adult	Yes	Yes	Yes	Aashif Khan	Aashif Khan	Bedoon Uddad	1	1,00e+06	1070000
6 Shuttlescock Boys	Normal	No	No	comedy	Yes	Yes	Yes	Aakar Kausik	Hemant Gaba	Airash Bagheri	10	5,00e+06	170000
7 Dirty Marriage	Holiday	No	No	adult	Yes	No	Yes	Ashish	Priyanka	Dharma	2	1,50e+06	25000
8 Future To Bright He II	Holiday	No	No	drama	No	No	Yes	Amir Bashir	Sergey Anar	Amir Ali	30	1,50e+07	825000
9 Gajini	Holiday	Yes	No	action	No	Yes	No	Amir Khan	A.R. Murugadasai	A.R. Rahman	1150	5,30e+08	194500000
10 Taare Zameen Par	Holiday	No	No	drama	No	Yes	No	Amir Khan	Amir Khan	Shankar - Dhason - Loy	300	1,80e+08	87570000
11 Mangal Pandey - The Rising	Holiday	No	No	drama	No	No	No	Amir Khan	Ketan Mehta	A.R. Rahman	675	3,70e+08	52570000
12 Fanaa	Normal	No	No	love_story	No	No	No	Amir Khan	Kumar Kohli	Jatin - Laili	650	3,00e+08	105400000
13 Dangal	Holiday	No	No	drama	No	No	No	Amir Khan	Nishu Tiwari	Pritam	4250	1,32e+09	7024750000
14 3 idiots	Holiday	No	No	drama	No	No	No	Amir Khan	Rakumar Hirani	Shantanu Moitra	1750	7,70e+08	3491035000
15 PK	Normal	No	No	drama	No	No	No	Amir Khan	Rakumar Hirani	Ajay - Alka	3600	1,20e+09	4160362500
16 Rang De Basanti	Holiday	No	No	drama	No	No	No	Amir Khan	Rakesh Omprakash Mehra	A.R. Rahman	600	2,80e+08	979060000
17 Talaash	Normal	No	No	thriller	No	No	No	Amir Khan	Reema Kagti	Ram Sampath	2500	7,10e+08	1729175000
18 Dhoom 3	Normal	No	Yes	action	No	No	No	Amir Khan	Vijay Krishna Acharya	Pritam	3650	1,75e+09	5243760000
19 Pyaar Mein Aisa Hota Hai	Holiday	No	No	thriller	Yes	Yes	Yes	Aamirika Bawa	Rajesh Kumar	Shabbir Azmi	5	5,50e+06	350000
20 Gurjar Aandolan - A Fight For Right	Normal	No	No	drama	Yes	Yes	Yes	Aarun Nagar	Aarun Nagar	Ashish Donadri	5	6,50e+06	575000
21 Yeh Hai High Society	Holiday	No	No	thriller	No	No	No	Aaryan	Ravi Sinha	S. Paul	20	6,50e+06	4000000
22 Say Yes To Love	Normal	No	No	love_story	Yes	No	No	Asaad Mirza	Mahrukh Mirza	Jatin Pandit	30	1,75e+07	525000
23 Dreams - Sapnay Sech Henge...	Holiday	No	No	love_story	Yes	Yes	No	Aashish Chandra	Aashish Chandra	Sajid - Waqid	25	1,25e+07	1375000
24 Three - Love Lies Betrayal	Normal	No	No	thriller	No	Yes	No	Aashish Chaudhary	Vishal Pandya	Chirantan Bratt	325	4,00e+07	1000000
25 Maasegon Me Gadbad Ghotala	Normal	No	No	comedy	Yes	Yes	No	Aash Alkabi	Anwar Shaikh	Sahil Pali	5	1,00e+06	300000
26 Kholi Sikka - Jaat Ke Tisrahi	Normal	No	No	comedy	Yes	Yes	Yes	Aash Kumar	Ram Prasad Singh	Uditvhar	50	9,00e+06	600000
27 Hi Fi Log	Holiday	No	No	adult	Yes	Yes	No	Ashay Bakshi	Baba Jagirdar	Baba Jagirdar	2	1,50e+06	500000
28 Dev D	Holiday	No	No	drama	No	No	No	Ashay Desai	Anurag Kashyap	Amir Trivedi	525	1,10e+08	20620000
29 One By Two	Normal	No	No	rom_com	No	Yes	No	Ashay Desai	Devika Bhagat	Shankar - Dhason - Loy	500	2,10e+08	37700000
30 Oye Lucky! Lucky Oye!	Normal	No	No	comedy	No	No	No	Ashay Desai	Dibakar Banerjee	Sneha Khanwalkar	400	1,20e+08	100620000
31 Sacha Na Tha	Holiday	No	No	rom_com	Yes	Yes	No	Ashay Desai	Imtiaz Ali	Sandeep Shandilya	110	5,00e+07	52800000
32 Happy Bhag Jayegi	Normal	No	No	rom_com	No	No	No	Ashay Desai	Muhammad Aziz	Suhail Sen	1000	2,00e+08	391500000
33 Manorama Six Feet Under	Normal	No	No	suspense	No	Yes	Yes	Ashay Desai	Nanddeep Singh	Jayesh Gandhi	75	3,25e+07	5800000
34 Honeymoon Travels Pvt. Ltd.	Normal	No	No	rom_com	No	Yes	No	Ashay Desai	Reema Kagti	Vishal - Shekhar	350	1,00e+08	195160000

2) Data Preprocessing

```
> #Data Pre Processing
> #B Handling the missing data
> movie$Budget.INR. = ifelse(is.na(movie$Budget.INR.),ave(movie$Budget.INR., FUN = function(x) mean(x, na.rm = 'TRUE')),movie$Budget.INR.)
>
> movie$Revenue.INR. = ifelse(is.na(movie$Revenue.INR.),ave(movie$Revenue.INR., FUN = function(x) mean(x, na.rm = 'TRUE')),movie$Revenue.INR.)
>
```

3) Data Visualization

Plots:

```
> #Designing the histogram
> hist(movie$Number.of.Screens, col = "red")
```

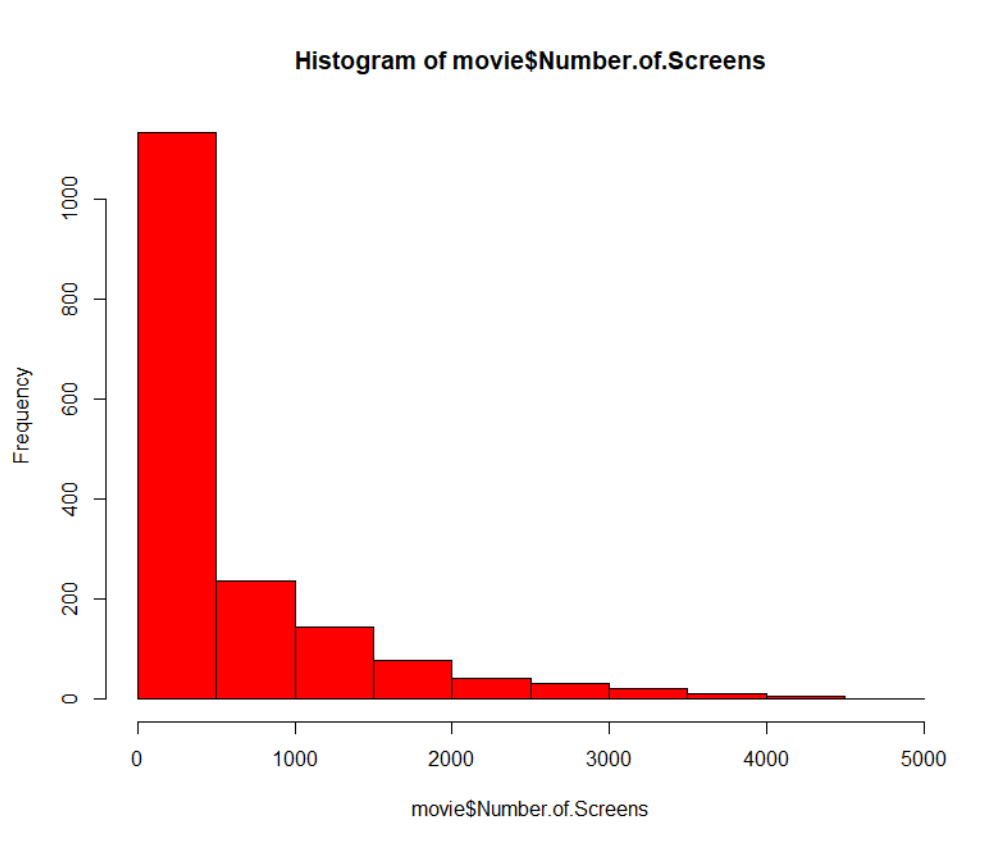


FIG-1 HISTOGRAM

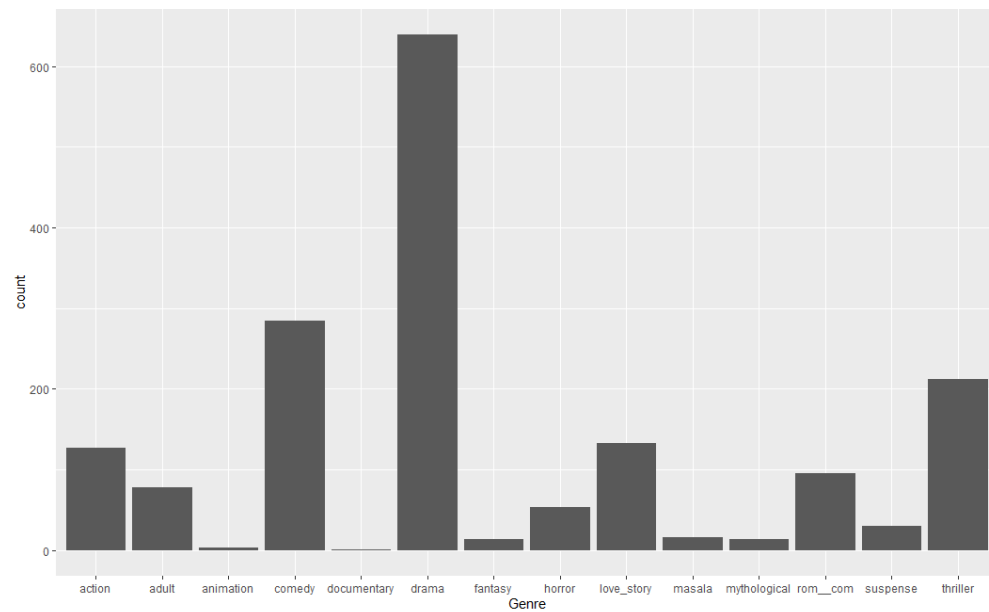


FIG-2 Bar Chart

```
> #b Design Histogram  
> ggplot(data = movie, aes(Budget.INR.))+geom_histogram(bins = 10)
```

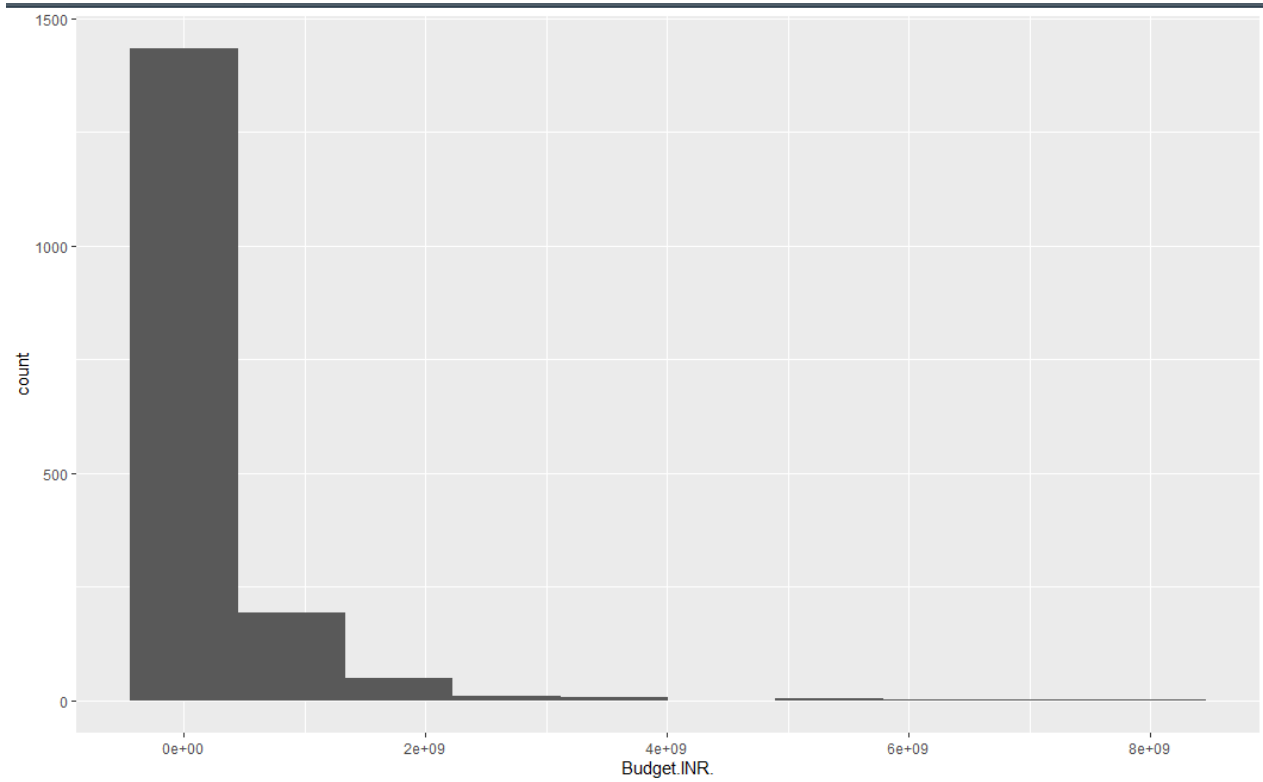


FIG-3 Histogram

```
> #c Design Scatter Plot  
> ggplot(data =movie, aes(x = Budget.INR., y = Revenue.INR., col = Genre))+geom_point()
```

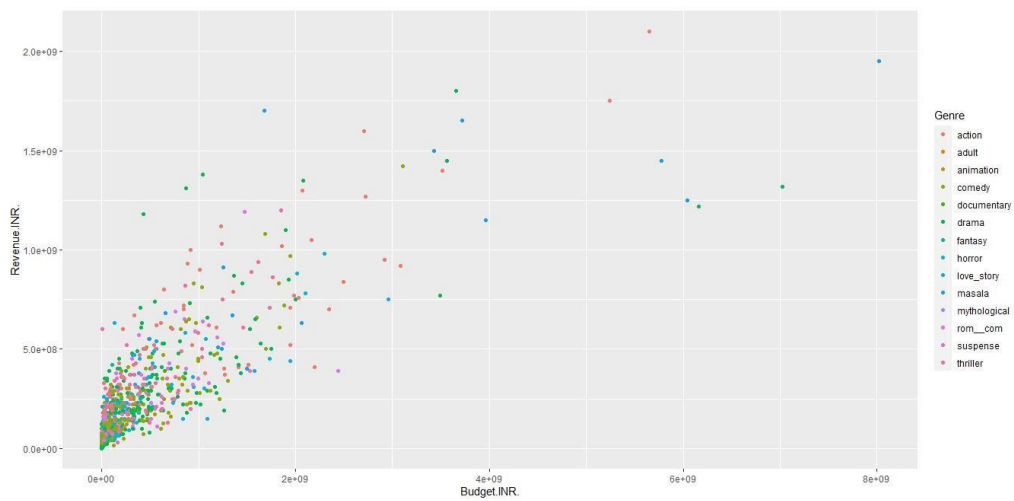


FIG-4 Scatter Plot

```
> #d Design Box Plot
> ggplot(data =movie, aes(fill = Genre, x = Budget.INR., y = Revenue.INR.))+geom_boxplot(notch=TRUE)
```

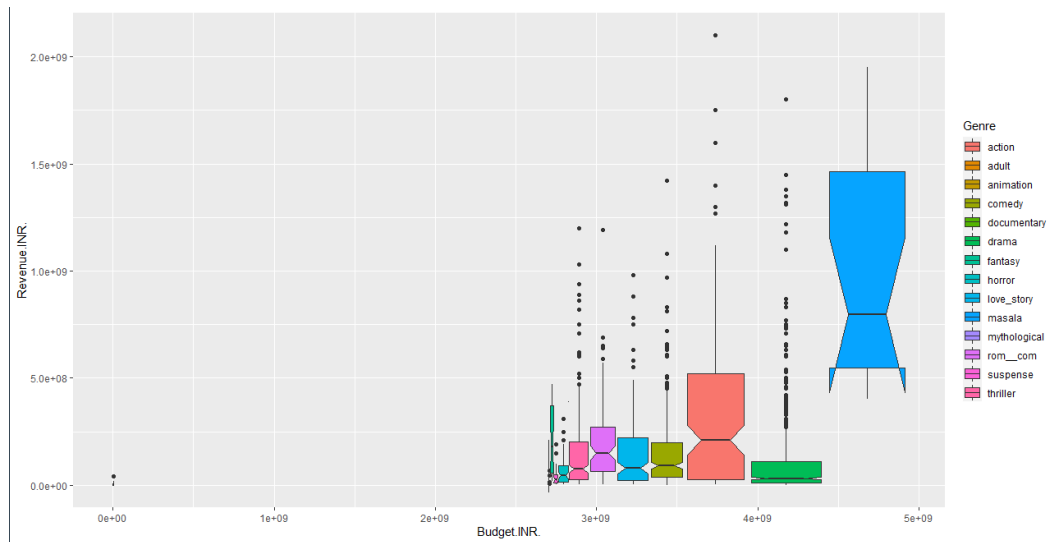


FIG-5 Box Plot

```
> #Factoring target feature
> movie$Whether.Remake = factor(movie$Whether.Remake,
+                               levels = c('No', 'Yes'),
+                               labels = c(0,1))
```

MovieName	ReleasePeriod	WhetherRemake	WhetherFranchise	Genre	NewActor	NewDirector	NewMusicDirector	LeadStar	Director	Music.Director	Number.of.Screens	Revenue.INR.	Budget.INR.
1. Gooden Boys	Normal	0	No	suspense	Yes	No	No	Jeff Gonsami	Ravi Varma	Baba Jagirdar	1	5.00e+06	85000
2. Kuchey Lubhoo	Holiday	0	No	drama	No	No	No	Kamla Shrivasth	Sagar Bhatia	Anandash Nigri	70	1.50e+07	825000
3. Not A Love Story	Holiday	0	No	thriller	No	No	No	Maha Gai	Bani Gosai Varma	Sandeep Chouda	525	7.50e+07	5670000
4. Quidam Band	Holiday	0	No	drama	Yes	No	No	Aadil Jain	Habib Talak	Anil Thedi	800	2.10e+08	4500000
5. Chaudhvi	Holiday	0	No	adult	Yes	Yes	Yes	Aadil Khan	Aadil Khan	Baboo Ustad	1	1.00e+06	1075000
6. Shuttlecock Boys	Normal	0	No	comedy	Yes	Yes	Yes	Aadil Kaushik	Hemant Gaba	Aamish Baghel	10	5.00e+06	170000
7. Dirty Marriage	Holiday	0	No	adult	Yes	No	Yes	Aakash	Priyanka	Dharma	2	1.50e+06	35000
8. Future To Bright Hai 8	Holiday	0	No	drama	No	Yes	Yes	Aamir Bashir	Sanjay Anwar	Anir Ali	30	1.50e+07	825000
9. Ghaghi	Holiday	1	No	action	No	Yes	No	Aamir Khan	A.R. Murugadas	A.R. Rahman	1550	5.20e+08	194820000
10. Sweet Caramel Pie	Holiday	0	No	drama	No	Yes	No	Aamir Khan	Aamir Khan	Shankar - Ehsaan - Loy	500	1.80e+08	671780000
11. Mangle Pandey - The Rising	Holiday	0	No	drama	No	No	No	Aamir Khan	Kunal Kohli	A.R. Rahman	870	3.70e+08	325780000
12. Faras	Normal	0	No	love_story	No	No	No	Aamir Khan	Kunal Kohli	Jeff - Lari	850	3.00e+08	105480000
13. Dargah	Holiday	0	No	drama	No	No	No	Aamir Khan	Nikhil Taseel	Pritam	4250	1.30e+09	702470000
14. 3 Idiots	Holiday	0	No	drama	No	No	No	Aamir Khan	Rajkumar Hirani	Shantanu Moitra	1750	7.70e+08	349105000
15. PK	Normal	0	No	drama	No	No	No	Aamir Khan	Rajkumar Hirani	Ajay - Abul	3600	1.22e+09	616082500
16. Rang De Basanti	Holiday	0	No	drama	No	No	No	Aamir Khan	Rakesh Omprakash Mehra	A.R. Rahman	630	2.80e+08	959080000
17. Talaash	Normal	0	No	thriller	No	No	No	Aamir Khan	Reema Kagti	Ram Sampath	2550	7.10e+08	1729717500
18. Dhoom 3	Normal	0	Yes	action	No	No	No	Aamir Khan	Vijay Krishna Acharya	Pritam	3650	1.75e+09	5243760000
19. Pyaar Mein Aisa Hota Hai	Holiday	0	No	thriller	Yes	Yes	Yes	Aasanka Sawa	Rajesh Kumar	Shabbir Azmi	5	5.00e+06	350000
20. Gully Boy - A Fight For Right	Normal	0	No	drama	No	No	Yes	Aamir Nigam	Aamir Nigam	Aashish Banoval	5	6.50e+06	375000
21. Veli Na High Society	Holiday	0	No	thriller	No	No	No	Aasani	Bani Sinha	S. Paul	30	6.50e+06	4000000
22. Say No To Love	Normal	0	No	love_story	Yes	No	No	Aasani	Mahesh Mittal	Jeff - Lari	30	1.75e+07	525000
23. Dreams - Sapna Sach Hange...	Holiday	0	No	love_story	Yes	Yes	No	Aashish Chandra	Aashish Chandra	Sajid - Wajid	25	1.25e+07	1375000
24. Three - Line Line Betrayal	Normal	0	No	thriller	No	Yes	No	Aashish Choudhary	Vishal Pandya	Chirantan Bharti	325	4.00e+07	10000000
25. Mangam Me Gadhral Ghotela	Normal	0	No	comedy	Yes	Yes	No	Aashish Choudhary	Anwar Shahid	Sahil Ravi	5	1.00e+06	300000
26. Khushi Sikha - Isak Ki Tharth	Normal	0	No	comedy	Yes	Yes	Yes	Aashish Choudhary	Ram Prasad Singh	Uditvish	50	6.00e+06	600000
27. Hi Fi Log	Holiday	0	No	adult	Yes	Yes	No	Aashish Choudhary	Baba Jagirdar	Baba Jagirdar	2	1.50e+06	500000
28. Dev D	Holiday	0	No	drama	No	No	No	Aashish Choudhary	Anurag Kashyap	Anil Thedi	525	1.10e+08	20020000
29. One By One	Normal	0	No	rom_com	No	No	No	Aashish Choudhary	Devika Baghel	Shankar - Ehsaan - Loy	500	2.10e+08	37180000
30. One Lucky Lucky Oye!	Normal	0	No	comedy	No	No	No	Aashish Choudhary	Shankar - Ehsaan - Loy	Shankar - Ehsaan - Loy	400	1.10e+08	100020000
31. Sacha Na Tha	Holiday	0	No	rom_com	Yes	Yes	No	Aashish Choudhary	Uditvish	Sandesh Shandilya	110	5.00e+07	5280000
32. Happy Bhag Singh	Normal	0	No	rom_com	No	No	No	Aashish Choudhary	Mudassar Aziz	Sohail Sen	1000	2.00e+08	391500000
33. Manorama Six Feet Under	Normal	0	No	suspense	No	Yes	Yes	Aashish Choudhary	Navdeep Singh	Jayesh Gandhi	75	3.25e+07	5800000
34. Homymoon Travels Pvt. Ltd.	Normal	0	No	rom_com	No	Yes	No	Aashish Choudhary	Reema Kagti	Vishal - Shashar	350	1.00e+08	195160000

```
> #Transforming data into class variable
> Remake_class<-ifelse(movie$Whether.Remake=="1", "Remaked", "Not Remaked");
> movie <- data.frame(movie, Remake_class)
> View(movie)
```

	MovieName	ReleasePeriod	Whether.Remake	Whether.Franchise	Genre	New.Actor	New.Director	New.Music.Director	Lead.Star	Director	Music.Director	Number.of.Screens	Revenue.INR.	Budget.INR.	Remake_class
1	Golden Boys	Normal	0	No	suspense	Yes	No	No	Jeet Gossami	Ravi Varma	Baba Jagirdar	5	5.00e+06	85000	Not Remaked
2	Kaccha Limboo	Holiday	0	No	drama	Yes	No	Yes	Karan Sharmachhi	Sagar Bairyati	Amardeep Nijjer	75	1.50e+07	625000	Not Remaked
3	Not A Love Story	Holiday	0	No	thriller	No	No	No	Mahir Gil	Ram Gopal Verma	Sandeep Chowta	525	7.50e+07	94700000	Not Remaked
4	Qaid Band	Holiday	0	No	drama	Yes	No	No	Aadar Jain	Habib Faizal	Amit Trivedi	800	2.10e+08	4500000	Not Remaked
5	Chashmal	Holiday	0	No	adult	Yes	Yes	Yes	Aadi Khan	Aadi Khan	Babloo Ullad	1	1.00e+06	1075000	Not Remaked
6	Shuttlcock Boys	Normal	0	No	comedy	Yes	Yes	Yes	Aakar Kaurthik	Hemant Gaba	Avinash Baghel	10	5.00e+06	170000	Not Remaked
7	Dirty Marriage	Holiday	0	No	adult	Yes	No	Yes	Aakash	Priyanka	Dharma	2	1.50e+06	35000	Not Remaked
8	Future To Bright Hai Ji	Holiday	0	No	drama	No	Yes	Yes	Aamir Bashir	Gargay Anwar	Amit Ali	30	1.50e+07	625000	Not Remaked
9	Ghagin	Holiday	1	No	action	No	Yes	No	Aamir Khan	A.R.Murugadas	A.R. Rahman	1550	5.20e+08	194550000	Remaked
10	Naam Caravan Par	Holiday	0	No	drama	No	Yes	No	Aamir Khan	Aamir Khan	Shankar - Ehsaan - Loy	500	1.80e+08	89570000	Not Remaked
11	Mange Pandey - The Rising	Holiday	0	No	drama	No	No	No	Aamir Khan	Karan Mehra	A.R. Rahman	675	3.70e+08	52705000	Not Remaked
12	Fanshi	Normal	0	No	love_story	No	No	No	Aamir Khan	Kunal Kohli	Jatin - Latt	850	3.00e+08	105400000	Not Remaked
13	Dargal	Holiday	0	No	drama	No	No	No	Aamir Khan	Nikhil Tawar	Pritam	4250	1.20e+09	700470000	Not Remaked
14	3 Idiots	Holiday	0	No	drama	No	No	No	Aamir Khan	Bhaskar Hirani	Shantanu Mohta	1750	7.10e+08	340105000	Not Remaked
15	PK	Normal	0	No	drama	No	No	No	Aamir Khan	Bhaskar Hirani	Ajay - Abul	3600	1.20e+09	616030000	Not Remaked
16	Kung De Bawari	Holiday	0	No	drama	No	No	No	Aamir Khan	Rakesh Omprakash Mehra	A.R. Rahman	600	2.80e+08	97900000	Not Remaked
17	Talaash	Normal	0	No	thriller	No	No	No	Aamir Khan	Neema Kagti	Ran Sanyal	2500	7.10e+08	1708717500	Not Remaked
18	Dhoom 3	Normal	0	Yes	action	No	No	No	Aamir Khan	Vijay Krishna Acharya	Pritam	3650	1.75e+09	524370000	Not Remaked
19	Pyar Mein Aisa Hota Hai	Holiday	0	No	thriller	Yes	Yes	Yes	Anamika Datta	Rajesh Kumar	Shabbir Asmi	5	5.50e+06	350000	Not Remaked
20	Gully Boydolan - A Fight For Right	Normal	0	No	drama	Yes	Yes	Yes	Aarun Nagar	Aarun Nagar	Ashish Donald	5	6.50e+06	575000	Not Remaked
21	Yeh Hai High Society	Holiday	0	No	thriller	No	No	No	Aaryan	Ravi Sinha	S. Paul	20	6.50e+06	4000000	Not Remaked
22	Say Yes To Love	Normal	0	No	love_story	Yes	No	No	Aasad Mirza	Mahesh Mirza	Jatin Pandit	30	1.75e+07	525000	Not Remaked
23	Dreams - Sapnay Sach Honga...	Holiday	0	No	love_story	Yes	Yes	No	Aashish Chavana	Aashish Chavana	Sajid - Waqid	25	1.25e+07	1375000	Not Remaked
24	Three - Love Lies Betrayal	Normal	0	No	thriller	No	Yes	No	Aashish Chauthary	Vishal Pandya	Chirantan Bharti	325	4.00e+07	10000000	Not Remaked
25	Maangam Me Gadbad Gholata	Normal	0	No	comedy	Yes	Yes	No	Aash Aliela	Anwar Sheikh	Sami Palli	5	1.00e+06	300000	Not Remaked
26	Khuda Sikka - Just Ki Thaath	Normal	0	No	comedy	Yes	Yes	Yes	Aash Kumar	Ram Pratap Singh	Uditvhar	50	9.00e+06	600000	Not Remaked
27	Hi Fi Log	Holiday	0	No	adult	Yes	Yes	Yes	Ashay Baskhi	Baba Jagirdar	Baba Jagirdar	2	1.50e+06	500000	Not Remaked
28	Dev D	Holiday	0	No	drama	No	No	No	Ashay Desai	Anurag Kashyap	Amit Trivedi	525	1.10e+08	20620000	Not Remaked
29	One By Two	Normal	0	No	rom_com	No	Yes	No	Ashay Desai	Devika Bhagat	Shankar - Ehsaan - Loy	500	2.10e+08	37700000	Not Remaked
30	Oye Lucky! Lucky Oye!	Normal	0	No	comedy	No	No	No	Ashay Desai	Dibakar Banerjee	Sneha Khanwalkar	400	1.20e+08	10062000	Not Remaked
31	Soccha Na Tha	Holiday	0	No	rom_com	Yes	Yes	No	Ashay Desai	Imtiaz Ali	Sandeep Shandilya	110	5.00e+07	52600000	Not Remaked
32	Happy Bhag Jayegi	Normal	0	No	rom_com	No	No	No	Ashay Desai	Mudassar Aziz	Sohail Sen	1000	2.00e+08	391500000	Not Remaked
33	Manorama Six Feet Under	Normal	0	No	suspense	No	Yes	Yes	Ashay Desai	Navdeep Singh	Jayesh Gandhi	75	3.25e+07	5800000	Not Remaked
34	Honeymoon Travels Pvt. Ltd.	Normal	0	No	rom_com	No	Yes	No	Ashay Desai	Reema Kagti	Vishal - Shekhar	350	1.00e+08	195160000	Not Remaked

```
> #Removing Whether.Remake from Data Frame
> movie <- movie[,-3]
```


4) Splitting the Dataset

```
> #Splitting the Dataset
> library(caTools)
Warning message:
package 'caTools' was built under R version 4.1.1
> set.seed(2)
> id <- sample(2, nrow(movie), prob=c(0.7, 0.3), replace=TRUE)
> print(id)
[1] 1 2 1 1 2 2 1 2 1 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 1 1 2 2 2 1 1 2 1 1 1 2 2 1 1 1 2 1
[56] 2 2 2 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[111] 1 2 1 1 2 2 2 1 2 1 1 2 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[166] 1 1 2 1 1 2 1 1 2 1 2 1 1 1 2 1 2 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1
[221] 1 2 1 1 1 2 2 1 2 2 1 1 2 1 1 1 1 1 2 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[276] 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[331] 1 2 2 1 1 1 1 1 2 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[386] 1 1 1 2 2 2 2 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[441] 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[496] 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[551] 2 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[606] 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[661] 2 1 1 2 1 1 2 1 1 1 2 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[716] 1 2 1 1 1 2 2 2 1 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[771] 1 1 1 1 1 2 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[826] 1 1 1 2 1 1 2 1 1 1 2 2 1 1 2 1 1 1 1 1 2 2 2 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[881] 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[936] 1 2 2 1 1 2 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[991] 1 2 1 1 1 2 1 1
[ reached getOption("max.print") -- omitted 698 entries ]
```

5) Training and Testing

```
> #Training and Testing
> movie_train <- movie[id==1,]
> movie_test <- movie[id==2,]
> print(movie_train)
```

	Movie.Name	Release.Period	Whether.Franchise	Genre	New.Actor	New.Director
1	Golden Boys	Normal	No	suspense	Yes	No
3	Not A Love Story	Holiday	No	thriller	No	No
4	Qaidi Band	Holiday	No	drama	Yes	No
7	Dirty Marriage	Holiday	No	adult	Yes	No
9	Ghajini	Holiday	No	action	No	Yes
10	Taare Zameen Par	Holiday	No	drama	No	Yes
11	Mangal Pandey - The Rising	Holiday	No	drama	No	No
12	Fanaa	Normal	No	love_story	No	No
14	3 Idiots	Holiday	No	drama	No	No
15	PK	Normal	No	drama	No	No
18	Dhoom 3	Normal	Yes	action	No	No
19	Pyaar Mein Aisa Hota Hai	Holiday	No	thriller	Yes	Yes
20	Gurjar Aandolan - A Fight For Right	Normal	No	drama	Yes	Yes
21	Yehi Hai High Society	Holiday	No	thriller	No	No
22	Say Yes To Love	Normal	No	love_story	Yes	No
24	Three - Love Lies Betrayal	Normal	No	thriller	No	Yes
25	Malegaon Me Gadbad Ghotala	Normal	No	comedy	Yes	Yes
26	Khota Sikka - Jaat Ke Thaath	Normal	No	comedy	Yes	Yes
27	Hi Fi Log	Holiday	No	adult	Yes	Yes
28	Dev D	Holiday	No	drama	No	No
30	Oye Lucky, Lucky Oye!	Normal	No	comedy	No	No

1) Building the model

```
Y               [,1]      [,2]
  Not Remaked 229272959 618319871
  Remaked    636630559 760895751

> pmodel <- predict(model, movie_test)
```

2) Building the Confusion Matrix

```
> #Building the confusion matrix
> table(pmodel, movie_test$Remake_class)

pmodel      Not Remaked Remaked
  Not Remaked      467      12
  Remaked          33      12
> (467+12)/(467+12+12+33)
[1] 0.9141221
```

ANALYSIS/CONCLUSION

- We concluded and analyzed whether the movie will be a remake or not according to the number of screens it was played as the people liked it and the revenue made by the movie by the help of the model.
- Movie played on more the number of screens more the revenue made by it.

REFERENCES

- 1) <https://en.wikipedia.org/>
- 2) <https://www.kaggle.com/>

