
# EXPLORATORY DATA ANALYSIS OF MOVIES

SUBMITTED TO:

PROF. SURAJ PATIL

SUBMITTED BY:

VED BHATKAR(B205)

ARIHANT JAIN(B220)

VARAD

NILAKHE(B238)

# INDEX

# 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 analysing the film genre, the release date around holidays, the release month of movies, the languages and country with more movies from the movie review dataset.

# LITERATURE REVIEW

- Much research has been undertaken into predicting movie box office revenue and success.
- Although the literature covers a varied range of research, this review will focus on the collection of movie data, techniques used to predict box office revenue and success, findings and results that have been found and benefits of movie box office revenue and success.
- Data is available after the release date of the movie. Hur et al. (2016) found that once a movie was released, data on audiences and weekly box office total were made available. It is clear that, after a movie has been released, there is some data that can be collected. It would make sense to use data that is released before a movie as it gives enough time to movie studios to plan and decide on marketing budgets. There may be plenty of data available after the release of a movie but it would be too late for movie studios to do anything with that information. It would be best to stick with pre-release data.

# METHODOLOGY

- 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 eachother 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.

# IMPLEMENTATION

- **Importing the dataset**

```
library(dplyr) movie<-read.csv("D:\\DM\\PROJECT\\movie.csv")
View(movie)
str(movie)
summary(movie)
```

| | Movie.Name | Release.Period | Whether.Remake | Whether.Franchise | Genre | New.Actor | New.Director | New.Music.Director | Lead.Star |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Golden Boys | Normal | No | No | suspense | Yes | No | No | Jeet Go |
| 2 | Kaccha Limboo | Holiday | No | No | drama | Yes | No | Yes | Karan B |
| 3 | Not A Love Story | Holiday | No | No | thriller | No | No | No | Mahie ( |
| 4 | Qaidi Band | Holiday | No | No | drama | Yes | No | No | Aadar J |
| 5 | Chaatwali | Holiday | No | No | adult | Yes | Yes | Yes | Aadil Kl |
| 6 | Shuttlecock Boys | Normal | No | No | comedy | Yes | Yes | Yes | Aakar K |
| 7 | Dirty Marriage | Holiday | No | No | adult | Yes | No | Yes | Aakash |
| 8 | Future To Bright Hai Ji | Holiday | No | No | drama | No | Yes | Yes | Aamir E |
| 9 | Ghajini | Holiday | Yes | No | action | No | Yes | No | Aamir K |
| 10 | Taare Zameen Par | Holiday | No | No | drama | No | Yes | No | Aamir K |
| 11 | Mangal Pandey - The Rising | Holiday | No | No | drama | No | No | No | Aamir K |
| 12 | Fanaa | Normal | No | No | love_story | No | No | No | Aamir K |
| 13 | Dangal | Holiday | No | No | drama | No | No | No | Aamir K |
| 14 | 3 Idiots | Holiday | No | No | drama | No | No | No | Aamir K |
| 15 | PK | Normal | No | No | drama | No | No | No | Aamir K |
| 16 | Rang De Basanti | Holiday | No | No | drama | No | No | No | Aamir K |
| 17 | Talaash | Normal | No | No | thriller | No | No | No | Aamir K |
| 18 | Dhoom 3 | Normal | No | Yes | action | No | No | No | Aamir K |
| 19 | Pyaar Mein Aisa Hota Hai | Holiday | No | No | thriller | Yes | Yes | Yes | Aanami |
| 20 | Gurjar Aandolan - A Fight For Right | Normal | No | No | drama | Yes | Yes | Yes | Aarun M |
| 21 | Yehi Hai High Society | Holiday | No | No | thriller | No | No | No | Aaryan |
| 22 | Say Yes To Love | Normal | No | No | love_story | Yes | No | No | Aasad M |
| 23 | Dreams - Sapnay Sach Honge... | Holiday | No | No | love_story | Yes | Yes | No | Aashish |
| 24 | Three - Love Lies Betrayal | Normal | No | No | thriller | No | Yes | No | Aashish |
| 25 | Malegaon Me Gadbad Ghotala | Normal | No | No | comedy | Yes | Yes | No | Aasif Al |
| 26 | Khota Sikka - Jaat Ke Thaath | Normal | No | No | comedy | Yes | Yes | Yes | Aatri Ku |
| 27 | Hi Fi Log | Holiday | No | No | adult | Yes | Yes | No | Abhay I |
| 28 | Dev D | Holiday | No | No | drama | No | No | No | Abhay I |
| 29 | One By Two | Normal | No | No | rom__com | No | Yes | No | Abhay I |
| 30 | Oye Lucky! Lucky Oye! | Normal | No | No | comedy | No | No | No | Abhay I |
| 31 | Socha Na Tha | Holiday | No | No | rom__com | Yes | Yes | No | Abhay I |
| 32 | Happy Bhaag Jayegi | Normal | No | No | rom__com | No | No | No | Abhay I |
| 33 | Manorama Six Feet Under | Normal | No | No | suspense | No | Yes | Yes | Abhay I |

## Data Preprocessing

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.)
```

## Data Visualization

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

#Designing All the Plots
library(ggplot2)

#a Display bar chart
ggplot(data=movie,aes(Genre))+geom_bar()
```

## Data Visualization

#b Design Histogram

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

#c Design Scatter Plot

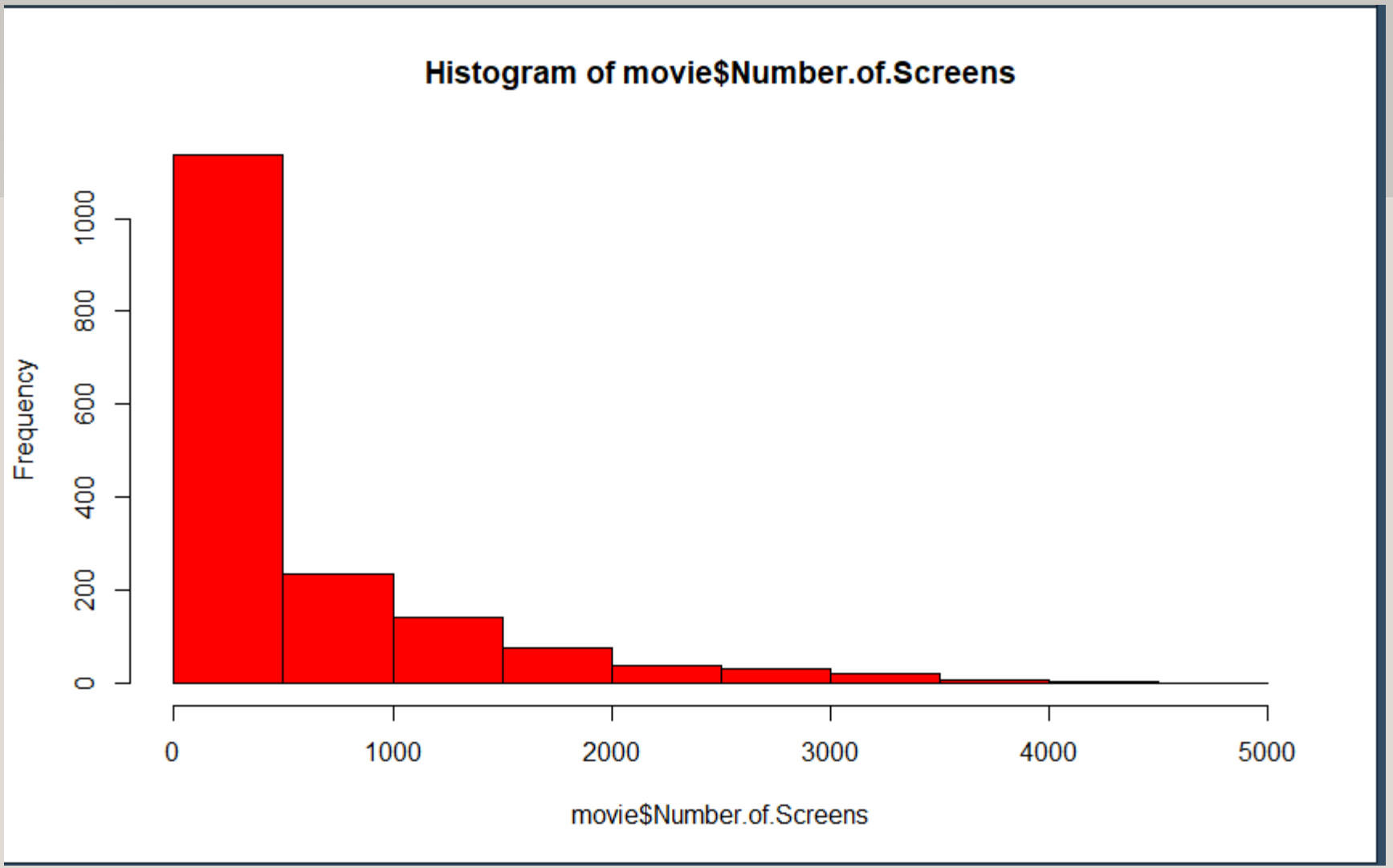
```
ggplot(data =movie, aes(x = Budget.INR., y = Revenue.INR., col = Genre))+geom_point()
```

#d Design Box Plot

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

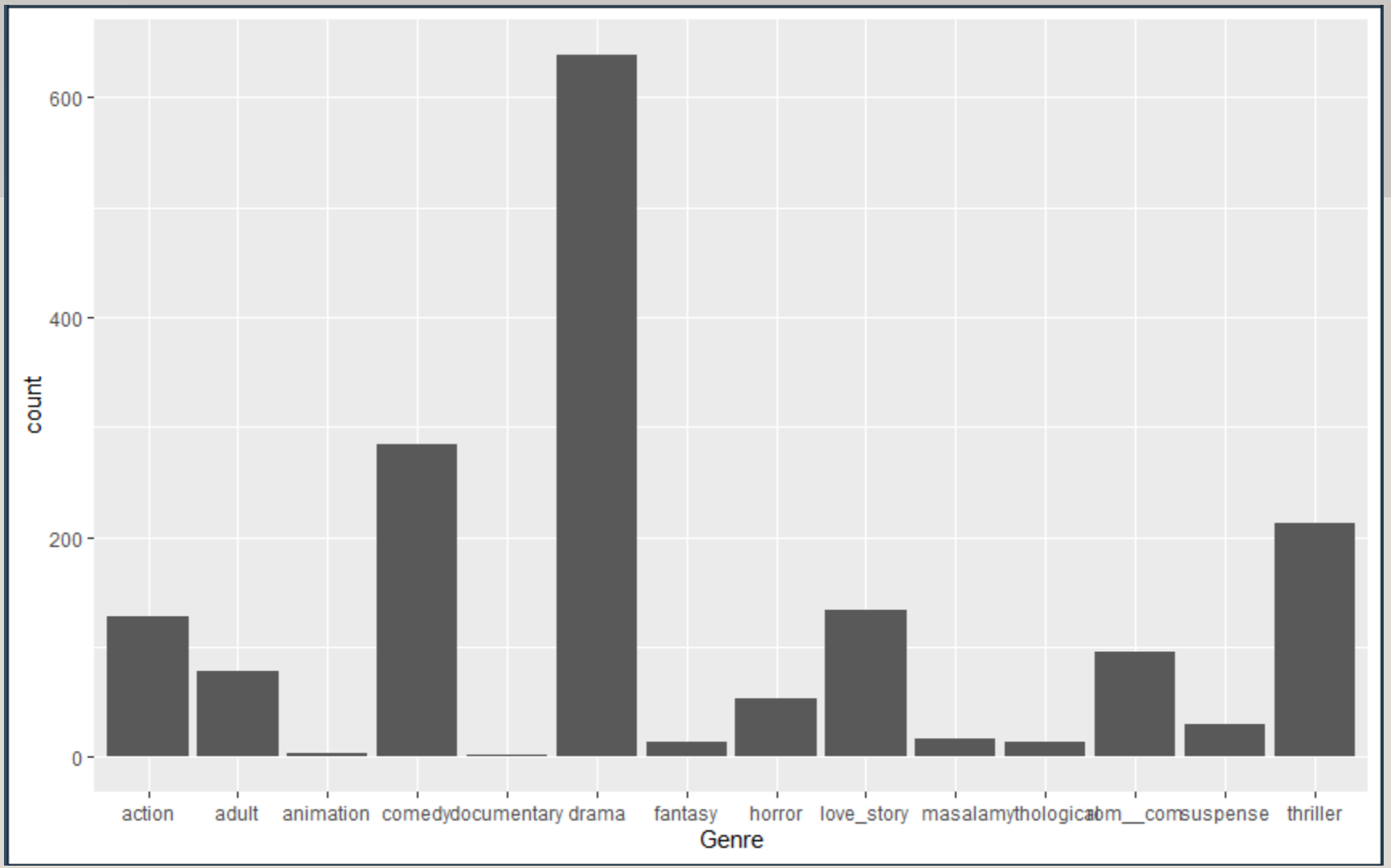
#Designing the histogram

hist(movie$Number.of.Screens, col = "red")



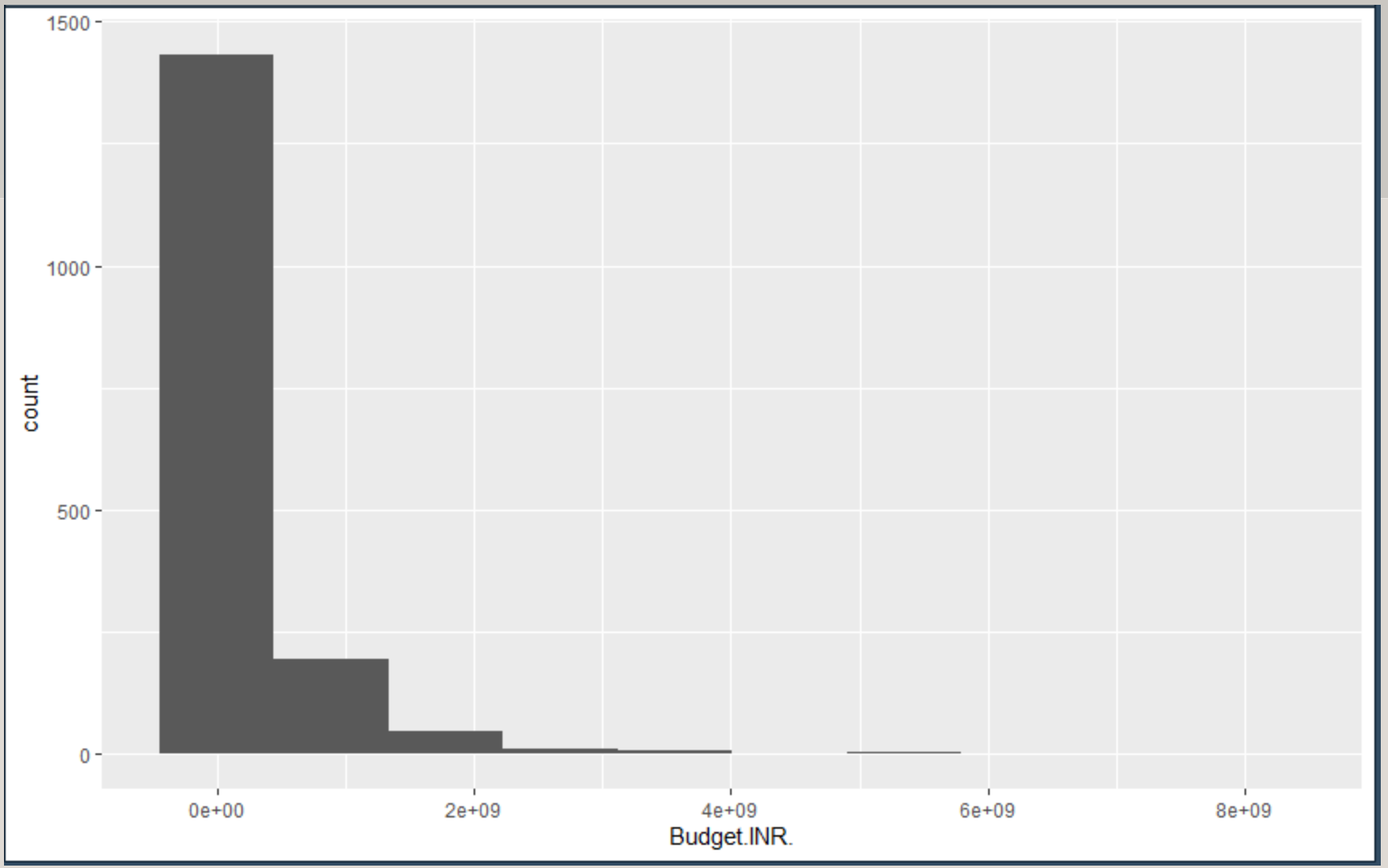
#a Display bar chart

ggplot(data=movie,aes(Genre))+geom_bar()



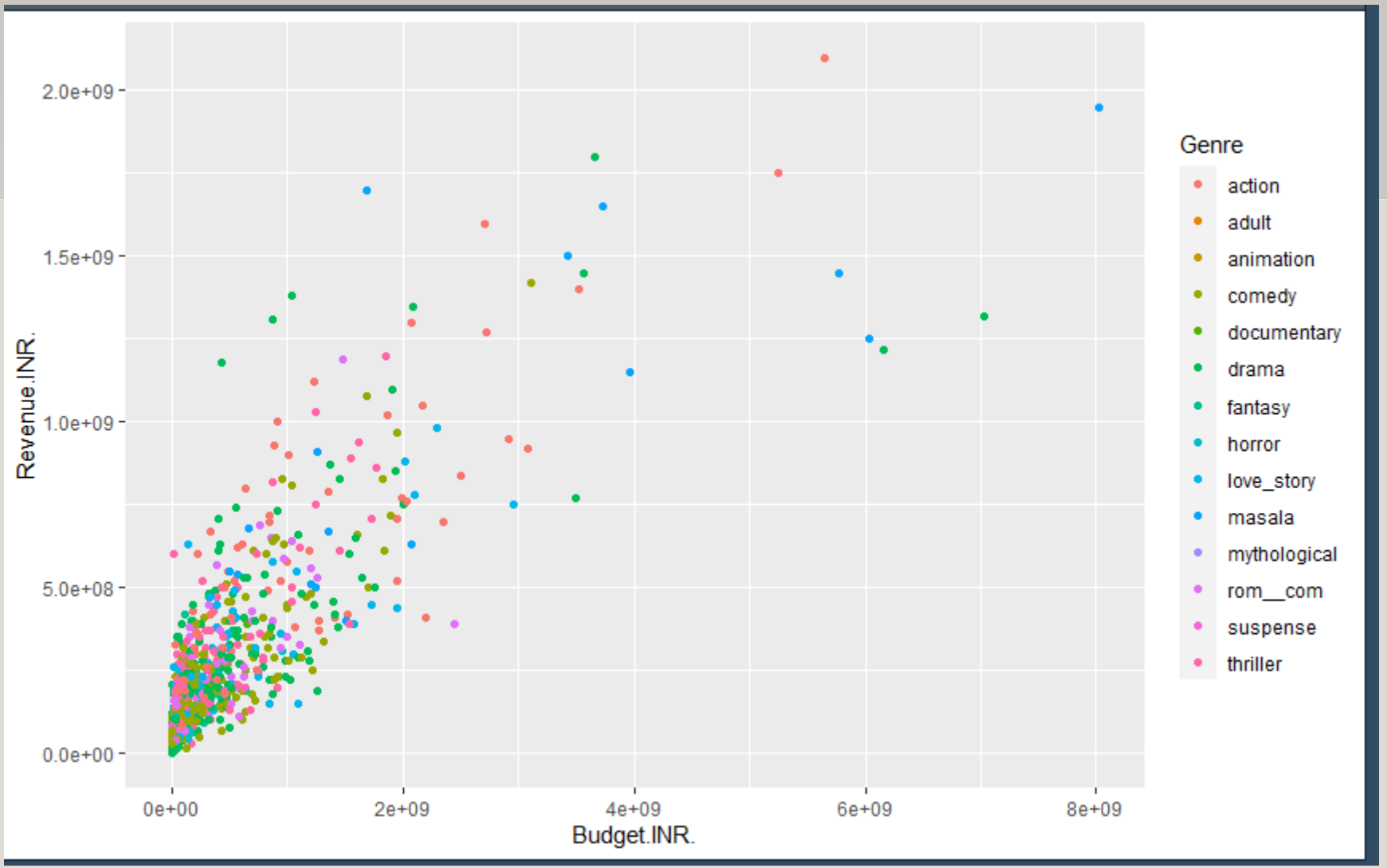
#b Design Histogram

ggplot(data = movie, aes(Budget.INR.))+geom_histogram(bins = 10)



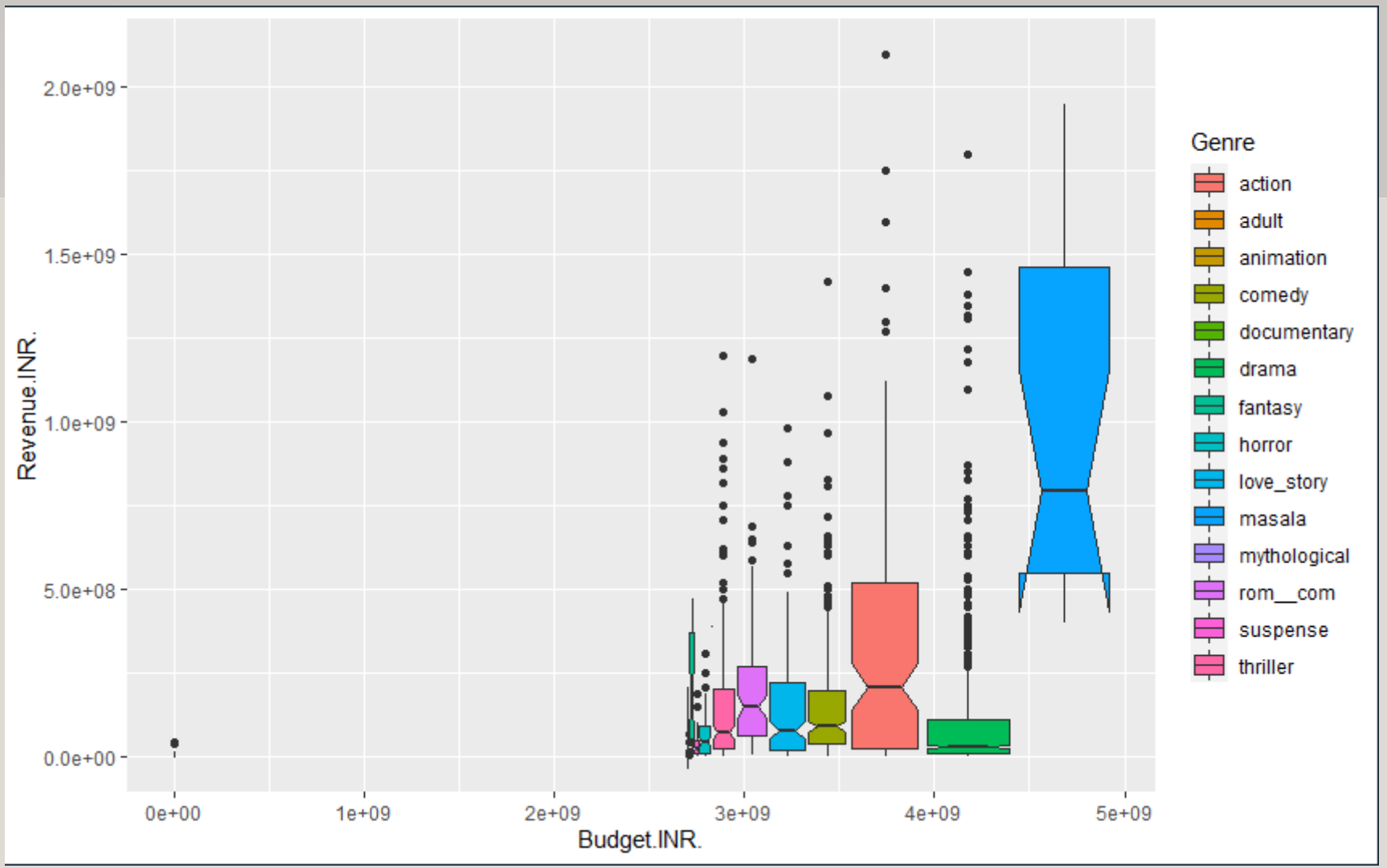
#c Design Scatter Plot

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]
```

Project.R | movie

Filter

| Actor | New.Director | New.Music.Director | Lead.Star | Director | Music.Director | Number.of.Screens | Revenue.INR. | Budget.INR. | Remake_class |
|---|---|---|---|---|---|---|---|---|---|
| | No | No | Jeet Goswami | Ravi Varma | Baba Jagirdar | 5 | 5.00e+06 | 85000 | Not Remaked |
| | No | Yes | Karan Bhanushali | Sagar Ballary | Amardeep Nijjer | 75 | 1.50e+07 | 825000 | Not Remaked |
| | No | No | Mahie Gill | Ram Gopal Verma | Sandeep Chowta | 525 | 7.50e+07 | 56700000 | Not Remaked |
| | No | No | Aadar Jain | Habib Faisal | Amit Trivedi | 800 | 2.10e+08 | 4500000 | Not Remaked |
| | Yes | Yes | Aadil Khan | Aadil Khan | Babloo Ustad | 1 | 1.00e+06 | 1075000 | Not Remaked |
| | Yes | Yes | Aakar Kaushik | Hemant Gaba | Avinash Baghel | 10 | 5.00e+06 | 170000 | Not Remaked |
| | No | Yes | Aakash | Priyanka | Dharma | 2 | 1.50e+06 | 35000 | Not Remaked |
| | Yes | Yes | Aamir Bashir | Sanjay Amar | Amir Ali | 30 | 1.50e+07 | 825000 | Not Remaked |
| | Yes | No | Aamir Khan | A.R. Murugadoss | A.R. Rehman | 1550 | 5.20e+08 | 1945820000 | Remaked |
| | Yes | No | Aamir Khan | Aamir Khan | Shankar - Ehsaan - Loy | 500 | 1.80e+08 | 875785000 | Not Remaked |
| | No | No | Aamir Khan | Ketan Mehta | A.R. Rehman | 675 | 3.70e+08 | 525785000 | Not Remaked |
| | No | No | Aamir Khan | Kunal Kohli | Jatin - Lalit | 650 | 3.00e+08 | 1054800000 | Not Remaked |
| | No | No | Aamir Khan | Nitish Tiwari | Pritam | 4250 | 1.32e+09 | 7024750000 | Not Remaked |
| | No | No | Aamir Khan | Rajkumar Hirani | Shantanu Moitra | 1750 | 7.70e+08 | 3491035000 | Not Remaked |
| | No | No | Aamir Khan | Rajkumar Hirani | Ajay - Atul | 3600 | 1.22e+09 | 6160362500 | Not Remaked |
| | No | No | Aamir Khan | Rakesh Omprakash Mehra | A.R. Rehman | 600 | 2.80e+08 | 979080000 | Not Remaked |
| | No | No | Aamir Khan | Reema Kagti | Ram Sampath | 2550 | 7.10e+08 | 1729717500 | Not Remaked |
| | No | No | Aamir Khan | Vijay Krishna Acharya | Pritam | 3650 | 1.75e+09 | 5243760000 | Not Remaked |
| | Yes | Yes | Aanamika Bawa | Rajesh Kumar | Shabab Azmi | 5 | 5.50e+06 | 350000 | Not Remaked |
| | Yes | Yes | Aarun Nagar | Aarun Nagar | Ashish Donald | 5 | 6.50e+06 | 575000 | Not Remaked |
| | No | No | Aaryan | Ravi Sinha | S. Paul | 20 | 6.50e+06 | 4000000 | Not Remaked |
| | No | No | Aasad Mirza | Mahrukh Mirza | Jatin Pandit | 30 | 1.75e+07 | 525000 | Not Remaked |
| | Yes | No | Aashish Chanana | Aashish Chanana | Sajid - Wajid | 25 | 1.25e+07 | 1375000 | Not Remaked |
| | Yes | No | Aashish Chaudhary | Vishal Pandya | Chirantan Bhatt | 325 | 4.00e+07 | 10000000 | Not Remaked |
| | Yes | No | Aasif Albela | Anwar Shaikh | Sahil Patil | 5 | 1.00e+06 | 300000 | Not Remaked |
| | Yes | Yes | Aatri Kumar | Ram Pratap Singh | Udbhav | 50 | 9.00e+06 | 600000 | Not Remaked |
| | Yes | No | Abhay Bakshi | Baba Jagirdar | Baba Jagirdar | 2 | 1.50e+06 | 500000 | Not Remaked |
| | No | No | Abhay Deol | Anurag Kashyap | Amit Trivedi | 525 | 1.10e+08 | 208200000 | Not Remaked |
| | Yes | No | Abhay Deol | Devika Bhagat | Shankar - Ehsaan - Loy | 500 | 2.10e+08 | 37700000 | Not Remaked |
| | No | No | Abhay Deol | Dibakar Banerjee | Sneha Khanwalkar | 400 | 1.20e+08 | 100620000 | Not Remaked |
| | Yes | No | Abhay Deol | Imtiaz Ali | Sandesh Shandilya | 110 | 5.00e+07 | 52800000 | Not Remaked |
| | No | No | Abhay Deol | Mudassar Aziz | Sohail Sen | 1000 | 2.00e+08 | 391500000 | Not Remaked |
| | Yes | Yes | Abhay Deol | Navdeep Singh | Jayesh Gandhi | 75 | 3.25e+07 | 5800000 | Not Remaked |

## ➢ Splitting the Dataset

```
library(caTools)
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 2
 [38] 1 1 1 2 1 1 1 2 2 2 1 1 2 1 1 1 2 1 2 2 2 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1
 [75] 1 2 1 2 1 1 1 1 1 1 1 2 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1
[112] 2 1 1 2 2 2 1 2 1 1 1 2 2 1 2 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1
[149] 1 1 1 1 2 1 2 1 2 1 1 1 2 2 1 2 1 1 1 2 1 1 2 1 1 2 1 2 1 1 1 1 2 1 2 2 1
[186] 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 2
[223] 1 1 1 2 2 1 2 2 1 1 2 1 2 1 1 1 1 1 2 1 1 2 2 2 1 1 1 1 2 1 1 1 2 1 1 1 1
[260] 1 2 1 1 1 2 1 1 1 2 1 2 2 2 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 2 1
[297] 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2
[334] 1 1 1 1 1 2 2 2 2 2 1 1 1 1 2 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1
[371] 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 1 1 2 1 2 1 1 1 1 2 2 1
[408] 1 1 2 1 2 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 1 2 2 1 1 1 2 1
[445] 1 1 2 1 1 1 1 2 1 1 2 2 1 2 2 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1
[482] 2 2 2 2 1 2 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1
[519] 1 2 2 1 1 2 1 1 1 1 1 2 1 2 2 1 1 2 2 1 1 2 1 2 1 2 1 1 1 2 1 2 2 1 1 1 2
[556] 1 1 1 1 2 2 1 2 1 2 1 1 1 2 1 2 1 2 2 1 2 2 1 1 1 1 1 2 1 1 2 1 1 2 2 1 1
[593] 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 2 1
[630] 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 2 1 1 2 1 1 2 1 2 1 2 1 2 1 1 1 2 1 1 2 1 1
[667] 2 1 2 1 1 1 2 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 2 2 2
[704] 1 1 1 2 1 2 1 2 1 1 1 2 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
[741] 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 2 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2
[778] 2 2 1 1 1 2 1 2 1 2 1 1 1 1 2 1 1 2 1 1 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 2
[815] 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 2 1 1 2 1 1 1 2 2 2 1 2 1 2
[852] 1 2 1 2 1 1 2 1 1 1 1 1 2 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[889] 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1
[926] 1 1 1 2 1 1 2 1 2 1 1 2 2 1 1 2 1 2 1 2 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2
[963] 1 1 1 2 1 2 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 2 1
[1000] 1
 [ reached getOption("max.print") -- omitted 698 entries ]
```

## Training and Testing

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


View(movie_train)
View(movie_test)
```

Project.R* | movie_train | movie

Filter

| | Movie.Name | Release.Period | Whether.Franchise | Genre | New.Actor | New.Director | New.Music.Director | Lead.Star | D |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Golden Boys | Normal | No | suspense | Yes | No | No | Jeet Goswami | R |
| 3 | Not A Love Story | Holiday | No | thriller | No | No | No | Mahie Gill | R |
| 4 | Qaidi Band | Holiday | No | drama | Yes | No | No | Aadar Jain | H |
| 7 | Dirty Marriage | Holiday | No | adult | Yes | No | Yes | Aakash | P |
| 9 | Ghajini | Holiday | No | action | No | Yes | No | Aamir Khan | A |
| 10 | Taare Zameen Par | Holiday | No | drama | No | Yes | No | Aamir Khan | A |
| 11 | Mangal Pandey - The Rising | Holiday | No | drama | No | No | No | Aamir Khan | K |
| 12 | Fanaa | Normal | No | love_story | No | No | No | Aamir Khan | K |
| 14 | 3 Idiots | Holiday | No | drama | No | No | No | Aamir Khan | R |
| 15 | PK | Normal | No | drama | No | No | No | Aamir Khan | R |
| 18 | Dhoom 3 | Normal | Yes | action | No | No | No | Aamir Khan | V |
| 19 | Pyaar Mein Aisa Hota Hai | Holiday | No | thriller | Yes | Yes | Yes | Aanamika Bawa | R |
| 20 | Gurjar Aandolan - A Fight For Right | Normal | No | drama | Yes | Yes | Yes | Aarun Nagar | A |
| 21 | Yehi Hai High Society | Holiday | No | thriller | No | No | No | Aaryan | R |
| 22 | Say Yes To Love | Normal | No | love_story | Yes | No | No | Aasad Mirza | M |
| 24 | Three - Love Lies Betrayal | Normal | No | thriller | No | Yes | No | Aashish Chaudhary | V |
| 25 | Malegaon Me Gadbad Ghotala | Normal | No | comedy | Yes | Yes | No | Aasif Albela | A |
| 26 | Khota Sikka - Jaat Ke Thaath | Normal | No | comedy | Yes | Yes | Yes | Aatri Kumar | R |
| 27 | Hi Fi Log | Holiday | No | adult | Yes | Yes | No | Abhay Bakshi | B |
| 28 | Dev D | Holiday | No | drama | No | No | No | Abhay Deol | A |
| 30 | Oye Lucky! Lucky Oye! | Normal | No | comedy | No | No | No | Abhay Deol | D |
| 31 | Socha Na Tha | Holiday | No | rom__com | Yes | Yes | No | Abhay Deol | In |
| 32 | Happy Bhaag Jayegi | Normal | No | rom__com | No | No | No | Abhay Deol | M |
| 35 | Ek Chalis Ki Last Local | Normal | No | comedy | No | Yes | No | Abhay Deol | S |
| 36 | Ahista Ahista | Holiday | No | love_story | No | Yes | No | Abhay Deol | S |
| 38 | Roar - Tigers Of The Sundarbans | Normal | No | thriller | Yes | Yes | Yes | Abhinav Shukla | K |
| 39 | Chakradhaar | Normal | No | action | Yes | Yes | No | Abhishek Anand | D |
| 40 | Players | Normal | No | action | No | No | No | Abhishek Bachchan | A |
| 42 | Khelein Hum Jee Jaan Sey | Normal | No | drama | No | No | No | Abhishek Bachchan | A |
| 43 | Drona | Holiday | No | action | No | No | No | Abhishek Bachchan | G |
| 44 | Guru | Holiday | No | drama | No | No | No | Abhishek Bachchan | M |
| 48 | Dum Maaro Dum | Holiday | No | thriller | No | No | No | Abhishek Bachchan | R |
| 49 | Bunty Aur Babli | Normal | No | comedy | No | No | No | Abhishek Bachchan | S |

Project.R* | movie_test | movie_train | movie

Filter

| | Movie.Name | Release.Period | Whether.Franchise | Genre | New.Actor | New.Director | New.Music.Director | Lead.Star | Dire |
|---|---|---|---|---|---|---|---|---|---|
| 2 | Kaccha Limboo | Holiday | No | drama | Yes | No | Yes | Karan Bhanushali | Sa |
| 5 | Chaatwali | Holiday | No | adult | Yes | Yes | Yes | Aadil Khan | Aa |
| 6 | Shuttlecock Boys | Normal | No | comedy | Yes | Yes | Yes | Aakar Kaushik | He |
| 8 | Future To Bright Hai Ji | Holiday | No | drama | No | Yes | Yes | Aamir Bashir | Sa |
| 13 | Dangal | Holiday | No | drama | No | No | No | Aamir Khan | Ni |
| 16 | Rang De Basanti | Holiday | No | drama | No | No | No | Aamir Khan | Ra |
| 17 | Talaash | Normal | No | thriller | No | No | No | Aamir Khan | Re |
| 23 | Dreams - Sapnay Sach Honge... | Holiday | No | love_story | Yes | Yes | No | Aashish Chanana | Aa |
| 29 | One By Two | Normal | No | rom__com | No | Yes | No | Abhay Deol | De |
| 33 | Manorama Six Feet Under | Normal | No | suspense | No | Yes | Yes | Abhay Deol | Na |
| 34 | Honeymoon Travels Pvt. Ltd. | Normal | No | rom__com | No | Yes | No | Abhay Deol | Re |
| 37 | Lottery | Holiday | No | drama | Yes | Yes | Yes | Abhijeet Sawant | He |
| 41 | Game | Holiday | No | thriller | No | Yes | No | Abhishek Bachchan | Ab |
| 45 | Raavan | Normal | No | action | No | No | No | Abhishek Bachchan | Ma |
| 46 | Delhi 6 | Holiday | No | drama | No | No | No | Abhishek Bachchan | Ra |
| 47 | Bluffmaster | Normal | No | thriller | No | No | No | Abhishek Bachchan | Ro |
| 50 | Jhoom Barabar Jhoom | Normal | No | rom__com | No | No | No | Abhishek Bachchan | Sh |
| 54 | Naya Pata | Holiday | No | drama | No | Yes | Yes | Abhishek Sharma | Pa |
| 56 | Hum Hai Raahi Car Ke | Holiday | No | rom__com | No | No | No | Adah Sharma | Jy |
| 57 | Strings - Bound By Faith | Normal | No | drama | Yes | No | Yes | Adam Bedi | Sa |
| 58 | Contract | Normal | No | thriller | Yes | No | No | Adhvik Mahajan | Ra |
| 60 | Dehraadun Diary | Normal | No | thriller | No | No | No | Adhyayan Suman | Mi |
| 61 | Jashnn | Normal | No | drama | No | No | No | Adhyayan Suman | Ra |
| 62 | Heartless | Normal | No | thriller | No | Yes | No | Adhyayan Suman | Sh |
| 65 | Lessons In Forgetting | Holiday | No | drama | No | Yes | No | Adil Hussain | Ur |
| 76 | Ek Aur Murder | Normal | No | adult | Yes | No | No | Affi Khan | Su |
| 78 | 1920 - Evil Returns | Holiday | Yes | horror | No | Yes | No | Aftab Shivdasani | Bh |
| 86 | Red - The Dark Side | Normal | No | thriller | No | No | No | Aftab Shivdasani | Vi |
| 87 | My Friend Ganesha | Normal | No | drama | No | No | Yes | Ahsaas Channa | Ra |
| 89 | Provoked | Normal | No | drama | No | No | No | Aishwarya Rai | Ja |
| 91 | U Me Aur Hum | Holiday | No | love_story | No | Yes | No | Ajay Devgn | Aj |
| 103 | Dil Toh Baccha Hai Ji | Normal | No | comedy | No | No | No | Ajay Devgn | Ma |
| 104 | Tango Charlie | Holiday | No | action | No | No | No | Ajay Devgn | Ma |

## Building the model

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

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

```
             Number.of.Screens
Y                      [,1]      [,2]
  Not Remaked  539.0515  778.7902
  Remaked     1237.4468 1039.5907

             Revenue.INR.
Y                      [,1]      [,2]
  Not Remaked 146742924 249254228
  Remaked     344202128 323090589

             Budget.INR.
Y                      [,1]      [,2]
  Not Remaked 229272959 618319871
  Remaked     636630559 760895751

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

# RESULT ANALYSIS

- There are attributes (country, languages, genre, movie release date, budget and revenue) taken from the dataset and the derived attributes (release month of the movie derived from release date of movie and profit from budget and revenue) is analyzed to determine the movie performance. The analyzed data is plotted in graphs for statistical observation of the movie success.

- **Building the Confusion Matrix**

table(pmodel, movie_test$Remake_class)

```
pmodel          Not Remaked Remaked
  Not Remaked           467       12
  Remaked                33       12
> accuracy = (467+12)/(467+12+12+33)
> (467+12)/(476+12+12+33)
[1] 0.8986867
```

# CONCLUSION

- We concluded 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


- https://en.wikipedia.org/
- https://www.kaggle.com/
- https://www.bartleby.com/