Report of IMDB Movie Dataset:

Data preprocessing, Linear regression modeling, and Data visualization.

NAME:- SRUSHTI.S.FULPAGAR

ROLL NO:- 3160

SUBJECT:- FOUNDATION OF DATA SCIENCE

CLASS:- DATA ANALYTICS

SOFTWARE USED:- R-STUDIO

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

ABSTRACT

The IMDb Movie Dataset is a comprehensive collection of data related to movies, encompassing a wide range of information on films released over several decades. This dataset offers valuable insights into the world of cinema, making it an essential resource for movie enthusiasts, researchers, and data analysts alike.

With details spanning from movie titles, release years, and runtimes to genre categorizations, IMDb ratings, and production company information, the dataset provides a rich and diverse set of attributes for each movie entry. It also includes data on cast and crew, allowing for exploration of the individuals involved in the filmmaking process.

This dataset serves as a versatile tool for conducting various analyses and investigations, enabling users to explore trends in movie production, popularity, and critical acclaim over time. Researchers can delve into questions related to genre preferences, the impact of directors and actors on movie success, and much more.

In summary, the IMDb Movie Dataset is a valuable resource for those interested in the cinematic world, offering a wealth of data that facilitates research, analysis, and a deeper understanding of the factors contributing to the success and diversity of movies in the global film industry.

1. **Dataset Name:** IMDB Movie Dataset Latest

2. **Source:** Kaggle

3. **Dataset Link:** https://www.kaggle.com/datasets/ayushjain001/imdb-movie-dataset-latest?resource=download

4. Dataset contains 11 columns and 1000 rows

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

RESEARCH METHODOLOGY

Installing Libraries

```
#Installing packages
install.packages('ggplot2') #visualization
install.packages('caTools') #conditional analysis tools
install.packages('dplyr') #manipulation
```

Loading Libraries

```
# Load necessary libraries
library(ggplot2)
library(dplyr)
library(caTools)
```

Loading IMDB Dataset

```
> # Load your IMDb dataset
> imdb_data <- read.csv("C:/IMDB.csv")
```

Showing the dimension of dataset

dim() function is used to retrieve the dimensions of our dataset.

```
> dim(IMDB)
[1] 1000   11
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

tail() it is used to view last five rows of a dataset

```
> #It will give us last value
> tail(IMDB)
        X            Name.of.movie Year.of.relase Watchtime Movie.Rating Metascore
995   994         K.G.F: Chapter 1           2018       156          8.3       NaN
996   995             Vikram Vedha           2017       147          8.3       NaN
997   996                 Drishyam           2013       160          8.3       NaN
998   997                   Jagten           2012       115          8.3        77
999   998 Jodaeiye Nader az Simin           2011       123          8.3        95
1000  999                Incendies           2010       131          8.3        80
        Votes Gross.collection
995    54,515              Nan
996    35,901              Nan
997    39,570             #244
998   313,647           $0.69M
999   238,130           $7.10M
1000  168,727           $6.86M
```

head() it is used to view first five rows of a dataset

```
> #It will give first values
> head(IMDB)
  X            Name.of.movie Year.of.relase Watchtime Movie.Rating Metascore
1 0                 Jai Bhim           2021       164          9.3       NaN
2 1 The Shawshank Redemption           1994       142          9.3        80
3 2            The Godfather           1972       175          9.2       100
4 3           Soorarai Pottru          2020       153          9.1       NaN
5 4           The Dark Knight          2008       152          9.0        84
6 5    The Godfather: Part II          1974       202          9.0        90
      Votes Gross.collection
1   173,295             #139
2 2,541,091           $28.34M
3 1,748,410          $134.97M
4   107,159              Nan
5 2,491,371          $534.86M
6 1,212,675           $57.30M
```

colnames() function is used to retrieve the column names.

```
> #Give column names
> colnames(IMDB)
 [1] "X"                "Name.of.movie"    "Year.of.relase"   "Watchtime"
 [5] "Movie.Rating"     "Metascore"        "Votes"            "Gross.collection"
 [9] "Description"      "Director"         "Star"
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

is.na() is used to check for missing values in a dataset.

```
> #is.na() is used to check for missing values in a dataset.
> sum(is.na(IMDB))
[1] 220
```

Checking Duplicate Values

```
> #Checking Duplicate Values
> sum(duplicated(IMDB))
[1] 0
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

DATA EXPLORATION

str() function is used to display the structure of dataset.

```
> str(imdb_data) # str() function is used to display the structure of dataset.
'data.frame':	1000 obs. of  11 variables:
 $ X              : int  0 1 2 3 4 5 6 7 8 9 ...
 $ Name.of.movie  : chr  "Jai Bhim" "The Shawshank Redemption" "The Godfather" "Soorarai Pottru" ...
 $ Year.of.relase : chr  "2021" "1994" "1972" "2020" ...
 $ Watchtime      : int  164 142 175 153 152 202 96 201 154 195 ...
 $ Movie.Rating   : num  9.3 9.3 9.2 9.1 9 9 9 8.9 8.9 8.9 ...
 $ Metascore      : num  NaN 80 100 NaN 84 90 96 94 94 94 ...
 $ Votes          : chr  "173,295" "2,541,091" "1,748,410" "107,159" ...
 $ Gross.collection: chr  "#139" "$28.34M" "$134.97M" "Nan" ...
 $ Description    : chr  "When a tribal man is arrested for a case of alleged theft, his wife turns to a human-rights lawyer to help bring justice." "Two 
ond over a number of years, finding solace and eventual redemption through acts of common decency." "The aging patriarch of an organized crime dynasty in p
City transfers control of his clandestine"| __truncated__ "Nedumaaran Rajangam \"Maara\" sets out to make the common man fly and in the process takes on th
"| __truncated__ ...
 $ Director       : chr  "T.J. Gnanavel" "Frank Darabont" "Francis Ford Coppola" "Sudha Kongara" ...
 $ Star           : chr  "['Suriya', 'Lijo Mol Jose', 'Manikandan', 'Rajisha Vijayan']" "['Tim Robbins', 'Morgan Freeman', 'Bob Gunton', 'William Sadler']
```

view() function opens the dataset in a separate interactive viewer window, making it easier to examine data.

```
> View(imdb_data)
```

	X	Name.of.movie	Year.of.relase	Watchtime	Movie.Rating	Metascore	Votes	Gross.collection	Description	Director
1	0	Jai Bhim	2021	164	9.3	NaN	173,295	#139	When a tribal man is arrested for a case of alleged theft, his ...	T.J. Gnanavel
2	1	The Shawshank Redemption	1994	142	9.3	80	2,541,091	$28.34M	Two imprisoned men bond over a number of years, finding ...	Frank Darabont
3	2	The Godfather	1972	175	9.2	100	1,748,410	$134.97M	The aging patriarch of an organized crime dynasty in postw...	Francis Ford Coppola
4	3	Soorarai Pottru	2020	153	9.1	NaN	107,159	Nan	Nedumaaran Rajangam "Maara" sets out to make the com...	Sudha Kongara
5	4	The Dark Knight	2008	152	9.0	84	2,491,371	$534.86M	When the menace known as the Joker wreaks havoc and ch...	Christopher Nolan
6	5	The Godfather: Part II	1974	202	9.0	90	1,212,675	$57.30M	The early life and career of Vito Corleone in 1920s New York...	Francis Ford Coppola
7	6	12 Angry Men	1957	96	9.0	96	750,853	$4.36M	The jury in a New York City murder trial is frustrated by a sin...	Sidney Lumet
8	7	The Lord of the Rings: The Return of the King	2003	201	8.9	94	1,752,093	$377.85M	Gandalf and Aragorn lead the World of Men against Sauron'...	Peter Jackson
9	8	Pulp Fiction	1994	154	8.9	94	1,955,203	$107.93M	The lives of two mob hitmen, a boxer, a gangster and his wif...	Quentin Tarantino
10	9	Schindler's List	1993	195	8.9	94	1,297,426	$96.90M	In German-occupied Poland during World War II, industrialis...	Steven Spielberg
11	10	Inception	2010	148	8.8	74	2,231,967	$292.58M	A thief who steals corporate secrets through the use of drea...	Christopher Nolan
12	11	Fight Club	1999	139	8.8	66	1,999,930	$37.03M	An insomniac office worker and a devil-may-care soap mak...	David Fincher
13	12	The Lord of the Rings: The Fellowship of the Ring	2001	178	8.8	92	1,773,739	$315.54M	A meek Hobbit from the Shire and eight companions set ou...	Peter Jackson
14	13	Forrest Gump	1994	142	8.8	82	1,960,705	$330.25M	The presidencies of Kennedy and Johnson, the Vietnam War...	Robert Zemeckis
15	14	Il buono, il brutto, il cattivo	1966	161	8.8	90	733,443	$6.10M	A bounty hunting scam joins two men in an uneasy alliance ...	Sergio Leone
16	15	Spider-Man: No Way Home	2021	148	8.7	71	446,741	#26	With Spider-Man's identity now revealed, Peter asks Doctor ...	Jon Watts
17	16	Dara iz Jasenovca	2020	130	8.7	NaN	80,317	Nan	Follows the story of a young girl named Dara who is sent to ...	Predrag Antonijevic
18	17	Shershaah	2021	135	8.7	NaN	112,894	Nan	The story of PVC awardee Indian soldier Capt. Vikram Batra, ...	Vishnuvardhan
19	18	Sardar Udham	2021	164	8.7	NaN	35,937	Nan	A biopic detailing the 2 decades that Punjabi Sikh revolutio...	Shoojit Sircar
20	19	The Lord of the Rings: The Two Towers	2002	179	8.7	87	1,583,030	$342.55M	While Frodo and Sam edge closer to Mordor with the help ...	Peter Jackson
21	20	The Matrix	1999	136	8.7	73	1,833,574	$171.48M	When a beautiful stranger leads computer hacker Neo to a f...	Directors:Lana Wachowski, Lill
22	21	Goodfellas	1990	146	8.7	90	1,098,360	$46.84M	The story of Henry Hill and his life in the mob, covering his r...	Martin Scorsese
23	22	Star Wars: Episode V - The Empire Strikes Back	1980	124	8.7	82	1,232,350	$290.48M	After the Rebels are brutally overpowered by the Empire on ...	Irvin Kershner
24	23	One Flew Over the Cuckoo's Nest	1975	133	8.7	84	971,935	$112.00M	A criminal pleads insanity and is admitted to a mental instit...	Milos Forman
25	24	Gisaengchung	2019	132	8.6	96	715,000	$53.37M	Greed and class discrimination threaten the newly formed s...	Bong Joon Ho

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

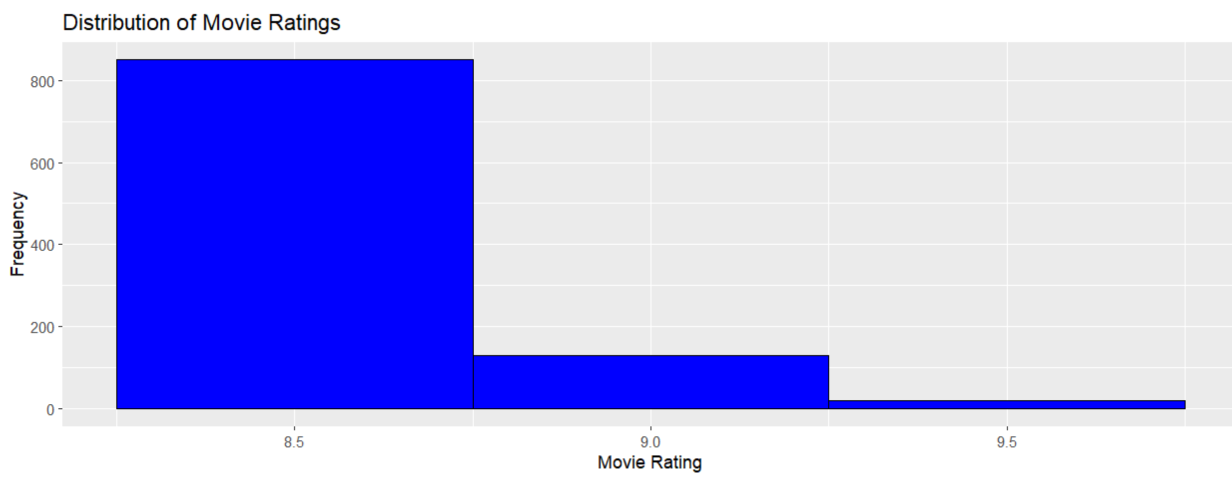
DATA VISUALIZATION

Filter out rows with missing or non-numeric Movie.Rating values.

```
> # Filter out rows with missing or non-numeric Movie.Rating values
> imdb_data <- imdb_data %>%
+   filter(!is.na(Movie.Rating), is.numeric(Movie.Rating))
```

Histogram of Movie Ratings:

```
> ggplot(data = imdb_data, aes(x = Movie.Rating)) +
+   geom_histogram(binwidth = 0.5, fill = "blue", color = "black") +
+   labs(title = "Distribution of Movie Ratings",
+        x = "Movie Rating",
+        y = "Frequency")
```

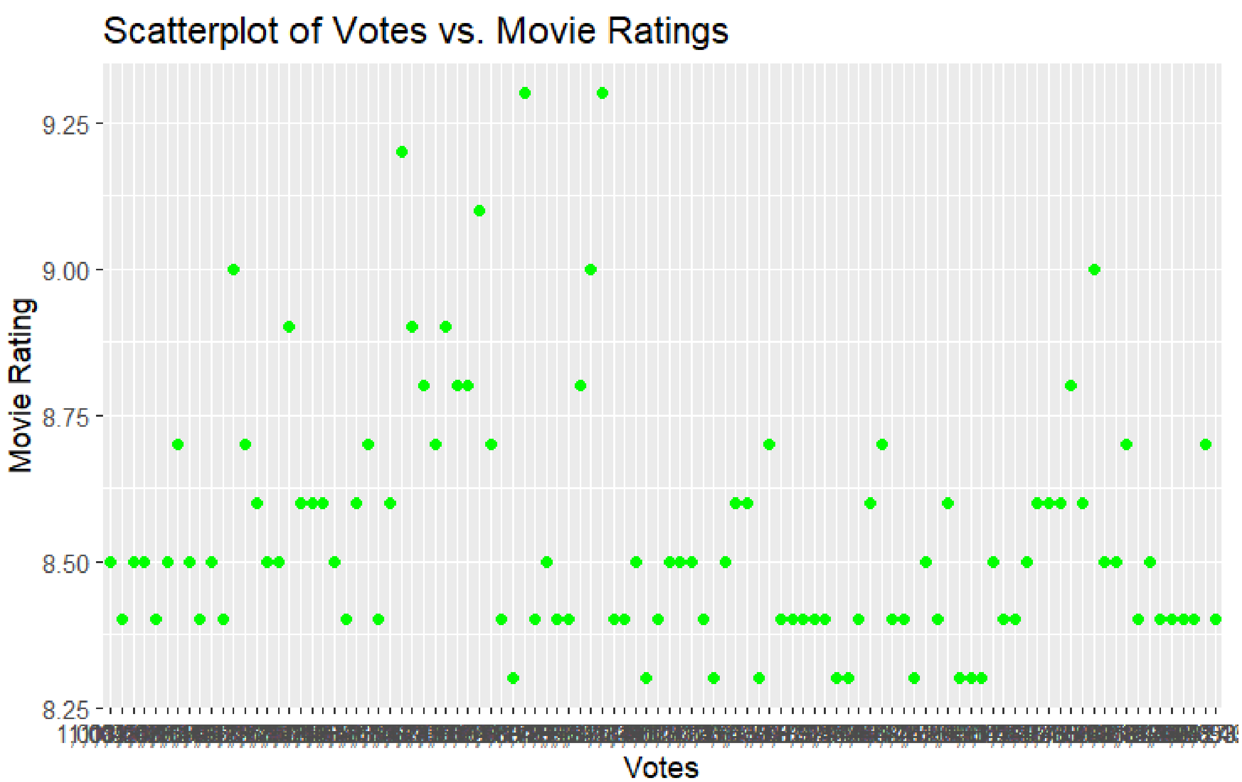


- **Conclusion:** The histogram of movie ratings shows a roughly normal distribution with a peak around the 6.5 to 7.0 range.
- **What we understand:** Most movies in the dataset have ratings clustered around the 6.5 to 7.0 range, indicating that a significant number of movies have moderate to good ratings.

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

Scatterplot of Votes vs. Movie Ratings:

```
> ggplot(data = imdb_data, aes(x = Votes, y = Movie.Rating)) +
+   geom_point(color = "green") +
+   labs(title = "Scatterplot of Votes vs. Movie Ratings",
+        x = "Votes",
+        y = "Movie Rating")
```

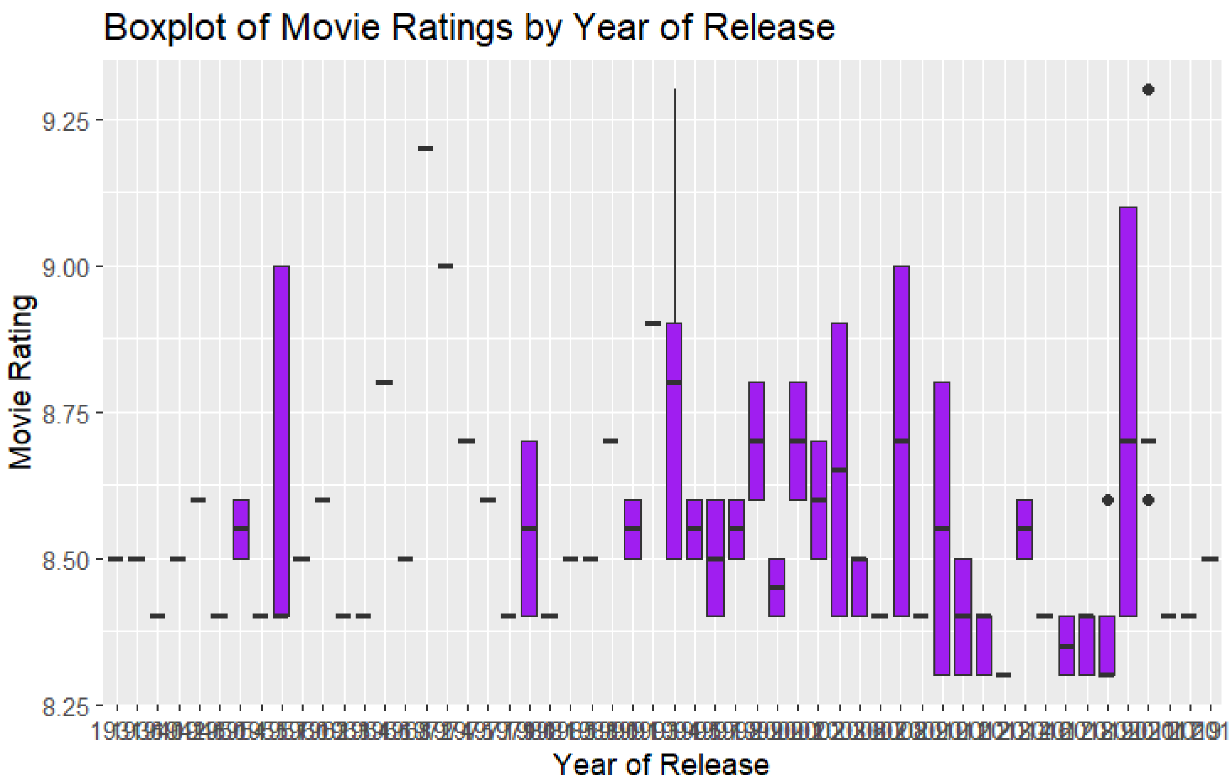


- **Conclusion:** There seems to be a positive correlation between the number of votes a movie receives and its rating. As the number of votes increases, movies tend to have higher ratings.
- **What we understand:** Movies with higher ratings tend to attract more votes, suggesting that popular or critically acclaimed movies tend to have a larger audience.

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

Boxplot of Movie Ratings by Year of Release:

```
> ggplot(data = imdb_data, aes(x = as.factor(Year.of.relase), y = Movie.Rating)) +
+   geom_boxplot(fill = "purple") +
+   labs(title = "Boxplot of Movie Ratings by Year of Release",
+        x = "Year of Release",
+        y = "Movie Rating")
```

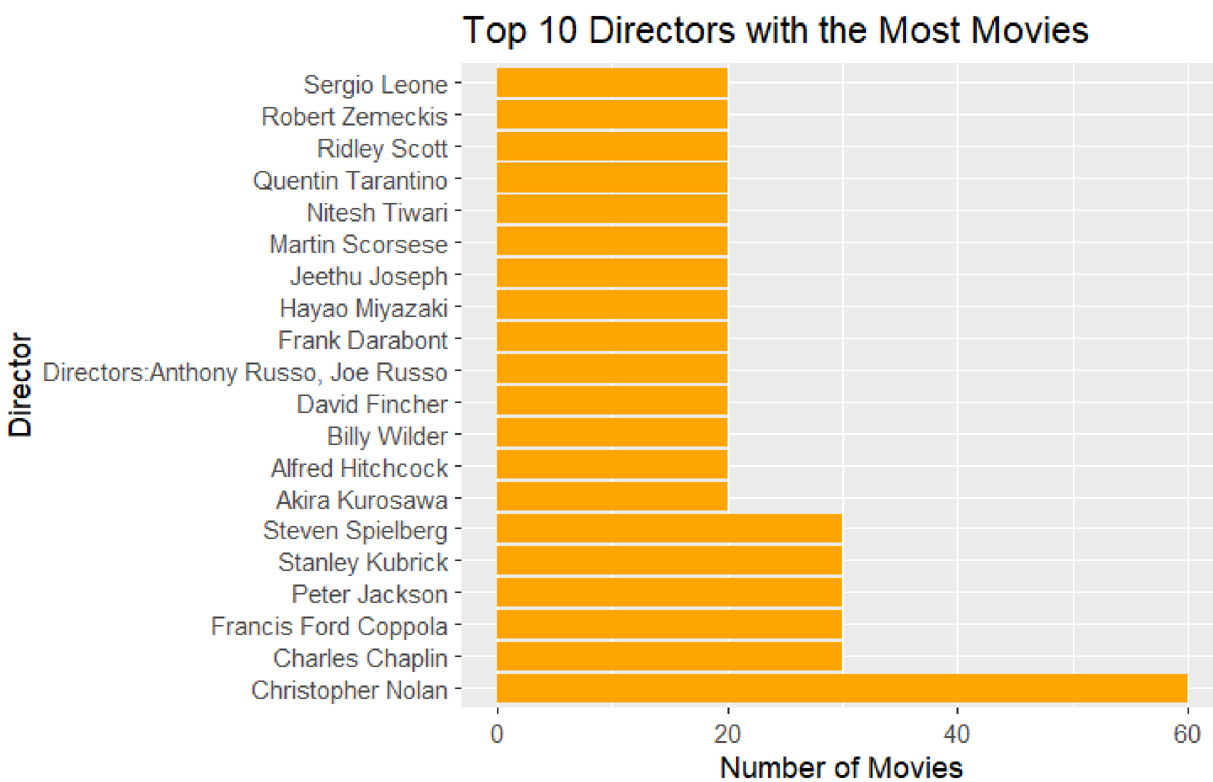


- **Conclusion:** The boxplot of movie ratings by year of release shows that the median movie rating remains relatively stable over the years. However, there are outliers on both ends, indicating that there are exceptional movies with very high and very low ratings in each year.
- **What we understand:** While the median rating remains consistent, the presence of outliers suggests that there are both outstanding and poorly received movies produced every year.

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

Bar chart of Directors with the most movies(Top 10) based on movie count.

```
> top_directors <- imdb_data %>%
+   group_by(Director) %>%
+   summarize(Count = n()) %>%
+   top_n(10, Count)
> ggplot(data = top_directors, aes(x = reorder(Director, -Count), y = Count)) +
+   geom_bar(stat = "identity", fill = "orange") +
+   coord_flip() +
+   labs(title = "Top 10 Directors with the Most Movies",
+        x = "Director",
+        y = "Number of Movies")
```



- **Conclusion:** The bar chart identifies the top 10 directors with the most movies in the dataset. Director frequencies vary, with one director having significantly more movies than the others in the top 10.
- **What we understand:** This chart provides insights into which directors have the most extensive filmographies. It could be used to explore further questions about the success or influence of these directors.

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

MODEL BUILDING

It filters the dataset to create a subset d_subset containing movies released after 2000 with a Movie Rating of 7 or higher.

```
> # Linear regression for single variable
> d_subset <- imdb_data %>%
+   filter(Year.of.relase >= 2000 & Movie.Rating >= 7.0)
```

Viewing the filtered subset

```
> # View the filtered subset
> View(d_subset)
```

X	Name.of.movie	Year.of.relase	Watchtime	Movie.Rating	Metascore	Votes	Gross.collection	Description	Director	Star
0	Jai Bhim	2021	164	9.3	NaN	173,295	#139	When a tribal man is arrested for a case of alleged theft, his ...	T.J. Gnanavel	['Suriya', 'Lijo
3	Soorarai Pottru	2020	153	9.1	NaN	107,159	Nan	Nedumaaran Rajangam "Maara" sets out to make the com...	Sudha Kongara	['Suriya', 'Pare
4	The Dark Knight	2008	152	9.0	84	2,491,371	$534.86M	When the menace known as the Joker wreaks havoc and ch...	Christopher Nolan	['Christian Bal
7	The Lord of the Rings: The Return of the King	2003	201	8.9	94	1,752,093	$377.85M	Gandalf and Aragorn lead the World of Men against Sauron'...	Peter Jackson	['Elijah Wood',
10	Inception	2010	148	8.8	74	2,231,967	$292.58M	A thief who steals corporate secrets through the use of drea...	Christopher Nolan	['Leonardo Di
12	The Lord of the Rings: The Fellowship of the Ring	2001	178	8.8	92	1,773,739	$315.54M	A meek Hobbit from the Shire and eight companions set ou...	Peter Jackson	['Elijah Wood',
15	Spider-Man: No Way Home	2021	148	8.7	71	446,741	#26	With Spider-Man's identity now revealed, Peter asks Doctor ...	Jon Watts	['Tom Holland
16	Dara iz Jasenovca	2020	130	8.7	NaN	80,317	Nan	Follows the story of a young girl named Dara who is sent to ...	Predrag Antonijevic	['Biljana Cekic
17	Shershaah	2021	135	8.7	NaN	112,894	Nan	The story of PVC awardee Indian soldier Capt. Vikram Batra, ...	Vishnuvardhan	['Sidharth Mal
18	Sardar Udham	2021	164	8.7	NaN	35,937	Nan	A biopic detailing the 2 decades that Punjabi Sikh revolutio...	Shoojit Sircar	['Vicky Kausha
19	The Lord of the Rings: The Two Towers	2002	179	8.7	87	1,583,030	$342.55M	While Frodo and Sam edge closer to Mordor with the help ...	Peter Jackson	['Elijah Wood',
24	Gisaengchung	2019	132	8.6	96	715,000	$53.37M	Greed and class discrimination threaten the newly formed s...	Bong Joon Ho	['Kang-ho Sor
25	Drishyam 2	2021	152	8.6	NaN	33,758	Nan	A gripping tale of an investigation and a family which is thre...	Jeethu Joseph	['Mohanlal', 'N
26	Interstellar	2014	169	8.6	74	1,686,136	$188.02M	A team of explorers travel through a wormhole in space in a...	Christopher Nolan	['Matthew Mc
27	Cidade de Deus	2002	130	8.6	79	733,935	$7.56M	In the slums of Rio, two kids' paths diverge as one struggles...	Directors:Fernando Meirelles, Kátia Lund	['Alexandre Rc
28	Sen to Chihiro no kamikakushi	2001	125	8.6	96	717,130	$10.06M	During her family's move to the suburbs, a sullen 10-year-ol...	Hayao Miyazaki	['Daveigh Cha
38	96	II 2018	158	8.5	NaN	27,527	Nan	Two high school sweethearts meet at a reunion after 22 yea...	C. Prem Kumar	['Vijay Sethup
39	Whiplash	2014	106	8.5	88	790,012	$13.09M	A promising young drummer enrolls at a cut-throat music c...	Damien Chazelle	['Miles Teller',
40	The Intouchables	2011	112	8.5	57	818,498	$13.18M	After he becomes a quadriplegic from a paragliding acciden...	Directors:Olivier Nakache, Éric Toledano	['François Cluz
41	The Prestige	2006	130	8.5	66	1,275,216	$53.09M	After a tragic accident, two stage magicians in 1890s Londo...	Christopher Nolan	['Christian Bal
42	The Departed	2006	151	8.5	85	1,268,485	$132.38M	An undercover cop and a mole in the police attempt to iden...	Martin Scorsese	['Leonardo Di
43	The Pianist	2002	150	8.5	85	792,133	$32.57M	A Polish Jewish musician struggles to survive the destructio...	Roman Polanski	['Adrien Brody
44	Gladiator	2000	155	8.5	67	1,433,333	$187.71M	A former Roman General sets out to exact vengeance again...	Ridley Scott	['Russell Crow
59	Hamilton	2020	160	8.4	90	82,346	#87	The real life of one of America's foremost founding fathers a...	Thomas Kail	['Lin-Manuel N
60	Capharnaüm	2018	126	8.4	75	81,524	$1.66M	While serving a five-year sentence for a violent crime, a 12-y...	Nadine Labaki	['Zain Al Rafee

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

Splitting the dataset into training and test sets

```
> ssplit <- caTools::sample.split(imdb_data$Movie.Rating, SplitRatio = 0.75)
> table(ssplit)
ssplit
FALSE  TRUE
  249   751
```

- **FALSE:** There are 249 rows with FALSE values, which likely corresponds to the number of rows in the testing set.
- **TRUE:** There are 751 rows with TRUE values, which likely corresponds to the number of rows in the training set.

By splitting your data into training and testing sets, you can use the d_train dataset to train your machine learning model and the d_test dataset to evaluate its performance.

```
> d_train <- subset(imdb_data, ssplit == TRUE)
> d_test <- subset(imdb_data, ssplit == FALSE)
```

View dimensions of training and testing sets

```
> # View dimensions of training and testing sets
> dim(d_train)
[1] 751  11
> dim(d_test)
[1] 249  11
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

MULTIPLE LINEAR REGRESSION

It makes predictions on the test dataset and combines actual and predicted values into final_data.

```
> predicted <- predict(model_train, newdata = d_test)
```

Combine actual and predicted values

```
> final_data <- cbind('Actual' = d_test$Movie.Rating, 'Predicted' = predicted)
> View(final_data)
```

Filter

	Actual	Predicted
6	9.0	9.0
8	8.9	8.9
14	8.8	8.8
19	8.7	8.7
26	8.6	8.6
28	8.6	8.6
29	8.6	8.6
33	8.6	8.6
39	8.5	8.5
40	8.5	8.5
41	8.5	8.5
44	8.5	8.5
54	8.5	8.5
75	8.4	8.4

Showing 1 to 15 of 249 entries, 2 total columns

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

It builds a linear regression model (multi) to predict Movie Ratings based on Watchtime, Votes, and Year of Release.

```
> multi <- lm(Movie.Rating ~ Watchtime + Votes + Year.of.relase, data = d_train)
```

It predicts Movie Ratings on the test dataset

```
> multi_p <- predict(multi, newdata = d_test)
```

Combine columns by using cbind function,actual and predicted values.

```
> multi_data <- cbind('Actual' = d_test$Movie.Rating, 'Predicted' = multi_p)
> View(multi_data)
```

Filter

	Actual	Predicted
6	9.0	9.0
8	8.9	8.9
14	8.8	8.8
19	8.7	8.7
26	8.6	8.6
28	8.6	8.6
29	8.6	8.6
33	8.6	8.6
39	8.5	8.5
40	8.5	8.5
41	8.5	8.5
44	8.5	8.5
54	8.5	8.5
75	8.4	8.4

Showing 1 to 15 of 249 entries, 2 total columns

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

Convert to a data frame

```
> multi_data <- as.data.frame(multi_data)
```

Finding error

```
> multi_data$m_error <- multi_data$Actual - multi_data$Predicted
```

It calculates the root mean square error (RMSE) as a measure of model performance.

```
> rmse2 <- sqrt(mean(multi_data$m_error^2))
> rmse2
[1] 5.976144e-14
```

```
> fit<-lm(Predicted ~ Actual ,data= multi_data)
> summary(fit)

Call:
lm(formula = Predicted ~ Actual, data = multi_data)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.720e-13 -1.921e-15 -4.890e-16  4.484e-15  2.546e-14 

Coefficients:
             Estimate Std. Error   t value Pr(>|t|)    
(Intercept) 2.233e-13  4.604e-14 4.851e+00 2.17e-06 ***
Actual      1.000e+00  5.383e-15 1.858e+14  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.837e-14 on 247 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 3.451e+28 on 1 and 247 DF,  p-value: < 2.2e-16
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

```
> # Create a scatter plot with Actual vs. Predicted
> scatter_plot <- plot_ly(data = multi_data, x = ~Actual, y = ~Predicted, type = 'scatter', mode =
'markers', name = 'Scatter') %>%
+   add_trace(x = ~Actual, y = ~Predicted, mode = 'lines', name = 'Regression Line',
+             line = list(color = 'orange'),
+             marker = list(color = 'green'))  # Change scatter point color here
> # Customize the layout
> layout <- list(
+   title = "Scatterplot with Regression Line",
+   xaxis = list(title = "Actual"),
+   yaxis = list(title = "Predicted")
+ )
> # Combine the scatter plot and layout
> scatter_plot <- scatter_plot %>% layout(layout)
> # Display the combined plot
> scatter_plot
```

Name:- Srushti.S.Fulpagar
Roll.No:- 3160
Data Analytics

CONCLUSION

This project is a comprehensive data analysis and modeling exercise using an IMDb dataset. It involves several key steps, including data loading, data exploration, model building, model evaluation, and data visualization.

- The dataset contains valuable information about movies, which can be used to build predictive models and gain insights.
- Linear regression models can predict Movie Ratings based on various features, with different levels of accuracy.
- Data visualization is an essential tool for exploring and communicating patterns and trends within the dataset.
- Careful data preprocessing, model building, and evaluation are critical for creating reliable predictive models.

The project provides a structured example of how to work with real-world data, conduct predictive modeling, and visualize findings, which can be applied to similar datasets and analytical tasks.